Modeling Structures and Transport Phenomena of Transmembrane Channels and Transporters

Thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
in
Computational Natural Sciences (Biophysics)

by

Siladitya Padhi
201166647
siladitya.padhi@research.iiit.ac.in

Center for Computational Natural Sciences and Bioinformatics
International Institute of Information Technology, Hyderabad
(Deemed to be University)
Hyderabad 500032, India

October 2016

International Institute of Information Technology

Hyderabad, India

DECLARATION OF AUTHORSHIP

I, Siladitya Padhi, declare that the thesis titled "Modeling Structures and Transport Phenomena of Transmembrane Channels and Transporters", and the work presented in it, are my own. I confirm that this work was done wholly or mainly while in a candidature for a research degree at this university. Whenever I have consulted the published work of others, it will always be clearly mentioned.

13.10.16

Date

Siladitya Padhi

Signature of the candidate

iii

International Institute of Information Technology

Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Modeling Structures and Transport Phenomena of Transmembrane Channels and Transporters" by Siladitya Padhi, has been carried out under my supervision, and is not submitted elsewhere for the award of any degree or diploma.

13.10.16

Date

U. Deva Priyakumar

Advisor: Dr. U. Deva Priyakumar

Acknowledgments

I must start by thanking my supervisor, Dr. U. Deva Priyakumar, for giving me an opportunity to work with him, and for making all this possible. Having had a similar approach towards work, I was pretty comfortable working with him. He has always been active and cooperative, and I always had the freedom to walk into his office any time and start discussing business. He has given me a lot of opportunity, the most noteworthy one being my visit to Germany for a 10-day workshop at Ruprecht-Karls-Universität Heidelberg, followed by a two-month stay at Westfälische Wilhelms-Universität Münster. I also had a lot to learn working as a teaching assistant with him.

There has been a lot to gain from faculty members at the Center for Computational Natural Sciences and Bioinformatics (CCNSB) at IIIT. Dr. Prabhakar Bhimalapuram has constantly given suggestions, right from my initial days, when he explained the basics of enhanced sampling methods to me, to my pre-PhD defense, during which he gave valuable and critical inputs. I have benefited immensely from the well-structured Statistical Mechanics course offered by Dr. Marimuthu Krishnan. The course started with a thorough brush-up of basic stat mech and ended with students writing codes – from scratch – for performing molecular dynamics and Monte Carlo simulations. Working as a teaching assistant with Prof. Harjinder Singh was a very nice experience. I am grateful to Prof. Singh also because he introduced us to our experimental collaborators at ICGEB Delhi as well as RCB Delhi.

Working with Dr. Shahid Jameel (formerly at ICGEB Delhi and now CEO, Wellcome Trust/DBT India Alliance) taught me to look at things from an experimentalist's perspective, and to describe things in a way that would interest an experimentalist. The work being done with Dr. Avinash Bajaj (RCB Delhi) for the last few months was also a nice learning experience. I must seriously thank Dr. Mark Waller, my boss at Universität Münster. His hospitality was admirable, and his patience with me was commendable! I really appreciate the method development work he is doing.

I am grateful to the Department of Biotechnology, Government of India, and the Department of Atomic Energy, Government of India, for funding our work. I am also grateful to Universität Heidelberg (Prof. Peter Comba, in particular) for funding my visit to Germany.

I must now thank the scientific/non-scientific staff and my colleagues at CCNSB, IIIT. Dr. Semparithi Aravindan is awesome: he has a solution for just about anything to do with the clusters. Umesh has been the unsung hero of CCNSB. Tanashree, Sandhya, Siddharth, Indrajit, Navneet, and Chinmayee have been wonderful friends, Aditya has been a funny entertainer, and Suresh and Ramakrishna played a very important role by assisting me with the scripts in my initial days. Friends from other research groups at CCNSB without whom the list would be incomplete: Antarip, Bipin(ji), Broto, Kartheek, Mohan, and the others.

Finally, I owe all my gratitude to my mother and my father, who always gave me the freedom to do what I want. The constant support of my parents and my brother (and his family) has been instrumental in making possible whatever I have done. It is their love, patience, and strength that have been my pillars of support.

Abstract

Membrane proteins account for more than half of present day drug targets, but an understanding of structure-function relationships in these proteins continues to be a challenge, owing to the difficulties associated with experimental investigations of these proteins. The work in this thesis makes use of all-atom molecular dynamics simulations to address important problems pertaining to membrane proteins. A major problem in drug design targeting membrane proteins is limited availability of experimental structures. Toward this end, an approach has been proposed here for modeling the structures of α-helical membrane proteins. While the approach is able to provide structural models for a number of channel proteins, possible improvements, that would make the approach more suitable for other membrane proteins, are proposed. The other aspect of membrane proteins that has been difficult to investigate experimentally is an atomistic level understanding of the conduction and selectivity mechanism in channels and transporters. This thesis provides insights into the mechanism of transport in a number of channels and transporters, including two viral ion channels, a mammalian aquaporin, and a bacterial urea transporter.

In the first part of the thesis, it is shown that the protein Vpu from human immunodeficiency virus type 1 (HIV-1) exists in a pentameric state, and a structural model for the oligomeric form of the protein is proposed. Free energy calculations performed on the structural model are able to provide a thermodynamic basis for the weak ion channel activity of the protein. In the next part of the thesis, a structure modeling approach is described that attempts to predict the structures of α-helical membrane proteins by minimizing unfavorable contacts. The approach works well for α-helical channels, and can be extended to all α-helical membrane proteins. This is followed by an investigation of ion conduction and selectivity in the p7 channel from hepatitis C virus (HCV). It is shown that a hydrophobic stretch in the channel preferentially allows the singly charged K^+ rather than the doubly charged Ca^{2+} to pass through, and that interionic repulsion is crucial for ion conduction through the pore. Water transport through the human aquaporin AQP2 is then examined, revealing an almost barrierless permeation profile for water transport. The implications of the restriction of water orientation in the middle region of the pore are discussed.

Finally, the mechanism of urea transport through the bacterial urea transporter dvUT is studied, and it is shown that permeating urea molecules are able to overcome a hydrophobic barrier arising from the existence of pore-facing phenylalanine residues by forming stacking interactions with these residues. While the work on structure modeling reveals some “rules” that membrane proteins follow at the time of oligomerization, the studies on transport across membrane proteins suggest that transport proteins have their unique mechanisms for determining selectivity, depending on what chemical species they conduct.

Contents

List of figures

List of tables

1 Introduction

Membrane proteins are proteins that are found in the organellar and plasma membranes of living cells. These proteins carry out some of the most critical functions in the cell, such as enabling selective transport of ions and metabolites across cells and initiating signaling pathways inside the cell [1]. Owing to this physiological significance, these proteins form some of the most important targets for intervention in diseased states [2]. However, difficulties associated with investigating structures of these proteins experimentally mean that there is limited understanding of the structural dynamics and structure-function relationships of these proteins [1,3]. These difficulties make the use of computational approaches inevitable, and, indeed, computational modeling has been used to investigate several different kinds of phenomena involving membrane proteins [4,5]. The work presented in this thesis aims to further our understanding of membrane proteins by employing molecular dynamics simulations. There are two broad problems that have been addressed here: (1) modeling the structures of helical membrane proteins, which is done by proposing a novel approach for assembling oligomeric forms of these proteins, and (2) investigating the mechanism of transport through channels and transporters, which is done by employing extensive molecular dynamics-based free energy calculations. While the work on structure modeling proposes a workflow to assemble individual α-helical structures into oligomeric forms, the work on transport proteins reveals interactions that drive conduction through – and render selectivity to – these proteins. This chapter starts with structural and functional classifications of membrane proteins, then describes different membrane models used in present day simulations, and finally provides a comparison of the proteins investigated in this thesis.

1.1 Membrane proteins could be helices or barrels

Membrane proteins in general are made up of α-helices or β-strands at the secondary structure level, and, depending on this secondary structure content, they occur either as helical structures or β-barrels (Figure 1) [1].

1.1.1 Helical structures

The majority of membrane proteins fall into this category of structures, and these have, consequently, gained more attention [1,6]. Helical membrane proteins can traverse the membrane either single or multiple times, and are accordingly named as single pass or multipass membrane proteins [1]. Single pass membrane proteins typically have a single α-helix extending across the membrane, with the N-terminal end on one side of the membrane and the C-terminal end on the other. Multipass membrane proteins have multiple α-helices connected by loop regions such that the helical regions lie in the membrane and the connecting loops lie in the aqueous phase on the two sides of the membrane.

The folding of helical membrane proteins into their native structure has generated a lot of interest because of its complexity [7,8]. The folding process is divided into two stages: insertion of the protein into the membrane guided by a translocon complex, and assembly and packing of the helical segments to result in the native tertiary/quarternary structure of the protein. Examples of helical membrane proteins include most channels, transporters, receptor proteins, and enzymes. Figure 1.1 (A) shows a G protein-coupled receptor (GPCR), which has seven helices traversing the membrane.

1.1.2 β-barrels

β-barrels consist of a number of intertwined β-strands, and are found only in the outer membranes of a few bacteria, and in mitochondrial and chloroplast membranes of a few eukaryotic cells [1,6]. Among proteins with β-barrel structure are porins (Figure 1.1 (B)), which are channels in the outer membranes of bacteria that allow a wide variety of solutes to pass through. The work in this thesis focuses on the more abundant α-helical transmembrane proteins, rather than the β-barrel structures.

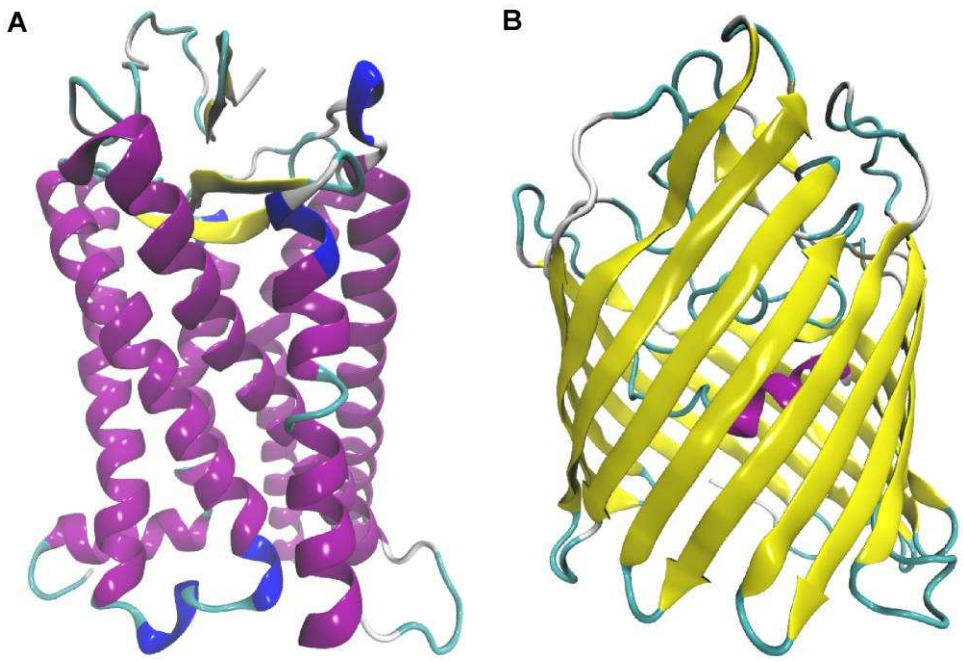

Figure 1.1 (A) An α-helical protein. The protein shown is metarhodopsin II, a G protein-coupled receptor (PDB ID: 3PQR). (B) The β-barrel-shaped outer membrane protein, OmpF, from Salmonella typhi (PDB ID: 3NSG).

1.2 Functional classification of membrane proteins

As far as function is concerned, the broad classes that membrane proteins can be classified into are transport proteins, receptors, and enzymes, besides a number of other smaller classes [1]. These classes, however, are not mutually exclusive, and there are a number of proteins that could fall into more than one of these classes.

1.2.1 Transport proteins

Transport proteins include channels and transporters, and they possess a pore through which they conduct solutes into and out of the cell [9,10]. What makes these proteins remarkable is their ability to allow certain molecules and ions to selectively pass through at a very high rate while at the same time prohibiting the passage of other molecules. By doing so, channels/transporters ensure that an optimum concentration of ions and metabolites is retained both inside and outside the cell. Between channels and transporters, there are differences in the manner in which the two classes exhibit their activity. Transporters usually have lower rates of conductance than channels, and the transport of solutes across transporters is typically coupled to conformational changes in

the transporter, unlike in the case of channels, which exist in an "open" state prior to the transport of solute, and remain more or less rigid during the process of transport.

Channels could exist in open or closed states, apart from states that are intermediate between the two [9,11]. The open state is characterized by a wide and continuous pore that is permeable to the solute. Closed states, on the other hand, have a discontinuous or very narrow pore, thereby prohibiting the passage of the solute. The transition of a channel from a closed state to an open state (or vice versa) is known as gating. Based on the type of gating they undergo, channels could be voltage-gated or ligand-gated. As might be inferred from their names, voltage-gated channels open in response to a change in transmembrane potential, and ligand-gated channels open in response to the binding of a ligand. All channels selectively allow certain molecules to pass through while prohibiting the permeation of others. Such selective behavior is made possible by certain residues facing the pore, and these residues constitute what is known as the selectivity filter of the channel.

Transporters do not have a continuous pore extending across the membrane, but they rather exist in states that are open to only one side of the membrane [10]. Depending on the side of the membrane on which the open face of the pore is found, the conformation of a transporter could be described as inward-facing or outward-facing. The binding of the substrate induces the transporter to go from one conformation to the other. The nature of transport across transporters could be either uniport, which involves movement of one molecule across the pore, or cotransport, which involves the movement of two molecules through the pore at a given instant. Cotransport, in turn, could occur via symport, in which the movement of two molecules moving in the same direction is coupled, or via antiport, in which the movement of two molecules moving in opposite directions is coupled.

1.2.2 Receptors

Receptors trigger certain signaling or metabolic pathways inside the cell upon the binding of a ligand [12]. One of the most receptors are the GPCRs, which have an extracellular face that binds to ligands, and a cytoplasmic face that activates G proteins. All GPCRs are seven-pass transmembrane proteins, meaning that they have seven helical segments that pass through the

membrane, with adjacent helical segments being connected by loop regions. Although all GPCRs share this structural similarity, these proteins show great diversity as far as their specificity toward ligands is concerned. Furthermore, different GPCRs are involved in the activation of different proteins downstream in the signaling pathway, which is why these proteins show great diversity in the function they perform [12]. Certain receptor proteins are linked to ion channels, the most noteworthy examples being neurotransmitter receptors like the nicotinic acetylcholine receptor (nAChR) and γ-aminobutyric acid (GABA) receptor.

1.2.3 Enzymes

Enzymes occurring in the membrane play an important role in a number of metabolic processes [1]. Membrane enzymes have diverse functions, and are involved in lipid metabolism, electron transport chains, and substrate phosphorylation, to name a few. Cytochromes are an important family of membrane enzymes that take part in electron transport chains leading to the formation of adenosine triphosphate (ATP). Certain membrane proteins serve as both receptors and enzymes, and the most noteworthy example is the family of receptor tyrosine kinases [1]. These proteins are cell-surface receptors having a hormone-binding domain on the extracellular side and a tyrosine kinase domain on the intracellular side. The tyrosine kinase domain phosphorylates tyrosine residues on the substrate, which leads to activation of signaling pathways. In general, membrane enzymes are highly sensitive to lipid composition, and in fact, certain enzymes require the presence of specific lipids for their activity [1].

1.2.4 Other membrane proteins

The three classes of membrane proteins described above belong to a category of proteins that span the entire cross-section of the membrane, and are therefore described as integral membrane proteins. There also exist proteins that are present on the surface of membranes, and these proteins are known as peripheral proteins [1]. Certain proteins, like toxins and antimicrobial peptides, can insert into the membrane and disrupt it, eventually leading to the formation of pores [1].

1.3 The lipid bilayer

1.3.1 Bilayer structure and explicit bilayer models

The fundamental unit of a cell membrane is a lipid, and the lipids that make up the membrane could be phospholipids, sphingolipids, glycoplipids, or sterols [13]. In order to illustrate the chemical nature of a typical lipid, the structure of a lipid called 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) is shown in Figure 1.2 (A,B). The molecule has a long hydrophobic tail made up of alkyl chains, and a hydrophilic head made up of a tertiary amine group, a phosphate group, and carbonyl groups. The lipid bilayers that constitute cell membranes have two layers of lipids arranged so that the hydrophobic moieties of the two layers face each other, while the hydrophilic heads face the aqueous phase on the two sides of the membrane. The two sides of the membrane correspond to the exterior and the interior of a cell (or an organelle, if it is an organellar membrane). Figure 1.2 (C) shows a simple lipid bilayer of POPC molecules with a protein embedded in the bilayer. The hydrophilic heads face the aqueous phase, while the hydrophobic tails face the interior of the membrane to form what is known as the hydrophobic core of the membrane. The bilayer shown is a homogeneous POPC bilayer, and it differs from the heterogeneous environment that is typically seen in membranes. However, such POPC bilayers have satisfactorily been used for modeling membranes [4,5], and it is a fair approximation to use POPC-only membranes, since POPC is the most abundant lipid in animal cell membranes [14,15]. All simulations in this thesis, unless mentioned otherwise, were performed in POPC bilayer models.

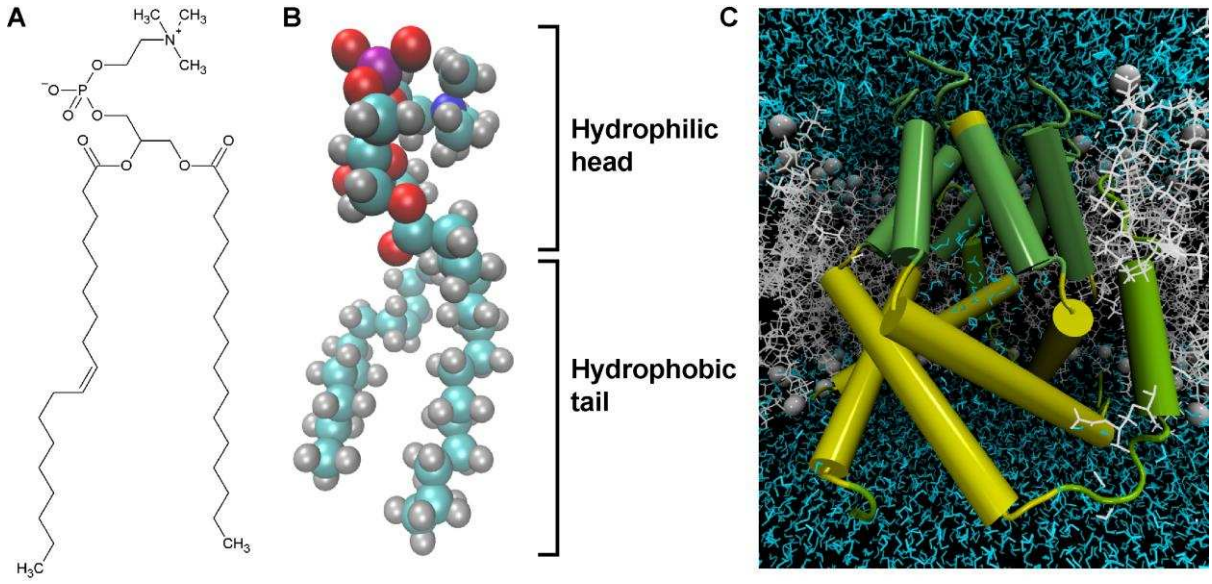

Figure 1.2 (A) Chemical structure of a POPC molecule. (B) Three-dimensional structure of POPC. Carbon atoms are shown in cyan, hydrogens in silver, oxygens in red, nitrogen in blue, and phosphorus in purple. (C) An ion channel in a lipid bilayer of POPC molecules. Water is seen on the two sides of the membrane, and in the pore of the channel.

1.3.2 Implicit membrane models

Implicit membrane models make it possible to take into account the electrostatic nature of a membrane while at the same time enhancing computing speed by reducing the number of particles in the system [16]. In this thesis, the generalized Born (GB) implicit membrane model has been used in the structure modeling studies for performing an initial conformational search of the proteins being modeled [17]. The GB model models the membrane as a low dielectric slab, and the aqueous phase as a high dielectric region, with a smoothing function ensuring a smooth transition from membrane phase to aqueous phase. The solvation free energy is calculated as

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{np} \ , \tag{1.1}$$

where ΔG_{elec} is the electrostatic solvation energy and ΔG_{np} is the nonpolar solvation energy. The electrostatic solvation energy is given by

$$\Delta G_{elec} = \frac{1}{2}\sum_{\alpha} q_{\alpha}\phi_{rf}(r_{\alpha}) \ , \tag{1.2}$$

where $\{q_{\alpha}\}$ denotes the charges on the solute and $\phi_{rf}(r_{\alpha})$ is the reaction field potential. The nonpolar solvation energy is

$$\Delta G_{np} = \gamma S \ , \tag{1.3}$$

where γ is the surface tension coefficient and S is the solvent accessible surface area of the solute.

1.4 Proteins studied in this thesis

The proteins studied in this work are all transport proteins, with the exception of ErbB2, which is a receptor protein. A brief introduction to these proteins is given below, and a comparison of all these proteins is made in Table 1.1.

1.4.1 Viroporins

The proteins Vpu from human immunodeficiency virus type 1 (HIV-1), p7 from hepatitis C virus (HCV), M2 from influenza A virus, and M2 from influenza B virus belong to a class of ion channels called viroporins [18,19]. Viroporins are ion channels encoded by viruses inside infected host cells, and they aid in the propagation of the virus by altering ionic concentrations inside host cells, which makes them potential targets for antiviral therapy [18,19]. Although viroporins have a transmembrane domain (TMD) as well as a cytoplasmic domain, the investigations performed here deal only with the TMD, since it is the TMD that is associated with ion channel activity. Viroporins conduct cations or protons, and they have a pore that is more hydrophobic than other ion channels. Furthermore, they usually lack the TTVGYGD motif that is highly conserved in the selectivity filter of eukaryotic potassium channels [20]. Thus, they are expected to possess selectivity and conduction mechanisms that are different from other cation channels.

1.4.2 Channels for polar solutes

Among transport proteins that conduct polar solutes, the human aquaporin AQP2 and the bacterial urea transporter dvUT (from Desulfovibrio vulgaris) have been studied here. Aquaporins are water-conducting channels that have implications in tumor angionesis, oedema, and kidney diseases, besides other minor diseases [21]. They exist as tetramers in cell membranes, with each monomeric unit possessing a pore filled with a single chain of water molecules. The pores have two highly conserved NPA motifs (“NPA” for asparagine, proline, and alanine) and an ar/R constriction, which is a constricted region with aromatic residues and arginines. Urea transporters (UTs), on the other hand, typically exist as trimers, and are seen in

kidney cells [22]. UTs exhibit channel-like rather than transporter-like behavior, since they do not undergo large scale conformational changes during the transport process, and conduct urea at rates comparable to any channel. dvUT is a bacterial homolog of kidney urea transporters, and is an ideal system for obtaining insights about UTs.

Table 1.1 Comparison of the proteins investigated in this thesis

Protein	**Organism**	**Functional class**	**Number of helices in monomer (excluding cytoplasmic domains, if any)**	**Oligomeric state**	**Species transported**
Vpu	HIV-1	Ion channel	1	Pentamer (?)	Cations
M2	Influenza A	Ion channel	1	Tetramer	Protons
BM2	Influenza B	Ion channel	1	Tetramer	Protons
p7	HCV	Ion channel	3	Hexamer	Cations
AQP2	Human	Water channel	8	Tetramer	Water
dvUT	Desulfovibrio vulgaris	Urea transporter	10	Trimer	Urea
ErbB2	Human	Receptor kinase	1	Dimer	-

References

1. Luckey M (2008) Membrane Structural Biology: With Biochemical and Biophysical Foundations Cambridge University Press, New York.
2. Rask-Andersen M, Almén MS, Schiöth HB (2011) Trends in the exploitation of novel drug targets. Nat Rev Drug Discov 10(8):579–590.
3. Deniaud A, Moiseeva E, Gordeliy V Pebay-Peyroula E (2010) In Membrane Protein Structure Determination: Methods and Protocols Lacapére JJ, Ed.; Methods in Molecular Biology Series, Springer Protocols, Humana Press.
4. Stansfeld PJ, Sansom MS (2011) Molecular simulation approaches to membrane proteins. Structure 19(11):1562–1572.
5. Biggin PC, Bond PJ (2015) Molecular Dynamics Simulations of Membrane Proteins. Methods in Molecular Biology. Molecular Modeling of Proteins 147–160.
6. Bigelow HR, Petrey DS, Liu J, Przybylski D, Rost B (2004) Predicting transmembrane beta-barrels in proteomes. Nucleic Acids Res 32(8):2566–2577.

7. Popot J-L, Engelman DM (2000) Helical membrane protein folding, stability, and evolution. Annu Rev Biochem 69(1):881–922.
8. Bowie JU (2005) Solving the membrane protein folding problem. Nature 438(7068):581–589.
9. McCoy JG, Nimigean CM (2012) Structural correlates of selectivity and inactivation in potassium channels. Biochim Biophys Acta, Biomembr 1818(2):272–285.
10. Diallinas G (2014) Understanding transporter specificity and the discrete appearance of channel-like gating domains in transporters. Front Pharmacol 5.
11. MacKinnon R (2003) Potassium channels. FEBS Lett 555(1):62-65.
12. Katritch V, Cherezov V, Stevens RC (2013) Structure-Function of the G Protein–Coupled Receptor Superfamily. Annu Rev Pharmacol Toxicol 53(1):531–556.
13. Meer GV, Voelker DR, Feigenson GW (2008) Membrane lipids: where they are and how they behave. Nat Rev Mol Cell Biol 9(2):112–124.
14. Li S, Lin H, Wang Z, Huang C (1994) Identification and characterization of kink motifs in 1-palmitoyl-2-oleoyl- phosphatidylcholines: a molecular mechanics study. Biophys J 66(6):2005–2018.
15. Rog T (2003) Effects of phospholipid unsaturation on the bilayer nonpolar region: a molecular simulation study. J Lipid Res 45(2):326–336.
16. Grossfield A (2008) Implicit modeling of membranes. Curents Topics in Membranes. Computational Modeling of Membrane Bilayers 60:131-157.
17. Im W, Feig M, Brooks CL (2003) An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins. Biophys J 85(5):2900–2918.
18. Nieva JL, Madan V, Carrasco L (2012) Viroporins: structure and biological functions. Nat Rev Microbiol 10(8):563-574.
19. Ouyang B, Chou JJ (2014) The minimalist architectures of viroporins and their therapeutic implications. Biochim Biophys Acta, Biomembr 1838(4):1058–1067.
20. Heginbotham L, Lu Z, Abramson T, MacKinnon R (1994) Mutations in the K channel signature sequence. Biophys J 66(4):1061–1067.
21. Verkman AS, Anderson MO, Papadopoulos MC (2014) Aquaporins: important but elusive drug targets. Nat Rev Drug Discov 13(4):259–277.

22. Esteva-Font C, Anderson MO, Verkman AS (2014) Urea transporter proteins as targets for small-molecule diuretics. Nat Rev Nephrol 11(2):113–123.

2 Methods

Computation of the energetic and dynamic properties of a molecule involves solving numerous mathematically complex equations. The principles involved in these computations are outlined below.

2.1. Force fields

Molecules may be modeled based on the principles of quantum mechanics or classical mechanics. Classical mechanics models – more popularly known as molecular mechanics (MM) models – do not explicitly consider the distribution of electrons, but use approximations to take electronic effects into account [1]. This simplicity makes MM models the preferred choice for modeling large macromolecular systems. These models treat atoms in a molecule as spheres connected by springs, with the spheres corresponding to the positions of the nuclei. The calculation of the energy of the system is made possible by a potential energy function, which is expressed as a function of nuclear coordinates. Potential energy functions, together with certain parameters (described below) that ensure that the modeling is as accurate as possible, constitute what are known as force fields.

Additive force fields are force fields that compute the potential energy as a sum of the contributions arising from various interactions, and most computations performed today employ this class of force fields. The CHARMM force field [2-6], which has been used for all studies in this work, has a force field of the form given below.

$$U(r) = \sum_{\text{bonds}} k_b(b - b_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi[1 + \cos(n\phi + \delta)] + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left[\frac{q_i q_j}{\varepsilon\, r_{ij}} + \varepsilon_{ij}\left(\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right)\right] \quad (2.1)$$

U(r) is the potential energy expressed as a function of the positions of the atoms in the molecule. The terms on the right hand side of the equation are described in the following sections.

2.1.1 Bond stretching

The first term on the right hand side accounts for changes in energy arising from stretching of bonds. A bond between two atoms is treated as a system of two rigid spheres connected by a spring (Figure 2.1 (A)). k_b denotes the spring constant of the spring connecting the two atoms, b_0 is the equilibrium bond length, and b is the bond length at any instant. Figure 2.1 (B) shows the variation of the energy due to bond stretching as a function of bond length. The curve has the form of a harmonic function.

2.1.2 Angle bending

Like bond stretching, angle bending (Figure 2 (C)) is also treated by a harmonic term. This is given by the second term in equation 2.1, in which θ_0 is the equilibrium bond angle, θ is the instantaneous bond angle, and k_θ is the force constant.

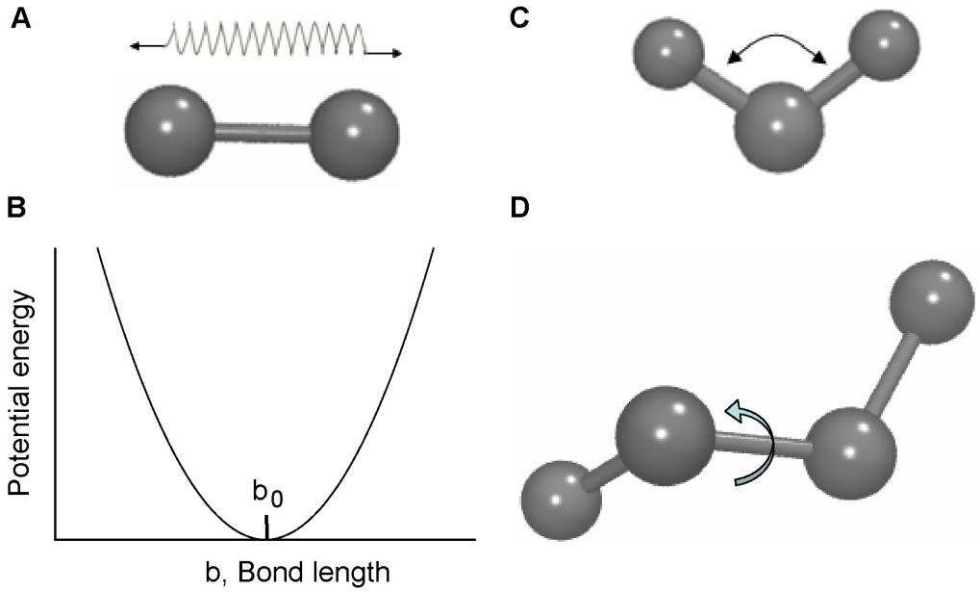

Figure 2.1 (A) Schematic representation of bond stretching. (B) The potential energy curve for harmonic functions. (C) Schematic representation of angle bending. (D) The dihedral angle about a bond quantifies the rotation about the bond.

2.1.3 Dihedrals

All molecules undergo rotation about single bonds; such rotation is quantified by a geometric measure called the dihedral angle. The dihedral angle term (also called torsion term) has the form of the third term in equation 2.1, where ϕ is the dihedral angle, k_ϕ accounts for the height of the potential energy curve, n determines the number of maxima/minima that occur as ϕ varies from -

π to π, and δ gives the phase shift for ϕ. These parameters can be modulated to model the potential energy curve for any dihedral angle. Figure 2.2 illustrates how variation of these parameters alters the nature of the curve.

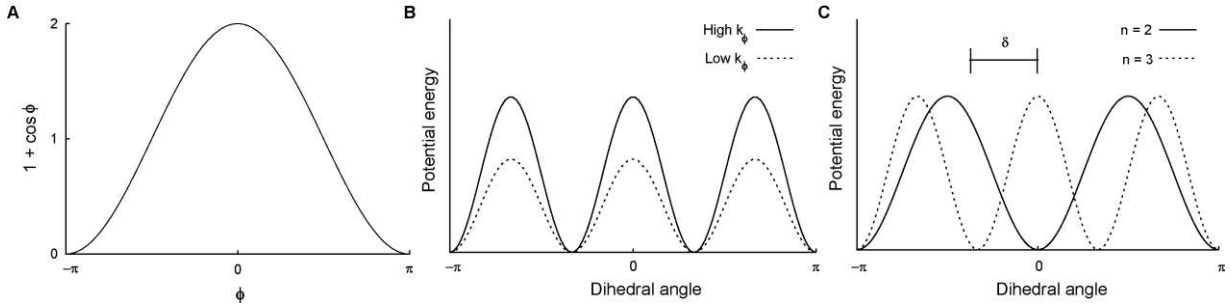

Figure 2.2 (A) The function 1 + cos ϕ as a function of ϕ. (B) The energy function for different values of k_ϕ. (C) Variation of the energy with n and δ.

2.1.4 Electrostatic interactions

The fourth term in equation 2.1 is that due to electrostatic interactions. For calculating the pairwise interaction between two charged particles i and j, a Coulombic potential is used. In equation 2.1, q_i and q_j denote the charge on particles i and j, respectively; ε is the permittivity in vacuum, and r_{ij} is the distance between the particles i and j. The pairwise interactions between all such pairs are calculated and summed. It must be noted that interactions between covalently connected atoms or atoms separated by two covalent bonds are not taken into account, since these interactions are accounted for by bond stretching and angle bending terms. The electrostatic interaction energy as a function of the distance between the charges is shown in Figure 2.3 (A).

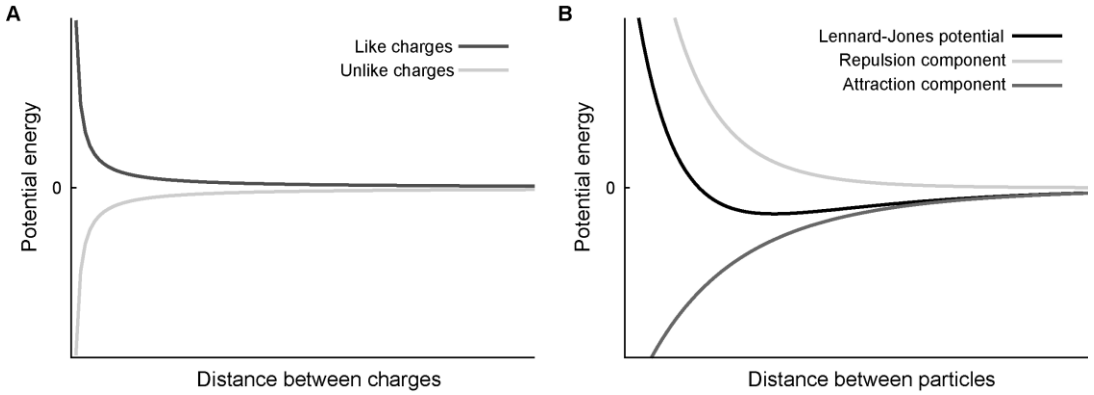

Figure 2.3 (A) Electrostatic interaction energy between two charged particles as a function of distance between the charges for opposite charges and like charges. (B) The attractive and repulsive components of the LJ potential shown together with the complete potential.

2.1.5 van der Waals interactions

Atoms that do not carry a substantial charge are also able to interact non-covalently, since they form instantaneous dipoles. At close enough distances, however, they repel each other because of internuclear and electron-electron repulsion. These kind of interactions, called van der Waals interactions, therefore include an attraction component and a repulsion component. The repulsion component comes into play at close separations, while the attraction component dominates at larger separations. van der Waals forces are commonly modeled by the Lennard-Jones (LJ) potential, which is given by the last term in equation 2.1. r_{ij} is the distance between atoms i and j. The repulsive force is proportional to $(1/ r_{ij})^{12}$, while the attractive force is proportional to $(1/ r_{ij})^{6}$. At low separations, $(1/ r_{ij})^{12} >> (1/ r_{ij})^{6}$, so the repulsive force dominates. At a certain distance called the collision diameter, $(1/ r_{ij})^{12} = (1/ r_{ij})^{6}$, and the energy due to the LJ potential is zero. At high separations, $(1/ r_{ij})^{12} << (1/ r_{ij})^{6}$, so the attractive forces dominate. The energy reaches a minimum at a distance $R_{min,ij}$. The depth of the energy at this point is denoted as ε. Figure 2.3 (B) shows the energy curve for the repulsive component, the attractive component, and the complete LJ potential.

2.1.6 Other terms

In addition to the terms discussed above, a few more terms are included in the CHARMM force field. Improper torsions are included to ensure that atoms that form part of a ring all lie in a plane. For example, four atoms that lie on opposite edges of a phenyl ring are commonly treated with a potential that ensures that the atoms lie in one plane. Interactions between the 1,3 atoms that form a bond angle are modeled using a Urey-Bradley potential, which treats 1,3 interactions as a harmonic function of the distance between the two atoms. Most force fields, including CHARMM, do not include an explicit term for hydrogen bonds, and instead use the Coulombic and LJ potentials to account for it.

2.2. Integration algorithms

Potential energy functions calculate the potential energy of a system with N atoms as a function of the 3N atomic coordinates [7]. The force acting on a given atom along a component (x-, y-, or

z-axis) can then be calculated by taking the first derivative of the potential energy with respect to the position of the atom along that component:

$$F_i^{\alpha} = -\frac{dU}{dr_i^{\alpha}} \tag{2.2}$$

Here F_i^{α} is the force acting along the i^{th} component for atom α, U is the potential energy, and r_i^{α} is the position of atom α along the i^{th} component (a given atom α can have three components x, y, and z). The acceleration for each particle is then given by

$$a_i^{\alpha} = \frac{F_i^{\alpha}}{m^{\alpha}} \ , \tag{2.3}$$

where m^{α} is the mass of atom α. It follows that, given an initial set of coordinates and velocities for a molecular system, and given that the force acting on each atom is known, it should be possible to calculate the coordinate for each atom at a later instant. This is achieved via a simple Taylor expansion:

$$r_i^{\alpha}(t+\Delta t) = r_i^{\alpha}(t) + \Delta t \frac{dr_i^{\alpha}(t)}{dt} + \frac{(\Delta t)^2}{2}\frac{d^2 r_i^{\alpha}(t)}{dt^2} \ , \text{ or}$$

$$r_i^{\alpha}(t+\Delta t) = r_i^{\alpha}(t) + \Delta t\, v_i^{\alpha}(t) + \frac{(\Delta t)^2}{2} a_i^{\alpha}(t) \ , \tag{2.4}$$

where t is the time point at which the expansion is carried out, Δt is the time step, and v_i^{α} is the velocity of atom α along component i. In molecular dynamics calculations, the Taylor expansion is typically implemented in a modified form given by Verlet [8,9]. The method carries out two Taylor expansions – one at t+Δt, and the other at t-Δt – and then adds these two expansions to arrive at the following expression:

$$r_i^{\alpha}(t+\Delta t) + r_i^{\alpha}(t-\Delta t) = 2r_i^{\alpha}(t) + (\Delta t)^2 a_i^{\alpha}(t) \ , \text{ or}$$

$$r_i^{\alpha}(t+\Delta t) = 2r_i^{\alpha}(t) - r_i^{\alpha}(t-\Delta t) + (\Delta t)^2 a_i^{\alpha}(t) \tag{2.5}$$

A widely used variation of the Verlet algorithm is the leapfrog algorithm [10], which calculates the velocity at t+(Δt/2) to obtain the position at t+Δt:

$$v_i^{\alpha}\left(t+\frac{\Delta t}{2}\right) = v_i^{\alpha}\left(t-\frac{\Delta t}{2}\right) + \Delta t\, a_i^{\alpha}(t) \tag{2.6}$$

$$r_i^{\alpha}(t+\Delta t) = r_i^{\alpha}(t) + \Delta t\, v_i^{\alpha}\left(t+\frac{\Delta t}{2}\right) \tag{2.7}$$

2.3. Periodic boundary conditions

Biomolecular simulations are typically carried out in an aqueous environment, reflecting their physiological conditions. This means simulations need to be carried out in a solvation box. It follows that molecules that lie near the edges of the box have fewer interacting neighbors than molecules that lie near the centre of the box. This gives rise to surface effects, and the solvent does not have the properties of bulk solvent. Furthermore, molecules near the edges of the box are likely to leave the box, leading to gradual evaporation of the system. These effects are taken care of by employing periodic boundary conditions (PBC), so that the primary simulation box is surrounded by image boxes on all sides. As illustrated in Figure 2.4, when the particle labeled 1 escapes the primary box from the right hand side edge, its image particle (particle 2) in the left hand side image box moves into the primary box. Thus, the number of solvent molecules in the box remains the same. Another important feature of PBC is that molecules in the image box are allowed to interact with nearby molecules in the primary box. This makes sure that a molecule near the edge of the box has interacting partners present on all sides, as would be the case in bulk solvent, and there are no surface effects.

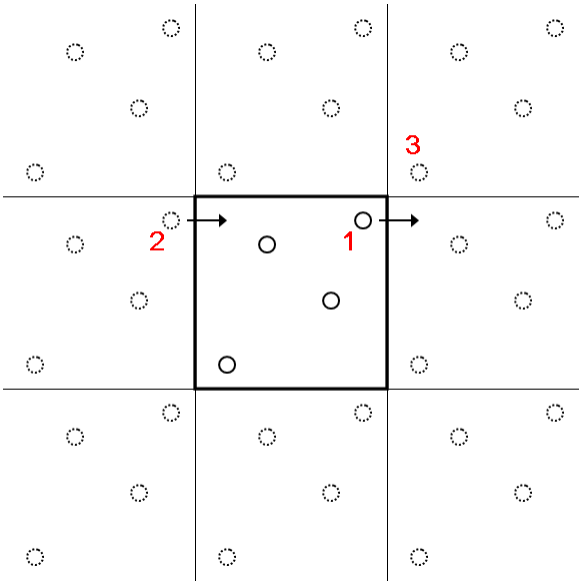

Figure 2.4 The image boxes around a primary simulation box. The boxes that are supposed to lie above and below the plane of the paper have been omitted for clarity. The number labels show (1) a particle that is about to leave the primary box, (2) its corresponding particle in an image box that is about to enter the primary box, and (3) a possible interacting partner in an image box for particle 1.

2.4. Ensembles

MD simulations compute the evolution of a molecular system over time. At a given instant, a molecular system with N atoms can be described by the N atomic positions and N atomic momenta (or atomic velocities), with the positions and momenta having three components each (for the x-, y-, and z-axes, respectively) [11]. Thus, a system has a 6N-dimensional space, called the phase space, in which each point (called a phase point) can be represented by a 6N-dimensional vector. MD simulations sample different points in the phase space of a system. The collection of phase points sampled by an MD simulation in the phase space forms an ensemble. The conformations thus sampled depend on the thermodynamic state at which the simulation is being performed. The most common ensembles in MD simulations are (i) microcanonical (also called NVE), which corresponds to a system at constant volume and energy, (ii) canonical (NVT), which corresponds to a system at constant volume and temperature, and (iii) isobaric-isothermal (NPT), which is performed at constant pressure and temperature. The "N" in NVE, NVT, and NPT indicates that the number of atoms is fixed.

The average value of a property A over an ensemble, known as the ensemble average, can be calculated as

$$< A> = \int dr' dp' A(r', p')\rho(r', p') \,, \qquad (2.8)$$

where *r'* and *p'* denote the positions and momenta, respectively, of all the atoms in the system, and $\rho(r',p')$ is the probability of finding the system in a particular phase point. The above integration is performed over the 3N components of atomic positions and the 3N components of atomic momenta, so the above expression is actually a 6N-dimensional integral. The integration is carried out over all phase points of the system. Alternatively, the average might also be calculated from a simulation by averaging the property over the simulation time:

$$< A> = \lim_{\tau\to\infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(r'(t), p'(t))\, dt \qquad (2.9)$$

Here τ is the total simulation time. An average calculated in this manner is known as the time average. This brings us to one of the central tenets of molecular dynamics, called the Ergodic hypothesis. It states that, if a simulation is carried out long enough to sample all possible phase points of a molecule, the time average of a property would be equal to its ensemble average. It

follows that the average values of macroscopic properties like energy, pressure, and heat capacity can be calculated from time averages over MD trajectories.

2.5 Enhanced sampling methods

Many biological phenomena, like the transport of ions across channels and large scale motions of domains in transporters, occur at timescales of microseconds. It is a challenging and computationally expensive task to sample all relevant states that are seen during such phenomena with the use of regular MD simulations. Modeling such phenomena is made possible by enhanced sampling methods like replica exchange and umbrella sampling.

2.5.1 Replica exchange

Replica exchange molecular dynamics (REMD) involves performing simulations on a number of replicas of a system at a range of temperatures, with replicas at adjacent temperatures being swapped at regular intervals [12]. Since the replicas at the higher temperatures are able to overcome energy barriers owing to high internal energy, they are able to sample conformations from a number of energy wells. Thus, swapping between replicas at different temperatures greatly enhances the sampling for the replicas at the lower temperatures.

In a replica exchange simulation, attempted swaps between replicas are accepted or rejected based on the Metropolis criterion with the probability given by

$$p = \min\{1, e^{(\beta_i - \beta_j)(E_i - E_j)}\} \ , \tag{2.10}$$

where i and j denote two adjacent replicas, and E_i and E_j are the total energies of the two replicas. After an exchange, the momentum for any particle α at the new temperature is given by

$$p_j^{\alpha} = \sqrt{\frac{T_{new}}{T_{old}}}\, p_i^{\alpha} \ , \tag{2.11}$$

where p_j^{α} is the momentum of the particle α along a given component (x, y, or z) after the exchange, p_i^{α} is the momentum along the same component before the exchange.

2.5.2 Umbrella sampling

Umbrella sampling enables sampling along a reaction coordinate by applying a bias on the system in the form of a harmonic potential [13]. A number of windows are defined along the reaction coordinate, and independent simulations are performed for each window. For investigating ion transport (or urea transport, for that matter), the biasing potential is applied on the ion so that different positions of the ion along the pore axis are sampled. For a window i, the biasing potential has the form

$$w_i(z) = \frac{1}{2} k(z - z_{i,0})^2 \tag{2.12}$$

where k is the harmonic constant, $z_{i,0}$ is the center of the window i, and z is the position of the ion along the reaction coordinate at any given instant. The potential energy function then has the form

$$E^b(r^N) = E^u(r^N) + w_i(z) \tag{2.13}$$

where r^N denotes the positions of all the atoms in the system, and $E_u(r^N)$ is the unbiased potential energy function. The unbiased probability distribution of the ion along the pore axis $P_i^u(z)$ can be computed from the biased probability distribution $P_i^b(z)$, and is given by

$$P_i^u(z) = P_i^b(z) e^{\beta w_i(z)} \langle e^{-\beta w_i(z)} \rangle \tag{2.14}$$

The free energy along the reaction coordinate is then given by

$$A_i(z) = -\frac{1}{\beta} \ln P_i^b(z) - w_i(z) + F_i \ , \tag{2.15}$$

where F_i is a constant for window i. There are several approaches for determining F_i, of which the most popular is the weighted histogram analysis method (WHAM) [14-16]. It calculates the unbiased global distribution $P^u(z)$ using the relation

$$P^u(z) = \sum_i p_i(z) P_i^u(z) \ , \tag{2.16}$$

where $p_i(z)$ is the weight for window i which ensures that the statistical error for $P_u(z)$ is minimized. It is given by

$$p_i = N_i \, e^{-\beta w_i(z) + \beta F_i} \ , \tag{2.17}$$

where N_i is the number of steps sampled for window i. F_i is calculated using the relation

$$e^{-\beta F_i} = \int P^u(z) e^{-\beta w_i(z)} dz \tag{2.18}$$

$P^u(z)$ and F_i are calculated in a self-consistent manner using the above equations.

2.6 Setting up membrane protein simulations

Setting up and running membrane protein simulations is more complicated than that for soluble proteins, owing to the presence of the lipid bilayer.

2.6.1 Membrane setup

The protein is usually aligned along the z-axis, so that the membrane can later be set up in the xy-plane. If the protein possesses a pore, as any transport protein would do, it is filled with water molecules. The next step is to model the protein in the bilayer, and this can be done by the insertion method or the replacement method [17]. In the insertion method, radially outward forces are applied on a lipid bilayer in order to generate a hole which is about the size of the protein. Since the forces are directed radially, this method is suited for proteins that have a cylindrical shape. In the replacement method, lipid molecules are built around the protein, so the manner in which lipid molecules are placed around the protein is sensitive to the shape of the protein. Consequently, this method is suitable for proteins of all kinds of shapes, but is slower than the replacement method.

2.6.2 Equilibration and production

In equilibration of membrane proteins, there should necessarily be restraints to retain lipid headgroups in one plane, in addition to restraints on the heavy atoms in the protein. All restraints are gradually removed over the course of the equilibration. The initial phase of the equilibration is done in the NVT ensemble, and the latter phase in the NPAT ensemble (where “A” in NPAT indicates constant area in the xy-plane). Production runs are usually performed in the NPAT ensemble without any restraint on the system (as described later, many of the simulations described here have been performed with a center of mass restraint on the protein to prevent overall drifting of the protein).

References

1. MacKerell AD (2004) Empirical force fields for biological macromolecules: Overview and issues. J Comput Chem 25(13):1584–1604.

2. Best RB, et al. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. J Chem Theory Comput 8(9):3257–3273.
3. MacKerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25(11):1400–1415.
4. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616.
5. Hart K, et al. (2012) Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. J Chem Theory Comput 8(1):348–362.
6. Denning EJ, Priyakumar UD, Nilsson L, MacKerell AD (2011) Impact of 2′-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. J Comput Chem 32(9):1929–1943.
7. Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev 106(5):1589-1615.
8. Verlet L (1967) Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys Rev 159(1):98–103.
9. Verlet L (1968) Computer "experiments" on classical fluids. II. Equilibrium correlation functions. Phys Rev 165(1):201–214.
10. Hockney RW (1970) Methods in Computational Physics. Alder, B:136-211.
11. Leach A (2001) Molecular modelling: Principles and applications (2e) Pearson.
12. Earl DJ, Deem MW (2005) Parallel tempering: Theory, applications, and new perspectives. Phys Chem Chem Phys 7(23):3910.
13. Kästner J (2011) Umbrella sampling. WIREs Comput Mol Sci 1(6):932–942.
14. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13(8):1011–1021.
15. Roux B (1995) The calculation of the potential of mean force using computer simulations. Comp Phys Comm 91(1-3):275–282.

16. Grossfield A “WHAM: the weighted histogram analysis method”, version 2.0.6, http://membrane.urmc.rochester.edu/content/wham.
17. Jo S, Kim T, Im W (2007) Automated builder and database of protein/membrane complexes for molecular dynamics simulations. PLoS ONE 2(9):e880.

3 Oligomeric State of HIV-1 Vpu

3.1 Introduction

The human immunodeficiency virus type-1 (HIV-1) employs a range of viral proteins to successfully establish and propagate infection in the host. These include the structural envelope (Env gp120, gp41), capsid ($p24^{CA}$) and matrix ($p17^{MA}$) proteins, the enzymes reverse transcriptase, ribonuclease H, integrase and protease, two regulatory proteins (Rev, Tat), and four accessory proteins (Nef, Vif, Vpr and Vpu) [1]. Of these, the accessory proteins are not required for viral replication in vitro but are indispensable for the establishment and persistence of HIV infection and pathogenesis [2]. The viral protein U (Vpu) is an 81-amino acid transmembrane (TM) protein encoded by HIV-1 that increases virus release from host cells [3,4]. The protein, however, is not encoded by the less virulent human immunodeficiency virus type-2 (HIV-2) and simian immunodeficiency virus (SIV) [4]. Knowledge of the three-dimensional structure of the oligomeric form of Vpu is expected to further our understanding of its functional mechanisms, and this can possibly be exploited as a drug target [5-7].

Vpu is a type I integral membrane protein with an N-terminal transmembrane domain (TMD) and a C-terminal cytoplasmic domain [8]. The cytoplasmic domain contains two α-helices [9,10] and is involved in the degradation of CD4 molecules at the endoplasmic reticulum [11,12]. Between the α-helical domains are two serine residues, Ser52 and Ser56 [13], which must necessarily be phosphorylated for Vpu to exhibit its CD4 degradation activity [14]. The TMD helps in viral release from host cells [12], which is brought about by the degradation of tetherin, an antiviral protein encoded by host cells that causes retention of virions on the cell surface [15]. Three residues in the Vpu TMD, Ala14, Ala18, and Trp22, have been shown to be important for this activity [16]. The ability of Vpu to oligomerize [8] allows it to form cation-selective ion channels [17]. Such channels can form both in planar lipid bilayers and in the plasma membrane of Escherichia coli in vivo [17], and they are known to facilitate virus release [5]. The ion channel activity ascribed to the TMD of Vpu [5] appears to be rather weak and the characteristics of the channel almost resemble those of a pore [18]. Gel permeation chromatography studies

show Vpu to be a pentamer [19], but recent photo-induced cross-linking studies indicate that a variety of oligomeric states might exist [20].

Solution NMR studies in lipid micelles on Vpu_{2-37}, a truncated form of Vpu containing the N-terminal TMD, showed an α-helix spanning residues 9 to 29 [21]. Similar studies on Vpu_{2-30}, a peptide containing residues 2 to 30 from Vpu with a 6-residue solubility tag, revealed an α-helix spanning residues 8 to 25 [22]. The helix has a kink around Ile17, and is tilted at an angle of 13° with respect to the membrane. Fourier transform infrared (FTIR) spectroscopy on the first 31 N-terminal residues of Vpu indicate an α-helix with a tilt of (6.5±1.7)° and a rotational pitch angle of (283±11)° around Val13 [23]. Simulated annealing with restrained molecular dynamics on rotationally symmetrical tetramers, pentamers, and hexamers of Vpu shows that only a pentamer has a rotational pitch angle for Val13 close to the experimental value [23]. Molecular dynamics (MD) simulations restraining the motion of ions to the axis of the pore show that the conductance of the pentamer is closer to the experimentally observed value than either the tetramer or the hexamer [24]. MD simulations performed using an octane layer for mimicking the properties of a lipid bilayer have shown that a helix is expelled from a hexameric arrangement; the same was not observed for a pentamer [25,26]. A number of modeling and simulation studies have been carried out by modeling the channel as a homo-pentamer [27-32]. Pentamer models have also been generated using pre-equilibrated monomers and these show the lumen of the pore to be a hydrophobic stretch [32]. However, all of the studies mentioned above were carried out with the assumption that the native oligomeric state is a pentamer. Notably, the modeling studies that originally suggested Vpu to form a pentamer did not take into consideration a fully hydrated lipid bilayer environment [24-26]. A systematic study of the Vpu TMD tetramers, pentamers and hexamers, taking into account the explicit lipid environment would provide atomistic details on the oligomeric structure, and the factors that determine the stability of the native structure.

A useful approach for evaluating the stability of different oligomeric states of a membrane protein is the use of replica-exchange molecular dynamics (REMD) in an implicit membrane environment [33,34]. REMD overcomes the problem of entrapment in local minima [35], thereby making the sampling of regions of phase space possible that are otherwise not accessible to constant temperature molecular dynamics. Implicit membrane models take into consideration

the physical environment in which a membrane protein finds itself without having real membrane and solvent molecules, making the approach computationally efficient [33,36]. Combining implicit membrane models with REMD has made it possible to estimate the stability of different oligomeric states in terms of both potential energy and free energy [33,34]. This study was carried out to model the possible oligomeric states of the Vpu TMD, and to understand the structural and energetic factors that make one oligomeric state more stable than others. Briefly, REMD simulations have been used with an implicit membrane model for sampling varied conformations of different possible oligomeric states of Vpu. Representative structures have then been selected for more extensive studies in fully hydrated lipid bilayers. The pentameric state is shown to be the most favored state, and structural features of the protein are described that help explain its function.

3.2 Methods

3.2.1 Modeling of TMD and replica-exchange molecular dynamics

The TMD of Vpu is important for its ion channel activity, tetherin degradation and virus release, and hence only this domain was considered for our modeling studies. Since the TMD is known to have a helical structure [21-23,37-39] the first 32 residues of HIV-1 NL4-3 Vpu (MVPIIVAIVALVVAIII AIVVWSIVIIEYRKI) were modeled as an idealized α-helix using the molecular modeling program SYBYL7.2 (Tripos International, St. Louis, Missouri, USA; http://www.tripos.com). While previous studies have shown residues 5 to 29 to form the TMD [21,22,40] we have extended this by a few residues on either side to ensure that there are no destabilizing effects due to abrupt termination of the TMD. Previous modeling studies have shown that these TMD extensions develop α-helical conformation when modeled in a lipid bilayer environment, and that these extensions play a role in stabilizing the helix via nonbonded interactions with lipid headgroups [40]. Different oligomeric states of the TMD were modeled using the methodology described by Bu et al. [34]. The TMD was aligned along the z-axis, moved by a distance of 20 Å in the positive direction of the y-axis, and then given two, three, four, five, and sixfold rotational symmetries about the z-axis using the IMAGE facility in CHARMM [41,42] to give a dimer, trimer, tetramer, pentamer and hexamer, respectively. For comparison, a monomer aligned along the z-axis was also modeled. Although it is unlikely that

the dimer or trimer will form the channel, these forms were modeled to examine the structural and energetic changes accompanying sequential assembly from monomer to higher oligomeric states. Figure 3.1 shows the initial arrangement of the tetramer, the pentamer, and the hexamer. The helices were placed at large separations to encourage rearrangement of the helices.

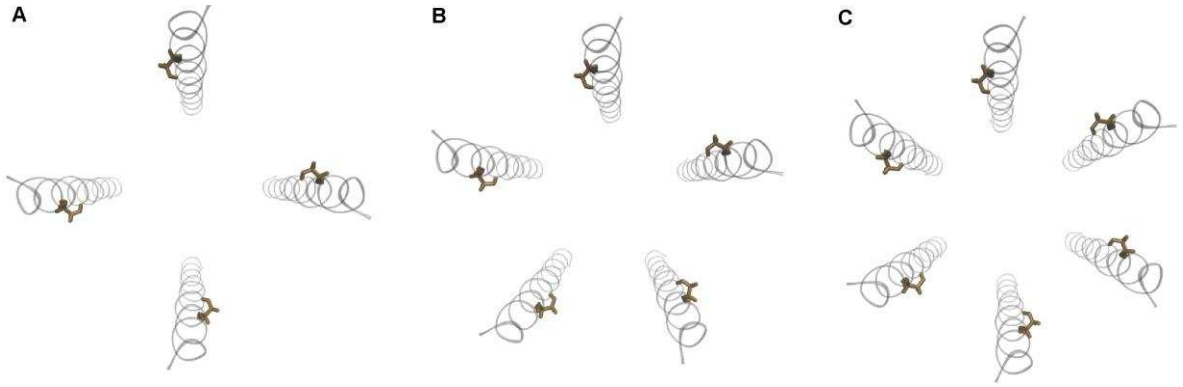

Figure 3.1 (A) Tetramer, (B) pentamer, and (C) hexamer models before the REMD simulations. The orientation of the hydrophilic Ser23 residue is also shown in the figure.

The oligomeric forms were first simulated in an implicit membrane environment. The generalised Born model with a simple switching function (GBSW) module [33,36] was used with a surface tension coefficient of 0.03 kcal mol^{-1} $Å^{-2}$. An implicit membrane with a hydrophobic core of thickness 35 Å was placed perpendicular to the z-axis. A smoothing region of thickness 0.5 Å was used on both sides of the hydrophobic core for a smooth transition from the hydrophobic implicit membrane to the hydophilic continuum solvent. REMD simulations were carried out for each oligomeric state using the CHARMM22 all-atom protein force field with CMAP corrections [43,44]. A total of eight replicas were distributed over an exponentially spaced temperature range from 300 K to 400 K. Temperatures above 400 K were not used in the study to avoid non-physical distortion of the structures. Since the oligomeric TMDs have a cylindrical geometry, a cylindrical harmonic restraint with force constant 1 kcal mol^{-1} $Å^{-2}$ and radius 25 Å was applied, thereby limiting the drifting of the TMDs. Langevin dynamics was used with a friction coefficient of 5 ps^{-1} for heavy atoms, and exchanges were attempted every 1 ps. Simulations were run for 10 ns, with the first 1 ns being considered as the equilibration period. The last 9 ns of the trajectories from simulations at the lowest temperature (i.e., 300 K) were used for analysis. The rotational and translational entropy terms were calculated from the

principal moments of inertia, while the vibrational entropy was calculated by first removing rotation and translation from all the frames, and then carrying out quasiharmonic mode analysis [45,46].

3.2.2 Selection of representative structures and further equilibration

Two model structures for each of the oligomeric states (tetramer, pentamer and hexamer) were chosen for further consideration for explicit membrane simulations, details of which are given in this section. The tilt angle of the helical principal axis with respect to the membrane normal was calculated for all the conformations sampled in the last 9 ns of simulation. The range of tilt angle values that occurred most frequently was determined from a probability distribution of tilt angles. For selecting a representative structure that may further be used in extensive MD simulations, a set of structures was chosen that had tilt angles with a high probability of occurrence. From this set for each oligomeric state, the structure with the lowest molecular mechanical energy was selected as the representative structure. Each of the representative structures was then equilibrated for 10 ns in a GBSW implicit membrane with the same parameters as given above but without any constraints on any part of the protein. The temperature was kept constant at 300 K using the Nosé-Hoover thermostat [47,48]. The configuration obtained after these 10 ns implicit membrane simulation of each of the three oligmeric states was taken as the other model. Since it is not straightforward to choose a reliable model without a reference structure as standard, we have considered the most sampled conformations (on the basis of tilt angle) as two different models for further calculations.

3.2.3 Molecular dynamics in a fully hydrated lipid bilayer

Since the representative dimer and trimer did not form a compact structure after the above equilibration step (see Results and Discussion), only the tetramer, pentamer and hexamer were considered for further studies. To investigate the stability of these oligomeric states in a membrane-like environment, simulations were carried out in a solvated lipid bilayer. For each of these three oligomeric states, two independent simulations were carried out, one with the representative structures after the 10 ns constant temperature simulation in an implicit membrane (henceforth referred to as “Model 1”), and the other with the structures before this 10 ns simulation (“Model 2”). The protein-lipid-solvent system was set up using the CHARMM-GUI

Membrane Builder [49,50]. The protein was first aligned along the z-axis and pore water was generated. A homogeneous lipid bilayer of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) was generated around all the oligomeric forms. The membrane was perpendicular to the z-axis, with its center at z = 0 Å. Bulk water of 15 Å thickness was placed above and below the membrane. Potassium and chloride ions were added to attain a salt concentration of 0.15 M KCl and a zero net charge on the system. Details about the number of each component in all the systems are shown in Table 3.1.

Table 3.1 Number of each component in the systems studied

System	Protein residues	Lipid (POPC) molecules	Water molecules	K^+ ions	Cl^- ions	Total number of atoms	System size along x-, y-, and z-axes (in Å)
Tetramer, Model 1	128	94	5307	13	17	30735	62 x 62 x 82
Pentamer, Model 1	160	95	5496	13	18	31984	63 x 63 x 82
Hexamer, Model 1	192	105	6848	16	22	37934	67 x 67 x 86
Tetramer, Model 2	128	85	4755	12	16	27871	60 x 60 x 80
Pentamer, Model 2	160	99	6050	15	20	34186	64 x 64 x 84
Hexamer, Model 2	192	89	5657	13	19	32211	64 x 64 x 82

The CHARMM22 all-atom protein force field including CMAP corrections [43,44], the CHARMM36 all-atom lipid force field [51], and the modified TIP3P water model [52] were used for the simulations. Periodic boundary conditions were set up, and the particle mesh Ewald method was used for calculating long-range electrostatic interactions [53]. Lennard-Jones interactions were modulated by a switching function between 10 Å and 12 Å, with all nonbonded interactions being truncated at 12 Å. The covalent bonds involving hydrogen atoms were

constrained using SHAKE [54]. The system was equilibrated using the six-step equilibration scheme proposed by Jo et al. [49]. Positional harmonic restraints were applied on the ions and the heavy atoms of the protein to hold them in place during the initial equilibration simulations. Planar harmonic restraints were applied on water molecules to ensure that no water molecule entered the hydrophobic region of the membrane. Also, lipid head groups were retained close to the membrane-water interface using planar harmonic restraints. The harmonic restraints on the different components were gradually reduced during the equilibration. The first two equilibration steps were carried out in the NVT ensemble (constant volume and temperature) and the last four in the NPT ensemble (constant pressure and temperature), keeping the temperature at 303.15 K using the Nosé-Hoover thermostat [47,48]. Production runs were carried out in the NPT ensemble for 10 ns with a timestep of 2 fs. The pentamer models were simulated for 30 ns to investigate the stability of the models. The molecular visualization program VMD [55] was used for rendering images and analyzing hydrogen bond interactions. Pore radius was measured using the HOLE2 program [56]. All MD simulations and other analyses were performed using the CHARMM program [41,42].

3.3 Results and Discussion

3.3.1 Higher oligomers display reduced tilt

The implicit membrane model that was used for the REMD simulations uses generalized Born electrostatics for modelling the solvent on both sides of the membrane [36]. The membrane hydrophobic core is represented using a low-dielectric slab with a fixed thickness [33]. Thus, when there is mismatch between the hydrophobic regions of a helical protein and the continuum solvent region, the membrane cannot respond by altering its thickness. The only mechanism by which such a mismatch is minimized is by tilting of the helices to ensure that as much of the hydrophobic part of the protein lies in the membrane as possible. Figure 3.2 (A) shows the probability distribution of the tilt angles of different oligomers over the last 9 ns of REMD. The tilt of the helices was seen to decrease with an increase in the number of helices in the system. In the higher oligomers (tetramer, pentamer and hexamer), the helices are packed closely together, allowing interhelical interactions to occur. These interactions are stabilizing, so the helices are able to overcome the destabilizing effect of hydrophobic mismatch. Thus, the protein does not

have to tilt too much to attain an energetically stable conformation. In the dimer and trimer, however, interhelical interactions are almost absent with hydrophobic mismatch being the major factor affecting the orientation of helices, and the protein tilts until most of its hydrophobic residues are buried in the membrane core.

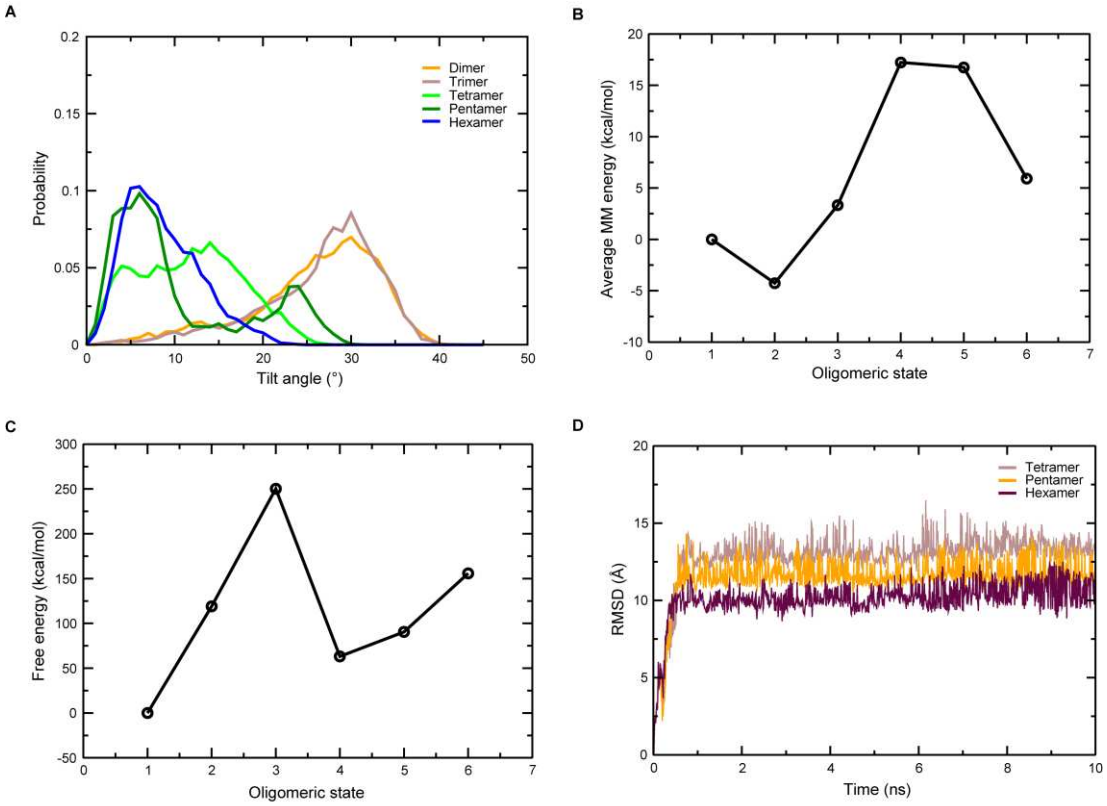

Figure 3.2 Replica-exchange molecular dynamics in an implicit membrane environment. (A) Probability distribution of the tilt angle for the conformations sampled at 300 K from the last 9 ns of replica-exchange molecular dynamics. (B) Average potential energy and (C) free energy of the different oligomeric states over the last 9 ns of replica-exchange molecular dynamics. The values shown are relative to the monomer. (D) RMSD of the tetramer, the pentamer, and the hexamer in the REMD simulations.

3.3.2 Tetramer, pentamer, and hexamer are possible oligomeric states

The average molecular mechanical potential energy of the TMD in different oligomeric states relative to the monomeric state is shown in Figure 3.2 (B). The energies of all the states were comparable, with the average energy of any given oligomeric form differing by less than 17 kcal mol^{-1} compared to the monomeric form. Based on this analysis, no particular oligomer could be identified as the native form. Bu et al. have suggested the necessity of considering entropic factors due to the assembly of helices [34]. The entropy loss accompanied by the formation of a

given oligomeric state was calculated by taking the entropy term for the monomer as the reference. The free energy values (Table 3.2 and Figure 3.2 (C)), which were obtained after taking into account entropy loss and stabilization arising from solvation, indicate the tetramer and pentamer to be stable oligomeric states. There is an increase in the free energy from monomer to dimer, and from dimer to trimer; this is followed by a remarkable decrease in the free energy in the trimer-tetramer transition, and a slight increase in going from tetramer to pentamer to hexamer. Figure 3.2 (D) shows the RMSD of the tetramer, the pentamer, and the hexamer in the REMD simulations. The RMSD changes rapidly in the initial period before being converged, indicating rearrangement of the helices until a stable conformation is attained.

Table 3.2 Entropy loss and free energies of the different oligomeric states in the replica-exchange MD simulations

Oligomeric state	**Potential energy (kcal/mol)**	**Solvation energy (kcal/mol)**	**Rotational entropy term, TS_{rot} (kcal/mol)**	**Translational entropy term, TS_{trans} (kcal/mol)**	**Vibrational entropy term, TS_{vib} (kcal/mol)**	**Free energy (kcal/mol)**
Monomer	487.32	-295.42	0	0	0	191.90
Dimer	483.06	-195.19	1.81	0.62	-25.46	310.90
Trimer	490.65	-108.26	2.34	0.98	-63.03	442.10
Tetramer	504.56	-232.64	1.86	1.24	13.79	255.03
Pentamer	504.07	-207.78	2.16	1.44	10.29	282.40
Hexamer	493.24	-152.71	2.42	1.60	-11.15	347.66

Equilibration of the representative structures sampled from REMD showed that the dimer and trimer do not form a compact structure, with the helices lying far apart (Figure 3.3). Since the dimer and trimer are not feasible oligomeric structures, the rest of this report is concerned with the tetramer, pentamer and hexamer, unless stated otherwise. The use of REMD ensures that much of the phase space of the various oligomeric states, and a majority of all possible conformations are sampled. As shown above, the tetramer and pentamer were identified as stable oligomeric states using REMD with an implicit membrane. Although it is able to model the physical characteristics of the membrane hydrophobic core and bulk solvent, a drawback of the

generalized Born implicit membrane model used here is that it does not take into account the hydrophilic nature of the pore region, which is central to the functioning of ion channels and strongly influences the orientation and behavior of residues lining the pore. A more realistic representation of the channel is therefore possible only with accurate modeling of the pore region. Thus, a comprehensive investigation of the stabilities of the tetramer, the pentamer, and the hexamer has been carried out in a hydrated lipid bilayer environment.

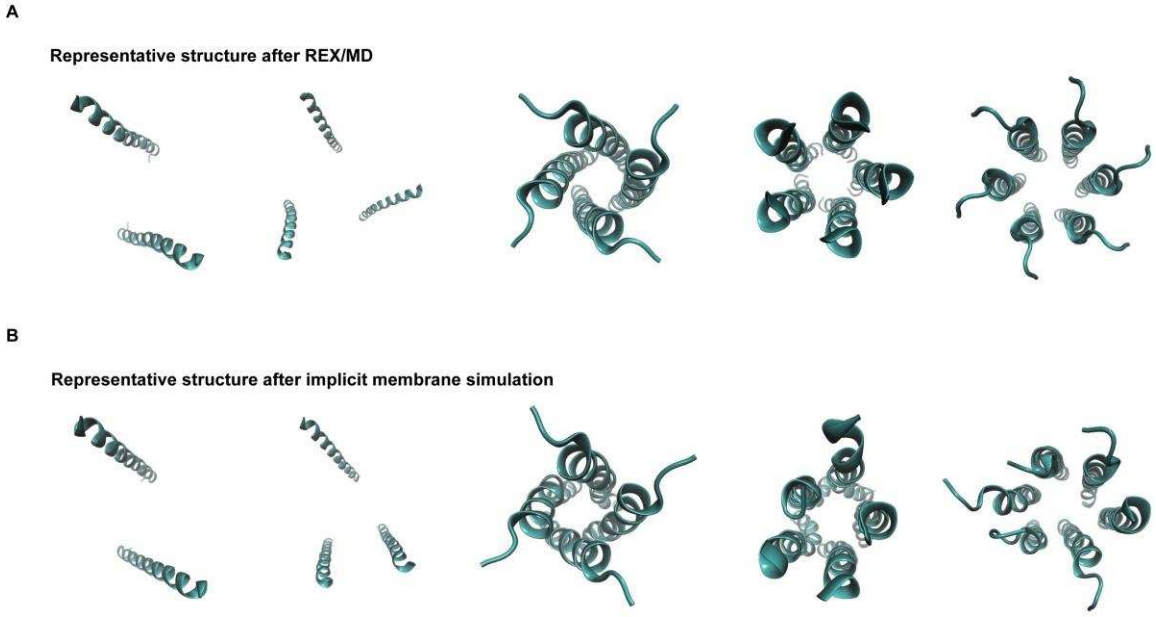

Figure 3.3 The representative structures for the different oligomeric states (A) before and (B) after 10 ns simulation in an implicit membrane environment are shown.

3.3.3 Explicit membrane MD simulations reveal the pentamer to be the most stable oligomeric state

Two independent simulations were carried out for each of the tetrameric, pentameric and hexameric states in the explicit bilayer environment (see Methods section). In both sets of simulations, visual inspection revealed that only the pentamer retained the rotational symmetry necessary for forming an ion channel (Figure 3.4 (A)). The root mean-square deviation (RMSD) values suggest that the pentameric forms attain equilibration in the first 2 ns, but the tetramers and the hexamers become distorted and completely lose their initial pore-like structure (Figure 3.4 (B)). Such distortion was seen in both the model structures for the tetrameric and hexameric states. However, model 1 of the tetramer exhibits RMSD values that are comparable to the pentamer. Interestingly, in one of the hexamer models (Model 1), a helix is expelled from the

bundle, leading to a pentameric structure. Such expulsion has been reported by Lopez et al. in studies on a hexameric form of the Vpu TMD in an octane environment [26], and is indicative of a propensity of the Vpu TMD to exist in a pentameric state. This is in agreement with earlier studies which suggest that the oligomer exists in a pentameric form [19,23]. The possible factors that favor the pentameric form over the other forms are elucidated below.

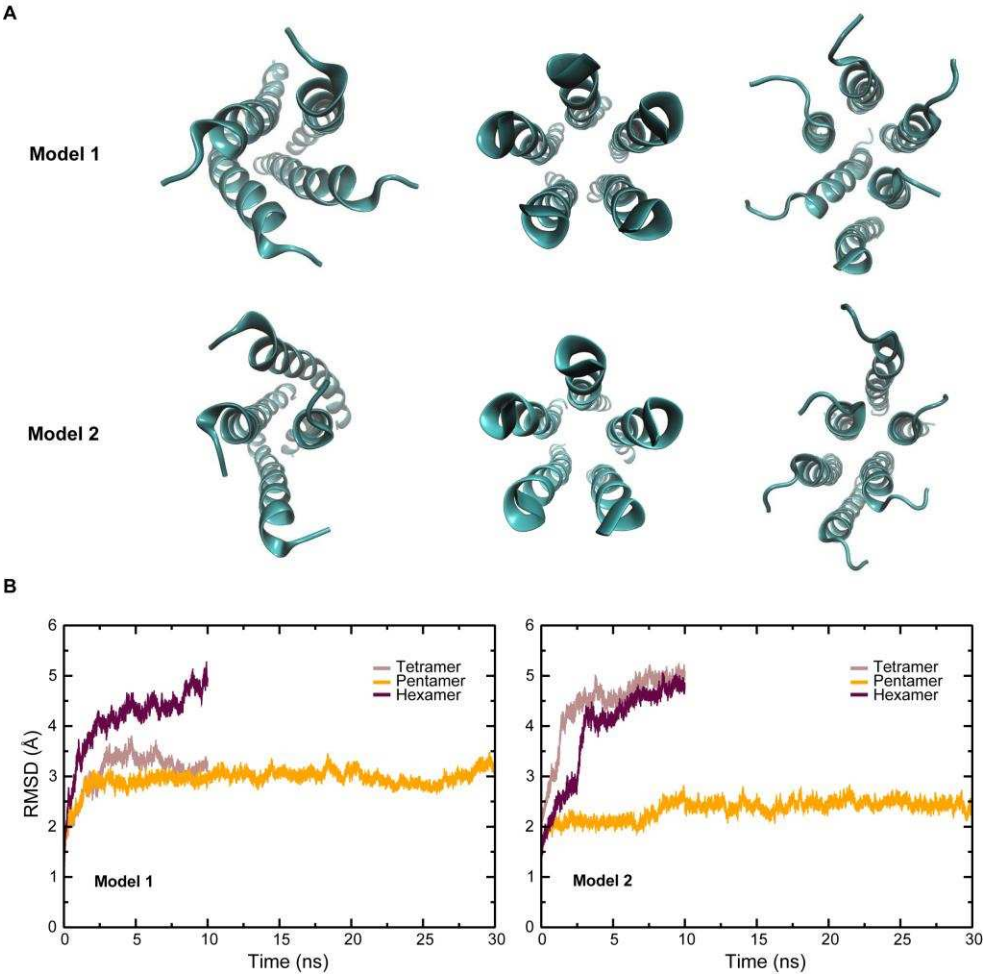

Figure 3.4 Molecular dynamics in an explicit membrane environment. (A) Models for the tetramer, pentamer and hexamer after simulation in a fully hydrated lipid bilayer. The images for the pentamer are after 30 ns, and those for the tetramer and hexamer are after 10 ns. The lipid bilayer and solvent molecules have been omitted for clarity. The pentamer retained a channel-like structure in both the simulations. (B) RMSD of the different oligomeric states.

3.3.4 The helices in the pentamer are held together by strong van der Waals interactions

We looked at the van der Waals interaction energy between neighbouring helices in the different oligomers. As can be seen clearly, interhelical van der Waals forces greatly stabilize the helices in the pentamer (Figure 3.5 (A)). The interhelical distances and rotational symmetry of the

pentameric state are probably better suited for van der Waals interactions than other oligomeric states. The significance of van der Waals interactions over other forms of nonbonded interactions between adjoining helices can be attributed to the fact that the residues in the interface between adjoining helices are non-polar, as apparent from the contact maps shown in Figure 3.5 (B). Since the structural features of models 1 and 2 are similar, only the figures for model 1 are shown hereafter. The non-polar residues form close contacts, making effective interhelical interactions. Most interhelical contacts occur on the N-terminal side of the channel, where the residues are all hydrophobic. A notable exception to the non-polar nature of these contacts is a salt bridge between Glu28 and Arg30 on neighboring helices, which is seen in the top right corner of the contact maps (Figure 3.5 (B)). One of the amino groups on the Arg30 side chain faces the carboxyl group on Glu28, while the other amino group points towards a phosphate oxygen in a nearby lipid molecule as shown in model 1 (Figure 3.5 (C)). The salt bridge is seen to occur between all pairs of neighboring helices (Figure 3.5 (C)), and it exists consistently in both the simulations. Such a salt bridge satisfies the hydrogen bond requirements for the two charged residues, thereby stabilizing the residues in an otherwise hydrophobic environment. In the tetramer and hexamer, however, the salt bridge is absent in some pairs of adjacent helices; this might be an important factor in stabilizing the oligomer.

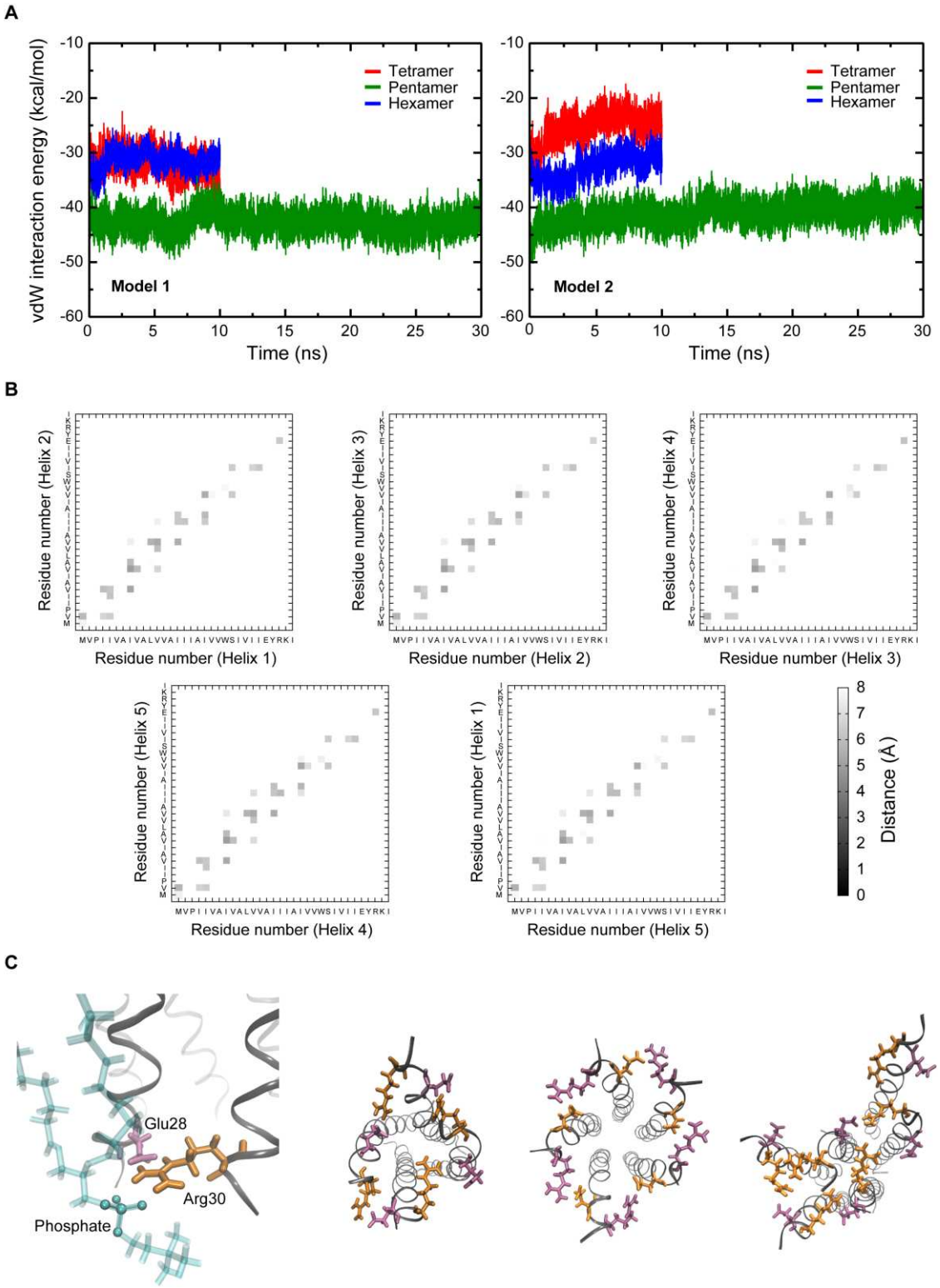

Figure 3.5 Interhelical interactions in the oligomers. (A) Interhelical van der Waals interaction energy per helix pair for the tetramer, the pentamer, and the hexamer in a lipid bilayer environment. The interhelical van der Waals interaction energy was calculated for all adjoining helix pairs in the oligomer and then divided by the number of

helices to give the average value per helix pair. (B) Contacts between residues on adjoining helices. Residue-residue distances have been averaged over time. The values shown are for Model 1. (C) Orientation of Arg30 ("licorice" representation, colored orange). One of the amino groups forms a salt bridge with Glu28 ("licorice" representation, colored mauve) on a neighboring helix, while the other interacts with headgroup oxygens (headgroup phosphate is shown in "CPK" representation). The TMD is shown in "ribbons" representation, and a POPC molecule is shown in "bonds" representation. The five salt bridges in the tetramer, pentamer and hexamer are also shown.

The probability distribution for interhelical distances between all adjoining helix pairs over the simulation was carried out, with the distance between the centers-of-mass of adjacent helices taken as the interhelical distance. The terminal residues were not considered in calculating the center-of-mass to avoid high fluctuations due to these residues. The narrow distribution for the pentamer suggests that the interhelical distance remains stabilised at a given distance (Figure 3.6 (A)). Furthermore, the occurrence of peaks centered around the same point for all helix pairs in the pentamer indicates that the interhelical distance is almost the same in all helix pairs. However, for the tetrameric and hexameric structures, the interhelical distances calculated for the helical pairs and the broad nature of some curves indicate structures that are not very stable (Figure 3.6 (A)). These observations support the highly symmetric nature of the pentameric structure, but not the tetrameric or hexameric structures.

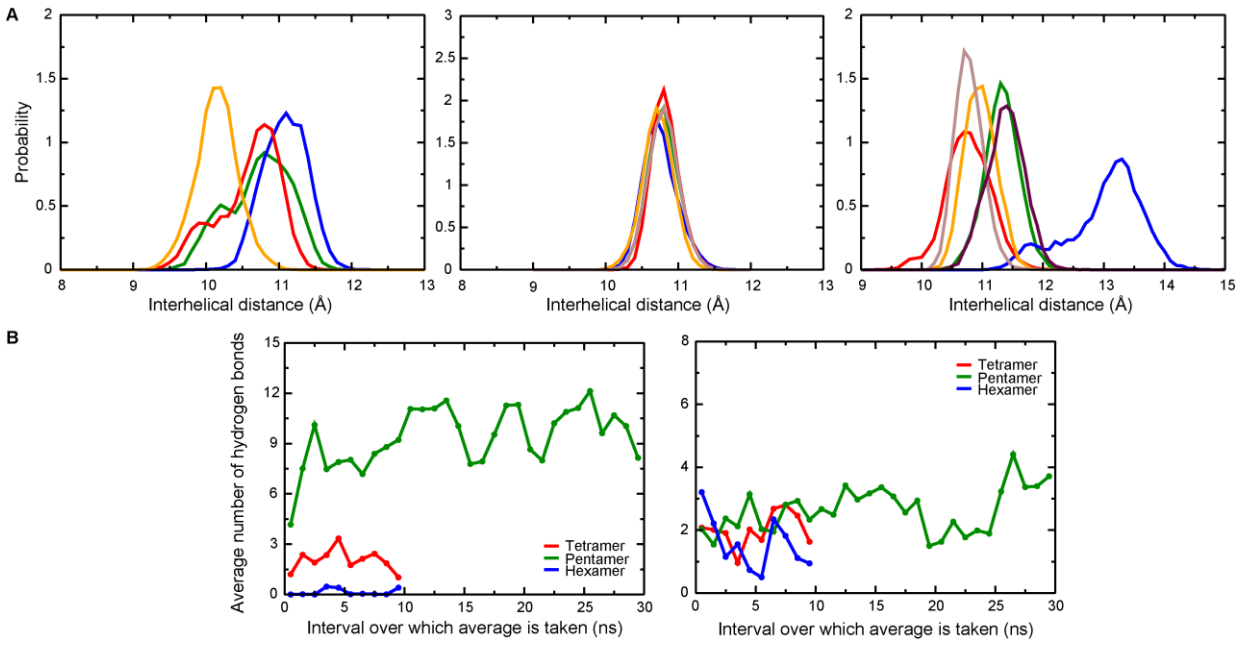

Figure 3.6 Interhelical distance and protein-lipid interactions. (A) Probability distribution of interhelical distance for tetramer, pentamer and hexamer. The distance between the centers-of-mass of adjoining helices was calculated. Only the helical backbone was considered, and the top three and bottom three residues were neglected. (B) Average

number of hydrogen bonds between lipid headgroups and polar residues for Arg30 and headgroup (left panel), and Tyr29 and headgroup (right panel). The cutoffs used were 3.5 Å for the donor-acceptor distance, and 45° for the donor-hydrogen-acceptor angle.

3.3.5 The hydrophilic and basic residues in the TMD interact with lipid headgroups

We then estimated the average number of hydrogen bonds between polar residues on the protein and the lipid headgroups (the cutoffs used in calculating the number of hydrogen bonds were 3.5 Å for the donor-acceptor distance, and 45° for the donor-hydrogen-acceptor angle). As described above, the arginine residues within the TMD have one of their amino groups facing the lipid headgroup, allowing the formation of hydrogen bonds between the side chain and the headgroup oxygen atoms. The number of such interactions with the headgroup is significantly higher in the pentamer (Figure 3.6 (B)), and might play a role in adhering the protein to the lipid bilayer. Tyrosine residues are also able to form hydrogen bonds, although the number of bonds is fewer in number than those due to arginine, especially in the pentamer. This might be due to tyrosine side chains lying slightly above the plane of the headgroup oxygens, while the arginine side chains lie in the plane of these oxygens. Moreover, the arginine side chains are oriented in a direction perpendicular to the axis of the lipid molecules, thereby making favorable hydrogen bond angles. Interactions between positively charged residues and lipid headgroups have previously been shown to be crucial in structures of ion channels as reported by both experimental [57,58] and computational studies [59,60].

3.3.6 A hydrophobic region occurs around the middle of the channel

The Ser23 residue faces the interior of the channel (Figure 3.7 (A)), providing a hydrophilic region in the pore around that residue. Notably, the initial orientations of the side chains were chosen to be random in which Ser23 was facing the exterior of the pore in the beginning of the REMD simulations (Figure 3.1). The explicit lipid bilayer simulations were started with a conformation where the pore was uniformly solvated, but most of the water molecules were expelled within the first 50 ps of the equilibration period. At the end of 30 ns of production run, much of the pore water was concentrated around the serine residue and toward the ends of the pore, near bulk water, rather than being spread uniformly across the pore (Figure 3.7 (B)). This is because the residues occurring in the middle of the protein are all hydrophobic, and the part of the channel lined by these residues, consequently, has a predominantly hydrophobic

environment. Furthermore, the pore is constricted towards the N-terminal, leaving less room for the accommodation of water molecules. As seen in a dynamical variation mapping of the pore radius, the narrowest part of the pore occurs in the middle around Val12 and Ile16 (Figure 3.7 (C)). The occurrence of hydrophobic residues along this narrow stretch is likely to impose an energy barrier to the transport of ions, and this region might play a role in controlling the kinetics of ion conduction. It is possible that the channel is in a closed conformation, since a large part of the pore is devoid of water molecules.

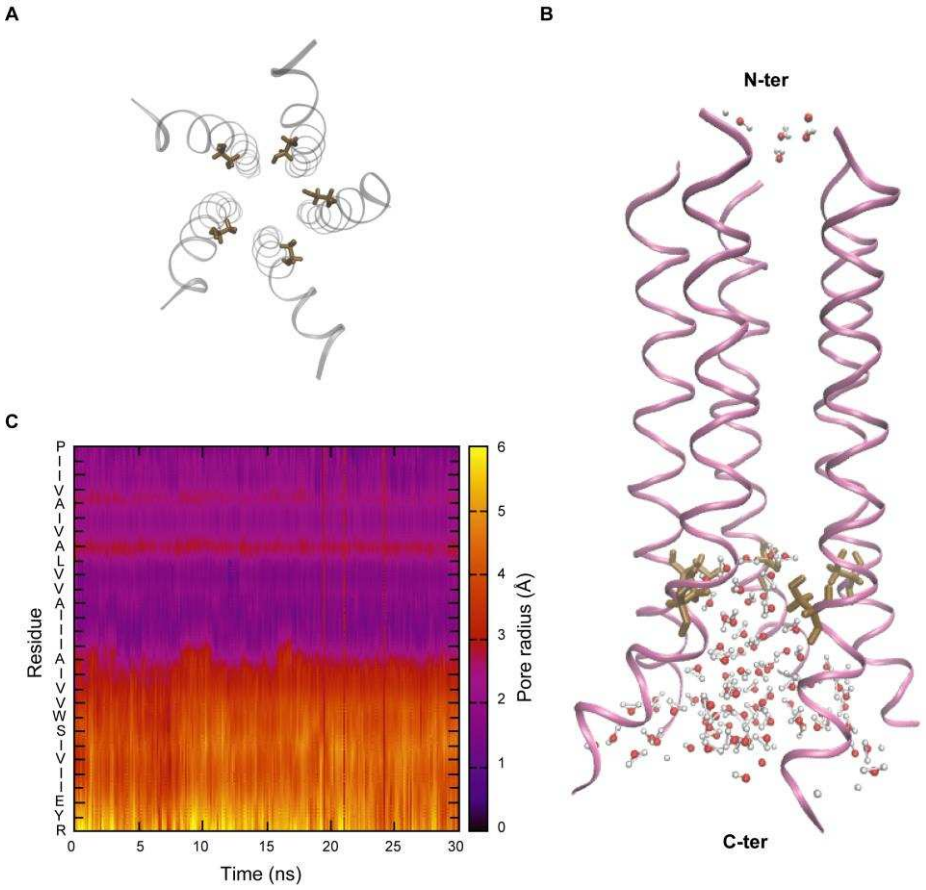

Figure 3.7 (A) View along the pore axis from the C-terminal showing the Ser23 residue in "licorice" representation. Serine faces the interior of the channel in the pentamer model. (B) Side view of the pentamer model showing the location of the Ser23 residue (in "licorice" representation) and water molecules in the pore. The N-terminal side is on the top and the C-terminal is at the bottom. (C) Pore radius across the axis of the pentamer model. The pore is constricted towards the N-terminal side (top half).

3.3.7 The structural model

In the absence of an experimentally characterized structure for the oligomeric Vpu TMD, the proposed models offer useful insights into structural features that govern channel behavior, and

into various intra- and intermolecular interactions that explain why the channel adopts a given structure. The two pentamer models discussed above are consistent with each other, with the RMSD between them differing by less than 3 Å (data not shown). The van der Waals interactions appear to be dominant between the helices, with the interhelical interface being formed by nonpolar residues. The importance of van der Waals forces in mediating helix-helix interactions in the membrane environment has previously been emphasized in solution NMR studies on the transmembrane peptide glycophorin A [61]. Certain structural features of the modeled structures are in remarkable agreement with available experimental data for the channel. Importantly, the kink observed in the helix around Ile17 in our model (Figure 3.8 (A)) is consistent with NMR studies [22]. The initial structure for the simulations had idealized α-helices without any kink, but this might be important for ion channel activity, as suggested by modeling studies on wild-type and mutant Vpu proteins [62]. Another important structural feature we observed was that the three residues known to interact with the tetherin transmembrane domain – Ala14, Ala18, and Trp22 [16] – all lie on the same face of the helix (Figure 3.8 (B)); this is consistent with a pentameric model generated using the structure reported in the above NMR study (PDB ID: 1PI7) [16,22]. It must be noted that the reported structure was obtained from the monomer structure followed by modeling the tetrameric and pentameric states on the basis of rotational symmetry. Such orientation of the residues is a necessary structural requirement, since, if tetherin is to bind to Vpu, it must bind to all of these residues. Moreover, the three residues lie on the exterior side of the channel, and they face the membrane rather than the pore. This has important implications. Firstly, the residues are available for binding to the tetherin transmembrane domain. Secondly, this allows the Trp22 residue to form hydrophobic contacts with lipid tails. Such hydrophobic interactions are important in stabilizing the protein in the hydrophobic lipid bilayer environment.

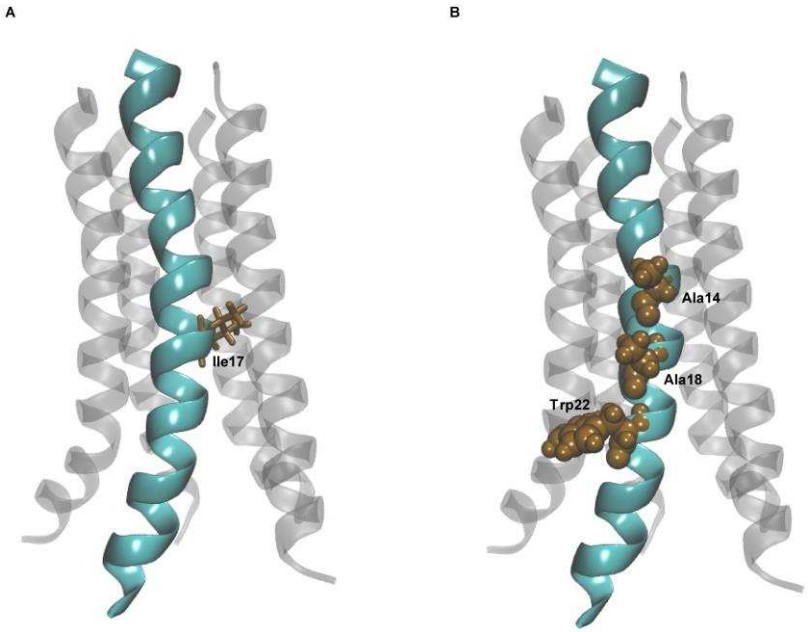

Figure 3.8 Structural features of the pentamer model. (A) Kink around the Ile17 residue in the pentamer model. (B) The three residues known to interact with tetherin shown in van der Waals representation.

The Vpu channel is equally selective toward K^+ and Na^+, and only weakly permeable to Cl^- [5,7]. The highly conserved motif TTVGYGD that is seen in the selectivity filter of K^+-specific channels [63,64] is absent in Vpu. However, because Vpu is able to discriminate cations over anions, it should have some structural motif that is responsible for this selectivity, which remains uncharacterized. Amino acids that differentially interact with ions, and thereby determine the selectivity of a channel, are likely to be charged/polar. The Vpu TMD has five such residues – Ser23, Glu28, Tyr29, Arg30, and Lys31. While Tyr29 faces the lipid headgroup, Arg30 is stabilized via interactions with both Glu28 and the lipid headgroup. Although Lys31 faces the pore, it is unlikely that a positively charged residue will determine the selectivity of a cation-specific channel like Vpu. Moreover, Lys31 lies on the C-terminal side, where the pore is broader, and the density of pore water is high enough to shield any electrostatic effect on a permeating ion. The only other polar residue facing the pore is Ser23, and we hypothesize that this plays an important role in ion permeation. This finds support in conductance studies showing that a mutant with the serine substituted for a leucine does not exhibit any ion-conducting activity [18]. The Vpu protein from HIV-1 subtype O and P viruses, however, has a tryptophan in place of serine at position 23. Given the importance of this serine residue, Vpu from these subtypes is not expected to show ion channel activity. It is interesting to note that assays on Vpu

from these subtypes report poor virus release from cells (Shahid Jameel and coworkers; unpublished results). This further supports a role for the Vpu ion channel activity in promoting viral propagation. The occurrence of serine at position 23, and the consequent channel activity, might offer a selective advantage to M-subtype viruses, thereby accounting for the predominance of this subtype.

It is not clear whether the modeled channel is in an open or closed conformation. Certain features of the model support a closed state. The presence of a constricted region lined by hydrophobic residues is not likely to encourage the passing of an ion. Besides, the largely dehydrated nature of the pore is characteristic of closed states. However, if the modeled channel represents an open conformation, the channel is likely to show very weak ion-conducting activity, owing to the structural features described above. This possibility cannot be ruled out, since conductance studies on Vpu have shown that the channel is indeed weakly conducting [18]. It is not possible to determine whether the modeled channel is in an open state or not unless the ion-conducting activity of the channel is investigated.

3.4 Conclusions

This study comprehensively examines the possibility of the Vpu TMD to exist in different possible oligomeric states in a hydrated lipid bilayer environment. The results suggest that the pentameric form is the most stable state, with the pentameric models possessing the symmetry that is typical of homo-oligomeric channels. The tetrameric and hexameric models, however, lose this symmetry over the course of the simulations. The major force stabilizing the pentameric form over the other forms is van der Waals interactions between adjoining helices. The pentamer is further stabilised via a salt bridge between Glu28 and Arg30, and via interactions between polar residues on the protein and lipid headgroups. These interactions are far weaker in the tetramer and hexamer, and might play a critical role in holding together the helical TMDs to the membrane. The structural features of the pentamer models from this study are able to account for much of the activity of Vpu observed in previous experimental studies. While the residues that bind to the tetherin transmembrane domain face the exterior (and are hence accessible), Ser23, which has been previously shown to be crucial in ion transport, faces the pore. The Vpu protein from O- and P-subtypes, however, is known to lack this serine at position 23, and is thus likely to

have reduced ion channel activity. We hypothesize that the predominance of M-subtype viruses might be facilitated by the ability of M-subtype Vpu to conduct ions. Ion channel activity, therefore, could possibly have a role in enhancing the replication fitness of the virus.

References

1. Greene WC, Peterlin BM (2002) Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. Nat Med 8(7):673–680.
2. Malim MH, Emerman M (2008) HIV-1 accessory proteins—Ensuring viral survival in a hostile environment. Cell Host Microbe 3(6):388–398.
3. Strebel K, Klimkait T, Martin M (1988) A novel gene of HIV-1, vpu, and its 16-kilodalton product. Science 241(4870):1221–1223.
4. Cohen EA, Terwilliger EF, Sodroski JG, Haseltine WA (1988) Identification of a protein encoded by the vpu gene of HIV-1. Nature 334(6182):532–534.
5. Schubert U, et al. (1996) Identification of an ion channel activity of the Vpu transmembrane domain and its involvement in the regulation of virus release from HIV-1-infected cells. FEBS Lett 398(1):12–18.
6. Ewart GD, Mills K, Cox GB, Gage PW (2002) Amiloride derivatives block ion channel activity and enhancement of virus-like particle budding caused by HIV-1 protein Vpu. Eur Biophys J 31(1):26–35.
7. Ewart GD, et al. (2004) Potential new anti-human immunodeficiency virus type 1 compounds depress virus replication in cultured human macrophages. Antimicrob Agents Chemother 48(6):2325–2330.
8. Maldarelli F, Chen MY, Willey RL, Strebel K (1993) Human immunodeficiency virus type 1 Vpu protein is an oligomeric type I integral membrane protein. J Virol 67(8):5056-5061.
9. Wray V, et al. (1995) Solution structure of the hydrophilic region of HIV-1 encoded virus protein U (Vpu) by CD and 1H NMR spectroscopy. Int J Pept Protein Res 45(1):35-43.
10. Federau T, et al. (1996) Solution structure of the cytoplasmic domain of the human immunodeficiency virus type 1 encoded virus protein U (Vpu). Int J Pept Protein Res 47(4):297-310.
11. Willey RL, Maldarelli F, Martin MA, Strebel K (1992) Human immunodeficiency virus

type 1 Vpu protein induces rapid degradation of CD4. J Virol 66(12):7193-7200.

12. Schubert U, et al. (1996) The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains. J Virol 70(2):809-819.
13. Schubert U, et al. (1992) Human-immunodeficiency-virus-type-1-encoded Vpu protein is phosphorylated by casein kinase II. Eur J Biochem 204(2):875-883.
14. Schubert U, Strebel K (1994) Differential activities of the human immunodeficiency virus type 1-encoded Vpu protein are regulated by phosphorylation and occur in different cellular compartments. J Virol 68(4):2260-2271.
15. Neil SJ, Zang T, Bieniasz PD (2008) Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature 451(7177):425-430.
16. Vigan R, Neil SJ (2010) Determinants of tetherin antagonism in the transmembrane domain of the human immunodeficiency virus type 1 Vpu protein. J Virol 84(24):12958-12970.
17. Ewart GD, Sutherland T, Gage PW, Cox GB (1996) The Vpu protein of human immunodeficiency virus type 1 forms cation-selective ion channels. J Virol 70(10):7108-7115.
18. Mehnert T, et al. (2008) Biophysical characterization of Vpu from HIV-1 suggests a channel-pore dualism. Proteins 70(4):1488-1497.
19. Hussain A, Das SR, Tanwar C, Jameel S (2007) Oligomerization of the human immunodeficiency virus type 1 (HIV-1) Vpu protein–a genetic, biochemical and biophysical analysis. Virol J 4(1):81.
20. Lu JX, Sharpe S, Ghirlando R, Yau WM, Tycko R (2010) Oligomerization state and supramolecular structure of the HIV-1 Vpu protein transmembrane segment in phospholipid bilayers. Protein Sci 19(10):1877-1896.
21. Ma C, et al. (2002) Expression, purification, and activities of full-length and truncated versions of the integral membrane protein Vpu from HIV-1. Protein Sci 11(3):546-557.
22. Park SH, et al. (2003) Three-dimensional structure of the channel-forming trans-membrane domain of virus protein "u" (Vpu) from HIV-1. J Mol Biol 333(2):409-424.
23. Kukol A, Arkin IT (1999) Vpu transmembrane peptide structure obtained by site-specific fourier transform infrared dichroism and global molecular dynamics searching. Biophys J 77(3):1594-1601.

24. Grice AL, Kerr ID, Sansom MSP (1997) Ion channels formed by HIV-1 Vpu: a modelling and simulation study. FEBS Lett 405(3):299-304.
25. Moore PB, Zhong Q, Husslein T, Klein ML (1998) Simulation of the HIV-1 Vpu transmembrane domain as a pentameric bundle. FEBS Lett 431(2):143-148.
26. Lopez CF, Montal M, Blasie JK, Klein ML, Moore PB (2002) Molecular dynamics investigation of membrane-bound bundles of the channel-forming transmembrane domain of viral protein U from the human immunodeficiency virus HIV-1. Biophys J 83(3):1259-1267.
27. Cordes FS, et al. (2001) The structure of the HIV-1 Vpu ion channel: modelling and simulation studies. Biochim Biophys Acta 1512(2):29-298.
28. Lemaitre V, Ali R, Kim CG, Watts A, Fischer WB (2004) Interaction of amiloride and one of its derivatives with Vpu from HIV-1: a molecular dynamics simulation. FEBS Lett 563(1-3):75-81.
29. Kim CG, Lemaitre V, Watts A, Fischer WB (2006) Drug-protein interaction with Vpu from HIV-1: proposing binding sites for amiloride and one of its derivatives. Anal Bioanal Chem 386(7-8):2213-2217.
30. Ulmschneider JP, Ulmschneider MB (2007) Folding simulations of the transmembrane helix of virus protein U in an implicit membrane model. J Chem Theory Comput 3(6):2335-2346.
31. Patargias G, Martay H, Fischer WB (2009) Reconstructing potentials of mean force from short steered molecular dynamics simulations of Vpu from HIV-1. J Biomol Struct Dyn 27(1):1-12.
32. Krüger J, Fischer WB (2009) Assembly of viral membrane proteins. J Chem Theory Comput 5(9):2503-2513.
33. Im W, Feig M, Brooks III CL (2003) An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. Biophys J 85(5):2900-2918.
34. Bu L, Im W, Brooks III CL (2007) Membrane assembly of simple helix homo-oligomers studied via molecular dynamics simulations. Biophys J 92(3):854-863.
35. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314(1):141-151.

36. Im W, Lee MS, Brooks III CL (2003) Generalized Born model with a simple smoothing function. J Comput Chem 24(14):1691-1702.
37. Wray V, et al. (1999) Solution structure and orientation of the transmembrane anchor domain of the HIV-1-encoded virus protein U by high-resolution and solid-state NMR spectroscopy. Biochemistry 38(16):5272-5282.
38. Marassi FM, et al. (1999) Correlation of the structural and functional domains in the membrane protein Vpu from HIV-1. Proc Natl Acad Sci USA 96(25):14336-14341.
39. Park SH, De Angelis AA, Nevzorov AA, Wu CH, Opella SJ (2006) Three-dimensional structure of the transmembrane domain of Vpu from HIV-1 in aligned phospholipid bicelles. Biophys J 91(8):3032-3042.
40. Fischer WB, Forrest LR, Smith GR, Sansom MSP (2000) Transmembrane domains of viral ion channel proteins: a molecular dynamics simulation study. Biopolymers 53(7):529-538.
41. Brooks BR, et al. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Phys 4(2):187-217.
42. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. J Comput Chem 30(10):1545–1614.
43. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616.
44. MacKerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25(11):1400–1415.
45. Schlitter J (1993) Estimation of absolute entropies of macromolecules using the covariance matrix. Chem Phys Lett 215(6):617-621.
46. Andricioaei I, Karplus M (2001) On the calculation of entropy from covariance matrices of the atomic fluctuations. J Chem Phys 115(14):6289-6292.
47. Nosé, S (1984) A unified formulation of the constant temperature molecular dynamics methods. J Chem Phys 81(1):511-519.
48. Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A 31(3):1695-1697.

49. Jo S, Kim T, Im W (2007) Automated builder and database of protein/membrane complexes for molecular dynamics simulations. PLoS ONE 2(9):e880.
50. Jo S, Lim JB, Klauda JB, Im W (2009) CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. Biophys J 97(1):50-58.
51. Klauda JB, et al. (2010) Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. J Phys Chem B 114(23):7830–7843.
52. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926-935.
53. Essmann U, et al. (1995) A smooth particle mesh Ewald method. J Chem Phys 103(19):8577.
54. Ryckaert J-P, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341.
55. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graphics 14(1):33–38.
56. Smart OS, Neduvelil JG, Wang X, Wallace B, Sansom MS (1996) HOLE: A program for the analysis of the pore dimensions of ion channel structural models. J Mol Graphics 14(6):354–360.
57. Zhang Y, Lewis RNAH, McElhaney RN, Ryan RO (1993) Calorimetric and spectroscopic studies of the interaction of Manduca sexta apolipophorin III with zwitterionic, anionic, and nonionic lipids. Biochemistry 32(15):3942-3952.
58. Liu F, Lewis RNAH, Hodges RS, McElhaney, RN (2004) Effect of variations in the structure of a polyleucine-based alpha-helical transmembrane peptide on its interaction with phosphatidylethanolamine bilayers. Biophys J 87(4):2470-2482.
59. Saiz L, Bandyopadhyay S, Klein ML (2004) Effect of the pore region of a transmembrane ion channel on the physical properties of a simple membrane. J Phys Chem B 108(8):2608–2613.
60. Deol SS, Bond PJ, Domene C, Sansom MSP (2004) Lipid-protein interactions of integral membrane proteins: a comparative simulation study. Biophys J 87(6):3737-3749.
61. MacKenzie KR, Prestegard JH, Engelman DM (1997) A transmembrane helix dimer: structure and implications. Science 276(5309):131-133.

62. Krüger J, Fischer WB (2010) Structural implications of mutations assessed by molecular dynamics: Vpu_{1-32} from HIV-1. Eur Biophys J 39(7):1069-1077.
63. Heginbotham L, Abramson T, MacKinnon R (1992) A functional connection between the pores of distantly related ion channels as revealed by mutant K^+ channels. Science 258(5085):1152-1155.
64. Heginbotham L, Lu Z, Abramson T, MacKinnon R (1994) Mutations in the K^+ channel signature sequence. Biophys J 66(4):1061-1067.

4 Ion Conduction through Vpu

4.1 Introduction

The viral protein U (Vpu) is one of the four accessory proteins encoded by HIV-1 that has an N-terminal transmembrane (TM) domain and a C-terminal cytoplasmic domain [1-10]. A schematic representation of the protein is shown in Figure 4.1 (A). The cytoplasmic domain has two α-helices [11,12], and is involved in the degradation of CD4 molecules in the host cell [13,14]. The TM domain, which is known to have a helical topology [15-20], has been suggested to enhance virus release via two mechanisms: formation of ion-conducting channels [21,22] and degradation of the antiviral protein tetherin [23]. The ion channel is formed via oligomerization of monomeric Vpu to form a pentamer [24-26,16]. Because of its importance in virus release, an understanding of the mechanism of ion permeation is of great importance. Vpu has been reported to show ion channel activity in lipid bilayers, Xenopus oocytes, and in the plasma membrane of Escherichia coli [21,22]. Vpu channel has been shown to be selective towards monocations (Na^+ and K^+), and cannot differentiate between the two [21,22].

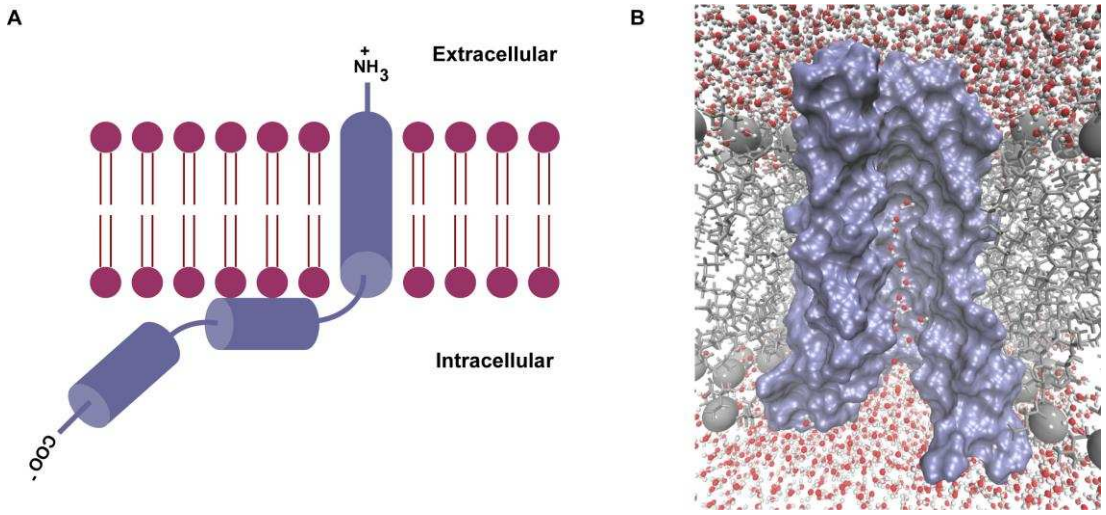

Figure 4.1 (A) Schematic illustration of the monomer. (B) The pentameric channel set up in a hydrated lipid bilayer. The monomeric unit on the front has been omitted to reveal the interior of the pore. Pore water molecules can be seen in the pore lumen.

Channel recordings on full length Vpu (Vpu_{1-81}) and the Vpu TM region (Vpu_{1-32}) excluding the cytoplasmic domain showed that the conductance states and ion-conduction recordings are

similar for the two [27]. The channel has been suggested to exist in a number of conductance states, which are believed to correspond to different open state conformations. Furthermore, the duration for which the channel remains open is independent of the voltage, and the channel shows biphasic voltage activation. The variation of conductivity with respect to salt concentration has been measured by Mehnert et al., which showed that ion conduction follows Michaelis-Menten behavior [28]. The ion transport activity is rather weak, though the channel shows pore-like characteristics. A mutant Vpu with Ser23 → Leu mutation in the TM domain does not conduct ions, suggesting a critical role for serine in ion transport [28]. Two possible roles have been ascribed to the serine residue. Firstly, it might act as a weak ion binding site, and secondly, it creates a hydrophilic environment in the vicinity, thereby reducing the length of the hydrophobic stretch that occurs inside the channel, thus making it easier for an ion to pass through. The passage of ions through the pore has been previously modeled using steered molecular dynamics (SMD) simulations [29]. The possibility of the existence of a number of conformations of oligomeric Vpu has recently been suggested by computational studies that have modeled the assembly of monomeric Vpu into oligomers [30]. This supports experimental channel recordings that suggested the existence of a number of conductance states of Vpu [27].

Molecular dynamics (MD) simulations have been successfully used to study transmembrane proteins in general [31-41]. A powerful approach is the use of umbrella sampling MD simulations, which are in general very helpful in understanding free energy changes corresponding to rare events [42-45]. This method has been used by Roux and coworkers for elucidating the energetics of ion permeation through channels [46-49]. This method also has been used to obtain a multi-ion free energy profile for the KcsA K^+ channel [46]. Results from that study showed that an ion trapped in the selectivity filter is able to proceed further into the channel because of electrostatic repulsion from ions that subsequently approach the selectivity filter. Based on this, it was proposed that rapid conduction in the channel is driven by interionic repulsion. The method has subsequently been used for calculating the free energy profile of ion transport through the gramicidin channel as a function of both axial and radial positions [47], and for elucidating how the KcsA channel preferentially allows K^+ rather than Na^+ ions to pass through [49].

The present study employs umbrella sampling MD simulations to elucidate the mechanism of ion conduction through the Vpu channel. We have recently reported a molecular model for the pentameric state of the TM domain of Vpu protein, which has been used here [24]. Available experimental information on Vpu was used to validate the proposed structure. Since only the TM domain is involved in channel activity [22] and the conductance properties of channels formed by the Vpu TM region are almost the same as those formed by full length Vpu [27], the cytoplasmic domain of Vpu has not been included in the model. This study explains the molecular and energetic basis of ion selectivity in the channel, and suggests that a hydrophobic stretch in the channel might control the kinetics of the ion permeation process. Finally, a mechanism for the transport of ions is proposed based on a multi-ion free energy landscape.

4.2 Methods

The Vpu channel structure used in the study is a pentameric form of the TM region modeled in a recent study [24]. In addition to the protein, the system includes 95 lipid (POPC) molecules, 5496 water molecules, 13 K^+ ions and 18 Cl^- ions, with a total of 31984 atoms. Prior to the current study, the system was equilibrated for 30 ns without any restraints and was found to be adequately equilibrated, as evidenced from the convergence of RMSD, the occurrence of several stabilizing intra-protein and protein-lipid interactions, and the retaining of the structural integrity of the channel, in general.

The transport of two different ions through the channel was studied, namely Na^+ and K^+. For modeling the permeation of a K^+ ion, the coordinates of a K^+ ion from bulk water were first swapped with a water molecule at the C-terminus of the channel. Conformations with the K^+ ion occurring at different positions along the axis of the pore were generated by pulling the ion along the pore axis using steered molecular dynamics (SMD) simulations. The ion was pulled at a constant velocity of 0.01 Å ps^{-1} by applying an external force on a dummy atom connected to the permeating ion via a spring with a harmonic constant of 10 kcal mol^{-1} $Å^{-2}$. The structural model for the channel was aligned along the z-axis [24], and hence the pulling force was applied along the z-axis, thereby making the ion move from the C-terminus to the N-terminus. Positional restraints with force constant 1 kcal mol^{-1} $Å^{-2}$ were applied on the heavy atoms of the protein to prevent the drifting of the helices during the pulling of the ion. Constant temperature and

pressure were employed by using Langevin dynamics and a Langevin pressure piston, respectively. It must be noted that the SMD simulations were used to obtain the initial structures for the umbrella sampling simulations and not for investigating the permeation mechanism. The NAMD program [50,51] was used for the SMD simulations with the CHARMM22 all-atom protein force field with CMAP corrections [52,53], the CHARMM36 all-atom lipid force field [54], optimized parameters for ions [55], and the TIP3P water model [56]. For the studies on the permeation of Na^+, the conformations were generated by replacing the K^+ ion in the pore with an Na^+ ion.

Umbrella sampling was performed by applying a biasing potential on the permeating ion, with the potential having a parabolic form. Independent simulations were performed for different positions of the ion along the pore axis, so that each simulation sampled the conformations of the system with the ion at the respective region. Conformations with the ion placed at different places along the pore axis were extracted from the SMD simulation, and were used as initial structures for the different windows in the umbrella sampling calculations. The ion position ranged from z = -30 Å to z = 36 Å, giving a total of 133 windows along the z-axis spaced 0.5 Å apart (the model structure used in the study had previously been aligned along the z-axis). A distribution of ion positions around the center of a given window was obtained by applying a biasing potential with a force constant of 20 kcal mol^{-1} $Å^{-2}$ on the ion using the MMFP module in the CHARMM program [60]. Positional restraints with force constant 1 $kcal^{-1}$ mol^{-1} $Å^{-2}$ were applied on the backbone C_α atoms of the protein to prevent the drifting of the channel. Ions other than the permeating ion were excluded from the pore region by using a repulsive sphere of radius 15 Å, with the repulsive force acting on ions only when they entered this sphere. Covalent bonds involving hydrogen were constrained using SHAKE [61]. Simulations were carried out in the NPT ensemble using the CHARMM program [60] with the same force field parameters as those used in the pulling simulations mentioned above. Simulations for each window involved 100 ps of equilibration followed by 1 ns of production, with a simulation time of (133 * 1.1) ns for each free energy profile. For calculating the potential of mean force (PMF) for the transport of the ion across the channel, the umbrella sampling trajectories were unbiased using WHAM [57-59].

The permeation of two K^+ ions through the channel was modeled by considering different positions of one ion relative to the second ion, with the spacing between the two ions varying from 4 Å to 10 Å (additional windows with 11 Å and 12 Å spacing between the ions were sampled for the middle region of the channel). A total of 395 windows spaced 1 Å apart were sampled with the two ions at different positions along the pore axis. A biasing harmonic potential of force constant 10 kcal mol^{-1} $Å^{-2}$ was applied on the two ions. Parameters for the other restraints used were the same as those used for the calculations with a single ion in the pore. Equilibration and production runs for each window were carried out for 100 ps and 200 ps, respectively. The total simulation time combining all the single ion and two-ion simulations was more than 0.4 μs, with a total of 661 independent simulations. All molecular images were rendered using the visual molecular dynamics (VMD) program [62]. An electrostatic surface representation of the channel was generated using the PDB2PQR server [63].

4.3 Results and Discussion

4.3.1 Atomistic structure of TM domain of Vpu, and features of the lumen of the pore

A representation of the channel modeled in a hydrated lipid bilayer is shown in Figure 4.1 (B). A few pore water molecules can be seen, especially in the bottom half of the pore, towards the C-terminal opening. The channel consists of five helical TM domains held together predominantly by van der Waals forces, with the interface between the helices being formed by nonpolar residues [24]. The channel is wider at the C-terminal end than at the N-terminal end, leading to a relatively greater degree of hydration at this end. The residues Arg30 and Tyr29 at the C-terminal end of the TM domain face the exterior side of the channel, and they stabilize the protein in the lipid bilayer by forming hydrogen bonds with phospholipid headgroups [24]. Although Lys31 is directed towards the interior of the channel, it lies in a region that is sufficiently solvated, so that the residue is shielded from exerting any influence on any ion passing through the channel (details of the structural model are presented in Chapter 3). A serine at position 23 faces the lumen of the pore, and it marks the end of the small hydrophilic space at the C-terminal entrance of the channel. Figure 4.2 (A) shows the serine residue facing the interior of the pentameric channel. The pore is constricted in the middle, which is a largely dehydrated region, owing to the occurrence of hydrophobic residues. The kinetics of ion

conduction through the pore is likely to be controlled by this long hydrophobic stretch. Initially, the nature of the pore was investigated by examining the electrostatic representation of the pore-lining residues in the channel shown in Figure 4.2 (B). In the surface representation, one monomer of Vpu has been omitted in the figure for clarity. The figure shows electronegative regions in red, and electropositive regions in blue. It can be seen that the top and middle region of the channel is constricted and lined entirely by nonpolar residues. Further down the channel, towards the C-terminal end, the pore widens, and the occurrence of a serine residue gives rise to a slightly hydrophilic region. The widening of the pore was shown to be due to the kink around the Ile17 residue (see Figure 3.8 (A)) [24], which is in agreement with experimental data [19]. The serine residue can be seen as a red region at the bottom of the channel. The occurrence of such a hydrophilic region is expected to stabilize the permeating ion, thereby assisting ion transport. It follows that the length of the hydrophobic stretch would have been greater in the absence of this serine residue. The serine therefore reduces the effective length of the hydrophobic stretch, and notably, it has been reported previously that the Ser23 → Leu mutant of Vpu is inactive towards ion permeation [28].

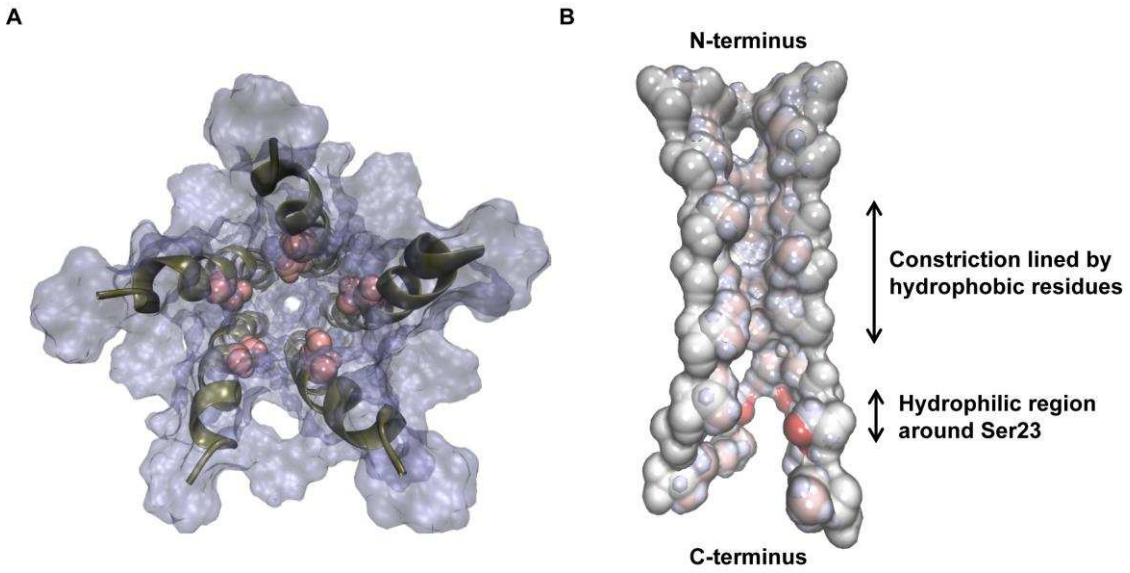

Figure 4.2 (A) The pentamer model used in the study shown with the Ser23 residue in van der Waals representation. (B) A representation of the electrostatic surface of pore-lining residues in the channel. The monomer unit on the front has been omitted for clarity. Red indicates negatively charged regions, blue indicates positively charged regions, and white indicates nonpolar regions. Only the sidechain atoms are shown in color, and the backbone atoms are in white to allow a clear illustration of the nature of pore-lining residues.

It is not practical to model events that occur at time scales longer than the ones that are practical with MD simulations. Additionally, insufficient sampling issues preclude calculation of free energies associated with such rare events. Use of umbrella sampling, which applies a harmonic potential along the predetermined reaction coordinate, allows sampling of intermediate states in the process of ion permeation with relatively less difficulty and with significantly less computing time, while at the same time allowing the calculation of the free energy change along the reaction coordinate using methods like WHAM [57-59]. An important step in employing umbrella sampling is the selection of the reaction coordinate. The axis of the pore has proved to be a valuable reaction coordinate in the study of ion conduction through channels [46-49,64,65]. Using the pore axis as the reaction coordinate, the present study illustrates the energetics and atomistic mechanism of the ion channel activity of Vpu.

4.3.2 Free energy profiles indicate weak ion conduction

The PMFs for the transport of Na^+ and K^+ ions are shown in Figure 4.3. The PMF along the pore axis is shown, with the C-terminus (or the intracellular side) on the left and the N-terminus (extracellular side) on the right. The residues lining the pore are also marked appropriately in the plots. The free energy profiles presented correspond to the ion permeating from the bulk water region outside the C-terminal end of the channel to the bulk water region on the other end of the channel. For determining the errors, the sampling data of last 1 ns/window was divided into five parts with 200 ps/window, and the PMF was computed separately for each part. The mean, standard deviation, and standard error were then calculated from the PMF obtained for the different intervals. The energy barriers for this process are about 43.9 kcal/mol for Na^+ and 43.3 kcal/mol for K^+. Such a high energy barrier is to be expected, given the highly hydrophobic environment that prevails inside the channel. The energy barrier observed is considerably higher than that for other known ion channels [46-49,64,65], and it suggests that the ion channel activity of Vpu is weaker than that of typical ion channels. This is in good agreement with conductance experiments suggesting that Vpu is expected to behave as a weakly conducting ion channel [28]. The high energy barrier encountered by the permeating ion is likely to make the process kinetically slow.

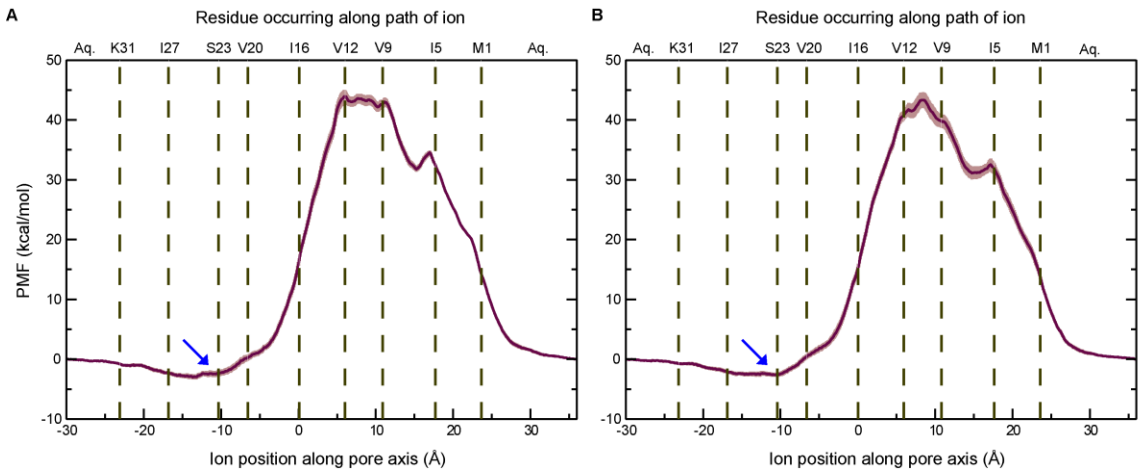

Figure 4.3 The potential of mean force (PMF) for the transport of permeating ions along the channel axis. (A) Na^+; (B) K^+. Residues facing the pore have been shown at their appropriate positions. The coordinates of these residues were calculated by determining the center of mass of their side chains. Error bars are shown as bands. The soft minimum around Ser23 is marked with an arrow.

The features of the two free energy profiles (Na^+ and K^+) were found to be very similar to each other. The movement of the ion into the channel from the C-terminal side is accompanied by the release of 2 kcal/mol of energy, with a shallow minimum occurring around z = -10 Å (the region is marked with an arrow in Figure 4.3). This corresponds to the position of the ring of Ser23 residues, and suggests the role of serine as a weak binding site. This is consistent with previous conductance studies on mutant forms of Vpu, which have proposed that the serine can act as a weak binding site for ions, since replacing the serine with a hydrophobic residue results in loss of ion channel activity. The fact that the minimum seen around the serine is shallow rather than deep indicates that the binding site is at best only a weak one. As the ion moves further into the channel, there is a surge in the free energies brought about by the unfavorable interactions that the positive ions have with the hydrophobic environment of the channel. The maximum is reached around two valine residues, namely Val9 and Val12, which occur in the middle of the hydrophobic stretch in the channel. The ions experience a soft minimum around 15 Å, which arises from the stabilization of the ion at this position by coordinating water molecules (see later). Figure 4.4 shows snapshots of the Na^+ ion at two places inside the channel. The first snapshot shows the ion near the ring of Ser23 residues, while the second snapshot shows the ion near the ring of Val12 residues. As can be seen in the figure, the latter corresponds to the narrow stretch of the channel, where the energy barrier reaches a maximum.

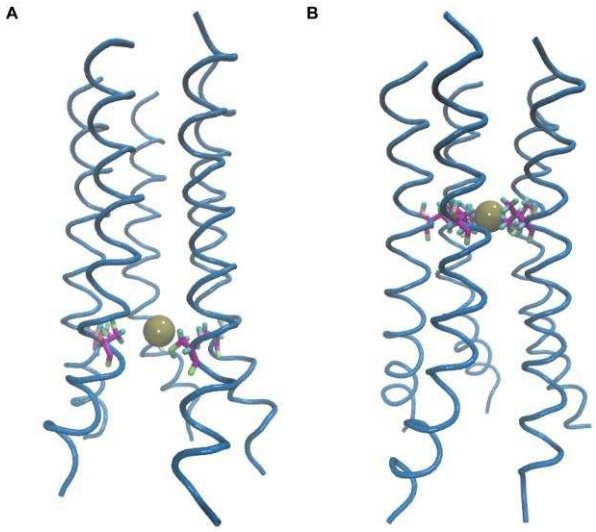

Figure 4.4 Conformations with the permeating ion inside the channel. Snapshots of the Na^+ ion near the ring of (A) Ser23 residues and (B) Val12 residues, respectively.

The remarkably similar free energy profiles corresponding to the permeation of Na^+ and K^+ explain why the channel does not discriminate Na^+ over K^+, as reported previously based on experimental studies [22]. Given that the windows are close to each other, 1.1 ns long MD simulations for each of the 133 windows per free energy profile is deemed adequate. The total simulation time for each of the two ions was therefore (1.1 ns * 133) ≈ 146 ns. The PMFs obtained at different durations of the simulations and for the whole duration are given in Figure 4.5. The PMFs calculated agree well with each other with only minor variations in the free energy values. Based on this, modeling of the multi-ion permeation was also performed using a similar protocol (see below).

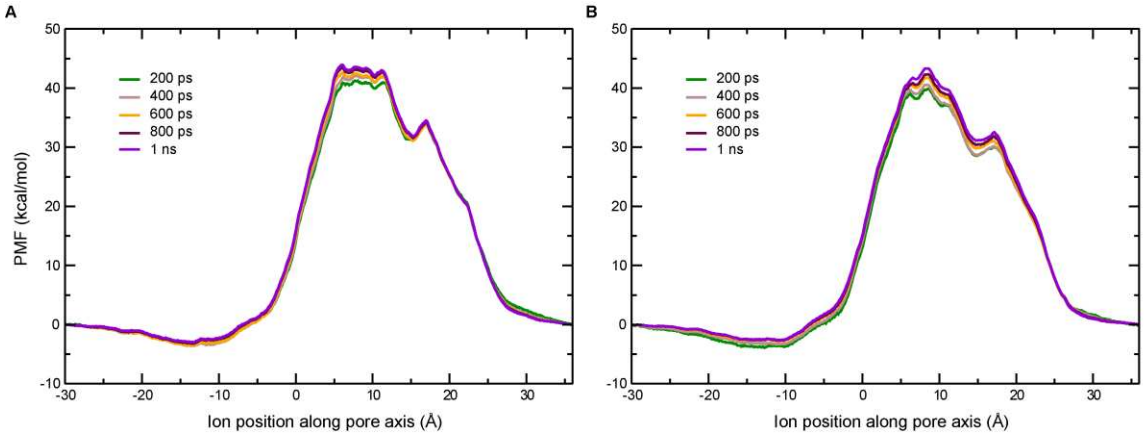

Figure 4.5 The PMF for ion transport for different sampling time periods. (A) Na^+ (B) K^+

4.3.3 Extent of solvation governs ion transport kinetics

Solvation of the ion during the permeation process and its relationship with the free energy profiles were examined based on the hydration number, the number of water molecules in the first solvation shell of the ion and the interaction energies (see below). Figure 4.6 shows the average hydration number of the ions with respect to the position along the channel. The hydration numbers were calculated based on cutoff values of the distance between the ion and the oxygen atom of water molecules (2.8 Å for Na^+ and 3.2 Å for K^+). These cutoff values were chosen accounting for the size of the respective ions [49]. As the ion enters the C-terminal side of the channel, it retains its hydration shell, and proceeds spontaneously into the channel. The hydrophilic environment in this region is created by the polar Ser23 sidechain (this corresponds roughly to the position -10 Å in the figure). As the ion proceeds towards the N-terminal side, it gradually loses waters of hydration, with the hydration number reaching a minimum in the middle region of the channel. The loss in hydration number leads to a destabilization of the ion in the middle region relative to the C-terminal region, which is reflected in the rise in PMF. This is consistent with the earlier discussion on the nature of the pore in this region that this region is rich in hydrophobic residues, which results in stripping of water molecules from the ion and drying of the whole region in general. It must be noted that there are relatively more number of water molecules around the ion when the ion is in the region between Ile5 and Val9 (corresponding to z = 15 Å). This is explained by a slight widening of the pore around this region [24], making it possible to accommodate more water molecules. The stabilization of the ion by this solvation explains the occurrence of the soft minimum seen in the PMF at this position (Figure 4.3). As the ion proceeds towards bulk water beyond the N-terminal end of the pore, the hydration number of the ion increases, and is associated with a sharp fall in the PMF.

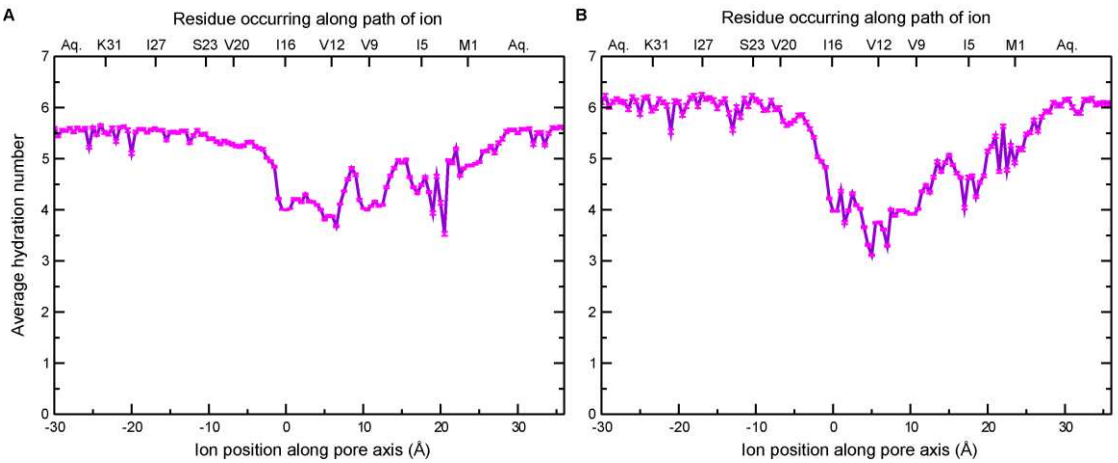

Figure 4.6 Average hydration number for permeating ions along the channel axis. (A) Na^+; (B) K^+. The hydration number was calculated by determining the number of water molecules within a certain cutoff distance of the permeating ion. The cutoff used was 2.8 Å for Na^+ and 3.2 Å for K^+. The hydration number values shown here have been averaged over the trajectory for the respective window. Error bars are shown in pink.

The permeating ion encounters two principal interacting partners inside the channel – the pore water molecules and the channel itself, and such interactions are expected to affect the free energy profiles significantly. Accordingly, interaction energies, which include the contributions of the electrostatic and Lennard-Jones (LJ) terms of the potential energy function, were calculated to examine the solvation energy. Mean values of the interaction energies between ion and water, and ion and protein are shown in Figure 4.7. The major contribution comes from the solvating water molecules, whose contribution towards ion stability is several orders of magnitude higher than that offered by the protein. It is to be noted that such interaction energy calculations have limited applicability, since these are calculated based on the potential energy equation only, and do not include entropy effects. The loss of solvation in the middle region, which is seen in the figure as an increase in the solvation energy, also explains the rise in PMF in this region. The energy barrier that is seen in the free energy profile therefore arises from the desolvation of the ion. This energy barrier, in turn, is likely to make the process of ion conduction kinetically slow, making Vpu a weakly conducting ion channel.

Previous biophysical studies have suggested that Ser23 might act as a weak binding site for ions [28]. The interaction of the ion with the protein shown in Figure 4.7 reveals a shallow minimum for both Na^+ and K^+ near the serine residue, suggesting weak binding ability. On the other hand,

there is a remarkable enhancement of stability due to solvation as the ion moves towards the serine; this is seen as a fall in the solvation energy of the permeating ion. The serine residue therefore enhances ion movement into the channel by giving rise to a sufficiently solvated region on the C-terminal side. There are thus two ways in which the Ser23 makes conditions conducive for ion transport: firstly, it acts as a weak binding site, and, secondly, it gives rise to a hydrophilic region around the entrance of the pore.

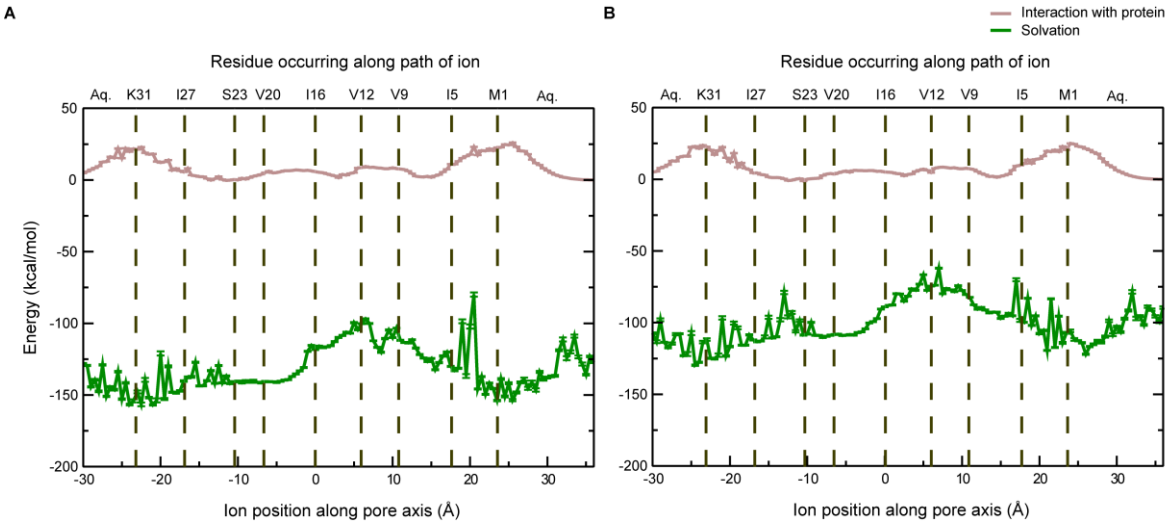

Figure 4.7 The average ion-protein interaction energy and average solvation energy for different positions of the permeating ion. (A) Na^+; (B) K^+. Values for each window have been averaged over the trajectory for the respective window, and are shown with error bars. The pore-lining residues are also shown at their respective position in the plot.

4.3.4 Absence of concerted ion transport explains the slow kinetics

To investigate if more than one ion can pass through the channel at a time, the free energy profile was calculated as a function of the positions of two permeating K^+ ions. The free energy landscape with respect to the position along the z-axis of the first ion and the distance of the second ion from the first is shown in Figure 4.8. The probability distributions obtained from the biased simulations were checked for adequate sampling along both the reaction coordinates prior to construction of the free energy surface. The lowest energy pathway with respect to the position of the first ion is also marked in the plot. The structures presented below the plot show the conformations of the channel corresponding to this path at three different positions. The free energy surface as a whole shows that permeation of two ions at the same time is not thermodynamically favorable compared to the movement of one ion. If one follows the minimum

energy path given in the figure, one can see that when the two ions are near the C-terminal side of the pore (bottom left corner of the figure), the ions are close in space. This is because this region is reasonably well solvated, and interactions between the two ions are shielded. As the two ions cross the hydrophilic region around Ser23 and enter the dehydrated middle region of the channel, interionic repulsion becomes significant. The reduced number of water molecules hydrating the ions, and the dry nature of this part of the pore leads to longer distances between the two ions. This means that two ions cannot enter the middle region of the channel simultaneously, and only one ion can pass at a time. Once the first ion reaches the N-terminal region, the second ion starts crossing the hydrophobic stretch in the pore. Thus a mechanism is hypothesized in which only one ion crosses the hydrophobic region of the channel at a given time while the next ion waits near the C-terminal region, until the first one reaches the N-terminal end of the pore. The kinetics of the process of ion transport is therefore seen to be controlled primarily by the hydrophobic stretch. Such a hypothesis is also consistent with previous experimental studies. Fischer and coworkers have proposed that the conductivity of Vpu follows Michaelis-Menten behavior upon increasing the salt concentration [28]. Such behavior is explained by the fact that a single ion can pass through the channel at a time, as suggested by the mechanism discussed here.

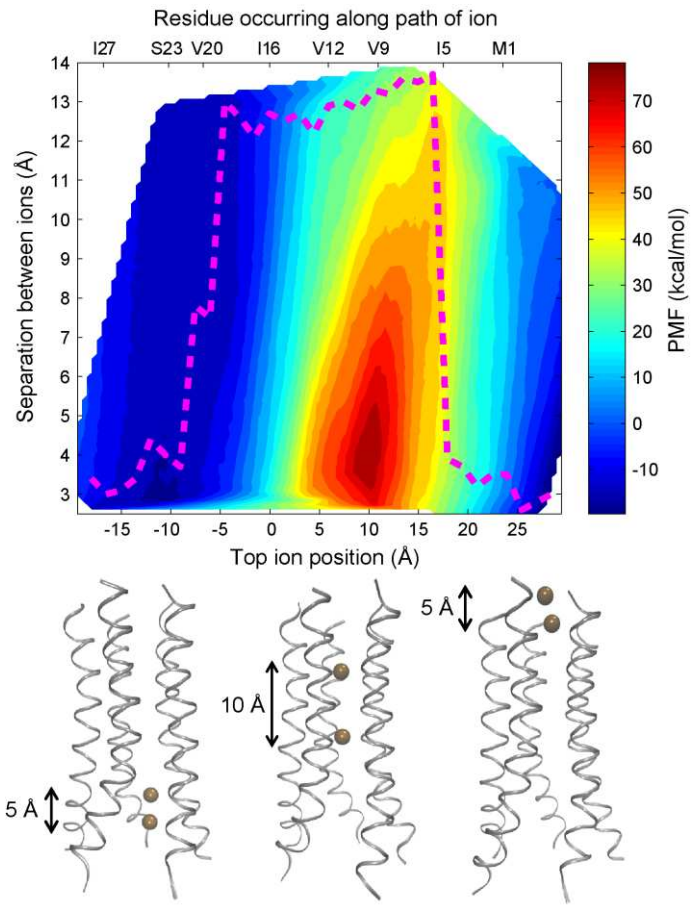

Figure 4.8 The PMF shown as a function of the positions of the two permeating ions. The x-axis labels at the bottom show the position of the top ion along the channel axis, while the labels at the top show the pore-lining residues at their respective positions along the channel axis. The pathway with the lowest free energy barrier is highlighted in maroon color. Images of ion positions corresponding to the pathway with the lowest free energy barrier are shown below the plot, together with the distance between the two ions at these positions.

It may be noted that, although the windows for the two ion permeation were chosen such that the spacing between the two ions varied from 4 Å to 12 Å, the conformations sampled during the simulations cover a range of 3 Å to 14 Å. While conformations with the spacing close to 3 Å are sampled in the terminal regions of the channel, conformations with the spacing increasing to 14 Å are seen in the middle region of the channel. This suggests a propensity of the two ions to stay close together while they are near the hydrated terminal regions of the channel. On the other hand, in the middle region of the pore, the ions tend to stay as far apart as possible, thereby minimizing interionic repulsion.

4.3.5 Perspectives

The high energy barrier reported here for the transport of K^+ and Na^+ through the Vpu channel can be attributed to the long stretch of hydrophobic residues which constrict the pore, and cause the dehydration of the pore lumen (Figure 4.1 (B)). It might be argued that the use of restraints on the backbone C_α atoms of the protein (to prevent drifting) during the umbrella sampling simulations might hinder relaxation of the protein, and thereby cause the PMF to rise. However, it must be mentioned that the protein structure used here was previously subjected to unrestrained equilibration simulations in explicit solvent and bilayer for 30 ns, which is expected to be long enough for the protein backbone and sidechain atoms to relax. Furthermore, unlike transporters, ion channels do not usually undergo large scale domain motions during the process of solute transport, since they undergo gating prior to solute transport [66]. Thus the use of weak restraints on C_α atoms during the umbrella sampling is not likely to affect the dynamics in a significant way.

An estimate of the rate of ion conduction can be made using the Arrhenius equation. An energy barrier of around 43 kcal/mol that is reported here suggests that the rate for the ion conduction to be several orders of magnitude lower than that estimated using the conductivity measurements previously reported by Mehnert et al. [28]. Thus, although the present study is able to qualitatively account for the weak channel activity of Vpu, it is not able to quantitatively reproduce the conductance reported in experiments. The difficulty in accurately reproducing experimentally determined conductances from the height of energy barriers derived from PMFs has been reported by Sansom and coworkers [65]. One factor that could lead to inaccuracies in the PMF is the non-polarizable nature of the force field used [47,48]. Apart from the protein, the polarizability of the water model is also greatly important, since the pore water molecules play a crucial role in stabilizing the ion inside the pore [47].

Another factor that could affect the quantification of properties of the channel is an inadequate representation (in simulations) of the conditions prevailing in electrophysiological experiments. Experimental channel recordings investigating selectivity in Vpu have been carried out under the application of a voltage [17,21,22,27,28]. It is likely that the presence of an electric potential brings about conformational changes in the channel, leading to different conductance states of

the channel. A possible approach for obtaining insights into these conductance states is to model the channel in the presence of a voltage. However, Vpu has been shown to exhibit a variety of conductance states even if a very small potential is applied [27]. Furthermore, the open time for the channel has been shown to be voltage-independent [27]. The factors leading to the occurrence of the different conductance states, therefore, are still not properly understood. It has been suggested that gating in the channel might be regulated by the lateral membrane pressure and the lipid composition [27]. Thus, an understanding of the different conductance states might be possible by modeling ion permeation through the channel in a number of phospholipid environments with varying lateral pressure. Further studies are necessary in understanding the role of applied external voltage on the structure and dynamics of Vpu, and the resultant changes in ion permeation thermodynamics/mechanism if any.

4.4 Conclusions

Free energy profiles for conduction of monovalent cations across the Vpu channel have been characterized using umbrella sampling free energy calculations. The calculations reveal reasonably high energy barriers for ion movement through the channel, owing to the highly hydrophobic environment of the pore. The high energy barriers suggest that ion transport across the pore is kinetically slow, making Vpu a weakly conducting channel. The weak ion conducting activity is consistent with experimental biophysical studies, and has been explained based on the free energy calculations and hydration properties. The energy barriers arise from the desolvation of the ion as it moves from bulk water to the middle hydrophobic region of the channel. The permeation of Na^+ and K^+ involves comparable energetic costs, explaining why the channel does not discriminate Na^+ over K^+. The results show that the Ser23 residue plays a role in ion channel activity by giving rise to a hydrophilic region in the pore, and by serving as a weak binding site. The possibility of multi-ion permeation has been investigated by computing the free energy as a function of the positions of two permeating ions. Results show that only one ion can pass at a time through the middle hydrophobic region of the channel. The mechanism proposed is able to account for the slow kinetics of ion transport across the channel, and it suggests a role for the hydrophobic stretch in controlling the kinetics of the process.

References

1. Cohen EA, Terwilliger EF, Sodroski JG, Haseltine WA (1988) Identification of a protein encoded by the vpu gene of HIV-1. Nature 334(6182):532–534.
2. Strebel K, Klimkait T, Martin M (1988) A novel gene of HIV-1, vpu, and its 16-kilodalton product. Science 241(4870):1221–1223.
3. Maldarelli F, Chen MY, Willey RL, Strebel K (1993) Human immunodeficiency virus type 1 Vpu protein is an oligomeric type I integral membrane protein. J Virol 67(8):5056-5061.
4. Malim MH, Emerman M (2008) HIV-1 accessory proteins—Ensuring viral survival in a hostile environment. Cell Host Microbe 3(6):388–398.
5. Tokarev A, Guatelli J (2011) Misdirection of membrane trafficking by HIV-1 Vpu and Nef. Cell Logist 1(3):90-102.
6. Wang K, Xie S, Sun B (2011) Viral proteins function as ion channels. Biochim Biophys Acta, Biomembr 1808(2):510-515.
7. Fischer WB, Hsu HJ (2011) Viral channel forming proteins—modeling the target. BBA - Biomembranes 1808(2):561-571.
8. Strebel K (2014) HIV-1 Vpu—an ion channel in search of a job. Biochim Biophys Acta, Biomembr 1838(4):1074-1081.
9. Fischer WB, Li LH, Mahato DR, Wang YT, Chen CP (2014) Viral channel proteins in intracellular protein–protein communication: Vpu of HIV-1, E5 of HPV16 and p7 of HCV. Biochim Biophys Acta, Biomembr 1838(4):1113-1121.
10. Ahn S, Lim G, Nam S, Lee J (2014) Structural modeling of Vpu from HIV-1 based on solid-state NMR observables. Chem Phys Lett 599:116-121.
11. Wray V, et al. (1995) Solution structure of the hydrophilic region of HIV-1 encoded virus protein U (Vpu) by CD and 1H NMR spectroscopy. Int J Pept Protein Res 45(1):35-43.
12. Federau T, et al. (1996) Solution structure of the cytoplasmic domain of the human immunodeficiency virus type 1 encoded virus protein U (Vpu). Int J Pept Protein Res 47(4):297-310.
13. Willey RL, Maldarelli F, Martin MA, Strebel K (1992) Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. J Virol 66(12):7193-7200.

14. Schubert U, et al. (1996) The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains. J Virol 70(2):809-819.
15. Wray V, et al. (1999) Solution structure and orientation of the transmembrane anchor domain of the HIV-1-encoded virus protein U by high-resolution and solid-state NMR spectroscopy. Biochemistry 38(16):5272-5282.
16. Kukol A, Arkin IT (1999) Vpu transmembrane peptide structure obtained by site-specific fourier transform infrared dichroism and global molecular dynamics searching. Biophys J 77(3):1594-1601.
17. Marassi FM, et al. (1999) Correlation of the structural and functional domains in the membrane protein Vpu from HIV-1. Proc Natl Acad Sci USA 96(25):14336-14341.
18. Ma C, et al. (2002) Expression, purification, and activities of full-length and truncated versions of the integral membrane protein Vpu from HIV-1. Protein Sci 11(3):546-557.
19. Park SH, et al. (2003) Three-dimensional structure of the channel-forming trans-membrane domain of virus protein "u" (Vpu) from HIV-1. J Mol Biol 333(2):409-424.
20. Park SH, De Angelis AA, Nevzorov AA, Wu CH, Opella SJ (2006) Three-dimensional structure of the transmembrane domain of Vpu from HIV-1 in aligned phospholipid bicelles. Biophys J 91(8):3032-3042.
21. Ewart GD, Sutherland T, Gage PW, Cox GB (1996) The Vpu protein of human immunodeficiency virus type 1 forms cation-selective ion channels. J Virol 70(10):7108-7115.
22. Schubert U, et al. (1996) Identification of an ion channel activity of the Vpu transmembrane domain and its involvement in the regulation of virus release from HIV-1-infected cells. FEBS Lett 398(1):12–18.
23. Neil SJ, Zang T, Bieniasz PD (2008) Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature 451(7177):425-430.
24. Padhi S, Khan N, Jameel S, Priyakumar UD (2013) Molecular dynamics simulations reveal the HIV-1 Vpu transmembrane protein to form stable pentamers. PLoS ONE 8(11):e79779.
25. Hussain A, Das SR, Tanwar C, Jameel S (2007) Oligomerization of the human immunodeficiency virus type 1 (HIV-1) Vpu protein–a genetic, biochemical and biophysical analysis. Virol J 4(1):81.

26. Lopez CF, Montal M, Blasie JK, Klein ML, Moore PB (2002) Molecular dynamics investigation of membrane-bound bundles of the channel-forming transmembrane domain of viral protein U from the human immunodeficiency virus HIV-1. Biophys J 83(3):1259-1267.
27. Mehnert T, et al. (2007) Towards a mechanism of function of the viral ion channel Vpu from HIV-1. J Biomol Struct Dyn 24(6):589-596.
28. Mehnert T, et al. (2008) Biophysical characterization of Vpu from HIV-1 suggests a channel-pore dualism. Proteins 70(4):1488-1497.
29. Patargias G, Martay H, Fischer WB (2009) Reconstructing potentials of mean force from short steered molecular dynamics simulations of Vpu from HIV-1. J Biomol Struct Dyn 27(1):1-12.
30. Li L-H, Hsu H-J, Fischer WB (2013) Assembling viral channel forming proteins: Vpu from HIV-1. Biopolymers 99(8):517-529.
31. Kandt C, Ash WL, Tieleman DP (2007) Setting up and running molecular dynamics simulations of membrane proteins. Methods 41(4):475-488.
32. Khalili-Araghi F, et al. (2009) Molecular dynamics simulations of membrane channels and transporters. Curr Opin Struc Biol 19(2):128-137.
33. Stansfeld PJ, Sansom MSP (2011) Molecular simulation approaches to membrane proteins. Structure 19(11):1562-1572.
34. Roux B, et al. (2011) Ion selectivity in channels and transporters. J Gen Physiol 137(5):415-426.
35. Johnston JM, Filizola M (2011) Showcasing modern molecular dynamics simulations of membrane proteins through G protein-coupled receptors. Curr Opin Struc Biol 21(4):552-558.
36. Shaikh SA, et al. (2013) Visualizing functional motions of membrane transporters with molecular dynamics simulations. Biochemistry 52(4):569-587.
37. Bu L, Im W, Brooks III CL (2007) Membrane assembly of simple helix homo-oligomers studied via molecular dynamics simulations. Biophys J 92(3):854-863.
38. Zhu F, Hummer G (2010) Pore opening and closing of a pentameric ligand-gated ion channel. Proc Natl Acad Sci USA 107(46):19814-19819.

39. Hsu HJ, Fischer WB (2012) In silico investigations of possible routes of assembly of ORF 3a from SARS-CoV. J Mol Modell 18(2):501-514.
40. de Jesus AJ, Allen TW (2013) The determinants of hydrophobic mismatch response for transmembrane helices. Biochim Biophys Acta, Biomembr 1828(2):851-863.
41. Lai CL, et al. (2013) Molecular mechanism of membrane binding of the GRP1 PH domain. J Mol Biol 425(17):3073-3090.
42. Kästner J (2011) Umbrella sampling. WIREs Comput Mol Sci 1(6):932–942.
43. Chipot C (2014) Frontiers in free-energy calculations of biological systems. WIREs Comput Mol Sci 4(1):71-89.
44. Priyakumar UD, MacKerell Jr AD (2006) Computational approaches for investigating base flipping in oligonucleotides. Chem Rev 106(2):489-505.
45. Priyakumar UD, MacKerell Jr AD (2006) NMR imino proton exchange experiments on duplex DNA primarily monitor the opening of purine bases. J Am Chem Soc 128(3):678-679.
46. Berneche S, Roux B (2001) Energetics of ion conduction through the K^+ channel. Nature 414(6859):73-77.
47. Allen TW, Andersen OS, Roux B (2004) Energetics of ion conduction through the gramicidin channel. Proc Natl Acad Sci USA 101(1):117-122.
48. Allen TW, Andersen OS, Roux B (2006) Ion permeation through a narrow channel: Using gramicidin to ascertain all-atom molecular dynamics potential of mean force methodology and biomolecular force fields. Biophys J 90(10):3447-3468.
49. Egwolf B, Roux B (2010) Ion selectivity of the KcsA channel: A perspective from multi-ion free energy landscapes. J Mol Biol 401(5):831-842.
50. Kale L, et al. (1999) NAMD2: greater scalability for parallel molecular dynamics. J Comput Phys 151(1):283-312.
51. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802.
52. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616.
53. MacKerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in

reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25(11):1400–1415.

54. Klauda JB, et al. (2010) Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. J Phys Chem B 114(23):7830–7843.
55. Beglov D, Roux B (1994) Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. J Chem Phys 100(12):9050-9063.
56. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926-935.
57. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13(8):1011–1021.
58. Roux B (1995) The calculation of the potential of mean force using computer simulations. Comp Phys Comm 91(1-3):275–282.
59. Grossfield A "WHAM: the weighted histogram analysis method", version 2.0.6, http://membrane.urmc.rochester.edu/content/wham.
60. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. J Comput Chem 30(10):1545–1614.
61. Ryckaert J-P, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341.
62. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graphics 14(1):33–38.
63. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res 32:W665-W667.
64. Fritsch S, Ivanov I, Wang H, Cheng X (2011) Ion selectivity mechanism in a bacterial pentameric ligand-gated ion channel. Biophys J 100(2):390-398.
65. Fowler PW, Abad E, Beckstein O, Sansom MS (2013) Energetics of multi-ion conduction pathways in potassium ion channels. J Chem Theory Comput 9(11):5176-5189.

66. Shaikh S, Wen PC, Enkavi G, Huang Z, Tajkhorshid E (2010) Capturing functional motions of membrane channels and transporters with molecular dynamics simulation. J Comput Theor Nanosci 7(12):2481.

5 Modeling Structures of Helical Membrane Proteins

5.1 Introduction

Membrane proteins account for more than 60% of drug targets, with receptors and transporters being the major classes of proteins being targeted [1]. However, in spite of the overwhelming medical importance of these proteins, the structures of only a very small fraction of these proteins have been solved experimentally, owing to the difficulty in solubilizing and crystallizing membrane proteins [2,3]. Thus, the major challenge in drug design targeting this class of proteins is limited availability of solved structures. This has necessitated the use of computational approaches in modeling the structures of membrane proteins.

As far as structure is concerned, the two major classes of membrane proteins are α-helical bundles and β-barrels, of which the latter are less abundant [4]. Consequently, most computational approaches that have attempted at predicting the structures of membrane proteins have focused on α-helical bundles. Brunger and coworkers modeled the structure of the pentameric protein phospholamban by combining a global search algorithm with mutagenesis data [5]. The structure of the pentamer was modeled using a two-body search as well as a pentamer search, and the structures from the two methods were found to be similar to each other. The use of the same search protocol together with simulated annealing was later applied on the glycophorin A (GpA) dimer [6]. A search protocol that optimizes the van der Waals interactions between the helices has been used to obtain models for GpA, neu/erbB-2, and phospholamban [7]. Conformational searches for transmembrane helices have also been performed in implicit membrane environments [8,9]. Generation of protein conformations in a implicit membrane environment using the Rosetta fragment assembly method has been a notable approach in membrane protein structure prediction [10].

Often, it has proved useful to incorporate available experimental information into the modeling protocol, thereby leading to the sampling of structures close to the native structure with a high probability. Kukol et al. have modeled the structures of the M2 proton channel from influenza A virus and the Vpu channel from HIV-1 by performing a molecular dynamics search guided by

the helical tilt angle value and the rotational pitch angle value, with the values for the two angles being obtained from site-directed infrared dichroism spectra [11,12]. The structure of GpA dimer has successfully been predicted using a potential smoothing algorithm, and thereby optimizing interhelical van der Waals interactions [13]. Symmetric homo-oligomers of phospholamban have been modeled by extensively searching the configuration space and taking into consideration only those structures that are consistent with NMR restraints [14]. Constraining helix-helix packing at contact sites identified from experiments or from sequence information has proven to be a useful approach in predicting structures of helical membrane proteins [15]. However, methods that incorporate experimental constraints are dependent on the availability of experimental information.

Although homology models can often provide accurate atomistic models for a protein, they suffer from the limitation that they can be used only if a suitable structural template is available. A useful way of sampling the conformational space of helical assemblies is the use of classical molecular simulations, in particular Monte Carlo (MC) simulations and molecular dynamics (MD) simulations. Sampling of possible conformations using multiple independent MC simulations followed by filtering of structures based on symmetry has successfully been employed to model transmembrane helices [16]. Replica exchange MC simulations have also been shown to be a useful approach for helical bundles [17]. Using coarse-grained MD simulations (CG-MD), Sansom and coworkers have modeled the assembly of monomeric units of M2 to form a tetramer, where the structural features of the assembled tetramer were seen to resemble those of a closed state [18]. The transmembrane domains (TMDs) of p7 channel from hepatitis C virus have been assembled by following two different protocols: (1) simulated annealing with restrained molecular dynamics, and (2) protein-protein docking [19]. Yet another approach for assembling helices has been equilibration of the monomeric unit followed by systematic variation of the interhelical distance, the tilt angle (with respect to membrane normal), and the spin angle of the equilibrated monomer to obtain the oligomer [20]. Energetically stable states are then identified from among the sampled conformations. Fischer and coworkers have investigated the assembly of helical transmembrane domains in the proteins SARS 3a and HIV-1 Vpu via two approaches: sequential assembly, in which the different helices are assembled one after the other, and simultaneous assembly, in which all helices are assembled together [21,22].

A successful approach for modeling the oligomeric forms of transmembrane helices is the use of replica-exchange MD (REMD) simulations in an implicit membrane environment [23]. Implicit membranes are able to mimic the physical nature of the lipid bilayer by using a low dielectric slab for modeling the hydrophobic core [24,25]. Thus there are no explicit lipid or water molecules, and the computing cost is greatly reduced. Using REMD with an implicit membrane environment, Brooks and coworkers have been able to predict the structures of homo-oligomers, given that the number of monomeric subunits in the oligomer is known. It has also been possible to predict the oligomeric state of the protein Vpu from HIV-1 by initially sampling possible conformations of different oligomeric states of the protein using REMD, and then subjecting a few representative conformations to extensive simulations in an explicit lipid bilayer environment [26]. The study shows that the pentameric form of Vpu is the preferred oligomeric state, and thereby proposes a structural model for the pentameric channel.

In modeling helical homo-oligomers, a critical factor that could affect the sampling of conformations and, consequently, the accuracy of the final model, is the orientation of the helices in the initial conformation for the simulation. The present study focuses on modeling the structures of transmembrane helices by optimizing the orientation of the hydrophilic residues in the helices. Two different approaches are then used on these oligomeric structures to sample possible conformations. Representative structures selected from among the sampled conformations based on geometric criteria are seen to be strikingly close to experimental structures.

5.2 Methods

The monomeric forms of the transmembrane domains (TMD) of influenza A M2 (M2), influenza B M2 (BM2), and ErbB2 were generated by modeling the residues as an idealized α-helix using the molecular modeling programs SYBYL7.2 (Tripos International, St. Louis, Missouri, USA; http://www.tripos.com) and Chimera [27]. The sequences of the three proteins are given in Table 5.1. The helix was aligned along the z-axis, and then translated by an appropriate distance along the xy-plane (9.7 Å for M2 and 10.5 Å for BM2; the distance was chosen such that it was sufficient for the helices to rotate about their principal axis, while at the same time being able to

interact with neighboring monomers). To obtain a tetramer model for M2 and BM2, three images of the monomeric unit were generated by applying four-fold rotational symmetry about the z-axis; the operation was performed using the IMAGE facility in the CHARMM program [28]. Tetrameric forms have been modeled here since previous experimental studies have shown M2 and BM2 to be tetrameric [29-32]. Similarly, taking into account the known oligomeric state of ErbB2 [33], a dimeric form of the protein was modeled by applying two-fold symmetry about the z-axis (the distance of the centre of mass of a given monomer from the origin being 8 Å). The initial orientation of the residues of the α-helices with respect to each other was chosen based on minimization of the exposed hydrophilic surface area of the oligomer (see Results and Discussion for more details on this approach).

Table 5.1 Sequences of M2, BM2, and ErbB2

Protein	**Sequence**
M2	SSDPLVVAASIIGILHLILWILDRL
BM2	MLEPFQILSISSFILSALHFIAWTIGHLNQIKR
ErbB2	SPLTSIISAVVGILLVVVLGVVFGILIKRRQ

Abbreviations: M2: influenza A M2; BM2: influenza B M2

With this initial conformation, two independent approaches were followed. In the first approach, a total of 15 independent simulations of the oligomer (with one real monomer and three image monomers) were carried out with different sets of initial velocities. Initially, the membrane was modeled in an implicit manner using the generalized Born model with a simple switching function (GBSW) [24,25]. The implicit membrane had a hydrophobic core of thickness 25 Å, and a smoothing region of thickness 0.5 Å was used on both sides of the hydrophobic core. Each simulation was performed for 6 ns, with the first 1 ns being regarded as equilibration period. Langevin dynamics was performed with a friction coefficient of 5 ps^{-1} for heavy atoms, and the temperature was maintained at 300 K. Structures from the last 5 ns of each trajectory were considered for clustering. The second approach was the use of REMD for sampling of the conformations in an implicit membrane environment. A total of eight replicas distributed exponentially over the range 300 K to 400 K were used, and exchanges were attempted every 1 ps. Langevin dynamics was used, and the simulation was performed for 15 ns, so the simulation time including all the replicas was (15*8) = 120 ns. Conformations sampled in the last 10 ns for

the replica at the lowest temperature were considered for clustering. All simulations and analyses in the two approaches described here were performed with the CHARMM program using the CHARMM36 all-atom protein force field [28,34-36]. In these simulations, all pairwise nonbonded forces were calculated explicitly irrespective of the distance between the concerned atoms. Covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm [37]. The total sampling time including all the simulations from the two approaches is (90+120)*4 = 840 ns.

For selecting a representative structure from the conformations sampled in the above two approaches, three geometrical features were considered: interhelical distance (the distance between a given monomer and its adjacent monomer), the tilt angle of a monomer with respect to the membrane normal, and the orientation angle of the hydrophilic residues in the transmembrane region of the protein. The protocol for selecting a representative structure was carried out separately for the above two approaches to obtain one representative structure for each approach. Structures were first filtered based on the above three parameters, and the filtered structures were then clustered on the basis of root mean square deviation (RMSD). The structure at the center of the largest cluster was chosen as the representative structure. A more elaborate discussion of the subtleties involved in the protocol can be found in the Results and Discussion section.

The stability of the representative structures was tested by modeling them in an explicit membrane. For this, the individual monomer units of the protein were modeled explicitly rather than using images. The structure was first set up in a hydrated lipid bilayer of POPC molecules using the CHARMM-GUI Membrane Builder [38,39]. Equilibration was performed in the presence of positional restraints on the protein heavy atoms and planar restraints on the lipid headgroups, with the restraints being gradually reduced over a period of 375 ps. This was followed by 20 ns of simulation in the NPT ensemble without any restraints. Nonbonded interactions were truncated by using a smoothing function between 10 Å to 12 Å, and long range electrostatic interactions were calculated using the particle mesh Ewald method [40]. The SHAKE algorithm was used for constraining bonds involving hydrogens [37]. The simulations were performed using the NAMD program with the CHARMM36 protein force field, the

CHARMM36 lipid force field, the TIP3P water model and optimized parameters for ions [34-36,41-44]. Structural snapshots were rendered using VMD [45].

5.3 Results and Discussion

The first part of the Results and Discussion section illustrates the application of the method on the M2 protein. Extension of the method for modeling BM2 and ErbB2 is described thereafter.

5.3.1 Assembling α-helices based on the minimum unfavorable contacts approach

Helices that span across the membrane have the majority of their exposed surface lying in the membrane hydrophobic core, and only the terminal region is exposed to the solvent on the two sides of the membrane. The TMDs in this study have been modeled such that the hydrophilic surface area on the exterior of the protein is minimized, thereby minimizing unfavorable contacts between hydrophilic residues and the hydrophobic core of the membrane. To achieve this, a monomeric unit was first aligned along the z-axis, and the solvent accessible surface area (SASA) of all residues was calculated. The SASA values were normalized with respect to the residue with the largest SASA (Trp41). Figure 5.1 (A) shows a polar plot depicting the orientation of the different residues in the monomer together with their sidechain SASA values. Hydrophobic residues are shown in red and hydrophilic residues in blue. The orientation shown for the residues is a projection of their sidechains in the xy-plane. The orientation angle for a residue is found out by calculating the angle formed between the x-axis and a line joining the center of mass of the helical backbone with the center of mass of the concerned residue. The length of the lines in the polar plot indicates the normalized SASA of the residue sidechains: the longer is the line, the larger is the SASA value. The terminal three residues on the two sides of the helix have been omitted in the figure (and in this analysis), since these lie in the bulk water region. Two hydrophilic residues can be seen in the figure: Ser31 and His37. It must be mentioned that His37 has been modeled in a neutral form here to model the channel in a deprotonated form. While Ser31 has a normalized SASA of 0.46, His37 has a normalized SASA of 0.85, meaning that the histidine has a much larger surface area. Ser31 and His37 lie on different faces of the helix, with the angle between the projections of the two residues in the xy-plane being 113°. Figure 5.1 (B) shows an N-terminal view of the helix with the two hydrophilic residues. It can be clearly seen that the two residues are oriented in different directions.

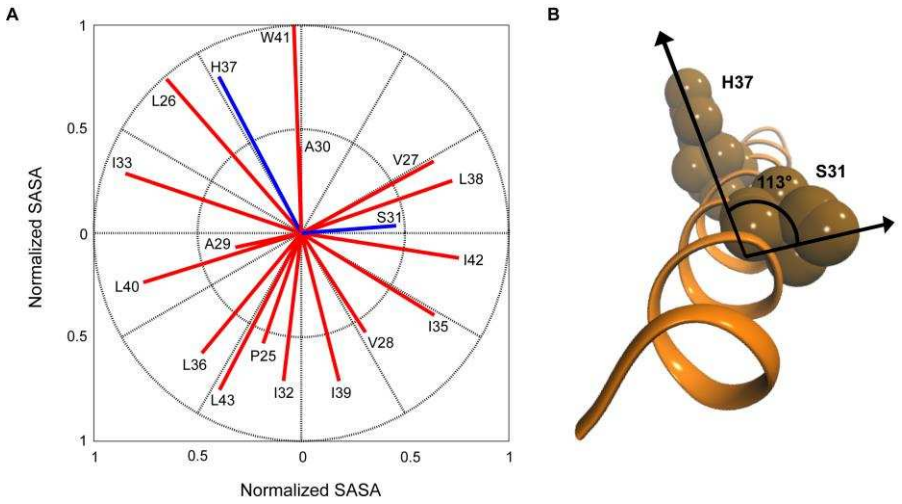

Figure 5.1 (A) Polar plot showing the orientation and solvent accessible surface area of the residues in the M2 monomer. The monomer has been aligned along the z-axis prior to the generation of the plot. Each residue is represented by a line, which is a projection in the xy-plane (the membrane plane) of a vector joining the center of mass of helical backbone with center of mass of the concerned residue sidechain. The SASA of all residues have been normalized with respect to Trp41, the residue with the largest surface area. The length of the lines in the plot indicates the normalized solvent accessible surface area of the residue sidechain. Hydrophobic residues are shown in red, and hydrophilic residues in blue. The terminal three residues on the two sides of a helix have been omitted in the figure. (B) N-terminal view of the helix showing the residues Ser31 and His37 in van der Waals representation. The angle formed between the projections in the xy-plane for the two residues is also shown.

The monomer was then translated a distance 9.7 Å along the xy-plane, and three images of the monomer were generated by applying a four-fold rotational symmetry about the z-axis. The next step was to rotate the helices such that the hydrophilic surface area on the exterior of the oligomer was minimized. It is not possible to orient the helices in such a way that both the polar residues are buried in the interior, since the two polar residues lie on different faces of the helix. Between Ser31 and His37, histidine has a larger sidechain SASA, so the hydrophilic surface area for the exterior of the oligomer can be minimized if the histidine rather than the serine faces the interior. Thus the helices were rotated such that the center of mass of the histidine sidechain on a given helix lied on a line joining the center of mass of the helix with the center of mass of the entire oligomer. Figure 5.2 (A) shows a polar plot for the tetramer after orienting the helices, while Figure 5.2 (B) shows the corresponding structure of the tetramer with the different sidechains. As can be seen clearly, the contacts formed by the polar residues with the hydrophobic exterior (the membrane hydrophobic core) have been minimized. It must be

mentioned that it is possible to attain an energetically stable arrangement of the oligomer if such unfavorable contacts are minimized. Since the native state of a protein is usually more stable than most of the other possible conformational states, modeling a structure that has optimum stability in terms of the orientation of polar and nonpolar residues is likely to lead to the sampling of conformations that are very close to the native structure. It follows that an unfavorable arrangement of residues in the initial conformation could make it difficult to sample structures that are close to the native conformation, even if the method used for sampling conformations is a rigorous one.

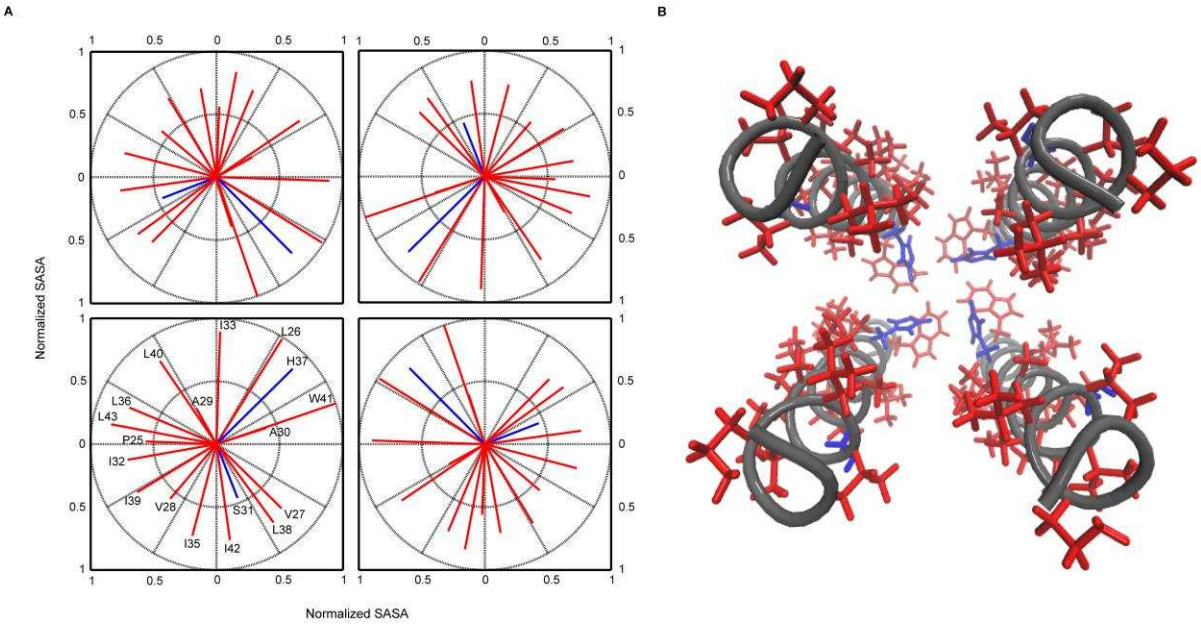

Figure 5.2 (A) Orientation and normalized solvent accessible surface area of the residues in the tetramer. The monomers are perpendicular to the xy-plane, and the four monomeric units lie in the four quadrants of the xy-plane. The tetramer has been generated by applying a four-fold rotational symmetry about the z-axis. The polar His37 sidechain is seen facing the center of the oligomer. (B) Structure of the tetramer model corresponding to the arrangement shown in the polar plot. Hydrophobic sidechains are shown in red and hydrophilic sidechains in blue. The terminal three residues on the two sides of the helix have been omitted.

It must be noted that oligomeric states other than the tetrameric form of M2 have not been modeled here, since M2 is known to be a tetramer [29,30]. Like most of the approaches described above for modeling oligomeric transmembrane helices (see Introduction), the approach used here takes into consideration the known oligomeric state of the protein. Furthermore, the oligomers have been modeled here using one real monomer and three image monomers, which are basically images of the real monomer generated by applying a four-fold rotational symmetry

about a central axis. The use of images preserves the symmetry of the tetramer about the principal axis. Such symmetry is characteristic of homo-oligomeric transmembrane helices, so the use of a constraint on the symmetry of the oligomer ensures that meaningful conformations are sampled.

5.3.2 Multiple simulations and REMD

The first approach employed for sampling conformations was the use of multiple simulations with different initial velocities. It has been shown that multiple short simulations that use different initial velocities are able to sample a greater variety of conformations compared to a single long simulation [46]. Sampling in a single brute-force simulation is likely to be confined to one energy well, and transitions between conformations are rare, owing to the difficulty in overcoming energy barriers. Performing a number of simulations with different initial velocities ensures that one moves on the potential energy surface in multiple directions, thereby making it possible to reach a number of local minima.

Figure 5.3 shows the RMSD of the different conformations sampled in the 15 simulations with respect to the initial modeled structure. The RMSD varies from 2.5 Å to 6 Å in the different simulations. Such a broad range is indicative of extensive sampling of the conformational space. If one considers RMSD from individual trajectories, it can be seen that the RMSD is converged in most of the cases after the initial equilibration time. Thus, there are hardly any transitions from one conformation to another in a given trajectory. It follows that it is rather improbable that conformations from multiple energy minima would be sampled using a single long simulation of this kind. Evidently, the use of multiple simulations makes it possible to sample a greater region of conformational space than that accessible to a single simulation.

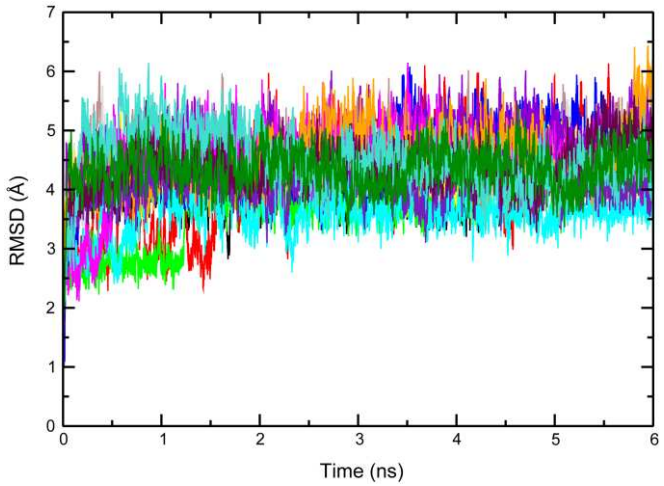

Figure 5.3 RMSD of the conformations sampled in the fifteen simulations performed with different initial velocities.

In an independent approach for sampling conformations, the same arrangement of the tetramer was subjected to REMD simulations. The use of REMD avoids the problem of getting entrapped in a local minimum, thereby allowing the sampling of a vast region of conformational space [23,47]. The RMSD of the structures attained converged values after the first 5 ns of simulation (Figure 5.4), so the conformations sampled in the last 10 ns were considered for further analysis. The RMSD is seen to fluctuate between two distinct conformations (one corresponding to RMSD values of about 3 Å, and the other corresponding to values of about 6.5 Å) during the last 10 ns; this is accounted for later.

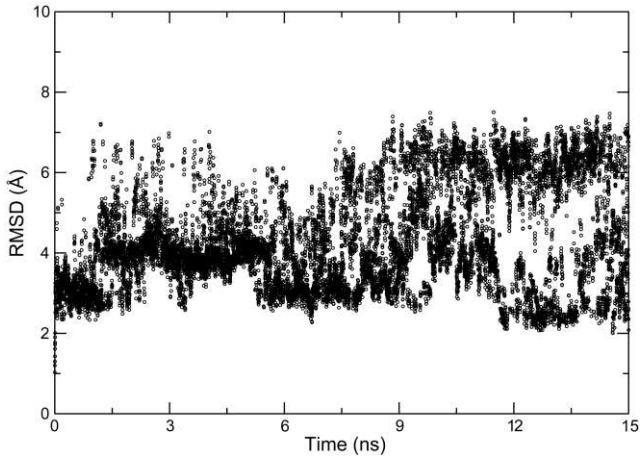

Figure 5.4 Backbone RMSD of the conformations sampled for M2 over the course of 15 ns of REMD.

5.3.3 Geometry-based selection

Three geometrical parameters were analyzed for the sampled conformations: interhelical distance, tilt angle of the helical axis with respect to membrane normal, and orientation angle of the His37 sidechain (which is defined as the angle formed by the center of mass of His37 sidechain on a monomer i, the center of mass of the helical backbone of monomer i, and the center of mass of the helical backbone of all four monomers taken together). These three parameters satisfactorily describe the overall structure of a helical bundle, and have been shown to be important in a number of previous modeling approaches [16,20]. Conformations chosen for the analysis were from the last 5 ns of each trajectory in the multiple simulation approach, and the last 10 ns of the trajectory at the lowest temperature (300 K) in the REMD simulations. This gave a total of 75,000 structures from the first method and 10,000 structures from the second. Structures from the two approaches were analyzed separately. The values for these parameters have been calculated for only one of the four helices (or one out of four helical pairs, in the case of interhelical distance), since the values for the other three helices would be exactly identical to the first helix, as these three helices are images of the first helix.

Figure 5.5 (A) shows the probability distribution of the interhelical distance from the multiple simulation approach and REMD. The distribution is broader in the case of REMD, which means that the helices are able to relax and adjust their distance from the neighboring helix, thereby making possible an extensive sampling of conformations. Figure 5.5 (B) shows the probability distribution of the tilt angle from the two approaches. Structures that have a left-handed arrangement of helices are assigned a positive tilt angle value, and those that are right-handed are given a negative value. Both approaches are able to sample left-handed as well as right-handed structures, although the number of right-handed conformations in the multiple simulation approach is rather small. As seen from the two distinct peaks, the REMD simulations are able to sample the two conformations almost equally, with a marginal preference for the left-handed structures (51.2% of structures being left-handed, as opposed to 48.8% right-handed structures). This means that the native structure could correspond to either of these two conformations. This is comparable to the sampling of both left-handed and right-handed conformations reported previously for the GpA homodimer and integrin αIIb/β3 heterodimer [48]. However, while the peak for left-handed structures is seen for conformations sampled from both the approaches, the

peak corresponding to the right-handed structures is almost negligible for the multiple simulation approach. This suggests a greater propensity of the structure to exist in a left-handed state, and it follows that the left-handed conformations are more likely to be closer to the actual structure.

The His37 sidechain orientation angles for multiple simulations and REMD are shown in Figure 5.5 (C). The orientation angle follows the spin of the His37 sidechain about the helical axis, with a value close to 0° suggesting that the sidechain faces the interior of the tetramer, and a value of 90° suggesting that it lies in the interhelical interface. The histidine is seen to occur in a wide range of orientations in the REMD, indicating a thorough search of the conformational space. The vertex of the peak, however, occurs at almost the same position for the two approaches. This suggests that, although sampling of the conformational space is better in the case of REMD, the most probable structures from the two methods are similar.

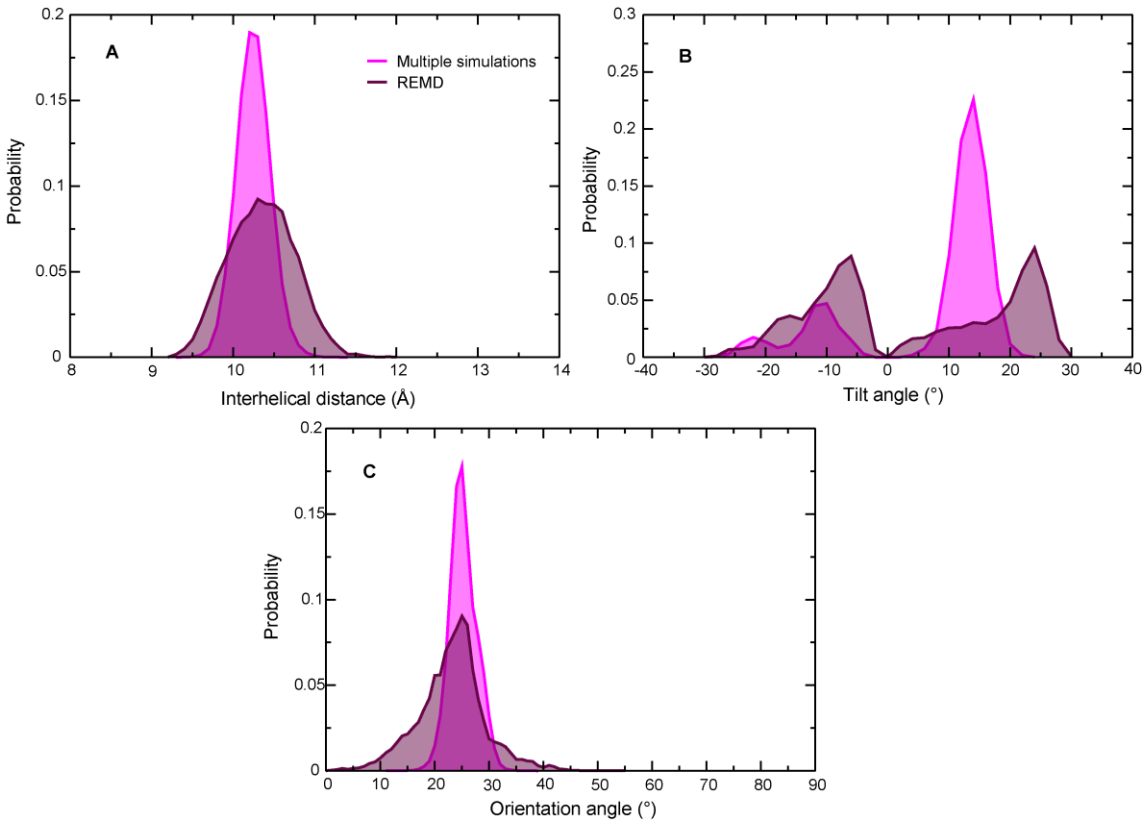

Figure 5.5 Probability distribution of (A) the interhelical distance, (B) the tilt angle, and (C) the His37 sidechain orientation angle for the structures sampled in the multiple simulation approach and REMD. Structures from the former approach have been chosen from the last 5 ns of each trajectory, and those from the latter have been chosen from the last 10 ns of the trajectory at the lowest temperature.

Conformations that are sampled more frequently during the simulations are likely to be preferred conformations. If one considers the three geometrical parameters discussed above, the conformations occurring with high probability are the ones that lie in the high probability regions of the probability distributions for all the three parameters. Thus, for further analysis, conformations that had values of interhelical distance, tilt angle, and orientation angle lying in the range corresponding to the peak of the probability distribution for the respective feature were chosen. The range for a given structural feature was determined by considering the region of the distribution that had a probability value of greater than 4/5 of the maximum probability. The range of values of the three parameters from which the structures were filtered is summarized in Table 5.2. It is important to note here that structures that had values falling into the mentioned range for all the three parameters have been selected. Structures that did not have the desired value for any one of the three parameters were rejected. For the tilt angle probability distribution for REMD, conformations corresponding to the peak for left-handed structures were selected, since there is a greater propensity of the structure to exist in a left-handed state, as described above. Table 5.3 shows the number of conformations selected from each trajectory in the multiple simulation approach (the table also includes details about the simulations on BM2 and ErbB2, which are described later in this chapter).

Table 5.2 Range of values of interhelical distance, tilt angle, and orientation angle from which the conformations were selected

Protein	**Sampling method**	**Interhelical distance (Å)**	**Tilt angle (°)**	**Orientation angle (°)**
M2	MS	10.0-10.5	11.5-16.0	23.5-26.0
	REMD	10.25-10.75	22.5-27.5	23.5-27.5
BM2 model 1	MS	10.0-10.5	10.0-13.5	10.0-13.5
	REMD	10.45-10.95	10.0-16.0	12.0-20.0
BM2 model 2	MS	10.4-10.9	11.0-16.0	32.0-36.0
	REMD	10.45-10.95	12.5-15.0	26.0-35.0
ErbB2	MS	9.25-9.75	2.0-5.0	39.0-47.0
	REMD	9.1-9.6	1.5-3.5	37.5-49.5

Abbreviations: MS: multiple simulations; REMD: replica exchange molecular dynamics; M2: influenza A M2; BM2: influenza B M2

Table 5.3 Number of conformations chosen for clustering from each trajectory in the multiple simulation method

Simulation number	M2	BM2 model 1	BM2 model 2	ErbB2
1	1418	0	0	1143
2	1040	687	0	0
3	384	0	0	993
4	1328	151	0	0
5	1163	980	0	889
6	1511	481	238	0
7	1448	682	11	94
8	1080	677	0	6
9	1319	897	196	759
10	0	0	0	813
11	0	2	94	0
12	0	1233	188	0
13	1001	2001	1	551
14	1522	1674	114	0
15	872	0	389	0
Total	14086	9465	1231	5248

5.3.4 Clustering structures

After the filtering step, there were 14086 structures from the multiple simulation approach and only 44 structures from the REMD approach. The selection of smaller number of structures from REMD can be attributed to the fact that the probability distributions for the three parameters are much wider in the case of REMD (Figure 5.5). Since the conformational sampling in the case of REMD is much more extensive, there are few structures lying in the desired range for the three parameters. The filtered structures were clustered on the basis of backbone RMSD. Structures from the multiple simulation approach gave a total of 11 clusters, and those from REMD gave a total of 3 clusters. While the largest cluster from the former approach had a total of 2644

structures, the same from the latter approach had 16 structures. The structure lying at the center of the largest cluster was chosen as a representative structure for validation. Further details about the clusters obtained are shown in Table 5.4.

Table 5.4 Clustering of conformations from the different trajectories

Protein	Sampling method	Number of conformations selected for clustering	Number of clusters generated	Number of conformations in largest cluster	Standard deviation from cluster center in largest cluster (Å)
M2	MS	14086	11	2644	0.03
	REMD	44	3	16	0.08
BM2 model 1	MS	9465	9	2182	0.03
	REMD	16	6	5	0.10
BM2 model 2	MS	1231	8	275	0.05
	REMD	18	3	11	0.08
ErbB2	MS	5248	5	1650	0.04
	REMD	400	5	112	0.04

Conformations have been clustered on the basis of RMSD. The radius of a given cluster was set at 0.25 Å.
Abbreviations: MS: multiple simulations; REMD: replica exchange molecular dynamics.

5.3.5 Representative structures and their validation

The structure and orientation of residues in the representative structures are shown in Figure 5.6 (A,B). The histidine residue is seen to face the pore, which is a characteristic feature of the closed state of M2. A similar structural model was proposed previously by Sansom and coworkers [49]. The accuracy of the representative structures from the two approaches was tested by comparing these structures with an experimental structure of the channel (PDB code: 2L0J), which is an NMR structure further refined using MD simulations in the presence of symmetry restraints [50]. Being symmetric, this experimental structure is ideal for validating the predicted structural models for M2. While the representative structure from multiple simulations differed from the experimental structure by a backbone RMSD of 2.76 Å, the structure from REMD differed by only 1.91 Å. The remarkable similarity between the representative structures and the experimental structure can be seen in Figure 5.6 (C), which shows structural superposition of the representative structures on to the experimental structure. It must be noted

that only the TMD of the experimental structure has been considered in performing the superposition and calculating the RMSD. To further test the accuracy of the predicted structures, contacts between adjacent helices were investigated. Figures 5.7 (A), (B), and (C) show interhelical contact maps for the experimental structure, and the predicted structures from the two approaches, respectively. The maps show the distance between the residues on a given monomer and the residues on the adjacent monomer. The maps are shown for a single helix pair, since all helix pairs are identical in the tetramer, which is rotationally symmetrical about the principal axis. The three maps are strikingly similar to one another, suggesting that the same residues form the interhelical interface in the three structures. Notably, the close contact between Ala30 and Ser31, and that between Ile33 and Ile35, is conserved in all the three structures.

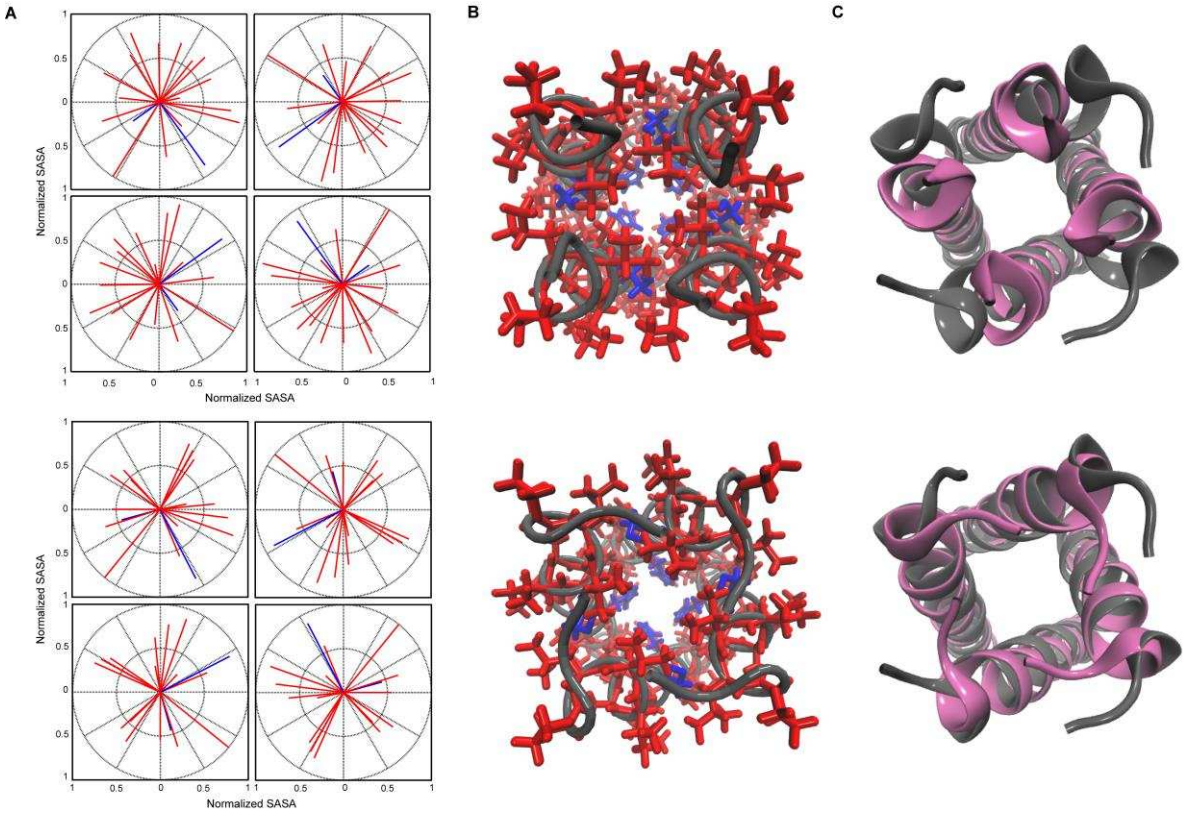

Figure 5.6 (A) Polar plots showing the orientation of residues in the representative structures obtained from the multiple simulation approach (above) and REMD (below). (B) Structure of the representative models with the hydrophilic residues shown in blue and hydrophobic residues in red. (C) The predicted structure from the multiple simulation approach (above) and the REMD approach (below) overlaid with the experimental structure. The predicted structure is shown in mauve and the experimental structure in gray.

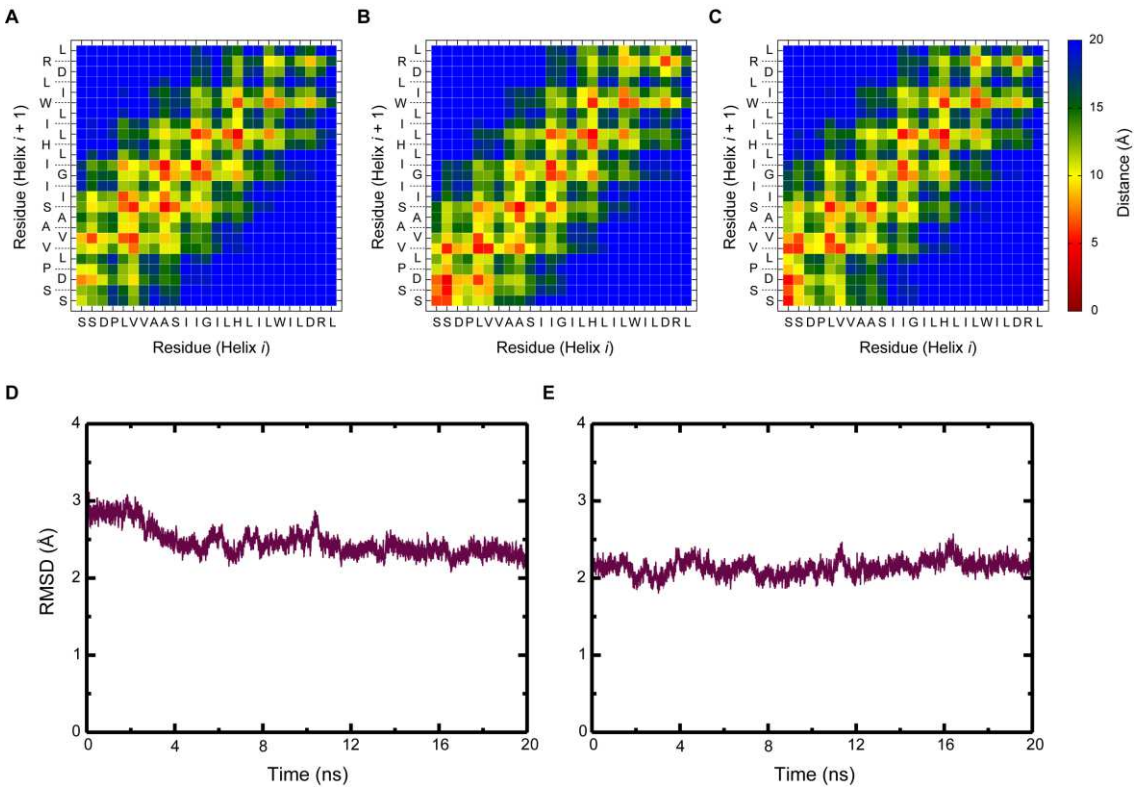

Figure 5.7 Validation of the predicted structures. Contact maps for (A) the experimental structure, (B) the representative structure from the multiple simulation approach, and (C) the representative structure from the REMD approach. The contact maps show distances between residues on a given monomer i and its adjacent monomer i+1. (D,E) RMSD with respect to the experimental structure of the representative structures from the multiple simulation approach and REMD, respectively.

The stability of the predicted structures was tested by modeling them in a fully solvated lipid bilayer environment. Figures 5.7 (D) and (E) show the backbone RMSD with respect to the experimental structure for the structures obtained from the multiple simulation method and REMD, respectively. The structure from the former method is seen to equilibrate within the first 5 ns. In fact, upon equilibration in the bilayer, the structure is further closer to the experimental structure, as indicated by the slight fall in RMSD. The structure from the REMD method, on the other hand, is quickly equilibrated, and it retains its RMSD difference of close to 2 Å with respect to the experimental structure. The RMSD results validate the stability of the structures in an explicit membrane. The structures are seen to retain their helicity throughout the simulations, and the orientation of the hydrophilic residues remains reasonably stable (data not shown).

5.3.6 Influenza B M2

The minimum unfavorable contacts approach was then applied on the BM2 protein. BM2 differs greatly from M2 in terms of the number of hydrophilic residues present in the TMD, and is therefore a suitable system for critically evaluating the validity of the approach. While M2 has only two hydrophilic residues in its TMD, BM2 has as many as ten hydrophilic residues (this is without consideration of the terminal three residues on the two sides of the TMD, which correspond to the solvent/lipid headgroup region). Figure 5.8 (A) shows a polar plot for the monomeric form of BM2 depicting the orientation angle and normalized SASA for the residues in the TMD. The residues Ser11 and Asn29 lie on a face of the helix that has very few hydrophilic residues, so this face is not likely to face the interior of the channel. The opposite face of the channel, on the other hand, has a number of hydrophilic residues. To ensure that the maximum possible hydrophilic surface area of the channel is buried, "clusters" of hydrophilic residues were identified. A cluster of hydrophilic residues was defined by selecting those residues on the polar plot that that lied within a cutoff angle of 90°. Two possible clusters were identified in this way: one with His19, Ser12, Gln30, Ser16, Ser9, and His27, and the other with Ser16, Ser9, His27, Thr24, and Gln6. Adding up the normalized SASA for the residues constituting the former cluster gives a total surface area of (0.78+0.35+0.69+0.41+0.28+0.47) = 2.98, and those for the latter cluster give a total of (0.41+0.28+0.47+0.33+0.53) = 2.02. Since the former cluster has higher surface area, residues belonging to this cluster were made to face the interior of the protein. Figure 5.8 (B) shows the location of the residues belonging to the cluster with the largest surface area on a monomeric form of BM2. The helices were orientated such that the centre of mass of the hydrophilic cluster on a given helix lied on a line joining the center of mass of the helix with the center of mass of the entire oligomer. Figure 5.8 (C) shows the orientation and normalized SASA of residues after orienting and assembling the monomeric forms to get a tetramer, while Figure 5.8 (D) shows the actual structure of this model. This model shall be referred to as "model 1".

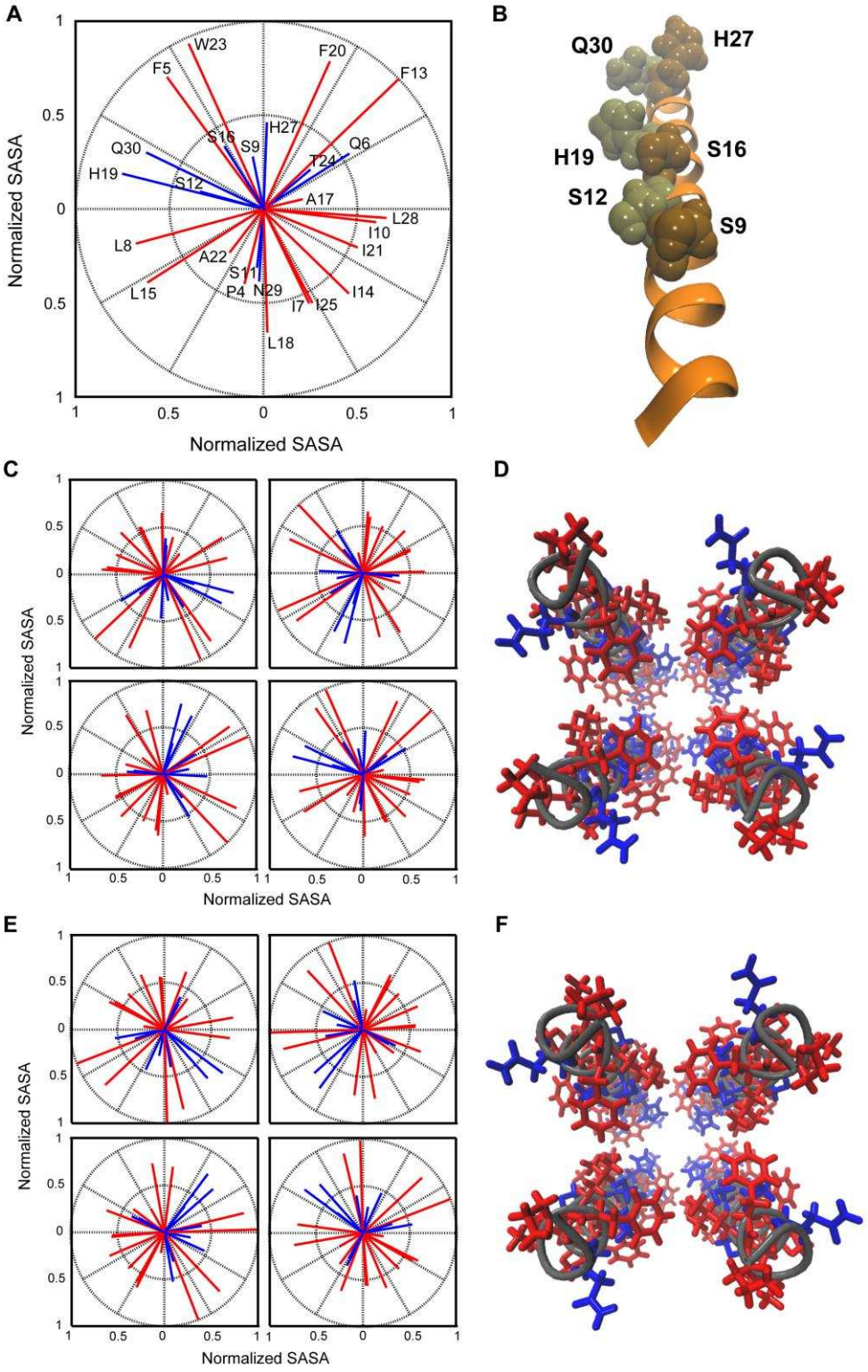

Figure 5.8 (A) Orientation and normalized SASA of the residues in the BM2 monomer. (B) Residues forming part of the hydrophilic cluster with the highest SASA in the monomer. (C,D) Orientation and structure of first BM2 model. (E,F) Orientation and structure of second BM2 model.

An alternate model for the BM2 tetramer was generated by defining a cluster of hydrophilic residues on the polar plot as one that contained residues that form an angle of less than 45° between them (as opposed to 90° for the previous model). A cluster formed by His19, Ser12, Gln30, and Ser16 had the highest surface area among all the possible clusters. A tetramer model was therefore generated that had these four residues facing the interior. The orientation and structure of this model is shown in Figures 5.8 (E) and (F). This model shall be referred to as "model 2".

Both the models were subjected to the two conformational sampling methods by following the protocol described previously for influenza A M2. Probability distributions for the interhelical distance, tilt angle, and the orientation angle of the concerned hydrophilic cluster are shown in Figure 5.9 (the orientation angle of a hydrophilic cluster on a given monomer was calculated by determining the center of mass of all the residue sidechains in the cluster taken together. The angle between the center of mass of the hydrophilic cluster, the center of mass of the concerned monomer, and the center of mass of the entire oligomer gave the orientation angle). The figures on the left side correspond to model 1, and the figures on the right correspond to model 2. In general, the distributions for conformations sampled using the multiple simulation approach are unimodal, while the distributions for REMD are bimodal. Similar to the observation for M2, two distinct conformations of BM2 are seen in the REMD simulations. Another important observation is that, the probability distributions for the two BM2 models differ from each other in terms of the location of the peaks, suggesting that the conformations sampled from the two BM2 models are different. This shows that the initial orientation of the residues is a major factor in determining which conformations are finally sampled.

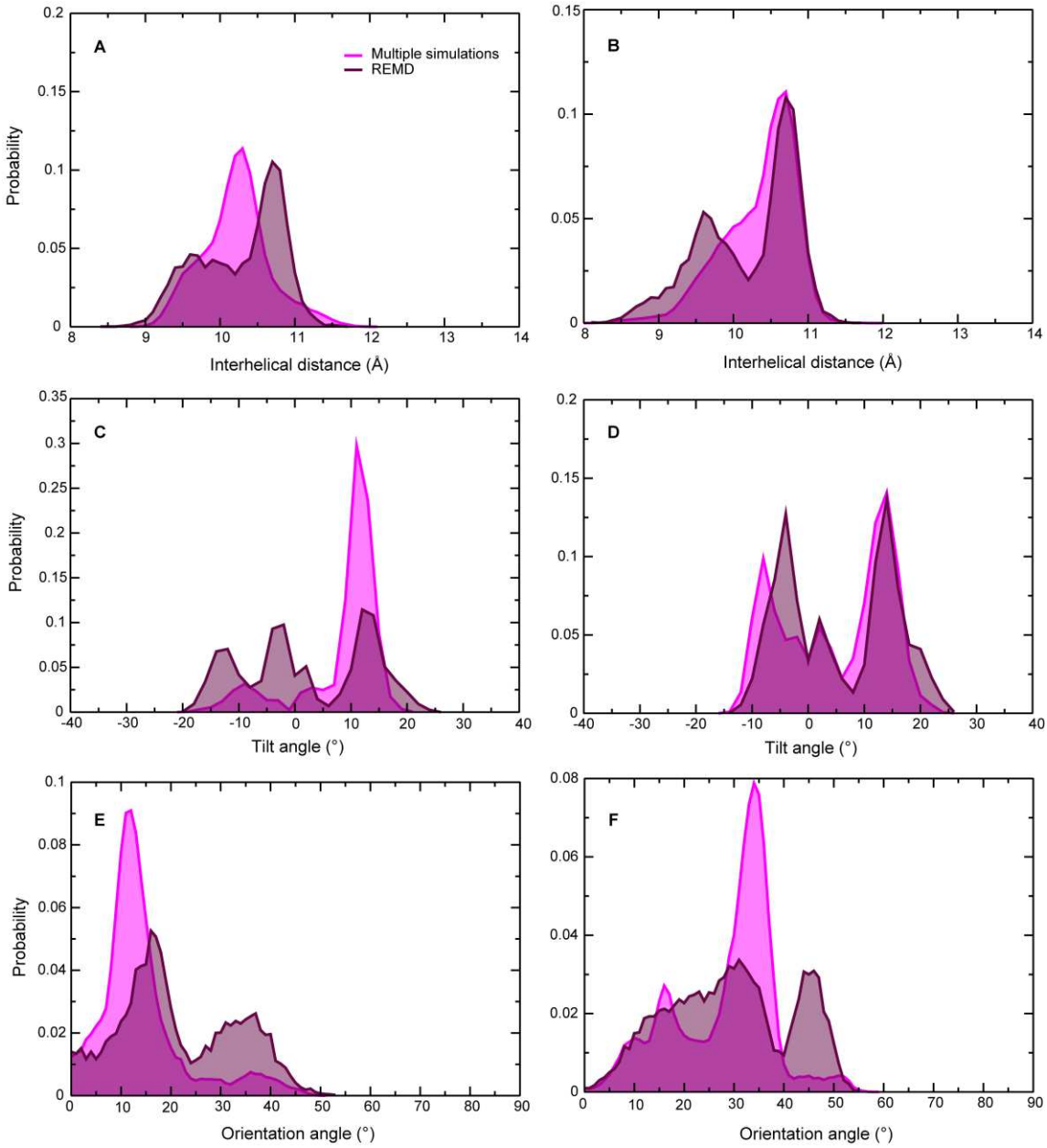

Figure 5.9 Probability distributions for the structural parameters over the course of the two sampling approaches applied to BM2. (A,B) Interhelical distance for the two models. (C,D) Tilt angle for the two models. (E,F) Orientation angle of the hydrophilic cluster for the two models. In calculating the orientation angle for a hydrophilic cluster, the center of mass of the hydrophilic cluster was determined by considering the sidechains of all the residues in the cluster taken together.

Upon following the clustering method described earlier, four representative structures were obtained for BM2: two structures each for the two BM2 models (one for multiple simulation approach and one for REMD). The RMSD of the representative structures with respect to an

experimental structure of BM2 (PDB code: 2KIX) [32] are as follows: 4.04 Å (from multiple simulations) and 3.51 Å (from REMD) for the structures obtained from model 1, and 3.64 Å (multiple simulations) and 4.76 Å (REMD) for the structures obtained from model 2. The representative structures are thus reasonably close to the experimental structure. Figure 5.10 (A) shows the predicted structures obtained from model 1; most of the hydrophilic residues (colored blue) are seen to face the pore. Figure 5.10 (B) shows the two predicted structures obtained from model 1 superposed with the experimental structure, and Figure 5.10 (C) shows contact maps for the experimental structure and the two predicted structures. The predicted models are seen to resemble the experimental structure both in terms of the structure of the backbone, and in terms of the contacts formed by the residues at the interhelical interface. For evaluating the stability of the predicted structures, the two structures obtained from model 1 of BM2 were subjected to equilibration in an explicit bilayer environment. Figure 5.10 (D) shows the RMSD of the two structures with respect to the experimental structure over the course of the equilibration. Both the structures are seen to retain their integrity in an explicit membrane. In fact, the structures are seen to be closer to the experimental structure after the equilibration. This suggests that equilibration in an explicit lipid bilayer could be used for refining the structures obtained after the clustering step.

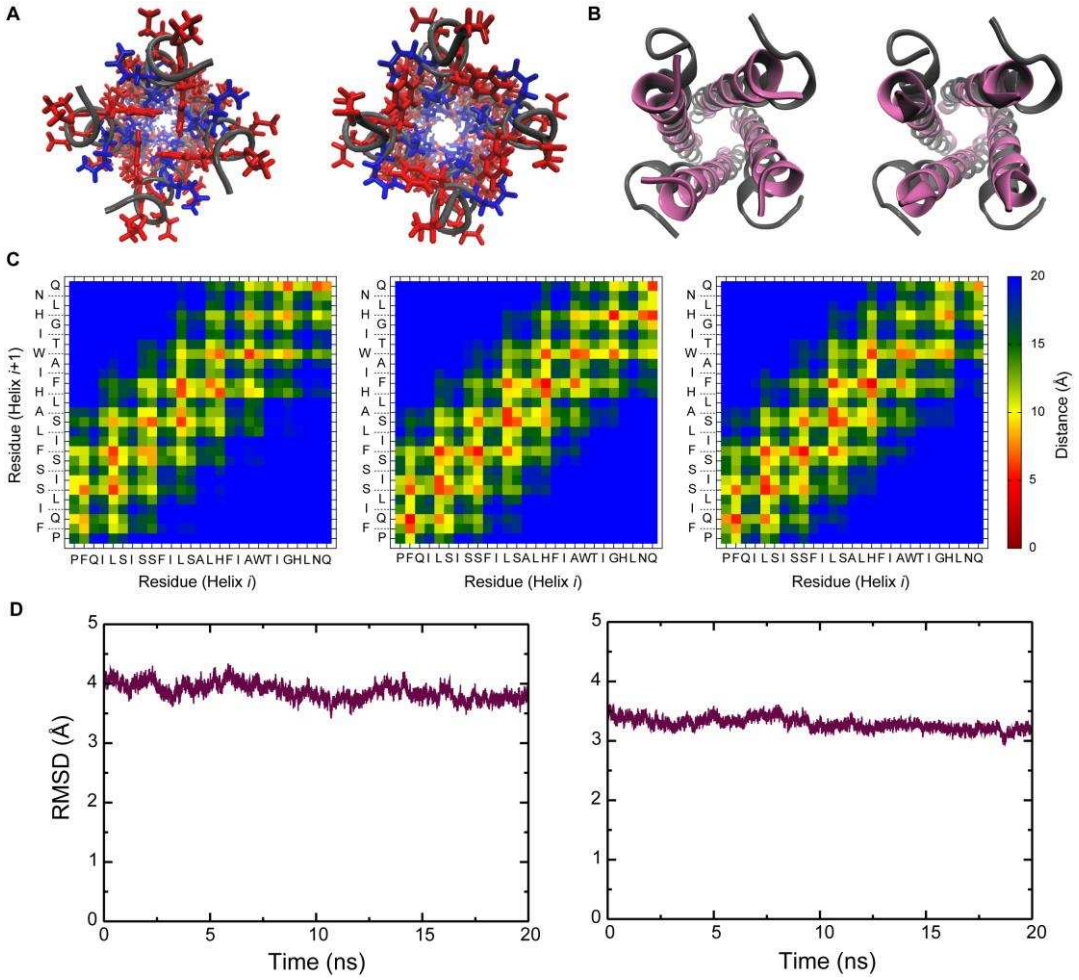

Figure 5.10 (A) The two predicted structures obtained from BM2 model 1. The structure on the left is from the multiple simulation approach, and the structure on the right is from REMD. (B) The two predicted structures aligned with the experimental structure. (C) Contact maps for the experimental structure (left), the predicted structure from multiple simulation approach (center), and the predicted structure from REMD approach (right). (D) RMSD of the two predicted models with respect to the experimental structure over the course of equilibration in explicit membrane environment.

5.3.7 ErbB2 dimer

The two proteins for which the structure has been successfully predicted are tetramers. To investigate if the method is applicable to a protein belonging to a different class of membrane proteins, and existing in a different oligomeric state, the dimeric form of the receptor tyrosine kinase ErbB2 was modeled. Only the TMD of the protein has been considered here, and the cytoplasmic domain has been omitted. A structure with minimum exposure of the hydrophilic residues to the hydrophobic core of the membrane was modeled by orienting the helices such

that the residues Thr652 and Ser656 were buried in the interior (Figure 5.11 (A, B)), and this structure was subjected to the two conformational sampling methods. The probability distributions for the interhelical distance, tilt angle, and orientation angle of the hydrophilic cluster are shown in Figure 5.11 (C, D, E). It must be mentioned that the analysis for the tilt angle of the helices in ErbB2 does not account for left-handedness and right-handedness, since it was not possible to determine the handedness of a number of sampled conformations, owing to a kink in the middle of the helix (data not shown). All tilt angle values therefore carry a positive sign. The probability distributions for the two sampling approaches are seen to be similar to each other for ErbB2, unlike M2 and BM2. Clustering of the structures from the two approaches gave two representative structures that differed from an experimental structure of the protein (PDB code: 2JWA) [33] by RMSDs of 5.68 Å (multiple simulations) and 5.70 Å (REMD). A superposition of the predicted structures with the experimental structure is shown in Figure 5.11 (F). The predicted structures differ from the experimental structures in that they possess a kink in the middle of the helix, which is absent in the experimental structure. Furthermore, the tilt angles of the helices in the predicted structures are different from the experimental structure.

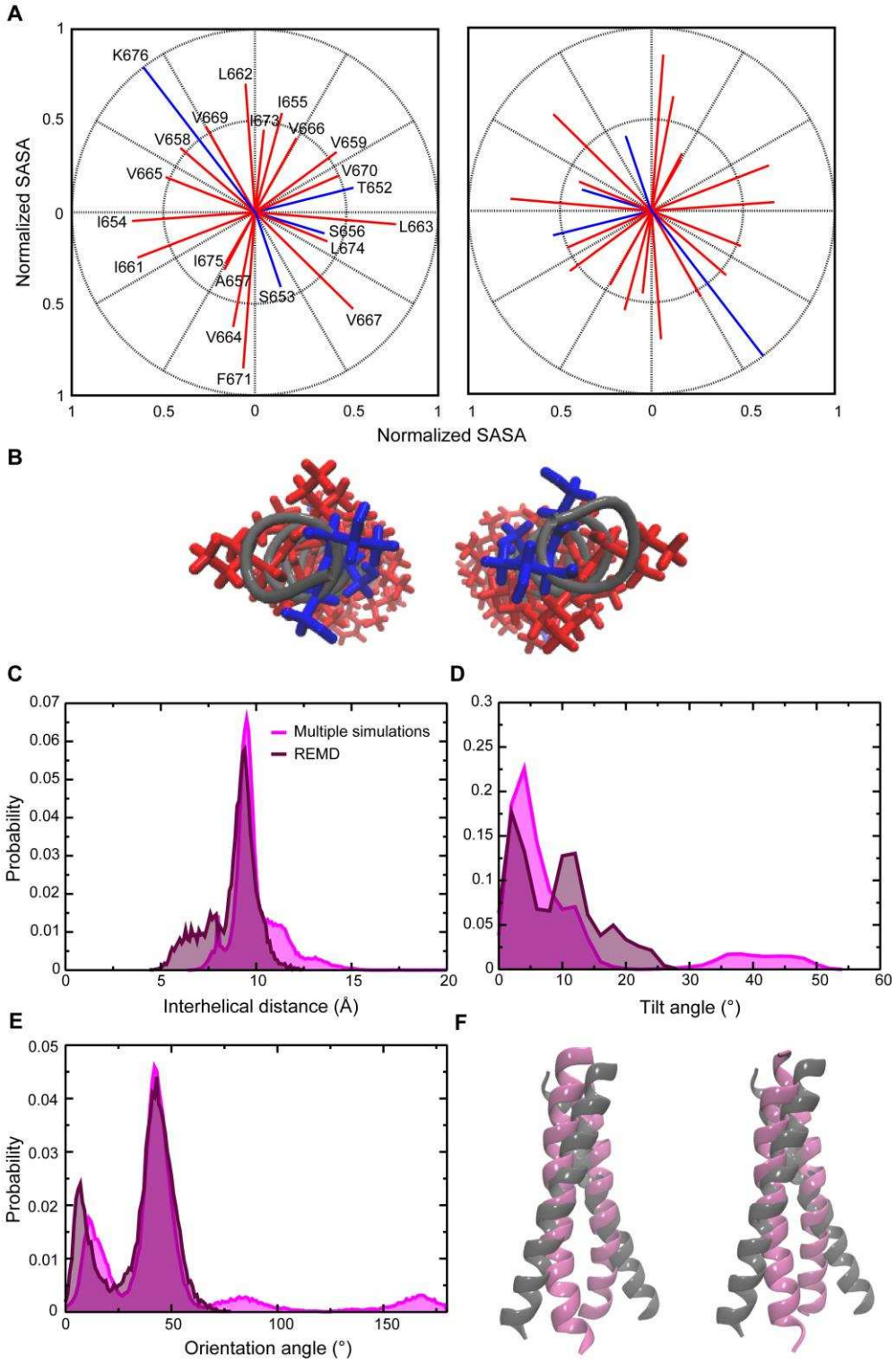

Figure 5.11 (A) Orientation of residues in the initial dimer model for ErbB2. (B) Structure of the model. (C,D,E) Probability distributions of interhelical distance, tilt angle, and orientation angle of hydrophilic cluster, respectively, during the multiple simulations and REMD for ErbB2. (F) Predicted structures (mauve) overlaid with experimental structure (gray). The structure on the left is from multiple simulations, and the one on the right is from REMD.

5.3.8 Alternative models for ErbB2

The fact that the predicted structures for M2 and BM2 are closer to the experimental structure than that for ErbB2 suggests that the minimum unfavorable contacts approach works better for channel-like proteins than for receptors. Since channel-like proteins usually have a hydrophilic pore, they are expected to possess a number of hydrophilic residues facing the interior of the channel. Thus, optimizing the orientation of the residues such that a maximum number of hydrophilic residues face the interior is likely to provide an optimized starting conformation. On the other hand, receptors may not necessarily possess hydrophilic residues in the interhelical interface. This suggests a need for improvements in the unfavorable contacts approach for modeling these classes of proteins. One important factor that could be considered is the orientation of the hydrophobic residues. Apart from ensuring that the contact between hydrophilic residues and the hydrophobic lipid tails is minimized, the unfavorable contacts approach could be modified to optimize the contacts formed by hydrophobic residues with one another, and with the hydrophobic core of the membrane.

An alternative approach was employed to assemble the ErbB2 dimer, and this involved optimizing hydrophobic packing in the interhelical interface of the dimer, as shown in Figure 5.12 (A). The optimized dimer was subjected to multiple simulations using the protocol described before. Figure 5.12 (B) shows the RMSD of the structures sampled during these simulations with respect to the experimental structure. It can be seen that the sampled structures are not very close to the experimental structure in any of the simulations. Visual inspection of these structures revealed that they do not possess the kind of helical tilt that is seen in ErbB2. Since helical tilt arises from hydrophobic mismatch between the protein and the membrane, it follows that the modeling approach is not able to accurately capture the effect of hydrophobic mismatch for dimeric proteins like ErbB2 (the tilt predicted for the tetrameric M2 and BM2 was close to the that predicted by experiments). This necessitates modifications in the approach for dimeric proteins, which have a relatively large surface area exposed to the membrane/solvent compared to the higher oligomeric forms. The use of restraints in determination of experimental structures, for example, the use of intermonomeric distance restraints in solving the structure of ErbB2 [33], could be another reason why this approach does not work for ErbB2.

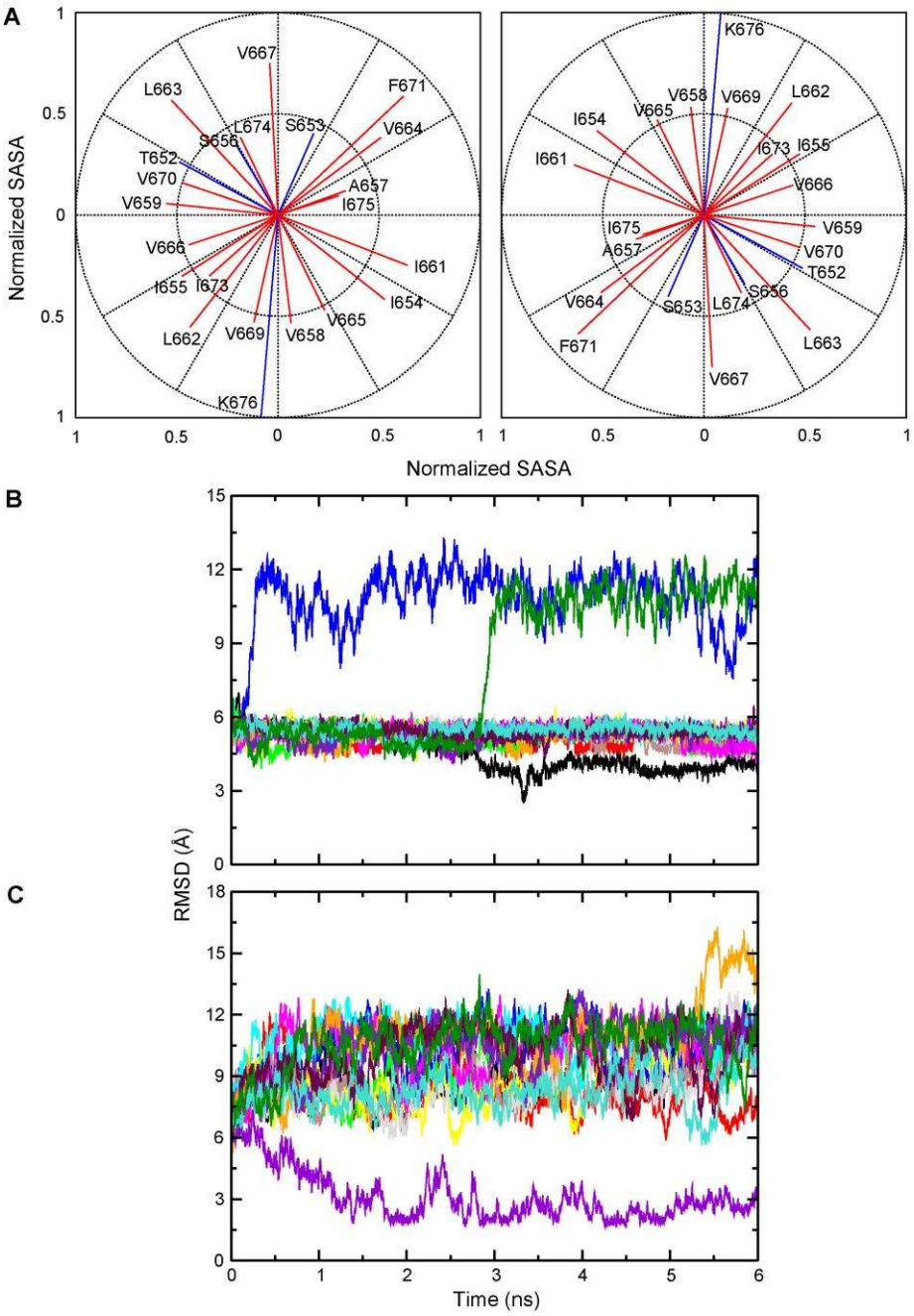

Figure 5.12 (A) Orientation of residues in the ErbB2 model with optimal hydrophobic packing. (B, C) Backbone RMSD of the sampled structures with respect to the experimental structure during the multiple simulations. Each line in the graph corresponds to an independent simulation. While (B) shows the simulations that were performed without an initial tilt, (C) shows the simulations that were performed with an initial tilt in the structure.

In order to overcome the limitations described above, yet another structural model was developed for ErbB2. A monomeric form of the protein was initially equilibrated in an implicit membrane for 5 ns, and the average of the tilt angle observed for this monomer over the last 4 ns

of simulation was used as the initial tilt angle for the dimeric form. Thus, the dimer was assembled by first orienting the individual monomers such that they have the orientation shown in Figure 5.12 (A), followed by assigning them the tilt estimated from the monomer simulation. This dimer model was subjected to mulltiple simulations, and the RMSD of the sampled structures with respect to the experimental structure is shown in Figure 5.12 (C). The sampled structures are close to the experimental structure in only one out of the 15 simulations. Although there are a greater number of sampled structures resembling the experimental structure compared to the simulations on the previous two models for ErbB2, the method still does not work well for the protein. The difficulty in modeling ErbB2 might be due to the lower interhelical contact area seen in dimers compared to higher oligomeric forms. Since the modeling approach works by optimizing contacts, it does not work well for proteins that have a less pronounced interhelical contact area. It would be noteworthy to apply the method on bigger receptor proteins like G protein-coupled receptors (GPCRs), which have significantly higher interhelical contacts.

5.3.9 Control simulations to test the minimum unfavorable contacts approach

The structure modeling approach described above is able to predict the structures of channels reasonably well. A major step in the modeling is determining the initial orientation of the transmembrane helices with maximum hydrophilic surface area buried in the interior. In order to test whether this step indeed plays a major role in the outcome of the approach, control simulations were performed where the maximum hydrophilic surface area was exposed to the hydrophobic exterior rather than the interior. Figure 5.13 (A) shows the model for M2, which has the His37 residue oriented such that it completely faces the membrane. Multiple simulations were performed on this model, followed by filtering and clustering to obtain the predicted structure. The predicted structure is shown overlaid with the experimental structure in Figure 5.13 (B). Although the predicted structure and experimental structures differ by a backbone RMSD of 4.77 Å, there is a notable difference in the orientation of the functionally important His37 residue. The residue faces the exterior in the predicted structure, which would prevent it from playing a role in the proton conduction activity of the channel. This reaffirms the role of the contact optimization step in providing more accurate structural models.

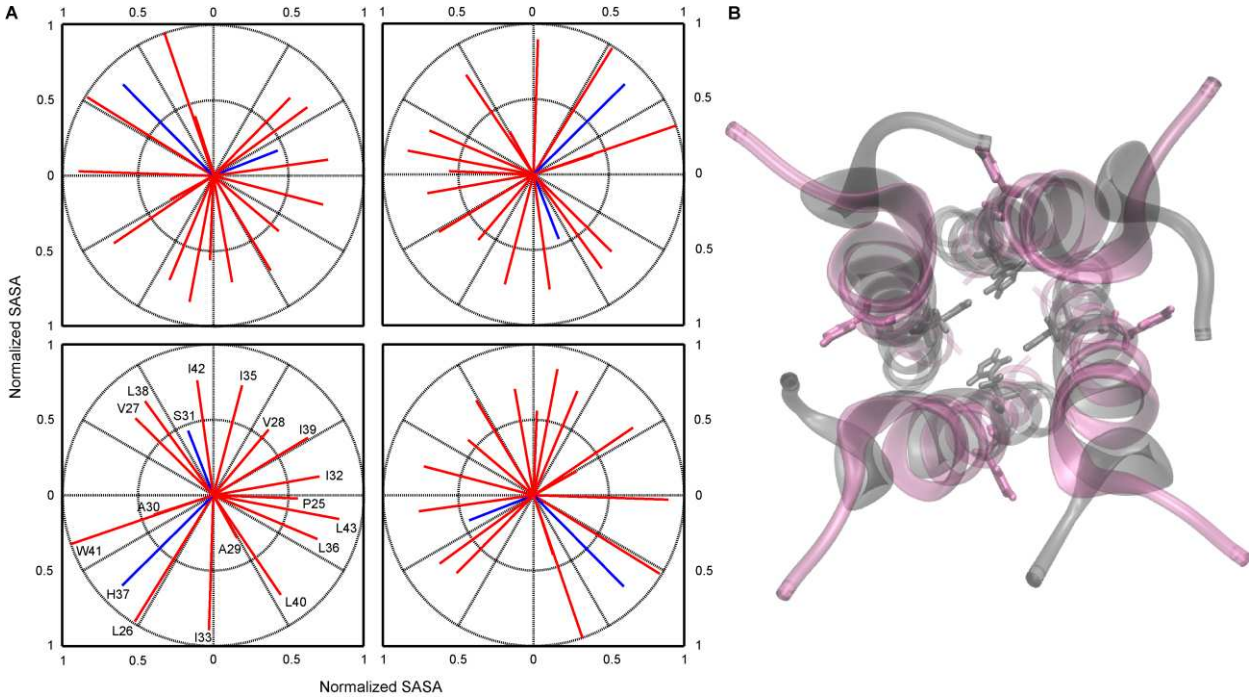

Figure 5.13 (A) Polar plot for the M2 model with maximum hydrophilic surface exposed to the exterior. (B) The predicted structure (in mauve color) overlaid with the experimental structure (gray). The His37 residues are shown in licorice representation.

5.4 Conclusions

The study suggests a novel protocol for modeling the initial structures of transmembrane helices prior to the use of any conformational sampling approach, using the influenza A M2 protein, influenza B M2 protein, and the receptor kinase ErbB2 as model systems. The helices have been oriented so that unfavorable contacts of the hydrophilic residues with the hydrophobic membrane environment are minimized. Starting from this initial structure, two approaches have been used for sampling the conformational space of the protein. The first method involves the use of multiple simulations with different initial velocities, while the second uses REMD. The sampling in both the approaches was performed in an implicit membrane environment to mimic the nature of a lipid bilayer. Conformations sampled from the two approaches have been filtered on the basis of three geometrical parameters: interhelical distance, tilt angle, and spin angle. For influenza A M2 and influenza B M2, representative structures selected from among these filtered structures are remarkably close to experimental structures, as indicated by RMSD and interhelical contact profiles. Furthermore, the stability of the predicted structures has been tested

in an explicitly hydrated lipid bilayer. The fact that there is a close resemblance of the representative structures obtained from the two approaches with the experimental structure suggests that the initial arrangement of the hydrophilic residues makes possible the sampling of conformations close to the native state, irrespective of the methodology used for sampling conformations. On the other hand, the predicted structure for ErbB2 differs from the experimental structure in terms of a kink and the helical tilt angle, suggesting that the method works better for channel-like proteins than for receptors. A possible reason why the predicted structures for the dimeric ErbB2 are different from the actual experimental structure could be the less pronounced interhelical contact area seen in dimeric membrane proteins. Since a fundamental step in this modeling approach is optimization of contacts, it works best when there is considerable interhelical contact.

References

1. Rask-Andersen M, Almén MS, Schiöth HB (2011) Trends in the exploitation of novel drug targets. Nat Rev Drug Discov 10(8):579–590.
2. Luckey M (2008) Membrane Structural Biology: With Biochemical and Biophysical Foundations Cambridge University Press, New York.
3. Deniaud A, Moiseeva E, Gordeliy V Pebay-Peyroula E (2010) In Membrane Protein Structure Determination: Methods and Protocols Lacapére JJ, Ed.; Methods in Molecular Biology Series, Springer Protocols, Humana Press.
4. Bigelow HR, Petrey DS, Liu J, Przybylski D, Rost B (2004) Predicting transmembrane beta-barrels in proteomes. Nucleic Acids Res 32(8):2566–2577.
5. Adams PD, Arkin IT, Engelman DM, Brünger AT (1995) Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. Nat Struct Biol 2(2):154–162.
6. Adams PD, Engelman DM, Brünger AT (1996) Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. Proteins 26(3):257–261.
7. Park Y, Elsner M, Staritzbichler R, Helms V (2004) Novel scoring function for modeling structures of oligomers of transmembrane α-helices. Proteins 57(3):577–585.

8. Mottamal M, Zhang J, Lazaridis T (2006) Energetics of the native and non-native states of the glycophorin transmembrane helix dimer. Proteins 62(4):996–1009.
9. Bordner AJ, Zorman B, Abagyan R (2011) Efficient molecular mechanics simulations of the folding, orientation, and assembly of peptides in lipid bilayers using an implicit atomic solvation model. J Comput Aided Mol Des 25(10):895–911.
10. Yarov-Yarovoy V, Schonbrun J, Baker D (2005) Multipass membrane protein structure prediction using Rosetta. Proteins 62(4):1010–1025.
11. Kukol A, Adams PD, Rice LM, Brunger AT, Arkin IT (1999) Experimentally based orientational refinement of membrane protein models: a structure for the Influenza A M2 H^+ channel. J Mol Biol 286(3):951–962.
12. Kukol A, Arkin IT (1999) vpu transmembrane peptide structure obtained by site-specific Fourier transform infrared dichroism and global molecular dynamics searching. Biophys J 77(3):1594–1601.
13. Pappu RV, Marshall GR, Ponder JW (1999) A potential smoothing algorithm accurately predicts transmembrane helix packing Nat Struct Biol 6:50-55.
14. Potluri S, Yan AK, Chou JJ, Donald BR, Bailey-Kellogg C (2006) Structure determination of symmetric homo-oligomers by a complete search of symmetry configuration space, using NMR restraints and van der Waals packing. Proteins 65(1):203–219.
15. Barth P, Wallner B, Baker D (2009) Prediction of membrane protein structures with complex topologies using limited constraints. Proc Natl Acad Sci USA 106(5):1409–1414.
16. Kim S, Chamberlain AK, Bowie JU (2003) A Simple Method for Modeling Transmembrane Helix Oligomers. J Mol Biol 329(4):831–840.
17. Kokubo H, Okamoto Y (2004) Prediction of membrane protein structures by replica-exchange Monte Carlo simulations: Case of two helices. J Chem Phys 120(22):10837.
18. Carpenter T, Bond PJ, Khalid S, Sansom MS (2008) Self-assembly of a simple membrane protein: Coarse-grained molecular dynamics simulations of the influenza M2 channel. Biophys J 95(8):3790–3801.
19. Patargias G, Zitzmann N, Dwek R, Fischer WB (2006) Protein−protein interactions: Modeling the hepatitis C virus ion channel p7. J Med Chem 49(2):648–655.
20. Krüger J., Fischer WB (2009) Assembly of viral membrane proteins. J Chem Theory Comput 5(9):2503–2513.

21. Hsu H-J, Fischer WB (2011) In silico investigations of possible routes of assembly of ORF 3a from SARS-CoV. J Mol Model 18(2):501–514.
22. Li L-H, Hsu H-J, Fischer WB (2013) Assembling viral channel forming proteins: Vpu from HIV-1. Biopolymers 99(8):517–529.
23. Bu L, Im W, Brooks CL (2007) Membrane assembly of simple helix homo-oligomers studied via molecular dynamics simulations. Biophys J 92(3):854–863.
24. Im W, Feig M, Brooks CL (2003) An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. Biophys J 85(5):2900–2918.
25. Im W, Lee MS, Brooks CL (2003) Generalized born model with a simple smoothing function. J Comput Chem 24(14):1691–1702.
26. Padhi S, Khan N, Jameel S, Priyakumar UD (2013) Molecular dynamics simulations reveal the HIV-1 Vpu transmembrane protein to form stable pentamers. PLoS ONE 8(11):e79779.
27. Pettersen EF, et al. (2004) UCSF Chimera-A visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612.
28. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. J Comput Chem 30(10):1545–1614.
29. Sugrue R, Hay A (1991) Structural characteristics of the M2 protein of influenza a viruses: Evidence that it forms a tetrameric channel. Virology 180(2):617–624.
30. Holsinger LJ, Alams R (1991) Influenza virus M2 integral membrane protein is a homotetramer stabilized by formation of disulfide bonds. Virology 183(1):32–43.
31. Balannik V, Lamb RA, Pinto LH (2007) The oligomeric state of the active BM2 ion channel protein of influenza B virus. J Biol Chem 283(8):4895–4904.
32. Wang J, Pielak RM, McClintock MA, Chou JJ (2009) Solution structure and functional analysis of the influenza B proton channel. Nat Struct Mol Biol 16(12):1267–1271.
33. Bocharov EV, et al. (2008) Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. J Biol Chem 283(11):6950–6956.
34. Best RB, et al. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. J Chem Theory Comput 8(9):3257–3273.

35. MacKerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25(11):1400–1415.
36. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616.
37. Ryckaert JP, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327-341.
38. Jo S, Kim T, Im W (2007) Automated builder and database of protein/membrane complexes for molecular dynamics simulations. PLoS ONE 2(9):e880.
39. Jo S, Lim JB, Klauda JB, Im W (2009) CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. Biophys J 97(1):50–58.
40. Essmann U, et al. (1995). A smooth particle mesh Ewald method. J Chem Phys 103(19):8577-8593.
41. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802.
42. Klauda JB, et al. (2010) Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. J Phys Chem B 114(23):7830–7843.
43. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926-935.

6 Ion Conduction through the p7 Channel

6.1 Introduction

Ion channels in cell membranes provide passages that are conducive for ion transport across the otherwise hydrophobic membrane, while at the same time ensuring that only certain specific ions are able to pass [1,2]. These channels exhibit differences in their conduction rates and selectivity, and molecular simulation studies have shown that different channels possess different mechanisms governing conduction and selectivity [3-6]. An ion channel that has drawn attention is the cation-selective p7 viroporin from hepatitis C virus [7-10], owing to its role in the propagation of the virus [11-13]. Several experimental and theoretical studies to resolve the structure of p7 have been reported. A number of nuclear magnetic resonance (NMR) spectroscopy studies have shown that the transmembrane region of the protein is primarily made up of α-helices [14-18]. The protein is a double-pass transmembrane protein, with the two transmembrane segments linked by a loop region [15-17]. The loop region allows the two transmembrane segments to orient with respect to each other, thereby providing flexibility to the protein [15]. Attempts to generate a model for the oligomeric form of the protein have been made in a number of molecular modeling studies [19-22].

A couple of recent studies have reported an unusual structure of the channel, making the channel an attractive system for investigating ion permeation [23-25]. The channel is seen to be a flower-shaped hexamer, with each monomer consisting of three α-helices [23], as shown in Figure 6.1 (A) (PDB code: 2M6X). The first two α-helices from each monomeric unit form the pore, while the third helix faces the membrane. As far as intramolecular contacts are concerned, contacting interfaces identified from the oligomeric models of the protein differ from those identified in the monomeric structures of the protein. While the structure for the monomeric form shows contacts between the second and third helices on a given monomer [15], the oligomeric form shows the third helix of a given monomer i forming contacts with the second helix of the i+2 monomer and first helix of the i+3 monomer [23]. This suggests that intramolecular interactions that stabilize the monomeric form of the protein are different from those that stabilize the oligomeric form. The entrance of the pore has a ring of arginine residues, and is followed by a hydrophobic stretch

made of W21, L24, and L28 (Figure 6.1 (B)). There is a hydrophilic stretch further inside the pore, and a ring of isoleucine residues at the exit of the pore.

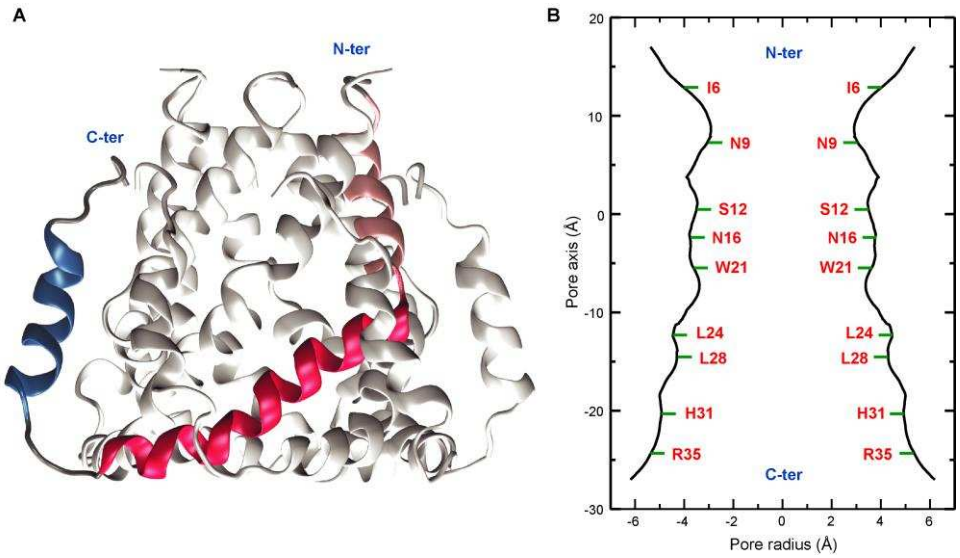

Figure 6.1 (A) The hexameric p7 channel shown with three helices from one of the monomers in different colors. (B) The pore radius shown with approximate positions of the pore-lining residues occurring at the corresponding regions.

Although an electrophysiology study proposed that the histidine residue at position 17 is pore-lining [26], based on the sensitivity of the channel towards pH, the recent NMR structure of the oligomeric form shows that this histidine partly faces the membrane, and partly the intermonomer interface [23]. This suggests that p7 is capable of ion conduction irrespective of the protonation state of the histidine; in fact, a number of studies have demonstrated the ion channel activity of p7 at a pH of about 7.4, which is close to neutral [9,10,15,23,27]. The channel activity of the protein has been shown to be blocked by adamantane derivatives in a number of studies [23,28-30]. The binding of adamantane derivatives to the membrane-facing side of the channel causes conformational changes, leading to a closed state [23]. Since the binding occurs close to the pore-lining Trp21 residue, it follows that proper orientation of the tryptophan residue is important in the conduction process. Ion transport through the channel is hypothesized to proceed from the wider intracellular side to the narrower extracellular side [23]. The channel is more selective toward monovalent rather than divalent cations [10], and the mechanism via which the channel achieves this selectivity remains elusive.

Umbrella sampling free energy calculations are invaluable for obtaining atomistic detailed insights into mechanisms that render selectivity to ion channels [3,4,31-35]. Typically, it is seen that the conduction of certain specific ions is carried out swiftly in a concerted fashion, while prohibiting the permeation of other ions. Roux and coworkers were the first to study multi-ion permeation using umbrella sampling simulations, and they showed that ion conduction through the K^+-selective KcsA channel proceeds via an ion-ion knock-on [3]. A very recent study showed that knock-on in K^+ channels proceeds via direct interactions between the ions, rather than via bridging water molecules [36]. The present study uses umbrella sampling molecular dynamics simulations to model both single ion and two ion permeation through the p7 channel. While the single ion permeation study identifies a hydrophobic barrier as a significant factor in determining the selectivity and kinetics of ion transport through the channel, the two ion permeation study provides a concerted mechanism of ion transport, where knock-on between ions makes possible a continuous flow of ions across the pore.

6.2 Methods

6.2.1 Single ion permeation

All molecular dynamics (MD) simulations reported here were carried out using the NMR structure of the p7 channel with PDB code 2M6X [23]. The channel was modeled in a hydrated bilayer of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) molecules using the CHARMM-GUI Membrane Builder [37,38] such that the axis of the pore was aligned along the z-axis, resulting in a system with a total of 68092 atoms. All the histidine residues in the protein, including the pore-lining His31, were modeled in a neutral form. His31 was not protonated in spite of being a pore-lining residue, since it has been suggested that the residue is not involved in pH sensing [15,39,40], and additionally His31 is very close to Arg35 (see Figure 6.1 (B)). Initially, equilibration of the system was performed in the presence of positional restraints on the protein heavy atoms and planar restraints on lipid headgroups to preserve their planar arrangement. The restraints were gradually reduced to zero over a period of 1 ns, with the NVT ensemble being used initially (50 ps with 1 fs timestep) followed by MD simulations in the NPAT ensemble (950 ps with 1 fs time step initially and 2 fs later). This was followed by 10 ns of equilibration in the NPAT ensemble without any restraints, using a timestep of 2 fs. The

results of this MD simulation were used for generating the initial structures for umbrella sampling. Lennard-Jones interactions were truncated by using a switching function between 10 Å and 12 Å, and long range electrostatic forces were computed using the particle mesh Ewald method [41]. The SHAKE algorithm was used to constrain the length of covalent bonds involving hydrogen atoms [42]. Langevin dynamics was employed to maintain the temperature at 303.15 K.

The structure obtained at the end of the above simulation was used as the initial structure for the umbrella sampling simulations, which used the position of the ion along the axis of the pore as the reaction coordinate. Two independent sets of calculations were performed for investigating the transport of K^+ and Ca^{2+} ions through the channel. Configurations with the concerned ion occurring at different positions were generated by replacing a water molecule occurring at the respective position with the concerned ion. The structures thereby generated were used as the initial configurations for the different windows in the umbrella sampling, and a harmonic potential with force constant 10 kcal mol^{-1} $Å^{-2}$ acting on the z-coordinate of an ion was used. During these simulations, other ions were excluded from the pore region using a repulsive potential with force constant 10 kcal mol^{-1} $Å^{-2}$. For windows corresponding to the ion being present in bulk water, a cylindrical restraint with radius 8 Å and force constant 10 kcal mol^{-1} $Å^{-2}$ was used to restrict the motion of the ion in the xy-plane. The drifting of the protein was restricted by applying a positional restraint (force constant 10 kcal mol^{-1} $Å^{-2}$) on its center of mass. The other simulation parameters used during the umbrella sampling calculations were the same as those used during the equilibration simulations described above. Each window was simulated for 1.1 ns, with the first 100 ps being considered as equilibration period. A total of 129 windows were sampled, so the cumulative simulation time was (129*1.1) ≈ 142 ns each for K^+ and Ca^{2+}.

The weighted histogram analysis method (WHAM) [43-45] was used to compute the potential of mean force (PMF) for the transport of the ion along the reaction coordinate using the relations described in Chapter 1. For calculating 2D-PMFs with respect to the position, z, of the ion along the pore axis and the distance r of the ion from the pore axis, histograms were computed for different values of r corresponding to a given value of z. The probabilities $\rho(r; z)$ thus obtained

were used to calculate the free energy given by –RT ln ρ(r; z), to which a constant C was added to ensure that the minimum value of the term for a given value of z was zero. The 1D-PMF, $A_i(z)$, for a given position on the pore axis was then used to compute the PMF for different values of r using the relation

$$A_i(z,r) = A_i(z) - RT \ln \rho(r;z) + C \quad (6.1)$$

This gave a 2D-PMF as a function of z and r. The bin size used was 0.1 Å for z and 0.5 Å for r. Similarly, a 2D-PMF as a function of z and the hydration number was calculated using bin size 0.1 Å for z and 1 for hydration number.

All simulations were performed using the NAMD program [46], with the CHARMM36 all-atom protein force field [47-49], the CHARMM36 all-atom lipid force field [50], optimized parameters for ions [51], and the TIP3P water model [52], and all trajectory analysis was performed using the CHARMM program [53]. Molecular images were rendered using the Visual Molecular Dynamics (VMD) program [54], and pore radii measurements were performed using the HOLE2 program [55].

6.2.2 Two ion permeation

The equilibrated structure obtained above was used to perform umbrella sampling simulations to model two ion permeation. Two ions (two K^+ or two Ca^{2+}) were placed in the pore region, and, for every position of a given ion along the pore axis, windows corresponding to different separations between the two ions were defined. Hence, the two dimensional PMF profiles that are calculated are with respect to the position of upper ion, and the separation between the two ions as the reaction coordinates (see later). The separation between the two ions was defined as the z-axis component of the distance between the two ions (the pore is aligned along the z-axis). A total of 43 x 9 = 387 windows each were sampled for K^+ and Ca^{2+}, with the separation between the two ions in the pore varying from 4 Å to 12 Å (at intervals of 1 Å). A harmonic potential of force constant 10 kcal mol^{-1} $Å^{-2}$ was applied on the two ions to restrain their position along the reaction coordinates, and the entry of other ions into the pore region was prevented by applying a repulsive force with force constant 20 kcal mol^{-1} $Å^{-2}$. For windows corresponding to the ion in bulk water, a cylindrical restraint of radius 8 Å was applied to restrict the lateral

displacement of the ion. A restraint with force constant 10 kcal mol^{-1} $Å^{-2}$ was applied on the center of mass of the protein to prevent an overall drifting.

Each window was sampled for 400 ps, with the first 100 ps being regarded as equilibration period, so that the total simulation time was ~ 155 ns each for K^+ and Ca^{2+}. The simulations were performed in the NPAT ensemble, with the position of the ion being saved every 0.2 ps, and that of the system being saved every 2 ps. The other simulation parameters were the same as those used in the single ion permeation study. The free energy profile for two ion permeation was calculated using the 2D-WHAM implementation of the WHAM program, which reads a distribution of the positions along the pore axis of the two ions [43-45].

The minimum free energy path (MFEP) for the permeation of two ions through the channel was calculated using a smoothening algorithm, which is described as follows. For every bin along the first reaction coordinate (position of upper ion along pore axis), the minimum energy configuration was identified among all configurations along the second reaction coordinate (separation between two ions). Following this, all configurations that had an energy difference of less than 0.5 kcal mol^{-1} relative to the minimum energy configuration for this bin were extracted. Among these extracted structures, a configuration was assigned as being a part of the MFEP such that, in moving from one bin along the first reaction coordinate to the adjacent bin, the change in the separation between the two ions is a minimum.

6.2.3 Two ion permeation with one-dimensional umbrella sampling

A different set of two ion permeation studies were performed with a restraining potential on only one of the ions. These simulations were designed to verify the transport mechanism inferred from the above two-dimensional umbrella sampling studies, which employed restraining potentials on both the ions. The system with two Ca^{2+} ions in the pore was chosen for this investigation, since both the hydrophobic barrier and the knock-on phenomenon would be more pronounced in the case of the doubly charged Ca^{2+} ion. All the windows in the simulations had a separation of 4 Å between the two ions in the initial configuration, and a positional restraint with force constant 10 kcal mol^{-1} $Å^{-2}$ was applied on the “lower” ion (see later), while allowing the

other ion to move freely. All other simulation parameters were the same as those in the two-dimensional umbrella sampling.

6.3 Results and Discussion

6.3.1 Ion permeation might be controlled by short hydrophobic stretches in the pore

Pore radii calculated along the whole stretch of the channel (Figure 6.1 (B)) based on average structures obtained from equilibration simulations (of the above mentioned crystal structure) indicate a wider pore compared to typical ion channels. It can be seen that the C-terminal side lined by a ring of Arg35 residues at the entrance is broader than the N-terminal side. The middle region of the channel includes a hydrophobic stretch comprising of two leucine residues (Leu24 and Leu28) and a tryptophan residue (Trp21) from each of the six monomeric units of the protein. Another relatively shorter hydrophobic region is formed by Ile6 near the N-terminal end. The narrowest region of the channel occurs around Asn9 and Ile6 followed by another constriction around the hydrophobic stretch in the middle of the pore (around Leu24 and Trp21). These two constrictions may possibly play a crucial role in the ion conduction mechanism. Overall, p7 has a wide pore with a largely hydrophilic environment, a feature that is expected to make it a strong ion-conducting channel.

6.3.2 Free energy profiles suggest p7 to be more permeable towards K^+ than Ca^{2+}

Potential of mean force (PMF) profiles for the conduction of K^+ and Ca^{2+} ions through the pore are shown in Figure 6.2. The left to right hand side in the figure corresponds to the movement of the ion from the C-terminal to the N-terminal side. The qualitative features of the two free energy curves corresponding to the permeation of K^+ and Ca^{2+} are very similar to each other (Figure 6.2). As the ion proceeds towards the N-terminal side, two distinct maxima along the reaction coordinate (~ -10 and 12 Å) are observed. The free energy maxima correspond to the two hydrophobic stretches in the pore, both of which are constricted regions. While the lower barrier corresponds to the narrowest constriction (near Ile6), the higher barrier is seen around the second narrowest constriction (near Leu24). The occurrence of two energy maxima near the two hydrophobic regions suggests that ion conduction through the pore is likely to be modulated by a hydrophobic gating mechanism. However, the inverse correlation of the width of the

constrictions and the height of the free energy barriers indicates that additional factors might also be involved in regulating ion conduction. A number of intermediate states are encountered between the two maxima, as indicated by shallow minimum energy states in the PMF. These intermediates are characterized as states where the ion interacts favorably with polar residues, namely, Asn16, Ser12 and Asn9. The free energy profiles were systematically verified and were found to be converged within the timescale of the simulations employed here (Figure 6.3).

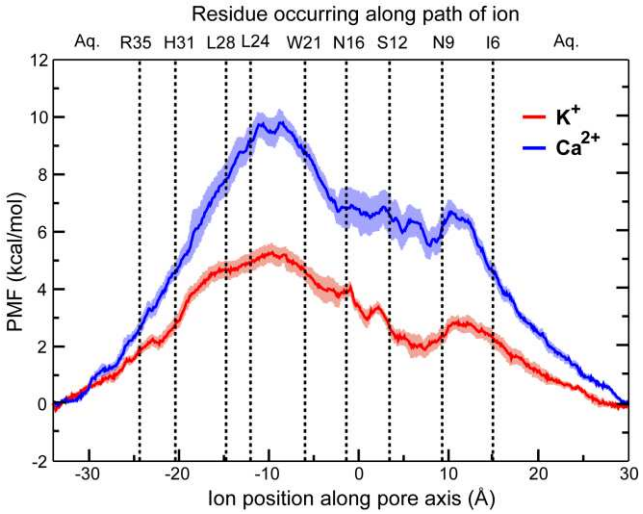

Figure 6.2 The potential of mean force (PMF) for the conduction of K^+ and Ca^{2+} is shown with standard errors. Residues occurring at corresponding regions in the pore lining are also shown. The standard errors have been calculated by dividing the 1 ns sampling data into five blocks of size 200 ps, and then calculating the free energy profile for each of these blocks.

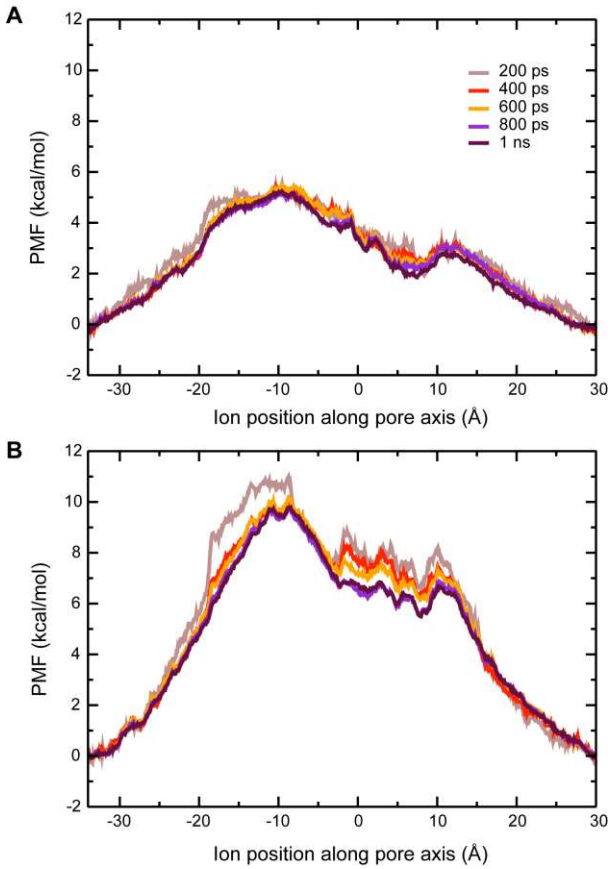

Figure 6.3 Convergence of the umbrella sampling calculations. The potential of mean force (PMF) corresponding to different sampling durations for (A) K^+ and (B) Ca^{2+} is shown.

As the permeating ion in bulk water enters the channel through the C-terminal end, it encounters a ring of Arg35 residues. Electrostatic repulsion between the permeating ion and these residues leads to a rise in the PMF. Once the ion crosses the His31 residues and moves further into the pore, the environment becomes less hydrophilic, leading to an additional rise in the PMF. While the rise in PMF for K^+ ion is gradual, the rise for Ca^{2+} is steep, with the peak of the energy barriers occurring around the Leu24 residue. The energy barrier is higher for Ca^{2+} than for K^+ (5.2 kcal/mol and 9.7 kcal/mol for K^+ and Ca^{2+}, respectively), thus providing the thermodynamic basis for the selectivity of the channel towards K^+ over Ca^{2+} that has been reported previously.[10] Similar to the differentiation of the free energy maxima, the intermediate states lie lower for K^+ compared to Ca^{2+}. Although Ile6 has been previously proposed to serve as a hydrophobic gate [23], the free energy profiles shown here indicate that the stretch around Leu28-Leu24-Trp21 is possibly the rate-limiting factor. Though the region of the channel around Ile6 is more

constricted than that around Leu28-Leu24-Trp21, the barrier is higher around the latter. This phenomenon can be accounted for by considering the residues that occur close to these hydrophobic constrictions. In the C-terminal end, the hydrophobic stretch is preceded by six arginine residues that are positively charged. The higher barrier in this region is the consequence of electrostatic repulsion between the ion and the ring of arginine residues. In contrast, the Ile6 residues are preceded by Asn9 residues to which the permeating cation can favorably bind to. Such a combination of two opposing factors leads to lowering of the barrier for the latter region, which corresponds to the narrowest region of the pore. It is also to be noted that the hydrophobic stretch near the C-terminal end is longer than that at the N-terminal end.

With an energy barrier of 9.7 kcal/mol for the conduction of Ca^{2+}, the channel is not impermeable towards Ca^{2+}, but poses a moderately high barrier for the permeation of the ion. Griffin et al. showed that HIS-tagged p7 is able to conduct Ca^{2+} ions in black lipid membranes [8]; however, the use of HIS tags is known to affect the functioning of channels [56]. A systematic investigation by Premkumar et al. based on conductance studies on p7 embedded in planar lipid bilayer membranes revealed that the conductance for Ca^{2+} is lower than that of K^+ [10]. The present results support the ability of p7 to be selective towards K^+ over Ca^{2+}.

6.3.3 K^+ exhibits more favorable interactions with pore-lining residues

While one-dimensional PMFs can provide insights into the energetics of ion conduction along the axis of the pore, the energetics associated with movements of the ion in a direction perpendicular to the pore axis can be understood by obtaining two-dimensional PMFs corresponding to both these degrees of freedom. Figure 6.4 shows the PMF landscapes for K^+ and Ca^{2+} as a function of the position of the ion along the pore axis and the radial distance of the ion from the axis of the pore (these 2D-PMFs are from the single ion permeation study). In the two landscapes, values close to zero for the radial position correspond to the ion occurring close to the pore axis, while higher values correspond to the ion occurring close to the pore-lining residues. In general, K^+ is seen to sample a wider region perpendicular to the pore axis, whereas Ca^{2+} is seen to occur closer to the pore axis for a large part of its path inside the pore. When the ion is in bulk water (bottom of the graph), it is equally likely to occur at different positions radially away from the pore axis, as suggested by the flat free energy landscape seen in this

region. As the ion enters the broad C-terminal entrance of the channel lined by the positively charged arginine residues, it moves slightly toward the axis of the pore. Upon entering the hydrophobic stretch, the permeating ion prefers to stay away from the pore-lining non-polar residues, so as to minimize unfavorable interactions with the leucine residues. In this stretch, the doubly charged Ca^{2+} is hardly seen to occur in positions close to the pore-lining residues, since the high charge on the ion prevents it from moving close to the hydrophobic lining. Once the ion enters the region formed by Asn16 and Ser12, it moves closer to the pore lining residues, indicating favorable interactions between the ion and these residues. These ion-dipole interactions are reflected in a fall in energy in the most probable path of the ion at this position. Around Asn16 and Ser12, K^+ is closer to the channel residues than Ca^{2+}. The Ca^{2+} is stabilized by coordinating water molecules in this region, and it retains a greater number of coordinating water molecules as it proceeds (see later). Around Asn9, the permeating ion moves slightly towards the pore axis, owing to the constriction of the pore at this region. As the channel widens slightly towards the N-terminal opening, the ion is able to move slightly away from the pore axis.

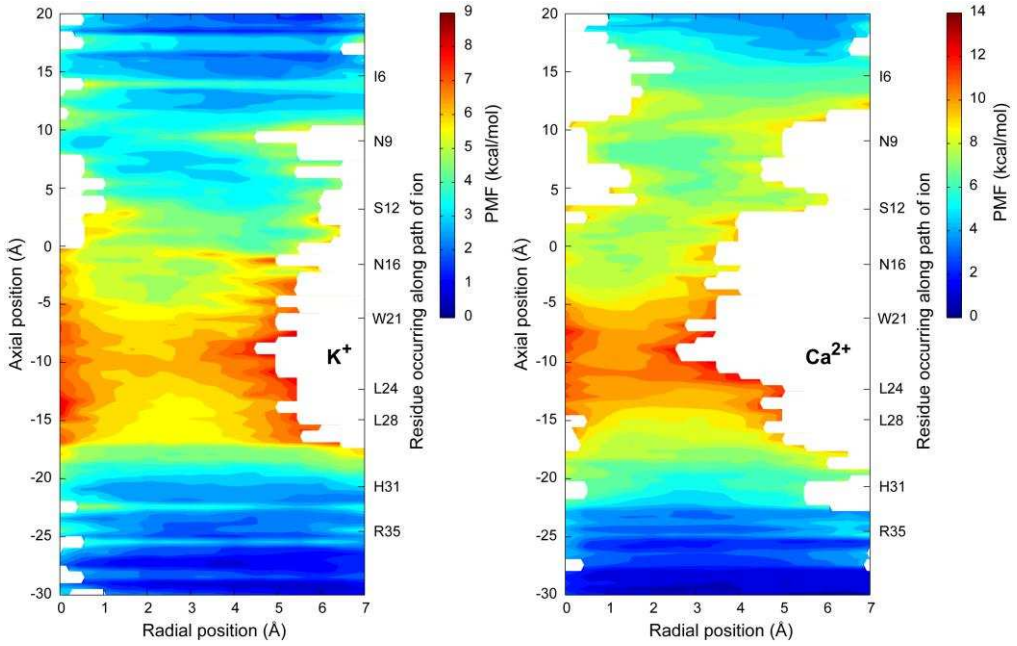

Figure 6.4 Two-dimensional PMF of the permeating ion as a function of the position of the ion along the pore axis and the radial distance of the ion from the pore axis. Only windows corresponding to the ion inside the pore region are shown in the two-dimensional PMF, and windows corresponding to the ion in bulk water have been omitted.

6.3.4 Selectivity towards K^+ rather than Ca^{2+} is provided by a hydrophobic stretch and arginine residues

The interaction energies between the permeating ion and the protein were calculated, and are shown in Figure 6.5 (A). The permeating ions experience a repulsive force as they enter the channel, which is attributed to a ring of positively charged Arg35 residues at the entrance of the pore. The repulsive electrostatic interactions between the cation and the arginine residues are reflected in the positive values of interaction energy in the C-terminal region of the pore. The Ca^{2+} ion experiences significantly greater repulsion in comparison to K^+, which is explained by the higher charge on the ion. Interactions with the protein continue to remain unfavorable as the ion crosses the stretch lined by hydrophobic residues (Leu28, Leu24 and Trp21). The higher positive charge on Ca^{2+} compared to K^+ makes it difficult for the former to cross the hydrophobic stretch. Once the K^+/Ca^{2+} ion reaches the hydrophilic lining of the pore, the ion is stabilized via favorable interactions with the protein. Since the residues asparagine and serine carry polar groups with partial negative charges, interactions between the cation, and Asn16, Ser12 and Asn9 are electrostatically favorable, leading to lowering of the interaction energy. Such stabilizing interactions between the permeating ion and the hydrophilic pore-lining residues give rise to intermediate structures as discussed above. It must be noted that the interaction of Ca^{2+} with the protein when the ion is found near the Asn9 residue, which is a major binding site, is found to be stronger than that of K^+ with the protein, owing to the higher charge on Ca^{2+}. Such strong interactions are consistent with the deeper minimum seen in the PMF for Ca^{2+} around the asparagine residue.

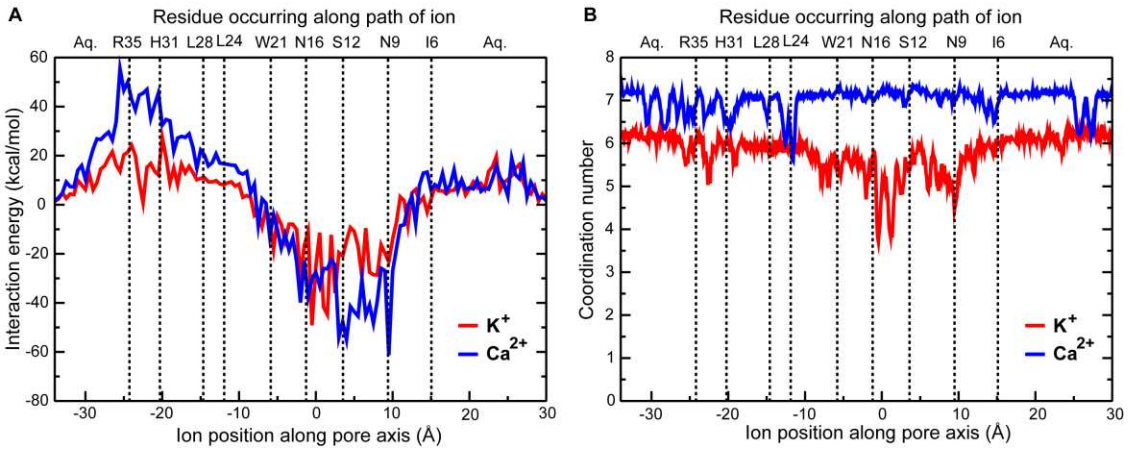

Figure 6.5 (A) Interaction energy between the permeating ion and the protein. Values shown are averages over the respective window. (B) Coordination number for the permeating ion along the reaction coordinate. The coordination

number includes both water molecules and Cl^- ions coordinating the permeating ion. The number of water molecules around the ion was calculated using a cutoff of 3.2 Å for K^+ and 2.9 Å for Ca^{2+}, and the number of coordinating Cl^- ions was calculated by using a cutoff of 3.6 Å for K^+ and 3.3 Å for Ca^{2+}.

OuYang et al. have proposed that the ring of Asn9 residues plays an important role in the ion conduction mechanism by acting as a selectivity filter [23]. Although the current study suggests that Asn9 serves as an ion-binding site, the discrimination between K^+ and Ca^{2+} cannot be explained based on this, since both can effectively bind to the asparagine residues. Instead, the selectivity is provided by the hydrophobic stretch consisting of Leu28, Leu24 and Trp21, and the Ca^{2+}, which carries a higher charge than K^+, encounters a greater energetic barrier in crossing this hydrophobic stretch. The mechanism of selectivity hypothesized here is significantly different from that found in typical K^+ channels, which have the GYG signature sequence in their selectivity filter [1,2,57]. In these channels, the larger size of the K^+ ion compared to smaller cations like Na^+ allows the ion to interact directly with the residues forming the selectivity filter.

6.3.5 Ion solvation dynamics and permeability

The solvation dynamics around the ions and their effect on the nature of interactions with sidechains play an important role in the permeability of ions through the channel. Pore regions that exhibit low water density may not be able to accommodate fully solvated ions, hence necessitating partial desolvation of the ions to permeate these regions. Figure 6.5 (B) shows the coordination number of the ion as a function of the reaction coordinate, with the coordination number including the total number of water molecules and Cl^- ions. Cl^- ions are included in the analysis to take into account the presence of these ions around the permeating K^+/Ca^{2+} ion in windows corresponding to the bulk water region. The ability of Ca^{2+} to remain in the fully solvated form is high compared to that of K^+, and high probability is observed corresponding to the Ca^{2+} ion coordinated to seven molecules. On the other hand, the K^+ ion samples states corresponding to a wider range of the number of coordinating molecules. The ions are seen to lose water molecules as they pass through the hydrophobic constrictions in the channel. The loss of coordinating water molecules becomes important in the hydrophobic stretch near Leu24, where the fall in the number of coordinating molecules around Ca^{2+} is acute. This loss of water molecules greatly destabilizes the Ca^{2+}, since there are no favorable interactions with the protein in this region, owing to predominance of hydrophobic residues. K^+, on the other hand, is seen to

lose water molecules around the hydrophilic stretch, which is compensated for via interactions with Asn16 and Asn9. Thus, the loss of coordinating water molecules occurs at different positions for K^+ and Ca^{2+}, and has contrasting effects on the two ions. To sum up, the smaller size of Ca^{2+} than K^+ (by about 0.3 Å), making the solvated calcium ion smaller, and the strong interaction of Ca^{2+} with water compared to K^+, are two possible factors that govern solvation states.

6.3.6 Ions pass one at a time through the hydrophobic stretch

This section and the following two sections discuss two ion permeation through the channel. Throughout this discussion, the two ions in the pore will be referred to as "upper" and "lower" ions on the basis of their relative position in the pore. The upper ion is closer to the N-terminal opening of the pore (and therefore appears at the top in the figures shown here). Free energy profiles with respect to the position of the upper ion and the separation between the two ions are shown in Figure 6.6. For both K^+ and Ca^{2+}, the permeating ions maintain a high separation (interionic separation less than 11 Å is avoided) as they enter the pore from the C-terminal side. The ring of positively charged Arg35 residues, followed by the hydrophobic stretch, are regions where the accommodation of more than one cation is highly unfavorable, and, consequently, only one cation passes through this region at a time. While the energy barrier along the MFEP is 5.4 kcal mol^{-1} for K^+, that for Ca^{2+} is 19.4 kcal mol^{-1}. This accounts for the selectivity of the channel toward K^+ over Ca^{2+} reported in experimental studies [8]. Although the energy barrier for K^+ seen here is almost the same as that observed previously in the single ion permeation study [Figure 6.2], the energy barrier for Ca^{2+} is significantly higher. This suggests that, while multiple K^+ ion transport through the hydrophobic stretch is possible, permeation of more than one Ca^{2+} ion is not favorable, further confirming the greater permeability of the channel toward K^+ (in fact, the ions prefer to maitain a separation of more than 12 Å in this region, as described later in section 6.3.9).

Once the upper ion crosses the hydrophobic stretch and binds to the ring of Ser12 residues, the lower ion starts moving from the ring of Leu24 residues. The upper ion remains bound to Ser12 until the lower ion enters the hydrophilic region and binds to Asn16, so that the two ions are very close together. While the separation between the two K^+ ions is ~ 5 Å, that between the Ca^{2+} ions

is ~ 7 Å when the two ions are bound to Ser12 and Asn16. As will be shown later, further movement of the upper ion toward the N-terminal end is facilitated by repulsion between the two ions. This is followed by a stable intermediate in which one ion is bound to Asn9 while the other is bound to Asn16. Further movement of the two ions is expected to be driven by a new incoming ion. It must be mentioned that, although the MFEP for Ca^{2+} suggests that the two ions are close together as they exit the pore region, the free energy values corresponding to different separations between the ions at this point are almost the same, and, hence, any of these states is equally likely to occur.

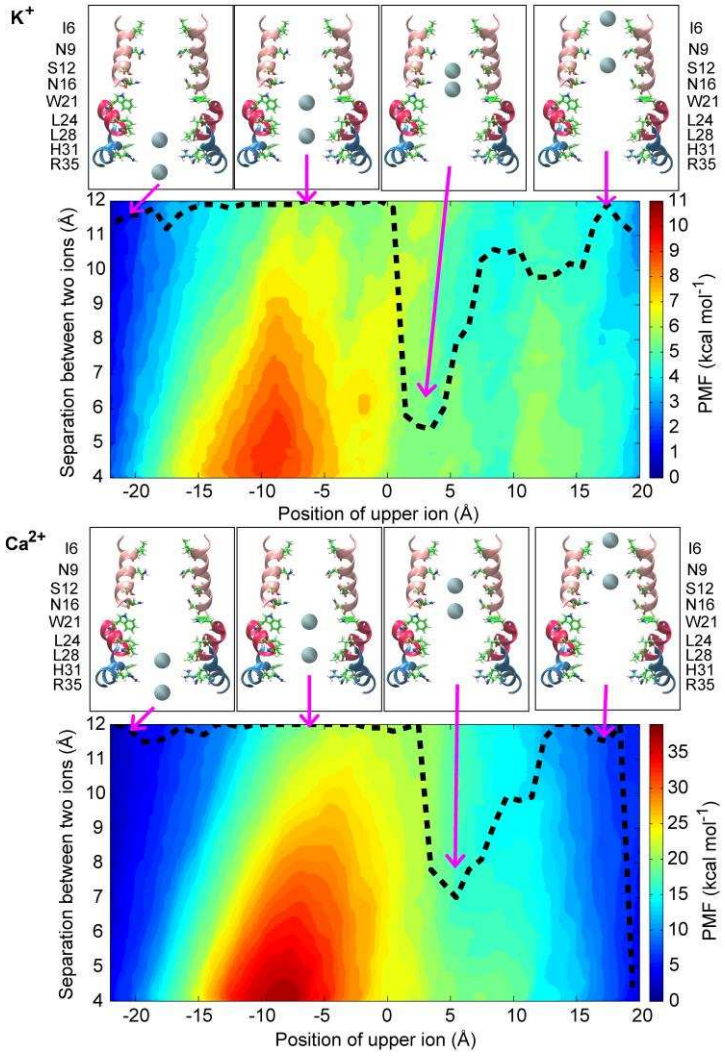

Figure 6.6 PMF for the permeation of two ions through the channel as a function of the position of the upper ion and the separation between the two ions. The minimum free energy paths (MFEP) are shown by black dashed lines. Representative configurations along the MFEP are also shown.

6.3.7 Ion passage through the hydrophilic region is driven by knock-on

The initial barrier in the PMFs due to the Arg35 residues and the hydrophobic stretch is expected (Figure 6.6). However, the local minimum corresponding to the ion binding sites become shallower compared to the single ion permeation. To explore the role of knock-on effect, the interaction energy between the two permeating ions was calculated. Figure 6.7 (A) shows the interaction energy between the two ions averaged over bins along the two reaction coordinates. Near the ring of positively charged arginine residues and in the hydrophobic region, the two ions avoid unfavorable interactions by maintaining a large separation. When the ions are inside the hydrophilic region formed by Asn9, Ser12, and Asn16, interionic repulsion becomes significant, with the interaction energy between the two permeating ions going up to more than 20 kcal/mol. For Ca^{2+}, although the interionic repulsion is seen only for a very short stretch along the MFEP, the magnitude of the repulsion is higher than that of K^{+}, explaining the greater separation between the permeating Ca^{2+} ions compared to the K^{+} ions. The favorable interaction between the permeating ion and the binding site is countered by ion-ion repulsion, which facilitates the further movement of the ion. Such a knock-on process has previously been shown to drive ion transport in the K^{+}-selective KcsA channel [3]. In the present study, knock-on is seen to involve direct repulsion between the two ions, without any bridging water molecule in between (data not shown), which is much like the knock-on mechanism reported recently for the KcsA channel [36].

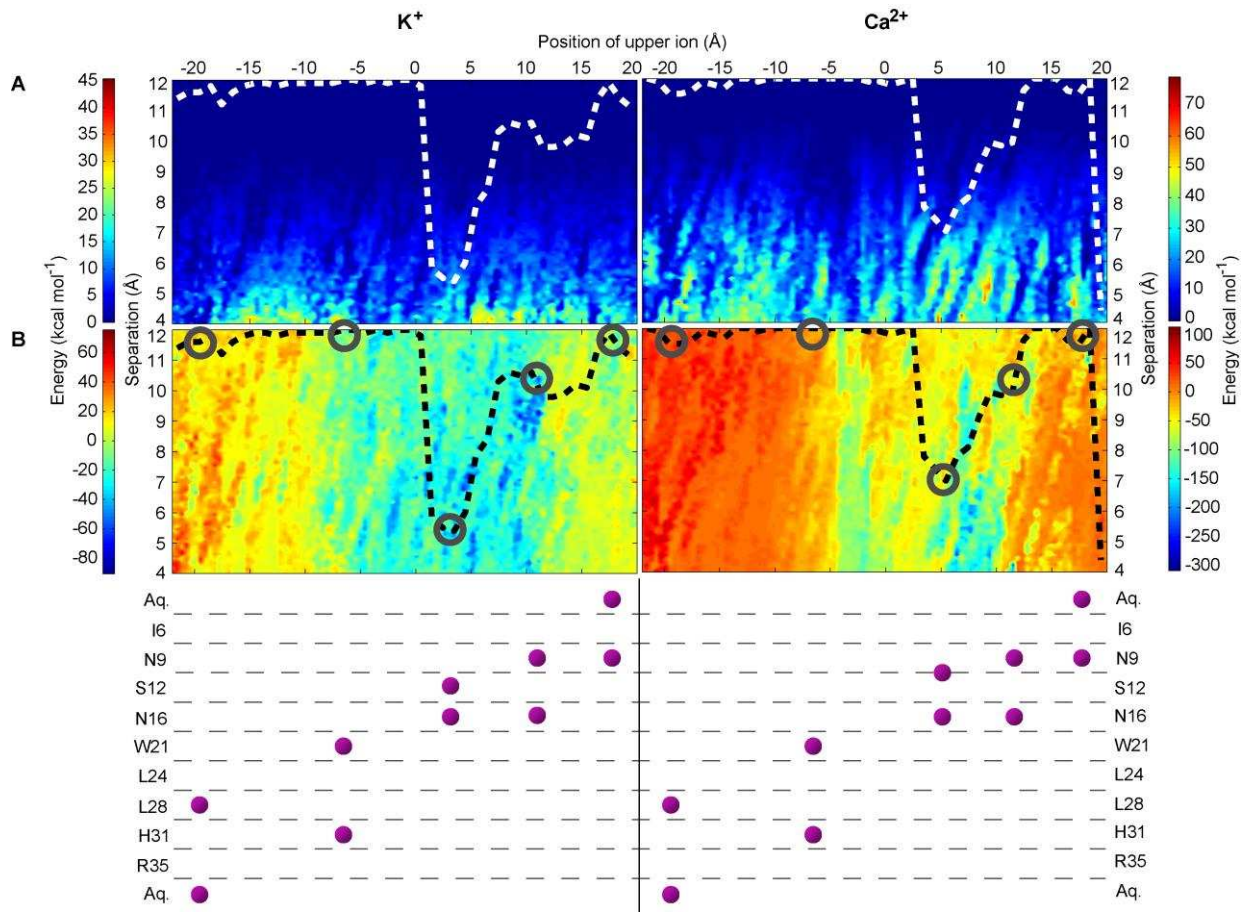

Figure 6.7 (A) Interaction energy between the two permeating ions with respect to the position of the upper ion and the interionic separation. MFEP is shown as dashed lines. (B) Interaction energy between the upper ion and the protein. Important configurations along the MFEP are encircled, and the approximate locations of the two ions (relative to pore-lining residues) for these configurations are shown at the bottom.

The interactions between the permeating ion and the protein were further analyzed by calculating the interaction energies (Figure 6.7 (B)). Once the upper ion crosses the hydrophobic stretch and is inside the hydrophilic region, it interacts strongly with the pore-lining residues; this behavior is observed for K^+ and Ca^{2+} alike. It can be seen that the upper ion is bound to Ser12 residue as the lower ion moves from the hydrophobic region to Asn16. The strong interaction between the upper ion and the serine is overcome only after the lower ion binds to the ring of Asn16 residues, and reaches a position where it repels the upper ion (this corresponds to the minimum separation between the two ions along the MFEP). Thus, ion-ion repulsion makes possible further movement of the upper ion towards the exit of the pore. Another stable configuration is seen where the upper ion is bound to Asn9 while the lower ion is still bound to Asn16. Further

movement of the upper ion from this configuration is likely to be driven by complete solvation of the ion as it approaches bulk water.

It is to be noted that the knock-on mechanism facilitates ion conduction by ensuring that ions do not remain bound in the hydrophilic region. In other words, this mechanism leads to a shallow energy minimum corresponding to the binding site, thereby preventing the system from getting trapped in this local minimum. The mechanism does not, however, lower the energy barrier for the ion transport process, and, in fact, the energy barrier for Ca^{2+} reported here is higher than that observed previously for single Ca^{2+} ion permeation. This suggests that it is highly unlikely for two Ca^{2+} ions to pass through the hydrophobic stretch at the same time. However, as far as the hydrophilic region is concerned, the existence of two Ca^{2+} ions in this region becomes a necessary phenomenon for knock-on to occur. It is with this knock-on that ions are driven out of the binding site, thereby ensuring a continuous flow of ions across the channel.

6.3.8 K^+ and Ca^{2+} exhibit different modes of binding in the binding site

Free energy calculations that have been performed above to investigate single ion permeation in p7 have reported Asn9 as a major binding site, and Asn16 and Ser12 as minor binding sites. In the present study, a particularly stable configuration that is seen is one with the upper ion bound to Asn9, and the lower ion bound to Asn16. The mode of binding of K^+ and Ca^{2+} to these sites has been elucidated in detail here based on a number of analyses. Figure 6.8 (A) shows a probability distribution of the distance between the permeating ions and the closest oxygen atom of the ring of asparagine residues for configurations corresponding to the two ions in the respective binding sites (the distance is measured between Asn9 and the upper ion, and between Asn16 and the lower ion). For both the upper ion and the lower ion, K^+ is seen to occur closer to any of the amide oxygens in comparison to Ca^{2+}. In spite of the lower size of Ca^{2+}, it is seldom seen to occur at a distance of less than 3 Å of the amide oxygens. It is also to be noted that, while a few configurations are seen where K^+ is not in the vicinity (> 4 Å) of the amide oxygens, thereby broadening the probability distribution, Ca^{2+} ions are seen to have a narrow distribution, suggesting that the latter sample a smaller space while they are inside the ring of asparagine residues. Such behavior is evident in Figure 6.8 (B), which shows snapshots of the K^+ and Ca^{2+} ions interacting with the ring of Asn9 residues. Inspection of the trajectories shows that Ca^{2+} is

completely solvated, and its interaction with the asparagines is via bridging water molecules only. K^+, on the other hand, is able to interact directly with the asparagines, as shown in Figure 6.8 (C), thereby explaining why it is seen in the vicinity of Asn9. Higher desolvation penalty expected for the dication and the size differences between the two ions are possible factors that necessitate this kind of behavior.

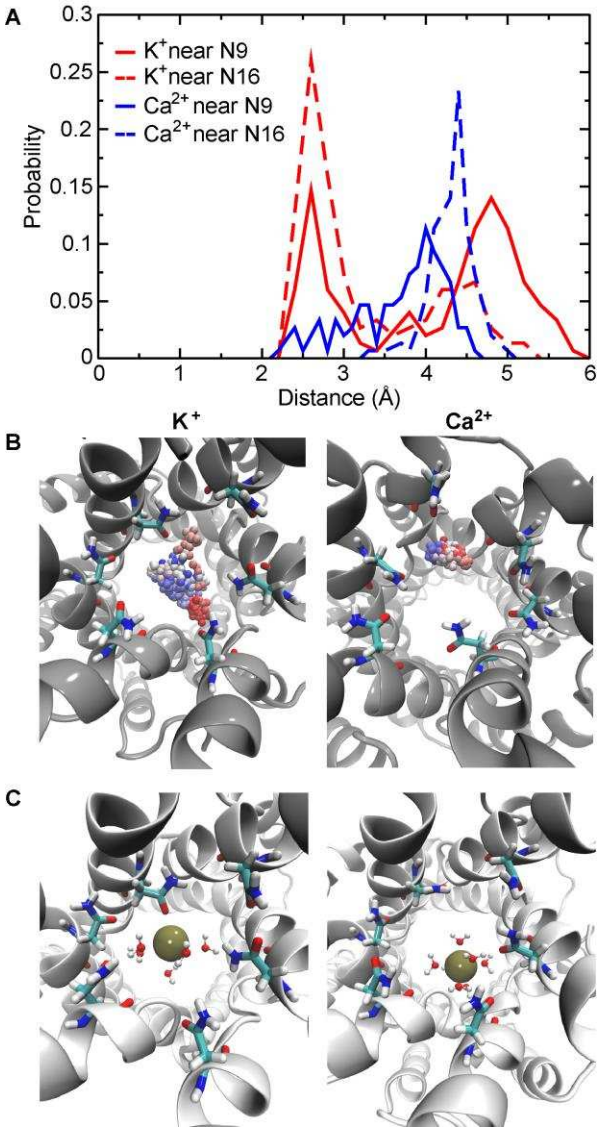

Figure 6.8 (A) Probability distribution of the distance between the permeating ion and the closest Asn9 residue. The data is taken from a window that corresponds to the upper ion near Asn9 and the lower ion near Asn16. (B) Snapshots of the upper ion in the binding site formed by the ring of Asn9 residues. All positions sampled by the ion in the window are shown. The color gradient from red to blue varies with the timestep of the frames in the trajectory. (C) Representative snapshots for the upper ion in the vicinity of the Asn9 are shown with coordinating water molecules.

6.3.9 One-dimensional umbrella sampling studies modeling two ion permeation reveal a similar mechanism

In order to further verify the mechanism inferred from the above umbrella sampling studies, two ion permeation studies were designed such that the restraining potential was applied on only one of the ions, namely the lower ion. The upper ion was allowed to move freely in these simulations, so that it occupies the most favorable position along the pore with respect to the lower ion in terms of the free energy. Figure 6.9 shows probabilities of different separations between the two ions with respect to the position of the lower ion. When the lower ion is in the vicinity of the arginine ring or in the hydrophobic stretch (between -25 Å and -6 Å), the separation between the two ions goes beyond 12 Å, sometimes as high as 18 Å. However, once the upper ion is in the hydrophilic region (corresponding to lower ion lying between -12 Å and 7 Å), the two ions are seen close together. The tendency of the two ions to stay close together in this region is indicative of the knock-on effect. The separation between the two ions increases rapidly once the first ion goes beyond the hydrophilic region (in some of these simulations, the upper ion is seen to exit the pore; consequently, it moves freely, and often lies at a distance of beyond 20 Å from the lower ion). The mechanism described here is consistent with the mechanism inferred from the two-dimensional umbrella sampling calculations. In brief, only one ion passes through the hydrophobic region at a time (the interionic separation never falls to below 12 Å in this region), but when the ions enter the hydrophilic region, they stay close together, so that a knock-on is possible.

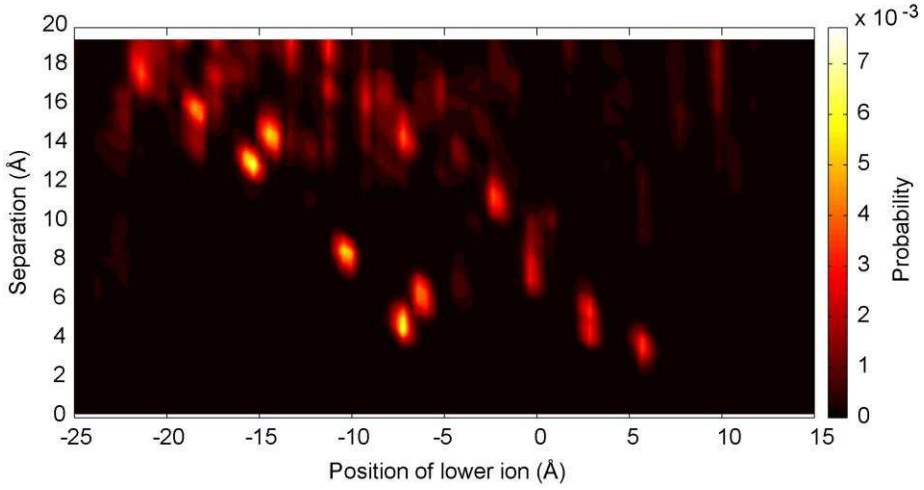

Figure 6.9 Probabilities for different separations between the two ions with respect to the position of the lower ion.

6.4 Conclusions

The p7 viroporin is seen to be more permeable to K^+ ions than to Ca^{2+} ions, in good agreement with experimental studies. A thermodynamic basis for the selectivity of the channel is provided by free energy profiles for the conduction of K^+ and Ca^{2+}. The free energy profiles for single ion permeation reveal energy barriers corresponding to the hydrophobic stretch, with the height of the barrier being 5.2 kcal/mol for K^+ and 9.7 kcal/mol for Ca^{2+}. A hydrophobic stretch formed by Leu28, Leu24 and Trp21 governs the selectivity of the channel by giving rise to a moderately hydrophobic region, which makes it difficult for the Ca^{2+} ion to pass through, since it carries a higher charge compared to K^+. Free energy profiles computed as a function of axial and radial position of the permeating ion position show that K^+ exhibits greater lateral movement as it passes through the pore, in comparison to Ca^{2+}. This suggests stronger interaction of the K^+ with pore-lining residues, which is also reflected in the interaction energy between the protein and the ion. The two ion permeation study shows that ion-ion knock-on drives a continuous flow of ions along the pore axis, enabling ions to overcome the energy barrier that they encounter in moving beyond the binding site. While K^+ binds directly to the asparagines forming the binding site, Ca^{2+} interacts with the asparagines via bridging water molecules. Thus, a novel cooperation of different phenomena and subtle differences in the response of the two ions resolve ion selectivity.

References

1. Roux B, et al. (2011) Ion selectivity in channels and transporters. J Gen Physiol 137(5):415-426.
2. Nimigean CM, Allen TW (2011) Origins of ion selectivity in potassium channels from the perspective of channel block. J Gen Physiol 137(5):405–413.
3. Allen TW, Andersen OS, Roux B (2003) Energetics of ion conduction through the gramicidin channel. Proc Natl Acad Sci USA 101(1):117–122.
4. Boiteux C, Vorobyov I, Allen TW (2014) Ion conduction and conformational flexibility of a bacterial voltage-gated sodium channel. Proc Natl Acad Sci USA 111(9):3454–3459.
5. Fowler PW, Beckstein O, Abad E, Sansom MSP (2013) Detailed Examination of a Single Conduction Event in a Potassium Channel. J Phys Chem Lett 4(18):3104–3109.

6. Padhi S, Burri RR, Jameel S, Priyakumar UD (2014) Atomistic Detailed Mechanism and Weak Cation-Conducting Activity of HIV-1 Vpu Revealed by Free Energy Calculations. PLoS ONE 9(11): e112983.
7. Atoom AM, Taylor NG, Russell RS (2014) The elusive function of the hepatitis C virus p7 protein. Virology 462-463:377–387.
8. Griffin SD, et al. (2002) The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. FEBS Lett 535(1-3):34–38.
9. Pavlovic D, et al. (2003) The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. Proc Natl Acad Sci USA 100(10):6104–6108.
10. Premkumar A, Wilson L, Ewart G, Gage P (2003) Cation-selective ion channels formed by p7 of hepatitis C virus are blocked by hexamethylene amiloride. FEBS Lett 557(1-3):99–103.
11. Sakai A, et al. (2003) The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences. Proc Natl Acad Sci USA 100(20):11646–11651.
12. Steinmann E, et al. (2007) Hepatitis C Virus p7 Protein Is Crucial for Assembly and Release of Infectious Virions. PLoS Pathog 3(7):e103.
13. Jones CT, Murray CL, Eastman DK, Tassello J, Rice CM (2007) Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus. J Virol 81(16):8374–8383.
14. Cook GA, Dawson LA, Tian Y, Opella SJ (2013) Three-dimensional structure and interaction studies of hepatitis C virus p7 in 1,2-Dihexanoyl-sn-glycero-3-phosphocholine by solution nuclear magnetic resonance. Biochemistry 52(31):5295–5303.
15. Montserret R, et al. (2010) NMR structure and ion channel activity of the p7 protein from hepatitis C virus. J Biol Chem 285(41):31446–31461.
16. Cook GA, Opella SJ (2011) Secondary structure, dynamics, and architecture of the p7 membrane protein from hepatitis C virus by NMR spectroscopy. Biochim Biophys Acta, Biomembr 1808(6):1448–1453.

17. Cook GA, Zhang H, Park SH, Wang Y, Opella SJ (2011) Comparative NMR studies demonstrate profound differences between two viroporins: p7 of HCV and Vpu of HIV-1. Biochim Biophys Acta, Biomembr 1808(2):554–560.
18. Cook GA, Opella SJ (2009) NMR studies of p7 protein from hepatitis C virus. Eur Biophys J 39(7):1097–1104.
19. Patargias G, Zitzmann N, Dwek R, Fischer WB (2006) Protein−protein interactions: Modeling the hepatitis C virus ion channel p7. J Med Chem 49(2):648–655.
20. Chandler DE, Penin F, Schulten K, Chipot C (2012) The p7 protein of hepatitis C virus forms structurally plastic, minimalist ion channels. PLoS Comput Biol 8(9):e1002702.
21. Wang Y-T, Schilling R, Fink RH, Fischer WB (2014) Ion-dynamics in hepatitis C virus p7 helical transmembrane domains — a molecular dynamics simulation study. Biophys Chem 192:33–40.
22. Kalita MM, Griffin S, Chou JJ, Fischer WB (2015) Genotype-specific differences in structural features of hepatitis C virus (HCV) p7 membrane protein. Biochim Biophys Acta, Biomembr 1848(6):1383–1392.
23. OuYang B, et al. (2013) Unusual architecture of the p7 channel from hepatitis C virus. Nature 498(7455):521–525.
24. Dev J, Brüschweiler S, Ouyang B, Chou JJ (2015) Transverse relaxation dispersion of the p7 membrane channel from hepatitis C virus reveals conformational breathing. J Biomol NMR 61(3-4):369–378.
25. Luik P, et al. (2009) The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy. Proc Natl Acad Sci USA 106(31):12712–12716.
26. Chew CF, Vijayan R, Chang J, Zitzmann N, Biggin PC (2009) Determination of pore-lining residues in the hepatitis C virus p7 protein. Biophys J 96(2):L10-L12.
27. Wozniak AL, et al. (2010) Intracellular proton conductance of the hepatitis C virus p7 protein and its contribution to infectious virus production. PLoS Pathog 6(9):e1001087.
28. StGelais C, et al. (2007) Inhibition of hepatitis C virus p7 membrane channels in a liposome-based assay system. Antiviral Res 76(1):48–58.
29. StGelais C, et al. (2009) Determinants of hepatitis C virus p7 ion channel function and drug sensitivity identified in vitro. J Virol 83(16):7970–7981.

30. Li H, et al. (2012) Genetic and functional heterogeneity of the hepatitis C virus p7 ion channel during natural chronic infection. Virology 423(1):30–37.
31. Furini S, Domene C (2009) Atypical mechanism of conduction in potassium channels. Proc Natl Acad Sci USA 106(38):16074–16077.
32. Egwolf B, Roux B (2010) Ion selectivity of the KcsA channel: a perspective from multi-ion free energy landscapes. J Mol Biol 401(5):831–842.
33. Corry B (2013) Na^{+}/Ca^{2+} selectivity in the bacterial voltage-gated sodium channel NavAb. PeerJ 1:e16.
34. Fowler PW, Abad E, Beckstein O, Sansom MSP (2013) Energetics of multi-ion conduction pathways in potassium ion channels. J Chem Theory Comput 9(11):5176–5189.
35. Wang Y, Chamberlin AC, Noskov SY (2014) Molecular strategies to achieve selective conductance in NaK channel variants. J Phys Chem B 118(8):2041–2049.
36. Kopfer DA, et al. (2014) Ion permeation in K channels occurs by direct Coulomb knock-on. Science 346(6207):352–355.
37. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem 29(11):1859–1865.
38. Wu EL, et al. (2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J Comput Chem 35(27):1997–2004.
39. Carrere-Kremer S, et al. (2002) Subcellular localization and topology of the p7 polypeptide of hepatitis C virus. J Virol 76(8):3720–3730.
40. Brohm C, et al. (2009) Characterization of determinants important for hepatitis C virus p7 function in morphogenesis by using trans-complementation. J Virol 83(22):11682–11693.
41. Essmann U, et al. (1995) A smooth particle mesh Ewald method. J Chem Phys 103(19):8577.
42. Ryckaert J-P, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341.

43. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13(8):1011–1021.
44. Roux B (1995) The calculation of the potential of mean force using computer simulations. Comp Phys Comm 91(1-3):275–282.
45. Grossfield A "WHAM: the weighted histogram analysis method", version 2.0.6, http://membrane.urmc.rochester.edu/content/wham.
46. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802.
47. Best RB, et al. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. J Chem Theory Comput 8(9):3257–3273.
48. MacKerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25(11):1400–1415.
49. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616.
50. Klauda JB, et al. (2010) Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. J Phys Chem B 114(23):7830–7843.
51. Beglov D, Roux B (1994) Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. J Chem Phys 100(12):9050-9063.
52. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926-935.
53. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. J Comput Chem 30(10):1545–1614.
54. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graphics 14(1):33–38.
55. Smart OS, Neduvelil JG, Wang X, Wallace B, Sansom MS (1996) HOLE: A program for the analysis of the pore dimensions of ion channel structural models. J Mol Graphics 14(6):354–360.

56. Desplantez T, Halliday D, Dupont E, Severs NJ, Weingart R (2011) Influence of v5/6-His tag on the properties of gap junction channels composed of connexin43, connexin40 or connexin45. J Membrane Biol 240(3):139–150.
57. Doyle DA (1998) The structure of the potassium channel: Molecular basis of K^+ conduction and selectivity. Science 280(5360):69–77.

7 Water Transport through Aquaporin AQP2

7.1 Introduction

Water, being the most abundant molecule in cells, necessitates the existence of specialized mechanisms for regulating its concentration inside and outside cells. This is achieved by a family of proteins called aquaporins, which rapidly conduct water while at the same time ensuring that bigger solute molecules and charged species like cations do not pass through, thereby preserving homoeostasis in cells [1]. Aquaporins are found in the cell membranes of a number of tissues across the body, and they typically conduct water at the rate of about 10^9 permeation events every second [1,2]. There are a total of thirteen different aquaporins in mammals, and lack of expression of these aquaporins can lead to several diseased states, including cataract, diuresis, and nephrogenic diabetes insipidus (NDI). Furthermore, higher levels of aquaporins have been reported in certain cancers, and deletion of aquaporins has been associated with reduction in tumor angiogenesis [3].

Human aquaporin 2 (AQP2) is an aquaporin found in the collecting ducts of the kidney, where it concentrates urine by enhancing water transport across cell membranes [4,5]. Certain mutations in AQP2 can prevent the targeting of AQP2 to these cells, and, consequently, the absence of AQP2 in the membranes of these cells leads to NDI [6,7]. The recently solved crystal structure of AQP2 has provided insights into how certain mutations in the protein can eventually lead to NDI [8]. Most of these mutations occur in the transmembrane region, and they lead to misfolding of the protein, followed by retainment in the endoplasmic reticulum. The subsequent absence of AQP2 in the apical membrane of principal cells of the collecting duct leads to NDI, which involves severe dehydration. However, since the above study focuses on the effect of NDI-causing mutations, it does not provide insights into the mechanism, kinetics, and energetics of water transport through the pore of the channel.

Molecular dynamics (MD) simulations have been helpful in understanding the mechanistic details of water transport at a microscopic level that is usually not accessible to experiments [9,10]. Since water permeation through aquaporins does not involve high energy barriers

(barriers are typically less than 5 kcal mol^{-1}), it becomes possible to observe complete permeation events in simulations, which, in turn, provide atomistic details of the permeation process. In a pioneering study, water transport through AQP1 was characterized in atomistic detail using unbiased MD simulations [11]. The study revealed the free energy profile for water permeation through AQP1, and provided an estimate of the number of permeation events that could possibly occur in unit time. The kinetics of water transport through AQP1 was later characterized by calculating the diffusion permeability and osmotic permeability for single-file water transport [12]. These studies were followed by a number of MD simulation studies on different aquaporins that provided insights into mechanisms of gating and selectivity, in addition to a quantification of permeability [13-25]. A feature of aquaporins that has been intriguing is the mechanism they adopt in order to prevent the permeation of protons; this has been the subject of a number of studies [26-33]. The most significant barrier to proton transport is encountered around the conserved Asn-Pro-Ala (NPA) motif of aquaporins, and this barrier is caused primarily by the electrostatic field of this motif, with the arrangement of the water chain in this region playing a secondary role [27,28,30].

The first MD study on AQP2 was reported soon after the structure of the protein was solved, and it quantified the diffusion permeability and osmotic permeability for water transport through the channel [34]. However, the energetics and selectivity mechanisms associated with water transport through the channel are still not understood. The current study elucidates the transport properties of AQP2 using microsecond-timescale MD simulations. In addition to the quantification of permeability, the study reveals the free energy profile for water transport. The regulation of the orientation of the water chain by the pore lumen, and its possible effect on proton transport, is discussed. The study complements experimental studies by revealing transport mechanisms in AQP2 that have not been possible to examine using experiments, and provides direction to new experiments that could attempt to modulate the permeation.

7.2 Methods

7.2.1 Simulation details

The model system used in this study is derived from the crystal structure of AQP2 with PDB ID 4NEF [8]. Since the channel is a homotetramer, and the four monomeric units are similar, only one of the four monomers was considered for investigating the transport properties and dynamics. The protein was set up in a hydrated lipid bilayer of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate (POPC) using the CHARMM-GUI Membrane Builder [35,36], giving rise to a system with 57641 atoms. The system was equilibrated in the presence of positional restraints on the protein and planar restraints on the lipid headgroups, with the restraints being gradually turned off over the course of equilibration. The equilibration was performed initially in the NVT ensemble and later in the NPAT ensemble for a total of 400 ps. This was followed by production run in the NPAT ensemble for 1 μs with a timestep of 2 fs, and trajectories were saved every 2 ps. Throughout the simulations, a harmonic restraint with force constant 10 kcal mol^{-1} $Å^{-2}$ was applied on the center of mass of the protein to prevent drifting within the lipid bilayer. Long range electrostatics was treated using the particle mesh Ewald method [37], and Lennard-Jones interactions were truncated by using a switching function in the range 10-12 Å. Covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm [38]. The NAMD program was used for performing the simulations with the CHARMM36 force field for proteins and lipids and the TIP3P water model [39-44]. Structural images were rendered using VMD, pore radii were calculated using HOLE2, and trajectory analyses were performed using CHARMM [45-47]. An N184A mutant of the protein was also modeled, and simulations for this system were carried out for 100 ns.

7.2.2 Free energy profiles

The pore axis was divided into cylindrical bins of height 0.5 Å and radius 6 Å, and the number of water molecules in each bin was calculated to obtain the number density of water molecules for each bin. The number densities were multiplied by a scaling factor so that the density of water in bulk was 1. The free energy corresponding to each bin was then obtained using the relation

$$G = -k_B T \ln n_i \,, \qquad (7.1)$$

where n_i is the normalized number density for bin i.

Two-dimensional free energy profiles were calculated with respect to water position along the pore axis and the angle made by the water dipole with the pore axis. For every bin i along the pore axis, the angle between the water dipole and pore axis was divided into bins of size 3°, and the probability for each such bin was calculated. The free energy was then obtained as

$$G(n_i, \theta) = G(n_i) - k_B T \ln \rho(\theta; n_i) + C_i \ , \quad (7.2)$$

where $\rho(\theta; n_i)$ is the probability of occurrence of angle θ for a certain n_i, and C_i is a constant for a given value of i. The last 990 ns from the trajectory was used for obtaining the free energy profiles, with the first 10 ns being treated as equilibration phase (for the N184A mutant, the last 90 ns from the trajectory was used for computing the free energy profiles).

7.2.3 Permeability coefficients

The diffusion permeability p_d for single-file water molecules permeating through a channel is given by

$$p_d = v_w q_0 \ , \quad (7.3)$$

where v_w is the volume of a single water molecule and q_0 is the number of complete permeation events in one direction across the channel in unit time [12]. If the average number of water molecules in the lumen of the channel is N, the osmotic permeability constant can be calculated as

$$p_f / p_d = N + 1 \quad (7.4)$$

7.3 Results and Discussion

7.3.1 The ar/R constriction could act as a size exclusion selectivity filter

The structure of a monomeric unit of AQP2 is shown along with a representation of the pore region in Figure 7.1 (A). Crucial residues that line the pore lumen and are likely to play an important role in water transport are shown in Figure 7.1 (B). The pore is the narrowest around the ar/R region, which is formed by the residues F48, H172 and R187. The narrow region is likely to serve the purpose of achieving selectivity by prohibiting the entry of big solutes like glycerol. The exclusion of glycerol distinguishes AQP2 from aquaglyceroporins, which allow the

permeation of water as well as glycerol, and consequently have slightly different functional roles [1]. Figure 7.2 shows the pore radius profile of the channel averaged over the simulation as a function of position along the pore axis. It can be seen that the pore radius decreases to ~ 1 Å around the ar/R region, which is much less than the radius of a glycerol molecule. Such a mechanism of filtering out glycerol molecules via size exclusion is known to exist in other aquaporins [14]. The highly conserved NPA motif of aquaporins occurs at two places in the protein, one extending across residues 68-70, and the other across residues 184-186 [8]. The asparagines from the two NPA motifs (N68 and N184) can be seen in Figure 7.1 (B) with their NH_2 group facing the pore, and these could be involved in the formation of hydrogen bonds with the oxygen atoms of water. AQP2 does not possess tyrosine residues at the two ends of the pore lumen, unlike AQP0. The absence of such a "phenolic barrier" that would have arisen from such tyrosines implies that AQP2 might be more permeable to water than AQP0.

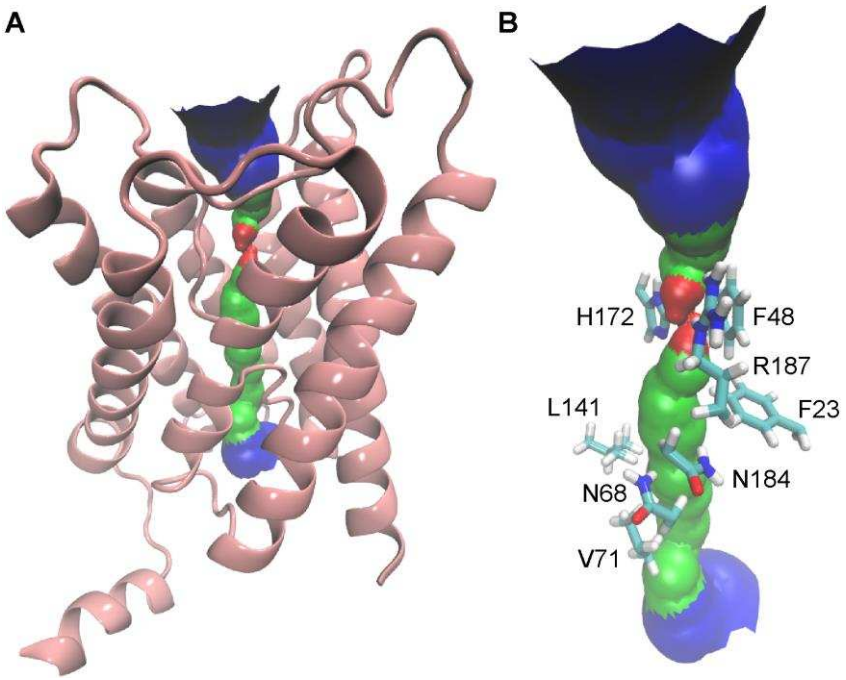

Figure 7.1 (A) The pore of the channel is shown along with a representation of the protein backbone. The color of the pore is an indication of the thickness of the pore at a given region, with blue > green > red. (B) Important residues lining the pore.

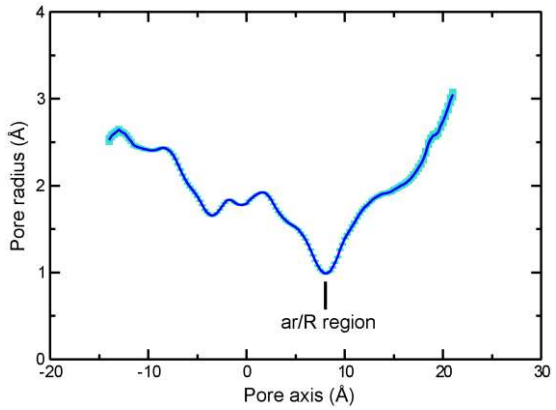

Figure 7.2 Pore radius of the channel averaged over the simulation. Error bars are shown in cyan.

Figure 7.3 (A) shows the root mean square fluctuation (RMSF) of the different residues during the simulations. While the pore-lining helices (in particular, the helices spanning residues 68-76, 126-149, and 184-196) exhibit very low fluctuation, the C-terminal helix (residues 229-239), which lies on the cytoplasmic side, fluctuates a lot. The variable position of the C-terminal helix has been implicated in the crystal structure of AQP2, where the helix exists in different orientations in the four monomeric units. The RMSD of the protein backbone shows that, while the transmembrane region of the protein is equilibrated rapidly, the C-terminal helix fluctuates during the first 100 ns, as indicated by the RMSD of the complete protein (Figure 7.3 (B)). The different conformations that are exhibited by this helix over the course of the simulation are shown in Figure 7.3 (C). The helix faces away from the pore lumen initially, and later orients itself so that it is directed toward the pore lumen.

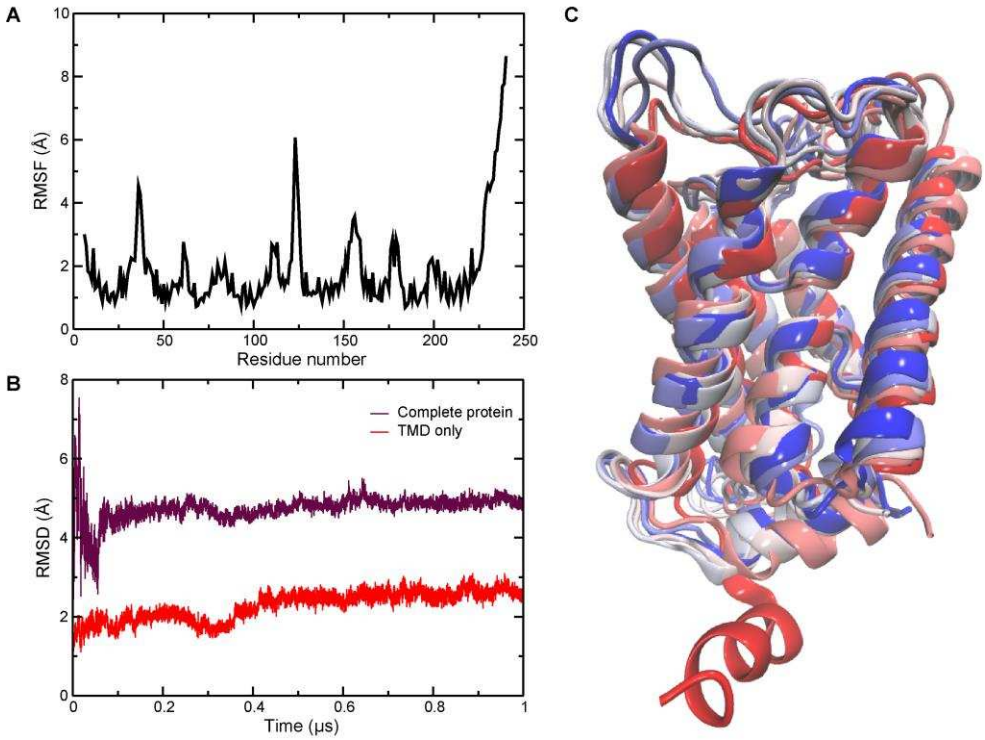

Figure 7.3 (A) Root mean square fluctuation (RMSF) of the residues over the course of the simulation. (B) Root mean square deviation (RMSD) of the protein backbone for the complete protein and for the transmembrane region of the protein. (C) Snapshots showing the different conformations of the protein over the course of the simulation. Color red corresponds to the initial phase of the simulation, and blue corresponds to the final phase of the simulation.

7.3.2 Water transport through the pore lumen is single-file and diffusion-limited

Figure 7.4 (A) shows the number density of water molecules as a function of position along pore axis. The water density decreases as one goes from bulk water to the pore lumen (the region between -8 Å to 12 Å is referred to as the pore lumen on account of the low pore radius of this region), with the density being the lowest around F23 and H172. For any given cross-section in the pore lumen, the number of water molecules is seen to be less than or equal to 1, suggesting single-file water transport in the lumen region. Figure 7.4 (A) also shows the minimum separation between a water oxygen atom at a given position along the pore and the oxygen atoms from the neighboring water molecules. The separation increases slightly around the NPA motif and the ar/R constriction, as a result of the water molecules orienting their principal axis parallel to the pore lumen in these regions (see later). However, the distance does not go beyond 3.5 Å, suggesting that water molecules are close enough to be able to interact with one another. These

are weak interactions, but they are important here because of the close separation between the water molecules. It follows that the continuous, interacting water chain makes it possible to have concerted water movement, which is an essential feature of single-file water transport. Figure 7.4 (B) shows a snapshot of the protein with a single-file water chain in the pore lumen.

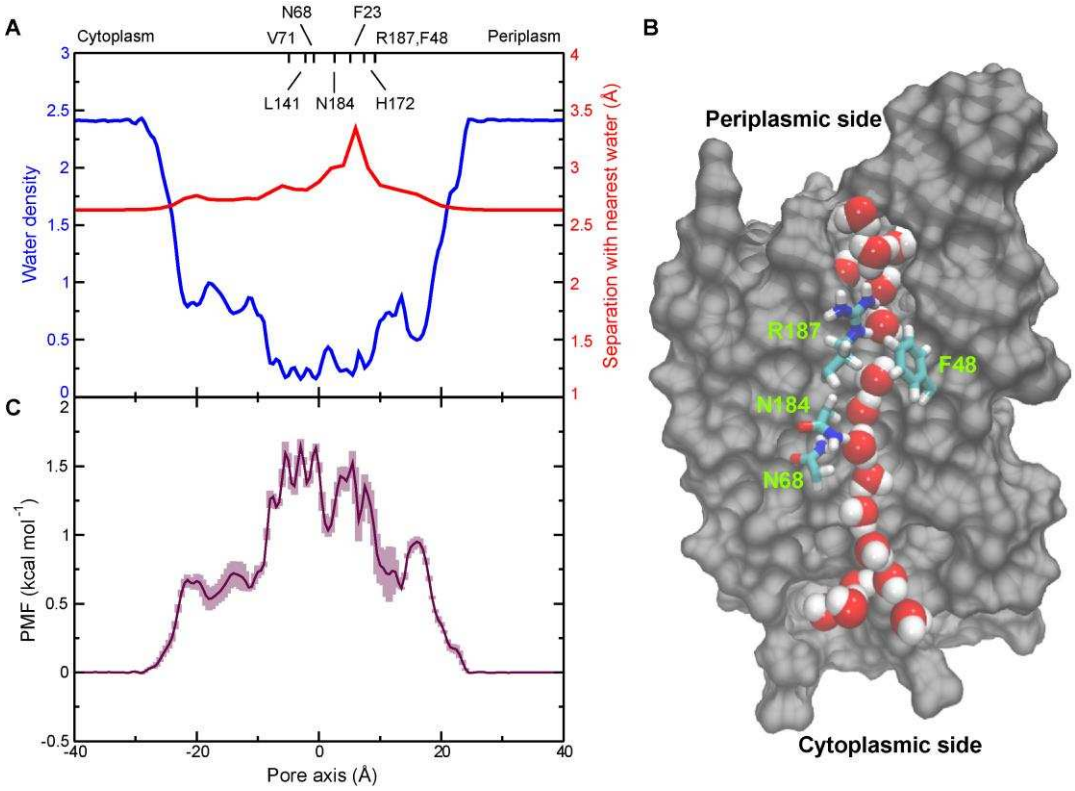

Figure 7.4 (A) Number density of water molecules (in blue) and separation between water oxygen atoms at a given position along the pore axis and the nearest water oxygen atom (red). The number density of water molecules was calculated by dividing the pore region into cylindrical elements of height 0.5 Å and radius 6 Å, and then calculating the number of water molecules in each element. Important residues lining the pore lumen are also shown at their respective position along the pore axis. (B) Snapshot showing a continuous water chain in the pore lumen. (C) Potential of mean force (PMF) for the transport of water molecules along the pore axis. The error bars were calculated by performing block analysis.

The potential of mean force (PMF) corresponding to the transport of water along the pore is shown in Figure 7.4 (C). The energy barrier is only 1.6 kcal mol^{-1}, suggesting that the transport process is almost diffusion-limited. The virtually barrierless free energy profile permits rapid rates of transport, as shown later. The peaks in the PMF occur near the short hydrophobic stretches formed by V71/L141 and F23/H172. Between these two peaks, there is a minimum near N184, leading to a release of 0.6 kcal mol^{-1} of energy. The amide group on the asparagine is

able to form hydrogen bonds with water molecules, leading to stable configurations of water in this region. Although the ar/R region is the narrowest region in the channel, it does not pose any significant barrier to the permeation of water. While the R187 sidechain can form hydrogen bonds with the water oxygen, it is likely to repel positively charged species like cations via electrostatic repulsion. This further corroborates the role of the ar/R region in governing the selectivity of the channel [8].

7.3.3 The orientation of the water dipole is restricted as it passes through the NPA motif

In order to find out how the orientation of the water dipole might be governed by the nature of the pore lumen, a two-dimensional PMF was computed as a function of water position along pore axis and the angle between the dipole moment of water and the pore axis, as shown in Figure 7.5. Water molecules in bulk are seen to exhibit a number of orientations, all of which are almost equally favorable. However, water molecules lying in the hydrophobic stretch formed by V71 and L141 exhibit a slight preference to orient themselves along the pore axis. Although there are no hydrogen bonding partners on the protein in this region, water molecules in this region orient themselves so as to make possible the formation of hydrogen bonds with asparagine sidechains in the neighboring NPA motif (Figure 7.6). In fact, when the water molecules are in the region corresponding to the NPA motif, the propensity to align themselves along the pore axis is the highest. Such an arrangement can allow optimal hydrogen bond formation between the asparagine amide group and water molecules. Similar orientation is also seen in the ar/R region, and this orientation could ease water passage through ar//R in two ways: (1) this would minimize the repulsion between water hydrogen atoms and the arginine sidechain, and (2) this would allow water molecules to pass through the narrow constriction with minimal steric hindrance. This channel does not, however, cause an inversion in the orientation of the water dipole as it passes through the NPA motif, which is unlike the case in the aquaglyceroporin GlpF [26,32].

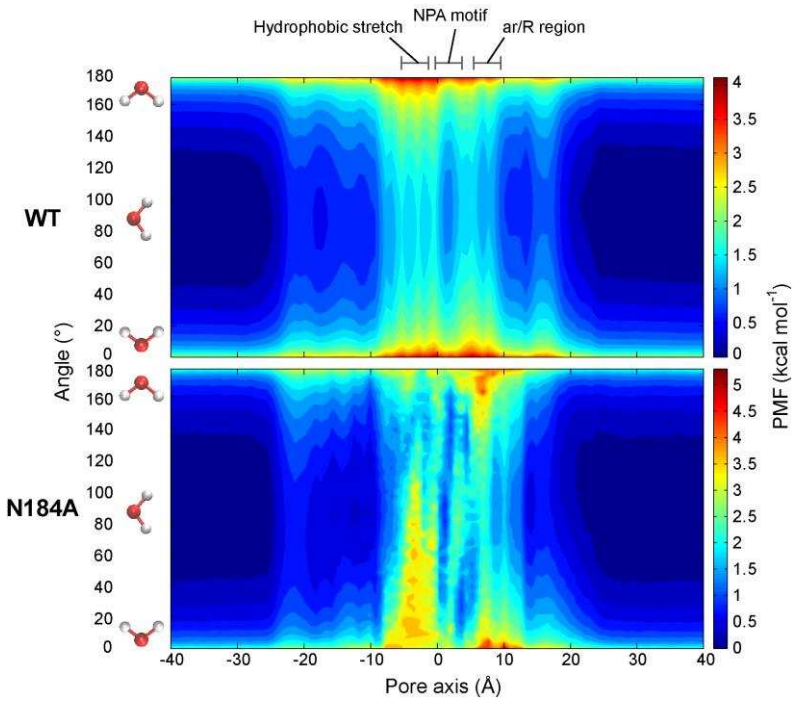

Figure 7.5 PMF for water transport along pore axis as a function of position along pore axis and angle between dipole moment of water and pore axis. The orientation of water molecules corresponding to angles of 0°, 90°, and 180° with the pore axis are shown on the left hand side of the figure.

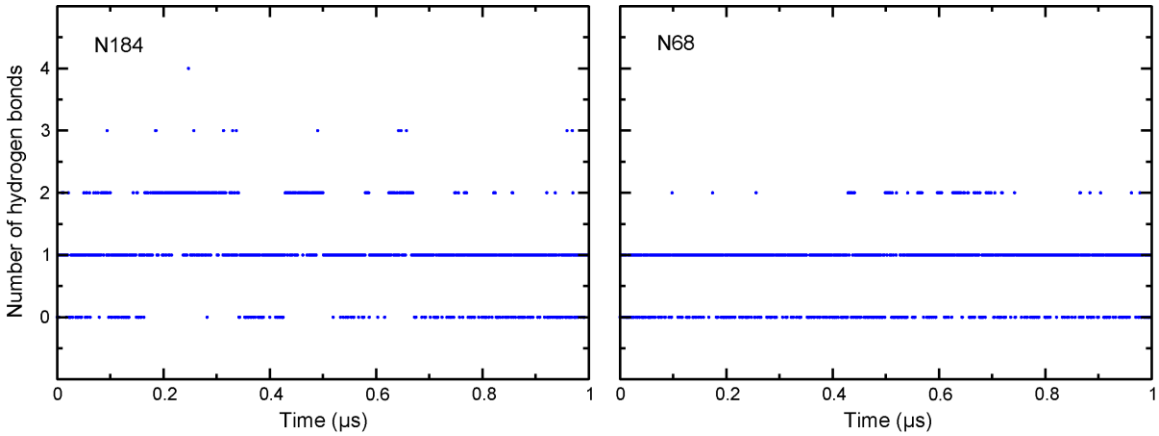

Figure 7.6 Number of hydrogen bonds between the pore-lining asparagines and water in wild-type AQP2.

In order to examine whether the restriction of the orientation of water around the NPA motif was indeed due to interactions with the N184 residue, an N184A mutant was investigated, and a 2D-PMF was calculated in a similar fashion (the 1D-PMF for this mutant is shown below in Figure 7.7; the energy barriers for the wild-type and mutant differ by a very marginal value of 0.4 kcal mol^{-1}). It is seen that, around the region corresponding to the NPA motif, the water molecules no

longer exhibit a propensity to align themselves along the pore axis, thereby confirming the role of the N184 residue in modulating water orientation. Thus, it is the electrostatics around the NPA motif in the wild-type form that leads to a rotational confinement of water molecules. The existence of this mechanism is likely to prevent the passage of protons through the channel. This is because, firstly, proton transport via a Grotthuss mechanism is not feasible, with the single-file water molecules being aligned such that the H-H vector of these water molecules is parallel to the pore axis. Secondly, the water molecules are already saturated with hydrogen bonds, owing to their interaction with the asparagine. Consequently, these water molecules will not be able to accept protons from neighboring water molecules. However, with the introduction of the N184A mutation, the above two mechanisms for proton exclusion no longer exist. It follows that this mutant is expected to show greater permeability toward protons than the wild-type form. It is proposed that experimental studies on this mutant can verify the enhanced proton permeability that has been discussed here.

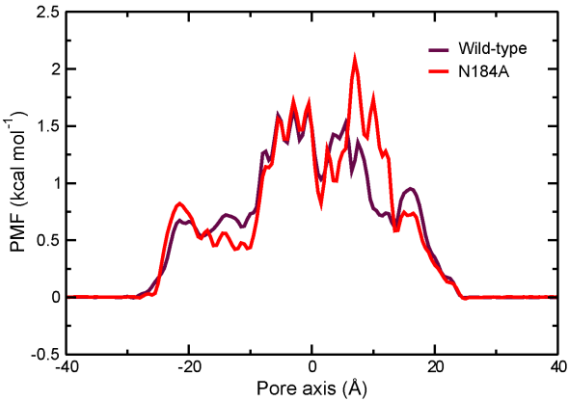

Figure 7.7 PMF for water transport along the pore axis of wild-type and mutant AQP2.

7.3.4 The channel exhibits greater permeability to water than other aquaporins

A complete permeation event was defined as the movement of a water molecule from one end of the pore lumen to the other end (the pore lumen is the region exhibiting single-file water transport, extending from -8 Å to 12 Å along the pore axis). Figure 7.8 (A) shows the cumulative number of complete permeation events in the wild-type form, and it includes permeation events in two directions, i.e., cytoplasmic side to periplasmic side and vice-versa. The number of permeation events increases steadily in the first 300 ns, then plateaus between 300ns and 400 ns, and finally increases gradually after 400 ns. As shown in Figure 7.8 (B), the number of water molecules in the pore lumen remains more or less the same in the first 400 ns. There is a slight

increase in the number of water molecules after 400 ns, followed by a period in which the number does not vary much. The occasional plateauing of the number of permeation events, and the slight increase in the number of water molecules, suggests that the channel exhibits different conductance states over the course of the simulation. The transition between the different states is likely to be regulated by the ar/R selectivity filter, since this region forms the narrowest part of the pore lumen. Visual inspection revealed that the F48 residue from the ar/R selectivity filter undergoes conformational changes over the course of the simulation. Snapshots of the phenylalanine sidechain over the course of the simulation are shown in Figure 7.8 (C). One can see conformations in which the aromatic ring is directed toward the pore lumen, thereby occluding it partially. The sidechain dihedral angles of F48 were examined, and these are shown in Figure 7.8 (D). The phenylalanine is seen to fluctuate between two conformations, one in which the χ_2 dihedral angle varies between -120° to -60°, and one in which χ_2 varies between 60° to 120°. After about 280 ns, the residue remains in the latter conformational state until about 430 ns. This is around the time when the permeation events slow down, indicating that the reduced permeability could be a consequence of the conformational change in the phenylalanine. After 430 ns, the phenylalanine remains in the alternative conformation for most of the time.

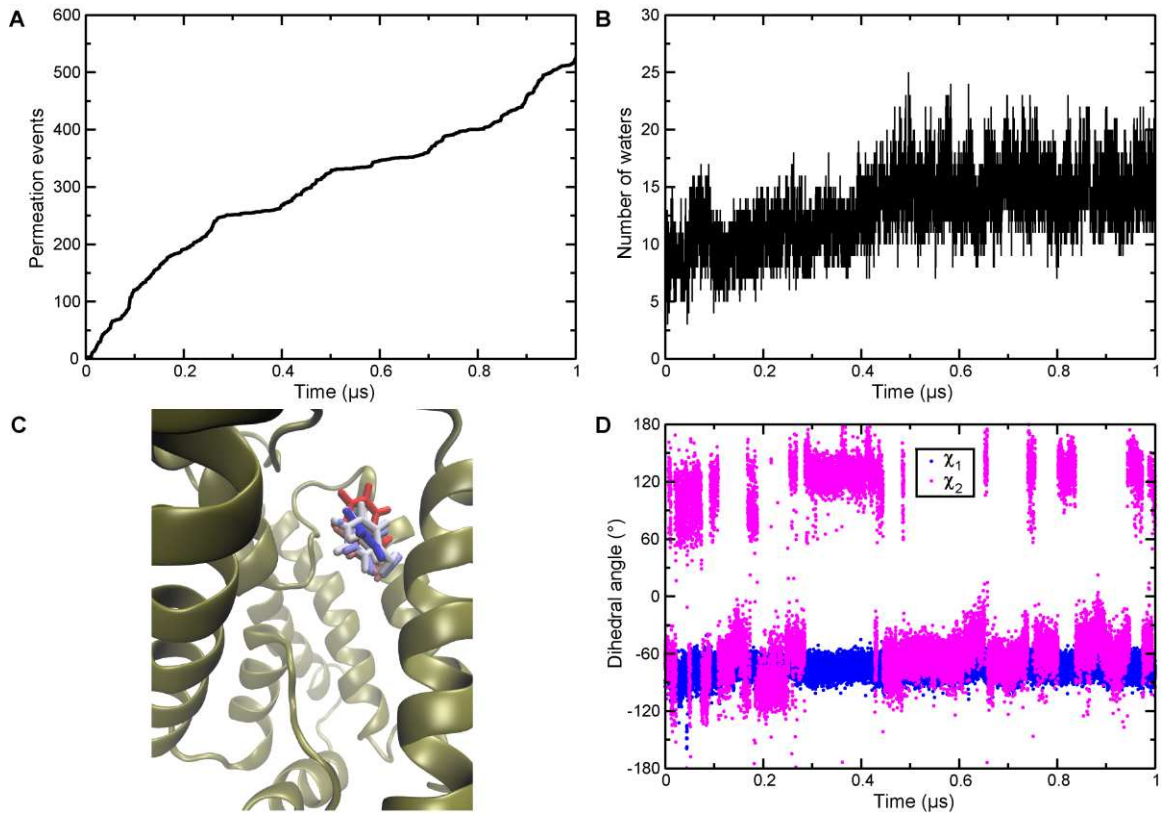

Figure 7.8 (A) Cumulative number of complete permeation events. (B) Number of water molecules in the pore lumen. (C) Snapshots of F48 over the course of the simulation. The pore axis extends along the vertical plane in the figure. Color blue corresponds to the initial phase of the simulation, and red corresponds to the final phase of the simulation. (D) Variation in sidechain dihedral angles of F48.

Table 7.1 lists the total number of permeation events, the average number of pore water molecules, and the permeability coefficients for wild-type and mutant AQP2. For the wild-type channel, the total number of permeation events in the two directions was 524 over the course of 1 μs of simulation (Figure 7.7 (A)), so the number of events in a single direction was considered as 524/2 = 262. The number of permeation events per second would then be $262 \times 10^6 \approx 0.3 \times 10^9$, which is the rate at which aquaporins typically conduct water [2]. Based on this, the diffusive and osmotic permeability coefficients were calculated as described in the Methods section, and were found to be 0.8×10^{-14} cm^3 s^{-1} and 10.9×10^{-14} cm^3 s^{-1}, respectively. As far as the permeability of the mutant is concerned, the average number of pore water molecules is less than that of the wild-type, but the permeability coefficients are higher for the mutant. The increased permeability could be attributed to free movement of water inside the pore without any rotational confinement. For both wild-type and mutant AQP2, the values are higher than those for other

aquaporins, as can be seen in Table 7.2, which lists permeability coefficients for other aquaporins from previous studies. The high permeability for AQP2 can be attributed to the relatively wide pore lumen and the absence of a phenolic barrier at the two ends of the pore lumen, as discussed before. The fact that such rapid water transport across membranes is possible because of AQP2 is indicative of the functionally significant role played by the channel in kidney cells. The permeability coefficients for AQP2 determined in the present study differ from those calculated by Binesh et al. [34]. This is possibly due to a difference in the length of the region that has been defined as pore lumen in the two studies, which leads to a different estimate of the number of water molecules in the pore lumen. Moreover, the study by Binesh et al. is based on a 20 ns simulation, and it is highly unlikely that the properties reported correspond to those of a channel in equilibrium.

Table 7.1 Comparison of permeabilities of wild-type and mutant AQP2

Property*	**Wild-type**	**N184A mutant**
N	12.9	10.4
q_0 (10^{9} s^{-1})	0.3	0.5
p_d (10^{-14} cm^3 s^{-1})	0.8	1.3
p_f (10^{-14} cm^3 s^{-1})	10.9	15.4

*These quantities are defined in Section 7.2

Table 7.2 Permeability coefficients of different aquaporins (p_d and p_f are in 10^{-14} cm^3 s^{-1})

Channel	**Study**	**N**	**p_d**	**p_f**
AQP0	Hashido et al.*	5.9	0.0	0.2
	Jensen et al.†	5.0	0.1	0.3
AQP1	Hashido et al.	6.9	1.4	10.1
AQP4	Hashido et al.	6.9	1.0	7.0
AQP5	Janosi et al.‡	7.1	0.7	3.5
AqpZ	Jensen et al.	7.2	0.4	4.2
	Hashido et al.	6.6	2.0	15.6

*The data shown corresponds to a lumen of length 18 Å.

†This data is for a tetrameric non-junctional form. Two other forms, tetrameric junctional and octameric junctional, were considered in the study, and they showed similar permeability.

‡Values are for the wide open state of the channel.

7.4 Conclusions

The transport properties and conformational dynamics of AQP2 have been studied using extensive unbiased MD simulations. The presence of a continuous water chain in the pore lumen ensures concerted, rapid transport of water molecules. With an energy barrier of only 1.6 kcal mol^{-1}, the transport process is almost diffusion-limited. Water molecules do not encounter a significant barrier around the ar/R region, which is the narrowest part of the channel, and the presence of the large phenylalanine sidechain and the positively charged arginine in the ar/R region suggests the role of this region determining the selectivity of the channel. Water molecules interact electrostatically with the NPA motif, which causes the water molecules to orient themselves such that their dipole moment is perpendicular to the pore axis; this is further confirmed by designing an N184A mutant. With the introduction of the non-polar alanine residue, the constraint on water rotation is removed, and it is proposed that this mutant should allow the transport of protons via a Grotthuss mechanism. Permeability coefficients have been determined, and these show the channel to be more permeable than other aquaporins. The study complements experimental structural biology studies by elucidating how structural features of the channel govern its activity. Furthermore, it motivates the design of functional studies on the channel by suggesting strategies for modulating the permeability of the channel.

References

1. Verkman AS, Anderson MO, Papadopoulos MC (2014) Aquaporins: important but elusive drug targets. Nat Rev Drug Discov 13(4):259–277.
2. Stroud RM, et al. (2003) Selectivity and conductance among the glycerol and water conducting aquaporin family of channels. FEBS Lett 555(1):79–84.
3. Verkman AS (2013) Aquaporins. Curr Biol 23(2):R52-R55.
4. Noda Y, Sohara E, Ohta E, Sasaki S (2010) Aquaporins in kidney pathophysiology. Nat Rev Nephrol 6(3):168-178.
5. Eto K, Noda Y, Horikawa S, Uchida S, Sasaki S (2010) Phosphorylation of aquaporin-2 regulates its water permeability. J Biol Chem 285(52):40777–40784.
6. Deen P, et al. (1994) Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine. Science 264(5155):92–95.

7. Moeller HB, Rittig S, Fenton RA (2013) Nephrogenic diabetes insipidus: Essential insights into the molecular background and potential therapies for treatment. Endocr Rev 34(2):278–301.
8. Frick A, et al. (2014) X-ray structure of human aquaporin 2 and its implications for nephrogenic diabetes insipidus and trafficking. Proc Natl Acad Sci USA 111(17):6305–6310.
9. Wang Y, Tajkhorshid E (2007) Molecular mechanisms of conduction and selectivity in aquaporin water channels. J Nutr 137(6):1509S-1515S.
10. Hub JS, Grubmüller H, de Groot BL (2009) Dynamics and energetics of permeation through aquaporins. What do we learn from molecular dynamics simulations? In Handbook of Experimental Pharmacology. Aquaporins. 190:57-76. Springer Berlin Heidelberg.
11. de Groot BL, Grubmüller H (2001) Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. Science 294(5550):2353-2357.
12. Zhu F, Tajkhorshid E, Schulten K (2004) Theory and simulation of water permeation in aquaporin-1. Biophys J 86(1):50-57.
13. Hashido M, Ikeguchi M, Kidera A (2005) Comparative simulations of aquaporin family: AQP1, AQPZ, AQP0 and GlpF. FEBS Lett 579(25):5549-5552.
14. Wang Y, Schulten K, Tajkhorshid E (2005) What makes an aquaporin a glycerol channel? A comparative study of AqpZ and GlpF. Structure 13(8):1107-1118.
15. Jensen MØ, Mouritsen OG (2006) Single-channel water permeabilities of Escherichia coli aquaporins AqpZ and GlpF. Biophys J 90(7):2270-2284.
16. Yu J, Yool AJ, Schulten K, Tajkhorshid E (2006) Mechanism of gating and ion conductivity of a possible tetrameric pore in aquaporin-1. Structure 14(9):1411-1423.
17. Hashido M, Kidera A, Ikeguchi M (2007) Water transport in aquaporins: osmotic permeability matrix analysis of molecular dynamics simulations. Biophys J 93(2):373-385.
18. Jensen MO, et al. (2008) Dynamic control of slow water transport by aquaporin 0: Implications for hydration and junction stability in the eye lens. Proc Natl Acad Sci USA 105(38):14430–14435.

19. Hub JS, de Groot BL (2008) Mechanism of selectivity in aquaporins and aquaglyceroporins. Proc Natl Acad Sci USA 105(4):1198-1203.
20. Aponte-Santamaría C, Hub JS, de Groot BL (2010) Dynamics and energetics of solute permeation through the Plasmodium falciparum aquaglyceroporin. Phys Chem Chem Phys 12(35):10246-10254.
21. Cui Y, Bastien DA (2011) Water transport in human aquaporin-4: molecular dynamics (MD) simulations. Biochem Biophys Res Commun 412(4):654-659.
22. Aponte-Santamaria C, Briones R, Schenk AD, Walz T, de Groot BL (2012) Molecular driving forces defining lipid positions around aquaporin-0. Proc Natl Acad Sci USA 109(25):9887–9892.
23. Janosi L, Ceccarelli M (2013) The gating mechanism of the human aquaporin 5 revealed by molecular dynamics simulations. PLoS ONE 8(4):e59897.
24. Stansfeld PJ, Jefferys EE, Sansom MS (2013) Multiscale simulations reveal conserved patterns of lipid interactions with aquaporins. Structure 21(5):810-819.
25. Kaptan S, et al. (2015) H95 is a pH-dependent gate in aquaporin 4. Structure 23(12):2309–2318.
26. Tajkhorshid E, et al. (2002) Control of the selectivity of the aquaporin water channel family by global orientational tuning. Science 296(5567):525-530.
27. de Groot BL, Frigato T, Helms V, Grubmüller H (2003) The mechanism of proton exclusion in the aquaporin-1 water channel. J Mol Biol 333(2):279-293.
28. Jensen MØ, Tajkhorshid E, Schulten K (2003) Electrostatic tuning of permeation and selectivity in aquaporin water channels. Biophys J 85(5):2884-2899.
29. Burykin A, Warshel A (2003) What really prevents proton transport through aquaporin? Charge self-energy versus proton wire proposals. Biophys J 85(6):3696-3706.
30. Chakrabarti N, Tajkhorshid E, Roux B, Pomès R (2004) Molecular basis of proton blockage in aquaporins. Structure 12(1):65-74.
31. Burykin A, Warshel A (2004) On the origin of the electrostatic barrier for proton transport in aquaporin. FEBS Lett 570(1):41-46.
32. Chakrabarti N, Roux B, Pomès R (2004) Structural determinants of proton blockage in aquaporins. J Mol Biol 343(2):493-510.

33. Ilan B, Tajkhorshid E, Schulten K, Voth GA (2004) The mechanism of proton exclusion in aquaporin channels. Proteins 55(2):223–228.
34. Binesh AR, Kamali R (2015) Molecular dynamics insights into human aquaporin 2 water channel. Biophys Chem 207:107-113.
35. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem 29(11):1859–1865.
36. Wu EL, et al. (2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J Comput Chem 35(27):1997–2004.
37. Essmann U, et al. (1995) A smooth particle mesh Ewald method. J Chem Phys 103(19):8577.
38. Ryckaert J-P, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341.
39. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802.
40. Best RB, et al. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. J Chem Theory Comput 8(9):3257–3273.
41. MacKerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25(11):1400–1415.
42. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616.
43. Klauda JB, et al. (2010) Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. J Phys Chem B 114(23):7830–7843.
44. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926-935.
45. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graphics 14(1):33–38.

46. Smart OS, Neduvelil JG, Wang X, Wallace B, Sansom MS (1996) HOLE: A program for the analysis of the pore dimensions of ion channel structural models. J Mol Graphics 14(6):354–360.
47. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. J Comput Chem 30(10):1545–1614.

8 Urea Transport through dvUT

8.1 Introduction

Urea plays a central role in the physiological processes of bacteria as well as mammals, and is seen to be involved in diverse processes [1,2]. While it serves as a source of nitrogen in bacteria, in mammals it acts as an osmolyte and aids in water reabsorption in the kidney. The importance of urea in these phenomena has, consequently, necessitated an understanding of the mechanisms by which urea enters and leaves cells. Urea transporters (UT) are a family of transmembrane proteins that allow urea to pass through cell membranes at rates much higher than that seen in typical transporters [3-6]. These are typically trimeric proteins, with each monomeric unit having one continuous pore that spans the entire membrane. Mammalian UTs can also conduct water and homologues of urea like formamide, acetamide, and methylurea [7-10]. Although most UTs occur in kidney cells, certain UTs are known to be expressed in other tissues as well [11-18]. The modulation of urea transport through UTs has been proposed to be a strategy for treating disorders associated with imbalances in water and salt concentration in the kidney [1]. The inhibition of urea transport can lead to an enhancement in the excretion of water as well as electrolytes through urine. This makes inhibitors of UTs potential diuretic agents for the treatment of disorders like congestive heart failure, hypertension, and hepatic cirrhosis.

Molecular simulation studies on several transport proteins have shown that, in general, channels conduct solutes at rates much faster transporters [19-26].UTs allow urea to pass through at very high rates that resemble channel-like behavior, rather than that of transporters [3-5]. The bacterial urea transporter dvUT from the bacterium Desulfovibrio vulgaris has been regarded as an ideal system for investigating urea transport, owing to its striking structural similarity with mammalian urea transporters [6]. Like other urea transporters, dvUT is trimeric, with three identical monomeric units that possess a narrow pore [6]. This narrow region forms the selectivity filter of the transporter, and has three distinct sites that are believed to play a crucial role in the transport process: S_o, S_m, and S_i, where the subscripts denote “outer”, “middle”, and “inner” regions of the pore, respectively (see later). While the S_o and S_i sites are characterized by two phenylalanine residues each, the S_m site has two threonines, in addition to two leucine

residues. The two threonines are conserved across all known urea transporters [6], indicating that they might play a critical role in the transport mechanism of urea. The crystal structure of dvUT has two dimethylurea (DMU) molecules bound in the pore region, one each in the S_o and S_i sites, suggesting that these are likely to be binding sites for urea.

The orientation of the two pairs of phenylalanine rings inside the selectivity filter is such that the phenylalanines from a given pair face each other with the rings almost parallel to each other, separated by a distance of about 7 Å. Such positioning of the rings is expected to be conducive for interactions with permeating urea molecules. The pore of the transporter is, in general, less hydrophilic than urea channels like the proton-gated urea channel HpUreI from Helicobacter pylori [27,28], and is therefore expected to have a higher energy barrier for the permeation of urea. Indeed, a recent molecular dynamics simulation study based on the adaptive biasing force approach showed that the barrier is higher for dvUT, and that the transport is driven by interactions of the urea with pore-facing oxygen atoms in the transporter [29]. However, one aspect that remains to be elucidated is how urea is able to pass through the region of the selectivity filter formed by the phenylalanines without a significant free energy barrier, and how passage of urea through this region influences its orientation and intermolecular interactions in the rest of the pore. In order to provide an explanation for the above phenomenon and further understand the transport mechanism, the present study investigates the mechanism of urea transport through dvUT using umbrella sampling molecular dynamics calculations. Stacking interactions between urea and the phenylalanine rings near the entrance and exit of the selectivity filter are seen to modulate the transport mechanism by governing the dynamics and orientation of the permeating urea molecule. The permeation of multiple urea molecules has also been modeled, revealing a concerted mechanism in which the dynamics of the pore-lining residues and the orientation of urea are intricately related.

8.2 Methods

The crystal structure of dvUT monomer bound to DMU (PDB code: 3K3G) [6] was modeled in a solvated bilayer of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) molecules using the CHARMM-GUI Membrane Builder [30,31], such that the lipid bilayer was in the xy-plane. The system had a total of 59708 atoms, and the protein included all the 337 residues that occur in

a monomeric unit, with missing residues being added using the CHARMM program [32]. The DMU molecules in the S_o and S_i sites of the selectivity filter were replaced with urea molecules such that the coordinates of the urea heavy atoms were the same as the corresponding atoms in DMU. A third urea molecule was modeled in the S_m site such that the molecule was sandwiched between Leu84 and Leu247, and was close enough to the hydroxyl groups on Thr130 and Thr294 so as to be able to form hydrogen bonds; this is in accordance with the location of urea in the S_m site that was proposed in the crystallography study [6].

The system was initially equilibrated with positional restraints on the heavy atoms of the protein and planar restraints on the lipid headgroups, with the restraints being gradually over a period of 400 ps (the first 50 ps of simulation was performed in the NVT ensemble and the later 350 ps in the NPAT ensemble). This was followed by a simulation without any restraints on the protein and lipid for a period of 25 ns in the NPAT ensemble. Positional restraints with force constant 10 kcal mol^{-1} $Å^{-2}$ were applied on the three urea molecules in the above simulations (in the initial 400 ps as well as the following 15 ns) to preserve the orientation of urea inside the transporter. Furthermore, positional restraints (force constant 10 kcal mol^{-1} $Å^{-2}$) were applied on the center of mass of the protein to prevent the drifting of the protein. The temperature of the system was maintained at 303.15 K by performing Langevin dynamics, and covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm [33]. Lennard-Jones interactions were truncated at 12 Å by employing a switching function over a range of 2 Å, and long range electrostatic interactions were computed using the particle mesh Ewald method [34].

Using the above equilibrated structure as the starting structure, umbrella sampling calculations were initially performed with a single urea molecule inside the pore by defining the position z of the urea carbon atom along the pore axis (which is parallel to the z-axis) as the reaction coordinate. For generating the initial conformation for a given window, the urea molecule closest to the concerned window was translated by an appropriate distance to bring it to the center of the window. The remaining two urea molecules were deleted, and two water molecules from bulk water were translated into the pore to fill the space occupied by the deleted urea molecules. For generating windows corresponding to urea lying in bulk water, the carbon atom of the urea molecule was pulled along the pore axis into the bulk water region at a constant velocity of 0.01

Å ps^{-1} using steered molecular dynamics. A total of 135 windows spaced 0.5 Å apart were used, with a harmonic potential with force constant 10 kcal mol^{-1} $Å^{-2}$ being applied on the z-coordinate of the urea carbon atom. Lateral motions of the urea molecule beyond a distance of 8 Å from the pore axis were restrained by applying a harmonic potential with force constant 10 kcal mol^{-1} $Å^{-2}$ on the urea carbon atom, and the drifting of the protein was prevented by applying a restraint on the center of mass of the protein. Each window was simulated for 5 ns, so that the cumulative simulation time was 135*5 ns = 675 ns. The last 4.5 ns from each window was considered for computing the potential of mean force (PMF) using the weighted histogram analysis method (WHAM) [35-37]. The 1D-PMF G(z) thereby computed along the reaction coordinate z was used to compute a 2D-PMF, G(z,r), as a function of the original reaction coordinate z and the radial distance r of the urea carbon atom from the pore axis using the relation

$$G(z,r) = G(z) - RT \ln \rho(r;z) + C \ , \qquad (8.1)$$

where ρ(r; z) denotes the probability of finding the urea carbon atom at different radial distances r for a given value of z, and C is a constant for a given value of the reaction coordinate.

In order to model the permeation of multiple urea molecules, umbrella sampling simulations with two urea molecules inside the pore were performed by defining windows that corresponded to different separations between the two urea molecules for a given position of one of the molecules. For 44 different positions of one urea molecule (z varying from -31 Å to +12 Å with 1 Å spacing), 9 positions of the second urea molecule were sampled, with the separation varying from 6 Å to 14 Å. Each window was sampled for 1.5 ns, so that the total sampling time was 44*9*1.5 ns = 594 ns. The last 1 ns from each window was used for computing a 2D-PMF using two-dimensional WHAM [37].

All simulations were performed using the NAMD program [38] with the CHARMM36 protein force field [39-41], the CHARMM36 lipid force field [42], the CHARMM36 general force field (for urea) [43], optimized parameters for ions [44], and the TIP3P water model [45]. Trajectory analysis was performed using the CHARMM program [46], pore radius profiles were determined using the HOLE2 program [47], and snapshots of molecules were rendered using Visual Molecular Dynamics (VMD) [48].

8.3 Results and Discussion

8.3.1 The selectivity filter is largely dehydrated, with short stretches of polar regions

The transporter is a multi-pass helical transmembrane protein, as can be seen from Figure 8.1 (A), which shows a snapshot of the protein depicting a structure averaged over the last 4.5 ns of simulation from a window corresponding to the urea molecule lying in the bulk water region. The figure also shows a three-dimensional representation of the pore of the protein, which reveals the shape and size of the pore. The pore of the transporter is not linear, and is twisted at a number of places along the principal axis of the transporter. Since the shape of the pore gives an approximate representation of the path that a permeating urea molecule is likely to follow, the twisted nature of the pore indicates that a permeating urea molecule must necessarily follow a non-linear path through the pore (additionally, interactions between urea and pore-lining residues also influence the path of urea). In fact, the umbrella sampling simulations performed here show that the permeating urea molecule follows a non-linear path as described later. It must also be noted that the pore of the transporter is continuous, which is a feature typical of channels, and is therefore different from typical transporters, where either one or both ends of the pore are occluded [25]. Such a pore without any occlusion is also seen in the mammalian urea transporter UT-B and the bacterial proton-gated urea channel HpUreI [5,27,28].

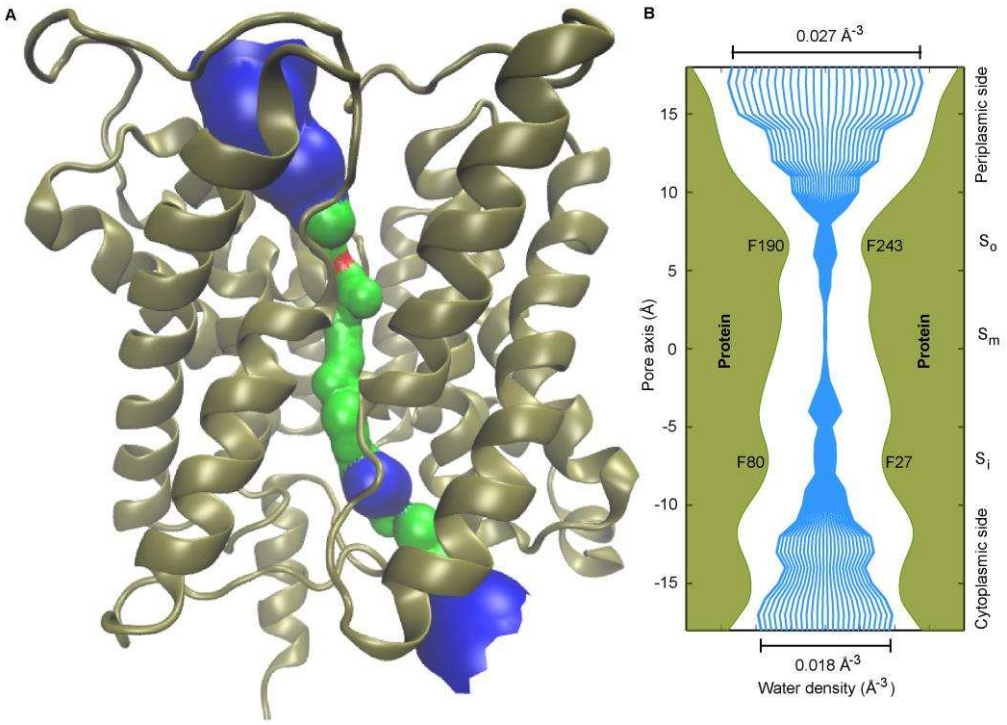

Figure 8.1 The pore of the transporter. (A) A three-dimensional representation of the pore region generated using the HOLE2 program [Smart et al., 1996]. The color of the pore region at a given position indicates the thickness of the pore at that position, with the thickness decreasing in the order blue > green > red. (B) Number density of water (in blue) molecules and pore radius (in green, not to scale). The S_o, S_m, and S_i sites of the selectivity filter are also shown in the figure.

A measure of the degree of solvation in the different parts of the pore is given by the number density of water molecules, which is basically the number of water molecules in a unit cross-section of the pore. The number density was determined by calculating the number of water molecules in cylindrical units of radius 8 Å and height 1 Å along the axis of the pore. Figure 8.1 (B) shows the number density together with the pore radius (not to scale) at the corresponding regions. The regions in the selectivity filter that are crucial in the permeation process, namely S_o, S_m, and S_i. are marked in the figure. There is a decrease in the number density in the constricted region in the middle of the pore, and the decrease in number density is evident around the two pairs of phenylalanines, which form the two major narrow constrictions in the pore. The number density decreases to nearly zero in the middle region of the pore, which has two leucines (Leu84 and Leu247) and two threonines (Thr130 and Thr294), indicating that permeating urea molecules are likely to be dehydrated in this region. However, the two threonines are likely to stabilize urea molecules in this region via formation of hydrogen bonds.

8.3.2 The rate-determining step involves movement of urea from one major binding site to another via a dehydrated region

The residues constituting the selectivity filter are shown in Figure 8.2 (A), and the PMF calculated as a function of the position of the urea carbon atom along the pore axis is shown in Figure 8.2 (B). Since urea transport through bacterial urea transporters usually proceeds from the periplasmic side of the transporter to the cytoplasmic side [4], the figure shows the periplasmic side (starting from z = 29 Å) on the left and the cytoplasmic side on the right (ending at z = -38 Å). The error bars have been calculated by dividing the data from the last 4.5 ns of each window into six blocks, computing the PMF for each block, and then estimating standard deviation and error from the PMFs for the different blocks. Additionally, the PMFs were calculated for different sampling durations, which showed that the PMF is adequately converged within the first 2.25 ns (Figure 8.3).

The PMF has multiple free energy minima, suggesting the presence of multiple urea-binding sites in the pore. The two major binding sites occur around z = 4.5 Å and z = -3.5 Å, and these correspond to an energy release of 3 kcal/mol and 6 kcal/mol, respectively. Movement from the former site to the latter comprises the major rate-determining step (z = 0.5 Å), which involves an energy barrier of 5 kcal/mol. The major urea-binding sites identified from the PMF here are close to the binding sites proposed from the crystal structure of the protein in complex with DMU. Snapshots depicting interactions between the urea and the protein residues occurring at the different minima and maxima are shown in Figure 8.2 (B).

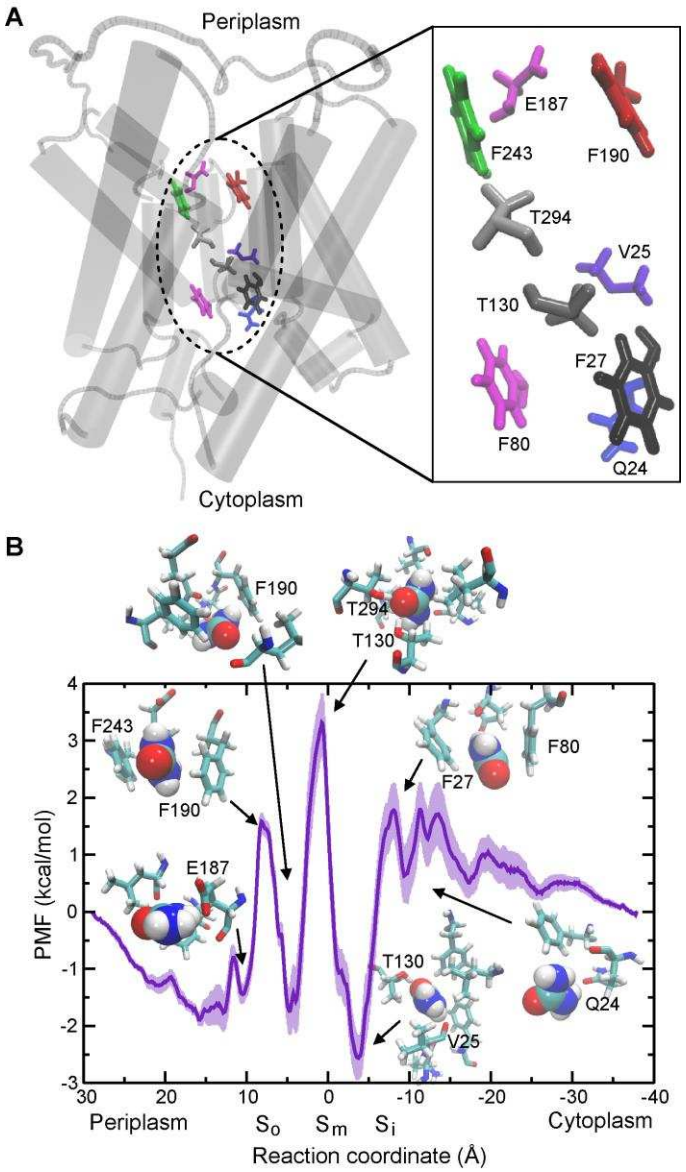

Figure 8.2 (A) Residues lining the selectivity filter. (B) Potential of mean force (PMF) for the transport of urea along the pore of the transporter shown with error bars. Snapshots corresponding to important energy minima and energy maxima are shown around the PMF profile.

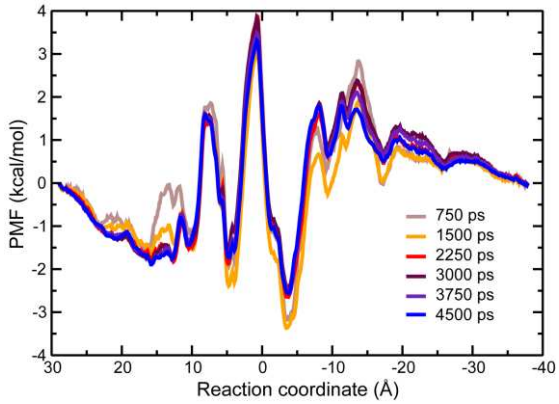

Figure 8.3 PMF corresponding to the permeation of a urea molecule through dvUT for different sampling durations. The first 500 ps have been ignored as equilibration period in all the cases.

As the urea enters the transporter from the periplasmic side, it hydrogen bonds with a sidechain oxygen on Glu187 (z = 10.5 Å). Upon entering the selectivity filter, the urea molecule passes through the narrow hydrophobic region formed by Phe190 and Phe243, which entails the first major energy barrier for the permeating molecule (z = 8 Å). Upon crossing this region, the urea enters a stable arrangement where it is involved in either or both of the following interactions: NH...π interaction involving the aromatic ring of Phe190 and the hydrogen atoms of urea, and hydrogen bond between the backbone oxygen atom of Val188 and the hydrogen atoms of urea (z = 4.5 Å). As the urea crosses the region corresponding to the energy barrier around ~ 0.5 Å, the urea oxygen atom is able to form two hydrogen bonds with sidechain hydroxyl groups from Thr294 and Thr130, but the hydrogen bonds formed by the urea hydrogen atoms with the backbone oxygen atom of Val188 are lost. Furthermore, this region is completely dehydrated (Figure 8.1 (B)), so any stabilization of urea arising from hydration is completely absent. Additionally, there are other factors, such as restriction of the overall rotation of urea imposed by the formation of urea-protein hydrogen bonds, that lead to the rise in PMF at this position, and these are discussed in the later sections. Further movement of urea is accompanied by the formation of hydrogen bonds between urea hydrogen atoms and the backbone oxygen atom of Val25 (z = -3.5 Å). While the hydrogen bond between the urea oxygen atom and Thr130 is still retained at this position, interactions between urea and Thr294 are lost. Another major energy barrier is encountered by the urea as it passes through the second hydrophobic constriction formed by Phe80 and Phe27. The last major binding site for the urea occurs around Gln24, where the urea hydrogen atoms interact with the sidechain oxygen atom of the glutamine (z = -9.5 Å), before exiting the pore.

8.3.3 Lateral motions of the urea become confined as it passes through the constricted middle region

To determine the exact path of urea through the pore of the transporter, the PMF was computed as a function of the original reaction coordinate and the radial distance of the urea carbon atom from the axis of the pore (the radial distance was calculated by finding out the distance between

the location of the urea carbon atom in the xy-plane and the pore axis, the latter being given by (x,y) = (0,0) in the same plane). The two-dimensional PMF (2D-PMF) thereby calculated using equation 8.1 is shown in Figure 8.4, and details of the computation of the 2D-PMF are described in the Methods section. As the urea enters the pore from the periplasmic side, it prefers to stay away from the pore axis to facilitate interactions with the pore-lining residues. However, once it is inside the narrow region of the pore, movements of the molecule in a direction perpendicular to the pore axis are restricted. Two distinct minima can be seen in this stretch, and the location of these minima is consistent with the minima seen in the 1D-PMF. Around z = -18 Å, as the urea leaves the pore, it is seen to move away from the pore axis, which is reflected in the 2D-PMF by the complete absence of configurations with urea close to the pore axis. Figure 8.5 (A) shows snapshots showing positions of the urea carbon atom in the last frame of each trajectory in the umbrella sampling simulations. The loop obstructing the path of the urea at the cytoplasmic end of the pore can be seen at the bottom of the figure, with the urea molecule deviating from its path along the pore axis. The exact path of the urea carbon atom can be seen in Figure 8.5 (B), which shows the most probable position of the atom along the x- and y-axes for different locations along the pore axis (z-axis). Consistent with the observation described above, the urea molecule's movements radially away from the pore axis (x, y = 0, 0) are confined in the region -8 < z < 8, which is a constricted region.

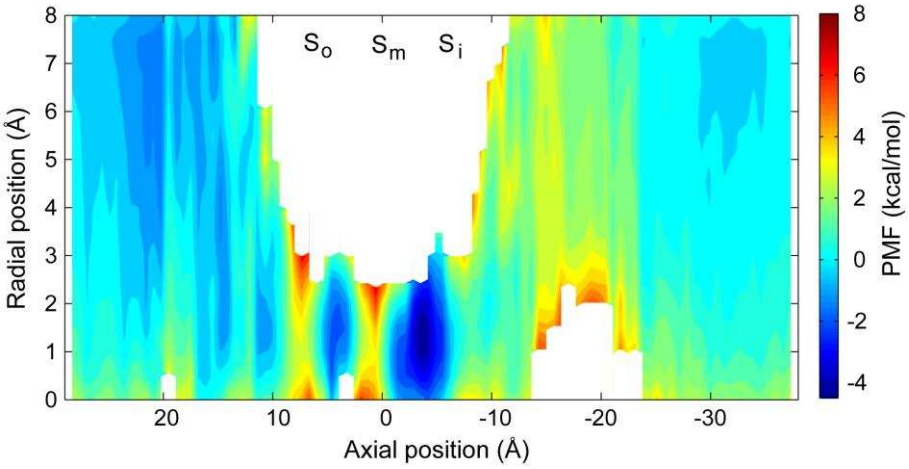

Figure 8.4 2D-PMF as a function of the original reaction coordinate (axial position of urea carbon atom) and radial distance of urea carbon atom from pore axis. The axial positions of the S_o, S_m, and S_i sites is also indicated.

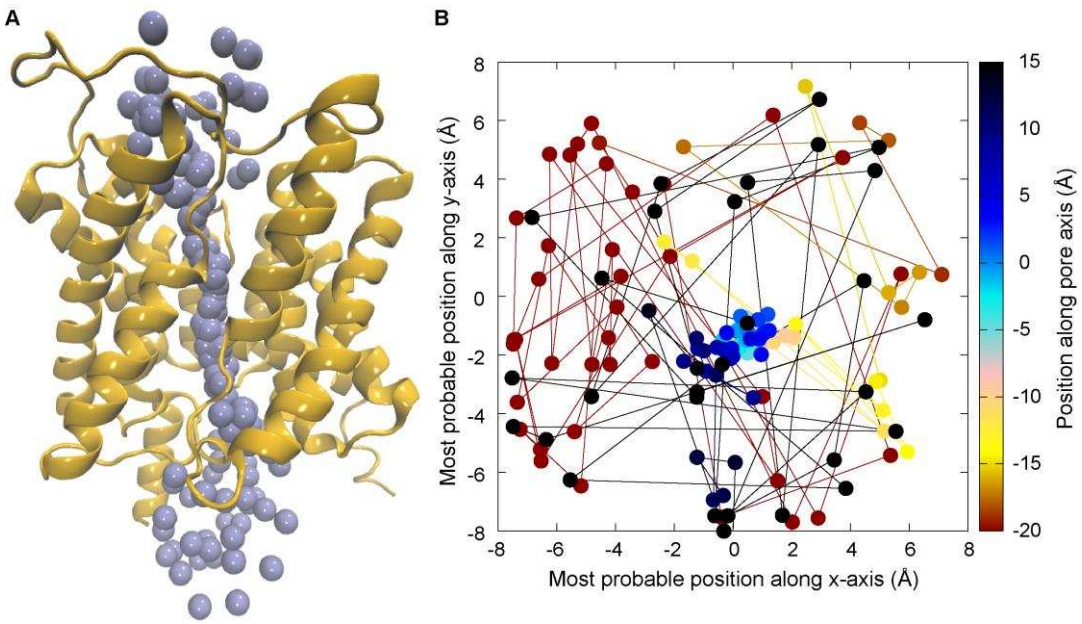

Figure 8.5 (A) Location of the urea carbon atom in the final frame of each trajectory. (B) Most probable position of the urea carbon atom along the x- and y-axes for each window.

8.3.4 Hydrogen bonds influence the energetics of transport

Figure 8.6 (A) shows the average interaction energy between urea and the protein for all positions along the reaction coordinate. The interaction energy becomes favorable as the urea molecule enters the pore, owing to the presence of a number of polar residues in the pore lumen. Particularly strong interactions between urea and the protein occur in the vicinity of Glu187 (z ~ 10.5 Å), Phe190, Val188 (~ 4 Å), Thr130, and Val25 (~ -3.5 Å). Decomposition of the urea-protein interaction energy into electrostatic and van der Waals components shows that electrostatic interactions are dominant, determining the location of the major binding sites, but the van der Waals component reveals an interesting role of the Phe residues lining the pore. An interesting point to note is the occurrence of two minima in the van der Waals component corresponding to positions of urea in the vicinity of Phe190/Phe243 and Phe27/Phe80, indicating dispersion interactions between urea and the aromatic rings. The possibility of stacking interactions between urea and the aromatic rings is later verified by means of appropriate geometric criteria (see later). However, since these interactions do not correspond to a minimum energy state in the free energy profile (Figure 8.2), it follows that the stacking influences the transport mechanism by determining the orientation of the urea molecule inside the pore, aiding selectivity. For investigating the effect of water on the energetics of transport, the interaction

energy between urea and water was computed, as shown in Figure 8.6 (B). It can be seen that the movement of urea from bulk water to the interior of the pore is accompanied by significant desolvation, owing to the low water density inside the pore. However, this desolvation is compensated for by interactions between urea and the protein, which greatly stabilize the urea inside the pore.

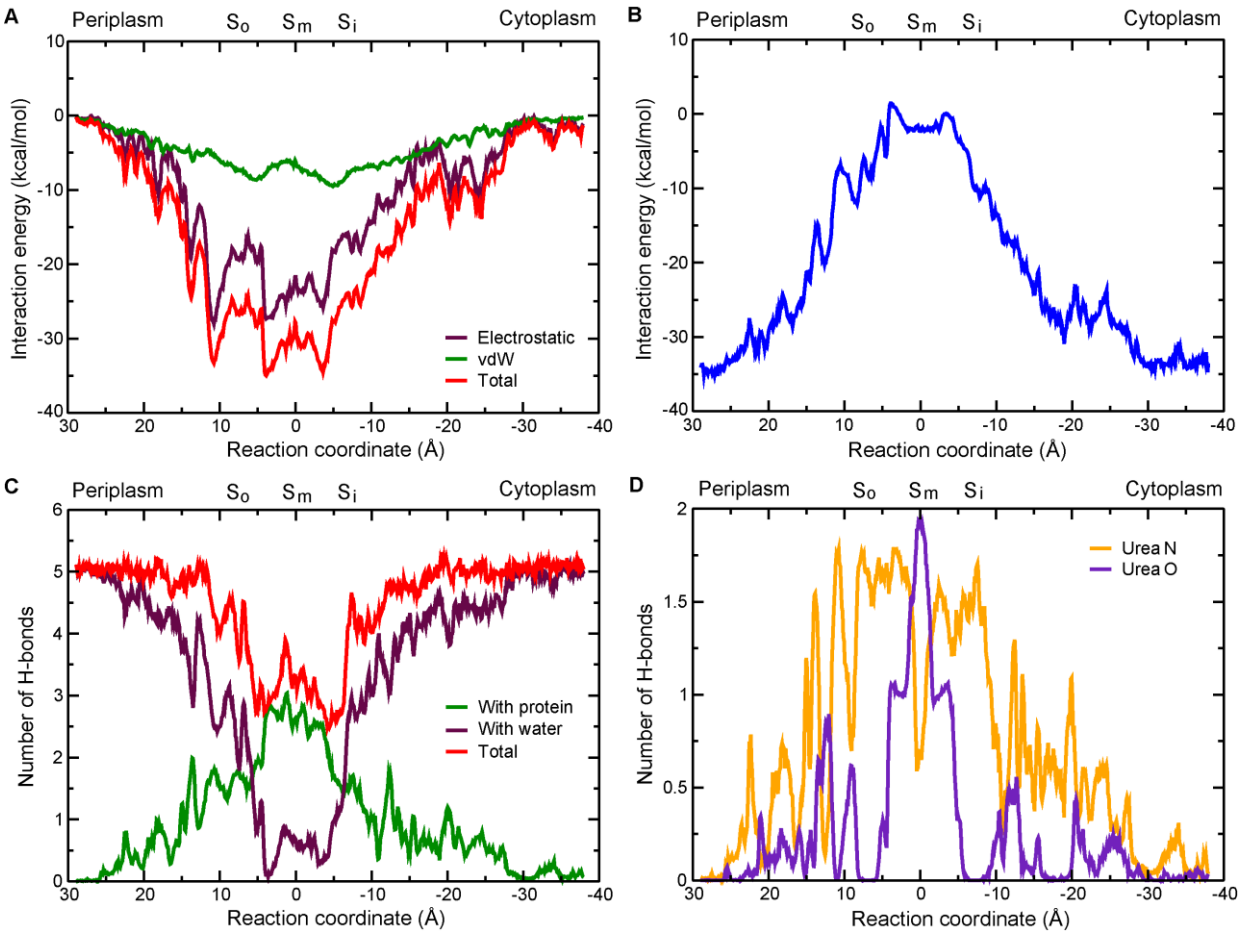

Figure 8.6 Nonbonded interactions between urea and protein/water. (A) Interaction energy between urea and protein, shown with the electrostatic and van der Waals components. (B) Interaction energy between urea and water. (C) Number of hydrogen bonds between urea and protein/water. (D) Number of hydrogen bonds between protein and nitrogen or oxygen atoms of urea. Data for all the windows has been combined, divided into bins along the reaction coordinate, and then averaged for each bin.

The fact that the nature of the interaction between urea and the protein is largely electrostatic suggests the presence of a considerable number of hydrogen bonds between urea and the protein. The number of hydrogen bonds formed by urea was therefore investigated, and is shown in Figure 8.6 (C). A hydrogen bond was said to occur if the distance between the hydrogen atom

and the hydrogen bond acceptor was less than 2.4 Å. As is evident from the figure, hydrogen bonds with water are almost completely lost in the middle region of the pore, but the energetic penalty arising from this is compensated for by the formation of hydrogen bonds with the protein. To find out if these hydrogen bonds with the protein are formed by the urea amide group or the urea carbonyl group, a decomposition of the number of hydrogen bonds formed by these two groups was carried out using the distance-based criterion for hydrogen bond formation described above. For most of the time, it is the amide group on urea that is involved in hydrogen bond formation, as shown in Figure 8.6 (D). The bond acceptors are usually sidechain or backbone oxygen atoms on the protein, as discussed earlier in the section on the PMF. However, when the urea is in the vicinity of Thr130 and Thr294, interactions between the urea oxygen atom and the hydroxyl group on the threonines become more important, with hydrogen bonds being formed with both the threonine groups for a short stretch inside the pore ($z \approx 0$ Å). Interestingly, this arrangement corresponds to a peak in the PMF, in spite of the occurrence of two hydrogen bonds. In addition to the dehydration of urea that occurs here, another possible reason for this peak in the free energy profile is that, for the urea to preserve the two hydrogen bonds, it must stay in a fixed arrangement with respect to the two threonines. This would restrict the motions of the urea molecule, which is entropically unfavorable, and leads to a rise in the free energy.

8.3.5 Stacking interactions between phenylalanine and urea are crucial for selectivity

The occurrence of minima in the van der Waals component of protein-urea interaction energy at positions corresponding to the pore-facing phenylalanine rings suggests stacking interactions between urea and the rings. To determine if urea actually stacks with the phenylalanine rings or not, the angle between the normal to the plane of urea and the normal to the plane of each of the phenylalanine rings was measured. Figure 8.7 shows these angles for the four phenylalanines along with the distances between the urea carbon atom and the corresponding phenylalanine ring. In these figures, the range of values shown for the reaction coordinate includes only those values that correspond to urea lying in the vicinity of the respective ring (i.e., +10 to 0 for Phe190/Phe243, and 0 to -10 for Phe80/Phe27). When the urea is away from a given ring (the distances between urea carbon and the ring show how far away the urea is), the angle between the two normals is able to sample a broad range of values. However, as the urea approaches the

aromatic ring, the angle between the two normals decreases to about 10°-20°. Thus, the plane of urea and the plane of the aromatic ring are almost parallel to each other, clearly indicating stacking arrangement between the two. The angle between the N-N vector of urea and the pore axis was also measured, and is shown in Figure 8.8 (A). The urea is seen to be aligned along the pore axis while it is inside the selectivity filter, an arrangement that makes it possible for the urea to stack with the phenylalanine rings (the urea makes an angle with the pore axis only around ~ 4.5 Å and ~ -5.5 Å, where it tilts to form NH...π interactions with the phenylalanine rings). Figure 8.8 (B) shows the angle between the C-O vector of urea and the x-axis, and quantifies the rotation of the urea molecule in the xy-plane. It is seen that the rotation of the urea molecule is hindered when it is inside the selectivity filter, which suggests that the phenylalanine rings govern the orientation of urea as it enters the selectivity filter and permeates through it.

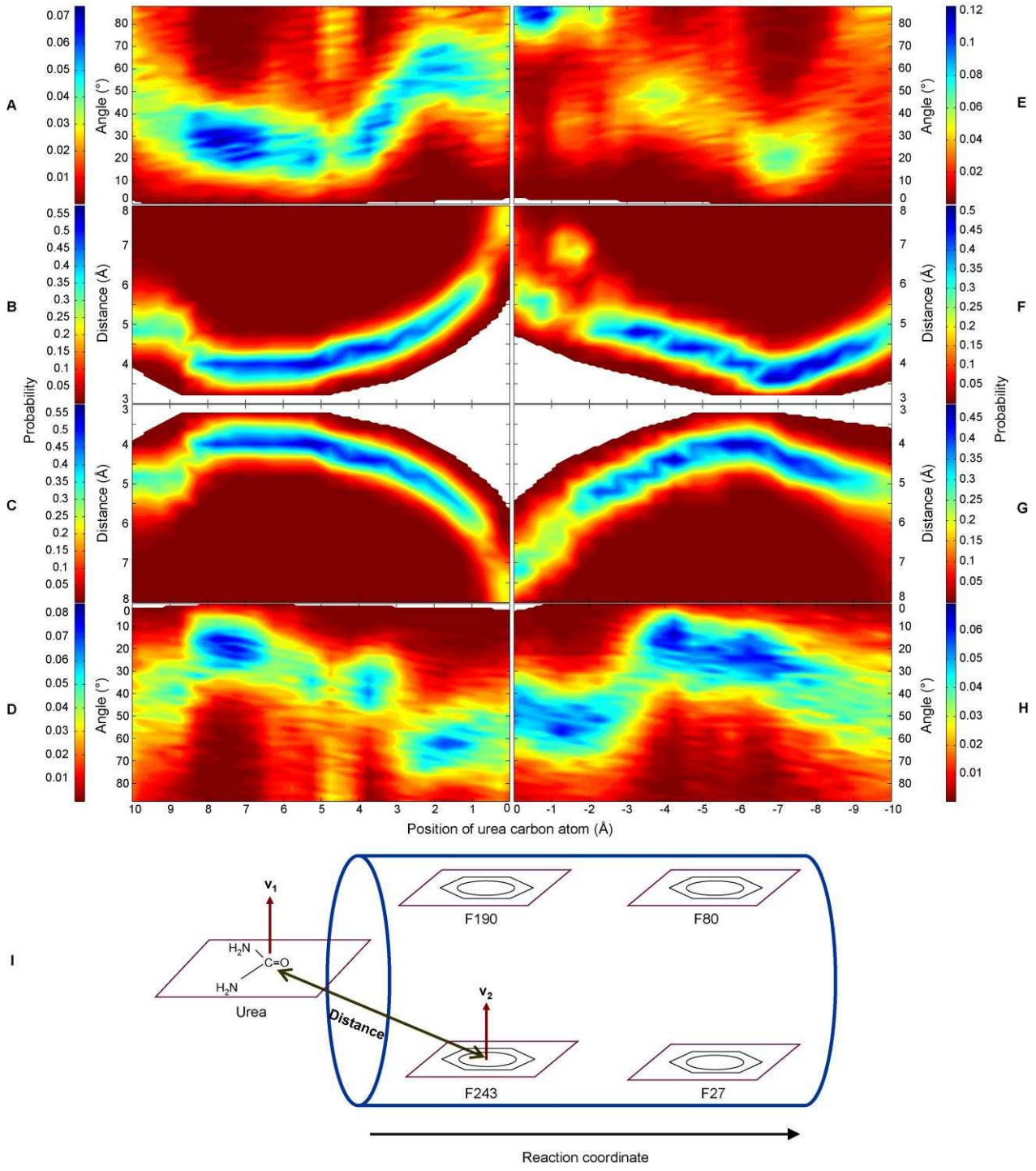

Figure 8.7 Probability distributions for the angle between the normal to the plane of urea and the normal to the plane of the different phenylalanine rings, and for the distance between the urea carbon and the center of mass of the different phenylalanine rings. (A, B) Phe190 (C, D) Phe243 (E, F) Phe80 (G, H) Phe27 (I) Schematic showing the distance and the angle that are being measured here. While v_1 is a vector perpendicular to the plane of urea, v_2 is a vector perpendicular to the plane of the phenylalanine ring. The angle between v_1 and v_2 is being measured here.

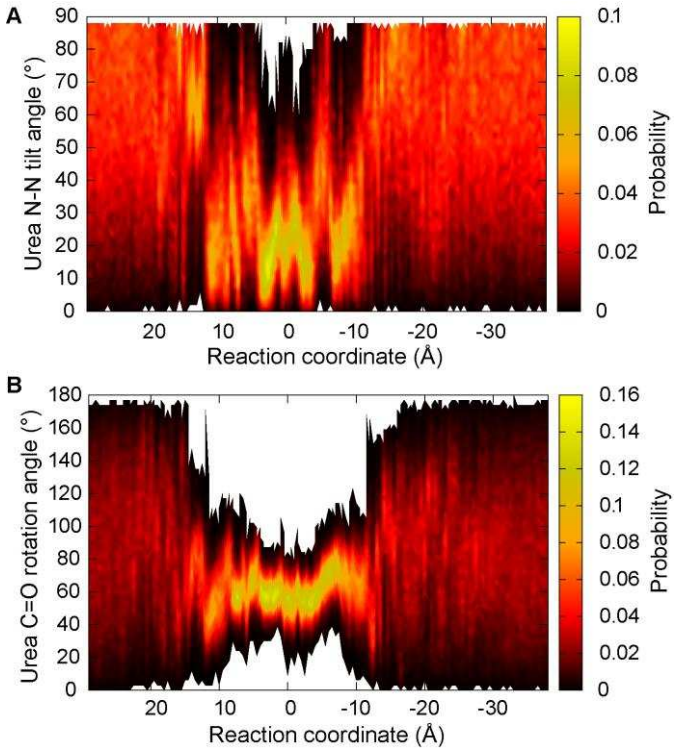

Figure 8.8 (A) The tilt angle of the N-N vector in urea with respect to the pore axis. (B) Angle between the C=O vector in urea and the x-axis.

To elucidate how the approach of urea affects the overall dynamics of the phenylalanine rings in the pore, two quantities were computed: the distance between the centers of mass two rings for a given pair of phenylalanines (Phe190-Phe243 and Phe80-Phe27), and the angle between the normals to the plane of the ring for a given pair of phenylalanines. Figure 8.9 shows these distances and angles as a function of the reaction coordinate z (position of urea carbon atom), where values of $z \approx 7$ correspond to urea near Phe190/Phe243, and values of $z \approx -7$ correspond to urea near Phe80/Phe27. When there is no urea in the vicinity of Phe190/Phe243, the distance between these two rings is 6 Å – 6.5 Å. As urea enters the space between these two rings, the distance between the two rings increases to about 8 Å, showing that the rings reorient to effectively widen the pore, and allow the urea to pass through. Similar widening of the pore is seen when urea passes in between Phe80/Phe27, leading to an increase in the distance between the two rings. Such a breathing motion of the phenylalanine residues seems to control the selectivity of the transporter. The existence of such a possible mechanism in the other urea transporters is discussed in the following section. The widening of the pore is evident in Figure

8.10, which shows pore radius profiles for the transporter corresponding to urea occurring at different regions inside the transporter. As far as the angle between the rings is concerned, the angle between Phe190 and Phe243 varies from 60° to 70°, and the angle between Phe80 and Phe27 varies from 35° to 60°. The fact that the angle between a given pair of rings is confined to a defined range of values suggests that two rings in a given pair retain a defined orientation with respect to each other even in the absence of urea. As the urea approaches the rings, the angle between a given pair of rings falls to about 40° for Phe190/Phe243 as well as Phe80/Phe27. Thus, the rings become somewhat parallel to each other, so that the urea molecule stacks with both rings at the same time.

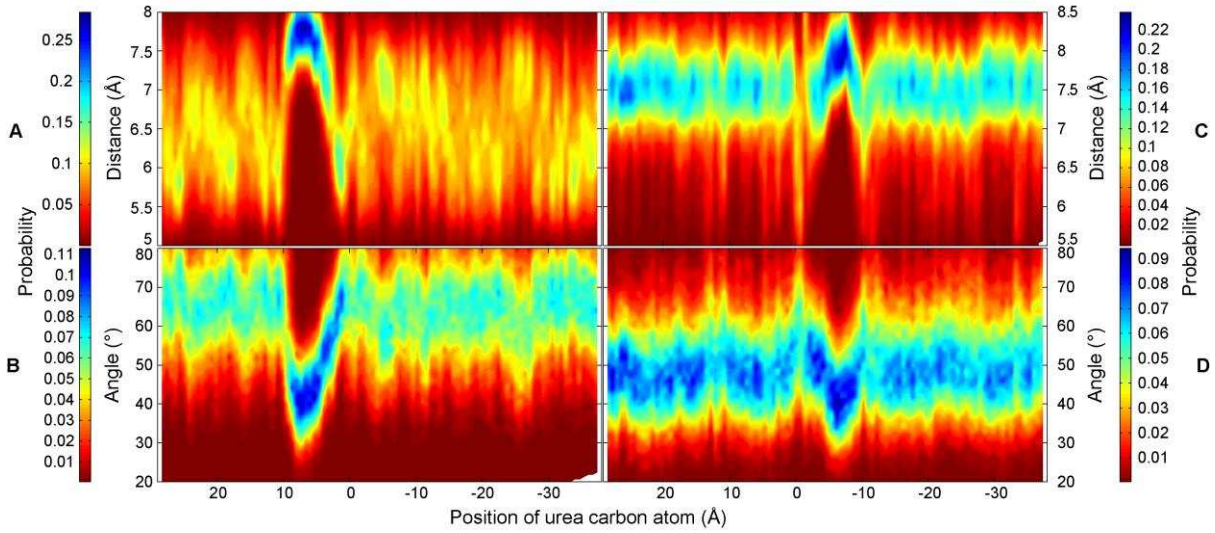

Figure 8.9 Probability distributions of the distance between the centers of mass of two rings in a given pair of phenylalanines, and the angle between normals to the planes of two rings in a given pair of phenylalanines. (A, B) Phe190/Phe243 (C, D) Phe80/Phe27.

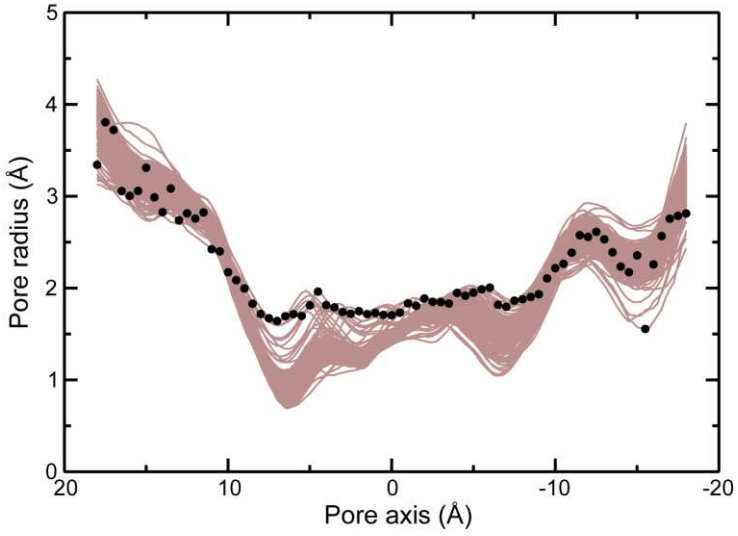

Figure 8.10 Pore radius profiles for all the windows in the umbrella sampling calculations. An average structure for the transporter has been used for each window, and a given line corresponds to a single window. Each line has a black dot, which denotes the position of the urea carbon atom (along the pore axis) for that particular window. It can be seen that, in the windows in which urea is in the middle region of the pore, the constricted region has widened in comparison to windows in which urea is in terminal positions.

Such a parallel arrangement of pore-facing aromatic rings is also seen in the mammalian urea transporter UT-B and the bacterial urea channel HpUreI [5,27], as shown in Figure 8.11. The conservation of such an arrangement across the structures of all known urea-conducting transport proteins suggests a role for these residues in the transport of urea. The distance between the aromatic rings of these residues lies in the range 6 Å – 10 Å, so that a permeating urea molecule can interact with both rings at the same time. The roughly parallel arrangement of the rings makes it possible for urea to form a stacking arrangement with these rings. It follows that stabilization of a permeating urea molecule via such stacked arrangements involving dispersion forces is possibly a selectivity mechanism. Other species that are commonly transported across the membrane, like cations, for instance, cannot possibly form such a stable arrangement in the region lined with these aromatic rings, and such species would instead encounter a significant hydrophobic barrier in this region.

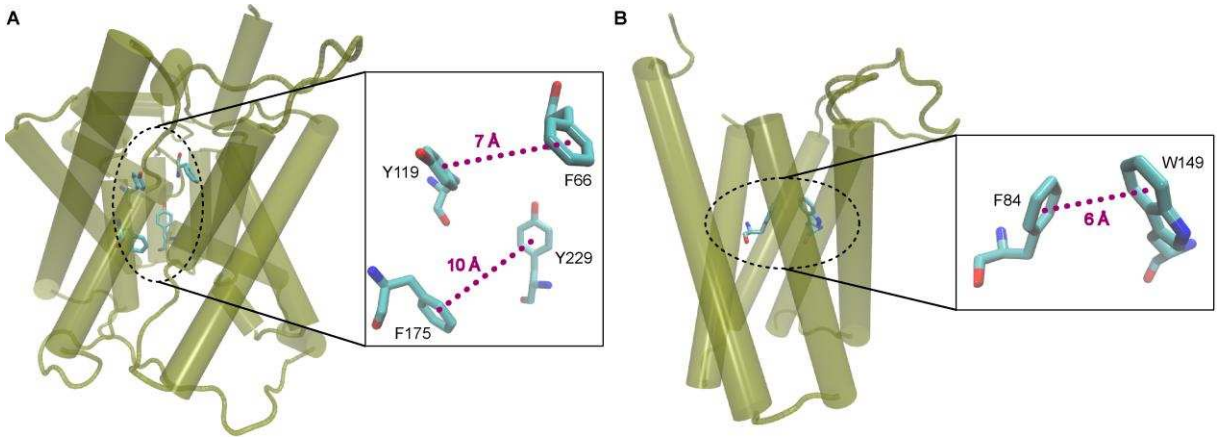

Figure 8.11 Pore-lining aromatic residues with rings exhibiting parallel arrangement in (A) mammalian urea transporter UT-B (PDB ID: 4EZD) and (B) bacterial urea channel HpUreI (PDB ID: 3UX4). The distance between the rings is also shown.

8.3.6 The possibility of permeation of multiple urea molecules

The PMF for the permeation of two urea molecules was computed to investigate the possibility of multiple urea permeation. Figure 8.12 shows the free energy landscape as a function of the position z of the carbon atom of the first urea molecule, and the separation d between the carbon atoms of the two urea molecules. The windows sampled include all configurations intermediate between the two urea molecules entering the pore from the periplasmic side, and the two molecules exiting from the cytoplasmic side. The path with the minimum free energy involves an energy barrier of about 9 kcal/mol, indicating that it is possible for the selectivity filter to accommodate more than one urea molecule. This is in good agreement with the original crystallographic structure, which has two DMU molecules in the pore of the transporter [6]. As described below, more than one urea is allowed to remain inside the selectivity filter at a given instant as a result of a number of stabilizing interactions with the pore-lining residues. Thus, in spite of possessing a largely dehydrated pore, the transporter allows the permeation of more than one molecule at a time, making possible rapid conduction of molecules.

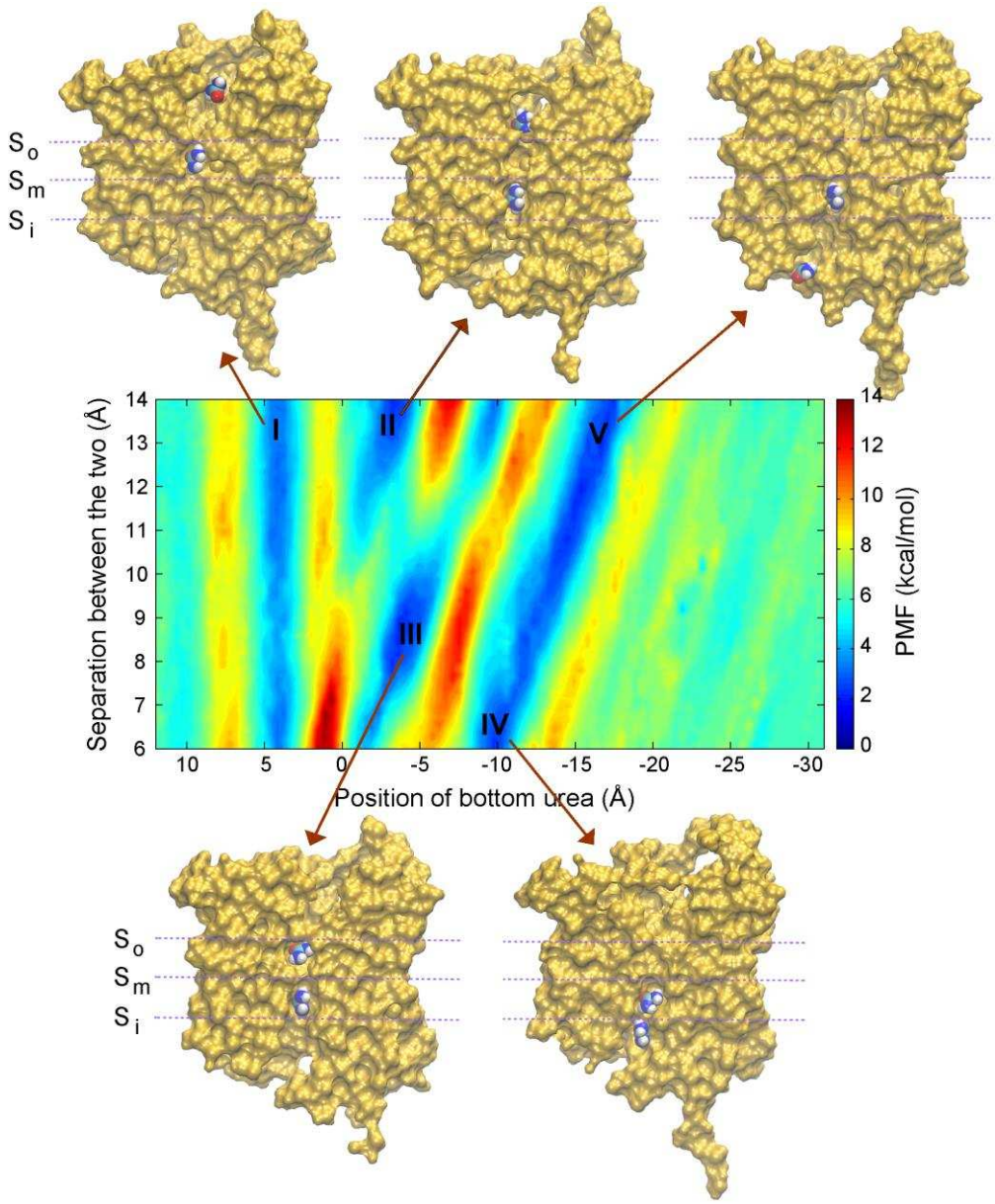

Figure 8.12 PMF shown as a function of the position of the bottom urea molecule, and the separation between the top urea molecule and the bottom urea molecule. Snapshots depicting the location of the two urea molecules are shown for the major intermediates in the process.

For configurations corresponding to the first urea at the entrance of the pore and the second urea in bulk water ($z > 10$, extreme left in Figure 8.12), the free energy landscape is flat, indicating that different separations between the two urea molecules is possible. The first stable intermediate is encountered when the first urea molecule reaches Thr130 in the middle of the selectivity filter ($z \approx -4$ Å). The first urea molecule is later stabilized by the backbone oxygen

atom of Val25, with the second urea hydrogen bonding with Glu187 at the same time ($z \approx -3$ Å, $d \approx 14$ Å). Slight rearrangement of the first urea is then accompanied by movement of the second urea from Glu187 to the region below the aromatic rings of Phe190/Phe243 ($z \approx -4$ Å, $d \approx 9$ Å). While the first urea is seen hydrogen bonding with Thr130 at this position, the second urea interacts with Phe190 via NH...π interactions. As the first urea molecule approaches and forms hydrogen bonds with Gln24 near the exit of the selectivity filter, the second urea molecule moves closer to the first urea to be able to interact with Thr130 ($z \approx -10$ Å, $d \approx 6$ Å). The second urea remains in this hydrogen bonded state until the first urea has exited the pore ($z \approx -17$ Å, $d \approx 14$ Å). The permeation process thus follows a concerted mechanism in which the two urea molecules move from one stable state to another, with the separation between the two urea molecules being governed by the location of the residues interacting with these urea molecules.

8.4 Conclusions

A detailed mechanism of urea transport through the urea transporter dvUT has been proposed based on umbrella sampling free energy calculations. The free energy profile reveals binding sites close to the S_o and S_i sites in the selectivity filter, which consist of two phenylalanine rings each with their planes arranged parallel to each other. The urea is shown to stack with these rings at the entrance of the selectivity filter, resulting in an orientation for the urea that is preserved as it permeates through the rest of the pore. However, the stacking interactions have a marginal effect on the energetics of transport, meaning that these interactions influence the transport process by controlling the dynamics rather than the energetics of urea inside the pore. Lateral motions of urea are seen to be confined as it passes through the narrow selectivity filter, as revealed by a two-dimensional free energy landscape. The possibility of multiple urea permeation has been investigated, and results show that the selectivity filter can accommodate more than one urea molecule, as proposed previously in experimental studies [6]. The permeation of multiple urea molecules is reported to follow a mechanism in which the permeating molecules move from one binding site to another in a synchronized manner. The findings from the study report the significance of stacking interactions between urea and aromatic side chains for the first time for this class of transporters, and show how it governs the overall transport process.

References

1. Esteva-Font C, Anderson MO, Verkman AS (2014) Urea transporter proteins as targets for small-molecule diuretics. Nat Rev Nephrol 11(2):113–123.
2. Shayakul C, Clémençon B, Hediger MA (2013) The urea transporter family (SLC14): Physiological, pathological and structural aspects. Mol Aspects Med 34(2-3):313–322.
3. MacIver B, Smith CP, Hill WG, Zeidel ML (2008) Functional characterization of mouse urea transporters UT-A2 and UT-A3 expressed in purified Xenopus laevis oocyte plasma membranes. Am J Physiol-Renal 294(4):F956-F964.
4. Raunser S, et al. (2009) Oligomeric structure and functional characterization of the urea transporter from Actinobacillus pleuropneumoniae. J Mol Biol 387(3):619–627.
5. Levin EJ, et al. (2012) Structure and permeation mechanism of a mammalian urea transporter. Proc Natl Acad Sci USA 109(28):11194–11199.
6. Levin EJ, Quick M, Zhou M (2009) Crystal structure of a bacterial homologue of the kidney urea transporter. Nature 462(7274):757-761.
7. Yang B (1998) Urea transporter UT3 functions as an efficient water channel. Direct evidence for a common water/urea pathway. J Biol Chem 273(16):9369–9372.
8. Yang B, Verkman AS (2002) Analysis of double knockout mice lacking aquaporin-1 and urea transporter UT-B: Evidence for UT-B-facilitated water transport in erythrocytes. J Biol Chem 277(39):36782–36786.
9. Azouzi S, et al. (2013) Energetic and molecular water permeation mechanisms of the human red blood cell urea transporter B. PLoS ONE 8(12): e82338.
10. Zhao D, Sonawane N, Levin MH, Yang B (2007) Comparative transport efficiencies of urea analogues through urea transporter UT-B. Biochim Biophys Acta, Biomembr 1768(7):1815–1821.
11. Timmer RT, Klein JD, Bagnasco SM, Doran JJ, Verlander JW, Gunn RB, Sands JM (2001) Localization of the urea transporter UT-B protein in human and rat erythrocytes and tissues. Am J Physiol Cell Physiol 281(4):C1318-C1325.
12. Collins D, et al. (2009) Differential protein abundance and function of UT-B urea transporters in human colon. Am J Physiol Gastrointest Liver Physiol 298(3): G345-G351.

13. Inoue H, et al. (2004) Identification and characterization of a Kidd antigen/UT-B urea transporter expressed in human colon. Am J Physiol Cell Physiol 287(1): C30-C35.
14. Tsukaguchi H, et al. (1997) Cloning and characterization of the urea transporter UT3: localization in rat kidney and testis. J Clin Invest 99(7):1506–1515.
15. Kwun Y-S, et al. (2003) Immunohistochemical localization of urea transporters A and B in the rat cochlea. Hear Res 183(1-2):84–96.
16. Duchesne R, et al. (2001) UT-A Urea transporter protein in heart: Increased abundance during uremia, hypertension, and heart failure. Circ Res 89(2):139–145.
17. Fenton RA, Cooper GJ, Morris ID, Smith CP (2002) Coordinated expression of UT-A and UT-B urea transporters in rat testis. Am J Physiol Cell Physiol 282(6): C1492-C1501.
18. Klein JD, Timmer RT, Rouillard P, Bailey JL, Sands JM (1999) UT-A urea transporter protein expressed in liver upregulation by uremia. J Am Soc Nephrol 10(10):2076-2083.
19. Medovoy D, Perozo E, Roux B (2016) Multi-ion free energy landscapes underscore the microscopic mechanism of ion selectivity in the KcsA channel. Biochim Biophys Acta, Biomembr 1858(7):1722–1732.
20. Jiang T, Han W, Maduke M, Tajkhorshid E (2016) Molecular basis for differential anion binding and proton coupling in the Cl^-/H^+ exchanger ClC-ec1. J Am Chem Soc 138(9):3066–3075.
21. Cheng MH, Bahar I (2015) Molecular mechanism of dopamine transport by human dopamine transporter. Structure 23(11):2171-2181.
22. Pothula KR, Solano CJ, Kleinekathöfer U (2015) Simulations of outer membrane channels and their permeability. Biochim Biophys Acta, Biomembr 1858(7):1760-1771.
23. Samanta S, Scorciapino MA, Ceccarelli M (2015) Molecular basis of substrate translocation through the outer membrane channel OprD of Pseudomonas aeruginosa. Phys Chem Chem Phys 17(37):23867-23876.
24. Padhi S, Priyakumar UD (2016) Cooperation of hydrophobic gating, knock-on effect, and ion binding determines ion selectivity in the p7 channel. J Phys Chem B 120(19):4351-4356.
25. Padhi S, Priyakumar UD (2015) Ion hydration dynamics in conjunction with a hydrophobic gating mechanism regulates ion permeation in p7 viroporin from hepatitis C virus. J Phys Chem B 119(20):6204-6210.

26. Padhi S, Burri RR, Jameel S, Priyakumar UD (2014) Atomistic detailed mechanism and weak cation-conducting activity of HIV-1 Vpu revealed by free energy calculations. PLoS ONE 9(11):e112983.
27. Strugatsky D, et al. (2012) Structure of the proton-gated urea channel from the gastric pathogen Helicobacter pylori. Nature 493(7431):255–258.
28. McNulty R, Ulmschneider JP, Luecke H, Ulmschneider MB (2013) Mechanisms of molecular transport through the urea channel of Helicobacter pylori. Nat Commun 4:2900.
29. Wang Z, et al. (2015) Computation and simulation of the structural characteristics of the kidney urea transporter and behaviors of urea transport. J Phys Chem B 119(16):5124–5131.
30. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem 29(11):1859–1865.
31. Wu EL, et al. (2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J Comput Chem 35(27):1997–2004.
32. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. J Comput Chem 30(10):1545–1614.
33. Ryckaert J-P, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341.
34. Essmann U, et al. (1995) A smooth particle mesh Ewald method. J Chem Phys 103(19):8577.
35. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13(8):1011–1021.
36. Roux, B (1995) The calculation of the potential of mean force using computer simulations. Comp Phys Comm 91(1-3):275-282.
37. Grossfield, A (2012) "WHAM: the weighted histogram analysis method", version 2.0.9, http://membrane.urmc.rochester.edu/content/wham.
38. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802.

39. Best RB, et al. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. J Chem Theory Comput 8(9):3257–3273.
40. MacKerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25(11):1400–1415.
41. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616.
42. Klauda JB, et al. (2010) Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. J Phys Chem B 114(23):7830–7843.
43. Vanommeslaeghe K, et al. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 31(4):671-690.
44. Beglov D, Roux B (1994) Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. J Chem Phys 100(12):9050-9063.
45. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926-935.
46. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. J Comput Chem 30(10):1545–1614.
47. Smart OS, Neduvelil JG, Wang X, Wallace B, Sansom MS (1996) HOLE: A program for the analysis of the pore dimensions of ion channel structural models. J Mol Graphics 14(6):354–360.
48. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graphics 14(1):33–38.
49. Diallinas G (2014) Understanding transporter specificity and the discrete appearance of channel-like gating domains in transporters. Front Pharmacol 5:207.

9 Conclusions

This work has attempted to further our understanding of membrane protein oligomerization and transport phenomena across membranes. To summarize the findings briefly, the protein Vpu from HIV-1 was shown to exist in a pentameric state, and a structural model for the pentamer was thereby proposed. Furthermore, Vpu was shown to be a weakly conducting ion channel, the reason for the weak channel activity being the hydrophobic nature of the pore. A modeling approach was then proposed for predicting the structures of α-helical membrane proteins. The approach involved minimization of unfavorable contacts in the initial structure followed by conformational sampling and clustering of sampled conformations. As far as investigation of transport across membranes is concerned, three transport proteins were studied, namely p7 from hepatitis C virus, the human aquaporin AQP2, and the bacterial urea transporter dvUT. The selectivity of the p7 channel toward K^+ rather than Ca^{2+} can be attributed to a multitude of mechanisms, including a hydrophobic barrier, knock-on, and differential solvation states in the ion-binding site. While the orientation of permeating water molecules in AQP2 was seen to be governed by an asparagine residue, the orientation of urea molecules in dvUT reported here explains how NH...π and stacking interactions lower the energy barrier for permeation. On the whole, the investigations provide some fundamental insights into structure-function relationships in membrane proteins, as outlined below.

9.1 Perspectives

9.1.1 Rule of maximization of pore hydrophilicity

The minimum unfavorable contacts proposed here starts with an initial model that has the maximum possible hydrophilic surface area of the channel facing the pore region, and minimum contact between hydrophobic and hydrophilic regions. In the case of both M2 and BM2 channels, it is seen that the structures retain this arrangement during the simulations, and even the final predicted structure has such an arrangement, indicating that the arrangement is favorable. Interestingly, the initial structural model for Vpu had an arbitrary structure without any minimization of unfavorable contacts. Notably, Ser23 – the only hydrophilic residue in the

pore region of this protein – was buried in the interhelical interface in the initial model. However, over the course of the simulations, rearrangements in the structure ultimately led to structural models with the serine facing the pore, consistent with experimental data. There is thus an inherent tendency of homooligomeric ion channels to assemble in such a way that a maximum number of hydrophilic residues face the pore at the center. Thus, the hydrophilicity of the pore is maximized, leading to a pore that can accommodate as many water molecules as possible. This minimizes the enthalpic cost required to transport an ion from bulk water to the interior of the channel. We shall call this functional adaptation of channels the "rule of maximization of pore hydrophilicity". It is to be noted that this arrangement is best suited for retaining the integrity of channels. While an arrangement with maximal hydrophilic surface area facing the membrane would lead to unfavorable hydrophilic-hydrophobic contacts, an arrangement with hydrophilic residues buried in the interhelical interface would disrupt van der Waals interactions that hold together the helices.

9.1.2 Transmembrane helices require optimal van der Waals packing

The tetrameric, pentameric, and hexameric forms of Vpu protein from HIV-1 differ in their rotational symmetry, since they have different numbers of protomeric units constituting the oligomer. Consequently, this leads to different residues lying in the interhelical interface in the three oligomeric forms. It turns out that the pentamer is the most stable oligomeric form because it possesses the best van der Waals packing of residues in the interhelical interface, with the interhelical van der Waals interaction energy being stronger than in the tetramer or the hexamer. The requirement for optimal packing in transmembrane helices is further supported by the fact that the final predicted structures for the M2 and BM2 channels have a much more compact structure compared to the initial model.

9.1.3 Certain parts of channels exhibit single ion permeation, certain parts could accommodate multiple ions

In the p7 channel, the hydrophobic stretch is responsible for the highest energy barrier, and allows one ion at a time to pass through. The hydrophilic stretch, on the other hand, must necessarily accommodate two ions, so that there is repulsion between these ions. This repulsion is necessary to prevent an ion from remaining bound to this site, and it ensures a continuous flow

of ions across the pore. Thus, certain parts in a channel could follow a single ion mechanism, while other parts could have a concerted multi-ion permeation process.

9.1.4 Single ion permeation studies do not reveal the complete picture

Single ion permeation studies on p7 did not explain how permeating ions managed to move out from the binding site formed by the ring of asparagines. The explanation was provided by two ion permeation studies, which suggested that an ion in the binding site is propelled out of this site only when another ion approaches this site and starts to repel the bound ion. The energy minimum reported in the two ion permeation study is therefore shallower than that in the single ion permeation study. Thus, a complete picture of the concerted transport process can best be obtained from a multi-ion permeation study.

9.1.5 Roles of hydrophobic barrier and solvation in ion conduction

For both Vpu and p7, the highest energy barrier for the ion transport process corresponded to the hydrophobic region, implying that such regions are the rate-determining regions. The hydrophobic stretch in p7 has an additional functionality – that of selectivity – and ensures that the doubly charged Ca^{2+} ion encounters a higher barrier in passing through this region than the singly charged K^{+} ion. The height of such energy barriers in ion channels is determined by the extent of desolvation of the permeating ion. Solvation also becomes important in the ion-binding site of p7, where K^{+} sheds a water molecule to bind directly to the asparagine, while Ca^{2+} interacts with the asparagine via bridging water molecules.

9.1.6 The orientation and dynamics of the permeating molecule has implications in selectivity

Water molecules in AQP2 are seen to align themselves along the pore axis when they are in the vicinity of the N184 residue. Confinement of water molecules to such an arrangement is likely to serve as a mechanism for excluding the transport of protons along the water chain via a Grotthus mechanism. This explains why AQP2 does not allow protons to pass through in spite of possessing a water chain that is continuous. In dvUT, urea molecules orient to stack with the pore-facing phenylalanine rings as they pass through the pore. The stacking interactions lower the energy barrier for urea permeation through this hydrophobic region. It follows that this

lowering of energy in the hydrophobic region is not feasible for molecules that are not able to form stacking interactions.

9.1.7 Selectivity mechanisms differ from channel to channel

Selectivity toward K^+ rather than Ca^{2+} in p7 mainly involved differential hydrophobic barriers for the two ions. In AQP2, one species that could possibly permeate the pore, apart from water molecules, is the proton, since these pores have a continuous water chain. Protons are, however, excluded via two mechanisms: restriction of the orientation of water molecules in the middle region of the pore, and the presence of the positively charged arginine residue in the narrowest region of the pore. In dvUT, the phenylalanine residues lining the pore make the pore conducive for urea transport while posing a hydrophobic barrier for other solutes. Thus, all channels have their own mechanisms for discriminating between different chemical species.

9.2 Future directions

9.2.1 Extension of the minimum unfavorable contacts approach to multi-pass transmembrane proteins

The results from the structure modeling of Vpu, M2, and BM2 show that transmembrane helices optimize van der Waals packing while assembling. van der Waals contacts can therefore be optimized in order to model helical bundles like G protein-coupled receptors (GPCRs), which have seven transmembrane helices connected by loop regions. Tools meant for predicting transmembrane regions in proteins can initially be used to identify the transmembrane segments (from the amino acid sequence), and then obtain an optimized structure consisting of these segments. The loop regions could then be added to these segments, followed by an equilibration, so as to obtain a structural model for the complete protein. In fact, we have employed such an approach for modeling the structure of the multi-pass protein 3a from SARS coronavirus [1].

9.2.2 Extension of the approach to transmembrane proteins with cytoplasmic domains

For proteins that have a cytoplasmic domain as well as a transmembrane domain, experimental structures are often available for the cytoplasmic part, and the bottleneck in obtaining the structure of the complete protein is solving the structure of the transmembrane part. For such

proteins, the approach proposed here could be used to model the structure of the transmembrane region. The structure of the cytoplasmic region obtained from alternate methods such as homology modeling and experimental studies can then be integrated with this model so as to obtain a structural model for the complete protein.

9.2.3 Structural models can serve as starting points for drug discovery

The ultimate goal of making structural models available for membrane proteins is to make structure-based drug design feasible. Structural models for channels and receptors can be used for high throughput screening approaches like molecular docking, followed by molecular dynamics-based free energy calculations. Thus, the structure modeling approach would make it possible to overcome the biggest hurdle in drug design against membrane proteins: obtaining a structure for the protein that, in turn, could be used for identifying lead compounds.

9.2.4 Modeling gating in channels

Modeling gating and other large scale domain motions in channels could elucidate the steps involved in the opening/closing of these channels. The sensitivity of p7 to voltage could be tested by simulating the channel in the presence of an external electric field. The conformational preferences and transport activity of channels at different pH could be tested by modeling protonable residues in different protonation states. Since p7, APQ2, and dvUT all have protonable residues in their pore, it would be interesting to see the effect of pH on these channels. pH sensitivity can make it possible to alter the conductance of these channels when necessary (for example, in a diseased state) simply by changing the pH of the environment in which these channels exist. It would also be interesting to study the activity of these channels in different lipid environments, and find out if certain kinds of lipids enhance/inhibit the activity. Factors that are identified as being responsible for altering the conductance of these channels can then be tested in experimental conductance studies.

9.2.5 Oligomeric forms of water and urea channels

In order to find out if insights obtained into transport properties for monomeric forms of AQP2 and dvUT hold equally well for the oligomeric forms (AQP2 and dvUT are known to exist as tetramers and trimers, respectively), it is necessary to model these oligomeric forms. Such

systems are immensely huge (of the order of 10^6 atoms), but enhancements in computing power should make it possible to perform microsecond scale simulations on such systems.

References

1. Ramakrishna S, Padhi S, Priyakumar UD (2015) Modeling the structure of SARS 3a transmembrane protein using a minimum unfavorable contact approach. J Chem Sci 127(12):2159–2169.

Publications

Journal Publications

1. Padhi S, Khan N, Jameel S, Priyakumar UD (2013) Molecular dynamics simulations reveal the HIV-1 Vpu transmembrane protein to form stable pentamers. PLoS ONE 8(11):e79779.
2. Padhi S, Burri RR, Jameel S, Priyakumar UD (2014) Atomistic detailed mechanism and weak cation-conducting activity of HIV-1 Vpu revealed by free energy calculations. PLoS ONE 9(11):e112983.
3. Padhi S, Ramakrishna S, Priyakumar UD (2015) Prediction of the structures of helical membrane proteins based on a minimum unfavorable contacts approach. J Comput Chem 36(8):539-552.
4. Padhi S, Priyakumar UD (2015) Ion hydration dynamics in conjunction with a hydrophobic gating mechanism regulates ion permeation in p7 viroporin from Hepatitis C virus. J Phys Chem B 119(20):6204-6210.
5. Ramakrishna S, Padhi S, Priyakumar UD (2015) Modeling the structure of SARS 3a transmembrane protein using a minimum unfavorable contact approach. J Chem Sci 127(12):2159-2169.
6. Padhi S, Priyakumar UD (2016) Cooperation of hydrophobic gating, knock-on effect, and ion binding determines ion selectivity in the p7 channel. J Phys Chem B 120(19):4351-4356.
7. Padhi S, Priyakumar UD (2016) Urea–aromatic stacking and concerted urea transport: Conserved mechanisms in urea transporters revealed by molecular dynamics. J Chem Theory Comput 12(10):5190-5200.
8. Suresh G, Padhi S, Patil I, Priyakumar UD (2016) Urea mimics nucleobases by preserving the helical integrity of B-DNA duplexes via hydrogen bonding and stacking interactions. Biochemistry 55(40):5653-5664.
9. Padhi S, Priyakumar UD (submitted) Microsecond simulations of human aquaporin 2 reveal structural determinants of water orientation, permeability, and selectivity.

10. Bikkina S, Bhati AP, Padhi S, Priyakumar UD (submitted) Temperature dependence of the stability of ion pair interactions, and its implications on the thermostability of proteins from thermophiles.
11. Khan N, Padhi S, Patel P, Priyakumar UD, Jameel S (to be submitted) The HIV-1 Vpu transmembrane domain topology and formation of a hydrophobic interface with BST-2 are critical for Vpu-mediated BST-2 downregulation.
12. Reddy LK, Padhi S, Priyakumar UD (to be submitted) Structural and energetic basis of formate permeation through a bacterial formate transporter.

Oral Presentations

1. Padhi S, Priyakumar UD (2016) Modeling structures and transport phenomena of transmembrane channels and transporters. Annual Research Awards. Organized by Dr. K.V.Rao Scientific Society, Hyderabad, India.
2. Padhi S, Priyakumar UD (2016) Microsecond simulations reveal atomistic details of water and urea transport across membranes. CTCNS Symposium. Organized by IIIT, Hyderabad, India.
3. Padhi S, Priyakumar UD (2016) Structure modeling and ion channel activity of viroporins. ACS on Campus Meeting. Organized by the American Chemical Society at IIIT, Hyderabad, India.
4. Padhi S, Zheng M, Priyakumar UD, Waller M (2015) Urea-assisted denaturation of proteins investigated using a density-based adaptive QM/MM method. MCBR Meeting. Organized by IIT, Delhi, India.
5. Padhi S, Priyakumar UD (2015) Structure modeling and ion channel activity of viroporins (flash presentation). MCBR Meeting. Organized by Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany.

Poster Presentations

1. Padhi S, Singh H, Jameel S, Priyakumar UD (2012) Lipid-Protein Interactions in Membranes: Implications for Health and Disease. Organized by U.S. Biophysical Society at CCMB, Hyderabad, India.

2. Padhi S, Ramakrishna S, Ragini G, Burri RR, Priyakumar UD (2012) Theoretical Chemistry Symposium. Organized by IIT, Guwahati, India. **(Best Poster Award)**
3. Padhi S, Ramakrishna S, Ragini G, Burri RR, Priyakumar UD (2013) Indo-German Meeting on Modeling Chemical and Biological (Re)Activity (MCBR). Organized jointly by NIPER Mohali and IISER Mohali in Mohali, India.
4. Padhi S, Khan N, Jameel S, Priyakumar UD (2014) Chemistry with Computers (CWC). Organized jointly by IIIT Hyderabad and IICT Hyderabad in Hyderabad, India.
5. Padhi S, Ramakrishna S, Burri RR, Priyakumar UD (2014) Recent Advances in Modeling Rare Events (RARE2014). Organized by IIT Kanpur in Vayalar, India.
6. Padhi S, Ramakrishna S, Priyakumar UD (2014) Theoretical Chemistry Symposium. Organized jointly by NCL Pune and IISER Pune at Pune, India. **(Best Poster Award)**
7. Padhi S, Ramakrishna S, Priyakumar UD (2015) Indo-German Meeting on Modeling Chemical and Biological (Re)Activity (MCBR). Organized by Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany.
8. Padhi S, Priyakumar UD (2015) Computing Free Energy Across Disciplines: From Method Development to Applications. Organized by Westfälische Wilhelms-Universität Münster, Münster, Germany.