Molecular Dynamics Study
of the
human voltage-gated proton channel, $h$HV1

by

Kethika Kulleperuma

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Biochemistry
University of Toronto

© Copyright 2016 by Kethika Kulleperuma
A membrane protein, the human voltage-gated proton channel (hHV1) is crucial for many physiological processes of a wide range of cell types. Such processes include bacterial killing by white blood cells, allergic responses by basophils, airway pH regulation in respiratory tracts, and sperm maturation. hHV1 is homologous to the voltage sensor domains (VSD) of voltage-gated cation channels. As such, hHV1 is predicted to adopt a four-helix (S1-S4) bundle structure with three highly-conserved Arg residues (R1-R3) on S4 as well as a number of other charged residues on S1-S3. As a channel, hHV1 displays a remarkable diversity in ion selectivity upon mutating residue Asp\(^{112}\) on S1, the putative selectivity filter of the pore. hHV1 carrying single or double point mutations of Asp\(^{112}\) along with Val\(^{116}\) makes the channel either proton-selective, anion-selective, or nonconducting.

The molecular basis for ion and charge selectivity in hHV1 remains unexplained. The lack of an experimentally-resolved native structure for hHV1 has hindered not only the uncovering of the molecular basis for ion permeation and charge selectivity in hHV1, but also the advancement of its potential applications as a drug target for the treatment of cerebral damage from ischemic stroke as well as of colorectal and breast cancers.

Here, I investigate the molecular basis for ion and charge selectivity of hHV1, using molecular dynamics simulations. I first present ROMP, a general, computationally-efficient simulation approach to predict the solvation and the orientation of peptides and proteins in lipid bilayers. The method, which employs a biphasic membrane mimetic to accelerate the convergence of solvent properties, is shown to reproduce the solvation and structural fluctuations of various types of membrane proteins and peptides in lipid bilayers.

I then describe the construction and the validation of two homology models (HM) of hHV1 based on three VSD templates and differing in the registry of S4 with respect to S1-S3 helices. By incorporating ROMP, the relative quality of these two models is evaluated using massively-repeated simulations in the biphasic membrane mimetic. The structural properties of the better model are consistent with experimental accessibility data and \textit{a priori} features of a proton-conducting channel, with a putative
proton permeation pathway comprised of an intermittent hydrogen-bonded water chain in a narrow non-polar bottleneck connecting two water-filled crevices lined with charged residues.

Next, I carry out a comparative assessment of the structure and dynamics of wild-type and mutant channels in a lipid bilayer. This analysis reveals a reorganization of the salt-bridge network involving primarily R2 and mutations at positions 112 and/or 116. The presence or absence of charge compensation of R2 by mutations modulates the electrostatic properties of the pore consistently with the proton or anion selectivity of the mutants, respectively.

Finally, I use free energy simulations to investigate the thermodynamic basis for the movement of Na$^{+}$ and Cl$^{-}$ in the wild-type and two anion-selective mutant channels. Results indicate that the movement of non-proton ions in the wildtype is opposed at the constriction site of the pore, where both types of ions face a substantial desolvation penalty. Although the barrier for Cl$^{-}$ is significant even in the anion-selective mutants, limited structural rearrangements in the constriction site and, as previously shown by our electrostatic calculations, the absence of a negatively charged residue at position 112, lower the barrier for Cl$^{-}$ in mutants compared to the wildtype channel. Taken together, these findings provide clues to refine our structural model as well as insight that can be used to further the elucidation of ion and charge selectivity and the molecular mechanism of proton translocation in hHv1.
Dedication

To

_Sumedha, Tikiri, Chooty & Teesha_

for their boundless love and care.
Acknowledgements

First and foremost, I would like to thank Dr. Régis Pomès, my PhD supervisor. I am profoundly fortunate to have been mentored under you all these years – I have never felt more like a “budding scientist” in my academic path. Thank you for always pulling me back to the correct path whenever I was lost, confused and going astray.

I would like to extend my heartfelt thank you to the PhD students who preceded me; Drs. Chris Madill, Chris Neale, Sarah Rauscher, Marty Kurylowicz, Grace Li, John Holyoake and Loan Huynh. Thank you for being tough acts to follow. You have inspired me in many ways and thank you for being the role models I looked at during these years. I especially thank Drs. Neale and Holyoake for patiently guiding me step up in every possible way imaginable – from writing a simple bash script to thinking critically about what simulations mean. I was very lucky to have my workstation next to Dr. Sarah Rauscher who set a new definition for what discipline means for the rest of my life. For that, thank you very much, Dr. Rauscher.

Thank you David and Zuyi for making me realize that there is more to programming than lines and lines of abstract words. Thank you Dr. Li for always inspiring me, even to this day, on how to apply the learning experiences accumulated during the PhD to a world outside of lab. Dr. Chakrabarti, thank you for teaching me the value of going back to the drawing board and then cracking the problem whenever in doubt. Many thanks to Chris Ing, de facto partner-in-crime during the last years of my PhD. Your enthusiasm for science and literally anything under the sky kept research in lab all fanciful and adorned.

Ellen Li, Aditi Ramesh, Paul Huynh, and Jingjing Huang; thank you for being the best colleagues. Being with you made me a better person.

I would like to extend my sincerest thanks and appreciation to my committee members, Drs. Christine Bear and Emil Pai for insightful discussions and advice given to me throughout these years.

Drs. Susan Smith and Tom DeCoursey for collaborating in my project. To say your help was crucial to this project would be an understatement. The numerous discussions we had over the years not only helped me progress in the project, but also grow as an academic collaborator.

Thank you Compute Canada for the computer resources and technical expertise. I would also like to thank the department of biochemistry, University of Toronto and SickKids. It was a privilege to have been nurtured in a world’s leading university and research institute that provided me boundless opportunities to better myself.

I came to Toronto in 2009. Up until that point, all my teachers, professors and mentors of Musaeus College, University of Colombo (UOC) and Adelphi University paved a path full of knowledge, encour-
agement and inspiration for me. Very specifically, I owe my immense gratitude to Mrs. Subasinghe of Musaeus College, Dr. Lawrence Hobbie of Adelphi University and Mr. Tasil Samarasinha for believing in me to pursue graduate studies and opening the door to a world of scholarship. Radhika— thank you for pushing me to step beyond UOC.

Sohan, Dhinali, Lochana, Shihan, Buddhi, Sachcha, Niro, Priyani, Yoshani, Oshadhi, Rebekha, Mr and Mrs. De Silva, Mrs. Dee Redhka, Mrs. Shriyani Perera, Mr and Mrs. Upali Perera— thank you for being the family far from home.

Last and not least, I would like to extend my heartfelt gratitude to my family; amma, thatthi, akki, and baba, who were there through thick and thin. Thank you for being the shoulder that I could rest my head on when times were tough and rough. Thank you for your unconditional love.
# Contents

1 Introduction.................................................. 1
   1.1 Ions in biological systems ................................................. 2
      1.1.1 The transport of ions across cell membranes. .................. 2
      1.1.2 Properties of ion channels ........................................... 4
      1.1.3 The study of ion channels: the progression from electrophysiology to MD simulations ............................................. 6
   1.2 The voltage-gated proton channel, H\textsubscript{V}1 .............................................. 7
      1.2.1 Characterization of H\textsubscript{V}1 by electrophysiology .............. 8
      1.2.2 Physiological roles of hH\textsubscript{V}1 ................................... 9
      1.2.3 Human H\textsubscript{V}1 in pathological conditions .................... 10
      1.2.4 Sequence and structural overview of hH\textsubscript{V}1 .................. 11
   1.3 Proton transport .............................................. 14
      1.3.1 Proton hydration .................................................. 15
      1.3.2 Proton mobility .................................................. 15
      1.3.3 Proton transfer in biological systems ................................ 19
   1.4 Thesis objectives and organization .................................. 25
      1.4.1 Thesis objectives .................................................. 25
      1.4.2 Thesis organization .................................................. 25

2 Theory and methods............................................. 27
   2.1 Introduction .................................................. 28
   2.2 Quantum and Classical Mechanics ........................................... 28
      2.2.1 Quantum mechanics .................................................. 29
      2.2.2 Molecular mechanics .................................................. 31
   2.3 Molecular dynamics simulations ........................................... 32
      2.3.1 From potential energy to atomic motions: equations of motion .......... 32
      2.3.2 Temperature and pressure control .................................... 33
      2.3.3 Periodic boundary conditions ....................................... 34
   2.4 Statistical mechanics and free energy simulations ....................... 34
      2.4.1 Ensembles .................................................. 35
      2.4.2 Boltzmann Distribution Law ....................................... 35
      2.4.3 Entropy .................................................. 36
      2.4.4 Free energy .................................................. 36
3 ROMP: Rapid Orientation of Membrane Proteins

3.1 Introduction .......................................................... 49
3.2 Methods ................................................................. 52
  3.2.1 Test proteins and peptides ...................................... 52
  3.2.2 Definition of solute orientation in the membrane .......... 54
  3.2.3 Simulations of KvAP and PagP in octane ................. 54
  3.2.4 Simulations of hIAPP in octane ............................... 55
  3.2.5 Simulations of KvAP in lipids ................................. 55
  3.2.6 PagP simulations in lipids .................................. 55
3.3 Results ................................................................. 57
  3.3.1 Orientation of KvAP ........................................... 57
  3.3.2 Orientation of PagP systems .................................. 61
  3.3.3 PagP helix insertion and orientation ....................... 69
  3.3.4 IAPP helix dimerization .................................... 71
3.4 Discussion ............................................................ 75

4 Construction and validation of a homology model of the human Hv1

4.1 Introduction .......................................................... 79
4.2 Methods ................................................................. 82
  4.2.1 Sampling of taxa and phylogenetic methods .............. 82
  4.2.2 Gene expression ................................................ 82
  4.2.3 Electrophysiology .............................................. 83
  4.2.4 Homology model ............................................... 83
  4.2.5 MD simulations ............................................... 84
4.3 Results ................................................................. 86
  4.3.1 Multiple sequence alignment and phylogenetic analysis .. 86
  4.3.2 Homology modeling ............................................ 88
  4.3.3 Evaluation of alternative models ........................... 89
  4.3.4 Analysis of structural divergence ......................... 90
  4.3.5 Structural properties ......................................... 94
  4.3.6 Hydration profile ............................................. 96
  4.3.7 Energetics of ion translocation ............................. 100
  4.3.8 Accessibility studies ......................................... 101
  4.3.9 Strategy for Zn$^{2+}$ studies ................................. 102
  4.3.10 Behavior of the mutants in the absence of Zn$^{2+}$ .... 102
  4.3.11 External accessibility ....................................... 103
  4.3.12 Internal accessibility ....................................... 103
4.4 Discussion ............................................................ 105
4.4.1 Phylogeny ................................................. 105
4.4.2 Computational validation of the homology models ................................. 107
4.4.3 Hydration and mechanism of proton conduction ................................. 109
4.4.4 How far does S4 move during gating ? ........................................... 111
4.4.5 Comparison with Starace and Bezanilla’s His scanning studies. .......... 112
4.5 Conclusion .............................................................................. 113

5 Peregrination of the selectivity filter delineates the pore of the human hV1 115
5.1 Introduction ................................................................. 116
5.2 Methods ........................................................................... 117
5.2.1 Gene expression ............................................................ 117
5.2.2 Electrophysiology ......................................................... 118
5.2.3 MD simulations ............................................................. 118
5.3 Results ............................................................................. 119
5.3.1 The second site mutation V116D restores proton selectivity to nonconducting or anion-permeable mutants ................................................................. 120
5.3.2 Asp supports current only when facing the pore .................................. 122
5.3.3 At position 109, Asp plays a permissive role .................................... 124
5.3.4 The engineered Asp116 proton channel functionally resembles WT .......... 125
5.3.5 Mutations at position 116 (in D112V channels) mimic mutations at 112 .... 126
5.3.6 Zn2+ sensitivity of Arg→His mutants shows S4 position in open channels .... 128
5.3.7 Introducing Asp into S2 or S3 did not support proton conduction .......... 129
5.3.8 MD simulations reveal significant differences in the electrostatic properties of mutant channels ................................................................. 129
5.4 Discussion ......................................................................... 136
5.4.1 Nonconducting mutants ...................................................... 136
5.4.2 The rules of the game ....................................................... 136
5.4.3 Microenvironment of positions 112 and 116 ..................................... 137
5.4.4 Why does 109 not work ? .................................................. 138
5.4.5 Limited plasticity of the channel with respect to axial helix movement does not support a change of S1 registry in the mutants .................. 138
5.4.6 Parallels in other molecules ................................................. 139
5.5 Conclusion .......................................................................... 140

6 Free Energy Simulations of Ion Translocation through hHv1 142
6.1 Introduction ......................................................................... 143
6.2 Methods ............................................................................. 144
6.2.1 System setup and umbrella sampling ......................................... 144
6.2.2 Simulation protocol .......................................................... 145
6.3 Results .............................................................................. 145
6.3.1 Thermodynamic basis for the movement of Na+ ............................ 145
6.3.2 Thermodynamic basis for the movement of Cl− ............................ 155
6.4 Discussion .......................................................................... 164
7 Conclusions and Future Directions

7.1 Conclusions ................................................................. 168
7.2 Future Directions .......................................................... 170
  7.2.1 Refining the homology model ........................................ 170
  7.2.2 Testing drug binding to H\textsubscript{V}1 .............................. 170
  7.2.3 Conducting quantum simulations .................................... 170
Bibliography ................................................................. 172
## List of Tables

3.1 Summary of MD simulations ..................................................... 53
3.2 Summary of the consensus orientations of the PagP systems. The notation $(\tau,\rho)$
   $= (X \pm x^o, Y \pm y^o)$ represents the consensus orientation, with $X,Y$ denoting the center of
   the basin and the standard deviation computed from a gaussian fit of the largest peak in
   each 1D distribution. ............................................................... 63

5.1 Impermeability of D112V/V116D and D112V/V116D/R211H channels to ions 122
5.2 Effects on selectivity of amino acids at positions 109, 112 and 116 in the S1
   transmembrane segment of $h$Hv1 ............................................. 137
List of Figures

1.1 **Kv and Hv1 channel structures.** (A) Schematic representation of the S1-S6 helix topology of the monomeric Kv channels, modified from Knöpfel 2009. (B) Top-down view of the crystal structure of tetrameric Kv1.2/2.1 channel (PDB:2R9R) reprinted from Fowler & Sansom 2013. The helices of one monomer are coloured and labelled. (C) The crystal structure of KvAP (PDB:1ORS). Arg residues on S4 helix and acidic residues on S1-S3 residues are in licorice representation. (D) Schematic representation of the S1-S4 helix topology of the monomeric Hv1 channels, modified from Knöpfel 2009. Images are reproduced/modified with permission from authors. 13

1.2 **Schematic drawing showing (A) the Eigen cation, a hydronium ion with its hydration, and (B) the Zundel cation, modified from Codorniu-Hernández & Kusalik 2013.** Images are modified with permission from authors. 15

1.3 **Schematic depiction of the Grotthuss mechanism in bulk water modified from Agmon 1999.** (A) An Eigen cation centering a H3O⁺ on H2Oₐ is coordinated to three water molecules, one of which is labelled H2Oₙ. (B) A hydrogen-bond cleavage between H₂Oₙ and H₂Oₖ is followed by a Zundel cation formation. (C) A new hydrogen-bond is formed between H₂Oₙ and H₂Oₖ localizing the excess proton on Oₙ. Hydrogen atoms (black), protonated oxygen atoms (white) and unprotonated oxygen atoms (red) are shown. Black dash and thick lines represent hydrogen-bonds and bonds within the Zundel cation, respectively. The image is modified with permission from authors. 17

1.4 **Schematic depiction of the Grotthuss mechanism reprinted from Pomés & Roux 2002.** An excess proton is incorporated at one end of the water chain followed by a succession of protonation-dissociation reactions between adjacent water molecules or “proton hop”, a proton is released at the other end of the chain and water molecules are inverted (“turn”) such that the dipoles of water molecules are reoriented to incorporate the next proton. The image is reproduced with permission from authors. 18

1.5 **Schematic depiction of the β-helix structure of the gA dimer reprinted from Pomés & Roux.** The narrow cylindrical pore is lined with peptide bonds and accommodates a single-file water chain of water molecules, depicted here as red spheres. The image is reproduced with permission from authors. 20
1.6 Schematic depiction of AQP reprinted from Chakrabarti et al. 2004. The monomeric GlpF channel is shown in ribbon representation (a) together with a close-up of the pore region (b). Three conserved residues, Arg206 in the selectivity filter, as well as Asn68 and Asn203 from the signature Asn-Pro-Ala (NPA) motifs, are highlighted along with the two half-membrane-spanning helices, M3 and M7. The nine single-file water molecules embedded in the pore are shown in their preferred bipolar organization. Water molecules lying on the periplasmic and cytoplasmic side of the pore are shown at the top and at the bottom, respectively. The image is reproduced with permission from authors.

2.1 Schematic representation of US windows along $\xi$. In the top panel, the global free energy (thick solid black curve) and the biased potential energy functions of seven distinct harmonic restraining potentials are overlaid on the global free energy line, where the first three potentials are shown in red, blue and green thin lines and the rest in black thin lines. In the bottom panel, the normalized histogram of biased probability distributions corresponding to each harmonic restraining simulation are shown.

2.2 Biased probability distributions of representative windows along $\xi$. Only three normalized histograms of biased probability distributions of overlapping windows are shown in red, green and blue. The histogram in black is the biased probability distribution of the last window. For example, the biased probability distribution of the $i$ window (red) is $\rho_i(\xi)^b$. Subscript $j$ means $j^{th}$ bin in $i^{th}$ simulation. The biased probability of finding the system at $\xi = \xi_j$ in window $i$, $\rho_i(\xi_j)^b$ is coloured in orange.

2.3 Schematic representation of “stitching” of PMFs of 3 umbrella windows to construct the global $W(\xi)$. The three colours represent those of the normalized histograms in Fig. 2.2. The two black arrows represent adjusting of $F_i$ values.

2.4 Homology modelling flowchart. Adopted from Venclovas, 2012. The image is modified with permission from authors.

2.5 The two zones of sequence alignments. Adopted from Krieger et al. 2003. Two sequences are likely to fold into the same structure if their length and percentage sequence identity fall into the region marked as “safe.” The image is reproduced with permission from authors.

2.6 Alignment between $\alpha$Hv1 and its 3 templates. The notations; 1ORSk, NaVBkA and 2R9R denote the sequences of VSD templates. The colours of the sequence alignment emulates the default colourscheme used for alignments in Clustal X, a graphical interface for the ClustalW multiple sequence alignment program. Three standard annotations; conservation, quality and consensus are depicted below the sequence alignment. Conservation is a numerical index reflecting the conservation of physico-chemical properties in the alignment. The color shading from most intense (bright yellow) to palest (dark brown) reflects the most conserved to least conserved residues. The alignment quality is measured based on the substitution matrix of BLOSUM62 and is plotted on a scale from 0 (dark brown) to 1 (bright yellow). A high alignment quality score for a column would suggest that there are no mutations. Consensus gives the commonest residues and their percentage for each column as reflected by the heights of the black bars.
3.1 The model systems. A) KVAP; B) PagP-Flexible Linker (PagP-FL): the N-terminal α-helix is flexible with respect to the β-barrel. The 2 unrestrained residues are highlighted in orange licorice. C) PagP-Fully Rigid (PagP-FR): pairwise distances between all the Cα atoms are fixed. D) PagP-Barrel Only (PagP-BO) E) PagP-Helix Only (PagP-HO). The backbone atoms are represented in ice blue ribbons. The Cα atoms used to restrain the systems are highlighted in grey van der Waals and are connected with black lines indicating the distance restraints. ................................................................. 53

3.2 Orientation of KVAP. A) 1D histogram of ρ angle distribution B) 1D histogram of τ angle distribution C) 2D histogram between ρ and τ angle distributions. D) Representative snapshots of the corresponding basins defined as a-d in plot C. Protein backbone (blue ribbons), octane (grey lines) and water (red spheres) are highlighted. .................. 58

3.3 Convergence of orientation of the KVAP and PagP-FL. A) Percentage of replicas that reached the consensus structure vs. time for PagP-FL (top) and KVAP (bottom) for n number of samples, where n = 48 (grey), 240 (red), and 480 (green). B) Fraction of replicas that reached the consensus structure (black) vs. starting orientation for PagP-FL (top) and KVAP (bottom). The data is normalized to 360°. The unconverged fraction is depicted in blue. ................................. 59

3.4 Change in secondary structure of unrestrained vs. rigid KVAP in lipid (top panel) vs. octane (bottom panel), respectively. Data is shown from 390-400 ns of the lipid simulation and 0-10 ns of one of the representative replica of the octane simulations. The TM region used for the definition of orientation is from residue numbers: 30-43, 58-73, 90-101 and 123-138. DSSP secondary structure assignment is shown for coil (white), bend (green), turn (yellow), α-helix (blue) and 3_{10} helix (grey). ..................... 60

3.5 Orientation of PagP systems. A) The 2D histogram of ρ and τ angles of the full data set of PagP-FL. B) Representative snapshots of the corresponding basins labelled as a-e in panel A. The protein backbone (blue ribbons), octane (grey lines) and water (red spheres) are highlighted. C) 2D PMF of PagP orientation, W(τ,ρ): only the surface below the isoenergetic contour W=1 kcal/mol is shown, with the most likely orientation at W=0. ............................................................ 62

3.6 Solvation of PagP-POPC vs PagP-FL. A) Average number of hydrogen bonds with water molecules as a function of residue number of PagP. Data were generated from the last 2 ns of the replicas that reached the consensus orientation. B) The difference map of the average number of hydrogen bonds with water, indicated by avg.# in PagP-FL - avg.# in PagP-POPC (grey bars) for each residue. The overlaying purple bars indicate the average number of hydrogen bonds with lipid headgroups in PagP-POPC simulations. C) Average number of hydrogen bonds with water molecules and POPC headgroups as a function of time in PagP-POPC and PagP-FL simulations. .............................. 65
3.7 Representative snapshot of PagP-POPC highlighting the hydrogen bonding difference between PagP-POPC and PagP-FL. Water molecules within the cut-off radius of 3.2 Å of polar atoms of the PagP-POPC are depicted in blue licorice. Residues with hydrogen bonding difference greater than 0.5 and less than -0.5 in Fig. 3.6B are highlighted in red (more hydrogen bonding with PagP-FL) and yellow (more hydrogen bonding with PagP-POPC) ribbons. The rest of the protein backbone is represented as a transparent cyan ribbon.

3.8 Sidechain dihedral angle distribution of aromatic residues located at the outer and inner membrane-water interfaces of PagP-POPC (black), PagP-FL (red), PagP-FR (blue) and PagP-BO (green). Both $\chi_1$ and $\chi_2$ angle distribution are depicted on left and right sides respectively. A) Residues located at the inner interface. B) Residues located at the outer interface. C) Representative structure of PagP-FL (red) is superimposed on a representative structure of PagP-FR (green) with aromatic residues highlighted at the inner (left) and outer (right) interfaces.

3.9 Orientation of the N-terminal $\alpha$-helix of PagP. A) 2D PMF of orientation of PagP-HO (pink) and PagP-POPC (red). B) Representative snapshots of basins indicated from a-b as defined in panel A. The protein backbone (blue ribbon), octane (grey lines), and water molecules (red spheres) are highlighted.

3.10 Fraction of replicas of PagP-HO reached one of the 2 orientational basins vs. their starting orientations. The basins converged to the orientation a and b in Figure 3.9 are shown in green and red, respectively. Fraction of replicas reaching orientations other than that of a and b is in blue.

3.11 Time evolution of the population of PagP-HO orientations, a and b in green and red lines, respectively.

3.12 Dimerization of hIAPP helices. 2D pmf of hIAPP dimerization as defined by the distance between the N-terminal ends of the two helices (r) vs. inter helical angle ($\theta$) in Nath et al. The plot only shows isoenergetic contours at $W=0.5$ and 1 kcal/mol. Four conformational basins labelled 1a, 1b, 2 and 3 match the orientation details provided in the text. Representative snapshots of (blue, yellow) peptide backbone and (grey) octane molecules in each basin are shown as inserts.

3.13 Population of pairwise contacts in the IAPP dimers. Increasing number of contacts are depicted in the color scale starting from white to blue for each conformational basin 1a, 1b, 2 and 3 as defined in Figure 3.12.

4.1 A subset of a larger multiple sequence alignment of the four TM regions of VSDs informed by alignment of the open state produced by structural superposition of crystal structures of the paddle chimera (green ribbon) and $Na_V$AB (white ribbon). The arginines of S4 (blue sticks) are labeled according to their position on the paddle chimera. Note that the most extracellular R of $Na_V$AB corresponds to R2 of the paddle chimera, and the most intracellular R of $Na_V$AB corresponds to K5 of the paddle chimera. The completely conserved phenylalanine on S2 is shown and labeled as “CTC” The following residues are indicated with symbols over the alignment: %, CTC; *, R1 of Shaker; #, R2 of Shaker; ^, R3 of Shaker.
4.2 **Phylogenetic tree constructed from the multiple sequence alignment of VSDs exemplified in Fig. 4.1.** Branches and names of subfamilies of VSDs are color coded. 
K<sub>V</sub>-E, eukaryotic potassium channel; K<sub>V</sub>-B, prokaryotic potassium channel; K<sub>V</sub>-H, H family potassium channel; HCN, eukaryotic hyperpolarization-activated, cyclic nucleotide-gated; CNG, cyclic nucleotide-gated channel; Trp, transient receptor potential channel; 
Na<sub>V</sub>-E, eukaryotic sodium channel; Na<sub>V</sub>-B, prokaryotic sodium channel; Ca<sub>V</sub>-E, eukaryotic calcium channel; H<sub>V</sub>1, voltage-gated proton channel; VSP, voltage-sensitive phosphatase; C15orf27, homologues of human C15orf27, a protein of unknown function.  

88

4.3 **Energy-minimized starting structures of the five models of R2D (A) and R3D (B) are superimposed.** D112 in S1 and the three Arg residues in S4 are shown in red and blue, respectively. The channels are viewed from the side (membrane), and the external end is at the top.  

89

4.4 **A plot of the root mean square deviation (RMSD) versus time for K<sub>V</sub>AP in a phospholipid bilayer and in octane.**  

89

4.5 **Fraction of structures in the most populated cluster, shown as a percentage of the total population at each clustering cutoff value, R<sub>c</sub>, for each of the three VSD templates and the two homology models.**  

90

4.6 **Analysis of structural relaxation from massively repeated sampling.** Relative probability distributions of (left) RMSD values (see Results) and (right) change in the number of α-helical residues, ΔN<sub>α</sub>, in the most populated cluster at each cutoff value (in nanometers, top right), relative to the starting structure of the simulations. The data are normalized to the total population.  

93

4.7 **Representative conformations of the two homology models.**  

94

4.8 **An analysis of salt-bridge populations of the two homology models (R2D: black, R3D: grey) in the most populated cluster at each cutoff value (in nanometers, right on each panel).**  

95

4.9 **Salt-bridge formation in the (A) R2D and (B) R3D model.** Acidic and basic residues in the α-helical TM region (red, S1; yellow, S2; green, S3; blue, S4) are shown in licorice representation. The salt bridges present in >10% of the most populated cluster obtained with R<sub>c</sub> = 0.1 nm and are highlighted in thick dotted lines.  

95

4.10 **Structural fluctuations of the two channel constrictions in model R2D.** (A) Configurations of the D112-R208 salt bridge (from left to right: bidentate, monodentate, and open). Hydrogen bonds are shown as orange lines. (B) Probability distribution of d<sub>1</sub>, the distance between atom C<sub>γ</sub> of D112 and atom Cz of R208. (C) Closed (left) and open (right) conformations of F150 and R211. (D) Probability distribution of d<sub>2</sub>, the distance between the center of the benzyl ring of F150 and atom Cz of R211. (E) Two-dimensional probability distribution of d<sub>1</sub> and d<sub>2</sub>, with increasing probability from purple to yellow.  

98
4.11 Pore hydration in the R2D model. Representative snapshots are shown successively for (A) the most constricted conformational state defined as \( d_1 < 0.42 \) nm and \( d_2 < 0.55 \) nm and (B) least constricted conformational state defined as \( d_1 > 0.52 \) nm and \( d_2 > 0.55 \) nm (see Fig. 4.10 E). Residues D112, F150, R208, and R211 are shown in licorice representation. Water molecules within 5 Å of these residues are colored in red and white, and hydrogen bonds between these molecules are shown with orange lines. (C) Water density within a 0.7 nm radius of the mean axis of the pore, normalized to the bulk water density in the most constricted and (D) least constricted conformational states, respectively. (E) Comparison of the water density profiles from four control simulations of 100 ns, successively in n-octane (blue) and in a POPC bilayer (magenta). Error bars represent the standard deviation.

4.12 Effect of the charge distribution of the channel on the energetics of ion translocation. (A) Representative snapshot of the water-filled open conformation of H\(_\text{V1}\) . Water molecules and charged residues in the pore are shown in licorice representation; the extracellular end is to the right. (B) Static field energy for the transfer of a positive point charge in (blue) WT and (red) neutral-D112 forms of H\(_\text{V1}\) . Two thick lines represent the mean of five running averages, each computed from a dataset of 30 snapshots using a five-point moving window. Blue shading and orange error bars represent the SEM.

4.13 Evaluation of external accessibility of the three Arg residues in the S4 helix of h\(\text{H\(_\text{V1}\)}\). All measurements were made in a construct designed to have low sensitivity to Zn\(^{2+}\), H140A/H193A/K221stop. (A) Currents during identical families of pulses in 10 mV increments up to +60 mV in the absence construct in whole-cell configuration, in the presence of 0, 10, or 100 \( \mu \)M Zn\(^{2+}\) in the bath solution. (B) Whole-cell currents in the R211H mutant during identical families of pulses in 10 mV increments up to +80 mV, in the presence of 0, 10, or 100 \( \mu \)M Zn\(^{2+}\). (C) Currents during identical pulse families in 20 mV increments up to +60 mV in the R208H mutant in the absence (black) or presence of 10 \( \mu \)M Zn\(^{2+}\) (red). (D) Currents during identical pulse families in 10 mV increments up to +30 mV in the R205H mutant in the absence (black) or presence of 10 \( \mu \)M Zn\(^{2+}\) (red). (E) Currents during three consecutive pulses to +90 mV in a cell transfected with R205H. At or slightly before the arrow, 10 \( \mu \)M Zn\(^{2+}\) was applied to the bath, reducing the current during the pulse (green). The next pulse (red) shows the steady-state extent of Zn\(^{2+}\) inhibition. (F) Later in the same experiment, EGTA was added during a pulse at or before the arrow, rapidly relieving Zn\(^{2+}\) effects. For all parts, pH\(_o\) was 7.0; pH\(_i\) was 7.0 for A, B, E, and F; and pH\(_i\) was 5.5 for C and D.

4.14 Evaluation of internal accessibility of Arg\(^{211}\) in the S4 helix of h\(\text{H\(_\text{V1}\)}\). All measurements are in inside-out patches at pH\(_o\) 7.0, pH\(_i\) 6.0. (A) Control families in the absence and presence of 0, 10, or 100 \( \mu \)M Zn\(^{2+}\) with pulses in 10 mV increments up to +60 mV. (B) Currents in an R211H patch during identical families of pulses in 10 mV increments up to +60 mV in the presence of 0 or 10 \( \mu \)M Zn\(^{2+}\). (C) Three consecutive pulses to +60 mV in an inside-out patch with the R211H mutant, first in the presence of 10 \( \mu \)M Zn\(^{2+}\) (red), and then with the addition of EGTA during the pulse (green), and finally in the absence of Zn\(^{2+}\) (black). The reversal of Zn\(^{2+}\) effects shows accessibility of R211H in the open state. For all parts, \( V_{\text{hold}} = -60 \) mV.
5.1 **The D112V mutation abolishes current, but V116D restores proton-specific current, an example of second-site suppression.** Whole-cell currents at pH$_o$ 7.0 and pH$_i$ 5.5 during pulses in 10 mV increments up to the indicated voltages for WT (A), D112V (B), or D112V/V116D (C), all expressed in COS-7 cells. Holding potential ($V_{hold}$) and pulse durations were -90 mV, 1 s (A), -40 mV, 3 s (B), and -60 mV, 2 s (C). In all cases, the voltage was returned to $V_{hold}$ after the pulse, which is why the tail currents are inward for A and outward for C. Cartoons in all figures indicate S1 and S4 helices, with color coding as follows: red, Asp or Glu; yellow, Val; blue, Arg; gray, other amino acids or nonpore-facing residues.

5.2 **Second site suppression in double mutants.** (A) Shifting the crucial aspartate from position 112 to 116 (V116D) restores proton selectivity to both the nonconducting D112V (red symbols) and the anion-permeable D112A (blue symbols) single mutants. Measurements in the same cell are connected by lines. For both, $V_{rev}$ measured over a wide range (for D112V/V116D pH$_o$ was 5.5, 6.0, 6.5, 7.0, 7.5, or 8.0, and pH$_i$ was 5.5, 6.0, 6.5, or 7.0; for D112A/V116D pH$_o$ was 4.5, 5.5, 6.0, 6.5, or 7.0, and pH$_i$ was 5.5, 6.0, or 6.5) falls close to the Nernst potential for H$^+$, $E_H$ (dashed green line). (B) Determination of $V_{rev}$ from tail currents for D112V/V116D in a COS-7 cell with pH$_i$ 5.5. Prepulses to +30 mV, pH$_o$ 5.5, or to -40 mV, pH$_o$ 7.0, activated the conductance, followed by repolarization to the indicated voltages in 10 mV increments, with the most positive labeled. (C) Determination of $V_{rev}$ from current families for D112A/V116D in a COS-7 cell with pH$_i$ 5.5. As indicated, $V_{hold}$ was -60 mV (left) or -70 mV (right); the voltage was returned to $V_{hold}$ after the pulses. Currents are shown during selected pulses that bracket $V_{rev}$ in 10 mV increments. Inward current was activated negative to $V_{rev}$, and the first outward current is labeled.

5.3 **Introducing Asp at position 113, predicted by our model to be in a nonpore-facing location, results in membrane expression of a functioning proton-selective channel.** The cartoon emphasizes that Asp$^{112}$ is still present. Families of currents are shown in a COS-7 cell in whole-cell configuration at pH$_o$ 5.5 (A) or pH$_i$ 5.5, with pulses applied from $V_{hold}$ as labeled to the indicated voltages in 10 mV increments. Cells were returned to $V_{hold}$ after pulses. (C) Proton selectivity is shown by the proximity of $V_{rev}$ to $E_H$ (dashed line). Insets show $V_{rev}$ determination (left) at pH$_o$ 5.5 and pH$_i$ 5.5 by reversal of current during a family of pulses in 10 mV increments ($V_{hold}$ = -60 mV) and at pH$_o$ 7.0 and pH$_i$ 5.5 (right) by tail currents. $V_{rev}$ was measured at pH$_o$ 4.5, 5.0, 5.5, 6.5, 7.0, and 7.5, and at pH$_i$ 5.5 or 6.5.

5.4 **Moving aspartate from position 112 to 109 results in anion currents.** Whole-cell currents in a COS-7 cell expressing D112A/V109D channels, all at pH$_o$ 5.5 and pH$_i$ 5.5, in symmetrical TMA$^+$ CH$_3$SO$_3^-$ (A and B) or with Cl$^-$ in the bath (C and D). Pulses applied in 10 mV increments. $V_{hold}$ was -40 mV (A and B) or -60 mV (C and D). Cells were returned to $V_{hold}$ after pulses. $V_{rev}$ determination from tail currents (B and D), with pulses in 10 mV increments.
5.5 Voltage and pH dependence of gating of hHV1 mutants with the selectivity filter shifter from 112 to 116. Families of currents in a COS-7 cell expressing D112A/V116D channels at pH\textsubscript{i} 6.5 and pH\textsubscript{o} 7.5 (A), 6.5 (B), or 5.5 (C) are shown in 10 mV increments as labeled, from $V_{\text{hold}}$ -60 mV (A) or -30 mV (B and C). Cells were returned to $V_{\text{hold}}$ after pulses. (D) Current-voltage relationships in this cell. Currents were fitted to a rising exponential function and extrapolated to infinite time. Note inward currents at pH\textsubscript{o} 5.5. (E) Conductance-voltage relationships for the same currents. Limiting slope conductance for the most negative voltages provides a gating charge estimate of 5.3-6.0 $e_0$. (F) Regulation of voltage gating by \( \Delta \text{pH} \). $V_{\text{threshold}}$ is plotted against $V_{\text{rev}}$ measured in the same cell and solution. Lines show linear regression fits defined by $V_{\text{threshold}} = 0.64 \ V_{\text{rev}} + 0$ mV (D112V/V116D; red circles) or $V_{\text{threshold}} = 0.63 \ V_{\text{rev}}$ -26 mV (D112A/V116D; blue diamonds).

5.6 The D112V/V116E mutant is a functional proton-selective channel. (A) Families of currents generated by D112V/V116E in a HEK-293 cell in 10 mV increments at pH\textsubscript{p} 5.5 and pH\textsubscript{i} 5.5 (A), and pH\textsubscript{p} 5.0 and pH\textsubscript{i} 5.5 (B). As indicated, $V_{\text{hold}}$ was -60 mV, and the membrane was returned to $V_{\text{hold}}$ after pulses. (C) $V_{\text{rev}}$ measured at pH\textsubscript{p} 5.0, 5.5, 6.5, 7.0, and 7.5, and at pH\textsubscript{i} 5.5 or 6.5, indicates that D112V/V116E is proton selective. Inset shows tail currents at 10 (uppermost), 5, and 0 mV at pH\textsubscript{p} 5.5 and pH\textsubscript{i} 5.5. Proton selectivity is imparted by either Asp or Glu at positions 112 or 116, indicating that side-chain length is not critical for this function.

5.7 The D112V/V116S mutant is permeable to Cl\textsuperscript{−}. Replacing CH\textsubscript{3}SO\textsubscript{3}\textsuperscript{−} with Cl\textsuperscript{−} increased outward current in families (A and C) and shifted the tail current reversal potential negatively (B and D). All measurements were in a COS-7 cell at symmetrical pH 5.5 with TMA\textsuperscript{+} CH\textsubscript{3}SO\textsubscript{3}\textsuperscript{−} solutions or external TMA\textsuperscript{+} Cl\textsuperscript{−} solutions, as indicated. Pulses were applied in 10 mV increments from -40 to 80 mV for families and in 10-mV increments for tail currents as indicated. The cell was held at -30 mV and returned to $V_{\text{hold}}$ after pulses. The same calibration bars apply to families (A and C) and tail currents (B and D).

5.8 The third arginine position in the S4 segment is accessible to the internal solution in the open state. (A) The black trace shows proton current in D112V/V116D/R211H in an inside-out patch from a HEK-293 cell. 10 \( \mu \)M Zn\textsuperscript{2+} was introduced into the internal solution during the red trace, and blue is the subsequent pulse in the continued presence of Zn\textsuperscript{2+}. (B) The blue record is from D112V/V116D/R211H in the presence of 0 \( \mu \)M Zn\textsuperscript{2+}, and the red record shows recovery from block on the addition of EGTA shortly after the start of the pulse. Black represents after washout. (C) The control mutant, D112V/V116D, in a COS-7 cell exhibits weak IC Zn\textsuperscript{2+} sensitivity: black, before; blue, in 0 \( \mu \)M Zn\textsuperscript{2+}; gray, after washout. All pulses are to +50 mV at pH\textsubscript{i} 7.0 and pH\textsubscript{o} 7.0 in inside-out patches of membrane.
5.9 Axial distribution of each of the four helical segments, S1S4, in each of the five systems (WT, VD, VS, D112V, and D112S) accumulated from the MD simulations in a lipid bilayer (solid line). Gaussian fits are also shown (dashed lines) for each of the curves. The five systems retained highly similar helical arrangements, with only small deviations in the average position of each helix relative to the rest of the protein. In all five systems, the average position of helix S1 was $0.73 \pm 0.06 \text{ Å}$ (SEM between the five systems), whereas helices S2, S3, and S4 were located at $0.34 \pm 0.05 \text{ Å}$, $-1.80 \pm 0.16 \text{ Å}$, and $0.53 \pm 0.06 \text{ Å}$, respectively. In other words, the registry of the four helices relative to one another was identical throughout the simulations of all five systems.... 130

5.10 Axial distribution of S1-S4 from the five systems (colored lines) are superimposed, and a unique Gaussian fit is obtained for the combined data for each helix (salmon surface). The distribution of S1 fits a Gaussian distribution at $0.75 \text{ Å}$ (standard deviation $\sigma = 0.4$ ). Likewise, the fits for helices S2, S3, and S4 were 0.3 ($\sigma = 0.5$), -1.8 ($\sigma = 0.7$), and 0.5 ($\sigma = 0.4$), respectively. The results emphasize the small amplitude of the axial fluctuations and show that S1 does not exhibit significant plasticity in the axial direction. 131

5.11 Pore hydration is similar in WT and several mutant channels despite very different selectivity. Average water density within a 0.7 nm radius of the mean axis of the pore is plotted, normalized to the bulk water density for 5,000 snapshots from each replica of different systems. The membrane boundaries are indicated by dashed lines, with the external surface to the right. The nadir is near Phe$_{150}$ in all cases. Average axial water density for: (A) WT (proton selective) and D112V (VAL, nonconducting); (B) WT and D112V/V116S (VS) and D112S (S), two anion-permeable channels; (C) WT and D112V/ V116D (VD), of which both are proton selective. 132

5.12 The EC salt-link network realigns in mutants. Representative snapshots of ionic networks in the WT and mutants, with the external end up. (A) WT protein in contact (left) and water-mediated (right) states of the D112R208 ion pair. (B) VD mutant, with R208 participating in (left) and free from (right) the EC salt-link network. (C) VS mutant. (D) D112S mutant. (E) D112V mutant. Acidic and basic side chains in the external and internal funnel, together with side chains of residues 112, 116, F150, and R211 are shown in licorice representation together with ribbon traces of the four $\alpha$-helical transmembrane regions: red, S1; yellow, S2; green, S3; blue, S4. Salt bridges are shown as orange lines. The volumetric surface of water within the pore is shown with a water radius of 0.14 nm. 133

5.13 Several configurations of Asp$_{116}$ in the D112V/V116D mutant. Two-dimensional histogram of the distances from the center of charge of D116 to that of R205 ($d_1$) and R208 ($d_2$), respectively, with increasing probability of ion pairing from purple to green. Snapshots surrounding the graph illustrate each type of interaction circled in the graph: (A) linked D116-R205 pair, 37% of the time; (B) open (no salt bridge), 18%; (C) D116 linked to R205 and R208 simultaneously, 33%; (D) D116-R208 salt bridge, 12%.... 134
5.14 Effect of the charge distribution of the channel on the energetics of ion translocation in WT and mutant channels. Orientation is as in Fig. 5.11, with the external solution to the right. Static-field energy for the transfer of a positive point charge in (A) WT, D112V, VD when R208 is unbound to the EC salt-link network, and two anion-permeable channels, VS and D112S. (B) WT and VD when R208 is forming salt link(s) (SL) with V116D or D185, or both.

6.1 Convergence of the PMF profiles for Na$^+$ translocation in wild-type and mutant $\delta$HV1. A) wild-type; B) single mutant D112S; C) double mutant D112V/V116S. In each plot, the average PMFs from four replicas were generated for five consecutive time intervals of 10 ns (thick coloured lines except black) from 0 to 50 ns. The average PMFs from 10-50 ns are depicted in black lines. The black error bars represent the standard error of the mean (SEM).

6.2 Comparison of the PMFs for Na$^+$ between wild-type and mutants. The average free energy profiles corresponding to the time interval between 10 and 50 ns (Fig. 6.1) are shown in thick coloured lines. Error bars represent the SEM.

6.3 Ligand coordination of Na$^+$ through the channel. Top: Representative snapshots highlighting Na$^+$ coordination in the wild-type channel within the specified regions. Water molecules (red and white licorice) and acidic residues (cyan and red licorice) in the first hydration shell of Na$^+$ are highlighted. Arg$^{208}$ and protein backbone are shown in licorice (cyan and blue) and ribbon (white), respectively. Bottom: A) wild-type. B) single mutant: D112S. C) double mutant: D112V/V116S. In each plot, the replica-averaged axial distributions of water O atoms (blue), polar heavy atoms of the channel (green) and counter-ions (orange) coordinating the Na$^+$ ion are represented in thick lines. The total coordination by water, channel and counter-ions is shown in black thick lines. Error bars represent the SEM. The PMF is shown in red dotted lines. The extent of the channel pore in the bilayer is demarcated by thin vertical red dotted lines.

6.4 Comparison of the coordination of Na$^+$ between wild-type and mutants. Top: Representative snapshots of Na$^+$ coordination at position 112. Bottom: Coordination by A) all the polar heavy atoms B) water O atoms C) polar heavy atoms of the wild-type (black) channel and in D112S (orange), and D112V/V116S (brown) mutants. The grey-shaded area demarcates the sidechain position of residue 112.

6.5 Coordination of Na$^+$ by channel residues. Top: The acidic residues coordinating Na$^+$ are highlighted in colored licorice. The protein backbone is in a white ribbon representation. Bottom: A) wild-type. B) single mutant: D112S. C) double mutant: D112V/V116S. In each plot, the replica-averaged coordination number of Na$^+$ by polar heavy atoms of the channel is shown in a green thick line. The decomposition of the channel coordination by acidic residues: D112 (yellow), E171 (red), E153 (green), D174 (blue), D185 (magenta), E119 (orange), D123 (violet), E196 (cyan), and E164 (brown) are shown in transparent filled curves. The extent of the channel pore is demarcated by thin vertical red dotted lines.

6.6 Differences in free energy and static field energy due to single- and double-point mutations.
6.7 Solvation of water molecules chelating Na\(^+\) inside the pore. A) Wild-type; B) single mutant D112S; C) double mutant D112V/V116S. In each plot, the replica-averaged coordination number of water molecules coordinating Na\(^+\), by water O atoms (blue), polar heavy atoms of the channel (green), and nonpolar heavy atoms of the channel (yellow), are shown in thick lines. The total coordination by polar heavy atoms is shown in a thick dotted black line. The coordination number by Phe\(^{150}\) is shown in a black filled-curve.

6.8 Nonpolar coordination of the hydrated Na\(^+\) in the region corresponding to \(-0.5 < z < 0\) nm. Side (top) and top-down (bottom) views of representative snapshots highlighting residues D112 (wild-type), D112S (single mutant), D112V (double mutant), R208, and R211 in cyan, red and/or blue licorice. Hydrophobic residues (yellow licorice) coordinating the hydrated (red and white licorice) Na\(^+\) (blue sphere) and protein backbone (white ribbon) are highlighted.

6.9 Convergence of the PMF profiles for Cl\(^-\) translocation in wild-type and mutant \(hHv1\). A) wild-type; B) single mutant D112S; C) double mutant D112V/V116S. In each plot, the average PMFs from four replicas were generated for five consecutive time intervals of 10 ns (thick coloured lines except black) from 0 to 50 ns. The average PMFs from 10-50 ns are depicted in black lines. The black error bars represent the SEM.

6.10 Comparison of the PMFs for Cl\(^-\) in wild-type and mutant \(hHv1\). The average free energy profiles corresponding to the time interval between 10 and 50 ns (Fig. 6.9) are shown in thick coloured lines. Error bars represent the SEM.

6.11 Ligand coordination of Cl\(^-\) through the channel. Top: Representative snapshots highlighting Cl\(^-\) coordination within the specified regions of the three systems. Counter ions, Na\(^+\) (blue sphere), water molecules (red and white licorice) and basic residues (cyan and blue licorice) in the first hydration shell of Cl\(^-\) are highlighted. Residues at position 112 and protein backbone are shown in licorice (cyan and/or red) and ribbon (white), respectively. Bottom: A) wild-type. B) single mutant: D112S. C) double mutant: D112V/V116S. In each plot, the replica-averaged axial distributions of water O atoms (blue), polar heavy atoms of the channel (green) and counter-ions (orange) coordinating the Cl\(^-\) ion are represented in thick lines. The total coordination by water, channel and counter-ions is shown in black thick lines. Error bars represent the SEM. The average free energy profile is shown in red dotted lines.

6.12 Comparison of the coordination of Cl\(^-\) in the wild-type and mutant channels. Coordination by A) all the polar heavy atoms B) water O atoms C) polar heavy atoms of the wild-type (black) channel and in D112S (orange), and D112V/V116S (brown) mutants. The grey-shaded area demarcates the sidechain position of residue 112.
6.13 **Coordination of Cl⁻ by polar and charged groups of the channel.** Top: The basic residues coordinating Cl⁻ are highlighted in colored licorice. The protein backbone is in a white ribbon representation. Bottom: A) wild-type. B) single mutant: D112S. C) double mutant: D112V/V116S. In each plot, replica-averaged coordination number of Cl⁻ by the polar atoms of the channel is shown in a green thick line. The decomposition of the channel coordination by basic residues: R208 (red), R211 (green), R205 (blue), R93 (magenta), K157 (orange), and K131 (violet) are shown in transparent filled curves. The free energy profile is shown in red dotted lines. The extent of the channel pore is demarcated by thin vertical red dotted lines.

6.14 **Solvation of water chelating Cl⁻ inside the pore** A) wild-type. B) single mutant: D112S. C) double mutant: D112V/V116S. In each plot, the replica-averaged solvation of water coordinating the Cl⁻, by water O atoms (blue), polar heavy atoms of the channel (green) and nonpolar heavy atoms of the channel (orange) are shown in thick lines. The total coordination by polar atoms, i.e from water O and channel, is shown in a thick dotted black line. The coordination number by Phe₁₅₀ is shown in a black filled-curve. The average free energy profile is shown in red dotted lines.

6.15 **Nonpolar coordination of partially hydrated Cl⁻ in the region corresponding to -0.5 < z < 0 nm.** Side (top) and top-down (bottom) views of representative snapshots highlighting residues D112 (wild-type), D112S (single mutant), D112V(double mutant), R208, and R211 in cyan, red and/or blue licorice. Hydrophobic residues (yellow licorice) coordinating the hydrated (red and white licorice) Cl⁻ (red sphere) and protein backbone (white ribbon) are highlighted.

6.16 **A representative snapshot highlighting the key residues responsible for nonpolar coordination of the second hydration shell of Cl⁻ or Na⁺ in the narrow region of the wild-type channel.** Nonpolar carbon atoms (yellow licorice) of residues Asp¹¹², Val¹⁰⁹, Val¹⁰⁵ on S1, Phe¹₅₀, Glu¹⁵₃ on S2, Val¹⁷₈ on S3 and Arg²⁰₈ and Arg²¹₁ on S4 and other charged residues (cyan, red, blue) of Figs. 6.5 and 6.13 are highlighted. The protein backbone is shown as a white ribbon.

6.17 **Comparison of the PMF profiles for Na⁺ (blue) and Cl⁻ (red).** The average free energy (see Fig. 6.1 and 6.9) are shown in thick coloured lines for A) wild-type, B) D112S, and C) D112V/V116S.
Chapter 1

Introduction

“Computers were made for biology: biology would never had advanced as it did without the dramatic increase in computer power and availability.”

Michael Levitt\textsuperscript{14},

A 2013 Nobel Laureate in chemistry, together with Martin Karplus and Arieh Warshel

for the development of multiscale models for complex chemical systems.

I present here the study of molecular basis for ion and charge selectivity of the human voltage - gated proton channel, hHV1, using molecular dynamics simulations as the main technique. To this end, I wish to take the reader through three main sections in this chapter. First, I begin by introducing the process of ion transport in biological systems and ion channels as one of the molecular machineries to transport ions across cells. Secondly, I describe the current understanding of hHV1 as characterized by electrophysiology, followed by its physiological relevance and an overview of the sequence and structure of hHV1. The final section entails a review of the general features of proton solvation and transport in water and proton wires followed by the current understanding of proton transfer in two well-studied channels. The chapter ends with motivation and aims of the thesis.
1.1 Ions in biological systems

Biological systems are reservoirs of charged particles, ions. Unicellular organisms from bacteria to protozoa as well as multicellular organisms from plants to human bodies choreograph a plethora of actions in which ions are involved. Life as we know it evolved relying on ions.\textsuperscript{15,16}

Biological ions can be categorized into two groups: organic and inorganic ions. Unlike the organic ions, inorganic ions do not contain carbon atoms bonded together. The simplest form of inorganic ions carry a single nucleus like, Ca\textsuperscript{2+}, K\textsuperscript{+}, Na\textsuperscript{+}, Mg\textsuperscript{2+}, H\textsuperscript{+} and Cl\textsuperscript{−}. Many ions of this group diffuse freely in the aqueous medium inside and surrounding cells, but not across the cell boundaries – the cell membrane, which is a 5-6 nm\textsuperscript{17} thick bilayer composed of amphipathic phospholipids. The hydrophobic core formed by the hydrocarbon chains of phospholipids restricts the free passage of polar or charged species and therefore presents almost an impenetrable barrier to ions. Nevertheless, cells require such species to enter and exit cells as and when required, for example as nutrients or waste. Nature’s solution to avoid permanent separation of polar and charged species by the lipid bilayer is membrane proteins. Embedded in the lipid bilayer, some of these protein are specialized to selectively transport ions across the membrane, thus establishing a transport mechanism critical for many cellular processes.

1.1.1 The transport of ions across cell membranes.

The transport across the lipid bilayer is mediated by three types of membrane proteins: channels and channels, passive transporters, and active transporters.\textsuperscript{17} In general, pores and channels provide a water-filled central passage through them for ions and small molecules to diffuse in a direction down a concentration gradient. This process does not require energy. (The term pore is used for bacteria, and channel for animals.)\textsuperscript{17} For example, the K\textsuperscript{+} channels in membranes of nerve tissues allow outward permeation of K\textsuperscript{+} ions more easily than Na\textsuperscript{+}.\textsuperscript{18} The transport via passive and active transporters relies on a cycle of conformational changes linked to solute binding on one side of the membrane and dissociation on the opposite side of the membrane.\textsuperscript{19} Passive transporters too do not require energy since the transport occurs down the concentration gradient of the solute. For example, glucose transporter 1 (or GLUT1) in erythrocytes is responsible for glucose uptake from the blood plasma where glucose is high in concentration after the intake of a meal. In contrast, active transporters require energy to move a solute against its concentration gradient. The most common energy source for this purpose is ATP. Active transporters like Na\textsuperscript{+}-K\textsuperscript{+}ATPase pumps are found in all animal cells. By hydrolysing one ATP molecule, a Na\textsuperscript{+}-K\textsuperscript{+}ATPase pumps three Na\textsuperscript{+} ions out of cells while pumping two K\textsuperscript{+} ions into the cells, against their concentration gradients, thus creating and maintaining ion concentration gradients across the
cell membrane. Likewise, plasma membrane Ca\(^{2+}\) ATPase (or PMCA) is involved in removing Ca\(^{2+}\) from all eukaryotic cells.

1.1.1.1 Ion channels

Ion channels, a ubiquitous class of specialized membrane proteins for passive diffusion are the broad interest of this thesis. Ion channels in plasma membrane are important for two major physiological functions: transepithelial transport and electrical excitability. In addition, ion channels found in membranes of internal organelles are involved in acidification of endosomes, with secondary effects on endocytosis.

Transepithelial transport can either be involved in absorption, i.e transport from lumen (via apical membranes) to blood, or secretion, i.e transport from blood (via basolateral membranes) to lumen. For example, cystic fibrosis transmembrane conductance regulator (or CFTR) responsible for producing mucus, sweat, saliva, tears, and digestive enzymes, secrets Cl\(^{-}\) into the lumen, leading to a net flow of water into the lumen. This process is crucial for the salt and water balance on epithelial surfaces and production of sweat, digestive fluids, and mucus. The epithelial sodium channel (or ENaC) present in apical membranes of distal nephron mediates the passive Na\(^{+}\) influx from urine into the cell, hence involved in Na\(^{+}\) reabsorption and in regulating blood pressure.

Electrical excitability (i.e the ability of a cell to react to an electric stimulus) of excitable cells such as neurons, cardiac and skeletal muscle is a fundamental cellular response. The membrane potential across plasma membrane is central to electrical excitability, which results from the unequal distribution of electrical charge, carried by ions, on the two sides of the membrane. A nerve impulse travels in the form of action potentials, large and rapidly reversible fluctuations in the membrane potential, propagating along the axon. Two major ion channels contribute to the action potential: voltage-gated Na\(^{+}\) and K\(^{+}\) channels. Action potentials are initiated by the opening of voltage-dependent Na\(^{+}\) channels and the resulting influx of Na\(^{+}\), which depolarizes the membrane. The membrane voltage is repolarized to its normal value by K\(^{+}\) efflux through voltage-dependent K\(^{+}\) channels. The long duration of cardiac action potential is due to the Ca\(^{2+}\) influx through voltage-gated Ca\(^{2+}\) channels that stay open longer for the contraction of the cardiac muscle.

1.1.1.2 The importance of understanding ion channels

Inevitably, functional defects of almost all ion channels result in impairments of cellular processes that may even lead to causing diseases. Some of the functional defects of the above-mentioned ion channels and their associated diseases include: (1) hyperexcitability due to slowed deactivation of voltage-gated Na\(^{+}\) and K\(^{+}\) channels, causing generalized epilepsy, (2) prolonged action potential of the voltage-gated
Ca\textsuperscript{2+} channel, leading to cardiac arrhythmia syndromes, (3) gain-of-function mutations in ENaC, causing increased renal Na\textsuperscript{+} uptake and leading to Liddle syndrome, (4) disruptions of CFTR function, causing thickening of the luminal secretions that ultimately leads to cystic fibrosis.\textsuperscript{22,25} These are just a few examples from an exhaustive list of human ion-channel diseases, ‘channelopathies’.\textsuperscript{22} In fact, the list of channelopathies is expanding so rapidly even including potential entries of certain types of cancer, psychiatric disorders, gastrointestinal diseases, and additional nervous system disorders.\textsuperscript{26,27}

Given the wide spectrum of cellular processes driven by ions channels, their ubiquitous presence in a diverse tissue distribution and membrane localization make ion-channel-targeted drug discovery highly compelling.\textsuperscript{28} The reality is, despite the fact that the human genome has more than 400 putative ion channels,\textsuperscript{28} our understanding of the intimate drug-channel interaction and subsequent therapeutic effects is limited only to a fraction of ion channels. Hence, there is an unwavering effort to understand how ion channels work and how they can be exploited for medical treatments.

In biophysics, this challenge is pursued by the investigation of the structure-function relationship in ion channels. To begin with, structures of ion channels are notoriously difficult to resolve. Determining the structure of ion channels by means of crystallography was once even regarded almost impossible.\textsuperscript{29} Progressing from structure, once it is resolved, demands methods of functional characterization with both high spatial and temporal resolutions. Eventually and more importantly, understanding the structure-function relationship of ion channels lead us to advancing in knowledge in molecular pathology of ion-channels-related diseases and better drug design strategies.

1.1.2 Properties of ion channels

In the study of ion channels, four main properties of ion channels are explored, providing the most fundamental understanding about how ion channels work: single channel conductance, ion selectivity, gating and pharmacology.\textsuperscript{24}

1.1.2.1 Single-channel conductance

This is a measure of the rate at which ions diffuse through an open channel. Typically, the single-channel conductance ranges from 0.1 to 100 pS (picosiemens).\textsuperscript{30}

1.1.2.2 Ion selectivity

Selectivity is the preference of a channel to conduct a specific ion over others. Ion diffusion via channels is affected by the chemical gradient as well as the electric gradient. This is because although ions move
down the chemical gradient in one direction, say from the extracellular side to the intracellular side (or vice versa), the gradual charge build-up in the intracellular side generates an electrical potential opposing further movement of ions into the intracellular side. When the chemical and electrical gradients are equal in magnitude, there is no net movement of ion and the system is said to be in electrochemical equilibrium. The electrical potential at which this equilibrium is reached is known as the Nernst equilibrium potential ($E_S$) and is calculated for each ionic species (S) using the Nernst equation (equation 1):

$$E_S = \frac{RT}{z_S F} \ln \frac{[S]_o}{[S]_i}$$  

(1)

where $R$ is the gas constant, $T$ is the absolute temperature, $z_S$ is the charge on the ionic species, $F$ is Faraday’s constant, $[S]_o$ is the external concentration of the ion and $[S]_i$ is the internal concentration.

The above equation is only applicable if there are not any other ion species present in either compartments separated by the membrane, but the permeating ion. The Goldman-Hodgkin-Katz (GHK) voltage equation (equation 2) is more appropriate to compute the reversal potential ($E_{rev}$) at which there is no net ion flux across the membrane when it is separating many different ionic species.$^{32-34}$ Also the GHK equation is a classical way of characterizing the selectivity of ion channels.

$$E_{rev} = \frac{RT}{F} \ln \frac{P_{Cl^-} [Cl^-]_i + P_{K^+} [K^+]_o + P_{Na^+} [Na^+]_o + P_{H^+} [H^+]_o}{P_{Cl^-} [Cl^-]_o + P_{K^+} [K^+]_i + P_{Na^+} [Na^+]_i + P_{H^+} [H^+]_i}$$  

(2)

$P_S$ is the permeability to ion S. The different Nernst potentials are weighted by the relative permeability of different ions. Measurements of $E_{rev}$ thus allow one to calculate permeability ratios.

### 1.1.2.3 Gating

*Gating* corresponds to the opening and closing of the channel under the influence of different kinds of modulators like binding/unbinding of ligands, variations in membrane potential, mechanical state of the membrane and environmental stimuli such as temperature and light.$^{18,35}$

### 1.1.2.4 Pharmacology

This refers to the response of the channel to various compounds that may influence channel properties. More specifically, such compounds help define the functional regions of the channel.
1.1.3 The study of ion channels: the progression from electrophysiology to MD simulations

The details of this section focuses particularly on the voltage-gated ion channels, but the notion of the so-called progression is generally true in the study of almost any ion channel.

Establishing the structure-function relationship for the ubiquitous family of voltage-gated ion channels is of crucial importance to understanding their involvement in a wide range of fundamental physiological functions and therapeutic applications. Current measurements typically reveal the existence of ion channels and in the case of the voltage-gated K\(^+\) and Na\(^+\) channel, this became evident in 1950s.\(^{36-40}\) High resolution structures of K\(^+\) and Na\(^+\) channels became available since 1998\(^{41}\), which thereafter with the use of MD simulations greatly contributed to the current understanding on the molecular basis of ion selectivity and ion permeation.\(^{42,43}\)

1.1.3.1 Electrophysiology

Because ions carry charges, their diffusion through channels can be measured using electrical recording techniques of electrophysiology. These experiments, providing macroscopic functional measurements, typically give the first glimpse on how ion channels work. Using voltage clamp experiments on a giant squid axon in 1950s, Hodgkin and Huxley were able to identify the first ever voltage-gated cation K\(^+\) and Na\(^+\) currents and deduce that they were responsible for the propagation of the action potential.\(^{44}\)

1.1.3.2 Structural studies

In 1998, the landmark determination of the three-dimensional crystal structure of the bacterial KcsA potassium channel paved the path, for the first time, to elucidating the underlying atomic-level origins of the functional observations—the selectivity and conduction of K\(^+\).\(^{41}\) On the basis of a static view of this structure, the molecular basis for the selectivity of K\(^+\) ion was probed, which also applies to the vertebrate voltage-gated K\(^+\) channels since the sequence of the pore region is nearly identical between the two classes of K\(^+\) channels.

Accordingly, the classical (and commonly accepted) explanation suggests that the ion selectivity is predominantly determined by structural factors, in that the backbone carbonyl oxygens of the narrow region of the pore where the selectivity is conferred, the selectivity filter, coordinate the nearly dehydrated K\(^+\), compensating for the cost of dehydration of a K\(^+\) ion. The selectivity filter is too wide and rigid in space to coordinate the much smaller Na\(^+\), resulting in a higher solvation penalty for Na\(^+\) in the pore than in water.\(^{45}\)
1.1.3.3 Molecular dynamics simulations

The basis for this simplified view on ion selectivity is a rigid pore that is solely optimized for coordinating $\text{K}^+$ ions and not for $\text{Na}^+$ ions. However, it long has been observed that proteins have a rather fluid-like, dynamic structure with rapid conformational fluctuations. Hence, the ion-protein interactions and their effect on ion permeation should not be limited to a single structure, but should be treated as a dynamic process with an atomic-level resolution. Molecular dynamics (MD) simulations are capable of generating an ensemble of conformations corresponding to a conducting channel model starting from the crystal structure. These conformations are generated on the basis of a body of physical principles which accounts for the experimentally-observed properties of biomolecules. Hence, MD simulations can provide a more descriptive and temporal view of the ion selectivity and mechanism of ion conductance with an atomic-level resolution.

With long enough time scales of simulations and triggers mimicking modulators of gating, MD simulations can also provide details about the molecular mechanism of gating. Furthermore, one of the most powerful applications of molecular dynamics simulations is the ability to quantify the underlying energetic basis of ion selectivity, conductance and gating. From the pharmacological point of view, MD simulations also enable the designing and optimizing of potential drugs that maximize the interaction with ion channel. Collectively, MD simulations are a powerful tool to understand the molecular basis for fundamental properties of ion channels.

I will now begin the story of $\text{hHV}_1$. This entails a brief historical timeline of the major discoveries, electrophysiological characterization, physiological relevance and an overview of the sequence and structure of $\text{hHV}_1$. The section ends with the key questions remaining about $\text{hHV}_1$ which I have explored in this thesis.

1.2 The voltage-gated proton channel, $\text{Hv1}$

The existence of a voltage-gated proton channel was first proposed by Fogel and Hastings in 1972 based on their observation of pH-regulated bioluminescence in unicellular eukaryotic marine algae, dinoflagellates. A decade later, Thomas and Meech reported large proton current recordings by depolarizing snail neurones. Subsequently, their work was confirmed to be the first voltage-clamp study of $\text{Hv1}$. What followed next, until the discovery of the genes coding for proton channels in 2006, was a proliferation of electrophysiological studies of $\text{Hv1}$, which were predominantly voltage-clamp experiments.
conducted under various cellular conditions (ex: pH, voltage) as different cell types expressing proton channel were discovered in human and other species.³⁴

1.2.1 Characterization of Hv1 by electrophysiology

Two properties that helped characterize Hv1 early on were selectivity and gating.⁵¹ Selectivity is the preference of a channel to conduct a specific ion over others. Electrophysiology characterizes selectivity by determining the Nernst equilibrium potential (a zero-current potential) and the permeability ratio between the ion of interest and a reference ion(s) from the reversal potential (a zero-current potential). Gating refers to conformation changes of the channel responsible for the opening and closing of the pore.¹⁸ In the case of voltage-gated ion channels, movements of certain charges, “gating charges”, as part of the conformational changes of gating elicit gating currents that can be measured and characterized to understand the gating mechanism.

1.2.1.1 Selectivity

For example, given that the extracellular (denoted by the subscript o) and intracellular (denoted by the subscript i) concentrations of protons in muscle cells are 4 x 10⁻⁵ and 1 x 10⁻³ mM respectively, the Nernst equilibrium potential for protons, E_H, is -86 mV.³⁵ The permeability ratios of P_H/P_S are in the orders of 10⁷ or 10⁸ in the presence of TMA or a combination of Cs⁺, K⁺, Na⁺.⁵² If the only permeable ion is proton, equation 2 simplifies to the Nernst equilibrium potential for proton. In the case of Hv1, the change of E_H due to changing [H⁺]ᵢ (or pHᵢ) is comparable to that of the V_rev, which remains the same when the concentrations of other ions are altered. This led to the conclusion that Hv1 is selective for protons.⁵²

1.2.1.2 Gating

The gating of voltage-gated proton channels is dependent on both membrane potential and pH. This dependence can be characterized in many ways. Current recordings elicited at changing potentials reveal the magnitude of the proton current increases with depolarizing potentials. With decreasing pHᵢ, i.e increasing cytoplasmic [H⁺] at constant pHₒ, the channel can be activated at increasing negative potentials.⁵³ Hence, the voltage at which the channel opens is determined by the pH gradient, ΔpH (ΔpH = pHₒ - pHᵢ). More specifically, an increase of pHₒ or decrease of pHᵢ by one unit shifts the current-voltage curve by 40 mV to more negative voltages.³⁴,⁵² The molecular mechanism by which ΔpH determines the voltage dependence of gating of Hv1 is not clear.⁵²
In addition to selectivity and gating, two other properties are characteristic of H\textsubscript{V}1. They are the (1) small single-channel conductance, (2) strong temperature dependance on both conductance and gating. The estimate of the single-channel conductance of H\textsubscript{V}1 is 10 fS at 100 mV, which corresponds to a flux of $\sim$6000 H\textsuperscript{+}/s.\textsuperscript{54} This ion flux of H\textsubscript{V}1 at physiological conditions is $\sim$10\textsuperscript{3} times smaller than that of most channels in similar conditions. For example, the K\textsuperscript{+} channel exhibits a conductance of 20 pS at 50 mV corresponding to a flux of $\sim$6x10\textsuperscript{6} K\textsuperscript{+}/s.\textsuperscript{55} The intracellular concentration of K\textsuperscript{+} is, however, 10\textsuperscript{6} times of H\textsuperscript{+} ([K\textsuperscript{+}] is $\sim$ 150 mM in mammalian skeletal muscle\textsuperscript{31}). Hence, compared to other voltage-gated cation channels, voltage-gated proton channels have a small conductance at physiological conditions.\textsuperscript{56} Rising temperature increases the magnitude of the proton current and the speed of gating (channel activation and deactivation) in H\textsubscript{V}1 in a way that is greater than in most other ion channels.\textsuperscript{57,58}

1.2.2 Physiological roles of hH\textsubscript{V}1

The presence of hH\textsubscript{V}1 in the human body is ubiquitous. Proton transfer (hereafter referred to as PT) in hH\textsubscript{V}1 is involved in a range of cellular functions of different cell types. The following subsections describe the well-studied cellular functions highlighting the role of PT in hH\textsubscript{V}1.

1.2.2.1 Phagocytosis by white blood cells

The human H\textsubscript{V}1 is found in various kinds of white blood cells such as neutrophils, eosinophils, macrophages, and other leukocytes.\textsuperscript{52} A key function common to all these cell types is the killing of pathogenic bacterial and other microbes in a process commonly named as phagocytosis.\textsuperscript{59} In the presence of a pathogen, the NADPH oxidase complex embedded in cell membrane is activated, generating NADP\textsuperscript{+} and H\textsuperscript{+}, and translocating electrons out of cells. These electrons reduce O\textsubscript{2} to reactive oxygen species that are responsible for killing pathogens. Thus, the overall effect of NADPH oxidase activity leads to membrane depolarization and acidification of the cytoplasm. To reduce the membrane depolarization due to electron efflux and reset the pH homeostasis, protons are extruded out of the cell via hH\textsubscript{V}1.\textsuperscript{52}

1.2.2.2 pH regulation in airway surface liquid

The airway surface liquid (ASL) is the thin layer of fluid covering the upper (nose, pharynx) and lower (larynx, trachea, bronchi, lungs) respiratory tracts.\textsuperscript{60} The ASL is thought to be involved in airway hydration and innate defense of which the proper function is pH regulated. The pH range of the ASL may vary from 5.5 to 8.3. It is shown that when ASL is alkalinized to a pH value $>$7, H\textsubscript{V}1 of the airway epithelium is activated to acidify the ASL.\textsuperscript{61} Since the airway epithelium has a relatively constant
membrane potential at around $\sim -20$ mV, the proton currents of HV1 in the airway epithelium seem to
depend more on the pH$_o$ and less on the membrane potential.$^{52}$

1.2.2.3 Histamine release by basophils

Histamine is an organic nitrogenous compound and histamine release is responsible for activating nu-
merous allergic inflammatory responses.$^{62}$ Basophils, a type of white blood cells, secrete histamine in
response to allergen binding to antibodies of basophils. The process of histamine release by allergen-
antibody stimulation results in cytosolic acidification and Ca$^{2+}$ influx. Proton extrusion via $hHV1$ thus
attenuates acidity and compensates the excess positive charge inside basophils.$^{52}$

1.2.2.4 Sperm capacitation

In the female reproductive tract, spermatozoa undergo capacitation—a functional maturation process
which enhances motility of sperm, and leads to burrowing through the oocyte coat.$^{63}$ It is proposed that
low pH$_i$ keeps sperm inactive until it enters the female reproductive tract and that high concentration of
Zn$^{2+}$, the most potent inhibitor of $hHV1$, in seminal fluid prevents protons efflux through $hHV1$.$^{63}$ As
the sperm enters more alkaline medium with low levels of Zn$^{2+}$ in the female reproductive tract, $hHV1$
are free to open and extrude protons, leading to activation of several pH-sensitive processes of sperm
capacitation such as sperm mobility, metabolism, and hyperactivation.$^{63}$

1.2.3 Human HV1 in pathological conditions

$hHV1$ has also been implicated in several pathological conditions. Of these, ischemic stroke, colorectal
and breast cancer metastasis are well examined, highlighting potential therapeutic applications of $hHV1$.

1.2.3.1 Ischemic stroke

Ischemic stroke is characterized by loss of blood flow to the brain due to an obstruction within a blood
vessel of the brain. The subsequent cellular response elevates the levels of reactive oxygen species (ROS)
which overwhelms antioxidant defences causing further damage to normal brain tissues.$^{65,66}$ Measures to
alleviate damage by ROS have been pursued after the antioxidant strategies failed to be fully effective.$^{67}$

NADPH oxidase of microglia, macrophages of the brain and spinal cord, contributes to ROS pro-
duction.$^{65}$ High levels of HV1 are found in human microglia that enables ROS production by phago-
cytosis.$^{68}$ It has been shown that mice lacking HV1 in microglial (HV1 knockdown by small hairpin
RNA - shRNA) reduces in situ ROS production after stroke and reduces neuronal damage compared
to wild-type after stroke. These observations led to the conclusion that \( \text{HV}_1 \) enhances brain damage from ischemic stroke and provided the rationale that \( \text{HV}_1 \) in microglia may serve as a more selective therapeutic target for treatment of brain injury during ischemia.\(^{68}\)

### 1.2.3.2 Cancer

The expression of \( \text{HV}_1 \) has been examined in both human cancer biopsies and established cancer cell lines of colorectal and breast cancers, using immunohistochemistry and immunofluorescence.\(^{69-71}\) High levels of \( \text{HV}_1 \) are found in highly-metastatic cancer lines compared to low levels of \( \text{HV}_1 \) in poorly-metastatic cancer lines. Further, down regulation of \( \text{HV}_1 \) (\( \text{HV}_1 \) knockdown by a small interfering RNA, siRNA) of highly metastatic cell lines shows decreased levels of \textit{in vitro} cell invasion and migration. Hence, it has been suggested that \( \text{HV}_1 \) plays a role in regulating pH\(_i\), that is important for cancer cell progression and metastasis. In addition, a strong correlation is shown to exist between \( \text{HV}_1 \) expression and clinicopathological characteristics (ex: tumor size, classification, clinical stage) of both breast and colorectal cancers.\(^{70,71}\) These findings collectively indicate that \( \text{HV}_1 \) might be a potential target for anticancer drugs as well as a biomarker for prognosis and diagnosis of these two cancers.\(^{35,71}\)

### 1.2.4 Sequence and structural overview of \( h\text{HV}_1 \)

About 30 years before the postulate of the existence of voltage-gated proton channels, the seminal work done by Hodgkin and Huxley reported the first ever voltage-gated cation (\( \text{K}^+ \) and \( \text{Na}^+ \)) currents which inspired the hypothesis of the existence of ion channels.\(^{33,36-40}\) The subsequent decades of research led to an enormous body of knowledge on voltage-gated \( \text{K}^+ \) (\( \text{K}_V \)) and \( \text{Na}^+ \) (\( \text{Na}_V \)) channels as well as voltage-gated \( \text{Ca}^{2+} \) (\( \text{Ca}_V \)) channels.

The sequences of voltage-gated cation channels as well as voltage-gated enzymes led the bioinformatic search resulting in discovering the genes of \( \text{HV}_1 \) in humans,\(^{72}\) mouse, and \textit{Cionaintestinali}.\(^{73}\) \( \text{HV}_1 \) (273 aa, 31.7 kDa in \( h\text{HV}_1 \)) contains three putative domains: a proline-rich N-terminal domain (<50 aa), a 4-helix transmembrane domain, and a C-terminal coiled-coil domain (<50 aa).\(^{72,74,75}\) \( \text{HV}_1 \) is homologous to voltage-gated cation channels in terms of sequence, structure, and even function, yet certain properties unique to \( \text{HV}_1 \) also exist. The following section is a sequence and structure comparison between \( \text{HV}_1 \) and voltage-gated cation channels.
1.2.4.1 HV1 in the context of other voltage-gated cation channels: sequence and structure

Almost three decades after the work of Hodgkin and Huxley, the eukaryotic genes of $\text{Nav}$ channel\textsuperscript{76} and $\text{Kv}$ channel\textsuperscript{77} were identified and cloned. Subsequently, crystal structures were resolved for prokaryotic and mammalian $\text{Kv}$ channels, revealing atomic details of the structural architecture.\textsuperscript{2,4,78} $\text{Kv}$ channels are homotetramers with each monomer containing a single polypeptide chain organized into six trans-membrane (6TM) helices referred as S1-S6 (Fig. 1.1 A). Helices S5-S6 of each monomer contribute to the formation of the centrally-located ion-conducting pore. Helices S1-S4 of each monomer are organized into a separate subunit consisting of a four-helical bundle that is located to peripherally the pore (Fig. 1.1 B). These subunits are coined “voltage-sensor domain (VSD)” due to their involvement in regulating the pore opening and closing by sensing changes in the membrane potential (Fig. 1.1 C).\textsuperscript{79} In contrast to $\text{Kv}$ channels, the tetramers of eukaryotic $\text{Nav}$ and $\text{CaV}$ are made out of a single polypeptide chain.\textsuperscript{80} Prokaryotic $\text{Nav}$ and $\text{CaV}$ on the other hand are homotetramers akin to $\text{Kv}$ channels.\textsuperscript{81–83}

A short helix between S5 and S6 helices, named P-domain, P-helix, or P-loop, bears the residues forming the canonical “selectivity filter (SF)”, a constriction site that determines the ion selectivity.\textsuperscript{84} A signature sequence of each monomer, Thr-Val-Gly-Tyr-Gly, constitutes the SF of $\text{Kv}$, which is highly conserved among prokaryotic and eukaryotic species. In contrast, the amino acid composition of the SF in $\text{Nav}$ channels is less conserved and less symmetric. Single amino acid from each monomer constitutes the SF of $\text{Nav}$, such that eukaryotic $\text{Nav}$ have EKEE, EEKE, DKEA, or DEKA in the SFs, while the bacterial $\text{Nav}$ exhibits symmetrical EEEE in the SFs.\textsuperscript{85}
Figure 1.1: **K<sub>V</sub>** and **H<sub>1</sub>V** channel structures. (A) Schematic representation of the S1-S6 helix topology of the monomeric K<sub>V</sub> channels, modified from Knöpfel 2009. (B) Top-down view of the crystal structure of tetrameric K<sub>V</sub>1.2/2.1 channel (PDB:2R9R) reprinted from Fowler & Sansom 2013. The helices of one monomer are coloured and labelled. (C) The crystal structure of K<sub>V</sub>AP (PDB:1ORS). Arg residues on S4 helix and acidic residues on S1-S3 residues are in licorice representation. (D) Schematic representation of the S1-S4 helix topology of the monomeric H<sub>1</sub>V channels, modified from Knöpfel 2009. Images are reproduced/modified with permission from authors.

Although the sequence composition of the SF greatly varies among the voltage-gated cation channels, their VSDs display a striking pattern in the conservation of charged residues, which is comprised of 5-7 positively charged Arg (labelled R1-R7) or Lys residues along the S4 helix and negatively charged residues on S1-S3 helices (Fig. 1.1 A&D). The positively charged residues on S4 helix contribute to gating charges. It is proposed that the gating charge-carrying Args in S4 are stabilized by sequential formation of ion pairs with negatively charged residues in S1-S3 helices during the channel activation.

Interestingly, it is these VSDs of voltage-gated cation channels, which are not channels but regulatory units of voltage sensing, that H<sub>1</sub>V shows strong sequence resemblance to. In particular, H<sub>1</sub>V displays more or less the same conservation pattern of charged residues (more details on sequence alignment is presented in Chapter 2). Hence, the putative transmembrane topology of H<sub>1</sub>V is a single VSD-like four helical bundle. The native expresive of H<sub>1</sub>V is a homodimer. Each monomer, however, functions as an independent voltage-sensing proton channel even when separated from the dimer. Due to the lack of a separate pore domain, the SF of H<sub>1</sub>V is supposedly located within the same VSD-like helix bundle. More specifically, the Asp<sup>112</sup> residue putatively located in the middle of S1 helix
is found to be crucial for proton selectivity. Single-point mutations of Asp\textsuperscript{112} to a neutral amino acid makes HV1 anion-selective.\textsuperscript{94}

### 1.2.4.2 HV1 in the context of other voltage-gated cation channels: function

Analogous to that of wild-type VSDs of KV and NaV channels, the Arg residues on the S4 helix of HV1 are shown to contribute to the gating charges.\textsuperscript{95} Interestingly, analogous to the function of HV1 - proton conduction - certain single-point mutations of VSDs were shown to conduct protons under specific cellular conditions: R1H, R4H and R2H or R3H mutants are proton selective at negative, positive and intermediate voltages, respectively.\textsuperscript{96–98} These observations suggest that VSDs and HV1 share a seemingly common architecture to conduct protons.

### 1.2.4.3 Key questions about HV1

Having only a putative VSD-like structure, HV1 is unique from the canonical function of other VSDs—the voltage-sensing, in that HV1 displays a dual function of proton conduction and voltage-sensing. Moreover, HV1 has been suggested to be the most selective ion channel known with no detectable permeability to other ions.\textsuperscript{34}

A landmark study discovered a single-mutation of the anionic Asp\textsuperscript{112}, supposedly located at the centre of the pore, into a small neutral residue turns HV1 into an anionic selective channel, conducting Cl\textsuperscript{−}.\textsuperscript{94} In the absence of an experimentally resolved structure, the molecular basis of HV1 and its mutants imparting charge and ion selectivity is not fully elucidated. Unfortunately, determining a native structure of HV1 has not been successful, impeding the progression of elucidating the structure-function relationship of HV1.

How a VSD-like HV1 exhibits channel properties with proton selectivity is a fundamental question that could potentially to be resolved by establishing the structure-function relationship of HV1.

### 1.3 Proton transport

The study of PT in biological transport systems like channels are pioneered and greatly supported by the understanding from theoretical and experimental studies of PT in water\textsuperscript{99–101}, which is still an active research interest.\textsuperscript{102,103} In fact, as discussed below, PT in biological systems invariably involves water. Hence, the atomic nature of protons in water, leading up to its mobility (in water) is first discussed here followed by a discussion on PT in the context of well-studied biological systems which involves both water and amino acids.
1.3.1 Proton hydration

Because of its small size and its bare positive charge resulting from the lack of an electron shell, H$^+$ is extremely reactive with polar molecules. This is demonstrated by hydration energy of proton, which exceeds that of any other monovalent ion in aqueous solution by more than 100 kcal mol$^{-1}$ reflecting strong interaction between proton and water. Hence, in water, proton does not exist by itself, but reacts with water molecules forming a hydrated proton. A hydronium ion, H$_3$O$^+$, is formed when a proton is located on a single water molecule. The text book description of H$_3$O$^+$ as the dominant hydrated proton structure is regarded too simplistic. Rather, the proton is shared between several water molecules. Two main forms of hydrated proton have been proposed and evaluated using both theory and experiments. The Eigen cation, H$_9$O$_4^+$, is a complex of three water molecules hydrating a H$_3$O$^+$ at its core via hydrogen bonds. The Zundel cation, H$_5$O$_2^+$, is a complex of two water molecules sharing a proton (Fig. 1.2).

![Figure 1.2: Schematic drawing showing (A) the Eigen cation, a hydronium ion with its hydration, and (B) the Zundel cation, modified from Codorniu-Hernández & Kusalik 2013. Images are modified with permission from authors.]

1.3.2 Proton mobility

The Nernst-Plank equation (3) relates ion flux across membrane to both electrical and concentration gradients. The following is one form of the Nernst-Plank equation:

$$J_p = -D \frac{\partial c}{\partial x} - zecu \frac{\partial \Psi}{\partial x}$$

where $J_p$ is the flux of particles, which is defined as the amount of particles passing through a unit area per unit time. The flux due to the concentration gradient is formulated using $D$ for diffusion coefficient of the charged particle, $c$ for its concentration, $x$ for position. The flux due to the electrical gradient is formulated using $z$ for valence of the charged particle, $e$ for elementary charge, $c$ for concentration, $u$ for ionic mobility, $\Psi$ for electrical potential. Thus, the Nernst-Plank equation shows that the ion flux
depends not only on electrical and chemical gradients, but also on the ion mobilities.

The ionic mobility, \( u \) is described using the following equation \( 4 \),

\[
    u = \frac{z e}{6 \pi \eta a}
\]

where \( \eta \) is the viscosity of the medium and \( a \) is the hydrodynamic radius, also known as the Stokes radius\(^{114} \), calculated from the diffusion coefficient using the Stokes-Einstein relation. For example, the mobility of Na\(^+ \) in water can be computed using \( z=1, e=1.602 \times 10^{-19} \) C, \( a=0.184 \) nm\(^{115} \), \( \eta=0.891 \) x \( 10^{-3} \) kg m\(^{-1} \) s\(^{-1} \)(at 298 K). Then \( u \approx 5.2 \times 10^{-8} \) m\(^2\)s\(^{-1}\)V\(^{-1} \). This value means that when there is a potential difference of 100 V across a solution of length 1 m, the drift speed at which Na\(^+ \) will travel is 5.2 \( \mu \)s\(^{-1} \).

The ionic mobility of protons in water at 298 K is 36.23 X \( 10^{-8} \) m\(^2\)s\(^{-1}\)V\(^{-1} \).\(^{105} \) Hence, the drift speed of H\(^+ \) is seven times that of Na\(^+ \) whose ionic radius (0.095 nm)\(^{115} \) is similar to that of H\(_3\)O\(^+ \) (0.100 nm).\(^{116} \) Therefore, proton transfer in water cannot be explained by simple diffusion of H\(_3\)O\(^+ \) or any other hydrated proton species and it is necessary to consider an alternative mechanism. Theodor von Grotthuss inspired such a mechanism for proton transfer, which is now celebrated with his name, the “Grotthuss mechanism”.\(^{99} \)

### 1.3.2.1 Grotthuss Mechanism

At a time when the correct composition of water was yet unknown, in 1806, Theodor von Grotthuss presented an original explanation of the electrolysis of water in Alessandro Volta’s voltaic pile, the first electrical battery.\(^{117} \) He described how water molecules arranged in a linear wire connecting cathode to anode can decompose into their electrolytes and instantly reform with that of the neighbouring water molecules, eliciting a current.\(^{99} \)

Details of the current view of the Grotthuss mechanism are revealed by both experiments\(^{118,119} \) and computational approaches.\(^{120,121} \) The consensus view of “proton jump”\(^{100} \) (commonly referred as “proton hop”\(^{122} \)) between two water molecules in a hydrogen-bonded network involves “structural diffusion”\(^{105} \), a process of proton-exchange between two water molecules as a result of the breaking and forming of a covalent O-H bond, coupled with a rearrangement of the hydrogen-bonds associated with those two water molecules.\(^{120} \) A proton jump between two water molecules is illustrated (Fig. 1.3) below involving the isomerization between Eigen and Zundel cations that is shown to occur during PT in bulk water.\(^{111,112,123,124} \) This process is also coined “Eigen-Zundel-Eigen” transition.\(^{125} \)

An Eigen cation centers a H\(_3\)O\(^+ \) carrying the excess proton. The oxygen atom of this H\(_3\)O\(^+ \) is
denoted O\textsubscript{a}. Each of the three water molecules hydrating H\textsubscript{3}O\textsuperscript{+} is hydrogen bonded, on average, by four ligands (e.g., H\textsubscript{2}O\textsubscript{b}), one of them is the H\textsubscript{3}O\textsuperscript{+} and three of them are water molecules in the second hydration shell of the H\textsubscript{3}O\textsuperscript{+}. Thermal fluctuations lead to a cleavage of a second-shell hydrogen bond. Consequently, the average bond lengths and bond angles readjust delocalizing the proton between the two oxygen atoms, O\textsubscript{a} and O\textsubscript{b}, forming a Zundel cation. Reformation of a second-shell hydrogen bond with the proton-leaving oxygen, O\textsubscript{a}, localizes the excess proton on O\textsubscript{b} forming a new Eigen cation whose oxygen atom of the H\textsubscript{3}O\textsuperscript{+} is O\textsubscript{b}. Likewise, a repeating exchange of Eigen-Zundel-Eigen transitions, in principle, may lead to long-range proton transport.

The PT via Grotthuss mechanism, thus, does not involve a translocation of a single nucleus of proton, but involves a series of successive protonation-dissociation reactions of water molecules in each of which the translocating proton is a new proton.

Kinetic and energetic details of the Eigen-Zundel-Eigen transition are as follows. The hydrogen-bond cleavage between the first and second hydration shell of H\textsubscript{3}O\textsuperscript{+} has been shown to be the rate-limiting step of PT in bulk water. Both experimental and computational evidence suggest that the Eigen cation is slightly more stable than the Zundel cation in liquid water. Nonetheless, PT does not appear to occur between well defined structures of Eigen or Zundel cations. Rather, quantum effects giving rise to numerous unclassifiable structures of Eigen or Zundel cations. Rather, quantum effects giving rise to numerous unclassifiable structures of Eigen or Zundel cations. Rather, quantum effects giving rise to numerous unclassifiable structures of Eigen or Zundel cations. Rather, quantum effects giving rise to numerous unclassifiable structures of Eigen or Zundel cations.

Hence, these two cations can be viewed only as “limiting” or “ideal” structures and by the same token the Zundel cannot be viewed as a typical transition state. A typical proton hop takes 1-2 picoseconds. The energetic barrier between two H\textsubscript{3}O\textsuperscript{+} cores opposing a proton hop under the quantum mechanical treatment is essentially nonexistent. Hence, PT via Grotthuss mechanism...
can be envisaged as a series of successive protonation-dissociation reactions in which reactants, products or any intermediates are roughly iso-energetic sharing the excess proton.\(^6\) This in effect means, PT may involve jumps of several protons simultaneously, giving rise to the notion that the proton could delocalize over extended hydrogen-bonded wires, which was first suggested by Eigen.\(^{104}\)

Next, PT via Grotthuss mechanism in model and real biological systems are discussed wherein protons can move even up to 40 times faster than through bulk water.\(^{51}\) The term “proton wire” emerged in the context of hydrogen-bonded chains formed by side chains of proteins, water or both, capable of biological proton transport.\(^{122,127}\) Although PT along proton wires may take place according to Grotthuss mechanism, certain details of the mechanism are markedly different from that of the bulk water. The following section highlights key features of PT in proton wires as revealed mainly by theoretical studies.

### 1.3.2.2 Proton wires

A single file of water molecules or a “water wire” that has been known to exist in biological systems, has been the primary object of computational study before the studies of PT in real and more complex biological systems.\(^7\) In the arrangement of a water wire, in addition to the proton hop between two water molecules, the inversion of the water wire should also take place in order for a second proton to be translocated in the same direction.

Figure 1.4 summarizes these dominant events involved in a single PT event along a polarized hydrogen-bonded water chain of five water molecules.

![Figure 1.4: Schematic depiction of the Grotthuss mechanism reprinted from Pomés & Roux 2002.\(^7\) An excess proton is incorporated at one end of the water chain followed by a succession of protonation-dissociation reactions between adjacent water molecules or “proton hop”, a proton is released at the other end of the chain and water molecules are inverted (“turn”) such that the dipoles of water molecules are reoriented to incorporate the next proton. The image is reproduced with permission from authors.](image)

The molecular basis for the hop-and-turn mechanism has been investigated using hydrogen-bonded
water chains embedded in model channels.\textsuperscript{128–132} These studies, in particular, highlighted the effects of thermal fluctuations of the water molecules, nuclear quantum fluctuations due to the light mass of the proton and model interactions between water and the protein, on PT along the water chain. Importantly, changes in oxygen-oxygen separation due to thermal fluctuations of the donor-acceptor modulate the energy profiles for the proton along the transfer coordinate. At the optimal oxygen-oxygen separation, there is no energetic barrier for proton hop. However, if the oxygen-oxygen separation is widened, an energy profile with a double-well appears indicating that proton sits preferentially close to either one of the water molecules.\textsuperscript{128} These observations led to the conclusion that “proton hop” is predominately governed by thermal fluctuations.\textsuperscript{7} Furthermore, the inclusion of nuclear quantum fluctuations alters the geometry of the water chain by dilating the lengths of the effective hydrogen bond length for PT.\textsuperscript{128} These effects, however, do not govern the “proton hop” event of the PT at least in equilibrium conditions.\textsuperscript{7}

It has also been shown that with increasing length of the proton wire, quasi concerted proton translocation is more apparent than the fully or near concerted proton translocation in which all hydrogen-bonding protons of the chain would hop in a single step.\textsuperscript{129,133} Such a single-step PT needs collective thermal fluctuations of the hydrogen-bond lengths over the entire chain at the same time.

The reorientation of a water molecule or the “turn” step creates a defect in the continuity of the hydrogen-bonded water wire and the overall reorientation process is described as the translocation of a “bonding defect”.\textsuperscript{134} It is demonstrated that the bonding defect in a single-file chain of nine water molecules in an inert cylindrical pore, mimicking an idealized “hydrophobic channel”, is essentially sequential.\textsuperscript{130} The activation free energy for the inversion of the total dipole moment water wire is 7 to 8 kcal/mol, suggesting dipole-dipole and hydrogen bonding interactions of the water molecules in the wire constitute the dominant forces that favor the polarized orientation of the chain and oppose the inversion of one stable polarized configuration to the other.\textsuperscript{130} Also, the rate limiting step of PT along a water wire is the bonding defect given it is a strongly activated process, compared to the activationless proton hop step.\textsuperscript{130}

### 1.3.3 Proton transfer in biological systems

In this section, I will describe two examples of channels have been widely used to understand the underlying physical principles of ion permeation. In particular, these examples are historically regarded as models of choice for the study of proton conduction in more complex proteins.\textsuperscript{7,135}
1.3.3.1 Gramicidin

Gramicidin is an antibiotic embedded in the cellular membranes of soil bacterial species. A 15-residue long polypeptide with alternating L- and D-amino acids adopts a right-handed $\beta^{6,3}$-helix structure that spans approximately one leaflet of a membrane bilayer with the N-terminus buried inside the bilayer (Fig. 1.5). With another monomer in the opposite leaflet, gramicidin forms a head-to-head dimer associated via hydrogen bonds. In this structural arrangement, hydrophobic side chains face out into the bilayer and the polar peptide backbone lines a cylindrical pore of 4 Å in diameter. The first crystal structure of gramicidin was reported in 1988\textsuperscript{136,137} and since then a large number of crystal structures of both uncomplexed and ion-complexed forms of gramicidin, as well as NMR structures became available.\textsuperscript{138}

![Fig. 1.5: Schematic depiction of the $\beta$-helix structure of the gA dimer reprinted from Pomés & Roux.](image)

Proton conductance: The Gramicidin channel is permeable to monovalent cations such as H$^+$, Cs$^+$, Na$^+$, K$^+$\textsuperscript{139} as well as ammonium.\textsuperscript{140} The diffusion coefficient for H$^+$ in gramicidin is about 20 times larger than for any other alkali metal cations\textsuperscript{140,141} and 8 times faster than the water molecules themselves.\textsuperscript{142} Furthermore, given the narrowness of the pore, the permeating water and alkali metal cations should move in a single file.\textsuperscript{143} These evidence suggests that unlike that of other cations, the diffusion of protons may take place according to a Grotthuss relay involving a hop-and-turn mechanism.

Early classical simulation studies revealed about 8 pore water molecules in a single file are arranged in a hydrogen-bonded network comprising water-water and water-channel hydrogen bonds,\textsuperscript{144} wherein a water molecule is most likely to hydrogen bond with two neighbouring water molecules and a channel backbone carbonyl oxygen, and the entire hydrogen-bonded chain is largely polarized. The inversion or the flipping of this polarized hydrogen-bonded chain of water molecules defines the “turn” step or “bonding defect” of the hop-and-turn mechanism. Structural comparison between gramicidin and hydrophobic channel models revealed that the presence of a polarized continuous water chain is not
caused by the polar pore of gramicidin since a longer and fully polarized chains are observed in the latter. Rather, it is an intrinsic property of the single-file water chain to optimize the dipole-dipole interactions between water molecules in the chain. Nonetheless, the presence of a local bonding defect of single water molecule, i.e. a flipping state intermediate of the two polarized states stabilized by two hydrogen bonds to two backbone carbonyl O atoms of channel, appears more frequently near the end of the chain in gramicidin.

Subsequent classical simulations with longer time scales not only demonstrated the dynamic fluctuations of the single-file water molecules in gramicidin, but also quantified the energetic basis for the propagation of the bonding defect along the chain. Although the initiation of a bonding defect near the ends of the chain is spontaneous, the progression of the bonding defect along the water wire is energy-activated, which requires 3-4 kcal/mol depending on the water model used. This is about half of the energy requirement for the propagation of the bonding defect in a water wire located in a idealized hydrophobic channel. This is because, in gramicidin, the partly-oriented flipping state is stabilized by the water-channel hydrogen bonding as observed near the end of the chain. End-to-end propagation of a bonding defect in the water wire is observed in the range of 100 ps or longer time scales.

Simulations with dissociable water models demonstrated the rapid exchange of H⁺ or hop along the water chain involving transitions between successive hydronium-like and Zundel-like arrangements of the water molecules, which occurs in picosecond or sub picosecond time scales. The propagation of the “hop” event or “ionic defect” is mediated by thermal fluctuations in the relative positions of proton-relaying water O atoms and is energetically barrierless similar to that was observed for water-wires in nonpolar channels. This also means the carbonyl O atoms arranged periodically lining the channel pore are not binding sites for protons. Collectively, the mechanism for the efficient transport of H⁺ in gramicidin is achieved by: (1) the disposition of water molecules in a single file, forming strong hydrogen bonds between them in order to support the rapid relay of H⁺ (2) the relatively slow reorganization of hydrogen bonds between single-file water molecules and channel backbone carbonyl groups reorienting proton-relaying water molecules.

Ion selectivity: Gramicidin is virtually impermeable to anions. The pore lining atoms do not carry a net charge, hence the mechanism for charge discrimination is not obvious from the structure alone. To elucidate the charge specificity of gramicidin, the free energy difference of K⁺ and Cl⁻ for the interior of the channel has been calculated using MD free energy simulations. Although K⁺ and Cl⁻ have solvation free energies on the order of -80 kcal/mol in bulk water, solvating K⁺ inside gramicidin is favoured by 60 kcal/mol over Cl⁻. The reason for this disparity was demonstrated by comparing the average radial distribution around the ions in bulk water and inside the channel. In bulk water, the K⁺
ion is surrounded by approximately 7.4 water molecules in the first solvation shell with water oxygen atoms pointing toward the ion, thus, providing a favorable electrostatic interaction energy. In the channel, oxygen atoms of the backbone carbonyl groups primarily solvate K\(^+\), giving rise to an average charge radial distribution that is remarkably similar to that in bulk water.

In the case of Cl\(^-\), water molecules form hydrogen bonds with the anion, providing a favorable electrostatic interaction energy. However, in the channel, similar hydrogen bonds with the NH amide backbone group are not easily formed. Thus, the asymmetry in the permanent charge distribution of the peptide backbone results in a favorable interaction energy for K\(^+\) and not for Cl\(^-\), leading to ion selectivity of cations over anions in gramicidin.

Channel gating: Gramicidin channel gating is known to occur by the association and dissociation of the two monomers. Gramicidin analogues wherein two monomers linked at their N-termini with a dioxolane ring provided one of the first applications of molecular dynamics methods to study the gating of an ion channel. These channels, like native gA, are selective for monovalent cations, but are long-lived channels. More specifically, the stereospecific isomers of the dioxolane-linked GA, the SS and RR forms, have strikingly different properties, such that interruptions or blocking of a K\(^+\) current (“flickers”) are more frequent in RR than in SS.

The channel blocking by the dioxolane ring was investigated using MD simulations and free energy profile along several probable reaction paths, describing the transition of the ring from the exterior to the interior of the channel. In particular, unlike in SS-linked GA channels, the gating transition of the RR-linked dioxolane ring is facilitated by the trans \(-\) to cis isomerization of peptide bond dihedral angle of Val\(^1\). The calculated blocking rate and average blocking time associated with this transition of RR-linked are in good accord with experimental observations. More importantly, the presence of a K\(^+\) ion inside the channel was necessary to reproduce the experimentally observed lifetime of the blocked channel. This observation reaffirms the importance of considering permeating ions when deducing gating kinetics and mechanisms of ion channels as inferred previously for ligand-gated and voltage-gated channels.

1.3.3.2 Aquaporins

Aquaporins (AQPs) are a family of membrane proteins in which all members act as water channels. In response to the osmotic imbalance across the cell membrane, AQPs transport billions of water molecules per second. Some subfamilies also transport uncharged solutes, such as glycerol, CO\(_2\), ammonia and urea.

At least fourteen different structures of AQP from a diverse range of organism have been determined primarily using electron microscopy and X-ray diffraction experiments. AQPs are homo-tetramers
with each monomer acting as an independent water channel. Each monomer defines an hour-glass pore formed by a bundle of six α-helices and two half-membrane spanning loops with their N-termini pointing head-to-head at the center of the channel, thus forming two macrodipoles inside the pore (Fig. 1.6). At the end of these two half helices, facing each other, are two conserved Asn-Pro-Ala (NPA) motifs. A cluster of aromatic amino acids and an arginine residue, ar/R, closer to the extracellular exit of the channel forms the narrowest part of the channel, the selectivity filter.

Molecular dynamics studies of water movement revealed that up to nine water molecules reside in the narrow part of the pore, which is 20 Å long. Confined by the ar/R selectivity filter, these water molecules assemble in a single file and exhibit a bipolar organization such that hydrogen atoms of roughly the first half of the single file (farthest from the cytoplasm) are oriented away from the cytoplasm while the second half of the single file is oriented in the opposite direction. The central water molecule located near the NPA motif is oriented perpendicularly to the channel axis, donating both hydrogen atoms to the two neighbouring water molecules that are polarized in opposite directions.\textsuperscript{155} Thus, this bipolar organization of water molecules are attributed to the opposing dipoles of the two half-helices.\textsuperscript{156,157} It was observed in the simulations that as the water molecules permeate through the pore, they rotate by 180° near the NPT motif maintaining the bipolar orientation. As a result, the single water file is prevented from forming a continuous hydrogen bond network inside the pore. It was also demonstrated that neutralizing the helical macrodipoles (turning off the partial charges) leads to a formation of a unidirectional water wire, which could be a potential proton wire.\textsuperscript{157} This led to the speculation that it is the bipolar water orientation that prevents the channel from proton permeating. Subsequently, a series of second generation simulations explicitly addressed the energetics and dynamics of an excess protons in the AQP channel, furthering the knowledge on the balance of physical forces governing structural diffusion of proton in biological systems.\textsuperscript{153} The current view of the molecular origin of proton blockage in AQP is as follows.
Molecular basis of proton blockage in AQP: The computational study of proton blockage in AQPs primarily focused on the effect of three main structural elements; the two macrodipoles, SF and NPA motif.

Almost all the studies computing the free energy profile of proton translocation reported a primary barrier at the NPA motif. The electrostatic profile (for the translocation of a positive point charge) along the pathway indicates a strong agreement with the above free energy profile, suggesting that charge distribution of the channel is the dominant factor opposing the proton movement. This is the single most agreed-upon conclusion of the mechanism of proton blockage in AQP, although the origin of the electrostatic barrier seems not yet resolved.

Turning off the partial charges of Asn of the NPA site only marginally lowers the static field barrier in addition to causing the bipolar arrangement to disappear, whereas neutralizing the macro dipoles significantly lowers the static barrier for the cation movement, while favoring any of the three polarization states (bipolar orientation and 2 fully polarized with water dipoles pointing either to or away from the cytoplasm) of the unprotonated water chain. Putting these results together suggests that (1) the bipolar arrangement is nucleated by the NPA site, which is stabilized by the macrodipoles (2) the proton at the NPA site is stabilized by the bipolar organization of the water chain, but is opposed by the repulsive
forces between the cation and the macro dipoles. The latter point is one of the two explanations for the origin of the electrostatic barrier. \cite{8,158}

The other explanation holds desolvation effects as the origin of the electrostatic barrier. It is suggested that the combined effects of low dielectric medium and few neighbouring water molecules partially stabilize the hydronium, causing a large energetic cost for moving the hydronium from bulk water into the channel. \cite{159,161}

In addition, neutralizing Arg lowers a secondary static barrier for the cation movement in the SF site, suggesting that the proton entry from the bulk water into the narrow pore is opposed by a repulsive electrostatic force between the cation and Arg in the SF. \cite{8,162}

1.4 Thesis objectives and organization

1.4.1 Thesis objectives

The primary objective of my research thesis is to determine the molecular basis for ion and charge selectivity of Hv1 using MD simulations as the main technique. To this end, as first step, constructing and validating a model for Hv1 is a prerequisite, which is pursued in this thesis. The central hypothesis for model construction is that Hv1 is homologous to VSDs of voltage-gated potassium and sodium channels and that structural characters that are common for VSDs are reliable indicators to determine the validity of a homology model for Hv1. In order to build and assess the structural quality of Hv1, I have (1) employed multiple-structure based homology building protocol to construct models for Hv1 and (2) designed a novel MD protocol to assess the model quality in comparison to the VSD templates.

Having obtained a valid homology model for Hv1, I then carried out systematic comparative studies of wild type and mutant variants of Hv1 to probe the molecular basis for their charge selectivity on the basis of structural (e.g. side chain fluctuations and pore hydration) characterization and effects of static interactions arising from the charge distribution of the residues lining the pore.

The next stage of simulations included free energy simulations for the movement of Na\(^+\) and Cl\(^-\) in wild type and mutant variants of Hv1 to further the understanding of the molecular and energetic basis of ion and charge selectivity.

1.4.2 Thesis organization

In chapter 2, I highlight the underlying theories of MD and quantum simulations, Statistical mechanics to derive thermodynamic properties from simulations, a sampling technique of MD simulations namely,
umbrella sampling and homology modelling.

In chapter 3, I describe a novel MD protocol to determine the solvation and orientation of proteins in lipid membranes. In addition, I demonstrate that this approach is a computationally inexpensive way to study the self-association of amphipathic helices and therefore is a promising protocol to study the self-association of helix bundles of membrane proteins.

In chapter 4, I describe the construction of homology models for Hv1 using experimentally resolved VSDs of voltage-gated potassium and sodium channels as templates, all prepared in a membrane-mimetic. Next, I present the development of a novel MD protocol to assess the model quality in comparison to the VSD templates. The chapter ends with a preliminary analysis probing the molecular basis of charge selectivity and proton translocation of the wild type Hv1.

In chapter 5, I investigate the molecular basis for charge selectivity of the wild-type Hv1 and four of its mutant variants prepared in an explicit lipid bilayer. Here, I perform structural characterization by conducting analysis on the side chain fluctuations, pore hydration and electrostatic properties of the pore between different systems.

In chapter 6, I carry out free energy simulations to determine the molecular basis for ion and charge selectivity with the presence of explicit cations and anions using biased simulations. I characterize the key molecular determinants modulating the movement of Na$^+$ and Cl$^-$ in wild-type and two anion-selective mutants of Hv1.
Chapter 2

Theory and methods

Abstract

This chapter first provides the theoretical framework of the computational methodologies to study biomolecular systems and the justification for the choice of molecular dynamics (MD) simulation as the main technique of study in this thesis. Next, key concepts of statistical mechanics and its use to derive a pertinent thermodynamic property for the study of ion permeation in channels, relative free energies, are described. Then, the underlying formalism of a sampling technique named umbrella sampling used in free energy simulations is discussed. Finally, since all MD simulations presented in this thesis are based on a homology model of the human Hv1 or hHv1, the final section of this chapter provides a brief theoretical background of homology modelling.
2.1 Introduction

Molecular systems can be investigated using quantum and classical simulation techniques. The next section of this chapter provides a summary of the underlying principles of molecular simulations, with an emphasis on the classical molecular mechanics and MD simulations. The main output of MD simulations consist of positions and velocities of all atoms comprising the system. Statistical mechanics provides a formal connection between these microscopic details and macroscopic properties such as pressure, temperature, etc. The most important macroscopic property to explain the energetic basis of a biological process like ion permeation through a channel is the free energy difference between two macrostates. The second part of this chapter provides a brief overview of statistical mechanics and an explanation of how free energy differences can be computed using the concepts of statistical mechanics on microscopic details of MD simulations. That section ends with an overview of umbrella sampling, a sampling technique used to improve the sampling of free energy simulations.

A simulation begins with an initial configuration of the system, which is comprised of a protein and its solvent for protein simulations. For folded proteins, in particular for membrane proteins, their starting configurations are often provided by experimentally-resolved structures. Since the discovery of the gene of hHV1 in 2006, the experimental determination of its structure has been pursued. The atomic structures of hHV1 resolved so far, however, have been limited in use as a starting configuration for MD simulations. Crystallography resulted in the determination of the cytosolic C-terminal fragment of the hHV1, which is a coiled-coil fold. In 2014, Okumara and Nakagawa groups resolved a new crystal structure of a full-length H$
u$1. That structure, however, is not representative of the native and “open” conformation of hHV1, since it is a chimera of three proteins from three species that is presumed to be in the closed (non-conducting) state. Alternatively, the construction of a computational model of hHV1 in the open (conducting) state has been pursued by our group and others following a homology (comparative) modelling procedure. The third and last section of this chapter provides an overview of the general homology modelling procedure followed by the illustration of the use of MODELLER, a computer program for comparative protein structure modelling that we used to construct a homology model for hHV1.

2.2 Quantum and Classical Mechanics

Mechanics is the branch of Physics that deals with the study of motion under the action of forces. The computational methodologies to study molecular systems belong to the two major sub-fields of mechanics;
quantum and classical mechanics. Both approaches are important to obtain a more accurate atomic view of biological systems.

2.2.1 Quantum mechanics

Quantum mechanics (QM) is required to account for the quantum nature of electrons and H\(^+\) due to their light mass. As such, QM methods embody the fundamental electronic and nuclear structures of atoms and molecules, which is necessary for the explicit treatment of chemical reactions such as proton transfer.

More specifically, QM delves into the behaviour of electrons under the influence of the electromagnetic force exerted by nuclear charges. Many methods of quantum mechanics have been developed. At the core of almost all these methods lies the Schrödinger equation, which describes how the electrons in a molecule behave.\(^{169,170}\) The Schrödinger equation is expressed in terms of a wavefunction, a mathematical expression to calculate the electron distribution. However, the Schrödinger equation cannot be solved exactly for any molecule with more than one electron. For this reason, approximations are required for systems with more than one electron. Quantum mechanical methods vary in the types of approximations and whether or not empirical parameters are used in their equations.

2.2.1.1 Ab initio methods

All ab initio methods rely on quantum mechanical calculations rigorously based on first principles without any inclusion of empirical data. Various ab initio methods (e.g. Hartree-Fock, post-Hartree-Fock) methods determine the wave function by incorporating the molecular orbital theory which approximates the positions of bonded electrons or molecular orbitals as a linear combination of atomic orbitals.\(^{171}\). Simulations based on ab initio methods are applicable to a system containing up to 100 atoms and last up to tens of picoseconds.

2.2.1.2 Semiempirical methods

Semiempirical methods (e.g. AM1, PM3) neglect computationally expensive calculations of certain integrals inherent to ab initio methods, instead use parameters derived from experimental data on certain properties of organic molecules such as heats of formation, dipole moments, etc.\(^{172}\). Therefore, semiempirical methods are less accurate than ab initio methods, but enable two or three orders of magnitude longer simulations.
2.2.1.3 Density functional theory (DFT) methods

These methods are based on approximating solutions of the Schrödinger equation, but not on the basis of the wave function. Instead, the electron probability density function serves as the basic descriptor of the electronic system. Central to DFT methods are the use of functionals (a rule that transforms a function into a number) and most common functionals use parameters derived from empirical data. Compared to *ab initio* methods, DFT methods are computationally more efficient.

2.2.1.4 QM/MM methods

The hybrid between quantum mechanics and molecular mechanics (QM/MM) rely on dividing the bimolecular system into regions: a reactive part which is treated by quantum mechanics and the remaining part which is treated by classical force fields (described in section 2.2.2). This approach requires a sophisticated treatment of the boundary between the QM and MM regions.

Collectively, the above QM methods provide, in principle, an accurate representation of a given reaction involving bond breakage/formation, but incur high computational cost. For a system of considerable size and complexity, obtaining long time scales and calculating statistically meaningful free energy profiles are not feasible. Furthermore, QM methods are highly sensitive to details of the conformations and the environment surrounding the reactants. Thus, having an accurate starting conformation for QM simulations is crucial. Therefore, obtaining a system that is well relaxed wherein slow fluctuations giving rise to orientational preferences of side chains and/or internal water molecules are sampled, is an important first step before embarking on QM simulations. To this end, non-reactive MD methods provide the benefit of computational efficiency that allows the system evolve over long time scales sampling conformations that are otherwise not accessible by QM simulations. In other words, classical MD simulations are almost a prerequisite before embarking on QM simulations if the 3D structure of a complex molecular system, such as a protein, is not known at high resolution. In the case of *hHv*1, conducting MD simulations is all the more appropriate due to the lack of an experimentally-resolved structure for the open-state channel. The following sections highlight the main theoretical background of MD simulations used for that purpose in the present thesis.
2.2.2 Molecular mechanics

The classical representation of a molecular system known as “molecular mechanics (MM)” treats the atoms as balls and the bonds as springs.\textsuperscript{178} In MM calculations, an empirical potential energy (PE) function describes how atoms interact with each other based on their relative position. The Born-Oppenheimer approximation allows the separation of nuclear and electron motions that happen in different time scales and therefore allows treating the energy of a system of atoms as a function of nuclear positions only. Because the distributions of electrons are not treated, bond formation and breakage, i.e. chemical reactions, cannot be studied with MM force fields.

Specifically, a MM function is a mathematical equation to compute the potential energy, $U$, of a system containing atoms and includes terms describing both covalent or bonded interactions: bond stretching, bond angle bending and bond twisting (torsions); and non-covalent (or non-bonded) interactions: van der Waals and electrostatic interactions (equation 2.1).

$$U = U(\text{bonds}) + U(\text{angles}) + U(\text{dihedrals}) + U(\text{vanderWaals}) + U(\text{electrostatic}) \quad (2.1)$$

Covalent bond stretching and bond angle bending terms are represented with a harmonic potential (equations 2.2, 2.3):

$$U(\text{bonds}) = \sum_{\text{bonds}} k_b (b - b_0)^2, \quad (2.2)$$

$$U(\text{angles}) = \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2. \quad (2.3)$$

In equations 2.2 and 2.3, $k_b$ and $k_\theta$ are force constants associated with the bond and angle terms, respectively, and the constants $b_0$ and $\theta_0$ are the reference bond length (for example, the value of the bond length in a harmonic oscillator in vacuum) and reference angle, respectively. $b$ and $\theta$ are variables representing the bond length and bond angles. The torsion energy is modeled by a simple periodic function as follows:

$$U(\text{dihedrals}) = \sum_{\text{dihedrals}} \sum_{n} k_\chi (1 + \cos(n \chi - \delta)). \quad (2.4)$$

The periodicity, $n$ indicates the number of cycles per 360° rotation about the dihedral. The phase, $\delta$, is the reference angle at which the dihedral energy is at its minimum.

The non-covalent interactions may be considered the most important for computational studies of macromolecules due to the large number of such interactions and strong influence of these interactions on the properties of the macromolecule.\textsuperscript{179} These interactions are calculated for pairs of atoms separated...
by 3 or more bonds and between atoms in different molecules. The van der Waals interaction between a pair of atoms \((i,j)\) is represented by the Lennard-Jones (LJ) potential as follows (equation 2.5):

\[
U_{vdw}(r_{ij}) = \epsilon_{ij} \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{6} \right].
\] (2.5)

The \(1/r^{12}\) term represents the repulsion between atoms associated with overlap of the electron clouds of the pair of atoms. The \(1/r^{6}\) term represents the London’s dispersion or instantaneous dipole-induced dipole interaction between the pair of atoms. The parameter \(\epsilon_{ij}\) indicates the depth of the LJ potential between \(i, j\) and \(R_{\text{min},ij}\) is the distance between atoms \(i\) and \(j\) at which \(\epsilon_{ij}\) occurs. \(U_{vdw}(r_{ij})\) is a function of the distance between the pair of atoms, \(r_{ij}\) and it decays very rapidly as \(r_{ij}\) increases.

The functional form of the PE, together with the associated parameters of each term of the PE function define a force field. These parameters are derived from a combination of experimental data, such as X-ray crystallography, NMR, vibrational spectroscopy and quantum mechanical calculations on tractable molecular fragments. The force field parameters are occasionally checked and improved as new and refined empirical calculations become available.

2.3 Molecular dynamics simulations

2.3.1 From potential energy to atomic motions: equations of motion

In a system composed of \(N\) particles interacting according to the rules of the force field, the force acting on each particle, \(i\), is estimated by the gradient of the potential energy with respect to atomic displacement (equation 2.6).

\[
\vec{F}_i = -\nabla_i U
\] (2.6)

When the forces have been assigned, Newton’s law of motion, \(\vec{F} = m\vec{a}\) (equation 2.7), can be used to numerically solve the motion of the particle \(i\).

\[
\vec{F}_i = m_i \vec{a}_i = m_i \frac{d\vec{v}_i}{dt} = m_i \frac{d^2\vec{r}_i}{dt^2}
\] (2.7)

In the above equation, \(m, a_i, v_i\) represent mass, acceleration, velocity, respectively, of the \(i\)th particle at time, \(t\). The variable \(\vec{r}_i\) represents the position vector of particle \(i\) in a cartesian coordinate system. The following relation (equation 2.8) gives the value of \(\vec{r}\) at time \(t + \delta t\) as a function of \(\vec{a}\), initial position, \(\vec{r}_i\) and the initial velocity, \(\vec{v}_i\) at time \(t\). Here, \(\delta t\) refers to the time step of the simulation, which typically
falls within the range of 1-4 femtoseconds (fs) in order to ensure that the change in forces over one time step is small.

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \vec{v}_i(t)\delta t + \frac{\vec{a}_i(t)\delta t^2}{2}$$  \hspace{1cm} (2.8)

Once the new position of each particle, $\vec{r}_i(t + \delta t)$, is computed, the interatomic forces are updated in order to calculate the subsequent positions after another time step. This entire process is repeated to obtain a trajectory of coordinates of the system for a finite time period. A number of numerical algorithms have been developed for integrating the equations of motion. In this thesis, the Leap-frog algorithm is employed.

2.3.2 Temperature and pressure control

In classical statistical mechanics, temperature is a direct measure of the average molecular kinetic energy and the definition of the temperature in a (classical) many-body system makes use of the equipartition theorem, which conceptually means that at thermal equilibrium, each independent degree of freedom has an equal amount of kinetic energy; for example, the average kinetic energy per degree of freedom in the translational motion of a molecule is equal to its rotational motions. Thus, according to the equipartition theorem, the average kinetic energy, $\langle K \rangle$, per degree of freedom is related to the thermodynamic temperature, $T$, in the following way (equation 2.9),

$$\langle K \rangle = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 = \frac{1}{2} k_B T, \hspace{1cm} (2.9)$$

where $m_i$, $v_i$, and $k_B$ represent the mass, velocity of the particle, and Boltzmann constant, respectively. In practice, the measure of the total kinetic energy normalized by the number of unrestrained degrees of freedom, $N_f$, is used to compute the instantaneous temperature, $T_i$ (equation 2.10):

$$T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{k_B N_f}. \hspace{1cm} (2.10)$$

Here, $v_i(t)$ refers to the velocity of particle, $i$ at time (t). In order to conduct simulations under constant temperature or constant pressure, various temperature and pressure coupling algorithms are employed. For example, a temperature coupling algorithm may involve velocity rescaling at each time step or stochastic impulsive forces that act occasionally on randomly selected particles. All approaches of pressure coupling algorithm involve the scaling of the volume of the simulation box. In this thesis, Nosé-Hoover algorithm and the Parrinello-Rahman algorithm were employed for temperature- and pressure-
coupling, respectively.

2.3.3 Periodic boundary conditions

Biomolecular simulations usually involve bulk systems, such as a solid crystal or protein in a solution. In MD simulations, a system containing the desired molecules is prepared in a finite simulation box due to limited computer memory and also to speed up all the calculations. By translating copies of the simulation box in 3D space, periodic boundary conditions allow the simulation of a bulk system and eliminate edge effects in a finite simulation box.

2.4 Statistical mechanics and free energy simulations

Experimental approaches often measure changes in macroscopic or thermodynamic properties like pressure, free energy, temperature, heat capacities, etc, that help understand the behaviour of thermodynamic systems. Such a system can be a reaction vessel, an electrochemical cell, a biological cell, and so on. In MD simulations, a system is an abstract model of a thermodynamic system. In contrast to macroscopic properties measured by experiments, MD simulations provide microscopic properties of the system by generating motion of individual particles that comprise the system. Statistical mechanics provides the link between the microscopic properties of a system and its macroscopic properties. MD simulations of a system involving biomolecules, thus, provide time evolution of all its molecules, from which the macroscopic properties of the system can be derived using statistical mechanics.

This section starts with a brief overview of main topics of statistical mechanics essential to compute a macroscopic property— the free energy, a measure of the available energy to do work, $W$. For example, $\text{hH}_V\text{1}$ conducts protons and not $\text{Na}^+$. From a thermodynamic standpoint, this phenomenon means that the free energy of $\text{H}^+$ in $\text{hH}_V\text{1}$ is low enough to promote proton conduction and not $\text{Na}^+$. A biological process like ion permeation involves an evolution of successive macrostates (defined in 2.4.2) of the system. The free energy difference between these macrostates is sufficient to explain the thermodynamic basis of a given biological process. If needed, as explained in section 2.4.5, special sampling techniques called free energy simulations are used to perform free energy calculations. One such sampling technique is umbrella sampling (US), which is described later in the following section together with the formalism used to compute free energy profiles in US calculations.
2.4.1 Ensembles

The term “ensemble” may refer to one of two meanings. First, it could be the entire collection of all the possible atomic arrangements of the system. Second, it could be specific conditions defining the thermodynamic state of the system. For example, the canonical ensemble is defined by constant “N,V,T”, meaning that the system has a constant number of particles, volume, and temperature.

2.4.2 Boltzmann Distribution Law

Central to statistical mechanics is the estimation of the probability distributions of the internal energy, which is the sum of configurational (potential) and motional (kinetic) energies of the system. Experiments measure averaged properties over those distributions. For example, the pressure of a gas, which is an experimental readout, is a result of the average force exerted by its molecules.

In MD simulations, the time evolution of a system can be represented by a multidimensional space of micro parameters: the position \( \vec{r}_i \) and velocity. Each point in this space represents a microstate. During its time evolution, the system only samples the accessible microstates – the ones that do not violate the conservation laws. A collection of microstates gives rise to a macrostate of a system that can be specified by its macroscopic parameters.

According to statistical mechanics, the Boltzmann distribution law gives the probability \( p_i \) of a given micro state \( i \) with internal energy \( \epsilon_i \) in equilibrium at constant particle number \( N \), volume \( V \), and temperature \( T \) (the canonical ensemble), which is proportional to \( e^{-\beta \epsilon_i} \), the Boltzmann factor:

\[
p_i = \frac{e^{-\beta \epsilon_i}}{\sum_i e^{-\beta \epsilon_i}}, \quad (2.11)
\]

where \( \beta = 1/k_B T \), \( k_B \) being Boltzmann’s constant and \( T \) being the absolute temperature of the system. The quantity in the denominator is called the partition function \( q \), which is a normalization factor for probabilities. The partition function adds up all of the Boltzmann factors of a system such that that \( \sum_i p_i = 1 \). The absolute \( p_i \) can be calculated if \( q \) is known. Using the above expression, the average internal energy \( \langle \epsilon \rangle \) can be expressed as follows,

\[
\langle \epsilon \rangle = \frac{1}{q} \sum_i \epsilon_i e^{-\beta \epsilon_i} \quad (2.12)
\]
2.4.3 Entropy

The entropy, $S$ is a macroscopic variable and is a measure of disorder of the system.\textsuperscript{180} From a statistical mechanics standpoint, the entropy is defined as follows:

$$S = -k_B \sum_i p_i \ln p_i \quad (2.13)$$

where $p_i$ is the probability that the system is in microstate $i$. The following Boltzmann’s entropy\textsuperscript{181} is the expression of entropy at thermodynamic equilibrium when all the accessible microstates of the system are equally likely. This occurs in a micro-canonical ensemble (constant number of particles, volume and internal energy) where all microstates have the same internal energy.

$$S = k_B \ln \Omega \quad (2.14)$$

Here, $\Omega$ is the number of microstates. According to this relation, a macrostate with more microstates has a higher entropy than a macrostate with fewer microstates.

2.4.4 Free energy

The free energy is a macroscopic equilibrium property of a system. The relative free energy $\Delta W$ of any two macrostates, 0 and 1, is related to their respective probabilities $p_0$ and $p_1$ via

$$\Delta W = -k_B T \ln \frac{p_0}{p_1} \quad (2.15)$$

The free energy is denoted as Helmholtz free energy, $A$, in the $NVT$ canonical ensemble, or Gibbs free energy, $G$ in the $NPT$ ensemble, where $N$, $V$, $T$ and $P$ represent constant number of particles, volume, temperature and pressure, respectively. Further, free energy unites the internal energy, $U$ and entropy in the following ways for the Helmholtz free energy and the Gibbs free energy, respectively:

$$A = U - TS \quad (2.16)$$

$$G = U + PV - TS \quad (2.17)$$

While in the gas phase and small molecular systems free energy changes can be computed by means of analytical expressions, for complex systems in the condensed phase, the potential of mean force (PMF)
for a given transformation or reaction is computed along a reaction pathway.

2.4.5 Potential of mean force

The potential of mean force (PMF) provides a quantitative description of free energy change for a biological process along a reaction pathway. The PMF is so named because its derivative with respect to the reaction coordinate gives the average force along a reaction coordinate (described next) at equilibrium.

The reaction coordinate, $\xi$, of the system can be an angle, a distance or a more complicated function of the cartesian coordinates of the system or conformational space, $R$. Therefore, $\xi$ is a function and a subset of $R$. The relative probability of finding the system in different macrostates as a function of $\xi$ can be used to compute the relative free energy of those macrostates.

First, let us consider a macrostate $A$ sampled during a simulation, and the probability of finding the system in a microstate of $A$ with energy $\epsilon_A$ as expressed by the Boltzmann distribution is,

$$\rho_A = \frac{e^{-\beta \epsilon_A}}{q}$$  \hspace{1cm} (2.18)

Solving for $-\beta \epsilon_A$ gives the following rearrangement,

$$-\beta \epsilon_A = \ln q + \ln \rho_A$$  \hspace{1cm} (2.19)

Similarly, the Boltzmann distribution and the rearrangement for a reference microstate $*$ would be,

$$\rho^* = \frac{e^{-\beta \epsilon^*}}{q}$$  \hspace{1cm} (2.20)

$$-\beta \epsilon^* = \ln q + \ln \rho^*$$  \hspace{1cm} (2.21)

Solving for $\epsilon_A$ by $-\beta \epsilon_A - (-\beta \epsilon^*)$ gives,

$$\epsilon_A = \epsilon^* - 1/\beta \ln \left[ \frac{\rho_A}{\rho^*} \right]$$  \hspace{1cm} (2.22)

The above expression is particularly useful to compute the relative probabilities and energy difference between two microstates, when computing $q$ is not feasible, which is generally the case. The expression for the PMF, $\mathcal{W}(\xi)$, along some coordinate $\xi$ takes the same form as equation 2.22 but with average probability distribution values where the angle brackets indicate averaging over all microstates at $\xi$ as.
follows,

\[ W(\xi) = W(\xi^*) - k_B T \ln \left( \frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right) \]  

(2.23)

The \( W \) notations replace \( \epsilon \) to more accurately symbolize the average energies. The average probability distribution value \( \langle \rho(\xi) \rangle \) is given by,\(^{183}\)

\[ \langle \rho(\xi) \rangle = \frac{\int dR \delta(\xi'[R] - \xi) e^{-\beta U(R)}}{\int dR e^{-\beta U(R)}} \]  

(2.24)

where \( U(R) \) represents the potential energy of the system as a function of the conformational space, \( R \). Thus, if \( \xi \) is a function (describing an angle or distance), only a subset of its values, \( \xi' \) becomes the user defined reaction coordinate. Therefore, in order to select only probabilities of \( \xi' \), Dirac delta function, \( \delta(\xi'[R] - \xi) \) is used. The above equation is non zero only when \( \xi' = \xi \).

However, computing direct conformational-space integrals and \( \langle \rho(\xi) \rangle \) straight from brute-force MD simulations is not always feasible.\(^{183,184}\) For example, within the available computer time, the configuration space around energy wells are typically well sampled, whereas regions separated from these energy wells by potential energy barriers significantly larger than \( k_B T \) are poorly sampled. Therefore, events that require transitions between such energy wells become rare events. Their frequency of occurrence or rates are extremely small. For example, a protein may spend much of its time in unfolded or misfolded conformations until the rare structural fluctuation advances the folding process into the native folded state. Similarly, such rare events occur in channels. For example channel hydration in the magnesium channel, CorA,\(^{185}\) and entry/exit of ions through the intracellular gate in the potassium channel, KcsA,\(^{186}\) are considered rare events whose times scales may exceed that of the simulation time, and thus are not captured by brute-force simulations.

Different techniques have been developed to sample rare events within the available computer time. One approach is the “Umbrella sampling (US) technique” developed by Torrie and Valleau.\(^{187,188}\)

### 2.4.6 Umbrella Sampling

In the US approach, the reaction coordinate is divided into “windows” each of which is a separate equilibrium simulation that samples the configuration space, \( d\xi \) around a reference value of \( \xi \) (windows shown in Fig. 2.1). In each window, the sampling is confined to a \( d\xi \) by a harmonic bias potential, which encourages barrier crossing and which also modifies the potential energy of the system. In post-processing at the end of simulations of all windows, the populations of sequential windows along the \( \xi \) are unbiased and combined to give the global free energy estimate, \( W(\xi) \) using the weighted histogram
Here, the formalism for conducting biased simulations and for recovering unbiased free energy differences from biased simulations is presented.

The bias potential \( w_i \) of a window \( i \) is an additional potential energy term which depends only on the reaction coordinate \( \xi \) and modifies the unbiased potential energy \( U \) as follows:

\[
W^b(R) = U^u(R) + w(\xi).
\] (2.25)

The superscript ‘b’ denotes biased quantities, whereas the superscript ‘u’ denotes unbiased quantities. The average probability distribution for an unbiased window can be presented using the following equation introduced in equation 2.24:

\[
\langle \rho_i(\xi) \rangle^u = \frac{\int dR \, \delta(\xi' [R] - \xi) e^{-\beta U(R)}}{\int dR \, e^{-\beta U(R)}}
\] (2.26)

Likewise, the expression for the average probability distribution for a biased window can be presented as follows:

\[
\langle \rho_i(\xi) \rangle^b = \frac{\int dR \, \delta(\xi' [R] - \xi) e^{-\beta [U(R) + w_i(\xi'(R))]} \, e^{-\beta \xi'(R)}}{\int dR \, e^{-\beta [U(R) + w_i(\xi'(R))]} e^{-\beta \xi'(R)}}
\] (2.27)
Rearranging $\langle \rho_i(\xi) \rangle^b$ to be expressed using $\langle \rho_i(\xi) \rangle^u$ results in:

$$\langle \rho_i(\xi) \rangle^b = (\langle \rho_i(\xi) \rangle^u e^{-\beta w_i(\xi)} \bigg) e^{-1}$$

(2.28)

A step-by-step formalism of the above expression can be found elsewhere. The expression for the unbiased PMF, $W(\xi)^u$ for window $i$ can be obtained using $\langle \rho_i(\xi) \rangle^b$ as follows:

$$W_i(\xi)^u = W(\xi^*)^u - k_B T \ln \left[ \frac{\langle \rho_i(\xi) \rangle^b}{\langle \rho(\xi^*) \rangle^u} \right] - w_i(\xi) + F_i,$$  

(2.29)

where $W(\xi^*)^u$ and $\langle \rho(\xi^*) \rangle^u$ represents the unbiased PMF and average unbiased probability distribution of the reference point $\xi^*$ respectively. The notation, $F_i$ is an undetermined constant that represents the free energy associated with introducing the window potential, which takes the following form:

$$e^{-\beta F_i(\xi)} = \left( e^{-\beta w_i(\xi)} \right).$$

(2.30)

$F_i$ connects the free energy curves obtained from separate windows. Now, the $\langle \rho(\xi) \rangle^u$ can be re-written using $F_i$:

$$\langle \rho_i(\xi) \rangle^u = \langle \rho_i(\xi) \rangle^b e^{-\beta w_i(\xi)} e^{-\beta F_i(\xi)}.$$  

(2.31)

The weighted histogram analysis method (WHAM) provides a scheme for obtaining a single unbiased property distribution from $F_i(\xi)$ values for each window and “stitching” a series of separate unbiased probability distributions, $\langle \rho_i(\xi) \rangle^u$.

2.4.7 Weighted Histogram Analysis Method (WHAM)

The main purpose of WHAM is to determine optimal $F$ values in equation (2.29) for combining unbiased probability distributions. How this is achieved essentially by solving two equations (2.33 and 2.34) iteratively is described below.

For convenience, the meaning of notations used in mathematical expressions are listed below that make reference to arbitrary histograms along $\xi$ in Fig. 2.2.
Figure 2.2: **Biased probability distributions of representative windows along** $\xi$. Only three normalized histograms of biased probability distributions of overlapping windows are shown in red, green and blue. The histogram in black is the biased probability distribution of the last window. For example, the biased probability distribution of the $i$ window (red) is $\rho_i(\xi)^b$. Subscript $j$ means $j^{th}$ bin in $i^{th}$ simulation. The biased probability of finding the system at $\xi = \xi_j$ in window $i$, $\rho_i(\xi_j)^b$ is coloured in orange.

- $\xi$: Reaction coordinate
- $i$: A simulation window
- $N_w$: Number of windows used along the entire $\xi$
- $\xi_i$: Reaction coordinate characterizing the center of window $i$
- $w_i(\xi)$: Biasing potential of the $i^{th}$ window
- $\rho(\xi)^u$: Unbiased probability of finding the system at $\xi$
- $\rho_i(\xi)^b$: Biased probability of finding the system at $\xi$ in window $i$
- $\rho_i(\xi_j)^b$: Biased probability of finding the system at $\xi = \xi_j$ in window $i$
- $n_i$: Number of conformations used to construct $\rho_i(\xi)^{(b)}$ in simulation $i$
- $F_i$: Free energy constant of window $i$
- $\beta$: $\frac{1}{k_B T}$
- $M$: Total number of bins ($M = 1 + \text{int}\left(\frac{\xi_{\text{max}} - \xi_{\text{min}}}{\Delta \xi}\right)$)

In US simulations, $N_w$ simulations with biasing potentials are performed such that $w_i(\xi)$ are centered on successive values, $\xi_i$, of the reaction coordinate, where $i = 1, 2, ..., N_w^\text{190}$. The most commonly used bias function is a harmonic function that is applied to keep the system close to the reference point $\xi_i^{\text{ref}}$:

$$w_i(\xi) = \frac{1}{2k}(\xi - \xi_i^{\text{ref}})^2$$  \hspace{1cm} (2.32)

where $k$ is the force constant which modulates the strength of the bias. For each simulation window $i$, the corresponding biased probability distribution $\rho_i(\xi)^b$, which is a normalized histogram is obtained from simulations (Fig. 2.2). The subsequent unbiasing and “stitching” for a one dimensional reaction
coordinate of WHAM is performed using essentially two equations, 2.33 and 2.34 based on $e^{-\beta F_i(\xi)}$ and $\rho_i(\xi)^u$ introduced in equations 2.30 and 2.31, respectively.

First, the optimal unbiased probability distribution for each bin is computed using $\rho_i(\xi_j)^b$, which denotes the biased probability of finding the system at $\xi = \xi_j$ in window $i$,

$$\rho_i(\xi_j)^u = \sum_{i=1}^{N_w} n_i \rho_i(\xi_j)^b \div \sum_{i=1}^{N_w} n_i e^{-\beta w_i(\xi_j)}$$  \hspace{1cm} (2.33)

where $j = 1, 2, \ldots, M$ ($j$ labels the histogram bins) and $\mu_i = e^{-\beta F_i}$. Starting with an initial guess for $\mu_i$ for each window, an estimate for the unbiased value of $\rho(\xi_j)^u$ is assigned. The estimate for $\rho_i(\xi_j)^u$, is then used to refine $\mu_i$ using:

$$\mu_k = \sum_{i=1}^{N_w} \rho_i(\xi_j)^u e^{-\beta w_i(\xi_i)}$$  \hspace{1cm} (2.34)

In equation (2.34), $k = 1, 2, \ldots, N_w$ ($k$ labels the window). The estimates for $\rho_i(\xi_j)^u$, is then used to obtain new values of $\mu_i(new)$. Equations (2.33) and (2.34) need to be solved iteratively until self-consistency is reached. Otherwise, $\mu_{(old)}$ is replaced with $\mu_{(new)}$ and equations (2.33) and (2.34) are solved continually. The iteration is continued until the difference between the constants in two consecutive iterations of a bin is below a threshold. Once self-consistency of equations 2.33 and 2.34 is achieved, the unbiased potential of mean force (PMF), $W(\xi)$, is computed as:

$$W(\xi_j) = -\frac{1}{\beta} \ln \rho_i(\xi_j)^u + A_0$$  \hspace{1cm} (2.35)

where $A_0$ is an arbitrary constant.

Figure 2.3: Schematic representation of “stitching” of PMFs of 3 umbrella windows to construct the global $W(\xi)$. The three colours represent those of the normalized histograms in Fig. 2.2. The two black arrows represent adjusting of $F_i$ values.
2.5 What is Homology modelling?

Homology modelling is a computational structure prediction procedure that relies on sequence and structural information of experimentally determined protein structures (templates) to generate a structural model of another protein (target) with a similar amino acid sequence. This is possible based on two main observations: (1) the primary amino acid sequence determines secondary and tertiary structures (2) the structure is more stable during evolution and changes much slower than the amino acid sequence. Despite a few exceptions, similar sequences adopt similar structures for the majority of cases. The modelling process typically comprises the following four steps (Fig. 2.4): (1) template selection, (2) target-template alignment, also referred as sequence-structure alignment (3) model building based on the sequence-structure alignment and (4) model quality validation.

![Homology modelling flowchart](image)  
Figure 2.4: Homology modelling flowchart. Adopted from Venclovas, 2012. The image is modified with permission from authors.

In template selection, the query sequence of the target is compared to libraries of sequences of experimentally determined protein structures. Rost, 1999 quantified a limit for selection of a template by plotting the length of the sequence against the percentage of identical residues (Fig. 2.5).
Figure 2.5: The two zones of sequence alignments. Adopted from Krieger et al. 2003. Two sequences are likely to fold into the same structure if their length and percentage sequence identity fall into the region marked as “safe.” The image is reproduced with permission from authors.

If the length of two sequences and the percentage of identical residues of a given template-target pair fall in the region marked as “safe” of the above plot, the two sequences supposedly adopt a similar structure. Specifically in the case of membrane proteins, it is shown that a sequence identity at least 30% or higher with a correct alignment yields acceptable models with Co-RMSD values of 2 Å (compared to the templates) or less for the transmembrane regions.

In step target – template alignment, one of two approaches can be followed: sequence-sequence alignment or sequence-structure alignment. In brief, the generation of the simplest pair-wise sequence alignment between two sequences with high sequence identity relies only on amino acid sequences of two proteins, a scoring matrix for residue substitutions and an algorithm to produce an alignment. As the sequence identity becomes low, more and more homologous sequences are used to generate a multiple sequence alignment (MSA) in order to improve the alignment between the target and template sequences. Here, matrices sensitive to conservation patterns, also called profiles, are used to iteratively refine the MSA.

Even the MSA may not be fully reliable. When the sequence identity is below 20%, it is very difficult to determine the homology between two sequences using only the sequence data and in such cases the homology can only be detected using structure alignment methods. Furthermore, the alignment should be as accurate as possible if the sequences are low in identity to produce a reasonable model. This is because the model quality is more affected by the alignment accuracy of sequences with low sequence identity than it would with high sequence identity. Hence, in addition to conservation patterns of the sequences, structural information from template(s) is used to improve the alignment, especially if the target-template sequence identity is low.

When multiple templates are available, a multiple-structure alignment is obtained before aligning the
target sequence. Multiple-structure alignment is an NP-hard (non-deterministic polynomial-time hard) problem. As a result, a variety of methods using heuristic approaches are developed to align structures. In brief, SALIGN, as implemented in MODELLER, employs a set of six features either calculated from their sequences and structures or defined by the user to obtain a multiple-structure alignment. These properties are residue type, interresidue distance, fractional side chain accessibility, secondary structure, local structure conformation, and any user-defined feature. Multiple-structure alignment is obtained following a progressive series of pairwise-structure alignments which starts by aligning the most closely related pair of sequences, then the next most similar one to that pair, and so on. Each pairwise structure alignment is constructed based on the contribution from the six features. Once the optimal multiple structure alignment is obtained, the target sequence is aligned to the multiple structure alignment taking into account structural information from the templates. This is different from a regular pairwise sequence alignment, because sequence-structure alignment is achieved through a variable gap penalty function that for example, avoids placing gaps in the alignment of secondary structure elements, buried regions or between two residues that are close in space.

When the alignment is ready, model generation is followed. For this step too, a variety of methods are available. The method as implemented in MODELLER includes extraction of spatial restraints from a few sources: homology-derived restraints on distances and dihedral angles are extracted from the aligned template structures, stereochemical restraints such as bond length and bond angle preferences are obtained from the parameters of CHARMM-22 molecular mechanics force field and statistical preferences of dihedral angles and non-bonded atomic distances are obtained from a representative dataset of experimentally resolved protein structures. An initial model, assigned to a scoring function, is then generated by minimizing the violations of the spatial restraints. Next, using molecular dynamics with simulated annealing, the scoring function is optimized resulting in an ensemble of homology models.

Once, generated, the model should be validated. There are multiple ways to perform model validation. A number of metrics are used to check different features of the model. For example, some of the common features assessed in validation include: stereochemical quality check of the distribution of the backbone dihedral angles assessed by Ramachandran plots, an RMSD between the model and templates — for membrane proteins, RMSD value below 3 Å implies a fairly good overall fold compared to the templates, packing quality of the model assessed by evaluating the non-local environment of each heavy atom and detection of the inside/outside distribution of hydrophilic and lipophilic residues. MODELLER incorporates an atomic distance-dependent knowledge-based potential, named Discrete Optimized Protein Energy (DOPE), which is simply a sum of interactions between pairs of atoms that can also be decomposed into a score per residue. This is useful to identify the problematic regions of
the model. Based on the outcomes of the model validation, the template-target alignment may be adjusted in order to generate new and improved models. Hence, homology modelling is an iterative refinement procedure until no further improvements are detected in the model.

2.5.1 Homology modelling of \( hH V_1 \)

For the homology modelling procedure of \( hH V_1 \), the first step of template selection was not necessary, since the selection of templates for \( hH V_1 \) was straightforward. \( hH V_1 \) is a structural homolog of voltage sensor domains (VSD) of voltage-gated K\(^+\) and Na\(^+\) channels and at the time there were 3 crystal structures of VSDs available in the PDB that were suitable as templates. In the case of \( hH V_1 \), sequence-structure alignment was more suitable than the widely-used sequence-sequence alignment \(^{165–167}\) due to the relatively low sequence identity (less than 30%) between \( hH V_1 \) and its 3 templates. Nonetheless, sequence profiles generated from 300 diverse VSD sequences were used to guide the structural alignment between the three templates using SALIGN\(^{201}\) of MODELLER (details of MSA is described in Methods section of Chapter 3). Once the best MSA is obtained, a sequence-structure alignment was performed between the target \( hH V_1 \) sequence and multiple structural alignment (Fig. 2.6). Next, an initial model was generated by minimizing the violations of the spatial restraints from which an ensemble of one hundred models were generated by randomizing the coordinates. During model evaluation, the lowest overall DOPE score as well as the DOPE score per residue were assessed to select the top five homology models.
Figure 2.6: Alignment between hV1 and its 3 templates. The notations; 1ORSk, NaVBkA and 2R9R denote the sequences of VSD templates. The colours of the sequence alignment emulate the default colours scheme used for alignments in Clustal X, a graphical interface for the ClustalW multiple sequence alignment program. Three standard annotations; conservation, quality and consensus are depicted below the sequence alignment. Conservation is a numerical index reflecting the conservation of physico-chemical properties in the alignment. The color shading from most intense (bright yellow) to palest (dark brown) reflects the most conserved to least conserved residues. The alignment quality is measured based on the substitution matrix of BLOSUM62 and is plotted on a scale from 0 (dark brown) to 1 (bright yellow). A high alignment quality score for a column would suggest that there are no mutations. Consensus gives the commonest residues and their percentage for each column as reflected by the heights of the black bars.
Chapter 3

ROMP: Rapid Orientation of Membrane Proteins

Contributions:
Kethika Kulleperuma conducted the MD simulations of all systems, except PagP in lipids, and wrote the relevant sections.
John Holyoake contributed to protocol design.
Chris Neale conducted the MD simulations of PagP in POPC and wrote the relevant Methods section.
Régis Pomès provided guidance for protocol design, analysis and editorial input.
Abstract

The function of membrane proteins and peptides is intricately coupled to their interaction with lipid bilayers, which modulates their structure and their interactions with other biomolecules. These properties depend on the solvation and the orientation of the proteic solute in the anisotropic environment of the membrane. Conventional structural methods, which usually rely on detergent micelles to solvate membrane proteins, do not provide information on their orientation in membranes, underlining the need for reliable predictive methods. However, predicting the orientation of proteins and peptides in lipid membranes is nontrivial. Atomistic computer simulations in explicit lipid bilayers are expensive and impeded by the slow relaxation of lipid bilayers, which presents a significant sampling bottleneck. Here we present a general and computationally-efficient molecular dynamics protocol based on an ensemble approach to predict the orientation of peptides and proteins in lipid membranes. In this approach, the proteic backbone of the solute is rigidified and the solute is inserted at random orientations in a membrane-mimetic hydrated octane slab. The method is applied to the voltage-sensing domain of the KvAP channel, a tetrameric helix bundle; PagP, a β-barrel protein with an amphipathic α-helix; and the human amylin peptide hIAPP. In all systems, the convergence of an ensemble of simulations to a consensus orientation is attained within a few nanoseconds. The predicted consensus orientations of the two proteins are shown to be identical to those obtained from extended simulations in explicit lipid bilayers and with available experimental evidence. The method is shown to work with a flexible interdomain linker in PagP. The application of the ROMP protocol to the dimerization of hIAPP helices suggests that this method is well suited to studying peptide-peptide interactions in lipid membranes.

3.1 Introduction

Proteins that span the biological membrane, transmembrane (TM) proteins, are an important class of biomolecules. Apart from their native roles in cell signalling, transport of nutrients, and energy production, TM proteins are also potent targets for many therapeutic drugs. The knowledge of these crucial physiological functions and the development of biomedical applications inevitably rely on the detailed description of TM proteins. Obtaining such details, however, remains a significant challenge compared to cytosolic proteins solvated in a homogeneous cellular medium that is mostly made up of water. This is because TM proteins are solvated in a heterogeneous membrane. Conditions that perturb this native environment (such as by purification, extraction, or crystallization methods) are likely to perturb the native conformation and/or the function of TM proteins. Such perturbations
are evident in assessments of solvation effects on structure and function of TM proteins, underlining the importance of understanding the solvation of TM proteins.\textsuperscript{211–213}

Lipid bilayers constitute an anisotropic and chemically-heterogeneous environment for the solvation of biomolecules.\textsuperscript{214} Solvating a peptide or protein in a lipid bilayer requires balancing the complex intermolecular interactions between lipids, water, and proteic solute as well as entropic effects, ultimately resulting in energetically-favourable configurations of the solute characterized by (a) specific conformation(s) with specific orientation(s). Therefore, a complete structural description of a TM protein is not limited to its spatial coordinates, but also requires atomistic-level details about its solvation and orientation(s) relative to the heterogeneous physico-chemical layers of the lipid bilayer. Each membrane protein in the PDB is likely to adopt well-defined orientations in the membrane that are not readily available from their coordinates.

Various experimental and computational methods are available to probe the solvation and orientation of TM proteins in native or membrane-mimetic environments. Experimental methods include spin-labeling, X-ray scattering, neutron diffraction, infrared spectroscopy, electron-cryomicroscopy, and nuclear magnetic resonance.\textsuperscript{215–217} In many cases, these methods provide high-resolution data for a subset of residues, but such data is limited due to the inherent technical difficulties in generalizing protocols to any TM protein.

A number of computer simulation methods are also available to predict the solvation and orientation of peptides and proteins in membranes. They include statistical,\textsuperscript{218} implicit,\textsuperscript{219–221} coarse-grained (CG)\textsuperscript{222} and all-atom\textsuperscript{223} simulation approaches, with the accuracy of the structural description and the required computation time increasing in that order. All these methods have inherent limitations. Statistical and implicit-solvent methods do not fully capture the physical basis for solvation, particularly the entropic contribution of the solvation properties. Explicit-solvent protocols involving CG and model membranes require conducting a so-called equilibration stage, a mandatory step, especially for explicit solvent models, to overcome nonequilibrium initial conditions.\textsuperscript{220} An important limitation in these approaches, however, is that the time required for equilibration cannot be predetermined.\textsuperscript{224} In addition to being highly variable among systems, the time scales for equilibration sometimes exceed the available simulation time, inhibiting the full equilibration of the system. The need for longer equilibration of lipid simulations is due to slow relaxation of lipid molecules. Solvating a peptide or protein in such a medium demands even longer time scales owing to the strong interactions between proteins and the hydrated lipid bilayer.

As a result, simulation methods with explicit solvent usually rely on an initial best guess for the orientation of the solute and rarely provide the (statistical) certainty of attaining the global minimum.
Unless the complete orientational space of the solute in the lipid bilayer is sampled, the global minimum for solvation cannot be ascertained. In principle, exhaustive sampling can be achieved by conducting a very long simulation allowing escape from local minima or by performing multiple simulations from an ensemble of different starting orientations allowing the exploration of the complete orientational space for a given solute. These approaches are computationally prohibitive with all-atom models. Achieving equilibrium sampling of multiple minima of indolicidin, a disordered antimicrobial peptide, in lipid bilayer was shown to require aggregated simulation times exceeding 1.5 milliseconds. 

Here we present an ensemble-based, all-atom MD protocol, abbreviated as ROMP (for Rapid Orientation of Membrane Proteins), to predict the favourable orientation(s) of peptides and proteins in membranes. The method is based on the premise that a given protein conformation adopts a preferential orientation in the membrane. To overcome the slow relaxation of lipid bilayers, the systematic sampling of orientation space is performed in a membrane mimetic \( n \)-octane slab, a simplified, fluid model of lipid bilayers.

Biphasic models consisting of an aqueous phase and an octane slab have long been used in simulation studies of membrane-associated processes, including studies of the gating mechanism of voltage-sensor domains (VSD),\(^{226}\) peptide adsorption and aggregation of \( \beta \)-sheet-forming polypeptide chains of amyloid fibrils,\(^{227}\) voltage dependent membrane-insertion of peptides,\(^{228}\) and the self-association of transmembrane helices.\(^{229}\) In addition, we have used the \( n \)-octane slab to assess the quality of homology models of a TM protein on the basis of large conformational ensembles (see Chapter 4).\(^{164}\) Specifically, we used massively-repeated simulations to generate conformational ensembles of the solute exploiting the high fluidity (fast relaxation) of the membrane mimic. Results demonstrated that the structure and fluctuations of voltage sensor domains (VSD) obtained from 400-ns-long simulations in an explicit, hydrated lipid bilayer were reproduced in the \( n \)-octane slab. In this study, we use an ensemble strategy to predict the preferred orientation(s) of proteic solutes in lipid membranes.

In the rest of this chapter, we first present the development and the validation of the ROMP protocol on two TM proteins with representative folds: the VSD of K\(^+\) channel KvAP, a four-helix bundle; and outer-membrane protein PagP, a \( \beta \)-barrel linked to an amphipathic helix. For the latter, we show that ROMP can be used to predict the solvation and orientation of the individual domains separately and with the presence of a rigid or flexible linker. We then apply the ROMP protocol to characterize the membrane-bound states of the human amylin peptide (also known as human Islet amyloid polypeptide, hIAPP) in its helical conformation. The spontaneous dimerization of hIAPP helices demonstrates the potential of the method for the study of peptide aggregation.
3.2 Methods

3.2.1 Test proteins and peptides

Since the TM proteins known to date belong to one of two distinct structural classes, α-helical bundles and β-barrels, test protein structures of each type were used to develop and validate the method: a simple four-helix bundle, the voltage-sensing domain (VSD) of K\(^+\) channel KvAP \(^4\) (PDB ID 1ORS, hereafter referred to as KvAP) and an eight-stranded β-barrel proceeded by a short N-terminal α-helix, PagP (PDB ID 3GP6). \(^{230}\)

For PagP, the crystallographic water and SDS molecules associated with the PDB structure were removed and the missing sidechains of residues 4, 35, 36, 38-45, 46, 47, 146, 147, and 148 were constructed as described elsewhere. \(^{230}\) Four variants of the PagP protein were constructed: barrel-only (BO, residues 22-162), helix-only (HO, residues 1-21), fully-rigid-linker (FR, residues 1-162), and flexible-linker (FL, residues 1-162). PagP-FR, BO, and HO systems were rigidified by restraining the distances between all the C-α atoms with a force constant of 1000 kJ mol\(^{-1}\)nm\(^{-2}\). The Cα atoms of the β-barrel (residues 22-162) and the N-terminal α-helix (residues 1-19) in PagP-FL were separately rigidified without any inter-domain restraints, leaving the residues 20-21 between these two domains unrestrained.

The hIAPP structures were obtained from NMR frames of human amylin (PDB 2KB8). \(^{231}\) In order to select one out of 30 NMR frames, an average structure was constructed using the Cα atoms. The frame with the lowest RMSD to the average structure for residues 1-21 was selected. The disordered C-terminal segment (residues 22-37) was deleted from the peptide. The C- and N- termini were capped using acetyl and NH\(_3^+\) groups, respectively. The distances between all the Cα atoms were rigidified as described above. A summary of the simulations performed is listed in Table 3.1. Snapshots of model systems of K\(^+\) channel KvAP and PagP mimicking distance restraints are shown in Fig. 3.1.
Table 3.1: Summary of MD simulations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Box size (nm³)</th>
<th># OC-T/Avg.</th>
<th># water molecules</th>
<th>Cα indices for axes definition</th>
<th># replicas</th>
<th>Total simulation time</th>
</tr>
</thead>
<tbody>
<tr>
<td>KVAP</td>
<td>6.7x6.7x8.3</td>
<td>124/6413</td>
<td></td>
<td>30-43,58-73,90-101, 123-138</td>
<td>420</td>
<td>4µs</td>
</tr>
<tr>
<td>KVAP</td>
<td>7x7x8</td>
<td>541/7266</td>
<td>20-151</td>
<td></td>
<td>420</td>
<td>4µs</td>
</tr>
<tr>
<td>PagP-POPC</td>
<td>10.5x9.8x7.1</td>
<td>260/10,024</td>
<td></td>
<td>23-28, 53-58, 66-72, 85-91, 103-110, 126-132, 136-141, 155-160</td>
<td>50</td>
<td>0.5 ms</td>
</tr>
<tr>
<td>PagP-FL</td>
<td>8x8x8</td>
<td>720/9661</td>
<td>same as PagP-POPC</td>
<td></td>
<td>420</td>
<td>4µs</td>
</tr>
<tr>
<td>PagP-FR</td>
<td>8x8x8</td>
<td>720/9661</td>
<td>same as PagP-POPC</td>
<td></td>
<td>48</td>
<td>192 ns</td>
</tr>
<tr>
<td>PagP-BO</td>
<td>8x8x8</td>
<td>720/9664</td>
<td>same as PagP-POPC</td>
<td></td>
<td>48</td>
<td>192 ns</td>
</tr>
<tr>
<td>PagP-HO</td>
<td>6x6x8</td>
<td>458/5493</td>
<td>1-22</td>
<td></td>
<td>48</td>
<td>192 ns</td>
</tr>
<tr>
<td>IAPP</td>
<td>6.8x6.8x13.8</td>
<td>678/12178</td>
<td>2-29</td>
<td></td>
<td>14</td>
<td>560 ns</td>
</tr>
</tbody>
</table>

The system descriptions are as follows; KVAP: Fully-rigid VSD, PagP-POPC: unrestrained, PagP-FL: flexible linker between rigid N-terminal α-helix and rigid β-barrel, PagP-FR: Fully-rigid PagP including the N-terminal α-helix and, PagP-BO: Rigid β-barrel only, PagP-HO: Rigid N-terminal α-helix only, IAPP: Fully-rigid α-helical segment.

Figure 3.1: The model systems. A) KVAP; B) PagP-Flexible Linker (PagP-FL): the N-terminal α-helix is flexible with respect to the β-barrel. The 2 unrestrained residues are highlighted in orange licorice. C) PagP-Fully Rigid (PagP-FR): pairwise distances between all the C-α atoms are fixed. D) PagP-Barrel Only (PagP-BO) E) PagP-Helix Only (PagP-HO). The backbone atoms are represented in ice blue ribbons. The Cα atoms used to restrain the systems are highlighted in grey van der Waals and are connected with black lines indicating the distance restraints.
3.2.2 Definition of solute orientation in the membrane

The spatial orientation of the peptides or proteins with respect to the bilayer normal is defined by its tilt (\(\tau\)) and rotation (\(\rho\)) angles, using a method previously described by Kim and Im.\(^{232}\) For this calculations, the principle axes were defined by a set of C\(\alpha\) atoms for a given system as listed in Table 3.1.

3.2.3 Simulations of K\(_V\)AP and PagP in octane

Each protein was assigned one of 48 starting orientations by combining one of 8 rotation angles (from 0\(^\circ\) to 315\(^\circ\) in 45\(^\circ\) increments) with one of 6 tilt angles (0\(^\circ\), 20\(^\circ\), 40\(^\circ\), 60\(^\circ\), 80\(^\circ\), 90\(^\circ\)). All systems were separately embedded in a hydrated octane slab of thickness 3.5 nm. Details of box size, number of octane and water molecules are listed in Table 3.1. Ions Na\(^+\) and Cl\(^-\) were added as needed to neutralize each system. First, each system was energy minimized using 1000 steps of steepest descent followed by a pre-equilibration phase of 1 ns with position restraints on protein backbone and water oxygen atoms using a force constant of 1000 kJ mol\(^{-1}\)nm\(^{-2}\). This procedure allowed octane molecules to relax around the protein and at the water-octane interface until the octane density stabilized. Afterwards, 10 independent replicas for each 48 orientation of K\(_V\)AP and PagP-FL were set up by randomizing the starting velocities to initiate the equilibration with position restraints on backbone protein heavy atoms for another 100 ps with a force constant of 1000 kJ mol\(^{-1}\)nm\(^{-2}\). For the other systems (PagP-FR, PagP-BO, PagP-HO), single replica was used for each orientation. The production run included proteins being treated as rigid body as described in section 3.2.1. Each replica of K\(_V\)AP and PagP-FL was run for 10 ns whereas PagP-FR and PagP-BO systems were run for 4 ns (see section Results for the justification for reduced production run time). PagP-HO was initially run for 4 ns, which was later extended to 10 ns. All simulations were conducted with version 4.0.5 of the GROMACS simulation package.\(^{233}\) All protein structures and octane were modeled by the OPLS-AA/L parameters.\(^{234,235}\) The water model was TIP3P.\(^{235}\) Periodic boundary conditions were enforced. The integration time step was 2 fs. Bond lengths in proteins and octane were constrained with the LINCS algorithm.\(^{236}\) The Lennard-Jones interactions were evaluated using a cutoff of 1 nm. Coulomb interactions were calculated using the smooth particle-mesh Ewald method\(^{237,238}\) with a real-space cutoff of 1 nm and a Fourier grid spacing of 0.15 nm. Simulation in the isothermal-isobaric (NpT) ensemble was achieved by semi-isotropic coupling to Parrinello-Rahman\(^{239}\) at 1 bar with coupling constants of 2 ps applied in x,y directions and zero compressibility in the z direction. The aqueous solution, octane, and the protein were coupled separately to a temperature bath at 300 K with a coupling constant of \(\tau_T = 0.1\) ps using the Nosé-Hoover algorithm.\(^{240}\)
3.2.4 Simulations of hIAPP in octane

The peptide was energy-minimized using 1000 steps of steepest descent and was replicated to generate a pair of monomers. The centers of mass of the two monomers were placed 8 Å apart and the relative arrangement of the monomers was randomized using the genconf command of GROMACS to rotate each monomer around any of the 3 axes by a random degree. This step was repeated 14 times to generate 14 independent starting conformations. The minimum end-to-end distance between heavy atoms of a given pair was 3 Å. Each of the 14 pairs of orientation and its copy were placed in the aqueous phase of a biphasic water-octane box such that the peptide dimers were near opposing octane-water interfaces, at a minimum peptide-octane heavy-atom distance of 5 Å. The full system was energy minimized using 20,000 steps of steepest descent followed by a pre-equilibration phase of 1 ns with backbone position restraints on protein and water oxygen atoms using a force constant of 1000 kJ mol$^{-1}$nm$^{-2}$. Restraints on water molecules were removed in the subsequent pre-equilibration phase of 1 ns. The production run included peptides treated as a rigid body as described above for 40 ns. The other parameters of the simulations were as described in previous sections.

3.2.5 Simulations of KvAP in lipids

KvAP system was prepared as described in Kulleperuma et al, (see Chapter 4)$^{164}$

3.2.6 PagP simulations in lipids

3.2.6.1 System preparation

Coordinates of PagP, including crystallographic water molecules within 0.5 nm of the protein, were taken from the crystal structure of PagP in sodium dodecyl sulfate/2-methyl-2,4-pentanediol (SDS/MPD; PDB ID 3GP6).$^{230}$ Missing residues were built as described in section 2.1. A single molecule of dodecyl phosphate (DP) was placed in the PagP binding pocket as a dodecyl sulphate analog according to the position of dodecyl sulphate in the crystal structure.$^{230}$ Parameters for DP were constructed based on those of dodecyl phosphocholine (DPC)$^{241}$ by removing the choline group and adding a partial charge of 0.1 to each of the four oxygen atoms. Two representations of PagP, each having a different L1 loop (residue No. 38-45) conformation, were overlaid on an equilibrated POPC bilayer containing 160 lipids per leaflet. The construction of L1 loop and the equilibrated POPC bilayer are described in our previous work.$^{230}$ The principle axis of the β-barrel of each molecule of PagP was oriented 1.5 nm to the bilayer normal, bringing the L3 loop toward the bilayer, in order to align the exposed hydrophobic/hydrophilic surfaces with the hydrophobicity profile of the bilayer as suggested by Ahn et al.$^{242}$ To reduce any
anisotropic effects that the mobile loop regions of the protein may inflict upon the relative surface tension of each leaflet, the proteins were inserted anti-parallel to one another. The dimensions of this system were 10.5 nm × 9.8 nm in the bilayer plane and 7.1 nm along the bilayer normal. The closest periodic distance between the centers of mass of the two PagP molecules in the bilayer plane was 6.9 nm and the closest periodic distance between any atom in one molecule of PagP and any atom in the other molecule of PagP was 2.9 nm. To create holes in the lipid bilayer to accommodate the overlaid molecules of PagP, phospholipids with any atoms within 0.02 nm of any protein atom were removed, additionally removing a minimal number of phospholipids adjacent to PagP to ensure that each leaflet contained the same number of lipids. This procedure produced a lipid bilayer with 130 lipids per leaflet. A surface representation of each protein was then constructed using MSMS\textsuperscript{243} and PagP-shaped holes were made in the bilayer according to the protocol of Faraldo-Gomez et al.\textsuperscript{244} in 3 segments of 20 ps each while applying a position restraint along the bilayer normal with a force constant of 1000 kJ mol\(^{-1}\)nm\(^{-2}\) on phosphorous atoms of all lipids. The strength of the hole-making force was 10, 100, and 500 kJ mol\(^{-1}\)nm\(^{-2}\) in the first, second and third segments, respectively. The composite system, composed of a solvated POPC bilayer, two protein molecules and associated DP molecule and crystal waters, was neutralized with 8 Na\(^+\) ions and subjected to 1000 steps of steepest-descent energy minimization. Both DP molecules and two Na\(^+\) ions were then removed. The final simulation system contained two molecules of PagP embedded in a POPC bilayer with 130 lipids per leaflet, 6 Na\(^+\) ions, and 10,024 water molecules, for a total of 48,821 non-virtual atoms. From this initial configuration, we conducted 50 repeat simulations, each 5 µs in duration. Each simulation was initialized with a different random seed for velocity Langevin integration. The total simulation time was 0.25 ms. Given that each simulation system contains two molecules of PagP, the total effective PagP simulation time was 0.5 ms.

### 3.2.6.2 Simulation conditions

The simulation system consisted of two molecules of PagP embedded in a hydrated 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine (POPC) bilayer. MD simulations were conducted with version 4.0.5 of the GROMACS simulation package.\textsuperscript{233} The water model was TIP4P.\textsuperscript{245} Lipids were modeled by the Berger parameters\textsuperscript{246} for POPC.\textsuperscript{247} PagP was modeled by the OPLS-AA/L parameters.\textsuperscript{234,235} Hydrogen atoms in PagP were modeled as virtual sites that have regular Lennard-Jones (LJ) and Coulombic interactions but no internal degrees of freedom.\textsuperscript{248} Berger and OPLS-AA/L parameter sets were combined self-consistently using the half-s double-pairlist method.\textsuperscript{249} Periodic boundary conditions were enforced via a rectangular unit cell. The integration time step was 5 fs. The nonbonded pairlist was updated every 20 fs. Water molecules were rigidified with SETTLE and bond lengths\textsuperscript{250} between non-virtual atoms in
both PagP and POPC were constrained with P-LINCS,\textsuperscript{251} which was conducted with a single iteration and a matrix expansion order of 6. LJ interactions were evaluated using a group-based twin-range cut-off\textsuperscript{252} calculated every step for separation distances less than 0.9 nm and every four steps for distances between 0.9 and 1.4 nm, when the nonbonded list was updated. Coulomb interactions were calculated using the smooth particle-mesh Ewald method\textsuperscript{237,238} with a real-space cutoff of 0.9 nm and a Fourier grid spacing of 0.12 nm. Simulation in the isothermal-isobaric (NpT) ensemble was achieved by semi-isotropic coupling to Berendsen barostats at\textsuperscript{253} 1 bar with coupling constants of 4 ps and temperature-coupling the simulation system using velocity Langevin dynamics at\textsuperscript{254} 310 K with a coupling constant of 1 ps. Molecular graphics of all systems are produced using Visual Molecular Dynamics (VMD).\textsuperscript{255}

### 3.3 Results

#### 3.3.1 Orientation of KVAP

##### 3.3.1.1 Consensus orientation

The $\tau$ and $\rho$ angle distributions of KVAP were computed from the last 2 ns of 480 independent rigid-body simulations of 10 ns.

The 1D histograms of Fig. 3.2A and 3.2B show the $\tau$ and $\rho$ angles, respectively, and the 2D histogram of $\tau$ and $\rho$ angles is depicted in Fig. 3.2C. The histograms reveal a few conformational basins, with the main basin centered at ($\tau$, $\rho$) = (15$\pm$ 2.5°, 186$\pm$ 12.8°) for the consensus orientation populated by 80% of the sample (see green line in Fig. S1, A-bottom panel). Here, 15° and 186° denote the center of the basin and the error is computed from a Gaussian fit of the largest peak in each 1D distribution. From the relatively broad distribution for the rotation angle, it appears that rotation space around the membrane normal is less restricted than the helix tilt. This consensus orientation consists of 4 membrane-spanning helices (referred as S1-S4) with a well-hydrated internal crevice and a relatively less-hydrated external crevice (Fig. 3.2D (A)). We then analyzed the starting orientations of the replicas that reached the consensus orientation (Fig. 3.3B). Almost all the replicas initiated at $\tau$$\leq$80°, irrespective of initial $\rho$ angle, converged to the consensus orientation. With a larger initial tilt angle, the polar residues on the inter helical loops become buried in the octane slab and non-polar side chains of TM regions are exposed to water. Nevertheless, 44% of replicas with initial $\tau$$\geq$80° reached the consensus orientation within 10 ns (see the black-shaded areas in Fig. 3.3B).
Figure 3.2: **Orientation of KvAP.** A) 1D histogram of $\rho$ angle distribution B) 1D histogram of $\tau$ angle distribution C) 2D histogram between $\rho$ and $\tau$ angle distributions. D) Representative snapshots of the corresponding basins defined as a-d in plot C. Protein backbone (blue ribbons), octane (grey lines) and water (red spheres) are highlighted.
Figure 3.3: Convergence of orientation of the KvAP and PagP-FL. A) Percentage of replicas that reached the consensus structure vs. time for PagP-FL (top) and KvAP (bottom) for $n$ number of samples, where $n = 48$ (grey), 240 (red), and 480 (green). B) Fraction of replicas that reached the consensus structure (black) vs. starting orientation for PagP-FL (top) and KvAP (bottom). The data is normalized to 360°. The unconverged fraction is depicted in blue.
We then compared the consensus orientation with that obtained from lipid simulations. Since, K\textsubscript{V}AP in lipids is conformationally unrestrained, the atom selection for the definition of orientation was restricted to a group of atoms in the TM region. The same definition was used to re-compute the orientation of the main basin of K\textsubscript{V}AP in n-octane, which yielded average ($\tau, \rho$) = (14±2.8°, 191±14°). The corresponding orientation of K\textsubscript{V}AP in lipids after 200 ns was observed at ($\tau, \rho$) = (20±3.5°, 193±12°). The difference in the tilt angle between the two solvent systems could be due to solvent effects including hydrophobic mismatch, change in secondary structure and/or structural relaxation or a combination of these effects. Notably, we observe a change in secondary structure in the TM region of unrestrained K\textsubscript{V}AP in the lipids (Fig. 3.4). Furthermore, the average RMS deviation of each helix from the starting structure is 0.1 (S1), 0.01 (S2), 0.14 (S3) and 0.18 (S4) nm. Hence, in order to test the hypothesis that the change in secondary structure or structural relaxation has caused the significant change in orientation, we extracted the last frame of K\textsubscript{V}AP from lipids, fully rigidified and simulated it in n-octane for 7 ns. The resulting orientation, ($\tau, \rho$) = (19± 2.5°, 191± 10°), is highly comparable to that of the lipids, which confirms that the initial disparity (14° vs. 20° for tilt angle) observed between octane and lipid is predominantly due to the change in structure of K\textsubscript{V}AP.

Figure 3.4: Change in secondary structure of unrestrained vs. rigid K\textsubscript{V}AP in lipid (top panel) vs. octane (bottom panel), respectively. Data is shown from 390-400 ns of the lipid simulation and 0-10 ns of one of the representative replica of the octane simulations. The TM region used for the definition of orientation is from residue numbers: 30-43, 58-73, 90-101 and 123-138. DSSP secondary structure assignment is shown for coil (white), bend (green), turn (yellow), \(\alpha\)-helix (blue) and 3\textsubscript{10} helix (grey).

The effect of local slab/bilayer adaptation is measured by the average hydrophobic thickness. A
lipid or octane molecule with any of its heavy atoms within 0.43 nm from the protein heavy atom is considered as a local lipid or octane molecule. The local hydrophobic thickness of the octane slab and the lipid bilayer is defined as the maximum distance between any two C atoms of a pair of local octane molecules and any two C atoms of the lipid tail of a pair of local lipid molecules, respectively. The local hydrophobic thickness around KvAP for the three different instances, i.e fully-rigid in octane, unrestrained in lipid, and fully-rigid in octane after the transfer from lipid, are respectively 3.6±0.2 nm, 3.8±0.2 nm, and 3.7±0.2 nm. These changes in local lipid/octane hydrophobic thickness are modest. However, the trend in the changes seems to suggest that the local slab/bilayer adaptation is structure dependent and that that local slab/bilayer thicknesses for a given conformation are comparable.

### 3.3.1.2 Alternate orientations

Unlike the consensus orientation, all other orientational basins are characterized by the disruption of the octane slab caused by water defects (Fig. 3.2D (b-e)). These defects consist of one or more columns of water molecules partially or fully penetrating the octane slab as a result of solvating polar residues buried inside the octane slab. The starting orientations of almost all the simulations that ended up in these basins had a tilt angle \(\tau>60^\circ\) irrespective of rotation angle (see the blue-shaded areas in Fig. 3.3(B)). Furthermore, these water defects were created during the pre-equilibration phase in which water and octane molecules were free to rearrange themselves around the fixed protein (see Methods). At that stage, water molecules replaced octane molecules solvating the polar residues (especially in the interhelical loops) buried inside the slab as a result of the starting orientation. Taken together, these results indicate that the minor orientational basins correspond to metastable states resulting from arbitrary and unrealistic initial conditions.

### 3.3.2 Orientation of PagP systems

#### 3.3.2.1 Results for PagP-FL

The analysis of PagP-FL, which was prepared with multiple simulation replicas for each initial orientation, is presented first. The tilt and rotational angle distributions of PagP with a flexible linker between the rigid \(\beta\)-barrel and \(\alpha\)-helix domains (PagP-FL) were computed from the last 2 ns of 480 independent 10-ns simulations to identify its preferred orientation(s) (Fig. 3.3.2.1A). A strong consensus orientation corresponding to \((\tau,\rho) = (34.5\pm3.1^\circ,-44.4\pm6.4^\circ)\) was reached by 80% of the replicas within 10 ns (Fig. 3.3.2.1A top panel).
Figure 3.5: Orientation of PagP systems. A) The 2D histogram of $\rho$ and $\tau$ angles of the full data set of PagP-FL. B) Representative snapshots of the corresponding basins labelled as a-e in panel A. The protein backbone (blue ribbons), octane (grey lines) and water (red spheres) are highlighted. C) 2D PMF of PagP orientation, $W(\tau,\rho)$: only the surface below the isoenergetic contour $W=1$ kcal/mol is shown, with the most likely orientation at $W=0$. 
In the consensus orientation, the N-terminal α-helix is located at the planar octane-water interface and the octane slab is free of water defects (Fig. 3.3.2.1B(a)). Almost all the replicas initiated with $\tau \leq 60^\circ$ converged to the consensus orientation (Fig. 3.3B).

As in the case of KvAP, the farther the starting orientation is from the consensus orientation, the smaller the fraction of replicas reaching the consensus orientation. Most other conformational basins are characterized by water defects disrupting the octane slab (Fig. B(b-e)). In the search for a favourable orientation, these conformations appear to be trapped in local minima in the vicinity of the starting conformations (see black shaded areas in Fig. 3.3).

### 3.3.2.2 Rate of convergence to the consensus orientation

To identify the rate of convergence of the ensemble of simulations to the consensus orientation, the percentage of consensus orientation reached as a function of simulation time is shown in Fig. 3.3.2.1A for various sample sizes (i.e. for a total of 48, 240, and 480 simulations). Results show that replication has no effect on convergence efficiency since single replicas reach the consensus orientation just as quickly as 5 or 10 replicas per starting conformation. Although the ensemble of simulations does not reach equilibrium within the 10 ns time span, approximately 60%, 70%, and 80% of the replicas reach the consensus orientation within 1, 4, and 10 ns, respectively. Based on this analysis, the computational requirement for subsequent studies was reduced 25-fold by running a single 4-ns simulation per starting orientation. The subsequent PagP simulations in octane (PagP-FR and PagP-BO) were carried out according to this streamlined protocol. Each single replica of PagP-HO was run for 10 ns.

### 3.3.2.3 Comparison of PagP systems

To evaluate how the octane slab and the helical domain affect the orientation of the β-barrel domain of PagP, we compare the two-dimensional energy basins for $\tau$ and $\rho$ angles of PagP-POPC with all the PagP systems simulated in n-octane (Fig. 3.3.2.1C and 3.2). To highlight the consensus orientation, we focus on the lowest energy basin for each system, within 1 kcal/mol of the global minimum.

<table>
<thead>
<tr>
<th>System</th>
<th>$\tau = X \pm x^o$</th>
<th>$\rho = Y \pm y^o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PagP-POPC</td>
<td>34.4±4.2°</td>
<td>-41.4±9.0°</td>
</tr>
<tr>
<td>PagP-FL</td>
<td>34.5±3.1°</td>
<td>-44.4±6.4°</td>
</tr>
<tr>
<td>PagP-FR</td>
<td>37.5±3.0°</td>
<td>-50.1±4.6°</td>
</tr>
<tr>
<td>PagP-BO</td>
<td>29.6±3.7°</td>
<td>-45.9±6.2°</td>
</tr>
</tbody>
</table>

Table 3.2: Summary of the consensus orientations of the PagP systems. The notation $(\tau,\rho) = (X \pm x^o,Y \pm y^o)$ represents the consensus orientation, with $X,Y$ denoting the center of the basin and the standard deviation computed from a gaussian fit of the largest peak in each 1D distribution.
The basin corresponding to the consensus orientation of PagP-FL overlaps well with that obtained from extensive simulations of unrestrained PagP in a POPC bilayer, indicating that the orientations of the two systems are highly comparable.

To further characterize the consensus orientation, the solvation of PagP as measured by hydrogen bonding between the protein and water molecules is analyzed. The hydrogen bond analysis was performed geometrically using two distance cutoffs: 2.4 Å and 3.2 Å for distances between donor-hydrogen / acceptor and donor/acceptor, respectively. The cutoff distances are based on radial distribution function for each corresponding pair of atoms (data not shown).

The comparison of the residue-specific average number of hydrogen bonds with water molecules reveals that PagP in octane is more hydrated than that in lipids (Fig. 3.6A), which is clearer from the difference between the two hydration profiles (Fig. 3.6B). This difference further reveals that (1) the discrepancy in hydration is localized in the loop regions, interfaces and the N-terminal helix (Fig. 3.7), (2) the barrel pore is more hydrated in the lipid simulations (Fig. 3.7), (3) hydrogen bonding with lipid headgroups is fully or more compensated by water molecules in n-octane (see higher grey and lower magenta bars in Fig. 3.6B).
Figure 3.6: **Solvation of PagP-POPC vs PagP-FL.** A) Average number of hydrogen bonds with water molecules as a function of residue number of PagP. Data were generated from the last 2 ns of the replicas that reached the consensus orientation. B) The difference map of the average number of hydrogen bonds with water, indicated by avg.# in PagP-FL - avg.# in PagP-POPC (grey bars) for each residue. The overlaying purple bars indicate the average number of hydrogen bonds with lipid headgroups in PagP-POPC simulations. C) Average number of hydrogen bonds with water molecules and POPC headgroups as a function of time in PagP-POPC and PagP-FL simulations.
Figure 3.7: **Representative snapshot of PagP-POPC highlighting the hydrogen bonding difference between PagP-POPC and PagP-FL** Water molecules within the cut-off radius of 3.2 Å of polar atoms of the PagP-POPC are depicted in blue licorice. Residues with hydrogen bonding difference greater than 0.5 and less than -0.5 in Fig. 3.6B are highlighted in red (more hydrogen bonding with PagP-FL) and yellow (more hydrogen bonding with PagP-POPC) ribbons. The rest of the protein backbone is represented as a transparent cyan ribbon.

To further characterize the stronger hydrogen-bonding tendency with water in n-octane we compare the time evolution of the hydrogen bonds with water molecules in n-octane and with phosphate oxygen atoms of the POPC lipid headgroups and water in the lipid system (Fig. 3.6C). On average, a total of 265 water molecules participate in hydrogen bonding in the n-octane system, which is reduced by 15% in the lipid system. This reduction can be explained by the lower water density at the water/POPC interface of the lipid membrane. Hydrogen bonding with POPC headgroups makes up for 10% of the difference in hydrogen bonding. This can be explained by the fact that lipid head groups lack H-bond donors and hence cannot form hydrogen bonds with H-bond acceptors of the protein. Nonetheless, the discrepancy in the solvation properties of the two solvent systems is too small to cause a significant difference in protein orientation.

The consensus orientations of PagP-FR and PagP-BO systems also fall within the consensus orientation of PagP-POPC but with less agreement than that of PagP-FL (Table 3.2). Qualitatively, these deviations suggest that the N-terminal helix has a small effect on the \( \tau \) angle of the \( \beta \)-barrel.

However, the average consensus \( \tau \) angles of PagP-FL and PagP-POPC are within the values of PagP-FR and PagP-BO within a standard error of \( \pm 4^\circ \), suggesting that the effect of the N-terminal \( \alpha \)-helix on the positioning of the \( \beta \)-barrel in the membrane is trivial. All PagP systems in octane have a higher \( \rho \) angle than that of PagP-POPC and yet fall within the standard error of \( \pm 9^\circ \), and that the orientation of PagP systems in n-octane are comparable to that in lipids.
Previous studies have highlighted preferential orientations of aromatic residues at the membrane-water interface,\(^ {256}\) which have been suggested to act as protein anchors into the bilayer. Here, we compare the orientation of the inner and outer aromatic girdles defining the PagP surface at the membrane interfaces. To compare the rotameric states of these residues in all PagP systems, the distributions of $\chi_1$ and $\chi_2$ sidechain dihedral angles are plotted in Fig. 3.8.
Figure 3.8: Sidechain dihedral angle distribution of aromatic residues located at the outer and inner membrane-water interfaces of PagP-POPC(black), PagP-FL(red), PagP-FR(blue) and PagP-BO (green). Both $\chi_1$ and $\chi_2$ angle distribution are depicted on left and right sides respectively. A) Residues located at the inner interface. B) Residues located at the outer interface. C) Representative structure of PagP-FL (red) is superimposed on a representative structure of PagP-FR (green) with aromatic residues highlighted at the inner (left) and outer (right) interfaces.
The $\chi_1$ and $\chi_2$ distributions of all the residues agree well among all the PagP systems in n-octane. The side chain rotamers of all residues, except Phe$^{161}$ (next to the C-terminal), on the inner girdle are almost identical to that of the lipids. In contrast, the agreement is not as good for the aromatic residues of the outer girdle, such that Trp51, Tyr119, Tyr142, Tyr153 (4 out of 6 residues), which sample subpopulations of $\chi_1$ or $\chi_2$ in addition to rotamers populated in n-octane. Interestingly, these residues are located at the base of long extracellular loops (L1-L4) that are unrestrained in the lipid simulation. The rotameric space sampled in octane seems restricted due to the distance restraints limiting loop flexibility. Nonetheless, the main conformational basin of rotamers are significantly similar in both solvents, suggesting that the biphasic membrane mimetic is able to reproduce reasonably well the conformation of protein side chains at the membrane-water interface despite the absence of lipid headgroups.

3.3.3 PagP helix insertion and orientation

We then set out to investigate whether the same consensus orientation of the N-terminal $\alpha$-helix sampled in the full protein could be obtained as the isolated $\alpha$-helix. The tilt and rotation angle distribution of PagP-HO was analyzed after 4 ns simulation to identify its final orientation(s). The 2D-PMF of $\tau$ and $\rho$ angles is depicted in Fig. 3.9. Two conformational basins labelled $a$ and $b$ in Fig. 3.9A correspond to $(\tau,\rho) = (38.1\pm10^\circ,73.5^\circ\pm6^\circ)$ and $(139.0\pm21^\circ,29.5^\circ\pm7^\circ)$ and are populated 36% and 59% of the time, respectively.
Chapter 3. ROMP: Rapid Orientation of Membrane Proteins

Figure 3.9: Orientation of the N-terminal α-helix of PagP. A) 2D PMF of orientation of PagP-HO (pink) and PagP-POPC (red). B) Representative snapshots of basins indicated from a-b as defined in panel A. The protein backbone (blue ribbon), octane (grey lines), and water molecules (red spheres) are highlighted.

Orientation a corresponds to the peptide at the octane-water interface and is identical to that of the N-terminal helix in the non-rigid PagP-POPC (Fig. 3.9B). In contrast, in orientation b, the helix traverses the octane slab and adopts a transmembrane orientation despite its amphiphilic character. Although a few water molecules solvate the polar side chains within the membrane, the octane slab remains largely undisrupted. Although the dominant conformation is b, it is also clear that starting orientation with small τ angles favour this conformation (Fig. 3.10). In fact, the time evolution of the two conformations (Fig. 3.11) reveals that the population of b decays with time (which became clearer if the runs are extended from 4 to 10 ns) while that of conformation a is increasing, indicating that the simulations have not yet reached equilibrium and that orientation b is metastable and slowly decays to the surface-bound orientation.
Figure 3.10: **Fraction of replicas of PagP-HO reached one of the 2 orientational basins vs. their starting orientations.** The basins converged to the orientation \( a \) and \( b \) in Figure 3.9 are shown in green and red, respectively. Fraction of replicas reaching orientations other than that of \( a \) and \( b \) is in blue.

Figure 3.11: **Time evolution of the population of PagP-HO orientations, \( a \) and \( b \) in green and red lines, respectively.**

Interestingly, orientations \( a \) and \( b \) resemble amphipathic helices making surface-bound or transmembrane oligomeric assemblies. This observation motivated the systematic study of assembly of amphipathic helices. As first step, we assessed the application of ROMP to the human amylin peptide, or hIAPP, to examine its oligomeric surface-bound state.

### 3.3.4 IAPP helix dimerization

Human IAPP (hIAPP) is a 37-residue peptide hormone co-secreted with insulin by pancreatic beta-cells.\(^{257}\) The deposition of hIAPP as toxic amyloid fibrils is found in 90% of patients with type-II diabetes and is of wide interest in order to understand amyloid toxicity and design a treatment for diabetes.\(^{258}\) Monomeric hIAPP binds to the membrane as an amphipathic helix before aggregating on the surface to form oligomers, amyloid fibrils and causing membrane disruption.\(^{259}\) Here we study hIAPP
as a test case to apply ROMP in order to study the association of membrane-bound α-helices. Here, the application of ROMP on hIAPP does not include probing a detailed study of amyloidogenic process, which may involve lipid-induced changes in secondary structure. Instead, the goal here is to assess the use of ROMP to determine the membrane-bound state of the peptide dimer. Specifically, we examine whether hIAPP can form membrane-bound dimers similar to the non-amyloidogenic rat isomer of IAPP (rIAPP).

Whereas hIAPP peptides are amyloidogenic, rIAPP which only differs from hIAPP in 6 out of 21 residues is non-amyloidogenic. Rat IAPP have been found to form anti-parallel dimers on membrane surfaces as studied by FRET experiments, and the dimer interface has been predicted by Rosetta docking protocol.

In our simulations, most of the hIAPP peptide monomers initially placed in the aqueous phase rapidly diffused to the octane-water interface, although not necessarily to the nearest one. Subsequently, the diffusion of peptide monomers on the octane-water interface led to the formation of 4 stable dimers that were present for at least the last 10 ns of 40-ns-long simulation. Other monomers either remained in the monomeric state or made transient or stable head-to-tail contacts between monomers. The following analysis examines the details of these 4 helix dimers over the last 10 ns of simulation.

All four peptide dimers adopted antiparallel arrangements. We use the same criteria previously used to characterize the conformation of the rIAPP helix dimers to analyze the hIAPP dimers, i.e. distance between the N-terminal ends of the two helices, r, and the inter-helical angle, θ. The 2D map of these metrics reveals four conformational basins labeled 1a, 1b, 2, and 3 located at \((r, \theta) = (3.1 \pm 1 \text{ nm}, 166 \pm 4^\circ), (3.0 \pm 1 \text{ nm}, 157 \pm 7^\circ), (2.5 \pm 1 \text{ nm}, 169 \pm 3^\circ), \) and \((2.1 \pm 1 \text{ nm}, 164 \pm 4^\circ)\), respectively (Fig. 3.12). Basins 2 and 3 match basins 1 and 3 reported in Fig. 2A of Nath et al. Basins 1a and 1b are characterized by a larger offset of the N-terminal end of the 2 helices, whereas the conformer defined by \((r, \theta) = (2 \text{ nm}, 140^\circ)\) in Nath et al., which is characterized by a smaller inter-helical angle and separation, was not populated in our simulations.
Figure 3.12: **Dimerization of hIAPP helices.** 2D pmf of hIAPP dimerization as defined by the distance between the N-terminal ends of the two helices ($r$) vs. inter helical angle ($\theta$) in Nath et al.\textsuperscript{13} The plot only shows isoenergetic contours at $W=0.5$ and 1 kcal/mol. Four conformational basins labelled 1a, 1b, 2 and 3 match the orientation details provided in the text. Representative snapshots of (blue, yellow) peptide backbone and (grey) octane molecules in each basin are shown as inserts.

The 2D histogram of pairwise contacts between the two monomers in each of the four dimeric arrangements (Fig. 3.13) further confirms that dimers are antiparallel, as characterized by a broad strip of contacts between residues perpendicular to the main diagonal of the map. A hypothetical diagonal joining the most populated contacts was used to define the main registry of the interface. The midpoint of the diagonal varied from residue 14-15 in dimer 1a to 15, 12-13 and 11 in dimers 1b, 2 and 3 respectively. Dimers 2 and 3 make largest number of contacts since their midpoints are closest to the pivot point of the helix, which is between the residues 10 and 11. Fewer contacts, however, do not necessarily mean less symmetry about the midpoint as dimers 1a is more symmetric than dimer 3. By symmetry, we mean that contacts between the same residue pair (e.g., T9 of peptide 1 and L16 of peptide 2, and vice versa) are made about the midpoint. Likewise, dimers 1a and 2 are symmetric, while dimers 1b and 3 are asymmetric in their dimer interfacial contacts. Furthermore, Nath et al.\textsuperscript{13} reported 3 interfacial residues: L12, F15 and L16, all of which appear as strong dimer interfacial contacts in the symmetric dimers, in addition to T9 and S20.
Figure 3.13: Population of pairwise contacts in the IAPP dimers. Increasing number of contacts are depicted in the color scale starting from white to blue for each conformational basin 1a, 1b, 2 and 3 as defined in Figure 3.12.
Collectively, these results indicate that hIAPP favors anti-parallel dimers on the membrane interface as predicted before for rIAPP.\textsuperscript{13} The disagreement between the results presented by Nath \textit{et al} and the current study in the details of the dimer interface may be due to several reasons; exclusion vs. inclusion of the flexible C terminus, spFRET constraints vs. distance restraints on secondary structure, docking simulations without a membrane vs. the use of n-octane slab, respectively. In addition, the current study is limited in two main ways. First, there are conformers making head-to-tail contacts between monomers, which may be in the pathway of forming dimers or even oligomers of higher stoichiometry. Secondly, the last 10 ns is too short to assess whether or not the above four dimers have reached an equilibrium topology. Hence, the simulations need to be repeated and further extended for a quantitative statistical assessment of the above-mentioned properties. Extending the simulations or adding more replicas can be done inexpensively. Nonetheless, these results show that ROMP is suitable to study the self-assembly of helices at membrane-water interfaces, a task that remains challenging using atomistic models of lipid bilayer.

### 3.4 Discussion

ROMP is an MD simulation protocol that predicts the orientation of peptides and proteins of known three-dimensional structure in a membrane. By keeping the protein backbone rigid in an \textit{n}-octane slab and starting from a sparse grid in orientation space—the two main features of ROMP—the structures tested here reached energetically-favourable, suitably-solvated orientations within a few nanoseconds. In fact, more than 50\% of independent simulations initiated with systematic differences in protein orientation reached a consensus orientation within 1 ns. The rapid reorientation to the most favourable orientation is driven by the preferential partition of polar and non-polar residues into the aqueous and nonpolar phases, respectively.

Using two test integral membrane proteins with representative structure, $\alpha$-helical KvAP and $\beta$-barrel PagP, we demonstrated that ROMP recapitulates the orientations obtained in long lipid simulations. The method predicts a consensus orientation for KvAP and PagP while generating a few alternate, metastable orientations resulting from initial orientations far from TM spanning arrangements of the proteins in membranes.

We showed that specific lipid-protein interactions in PagP are largely replaced by protein-water interactions at the interface of the biphasic system, conserving the native orientation. Moreover, side chain rotamers at the biphasic interface are highly comparable to those sampled at the lipid bilayer interface. Lipid-dependent effects on orientation, however, may exist in other proteins. Thus, the main
drawback of the ROMP method is that it cannot account for lipid-dependent effects. Instead, the method is limited to generic solvation effects that can be captured in a biphasic system.

Furthermore, ROMP was used to assess the effect of inter-domain flexibility on the orientation of PagP as well as the orientations of those individual domains separately. We found that truncating the N-terminal helix of PagP or modulating its relative flexibility with respect to the $\beta$-barrel only has a moderate effect on the native orientation of the PagP barrel.

This observation is qualitatively consistent with the evidence on tilted insertion of PagP into the membrane and the fact that the N-terminal helix is not required for folding and membrane insertion of PagP but may act as a clamp, locking the orientation of PagP. The amphipathic N-terminal helix in isolation adopted two orientations, superficial and trans-membrane. The population of the trans-membrane orientation, which was initially biased by the starting orientations decayed with time. This results highlights the capacity of the method to discriminate between multiple orientations, even if metastable states are initially more populated and decay relatively slowly. In such cases, the much slower relaxation expected in lipid bilayers (2 to 3 orders of magnitude in time) may lead instead to the wrong conclusion regarding superficial vs. trans-membrane orientation.

Finally, we applied ROMP to study the self-aggregation of two $\alpha$ helices on the membrane surface. Human amylin peptide or hIAPP was used to selectively study the membrane-bound dimerization.

We found hIAPP spontaneously formed anti-parallel dimers, confirming the predicted dimer interface to a large extent. However, the simulations reported here were too short to assess the equilibrium properties of the dimer, for which longer simulation time with more replicas would be required. Future simulations will also focus on applying ROMP on self-aggregation of TM peptides in attempt to predict the tertiary structure of TM peptide bundles.

In summary, ROMP was shown to mimic the generic properties of peptides and proteins in lipid bilayers and to greatly accelerate their recovery from arbitrary initial orientations, thereby avoiding the need for simulating prohibitively slow rearrangement of lipid bilayers. The development and validation of ROMP presented here demonstrate that the method can be used as a time-efficient, tractable pre-equilibration phase as means of obtaining better starting points for lipid simulations.
Chapter 4

Construction and validation of a homology model of the human \( \text{ Hv1} \)

The contents of this section were adapted from an article published in the *Journal of General Physiology*.


This publication earned the Beckman Coulter – Molecular Devices Prize (2013) awarded by the Department of Biochemistry, University of Toronto.

*Contributions:*

Kethika Kulleperuma conducted the research relating to homology model building and MD, and wrote their sections.

Susan ME Smith conducted the research relating to sequence alignment and phylogeny, and wrote their sections.

Deri Morgan, Boris Musset, Vladimir V. Cherny and Thomas E. DeCoursey conducted the research relating to electrophysiology and wrote its section.

John Holyoake and Nilmadhab Chakrabarti contributed in MD research design and continuum electrostatic calculation respectively.
Régis Pomès provided guidance for design of simulations, analysis and editorial input.
Chapter 4. Construction and validation of a homology model of the human HV1

Summary

The topological similarity of voltage-gated proton channels (HV1s) to the voltage-sensing domain (VSD) of other voltage-gated ion channels raises the central question of whether HV1s have a similar structure. We present the construction and validation of a homology model of the human HV1 (hHV1). Multiple structural alignment was used to construct structural models of the open (proton-conducting) state of hHV1 by exploiting the homology of hHV1 with VSDs of K⁺ and Na⁺ channels of known three-dimensional structure. The comparative assessment of structural stability of the homology models and their VSD templates was performed using massively repeated molecular dynamics simulations in which the proteins were allowed to relax from their initial conformation in an explicit membrane mimetic. The analysis of structural deviations from the initial conformation based on up to 125 repeats of 100-ns simulations for each system reveals structural features consistently retained in the homology models and leads to a consensus structural model for hHV1 in which well-defined external and internal salt-bridge networks stabilize the open state. The structural and electrostatic properties of this open-state model are compatible with proton translocation and offer an explanation for the reversal of charge selectivity in neutral mutants of Asp112. Furthermore, these structural properties are consistent with experimental accessibility data, providing a valuable basis for further structural and functional studies of hHV1. Each Arg residue in the S4 helix of hHV1 was replaced by His to test accessibility using Zn²⁺ as a probe. The two outermost Arg residues in S4 were accessible to external solution, whereas the innermost one was accessible only to the internal solution. Both modeling and experimental data indicate that in the open state, Arg211, the third Arg residue in the S4 helix in hHV1, remains accessible to the internal solution and is located near the charge transfer center, Phe150.

4.1 Introduction

Voltage-gated proton channels (HV1) enable phagocytes to kill pathogens, basophils to secrete histamine, airway epithelia to control surface pH, sperm to capacitate and fertilize eggs, and B lymphocyte signaling, and may exacerbate breast cancer metastasis and ischemic brain damage. When the gene was discovered in 2006, however, the most remarkable feature of the human HV1 (hHV1) protein was its resemblance to the voltage-sensing domain (VSD) of other voltage-gated ion channels. The VSD is a protein module that confers the ability to respond to potential changes across a membrane. Classes of proteins with VSDs include many voltage-gated cation channels, HV1s, voltage-sensing phosphatases (VSPs), and C15orf27 proteins of unknown function. The VSD con-
tains four transmembrane (TM) segments, S1-S4, and intervening intracellular and extracellular loops. Voltage-sensitive cation channels contain one to several repeats of the fundamental unit comprising a VSD and a TM pore segment consisting of two TM regions; the fundamental unit may have additional N- and C-terminal domains that confer, for example, the ability to respond to cyclic nucleotides. Cation channels arrange themselves in the membrane so that four VSDs surround an ion pore that assembles from four pairs of TM helices. The conduction pathway in hHV1 is contained within S1-S4 and does not require accessory proteins. Our aim is to define the structure of the conducting (open) conformation of hHV1.

Several structural features characterize VSDs. The S4 helix contains two to seven positively charged residues (mostly arginine and less frequently lysine), each separated by two hydrophobic residues. S1-S3 contain negatively charged residues thought to form both intracellular and extracellular charge clusters together with the cationic charges in S4. Finally, VSDs contain a gating charge transfer center or hydrophobic center characterized by a highly conserved phenylalanine residue on S2, thought to delimit internal from external access. A wealth of evidence, including crystal structures of potassium and, more recently, sodium channels, indicates that the S4 helix of the VSD moves in response to the membrane potential, and that the mechanical transduction of this motion opens or closes the pore. As S4 moves, its arginines participate in salt bridges with intracellular and extracellular charge clusters, which are separated by the constriction at the charge transfer center.

We previously undertook a phylogenetic analysis that included VSDs not only from eukaryotic voltage-gated potassium (Kv), sodium (NaV), and calcium channels (CaV), but also from VSD homologues that lack an ion pore (HV1, VSP, and C15orf27). This analysis showed that the VSDs that lack an ion pore comprise a subfamily distinct from the VSDs of eukaryotic cation channels. Despite this subfamily occupying a separate branch of the phylogenetic tree, several lines of evidence indicate that its S4 moves qualitatively like the S4 of other VSDs. In VSPs, this motion presumably controls the activity of the phosphatase; in hHV1, this motion is thought to regulate proton conduction. The VSDs of CiVSP and of human C15orf27 do not appear to conduct protons or other ions. VSDs of voltage-gated potassium and sodium channels do not conduct ions but can be converted into voltage-gated proton channels by point mutations on the S4 segment. The hHV1s are VSD homologues that lack a separate ion pore but include a proton-specific conduction pathway within the VSD.

Recently, Musset et al. (2011) discovered that Asp112 is a crucial component of the proton selectivity filter in hHV1. Neutralizing mutants at position 112 not only exhibited impaired proton selectivity but were anion selective. Only the conservative Asp → Glu mutant remained proton selective. The presence
of a carboxyl side chain is not per se sufficient to provide selectivity, as Asp$^{185}$ is also in the pore but can be neutralized without affecting selectivity. These findings raise important questions regarding a mechanistic understanding of proton-specific conduction. How does Asp$^{112}$ modulate ionic selectivity? What other residues form the selectivity filter? Do titratable groups of the protein participate directly in proton relay, or does proton translocation occur via a water-mediated mechanism as in simple, water-filled pores such as the gramicidin channel. Explaining the molecular mechanism of proton conduction of hHV1 requires a deeper exploration of molecular structure.

Given the absence of an experimentally determined structure of hHV1, we used homology modeling to generate a structural model of hHV1 from its primary amino acid sequence using sequence alignments and structural templates of experimentally determined homologous VSDs. On the one hand, this task is made possible by the availability of high resolution atomic structures of homologous VSDs. On the other hand, below-average sequence identity between hHV1 and its homologues makes homology modeling a difficult task. A minimum 30% sequence identity is typically pursued in generating homology models. Homology models based on low sequence identity have been regarded as likely to be inaccurate. Therefore, sequence alignment for hHV1 is nontrivial.

Here, we focus on the structure of the open conformation of hHV1 as a starting point for future studies of closed states, gating, and other properties. We used a three-pronged approach to gain insight into the structural basis for the function of hHV1. First, a more comprehensive phylogenetic analysis of VSD sequences was undertaken, which provided a more reliable understanding of the relationships among VSDs and a more structurally informed alignment of the arginines that control S4 movement. Second, we used alternative alignments and multiple structural templates to develop several homology models of the VSD of hHV1. We performed structure-stability tests of two select models, one based on our phylogenetically informed alignment and the other based on another widely accepted alignment. To perform these tests, the computational efficiency afforded by a lipid membrane mimetic was combined with an ensemble approach consisting of many molecular dynamics (MD) simulation repeats to obtain meaningful statistics. The analysis of these simulations allowed us to choose a preferred homology model that was further validated by simulations in a phospholipid bilayer, and led to a testable hypothesis regarding the solvent accessibility of the S4 arginines of hHV1 in the open state. Finally, we used His scanning of the S4 Arg residues to evaluate accessibility by Zn$^{2+}$ effects on currents under voltage clamp. Previous studies have evaluated accessibility by mutating specific residues to Cys (Cys scanning) and then using MTS reagents as probes, detected either by Western blotting (PEGylation assay) or by altered channel gating. The Zn$^{2+}$ accessibility results support the hypothesis generated by the computational study and provide a new basis for further mechanistic investigations into hHV1.
4.2 Methods

4.2.1 Sampling of taxa and phylogenetic methods

Sequences of VSD homologues were identified by similarity using BLAST\textsuperscript{287} and HMMer\textsuperscript{288} searching, and an effort was made to include prokaryotic and fungal sequences as well as a diversity of metazoan sequences (e.g., plants, invertebrates, and vertebrates). VSD sequences were aligned using Pro-
malS3D\textsuperscript{289}, constraining the alignment to two VSD crystal structures, 2R9R (paddle chimera\textsuperscript{2}) and 3RVY (Na\textsubscript{V}Ab\textsuperscript{81}). Two other crystal structures of K\textsubscript{V} channels were deliberately omitted to avoid biasing the structural alignment toward a K\textsubscript{V}-type structure. The resulting alignment was inspected visually and analyzed with Maxalign.\textsuperscript{290} TM helices were identified by homology, and intervening loops were removed. Sequences were submitted to Prottest, which predicted the LG amino acid substitution model\textsuperscript{291} with I (proportion of invariant sites) and G model (gamma shape) parameters. A maximum likelihood phylogenetic tree was generated from this alignment using PhyML\textsuperscript{292} at the ATGC site (http://www.atgc-montpellier.fr) using the LG model, four substitution rate categories, estimated proportion of variable sites, estimated gamma shape parameter, and nearest neighbor plus subtree pruning and regrafting tree improvements. The sequence order in the TM-only alignment was randomized, and parsimony trees were generated from subsets of the randomized full alignment using Protpars (Phylogeny Inference Package [PHYLIP] version 3.5c) at the Mobyle site;\textsuperscript{293} maximum likelihood trees from the same subsets were also generated using PhyML. A different alignment was constructed from the full set of sequences using the default TCoffee program\textsuperscript{294} in which there is no structural constraint, and a maximum likelihood tree was generated using PhyML. Trees were visualized at ITOL\textsuperscript{295} and in FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

4.2.2 Gene expression

Site-directed mutants were created using the Stratagene QuikChange (Agilent Technologies) procedure according to the manufacturer’s instructions. Clones were sequenced commercially to confirm the mutations. COS-7 cells were maintained in 5% CO\textsubscript{2} and 95% air in a humidified incubator at 37°C in Dulbeccos modified Eagle’s medium containing 10% heat-inactivated fetal bovine serum, 2 mM l-glutamine, 100 IU/ml penicillin, and 100 µg/ml streptomycin. The coding sequence of hHv1 (Hven1) was cloned into pQBI25-FC3 (to make GFP-H\textsubscript{V}1) vectors as described previously.\textsuperscript{72} COS-7 cells were grown to \(\sim80\%\) confluency in 35-mm cultures dishes. Cells were transfected with \(\sim5\mu\text{g}\) of the appropriate cDNA using Lipofectamine 2000 (Invitrogen) or polyethylenimine (Sigma-Aldrich). After a 6-h...
incubation at $37^\circ C$ in 5% CO$_2$, cells were trypsinized and replated onto glass coverslips at low density in medium containing fetal bovine serum for patch-clamp recording.

4.2.3 Electrophysiology

GFP-tagged proton channels were studied. Fluorescent cells were identified using inverted microscopes with fluorescence capability (Nikon). Micropipettes were pulled using an automatic pipette puller (Flaming Brown; Sutter Instrument) from glass (PG52165-4; World Precision Instruments, Inc.) coated with Sylgard 184 (Dow Corning Corp.) and heat polished to a tip resistance ranging typically from 5 to 15 MΩ with the pipette solutions used. Electrical contact with the pipette solution was achieved by a thin sintered Ag-AgCl pellet (In Vivo Metric Systems) attached to a Teflon-encased silver wire, or simply a chlorided silver wire. A reference electrode made from a Ag-AgCl pellet was connected to the bath through an agar bridge made with Ringer solution (mM: 160 NaCl, 4.5 KCl, 2 CaCl$_2$, 1 MgCl$_2$, and 5 HEPES, pH 7.4). The current signal from the patch clamp (EPC-9; HEKA, or Axopatch 200B; Axon Instruments) was recorded and analyzed using Laboratory View SCB-68 (National Instruments), Pulse and PulseFit (HEKA), or P-CLAMP (Molecular Devices) software supplemented by Microsoft Excel, Origin 7.5, and Sigmaplot (SPSS Inc.). Seals were formed with Ringer solution in the bath, and the potential zeroed after the pipette was in contact with the cell. Where stated, reversal potentials were corrected for measured liquid junction potentials.

Inside-out patches were formed by forming a seal and then briefly lifting the pipette into the air or solution. For whole-cell and inside-out patch recording, bath and pipette solutions contained 60-100 mM buffer, 1-2 mM CaCl$_2$ or MgCl$_2$ (intracellular solutions were Ca$^{2+}$ free), 1-2 mM EGTA, and TMAMeSO$_3$ to adjust the osmolality to roughly 300 mOsm, titrated with TMAOH. Buffers used were Mes, pH 5.5-6.0, and PIPES, pH 7.0. We omitted EGTA from solutions for Zn$^{2+}$ measurements. No leak correction has been applied to any current records. Measurements were done at 21°C or at room temperature (20-25°C). The bath temperature was controlled by a system from Brooks Industries or by an in-house-built system using Peltier devices in a feedback arrangement, monitored by a resistance temperature detector element (Omega Scientific) immersed in the bath.

4.2.4 Homology model

The VSD of potassium channel KvAP (Protein Data Bank accession no. 1ORS$^4$), the paddle chimera (Protein Data Bank accession no. 2R9R$^2$), and the VSD of chain A of the Na$_V$Ab sodium channel (Protein Data Bank accession no. 3RVY$^8$) were used as templates. These structures are presumed to be in
the open state because they experience no potential during crystallization, so that their VSDs are presumed to be “activated.” The three template structures were structurally aligned using Modeller 9.7 to obtain a multiple-template, structure-based sequence alignment. The query sequence of hHV1 was then aligned to the template alignment using the large multiple sequence alignment of VSDs (abstracted in Fig. 4.1) as a guide; two alignments, “R3D” and “R2D” (differing in the register of the arginines in S4) were generated. Each alignment was then used to create a set of 100 homology models, and the best five models were selected by the smallest value of the normalized discrete optimized molecule energy (DOPE) score available in Modeller 9.7.

The S1-S2 loop is of variable length in the four sequences considered: 7, 8, 32, and 15 residues, respectively, in KvAP, NaVAb, the paddle chimera, and hHV1. In the modeling of hHV1, the S1-S2 loop was by far the least reliable part of the homology model, as quantified by a DOPE score above -0.03. For this reason, in each of the selected models (five from each alignment), the S1-S2 loop (HV1 sequence LDLKIIQPDKNNYAA) was reconstructed using SuperLooper. The starting structures were then energy minimized to correct the splice points. This procedure resulted in an acceptable DOPE score.

4.2.5 MD simulations

4.2.5.1 Octane simulations

Each template and five starting configurations of each model were then embedded in a hydrated octane slab with an approximate thickness of 35 Å. Depending on the size of the protein, large enough box sizes were used to avoid boundary condition artifacts. KvAP and the paddle chimera were in a 6 x 6 x 7.5 nm³ box size, whereas NaVAb was in a 5 x 5 x 7.5 nm³ box size, with 349 and 227 octane molecules, respectively. Both homology models were in a 6 x 6 x 7.5 nm³ cubic box with 381 octane molecules. For a given box size, the number of water molecules varies with the length and conformation of the loops. The KvAP, paddle chimera, and NaVAb systems contained 4,700, 4,509, and 3,085 water molecules, respectively. R2D and R3D systems had a mean water content of 4,420 and 4,418 water molecules, respectively. Enough Na⁺ and Cl⁻ ions were added to neutralize the systems and maintain an ionic concentration at 500 mM. The systems were first energy minimized in 1,000 steps, followed by an equilibration phase of 100 ps with backbone position restraints on protein and water oxygen atoms. This procedure allowed octane molecules to relax around the protein and at the water-octane interface until the octane density stabilized. After this equilibration step, 25 independent replicas for each system were set up by randomizing the starting velocities to initiate the equilibration with position restraints on backbone protein for another 100 ps. During an extra equilibration phase of 2 ns, proteins were
treated as rigid bodies by restraining the distance between C-α atoms with a force constant of 1,000 kJ mol⁻¹ nm⁻².

The production runs consisted of unrestrained simulations of 100 ns for each system, with subsequent extension to 200 ns for the R2D model. All simulations were generated using Gromacs (v4.0) using the OPLS-AA force field for protein and octane and the TIP3P model for water. The integration time step was 2 fs. A twin-range cutoffs of 10 Å for the van der Waals interactions and 10 Å for direct electrostatic interactions calculated by particle-mesh Ewald were used together with 10-Å neighbor lists updated every 10 steps. Constant NPT conditions were applied using Parrinello–Rahman semi-isotropic pressure coupling in XY directions, with a constant pressure of one bar applied via a coupling constant of \( \tau_p = 2.0 \) ps and zero compressibility in the z direction. The aqueous solution, octane, and the protein were coupled separately to a temperature bath at 300 K with a coupling constant of \( \tau_T = 0.1 \) ps using the Nosé–Hoover algorithm. The LINCS algorithm was used to constrain bond lengths.

### 4.2.5.2 Lipid bilayer simulations.

The x-ray structure of KvAP and the R2D model were simulated in a phospholipid bilayer. A configuration containing a 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) bilayer with 64 lipids per leaflet was obtained from and equilibrated at 300 K for 50 ns in the absence of any solute. The OPLS-AA protein force field was mixed with the Berger lipid parameters by applying the half-\( \epsilon \) double-pairlist method. InflateGRO was used to embed the proteins in the bilayer. The bilayer was then hydrated, and 500 mM NaCl was added as in the octane simulations. The final KvAP system contained 126 POPC and 5,862 water molecules in a 6.6 x 6.9 x 7.8 nm³ box. The system was first energy minimized for 2,500 steps, followed by an equilibration phase of 25 ns with backbone position restraints on protein and an unrestrained production run of 430 ns. All parameters were the same as those of the octane simulation except for pressure coupling. A reference pressure of one bar was imposed by semiisotropically pressure coupling separately in the directions of the bilayer plane and the z direction using a coupling time constant of 1.0 ps and equal compressibility.

Four representative conformations of R2D in its hydrated state (see Results), including pore water, were selected from octane simulations as the starting conformations for unrestrained 100 ns simulations in POPC after a 25 ns equilibration in the presence of position restraints on protein and pore water oxygen atoms. The 7.2 x 7.5 x 7.5 nm³ simulation cells contained 126 POPC and 5,877 water molecules on average. Simulations of the same four protein conformations were extended in octane to 100 ns. Molecular graphics were generated by VMD 1.8.7, and all trajectories were analyzed using Gromacs tools and in-house-built codes.
4.2.5.3 Continuum electrostatic calculations

We calculated the static field for the passage of a positive point charge through the \( hHV1 \) pore by solving the Poisson-Boltzmann (PB) equation using the PBEQ module of CHARMM (version 31\textsuperscript{301,302}), together with a set of optimized atomic radii for amino acids.\textsuperscript{303} The regions occupied by \( n \)-octane, protein, and water were assigned dielectric constants of 2, 4, and 80, respectively. 150 conformations were selected by picking every 50th snapshot from the combined trajectory segments in the fully hydrated conformation of the channel (III state in Fig. 4.10 E), in which a water pathway is present throughout the pore. For each of these snapshots, the coordinates of the oxygen atoms of water molecules along the pore were stored before deleting the water molecules along with ions and octane molecules. The probe point charge was placed at these O atom coordinates to calculate the static field for each protein conformation. The PB equation was solved on a grid of 115 x 115 x 141 points with an initial cell size of 1.0 Å, followed by focusing on a finer grid of 0.25 Å. The implicit membrane was assigned a thickness of 34 Å, and no membrane potential was applied. The calculation was repeated for the same set of snapshots after neutralizing the charge of D112 to mimic the effect of neutral mutants of that residue on charge transfer through the pore of \( hHV1 \).

4.3 Results

4.3.1 Multiple sequence alignment and phylogenetic analysis

A previous multiple sequence alignment and phylogenetic analysis including 123 genes showed that the VSDs of \( H\_V\_1s \), VSPs, and C15orf27s comprise a subfamily distinct from other, mainly eukaryotic VSDs\textsuperscript{94}. To get a more comprehensive view of the relationships among VSD sequences, a multiple sequence alignment of almost 300 diverse VSD sequences was created using PromalS3D,\textsuperscript{289} an algorithm that incorporates both structural and sequence information. The sequences were chosen to include preferentially VSDs known to respond directly to membrane potential and were drawn from a wide spectrum of organisms. The parameters were chosen so that the algorithm used the VSDs from the \( Na\_V\_Ab \) crystal structure (3RVY) and the paddle-chimera crystal structure (2R9R) to constrain the alignment structurally. A representative subset of the TM-only alignment is shown in Fig. 4.1.

The tree produced by maximum likelihood phylogenetic analysis of this multiple sequence alignment (Fig. 4.2) has very similar topology to trees previously constructed both from full-length ion channel sequences and from the sequences of the ion pores only.\textsuperscript{304,305} As was concluded in the previous full-length and pore-only alignments, it seems clear from this tree that \( K\_V\_ \) and \( Na\_V\_Ca\_V\)-type channels
separated in bacteria, at the latest approximately one billion years ago.

Figure 4.1: A subset of a larger multiple sequence alignment of the four TM regions of VSDs informed by alignment of the open state produced by structural superposition of crystal structures of the paddle chimera (green ribbon) and Na\textsubscript{V}AB (white ribbon). The arginines of S4 (blue sticks) are labeled according to their position on the paddle chimera. Note that the most extracellular R of Na\textsubscript{V}AB corresponds to R2 of the paddle chimera, and the most intracellular R of Na\textsubscript{V}AB corresponds to K5 of the paddle chimera. The completely conserved phenylalanine on S2 is shown and labeled as “CTC”. The following residues are indicated with symbols over the alignment: $\%$, CTC; $\ast$, R1 of Shaker; $\#$, R2 of Shaker; $\wedge$, R3 of Shaker.

The tree in Fig. 4.2 also agrees with topologies of previous trees obtained from pore alignments in that it shows VSDs from CNG-type channels as more closely related to VSDs from K\textsubscript{V} channels, and VSDs from transient receptor potential channels as more closely related to VSDs of Na\textsubscript{V}/Ca\textsubscript{V} channels. Interestingly, phylogenies created from various multiple sequence alignments (see below) of VSD sequences contain no indication of Na\textsubscript{V}/Ca\textsubscript{V}-type VSDs being associated with K\textsubscript{V} pores, or vice versa, indicating that swapping VSDs and pore domains among the ion channel families is extremely rare to nonexistent. The similarity of the phylogeny based on pore-only and the VSD-only (pore-excluded) sequences reinforces the validity of both.

We now focus on the charged residues in the S4 region that are considered to sense membrane potential. The full multiple sequence alignment (abstracted in Fig. 4.1) shows an alignment in S4 in which R4 (the fourth charged residue counting from the N terminus) of Na\textsubscript{V}Ab occupies a position equivalent to K5 of Shaker, in agreement with the structural alignment obtained by superposition of the crystal structures of Na\textsubscript{V}Ab and the paddle chimera (Fig. 4.1). This same alignment also predicts that R3 of H\textsubscript{V}1 is equivalent to R4 of Na\textsubscript{V}Ab and K5 of the paddle chimera. Similar to alignments produced by HHpred,\textsuperscript{90} PromalS3D alignments that included just Na\textsubscript{V}/Ca\textsubscript{V} and K\textsubscript{V} VSDs and that were guided by the Na\textsubscript{V}Ab and paddle-chimera crystal structures also confirmed the homology of R4 of Na\textsubscript{V}/Ca\textsubscript{V} with K5 of K\textsubscript{V} (not depicted). The consensus from the phylogenetic analysis and the structurally guided multiple sequence alignments leads to the conclusion that R3 of H\textsubscript{V}1 is equivalent to K5 of Shaker.
Figure 4.2: Phylogenetic tree constructed from the multiple sequence alignment of VSDs exemplified in Fig. 4.1. Branches and names of subfamilies of VSDs are color coded. KV-E, eukaryotic potassium channel; KV-B, prokaryotic potassium channel; KV-H, H family potassium channel; HCN, eukaryotic hyperpolarization-activated, cyclic nucleotide-gated; CNG, cyclic nucleotide-gated channel; Trp, transient receptor potential channel; Nav-E, eukaryotic sodium channel; Nav-B, prokaryotic sodium channel; Cav-E, eukaryotic calcium channel; Hv1, voltage-gated proton channel; VSP, voltage-sensitive phosphatase; C15orf27, homologues of human C15orf27, a protein of unknown function.

4.3.2 Homology modeling

These results, along with the availability of the NavAb crystal structure, suggested a need for the generation of a new homology model of hHv1. The model generated is presumed to be in the activated state (see Materials and methods). The two alternative sequence alignments of hHv1 to the template structural alignment, R2D and R3D (see Methods), differed significantly only in the Arg registry in the fourth TM helix. The Arg registry of R2D obtained from the analysis in Fig. 4.1 has been suggested in only two previous studies. The R3D Arg registry closely resembles that in most previous studies. As expected from the underlying alignment, the subsequent models from these two alignments also differed in the positioning of the Arg residues of S4. From this point, we shall refer to models in which D112 is positioned opposite the second and third Arg residue (i.e., R208 and R211) as derived from the R2D and R3D alignments, respectively (Fig. 4.3).
4.3.3 Evaluation of alternative models

To evaluate the quality of the two structural models, R2D and R3D, we performed MD simulations of the protein embedded in a hydrated membrane-mimetic $n$-octane slab. The principal advantage of using an octane slab rather than an explicit lipid bilayer is that the shorter relaxation time scale of the octane phase speeds up the conformational relaxation of the protein and the statistical convergence of the simulations. This computationally efficient feature of the $n$-octane slab allows the use of many independent samples for a statistically meaningful study (see Methods for more details).

To validate the use of octane as a membrane mimetic, we performed a 430-ns-long control simulation of the highest resolution VSD template, $K\alpha$AP, in a lipid bilayer and compared the results with those of octane/water simulations. The results, which are illustrated in Fig. 4.4, show that the average structure of the octane-solvated VSD is closer to that of the lipid-solvated VSD (RMSD of 0.12 nm) than either of these structures are to the initial crystallographic model (0.18 and 0.24 nm, respectively). These results demonstrate the validity of using this membrane mimetic for VSD-like domains.

Figure 4.4: A plot of the root mean square deviation (RMSD) versus time for $K\alpha$AP in a phospholipid bilayer and in octane.
25 simulations were performed on each of the top five models from each alignment (see Methods). In each of these simulations, the system was allowed to evolve in the octane slab at 300 K for 100 ns without any structural restraint. A clustering analysis was performed to characterize the conformational heterogeneity within an ensemble of structures generated by pooling together snapshots taken at 2 ns intervals during the last 20 ns of each of the 25 or 125 simulations for the templates and the homology models, respectively. The criterion for the clustering analysis was the RMSD of TM C-α atoms from the starting conformation of the protein, a measure of overall structural stability. The single-linkage clustering method of GROMACS clustering tool g.cluster was used to add a particular snapshot to a cluster when its RMSD value is less than a predefined cutoff value (Rc (nm)). The clustering of 1,250 snapshots from each homology model was repeated with increasing cutoff values ranging from 0.1 to 0.17 nm in 0.01 nm increments. The partition of the individual snapshots into clusters is a measure of structural similarity within the structural ensemble. The smaller the cutoff, the larger the number of clusters and the smaller the number of structures in each cluster (Fig. 4.5) illustrates the number of clusters and number of structures in the most highly populated cluster as a percentage of the total population at each cutoff value of the templates, R2D and R3D. Because model variation within R2D and R3D was found to be trivial, the structures from the top five models are combined for each R2D and R3D.

Figure 4.5: Fraction of structures in the most populated cluster, shown as a percentage of the total population at each clustering cutoff value, Rc, for each of the three VSD templates and the two homology models.

4.3.4 Analysis of structural divergence

The statistics of the clustering itself highlights structural divergence. As the cluster cutoff value increases, the number of clusters decreases, indicating that the boundary for grouping the related structures together becomes gradually less stringent. The number of clusters at any cutoff value is always higher
in the homology models than in the templates, indicating that structural divergence is greater in the homology models than in the templates. In $K_V AP$ and the paddle chimera, the population of the most highly populated cluster ($N_C$) at every cutoff value is a large percentage (from 50% at 0.1 nm to 100% at a 0.17 nm cutoff) of its total population, i.e., the total number of snapshots (Fig. 4.5). In contrast, $N_C$ for $Na_V Ab$ is always less populated, further indicating that its structure is more divergent than those of $K_V AP$ and the paddle chimera. The same interpretation applies to the homology models, with a significantly larger structural divergence in R3D than in R2D models.

Two measures of structural stability of the models compared to the templates are the RMSD itself and the change in secondary structure. At each cutoff value, the probability distribution of the RMSD values of individual conformations in the most highly populated cluster was plotted for both templates and homology models (Fig. 4.6, left). This analysis reveals several key features. First, the structures of the three templates, although different from the starting structure, converge to a common single point independent of the cutoff value. Second, there is a trend of increasing deviation among these structures as more and more structures are added onto a cluster (in other words, the cluster grows with increasing cutoff value). The deviation is minimal in $K_V AP$ and maximal in $Na_V Ab$. This property is correlated to the decreasing atomic resolution of the structural models from $K_V AP$ to the paddle chimera and to $Na_V Ab$ (1.9, 2.4, and 2.7 Å, respectively). For the homology models, deviation was greater for R3D than for R2D. Moreover, the R2D models converge to a single point, whereas R3D models lack such a feature. Finally, the distribution of RMSD values as the cluster grows retains its overall shape better in R2D than in R3D, indicating greater structural divergence in the latter model.

The change in secondary structure, as measured by the loss or gain of $\alpha$-helical residues in the TM region, is shown at different cutoff values in Fig. 4.6 (right). Qualitatively, $K_V AP$ undergoes both loss and gain of $\alpha$-helical structure, whereas the paddle chimera and $Na_V Ab$ undergo mainly a gain and a loss of $\alpha$-helical structure, respectively. Discrepancies in $\alpha$-helical content between the crystallographic models and the simulations of the VSD templates could be caused in part by inaccuracies in the OPLS-AA force field.

However, secondary structure is largely preserved in the structural ensembles of the two homology models, with distributions centered near zero (the greater discrepancy as a function of cluster cutoffs again reflecting the greater conformational diversity of R3D). As such, the main difference between the two homology models lies in the consistency of the relative arrangement of the TM helices, with significantly looser helix packing in R3D relative to R2D (Fig. 4.7).

Collectively, these data suggest that the R2D model is more stable and less divergent than R3D. Structural divergence of the templates increased with decreasing resolution of the crystal structures.
The fact that the R2D model was only moderately more divergent than the lowest resolution structure (NaVAb) supports its validity as a model.
Figure 4.6: **Analysis of structural relaxation from massively repeated sampling.** Relative probability distributions of (left) RMSD values (see Results) and (right) change in the number of α-helical residues, $\Delta N_\alpha$, in the most populated cluster at each cutoff value (in nanometers, top right), relative to the starting structure of the simulations. The data are normalized to the total population.
4.3.5 Structural properties

Having compared the overall structural stability and convergence of R2D and R3D, we examined the extent to which these models exhibit previously recognized structural features of VSDs. One “structural signature” of VSDs is two networks of electrostatic interactions (salt bridges), one internal and one external, between basic and acidic residues in the four helices.\textsuperscript{90,270,273} We therefore systematically screened R2D and R3D clusters to identify alignment-specific salt bridges adopted during simulations and evaluated the consistency of those signatures as the cluster grew. Fig. 4.8 illustrates the percentage of structures bearing a particular salt bridge in the highest populated cluster with increasing cutoff value in R2D and R3D. Fig. 4.9 highlights the salt bridges of both models that were present in >10% of the population of the most highly populated cluster at the 0.1 nm cutoff value.

As shown in Fig. 4.9 A, R2D exhibits two networks of salt bridges: an external network formed by R205 and R208 in S4, E119 and D112 in S1, and D185 in S3; and an internal network formed by R211 in S4, K157 and E153 in S2, and D174 and E171 in S3. Residue F150, homologous with the conserved phenylalanine of the charge transfer center described by\textsuperscript{274}, appears to mark the division between internal and external salt-bridge networks.
Figure 4.8: An analysis of salt-bridge populations of the two homology models (R2D: black, R3D: grey) in the most populated cluster at each cutoff value (in nanometers, right on each panel).

Figure 4.9: Salt-bridge formation in the (A) R2D and (B) R3D model. Acidic and basic residues in the α-helical TM region (red, S1; yellow, S2; green, S3; blue, S4) are shown in licorice representation. The salt bridges present in >10% of the most populated cluster obtained with $R_c = 0.1$ nm and are highlighted in thick dotted lines.
Note that the population of some prominent salt bridges drops as the cluster cutoff increases, implying that these salt bridges contribute to the observed structural convergence and structural stability. In particular, the salt bridge between D112 and R208 located approximately at the center of the pore, was present in four of the five starting models of R2D and appears to be the most populated salt bridge of the external network during the simulation. The salt bridge of R205 and D185 was not present in any of the starting structures and occurs with a lower probability during the simulation but is still present a large proportion of the time, indicating a probable role in stability. Other salt bridges of the external network do not seem to make a significant contribution to the structural stability and convergence. For example, the R208-D185 salt bridge is absent at low cutoff and increases only to 10% at the highest cutoff examined.

The formation of the internal salt-bridge network in the R2D model is even more striking. The branching interaction of R211 with both E153 and D174 was the only salt bridge present in the starting structures. The consensus conformation after simulation, however, reveals a highly populated five-member salt-bridge network involving E153, K157, E171, D174, and R211.

Comparison of the R3D model with R2D reveals several important differences (Fig. 4.9). An external salt-bridge network in R3D comprising a chain linking D112-R211-E119-R208 together is present in the initial structures. Structural relaxation leads to stabilization of the R211-D112 ion pair at the center of the helix bundle, whereas the other two Arg residues evolve to form a persistent chain of salt bridges, R205-E119-R208-D185, in the external funnel (Fig. 4.9 B). These interactions tend to become less populated with increasing cutoff values because of greater structural heterogeneity (Fig. 4.8). The internal salt-bridge network is missing in R3D. Only K157-E153 is significantly populated, with low occurrence of K157-D174 and K157-E171 ion pairs.

### 4.3.6 Hydration profile

Water penetration into VSDs via external and internal water-filled crevices has been suggested to focus the membrane electric field and to provide the structural basis for ion translocation in mutant VSDs; therefore, it must be considered a characteristic structural feature of VSDs.\textsuperscript{96,310,311} Given that hHV_{1} is a proton-selective channel that excludes the translocation of other ions, it is reasonable to expect it to contain a narrow pore or a constriction site within the TM region. Having identified R2D as the preferred homology model for hHV_{1}, we next characterize pore hydration from extended simulations of 200 ns for each of 125 replicas.

Specifically, we examine the conformational fluctuations of residues located at the narrowest part
of the channel (Fig. 4.10) and the time-averaged probability distribution of water along the mean axis of the helix bundle (Fig. 4.11) using conformations saved every 10 ps. The hydration profile reveals the presence of external and internal water-filled cones consistent with the structure of other VSDs. Between these two funnels, two bottlenecks, or “constriction sites,” delineate a narrow hydrated region at the pore center (Fig. 4.11). The external bottleneck corresponds to the location of the salt bridge between D112 and R208, whereas the internal bottleneck is at the interface between R211 and F150. The distances between these two residue pairs were monitored to characterize their spontaneous conformational fluctuations (Fig. 4.10, A and B). Three distinct populations emerge for the external constriction site (D112-R208). These populations represent the three configurations of the salt bridge: bidentate and monodentate conformations (involving two or one hydrogen bond, respectively), and open, in order of increasing separation. The separation between R211-F150 is essentially bimodal (Fig. 4.10, C and D). When the distance between the center of mass of the phenyl ring of F150 and the guanidinium Cz atom of R211 is \(<0.55\) nm, the guanidinium group makes a direct contact with the aromatic ring. Rapid conformational fluctuations modulate the size of each of the two constrictions independently of the other. Seven distinct conformational states are resolved in the two-dimensional map, of which five are labeled IV and illustrated in Fig. 4.10 E. The most constricted of these states (designated IV in Fig. 4.10 E) corresponds to a bidentate D112-R208 salt bridge and close nonpolar contact of R211-F150. In contrast, the least constricted state (III) corresponds to disrupted interactions for both D112-R208 and R211-F150 residue pairs. Although the latter residue pair is disrupted 87% of the time, the D112-R208 salt bridge is disrupted (or “open”) only 10% of the time (state III). Here, “open” refers to the conformation of the D112-R208 salt bridge, not that of the channel, which is presumably in the open conformation throughout the simulation.
Figure 4.10: **Structural fluctuations of the two channel constrictions in model R2D.** (A) Configurations of the D112-R208 salt bridge (from left to right: bidentate, monodentate, and open). Hydrogen bonds are shown as orange lines. (B) Probability distribution of $d_1$, the distance between atom C$\gamma$ of D112 and atom Cz of R208. (C) Closed (left) and open (right) conformations of F150 and R211. (D) Probability distribution of $d_2$, the distance between the center of the benzyl ring of F150 and atom Cz of R211. (E) Two-dimensional probability distribution of $d_1$ and $d_2$, with increasing probability from purple to yellow.
Computing the hydration profile separately for the most and least constricted conformations of the bottleneck (respectively, states IV and III in Fig. 4.10 E) confirms that pore hydration is modulated by the conformation of these two constriction sites (Fig. 4.11). The two distinct bottlenecks characterizing the constricted state (Fig. 4.11, A and C) disappear in the relaxed state, where water density becomes nearly homogeneous throughout the narrow, 1-nm-long pore (Fig. 4.11, B and D). Unlike the constricted state, the dilated state contains a continuous hydrogen-bonded column of water molecules with occasional branching. The formation of a continuous water chain is facilitated by the disruption of the D112-R208 salt bridge, with one third of these opening events lasting longer than 1 ns.

To evaluate the effect of the membrane mimetic on the homology model, four simulations of R2D were extended by another 100 ns, separately in octane and in a phospholipid bilayer. The RMSD between the average backbone structures in the two solvents is 0.15 nm, and the average hydration profile of the pore is similar in the octane and lipid simulations (Fig. 4.11 E), indicating that the overall structural properties of the model are conserved in a lipid bilayer.
Figure 4.11: **Pore hydration in the R2D model.** Representative snapshots are shown successively for (A) the most constricted conformational state defined as $d_1 < 0.42 \text{ nm}$ and $d_2 < 0.55 \text{ nm}$ and (B) least constricted conformational state defined as $d_1 > 0.52 \text{ nm}$ and $d_2 > 0.55 \text{ nm}$ (see Fig. 4.10 E). Residues D112, F150, R208, and R211 are shown in licorice representation. Water molecules within 5 Å of these residues are colored in red and white, and hydrogen bonds between these molecules are shown with orange lines. (C) Water density within a 0.7 nm radius of the mean axis of the pore, normalized to the bulk water density in the most constricted and (D) least constricted conformational states, respectively. (E) Comparison of the water density profiles from four control simulations of 100 ns, successively in n-octane (blue) and in a POPC bilayer (magenta). Error bars represent the standard deviation.

### 4.3.7 Energetics of ion translocation

As a first step toward understanding the energetics of ion permeation in $hHV_1$, we used continuum electrostatic calculations to compute the static field, i.e., the energetic contribution arising from the distribution of electric charge in the pore, for the transfer of a positive point charge through the water-filled state of the pore (corresponding to conformation III in Fig. 4.10E; Fig. 4.12A). In the WT channel, the static field profile is compatible with the transfer of a point charge through the narrow part of the pore. Overall, the static field can be approximately divided into three regions. The cytosolic end of the pore (-1.5 <z <-0.5 nm) contains an energy well induced by the net negative charge from the salt-bridge network (E153, K157, E171, D174, and R211) present in the internal water-filled crevice. Importantly,
the static field essentially cancels out throughout the narrowest region of the pore (-0.5 <z <0.8 nm), which contains D112 (0.25 <z <0.8 nm) and the flexible side chain of R208 (-0.5 <z <1 nm). The effect of the positive charge of R205 in the extracellular crevice is cancelled out by the proximity of D185.

Figure 4.12: Effect of the charge distribution of the channel on the energetics of ion translocation. (A) Representative snapshot of the water-filled open conformation of Hv1. Water molecules and charged residues in the pore are shown in licorice representation; the extracellular end is to the right. (B) Static field energy for the transfer of a positive point charge in (blue) WT and (red) neutral-D112 forms of Hv1. Two thick lines represent the mean of five running averages, each computed from a dataset of 30 snapshots using a five-point moving window. Blue shading and orange error bars represent the SEM.

Upon neutralization of residue D112, the static field increases by ~10 kcal/mol in the narrow region of the pore, where the presence of the unpaired positive charge of R208 presents a significant barrier to the movement of cations (Fig. 4.12 B). Collectively, these findings suggest why D112 is necessary for the movement of protons and why neutral mutants of D112 are selective to anions (see Discussion).

4.3.8 Accessibility studies

The MD results favor the R2D open-state model as more stable than the R3D open-state model. These results suggested that when hHv1 is in the open state, R211 should be accessible only to the cytoplasmic side, and that R205 should be accessible only to the extracellular side. We tested this prediction by
replacing each of the three Arg residues in the S4 TM segment with His. We probed the mutants with Zn$^{2+}$ on the assumption that Zn$^{2+}$ binding to the introduced His residue should affect the measured current. Because all three Arg residues studied here are located within the putative proton conduction region and face this pathway in existing models,$^{165,166,312}$ it would be surprising if Zn$^{2+}$ binding to any one of these sites had no detectable effect on current. Because a larger effect does not necessarily indicate better accessibility, we focus only on the qualitative effects of Zn$^{2+}$ on each introduced His residue. We cannot rule out the possibility that Zn$^{2+}$ might bind to a different site that is preferentially exposed by a particular mutation.

### 4.3.9 Strategy for Zn$^{2+}$ studies.

To minimize Zn$^{2+}$ binding to other parts of the channel, we replaced both His residues known to bind Zn$^{2+}$ in WT channels,$^{72,312}$ namely H140A and H193A. We also truncated the C terminus at position 221 to prevent dimerization,$^{91,93,275}$ because Zn$^{2+}$ binding at the dimer interface appears to interfere with channel opening.$^{312}$ In addition, using monomeric constructs precludes complications caused by interactions thought to occur between protomers during gating$^{275,313}$ but whose mechanism is unknown. For most measurements of external accessibility, Zn$^{2+}$ was applied at pH$_o$ 7.0, because H$^+$ and Zn$^{2+}$ compete for binding sites on proton channels.$^{314}$ Whole-cell measurements of mutants that generated small currents were done at pH$_i$ 5.5 to increase current amplitude. Measurements in inside-out patches were made at pH$_i$ 6.0, also to increase the current amplitude. Effects of Zn$^{2+}$ were weaker at pH$_i$ 6.0 than at pH$_i$ 7.0 but not dramatically so; the more profound competition in WT channels is caused by Zn$^{2+}$ coordination at multiple titratable sites.$^{314}$ We interpret effects of Zn$^{2+}$ as the result of binding to the His residue introduced by mutation; EGTA application removes Zn$^{2+}$ and should reverse the effect.

### 4.3.10 Behavior of the mutants in the absence of Zn$^{2+}$.

There were no drastic behavioral changes in voltage-clamp studies of the R→H mutants compared with WT channels. All three constructs were nonconducting at negative voltages and opened with time during depolarizing pulses. Specifically, we did not observe evidence for H$^+$ conductance activated at negative voltages, or for proton carrier activity at intermediate voltages, as might have been expected by analogy with the His scanning studies of Starace and Bezanilla in Shaker K$^+$ channels.$^{96–98}$ All R→H mutants retained proton selectivity. The change in $V_{rev}$ for a 1.5-U change in pH$_o$ averaged $-75 \pm 4.5$ mV (mean ± SEM; $n = 4$) for R205H, $-83.2 \pm 7.3$ mV ($n = 5$) for R208H, and $-75.1 \pm 2.0$ mV ($n = 8$) for R211H, compared with $-80.9$ mV in the WT channel.$^{94}$ A report that R211x mutants are permeable to
guanidinium\textsuperscript{315} evidently does not indicate a general widening of the pore, because smaller cations and anions were impermeant in R211H. The mean change in $V_{\text{rev}}$ (corrected for measured liquid junction potentials) when TMA$^+$ or CH$_3$SO$_3^-$ was replaced was $+0.3 \pm 2.0$ mV (mean $\pm$ SEM; $n = 5$) for K$^+$, $+2.3 \pm 0.4$ mV ($n = 5$) for Li$^+$, $+3.0 \pm 1.4$ mV ($n = 5$) for Na$^+$, and $+0.8 \pm 1.3$ mV ($n = 4$) for Cl$^-$. We interpret guanidinium permeation as a reflection of its propensity to modify hydrogen bonds and denature proteins.\textsuperscript{316}

4.3.11 External accessibility.

Fig. 4.13 A shows that the “Zn$^{2+}$-insensitive” background construct studied in whole-cell configuration was negligibly affected by externally applied Zn$^{2+}$ at 10 $\mu$M but was distinctly inhibited at 100 $\mu$M. Similar results were obtained in eight cells. Therefore, we adopted 10 $\mu$M Zn$^{2+}$ as a test concentration. Fig. 4.13 B, representative of 10 cells, shows that the inner-most Arg in S4, R211H, was not sensitive to externally applied Zn$^{2+}$.

4.3.12 Internal accessibility.

Control measurements in inside-out patches produced a similar picture for internal application (Fig. 4.14 A) as seen for external (Fig. 4.13 A); 10 $\mu$M Zn$^{2+}$ produced negligible effects ($n = 5$), and 100 $\mu$M Zn$^{2+}$ distinctly attenuated the current. Determining internal accessibility requires inside-out patches, and unfortunately, both R205H and R208H exhibited currents too small to permit clearly interpretable data. In four of five inside-out patches, R208H currents appeared to be inhibited to some extent by 10 $\mu$M Zn$^{2+}$ (not depicted). R211H was amenable to study, and we found that 10 $\mu$M Zn$^{2+}$ applied internally consistently inhibited the current ($n = 7$), indicating accessibility to the internal solution (Fig. 11 B). The application of 10 $\mu$M Zn$^{2+}$ during long pulses did not produce clear evidence of block during the pulse; instead, currents declined gradually over several pulses (1-2 min). However, as illustrated in Fig. 4.14 C, the application of 1 mM EGTA during depolarizing pulses did reverse Zn$^{2+}$ inhibition ($n = 3$), indicating open-state accessibility of R211H.
Figure 4.13: Evaluation of external accessibility of the three Arg residues in the S4 helix of hHV1. All measurements were made in a construct designed to have low sensitivity to Zn$^{2+}$, H140A/H193A/K221stop. (A) Currents during identical families of pulses in 10 mV increments up to +60 mV in the background construct in whole-cell configuration, in the presence of 0, 10, or 100 µM Zn$^{2+}$ in the bath solution. (B) Whole-cell currents in the R211H mutant during identical families of pulses in 10 mV increments up to +80 mV, in the presence of 0, 10, or 100 µM Zn$^{2+}$. (C) Currents during identical pulse families in 20 mV increments up to +60 mV in the R208H mutant in the absence (black) or presence of 10 µM Zn$^{2+}$ (red). (D) Currents during identical pulse families in 10 mV increments up to +30 mV in the R205H mutant in the absence (black) or presence of 10 µM Zn$^{2+}$ (red). (E) Currents during three consecutive pulses to +90 mV in a cell transfected with R205H. At or slightly before the arrow, 10 µM Zn$^{2+}$ was applied to the bath, reducing the current during the pulse (green). The next pulse (red) shows the steady-state extent of Zn$^{2+}$ inhibition. (F) Later in the same experiment, EGTA was added during a pulse at or before the arrow, rapidly relieving Zn$^{2+}$ effects. For all parts, pH$_o$ was 7.0; pH$_i$ was 7.0 for A, B, E, and F; and pH$_i$ was 5.5 for C and D.

The apparent contradiction between the effects of Zn$^{2+}$ and EGTA can be understood by considering that a substantial unstirred volume at the tip of the pipette insulates the inside of the membrane from the bath solution. The concentration of Zn$^{2+}$ near the membrane will increase gradually after a delay and must approach equilibrium to affect the current markedly. In contrast, the local EGTA concentration need only reach 1% of its final value to remove Zn$^{2+}$ because of the 100-fold excess of EGTA (1 mM) over Zn$^{2+}$ (10 µM). Higher concentrations of Zn$^{2+}$ (100 µM) did sometimes reduce the current when applied during pulses, but this result was ambiguous because this concentration has a distinct effect in the control (Fig. 11A). In summary, R211H is accessible to the internal solution even in the open state.
Chapter 4. Construction and validation of a homology model of the human Hv1

Figure 4.14: Evaluation of internal accessibility of Arg^{211} in the S4 helix of hHv1. All measurements are in inside-out patches at pH_o 7.0, pH_i 6.0. (A) Control families in the absence and presence of 0, 10, or 100 µM Zn^{2+} with pulses in 10 mV increments up to +60 mV. (B) Currents in an R211H patch during identical families of pulses in 10 mV increments up to +60 mV in the presence of 0 or 10 µM Zn^{2+}. (C) Three consecutive pulses to +60 mV in an inside-out patch with the R211H mutant, first in the presence of 10 µM Zn^{2+} (red), and then with the addition of EGTA during the pulse (green), and finally in the absence of Zn^{2+} (black). The reversal of Zn^{2+} effects shows accessibility of R211H in the open state. For all parts, V_{hold} = -60 mV.

4.4 Discussion

4.4.1 Phylogeny

The results that place the Hv1/C15orf27/VSP subfamily of VSDs in the same main branch as the NaV/CaV- type VSDs are robust to different tree construction methods, changes in the included taxa, and even different underlying multiple sequence alignments (Fig. 4.1). These results indicate that the VSD subfamily that is not associated with a canonical ion pore, typified by Hv1, is more closely related to the NaV/CaV family of VSDs and appears to have split from the Kv family of VSDs at least a billion years ago. Sequence similarity between the Hv1/C15orf27/VSP subfamily and the NaV/CaV subfamily
thus most likely represents leftover relationships from “history” rather than positive selection. In this view, the sequence similarity is evidence of the common ancestry of these groups but does not reflect any particular property that evolved in both groups. Despite the presence of Kir-type channels in bacteria, which are homologous to the S5-S6 TM helices of voltage-gated cation channels, the lack of discernible “poreless” VSDs in bacteria makes any assignment of the association of the ancestral VSD with the ancestral ion pore extremely speculative.

The gold standard for a “correct” sequence alignment is one derived from a structural superposition. The sequence alignment in S4 of VSDs is challenging precisely because this gold standard cannot be applied from sequence information alone; a particular arginine within the same VSD is thought to occupy different positions relative to the rest of the VSD, depending on whether the VSD is “activated,” “inactivated,” “resting,” or at some intermediate position. This point has been exemplified dramatically in the presumed inactivated-state crystal structures of Na\textsubscript{V}Ab\textsuperscript{82} and Na\textsubscript{V}Rh\textsuperscript{83} that appeared while this paper was in preparation. S4 of the Na\textsubscript{V}Rh crystal is displaced by one full helical turn compared with S4 of Na\textsubscript{V}Ab when using the conserved phenylalanine of the charge transfer center as a reference point;\textsuperscript{83} this difference is presumed to represent a different position of S4 rather than a different alignment of the four arginines in each S4. The sequence alignment methods that incorporate structural information, as do both HHpred and PromalS3D, provide the crucial comparison that leads to the “correct” representation of the registry of the arginine residues in S4 in the resulting multiple sequence alignments: constraining the alignment to the “activated”-state crystal structures, for example, provides the “activated”-state registry and thus an alignment that will be preserved in the other states of the VSD characterized by S4 motion. The alignment of the S4 arginine in the multiple sequence alignment shown in Fig. 4.1 is consistent with the structural superposition of the presumed active Na\textsubscript{V}Ab and paddle-chimera crystal structures, and also with a recent alignment of Na\textsubscript{V}/Ca\textsubscript{V} and K\textsubscript{V} VSDs obtained by using HHpred.\textsuperscript{90} The alignment used here also includes the H\textsubscript{V}1/C15orf/VSP subfamily of VSDs and provides a different alignment in S4 than has been used previously in other credible and useful structural analyses.\textsuperscript{73,165,275,286,308,315} Only two previous studies aligned R1 in H\textsubscript{V}1 with R3 in Shaker.\textsuperscript{306,307} These considerations led us to develop a new homology model of hH\textsubscript{V}1 and to compare it to another model based on a previous, widely accepted alignment. Alignments with significant differences in the placement of charged residues outside of S4 seem implausible on the basis of alignment scores. Only one such alignment\textsuperscript{308} places a charged residue (H\textsubscript{V}1 D185) more than one residue away compared with the alignment we use here. The latter alignment was obtained before the sodium channel crystal structures were available and disagrees at that position with the structurally informed alignment obtained by Yarov-Yarovoy \textit{et al.}(2012).
4.4.2 Computational validation of the homology models

MD simulations have been used previously both as assessment and refinement tools for homology models of proteins. For soluble proteins, these simulations typically involve all-atom explicit models in water; gradual model refinement using structural restraints; a simulation time spanning hundreds of nanoseconds to microseconds; and the analysis of ensemble structural properties, such as the conformational drift from the starting conformation, conservation of the secondary structure, and nonbonded residue contacts within the model and its templates. Similar efforts for homology models of membrane proteins, which must take into account the anisotropic membrane environment, have until now been limited to comparatively shorter simulation times. These simulations have used either a lipid bilayer or biphasic membrane mimetic n-octane slab, which have typically been limited to a single simulation. The advantage of using n-octane is that its low viscosity speeds up the conformational fluctuations and relaxation of the embedded protein compared with a lipid bilayer. This property results in a wider conformational ensemble of the protein within a relatively short simulation time. This improved sampling efficiency, together with multiple repeats, was exploited in the current protocol to achieve a broad conformational sampling in a computationally efficient manner. In addition, we used a clustering analysis to identify a consensus conformation for the protein. To our knowledge, this work is the first comparative homology modeling MD simulation based on a consensus-derived three-dimensional model for a membrane protein. The general protocol presented here can readily be applied to evaluate the quality of homology models of other membrane proteins.

Because homology modeling is derived from little or no experimental evidence, and particularly in light of the low sequence identity between crystallized VSDs and hHV1, it is paramount to assess the quality of the model before examining the structural basis of hHV1 function. Although other homology models of hHV1, including another generated by our group, have been proposed recently, the present study is the first to offer a comparative assessment of the quality of homology models. In this study, two principal criteria were used to assess and compare the structural stability of the models to its templates: overall structural drift of the TM segments and retention of α-helical structure from the starting conformation. A key aspect of our approach is the evaluation of statistical properties from multiple replicas and subsequent clustering analysis. It is noteworthy that in some cases (Fig. 4.6), the probability distribution of R2D and R3D models almost overlaps. With clustering, however, this source of ambiguity is resolved. Thus, our results indicate that combining massively repeated simulations with clustering made it possible to identify R2D as the better model for hHV1. This conclusion appears to be corroborated by the subsequent analysis of salt-bridge formation as well as hydration and solvent
accessibility (see below).

The validity of using a biphasic membrane mimetic (or simple n-octane slab) for the purpose of estimating the stability of membrane proteins in general, and of VSDs in particular, is demonstrated by the results obtained for the VSD templates and the control simulation in a lipid bilayer, which consistently retained their crystallographic conformation within a mere 0.2 nm RMSD over 100 ns MD trajectories. This value is consistent with what would be expected in lipid bilayers.\textsuperscript{322} Moreover, the analysis reported above shows that the magnitude of structural deviations from the starting point is correlated to the resolution of the crystallographic structure, which demonstrates the sensitivity of the approach to the quality of the structural model. As in any molecular simulation study, it is not possible to predict whether or not conformational changes not observed in a given simulation may occur on longer time scales. As such, we cannot dismiss the possibility that longer simulations of our systems may lead to conformational differences caused by approximations in the membrane representation or by other factors. Nevertheless, the simulations reported in this study represent over 45 $\mu$s of sampling, which is, to our knowledge, orders of magnitude longer than any previous validation study of a homology model of a membrane protein.

Although the stability of both homology models is lower than that of the lowest resolution template (the VSD from sodium channel Na$\textsubscript{V}$Ab) in terms of both the magnitude of deviation from the starting structure and the heterogeneity of the ensemble of structures generated by the simulations, they were clearly differentiated by the analysis. The resulting structural ensemble is much better clustered around a consensus structure for the R2D than for the R3D model, as manifested by the conserved location of the peak in the RMSD distribution, independent of clustering cutoff (Fig. 4.6). Thus, although the ensemble of structures obtained at 300 K deviates from the initial homology R2D model, these structures retain a high degree of similarity with one another.

These findings indicate that, unlike the R3D model, the R2D model corresponds to a unique conformational basin that is consistently represented by the thermal ensemble obtained at 300 K. Although the quality of the model is evidently inferior to that of the crystallographically derived structural models of the VSD templates from K$^+$ and Na$^+$ channels, the self-similarity of structures making up the R2D conformational ensemble approaches that obtained for the VSD of Na$\textsubscript{V}$Ab (Figs. 4.5 and 4.6). Moreover, the high degree of structural stability of salt bridges in the simulations of R2D is consistent with the proposal that charge pairing contributes to the stability of VSDs.\textsuperscript{270,271,273} As shown in Fig. 4.8, the population of ion pairs making up the salt-bridge networks in the conformational ensemble of R2D is essentially independent of cluster cutoff, whereas alternative salt bridges become populated with increasing cutoff values in the R3D ensemble. These results do not prove that the R2D model is accu-
rate. Nevertheless, the above results, collectively, support the validity of our systematic, ensemble-based computational approach in discriminating between two possible homology models.

4.4.3 Hydration and mechanism of proton conduction

The above structural analysis leads to various considerations regarding the putative functional mechanism of $hH_{V1}$. Importantly, the overall structural features of our R2D model are a priori consistent with the properties required of a selective ion channel in its open state. Although the helix bundle retains the overall “hourglass” shape of the VSDs of $K^+$ and $Na^+$ channels, the four helices of $hH_{V1}$ define a narrow, mostly water-filled pore. This narrow pore consists of a 1-nm-long constriction defined by two pairs of highly conserved residues, Arg$_{211}^{-}$–Phe$_{150}$ and Arg$_{208}^{-}$–Asp$_{112}$. In the above simulations, the residues lining this putative ion-translocation pathway undergo structural fluctuations resulting in the transient formation of a water chain throughout the pore (Fig. 4.11 D). Although the presence of a narrow, intermittent chain of water molecules that could serve as a putative proton pathway does not prove that this structure would be selective or even permeable to protons, these features can conceivably play a role in selective proton permeation, as discussed below.

The presence of a salt bridge involving D112 at the narrowest part of the channel suggests possible mechanisms for proton transport and selectivity. Either protons are translocated throughout the entire length of the pore in a water-mediated Grotthuss-like mechanism, as described in simpler, nonselective cation channels such as gramicidin, or one or more titratable groups of the protein act as proton-relay groups. Because proton transport in hydrogen-bonded water wires is very efficient, and because D185N and D185A retained proton selectivity, it is likely that protein-mediated proton relay, if it occurs, would be restricted to the D112-R208 bottleneck. Even if these residues are directly involved in proton relay, the ion pair would have to dissociate, if only transiently, to enable the changes in protonation states and in the organization of the hydrogen-bonded network inherent in vectorial proton movement. As a result, the fact that the ion pair undergoes reversible opening transitions in our simulations supports proton translocation a priori, although it does not allow us to discriminate between water-only and protein-mediated proton relay mechanisms.

Ultimately, discriminating between these two mechanisms will require detailed examination of the energetics and kinetics for the movement of an excess proton through the pore. In principle, both mechanisms may be invoked in the mechanism of proton selectivity. The rapid positional fluctuations of Arg$^{208}$ would be expected to lower the $pK_a$ of Asp$^{112}$ when they are close together, resulting in proton ejection. Alternatively, it is conceivable that fluctuations resulting in the formation of a transient water
column (~10% of the time) might allow the rapid translocation of protons but not the slower diffusion of other cations. The gramicidin channel contains a water wire essentially continuously, yet its proton conductance, extrapolated to neutral pH, is an order of magnitude lower than that of hHv1. The conductance of hHv1 might be increased, compared with that of gramicidin, by (a) a shorter narrow segment or (b) its negatively charged inner vestibule (Fig. 4.12). There is a possible parallel between the largely discontinuous water column in hHv1 and suggestions that other ion channels may “close” by hydrophobic collapse. However, our model is exclusively of the open conformation of hHv1; based on existing evidence, we imagine that the closed conformation differs significantly in the relative position of S4. In addition, interruption of the water wire by Asp112–Arg208 salt bridge formation occurs on a nanosecond timescale, much faster than macroscopic closing of the channel, which has time constants of milliseconds to seconds. Future simulation studies of hHv1 in a lipid bilayer, including free-energy calculations for the translocation of protons and other ions, will be performed in an effort to relate the microscopic structure of the channel to the mechanism of proton translocation and selectivity.

Nevertheless, it is pertinent and informative at the present model-validation stage to consider the competence of the model in terms of cation permeation by examining whether the electrostatic environment of the pore is compatible with the movement of a positive point charge (Fig. 4.12). A previous computational study of aquaporin has shown that the static field contribution of the electrostatic potential for the movement of a positive point charge is similar to the potential of mean force for the translocation of an excess proton in the single-file region of the pore. This agreement suggests that the free energy for the movement of H+ in the narrow region is largely determined by the distribution of polar and charged groups of the channel. As such, the static field profile affords a first glimpse into the energetics of ion permeation in hHv1. The cancellation of the static field in the bottle-neck region of the channel suggests that there is no intrinsic preference for cations over anions in the WT channel, and that both anions and cations other than H+ could be excluded by the same mechanism. In contrast, the large static field barrier induced by the neutralization of the charge of D112, which results in an excess positive charge (that of R208) in the bottle-neck, is consistent with the experimentally observed anion selectivity of neutral mutants of D112. On the one hand, reversing the sign of the probe point charge yields the opposite static field profile, turning the broad 10 kcal/mol energy barrier (Fig. 4.12) into a well of the same depth (not depicted). On the other hand, adding the desolvation penalty for bringing a finite-size ion from bulk water into the narrow part of the pore, a quantity that is always positive in sign, would further disfavor cations while bringing the electrostatic profile of anions closer to zero, consistent with anion selectivity.
4.4.4 How far does S4 move during gating?

In other voltage-gated ion channels, the membrane electrical field is concentrated and much of it drops across a narrow region that is at the focus of an hourglass of water. The Arg residues are thought to move sequentially past this constriction, altering their accessibility from inside to outside as they do so, and thus transferring gating charge across a large fraction of the membrane field at the expense of a relatively short actual movement. Recent evidence indicates that three or four S4 Arg residues move past the charge transfer center during Na⁺ or K⁺ channel opening and inactivation. Given the remarkable similarities between hHV1 and the VSD of other voltage-gated ion channels, the movement of S4 might be expected to be similar. However, our preferred homology model (R2D) suggests that in hHV1, R3 (R211) remains internal to F150, the external delimiter of the charge transfer center past which the S4 arginines of the VSD are envisioned to move during gating. Consistent with this prediction, when the Arg residues in S4 of hHV1 were mutated to His, the two outermost positions (R205H and R208H) were accessible to the external solution in the open state, but the third, R211H, was not. R211H was accessible to the internal solution even in the open state. The accessibility determined here with Zn²⁺ is somewhat greater than in previous studies using the MTS reagents as probes. PEGylation assays indicated that R3 of mHV1 was fully accessible to the internal solution, although the channel was presumably in the closed state. R3 in CiHV1 was accessible to internally applied MTS reagents in closed but not open states. The latter study reported external accessibility of R1 in open but not closed CiHV1 channels. R2 was inaccessible in both studies. The higher accessibility of the S4 Arg residues observed for Arg→His mutants, together with the higher Zn²⁺ affinity than that typically observed for binding to imidazole, may reflect favorable contributions of the external and internal charge networks identified here (Figs. 4.9 and 4.12) to coordinating Zn²⁺ at these intimate locations.

The conclusion from our accessibility studies that R211 remains internal to the charge transfer center in the open state is compatible with the finding that proton channel gating and selectivity persisted even when the C terminus was truncated, removing R3 altogether. The lowest energy structure of NaChBac predicted by also supports the internal accessibility of its R4 (the equivalent of R3 in our alignment of hHV1) in closed states and at least transiently in open states, although 9 of 10 low energy structures show R4 of NaChBac making contact with the extracellular charged nexus. The recent structure of NavRh shows that its R4 can ratchet “up” even one more turn in a presumably inactivated state, whereas our accessibility data suggest that the homologous position in hHV1 does not move past the charge transfer center, Phe. Perhaps the smaller number of charged groups on S4 of hHV1 prevents the full extent of movement predicted in NaChBac and NavRh, keeping R3 of hHV1 internally accessible.
Our results appear to limit the outward movement of S4 and suggest that S4 movement during gating is more restricted in hHV1 than in the K+ channel VSD. A speculative explanation for this difference is that unlike K+, Na+, and Ca2+ channels, hHV1 lacks a separate pore domain. The large displacement of S4 that is needed to open other channels is evidently not needed to produce a proton pathway.

These conclusions depend on the assumption that the mutations introduced here do not drastically affect the position of S4 in open or closed channels. Because in all cases charged residues were mutated, there is a strong possibility that the lowest energy position of the mutated residues will differ at least to some extent from the WT channel. On the other hand, the Arg→His mutation does not necessarily change net charge. Speaking against gross structural changes, the Arg→His mutants displayed relatively normal gating and were all proton selective. Although hHV1 appears to function as a dimer that exhibits cooperative gating, the present study does not address gating in general or the cooperative gating of the dimer. The main unique features of hHV1 proton-specific permeation and ΔpH-dependent gating occur both in monomeric and in dimeric hHV1 constructs. Nevertheless, we cannot rule out the possibility that dimeric channels might differ in ways that we cannot now predict.

### 4.4.5 Comparison with Starace and Bezanilla’s His scanning studies.

Our Arg→His mutations mirror those performed in the Shaker K+ channel VSD by Starace and Bezanilla, who mutated each of the first four Arg residues of S4 to His. The picture that emerged was of aqueous crevices on either side of a narrow bridge in the VSD that interrupts aqueous access from one side of the membrane to the other. The probes in these studies were effectively water molecules or, more accurately, hydronium ions. The four mutations (numbering the Arg residues starting from the outside) produced a hyper-polarization-activated proton channel (R1→His), a proton carrier (R2→His or R3→His), or a depolarization activation proton channel (R4→His). In the cases of R1→His or R4→His, the appearance is that in closed or open states, respectively, the His is positioned at a constriction where it is accessible to both sides of the membrane simultaneously. Strictly defined, “accessible” means capable of being protonated by H3O+ or donating a proton to H2O to form H3O+. Based on these studies, one might expect the R205H mutant of hHV1 to be a hyperpolarization-activated proton channel. The family of currents in Fig. 4.13 D shows that this expectation was not met. There is no significant inward current at negative voltages. We might further expect R208H to be a proton carrier, shuttling protons across the membrane at intermediate voltages where open↔closed transitions occur frequently. Even if this prediction were correct, it would be very difficult to distinguish H+ efflux mediated by channel versus carrier modes. At positive voltages, hHV1 functions as a proton channel (in contrast to the K+ channel
VSD that does not normally conduct current), and the rate of transport by a shuttle mechanism is unlikely to be as great as the flux in channel mode. Finally, we might predict that R211H would conduct protons at positive voltages. Because hHV1 does this already, we can think of no way to evaluate this possibility. In summary, the only definitive conclusion from the properties of the Arg→His mutants of hHV1 is that R205H does not act as a hyperpolarization-activated proton channel at negative voltages. Several interpretations are possible. R205H may not be located at the constriction of hHV1 in the closed state. Alternatively, R205H may be located at a constriction of hHV1 in the closed state, but, the geometry of the constriction does not allow $H_3O^+$ access from both sides. In any case, hHV1 appears to differ significantly from the Shaker VSD.

In summary, our model indicates that the open state of hHV1 is stabilized by an external charge cluster that includes a prominent Asp$^{112}$-Arg$^{208}$ salt bridge, and by a large internal charge network. Both model and His scanning indicate that the third S4 Arg, Arg$^{211}$, remains accessible to the internal solution in the open state. This result may be understood in light of hHV1 having only three Arg residues in S4, in contrast with approximately seven cationic residues in many other voltage-gated ion channels. In K$^+$ channels, after R1-R4 ratchet through the charge transfer center during opening, several cationic residues remain inside to interact with acidic groups. If S4 in hHV1 ratcheted up so that all three Arg residues in S4 were above the constriction at F150, none would remain inside to interact with internal acidic groups. Here, we propose that the middle Arg (R208) is poised at Asp$^{112}$ in the open state, with R1 outside and R3 inside, both able to interact with external and internal charge clusters, respectively, and stabilize the open state.

### 4.5 Conclusion

Human Hv1 catalyzes proton transport across cell membranes. In this study, we propose a homology model of the proton-conducting state of hHV1 guided by its structural homologous, VSDs of voltage-gated K$^+$ and Na$^+$ channels. The features of the model were validated using massively repeated unrestrained molecular dynamics simulations and tested against experimental accessibility experiments. Consensual structure produced by MD displayed stable, mostly $\alpha$-helical four TM regions (S1-S4) and a centrally located narrow pore delimited by well-defined external and internal salt-bridge networks each buried in a water crevice, similar to its homolog templates. Experimental data revealed the accessibility of the 3 Arg residues on S4 helix to one the 2 water crevices that is accurately predicted by the model. Furthermore, the electrostatic properties of the open-state model are compatible with both proton translocation in wild-type and anion selectivity of the mutant model where Asp$^{112}$ on S1 is neutralized. In addition,
the formation of a transient water-wire in the wild-type model supports, in priori, water-mediated Grotthuss-like proton conduction mechanism. Collectively, we present a structural model for hHV1 that is statistically and experimentally validated, which can be used to further the elucidation of molecular mechanism of proton translocation and charge selectivity.
Chapter 5

Peregrination of the selectivity filter delineates the pore of the human $H_V1$

The contents of this section were adapted from an article published in the *Journal of General Physiology*.


*Contributions:*
Deri Morgan, Boris Musset, Vladimir V. Cherny and Thomas E. DeCoursey conducted the research relating to electrophysiology and wrote the relevant section.
Kethika Kulleperuma conducted MD simulations and wrote the relevant sections.
Susan ME Smith made the constructs and wrote the relevant sections.
Régis Pomès provided guidance for simulation design, analysis and editorial input.
Abstract

Extraordinary selectivity is crucial to all proton-conducting molecules, including the human voltage-gated proton channel (hHV1), because the proton concentration is >106 times lower than that of other cations. Here we use “selectivity filter scanning” to elucidate the molecular requirements for proton-specific conduction in hHV1. Asp112, in the middle of the S1 transmembrane helix, is an essential part of the selectivity filter in wild-type (WT) channels. After neutralizing Asp112 by mutating it to Ala (D112A), we introduced Asp at each position along S1 from 108 to 118, searching for “second site suppressor” activity. Surprisingly, most mutants lacked even the anion conduction exhibited by D112A. Proton-specific conduction was restored only with Asp or Glu at position 116. The D112V/V116D channel strikingly resembled WT in selectivity, kinetics, and ΔpH-dependent gating. The S4 segment of this mutant has similar accessibility to WT in open channels, because R211H/D112V/V116D was inhibited by internally applied Zn2+. Asp at position 109 allowed anion permeation in combination with D112A but did not rescue function in the nonconducting D112V mutant, indicating that selectivity is established externally to the constriction at F150. The three positions that permitted conduction all line the pore in our homology model, clearly delineating the conduction pathway. Evidently, a carboxyl group must face the pore directly to enable conduction. Molecular dynamics simulations indicate reorganization of hydrogen bond networks in the external vestibule in D112V/V116D. At both positions where it produces proton selectivity, Asp frequently engages in salt linkage with one or more Arg residues from S4. Surprisingly, mean hydration profiles were similar in proton-selective, anion-permeable, and nonconducting constructs. That the selectivity filter functions in a new location helps to define local environmental features required to produce proton-selective conduction.

5.1 Introduction

Voltage-gated proton channels (Hv1s) enable phagocytes to kill pathogens and release histamine, and airway epithelia to control surface pH, as well as enable sperm motility and capacitation, and B lymphocyte signaling, and may exacerbate breast cancer metastasis and ischemic brain damage. All of these functions are predicated on the proton specificity of Hv1. The low concentration of H+ in biological fluids means that extraordinary selectivity is necessary even to ensure that H+ is the main conducted species. In fact, proton selectivity in Hv1 appears to be perfect.

An acidic group in the middle of the S1 transmembrane segment is critical to the proton specificity of Hv1 and is provided by Asp112 in human Hv1 and Asp51 in Hv1 from a dinoflagellate, Karlodinium.
veneficum.\textsuperscript{307} Despite only 15\% amino acid identity of the proteins, the conservative Asp→Glu mutation preserved proton specificity, whereas Ser, Ala, or His substitution for Asp at this position resulted in anion permeability in both species, strongly suggesting that the selectivity mechanism is widely conserved evolutionarily. The presence of an Asp facing the pore is not sufficient, however, because Asp\textsuperscript{185} can be neutralized without compromising proton selectivity, and does not preserve selectivity when Asp\textsuperscript{112} is neutralized.\textsuperscript{94} Other molecular elements that may be required are not known. Our homology model indicates that the second of three Arg residues in the S4 segment, Arg\textsuperscript{208}, forms a salt bridge with Asp\textsuperscript{112}, and that the resulting charge compensation is important for proton selectivity.\textsuperscript{164} To refine further the molecular requirements of the selectivity filter, we explore here the extent to which the critical Asp can be moved along the S1 segment. We find that an excellent proton channel is produced when Asp is shifted from position 112 to position 116. The mutant channel is proton specific, exhibits ΔpH-dependent gating characteristic of all Hv1s, and surprisingly, neutral amino acid substituents at this location produce anion permeability. Molecular dynamics (MD) simulations indicate that Asp\textsuperscript{116} forms a salt bridge with Arg\textsuperscript{205} and/or Arg\textsuperscript{208}, but the latter also pairs with Asp\textsuperscript{185}, reflecting reorganization of charge clusters in the mutant compared with WT channels. The results underline the importance of intramolecular charge compensation for proton selectivity.

5.2 Methods

5.2.1 Gene expression

Site-directed mutants were created using the Stratagene QuikChange (Agilent Technologies) procedure according to the manufacturer’s instructions. Transfection was done as described previously.\textsuperscript{164} Both HEK-293 cells and COS-7 cells were used as expression systems, the latter more frequently. In a previous study, we systematically compared the properties of hHv1 when expressed in these two cell lines and found no difference.\textsuperscript{267} Although currents that decayed at large positive voltages (presumed to be volume-regulated anion currents) were sometimes seen at the start of experiments, these disappeared over time. Occasional cells displayed a few unidentified single-channel currents superimposed on the macroscopic currents. The unitary conductance of hHv1 is just 140 fS at pH\textsubscript{i} 5.5;\textsuperscript{56} thus, visible unitary currents were considered extraneous and were ignored. No other voltage-or time-dependent conductances were observed under the conditions of this study. Thus, depolarization-activated time-dependent currents, which in many cells were orders of magnitude larger than any background currents, were assumed to reflect the transfected construct. Both cell lines sometimes exhibit small native Hv1 currents, which
could be distinguished from transfected mutant channels by their high \( \text{Zn}^{2+} \) sensitivity (see Results).

### 5.2.2 Electrophysiology

GFP-tagged proton channels were identified using inverted microscopes (Nikon) with fluorescence capability. Conventional patch-clamp techniques were used at 21°C or at room temperature (20-25°C).\(^{164}\) Bath and pipette solutions contained 60-100 mM of buffer, 1-2 mM \( \text{CaCl}_2 \) or \( \text{MgCl}_2 \) (intracellular [IC] solutions were \( \text{Ca}^{2+} \) free), 1-2 mM EGTA, and TMAMeSO\(_3\) to adjust the osmolality to \(~300\) mOsm, titrated with TMAOH. Buffers used were Mes at pH 5.5-6.0, HomoPIPES at pH 4.5, and PIPES at pH 7.0. For \( \text{Zn}^{2+} \) measurements, EGTA was omitted. Currents are shown without leak correction. Reversal potentials were corrected for measured liquid junction potentials. Unless stated otherwise, cells were held at a holding potential, \( V_{\text{hold}} \), before pulses and returned to \( V_{\text{hold}} \) after families of pulses.

Reversal potentials were determined by two methods, depending on the relative positions of \( V_{\text{threshold}} \) and \( V_{\text{rev}} \). For most constructs, \( V_{\text{threshold}} \) was positive to \( V_{\text{rev}} \), and the latter was determined by examining tail currents (e.g., Fig. 5.2 B). Because \( hHV1 \) currents were the only time-dependent conductance present, estimates of the amplitude and direction of current decay during deactivation were used to establish \( V_{\text{rev}} \).\(^{334}\) By this procedure, time-independent leak or other extraneous conductances do not affect \( V_{\text{rev}} \). For mutants in which \( V_{\text{threshold}} \) was negative to \( V_{\text{rev}} \) (for example, D112A/V116D in Fig. 5.2 C), it was possible to observe the reversal of the direction of currents activated during pulse families. Tail currents were not observed in nontransfected cells; for example, Fig. 5.1 B illustrates the absence of tail currents in a cell with the nonconducting D112V mutant.

### 5.2.3 MD simulations

MD simulations of the WT protein, single-point mutants D112V and D112S, and double-point mutants D112V/V116D (“VD”) and D112V/V116S (“VS”) were performed in a hydrated lipid bilayer based on the homology model constructed and validated in a recent study.\(^{164}\)

12 conformations of the WT protein with pore-associated water were used as initial structures. These snapshots correspond to the endpoint of 12 different 200-ns-long unrestrained simulations in a membrane-mimetic octane slab.\(^{164}\) Each conformation represents one of the three configurations of the D112-R208 salt bridge: bidentate and monodentate conformations (involving two or one hydrogen bond, respectively), and open, in order of increasing separation. A preequilibrated configuration of a 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) bilayer was obtained from a previous study.\(^{164}\) The OPLS-AA protein force field\(^{235}\) was mixed with the Berger lipid parameters\(^{335}\) by applying the half-\( \varepsilon \)
double-pairlist method\textsuperscript{288}. The TIP3P water model was used.\textsuperscript{245} InflatableGRO\textsuperscript{299} was used to embed the protein in the bilayer. The system was then hydrated, and 54 Na\textsuperscript{+} and 56 Cl\textsuperscript{−} ions were added to neutralize the charge of the system and yield an approximate ionic concentration of 500 mM. The resulting simulator cell consisted of 126 POPC and \~5,900 water molecules in a box of \~6.5 x 6.5 x 8 nm\textsuperscript{3}. The MD parameters used for this study are described elsewhere.\textsuperscript{164}

Each of the 12 WT systems was first energy-minimized using 50,000 steps of steepest descent, followed by an equilibration phase of 50 ns with position restraints on protein backbone and pore-associated water oxygen atoms. The production run consisted of 200-ns-long unrestrained simulations for each system. Snapshots of protein and pore-associated water molecules were selected from each of the 12 equilibrated WT protein systems at \(t=100\) ns to produce single- and double-point mutants. Mutations were introduced using an in-house script, followed by 1,000 steps of energy minimization. Asp\textsuperscript{112} and Val\textsuperscript{116} side chains were modified to Val, Ser, or Glu by either overwriting the WT heavy atoms or deleting some or both. Side-chain and backbone dihedral angles were checked after energy minimization. Extra ions were added to the solution as required to neutralize the system after mutations were introduced. Another 50,000 steps of energy minimization were performed before an equilibration phase consisting of an additional 25 ns with position restraints, as described above. 12 time trajectories of 200 ns differing in the initial conformation of the protein were generated for each mutant. The total production time was 12 \(\mu\)s. For the final analysis, eight replicas of WT and VD and seven replicas of VS, D112V, and D112S were selected after discarding replicas that displayed significant changes in secondary structure.

Snapshots saved every 20 ps during the last 100 ns of each selected production run were analyzed for each system. Molecular graphics were generated by VMD 1.8.7,\textsuperscript{300} and all trajectories were analyzed using Gromacs tools and in-house codes.

5.3 Results

We generated a series of mutants in which the Asp residue critical to proton specificity was effectively shifted up and down the S1 helix to each position from 108 to 118. In the initial series of experiments, we replaced Asp\textsuperscript{112} with Ala (D112A) to produce an anion-permeable channel,\textsuperscript{94} and then introduced a second mutation with the goal of restoring proton selectivity (“second site suppression”). Additional studies were done in the D112V background, which in a sense is more rigorous, because this single mutant does not conduct at all.\textsuperscript{94} All mutations were introduced into a Zn\textsuperscript{2+}-resistant background (H140A/H193A), so that spurious small native H\textsubscript{V1} currents often present in COS-7 or HEK-293 cells\textsuperscript{94} could be identified by their sensitivity to 10 \(\mu\)M Zn\textsuperscript{2+}. Distinct Zn\textsuperscript{2+}-insensitive currents were observed
only in mutants with Asp at positions 116 or 109 (D112A/V116D, D112A/V109D).

### 5.3.1 The second site mutation V116D restores proton selectivity to non-conducting or anion-permeable mutants

Fig. 5.1 shows that although the single-point mutation D112V eliminates current altogether, introducing Asp at position 116 restores robust proton current to the double mutant, D112V/V116D. Similarly, the D112A single mutant is anion permeable, but introducing Asp at position 116 restored proton-specific current to the double mutant, D112A/V116D.

**Figure 5.1:** The D112V mutation abolishes current, but V116D restores proton-specific current, an example of second-site suppression. Whole-cell currents at pH<sub>o</sub> 7.0 and pH<sub>i</sub> 5.5 during pulses in 10 mV increments up to the indicated voltages for WT (A), D112V (B), or D112V/V116D (C), all expressed in COS-7 cells. Holding potential (V<sub>hold</sub>) and pulse durations were -90 mV, 1 s (A), -40 mV, 3 s (B), and -60 mV, 2 s (C). In all cases, the voltage was returned to V<sub>hold</sub> after the pulse, which is why the tail currents are inward for A and outward for C. Cartoons in all figures indicate S1 and S4 helices, with color coding as follows: red, Asp or Glu; yellow, Val; blue, Arg; gray, other amino acids or nonpore-facing residues.

The proton selectivity of both double mutants, D112A/V116D and D112V/V116D, was confirmed by the proximity of their reversal potentials, V<sub>rev</sub>, to the Nernst potential for H<sup>+</sup>, EH (Fig. 5.2 A), over a wide pH range (pH<sub>o</sub> 4.5-7.5 and pH<sub>i</sub> 5.5-7.0). Fig. 5.2 B illustrates determination of V<sub>rev</sub> from tail currents in a cell with D112V/V116D channels at pH<sub>o</sub> 5.5 and 7.0, with pH<sub>i</sub> 5.5. As indicated by the arrows, V<sub>rev</sub> shifted from 0 to -77 mV, near EH of -87 mV. The D112A/V116D mutant activated in a more negative voltage range so that inward currents were observed negative to EH in families of currents. Fig. 5.2 C illustrates currents from pulses bracketing V<sub>rev</sub> that reveal an ~60 mV shift between pH<sub>o</sub> 5.0
and 6.0. In addition, substituting Na\(^+\), Li\(^+\), or K\(^+\) for TMA\(^+\), or Cl\(^-\) for CH\(_3\)SO\(_3\)\(^-\) had no effect on \(V_{rev}\) (Table 5.1). In summary, shifting Asp from position 112 to 116 moves the proton selectivity filter outward by one turn of the helix.

Figure 5.2: **Second site suppression in double mutants.** (A) Shifting the crucial aspartate from position 112 to 116 (V116D) restores proton selectivity to both the nonconducting D112V (red symbols) and the anion-permeable D112A (blue symbols) single mutants. Measurements in the same cell are connected by lines. For both, \(V_{rev}\) measured over a wide range (for D112V/V116D pH\(_o\) was 5.5, 6.0, 6.5, 7.0, 7.5, or 8.0, and pH\(_i\) was 5.5, 6.0, 6.5, or 7.0; for D112A/V116D pH\(_o\) was 4.5, 5.5, 6.0, 6.5, or 7.0, and pH\(_i\) was 5.5, 6.0, or 6.5) falls close to the Nernst potential for H\(^+\), EH (dashed green line). (B) Determination of \(V_{rev}\) from tail currents for D112V/V116D in a COS-7 cell with pH\(_i\) 5.5. Prepulses to +30 mV, pH\(_o\) 5.5, or to -40 mV, pH\(_o\) 7.0, activated the conductance, followed by repolarization to the indicated voltages in 10 mV increments, with the most positive labeled. (C) Determination of \(V_{rev}\) from current families for D112A/V116D in a COS-7 cell with pH\(_i\) 5.5. As indicated, \(V_{hold}\) was -60 mV (left) or -70 mV (right); the voltage was returned to \(V_{hold}\) after the pulses. Currents are shown during selected pulses that bracket \(V_{rev}\) in 10 mV increments. Inward current was activated negative to \(V_{rev}\), and the first outward current is labeled.
Table 5.1: **Impermeability of D112V/V116D and D112V/V116D/R211H channels to ions**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Ion</th>
<th>$V_{rev}$ mean $\pm$ SEM $(n)$</th>
<th>$V_{jet}$ correction</th>
<th>Corrected $V_{rev}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D112V/V116D</td>
<td>Na$^+$</td>
<td>$-1.8 \pm 1.9$ (4)</td>
<td>1.3</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>K$^+$</td>
<td>$-2.3 \pm 1.9$ (4)</td>
<td>4.8</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Li$^+$</td>
<td>$0.0 \pm 1.0$ (7)</td>
<td>-1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td></td>
<td>Cl$^-$</td>
<td>$-1.3 \pm 1.5$ (4)</td>
<td>-3.3</td>
<td>-4.6</td>
</tr>
<tr>
<td>D112V/V116D/R211H</td>
<td>Na$^+$</td>
<td>$-1.6 \pm 1.8$ (3)</td>
<td>1.3</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td>K$^+$</td>
<td>$-3.1 \pm 1.6$ (3)</td>
<td>4.8</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>Li$^+$</td>
<td>$1.5 \pm 1.4$ (7)</td>
<td>-1.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Cl$^-$</td>
<td>$3.0 \pm 2.6$ (3)</td>
<td>-3.3</td>
<td>-0.3</td>
</tr>
</tbody>
</table>

Measurements were made at symmetrical pH 7.0, with TMA$^+$ or CH$_3$SO$_3^-$ replaced by cations or anions at concentrations of $\sim$100 mM. The small changes in $V_{rev}$ were of the same order as variability and liquid junction potentials.

### 5.3.2 Asp supports current only when facing the pore

Surprisingly, moving Asp to positions other than 109 or 116 eliminated voltage-gated current altogether. All mutants were tagged with GFP, and transfected cells with green “halos” indicating membrane expression were selected under fluorescence for recording. Nonconducting Asp mutants included (mutation, number of cells): positions 108 (D112A/L108D, 10), 110 (D112A/V110D, 4), 111 (D112A/L111D, 4), 113 (D112A/A113D, 5), 114 (D112A/L114D, 4), 115 (D112A/L115D, 4), 117 (D112A/L117D, 7), and 118 (D112A/A118D, 4). Positions 116, 112, and 109 (Fig. 5.1, cartoon) all face the “pore” in the predicted open-state structure of hHV1.$^{164}$ Evidently, Asp at a nonpore-facing location fails to support conductance, confirming that positions 109, 112, and 116 line the pore of hHV1.

As a control for the possibility that nonconducting mutants did not fold properly, we created a single mutant, A113D, in which the native Asp$^{112}$ was preserved. If Asp at a nonpore-facing location caused the protein to misfold, A113D should not conduct. Instead, A113D displayed small voltage-gated currents that were kinetically different from WT currents but were unequivocally proton selective (Fig. 5.3). Thus, nonconducting mutants, including D112A/A113D, most likely were expressed but nonfunctional. If, despite the appearance of green protein in the membrane, some mutants misfolded, the fact remains that these proteins do not function as channels.
Figure 5.3: Introducing Asp at position 113, predicted by our model to be in a nonpore-facing location, results in membrane expression of a functioning proton-selective channel. The cartoon emphasizes that Asp$^{112}$ is still present. Families of currents are shown in a COS-7 cell in whole-cell configuration at pH$_o$ 5.5 (A) or 7.0 (B), with pH$_i$ 5.5, with pulses applied from $V_{hold}$ as labeled to the indicated voltages in 10 mV increments. Cells were returned to $V_{hold}$ after pulses. (C) Proton selectivity is shown by the proximity of $V_{rev}$ to $E_H$ (dashed line). Insets show $V_{rev}$ determination (left) at pH$_o$ 5.5 and pH$_i$ 5.5 by reversal of current during a family of pulses in 10 mV increments ($V_{hold}$ = -60 mV) and at pH$_o$ 7.0 and pH$_i$ 5.5 (right) by tail currents. $V_{rev}$ was measured at pH$_o$ 4.5, 5.0, 5.5, 6.5, 7.0, and 7.5, and at pH$_i$ 5.5 or 6.5.
5.3.3 At position 109, Asp plays a permissive role

Introducing Asp at position 109 into the nonconducting D112V background\textsuperscript{94} did not overcome the lack of conductance produced by the D112V single mutation (n= 8 cells). However, the D112A/V109D mutant exhibited distinct currents at pH\textsubscript{o} 5.5 and pH\textsubscript{i} 5.5 (Fig. 5.4 A), in contrast to the majority of mutants that did not conduct (D112A/L108D, D112A/ V110D, D112A/L111D, D112A/A113D, D112A/L114D, D112A/L115D, D112A/L117D, and D112A/A118D). Replacing CH\textsubscript{3}SO\textsubscript{3}\textsuperscript{-} with Cl\textsuperscript{-} increased the outward current (Fig. 5.4 C), reflecting Cl\textsuperscript{-} influx, and produced a large negative shift of \( V_{rev} \) (Fig. 5.4, B vs. D), confirming Cl\textsuperscript{-} permeability. The shift of \( V_{rev} \) when Cl\textsuperscript{-} replaced CH\textsubscript{3}SO\textsubscript{3}\textsuperscript{-} was \(-35.9 \pm 4.2 \text{ mV} \) (SEM; \( n = 5 \)), in the range reported for the D112A single mutant, \(-29 \text{ mV} \)\textsuperscript{94}. Evidently, introducing Asp at position 109 did not interfere with anion conduction seen in the D112A single mutant.\textsuperscript{94} Both the anion selectivity of D112A/V109D and the lack of conductance in D112V/V109D suggest that at position 109, Asp cannot mediate proton selectivity and has no discernible effect on the selectivity that is established elsewhere. The two positions where Asp produced proton selectivity are in the external vestibule in our model, outside the highly conserved charge transfer center delimiter Phe\textsuperscript{150}.\textsuperscript{274}

Figure 5.4: Moving aspartate from position 112 to 109 results in anion currents. Whole-cell currents in a COS-7 cell expressing D112A/V109D channels, all at pH\textsubscript{o} 5.5 and pH\textsubscript{i} 5.5, in symmetrical TMA\textsuperscript{+} CH\textsubscript{3}SO\textsubscript{3}\textsuperscript{-} (A and B) or with Cl\textsuperscript{-} in the bath (C and D). Pulses applied in 10 mV increments. \( V_{hold} \) was \(-40 \text{ mV} \) (A and B) or \(-60 \text{ mV} \) (C and D). Cells were returned to \( V_{hold} \) after pulses. \( V_{rev} \) determination from tail currents (B and D), with pulses in 10 mV increments.
5.3.4 The engineered Asp$^{116}$ proton channel functionally resembles WT

To what extent does moving the selectivity filter outward by one turn of the helix reinstate native hHv1 properties? The gating and pH dependence of the D112A/V116D and D112V/V116D mutants are illustrated in Fig. 5.5. The general appearance of the proton currents is unremarkable; the voltage dependence and gating kinetics are roughly similar to WT. In D112V/V116D, the time constant of tail current decay, $\tau_{tail}$, was $1.1 \pm 0.02$ s (mean ± SEM; n = 3) at -40 mV and pH 5.5 / / 5.5, which is close to the WT value of 0.81 s measured under the same conditions. 94 The activation time constant, $\tau_{act}$, was $2.3 \pm 0.3$ s (n = 3) at + 40 mV in D112V/V116D, compared with the WT value of 1.1 s. 94 Thus, marked changes in the gating of the mutant channels were not observed.

Figure 5.5: Voltage and pH dependence of gating of hHv1 mutants with the selectivity filter shifter from 112 to 116. Families of currents in a COS-7 cell expressing D112A/V116D channels at pH $i$ 6.5 and pH $o$ 7.5 (A), 6.5 (B), or 5.5 (C) are shown in 10 mV increments as labeled, from $V_{hold}$ -60 mV (A) or -30 mV (B and C). Cells were returned to $V_{hold}$ after pulses. (D) Current-voltage relationships in this cell. Currents were fitted to a rising exponential function and extrapolated to infinite time. Note inward currents at pH $o$ 5.5. (E) Conductance-voltage relationships for the same currents. Limiting slope conductance for the most negative voltages provides a gating charge estimate of 5.3-6.0 $e_0$. (F) Regulation of voltage gating by $\Delta$pH. $V_{threshold}$ is plotted against $V_{rev}$ measured in the same cell and solution. Lines show linear regression fits defined by $V_{threshold} = 0.64 V_{rev} + 0$ mV (D112V/V116D; red circles) or $V_{threshold} = 0.63 V_{rev} -26$ mV (D112A/V116D; blue diamonds).
A characteristic property shared by all known HV1 is tight regulation of the position of the $g_{H-V}$ relationship by the pH gradient, $\Delta p\text{H}$.$^{52,53}$ Because the mechanism remains mysterious, we examined this property. As evident in Fig. 5.5 (D and E), changes in $pH_o$ produced shifts in both $I_{H-V}$ and $g_{H-V}$ curves of roughly 40 mV/U, similar to the shifts observed in native and WT $hHV_1$s.$^{72,332}$ Plots of $V_{threshold}$ versus $V_{rev}$ for D112V/V116D (Fig. 5.5, D and E, red) and D112A/V116D (blue) had slopes (0.64 and 0.63, respectively) similar to those observed previously for WT $hHV_1$s.$^{332}$ The D112A/V116D mutant activated at voltages 26 mV more negative than D112V/V116D, suggesting relative stabilization of the open state by Ala versus Val at position 112. As a consequence of this shift, $V_{threshold}$ in D112A/V116D was often negative to $E_H$ so that inward currents were observed. Nevertheless, $\Delta p\text{H}$-dependent gating was preserved in both double mutants.

5.3.5 Mutations at position 116 (in D112V channels) mimic mutations at 112

That proton selectivity was restored to D112V and D112A mutants by introducing Asp at position 116 suggests that the intramolecular interactions at both positions that contribute to proton selectivity are similar. To explore the extent of equivalence of these positions, we compared effects of several other mutations at 116 with those at 112. Fig. 5.6 shows that robust proton-selective currents were observed when Glu replaced Asp at position 116 (in the D112V background), just as with replacement of Asp by Glu at position 112 (Musset et al., 2011). We found previously that neutral mutants of Asp112 were anion selective.$^{94}$ Astoundingly, both Asn and Ser at position 116 produced anion-permeable channels in the D112V background, which itself does not conduct. Fig. 5.7 illustrates a current family in a cell with D112V/V116S in symmetrical pH 5.5 TMACH$_3$SO$_3$ solutions (Fig. 5.7 A), with $V_{rev}$ near 0 mV (Fig. 5.7 B). Replacing bath CH$_3$SO$_3^-$ with Cl$^-$ increased the outward current (Fig. 5.7 C) and shifted $V_{rev}$ strongly negatively (Fig. 5.7 D): on average, by -37.0 ± 2.4 mV (mean ± SEM; $n=7$) for Asn and -35.8 ± 1.8 mV ($n=4$) for Ser. These values are similar to those obtained previously for single mutants D112N and D112S upon Cl$^-$ addition: -33.1 and -40.8 mV, respectively.$^{94}$ Thus, the introduction of Glu, Ser, or Asn at position 116 conferred permeability onto the non-conducting D112V mutant, in each case recapitulating the selectivity of the corresponding single mutations at position 112. Viewed in terms of the ability of a single-amino acid substituent to produce anion or proton selectivity, positions 112 and 116 were identical.
Chapter 5. Peregrination of the SF delineates the pore of the hHv1

Figure 5.6: **The D112V/V116E mutant is a functional proton-selective channel.** (A) Families of currents generated by D112V/V116E in a HEK-293 cell in 10 mV increments at pH_o 5.5 and pH_i 5.5 (A), and pH_o 5.0 and pH_i 5.5 (B). As indicated, \( V_{\text{hold}} \) was -60 mV, and the membrane was returned to \( V_{\text{hold}} \) after pulses. (C) \( V_{\text{rev}} \) measured at pH_o 5.0, 5.5, 6.5, 7.0, and 7.5, and at pH_i 5.5 or 6.5, indicates that D112V/V116E is proton selective. Inset shows tail currents at 10 (uppermost), 5, and 0 mV at pH_o 5.5 and pH_i 5.5. Proton selectivity is imparted by either Asp or Glu at positions 112 or 116, indicating that side-chain length is not critical for this function.

Figure 5.7: **The D112V/V116S mutant is permeable to \( \text{Cl}^- \).** Replacing CH_3SO_3^- with \( \text{Cl}^- \) increased outward current in families (A and C) and shifted the tail current reversal potential negatively (B and D). All measurements were in a COS-7 cell at symmetrical pH 5.5 with TMA^+ CH_3SO_3^- solutions or external TMA^+ Cl^- solutions, as indicated. Pulses were applied in 10 mV increments from -40 to 80 mV for families and in 10-mV increments for tail currents as indicated. The cell was held at -30 mV and returned to \( V_{\text{hold}} \) after pulses. The same calibration bars apply to families (A and C) and tail currents (B and D).
5.3.6 \( \text{Zn}^{2+} \) sensitivity of Arg→His mutants shows S4 position in open channels

It seemed possible that shifting the selectivity filter out-ward might alter the position of S4 in open channels. As a frame of reference, R211H currents were found previously to be sensitive to internal but not external \( \text{Zn}^{2+} \) during depolarizing pulses, suggesting that Arg\(^{211} \) remains internally accessible in the open state.\(^{164} \) The third Arg in S4 of the \( \text{Ciona} \) Hv1 was reported to be internally accessible in closed but not open channels,\(^{275} \) perhaps reflecting the large size of the MTSET probe. Internal accessibility of position 211 in the filter-shifted R211H/D112V/V116D mutant was evaluated by testing the \( \text{Zn}^{2+} \) sensitivity of inside-out patches. This construct was proton selective. In nine patches with R211H/D112V/V116D, 10 \( \mu \text{M} \) \( \text{Zn}^{2+} \) was introduced into the bath at pH\(_i\) 7.0, and test pulses were applied. The H\(^+\) current was inhibited by 90.0 ± 3.6\% (mean ± SEM), significantly (\( P < 0.0001 \) by unpaired \( t \) test) more than 29.7 ± 6.8\% in six control patches (D112V/V116D). These results show that the 211 position is accessible to internally applied \( \text{Zn}^{2+} \), but they do not distinguish whether the site is accessible in closed or open channels.

Accessibility of His\(^{211} \) in R211H/D112V/V116D channels in the open state was evaluated by adding \( \text{Zn}^{2+} \) or EGTA during pulses. In Fig. 5.8 A, shortly after the start of a pulse, 10 \( \mu \text{M} \) \( \text{Zn}^{2+} \) was introduced (red record), producing slowly progressing block. The subsequent pulse (Fig. 5.8 A, blue) illustrates the full extent of inhibition. Given that the probability of being open is high during large pulses—95\% at pH\(_i\) 5.5 and 75\% at pH\(_i\) 6.5\(^{56} \)—and that gating is slow, cycling of channels through closed states during the pulse seems unlikely. Because \( \text{Zn}^{2+} \) must diffuse through the unstirred volume at the tip of the pipette to reach the membrane, onset of block was slow and likely dependent on pipette and patch geometry, but was observed unequivocally in four patches during long pulses. The addition of EGTA during a pulse to remove \( \text{Zn}^{2+} \) resulted in rapid reversal of block (Fig. 5.8 B) in six patches. The application of 10 \( \mu \text{M} \) \( \text{Zn}^{2+} \) to inside-out patches from cells expressing the control construct D112V/V116D produced comparatively minor effects on the current (Fig. 5.8 C), reminiscent of its effects on native Hv1.\(^{314} \) Thus, His at position 211 in R211H/D112V/V116D was accessible to the internal solution even in the open state, indicating similar accessibility of the S4 Arg residues in open channels with the selectivity filter at either position 112 or 116.
Chapter 5. Peregrination of the SF delineates the pore of the hHv1

Figure 5.8: The third arginine position in the S4 segment is accessible to the internal solution in the open state. (A) The black trace shows proton current in D112V/V116D/R211H in an inside-out patch from a HEK-293 cell. 10 μM Zn²⁺ was introduced into the internal solution during the red trace, and blue is the subsequent pulse in the continued presence of Zn²⁺. (B) The blue record is from D112V/V116D/R211H in the presence of 0 μM Zn²⁺, and the red record shows recovery from block on the addition of EGTA shortly after the start of the pulse. Black represents after washout. (C) The control mutant, D112V/V116D, in a COS-7 cell exhibits weak IC Zn²⁺ sensitivity: black, before; blue, in 0 μM Zn²⁺; gray, after washout. All pulses are to +50 mV at pHᵢ 7.0 and pHₒ 7.0 in inside-out patches of membrane.

5.3.7 Introducing Asp into S2 or S3 did not support proton conduction

Given that Asp can produce proton selectivity at two locations on S1, we moved Asp to several positions in the S2 or S3 transmembrane segments, always in the nonconducting D112V background. We chose five locations in the outer vestibule (S143D, I146D, L147D, V178D, and S181D) that face the pore in our model and are located roughly between the levels of positions 112–116. Four failed to produce distinct current. D112V/S143D generated small Zn²⁺-insensitive currents that reversed at -17 ± 6.6 mV (mean ± SEM; n = 5) at pHₒ 7.0 and pHᵢ 5.5, well positive to E_H, which is -87 mV. In five cells, V_rev shifted by -24.3 ± 1.6 mV when CH₃SO₃⁻ was replaced with Cl⁻ at pHₒ 5.5 and pHᵢ 5.5, demonstrating Cl⁻ permeation. That Asp at position 143 (in the S2 segment) overcame the nonconduction of D112V is consistent with position 143 facing the pore, as predicted in our model, but in a location incompatible with its producing H⁺ selectivity.

5.3.8 MD simulations reveal significant differences in the electrostatic properties of mutant channels

To assess the structural impact of the filter shift, MD simulations were performed on the WT protein, single-point mutants D112V and D112S, and double mutants D112V/V116D (VD) and D112V/V112S (VS) in a hydrated lipid bilayer bathed in 500 mM NaCl, based on our homology model. The overall structure of the channel was preserved in the mutants. In particular, the average root-mean-square deviation between WT and mutants ranged from 2.0 Å (D112S) to 2.2 Å (D112V). In all five systems, the average axial position of helices S1–S4 relative to the rest of the bundle was 0.73 ± 0.06 Å, 0.34 ± 0.05 Å, -1.80 ± 0.16 Å, and 0.53 ± 0.06 AA (SEM), respectively, indicating that the registry of the
four helices was conserved in all the simulations (Fig. 5.9). Combining data from the five systems leads to axial distributions that fit a single Gaussian distribution for each of the four helices, S1–S4, with standard deviations $\sigma = 0.4$, 0.5, 0.7, and 0.4 Å, respectively, emphasizing the small amplitude of axial fluctuations (Fig. 5.10).

**Figure 5.9:** Axial distribution of each of the four helical segments, S1S4, in each of the five systems (WT, VD, VS, D112V, and D112S) accumulated from the MD simulations in a lipid bilayer (solid line). Gaussian fits are also shown (dashed lines) for each of the curves. The five systems retained highly similar helical arrangements, with only small deviations in the average position of each helix relative to the rest of the protein. In all five systems, the average position of helix S1 was $0.73 \pm 0.06$ Å (SEM between the five systems), whereas helices S2, S3, and S4 were located at $0.34 \pm 0.05$ Å, $-1.80 \pm 0.16$ Å, and $0.53 \pm 0.06$ Å, respectively. In other words, the registry of the four helices relative to one another was identical throughout the simulations of all five systems.
Figure 5.10: Axial distribution of S1-S4 from the five systems (colored lines) are superimposed, and a unique Gaussian fit is obtained for the combined data for each helix (salmon surface). The distribution of S1 fits a Gaussian distribution at 0.75 Å (standard deviation $\sigma = 0.4$). Likewise, the fits for helices S2, S3, and S4 were 0.3 ($\sigma = 0.5$), -1.8 ($\sigma = 0.7$), and 0.5 ($\sigma = 0.4$), respectively. The results emphasize the small amplitude of the axial fluctuations and show that S1 does not exhibit significant plasticity in the axial direction.

Furthermore, the average hydration profiles of the pore in WT and various mutants did not differ significantly from one another (Fig. 5.11). The pore is characterized by an hourglass shape with an ~13-Å-long narrow region comprising an IC bottleneck at Phe$^{150}$ and an extracellular (EC) bottleneck at Asp$^{112}$. It is noteworthy that average pore hydration was quite similar for proton-selective (WT, VD), anion-permeable (VS, S = D112S), and nonconducting mutants (VAL = D112V), indicating that the mean hydration profile is not a good predictor of selectivity.

Despite these overall similarities (Fig. 5.11), differences were evident in the local structure of the pore near the EC bottleneck (Fig. 5.12). In the WT (Fig. 5.12 A), the EC constriction usually consists of a salt bridge between Arg$^{208}$ and Asp$^{112}$ (or occasionally Asp$^{185}$), whereas the other charged residues in the EC vestibule, namely Asp$^{185}$ and Arg$^{205}$, usually form a spatially distinct ion pair. Consistent with our previous simulation study,$^{164}$ Asp$^{112}$–Arg$^{208}$ in WT is present as an ion pair most of the time but is occasionally disrupted by water molecules, resulting in the transient appearance of a water chain. In all four mutants considered, however, the absence of a charged side chain at position 112 led to the reorganization of the ionic network in the EC vestibule.
Figure 5.11: **Pore hydration is similar in WT and several mutant channels despite very different selectivity.** Average water density within a 0.7 nm radius of the mean axis of the pore is plotted, normalized to the bulk water density for 5,000 snapshots from each replica of different systems. The membrane boundaries are indicated by dashed lines, with the external surface to the right. The nadir is near Phe\textsuperscript{150} in all cases. Average axial water density for: (A) WT (proton selective) and D112V (VAL, nonconducting); (B) WT and D112V/V116S (VS) and D112S (S), two anion-permeable channels; (C) WT and D112V/ V116D (VD), of which both are proton selective.
In the proton-selective VD mutant (Fig. 5.12 B), various arrangements of ionic networks involving between two and four charged side chains from helices S1 (Asp<sup>116</sup>), S3 (Asp<sup>185</sup>), and S4 (Arg<sup>205</sup> and Arg<sup>208</sup>) were observed. Fig. 5.13 illustrates the most frequent configurations adopted by Asp<sup>116</sup>. Most of the time (∼82%), it interacts with Arg<sup>205</sup> (Fig. 5.13 A) or Arg<sup>208</sup> (Fig. 5.13 D) or both (Fig. 5.13 C). Similarly, in WT hHv1, Asp<sup>112</sup> was engaged in salt linkage 90% of the time but almost exclusively with Arg<sup>208</sup>. However, in WT hHv1, a continuous water chain was present only when the Asp112Arg<sup>208</sup> salt bridge was broken, which occurred ∼10% of the time. In contrast, water pathways were observed in all configurations of the VD mutant, reflecting the greater width of the pore at this level compared with position 112. That an aqueous pathway is not predictive of proton selectivity is not surprising given the example of aquaporin channels that conduct water at high rate but are impermeable to protons. In contrast with WT, pore hydration is not significantly modulated by the configurations of Asp<sup>116</sup> in the VD mutant.

Figure 5.12: The EC salt-link network realigns in mutants. Representative snapshots of ionic networks in the WT and mutants, with the external end up. (A) WT protein in contact (left) and water-mediated (right) states of the D112R208 ion pair. (B) VD mutant, with R208 participating in (left) and free from (right) the EC salt-link network. (C) VS mutant. (D) D112S mutant. (E) D112V mutant. Acidic and basic side chains in the external and internal funnel, together with side chains of residues 112, 116, F150, and R211 are shown in licorice representation together with ribbon traces of the four α-helical transmembrane regions: red, S1; yellow, S2; green, S3; blue, S4. Salt bridges are shown as orange lines. The volumetric surface of water within the pore is shown with a water radius of 0.14 nm.
Chapter 5. Peregrination of the SF delineates the pore of the hHV1

Figure 5.13: **Several configurations of Asp116 in the D112V/V116D mutant.** Two-dimensional histogram of the distances from the center of charge of D116 to that of R205 ($d_1$) and R208 (itd$_2$), respectively, with increasing probability of ion pairing from purple to green. Snapshots surrounding the graph illustrate each type of interaction circled in the graph: (A) linked D116-R205 pair, 37% of the time; (B) open (no salt bridge), 18%; (C) D116 linked to R205 and R208 simultaneously, 33%; (D) D116-R208 salt bridge, 12%.

In particular, Arg$^{208}$ was observed to form ion pairs with Asp$^{116}$, Asp$^{185}$, or neither, or both. In contrast to WT, the local hydration of the constriction at the bottom of the EC funnel does not depend on whether or not Arg$^{208}$ is engaged in ion pairing (Fig. 5.12 B). Furthermore, our models predict the presence of water pathways in the anionic mutants D112S and VS (Fig. 5.12, C and D), and even in the nonconducting D112V mutant (Fig. 5.12 E). In all systems considered, the channel contains a narrow bottleneck $\sim$0.5 nm in length ($$-0.68 < z < -0.18$$ nm) between the tips of the EC and IC funnels. Although this bottleneck is lined with nonpolar residues including Val$^{109}$, Phe$^{150}$, Val$^{178}$, and Val$^{209}$, it contains a single file of two water molecules at least 75% of the time, consistent with a putative ion
pathway between the EC and IC funnels.

As a first step toward characterizing the energetics of ion permeation in the channel, we computed the static-field energy for a virtual positive point charge along water pathways spanning the length of the channel. Fig. 5.14 A shows that D112V, D112S, and VS mutants contain an electrostatic barrier opposing the movement of a cation through the EC side of the bottleneck, which is lacking in the WT. In contrast, the energetic properties of the VD double mutant depend on the local arrangement of ionic residues in the EC funnel. Specifically, a static-field barrier to cation movement is present at the EC bottleneck when Arg\textsuperscript{208} does not take part in any salt link (Fig. 5.14 A, “VD unpaired”), but this barrier is reduced by at least one half when the guanidinium group of Arg\textsuperscript{208} is paired with the carboxylate group of either Asp\textsuperscript{116} or Asp\textsuperscript{185}, or both (Fig. 5.14 B, “SL,” salt linkage). In the latter cases, the static-field profile becomes comparable to that of WT. Intriguingly, the electrostatic profile seems to be relatively insensitive to the nature of the ionic pairing of Asp\textsuperscript{116} (illustrated in Fig. 5.13).

Figure 5.14: Effect of the charge distribution of the channel on the energetics of ion translocation in WT and mutant channels. Orientation is as in Fig. 5.11, with the external solution to the right. Static-field energy for the transfer of a positive point charge in (A) WT, D112V, VD when R208 is unbound to the EC salt-link network, and two anion-permeable channels, VS and D112S. (B) WT and VD when R208 is forming salt link(s) (SL) with V116D or D185, or both.
Collectively, the above results suggest that the lack of proton selectivity in D112V, D112S, and VS mutants is caused at least in part by the distribution of charged groups in the lumen, where the presence of an excess positive charge near the EC constriction would tend to favor anions over cations. This finding is consistent with the anionic selectivity of D112S and VS mutants. Inversely, the neutrality of the EC bottleneck region in WT and VD systems results in approximate cancellation of the static field, which, although it does not explain proton selectivity, is consistent with the fact that these two channels are permeable to a cation.

5.4 Discussion

5.4.1 Nonconducting mutants

When the critical Asp\textsuperscript{112} residue was moved along the S1 segment from position 108 through 118 (with D112A), 8 of 10 mutants did not conduct, proton current was seen at 116, and anion current was seen at 109. Positions 116, 112, and 109 all face the “pore” in the predicted open-state structure of hH\textsubscript{V1}.\textsuperscript{164} Asp at pore-facing position 143 in the S2 segment also exhibited anion current. Evidently, the Asp carboxyl group must face the pore to enable conductance of any kind, and with Asp at a nonpore-facing location, the S1 segment is not free to rotate enough for the carboxyl group to reach the pore. When Asp faces away from the pore, its pK\textsubscript{a} likely increases substantially, making it permanently neutral. When Lys was introduced at a series of locations in the acetylcholine receptor channel, its pK\textsubscript{a} was decreased, often drastically, when it did not face the pore directly.\textsuperscript{337} When ionizable amino acids are inserted inside proteins by mutation, their pK\textsubscript{a} generally shifts in the direction that promotes neutrality.\textsuperscript{338} In contrast, in native proteins, ionizable residues have evolved to establish interactions with their neighbors that favor ionization.\textsuperscript{339,340}

5.4.2 The rules of the game

The selectivity of various mutants summarized in Table 5.2 reveals “the rules of the game” for hH\textsubscript{V1}. In the mutants studied, the presence of Asp or Glu at position 112 or 116 was necessary and sufficient to produce H\textsuperscript{+} selectivity. The identity of the amino acid at position 109 had no effect; selectivity was determined entirely by positions 112 and 116. When Val was present at both critical positions, 112 and 116, the channel failed to conduct (an effect that was not overcome by introducing Asp at 109). That Val seems to disfavor permeation may reflect its relative hydrophobicity.\textsuperscript{341,342} The VxxDxxxV motif (or a closely conserved version in which Leu replaces Val\textsuperscript{109} and/or Val\textsuperscript{116}) appears in the S1 segment
of all eight species with electrophysiologically confirmed Hv1. In \( h \)Hv1, when positions 112 and 116 were occupied by Val and a small neutral amino acid (Ala or Ser, for example) in either order, the result was anion permeation.

Table 5.2: Effects on selectivity of amino acids at positions 109, 112 and 116 in the S1 transmembrane segment of \( h \)Hv1

<table>
<thead>
<tr>
<th>Construct</th>
<th>109</th>
<th>112</th>
<th>116</th>
<th>Selectivity</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT ( h )Hv1</td>
<td>Val</td>
<td>Asp</td>
<td>Val</td>
<td>H(^+)</td>
<td>Many</td>
</tr>
<tr>
<td>A113D</td>
<td>Val</td>
<td>Asp</td>
<td>Val</td>
<td>H(^+)</td>
<td>This</td>
</tr>
<tr>
<td>D112E</td>
<td>Val</td>
<td>Glu</td>
<td>Val</td>
<td>H(^+)</td>
<td>Musset et al., 2011</td>
</tr>
<tr>
<td>D112A/V116D</td>
<td>Val</td>
<td>Ala</td>
<td>Asp</td>
<td>H(^+)</td>
<td>This</td>
</tr>
<tr>
<td>D112V/V116D</td>
<td>Val</td>
<td>Val</td>
<td>Asp</td>
<td>H(^+)</td>
<td>This</td>
</tr>
<tr>
<td>D112V/V116E</td>
<td>Val</td>
<td>Val</td>
<td>Glu</td>
<td>H(^+)</td>
<td>This</td>
</tr>
<tr>
<td>D112V</td>
<td>Val</td>
<td>Val</td>
<td>Val</td>
<td>0</td>
<td>Musset et al., 2011</td>
</tr>
<tr>
<td>D112V/V109D</td>
<td>Asp</td>
<td>Val</td>
<td>Val</td>
<td>0</td>
<td>This</td>
</tr>
<tr>
<td>D112A, N, S, H, K, F</td>
<td>Val</td>
<td>Ala</td>
<td>Val</td>
<td>Cl(^-)</td>
<td>Musset et al., 2011</td>
</tr>
<tr>
<td>D112A/V109D</td>
<td>Asp</td>
<td>Ala</td>
<td>Val</td>
<td>Cl(^-)</td>
<td>This</td>
</tr>
<tr>
<td>D112V/V116S</td>
<td>Val</td>
<td>Val</td>
<td>Ser</td>
<td>Cl(^-)</td>
<td>This</td>
</tr>
<tr>
<td>D112V/V116N</td>
<td>Val</td>
<td>Val</td>
<td>Asn</td>
<td>Cl(^-)</td>
<td>This</td>
</tr>
</tbody>
</table>

Selectivity to H\(^+\) means \( V_{\text{rev}} \) was close to \( E_H \) at various \( \Delta p\text{H} \); 0 means no credible currents; Cl\(^-\) means \( V_{\text{rev}} \) shifted negatively when Cl\(^-\) replaced CH\(_3\)SO\(_3\)^- in the external solution. Acidic residues are bold; neutral residues other than Val are italicized.

5.4.3 Microenvironment of positions 112 and 116

With respect to selectivity, voltage dependence, kinetics, and \( \Delta p\text{H} \) dependence of gating, the filter-shifted D112V/V116D channels were quite similar to WT. In addition, the effects of point mutations at position 116 closely resembled those at 112. Glu replacing Asp at either position preserves proton specificity. Channels with neutral residues like Ser or Asn at either position (with Val occupying the other) conduct anions. These phenomenological parallels indicate that in terms of the molecular details required to establish proton (or anion) selectivity, the two positions are virtually indistinguishable.

However, the identities of individual neighbors of the critical aspartate differ substantially at the two positions. Side chains with atoms within 6 Å of Asp\(^{112}\) (WT) or Asp\(^{116}\) (in the D112V background) at least 50% of the time were identified from time-averaging of MD simulations. Intriguingly, in WT, the nearest neighbors to Asp\(^{112}\), excluding those on the S1 segment (Ala\(^{113}\), Leu\(^{111}\), Leu\(^{115}\), Val\(^{116}\), and Val\(^{109}\)), include residues on both S2 (Ile\(^{146}\), Ser\(^{143}\), and Phe\(^{150}\)) and S4 (Arg\(^{208}\)), whereas in the VD mutant, only S4 residues (Arg\(^{205}\) and Arg\(^{208}\); along with S1 residues Leu\(^{117}\), Glu\(^{119}\), Leu\(^{115}\), and Leu\(^{120}\)) are within 6 Å of Asp\(^{116}\). In the D112V/V116D mutant, the selectivity filter is roughly twice as far from Phe\(^{150}\) (the external delimiter of the charge transfer center\(^{274}\)) as Asp\(^{112}\) is in the WT channel.
Chapter 5. Peregrination of the SF delineates the pore of the hHv1

This result means that immediate proximity of the carboxyl group to the most hydrophobic point in the channel, which in all constructs occurs at Phe\textsuperscript{150} (Figs. 5.11 and 5.12), is not essential to establishing any of the main features of the channel, including proton specificity, gating kinetics, and \( \Delta p\text{H} \)-dependent gating.

**5.4.4 Why does 109 not work?**

The lack of influence of position 109 (Table 5.1) suggests that selectivity is established in the external vestibule, outside the charge transfer center boundary at Phe\textsuperscript{150}. All current models of hHv1 place Asp\textsuperscript{112} in the external vestibule.\textsuperscript{164–166,312} The MD simulations provide additional insight. In WT hHv1, Asp\textsuperscript{112} is frequently paired with Arg\textsuperscript{208}. During the 10\% of the time that these residues were unpaired, a continuous water wire was present. With Asp moved to 116, there was an electrostatic barrier to cation permeation when Arg\textsuperscript{208} was unpaired, but not when it was paired with either Asp\textsuperscript{185} or Asp\textsuperscript{116}. In support of these findings, an unpaired Arg residue near the EC mouth of aquaporin channels is essential to block the translocation of protons.\textsuperscript{68,343}

**5.4.5 Limited plasticity of the channel with respect to axial helix movement does not support a change of S1 registry in the mutants**

The interpretation of measurable effects of point mutations assumes conservation of the global protein conformation. Accordingly, we have modeled the structure of the mutants by substituting individual side chains in our model of the WT. An alternative interpretation for the native-like phenotype of the VD double mutant is that the S1 helix is shifted by one turn toward the interior of the cell, so that D112V and V116D take the place of V109 and D112, respectively. However, the analysis of collective fluctuations from the simulations shows that not only were the overall structure and hydration of the channel conserved, but there was also no significant axial displacement of individual helices in response to any of the mutations, and the plasticity of the channel with respect to axial helix movement remained limited in all the systems studied (Figs. 5.9 and 5.10). Although the possibility of a change in S1 registry with respect to the rest of the bundle cannot be ruled out, these results suggest that such a change is unlikely in any of the mutants considered in the present study.

Collectively, our current findings suggest that charge neutrality resulting from the presence of Asp (or Glu) vis-à-vis Arg\textsuperscript{208} at the external constriction is required for the charge selectivity of the channel; in particular, the presence of an excess positive charge at the narrow end of the external funnel in neutral Asp\textsuperscript{112} mutants leads to a barrier opposing cation movement but compatible with anion selectivity.
In addition, the presence of an ionic network involving Arg\textsuperscript{208} results in dynamic fluctuations of pore hydration and/or electrostatic properties that may contribute to the mechanism of proton selectivity in the WT and VD channels.

### 5.4.6 Parallels in other molecules

To our knowledge, h\textsubscript{HV}1 is the first example of a selective ion channel whose selectivity filter can be moved by a pair of point mutations. The acetylcholine receptor channel remains nonselective among cations when its ring of Glu residues is shifted by one turn of the α helix.\textsuperscript{344} Several other proton-conducting pathways permit shifts of critical amino acids, although in these molecules the portable function is primarily rapid proton flux rather than proton selectivity per se. For example, in F\textsubscript{1}F\textsubscript{o}-type ATP synthase (\textit{Escherichia coli}), proton translocation is preserved when Asp\textsuperscript{61} is shifted to position 24 on another helix.\textsuperscript{345} Asp\textsuperscript{61} can be replaced by Glu\textsuperscript{61}, but with diminished proton transport, suggesting that precise location of the carboxyl is critical.\textsuperscript{346} Portability of Asp shows that the precise structure of apolar neighbors of the carboxyl group is not critical.\textsuperscript{346} Another essential residue in ATP synthase is Arg\textsuperscript{210}, which is thought to lower the pKa of Asp\textsuperscript{61} transiently to ensure proton release.\textsuperscript{347,348}

As in ATP synthase, an Asp\textsuperscript{213} critical for proton translocation in nicotinamide nucleotide transhydrogenase can be replaced by Glu, but activity is decreased to 18%.\textsuperscript{349} In enzyme studies, activity that is rate-limited by proton translocation is assessed. The correlate in h\textsubscript{HV}1 is single-channel conductance, which was not examined here. Our criterion is perfect proton selectivity, which is preserved when Glu replaces Asp at either position 112 or 116 in h\textsubscript{HV}1. We cannot say whether Glu is equally efficient.

When Asp\textsubscript{135} in the proton entry channel in cytochrome \textit{bo} ubiquinol oxidase of \textit{E. coli} is neutralized, its function can be restored by shifting Asp to position 139 or 142.\textsuperscript{350} Neutralizing Asp\textsubscript{132}, the namesake of the D channel in cytochrome \textit{c} oxidase from \textit{Rhodobacter sphaeroides}, by mutations D132N or D132A nearly abolishes proton uptake,\textsuperscript{351} which is restored by repositioning the Asp at N139D.\textsuperscript{352,353} Intriguingly, proton uptake is also restored by removing subunit III\textsuperscript{354} and also in the D132N/N139T double mutant, showing that in the D channel, rapid proton uptake can be accomplished without an acidic group, although enzyme turnover remains impaired.\textsuperscript{355} Asn\textsuperscript{139} is thought to serve a special “gating” function in cytochrome \textit{c} oxidase that may normally limit WT H\textsuperscript{+} flux;\textsuperscript{356} Thr\textsuperscript{139} appears to optimize aqueous connectivity within the pore.\textsuperscript{355}

Finally, His\textsuperscript{64} shuttles protons from the catalytic center of human carbonic anhydrase II, and function is preserved with His shifted to His\textsuperscript{67} (H64A/N67H), but not His\textsuperscript{62} (H64A/N62H), despite crystal structures indicating that the side chains of both His\textsuperscript{62} and His\textsuperscript{67} extend into the active-site cavity at
distances from the catalytic zinc similar to His\(^{64}\).\(^{357}\)

These examples show that an amino acid side chain must be positioned correctly to maintain a high rate of proton transfer, but that in some cases a reasonable rate of proton transfer can be retained upon moving the side chain, especially if it is not moved too far. Given that proton transfer via titratable amino acid side chains is a way to achieve proton selectivity, proton transfer and proton selectivity may be the same process, and thus in some cases, a side chain essential for H\(^{+}\) selectivity may be moved without losing selectivity.

In summary, shifting Asp along the S1 segment identified three locations in hHV\(_1\) that line the pore and permit conduction: 109, 112, and 116. Asp produced proton specificity only at positions 112 and 116. When introduced at nonpore-facing positions, Asp abolished function. Glu at either position preserved selectivity, indicating leeway in side-chain length. We conclude that the minimal requirements for proton specificity of hHV\(_1\) include Asp or Glu, which must face the pore directly, and evidently must be located in the external vestibule, above the charge transfer center, Phe\(^{150}\). The portability of the selectivity filter indicates latitude in the requisite local environment. This observation seems consistent with the suggestion that ionizable groups that enable proton transport could have evolved by random mutation without the need to simultaneously develop a specialized micro-environment for charge stabilization.\(^{338}\) On the other hand, the inability of Asp to produce H\(^{+}\)-selective conductance at most positions tested argues that additional factors are involved. Modeling indicates frequent salt-bridge formation between Asp and several partners, which may contribute to selectivity, proton transfer, and conformational stability. Ionizable residues favor the incorporation of water molecules\(^{358}\) that may also facilitate proton translocation. However, neither the presence of a continuous water pathway nor the mean hydration profile was found to have any clear relationship with the selectivity of the channel. Detailed computational studies of H\(^{+}\) and other ions in the permeation pathway will be required to examine the molecular basis of ion movement and achieve a full understanding of the mechanism of proton selectivity in hHV\(_1\).

### 5.5 Conclusion

The role of human H\(_V\)\(_1\) is selective permeation of protons across the cell membrane. Having identified the pore-facing Asp\(_{112}\) on S1 helix a crucial determinant for proton selectivity that becomes anion selective upon D112x (x=A, N, S, H, K or F), we set out, in this study, to investigate where else on S1 can a carboxyl side chain be introduced that remained proton selective. Only two pore-facing locations, 116 and 109, each roughly one helical turn up and down of 112 respectively displayed channel
properties, whose charge selectivity was further determined by the chemical property at 112. Such that, double mutants D112V or A/V116D remained proton selective, while D112A/V109D was anion selective. However, there was no channel properties in D112V/V109D and D112V of which the molecular basis for non-conductance remains elusive. Like in WT, the charge neutralization continues to determine, at least in part, the charge selectivity as only when Arg^{208} is neutralized by the Asp introduced at position 116 via a salt-link, the D112V/V116D remains cation conducive. Otherwise, the unpaired Arg^{208} can lead to anion selectivity as seen in D112S and D112V/V116S that have an unpaired Arg^{208} in the extracellular crevice - a feature that may also explain anion selectivity of D112A/V109D and D112V/V116N. Collectively, from a series of mutations on S1 helix, our study identified new locations of Hv1 that display channel properties and for some locations, we were able to characterize the local environment that favors one charge for permeation over the other.
Chapter 6

Free Energy Simulations of Ion Translocation through $hHV_1$

Abstract

The selectivity for protons in the human voltage-gated proton channel, $hHV_1$, is remarkably high. Asp$^{112}$ located in the putative selectivity filter of the pore is crucial for imparting proton selectivity. Mutations involving Asp$^{112}$ turn $hHV_1$ into a non-conducting or anion-selective channel, or convert a non-conducting channel back to a proton-selective one. Although $hHV_1$ and its mutants exhibit drastic differences in ion permeation, the molecular basis of proton selectivity in the wild-type and of anion selectivity in mutants remains unexplained. In order to elucidate the ion and charge selectivity of $hHV_1$, we compute free energy profiles for the translocation of Na$^+$ and Cl$^-$ ions through the pore of a homology model of $hHV_1$ and of a single and a double mutant, both of which are anion selective. Results indicate that the movement of non-proton ions in the wild-type is opposed at the constriction site of the pore, where both types of ions face a substantial desolvation penalty. Although the barrier for Cl$^-$ is significant even in the anion-selective mutants, limited structural rearrangements in the constriction site and the absence of a negatively charged residue at position 112 lower the barrier for Cl$^-$ in mutants compared to the wild-type channel. Taken together, these findings provide clues to refine the structural model as well as insight that can be used to further the elucidation of ion and charge selectivity and the molecular mechanism of proton translocation in $hHV_1$. 
6.1 Introduction

Voltage-gated proton channels, Hv1, are structural homologs of voltage-sensor domain (VSD) of voltage-gated cations channels, responsible for sensing membrane potentials (commonly referred as voltage-sensing) to regulate channel-gating. In addition to voltage-sensing, the VSD-like four helical (S1-S4) arrangement of Hv1 confers selective permeation for protons. VSDs also display channel properties under certain conditions. With varying membrane potentials, replacing highly conserved Arg residues of the S4 helix to His elicits proton currents through the VSD of Shaker Kv. Similarly, mutations of Arg to small side chains lead to cations leaking through the VSD. These cations include Li\(^+\), K\(^+\), Cs\(^+\) and even guanidinium. Anion permeation in these VSDs has not been reported so far.

Hv1, on the other hand, seems more diverse in its ion and charge selectivity upon mutations involving Asp\(^{112}\) on S1. The mutation of Asp\(^{112}\) to Val abrogates channel properties. Replacing Asp\(^{112}\) by a smaller neutral residue such as Ser turns hHv1 into an anion selective channel that conducts Cl\(^-\), as does the double mutant D112V/V116S. Surprisingly, D112V/V116D is proton selective. The molecular basis for this spectrum of ion and charge selectivity in the wild-type hHv1 and its mutants remains unexplained. Here, we investigate the energetic basis specifically for the movement of Na\(^+\) and Cl\(^-\) in the wild-type hHv1 and two anion-selective mutants, D112S and D112V/V116S.

We have previously shown, based on a homology model of hHv1, that neutralizing the charge of Asp\(^{112}\) imposes an electrostatic barrier for cation movement, which is compatible with the anion selectivity of single point mutations of Asp\(^{112}\). The electrostatic barrier appears due to the presence of the unpaired positive charge of Arg\(^{208}\) on S4 that otherwise forms a salt-bridge with Asp\(^{112}\), delineating the extracellular (EC) boundary of the 1-nm-long narrow region of the pore. The intracellular (IC) boundary is delineated by the residue pair of Arg\(^{211}\) on S4 and highly conserved Phe\(^{150}\) on S2. Based on the hydration of the pore, the Arg\(^{211}\)–Phe\(^{150}\) pair appears to form the most constricted site/bottleneck of the pore (Chapter 4, Fig. 11).

Further treating the mutants explicitly, we have shown that the unpairing of Arg\(^{208}\) could be induced by replacing D112 with Ser or Val in D112S and D112V/V116S, respectively (Chapter 5, Fig. 14). Hydration of the pore was indistinguishable in the wild-type and these mutants, making it impossible to explain the selectivity (Chapter 5, Fig. 11). In addition, the brute force simulations of that study did not report any spontaneous movements of Na\(^+\) or Cl\(^-\) ions through the pore, making it impossible to quantify the energetics of the ion and charge selectivity of hHv1 and its mutants.

Using biased simulations in the current study, we demonstrate a comparative assessment of the molecular determinants and the energetics governing the movement of Na\(^+\) and Cl\(^-\) in hHv1. Deter-
Chapter 6. Free Energy Simulations of Ion Translocation through hHv1

mining the exclusion of non-protonic cations and charge discrimination in the wild-type and mutants would provide the essential first step towards thermodynamic validation underlying the ion movement (or non-movement) of the current model of hHv1 before embarking on computationally exhaustive simulations involving explicit protons. For such simulations, the current study would also provide insight on designing a reduced model limited to the most critical regions of proton translocation.

6.2 Methods

6.2.1 System setup and umbrella sampling

Molecular dynamics simulations of the wild-type protein are based on the homology model described in Kulleperuma et al. 2013 (Chapter 4 in this thesis). This model was then used to construct the single and double mutant systems: D112S and D112V/V112S, which, along with the wild-type, were separately simulated in a hydrated POPC bilayer for 200 ns using 12 independent replicas as described in Morgan et al., 2013 (Chapter 5 in this thesis). Here, these simulations are extended to 300 ns.

For each system of the above simulations, conformations from four replicas exhibiting the most hydration of the narrow region of the pore between 100 and 300 ns were used to extract the initial conformations for umbrella sampling (US) simulations. For each of the four replicas, a starting conformation which represents a single window was used to replace one of the pore-associated water molecules by a Na\(^+\) or Cl\(^-\) ion. This process was repeated for all the conformations for each of the four replicas, ensuring there was sufficient overlap between adjacent histograms (data not shown) of the configurations of the ion in the \(z\) axis such that the spacing between any two windows was \(\sim 0.05\) nm.

The lower and upper boundaries of the \(z\) position for windows varied based on the ion type. This is because previous brute force simulations revealed that Na\(^+\) spontaneously penetrated into the negatively charged extracellular (EC) and intracellular (IC) crevices of the channel, leaving only a narrow pore region to be sampled by US simulations. In contrast, the region for US sampling of Cl\(^-\) extended into both EC and IC crevices of the channel. More specifically, the \(z\) range for US simulations of Na\(^+\) and Cl\(^-\) ranged from -1 to 1 nm and -1.8 nm to 1.5 nm, respectively. For a replica of a given system, the total number of windows for Na\(^+\) and Cl\(^-\) were 35 and 60, respectively.

The final simulation cell consisted of 126 POPC and approximately 5900 water molecules in a box of approximately \(6.5 \times 6.5 \times 8\) nm\(^3\). Excess ions yielding an approximate ionic concentration of 500 mM were present in the brute force simulations and US simulations of Cl\(^-\). US simulations of Na\(^+\) with excess ions are currently being carried out.
Chapter 6. Free Energy Simulations of Ion Translocation through $h$HV1

6.2.2 Simulation protocol

MD simulations were conducted with version 4.0.5 of the GROMACS simulation package. The OPLS-AA protein force field was mixed with the Berger lipid parameters by applying the half-$\epsilon$ double-pairlist method. The water model was TIP3P. In each of each of wild-type, single, and double mutants, each US window from four replicas was subjected to 1000 steps of steepest-descent energy minimization, followed by a pre-equilibration phase of 2 ns with backbone position restraints on protein, using a force a constant of 1000 kJ/mol/nm$^2$. The $z$ position of the ion was harmonically restrained to that of the water oxygen atom it replaced with a force constant of 4000 kJ/mol/nm$^2$ in all steps of the simulations. Each US window was simulated for 50 ns of production run. The axial position of the permeating magnesium ion, $z$, was stored every 20 fs to generate independent PMFs using Alan Grossfield’s implementation of the weighted histogram analysis method (WHAM). The parameters for the extension of brute force simulations to 300 ns and production run of US simulations were identical to those described in Chapter 5. The total US simulation time for Na$^+$ and Cl$^-$ was 7 and 12 $\mu$s, respectively.

6.3 Results

6.3.1 Thermodynamic basis for the movement of Na$^+$

6.3.1.1 Convergence of PMFs and energetic barrier for Na$^+$

All PMFs were generated by combining data from US simulations for the pore regions as specified above and from brute force sampling for the rest of the system. First, for each of wild-type, single, and double mutants of $h$HV1, the convergence of the PMF for Na$^+$ along the pore was assessed by computing the PMF sequentially for five consecutive time intervals of 10 ns from 0 to 50 ns (Fig. 6.1) and averaging these 5 PMFs over the four replicas (Fig. 6.1). The barriers in the PMFs computed from the 0-10 ns time interval in all three systems are significantly higher than those of other time intervals, suggesting that part of the conformational biases originating from starting conformations is eliminated within the first 10 ns. Hence, the remaining analysis for all systems is based on the time interval between 10 and 50 ns for all windows.
Figure 6.1: Convergence of the PMF profiles for Na\(^+\) translocation in wild-type and mutant hH\(_{\text{V1}}\). A) wild-type; B) single mutant D112S; C) double mutant D112V/V116S. In each plot, the average PMFs from four replicas were generated for five consecutive time intervals of 10 ns (thick coloured lines except black) from 0 to 50 ns. The average PMFs from 10-50 ns are depicted in black lines. The black error bars represent the standard error of the mean (SEM).
Chapter 6. **Free Energy Simulations of Ion Translocation through hHv1**

The PMFs of all three systems consist of a single free energy barrier. The barrier region spans the central region of the pore, at \(-1.0 \leq z \leq 1.0 \) nm. In the wild-type protein, the top of the \(14\pm0.5 \) kcal/mol barrier is localized at \(-0.4 < z < -0.2 \) nm, which corresponds to the location of Phe\(^{150}\), with the Arg\(^{211}\)-Phe\(^{150}\) pair forming the apparent steric bottleneck of the pore. The PMF profiles of the two mutants are almost indistinguishable, with a barrier top of \(14.5\pm0.5 \) kcal/mol and are significantly broader, by \(3 \) Å compared to the wild-type PMF (Fig. 6.2). The barrier top has moved to \(z=0\), which corresponds to the center of the hydrophobic stretch spanning the region \(z = -0.4 \) to \(0.3 \) nm.

![Figure 6.2: Comparison of the PMFs for Na\(^+\) between wild-type and mutants.](image)

The average free energy profiles corresponding to the time interval between 10 and 50 ns (Fig. 6.1) are shown in thick coloured lines. Error bars represent the SEM.

### 6.3.1.2 Analysis of Na\(^+\) coordination

To understand the physical origin of the PMFs, the coordination of Na\(^+\) is analysed. The axial distribution of the coordination number of Na\(^+\) through the pore is shown in Fig. 6.3. The coordination number is defined as the number of heavy atoms in direct contact with the ion of interest, which varies with time in a dynamic system. Here, a polar heavy atom is considered to coordinate Na\(^+\) if it lies within \(0.32 \) nm from the ion. This distance cutoff corresponds to the first minimum of the radial distribution function of water O atoms around Na\(^+\) (data not shown).
Figure 6.3: Ligand coordination of Na\(^+\) through the channel. Top: Representative snapshots highlighting Na\(^+\) coordination in the wild-type channel within the specified regions. Water molecules (red and white licorice) and acidic residues (cyan and red licorice) in the first hydration shell of Na\(^+\) are highlighted. Arg\(^{208}\) and protein backbone are shown in licorice (cyan and blue) and ribbon (white), respectively. Bottom: A) wild-type. B) single mutant: D112S. C) double mutant: D112V/V116S. In each plot, the replica-averaged axial distributions of water O atoms (blue), polar heavy atoms of the channel (green) and counter-ions (orange) coordinating the Na\(^+\) ion are represented in thick lines. The total coordination by water, channel and counter-ions is shown in black thick lines. Error bars represent the SEM. The PMF is shown in red dotted lines. The extent of the channel pore in the bilayer is demarcated by thin vertical red dotted lines.
In bulk water, the average coordination number of Na\(^{+}\) is 5.6±0.2. In all three systems, as the ion enters from the bulk phase into the water-filled crevices on either sides of the channel, i.e \(-1.5 < z < -0.5\) nm in IC and \(0.8 < z < 1.7\) nm in EC (the demarcation is based on the previous study or Chapter 4), Na\(^{+}\) becomes partially dehydrated, an effect that is compensated by channel coordination.

The total coordination number, however, displays only small variations (Fig. 6.4 panel A), including in the narrowest region of the pore (\(-0.5 < z < 0.8\) nm). Nonetheless, the average channel and water coordinations notably differ where residue 112 is located (\(0 < z < 0.5\) nm) (grey shaded area in Fig. 6.4 panels B and C). The reduction in the channel coordination correlate with the polarity of the residue at position 112 (D>S>V). Reciprocally, the water coordination increases by 1 or 2 to compensate for loss of channel coordination in the single and double mutants, respectively (Fig. 6.4, top). Residue 116 one helical turn up from 112 is located in the well-hydrated EC funnel so that it has no effect on the solvation of the Na\(^{+}\) ion.
A number of ionic residues line the pore of \( hHv1 \).\(^{362} \) In order to assess their contribution to the ion coordination, we decomposed the total channel coordination into individual ionic residues (Fig. 6.5). In all three systems, the total coordination by acidic residues in the IC and EC crevices roughly adds up to at least 1, which suggests the permeating \( \text{Na}^+ \) ion is essentially neutralized by the channel coordination in those regions. The narrow region of the channel in the two mutants, however, is devoid of acidic residues for charge neutralization, which also correlates with the location of the EC-facing shoulder of the free energy profiles. In contrast, Asp\(^{112} \) in the wild-type protein extends into the narrow region (yellow filled curve in Fig. 6.5) and neutralizes the ion, effectively narrowing the barrier top to the internal bottleneck where the channel-coordination is at its minimum. Furthermore, the possibility of a salt link between Asp\(^{112} \) and Arg\(^{208} \) (Fig. 6.3, top right) does not in itself present a significant barrier to \( \text{Na}^+ \) movement.
Chapter 6. Free Energy Simulations of Ion Translocation through hHV1

Figure 6.5: Coordination of Na\(^+\) by channel residues. Top: The acidic residues coordinating Na\(^+\) are highlighted in colored licorice. The protein backbone is in a white ribbon representation. Bottom: A) wild-type. B) single mutant: D112S. C) double mutant: D112V/V116S. In each plot, the replica-averaged coordination number of Na\(^+\) by polar heavy atoms of the channel is shown in a green thick line. The decomposition of the channel coordination by acidic residues: D112 (yellow), E171 (red), E153 (green), D174 (blue), D185 (magenta), E119 (orange), D123 (violet), E196 (cyan), and E164 (brown) are shown in transparent filled curves. The extent of the channel pore is demarcated by thin vertical red dotted lines.
To further illustrate the effect of the presence or absence of a negative charge at position 112 on free energy profiles, we compared the difference of free energy profiles with that of the static field profiles (Chapter 5, Fig. 14) between the two mutants and the wild-type protein (Fig. 6.6). The qualitative agreement between these difference plots reaffirms that the broadening of the free energy profiles of the mutants is due to the absence of a negative charge at position 112. Collectively, these results suggest that the local charge neutralization by the channel-coordination modulates the energetics of Na$^+$ ion permeation in hHV1 and the anion-selective mutants.

Figure 6.6: Differences in free energy and static field energy due to single- and double-point mutations.

6.3.1.3 Solvation shell of water chelating Na$^+$

Since Na$^+$ is almost fully hydrated at and near the top of the free energy barrier, we then further characterized the local environment of the water molecules located in the internal bottleneck. To this end, the solvation shell of the water molecules chelating the ion is analyzed (Fig. 6.7). A distance of 0.35 nm corresponding to the first minimum of the water O-O radial distribution function is used to define the radius of the solvation shell. In a water-ion complex, this distance measured from the ion corresponds to 0.59 nm, which is slightly higher than the absolute radius of the second hydration shell of Na$^+$, 0.53 nm. Three kinds of ligands solvate the water molecules chelating Na$^+$: other water molecules as well as polar and nonpolar heavy atoms of the channel.
Figure 6.7: Solvation of water molecules chelating Na\(^+\) inside the pore. A) Wild-type; B) single mutant D112S; C) double mutant D112V/V116S. In each plot, the replica-averaged coordination number of water molecules coordinating Na\(^+\), by water O atoms (blue), polar heavy atoms of the channel (green), and nonpolar heavy atoms of the channel (yellow), are shown in thick lines. The total coordination by polar heavy atoms is shown in a thick dotted black line. The coordination number by Phe\(^{150}\) is shown in a black filled-curve.
The top of the free energy barrier matches the location of the minimum coordination by polar atoms in all three systems, suggesting that the energetic penalty is due to the desolvation of the hydrated Na\(^+\) ion. In addition, this region corresponds to the location of the nonpolar coordination overtaking the polar coordination of the channel. More specifically in the wild-type protein, the peak of the free energy barrier overlaps well with the location of the maximum coordination by nonpolar heavy atoms \((-0.5 < z < 0 \text{ nm})\) which also corresponds to the location of Phe\(^{150}\), the main contributor to the nonpolar coordination of the hydrated Na\(^+\) ion (Fig. 6.8). However, this coincidence is less striking in the mutants.

![Figure 6.8: Nonpolar coordination of the hydrated Na\(^+\) in the region corresponding to -0.5 <z <0 nm. Side (top) and top-down (bottom) views of representative snapshots highlighting residues D112 (wild-type), D112S (single mutant), D112V (double mutant), R208, and R211 in cyan, red and/or blue licorice. Hydrophobic residues (yellow licorice) coordinating the hydrated (red and white licorice) Na\(^+\) (blue sphere) and protein backbone (white ribbon) are highlighted.](image)

Furthermore, the nonpolar coordination overtaking the polar coordination is somewhat shifted towards the EC side in the double mutant channel. The net effect, however, seems insignificant as the height of the free energy profiles are almost equal among the three systems.

Altogether, this analysis reveals that despite the fact that the coordination number of Na\(^+\) is preserved, the first hydration shell of Na\(^+\) is surrounded by a nonpolar region in the narrow region of the pore.

In summary, the height of the barrier opposing Na\(^+\) movement is nearly identical in wild-type and anion-selective mutants and is due primarily to the desolvation penalty of the hydrated-Na\(^+\) in the
nonpolar narrow region of the channel. The propensity to form a salt bridge between Asp$^{112}$ and Arg$^{208}$ in the wild-type channel does not result in a significant free energy penalty to the movement of Na$^+$. However, the lack of charge neutralization of Na$^+$ by the carboxylic group of Asp$^{112}$ significantly increases the width of the barrier in the mutants.

6.3.2 Thermodynamic basis for the movement of Cl$^-$

6.3.2.1 Convergence of the PMF profiles and energetic barrier for Cl$^-$

The convergence of the PMF profiles is assessed as described above for Na$^+$. Similar to that of Na$^+$, the PMFs computed based on the 0-10 ns are significantly higher than those obtained from subsequent time intervals, hence the remaining analysis is based on the time interval between 10 and 50 ns for all windows of each of the three systems.
Figure 6.9: Convergence of the PMF profiles for Cl\(^-\) translocation in wild-type and mutant \(hHv1\). A) wild-type; B) single mutant D112S; C) double mutant D112V/V116S. In each plot, the average PMFs from four replicas were generated for five consecutive time intervals of 10 ns (thick coloured lines except black) from 0 to 50 ns. The average PMFs from 10-50 ns are depicted in black lines. The black error bars represent the SEM.
The barrier opposing Cl\(^-\) movement in all three channels peaks between -0.5 and -0.3 nm and the WT channel features a barrier top of 16.5 kcal/mol (Fig. 6.10) which is decreased by 2.5±0.5 kcal/mol in the two mutants which show similar topological features in the free energy profiles. Furthermore, the EC side of the barrier top (-0.3 nm > z) is lowered by about the same amount compared to the wild-type channel, which correlates with the disappearance of a negative charge at position 112 in the mutants.

![Figure 6.10: Comparison of the PMFs for Cl\(^-\) in wild-type and mutant hHV1. The average free energy profiles corresponding to the time interval between 10 and 50 ns (Fig. 6.9) are shown in thick coloured lines. Error bars represent the SEM.](image)

6.3.2.2 Analysis of Cl\(^-\) coordination

Ligand coordination of Cl\(^-\) is analyzed using a distance cutoff of 0.39 nm, which corresponds to the first minimum of the radial distribution function of water O atoms around Cl\(^-\) (data not shown).

In bulk water, the average coordination number of Cl\(^-\) is 7.2±0.2 (Fig. 6.11). In all three systems, as the ion enters from the bulk phase into the water-filled crevices of either sides of the channel, Cl\(^-\) becomes partially dehydrated, which is partly compensated by the channel coordination. In addition, an intermittent coordination by Na\(^+\) ions is visible throughout the pore. This effect is much more pronounced in the bottleneck of the wild-type channel compared to the mutants, with an average counterion coordination of 0.6±0.2 compared to 0.1±0.1 in the mutants, in the range -0.7 < z < 0.5 nm.
Figure 6.11: **Ligand coordination of Cl\(^{-}\) through the channel.** Top: Representative snapshots highlighting Cl\(^{-}\) coordination within the specified regions of the three systems. Counter ions, Na\(^{+}\) (blue sphere), water molecules (red and white licorice) and basic residues (cyan and blue licorice) in the first hydration shell of Cl\(^{-}\) are highlighted. Residues at position 112 and protein backbone are shown in licorice (cyan and/or red) and ribbon (white), respectively. Bottom: A) wild-type. B) single mutant: D112S. C) double mutant: D112V/V116S. In each plot, the replica-averaged axial distributions of water O atoms (blue), polar heavy atoms of the channel (green) and counter-ions (orange) coordinating the Cl\(^{-}\) ion are represented in thick lines. The total coordination by water, channel and counter-ions is shown in black thick lines. Error bars represent the SEM. The average free energy profile is shown in red dotted lines.
On average, in the narrow region corresponding to the range -0.5 < z < 0.8 nm, the total coordination drops to 5.9 ± 0.7, 5.4 ± 0.5, and 5.3 ± 0.5 in wild-type, D112S, and D116V/V116S, respectively, which corresponds to a 20-28% drop from coordination in the bulk (Fig. 6.12). Interestingly, the barrier top does not coincide with the minimum coordination of any ligand type. In fact, it is located more towards the bottleneck of the pore at z = ~0.4 nm (Fig. 6.11). This shift suggests that the energetic penalty due to the desolvation of the narrow nonpolar region of the pore is less than that due to the steric effects arising from the constriction of the bottleneck.

The decomposition of the total channel coordination into contributions from individual basic residues reveals that charge compensation occurs throughout most, but not the entire length of the pore (Fig. 6.13). This neutralizing effect seems more localized in the double mutant than in the wild-type and single-mutant channels.

Figure 6.12: Comparison of the coordination of Cl⁻ in the wild-type and mutant channels. Coordination by A) all the polar heavy atoms B) water O atoms C) polar heavy atoms of the wild-type (black) channel and in D112S (orange), and D112V/V116S (brown) mutants. The grey-shaded area demarcates the sidechain position of residue 112.
Figure 6.13: Coordination of Cl\(^-\) by polar and charged groups of the channel. Top: The basic residues coordinating Cl\(^-\) are highlighted in colored licorice. The protein backbone is in a white ribbon representation. Bottom: A) wild-type. B) single mutant: D112S. C) double mutant: D112V/V116S. In each plot, replica-averaged coordination number of Cl\(^-\) by the polar atoms of the channel is shown in a green thick line. The decomposition of the channel coordination by basic residues: R208 (red), R211 (green), R205 (blue), R93 (magenta), K157 (orange), and K131 (violet) are shown in transparent filled curves. The free energy profile is shown in red dotted lines. The extent of the channel pore is demarcated by thin vertical red dotted lines.
6.3.2.3 Solvation shell of water chelating Cl\(^-\)

The second hydration shell of Cl\(^-\) is analyzed based on the same criteria described for Na\(^+\) (Fig. 6.14). The free energy barrier roughly matches the maximum coordination of the nonpolar heavy atoms in the region \(-0.5 \text{ nm} < z < 0 \text{ nm}\) of the wild-type, where residue Phe\(^{150}\) acts as the main contributor to the nonpolar coordination in the second shell of the Cl\(^-\) ion.

![Figure 6.14: Solvation of water chelating Cl\(^-\) inside the pore](image)

A) wild-type. B) single mutant: D112S. C) double mutant: D112V/V116S. In each plot, the replica-averaged solvation of water coordinating the Cl\(^-\), by water O atoms (blue), polar heavy atoms of the channel (green) and nonpolar heavy atoms of the channel (orange) are shown in thick lines. The total coordination by polar atoms, i.e. from water O and channel, is shown in a thick dotted black line. The coordination number by Phe\(^{150}\) is shown in a black filled-curve. The average free energy profile is shown in red dotted lines.
Nonpolar coordination of partially hydrated Cl\(^-\) in the region corresponding to \(-0.5 < z < 0\) nm. Side (top) and top-down (bottom) views of representative snapshots highlighting residues D112 (wild-type), D112S (single mutant), D112V (double mutant), R208, and R211 in cyan, red and/or blue licorice. Hydrophobic residues (yellow licorice) coordinating the hydrated (red and white licorice) Cl\(^-\) (red sphere) and protein backbone (white ribbon) are highlighted.

Notably, the overlap between the location of the barrier top and that of maximum nonpolar coordination is poor in the mutants. Although Phe\(^{150}\) remains the main contributor to the nonpolar coordination of the anion, its distribution as a second-shell ligand of Cl\(^-\) is moderately shifted towards the EC side, possibly lowering its contribution to the steric effects at the bottleneck, as this effect correlates with the lower free energy barrier in the mutants.

Furthermore, the region over which the nonpolar coordination supersedes the polar heavy atom coordination is wider in the mutants than in the wild-type. More specifically, the Val residue at position 112 (z = ~0 nm) contributes to this expansion as does the looser contact with Arg\(^{208}\) in the double mutant. In turn, these effects may be responsible for the somewhat lower barrier for Cl\(^-\) movement, notably in the EC side of the double mutant relative to the single mutant.
Figure 6.16: A representative snapshot highlighting the key residues responsible for nonpolar coordination of the second hydration shell of Cl$^-$ or Na$^+$ in the narrow region of the wild-type channel. Nonpolar carbon atoms (yellow licorice) of residues Asp$^{112}$, Val$^{109}$, Val$^{105}$ on S1, Phe$^{150}$, Glu$^{153}$ on S2, Val$^{178}$ on S3 and Arg$^{208}$ and Arg$^{211}$ on S4 and other charged residues (cyan, red, blue) of Figs. 6.5 and 6.13 are highlighted. The protein backbone is shown as a white ribbon.

In summary, a significant free energy barrier, decreasing in the order WT>D112S >D112V/ V116S, opposes the movement of Cl$^-$. The barrier originates primarily due to the steric effects on the partially desolvated Cl$^-$ in the bottleneck of the channel.
Chapter 6. Free Energy Simulations of Ion Translocation through hHV1

6.4 Discussion

Mutations involving Asp$^{112}$, the putative selectivity filter of hHV1, exhibit drastic differences in ion permeation and charge selectivity.\cite{94,360} Here, we investigate the energetic basis for the movement of Na$^+$ and Cl$^-$ in the wild-type and in two anion-selective mutants of hHV1, D112S and D112V/V116S. Consistent with the experimental observation of the low conductance rates of anion permeation in these two mutants, ion movement did not occur spontaneously in microsecond-time-scale brute force simulations.\cite{164} Hence, biased simulations using umbrella sampling were employed to generate the free energy profiles for ion translocation.

The free energy barrier opposing Na$^+$ movement is large, $\Delta G \approx 14$ kcal/mol, and nearly identical in all three channels, which is consistent with the fact that none of the hHV1 variants are cation-selective, except for protons in the wild-type. Quantitatively, this value corresponds to approximately 8% of the

Figure 6.17: Comparison of the PMF profiles for Na$^+$ (blue) and Cl$^-$ (red). The average free energy (see Fig. 6.1 and 6.9) are shown in thick coloured lines for A) wild-type, B) D112S, and C) D112V/V116S.
hydration free energy of Na$^+$.\textsuperscript{364} Despite the fact that Na$^+$ remains nearly fully hydrated at the barrier top, the coordination number in the second hydration shell of Na$^+$ reaches a minimum, suggesting that the desolvation penalty in the narrow hydrophobic region of the channel is the primary factor opposing Na$^+$ movement.

The effect of charge neutralization in the EC and IC funnels in all three systems (and in particular at position 112 in the wild-type) is secondary, as the absence of a negative charge at position 112 in mutants only confines the free energy barrier for Na$^+$ movement to the narrow hydrophobic region.

The barrier height computed here is commensurate with those obtained in a previous study of Na$^+$ permeation in simplified models of hydrophobic pores with varying radii,\textsuperscript{363} which reports an energetic barrier of 11 kcal/mol for a hydrophobic pore of radius 0.35 nm, a radius that would allow the permeation of a Na$^+$ ion with its first hydration shell. This finding is consistent with the desolvation penalty being responsible for the impermeability of $h$HV$_1$ to Na$^+$.

Taken together, our results indicate that the free energy barrier opposing Na$^+$ in the wild-type and anion-selective mutants originates from the high energetic cost for Na$^+$ ion to shed its second hydration shell. In other words, non-protonic cations are discriminated in $h$HV$_1$ in the narrow hydrophobic region of the pore.

The barrier for Cl$^-$ movement is $\sim$16 kcal/mol in the wild-type channel and decreases by 13% and 19% in the anion-conducting mutants D112S and D112S/V116, respectively. The barrier peaks at or near the bottleneck located at the IC mouth of the narrow hydrophobic region of the pore.

Interestingly, partial desolvation of Cl$^-$ in the first solvation shell is maximal in the narrow hydrophobic region, yet the energetic penalty due to desolvation does not appear to be the primary reason for the free energy barrier as it does for Na$^+$. This is due to two main reasons. First, the hydration free energy of Cl$^-$ is -85 kcal/mol in contrast to -105 kcal/mol for Na$^+$,\textsuperscript{365} which indicates that water molecules are more tightly bound to Na$^+$ than to Cl$^-$. Second, the energetic penalty due to the desolvation of Cl$^-$ is compensated by charge pairing with basic Arg residues that extends well into the narrow hydrophobic region, where such charge pairing is lacking for Na$^+$.

Consistent with anion permeation in the mutants, the barrier to Cl$^-$ movement is lower in the mutants than in the wild-type channel as well as lower than the barrier for Na$^+$ in both wild-type and mutants. However, the barrier opposing the movement of Cl$^-$ in the mutants is still significant at the bottleneck of the pore, where Cl$^-$ remains both partly solvated and neutralized by basic residues of the channel, similar to the narrow hydrophobic region. Therefore, the fact that the free energy barrier peaks at the bottleneck and not in the narrow hydrophobic region suggests that the former is too constricted for the movement of Cl$^-$. 
The energetic barrier at the bottleneck is lower for Na\textsuperscript{+} than Cl\textsuperscript{-} by only 2-3 kcal/mol in the wild-type and mutants (Fig. 6.17). It is plausible that the steric effects of the bottleneck are more pronounced for Cl\textsuperscript{-} than for Na\textsuperscript{+} due to the larger ionic radius of the former. Furthermore, the free energy barrier for Cl\textsuperscript{-} movement is slightly lowered in the mutants in which Phe\textsuperscript{150} of the bottleneck is somewhat displaced towards the EC crevice compared to the wild-type. However, our results suggests that such a displacement is not sufficient to lower the barrier for Cl\textsuperscript{-} movement in the mutants from a value of \( \sim 14 \) kcal/mol to the expected experimental value of \( \sim 9.5 \) kcal/mol. This value is predicted by Eyring transition-state theory\textsuperscript{366} for a Cl\textsuperscript{-} ion flux of \( \sim 60,000 \) s\textsuperscript{-1} in anion-selective channels at a concentration of 0.09 M in the bulk solution (TE DeCoursey, personal communication). Achieving an experimentally compatible barrier height, at similar concentrations may require, at minimum, alternate placements of the sidechains of Phe\textsuperscript{150} and other neighboring residues possibly involved in imposing steric effects in the bottleneck.

According to the current homology model, alleviating these steric effects is unlikely to eliminate the barrier for Cl\textsuperscript{-} in the wild-type due to the remaining desolvation effect and since the presence of a negative charge at position 112 serves as an electrostatic barrier for anion movement. In addition, it can be speculated that the combined effects of a less constricted bottleneck and the absence of a salt-link between Asp\textsuperscript{112} and Arg\textsuperscript{208} may result in loser packing between S1 and S4 helices (at least locally), which could in turn potentially widen the narrow hydrophobic region in the mutants and further lower the free energy penalty arising from the desolvation of Cl\textsuperscript{-} in this region. The latter of these two effects is absent in the wild-type channel. Therefore, the possibility of a moderately wider hydrophobic region for mutants cannot be completely ruled out. However, the hydrophobic region should not be so wide as to eliminate the desolvation penalty for Na\textsuperscript{+}, which culminates in the hydrophobic region.

Furthermore, in the wild-type channel, Cl\textsuperscript{-} has a tendency to recruit a Na\textsuperscript{+} counterion throughout the narrow region of the pore, suggesting a higher penalty opposing anion movement. Accordingly, the free energy barrier for Cl\textsuperscript{-} movement is both broader and higher than for the movement of an (unpaired) Na\textsuperscript{+} ion, consistent with the fact that the wild-type channel is a cation channel.

Contrary to the wild-type results, which indicate a preference for cations almost throughout the channel, mutants prefer anions in the region of \(-0.3<z<0.9 \) nm which includes the hydrophobic region and residue 112. Cations are preferred in the rest of the channel, including the entire IC crevice and the bottleneck (Fig. 6.17).

A possible way to refine our homology model of hHV1 is to consider alternate side chain conformations for the residues located at or near the pore bottleneck, which could be achieved by simulated annealing conserving the overall channel conformation but allowing rotameric isomerization of selected side chains.
This approach would be consistent with the expectation that, at best, a homology model corresponds to a moderate-resolution x-ray structure, in which the secondary structure and overall fold are correct but uncertainties remain on the placement of side chains.\textsuperscript{164}

Despite these uncertainties and current limitations of the model, our results suggest that the two anion-conducting mutants present broader and lower free energy barriers to Na\textsuperscript{+} and Cl\textsuperscript{−} permeation, respectively, than the proton-conducting (wild-type) channel, a finding that is consistent with the role of the charged Asp\textsuperscript{112} side chain in modulating charge selectivity of the hHV\textsubscript{1} channel, as proposed in a previous study based on electrostatic calculations.\textsuperscript{164}

If the origin and the location of the primary barrier to Na\textsuperscript{+} and Cl\textsuperscript{−} motion suggested by the present study are qualitatively correct, the “selectivity filter” controlling ion permeation in the wild type channel is the narrow hydrophobic region rather than a fluctuating salt bridge between Asp\textsuperscript{112} and Arg\textsuperscript{208}, as proposed in a previous study on the basis of a reduced quantum model involving the Asp-Arg residue pair.\textsuperscript{367} The results shown in Fig. 6.5 show that this region does not present a significant free energy barrier to Na\textsuperscript{+} permeation. Instead, the current study, consistent with electrostatic calculations in our previous study (chapter 5)\textsuperscript{360}, suggests that the presence of a negative charge located at or above position 112 is crucial to impart selectivity for cations. This is affirmed by the only two cation selective forms of hHV\textsubscript{1}, namely, the proton selective wild-type and D112V-V116D.

Finally, the hydrophobic bottleneck of the channel characterized by the presence of an intermittent single-file chain of water molecules may support proton permeation as shown in previous studies of hydrophobic analogs of the gramicidin pore,\textsuperscript{141} synthetic hydrophobic cylindrical channels,\textsuperscript{368} and carbon nanotubes,\textsuperscript{369} where the hydrophobic wall provides an environment in which one-dimensional hydrogen-bond network is favoured over the three-dimensional networks often present in bulk water. In particular, it has been shown that the enhanced collective character of the dynamics of the protonated water wire in such a hydrophobic pore results in a (quasi-)concerted Grotthuss-like proton relay mechanism.\textsuperscript{141} In such a concerted process, the change of the excess proton is effectively delocalized through the length of the wire, dramatically lowering the electrostatic barrier to proton translocation. By analogy, the hydrophobic bottleneck of hHV\textsubscript{1}, may off a key advantage of H\textsuperscript{+} relative to other (cat)ions in its ability to selectively catalyze proton translocation through the hydrophobic bottleneck.
Chapter 7

Conclusions and Future Directions

7.1 Conclusions

The primary objective of my thesis was to determine the molecular basis for ion permeation and selectivity in hHv1, a voltage-gated proton channel that is ubiquitously present in the human body and a potential target for drug discovery. To this end, I have investigated the structure, dynamics and the energetic basis of wild-type and its mutants that either retain proton selectivity, display anion selectivity or nonconducting channel properties, using both unbiased and biased molecular dynamics simulations on the microsecond time scale.

In Chapter 3, I described a novel and tractable protocol for the study of solvation and orientation of membrane proteins in lipid membranes. The premise of the protocol is based on accelerating the convergence of an ensemble of protein structures differing in initial solvation conditions, to a consensus orientation. Further, I demonstrated that this approach is a computationally inexpensive way to study the self-association of amphipathic helices and therefore is a promising protocol to study the self-association of helix bundles of membrane proteins.

In Chapter 4, I described the development of the homology model and a novel MD protocol, also incorporating the method described in Chapter 3, to assess the model quality relative to the structural homologs of Hv1—the voltage-sensor domains (VSD) of K⁺ and Na⁺ channels. This protocol was based on a broad conformational sampling achieved in a computationally efficient manner. To the best of my knowledge, this thesis represents the most rigorously validated computational model of Hv1, using massively repeated molecular dynamics simulations. The sampling led to the identification of a consensus conformation for the protein structure. This general protocol can readily be applied to evaluate the
quality of homology models of other membrane proteins.

Furthermore, the structural properties predicted by this open-state model of Hv1 are consistent with experimental accessibility data, providing a valuable basis for further structural and functional studies in wild-type Hv1. Based on the electrostatic properties and the formation of a transient water-wire in this study, it was hypothesized that a water-mediated Grotthuss-like proton wire underlies the proton permeation mechanism of Hv1. In addition, I demonstrated, based on the predicted electrostatic properties, that the model is compatible with anion selectivity of the mutant model where Asp$^{112}$, the putative selectivity filter, is neutralized.

In Chapter 5, I presented a detailed comparative study of the structural and electrostatic properties of proton-selective, anion-permeable, and nonconducting mutants of Hv1 based on extensive, unbiased MD simulations. A key result of this study was that the salt-bridge network located in the vicinity of the selectivity filter becomes reorganized in the mutants, modulating the electrostatic properties of the pore consistently with the charge selectivity of the mutants. However, spontaneous ion permeating events were not observed in any of the systems in the microsecond time scale of the simulations, motivating the study presented in Chapter 6, in which free energy profiles for ionic movement through the pore were computed using biased sampling.

Free energy simulations of Na$^+$ and Cl$^-$ movement were conducted in wild-type and two anion-selective mutants. The movement of Na$^+$ is opposed in all three systems, which is qualitative consistent with the fact that none of the Hv1 variants are cation selective except for protons in the wild-type. The energetic barrier to Na$^+$ movement is primarily due to the desolvation penalty of the water-Na$^+$ complex at the narrowest point of the channel, which corresponds to the location of the side chain of Phe$^{150}$, which is highly conserved in VSDs. A significant barrier, yet decreasing in the system order of WT$>$D112S$>$D112V/V116S, opposes the movement of Cl$^-$. The barrier originates primarily due to the partially-dehydrated Cl$^-$ at the narrowest point of the channel. The notable displacement of the side chain of Phe$^{150}$ of D112V/V116S seems to be somewhat relaxing the constriction site, lowering the barrier to Cl$^-$ movement.

Collectively, my work provides insight that can be used to further the elucidation of molecular mechanism of proton translocation and for the development of inhibitory drugs against Hv1.
7.2 Future Directions

7.2.1 Refining the homology model

There are two main approaches to refine the homology model.

As demonstrated in Chapter 6, the hydrophobic constriction at residues Phe$^{150}$ may be too narrow to let Cl$^-$ through in the mutants. A possible way to refine the homology model is to consider alternate side chain conformations for residues Phe$^{150}$, using simulated annealing. These simulations would help sample side chain conformations of Phe$^{150}$ that were not accessible in studies described in this thesis. Using these potential basins of conformations, a systematic study of free energy simulations would help determine the role of Phe$^{150}$ in ion movement in Hv1.

Second approach entails the homology modelling procedure. There were three crystal structures of VSDs available at the time that were used as the templates for the development of the homology model for Hv1. Since then, a number of crystal structures of voltage-gated Na$^+$ and channels from different organisms and in different states of the channel have become available. VSDs of these structures could be potential templates for refining the homology model. However, it has been shown that the addition of multiple templates does not always guarantee improved homology models.$^{370}$ Therefore, new models can be generated with different sets of templates and ranked these models based on the protocol introduced in Chapter 4.

7.2.2 Testing drug binding to Hv1

A guanidine derivative, 2-guanidinobenzimidazole (2GBI), has been shown to bind Hv1 from the intracellular side of the membrane and act as a potential channel blocker. The residues responsible for the binding of 2GBI have been experimentally characterized.$^{371,372}$ In addition, the binding affinity of 2GBI to Hv1 has been quantified.$^{372}$ MD simulations can be performed to recapitulate these binding properties of 2GBI to Hv1, and thereby further the model validation of Hv1. More importantly, such simulations would potentially be insightful in providing molecular-level details for designing novel and higher-efficacy 2GBI analogs that are more potent.

7.2.3 Conducting quantum simulations

As described in Chapter 2, quantum simulations enable, in principle, the explicit treatment of chemical reactions such as the proton transfer. In addition, dissociable models such as the PM6 model used by Pomès & Roux$^7,141$ and in the Pomès lab before can be used to study Grotthuss translocation in the
water chain. This water model is more expensive than conventional forcefields but much less expensive than quantum simulations. Although simulations using PM6 would not take into account the proton relaying by D112, it is possible to obtain a PMF for protons through the hydrophobic bottleneck and determine whether or not the charge delocalization of the excess charge afforded by a concerted (or semi-concerted) proton hopping would result in a lower free energy barrier.
Bibliography


[74] SJ Li, Q Zhao, Q Zhou, and Y Zhai. Expression, purification, crystallization and preliminary crystallographic study of the carboxyl-terminal domain of the human voltage-gated proton channel.


[269] M Capasso, MK Bhamrah, T Henley, RS Boyd, C Langlais, K Cain, D Dinsdale, K Pulford, M Khan, B Musset, VV Cherny, D Morgan, RD Gascoyne, E Vigorito, TE DeCoursey, IC MacLen-


[305] FH Yu and WA Catterall. The VGL-chanome: a protein superfamily specialized for electrical

[306] AA Alabi, MI Bahamonde, HJ Jung, JI Kim, and KJ Swartz. Portability of paddle motif function

[307] SME Smith, D Morgan, B Musset, VV Cherny, AR Place, JW Hastings, and TE DeCoursey.
Voltage-gated proton channel in a dinoflagellate. *Proceedings of the National Academy of Sciences*,


[310] HP Larsson, OS Baker, DS Dhillon, and EY Isacoff. Transmembrane movement of the *Shaker* K+

[311] N Yang, AL George, and R Horn. Molecular basis of charge movement in voltage-gated sodium

[312] B Musset, SME Smith, S Rajan, VV Cherny, S Sujai, D Morgan, and TE DeCoursey. Zinc
inhibition of monomeric and dimeric proton channels suggests cooperative gating. *The Journal of

[313] F Tombola, MH Ulbrich, SC Kohout, and EY Isacoff. The opening of the two pores of the *Hv1*
voltagated proton channel is tuned by cooperativity. *Nature Structural & Molecular Biology*,

[314] VV Cherny and TE DeCoursey. pH-dependent inhibition of voltage-gated H+ currents in rat
alveolar epithelial cells by Zn2+ and other divalent cations. *The Journal of General Physiology*,

2011.

[316] JL England and G Haran. Role of solvation effects in protein denaturation: from thermodynamics


