# Computer Modelling the Root Cause of Cystic Fibrosis

by

Miro Alexander Astore

*A thesis submitted in fulfilment of the
requirements for the degree of*

Doctor of Philosophy

School of Physics
Faculty of Science
The University of Sydney

2022

Declaration of Original contribution

of the dissertation submitted by

Miro Alexander Astore

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

*Miro Alexander Astore*, Author                                   Date

# Abstract

placeholder text

# Acknowledgments

*"Fear cuts deeper than swords."*

Arya Stark

# List of Publications

MA - Miro Alexander Astore
SK - Serdar Kuyucak

1. placeholdertext

# Publication Authorship Attribution

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

_____                    _____

*Miro Alexander Astore*, Student                                                Date

As the supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

_____                    _____

*Serdar Kuyucak*, Supervisor                                                Date

# Contents

# List of Abbreviations

| | |
|---|---|
| $AMBER$ | Assisted Model Building with Energy Refinement |
| $BAR$ | Bennett-Acceptance-Ratio |
| $CF$ | Cystic Fibrosis |
| $CFTR$ | Cystic Fibrosis Transmembrane Conductance Regulator |
| $CHARMM$ | Chemistry at Harvard Macromolecular Mechanics |
| $COM$ | Centre of Mass |
| $CV$ | Collective Variable |
| $FEP$ | Free-Energy Perturbation |
| $gA$ | Gramicidin A Ion Channel |
| $Glt_{Ph}$ | Glutamate Transporter - *Pyrococcus horikoshii* |
| $GROMACS$ | GROningen MAchine for Chemical Simulations - MD program |
| $GROMOS$ | GROningen MOlecular Simulation - MD program |
| $LJ$ | Lenard-Jones Potential |
| $MBAR$ | Multistate Bennett-Acceptance-Ratio |
| $MD$ | Molecular Dynamics |
| $MetaD$ | Meta Dynamics |
| $NAMD$ | Nanoscale Molecular Dynamics - MD Program |
| $NBD$ | Nucleotide Binding Domain |
| $NPT$ | Constant number of Particles, Pressure and Temperature |
| $NVT$ | Constant number of Particles, Volume and Temperature |
| $OpenMM$ | Open Molecular Mechanics - MD Program |
| $OPLS$ | Optimised Potentials for Liquid Simulations |
| $PBC$ | Periodic Boundary Condition |
| $PCA$ | Principal Component Analysis |
| $PDB$ | Protein Data Bank |
| $PMF$ | Potential of Mean Force |
| $PME$ | Particle Mesh Ewald - Long-range Electrostatics Method |
| $POPC$ | 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine |
| $POPE$ | 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine |
| $RMSD$ | Root-Mean-Square Deviation |
| $TI$ | Thermodynamic Integration |
| $TICA$ | Time-lagged Indepenent Component Analysis |
| $US$ | Umbrella Sampling |
| $VMD$ | Visual Molecular Dynamics - MD Visualisation Program |
| $WHAM$ | Weighted Histogram Analysis Method |

# List of Figures

# List of Tables

*"Research is what I'm doing when I don't know what I'm doing."*

Wernher von Braun

# Chapter 1

# Introduction

## 1.1 Physics in a test tube

Why can't I write down an equation will tell me how long I will live? Or how many hairs I will grow?

This might seem like an inane question but if you asked a physicist for the formula for how long it takes a radioactive material to decay or how long it will take an object to fall into a black hole they will be able to answer easily.

What makes the first set of questions so much more difficult to answer?

I posit that it is the diversity of components that makes biological questions so difficult to ask and answer. Biology distinguishes itself amongst scientific disciplines requiring the study of systems that are both complex and heterogeneous. In the study of more simple physical systems a simple analogy such as a mass on a spring or a gas of hard spheres can be extremely successful in explaining macroscopic phenomena. For biological systems there appears to be too much complexity for such analogies to have the same level of success. They may struggle to answer questions such as "If this gene mutates how will that affect lung function?" "If this drug were given at a higher dosage what would its effect be?" "What if we change this chemical moiety?" At the moment, a trained chemist needs to go and answer these questions pipette in hand, the physicist with their notebook is hopeless.

It seems like a silly question but it seems important to ask why we can't just use a device similar to a harmonic oscilator or a perfect black body to speculate at useful answers for these quantitative questions. The answer is just as silly. If you look with your naked eye at your arm, you will notice hair, pores, dry skin, dead skin, perhaps even tendons and muscles under the the skin. If you take a miscroscope you will notice the 3 layers to your skin with different functions and composition. If you were to take a single cell from any of those layers and stain it to distinguish features in an electron

miscroscope you would notice all sorts of complex structures and the size and number of these structures would vary depending on where you took the cell from in the body. Within and between each those structures is a salty, wet dance of moelcules large and small. This heterogeneity on length scales hints at the reasons behind biology's physical complexity. Plasma physics is often characterised by the density of the plasma studied. This parameter may span 28 orders of magnitude from a dense stellar core to the sparse intergallactice nebulae. The same mathematical tools can be used to map any plasma in these energy scales. Would that we were so lucky in biology. We struggle to apply same physical models to deal with phenomena across a single order of magnitude.

Thus, in order to move towards more predictive theories of biology it is necessary to consider much more of the fundamental physical processes occuring within biological systems than simply searching for statistical trends. One form of this from fundamentals approach is the simulation of every atom in a biological system. Although computationally expensive, this approach appears necessary due to the heterogenous nature of biological systems.

## 1.2   What is Physics?

Personally I have always given answers along the lines of "the study of the movement of energy within a system" or when I was in high school "The study of how things move". Although adequate for a layman these might obscure the fundamental structure within physics that make it such a powerful tool. It is the conception of some causal unit in a system and the ability to scale up the behaviour of that unit to make predictions about measurable phenomena.

This might take a few different forms at different scales, it's what makes physics feel like the most "fundamental" of the sciences.

Examples include:

Newton's laws of gravitation to explain the organisation of the solar system.

Einstein's theories employing Reimannian geometry to track the motions of galaxies and black holes.

The conception of atoms as hard spheres used to derive the macroscopic behaviour of gasses.

The photon

The schrodinger wave function to find the structure of atoms, which can then be integrated further up to find their macroscopic organisations. More on this later.

Biological systems exhibit such a problem for the physicist because unlike the above problems it is extremely hard to pick out a fundamental unit to even begin our upwards journey. An evolutionary biologist might say to choose the "gene" but this is actually far too high in our spatial heirarchy already. Really a gene is only meaningful to the dance of life if it has partners to dance with. Genes of hard spheres ?

A coil of DNA in water doesn't really do much in solution except decay without machinary that can preserve, read, translate and replicate it. The gene is an emergent property, we have to go deeper.

So, what creates the gene?

A slew of biological machinary that mostly take the form of proteins. These proetins are then coded for by the DNA in a strange loop.

This self referential loop is one of the reasons biology is so difficult. Since we know that this strange loop is kicked off by atomic interactions we will start there. As we are taking a physical, pragmatic approach here it would make sense to begin with the protein, after all, they stave off the march of entropy constantly trying to eat up all of your cells. It also just so happens that they are much easier to understand computationally since their motions are faster and more flexible.

The first level sub cellular organisation is perhaps the most intimdating first step for me personally after spending 4 years simulating a single protein. Glimpsing the complexity within a single one of these molecules has been one of the most existential experiences of my life but the knowledge that there are astronomical nubmers of these things inside me all of the time

It is hoped that illustrating the monumental task in both intellectual effort and resources of incrementally increasing the understanding of a single protein amongst tens of thousands will give the reader and understanding of how we might continue our quest to understand the molecular dance that plays within all of us.

This makes sense if we think about it Somewhere on the scale between a single protein and a single cell this is what we consider "life". We have single unicellular organisms but we don't have uniproteomic organisms. So the fundamental length scale of life is somewhere between $10^{-10}m$ and $10^{-3}m$. This is the first loop in our strange loop.

After this things start to run away from me with my handful of GPUs and limited patience. Once we move from prokaryotes to eukaryotes we have gone a few levels deeper. There is of course unicellular eukaryotes but how did we get from P to E? I'll have to leave that one for evolutionary cell biologists. Certainly there is something strangely loopy about the appropriation of cells by other cells. Then we have something more interesting, cellular collectivisation.

Cells clump together and act in unison to give us colonial organisms. (Self-similar colony morphogenesis by gram-negative rods as the experimental model of fractal growth by a cell population). Like any advanced economy cells .

Biological strange loops would not seem to be as self similar as the clean nice logics in the strange loop of the Godelian knot. Why is this?

## 1.3   Why Cystic Fibrosis

The sad truth of this debiletating disease is that those afflicted are extremely unlucky. A single, small change to the genome and their lungs fill with sticky mucus and become

infected with bacteria, making every breath cumbersome. Personally, I've not met somebody who has this disease. I have consistently wondered what perpsective I'm missing by not suffering myself from such a condition or even knowing somebody with it. I'm a relatively healthy well adjusted Male. I have not been trained in the ethics of studying medicine and my undergraduate professors were only concerned with what was morally acceptable when it came to mathematical theorems.

In this way, my motivations for studying this disease aren't wholly humanitarian. There is a perspective on protein evolution which states that the primary sequence of a particular gene contributes to the overall fitness of an organisms by a formula. []

It just so happens that the CFTR gene sits at the precipice of a daunting cliff in sequence space. So by taking small steps in sequence space and plunging down this cliff we can try to understand how we might push the ball back up the cliff and retain functionality.

Moreover, by learning the nuts and bolts of what goes wrong with CFTR we can start to think about where some of these cliffs might be in other places in the proteome, to gain function and avoid disease and debiletation..

# Chapter 2

# Theory and Methods

> *"In the real world, this could eventually mean that most chemical experiments are conducted inside the silicon of chips instead of the glassware of laboratories. Turn off that Bunsen burner; it will not be wanted in ten years."*
>
> The Economist, on the 1998 Chemistry Nobel Prize Awardees
> Walter Kohn and John Pople

## 2.1  Quantum Mechanical Origins

For a microscopic system, the nature of the interactions between particles/atoms is fundamentally quantum mechanical. This quantum mechanical system contains both nuclei $N_n$ and electrons $N_e$. Suppose $i$ and $\alpha$ represent the indices for individual nuclei and electrons, then the coordinates can then be written as $\mathbf{R}_i$ and $\mathbf{r}_\alpha$ respectively. The interaction of this many-body system is described with the time-dependent Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\Psi(\{\mathbf{R}_i, \mathbf{r}_\alpha\}, t) = H\Psi(\{\mathbf{R}_i, \mathbf{r}_\alpha\}, t), \tag{2.1}$$

where $\Psi$ and $H$ is the wave function and Hamiltonian of the nuclei-electron system, respectively. Solving this equation can be extremely complicated if not impossible; thus, some approximations are required. Instead of solving the equation above the time-independent Schrödinger equation can be solved instead

$$H\Psi(\{\mathbf{R}_i, \mathbf{r}_\alpha\}) = E\Psi(\{\mathbf{R}_i, \mathbf{r}_\alpha\}), \tag{2.2}$$

where $E$ is the total energy. This is valid only for stationary states (i.e. excluding excited states). The Hamiltonian for the system of interacting nuclei and electrons is written as the sum of several components

$$H = T_n + T_e + V_{n\text{-}n} + V_{e\text{-}e} + V_{n\text{-}e}. \tag{2.3}$$

The first two terms are the kinetic energies of nuclei and electrons, and the last three terms are the nucleus-nucleus, electron-electron and nucleus-electron potential energy.

The potential energies can be represented as Coulomb interactions, and thus the individual terms above are expanded to the set of equations below,

$$T_n = -\sum_i \frac{\hbar^2}{2M_i}\nabla_i^2, \qquad\qquad V_{n\text{-}n} = \sum_{i>j} \frac{z_i z_j e^2}{|\mathbf{R}_i - \mathbf{R}_j|},$$

$$T_e = -\sum_\alpha \frac{\hbar^2}{2m}\nabla_\alpha^2, \qquad\qquad V_{e\text{-}e} = \sum_{\alpha>\beta} \frac{e^2}{|\mathbf{r}_\alpha - \mathbf{r}_\beta|},$$

$$V_{n\text{-}e} = \sum_{i,\alpha} \frac{z_i e^2}{|\mathbf{R}_i - \mathbf{r}_\alpha|}. \tag{2.4}$$

Here $M_i$ and $m$ are the masses of the nuclei and electrons, respectively. The total wave function $\Psi(\{\mathbf{R}_i, \mathbf{r}_\alpha\})$ can be separated into the product of the nuclei $\Psi_n(\{\mathbf{R}_i\})$ and electrons $\Psi_e(\{\mathbf{R}_i, \mathbf{r}_\alpha\})$ wave functions. This decoupling is valid since the nuclei are much heavier than the electrons, so the motion of the nuclei is independent of the electrons. This *ansatz* is known as the adiabatic or Born-Oppenheimer approximation [63]. However, the motion of the electrons will still be dependent on the nuclei due to the electron being localised to the nuclei. With this approximation, the time-independent Schrödinger equation for the electrons and nuclei respectively are

$$[T_e + V_{e\text{-}e} + V_{n\text{-}e}]\Psi(\{\mathbf{R}_i, \mathbf{r}_\alpha\}) = E_e(\{\mathbf{R}_i\})\Psi_e(\{\mathbf{R}_i, \mathbf{r}_\alpha\}), \tag{2.5a}$$

$$[T_n + V_{n\text{-}n} + E_e(\{\mathbf{R}_i\})]\Psi(\{\mathbf{R}_i\}) = E\Psi_n(\{\mathbf{R}_i\}. \tag{2.5b}$$

Eq. (2.5a) is the starting point for any *ab initio* electronic calculations and $E_e(\{\mathbf{R}_i\})$ is the total energy of the electrons (usually taken at the ground state). To further simplify the problem, the nuclei interactions are treated classically. Following the definition of the conservative force $F = -\nabla V$ (force depending only on position) the motion of the nuclei is dictated by Newton's equation of motion

$$M_i\ddot{\mathbf{R}}_i = -\nabla_i V(\{\mathbf{R}_i\}), \tag{2.6}$$

where $V(\{\mathbf{R}_i\})$ is a potential energy function. Using the definitions from Eqs. (2.4) and (2.5b) the potential energy function that describes nuclei interactions is

$$V(\mathbf{R}_i) = \sum_{i>j} \frac{z_i z_j e^2}{|\mathbf{R}_i - \mathbf{R}_j|} + E_e(\{\mathbf{R}_i\}), \tag{2.7}$$

which is the nucleus-nucleus Coulomb interaction plus the ground state electronic energy. The potential energy function above thus contains both classical and quantum mechanical description. Eqs. (2.5a), (2.6) and (2.7) are the foundations for *ab initio* molecular dynamics (AIMD) simulations. Although AIMD gives a very accurate description of the microscopic interactions, simulating biomolecular systems (consisting of $\sim 10^5$ atoms) is still unreachable with today's computing hardware. This is true even with the density functional theory (DFT) approximation [64, 65] coupled with Car-Parrinello MD (CPMD) [66], which is the "state-of-the-art" AIMD method. CPMD can simulate system sizes of $\sim 10^3$ atoms but only reaching time scales of $\sim 10^2$ picoseconds with the current hardware. Therefore, a classical approximation of the electronic interactions Eq. (2.7) is needed for simulating biomolecular systems.

## 2.2  Classical Molecular Dynamics

The quantum mechanical part of Eq. (2.7) can be decomposed into three components: *repulsion*, *dispersion* and *polarisation*. Electron repulsion is a consequence of the Pauli exclusion principle where two electrons cannot occupy the same quantum state. The repulsive force is short-range and ensures that electron orbitals do not overlap. Dispersion forces are attractive and arise from weak dipole-dipole interactions (van der Waals or London dispersion forces) that have a $1/r^6$ dependence. Lastly, polarisation is the ability of electron orbitals to be distorted under the influence of an external electric field. All three of these interactions can be approximated classically through the use of empirically fitted analytical functions.

### 2.2.1  Empirical Force Fields

Most MD simulations of biomolecules employ a force field to describe the potential energy function of the system. Some commonly used force field packages include CHARMM [9], AMBER [67], OPLS [68] and GROMOS [69]. These force field packages describe atomic interactions using analytical functions that are fitted to some target data. The target data may be from experiment or accurate *ab initio* calculations (or a mixture of both). In this thesis, all simulations utilise the CHARMM additive force field. Firstly, the electrostatic interactions between two atoms $i$ and $j$ are calculated with Coulomb's law

$$V_{Elec}(r_{ij}) = k_c \frac{q_i q_j}{r_{ij}}, \tag{2.8}$$

where $k_c$ is the Coulomb constant. This potential is essentially the classical part of Eq. (2.7). For the quantum mechanical part, the repulsive and dispersive forces can be combined into one equation known as the 12-6 Lennard-Jones (LJ) potential

$$V_{LJ}(r_{ij}) = \epsilon_{ij} \left[ \left( \frac{R_{\min_{ij}}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\min_{ij}}}{r_{ij}} \right)^6 \right], \tag{2.9}$$

where $\epsilon_{ij}$ is the well depth and $R_{\min_{ij}}$ is the position of the potential at the minimum. In the literature, the LJ potential is usually written terms of $\sigma_{ij}$, which is the distance at which the energy is zero, and the $\epsilon_{ij}$ has a constant of 4 (the constant of 2 inside is removed). The relationship between the two distances is $R_{\min} = 2^{1/6}\sigma$. While the dispersive term comes out of quantum mechanics, the $1/r^{12}$ dependence for the repulsive term was adopted for simplicity. This term was chosen due to the hardware available when John Lennard-Jones first proposed the potential function, and it is easier to calculate $1/r^{12}$ than some other term because it is the square of $1/r^6$.[1] The two constants in the equation above are parameters that need to be optimised for each atom type. The LJ parameters between pairs of atoms are combined using the Lorentz-Berthelot combination rule [1]. This rule states that for any pairs of atoms, $\epsilon_{ij}$ and $R_{\min_{ij}}$ is equal to the geometric and arithmetic mean of the individual components, respectively. Mathematically this is expressed as

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}, \qquad R_{\min_{ij}} = \frac{1}{2}(R_{\min_i} + R_{\min_j}). \tag{2.10}$$

---

[1]Some force fields like the AMOEBA force field uses a 14-7 LJ [70].
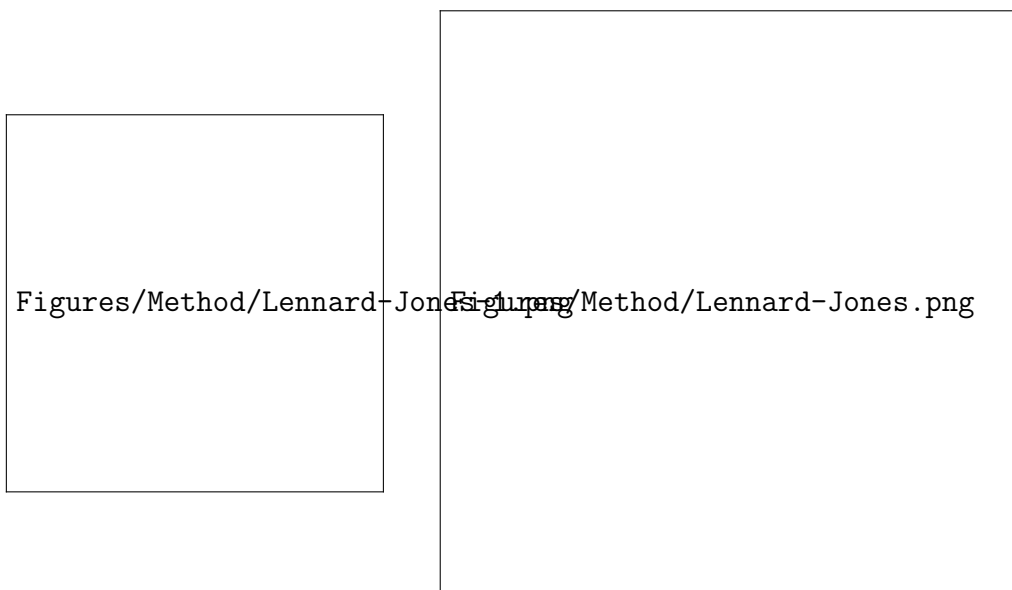
**Figure 2.1:** (*Left*) Two different atoms separated by $r_{ij}$ and (*Right*) the LJ potential between the two atoms. The total potential is the sum of the repulsive and dispersive (attractive) terms. (Image on the left is designed in Tikz and the plot on the right is generated in Matlab)

The rule above is based on the hard-sphere model of particles and prevents any overlap between atoms. Fig. 2.1 illustrates the dual nature of the LJ potential (dispersive and repulsive). The effect of polarisation is not included in the CHARMM additive force field; hence, it is classified as a non-polarisable force field. Polarisation in classical force fields can be included by using the classical Drude model of charge on a spring or direct induced polarisation with multipole expansion [71]. Packages that include explicit polarisation using these methods are CHARMM Drude force field [72] and the AMOEBA force field [70]. Although polarisable force fields are available, they are generally more computationally expensive than non-polarisable force fields (though the implementation of the Drude polarisable force field in NAMD only increases the computational cost not more than a factor of two compared to non-polarisable force fields [73]). Also, compared to polarisable force fields, non-polarisable force field has been in development since the 1980s. Thus the validity of their chemical accuracy is better understood than polarisable force fields. The potentials $V_{Elec}$ and $V_{LJ}$ describe interactions between pairs of atoms; however, they do not capture the topology of molecules. Thus the sum of these two potentials is referred to as *non-bonded* interactions, $V_{nonbonded}$.

For molecules, extra potential energy functions are needed to describe the many-body interactions. The potential energy function of Eq. (2.6) can be expanded as an infinite series

$$V(\{r_i\}) = \sum_{i>j} V_{\text{2-body}}(r_i, r_j) + \sum_{i>j>k} V_{\text{3-body}}(r_i, r_j, r_k) + \sum_{i>j>k>l} V_{\text{4-body}}(r_i, r_j, r_k, r_l) + ...$$

$$(2.11)$$

Almost all force field packages available include the expansion above only up to the fourth term to simplify the problem. The 2-body interactions, i.e. bonds between two atoms, is described with a harmonic oscillator (Fig. 2.2 A) with a potential energy

function in the form of

$$V_{bonds} = K_b(b - b_0)^2, \tag{2.12}$$

where $b$ is the bond length (i.e. $|r_i - r_j|$) and $b_0$ is the equilibrium position. The spring constant above contains the factor of $1/2$ as per the convention used in CHARMM (i.e. $K = k/2)^2$.

The 3-body interactions describe the bending of two bonds formed between three atoms. For example, the H-O-H angle of a water molecule is 104.5°, and a potential function is required to preserve the shape of the molecule. In CHARMM there are two different types the 3-body potentials as shown below

$$V_{angles} = K_\theta(\theta - \theta_0)^2 + K_{ub}(S - S_0)^2, \tag{2.13}$$

where $\theta$ is the angle between atoms $i$-$j$-$k$, $S$ is the distance between atoms $i$-$k$ and $\theta_0$ and $S_0$ are the respective equilibrium positions. The first term is a harmonic potential applied on the angle between two bonds as illustrated in Fig. 2.2 B, while the second part is a harmonic potential applied to the distance between atoms 1 and 3. This distance potential is known as the *Urey-Bradley* potential, which is used to include angle bending in some molecules.

The 4-body interactions are called dihedral potentials, which determines the flexibility of molecules. The dihedral potentials used in CHARMM is in the form of

$$V_{dihedrals} = K_{proper}(1 + cos(n\phi - \delta)) + K_{improper}(\varphi - \varphi_0)^2. \tag{2.14}$$

Here $n$ is the multiplicity, $\delta$ is the phase angle, and $\phi$ and $\varphi$ are the *proper* and *improper* dihedral angles respectively. The first potential above describes proper dihedrals, which tries to capture molecular conformations that is a consequence of $\pi$-bonding, etc., and the constant $K_{proper}$ is the energy barrier height instead of a spring constant. Without proper dihedrals, molecules like polymers would not be able to rotate or change conformations. The second 4-body potential is the improper dihedrals, which keeps a group of atoms flat on a plane. One example where this might be used is in a benzene molecule. Since the carbon atoms are $sp^2$ hybridised, the benzene molecule must be planar and improper dihedrals, try to mimic this behaviour. The dihedral potentials are illustrated in Figs. 2.2 C–D. The sum of the bond, angles and dihedral potentials is equal to the *bonded* potential $V_{bonded}$ that describes the topology of a molecule. The potential energy function that is used in Eq. (2.6) to replace the quantum mechanical potential is

$$V_{System} = V_{bonded} + V_{nonbonded},$$

$$V_{bonded} = \begin{cases} V_{bonds}(b) \\ V_{angles}(\theta, S) \\ V_{dihedrals}(\phi, \psi) \end{cases} + V_{nonbonded} = \begin{cases} V_{LJ}(r_{ij}) \\ V_{Elec}(r_{ij}) \end{cases} \tag{2.15}$$

Newton's equation of motion is numerically integrated with the potential energy function above, and, if appropriately parametrised, should be able to describe biomolecular systems accurately in comparison to experiment or *ab initio* calculations.[3]

---

[2]This Convention is kept from the 1980s since it removes the need to calculate $k/2$, reducing the amount of unnecessary computation at each integration step.

[3]Although some systems of interest may require the explicit treatment of polarisation.

## 2.2.2   Numerical Integration

Given the potential energy function of the system, we can determine the force acting on each atom. If the force on each atom is known, then we can determine the motion of each atom following Newton's equation of motion. Using Newton's second law and the conservative force we arrive at

$$F_i = m_i \frac{d^2 r_i}{dt^2} = -\nabla_i V(\mathbf{r}_i), \tag{2.16}$$

where $m$ is the mass of atom $i$ and $r_i$ is the current position of the atom. The positions and velocities of each atom are obtained by integrating the acceleration $a_i = d^2 r_i/dt^2$. The integration is performed numerically by discretising the time $dt$ with a small timestep $\Delta t$. The time step $\Delta t$ needs to be chosen carefully as a large integration step will lead to numerical instabilities. A time step of 1 fs is needed to capture the vibrational motion of hydrogen atoms, however, constraint algorithms like SHAKE [74] and LINCS [75] reduces the hydrogen degrees of freedom (rigid water molecules) and a time step of 2 fs can be used instead. A further trick that can be used to increase the time step is to increase the mass of the hydrogen atoms. This is known as hydrogen mass repartitioning (HMR) and allows an increase in the time step to 4 fs [76]. However,



**Figure 2.2:** Bonded terms describing the (A) 2-body bonds (B) 3-body angle and 4-body (C) proper and (D) improper dihedral interactions used in force fields. (Images are designed in Tikz)

the larger integration step is associated with a larger numerical error. All simulations performed in the investigations carried out in this thesis use a time step of 2 fs by employing the SHAKE algorithm.

There are a few computer algorithms that are used for numerically integrating the acceleration. The NAMD program uses an efficient *velocity-Verlet* method [77]. This method requires the evaluation of velocity twice, but only one calculation is needed for the position and force. The position, force and velocity is updated at every time step following the algorithm below

$$v_{n+\frac{1}{2}} = v_n + \frac{F_n}{m}\frac{\Delta t}{2}, \tag{2.17a}$$

$$r_{n+1} = r_n + v_{n+\frac{1}{2}}\Delta t, \tag{2.17b}$$

$$F_{n+1} = F(r_{n+1}), \tag{2.17c}$$

$$v_{n+1} = v_{n+\frac{1}{2}} + \frac{F_{n+1}}{m}\frac{\Delta t}{2}. \tag{2.17d}$$

The most expensive calculation in the steps above is the force evaluation (2.17c) as this requires the computation of the interactions between all atoms. The computation time increases with a factor of $N^2$ where $N$ is the number of atoms. One way of reducing computation time is the use of a cut-off distance $r_c$ when evaluating atomic interactions. In all simulations performed in this thesis, a cut-off of 12 Å with a switching distance $r_s$ of 10 Å is used for the LJ interactions to limit the computational cost. A switching function replaces the potential with a smooth function so that the energy approaches zero at 12 Å starting from 10 Å (see Fig. 2.3 for an illustration). The use of a cut-off distance is justified because the LJ interactions drop by a factor of $1/r^6$, and beyond 12 Å the LJ interactions become negligible. However, the Coulomb interactions have a $1/r$ dependence[4], and a 12 Å cut-off is too short to capture this interaction accurately. Thus instead of increasing the cut-off distance, the long-range electrostatic interactions are determined using the particle-mesh Ewald (PME) summation method [78]. This method requires the use of periodic boundary conditions (PBC). With PME, the simulation box is divided into a finite grid, and the charges in the system are mapped onto the grid points. The fast Fourier transform (FFT) method is used to calculate the long-range electrostatic force of the charges on the grid. Calculating the long-range electrostatic force this way gives an $N.\log(N)$ scaling instead of $N^2$.

### 2.2.3   Thermodynamic States

An important aspect to consider when running MD simulations is the thermodynamic state of the system. A thermodynamic state can be defined by parameters such as the number of particles $N$, temperature $T$, pressure $P$ and chemical potential $\mu$. Fixing these parameters, i.e. making them constant throughout the simulation, will set the thermodynamic state or the *ensemble* of the system. During MD simulations, points are generated in phase space (momentum-position space) and the volume of these points gives the total number of states (permutations) of the system. Thus the ensemble of a system relates the possible permutation of the microscopic system to the observable

---

[4]This dependence is from the monopole term of point charges
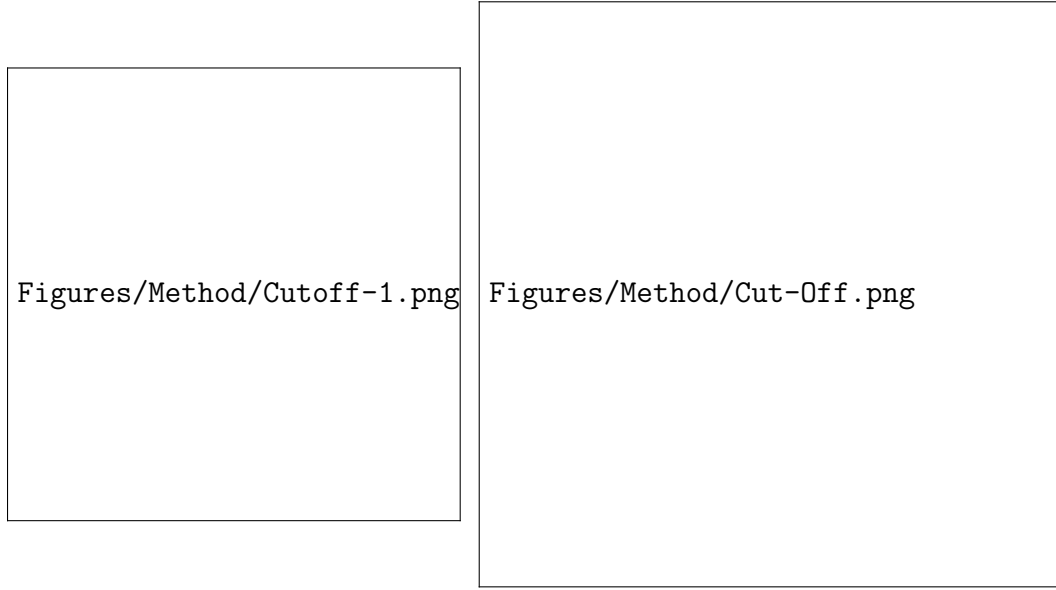
**Figure 2.3:** (Left) The cut-off distance $r_c$ used in MD simulations to reduce computational cost. A switching function is used, starting from $r_s$ to $r_c$ to smooth the potential to zero (Right). (Image on the left is designed in Tikz and the plot on the right is generated in Matlab)

macroscopic states. Integrating Eq. (2.16) will intrinsically give a constant number of atoms $N$, volume $V$ and energy $E$ (NVE). MD simulations with this type of thermodynamic state will be in a *microcanonical* ensemble, which represents an isolated system. Since biomolecular systems in real life are not isolated, modification to the equation of motion is required to incorporate other thermodynamic states.

**Temperature Control**

Temperature is a human-derived quantity that is used for convenience to relate how "hot" or "cold" something is. In thermodynamics, however, the temperature is related to the average kinetic energy of all particles in the system. To keep the temperature constant in MD simulations, a thermostat is needed that "kicks" the particles back to the desired kinetic energy. One way to do this is to employ Langevin dynamics that is based on the Langevin equation

$$\frac{d^2 r_i}{dt^2} = \frac{F_{i,\text{sys}}(r)}{m_i} - \gamma \frac{dr_i}{dt} + \sqrt{\frac{2\gamma k_\text{B} T}{h}} R_i(t), \tag{2.18}$$

where $F_\text{sys}(r)$ is the total force on an atom from all other atoms in the system, $\gamma$ is a friction coefficient, $h$ is Planck's constant and $R(t)$ is a univariate Gaussian noise. This equation is essentially a damped harmonic oscillator with random "kicks" that follows a Gaussian distribution. The noise term represents small Brownian particles that collide with the atoms in the system to push the temperature of the system (i.e. kinetic energy) to the set constant value. With the temperature held constant, the system running under Langevin dynamics is in the *canonical* ensemble (NVT).

**Pressure Control**

In order to control the pressure of the system, the volume of the box must change. This is necessary to run MD simulations in an isobaric-isothermal ensemble (NPT). The barostat[5] employed in NAMD is a hybrid Nosé-Hoover Langevin piston method [79]. In the Nosé-Hoover method, a fictitious particle $e$ is used to represent a "piston" acting on the system box [80]. The piston can compress and expand the box isotropically or anisotropically with an oscillation period of $\tau$. The motion of the piston is governed by the Langevin equation and thus the dynamics of the piston is given by

$$\frac{de}{dt} = \frac{1}{3V}\frac{dV}{dt}, \tag{2.19a}$$

$$\frac{d^2e}{dt^2} = \frac{3V}{W}(P - P_0) - \gamma_e\frac{de}{dt} + \sqrt{\frac{2\gamma_e k_{\mathrm{B}}T}{h}}R_e(t), \tag{2.19b}$$

$$W = 3N\tau^2 k_B T, \tag{2.19c}$$

where $V$ is the volume of the box, $P$ is the pressure, $W$ is the mass of the fictitious particle (i.e. the piston), and $R_e(t)$ is a similar Gaussian noise in Eq.(2.18) for the piston. The beauty of this method is that only a small modification to the equation of motion is required for the velocity and acceleration terms. The velocity and acceleration of atoms are modified by $(de/dt)r$ and $(de/dt)v$ respectively. Therefore, the positions of each atom are scaled to the change in box dimensions automatically as a result of the piston motion.

The friction or damping coefficients $\gamma$ and $\gamma_e$ determines the coupling strength of the simulation to the thermostat and barostat, respectively. This parameter is set by the user in units of ps$^{-1}$ and the oscillation period $\tau$ scales by $1/\gamma_e$. This ensures that the piston dynamics is close to being critically damped. Using a large value for the decay might affect the properties like diffusion of molecules as it prioritises equilibration. Therefore, an appropriate value needs to be used to give a stable simulation run.

## 2.3   Free Energy Calculations

> *"Thermodynamics is a funny subject. The first time you go through it, you don't understand it at all. The second time you go through it, you think you understand it, except for one or two small points. The third time you go through it, you know you don't understand it, but by that time you are so used to it, it doesn't bother you any more."*
>
> Arnold Sommerfeld

Simulations of biomolecules with MD can give some information about the behaviour and dynamics of the system. However, in most cases, the time scales of processes like protein-ligand binding, ion diffusion through ion channels and protein folding is unreachable with brute-force MD. This is where free energy calculations are used to extract useful information about the system. Free energy methods can give quantities

---

[5]Analogous to thermostat for pressure

like solvation and binding free energies, which can be directly compared to experiments. Such a tool is valuable for the validation of force field models/parameters or complements experimental results to elucidate some processes. There are many different ways of computing free energies, and they all can be classified into two categories *path-independent* and *path-dependent*.

## 2.3.1 Alchemical Perturbation

Path-independent free energy methods are focused on obtaining the free energy difference between two states. For example, the binding free energy of a ligand to a protein host is equal to the change in free energy between the ligand in bulk water (point **A**) and inside the protein (point **B**). The path the ligand takes from **A** to **B** is not considered in the calculation; however, if the path is an essential factor in capturing the dynamics, then a path-dependent method is required (see Section 2.3.2). To obtain the free energy between two points, the ligand or molecule is "*alchemically*" transformed into a void or another species. This means that the nonbonded interactions of the molecule with the environment are switched on and off. The alchemical transformation is achieved by introducing a hybrid potential energy in the form of

$$U(\lambda) = (1 - \lambda)U_0 + \lambda U_1, \tag{2.20}$$

where $U_0$ and $U_1$ are the potential energies for the initial and final states, respectively. The variable $\lambda$ is a discreet coupling parameter that ranges from 0 to 1 and determines the degree of perturbation of the system. As an example, $\lambda = 0$ may represent a positive ion surrounded by water, and $\lambda = 1$ is the state where the ion is completely annihilated (illustrated in Fig. 2.4). At each $\lambda$ value, the system is sampled with the modified potential energy. In practice, a certain amount of data is discarded as the equilibration phase[6] and simulation data after this is taken as the production data. Simulations that use this modified potential energy are referred collectively as free energy perturbation (FEP).

**Exponential Averaging**

Given the modified potential energy of the system, the free energy can be obtained using several methods. The earliest method to obtain free energy of a perturbed system is *exponential averaging*[7] (EXP) [81]. In essence, at each time step of $\lambda_i$, the energy at both $\lambda_i$ and $\lambda_{i+1}$ is calculated. The energy at $\lambda_{i+1}$ is calculated with the coordinates of $\lambda_i$. The Boltzmann probability of the energy difference between these two states is averaged over the number of samples at each state. Then, the Gibbs free energy at $\lambda_i$ is given by

$$\Delta G_i = -k_\mathrm{B}T \ln \left\langle \exp\left[-\frac{U(\lambda_{i+1}) - U(\lambda_i)}{k_B T}\right] \right\rangle_{\lambda_i}, \tag{2.21}$$

---

[6]In Monte Carlo simulations this is referred as the "burn" phase
[7]Also known as the Zwanzig relation

**Figure 2.4:** An ion surrounded by water molecules is alchemically destroyed starting from $\lambda = 0$ to 1. (Image is designed in Tikz)

where $k_\mathrm{B}$ is the Boltzmann constant and $T$ is the temperature. The sum gives the free energy of the whole transformation (i.e. $\lambda$ from 0 to 1)

$$\Delta G = \sum_{i}^{N} \Delta G_i. \tag{2.22}$$

Choosing the number of $\lambda$ points is tricky and is system dependent. For example, a previous study showed that 130 equally spaced $\lambda$ values were needed to obtain convergence for the free energy of translocating a charged ligand from the binding site in $\mathrm{Glt}_{\mathrm{Ph}}$ to bulk water [82]. However, using exponentially spaced values instead reduces the number of windows to 66. For calculating solvation free energies of small molecules, however, a smaller number of $\lambda$ values is needed. Different strategies may be adopted to obtain convergence, and one recipe may not apply to all systems. The work in this thesis follows the strategy given in Ref [82] as the same system (i.e., $\mathrm{Glt}_{\mathrm{Ph}}$) is investigated.

**Thermodynamic Integration**

The large number of $\lambda$ values required with EXP can make it very expensive to compute. An alternate way to calculate free energy with the alchemical route is to use thermodynamic integration (TI) [81]. In this method, the ensemble average of the derivative $\partial U/\partial \lambda$ is calculated instead. The free energy difference is obtained by numerically integrating the derivative of the hybrid potential energy with respect to $\lambda$,

$$\Delta G = \int_0^1 \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda. \tag{2.23}$$

With TI, the number of $\lambda$ values can be significantly reduced. For small neutral molecules, 11 windows are sufficient to capture the free energy profile accurately. For charged molecules, the free energy profile is quadratic[8], and Gaussian quadrature can be used for the integration [83]. Gaussian quadrature is a numerical method that approximates an integral with weighted sums of function values at the specified point, i.e.

$$\int_0^1 f(x) \approx \sum_i^n \omega_i f(x_i), \tag{2.24}$$

where $\omega_i$ is the weight value at a point $i$ and $f(x_i)$ is the function value at $x_i$. Applying this to TI, the function $f(x_i)$ represents the free energy derivative at $\lambda_i$. In a previous study of ion binding in gA with TI, a seven-point quadrature was found to be sufficient [40]. Therefore, more extensive sampling can be done for each $\lambda$ value allowing a more accurate estimate of the free energy.

**Bennett Acceptance Ratio**

Although TI is a good alternative, obtaining the free energy is very different from the EXP method. If the difference in total energy is preferred than the derivative, the Bennett-Acceptance-Ratio (BAR) [84] can be used instead to get a much better estimate of the free energy. In essence, the method requires the calculation of the potential energy difference $\Delta U = U_B - U_A$ between two neighbouring lambda points like in Eq. (2.21) or between the forward and backward transformation at $\lambda_i$. The free energy difference is recovered by calculating the logarithm of the ratio of the partition function $Q$ between the two states, i.e.

$$\Delta G = -k_{\mathrm{B}} T \ln \frac{Q_1}{Q_0}. \tag{2.25}$$

The partition function is obtained by using a maximum likelihood estimator of the potential energy difference and takes the form of the Fermi-Dirac distribution

$$f(x) = \frac{1}{1 + \exp(x)}. \tag{2.26}$$

Considering Eqs. (2.25) and (2.26), the free energy difference is estimated using the equation below

$$\Delta G = -k_{\mathrm{B}} T \ln \left[ \frac{\langle f[-(\Delta U - C)/k_B T] \rangle_1}{\langle f[(\Delta U - C)/k_B T] \rangle_0} \right]. \tag{2.27}$$

---

[8]This comes from nature of Coulomb interaction, i.e. $q^2$

The constant $C$ and the free energy $\Delta G$ are two unknown variables; thus Eq (2.27) is solved iteratively. The constant $C$ is initially given a value of zero, and the free energy $\Delta G$ is calculated using the equation above, which takes in the data from the alchemical simulation. The calculated value of $\Delta G$ in step $i$ is used as the value for $C$ in the next step $i + 1$, i.e.

$$C_{i+1} = \Delta G_i. \tag{2.28}$$

This process is repeated until the value of $\Delta G_i$ and $C_{i-1}$ is equal to each other (or numerically the same up to a given tolerance value like $1 \times 10^{-5}$). A BAR estimator called `ParseFEP` [85] is implemented as a VMD plugin and can be used to process files generated by NAMD. BAR gives a more accurate estimate of $\Delta G$ than EXP since both forward and backward transformations at each $\lambda$ point are considered in the estimator. In comparison, EXP only considers the average of the endpoints (i.e. at $\lambda = 1$) when combining the forward and backward data.

**Multistate Bennett Acceptance Ratio**

The BAR method, as described previously, can be generalised to include all states from $\lambda = 0$ to 1 using the multistate Bennett-Acceptance-Ratio (MBAR) [86]. MBAR reduces to BAR when only two states are considered and hence the given name. In this method, a *reduced potential* for thermodynamic state $i$ is defined depending on the ensemble of the system

$$u_i(\boldsymbol{x}) = [U_i(\boldsymbol{x}) + p_i V(\boldsymbol{x}) + \mu_i^T n(\boldsymbol{x})]/k_\mathrm{B} T, \tag{2.29}$$

where the first term is the potential energy function and the second and third terms are energies due to pressure and chemical potentials, respectively. To generate the reduced potential $u(\boldsymbol{x})$ in FEP calculations, the results from different states are combined. At $\lambda_j$, the energy of the system is calculated in the same way as in the EXP method. However, the energy at different $\lambda$ states is also calculated using the same coordinates used to calculate the energy at $\lambda_j$. This procedure is continued for the next time step up to $N$ trajectory steps and repeated for the next state $\lambda_{j+1}$ up to $K$ states. Fig. 2.5 gives an illustration of the reduced potential in matrix form. Once $u(\boldsymbol{x})$ is defined, the free energy is recovered using the equation below[9]

$$\Delta G_i = -k_\mathrm{B} T \ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{\exp[-u_i(\boldsymbol{x}_{jn})]}{\sum_{k=1}^{K} \exp[\Delta G_k - u_k(\boldsymbol{x}_{jn})]}, \tag{2.30}$$

where $j$ and $k$ are indices over $\lambda$ states and $n$ is an index over frames/trajectory data. Similar to BAR, $\Delta G_k$ is unknown initially and is solved with an iterative method. The final free energy from $\lambda = 0$ to 1 is calculated using Eq. (2.22). A python implementation called *pymbar* is available from https://github.com/choderalab/pymbar. Since the results from all states are combined, MBAR is regarded as a much better estimator than EXP or BAR and gives the smallest variance compared to all other FEP methods [86]. However, this can be computationally expensive if many states are considered, and the data is large. In this case, it is better to use the uncorrelated data in Eq. (2.30) instead and is determined with the `timeseries` module as part of the *pymbar* package [87].

---

[9]In the original derivation the free energy is unitless and written with the symbol $f$.

Figures/Method/MBAR-Potential.png

**Figure 2.5:** Diagram illustrating the reduced potential $u(\boldsymbol{x})$ used in free-energy calculations with MBAR. $K$ is the number of $\lambda$ states and $N$ is the number of frames. The rows represent $\Delta U$ at each $\lambda$ states and the columns represent $\Delta U$ at $\lambda_k$ with the coordinates at $\lambda_j$. (Image is designed in Tikz)

## 2.3.2  Potential of Mean Force

When performing path-dependent free energy calculations, we are interested in describing the potential energy surface of some process. On this end, path-dependent methods require the determination of the potential of mean force (PMF). In simple terms, this is the work done $W$ for moving a particle from point A to B along some reaction coordinate[10],

$$PMF(\xi) = W(\xi) = \int_A^B \langle F(\xi) \rangle_\xi \, d\xi, \tag{2.31}$$

---

[10]In some literature text this is referred to as collective variables.

where $\xi$ is the reaction coordinate and $\langle F(\xi) \rangle_\xi$ is the average or mean force at point $\xi$. Using protein-ligand as an example, the ligand can escape the binding site if we simulate to experimental time scales. The PMF can then be obtained by constructing a histogram distribution $\rho(\xi)$ along the reaction coordinate

$$PMF(\xi) = -k_\mathrm{B} T \ln[\rho(\xi)] + C, \tag{2.32}$$

where $C$ is a constant value[11]. However, experimental time scales are still not reachable with the current hardware. Therefore, to construct the PMF, we need to force the ligand out of the binding site by applying a "bias" that is corrected *a posteriori*.

---

[11]$C$ is an offset as a result of integration



**Figure 2.6:** A molecule is moved from point A to B by applying an external harmonic potential. The path is discretised in separate simulations, and the PMF can be constructed by solving the WHAM equations. (Image is designed in Tikz)

The most commonly used method for biasing a molecule over a reaction coordinate is the umbrella sampling (US) method [88]. In essence, the path from **A** to **B** is discretised into multiple windows (stratification) by applying an external potential. The external potential is usually a harmonic potential as it is easier to work with due to its mathematical properties. The external harmonic potential results in a Gaussian distribution of the ligand's centre of mass (COM) on the reaction coordinate. The PMF profile can be recovered from the biased distributions using the weighted histogram analysis method (WHAM)[12] [89] (see Fig. 2.6). The biased distributions are combined using what is referred to as the WHAM equations

$$\rho^{(unbiased)}(\xi) = \frac{\sum\limits_{i=1}^{N_w} n_i \rho_i^{(biased)}(\xi)}{\sum\limits_{j=1}^{N_w} n_j e^{-(V_j(\xi) - \Delta G_j)/k_\mathrm{B}T}}, \qquad (2.33a)$$

$$e^{-\Delta G_i/k_\mathrm{B}T} = \int e^{-V_i(\xi)/k_\mathrm{B}T} \, \rho^{(unbiased)}(\xi) \, d\xi. \qquad (2.33b)$$

Here $N_w$ is the number of windows, $n$ is the number of samples at each window, $V(\xi)$ is the harmonic biasing potential, and $\rho^{(biased)}$ and $\rho^{(unbiased)}$ are the biased and unbiased probability distributions respectively. Since $\Delta G$ is the unknown quantity present in both equations, the WHAM equations are solved iteratively until convergence is reached. A C++ implementation of the WHAM equations is distributed by the Grossfield Lab [90]. To ensure convergence of the PMF, the distributions of the reaction coordinate(s) between neighbouring windows must have sufficient overlap between them. In a study of ChTx toxin-binding to KcsA ion channel 5% overlap between neighbouring windows was found to be sufficient [91]. However, different systems might require different criteria for convergence. There are several other methods available that can give an estimate of the PMF like the adaptive-biasing force (ABF) [92] and metadynamics [93]. These methods are based on non-equilibrium statistical mechanics, while US-WHAM is based on equilibrium simulations. All of these methods essentially changes the underlying potential energy to speed up a slow process that otherwise cannot be observed by brute-force MD simulations alone.

---

[12]WHAM is just one method to recover the PMF

# Chapter 3

# Role of Methionine in Na2 Binding in Glt$_{\text{Ph}}$

ABSTRACT

Glutamate transport through the excitatory amino acid transporters is coupled to the co-transport of three Na$^+$ ions, the binding sites (Na1–Na3) of which are conserved from archaea to mammalians. Molecular dynamics (MD) simulations reproduce Na1 and Na3 binding sites observed in the crystal structures but fail in the case of Na2. A distinguishing feature of the Na2 site is that an S atom from a conserved methionine residue is in the coordination shell of Na$^+$. We perform MD simulations on the recent Glt$_{\text{Tk}}$ structure and show that the problem with the Na2 site arises from using an inadequate partial charge for S. When methionine is appropriately parametrised, both the position and the binding free energy of Na$^+$ at the Na2 site can be reproduced in good agreement with the experimental data. Other properties of methionine, such as its dipole moment and the solvation energy of its side chain analogue, also benefit from this reparametrisation. Thus, the Na2 site in glutamate transporters provides a good opportunity for a proper parametrisation of methionine in MD force fields.

## 3.1  Introduction

Glutamate is transported across membranes by excitatory amino acid transporters (EAATs), which clear excess extracellular glutamate at synapses [14]. A unifying feature of the glutamate/aspartate transport is that it is coupled to the co-transport of three Na$^+$ ions [98, 107] binding sites of which (Na1–Na3) appear to be conserved from archaea to mammalians [13]. While there are no crystal structures of EAATs for mammalians yet, several structures have been determined for archaeal homologues, e.g., Glt$_{Ph}$ from *Pyrococcus horikoshi* [24, 31, 102, 103, 144] and Glt$_{Tk}$ from *Thermococcus kodakarensis* [33, 136]. In the GltPh structures, Tl$^+$ ions were substituted for Na$^+$ to increase the visibility of the bound ions, which allowed identification of the Na1 and Na2 sites but not Na3. The latest high-resolution Glt$_{Tk}$ structure was obtained in the presence of Na$^+$ ions, and all three ion binding sites were identified [33]. There was a good agreement for the Na1 and Na2 sites between the Glt$_{Ph}$ and Glt$_{Tk}$ structures, confirming that Tl$^+$ was a good substitute for Na$^+$. A possible explanation for the non-observation of the Na3 site in the Glt$_{Ph}$ structures is that the path leading to this site is too narrow, which prevents Tl$^+$ substitution for Na$^+$.

Many molecular dynamics (MD) simulations of Glt$_{Ph}$ have been performed to study the various steps involved in the transport mechanism from the initial binding of ions and substrate in the outward-facing conformation to their release in the inward-facing conformation (see Ref. [94] for a recent review). There was general agreement between the simulation results and experimental observations with regard to the Asp and Na1 binding sites [82, 105, 106, 113, 120, 121, 143], but the description of the Na2 site remained problematic. In most MD simulations, the Na$^+$ ion left the Na2 site within a few nanoseconds [82, 113, 115, 121, 143], and its binding free energy was found to be positive [82, 113]. The Tl$^+$ substitution for Na$^+$ was proposed as a possible source of this discrepancy. The missing Na3 site in the Glt$_{Ph}$ structure was also searched for with MD simulations, and three different sites were proposed [32, 109, 111]. Observation of the Na3 site in the latest Glt$_{Tk}$ structure [33] has settled this issue, confirming that the site proposed in Ref. [32] is the correct one.

The latest Glt$_{Tk}$ structure has also provided important clues for the resolution of the Na2 problem in MD simulations. By firmly establishing that the S atom of methionine in the conserved NMDGT motif is in the coordination shell of Na$^+$ at the Na2 site, it has focused attention on the methionine parameters and, in particular, the S atom. The partial charge on S widely differs among the common force fields, varying from $-0.09e$ in CHARMM [9], to $-0.27e$ in AMBER [67], and $-0.33e$ in OPLS [68]. The lack of a simple system which could be used to determine the partial charge on S is the main reason for the uncertainty in the charge value. The fact that methionine can assume a variety of roles in proteins from non-polar to polar further complicates its parametrisation. In this regard, the Na2 site in glutamate transporters, with a well-determined coordination shell and binding free energy for the Na$^+$ ion, offers a good opportunity to optimise the partial charge on the S atom of methionine. We note that a conserved methionine is also found in the ion-binding site of the Nramp family of divalent metal transporters [145], and its role in the selectivity for transition-metal ions was investigated in MD simulations [146]. However, the divalent ions are poorly

described in the current non-polarisable force fields, which prevents using this system to determine the partial charges on methionine.

Here we perform MD simulations on the fully bound $\text{Glt}_{\text{Tk}}$ structure [33] using the CHARMM force field. The charge on the methionine S atom is systematically varied while the neutrality of its side chain is maintained. The coordination of $\text{Na}^+$ and its binding free energy at the Na2 site are determined from MD simulations and compared to the experimental data to find an optimal value for the charge on S. The effect of the modified parameters on the dipole moment of methionine and the solvation energy of its side chain analogue are also considered for further tests of their suitability.

## 3.2 Method

### 3.2.1 Model System and MD Simulations

The recently resolved crystal structure of $\text{Glt}_{\text{Tk}}$ with all three $\text{Na}^+$ and aspartate bound (PDB ID 5E9S) [33] is used in MD simulations. The trimer is embedded in a 1-palmitoyl-2-oleoylphosphatidylethanolamine (POPE) phospholipid bilayer and solvated in a box of TIP3P molecules. The system is neutralised and ionised using 0.15 M NaCl. The simulation system is prepared using the VMD software package [147] and contains a total of $\sim$100,000 atoms. MD simulations are performed using NAMD (version 2.11) [148] with the CHARMM36 force field [149]. We consider a total of eight different charges on the S atom of methionine, including the CHARMM default value $q = \{-0.09, -0.15, -0.20, -0.25, -0.30, -0.40, -0.50\}e$. In order to maintain the neutrality of the side chain of methionine, the charges on the neighbouring $\text{C}_\gamma$ and $\text{C}_\epsilon$ atoms are increased commensurately. The temperature of the system is kept constant at 300 K with a Langevin damping of 5 $\text{ps}^{-1}$, and the pressure is maintained at 1 atm using the Langevin piston method with a damping coefficient of 20 $\text{ps}^{-1}$ (NPT ensemble) [79]. The particle-mesh Ewald (PME) method is used with periodic boundary conditions [78]. Non-bonded interactions are truncated at 12 Å and replaced with a switching function starting from 10 Å. A time step of 2 fs is used throughout the MD simulations.

### 3.2.2 Free Energy Calculations

The standard binding free energy of an ion is expressed as [150]

$$\Delta G_{\text{b}} = \Delta G_{\text{int}} + \Delta G_{\text{tr}}. \tag{3.1}$$

The first term describes the free energy change in translocating the ion from bulk to the binding site, while the second term represents the loss in translational entropy during this process. The translational free energy difference is evaluated from the fluctuations of the ion positions in the binding site in unrestrained MD simulations. Assuming a Gaussian distribution for the fluctuations, the free energy difference is given by [151]

$$\Delta G_{\text{tr}} = -k_{\text{B}}T \ln \left[ \frac{(2\pi e)^{3/2} \sigma_x \sigma_y \sigma_z}{V_0} \right], \tag{3.2}$$

where $\sigma$'s are the principal root mean square fluctuations of Na$^+$, $V_0 = 1661$ Å$^3$ is the volume for the standard concentration of 1 M, and $e$ is Euler's number. The free energy of translocation is calculated using the free energy perturbation (FEP) method [152]. We used exponentially spaced lambda values to reduce the number of windows to 66 instead of 130, which is required for FEP calculations when a charge is created/annihilated [82]. Each window is equilibrated for 40 ps followed by 40 ps of production run. Following previous FEP calculations [82], we transform the Na$^+$ ion to a water molecule at the Na2 site, while the reverse transformation is performed in bulk simultaneously in the same system. We transform the Na$^+$ into water because, in unrestrained MD simulations without a Na$^+$ ion at the Na2 site, a water molecule fills this site within a short simulation time [82]. Performing the site/bulk FEP calculations simultaneously in the same system preserves the charge neutrality of the system and also avoids simulation artefacts arising from the use of different systems for site and bulk.

### 3.2.3 Methionine Side Chain Analogue

Because methionine is oxidised in solution, we consider its side chain analogue (methyl ethyl sulphide) to examine the effect of the modified parameters on the solvation energy. The side chain analogue was built by removing the acid part of methionine and terminating the C$_\beta$ atom with a third hydrogen atom. The partial charge of C$_\beta$ is adjusted to maintain neutrality. The system is solvated with TIP3P water in a box with dimensions $50 \times 50 \times 50$ Å$^3$ (~11,700 atoms). The solvation free energy is calculated using the FEP method, where the molecule in the system is annihilated by turning off the electrostatic and Lennard-Jones (LJ) interactions. To improve convergence, we split the two interactions and calculate them separately. Thus, the solvation free energy is the sum of the two free energy terms

$$\Delta G_{\text{solv}} = \Delta G_{\text{Elec}} + \Delta G_{\text{LJ}}. \tag{3.3}$$

The FEP calculations are performed with 31 lambda values, concentrating more points in the end-points. At each lambda value, we perform 20 ps of equilibration and 20 ps of production run. We apply an analytical correction to the long-range part of the LJ calculations to account for the long-range part that is lost due to the use of a switching function and cut-off in the LJ potential. A long-range correction is an option available in NAMD with the keyword `LJcorrection`, and the implementation follows that of Ref. [153].

## 3.3 Results and Discussion

### 3.3.1 Characterisation of the Na2 Site: Experiments vs MD Simulations

The residues that coordinate the Na$^+$ ion at the Na2 site of Glt$_{Tk}$ are shown in Fig. 3.1 (PDB ID: 5E9S) [33]. The same residues have also been found to coordinate the Tl$^+$ ion at the Na2 site of Glt$_{Ph}$ (PDB ID: 2NWX) [31]. To make the comparison more

**Figure 3.1:** Residues coordinating the Na$^+$ ion (yellow ball) at the Na2 site of Glt$_{\text{Tk}}$ crystal structure [33]. The Na2 site is in between transmembrane helix 7 (TM7, green) and hairpin 2 (HP2, violet). In MD simulations with the CHARMM force field, Na$^+$ moves away from the S atom of M314 to a new site (orange ball). When the charge on the S atom is boosted from $q = -0.09e$ to $q = -0.03e$, Na$^+$ remains at the Na2 site (blue ball overlapping with the yellow ball).

quantitative, we list the distances between the Tl$^+$/Na$^+$ ion and the coordinating atoms in Table 3.1. Allowing for the differences between the ionic radii of Tl$^+$ and Na$^+$ (radius of Tl$^+$ is 0.5 Å larger than that of Na$^+$) [154] and the resolutions of the crystal structures (2NWX, 3.5 Å vs 5E9S, 2.8 Å), there is good agreement between the two Na2 sites. In particular, the Tl$^+$−S and Na$^+$−S distances match very well with the contact distances obtained from the ionic radii (3.34 Å and 2.86 Å, respectively) [154], which leaves no doubt that the S atom is in the coordination shell of the respective ion in both crystal structures.

The inability of the MD simulations to hold the Na$^+$ ion at the Na2 site of Glt$_{\text{Ph}}$ has already been noted in the Introduction. A similar result has been obtained for the Na2 site of Glt$_{\text{Tk}}$ using the AMBER force field [33]. Performing MD simulation on Glt$_{\text{Tk}}$ using the CHARMM force field, we also reach the same conclusion: the Na$^+$ ion immediately moves away from the S atom of M314 to an alternative site at the protein-water interface, coordinated with about two water molecules (Fig. 3.1 and the fourth column in Table 3.1). With longer MD simulations, the Na$^+$ ion leaves this site as well. Further evidence for the instability of Na$^+$ at this site is provided by the binding free energy calculations, which yield a positive value (see below).

**Table 3.1:** Average Distances (in Å) of the Atoms Coordinating the Tl$^+$/Na$^+$ Ion at the Na2 Site.[a]

| | | | Na$^+$ | |
|---|---|---|---|---|
| helix-residue | Glt$_{Ph}$ Tl$^+$ | Glt$_{Tk}$ Na$^+$ | $q = -0.09e$ | $q = -0.03e$ |
| TM7–T311 (O) | 2.6 | 2.3 | $2.3 \pm 0.1$ | $2.3 \pm 0.1$ |
| TM7–M314 (S$_\delta$) | 3.4 | 2.9 | $6.2 \pm 0.4$ | $2.9 \pm 0.2$ |
| HP2–S352 (O) | 2.1 | 2.4 | $4.7 \pm 0.3$ | $2.3 \pm 0.1$ |
| HP2–I353 (O) | 3.2 | 2.8 | $2.4 \pm 0.3$ | $2.9 \pm 0.2$ |
| HP2–T355 (O) | 2.2 | 2.2 | $2.3 \pm 0.1$ | $2.3 \pm 0.1$ |

[a]The second column is for the Tl$^+$ ion in Glt$_{Ph}$ (2NWX) and the third column is for the Na$^+$ ion in Glt$_{Tk}$ (5E9S). The fourth column is obtained from 5 ns unrestrained MD simulations with the CHARMM charge of $-0.09e$ on the S atom of M314, which is boosted to $-0.30e$ in the last column. The residue numbers refer to Glt$_{Tk}$ (those of Glt$_{Ph}$ are three less). Side chain atoms are indicated with a subscript.

### 3.3.2  Optimising the Partial Charges on Methionine

The above results indicate that the charge on the S atom of methionine is inadequate for binding Na$^+$ at the Na2 site and needs to be boosted. To find a better value for the charge on the S atom, we perform several MD simulations with the charge varying from $-0.09e$ to $-0.50e$ (listed in Methods). For each system, we perform 10 ns unrestrained MD simulation and determine the average distances between Na$^+$ and the coordinating atoms. Variation of the coordination distances with the charge on S is shown in Fig. 3.2. The coordination distances are normalised with those from the Glt$_{Tk}$ crystal structure (third column in Table 3.1) so that 1 corresponds to perfect agreement with the experimental data. It is seen that for $q = -0.30e$, there is an excellent agreement between the coordination distances obtained from the MD simulations and the Glt$_{Tk}$ crystal structure (see columns 3 and 5 of Table 3.1 for a comparison of the actual distances and Fig. 3.1 for the position of Na$^+$). When the charge is increased from $-0.30e$ toward $-0.09e$, the Na$^+$ ion moves away from M314 (S$_\delta$) and the neighbouring S352 (O) while contact with the other three carbonyl oxygen atoms are kept. The inset in Fig. 3.2 shows that the actual Na$^+$–S$_\delta$ distance increases by more than 3 Å. Reducing the charge from $-0.30e$ toward $-0.50e$ causes the Na$^+$ ion to be more tightly bound to M314 (S$_\delta$). As the pair are already at the contact distance, the Na$^+$–S$_\delta$ distance is reduced only slightly. Nevertheless, this is sufficient for I353 (O) on the opposite side to flip away from Na$^+$ (Fig. 3.2), presumably because its interaction with Na$^+$ is weaker than the other carbonyl oxygen atoms (Table 3.1). In contrast, the strongly interacting T311 (O) maintains perfect contact with Na$^+$, regardless of the charge used thanks to the ability of the carbonyl group to track the movement of Na$^+$. S352 (O) and T355 (O) also retain contact with Na$^+$ for $q < -0.3e$. Thus, due to the weaker interaction of I353 (O) with Na$^+$, the corresponding coordination distance remains very sensitive to the charge on S, which suggests an optimal value of $q = -0.30e$ (Fig. 3.2).

We next perform a similar study for the binding free energy of Na$^+$ at the Na2 site, which was measured in Glt$_{Ph}$ as $-3.3$ kcal/mol [104]. Because the Glt$_{Ph}$ and Glt$_{Tk}$ structures share 77% sequence identity and the ion and substrate binding sites are conserved, we expect a similar value for the binding free energy in Glt$_{Tk}$. Again the

**Figure 3.2:** Variation of the $Na^+-O/S_\delta$ distances with the charge in the S atom of M314. The average distances are obtained from 5 ns MD simulations and normalised with those from the $Glt_{Tk}$ structure. The inset shows the actual $Na^+-S_\delta$ distances.

charge on the S atom of M314 is varied from $-0.09e$ to $-0.50e$, and the binding free energy of $Na^+$ is determined from the FEP calculations as described in Methods. The results are shown in Fig. 3.3 A. As a general trend, the binding free energy of $Na^+$ decreases with the charge on S, and the binding free energy curve crosses the dashed line representing the experimental value at $q = -0.30e$. This provides an independent confirmation for the proposed charge on S determined from the crystal structure of $Glt_{Tk}$.

As anticipated from the MD simulations with the CHARMM force field, the binding free energy of $Na^+$ is positive for $q = -0.09e$. Exceptional behaviour of the binding free energies is observed between $q = -0.30e$ and $-0.25e$ and also between $q = -0.15e$ and $-0.09e$, where the increase in the free energy is much smaller than the trend or even reduced in the latter case (Fig. 3.3 A). In both cases, an extra water molecule enters in the coordination shell of $Na^+$ at the larger $q$ value, which dampens the increasing trend in the binding free energy by compensating for the atoms departing the coordination shell and the missing charge on S.

Changing the partial charges on a molecule may affect its solution properties, which can be compensated by slightly adjusting the LJ parameter $R_{min}$ [50]. Unfortunately, such data are not available for methionine due to its oxidation in solution. Nevertheless, we have considered the effect of changing the $R_{min}$ value of the S atom on the $Na^+$ binding free energy at the Na2 site (Fig. 3.3 B). To compensate for a charge boost, $R_{min}$ is expected to be increased. From Fig. 3.3 B, it is seen that such an adjustment of $R_{min}$ will have a relatively small effect on the $Na^+$ binding free energy. We note that

a small increase in the Na$^+$ binding free energy due to a larger $R_{\min}$ value can be easily accommodated by slightly decreasing the charge on the S atom (Fig. 3.3 A).

### 3.3.3  Effect of the Modified Charges on Other Properties of Methionine

It is of interest to see how other properties of methionine are affected by the proposed changes. The dipole moment of methionine in the gas phase is measured as 1.6 D [155], while the current CHARMM parameters yield a value of 1.1 D. With the modified charges, the dipole moment of methionine is increased to 2.4 D. In a fixed-charge model, the dipole moment of a molecule needs to be boosted from its gas phase value in order to take the polarisation effects into account. Thus, the modified charges provide a more realistic description for the dipole moment of methionine in solution.

We next consider the solvation free energy for the methionine side chain analogue, which was measured as $-1.48$ kcal/mol [156]. Previous calculations using CHARMM yielded positive values for the solvation free energy, e.g., $+0.93$ [157] and $+1.08$ kcal/mol [158], which is attributed to the weak partial charge on the S atom. Boosting the partial charge on S to $-0.30e$, we obtain a solvation free energy of $-1.10$ kcal/mol, which has the correct sign and is much closer to the experimental value. The effect of changing the $R_{\min}$ value of the LJ potential on the solvation free energy while the charge is kept constant at $q = -0.30e$ is shown in Fig. 3.4. Again increasing the $R_{\min}$ value leads to a relatively small change in the solvation free energy, which can be accommodated by a slight adjustment of the charge on the S atom. Further fine-tuning of the methionine parameters is clearly possible, but this should preferably be done using a larger set of methionine interactions, including, e.g., aromatic interactions [159].

### 3.3.4  Discussion

Ion coordination by a methionine side chain is a very rare occasion, which raises the question, how did methionine at the Na2 site survive the evolutionary pressure and remain intact from archaea to mammalians? Mutation experiments performed on M367 in EAAT3 (the equivalent of M314 in Glt$_{Tk}$) provide a plausible answer to this question [160]. The glutamate affinity in M367L, M367C, and M367S mutants of EAAT3 was reduced by $10-20$-fold compared to the wild type. The effect of these mutations on the Na$^+$ affinity was found to be much less, i.e., only $2-3$-fold reductions compared to the wild type [160]. Thus, the steric role the methionine side chain plays in fitting the substrate into the binding pocket appears to be much more important than the role of the S atom in coordinating the Na$^+$ ion at the Na2 site and must be the main reason for the conservation of methionine in the NMDGT motif of glutamate transporters.

The boosting of the charge on the S atom and the corresponding increase of charges on the neighbouring C$_\gamma$ and $C_\epsilon$ atoms is expected to have a substantial effect on MD simulations of proteins where methionine residues play functional roles. For example, analysis of the protein structures in the Protein Data Bank indicates that methionine commonly interacts with aromatic residues, but the interaction energy estimated using the CHARMM force field is quite low ($1-3$ kcal/mol) [159]. A properly parametrised

**Figure 3.3:** (A) Binding free energy of Na$^+$ at the Na2 site as a function of the charge on the S atom of M314 using the default LJ parameter $R_{\mathrm{min}}$= 4.00 Å. (B) The effect of varying $R_{\mathrm{min}}$ of the S atom on the Na$^+$ binding free energy for $q = -0.30e$. The dashed lines indicate the experimental value of $-3.3$ kcal/mol measured in Glt$_{\mathrm{Ph}}$ [104].

methionine side chain could increase this interaction energy substantially, as demonstrated by the Na$^+$ binding free energy here. This would provide a better understanding of the role of the methionine-aromatic residue interactions in stabilising the protein structure. Methionine mutations are involved in some diseases, such as Alzheimer's and von Willebrand disease, so an improved description of methionine in MD simulations will also result in better characterisation of the effect of such mutations, which in turn could help in the rational design of therapeutics targeting these diseases.

**Figure 3.4:** Solvation energy of the methionine side chain analogue. The black and red lines show the results obtained using the default CHARMM parameters and modified ones, respectively. The dashed line indicates the experimental value.

## 3.4 Conclusion

In conclusion, using the latest Glt$_{Tk}$ structure as a guide, we have argued that the S atom of M314 is in the coordination shell of the Na$^+$ ion at the Na2 site, and the inability of MD force fields to hold Na$^+$ at this site is due to the inadequate charge on S. Using the coordination data from the Na2 site and the binding free energy of Na$^+$, we have shown that the charge on S needs to be boosted from $q = -0.09e$ to $-0.30e$ in the CHARMM force field. The proposed change also yields a better description of the dipole moment of methionine and the solvation free energy of its side chain. Although the AMBER and OPLS force fields have substantially larger charges on the S atom of methionine compared to CHARMM, MD simulations of Glt$_{Tk}$ and Glt$_{Ph}$ with these force fields also failed to bind the Na$^+$ ion at the Na2 site [33, 115]. Thus, it would be useful to optimise the methionine charges for these force fields as well using the Na2 binding site in Glt$_{Tk}$ as demonstrated here. A proper parametrisation of methionine will hopefully lead to a better understanding of its diverse functional roles in proteins.

# Chapter 4

# Mechanism of Ligand Binding in Glt$_\mathrm{Ph}$

ABSTRACT

Glutamate transporters clear up excess extracellular glutamate by co-transporting three Na$^+$ and one H$^+$ with the counter-transport of one K$^+$. The archaeal homologs are selective to aspartate and only co-transport three Na$^+$. The crystal structures of Glt$_\mathrm{Ph}$ from archaea have been used in computational studies to understand the transport mechanism. While some progress has been made with regard to the ligand-binding sites, a consistent picture of transport still eludes us. A major concern is the discrepancy between the computed binding free energies, which predict high-affinity Na$^+$– low-affinity aspartate binding, and the experimental results where the opposite is observed. Here we show that the binding of the first two Na$^+$ ions involve an intermediate state near the Na1 site, where two Na$^+$ ions coexist and couple to aspartate with similar strengths boosting its affinity. Binding free energies for Na$^+$ and aspartate obtained using this intermediate state are in good agreement with the experimental values. Thus the paradox in binding affinities arises from the assumption that the ligands bind to the sites observed in the crystal structure following the order dictated by their binding free energies with no intermediate states. In fact, the presence of an intermediate state eliminates such a correlation between the binding free energies and the binding order. The intermediate state also the facilitates transition of the first Na$^+$ ion to its final binding site via a knock-on mechanism, which induces substantial conformational changes in the protein consistent with experimental observations.

## 4.1  Introduction

Glutamate transporters (Glts) are membrane proteins responsible for removing excess extracellular glutamate in the synaptic cleft. The transport highly depends on the electrochemical gradient of Na$^+$ ions [14]. In mammals, Glts are called the excitatory amino-acid transporters (EAATs). In one cycle, they co-transport three Na$^+$ and a H$^+$ with glutamate and counter-transport one K$^+$ [13]. Failed regulation of Glts can lead to cell death due to overexcitation of receptors (excitotoxicity), which has been implicated in neurological diseases such as Alzheimer's and amyotrophic lateral sclerosis [14]. As in ion channels, the first crystal structure of Glts was resolved in prokaryotes, namely, Glt$_{Ph}$ from the archaeal homologue *Pyrococcus horikoshii* [24, 31]. Although Glt$_{Ph}$ has only 36% amino-acid sequence identity with EAATs overall, the sequence identity is much higher in the binding pocket, and they share some common properties. As in EAATs, Glt$_{Ph}$ requires coupling of three Na$^+$ ions but does not require the co-transport of H$^+$ and the counter-transport of K$^+$ [107]. In addition, Glt$_{Ph}$ is selective for Asp over Glu, which is not seen EAATs [13].

In the initial crystal structures of Glt$_{Ph}$, the binding sites for two Na$^+$ ions (dubbed Na1 and Na2) and Asp were resolved but not the Na3 site for the third Na$^+$ ion. It is likely that the substitution of Tl$^+$ ions for Na$^+$ to obtain stronger signals has prevented the larger Tl$^+$ ions from entering the Na3 binding site. Several computational studies have been performed to determine the third sodium binding site labelled as Na3 [32, 108–111]. In the recent crystal structure of Glt$_{Tk}$ [33], all the Na$^+$ binding sites were resolved with Na$^+$ instead of Tl$^+$, which confirmed the Na3 site predicted from molecular dynamics (MD) simulations and mutagenesis experiments [32]. In addition to the Na3 binding site, the various steps in ligand binding in the outward-facing conformation [82, 105, 106, 120], and release in the inward-facing conformation [113, 121, 122, 143] have also been investigated with MD simulations (see Ref. [94] for a recent review). In general, MD simulations provide a satisfactory description of the substrate and Na$^+$ binding sites. A remaining discrepancy about the coordination of Na$^+$ at the Na2 site [82, 115] has been recently clarified using the Glt$_{Tk}$ structure [33] as a guide, which indicated that the S atom of a methionine (M311) coordinated Na$^+$. The discrepancy arose from the undercharging of S in the CHARMM force field, and once it was properly charged, agreement with the experimental coordination shell and the binding free energy of Na$^+$ at the Na2 site was restored [161].

In contrast, the situation with regard to the binding free energies of the other ligands has remained baffling. The initial free energy calculations of ligand binding in Glt$_{Ph}$ gave the order of binding as Na3 ($-18.7$ kcal/mol), Na1 ($-7.1$ kcal/mol), and Asp ($-3.8$ kcal/mol), following the order of the binding free energies quoted in parenthesis [82]. Similar results were obtained for the release of these ligands in the inward-facing conformation [113]. Thus, two Na$^+$ ions were predicted to bind before Asp, consistent with the earlier observations [13, 110]. However, in subsequent experiments, almost the opposite results were obtained for the binding free energies with Na$^+$ binding with low affinity and Asp with high affinity [116–118]. Given that the free energy calculations performed using crystal structures usually retain chemical accuracy, the size of the discrepancy is rather unsettling and has compelled us to look beyond the

static picture of binding provided by the crystal structure. For guidance, it is best to focus on Asp as there is a unique and well-defined binding site. We know that Asp does not bind in the absence of $Na^+$ [13]. Free energy calculations indicate that Asp can bind with one $Na^+$ when it is at Na1 but not when it is at Na3 [82, 113]. That is, $Na^+$ is strongly coupled to Asp at the Na1 site but not at the Na3 site. This suggests that the Asp affinity can be significantly boosted if there is an intermediate site for the first $Na^+$ in the vicinity of the Na1 site (dubbed Na1′), and two $Na^+$ ions can coexist at the Na1 and Na1′ sites. The Coulomb repulsion between the two $Na^+$ ions in close proximity would then naturally explain the low affinity and slow binding of the second $Na^+$. A strong candidate for such a Na1′ site is the one found by Huang and Tajkhorshid while searching for the Na3 site, where $Na^+$ is coordinated simultaneously by the D312 and D405 side chains [111].

Here we investigate the feasibility of such a scenario by performing MD simulations and free energy calculations, where the Na1′ site is included as an intermediate state during the ligand-binding process. We show that Na1′ provides the deepest binding site so it will trap the first $Na^+$ and the second $Na^+$ can still bind to the Na1 site while Na1′ is occupied. We then probe Asp binding to $Glt_{Ph}$ for different ion configurations and draw inferences to resolve the discrepancy between the computed and experimental affinities.

## 4.2 Method

### 4.2.1 Model System and Simulation Details

The crystal structure of the outward-facing conformation of $Glt_{Ph}$ with an Asp and two $Na^+$ is used in MD simulations (PDB ID: 2NWX) [31]. The third $Na^+$ is placed at the Na3 site as determined in Ref. [32] and the crystal structure of $Glt_{Tk}$ [33]. The simulation system is prepared using the VMD software package [147]. First, the trimer structure is embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylethanolamine (POPE) phospholipid bilayer. The system is then solvated in a box of TIP3P water molecules [162, 163] with 0.15 M NaCl and neutralised with extra $Cl^-$ ions. The final system contains a total of $\sim$100,000 atoms in a simulation box of size $110 \times 110 \times 100$ $\mathring{A}^3$. All MD simulations are performed using the NAMD program (version 2.12) [148] together with the CHARMM36 force field [164], where the NBFIX correction for $Na^+$ interactions is the default option. The default (pH 7) values are used for the charge states of the amino acids. The temperature is kept constant at 300 K using Langevin dynamics with a damping coefficient of 1 $ps^{-1}$. The pressure of the system is maintained at 1 atm using the Langevin piston method [165] with a coefficient of 20 $ps^{-1}$. Short-range Lennard-Jones (LJ) interactions are truncated at 12 Å and a smooth switching function is used starting at 10 Å. Periodic boundary conditions are employed, and the long-range interactions are calculated with the particle-mesh Ewald method [78]. A time step of 2 fs is used in all MD simulations.

## 4.2.2   Free Energy Calculations

The standard binding free energy of an ion to a protein is expressed as

$$\Delta G_{\mathrm{b}}^0 = \Delta\Delta G_{\mathrm{int}} + \Delta\Delta G_{\mathrm{tr}}^0 + \Delta\Delta G_{\mathrm{c,P}}, \tag{4.1}$$

where the first term is the free energy change in translocating the ion from bulk water to the binding site, and the second term represents the loss of translational entropy during this process. The last term is used when the protein residues coordinating the ligand are restrained to improve convergence in the calculation of $\Delta G_{\mathrm{int}}$. If no restraints are applied, $\Delta\Delta G_{\mathrm{c,P}} = 0$. Assuming a Gaussian distribution for the ion positions in the binding site, the translational free energy difference is estimated using [151]

$$\Delta\Delta G_{\mathrm{tr}}^0 = -k_{\mathrm{B}}T \ln\left[\frac{(2\pi e)^{3/2}\sigma_x\sigma_y\sigma_z}{V_0}\right], \tag{4.2}$$

where $V_0 = 1661\,\mathrm{\mathring{A}}^3$ is the reference volume for the standard concentration of 1 M, $e$ is Euler's number, and the $\sigma$'s are the principal root mean square fluctuations of the ion positions. The restraint-free energy is calculated using the thermodynamic integration (TI) approach of Cecchini et al. [166]

$$\Delta G_{\mathrm{c}} = \frac{1}{2}\int_0^k \langle|\boldsymbol{X} - \boldsymbol{X}_0|^2\rangle_k \, dk. \tag{4.3}$$

Here, $|\boldsymbol{X} - \boldsymbol{X}_0|^2$ is the mean squared displacement of the restrained residues from a reference structure at a given spring constant $k$. We find that a value of 1.0 kcal/mol/$\mathrm{\mathring{A}}^2$ per atom is sufficient to reduce the fluctuations of the coordinating residues. The TI calculations are performed using 11 windows, and each window is simulated 200 ps for equilibration and 800 ps for production. To obtain $\Delta\Delta G_{\mathrm{c,P}}$, two separate calculations are performed, one with the ion at the binding site and another with the ion in bulk, i.e., $\Delta\Delta G_{\mathrm{c,P}} = \Delta G_{\mathrm{c,P}}^{\mathrm{ion-bulk}} - \Delta G_{\mathrm{c,P}}^{\mathrm{ion-site}}$.

The free energy of translocation is calculated using both free energy perturbation (FEP) and TI [152]. The FEP calculation is performed first, and the trajectory is used as the starting point for the TI calculations. As in the restraint-free energy, the free energy change between the states with the ion at the site and in bulk are calculated, i.e., $\Delta\Delta G_{\mathrm{int}} = \Delta G_{\mathrm{int}}^{\mathrm{ion-bulk}} - \Delta G_{\mathrm{int}}^{\mathrm{ion-site}}$. However, only one calculation is needed if we create and destroy the ion simultaneously in the same system. For the FEP calculations, we use 66 exponentially spaced $\lambda$ values as this was shown in previous studies to be adequate for charged molecules [82]. At each $\lambda$ value, we perform 50 ps of equilibration, followed by a further 50 ps of production run. In the TI method, the ensemble of the free energy derivative is calculated at each $\lambda$ value. The free energy is recovered by integrating the free energy derivative with respect to the $\lambda$ values

$$\Delta G_{\mathrm{int}} = \int_0^1 \left\langle\frac{\partial U(\lambda)}{\partial\lambda}\right\rangle_\lambda d\lambda. \tag{4.4}$$

For charged molecules, the integral can be done using Gaussian quadrature. A seven-point quadrature is used as this was shown to be adequate in a previous study [43]. The

starting point for each TI window is obtained from the corresponding FEP calculation, and each window is run for a total of 1 ns with the first 200 ps discarded as the equilibration phase. The convergence of the TI calculations for the ions is shown in Fig. 4.A.1.

For the substrate, the standard binding free energy is expressed as

$$\Delta G_{\text{b}}^0 = \Delta\Delta G_{\text{int}} + \Delta\Delta G_{\text{tr}}^0 + \Delta\Delta G_{\text{rot}} + \Delta\Delta G_{\text{c,L}}, \tag{4.5}$$

which includes the free energy loss due to rotational entropy $\Delta\Delta G_{\text{rot}}$, and the conformational restraints are applied on the substrate instead of the protein as indicated in the last term, $\Delta\Delta G_{\text{c,L}}$. The rotational free energy difference is calculated in a similar manner to Eq. (4.2) [151]

$$\Delta\Delta G_{\text{rot}} = -k_{\text{B}}T \ln\left[\frac{(2\pi e)^{3/2}\sigma_{\phi_1}\sigma_{\phi_2}\sigma_{\phi_3}}{8\pi^2}\right]. \tag{4.6}$$

Here $\sigma_\phi$'s are the rotational root mean square fluctuations of the substrate at the binding site calculated using the quaternion representation. For the conformational restraint, a spring constant of 5.0 kcal/mol/Å$^2$ is applied to the heavy atoms of the substrate. The bulk calculation is performed separately with the substrate placed in a 25 Å cubic water box with a neutralising ion. As in previous calculations [82, 113], the translocation free energy is split into three stages to improve convergence

$$\Delta\Delta G_{\text{int}} = \Delta\Delta G_{\text{elec}} + \Delta\Delta G_{\text{LJ–bb}} + \Delta\Delta G_{\text{LJ–sc}}. \tag{4.7}$$

The first term on the right-hand side is the electrostatic contribution, and the last two terms are the LJ contributions. The LJ term is split into the backbone (LJ–bb) and the side chain (LJ–sc) parts in order to prevent a water molecule getting trapped behind the substrate in the binding site as noted previously [82]. For the electrostatic free energies, we report the values obtained with the TI method with 0.5/1.0 ns for equilibration/production. For the LJ free energies, we report the FEP values instead due to the lack of convergence with quadrature windows. A total of 52 and 37 $\lambda$ values are used for the LJ–bb and LJ–sc calculations, respectively. For each window, we run a 100/100 ps for equilibration/production (see Fig. 4.A.2 for the convergence of the forward and backward transformation for both electrostatic and LJ free energies). A soft-core potential is used in the LJ calculations with a shifting coefficient of 6.0 [167].

Umbrella sampling simulations are performed for the Na1$'$ → Na3 transition of Na$^+$ to find the potential of mean force (PMF) and study the conformational changes caused by this transition. The reaction coordinate is chosen along the vector from the Na1$'$ site to the Na3 site, and umbrella windows are generated at 0.5 Å intervals using $k = 10$ kcal/mol/Å$^2$ for the harmonic biasing potential. Overlaps between the samples of neighbouring windows are found to be $> 10\%$, which assures a robust construction of the PMF using the weighted histogram analysis method. Each window is simulated for 8 ns. Convergence of the PMF up to the barrier is obtained after 3 ns, so the final PMF is constructed from the last 5 ns.

## 4.3 Results and Discussion

### 4.3.1 Characterisation of the Na$^+$ Binding Sites

Here we use the 2NWX structure for convenience, but it could be argued that the recent apo Glt$_{Ph}$ crystal structure 4OYE [144] may provide a more appropriate starting point for the binding free energy calculations. Thus, it is prudent to show that there is no difference between the ligand-free 2NWX and 4OYE systems once they are equilibrated. To this end, we have performed 60 ns MD simulations of the apo-2NWX and 4OYE systems under the same conditions. Superposition of the final snapshots of the two systems shows that there is a good overlap between the two structures (Figs. 4.A.3 A). For a more quantitative analysis, we compare the time series of the RMSDs of the two systems relative to 4OYE (Figs. 4.A.3 B). The apo-2NWX system is seen to relax to that of 4OYE within 25 ns. After the relaxation, the average RMSDs of the apo-2NWX and 4OYE systems relative to 4OYE are 1.09 Å and 1.06 Å, respectively. Thus either crystal structure can be used in binding free-energy calculations as long as the simulation system is equilibrated appropriately.

Starting with the fully bound equilibrated structure from a previous study [161], we remove Asp and Na$^+$ at Na2. Following a 50 ns MD simulation, Na$^+$ ions at Na1 and Na3 are stable with a water molecule coordinating the ion at Na1. The binding modes listed in Table 4.1 (columns 3 and 4) are in good agreement with those observed in the crystal structures [31]. We note that the Na$^+$ coordination at the Na3 site is slightly different when no ion is present at Na1 (Table 4.1, column 2). The ion moves away from the S93 residue but is more tightly coordinated by both O atoms of D312. To generate the intermediate site Na1$'$, we remove the Na$^+$ ion from Na3 and perform MD simulations of the system for another 50 ns. As seen in earlier MD simulations of Glt$_{Ph}$ without a Na$^+$ at Na3 [32, 111], the side chains of N310 and D312 swing towards the Na1 site and start coordinating the Na$^+$ ion there at a somewhat different site called Na1$'$ (Fig. 4.1 A). To achieve the coordination by N310 and D312, the Na$^+$ ion moves away from G306 and N310 by about 2 Å. Comparing the binding mode of the Na1$'$ site with that of Na1 (Na3) (Table 4.1, columns 5 and 4) , they appear qualitatively similar. However, having two Asp side chains instead of one, confers a distinct advantage to the Na1$'$ site as will be demonstrated with the binding free energies in the next section.

The prevailing assumption at this stage is that the first Na$^+$ ion moves from Na1$'$ to Na3 before a second Na$^+$ ion can bind to Na1 [32, 94, 111]. But, as discussed in the Introduction, the binding free energies of Na$^+$ and Asp calculated with this assumption are in conflict with the experimental results. Thus we are compelled to examine this assumption more closely and look for alternative scenarios of binding. MD simulations lasting 40 ns indicate that the Na$^+$ ion at the Na1$'$ site is very stable and exhibits no tendency to move towards Na3. We will show in the next section that such a transition is, in fact, energetically not feasible. Thus the presence of a second Na$^+$ in the vicinity of Na1$'$ is necessary to enable the transition of the first Na$^+$ from Na1$'$ to Na3. To search for such a site, we place the second Na$^+$ at the tip of the TM7a helix near the beginning of the NMDGT motif, where the carbonyl O atoms of G306, A307, T308, and N310 provide an oxygen-rich environment (Fig. 4.1 B). The ion is placed at the centre of these carbonyl O atoms and harmoni-

**Table 4.1:** Binding modes of the $Na^+$ ions in $Glt_{Ph}$ for various states of occupation.[a]

| Helix-Residue | Na3 | Na3 (Na1) | Na1 (Na3) | Na1' | Na1' (Na1) | Na1 (Na1') |
|---|---|---|---|---|---|---|
| TM3-Y89 (O) | $2.6 \pm 0.2$ | $2.4 \pm 0.2$ | - | - | - | - |
| TM3-T92 (OH) | $2.4 \pm 0.1$ | $2.4 \pm 0.1$ | - | - | - | - |
| TM3-S93 (OH) | $3.4 \pm 0.3$ | $2.6 \pm 0.3$ | - | - | - | - |
| TM7-G306 (O) | - | - | $2.5 \pm 0.2$ | - | - | $2.3 \pm 0.1$ |
| TM7-N310 (O) | - | - | $2.4 \pm 0.1$ | - | - | $3.7 \pm 0.2$ |
| TM7-N310 ($O_\delta$) | $2.3 \pm 0.1$ | $2.2 \pm 0.1$ | - | $2.4 \pm 0.1$ | $2.3 \pm 0.1$ | - |
| TM7-D312 ($O_1$) | $2.2 \pm 0.1$ | $2.4 \pm 0.5$ | - | $2.4 \pm 0.2$ | $2.4 \pm 0.1$ | - |
| TM7-D312 ($O_2$) | $2.3 \pm 0.1$ | $2.9 \pm 0.6$ | - | $2.7 \pm 0.3$ | $2.7 \pm 0.3$ | - |
| TM8-N401 (O) | - | - | $2.6 \pm 0.3$ | $2.6 \pm 0.2$ | $2.7 \pm 0.2$ | $3.5 \pm 0.3$ |
| TM8-D405 ($O_1$) | - | - | $2.3 \pm 0.1$ | $2.4 \pm 0.1$ | $2.4 \pm 0.1$ | $2.5 \pm 0.2$ |
| TM8-D405 ($O_2$) | - | - | $2.2 \pm 0.1$ | $3.8 \pm 0.3$ | $4.4 \pm 0.1$ | $2.3 \pm 0.1$ |
| H2O (1) | - | - | $2.3 \pm 0.1$ | $2.5 \pm 0.2$ | - | $2.4 \pm 0.1$ |
| H2O (2) | - | - | - | - | - | $2.5 \pm 0.4$ |
| H2O (3) | - | - | - | - | - | $2.4 \pm 0.1$ |
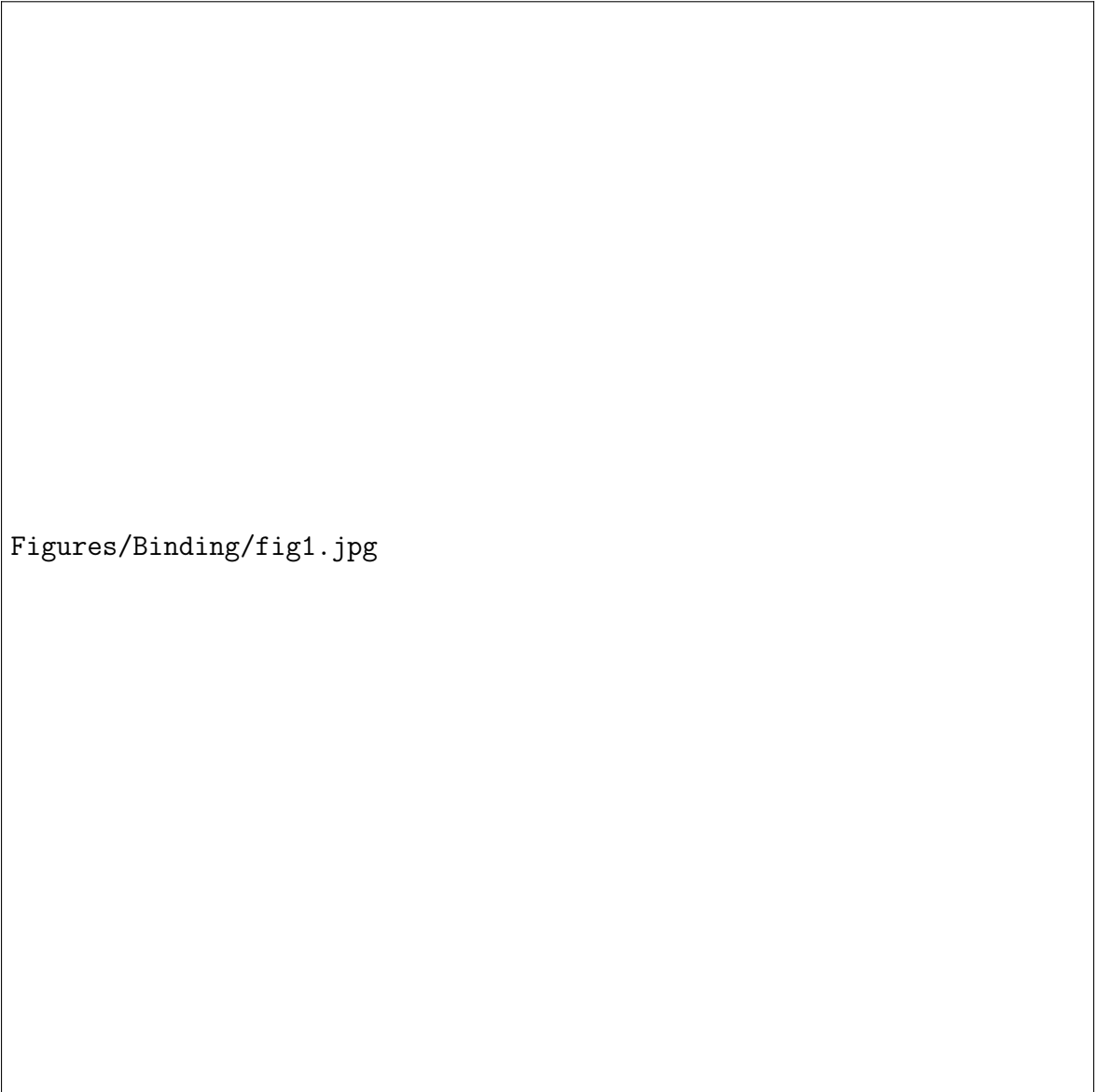
[a]Average distances (in Å) of protein atoms coordinating $Na^+$ ions. Other ions present are indicated in parenthesis at the top row. The average distances and errors are estimated from a 5 ns unrestrained MD simulation.

cally restrained with a spring constant of 5 kcal/mol/Å². We release the ion in 9 steps with $k = \{5.0, 4.0, 3.0, 2.0, 1.0, 0.5, 0.2, 0.1, 0.0\}$ kcal/mol/Å², performing 1 ns MD simulation at each step. The two $Na^+$ ions are initially separated by $\sim$6.5 Å and remain in this position up to $k = 2.0$ kcal/mol/Å². At $k = 1.0$ kcal/mol/Å², the second $Na^+$ moves closer to Na1' with a water molecule separating the two ions and is weakly coordinated by G306 (O) and N310 (O) (Figs. 4.1 E–D). When the restraint is reduced to 0.1 kcal/mol/Å², the strong negative charge from D405 attracts $Na^+$ and knocks out the water molecule between the two ions (Fig. 4.1 C). The distance between the two $Na^+$ ions in this state is $\sim$3.5 Å (Fig. 4.1 F). We subject the unrestrained system to a further 45 ns simulation to check the stability of the two $Na^+$ ions in close proximity. As shown in Figs. 4.1 D–F, both $Na^+$ ions remain well-coordinated and do not exhibit any tendency to break away from the coexistent state.

Firmness of the Na1' site becomes more evident after the binding of the second $Na^+$, which has hardly made any dent on the Na1' coordination despite being so close (Table 4.1, columns 5 and 6). In contrast, the presence of $Na^+$ at Na1' has a detrimental effect on Na1 coordination as seen from the comparison of the binding modes in columns 4 and 7 of Table 4.1. The stress on $Na^+$ at Na1 will be relieved after the transition of the first $Na^+$ ion from Na1' to Na3, which will be discussed further after the binding free energy calculations.

## 4.3.2 Binding Free Energies

Here we calculate the binding free energies of ligands in various states of occupation to turn the qualitative insights gathered in the last section to quantitative predictions. The results for the binding free energies of the $Na^+$ ions are summarised in Table 4.2. As shown in Table 4.1, there is no stable Na1 site in the apo $Glt_{Ph}$, therefore it is not

**Figure 4.1:** (*Left column*) Coordination of the first two Na$^+$ ions in Glt$_{Ph}$: (*A*) The first ion binds to Na1′ coordinated by both D312 and D405, (*B*) the initial position of the second Na$^+$ held with a harmonic restraint, and (*C*) the final position of the second Na$^+$ in the two ion-state configuration after it is released (Na1). (*Right column*) Time evolution of distances (in Å) between atoms with the first 9 ns representing the release of restraints: (*D*) Na1′ and protein atoms, (*E*) Na1 and protein atoms and (*F*) Na1 and Na1′. For clarity, the running averages of the distances are plotted.

included in the binding free energy calculations for the first Na$^+$. The first Na$^+$ can bind to either Na1′ or Na3. The binding free energies in Table 4.2 clearly favour the Na1′ site by a margin of −3.3 kcal/mol. As mentioned before, simultaneous coordination by the D312 and D405 side chains is responsible for the deeper binding and extreme stability of the first Na$^+$ at the Na1′ site.

We next consider the binding of the second Na$^+$ ion. The crucial question here is whether the second Na$^+$ can bind to the Na1 site while Na1′ is occupied so that the two

**Figure 4.2:** (*Top row*) Asp bound to $Glt_{Ph}$ with the $Na^+$ ions in the (*A*) Na1–Na3 state, and (*B*) Na1–Na1′ state. The dashed lines represent the hydrogen-bond network between Asp and the Na1 site bridged with a water molecule. (*Bottom row*) The distance between the $Na^+$ ions and the $C\alpha$ atom of Asp when it is bound to $Glt_{Ph}$ in the (*C*) Na1–Na3 state, and (*D*) Na1–Na1′ state.

$Na^+$ ions can coexist in close proximity. In order to calculate this binding free energy, we have included $Na^+$ at Na1′ as part of the surrounding atoms restrained. Otherwise the FEP calculation becomes numerically unstable as the charge at Na1 is alchemically turned off. The effect of this restraint is included as part of $\Delta\Delta G_{c,P}$ as described in the Methods section. The binding free energy of the second $Na^+$ at Na1 is found to be $-1.8$ kcal/mol. So the answer to the coexistence question is affirmative. This value is also in good agreement with the experimental values from Ewers et al. [117], Reyes at al. [118], and Hänelt et al. [116] ($-2.3$, $-1.4$, and $-1.3$ kcal/mol, respectively). We stress that the slow binding of the second $Na^+$ with low affinity is made possible by

**Figure 4.3:** Time evolution of the centre of mass of Asp in Glt$_{Ph}$ with the two ion state Na1–Na1′. Black, red, and green represent Asp in chains A, B, and C, respectively. The running averages of the distances are plotted.

the Coulomb repulsion of the first Na$^+$ at Na1′. If we use the alternative scenario as in previous calculations and assume that the first Na$^+$ is at Na3, the binding free energy of the second Na$^+$ to Na1 becomes −6.9 kcal/mol, which is in conflict with the experimental values. Thus we can infer from the free energy simulations that the Na$^+$ ions must be in the intermediate Na1–Na1′ state during the measurements instead of the final Na1–Na3 configuration.

Having resolved the discrepancy in Na$^+$ binding affinity, we turn to the discrepancy in Asp affinity for Glt$_{Ph}$. To present an unbiased view, we consider all possible ion configurations in the binding free energy calculations for Asp, which are listed in Table 4.3. Systems consisting of only Asp and one Na$^+$ at Na1 or Na3 are created by removing

**Table 4.2:** Binding free energies of Na$^+$ ions to Glt$_{Ph}$ in different states of occupation.[a]

| Ion State | $\Delta\Delta G_{\text{int}}$ | $\Delta\Delta G_{\text{tr}}^0$ | $\Delta\Delta G_{\text{c,P}}$ | $\Delta G_{\text{b}}^0$ |
|---|---|---|---|---|
| Na1′ | −26.7 ± 1.3 | +4.6 | +2.0 | −20.1 ± 1.3 |
| Na3 | −21.4 ± 1.4 | +4.6 | - | −16.8 ± 1.4 |
| Na1 (Na1′) | −10.8 ± 1.5 | +4.8 | +4.2 | −1.8 ± 1.5 |
| Na1 (Na3) | −11.8 ± 1.5 | +4.9 | - | −6.9 ± 1.5 |
| Na1′ (Asp, Na1) | −24.2 ± 1.0 | +4.0 | +7.0 | −13.2 ± 1.0 |
| Na3 (Asp, Na1) | −24.0 ± 1.1 | +4.0 | +2.5 | −17.6 ± 1.1 |

[a]Labels inside the parenthesis indicate other ligands present. $\Delta\Delta G_{\text{int}}$ is the average of the forward and backward transformation. Convergence of the transformation are shown in Fig. 4.A.1. No restraints are needed for the Na3 and Na1 (Na3) calculations. All energies are reported in kcal/mol.

the ion at Na1 or Na3 from, respectively, from the Na1–Na3 configuration. The state with one Na$^+$ at Na1′ is obtained similarly from the Na1–Na1′ configuration. Finally, both Na$^+$ ions are removed for Asp binding to the apo Glt$_{Ph}$. Inspection of Table 4.3 shows that positive binding free energies are obtained for the apo Glt$_{Ph}$ and also with one Na$^+$ at Na3 or Na1′. A negative free energy is obtained for a Na$^+$ at Na1, but this site does not form with only one Na$^+$. Thus two Na$^+$ ions are required for the binding of Asp consistent with experiments [13, 110, 116–118].

The possible ion configurations with two Na$^+$ are Na1–Na3 and Na1–Na1′. The Na1–Na3 configuration comes from the crystal structure and results in a low-affinity binding of Asp (−3.6 kcal/mol) as found in previous calculations [82, 113]. This is in conflict with experiments where Asp is determined to bind with high affinity, e.g., −7.3 kcal/-mol [117] and −6.8 kcal/mol [116] (both were measured at 6 °C). The binding free energy of Asp is significantly boosted when the calculation is done using the Na1–Na1′ configuration. The calculated value (−7.4 kcal/mol) is in very good agreement with the experimental values quoted above. The difference between the binding free energies of the two configurations arises from the distance of the first Na$^+$ to Asp at the Na3 and Na1′ sites, which is depicted in Figs. 4.2 C and D, respectively. The distance between the Na$^+$ at Na1 and the Cα atom of Asp is about 7 Å for both systems. The Cα distance with the first Na$^+$ ion, however, is 9.5 Å at Na1′ and increases to 12.5 Å at Na3. This 3 Å increase in the distance significantly reduces the electrostatic attraction, leading to a smaller binding free energy of Asp for the Na1–Na3 configuration.

As in the case of the binding of the second Na$^+$, the free energy simulations imply that Asp must bind while the first Na$^+$ is still at the Na1′ site. That is the only way one can explain the high-affinity binding of Asp. We expect that the binding of the second Na$^+$ to Na1 will sufficiently destabilise the first Na$^+$ at Na1′ so that it will not be the deepest binding site anymore. To confirm this, we have repeated the free energy calculations for the binding of Na$^+$ to the Na1′ and Na3 sites in the presence of the second Na$^+$ at Na1 and bound Asp. As shown in the last two rows of Table 4.2, the Na3 site is now 4.4 kcal/mol deeper than the Na1′ site, so the first Na$^+$ will eventually move about ∼4 Å from Na1′ to Na3. But this transition is clearly delayed by an energetic barrier caused by the detachment of the D405 side chain, and the flipping of the D312 and N310 side chains which coordinate Na$^+$ at both sites (cf. columns 2 and 6 of Table 4.1).

**Table 4.3:** Binding free energies of Asp bound to Glt$_{Ph}$ with different ion configurations.[a]

| Substrate State | $\Delta\Delta G_{\mathrm{elec}}$ | $\Delta\Delta G_{\mathrm{LJ-bb}}$ | $\Delta\Delta G_{\mathrm{LJ-sc}}$ | $\Delta\Delta G_{\mathrm{tr}}^0$ | $\Delta\Delta G_{\mathrm{rot}}$ | $\Delta\Delta G_{\mathrm{c,L}}$ | $\Delta G_{\mathrm{b}}^0$ |
|---|---|---|---|---|---|---|---|
| Asp | −4.7 ± 0.8 | +0.5 ± 0.3 | +0.8 ± 0.2 | +2.8 | +3.4 | +0.2 | +3.0 ± 0.9 |
| Asp (Na3) | −8.6 ± 1.5 | +1.8 ± 0.3 | +1.0 ± 0.3 | +3.9 | +4.1 | −0.1 | +2.1 ± 1.6 |
| Asp (Na1′) | −10.7 ± 1.1 | +1.5 ± 0.3 | +0.9 ± 0.3 | +3.8 | +3.6 | +1.7 | +0.8 ± 1.2 |
| Asp (Na1) | −16.2 ± 1.2 | +3.1 ± 0.3 | +1.3 ± 0.3 | +4.1 | +4.1 | +1.2 | −2.4 ± 1.3 |
| Asp (Na1, Na3) | −16.6 ± 1.3 | +4.4 ± 0.4 | +1.1 ± 0.3 | +4.2 | +4.1 | −0.8 | −3.6 ± 1.4 |
| Asp (Na1, Na1′) | −22.5 ± 0.8 | +4.4 ± 0.3 | +0.6 ± 0.3 | +4.0 | +4.2 | +1.9 | −7.4 ± 0.9 |

[a]Labels inside the parenthesis indicate the ion states. The electrostatic and LJ energies are averages of the forward and backward transformation. Convergence of the transformation for the electrostatic and LJ interactions are shown in Fig. 4.A.2. All energies are reported in kcal/mol.

In order to get more insights for this barrier, we have performed potential of mean force (PMF) calculations for the Na1′ → Na3 transition of the first Na$^+$. The D312 side chain is found to chaperone Na$^+$ throughout the transition while the N310 side chain flips at the top of the barrier. The flipping of the N310 side chain is an unfavourable motion and causes some stress on the NMDGT motif, which has to move to allow the transition. Unfortunately, we could not obtain a converged PMF for the whole path due to the sampling problems caused by the large conformational changes in the protein. Because the existence of this barrier is essential for the stability of the proposed (Na1, Na1′) state, we show the initial part of the PMF, where convergence is not an issue yet (Figs. 4.4 A). As the Na$^+$ ion is pushed from the Na1 site towards Na3, the PMF steeply rises, reaching a height of over 10 kcal/mol, which is sufficient to trap the first Na$^+$ at Na1′.

The residue-specific RMSDs obtained from the initial and final umbrella windows show that the movement of the Na$^+$ ion induces substantial conformational changes in the NMDGT motif and the neighbouring residues when the ion is at the top of the barrier (Figs. 4.4 B, red bars). In contrast, the RMSDs of the same residues obtained from the equilibrium simulations of the (Na1, Na3) state indicates that they are completely relaxed after the Na$^+$ ion moves to the Na3 site (Figs. 4.4 B, black bars). The PMF and RMSD results in Fig. 4.4 show that the protein is in a slightly excited in the (Na1, Na1′) state and is driven into a highly excited state during the Na1′ → Na3 transition of Na$^+$. After Na$^+$ binds to Na3, the protein relaxes to a minimum free energy state and the conformational energy stored in the protein is completely dissipated. The above depiction of events during the Na1′ → Na3 transition is consistent with the experimental observations that the Na$^+$ binding induces structural changes in the protein [116, 118], and provides a physical mechanism for how these changes are induced. Here we stress the role of the (Na1, Na1′) intermediate state in facilitating the Na1′ → Na3 transition of the first Na$^+$. Without the Coulomb repulsion of the second Na$^+$ ion at Na1, the first Na$^+$ would have a much harder time pushing through the conformational energy barrier. Such a knock-on effect is well-known in ion channels. Nature seems to have exploited it in glutamate transporters as well.

From the binding free energy results, we propose the following mechanism for the binding of the ligands to Glt$_{Ph}$ (illustrated in Fig. 4.5). Starting with the outward-facing apo state of Glt$_{Ph}$, the first Na$^+$ ion binds to the Na1′ site with a very high affinity, where it is coordinated by both D405 and D312. This is followed by the entry of the second Na$^+$ ion (Asp cannot bind yet), which binds to the Na1 site, where it is coordinated by D405. The distance between the two ions is about 3.5 Å, which results in low-affinity binding of the second Na$^+$. In this two ion Na1–Na1′ state, the Asp substrate binds to Glt$_{Ph}$ with high affinity. Finally, the first Na$^+$ moves from Na1′ to Na3 through conformational changes involving the N310 and D312 side chains, and the NMDGT motif. Although we couldn't resolve this transition, both experiments [116, 118] and preliminary PMF calculations indicate that it will be associated with a large energy barrier, and hence it will be a slow process. In the final state, the HP2 gate is closed and locked by the binding of the third Na$^+$ to the Na2 site. MD simulations and binding free energies [104, 120, 161] suggest that both of these processes are relatively fast, so it is likely that the binding of the first Na$^+$ to its final destination at the Na3
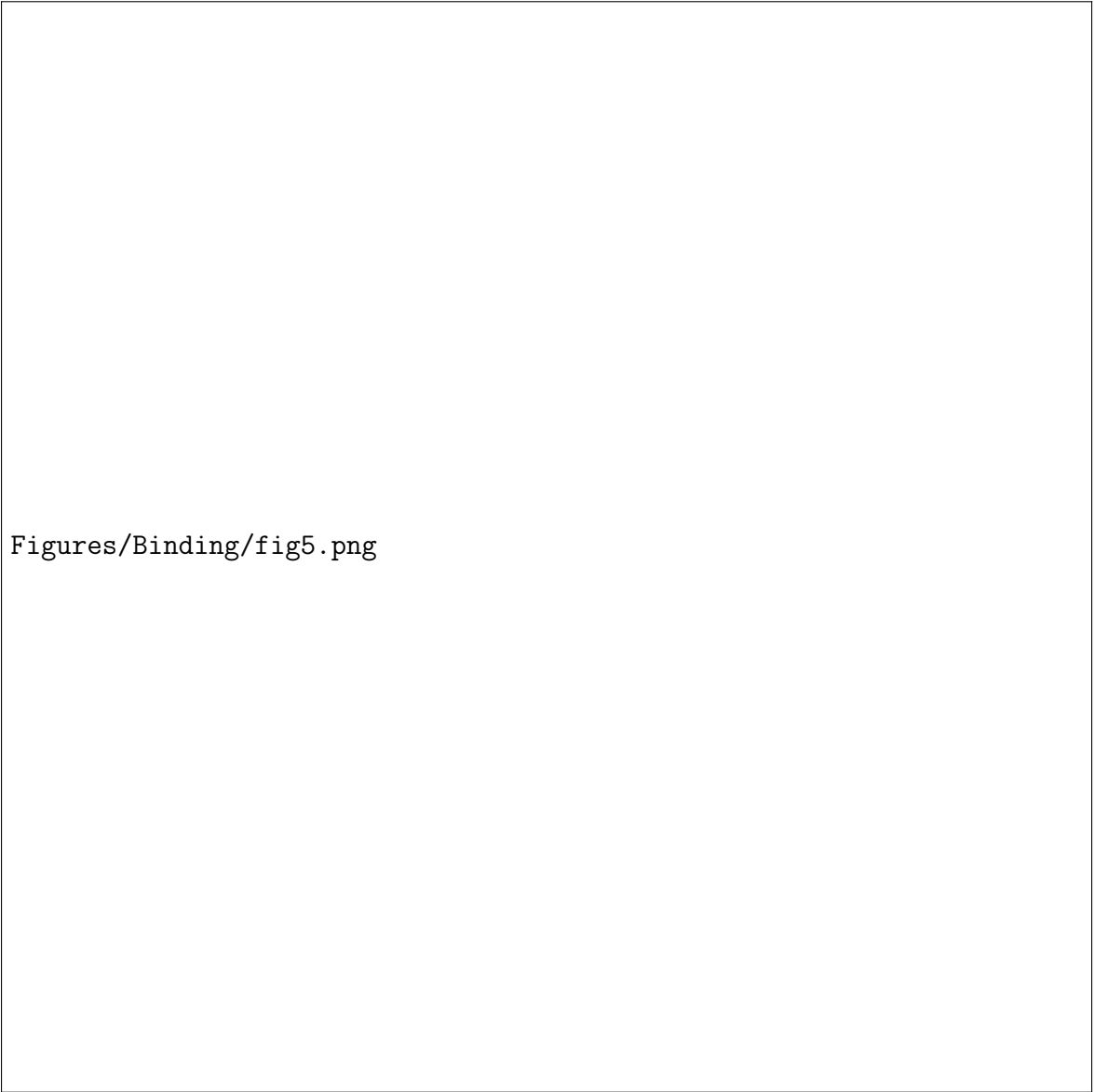
**Figure 4.4:** (A) The PMF of the $Na^+$ ion at Na1′ as it is pushed towards Na3. Only the converged part of the PMF is shown. (B) Residue-specific RMSDs of the NMDGT (310-314) and neighbouring residues (306-309) affected by the Na1′ → Na3 transition with respect to the 2NWX crystal structure. The RMSDs obtained from equilibrium simulations when the two $Na^+$ ions are at the (Na1, Na1′) and (Na1, Na3) states are shown with blue and black, respectively. The RMSDs obtained from the last umbrella window in the PMF, which indicates the barrier position, is shown with red.

site will occur last.

## 4.4 Conclusion

In order to resolve the large discrepancies found in the calculation of the binding free energies of ligands, we have re-examined the ligand-binding mechanism in the outward-facing conformation of $Glt_{Ph}$. We have included the intermediate $Na^+$ binding site

**Figure 4.5:** Schematic diagram of the ligand-binding order in Glt$_{Ph}$. (*A*) Initially, Glt$_{Ph}$ is in the open apo state. A Na$^+$ ion binds to the Na1′ site, coordinated by the D405 and D312 residues. (*B*) A second Na$^+$ ion pushes the first ion slightly and binds to the Na1 site coordinated by the D405 residue, thus forming the Na1–Na1′ two ion state. (*C*) The Asp substrate binds to Glt$_{Ph}$ in this two ion configuration. (*D*) The first Na$^+$ ion moves from Na1′ to Na3 chaperoned by the D312 side chain. The flipping of the N310 side chain oxygen from the Na1′ site to the Na3 site is the critical event that enables this transition.

labelled Na1′, and found it to be very stable. The coordination of the Na1′ site is tighter compared to Na3, and as evidenced by the binding free energy calculations, the Na1′ site is at a lower energy state than Na3. In addition, the transition of Na$^+$ from Na1′ to Na3 requires flipping of the N310 side chain, which is an unfavourable motion due to the stress it causes on the NMDGT motif. We have then considered the possibility that the second Na$^+$ ion could bind to Na1 while the first Na$^+$ is still at Na1′. The two ions at the Na1–Na1′ sites are found to be stable, separated by 3.5 Å. The

Coulomb repulsion of the first $Na^+$ at Na1' ensures that the second $Na^+$ binds to Na1 with a low affinity in good agreement with the experimental measurements [116–118]. We have next calculated the binding free energy of Asp for all ion occupation states and found that the lowest free energy is obtained when the two $Na^+$ ions are at the Na1–Na1' sites. The calculated binding free energy is again in good agreement with the experimental measurements, which indicate high-affinity Asp binding [116, 117].

The intermediate Na1' site, which has helped to resolve the low-affinity $Na^+$– high-affinity Asp binding paradox, is identified through the free energy simulations in this work. This site has not been observed in previous crystal structures, because it disappears in the fully-bound $Glt_{Ph}$. However, if the transition of the first $Na^+$ from Na1' to Na3 can be prevented through mutations, $Glt_{Ph}$ can be trapped in a state with two $Na^+$ ions at the Na1–Na1' sites. Inspection of Table 4.1 shows that two such mutations are T92A and S93A. In fact, these mutations have been considered before to identify the Na3 site [32]. Binding free energy calculations indicate that these mutations will reduce the $Na^+$ affinity to Na3 by 6–7 kcal/mol, which is sufficient to prevent the occupation of Na3. Thus the proposed Na1–Na1' intermediate binding site can be resolved if $Glt_{Ph}$ is crystallised with one of the above mutations.

The existence of the Na1' site will also affect the release mechanism of the ligands in the inward conformation of $Glt_{Ph}$, which has to be reconsidered. More interestingly, the Na1–Na1' sites could help to clarify the role of the $K^+$ ion in the transport cycle of EAATs. It has been hypothesised that the release of the last $Na^+$ ion is a rate-limiting step in $Glt_{Ph}$, and its exchange with $K^+$ in EAATs avoids this slow process [113, 143]. But no concrete mechanism could be proposed for such a $K^+$–$Na^+$ exchange. The Na1–Na1' sites provide a natural location for the $K^+$–$Na^+$ exchange to occur in EAATs. Finally, such an intermediate $Na^+$ binding site could exist in other $Na^+$– coupled transporters. Because they are not likely to be observed in crystal structures, one may have to rely on discrepancies in ligand binding free energies to infer their existence. We hope to report on the implications of the intermediate Na1' site on the transport cycle of EAATs and other transporters in feature publications.

The existence of the Na1' site will also affect the release mechanism of the last $Na^+$ from the Na3 site in the inward conformation of $Glt_{Ph}$, which has to be reconsidered. More interestingly, the Na1–Na1' sites could help to clarify the role of the $K^+$ ion in the transport cycle of EAATs, which have much faster turnover rates compared to $Glt_{Ph}$. It has been hypothesised that the release of the last $Na^+$ ion—without any external help—could be too slow for EAATs, and its exchange with $K^+$ could avoid this slow process [143]. But no concrete mechanism could be proposed for such a $K^+$–$Na^+$ exchange. The Na1–Na1' sites provide a natural location for the $K^+$–$Na^+$ exchange to occur in EAATs. Finally, such an intermediate $Na^+$ binding site could exist in other $Na^+$–coupled transporters. Because they are not likely to be observed in crystal structures, one may have to rely on discrepancies in ligand binding free energies to infer their existence. We hope to report on the implications of the intermediate Na1' site on the transport cycle of EAATs and other transporters in feature publications.

# 4.A    Appendix for Chapter 4



**Figure 4.A.1:** The convergence of TI results for the binding free energies of Na$^+$ ions. The black and red lines are the forward and backward transformations, respectively.

**Figure 4.A.2:** The convergence of the forward (black) and backward (red) transformations of Asp. The first row is the electrostatic interaction presented as the running averages of the TI calculations. The middle and bottom rows show the FEP calculations for the LJ–bb and LJ–sc interactions. These plots show the FEP transformation as a function of the coupling parameter $\lambda$.
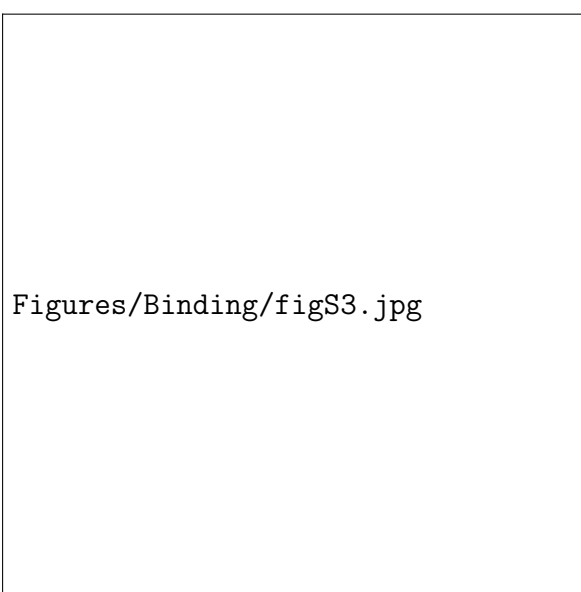
**Figure 4.A.3:** (A) Superposition of the ligand-free 2NWX and 4OYE Glt$_{Ph}$ structures after 60 ns of MD simulations, represented in orange and yellow, respectively. (B) The time evolution of the RMSDs of the two structures with respect to the 4OYE crystal structure. The average RMSDs after equilibration are 1.09 Å and 1.06 Å for the apo-2NWX and 4OYE systems, respectively. The HP2 gate moves freely in the apo state, and therefore excluded from the RMSD calculations.

# Chapter 5

# Escape of Na$^+$ from Na3 in Glt$_{\mathrm{Ph}}$

ABSTRACT

Glutamate/Aspartate transport through the mammalian excitatory amino acid transporters is coupled to the co-transport of three Na$^+$, one H$^+$ and the counter-transport of K$^+$. The archaeal homologue Glt$_{\mathrm{Ph}}$ is coupled to only the co-transport of three Na$^+$. The first two Na$^+$ sites (Na1,Na2) were resolved from both the outward and inward the crystal structure. The third site (Na3) was determined through a combination of mutagenesis experiments and molecular dynamics simulations and later verified in the crystal structure of Glt$_{\mathrm{Tk}}$. Previous studies using path-independent free energy methods have suggested unbinding of the last Na$^+$ from Na3 is the rate-limiting step. Here we use path-dependent methods to determine the path taken and energy required for Na$^+$ to leave the Na3 site to the bulk region to the bulk region. A detailed characterisation of structural changes of the binding pocket as Na$^+$ leaves the binding site are given. In addition, we estimate the release time of Na$^+$ using Smoluchowski mean-first-passage-time theory. Our observations of structural changes made along the path and estimated time of release are consistent with known experimental data. The results presented here indicate that the release of Na$^+$ in the third binding site is a slow process but not the rate-limiting step of the transport cycle. This investigation may shed light on the ligand release mechanism in the mammalian transporters.

## 5.1   Introduction

The mammalian glutamate transporters, also known as the excitatory amino acid transporters (EAATs), are responsible for clearing excess glutamate released at the synapses. Disruption of EAATs can lead to increased concentrations of glutamate, causing excitotoxicity of receptors. Such effects have been associated with many pathological conditions, including Alzheimer's disease, cerebral ischemia and amyotrophic lateral sclerosis [14]. EAATs function by cycling between the extracellular and intracellular medium. In each cycle, EAATs bind and transport the substrate (Asp or Glu) across the membrane by coupling three Na$^+$ and one H$^+$ ions [98]. After the release of the ligands, one K$^+$ ion is counter-transported to complete the cycle. While much has been learned about the functional properties of glutamate transporters from mutagenesis experiments, a mechanistic understanding of the transport process was not possible in the absence of a crystal structure.

The first crystal structure of aspartate/glutamate transporters was that of Glt$_{Ph}$ from Pyrococcus horikoshii—an archaeal homologue of EAATs [24]. In successive iterations of the Glt$_{Ph}$ crystal structure, first, the binding sites for the substrate and two Na$^+$ ions (called Na1 and Na2) were resolved in the outward-facing (OF) state [31], followed by the determination of the inward-facing (IF) [102], and intermediate conformations [103]. Physiological studies of Glt$_{Ph}$ revealed that it transported Asp (and not Glu) by coupling three Na$^+$ ions, but without the co-transport of a H$^+$ ion and the counter-transport of a K$^+$ ion [104, 107]. The third Na$^+$ site (called Na3) could not be resolved in the crystal structures. Several Na3 site have been proposed from electrostatic calculations [108] and molecular dynamics (MD) simulations [32, 109, 111]. The Na$^+$ coordination shell proposed in the last reference (T92, S93, N310, and D312 side chains and Y89 backbone) was consistent with the available mutagenesis data, and tests via the T92A and S93A mutations provided further support for this site is the Na3 site [32]. This was later verified with the recent crystal structure of Glt$_{Tk}$ that shows Na$^+$ bound to the Na3 site proposed earlier [33].

Although the overall sequence identity between Glt$_{Ph}$ and EAATs is low (37%), the homology for the binding pocket is close to 60%. In addition, the residues involved in ligand binding in Glt$_{Ph}$ are conserved in EAATs [13, 24, 31, 168]. Thus, Glt$_{Ph}$ provides a good starting model for mechanistic studies of the transport mechanism in glutamate transporters. Several computational investigations of Glt$_{Ph}$ have been performed so far which include: MD simulations of ligand binding and gating in the OF [105, 106, 111] and IF states [122], free energy calculations of ion and substrate binding in the OF [82, 109] and IF states [113], metadynamics simulations of substrate binding and release [120], and study of the transition from the OF to IF state using the anisotropic network models in combination with MD simulations [26, 125]. A homology model for EAAT3 was recently constructed using the Glt$_{Ph}$ structure, which provided further insights on the K$^+$ binding site [137] as well as elucidating the mechanism of H$^+$ transport [143].

Free energy calculations performed using the crystal structures of Glt$_{Ph}$ indicate that the binding order of ligands is Na3, Na1, Asp, and Na2 in the OF state [82], and they are released in the reverse order in the IF state [113]. While the predicted binding order is consistent with the recent experimental studies of Na$^+$ and Asp binding [116–

118], there are sizeable discrepancies in the binding free energies. These discrepancies are addressed in the Chapter 4 where the issue lies in not considering an intermediate state (Na1$'$) in the free energy calculations. The transition of Na1$'$ to Na3 requires conformational changes in the protein. Experimentally, substantial conformational changes are observed in the protein during ligand binding as large activation energies are obtained from calorimetric studies of Glt$_{Ph}$ [116–118]. More direct experimental evidence for such conformational changes is provided by the comparison of the recent apo structures of Glt$_{Ph}$ [144] and Glt$_{Tk}$ [136] with the fully bound Glt$_{Ph}$ structures [31, 102]. The path-independent free energy methods are clearly inadequate for description of ligand binding where conformational changes occur, and one has to appeal to path-dependent methods to capture the effect of such changes on the free energy of a ligand along the reaction coordinate.

In this work, we attempt to describe the conformational changes that occur in the binding pocket during the release of the last Na$^+$ ion from the Na3 site and provide an estimate for its release time. We do not consider the other ligands because they are observed to unbind during microseconds simulations, indicating that their release time is fast [122]. Using the crystal structure of the IF state of Glt$_{Ph}$ [102], we determine the path taken by the Na$^+$ ion from the Na3 site to the bulk region. We then perform umbrella sampling MD simulations to calculate the free-energy profile and the diffusion coefficient of the ion along this path. The free-energy profile and diffusion coefficient results are used in the mean first-passage solution of the Smoluchowski equation to get an estimate for the release time of the Na$^+$ ion. We also use the trajectory data from the umbrella sampling MD simulations to quantify the conformational changes that occur in the protein as the Na$^+$ ion moves from the Na3 site to bulk. Finally, we discuss the implications of our results for the transport mechanism in EAATs, which is much faster than that of Glt$_{Ph}$.

## 5.2 Method

### 5.2.1 Model System and Simulation Details

In this study, we use the crystal structure of Glt$_{Ph}$ in the inward-facing closed conformation (PDB ID: 3KBC). The crystal structure consists of the protomer with two Na$^+$ and Asp bound. The third Na$^+$ was added to the Na3 site observed in Glt$_{Tk}$ [33]. We embed the Glt$_{Ph}$ trimer in a 1-palmitoyl-2-oleoyl-phosphatidylethanolamine (POPE) phospholipid bilayer using the software VMD [147]. We then solvate this protein-lipid complex with 16204 water molecules along 34 Na$^+$ and 43 Cl$^-$ neutralising ions. After following the same standard equilibration protocol as in [113] we then remove the Na2, Asp and Na1 from the monomers simulating for 5 ns after the removal of each ligand. We further equilibrate the system for 50 ns without any restraints making sure the gate between HP1 and HP2 is open, and water molecules fill the binding pocket.

All MD simulations are performed using NAMD package (version 2.10) [148] with the CHARMM force field [164]. We use the NPT ensemble keeping the temperature constant at 300 K using the Langevin thermostat with a damping factor of 5 ps$^{-1}$. The pressure is maintained at 1 atm using the Langevin piston method with a damping

factor of 20 ps$^{-1}$ [79]. We utilise periodic boundary conditions, and electrostatic interactions are calculated using the particle-mesh Ewald (PME) method [78] without truncation. Non-bonded interactions are truncated at 12 Å and replaced with a smooth switching function starting from 10 Å. In all simulations, a time step of 2 fs is employed for the integrator.

## 5.2.2 Umbrella Sampling and Free Energy Calculations

In order to perform path-dependent free energy calculations, we need to establish an appropriate reaction coordinate (RC) for the Na$^+$ ion from the Na3 site to bulk. In most systems studied, the RC follows a straight line and can be aligned with one of the Cartesian axes, which simplifies the construction of the window positions for umbrella sampling. However, for the simulation system at hand, the path for the Na$^+$ ion follows a curved trajectory, and such a simple choice is not possible. To facilitate the construction of the umbrella windows, we exploit the presence of two intermediate binding sites along the RC to be denoted by Na1$'$ and Na1$''$ (details of these sites are discussed in Results). The path between any two neighbouring sites is approximately linear, thus they provide an initial path to build the actual RC. Three such vectors are connecting Na3–Na1$'$, Na1$'$–Na1$''$, and Na1$''$–bulk. To find the RC, we first perform steered MD (SMD) simulations along the direction of the Na3–Na1$'$ vector using a stiff-spring constant of 15 kcal/mol/Å$^2$ and a relatively slow speed of 1 Å/ns. A second SMD simulation with different initial conditions is observed to yield essentially the same path. Therefore, the average of the two trajectories is taken as the RC (more SMD trajectories would be included in the average if the paths were not similar). The same procedure is used for the vectors between Na1$'$–Na1$''$ and Na1$''$–bulk to obtain the complete RC. In the following, we will denote the curvilinear RC with $\zeta$ and its projection on the vectors between the binding sites by $z$. For convenience, umbrella potentials are constructed at 0.5 Å intervals along the $z$-coordinate at points $z_i$, and applied at the corresponding points $\zeta_i$ along the RC. That is, each window experiences a harmonic potential in the form

$$U = \frac{1}{2}k(\zeta - \zeta_i)^2, \tag{5.1}$$

where $\zeta$ is obtained from the projection of the position of the Na$^+$ ion on to the RC. A spring constant of 20 kcal/mol/Å$^2$ is used in the direction of the RC (defined by the tangent to the RC at $\zeta_i$), but no restraints are applied to the ion in the orthogonal plane. The reason for using a relatively large spring constant is to facilitate the calculation of the diffusion coefficient of the ion (see the section below for more details). In order to keep the reference to the RC absolute, we apply small restraints to the C$\alpha$ atoms of the protein residues outside the binding pocket with a spring constant of 0.1 kcal/mol/Å$^2$. The decision to exclude the C$\alpha$ atoms in the binding pocket from restraining follows from test umbrella sampling calculations, which have shown that the free-energy profile is affected when these residues are restrained. The trajectories obtained from the umbrella sampling MD simulations in each window are unbiased and combined using the weighted histogram analysis method (WHAM) [89] to construct the free-energy profile between two binding sites. For the Na1$''$ → bulk transition, we apply a flat-bottom funnel potential to reduce the phase space in the bulk [169]. The

funnel parameters employed are 12 Å, 0.6 rad, 1 Å and 1 kcal/mol/Å² for $R_{cyl}$, $Z_{cc}$, $\alpha$ and $k_{cyl}$ respectively. The funnel potential is implemented as a `tclForces` script in NAMD (see Appendix B for the script).

### 5.2.3 Diffusion Coefficient and Mean First-passage Time

In a previous work, Kramer's rate theory was used to estimate the release time of the Na$^+$ ion from the Na3 binding site, assuming that the escape path can be represented by a single potential well [113]. As already alluded, the escape path is more complicated than a single well, so a more sophisticated treatment of the release time is required. Here we employ a method that uses the results from the path-dependent umbrella sampling simulations as input. The time scale for the release time of Na$^+$ is much larger compared to the relaxation time of the ion's velocity correlations. This means that the inertial effects can be neglected and the motion of Na$^+$ inside Glt$_{Ph}$ can be described by the Smoluchowski equation [170]. The mean first-passage time solution of the Smoluchowski equation is given by [171]

$$\tau(a \rightarrow b) = \int_a^b d\zeta \frac{e^{W(\zeta)/k_B T}}{D(\zeta)} \times \int_a^\zeta d\zeta' e^{-W(\zeta')/k_B T}. \tag{5.2}$$

Here $W(\zeta)$ is the free-energy along the RC, $D(\zeta)$ is the path-dependent diffusion co-efficient, and $a$ and $b$ denote the initial and final states, respectively, which we choose as the bottom of the well and top of the barrier in the free-energy profile.

The diffusion coefficient is usually calculated using the velocity autocorrelation function (Green-Kubo relation), and the value obtained is isotropic. For path-dependent calculations, Woolf and Roux derived an expression for the diffusion coefficient by taking the Laplace transformation of the velocity autocorrelation function [172]. The diffusion coefficient is obtained via a numerical extrapolation process, which is not straightforward. A more convenient expression was derived by Hummer, who showed that the position, instead of the velocity autocorrelation function, can be used in the calculation of the diffusion coefficient [173]

$$D(\zeta_i) = \frac{var(\zeta_i)}{\tau_i}, \tag{5.3}$$

where $var(\zeta_i) = \langle \delta\zeta^2 \rangle_i = \langle \langle \zeta^2 \rangle - \langle \zeta \rangle^2 \rangle_i$ is the variance of the ion position along the RC at the $i^{th}$ window, and $\tau_i$ is the characteristic time of the normalised autocorrelation function of $\zeta$ at the $i^{th}$ window

$$\tau_i = \frac{\int_0^\infty \langle \delta\zeta(t).\delta\zeta(0) \rangle_i dt}{var(\zeta_i)}. \tag{5.4}$$

This expression is valid provided the system behaves like an overdamped harmonic oscillator. This behaviour can be enforced on the ion by using a sufficiently strong spring constant in the direction of the RC. We find that $k = 20$ kcal/mol/Å² is sufficient for this purpose. For smaller values of $k$, the autocorrelation function does not behave like an exponentially decaying function. The integration of the position autocorrelation function is done up to 1 ps using a total of 4 ns trajectory data for the ensemble average (Appendix C list the Fortran source code).

## 5.3 Results and Discussion

### 5.3.1 Escape Path of Na$^+$ from the Na3 Site

In order to search for potential binding sites, the Na$^+$ ion is steered from the Na3 site towards the Na1 site in SMD simulations. Several SMD simulations are performed with different initial conditions to ensure adequate sampling. At the end of the steering, we allow the system to relax by performing 10 ns MD simulation. At the end of the equilibrium simulation, we find a stable site between the Na3 and Na1 sites, which will be referred to as Na1$'$. SMD simulations between the Na3 and the Na1$'$ sites are repeated to make sure that the proposed Na1$'$ site is unique. The Na1$'$ site is coordinated by both the D312 and D405 side chains (see Table 5.1 for the coordinating residues and the average Na$^+$–O distances). A similar intermediate state was also observed in MD simulations of the OF conformation of Glt$_{Ph}$ [111]. We note that, as in the Na3 site, there are no water molecules in the coordination shell of Na$^+$ in the Na1$'$ site.

Superposing the current configuration with a system containing two Na$^+$ ions bound to the Na1 and Na3 sites, we observe about 3 Å distance between the Na1$'$ and Na1 sites. This suggests that there may be another binding site in the vicinity of the Na1 site. Again we perform SMD simulations, steering the Na$^+$ ion from the Na1$'$ site towards the Na1 site. After equilibration, a stable binding site is found at the location of the Na1 site, which will be called Na1$''$. The coordination shell of Na1$''$ is very similar to

**Table 5.1:** Residues coordinating the Na$^+$ ion in the Na3, Na1$'$, and Na1$''$ sites, and at the transition states (TS) in between.$^a$

| Helix-Residue | Na3 | TS1 | Na1$'$ | TS2 | Na1$''$ | TS3 |
|---|---|---|---|---|---|---|
| TM3–Y89 (O) | $2.3 \pm 0.1$ | $5.2 \pm 0.2$ | - | - | - | - |
| TM3–T92 (OH) | $2.4 \pm 0.1$ | $3.8 \pm 0.2$ | - | - | - | - |
| TM3–S93 (OH) | $2.4 \pm 0.1$ | $5.0 \pm 0.3$ | - | - | - | - |
| TM7–D312 (O1) | $3.5 \pm 0.2$ | $2.4 \pm 0.1$ | $2.3 \pm 0.1$ | $2.2 \pm 0.1$ | - | - |
| TM7–D312 (O2) | $2.1 \pm 0.1$ | $2.6 \pm 0.2$ | $2.3 \pm 0.1$ | $3.2 \pm 0.5$ | - | - |
| TM7–N310 (O$_\delta$) | $2.2 \pm 0.1$ | $2.2 \pm 0.1$ | $2.3 \pm 0.1$ | $2.7 \pm 0.3$ | $2.3 \pm 0.1$ | $8.6 \pm 0.4$ |
| TM7–N310 (O) | $6.2 \pm 0.2$ | $5.0 \pm 0.3$ | $6.3 \pm 0.2$ | $5.9 \pm 0.3$ | $5.2 \pm 0.6$ | $7.0 \pm 0.8$ |
| TM8–N401 (O) | - | $4.8 \pm 0.2$ | $2.4 \pm 0.1$ | $2.4 \pm 0.2$ | $2.7 \pm 0.4$ | $8.7 \pm 1.0$ |
| TM8–D405 (O1) | - | $4.4 \pm 0.2$ | $3.9 \pm 0.2$ | $3.0 \pm 0.4$ | $2.2 \pm 0.1$ | $7.4 \pm 1.1$ |
| TM8–D405 (O2) | - | $5.9 \pm 0.2$ | $2.2 \pm 0.1$ | $2.2 \pm 0.1$ | $2.3 \pm 0.1$ | $7.0 \pm 1.3$ |
| TM7–G306 (O) | - | - | - | $4.1 \pm 0.5$ | $2.5 \pm 0.3$ | $4.9 \pm 1.6$ |
| TM7–A307 (O) | - | - | - | - | $5.7 \pm 0.4$ | $3.6 \pm 1.2$ |
| H2O (1) | - | - | - | - | $2.4 \pm 0.3$ | $2.4 \pm 0.2$ |
| H2O (2) | - | - | - | - | - | $2.3 \pm 0.1$ |
| H2O (3) | - | - | - | - | - | $2.4 \pm 0.3$ |
| H2O (4) | - | - | - | - | - | $2.6 \pm 0.5$ |
| H2O (5) | - | - | - | - | - | $2.7 \pm 0.5$ |

$^a$The average Na$^+$–O distances calculated from 2 ns MD simulations are presented in columns 2–7 (in Å). The Na1 coordination shell differs from that of Na1$''$ by the replacement of N310 (O$_\delta$) with N310 (O).

that of Na1 (see Table 5.1)—the only difference is that the carbonyl oxygen of N310 is replaced by the side-chain oxygen of N310. In this site, D312 no longer coordinates $Na^+$, and the coordination shell is completed by the G306 carbonyl oxygen and a water molecule. To see if the same coordination shell as in the Na1 site can be achieved, we generate a new configuration starting with two $Na^+$ ions present in Na1 and Na3 sites. After removing the $Na^+$ ion from the Na3 site and equilibrating for 20 ns, we see the $Na^+$ ion moving from the Na1 site to the Na1' site instead. Thus it appears that the Na1'' configuration, where $Na^+$ is coordinated by N310 ($O_\delta$), is only accessible when $Na^+$ is steered from the Na3/Na1' sites to the Na1 position.

From the Na1'' site, there is only one obvious path for $Na^+$ to take. The Na1'' site is solvent-exposed, and we expect it to be the last site before $Na^+$ comes to a bulk-like environment outside the binding pocket. We steer the $Na^+$ ion from the Na1'' site towards the HP1-HP2 gate. Along the SMD trajectory, $Na^+$ has a brief contact with the A307 (O) along with G354 (O) and S279 (OH), which are residues part of HP1-HP2, but does not interact with any other residues in the binding pocket. Hence, we deduce that Na1'' is the last binding site before bulk. Summarising the steering simulations, we predict the last $Na^+$ ion to leave $Glt_{Ph}$ through the following path: Na3 $\rightarrow$ Na1' $\rightarrow$ Na1''$\rightarrow$ Bulk. The full transition path of the $Na^+$ ion from the Na3 site to bulk and the $Na^+$ binding sites are shown in Fig. 5.1.

It is clear from Table 5.1 and Fig. 5.1 that the residues N310, D312, and to a lesser degree, D405 must undergo substantial conformational changes during the release of the $Na^+$ in order for their side-chain oxygen atoms to keep coordinating it. To help
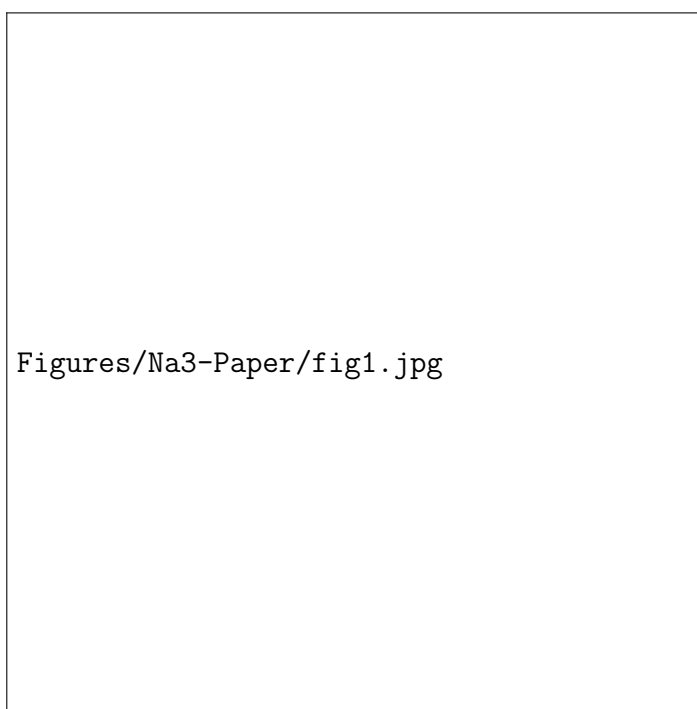


**Figure 5.1:** Full transition path of the $Na^+$ ion from the Na3 site to bulk (black line). The binding sites are indicated with yellow balls. The snapshot shows the positions of the key residues when the $Na^+$ ion is in the Na3 site.

visualising these changes, we show snapshots of the three transitions, Na3 $\rightarrow$ Na1′, Na1′ $\rightarrow$ Na1″, and Na1″ $\rightarrow$ Bulk, in Fig. 5.2. In each transition, the final conformations of the key residues (N310, D312, and D405) are superimposed on the initial ones using different colours, to indicate the extent of the conformational changes. The role of these changes in facilitating the release of the Na$^+$ ion from the Na3 site will be further discussed and quantified when we analyse the trajectory data obtained from the umbrella sampling simulations. It is important to note that the path described above may not be the lowest free energy path and that there may be other paths. However, based on the trial SMD simulations, this is the most likely path that is available.

### 5.3.2 Free Energy from Umbrella Sampling Simulations and Conformational Changes

Using the RC described in the methods, we perform umbrella sampling simulations along the three transition paths, A) Na3 $\rightarrow$ Na1′, B) Na1′ $\rightarrow$ Na1″, and C) Na1″ $\rightarrow$ Bulk. For each transition path, a free-energy profile is constructed from the RC of the Na$^+$ ion using WHAM (Fig. 5.3, top). Because a large spring constant is used, the overlap of the RC distributions between the neighbouring windows is observed to drop below the critical value (<5%) in several places. Extra windows are inserted in such places to avoid any simulation artefacts in the free-energy profile due to poor sampling. This was a problem, especially near the transition states, where window separations as small as 0.1 Å had to be used. The windows are obtained by performing SMD simulations from the closest umbrella window available. Evidence for the convergence



**Figure 5.2:** Transition path of the Na$^+$ ion: (A) Na3 $\rightarrow$ Na1′, (B) Na1′ $\rightarrow$ Na1″, and (C) Na1″ $\rightarrow$ Bulk. Only the important residues (N310, D312 and D405) are shown for clarity. Red, blue, and black balls represent the O, N, and C$\alpha$ atoms of these residues. Conformations of these residues in the initial and final states are shown in blue and orange, respectively. Yellow balls indicate the average position of the Na$^+$ ion in each umbrella window. Interactions of N310 (N$_\delta$) with the D312 and D405 side chains are indicated by dotted lines.

of the free-energy profiles is presented in Fig. 5.A.1 in the Supporting Material.

The trajectory data from the umbrella sampling simulations are also used to track the changes in the coordination shell of the $Na^+$ ion as it moves along each transition path (Fig. 5.3, middle), and to calculate its diffusion coefficient using Eq. (5.3) (Fig. 5.3, bottom). The side chains oxygens of N310 and D312 from TM7, and N401 and D405 from TM8 remain in the coordination shell of the $Na^+$ ion over substantial parts of the transition paths. It is of interest to find out whether the backbone atoms near these residues also move to help coordinating the $Na^+$ ion besides the side chains. For this purpose, we have calculated the RMSD of the $C\alpha$ atoms of the N310–T314 residues in TM7 (NMDGT motif), and the N401–D405 residues in TM8 using the configuration when $Na^+$ is in Na3 as reference (Fig. 5.4).

Below we discuss the critical events in each transition and correlate the energetic features of the free-energy profile with the conformational changes that occur in the protein using Figs. 5.2–5.4.

A) Na3 → Na1′ (Fig. 5.3 A): The two sites are separated by about 5.5 Å, and the transition state (TS1) is near $z = 3.1$ Å. After the $Na^+$ ion leaves the Na3 site, first S93 (OH) departs from its coordination shell followed by Y89 (O) and then T92 (OH). This is partly compensated by the full integration of D312 (O1) in the shell. Both the N310 and D312 side chains faithfully track the $Na^+$ ion in this region. The ion remains under-coordinated while dragging the N310 and D312 side chains, which results in a steeply rising free-energy profile up to TS1. The height of the energy barrier at TS1 is 17.3 kcal/mol, which is an extremely high barrier for an ion to cross. Entry of N401 (O) to the coordination shell at TS1 provides some relief to the free-energy profile. Finally, with the entry of D405 (O1) to the coordination shell, the Na1′ binding site is formed. Inspection of Table 5.1 shows that the quality of the coordination shell of $Na^+$ at Na1′ is at least as good as that at Na3, yet the $Na^+$ free energy at Na1′ is 6.9 kcal/mol higher than that at Na3. We attribute this to the conformational changes that occur both in the N310 and D312 side chains (Fig. 5.2 A), and in the backbone of the NMDGT motif (Fig. 5.4 A) during the transition. The RMSDs of the $C\alpha$ atoms in NMDGT change little between Na3 and TS1 but exhibit larger changes between TS1 and Na1′.

The diffusion coefficient profile of $Na^+$ along this path remains below the bulk value (i.e. $1.33 \times 10^{-9}$ m²/s [174]) even at the transition state. This is because the ion is well coordinated by oxygen atoms through the transition (Fig. 5.3 A). If the ion is not well coordinated, then the environment will result in a smaller friction coefficient for the ion and the diffusion profile increases. We next combine the free-energy profile and diffusion coefficient results in Eq. (5.2) to estimate the escape time for the Na3 → Na1′ transition, which yields about 7 sec. This is a very slow process for an ion but is a small fraction (4%) of the observed turnover time of ∼3 min in $Glt_{Ph}$ [104].

An interesting question here is the chaperone role played by the D312 side chain, and how much this helps the $Na^+$ ion to cross the Na3 → Na1′ barrier. To address this question, we have constructed another free-energy profile for this transition while keeping the D312 residue restrained in its initial position at the Na3 site. As shown in Fig. 5.5, the energy barrier faced by the $Na^+$ ion is nearly doubled when the D312
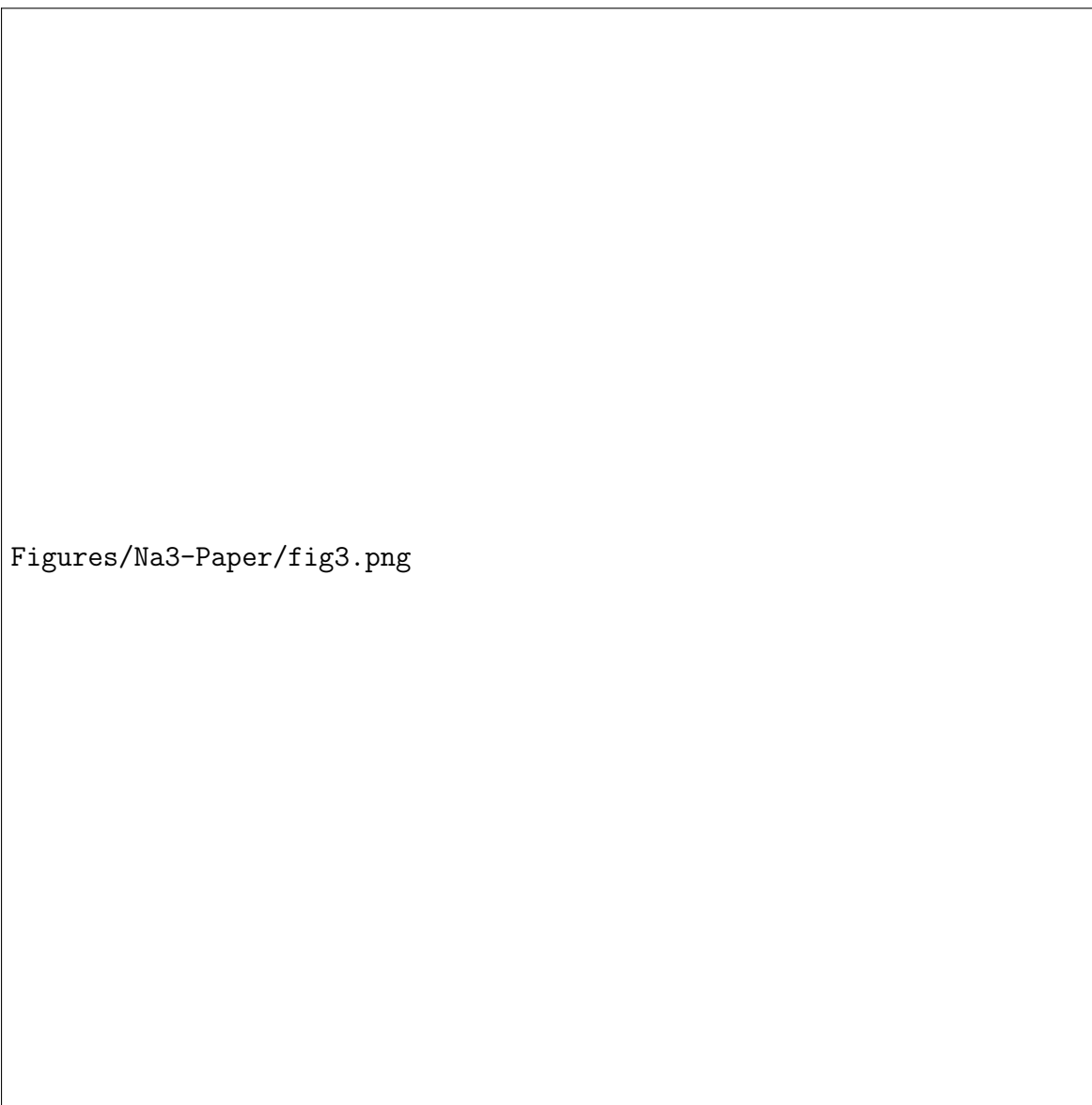
**Figure 5.3:** Free-energy profiles (top), the Na$^+$–O distances (middle), and the relative diffusion coefficients (bottom) for the transitions (A) Na3 → Na1′, (B) Na1′ → Na1″, and (C) Na1″ → Bulk. The bulk diffusion coefficient of Na$^+$ ions is taken as $D_0 = 1.33 \times 10^{-9}$ m$^2$/s [174].

residue is restrained. To see the effect of the higher barrier on the escape time, we repeat the above calculation using the restrained free-energy profile in Fig. 5.5. We find that the escape time has increased from 7 sec to $10^{13}$ min or 18 million years. Thus chaperoning by the D312 side chain plays an essential role in facilitating the escape of the Na$^+$ ion from the Na3 site. While the N310 side chain also tracks the Na$^+$ ion, this is not solely due to the N310 (O$_\delta$)–Na$^+$ interaction. As shown in Figs. 5.2 A–B, N310 (N$_\delta$) forms an ionic bond with D405 (O1), which provides an extra incentive for the movement of the N310 side chain. Presumably, this interaction is also responsible for the exclusion of D405 (O1) from the coordination shell of Na$^+$ at the Na1′ site (Table 5.1).

Figures/Na3-Paper/fig4.jpg

**Figure 5.4:** RMSD of the C$\alpha$ atoms of residues in TM7 (A) and TM8 (B). The reference frame is chosen as the state with Na$^+$ in the Na3 site. RMSDs are calculated at all Na$^+$ binding sites, and the transition states in between.

B) Na1$'$ $\rightarrow$ Na1$''$ (Fig. 5.3 B): The sites are separated by 2 Å, and TS2 is at $z = 1.2$ Å. The Na$^+$ ion remains well-coordinated throughout this region. In particular, at TS2 both D312 and D405 side chains coordinate the ion. After TS2, D312 completely decouples from the ion with its backbone and side chain relaxing back while both side-chain oxygens of D405 firmly couple to the ion (Fig. 5.2 B). But functionally the most significant event in this transition is the conformational change exhibited by N310— position of its backbone undergoes a large shift and its side chain flips (Fig. 5.2 B). It is seen from Fig. 5.4 that the stress created on the NMDGT motif by the binding of the Na$^+$ ion to the Na1$'$ site becomes maximal at TS2 and is relieved only after the ion binds to Na1$''$. Flipping of the N310 backbone is not driven by ion coordination, which actually gets a bit worse at TS2. Rather, it enables N310 (N$_\delta$) to switch its

H-bond partner from D405 (O1), which is needed for the coordination of Na$^+$, to D312 (O1) which has become free after TS2 ([Fig. 5.2 B](#)). In this position, N310 (N$_\delta$) also



**Figure 5.5:** Effect of the restraining of the D312 residue on the Na3 $\to$ Na1$'$ free-energy profile. The restraint-free profile from [Fig. 5.3 A](#) (black) is compared to the profile obtained while D312 is restrained (red).
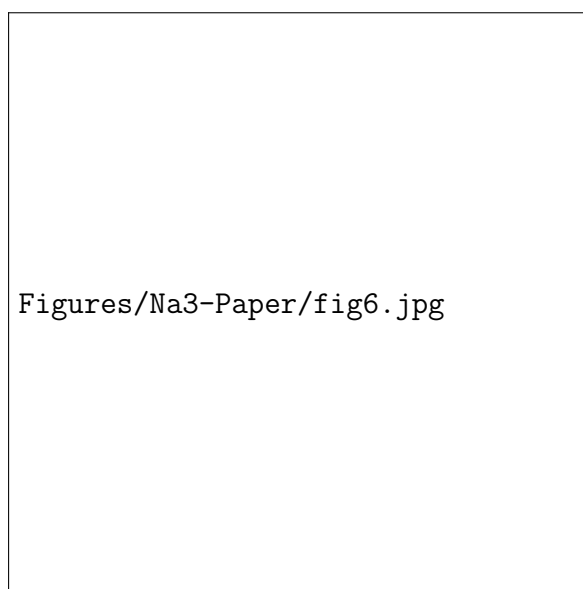


**Figure 5.6:** Ratchet-like behaviour of N310 as Na$^+$ moves from the Na1$'$ site (A) to the Na1$''$ site (B). N310 (N$_\delta$) switches its H-bond from D405 in (A) to D312 in (B), which sets up the H-bond network that prevents Na$^+$ from falling back to Na1$'$. H-bonds are indicated with dashed lines.

**Figure 5.7:** Comparison of the free-energy profiles for the forward (black) and reverse (red) transitions between the Na1′ and Na1″ sites. The reverse energy barrier exceeds 16 kcal/mol, which is well over the barrier for the Na1″ → Bulk transition.

makes an H-bond with N401 (O), thus setting up the H-bond network, D312 (O1)–N310 ($N_\delta$)–N401 (O)–D405 (N) (Figs. 5.6 A–B). The snapshot in Fig. 5.6 B suggests that this H-bond network may prevent the backward motion of the Na$^+$ ion from Na1″ to Na1′.

Inspection of the free-energy profile in Fig. 5.3 shows that this is a real possibility for the ion at Na1″. For Na$^+$ at Na1′, the forward and backward barriers are 4.9 and 10.4 kcal/mol, respectively, so the forward motion is much more likely. But for Na$^+$ at Na1″, the forward and backward barriers are 14.3 and 2.4 kcal/mol respectively, which clearly prefers the backward motion. However, due to the formation of the H-bond network after the ion goes over TS2, the Na1′ → Na1″ transition is not reversible, and the free-energy profile in Fig. 5.3 B is valid only in the forward direction. To show what happens in the reverse direction and demonstrate the effect of the irreversibility quantitatively, we have constructed another free-energy profile for the reverse transition Na1″ → Na1′. As shown in Fig. 5.7, there is a rising free-energy profile in the reverse direction well beyond TS2. Thus the Na$^+$ ion cannot breach the H-bond network, and it has to move forward to bulk. A clearer illustration of the ratchet-like behaviour of N310 during the Na1′ → Na1″ transition is provided in Fig. 5.6. The ratchet function of N310 is as important in facilitating the release of the Na$^+$ ion as its chaperoning by D312 during the Na3 → Na1′ transition. However, it is important to note that the ratchet process may be a result of micro-irreversibility due to limited sampling. Long-time MD simulations in the scale of microseconds may reveal a different picture.

A similar H-bond network but without the involvement of D312 was also observed in

MD simulations of the OF conformation of Glt$_{Ph}$ [111]. It was noted that this H-bond network prevented hydration of the Na3 site, but the presence of a Na$^+$ ion at the Na1 site was sufficient to break the H-bond network and access the Na3 site [111]. As mentioned earlier (Table 5.1), Na$^+$ is coordinated by N310 (O) at the Na1 site, and not by the N310 side chain, which makes only one H-bond with N401 (O). Thus it is relatively easy for Na$^+$ to break this single H-bond as observed in MD simulations. In contrast, at the Na1″ site of the IF conformation, the N310 side chain is involved in two H-bonds besides coordinating Na$^+$ (Fig. 5.6 B). As demonstrated by the reverse free-energy profile (Fig. 5.7), it is much harder to break this H-bond network, which is fortified by the addition of the D312 side chain.

In other respects, the Na1′ → Na1″ transition is rather ordinary. The diffusion coefficient of Na$^+$ remains near the bulk values, and the energy barrier at TS2 (4.9 kcal/mol) is rather low. The transition time estimated using the free-energy profile and diffusion coefficient results in Fig. 5.3 B yields about 10 ns. This is nine orders of magnitude smaller than the escape time from the Na3 site and hence completely negligible compared to it.

C) Na1″ → Bulk (Fig. 5.3 C): Here, the transition state on the free-energy profile extends to about 4.5 Å. As the Na$^+$ ion moves towards bulk, N401 (O) is the first to depart its coordination shell. The N310 and D405 side chain oxygens track Na$^+$ for about 4 Å and start decoupling after that. During the initial pulling stage from N310 and D405, a second water molecule enters the coordination shell of Na$^+$. We note that the TM8 backbone remains relatively rigid during the release of Na$^+$ (Fig. 5.4), which partly explains the limited chaperoning role played by D405 compared to D312. The under-coordination of Na$^+$ while pulling the N310 and D405 side chain is responsible for the steep rise in the profile in the region $z = 0$–$5$ Å. After decoupling of N310 and D405 (i.e. from $z = 5$ Å) the ion is coordinated by about 5 water molecules as shown in Fig. 5.8. At this point, Na$^+$ is in the third transition state TS3 and is coordinated by one shell of water molecules. It is important to note that the Na$^+$ ion is not out the bulk medium yet as it is still inside the binding pocket even though it is coordinated with 5 water molecules. The number of water molecules coordinating Na$^+$ decreases to three as it passes through the HP1-HP2 gate. From 6–10 Å the Na$^+$ weakly interacts with A307 (O), G354 (O) and S279 (OH) residues, which is responsible for the dip in the free-energy profile from the transition state and reduced number of water molecules. From 10 Å onwards Na$^+$ detaches itself from any protein oxygens with S279 (OH) being the last and is coordinated by about 6 water molecules, Fig. 5.8. As shown in the free-energy profile beyond 10 Å the energy is flat signalling that the ion has reached a bulk-like medium.

The peak value of the free-energy profile is 14.3 kcal/mol, which is a substantial energy barrier for the ion to surmount and exit into the bulk. At the bulk (the flat region in the plot) the energy is 11.1 kcal/mol. The diffusion coefficient of the ion fluctuates around half the bulk value and approaches the bulk value near the flat region of the profile reflecting the exit and entry of various protein and water oxygens to the coordination shell of the ion during the extended transition region. The escape time obtained from the free-energy profile and diffusion coefficients in Fig. 5.3 C is about 0.5 s. While this is slower than the Na1′ → Na1″ transition, it is only one order of magnitude faster

**Figure 5.8:** The average number of water molecules entering the coordination shell of $Na^+$ during the $Na1'' \rightarrow$ Bulk transition. The results are obtained from the umbrella sampling windows using a cutoff radius of 3 Å for the $Na^+$–O distance.

compared to the escape time from the Na3 site. Thus the slowest step for the release of $Na^+$ from Na3 is the Na3 $\rightarrow$ Na1′ transition.

### 5.3.3 Implications for $Glt_{Ph}$

The detailed discussion of the release of the $Na^+$ ion from the Na3 site to bulk shows that the protein does not just provide a passive conduit but is actively involved in the release of the ion through the chaperon and ratchet functions of the D312 and N310 residues in the NMDGT motif. This is consistent with the calorimetric studies [116–118] and comparison of the apo and bound structures of $Glt_{Ph}$ [144], which indicate that substantial conformational changes occur in the protein during $Na^+$ binding. An important consequence of these observations is that binding free energy calculated using path-independent methods with unvarying $Glt_{Ph}$ structures cannot provide an accurate description of the ligand binding/unbinding process. For example, the energy for the translocation of $Na^+$ from the Na1″ site differs between the two methods. The free-energy profile for this transition yields an energy difference of −11.1 kca/mol (Fig. 5.3 C) while TI calculations give −17.5 kcal/mol (Fig. 5.A.2). This difference is due to the conformational changes as a result of doing work on the protein as the ion unbinds. In this case, the chaperoning of the $Na^+$ ion by the N310 and D405 side chain up to TS3 helps reduce the energy barrier in the profile and hence the translocation energy. Therefore, to obtain the correct value with path-independent methods, the conformational energy needs to be taken into account. As mentioned previously, the estimated escape time of $Na^+$ is around 7 sec, yet the measured turnover

rate is 3 min [104]. Thus even though the release process is it is not be the rate-limiting step in the transport cycle. The rate-limiting step in the transport cycle is most likely the conformational transition of the protein from the OF to IF state. Work has been done on this transition using the anisotropic network model (ANM) with coarse-grained models of Glt$_{Ph}$ [125]. However, an estimate of the transition time has not been calculated, and further studies are needed.

### 5.3.4  Implications for EAATs

Transport rates in EAATs are much faster than Glt$_{Ph}$, e.g., the turnover time for EAAT3 is about 0.01 s [175], which is 2000 times faster than Glt$_{Ph}$. Because the residues involved in ligand binding are conserved between Glt$_{Ph}$ and EAATs, one has to consider other factors for the speed up. Computational studies on a homology model of EAAT3 have shown that a K$^+$ ion can bind to a site near the Na1$'$–Na1$''$ sites [137, 143]. As discussed in the Na3 $\rightarrow$ Na1$'$ transition, the tunnel connecting the two sites is too narrow to allow any exchange of the Na$^+$ and K$^+$ ions. This also appears to be the reason why Na$^+$ at the Na3 site could not be exchanged with Tl$^+$ in Glt$_{Ph}$ crystal structures [31, 102]. Inspection of the EAAT1 crystal structure [130] shows that this tunnel is larger than that in Glt$_{Ph}$. Thus the barrier of Na3 $\rightarrow$ Na1$'$ is expected to decrease in EAATs. In addition, the exchange of K$^+$ and Na$^+$ will further reduce the transport cycle time, which could explain the extremely fast process compared to Glt$_{Ph}$. The exchange process of Na$^+$ and K$^+$ requires further experimental and computational studies to clarify this hypothesis.

## 5.4  Conclusion

Compelled by the recent experimental findings that conformational changes accompany binding of Na$^+$ ions in Glt$_{Ph}$, we have performed umbrella sampling calculations for the release of the last Na$^+$ ion from the Na3 site to bulk in the IF state of Glt$_{Ph}$. Identifying the rate-limiting steps in Glt$_{Ph}$ is expected to provide important clues on how the coupling of H$^+$ and Na$^+$ ions in EAATs could speed up the transport rate by up to three orders of magnitude. Our estimate for the release time of the last Na$^+$ ion, obtained from the free-energy profile and diffusion coefficient results, is about 7 sec. This is a small fraction of the observed turnover time of 3 min (about 4%), which suggests that the release of the last Na$^+$ ion is a very slow process but is not the rate-limiting step in Glt$_{Ph}$. The transition from the OF to IF is expected to be a slower process than the release of Na$^+$.

The umbrella sampling simulations, performed along the three transition paths identified between the Na3 site and bulk, have revealed that the D312 and N310 residues in the NMDGT motif undergo substantial conformational changes, and thereby play crucial roles in facilitating the release of the Na$^+$ ion. As shown in Fig. 5.5, it would be impossible for the Na$^+$ ion to escape from the Na3 site without the chaperoning of the D312 side chain. Similarly, the N310 side chain chaperones the Na$^+$ ion from the Na3 site all the way to the protein/bulk interface. Another contribution of N310 to the release of the Na$^+$ ion is its ratchet function during the Na1$'$ $\rightarrow$ Na1$''$ transition, which

prevents the ion from going back in the short timescale available in the current MD simulations (Figs. 5.6 and 5.7). Much longer simulations in the microseconds timescale may demonstrate the reversibility of the N310 side chain. Also, the path investigated may or may not be the lowest free energy path. Further investigation is required, and more advanced computational method like the string-method with swarms of trajectories [176] may be needed.

# 5.A    Appendix for Chapter 5



**Figure 5.A.1:** The convergence of the free-energy profile shown in Fig. 5.3 from block data analysis. In each case, 3 ns of production is collected after equilibration, which took up to 2 ns (not shown). The free-energy profile, obtained from 1 ns blocks, is seen to fluctuate around the baseline profile obtained from the total production data.
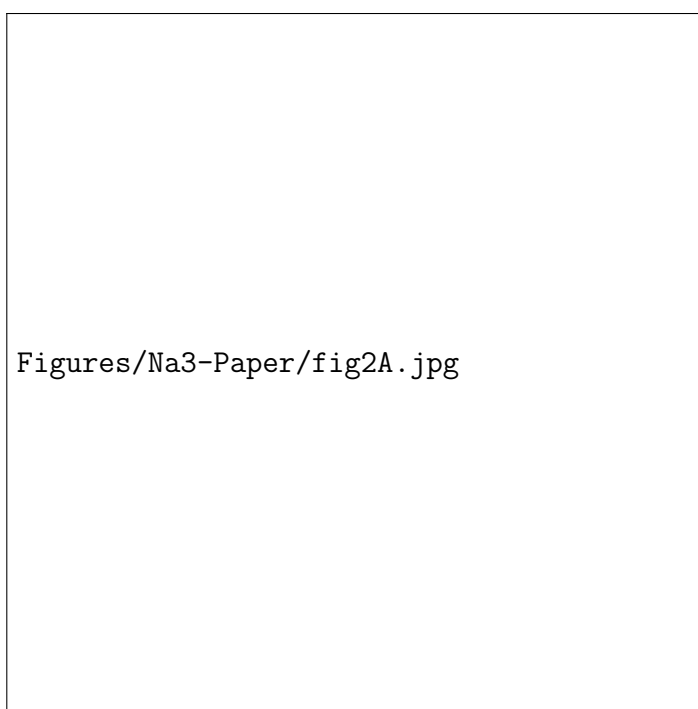
**Figure 5.A.2:** The convergence of TI calculations for translocating the $Na^+$ ion from bulk to the $Na1''$ site. Calculations are performed using the protocols in Refs. [82, 113]. The difference between the forward and backward TI results is less than a kcal/mol, indicating negligible hysteresis. The average of the forward and backward results yields $-17.5$ kcal/mol for the translocation free energy of $Na^+$ from bulk to the $Na1''$ site.

# Chapter 6

# Effect of Lipid Membranes on $K^+$ Permeation in gA

ABSTRACT

Membrane proteins are embedded in a lipid bilayer and interact with the lipid molecules in subtle ways. This can be studied experimentally by examining the effect of different lipid bilayers on the function of membrane proteins. Understanding the causes of the functional effects of lipids is difficult to dissect experimentally but more amenable to a computational approach. Here we perform molecular dynamics simulations and free energy calculations to study the effect of two lipid types (POPC and NODS) on the conductance of the gramicidin A (gA) channel. A larger energy barrier is found for the $K^+$ potential of mean force in gA embedded in POPC compared to that in NODS, which is consistent with the enhanced experimental conductance of cations in gA embedded in NODS. Further analysis of the contributions to the potential energy of $K^+$ reveals that gA and water molecules in gA make similar contributions in both bilayers but there are significant differences between the two bilayers when the lipid molecules and interfacial waters are considered. It is shown that the stronger dipole moments of the POPC head groups create a thicker layer of interfacial waters with better orientation, which ultimately is responsible for the larger energy barrier in the $K^+$ PMF in POPC.

## 6.1   Introduction

Gramicidin A (gA) is an antibiotic peptide that disrupts bacteria by allowing unimpeded flow of cations into the plasma membrane. This small hydrophobic peptide is composed of two identical helical subunits with 16 residues on each. The two subunits form a stable compound only inside a lipid bilayer environment. When stable, the gA dimer forms a narrow cylindrical hole across the bilayer, through which water and monovalent cations can permeate near diffusion rates. The channel structure was determined using solution [177] and solid-state NMR [178], and consists of a single-stranded, right-handed $\beta$-helical dimer. Each subunit is made up of formyl-VGA<u>L</u>A<u>V</u>V<u>V</u>W<u>L</u>W<u>L</u>W<u>L</u>W-ethanolamine (underlined residues indicate D-amino acids and L-amino acids otherwise) [179]. The alternating L-D amino acid sequence allows the peptide to fold into a helix with the side chains aligned on the exterior of the helix [180]. There are two high-resolution structures available, PDB:1MAG [36] and PDB:1JNO [37]. These two structures have been used in a large number of computational studies, investigating the ion permeation properties of gA [40–42, 181]. Due to its simple structure, gA has often been used as a model for membrane proteins. For example, it was used as a prototype ion channel model long before the first potassium ion channel protein was crystallised [182, 183]. It has also been used as a testing model for developing and validating computational methods from continuum theories to *ab initio* molecular dynamics (MD) [41, 184–187].

Biological membranes play an important role in cell biology. They consist of lipid molecules organised in a bilayer formation, leaving a $\sim$30 Å hydrophobic layer between the intracellular and extracellular environment of cells [188]. Membrane proteins embedded in lipid bilayers may have their functions modulated by the surrounding environment as shown in some studies [189–192]. Because gA is a small protein, the effects of lipids on its function is expected to be greater than those of larger membrane proteins. Hence, gA is a good candidate to investigate the effects of lipid environment on protein function.

The effect of membrane environment on gA function have been investigated in several studies previously [46, 47, 193–195]. Of particular interest is the work of Cukierman and co-workers, where gA is embedded in four different lipid environments; monoglycerides, ceramides, sphingolipids and phospholipids [46]. The experiments showed a markedly decreased ion conductance for gA in a phospholipid environment compared to ceramides. On the other hand, the proton conductance exhibits the opposite behaviour with phospholipids enhancing conductance [47]. Enhancement in proton transport is attributed to the orientation of water molecules at the membrane-water interface due to the polar phosphate head groups [193]. Proton transport involves jumps between hydronium and water molecules connected through hydrogen-bonds [194]. Alignment of water molecules at the interface, due to the membrane dipole potential [195], reduces entropy, thereby decreasing the energy barrier for proton transport. Ion transport, however, proceeds through hydrodynamic diffusion. The role of the phosphate head groups and the aligned water molecules at the lipid interface in ion transport across gA is still not clear.

In this article, we attempt to explain the origin of the effects of the lipid environment

**Figure 6.1:** Molecular structure of phospholipids (POPC) (left) and ceramides (NODS) (right) lipid molecules. Hydrogen atoms are not shown for clarity.

on ion conductance of gA channel. We test two different lipid molecules, comparing phospholipids (POPC) and ceramides (NODS) (Fig. 6.1). This is accomplished using MD simulations and free energy calculations. Our results show that ion conductance is indeed enhanced by ceramide lipids consistent with the experimental observations. The reduction in conductance for phospholipids comes from two sources, the phosphate head groups and the orientation of water molecules in the interface layer.

## 6.2 Method

### 6.2.1 Model System

The two lipid molecules considered are 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidyl choline (POPC) and N-oleoyl-D-erythrosphingosine or ceramide C18 (NODS). The structure of both molecules are generated using the Avogadro software [196]. For POPC, we use the standard topology and parameters available in CHARMM36 [149]. The topology for NODS is generated from the ParamChem server [197, 198] and is derived from sphingomyelin parameters available in CHARMM36 [199]. One dihedral parameter with a penalty higher than 10 is optimised using the force-field toolkit (FFTK) [200] available in VMD [147]. The target data for optimisation are generated using Gaussian 09 [201].

The 1NJO structure [37] is used for the gA dimer embedded in a lipid bilayer. The gA dimer isplaced in a hexagonal cell with 20 lipid molecules placed around the peptide per layer. We solvate the gA–membrane complex with TIP3P water molecules and neutralise the system with 0.15 M of KCl. The system is equilibrated in two stages to ensure the stability of the peptide-membrane complex. First, the gA atoms are

fixed, and the cell is allowed to fluctuate isotropically until the correct lipid and water densities are obtained. The cell in the $x$-$y$ plane is then fixed, and only fluctuations in the $z$-direction are allowed. In the second stage, the gA atoms are gradually relaxed over 3 ns of equilibration. The system is further equilibrated without any restraints for 10 ns. We have performed the production runs without any restraints on the gA atoms because the flexibility of the protein affects ion permeation as demonstrated in a previous work [185]. To keep the protein stable inside the bilayer, we apply harmonic restraints on the orientation and position of the centre of mass of the protein. The restraint on the centre of mass does not affect the internal degrees of freedom of the protein.

### 6.2.2 PMF Calculations

We calculate the potential of mean force (PMF) of potassium ion along the gA channel using the equilibrium umbrella sampling simulations [88] with the weighted histogram analysis method [89]. We use the same protocols as in an earlier work [43] for configuring the calculations, which are briefly described here. The reaction coordinate for the ion is its distance from the gA centre along the channel axis. We perform the PMF calculations in the range [0, 20] Å, assuming symmetric behaviour between the two monomers. For umbrella sampling calculations, we use a total of 41 windows at 0.5 Å intervals. The umbrella windows are generated via steered MD simulations, starting from the equilibrated system with the $K^+$ ion at the centre of gA. We apply a harmonic potential with a force constant of 10 kcal/mol, which is reduced to 7 kcal/mol outside the pore ($z > 15$ Å) to improve sampling between windows. In the bulk region, a harmonic potential of 1 kcal/mol is applied in the radial direction to prevent the ion drifting away from the central axis. Each window is run for 3 ns with 1 ns for equilibration, giving a total simulation time of 123 ns.

### 6.2.3 MD Simulation

All MD simulations in this work are performed with the NAMD package (version 2.10) [148] with the CHARMM force field [149]. We employ the NPT ensemble and kept the simulation temperature constant at 300 K using the Langevin thermostat with a damping factor of 1 $ps^{-1}$. The pressure is kept constant at 1 atm using the Langevin piston method with a damping factor of 50 $ps^{-1}$ [79]. Periodic boundary conditions are used, and electrostatic interactions are calculated using the particle-mesh Ewald (PME) method [78] without truncation. Non-bonded interactions are truncated at 12 Å and replaced with a smooth switching function starting from 10 Å. In all simulations, a time step of 2 fs is employed for the integrator.

## 6.3 Results and Discussion

### 6.3.1 Membrane Bilayer

Before investigating ion conductance in gA we studied the behaviour of the lipid bilayers with POPC and NODS. The molecular structures of POPC and NODS are shown

in Fig. 6.1. For the purpose of investigating pure bilayers, we used a simulation system consisting of 64 lipid molecules per leaflet. To compare the thickness, we calculated the mass density of the head groups for each lipid type and plotted the densities in Fig. 6.1. For POPC, we chose the phosphorus atom as the reference and the oxygen on the hydroxyl group in NODS. From 4 ns of MD simulations, we sampled the location of the head groups along the bilayer normal ($z$-axis). From the maximums, it can be seen that the oxygen atoms are located slightly further away from the bilayer centre compared to the phosphorus atoms (21.0 and 20.3 Å, respectively). Thus, the average bilayer thickness is about 41 Å and 42 Å for POPC and NODS, respectively, in agreement with experiments and simulations [203, 204]. It has been shown that the asymmetric hydrocarbon chains of ceramide results in a thicker bilayer formation [205]. We observed a larger difference in the horizontal densities (i.e., area per lipid) of POPC and NODS lipids. For POPC, we found an area of ∼60 Å$^2$ per lipid in good agreement with the experimental values [206]. For NODS bilayer, the area per lipid we obtained is ∼46 Å$^2$, again very similar to the experimental value [207]. These give confidence to the force field parameters used in the simulations. The large difference between POPC



**Figure 6.1:** Analysis of the pure bilayer systems: (a) mass density of head groups; (b) water number density; and (c) average orientation of water molecules with respect to the bilayer plane (the angle between the dipole vector and the $x$-$y$ plane). Black and red lines represent systems with POPC and NODS lipids, respectively. Densities are calculated using the density profile extension in VMD [202].

and NODS densities is simply due to the larger head group in POPC. We also plotted the water density along the bilayer normal in Fig. 6.1. The water density profiles show a thicker layer of water molecules at the interface for POPC than NODS. This behaviour is a direct result of the tighter packing of NODS molecules in the bilayer formation.
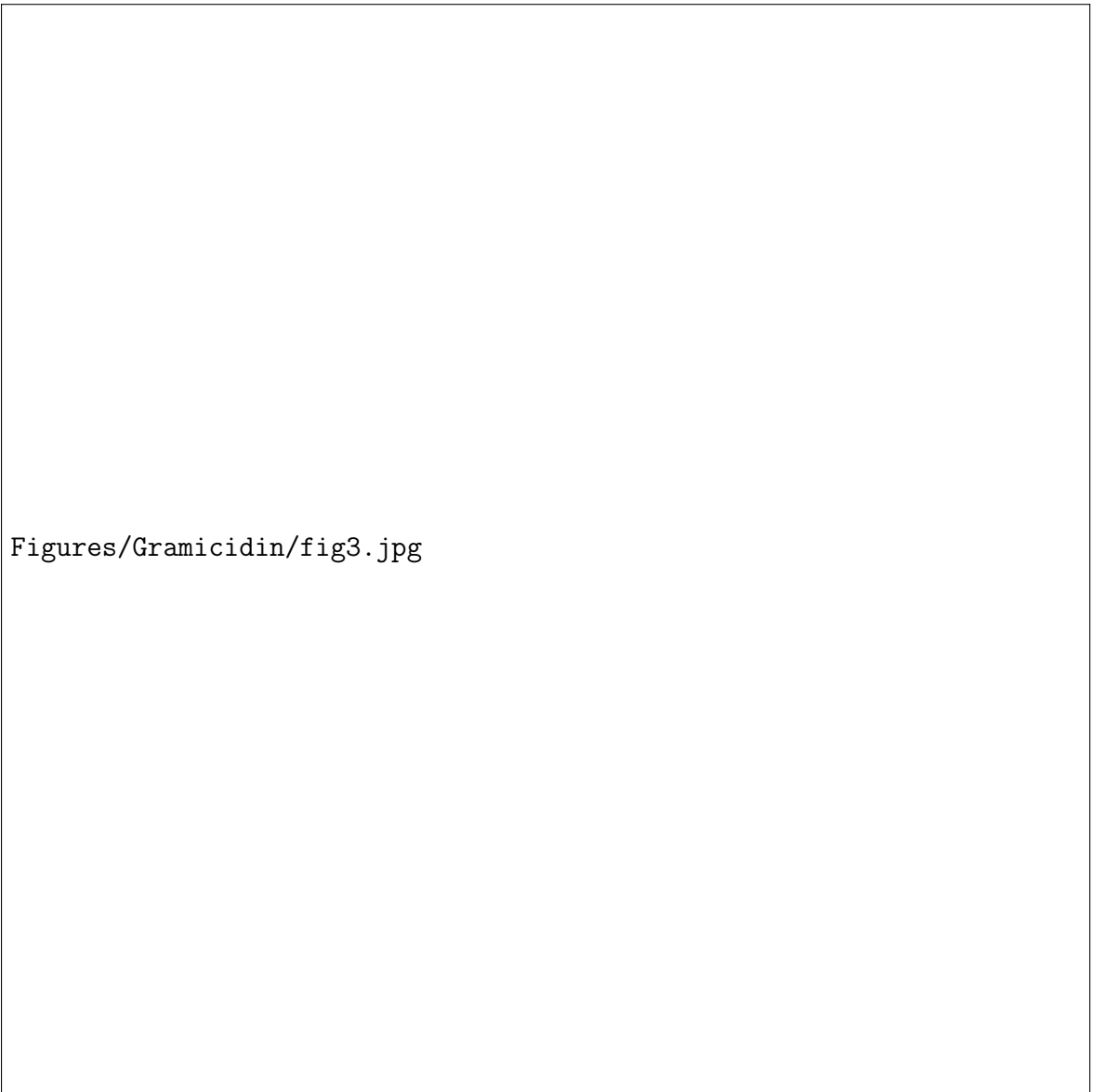
To further analyse the previous observation, we plotted the average orientation of water molecules as a function of the bilayer normal. The orientation is defined as the angle between the dipole moment of water molecules and the bilayer plane. The angles are averaged over a 2 ns trajectory collected in 0.5 Å bins. The peak of the profiles shown in Fig. 6.1 is roughly 20° and 10° for POPC and NODS, respectively. Qin et al. obtained a peak close to 30° for a DiPhPC bilayer [193]. The difference between our results and Ref. [193] for the PC head groups can be attributed to polarisation effects that is included semi-empirically in the multistate empirical valence bond model [208]. However, even without polarisation, our results clearly show that the PC head groups orient water molecules over a wider range than for ceramide. Thus the larger head groups in POPC not only increases the area per lipid but also results in a greater dipole potential. It is interesting to compare the water orientation with GMO bilayers [193], where water molecules exhibit close to a random orientation on average at the interface. Experimental measurements of caesium conductance through gA embedded in GMO bilayers show very similar behaviour to NODS at concentrations greater than 250 mM [46]. The orientation of water molecules at the interface may not be the only factor contributing to the enhanced conductance.

## 6.3.2   gA Embedded in Membrane Bilayer

After embedding the gA in bilayers, we have simulated the equilibrated system for 10 ns with only water molecules inside. The resulting RMSD, calculated by excluding the ethanolamine groups, is shown in Fig. 6.2. The RMSD fluctuates around 0.5–0.8 Å, and there are some slight changes in RMSD over the trajectory, but there are no perceptible differences between the RMSDs of gA embedded in POPC and NODS. When we include ethanolamine in the calculations, the RMSD time series fluctuates closer to 1 Å. To show this effect quantitatively, we plot the average RMSD per residue, including the ethanolamine, which is also shown in Fig. 6.2. For POPC, all the residues have RMSD values less than 0.75 Å. In NODS, however, the residues near the entrance of the channel fluctuate more than POPC because the gA side chains have much weaker interactions with the NODS head groups compared to those of POPC. In particular, the ethanolamine group swings instantaneously and this is also observed in MD simulations for gA with other bilayers [209]. The swing motion perturbs the neighbouring residues thereby explaining the slight turns in the RMSD time series even without ethanolamine. The swing motion of ethanolamine is due to loss of hydrogen-bonds [210], and we circumvent this problem by sampling the systems longer. Despite these minor differences, gA is essentially stable inside the hydrophobic environment of both lipids. With gA embedded inside the membrane, we find that the bilayer thickness of the lipid molecules in the first shell decreases appreciably as observed by Kim et al. previously [209]. The bilayer thickness is recovered in the second lipid shell, and hence only lipids in the first shell are affected. Because we have embedded gA in 20 lipid

molecules per leaflet configured in a hexagonal cell, we report only the first shell results here. The bilayer thickness, as defined in the previous subsection, decreases from 41 to 31 Å for POPC, and from 42 to 37 Å for NODS. The larger decrease in the thickness of a POPC bilayer is again related to the much stronger coupling of the POPC head groups with the gA side chains compared to those of NODS.

In the previous trajectory of the gA system with water molecules, we have replaced a water molecule near the centre of gA with a $K^+$ ion. We have applied a harmonic restraint to keep the $K^+$ ion at the centre and sampled the systems for a further 4 ns (the ethanolamine group is stable at this point). To gain further insights into the potential contributors to the ion conductance, we have analysed the behaviour of water molecules inside the channel. When the $K^+$ ion is positioned at the centre of gA, there



**Figure 6.2:** RMSD of gA embedded in POPC and NODS lipids: (a) total and (b) average per residue. Error bars are not shown for RMSD per residue for clarity.

**Figure 6.3:** Water molecules inside the channel with $K^+$ located at the centre of gA. (a) Structure showing $K^+$ positioned in the centre of gA with aligned water molecules (lipids not shown); and (b) average orientation of water molecules. The orientation is defined as the angle between the dipole vector and the x-y-plane. Labels for water molecules are shown on the diagram.

are three water molecules aligned on either side of the ion inside the channel. Similar to the results in Fig. 6.1, we have calculated the average orientation of the six water molecules in gA with respect to the bilayer plane (Fig. 6.3). As expected, the alignment of the water dipoles with the channel axis, and hence the strength of the interaction between the ion and water molecules, decreases as a function of the distance from the ion. We note that the water dipoles are not aligned closer to 90° with the plane because they make hydrogen bonds with the carbonyl groups of gA. Taking into account the statistical fluctuations, the interaction between the $K^+$ ion and water molecules inside the channel are very similar for the POPC and NODS bilayers.

### 6.3.3   Potential of Mean Force of K$^+$ in gA

We have performed umbrella sampling simulations to determine the PMFs of a K$^+$ ion along the central axis of gA embedded in the POPC and NODS bilayers. Details of the umbrella sampling simulations are described in the Methods section. Each umbrella window is simulated for 3 ns. Using the stability of the ethanolamine group and the convergence of the PMFs as criteria, the first 1 ns of data are discarded for equilibration, and the PMFs are constructed from the final 2 ns of production data. The PMFs obtained with the POPC and NODS bilayers are shown in Fig. 6.4. The PMFs exhibit similar behaviour from the bulk region up to ∼7.5 Å inside the channel but start diverging from there to the centre of gA. There is also some difference in the binding free energies at the binding pocket (∼11.3 Å). For the POPC bilayer, we obtain a well depth of 2.5 kcal/mol at the binding site with respect to the bulk. In previous PMF calculations for a K$^+$ ion, well depths in the range of 2–3 kcal/mol were obtained for PC bilayers [40, 41, 43, 181]. As the accuracy of the PMF calculations is about 1 kcal/mol, the present result for the well depth is consistent with those earlier results. In the case of NODS, the well depth at the binding site is 3.7 kcal/mol, which is 1.2 kcal/mol deeper compared to the POPC bilayer. Thus we predict an eight-fold difference between the binding constants of K$^+$ ions for gA in POPC vs NODS bilayers, which can be easily distinguished in experiments.

The two PMFs start diverging around 7.5 Å, and the difference becomes quite substantial near the centre of gA. The energy barrier measured from the binding site to the peak in the PMF is 10.9 kcal/mol for POPC and 8.2 kcal/mol for NODS. The barrier



**Figure 6.4:** Potential of mean force (PMF) profiles of a K$^+$ ion along the gA channel axis.

value for POPC is again in agreement with the previous PMF calculations in gA [40, 41, 43, 181]. The lower energy barrier observed in the NODS bilayer compared to that in the POPC bilayer is consistent with the experimental data which shows a four-fold increase in cation conductance of gA embedded in NODS compared to POPC [46]. We stress that polarisation plays a significant role in ion permeation across the gA channel, hence MD simulations with a non-polarisable force field provide only qualitative results

It is of interest to find out how the change in the lipid bilayer affects the ion PMFs. To determine the cause for this difference in the PMFs, we calculate the average potential energy acting on the $K^+$ ion from four different components of the system individually. The potential energy is calculated from the umbrella sampling trajectory data at 0.5 Å intervals. The four components we chose to calculate are the protein, the lipid molecules, and the water molecules inside the channel and at the lipid interface. As shown in Figs. 6.1 and 6.3, these water molecules exhibit alignment with the channel axis or the bilayer normal, hence will contribute to the potential energy of the ion. The channel water molecules are defined as the water molecules inside the region $[-10, 10]$ Å and interfacial water is defined by the range of molecules oriented by the lipid head groups as shown in Fig. 6.1. The $K^+$ potential energies due to these four components are plotted as a function of gA channel axis in Fig. 6.5. We note that the potential energy does not include entropy effects, hence we focus on qualitative rather than quantitative results.

We first consider the potential energy of the $K^+$ ion due to gA. As expected, this potential energy approaches zero as the ion moves further into the bulk region. As the ion moves closer to the binding site the potential energy decreases to a minimum and then slowly increases as it reaches the centre. This trend is consistent with Allen et al. [42], but the values we calculate are of a different scale because we consider the potential energy rather than integrating the mean force. For both POPC and NODS, the potential energy profiles are very similar, as evident from Fig. 6.5. This result indicates that the lipid environment does not directly affect the interaction between gA and the ion inside the channel. Kim et al. investigated the behaviour of gA in different phospholipid environments and observed variations in the dynamics of gA [209]. From our results, we conjecture that although the lipid environment can perturb gA dynamics due to hydrophobic mismatch, these perturbations may not be significant enough to change the nature of the interaction between gA and $K^+$.

Considering next the potential energy of the $K^+$ ion due to the lipid molecules, we see a noticeable difference between the POPC and NODS bilayers Fig. 6.5. The potential energy due to NODS remains close to zero throughout, with a slight change from negative near the binding site to positive as the ion moves towards the bulk. This is mainly due to the interaction between $K^+$ and the NODS head groups, which have small dipole moments and a short-range. For POPC, however, the potential energy is quite substantial, which is the result of the strong dipole potential of the PC head groups. The potential energy remains negative within gA and approaches zero as the ion moves towards the bulk as expected. On average, the difference in energy between the two lipid bilayers is $\sim$20 kcal/mol within gA, and it favours POPC relative to NODS. This is clearly in the opposite direction to the PMF profiles, where the NODS PMF is lower than the POPC PMF (Fig. 6.4). Thus other contributions to the $K^+$

potential energy are needed to explain the difference in the PMFs.

The other components that could influence the potential energy of the K$^+$ ion in gA are water molecules. Water molecules can be separated into three groups based on their orientation in the system. Firstly, water molecules with random orientation are located in the bulk region, where interactions with protein and membrane are negligible. Next, as discussed earlier, a layer of water molecules tend to orient away from the bilayer due to membrane dipole potential. Finally, the single-file water molecules inside the channel align their dipole moments with the electric field of the ion, which is modulated by the formation of hydrogen bonds with the carbonyl oxygens of gA. Here we are interested only in water molecules with a specific orientation, and hence consider only the potential energy due to the channel and interfacial water molecules.

We start with the interaction of the channel waters with the K$^+$ ion. We have stated earlier that the behaviour of the dipole moments of water molecules in gA, in the presence of a K$^+$ ion at the centre, is approximately the same in the NODS and POPC
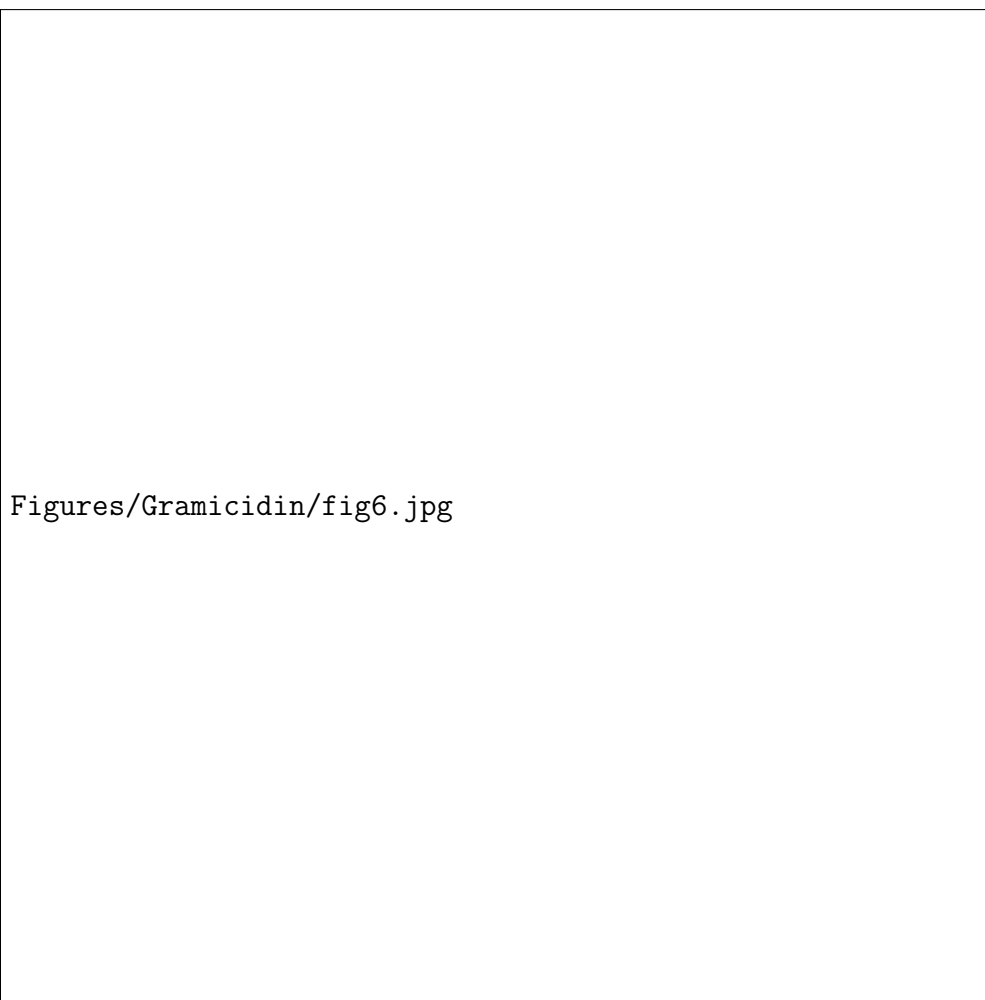


**Figure 6.5:** Potential energy acting on K$^+$ as a function of ion positions in the gA channel. Interactions calculated include protein, lipid molecules, channel water and interfacial water. The red and black line represents POPC and NODS lipid molecules respectively.

bilayers (Fig. 6.2). This behaviour is seen to be maintained for other positions of the $K^+$ ion in gA—the potential energies of the $K^+$ ion due to the channel waters are seen to overlap well for the two lipid molecules (Fig. 6.5). The small differences observed when the ion is outside gA is likely arising from the flipping of the dipole moments of water molecules. This verifies that the water molecules inside the channel do not directly contribute to the attenuation and enhancement of ion conductance in gA embedded in POPC and NODS bilayers, respectively. It is interesting to note the trend of the interaction inside the channel. There are three different stages in the range [0, 20] Å where the interaction changes. Starting from 20 Å, the potential energy is on average close to zero as expected. As the ion is moved closer to the binding site, the potential decreases to $-30$ kcal/mol. This sudden change in potential is the result of one water molecule being pushed out of the channel on the other side to accommodate the $K^+$ ion. From the binding site to 7.5 Å, the potential energy remains steady. However, as the ion is pushed further to the centre, there is a second dip in the potential energy, bringing it down to about $-50$ kcal/mol. We have observed that this results from the reorientation of water molecules inside the channel. The potential energy comes to a minimum at the centre with $-60$ kcal/mol, where the dipole moments of three water molecules are aligned in the direction of the ion's electric field on either side of the channel in a symmetric configuration. The potential energy profile demonstrates the stages that take place as the ion moves through gA, which cannot be traced from the PMF alone.

Lastly, we consider the interaction of $K^+$ with the interfacial water molecules, which exhibits a significant difference in the potential energy of $K^+$ with the NODS and POPC bilayers. For NODS, the potential energy is zero at the centre and decreases to about $-10$ kcal/mol at the binding site. For POPC on the other hand, the potential energy is around $+25$ kcal/mol in gA and only starts to decrease around the binding site. As the ion is moved further away from gA, the potential energy of both systems converges to the same level. We note that the potential due to the interfacial waters will be strongly screened by bulk water, and the free energy of $K^+$ will vanish in bulk as seen in the $K^+$ PMF (Fig. 6.4). To understand the difference between the POPC and NODS results, we refer to Fig. 6.1 which shows that there is a thicker layer of water molecules in POPC, better oriented by the PC head groups with larger dipole moments. This results in a much stronger dipole potential due to the interfacial waters when gA embedded in a POPC bilayer compared to that of NODS. Comparing the lipid contribution to the potential energy of $K^+$ with that of the interfacial waters (Fig. 6.5), it is seen that the latter contribution more than compensates the former in POPC, resulting in a positive potential energy in gA. In contrast, the sum of the two contributions remain near zero but slightly negative in NODS. Thus our results indicate that the larger energy barrier in the POPC–PMF relative to the NODS–PMF is most likely due to the better orientation of the interfacial water molecules by the stronger dipole potential of the POPC head groups.

## 6.4 Conclusion

Our objective in this work is to understand the cause of the enhancement of ion conductance in gA embedded in a NODS bilayer compared to that of POPC. For pure bilayer systems, we have observed different lipid–water interactions with the result that water molecules tend to be more structured in phospholipids than ceramide. This behaviour was also demonstrated in comparisons of phospholipids with monoglycerides [193]. Our simulations of the gA system show that the peptide behaves in a similar fashion when it is embedded in both lipid environments. However, the PMF of a K$^+$ ion in gA embedded in POPC results in a distinctly larger energy barrier compared to that of NODS, which is consistent with the observed enhancement of the conductance in NODS [46]. To understand the origin of this enhancement, we have analysed the different contributions to the potential energy of the K$^+$ ion from gA, lipid molecules and water molecules within gA and at the lipid interface. We find that the interaction of gA and water molecules inside gA with K$^+$ are virtually identical for both lipid molecules. But there are substantial differences between the POPC and NODS bilayers when we consider the interaction of lipid molecules and interfacial waters with K$^+$. In NODS, the weak dipole moments of the head groups result in loosely structured interfacial water, and the contribution from either group to the ion's potential energy is very small. In POPC, the strong dipole moments of the head groups give rise to a thicker layer of interfacial waters with better orientation. Thus, both groups make a substantial contribution to the ion's potential energy, but the positive contribution from the interfacial waters overcomes the negative one from lipids. This difference is consistent with the higher energy barrier found in the K$^+$ PMF in POPC compared to the K$^+$ PMF in NODS.

# Chapter 7

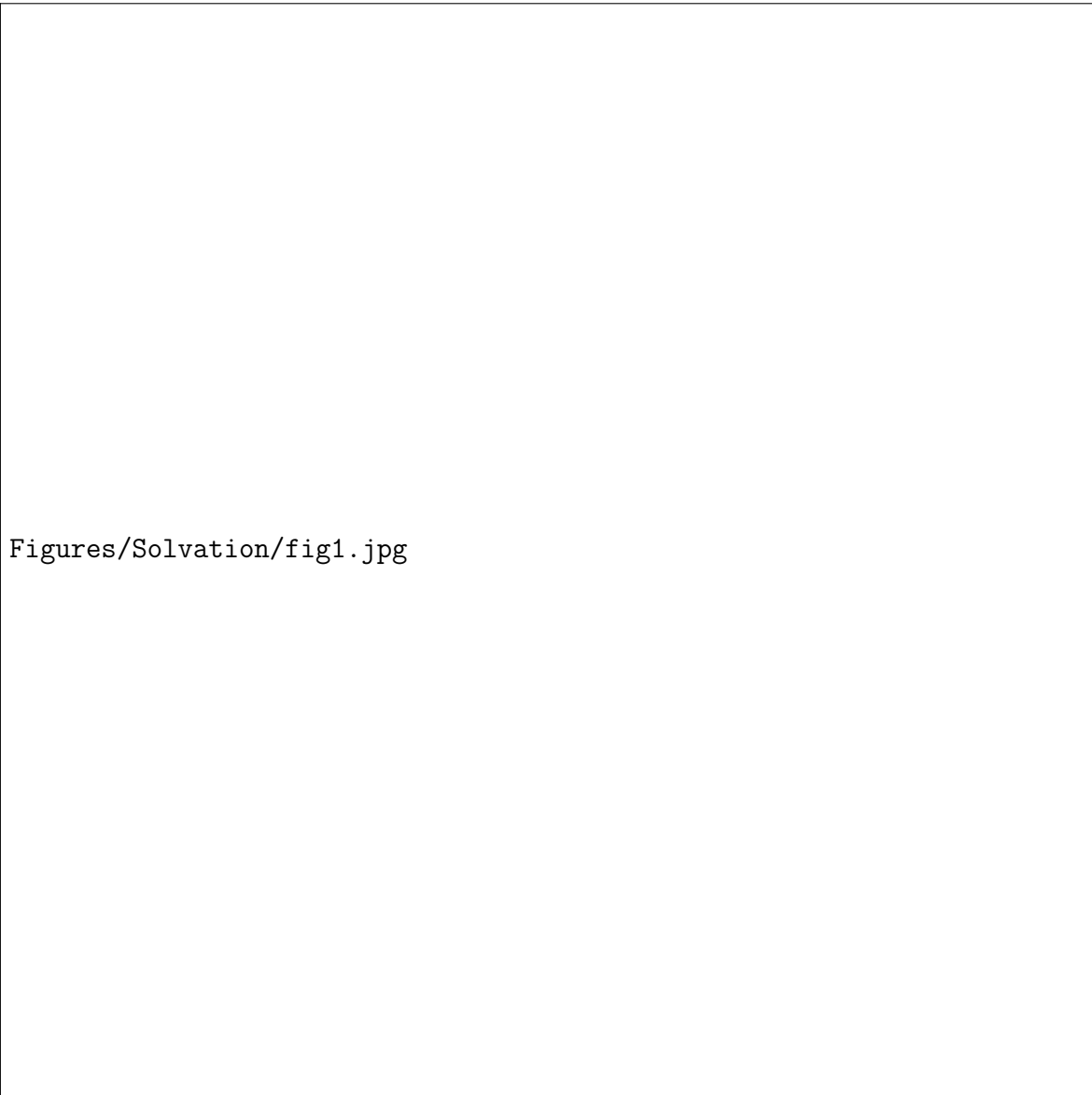# Ion Solvation Energy with Spherical Boundary Conditions

ABSTRACT

Solvation free energies of ions are difficult to determine from molecular dynamics simulations due to the long-range Coulomb interactions. Various approximations and corrections are introduced to enable their calculation in small systems, which however raises issues of robustness. We show that solvation free energies of ions can be calculated using the spherical boundary conditions without introducing any corrections at the boundary via a buffer zone. The results are shown to converge for a droplet size of 21 Å and are independent of the parameters used for confining water or restraining the ion. The proposed method thus resolves the robustness issues in solvation free energy calculations of ions and can be used with confidence to determine force field parameters from such calculations. We apply the method to calculate the solvation free energies of the side chain analogues of charged amino acids. Tests using periodic boundary conditions show that similar results are also obtained in that case.

# 7.1    Introduction

Solvation of ions in water and other solvents is a fundamental phenomenon in physical chemistry with applications in many other fields such as biochemistry, molecular biology and physiology [211, 212]. Properties of solvated ions can be described at several levels depending on the desired level of accuracy and detail. Two of the most common approaches are the Poisson-Boltzmann equation with implicit solvent and molecular dynamics (MD) simulations with explicit solvent molecules. Thanks to the improvements in computer hardware and software, MD simulations have now become the primary tool for atomic-level description of biomolecules and their interactions. The accuracy of simulation results depends on the quality of the force fields used in MD simulations. In this regard, accurate calculation of the solvation free energies of solutes is of primary importance because the balance between the solute–solute and solute–solvent interactions determines whether two solute molecules bind or not. For this reason, solvation free energies are routinely used for determining the parameters in force fields [10, 68, 213, 214]. The long-range Coulomb potential has been the main obstacle in the calculation of the solvation free energies of ions from MD simulations. It was shown early on that the use of a cut-off radius in the calculation of the Coulomb interactions gave spurious results and had to be avoided [215]. Because the size of simulation systems was limited by the available computer power, the only alternatives were using either a simulation box with periodic boundary conditions (PBC) and Ewald sum for long-range interactions, [211, 216] or a droplet with spherical boundary conditions (SBC) and a buffer zone with implicit solvent beyond [48, 57]. In the former method, a number of corrections arising from periodicity and finite-size effects have to be applied to the raw solvation free energy obtained from MD simulations, which raises doubts about the robustness of the results [58]. Using a droplet with SBC avoids these issues, but the introduction of a buffer zone, where external forces are applied to water molecules to counter the effect of the vacuum outside, brings back the robustness question. It is difficult to quantify how successful the buffer zone is in mimicking the actual behaviour of water molecules embedded in a much larger system.

The issue with the robustness of the computations in the SBC method could be resolved if a large enough system is used so that one can dispense with the buffer zone and just embed the droplet in a cavity of implicit water (Fig. 7.1). The currently available computer power certainly allows consideration of much larger simulation systems compared to two decades ago. All one needs to do is to show that the computed solvation energy of an ion converges to a unique value with increasing droplet size independent of the parameters used for the confinement of water molecules or restraining of the ion. Here we demonstrate the feasibility of this simpler approach by computing the solvation free energies of $Na^+$ and $Cl^-$ ions using a droplet embedded in a cavity whose size is varied. We show that the correction terms arising from the confinement of the droplet and its interface with vacuum become negligibly small for droplet radii greater than 20 Å. We then generalise the cavity energy to a molecular ion with arbitrary partial charges and use our method to compute the solvation free energies of the side chain analogues of the charged amino acids. Tests with PBC show that convergent solvation free energies are also obtained in that case when simulation boxes with volumes similar to droplets are used. Finally, we discuss the implications of the results for the current force fields.

**Figure 7.1:** (A) The full system of an ion embedded in a bath of explicit water. A spherical boundary with radius $R$ is placed around the ion to split the system into two parts. (B) A water droplet with explicit water is cut out from the full system and placed in a vacuum. (C) A hollow spherical cavity (vacuum) embedded in an implicit water with the ion at the centre. The full ion-water interaction in (A) is obtained by adding the ion-water interactions in (B) and (C).

## 7.2 Method

### 7.2.1 Solvation Free Energy with SBC

In the SBC method, the simulation system is divided into two parts using a spherical boundary around the ion, and the water outside the boundary is represented implicitly (Fig. 7.1). In previous applications of the SBC method [48, 57], a buffer zone is used to rectify the effect of vacuum on water molecules near the boundary. This was

necessary for the relatively small spheres ($R \sim 10$ Å) employed in these work because the structural order imposed by the vacuum boundary could affect the hydration shells of the ion. This would not be a problem when a sufficiently large sphere is employed. A second artefact introduced by the vacuum boundary is the monopole ($l = 0$) and higher-order contributions to the solvation energy, which depend on $R$ as $1/R^{l+1}$ and would be negligible for large $R$. We note that the monopole term is cancelled by the interaction of the ion with the cavity, given by the Born energy

$$\Delta G_{\text{cav}} = -\frac{(1 - 1/\epsilon)q^2}{8\pi\epsilon_0 R}, \tag{7.1}$$

where $q$ is the charge of the ion and $\epsilon$ is the dielectric constant of water. The remaining artefacts fall off as $1/R^2$ or faster so with an appropriate choice of $R$, they can be made negligibly small.

Here we investigate the feasibility of this approach for practical applications. The system is decomposed into a water droplet and a cavity part as shown in Fig. 7.1, and the solvation free energy of the ion is calculated from

$$\Delta G_{\text{solv}} = \Delta G_{\text{drop}} + \Delta G_{\text{cav}}. \tag{7.2}$$

Here $\Delta G_{\text{drop}}$ is the solvation free energy of the ion in the droplet, and $\Delta G_{\text{cav}}$ is the self-energy of the ion in the cavity, which is given by the Born energy, Eq. (7.1), for a central ion. The solvation free energy is usually calculated by an alchemical route, i.e., by slowly turning off the interaction between the ion and water molecules in reference to vacuum. To improve convergence, the contribution of the electrostatic and Lennard-Jones (LJ) interactions are calculated separately

$$\Delta G_{\text{drop}} = \Delta G_{\text{drop-el}} + \Delta G_{\text{drop-LJ}}. \tag{7.3}$$

Note that due to the $1/r^6$ dependence of the LJ interactions, their contribution to the cavity term can be neglected. The electrostatic and LJ interactions of the ion with the water molecules inside the droplet can be calculated by either free energy perturbation (FEP) or thermodynamic integration (TI) [152]. In both methods, the ion-water interaction is coupled/decoupled over $n$ simulation windows chosen between $\lambda = 0$ and 1. The sum of the change in free energy or the integral of the free energy gradient over $\lambda$ gives the final free energy for FEP and TI, respectively.

As mentioned above, the monopole contribution to $\Delta G_{\text{drop-el}}$ is cancelled by $\Delta G_{\text{cav}}$, which removes the $1/R$ dependence from the solvation free energy. In contrast, there is no $1/R$ dependence in $\Delta G_{\text{drop-LJ}}$. Because this will be important when we study the convergence of the solvation free energy with $R$, we rewrite the solvation free energy of the ion in the form

$$\Delta G_{\text{solv}} = \Delta G_{\text{el}} + \Delta G_{\text{LJ}}, \tag{7.4}$$

where

$$\Delta G_{\text{el}} = \Delta G_{\text{drop-el}} + \Delta G_{\text{cav}} \tag{7.5}$$

represents the total electrostatic contribution to the free energy, and we have removed the "drop" label from the LJ term to simplify the notation. Representation of the cavity term with the Born energy is valid only for a central ion. For a molecule with distributed partial charges, the self-energy contribution due to the cavity needs to be generalised, which we consider next.

## 7.2.2  Self-energy of Ions and Molecules in a Cavity

We first consider the self-energy of a point charge $q$ placed at $d\hat{\mathbf{z}}$ in a spherical cavity of radius $R$ (see Fig. 7.1). The dielectric constant outside the sphere is $\epsilon = 80$ (implicit water), and inside it is 1 (vacuum). The potentials inside and outside the cavity can be determined by expanding the potentials in terms of Legendre polynomials and applying the appropriate boundary conditions at $r = R$ [217]. The potential inside the cavity due to the induced charges at the boundary (reaction field) is given by

$$\varphi_{\mathrm{RF}} = -\frac{q}{4\pi\epsilon_0} \sum_l \frac{(l+1)(\epsilon-1)}{(l+1)\epsilon + l} \frac{d^l r^l}{R^{2l+1}} P_l(\theta). \tag{7.6}$$

Because $\epsilon \gg 1$, we can neglect the second term in the denominator of Eq. (7.6) without much error. Then the factors of $l+1$ from the numerator and denominator cancel, yielding

$$\varphi_{\mathrm{RF}} \simeq -\frac{(qR/d)\,(\epsilon-1)}{4\pi\epsilon_0}\,\frac{1}{\epsilon} \sum_l \frac{r^l}{(R^2/d)^l} P_l(\theta). \tag{7.7}$$

Here, the sum over $l$ corresponds to the expansion of $1/|\mathbf{r}-(R^2/d)\hat{\mathbf{z}}|$. Thus the reaction potential in Eq. (7.7) can be represented by an image charge of magnitude $q' = -(1 - 1/\epsilon)qR/d$ placed at $R^2/d$ on the $z$-axis. Generalising this result to a charge $q$ at an arbitrary position $\mathbf{r}'$ in the cavity, we obtain

$$\varphi_{\mathrm{RF}} = -\frac{(1-1/\epsilon)}{4\pi\epsilon_0}\,\frac{qR/r'}{|\mathbf{r}-(R/r')^2\mathbf{r}'|}. \tag{7.8}$$

The self-energy of the charge $q$ in the cavity follows from

$$
\begin{aligned}
\Delta G_{\mathrm{cav}} &= \frac{1}{2}q\varphi_{\mathrm{RF}}(\mathbf{r}=\mathbf{r}') \\
&= -\frac{(1-1/\epsilon)}{8\pi\epsilon_0}\,\frac{q^2 R}{R^2 - r'^2},
\end{aligned}
\tag{7.9}
$$



**Figure 7.1:** A spherical cavity with radius $R$ is embedded in implicit water medium. A charge $q$ is placed on the $z$-axis at a distance $d$ from the origin.

which reduces to the Born energy for $r' = 0$. Thus the approximation made to obtain Eq. (7.7) does not affect the leading monopole term; the $\simeq 1\%$ error affects only the higher-order terms.

We next use the superposition principle to generalise the above results to an ionic molecule in a cavity, which has $N$ partial charges $q_i$ at positions $\mathbf{r}_i$. From Eq. (7.8), the reaction potential in the cavity is given by

$$\varphi_{\mathrm{RF}}(\mathbf{r}) = -\frac{(1 - 1/\epsilon)}{4\pi\epsilon_0} \sum_i \frac{q_i R/r_i}{|\mathbf{r} - (R/r_i)^2 \mathbf{r}_i|}, \tag{7.10}$$

and the self-energy of the molecule follows from

$$
\begin{aligned}
\Delta G_{\mathrm{cav}} &= \frac{1}{2} \sum_j q_j \varphi_{\mathrm{RF}}(\mathbf{r} = \mathbf{r}_j) \\
&= -\frac{(1 - 1/\epsilon)}{8\pi\epsilon_0} \sum_{ij} \frac{q_i q_j R}{|\, R^2 \hat{\mathbf{r}}_i - r_i r_j \hat{\mathbf{r}}_j|}.
\end{aligned}
\tag{7.11}
$$

Separating the diagonal and cross-terms, the self-energy can be written in the form

$$\Delta G_{\mathrm{cav}} = -\frac{(1 - 1/\epsilon)}{8\pi\epsilon_0} \left[ \sum_i \frac{q_i^2 R}{R^2 - r_i^2} + \sum_{i>j} \frac{2 q_i q_j R}{|R^2 \hat{\mathbf{r}}_i - r_i r_j \hat{\mathbf{r}}_j|} \right]. \tag{7.12}$$

The distance in the denominator of the cross term can be calculated using

$$|R^2 \hat{\mathbf{r}}_i - r_i r_j \hat{\mathbf{r}}_j| = \left[ R^4 + (r_i r_j)^2 - 2 R^2 \mathbf{r}_i \cdot \mathbf{r}_j \right]^{1/2}. \tag{7.13}$$

Eq. (7.12) generalises the self-energy of an ion in a cavity to a molecule with arbitrary partial charges on its atoms. It will be used in calculating the cavity contribution to the solvation free energy of a molecule.

## 7.2.3 Computational Details for SBC

We first consider the solvation free energies of Na$^+$ and Cl$^-$ to establish the bona fides of the proposed method. The method is then used to investigate the solvation free energies of the side chain analogues of the charged amino acids. Side chain analogues are built by removing the backbone of the amino acids and terminating the dangling carbon atom with a hydrogen atom. The partial charge of the dangling atom is adjusted to preserve the total charge of the side chain analogue. The topologies are available as part of the CHARMM36 CGenFF [149] distribution. The solute molecule is solvated in a sphere of a given radius $R$ with TIP3P [162, 163] molecules using the PACKMOL program [218]. All MD simulations are performed in OpenMM version 7.2.2 [219] with a time step of 2 fs, and the temperature is kept constant at 300 K with a Langevin damping of 1 ps$^{-1}$. Calculations are run on V100 Tesla graphical processing units (GPU) at the high-performance computing facility *Artemis* with mixed precision (a test with double precision gave virtually the same result).

In order to maintain the density of the water droplet, a half-harmonic potential is applied to the oxygen atom of water molecules

$$U_{\text{sphere}} = \begin{cases} \frac{1}{2}k_s(r-r_0)^2, & \text{if } r > r_0, \\ 0, & \text{otherwise,} \end{cases} \tag{7.14}$$

where $k_s$ is the spring constant and the equilibrium $r_0$ is defined as the droplet radius $R$ minus the average fluctuation from the restraint at temperature $T$, i.e., $r_0 = R - \sqrt{k_{\text{B}}T/k_s}$. The offset ensures that the monopole contribution to the solvation free energy from the droplet at the boundary is cancelled by the cavity term. This is similar to the potential used in the MD program Q [220] but without an attractive Morse potential. A spring constant of 10 kcal/mol/Å$^2$ is employed in most of our simulations. In addition to the confinement potential, a harmonic restraint of 10 kcal/mol/Å$^2$ is applied to the centre of charge of the solute to keep it close to the origin. The choice of these two parameters does not affect the final solvation free energy significantly as will be shown in the results and discussion section. In previous calculations using the SBC method, angular restraints were applied on water molecules in the boundary region to prevent their ordering at the vacuum interface so that their behaviour was consistent with those in a much larger droplet [48, 57]. Here we do not use any angular restraints. As will be shown in the results section, such boundary effects become negligible when the system is chosen large enough.

FEP calculations are performed to obtain the free energy terms $\Delta G_{\text{drop-el}}$ and $\Delta G_{\text{LJ}}$. For both the droplet electrostatic and LJ terms, we use 21 equally spaced windows with a sampling time of 0.1 ns/1.0 ns for equilibration/production at each $\lambda$ value. To improve convergence in the LJ calculations, we use a soft-core potential with a radius-shifting coefficient of 5.0 Å. The free energy is calculated with the multistate Bennett-Acceptance-Ratio (MBAR) method [86] using the python implementation *pymbar* available at https://github.com/choderalab/pymbar. The cavity correction $\Delta G_{\text{cav}}$ is calculated using Eq. (7.9) for an ion and Eq. (7.12) for a molecule, and averaged over a 2 ns MD simulation for each system. See Appendix A for the python implementation.

### 7.2.4 Solvation Free Energy with PBC

The intrinsic solvation free energy with PBC is defined as

$$\Delta G_{\text{solv}}^{\text{intr}} = \Delta G_{\text{el}} + \Delta G_{\text{LJ}}, \tag{7.15}$$

where the two terms represent the electrostatic and LJ contributions, respectively. In order to compare the experimental results, a surface term needs to be added to the intrinsic solvation free energy due to the absence of a vacuum interface in PBC. Thus the "real" solvation free energy is given by [58]

$$\Delta G_{\text{solv}}^{\text{real}} = \Delta G_{\text{el}} + \Delta G_{\text{LJ}} + zF\phi_{\text{lv}}, \tag{7.16}$$

where $z$ is the valence of the ion, $F$ is the Faraday constant, and $\phi_{\text{lv}}$ is the electrostatic potential that arises from a liquid-vacuum interface. For the TIP3P water model used in this work, $\phi_{\text{lv}} = -0.52$ V [58], so $-12.0z$ kcal/mol needs to be added to the intrinsic

value to obtain the real solvation free energy. The electrostatic contribution is split into two terms in the spirit of Eq. (7.5),

$$\Delta G_{\text{el}} = \Delta G_{\text{box-el}} + \Delta G_{\text{SE}}, \tag{7.17}$$

where $\Delta G_{\text{box-el}}$ gives the electrostatic contribution of the water molecules inside the PBC box and $\Delta G_{\text{SE}}$ describes the self-energy of the ion in the box, similar to the $\Delta G_{\text{cav}}$ term in Eq. (7.5). For a cubic box with size $L$, the self-energy is given by [55]

$$\Delta G_{\text{SE}} = -\frac{2.837297}{4\pi\epsilon_0}\frac{q^2}{2L}. \tag{7.18}$$

Like the Born energy (7.1), the self-energy drops as $1/L$ and is not negligible even for relatively large box sizes.

Simulations of ion solvation with PBC are done using the NAMD (version 2.12) [148] program that implements a particle mesh Ewald (PME) summation with a tinfoil boundary. A cut-off of 12 Å for the non-bonded interactions and a smooth switching is applied at 10 Å. The long-range LJ interactions between the ion and water are included using an analytical correction [153]. A constant temperature of 300 K is set using the Langevin thermostat with a coefficient of 1 ps$^{-1}$. The pressure of the box is maintained at 1 bar using the Langevin piston method with a damping coefficient of 20 ps$^{-1}$ (NPT ensemble) [79]. The ion is restrained at the centre of the box with a spring constant of 10 kcal/mol/Å$^2$ to prevent the ion drifting to the boundary. Both the forward and backward FEP transformations are combined with the Bennett acceptance ratio (BAR) to get the final free energy implemented using the VMD `ParseFEP` plugin [85].

## 7.3    Results and Discussion

### 7.3.1    Ion Solvation

Solvation free energies of ions have been measured using different methods, and therefore exhibit some variation, as shown in Table 7.1. The sum of the solvation free energy for a cation–anion pair is determined from the heat of formation measurements, which are reasonably well known. The main source of the uncertainty comes the absolute solvation free energy of the proton, $\Delta G^{\circ}[\text{H}^+]$, which is used as a reference in finding the individual solvation free energies of ions. For example, the large difference of the Gomer [221] and Marcus [62] values from the others is due to the use of substantially larger $\Delta G^{\circ}[\text{H}^+]$ values (11 and 12 kcal/mol, respectively) in these work compared to that of Tissandier et al. [60] The more recent $\Delta G^{\circ}[\text{H}^+]$ value determined by Tissandier et al. [60] ($-264$ kcal/mol) is generally considered more accurate and adapted in other work [222]. The $+1.9$ kcal/mol difference between the values of Tissandier et al. and Kelly et al. is simply due to the choice of the standard state—in the former, a concentration of 1 bar in the gas phase and 1 M in the aqueous phase is used, whereas a concentration of 1 M is used in both phases in the latter. Because the latter condition is assumed in the simulations, we will use the solvation free energies from Kelly et al. [222] in comparisons with the calculated values below.

In order to discuss the issues with using small water droplets in solvation free energy calculations and motivate the proposed method, we first consider the variation of the average charge density around the $Na^+$ and $Cl^-$ ions in a water droplet (Fig. 7.1). The charge densities are obtained from 30 ns MD simulations of a central ion in a water droplet with $R = 24$ Å. It is seen that the effect of both ions on the charge density in the water extends up to 9 Å from the centre. The charge density then remains flat until about 6 Å from the boundary, where it again exhibits variations due to the ordering of water molecules by the vacuum interface. For water droplets with $R < 15$ Å, the two regions overlap, and the water molecules in the inner shells are affected by vacuum. This perturbs the ion-water interactions in the inner shells, which are clearly not acceptable. Thus the use of angular restraints to mimic the polarisation of water molecules in a larger system is necessary when small droplets with $R < 15$ Å are employed [56]. Use of such small systems was unavoidable 2-3 decades ago due to the limited computational resources. However, this is no more the case, and one can avoid the overlap problem by using droplets with $R > 15$ Å.

**Table 7.1:** A compilation of experimental solvation free energies of $Na^+$ and $Cl^-$ from various sources.

|        | Noyes [223] | Gomer [221] | Klots [224] | Marcus [62] | Tissan. [60] | Kelly [222] |
|--------|-------------|-------------|-------------|-------------|--------------|-------------|
| $Na^+$ | $-97.0$     | $-90.6$     | $-100.1$    | $-87.2$     | $-101.3$     | $-103.2$    |
| $Cl^-$ | $-74.8$     | $-81.4$     | $-73.9$     | $-81.3$     | $-72.7$      | $-74.5$     |
| Sum    | $-171.8$    | $-172.0$    | $-174.0$    | $-168.5$    | $-174.0$     | $-177.7$    |



**Figure 7.1:** The radial dependence of the charge density in a water droplet of size $R = 24$ Å for $Na^+$ (black) and $Cl^-$ (red) ions.

**Table 7.2:**  Break down of the solvation free energies as a function of the droplet size $R$ for both $Na^+$ and $Cl^-$. Water molecules are confined using a half-harmonic potential with $k_s = 10$ kcal/mol/Å$^2$, and the ion is restrained at the centre with 10 kcal/mol/Å$^2$. All energies are reported in kcal/mol, and the maximum statistical error for the solvation free energies is 0.3 kcal/mol.

| | $Na^+$ | | | | | $Cl^-$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $R$ (Å) | $\Delta G_{\text{drop-el}}$ | $\Delta G_{\text{cav}}$ | $\Delta G_{\text{el}}$ | $\Delta G_{\text{LJ}}$ | $\Delta G_{\text{solv}}$ | $\Delta G_{\text{drop-el}}$ | $\Delta G_{\text{cav}}$ | $\Delta G_{\text{el}}$ | $\Delta G_{\text{LJ}}$ | $\Delta G_{\text{solv}}$ |
| 6 | $-81.7$ | $-27.4$ | $-109.1$ | $3.2$ | $-105.9$ | $-60.2$ | $-27.4$ | $-87.6$ | $6.0$ | $-81.6$ |
| 9 | $-88.2$ | $-18.2$ | $-106.4$ | $2.8$ | $-103.6$ | $-68.0$ | $-18.2$ | $-86.2$ | $5.6$ | $-80.6$ |
| 12 | $-91.5$ | $-13.6$ | $-105.1$ | $2.7$ | $-102.4$ | $-72.3$ | $-13.6$ | $-86.0$ | $5.3$ | $-80.7$ |
| 15 | $-93.6$ | $-10.9$ | $-104.5$ | $2.6$ | $-101.9$ | $-74.7$ | $-10.9$ | $-85.6$ | $5.2$ | $-80.4$ |
| 18 | $-95.1$ | $-9.1$ | $-104.2$ | $2.5$ | $-101.7$ | $-76.3$ | $-9.1$ | $-85.4$ | $5.0$ | $-80.4$ |
| 21 | $-96.0$ | $-7.8$ | $-103.8$ | $2.6$ | $-101.2$ | $-77.6$ | $-7.8$ | $-85.4$ | $5.0$ | $-80.5$ |
| 24 | $-96.9$ | $-6.8$ | $-103.7$ | $2.5$ | $-101.2$ | $-78.5$ | $-6.8$ | $-85.3$ | $4.7$ | $-80.6$ |
| 30 | $-97.9$ | $-5.5$ | $-103.4$ | $2.4$ | $-101.0$ | $-79.9$ | $-5.5$ | $-85.4$ | $4.5$ | $-80.8$ |
| 36 | $-98.9$ | $-4.5$ | $-103.4$ | $2.4$ | $-101.0$ | $-80.5$ | $-4.5$ | $-85.0$ | $4.4$ | $-80.7$ |

For larger droplets with $R > 15$ Å, we only need to worry about the artefacts introduced by the variations in the charge density near the boundary. Integration of the charge density at the boundary region shows that there is a net charge given by $(1 - 1/\epsilon)q$, consistent with the continuum calculations. This monopole contribution to the ion's interaction energy is cancelled by the cavity term, which leaves the dipole and higher-order interactions behind. Because these terms fall as $1/R^2$ or faster, it should be possible to make them negligibly small by choosing a large enough $R$. To this end, we investigate the convergence of the solvation free energies with $R$ by varying it from 6 to 36 Å for both $Na^+$ and $Cl^-$. The results are shown in Table 7.2, where individual contributions to the free energies are listed separately. For both ions, the convergence of $\Delta G_{\text{solv}}$ occurs around $R = 21$ Å. The $\Delta G_{\text{solv}}$ values for $R \geq 21$ Å differ from that at $R = 36$ Å by 0.1–0.2 kcal/mol, which is within the statistical errors in the calculations. Inspection of the individual contributions in Table 7.2 shows that cancellation of the $1/R$ dependent term in $\Delta G_{\text{drop-el}}$ with $\Delta G_{\text{cav}}$ ensures the fast convergence of the results observed in $\Delta G_{\text{el}}$.

The converged $\Delta G_{\text{solv}}$ values in Table 7.2 are in good agreement with previous calculations where the spherical solvent boundary potential method [57] was used with the CHARMM parameters to obtain $-101.1$ kcal/mol [58] for $Na^+$ and $-80.0$ kcal/-mol [225] for $Cl^-$. Comparison of the calculated $\Delta G_{\text{solv}}$ values with the experimental ones from Kelly et al. in Table 7.1 yields mixed results. There is a small deviation for $Na^+$ because the earlier experimental values were targeted in the calculations where $Na^+$ parameters were determined. A larger discrepancy (6 kcal/mol) occurs for $Cl^-$. Presumably, $Cl^-$ suffers from the neglected cousin syndrome as it plays a relatively minor role in biochemical processes compared to $Na^+$. But its proper parametrisation should nevertheless be considered in future versions of the force field.

We next demonstrate the robustness of the results by showing that they are not affected by the parameters of the potentials used in confining the water molecules in the droplet and restraining the ion at the centre. For the confining potential, we repeat the previous calculations with $k_s = 10$ kcal/mol/Å$^2$ using $k_s = 5$ and 20 kcal/mol/Å$^2$. The solvation

**Table 7.3:** The effect of the harmonic restraint (kcal/mol/Å$^2$) applied to the Na$^+$ ion on its solvation free energy. Calculations are performed in a droplet with $R = 24$ Å. All energies are reported in kcal/mol.

| $k$ | $\Delta G_{\text{drop-el}}$ | $\Delta G_{\text{cav}}$ | $\Delta G_{\text{el}}$ | $\Delta G_{\text{LJ}}$ | $\Delta G_{\text{solv}}$ |
|------|------|------|------|------|------|
| 0.1 | −97.0 | −7.0 | −104.0 | 2.5 | −101.5 |
| 0.5 | −96.8 | −6.9 | −103.7 | 2.5 | −101.2 |
| 1.0 | −97.0 | −6.8 | −103.8 | 2.5 | −101.3 |
| 10.0 | −96.9 | −6.8 | −103.7 | 2.5 | −101.2 |

free energy results are shown in Fig. 7.2 for both Na$^+$ and Cl$^-$. The solvation free energy differs by a maximum of 0.3 kcal/mol between the three $k_s$ values, which is within the statistical errors. Statistical nature of the variations is also apparent from the fact that there is no clear trend in the $\Delta G_{\text{solv}}$ values with increasing $k_s$. Thus the strength of the confinement potential does not have an appreciable effect on the stability of the results. The value of 10 kcal/mol/Å$^2$ is sufficient to confine the water molecules inside the droplet and will be used in the following calculations. The second parameter used in the calculations is the harmonic restraint that keeps the ion at the centre of the droplet. We test the use of different centre of charge restraints for Na$^+$ only, and the results are summarised in Table 7.3. Overall, the solvation energy differs by only 0.1–0.3 kcal/mol. The largest deviation occurs for the 0.1 kcal/mol/Å$^2$ restraint, which is presumably a consequence of under-sampling due to the larger phase-space resulting from the weaker restraint.

For a molecule with distributed partial charges, most of the charges will necessarily be at off-centre positions when the centre of the charge is restrained at the origin. In order to ensure that the proposed method can be applied to molecules, we need to show that its accuracy is maintained for off-centre positions of charges, and find the maximum distance of charges from the centre for which the method still gives an acceptable result. Such information will be useful for choosing an appropriate system size when larger molecules are considered without the need to check the system size dependence for each molecule. We place the ion at off-centre positions by applying a harmonic restraint of 10 kcal/mol/Å$^2$ along the $z$-axis. Calculations are performed using a droplet size of $R = 24$ Å and the results are shown in Table 7.4. We see that the converged value of the solvation free energy is maintained up to 4 Å from the centre. The maximum change in the energy is about 0.3 kcal/mol for Na$^+$ and 0.1 kcal/mol for Cl$^-$, which are within the statistical errors. At 8 Å the ion solvation free energy is still within 1 kcal/mol, but at 12 Å the inner and outer shells start overlapping (Fig. 7.1), which has a larger effect on the solvation free energy. These results suggest that we can use a droplet size of $R = 24$ Å to determine the solvation free energy accurately for solute molecules up to 8 Å in length. Even for molecules of length 16 Å, the expected accuracy is about 1 kcal/mol. Thus using a droplet with radius 24 Å should be adequate for the charged amino acid side chain calculations performed in Section 7.3.2.

A droplet of size 24 Å contains about 2000 water molecules, which is not a large system for today's standards. However, repeated calculations of solvation free energies

**Figure 7.2:** Solvation free energy as a function of the droplet radius $R$ for Na$^+$ (top) and Cl$^-$ (bottom). Three sets of calculations are compared in the plot, corresponding to the three different spring constants used for the confinement of water molecules, $k_s = \{5, 10, 20\}$ kcal/mol/Å$^2$.

for parameter fitting purposes can still turn this into a computationally expensive exercise. This is one of the main reasons why a much smaller droplet radius has been used in previous calculations of solvation free energies. Nevertheless, running these calculations on a GPU can significantly reduce the computation time. This is made possible through GPU optimised codes like OpenMM. For example, for a droplet with radius 21 Å, MD simulations can be run at about 800 ns/day on a Tesla V100 GPU. Increasing the radius to 24 Å, one can still get 600 ns/day, which is sufficient to perform 13 solvation free energy calculations per day. On a consumer-grade GPU card (GTX 1080 Ti), the corresponding run times are 570 and 450 ns/day for 21 Å and 24 Å droplets, respectively. Thus, even using a system size of 24 Å, 10–13 calculations can be performed per day on a GPU, which should be sufficient for optimisation of force

**Table 7.4:** Same as Table 7.2 but as a function of the ion off-centre positions along the $z$-axis with a droplet size of $R = 24$ Å.

| | Na$^+$ | | | | | Cl$^-$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $z$ (Å) | $\Delta G_{\text{drop-el}}$ | $\Delta G_{\text{cav}}$ | $\Delta G_{\text{el}}$ | $\Delta G_{\text{LJ}}$ | $\Delta G_{\text{solv}}$ | $\Delta G_{\text{drop-el}}$ | $\Delta G_{\text{cav}}$ | $\Delta G_{\text{el}}$ | $\Delta G_{\text{LJ}}$ | $\Delta G_{\text{solv}}$ |
| 0 | −96.9 | −6.8 | −103.7 | 2.5 | −101.2 | −78.5 | −6.8 | −85.3 | 4.7 | −80.6 |
| 1 | −96.8 | −6.8 | −103.6 | 2.5 | −101.2 | −78.5 | −6.8 | −85.4 | 4.6 | −80.7 |
| 2 | −97.0 | −6.9 | −103.8 | 2.5 | −101.3 | −78.7 | −6.9 | −85.5 | 4.8 | −80.8 |
| 4 | −96.9 | −7.0 | −104.0 | 2.4 | −101.5 | −78.5 | −7.0 | −85.5 | 4.7 | −80.8 |
| 8 | −96.8 | −7.7 | −104.5 | 2.5 | −101.8 | −78.5 | −7.7 | −86.2 | 4.6 | −81.6 |
| 12 | −97.0 | −9.1 | −106.1 | 2.5 | −103.6 | −78.7 | −9.1 | −87.8 | 4.8 | −83.0 |

fields. These numbers are for 1 ns production time per window, which is probably excessive. For large scale calculations, production times can be optimised which will shorten the computation time further.

Arguments made for the convergence of solvation free energies using SBC can also be made for PBC. Thus it is of interest to repeat the calculations for the Na$^+$ and Cl$^-$ ions using PBC with a variable box size. The results are presented in Table 7.5 for box sizes varying from 20 to 60 Å. The convergence of the solvation free energies occurs around $L = 35$ Å, which has a similar volume to a droplet with $R = 21$ Å, where the convergence occurs in a droplet. Thus at those volumes, boundary artefacts cease to contribute to the solvation free energy of the ion regardless of the specific geometry used. Comparing the SBC and PBC results in Tables 7.2 and 7.5, we see that there is very good agreement between the two methods for Na$^+$, but they differ by about a kcal/mol for Cl$^-$.

## 7.3.2 Charged Amino Acids

As an application of the method, we consider the solvation free energies of the side chain analogues for the four charged amino acids, which play important roles in protein interactions. Calculations are performed using a droplet of size $R = 24$ Å and the same confinement potential with $k_s = 10$ kcal/mol/Å$^2$. The centre of charge of the molecules is restrained at the centre with 10 kcal/mol/Å$^2$. The results are listed in Table 7.6, together with the experimental values [222, 226]. To ensure that convergent results are also obtained for molecules, we have repeated the acetate calculations using droplet sizes $R = 36$ and 48 Å. The results differed from that of $R = 24$ Å by at most 0.1 kcal/mol, which is within the statistical errors. We have also performed the acetate calculations using PBCs with box sizes varying from 20 to 60 Å and obtained a converged value of $-84.7$ kcal/mol. As in the case of Cl$^-$, there is about a kcal/mol difference between the SBC and PBC results. This is presumably caused by the interfacial potential, which shows small variations depending on the model used in its computation. One consequence of using a relatively large droplet is that the multipole contributions to the cavity term envisaged in Eq. (7.12) become negligibly small so that this term is essentially given by the Born energy in Eq. (7.1).

**Table 7.5:** Break down of the solvation free energies as a function of the box length $L$ for both Na$^+$ and Cl$^-$. $\Delta G_{\text{solv}}^{\text{real}}$ is obtained from $\Delta G_{\text{solv}}^{\text{intr}}$ by adding $-12.0z$ kcal/mol for the interfacial potential. All energies are reported in kcal/mol, and the maximum statistical error for the solvation free energies is 0.3 kcal/mol.

| | Na$^+$ | | | | | | Cl$^-$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $L$ (Å) | $\Delta G_{\text{box-el}}$ | $\Delta G_{\text{SE}}$ | $\Delta G_{\text{el}}$ | $\Delta G_{\text{LJ}}$ | $\Delta G_{\text{solv}}^{\text{intr}}$ | $\Delta G_{\text{solv}}^{\text{real}}$ | $\Delta G_{\text{box-el}}$ | $\Delta G_{\text{SE}}$ | $\Delta G_{\text{el}}$ | $\Delta G_{\text{LJ}}$ | $\Delta G_{\text{solv}}^{\text{intr}}$ | $\Delta G_{\text{solv}}^{\text{real}}$ |
| 20 | -67.8 | -23.5 | -91.3 | 2.0 | -89.2 | -101.2 | -71.8 | -23.5 | -95.3 | 3.3 | -92.0 | -80.0 |
| 25 | -72.3 | -18.8 | -91.1 | 2.0 | -89.1 | -101.1 | -76.5 | -18.8 | -95.3 | 3.4 | -91.8 | -79.8 |
| 30 | -75.5 | -15.7 | -91.1 | 2.0 | -89.2 | -101.2 | -79.3 | -15.7 | -95.0 | 3.6 | -91.4 | -79.4 |
| 35 | -77.5 | -13.4 | -90.9 | 2.0 | -88.9 | -100.9 | -81.6 | -13.4 | -95.0 | 3.6 | -91.4 | -79.4 |
| 40 | -79.2 | -11.8 | -90.9 | 2.0 | -88.9 | -100.9 | -83.2 | -11.8 | -95.0 | 3.6 | -91.4 | -79.4 |
| 50 | -81.6 | -9.4 | -91.0 | 2.0 | -89.0 | -101.0 | -85.4 | -9.4 | -94.8 | 3.4 | -91.5 | -79.5 |
| 60 | -83.0 | -7.8 | -90.8 | 2.0 | -88.9 | -100.9 | -87.1 | -7.8 | -94.9 | 3.5 | -91.4 | -79.4 |

**Table 7.6:** Solvation free energies of the charged amino-acid side chain analogues obtained using a droplet size of $R = 24$ Å. All energies are in kcal/mol, and the maximum statistical error for the solvation free energies is 0.3 kcal/mol.

| Amino acid | Analogue | $\Delta G_{\text{drop-el}}$ | $\Delta G_{\text{cav}}$ | $\Delta G_{\text{el}}$ | $\Delta G_{\text{LJ}}$ | $\Delta G_{\text{solv}}$ | Expt.[a] |
|---|---|---|---|---|---|---|---|
| arginine | guanidinium | −62.9 | −6.8 | −69.7 | 1.4 | −68.3 | −67.3 |
| | methyl-guanidinium | −57.1 | −6.8 | −64.0 | 1.9 | −62.1 | |
| | ethyl-guanidinium | −56.7 | −6.8 | −63.5 | 2.5 | −61.0 | |
| | n-propyl-guanidinium | −56.6 | −6.8 | −63.4 | 3.3 | −60.1 | |
| lysine | ammonium | −75.8 | −6.8 | −82.6 | 2.5 | −80.1 | −85.2 |
| | methyl-ammonium | −66.4 | −6.8 | −73.2 | 2.9 | −70.3 | −76.4 |
| | ethyl-ammonium | −65.0 | −6.8 | −71.8 | 3.2 | −68.6 | |
| | n-propyl-ammonium | −64.3 | −6.8 | −71.1 | 3.8 | −67.3 | −71.5 |
| | n-butyl-ammonium | −64.4 | −6.8 | −71.2 | 4.4 | −66.8 | |
| aspartate | acetate | −82.9 | −6.8 | −89.7 | 3.6 | −86.1 | −77.6 |
| | acetate (OPLS-AA)[b] | −76.8 | −6.8 | −83.6 | 1.9 | −81.7 | |
| glutamate | propanoate | −82.3 | −6.8 | −89.2 | 4.1 | −85.1 | −76.2 |

[a]All experimental values are from Kelly et al. [222] except for guanidinium which is taken from Reif et al. [226] The guanidinium value is adjusted to make it consistent with the Kelly et al. set [222], i.e., 1.9 kcal/mol is subtracted for conversion of the state from 1 bar to 1 M in the gas phase and 1.1 kcal/mol is subtracted to shift the $\Delta G°[\text{H}^+]$ value to that of Tissandier et al. [60].
[b]Acetate parameters from Ref. [228] (q,$\sigma$,$\epsilon$): O=(−0.67$e$, 2.96 Å, 0.21 kcal/mol), C=(0.44$e$, 3.75 Å, 0.105 kcal/mol).

Comparing the calculated solvation free energies with the experimental values in Table 7.6, it is seen that good agreement is obtained for the arginine analogue guanidinium. For the three lysine analogues listed, the discrepancy between the experimental and calculated values varies from 4–6 kcal/mol. The solvation free energies are systematically under-predicted, suggesting that the ammonium partial charges need to be boosted slightly in the lysine analogues. The discrepancy grows to about 9 kcal/mol for the aspartate and glutamate side-chain analogues. The solvation free energies are over predicted in both cases. Thus the partial charges on the O atoms of these side chains need to be reduced to improve the agreement with the experiments. We note that in earlier calculations, the experimental values from Sitkoff et al. [227] were targeted where $\Delta G°[\text{H}^+] = -261$ kcal/mol is used. This shifts the solvation free energies by 3 kcal/mol up for the positive side chains and down for the negative ones, making the calculated lysine results more agreeable with experiment but sizeable discrepancies remain for the negative side chains. To indicate the size of the correction needed, we have performed the acetate calculation using the optimised OPLS-AA force field [228], which has a lower O charge (−0.67$e$ compared to −0.76$e$ in CHARMM). As seen in Table 7.6, the discrepancy is reduced to about 4 kcal/mol, so there is further room for improving the aspartate and glutamate charges.

The above observations about the charges on the side chains of the amino acids Arg, Lys, Asp, and Glu are consistent with the observed behaviour of the side-chain interactions during dissociation of peptide ligands from proteins [229, 230]. The N–O distances between two side chains, monitored during potential of mean force calcula-

tions, reveal that Arg–Asp and Arg–Glu contacts can persist beyond 10 Å but such an unexpected behaviour does not occur for Lys contacts. The persistence of the former contacts can be explained by the overcharging of the O atoms in the Asp and Glu side chains which, therefore can stick to the Arg side chain longer than expected. The undercharging of the Lys side chain apparently counters the stickiness of the Asp and Glu side chains so that such a persistence of contacts is not seen in Lys interactions.

## 7.4   Conclusion

A spherical droplet provides the simplest and physically most transparent system for calculation of solvation free energies of ions. We have shown that artefacts arising from the boundary with vacuum become negligibly small for droplet sizes of $R = 21$ Å so that one can obtain uniquely convergent solvation free energies independent of the parameters used for confining water in the droplet or restraining the ion at the centre. The computational effort required is rather small for a single calculation, and the method can be easily upscaled for optimisation of force fields using GPUs and clusters.

Application of the method to the calculation of solvation free energies of ions and the side chain analogues of charged amino acids have revealed discrepancies with experiments in several cases. A sizeable portion of the discrepancies is caused by the older experimental values used in fitting the force fields, which have been shifted in the new compilations by the choice of $\Delta G^\circ[H^+]$ [60, 222]. The free energy calculations can be performed with chemical accuracy nowadays, but this would be of little value if the results are affected by inaccuracies in force fields. So it is important to optimise force fields using the latest solvation free energies. The uniquely convergent solvation free energies obtained using our method could help to make this process more robust.

# Chapter 8

# Concluding Remarks

*"There is no real ending. It's just the place where you stop the story."*

Frank Herbert

## 8.1 Summary of Research

The use of atomistic computer models provides a valuable tool to probe systems at the molecular level that cannot be reached by current experimental methods. The rise of and continued improvements to computer technology, both hardware and software, is paving the way for research avenues that seemed impossible only a few decades ago. With current hardware technology, it is now possible to perform advanced free energy calculations to predict quantities that can be compared to experiments, which is one of the goals of computer simulations. This thesis presents the investigation of using free energy methods to explain observations in biological molecules.

The first three results chapters (4-6) investigate the interaction of ligands in glutamate transporters. Due to the availability of the crystal structure, the focus is on the archaeal homologue. The first study investigates the apparent discrepancy between experiment and simulation of the Na2 site. $Na^+$ ion is observed in simulations to be unstable in the Na2 site and always leaves the binding site. The root cause of this issue is the undercharging of the S atom in the methionine residue ($Glt_{Ph}$:M311 and $Glt_{Tk}$:M314). Increasing the partial charge on the S atom by a factor of three from the default value brings the ion closer to the binding site. Calculation of the binding free energy also agrees with the experimental value, which was previously estimated as a positive value. The work done in this project reconcile the discrepancy between simulation and experiment for the Na2 site. In addition, the methionine residue is conserved throughout the SLC1 family, and methionine mutations are involved in diseases like Alzheimer's disease. Thus this study highlights the importance of having a properly parametrised force field.

The second study of glutamate transporters focuses on the ligand-binding mechanism before the closing of the HP1-HP2 gate and the binding of the second ion at Na2. Previous binding free energy calculations indicated high-affinity binding for the second

Na$^+$ ion and low-affinity binding for Asp. Experimental results published after these computational studies show the opposite behaviour. In this second study, other potential ion binding states are searched. Starting with an intermediate site called Na1$'$, observed in previous computational studies, the possibility of a second Na$^+$ binding to Glt$_{Ph}$ at this state is probed. Remarkably the Na$^+$ ion binds to the D405 residue that was initially bound to Na1$'$. The two ions are each bound to an aspartate residue (D405 and D312) with a separation distance of 3.5 Å. The binding free energy of the outermost ion in this state (Na1–Na1$'$) is less than the value in the previous state (Na1–Na3), and this value agrees with experimental results. In addition, Asp also binds to Glt$_{Ph}$ with the Na1–Na1$'$ ion state with a binding free energy in good agreement with the experimental value, resolving the discrepancy in previous calculations. This suggests that the substrate binds to the protein with two ions bound at the Na1–Na1$'$ state, which is not a state known before in glutamate transporters.

Previous studies established that the Na3 site is the deepest binding site and hypothesised as the final ligand to leave the protein. The final work in Glt$_{Ph}$ reported in Chapter 5 attempts to estimate the release time of Na$^+$ from the Na3 site to the bulk. Previous calculations employed path-independent free energy methods; however, recent experimental observations indicate ligand binding triggers conformational changes in the protein. Thus a path-dependent method is used to capture the effect of protein conformation. The PMF profile for Na3 $\rightarrow$ Na1$'$ reveals a 17 kcal/mol energy barrier. Along the transition, significant conformational changes are observed in the NMDGT motif primarily in the N310 and D312 residues. Inspection of the umbrella windows reveals that D312 plays a chaperoning role for Na$^+$ while N310 behaves like a ratchet that prevents Na$^+$ from falling back to the Na3 site. The transition time estimated with the Smoluchowski equation is around 7 sec, which is about 4% of the total turnover time of Glt$_{Ph}$ at 3 min. Two other transitions take place after this before Na$^+$ reaches the bulk environment, and the transition times are significantly faster. This study reveals that the release of the last Na$^+$ is definitely a slow process, but it is not the rate-limiting step in the transport cycle. In addition, this finding may explain the need for a K$^+$ ion to speed the process in the mammalian variants.

The first side project in Chapter 6 attempts to investigate the cause for the difference in ion conductance of gA embedded in different lipids as observed in experiments. Indeed, the PMF profile obtained from umbrella sampling indicates a difference in energy barrier in K$^+$ ion permeation in gA between POPC and NODS lipids. A breakdown of the potential energy reveals significant differences in the interactions of the K$^+$ ion with the lipid head group and water molecules at the interface. The phosphate head group in POPC has a stronger dipole moment compared to NODS, and this orients water molecules at the lipid-water interface. The orientation of the water molecules at the interface produces the larger energy barrier in the PMF and explains the apparent impedance in ion conductance.

The last project reported in Chapter 7 deals with the simplest system investigated, the solvation of ions and small molecules. The purpose of this project is to compare results for solvation free energy using two different boundary conditions; PBC and SBC. The investigation found that the energies converge to a unique value above a critical volume, which is similar in both systems. In particular, for SBC this means that the boundary

effects (water polarisation at the liquid-vacuum boundary) become negligible. This was a problem back in the 1990s as the hardware technology limits the calculations to very small systems (radius of 8–12 Å). With small systems, the polarisation of water molecules at the boundary affects the final solvation free energy, and thus angular restraints are applied, creating a buffer zone. Increasing the radius to >21 Å reduces the boundary effects but increases the computation time. Running these calculations on the GPUs, however, removes this limitation and allows for the calculations of larger systems. In addition to the solvation of ions, the solvation of amino acid side chain analogues reveals problems with the current force field parameters. The results show an overestimation of the energies for the negative amino acids by about 9 kcal/mol. This explains the apparent stickiness between aspartate/glutamate with arginine during peptide-protein dissociation.

## 8.2   Potential Future Research

For glutamate transporters, there are a number of avenues for future work. As noted in Chapter 4 it is not possible to estimate the transition time from Na1′ to Na3 in the presence of Asp and Na1 with the MD/FEP. A potential project would be to use a path-dependent method to obtain the PMF profile between the two states. One difficulty with this, however, is the flipping of N310 and conformational changes in the protein, which requires the construction of a multidimensional PMF. Although this increases the computation time of the calculations, future generation hardware technology will definitely make it easier. In addition to ligand binding, the transition time for the translocation of the transport domain from the outward to inward conformation is an important avenue to research. An estimate of such transition time may reveal the rate-limiting step in the transport cycle. One hurdle with calculating this is the choice of collective variable (or reaction coordinate) as a simple distance with respect to the membrane normal may not be adequate. Recently, a new method for choosing appropriate collective variables with machine learning is reported [231]. In this method, the algorithm studies the endpoints between transitions and identifies suitable collective variables by supervised learning. More than one collective variable might be needed to describe the motion. A dimensionality reduction representation can be used in this case with either the string method [176, 232] or the path-collective variables (PCV) [233]. The string method coupled with quaternion based collective variables has been used to describe the transition of glycerol-3-phosphate: phosphate antiporter (GlpT) from the OF to IF conformation [234]. Thus estimating the transition energy from the OF to IF in $Glt_{Ph}$ may not be an impossible task. Given the PMF, it is also possible to estimate the diffusion profile more accurately with a Bayesian estimator based on the ABF method [235]. A comparative study with the EAATs will be an appropriate study to explain the faster turnover rate of EAATs potentially. The recent crystal structure of an engineered EAAT1 protein is a potential candidate for the template of homology models [130]. As mentioned previously in the conclusion of Chapter 4, the two-ion intermediate state may be the key to the $K^+$-$Na^+$ exchange process, which may explain the faster turnover rate. It may also be possible to use the current methods used in this thesis to study the effect of certain mutations in neurological diseases. For example, if a certain mutation in EAATs is linked to a neurological disease, we can study the

effect of the mutation to ligand binding. If the mutation disrupts the ligand binding free energy, then the signals in the nervous system are impaired. The predictions from simulations can be compared and verified with physiological experiments. This is just one way to study the impact of mutations on certain neurological diseases. Also, once the effect of the mutation is defined, it is possible to use this as a target for drug design (can be studied using the free energy methods used in this thesis).

One of the significant problems with computational studies of gA is the lack of explicit polarisation in classical force fields. A recent computational study with the polarisable AMOEBA force field reduces the barrier to about 5 kcal/mol [45]. This improvement comes from an increase in the dipole moment of the neighbouring water molecules from 2.59 D to 2.70 D. However, QM/MM simulations indicate that the dipole moment can vary from 1.80 to As mentioned previously2.85 D [187]. Although the AMOEBA force field improves the PMF and ultimately the conductance, the value is still about 30% off. Thus, further study can be done in investigating the effect of polarisation in gA, and the extension to divalent ions like $Ca^{2+}$ can be studied with the AMOEBA force fields. Besides parametrisation, gA can be used as a benchmark system to compare the current polarisable force field models (the fluctuating charge, Drude oscillator and induced polarisation) or potential future models that can determine the robustness of the models.

A major issue with using SBC for MD simulation is the internal pressure that arises due to the vacuum that is not investigated in this thesis. The water molecules minimise its surface area, much like a soap bubble in the air. Many authors have attempted to fix this issue by applying an empirical potential that pushes water molecules out of the centre [157, 220, 236]. However, these potentials need to be parametrised for specific droplet sizes. Thus one possible future project is to investigate the pressure as a function of droplet radius. An analytical correction to the pressure may then be constructed based on this study. Another approach to correct for the pressure is to apply a potential on the water molecules that is independent of the droplet size. This will require testing on different potential functions that can recover the radial density of water profile to that of the experimental value. In addition to the pressure issue in SBC, the negatively charged amino acids in the CHARMM force field overestimates the solvation energies compared to the experimental values from Kelly et al. [222]. This has a crucial implication for not only CHARMM but for other force fields. Taking acetate as an example, the CHARMM force field applies NBFix that replaces the normal LJ values with different values between pairs to match the osmotic pressure from experimental data [237]. It might be possible to obtain parameters without NBFix by first calculating the solvation free energy with different parameters for the O atom ($q$, $\epsilon$, $R_{min}$). Since there may be more than one minimum in the parameter set from the solvation energy, the global minimum must be chosen by fitting the osmotic pressure for $Na^+$ or $K^+$ to experimental data [238]. Data for arginine and lysine side-chain analogues also exists. Thus it might be possible to obtain a self-consistent parameter set without the need for NBFix. This may resolve the issue with the stickiness observed in peptide-protein interactions, and the work can be extended to polarisable force fields.

# Appendix A

# Ion Self-Energy Python Script

The self-energy of ions in Eq. (7.12) is implemented in `Python` using the `MDAnalysis` module to read in the MD trajectory [239, 240]. Since `MDAnalysis` can load trajectories written by a number of MD programs, the code below can be used with any of the programs `MDAnalysis` supports. The scripts used to run the solvation free energy calculations in OpenMM are uploaded in GitHub (https://github.com/jeff231li/solvation-openmm).

* self-energy.py

# Appendix B

# Funnel Potential tclForces Script

The source code below is a Tcl implementation of the funnel potential that was invented by Parrinello and co-workers for use with metadynamics [169]. NAMD provides a high-level Tcl scripting interface for user-defined potentials without the need to recompile the program.

In the NAMD configuration file, the section below defines the funnel potential parameters. The script is written so that the user can define the endpoints **A** and **B** arbitrarily so that the potential is not limited to the three Cartesian axes. The atoms involved in the restraint is chosen by reading a PDB file with columns `beta` or `occupancy` filled with the given user value (similar interface to the `colvar` module). The script also prints the positions of the molecule along the $r_{xy}$ and $r_z$ vectors. This code is used in Chapter 5 for the Na1″ $\rightarrow$ Bulk transition. These scripts are available on GitHub (https://github.com/jeff231li/funnel_potential).

```
* In NAMD config file:  * funnel.tcl
```

# Appendix C

# Position-Dependent Diffusion Coefficient

*"FORTRAN's tragic fate has been its wide acceptance, mentally chaining thousands and thousands of programmers to our past mistakes."*

Edsger W. Dijkstra

The position-dependent diffusion coefficient for a system restrained in a harmonic potential was first derived by Woolf and Roux [172], which requires a Laplace transformation of the velocity autocorrelation. Hummer simplified the procedure by using the position autocorrelation instead [173]. The code below is a Fortran implementation for calculating the diffusion coefficient using the method derived by Hummer. This is used to calculate the diffusion coefficient in Chapter 5. This code is available on GitHub (https://github.com/jeff231li/position_diffusion).

* x_acf.f90

# Bibliography

(1) Allen, M. P.; Tildesley, D. J., *Computer Simulations of Liquids*; Oxford University Press: 2017, DOI: `10.1093/oso/9780198803195.001.0001`.

(2) Radhakrishnan, R. Multiscale Modeling, 2018, `https://fling.seas.upenn.edu/~biophys/dynamic/wordpress/?page_id=51`.

(3) Alder, B. J.; Wainwright, T. E *Journal of Chemical Physics* **1964**, *136*, 405–411, DOI: `10.1063/1.1730376`.

(4) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *Journal of Chemical Physics* **1953**, *21*, 1087–1092, DOI: `10.1063/1.1699114`.

(5) Rahman, A *Physical Review* **1964**, *136*, A405, DOI: `10.1103/PhysRev.136.A405`.

(6) Rahman, A.; Stillinger, F. *Journal of Chemical Physics* **1971**, *55*, 3336–3359, DOI: `10.1063/1.1676585`.

(7) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590, DOI: `10.1038/267585a0`.

(8) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713, DOI: `10.1021/acs.jctc.5b00255`.

(9) MacKerell, A. D. et al. *Journal of Physical Chemistry B* **1998**, *102*, 3586–3616, DOI: `10.1021/jp973084f`.

(10) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *Journal of Computational Chemistry* **2004**, *25*, 1656–1676, DOI: `10.1002/jcc.20090`.

(11) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. *Journal of Chemical Theory and Computation* **2015**, *11*, 3499–3509, DOI: `10.1021/acs.jctc.5b00356`.

(12) Splettstoesser, T. Action potentials and synapses, 2017, `https://qbi.uq.edu.au/brain-basics/brain/brain-physiology/action-potentials-and-synapses`.

(13) Vandenberg, R. J.; Ryan, R. M. *Physiological reviews* **2013**, *93*, 1621–57, DOI: `10.1152/physrev.00007.2013`.

(14) Danbolt, N. C. *Progress in Neurobiology* **2001**, *65*, 1–105, DOI: `10.1016/S0301-0082(00)00067-8`.

(15) Hynd, M. R.; Scott, H. L.; Dodd, P. R. *Neurochemistry International* **2004**, *45*, 583–595, DOI: `10.1016/j.neuint.2004.03.007`.

(16) Rothstein, J. D.; Martin, L. J.; Kuncl, R. W. *New England Journal of Medicine* **1992**, *326*, 1464–1468, DOI: 10.1056/NEJM199205283262204.

(17) Rossi, D. J.; Oshima, T; Attwell, D *Nature* **2000**, *403*, 316–321, DOI: 10.1038/35002090.

(18) During, M. J.; Spencer, D. D. *Lancet* **1993**, *341*, 1607–1610, DOI: 10.1016/0140-6736(93)90754-5.

(19) Arriza, J. L.; Kavanaugh, M. P.; Fairman, W. A.; Wu, Y. N.; Murdoch, G. H.; North, R. A.; Amara, S. G. *Journal of Biological Chemistry* **1993**, *268*, 15329–15332.

(20) Levy, L. M.; Warr, O; Attwell, D *Journal of Neuroscience* **1998**, *18*, 9620–9628, DOI: 10.1523/JNEUROSCI.18-23-09620.1998.

(21) Owe, S. G.; Marcaggi, P.; Attwell, D. *Journal of Physiology* **2006**, *577*, 591–599, DOI: 10.1113/jphysiol.2006.116830.

(22) Fairman, W. A.; Vandenberg, R. J.; Arriza, J. L.; Kavanaugh, M. P.; Amara, S. G. *Nature* **1995**, *375*, 599–603, DOI: 10.1038/375599a0.

(23) Jardetzky, O. *Nature* **1966**, *211*, 969–970, DOI: 10.1038/211969a0.

(24) Yernool, D.; Boudker, O.; Jin, Y.; Gouaux, E. *Nature* **2004**, *431*, 811–8, DOI: 10.1038/nature03018.

(25) Kavanaugh, M. P.; Bendahan, A.; Zerangue, N.; Zhang, Y.; Kanner, B. I. *Journal of Biological Chemistry* **1997**, *272*, 1703–1708, DOI: 10.1074/jbc.272.3.1703.

(26) Jiang, J.; Shrivastava, I. H.; Watts, S. D.; Bahar, I.; Amara, S. G. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 15141–15146, DOI: 10.1073/pnas.1112216108.

(27) Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. *Annual Review of Biophysics* **2010**, *39*, 23–42, DOI: 10.1146/annurev.biophys.093008.131258.

(28) Stolzenberg, S.; Khelashvili, G.; Weinstein, H. *Journal of Physical Chemistry B* **2012**, *116*, 5372–5383, DOI: 10.1021/jp301726s.

(29) Akyuz, N.; Altman, R. B.; Blanchard, S. C.; Boudker, O. *Nature* **2013**, *502*, 114–8, DOI: 10.1038/nature12265.

(30) Akyuz, N.; Georgieva, E. R.; Zhou, Z.; Stolzenberg, S.; Cuendet, M. A.; Khelashvili, G.; Altman, R. B.; Terry, D. S.; Freed, J. H.; Weinstein, H.; Boudker, O.; Blanchard, S. C. *Nature* **2015**, *518*, 68–73, DOI: 10.1038/nature14158.

(31) Boudker, O.; Ryan, R. M.; Yernool, D.; Shimamoto, K.; Gouaux, E. *Nature* **2007**, *445*, 387–93, DOI: 10.1038/nature05455.

(32) Baştuğ, T.; Heinzelmann, G.; Kuyucak, S.; Salim, M.; Vandenberg, R. J.; Ryan, R. M. *PLOS One* **2012**, *7*, e33058, DOI: 10.1371/journal.pone.0033058.

(33) Guskov, A.; Jensen, S.; Faustino, I.; Marrink, S. J.; Slotboom, D. J. *Nature Communications* **2016**, *7*, 1–16, DOI: 10.1038/ncomms13420.

(34) Urry, D. W. *Proceedings of the National Academy of Sciences* **1971**, *68*, 672–676, DOI: 10.1073/pnas.68.3.672.

(35) Lundbaek, J. A.; Collingwood, S. A.; Ingólfsson, H. I.; Kapoor, R.; Andersen, O. S. *Journal of the Royal Society, Interface* **2010**, *7*, 373–395, DOI: 10.1098/rsif.2009.0443.

(36) Ketchem, R. R.; Lee, K. C.; Huo, S; Cross, T. A. *Journal of Biomolecular NMR* **1996**, *8*, 1–14, DOI: 10.1007/BF00198135.

(37) Townsley, L. E.; Tucker, W. A.; Sham, S.; Hinton, J. F. *Biochemistry* **2001**, *40*, 11676–11686, DOI: 10.1021/bi010942w.

(38) Allen, T. W.; Andersen, O. S.; Roux, B. *Journal of the American Chemical Society* **2003**, *125*, 9868–9877, DOI: 10.1021/ja029317k.

(39) Kelkar, D. A.; Chattopadhyay, A. *Biochimica et Biophysica Acta* **2007**, *1768*, 2011–25, DOI: 10.1016/j.bbamem.2007.05.011.

(40) Baştuğ, T.; Kuyucak, S. *Biophysical Journal* **2006**, *90*, 3941–3950, DOI: 10.1529/biophysj.105.074633.

(41) Allen, T. W.; Baştuğ, T.; Kuyucak, S.; Chung, S.-H. *Biophysical Journal* **2003**, *84*, 2159–68, DOI: 10.1016/S0006-3495(03)75022-X.

(42) Allen, T. W.; Andersen, O. S.; Roux, B. *Proceedings of the National Academy of Sciences* **2004**, *101*, 117–122, DOI: 10.1073/pnas.2635314100.

(43) Baştuğ, T; Kuyucak, S *Journal of Chemical Physics* **2007**, *126*, 105103, DOI: 10.1063/1.2710267.

(44) Hille, B., *Ion Channels of Excitable Membranes*; Sinauer Sunderland, MA: 2001, DOI: 10.1016/S0006-3495(78)85489-7.

(45) Peng, X.; Zhang, Y.; Chu, H.; Li, Y.; Zhang, D.; Cao, L.; Li, G. *Journal of Chemical Theory and Computation* **2016**, *12*, 2973–2982, DOI: 10.1021/acs.jctc.6b00128.

(46) De Godoy, C. M. G.; Cukierman, S. *Journal of Physical Chemistry B* **2011**, *115*, 5026–5031, DOI: 10.1021/jp2010716.

(47) Wyatt, D. L.; Godoy, C. M.G. D.; Cukierman, S. *Journal of Physical Chemistry B* **2009**, *113*, 6725–6731, DOI: 10.1021/jp900087g.

(48) Åqvist, J. *Journal of Physical Chemistry* **1990**, *94*, 8021–8024, DOI: 10.1021/j100384a009.

(49) Jensen, K. P.; Jorgensen, W. L. *Journal of Chemical Theory and Computation* **2006**, *2*, 1499–1509, DOI: 10.1021/ct600252r.

(50) Luo, Y.; Roux, B. *Journal of Physical Chemistry Letters* **2010**, *1*, 183–189, DOI: 10.1021/jz900079w.

(51) Timko, J; De Castro, A.; Kuyucak, S. *Journal of Chemical Physics* **2011**, *134*, 204510, DOI: 10.1063/1.3595261.

(52) Born, M. *Zeitschrift für Physik* **1920**, *1*, 45–48, DOI: doi.org/10.1007/BF01881023.

(53) Kirkwood, J. G. *Journal of Chemical Physics* **1934**, *2*, 351–361, DOI: 10.1063/1.1749489.

(54)   Hummer, G.; Pratt, L. R.; García, A. E. *Journal of Physical Chemistry* **1995**, *99*, 14188–14194, DOI: 10.1021/j100038a062.

(55)   Hummer, G.; Pratt, L. R.; García, A. E. *Journal of Physical Chemistry* **1996**, *100*, 1206–1215, DOI: 10.1021/jp951011v.

(56)   King, G.; Warshel, A. *Journal of Chemical Physics* **1989**, *91*, 3647–3661, DOI: 10.1063/1.456845.

(57)   Beglov, D.; Roux, B. *Journal of Chemical Physics* **1994**, *100*, 9050–9063, DOI: 10.1063/1.466711.

(58)   Lin, Y. L.; Aleksandrov, A.; Simonson, T.; Roux, B. *Journal of Chemical Theory and Computation* **2014**, *10*, 2690–2709, DOI: 10.1021/ct500195p.

(59)   Lamoureux, G.; Roux, B. *Journal of Physical Chemistry B* **2006**, *110*, 3308–3322, DOI: 10.1021/jp056043p.

(60)   Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *Journal of Physical Chemistry A* **1998**, *102*, 7787–7794, DOI: 10.1021/jp982638r.

(61)   Schmid, R.; Miah, A. M.; Sapunov, V. N. *Physical Chemistry Chemical Physics* **2000**, *2*, 97–102, DOI: 10.1039/A907160A.

(62)   Marcus, Y. *Biophysical Chemistry* **1994**, *51*, 111–127, DOI: 10.1016/0301-4622(94)00051-4.

(63)   Born, M.; Oppenheimer, R. *Annalen der Physik* **1927**, *389*, 457–484, DOI: 10.1002/andp.19273892002.

(64)   Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, A1133, DOI: 10.1103/PhysRev.140.A1133.

(65)   Hohenberg, P.; Kohn, W. *Physical Review* **1964**, *136*, B864, DOI: 10.1103/PhysRev.136.B864.

(66)   Car, R.; Parrinello, M. *Physical Review Letters* **1985**, *55*, 2471–2474, DOI: 10.1103/PhysRevLett.55.2471.

(67)   Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Structure, Function and Bioinformatics* **2006**, *65*, 712–725, DOI: 10.1002/prot.21123.

(68)   Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236, DOI: 10.1021/ja9621760.

(69)   Horta, B. A.; Fuchs, P. F.; van Gunsteren, W. F.; Hünenberger, P. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 1016–1031, DOI: 10.1021/ct1006407.

(70)   Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. a.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *Journal of Physical Chemistry B* **2010**, *114*, 2549–2564, DOI: 10.1021/jp910674d.

(71)   Baker, C. M. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2015**, *5*, 241–254, DOI: 10.1002/wcms.1215.

(72) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2013**, *9*, 5430–5449, DOI: 10.1021/ct400781b.

(73) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D.; Schulten, K.; Roux, B. *Journal of Physical Chemistry Letters* **2011**, *2*, 87–92, DOI: 10.1021/jz101461d.

(74) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327–341, DOI: 10.1016/0021-9991(77)90098-5.

(75) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G.E. M. *Journal of Computational Chemistry* **1997**, *18*, 1463–1472, DOI: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(76) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. *Journal of Chemical Theory and Computation* **2015**, *11*, 1864–1874, DOI: 10.1021/ct5010406.

(77) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *Journal of Chemical Physics* **1982**, *76*, 637–649, DOI: 10.1063/1.442716.

(78) Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* **1993**, *98*, 10089–10092, DOI: 10.1063/1.464397.

(79) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *Journal of Chemical Physics* **1995**, *103*, 4613–4621, DOI: 10.1063/1.470648.

(80) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *Journal of Chemical Physics* **1994**, *101*, 4177–4189, DOI: 10.1063/1.467468.

(81) Beveridge, D. L.; DiCapua, F. M. *Annual Review of Biophysics and Biophysical Chemistry* **1989**, *18*, 431–492, DOI: 10.1146/annurev.bb.18.060189.002243.

(82) Heinzelmann, G.; Baştuğ, T.; Kuyucak, S. *Biophysical Journal* **2011**, *101*, 2380–8, DOI: 10.1016/j.bpj.2011.10.010.

(83) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P., *Numerical Recipes 3rd Edition: The Art of Scientific Computing*; Cambridge university press: 2007.

(84) Bennett, C. H. *Journal of Computational Physics* **1976**, *22*, 245–268, DOI: 10.1016/0021-9991(76)90078-4.

(85) Liu, P.; Dehez, F.; Cai, W.; Chipot, C. *Journal of Chemical Theory and Computation* **2012**, *8*, 2606–2616, DOI: 10.1021/ct300242f.

(86) Shirts, M. R.; Chodera, J. D. *Journal of Chemical Physics* **2008**, *129*, 124105, DOI: 10.1063/1.2978177.

(87) Chodera, J. D. *Journal of Chemical Theory and Computation* **2016**, *12*, 1799–1805, DOI: 10.1021/acs.jctc.5b00784.

(88) Torrie, G. M.; Valleau, J. P. *Journal of Computational Physics* **1977**, *23*, 187–199, DOI: 10.1016/0021-9991(77)90121-8.

(89) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021, DOI: 10.1002/jcc.540130812.

(90)    Grossfield, A. WHAM: The Weighted Histogram Analysis Method, 2013.

(91)    Chen, P. C.; Kuyucak, S. *Biophysical Journal* **2011**, *100*, 2466–2474, DOI: `10.1016/j.bpj.2011.03.052`.

(92)    Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. *Journal of Chemical Physics* **2008**, *128*, 1–13, DOI: `10.1063/1.2829861`.

(93)    Laio, A.; Gervasio, F. L. *Reports on Progress in Physics* **2008**, *71*, 126601, DOI: `10.1088/0034-4885/71/12/126601`.

(94)    Setiadi, J.; Heinzelmann, G.; Kuyucak, S. *Biomolecules* **2015**, *5*, 3067–3086, DOI: `10.3390/biom5043067`.

(95)    Lisman, J. E.; Raghavachari, S.; Tsien, R. W. *Nature Reviews Neuroscience* **2007**, *8*, 597–609, DOI: `10.1038/nrn2191`.

(96)    El Mestikawy, S.; Wallén-Mackenzie, A.; Fortin, G. M.; Descarries, L.; Trudeau, L.-E. *Nature Reviews Neuroscience* **2011**, *12*, 204–16, DOI: `10.1038/nrn2969`.

(97)    Herman, M. A.; Jahr, C. E. *Journal of Neuroscience* **2007**, *27*, 9736–9741, DOI: `10.1523/JNEUROSCI.3009-07.2007`.

(98)    Zerangue, N.; Kavanaugh, M. P. *Nature* **1996**, *383*, 634–637, DOI: `10.1038/383634a0`.

(99)    Slotboom, D. J.; Konings, W. N.; Lolkema, J. S. *Microbiology and Molecular Biology Reviews* **1999**, *63*, 293–307.

(100)   Shafqat, S.; Tamarappoo, B. K.; Kilberg, M. S.; Puranam, R. S.; McNamara, J. O.; Guadaño-Ferraz, A.; Fremeau, R. T. *Journal of Biological Chemistry* **1993**, *268*, 15351–15355.

(101)   Utsunomiya-Tate, N; Endou, H; Kanai, Y *Journal of Biological Chemistry* **1996**, *271*, 14883–14890, DOI: `10.1074/jbc.271.25.14883`.

(102)   Reyes, N.; Ginter, C.; Boudker, O. *Nature* **2009**, *462*, 880–5, DOI: `10.1038/nature08616`.

(103)   Verdon, G.; Boudker, O. *Nature Structural and Molecular Biology* **2012**, *19*, 355–7, DOI: `10.1038/nsmb.2233`.

(104)   Ryan, R. M.; Compton, E. L. R.; Mindell, J. A. *Journal of Biological Chemistry* **2009**, *284*, 17540–17548, DOI: `10.1074/jbc.M109.005926`.

(105)   Shrivastava, I. H.; Jiang, J.; Amara, S. G.; Bahar, I. *Journal of Biological Chemistry* **2008**, *283*, 28680–28690, DOI: `10.1074/jbc.M800889200`.

(106)   Huang, Z.; Tajkhorshid, E. *Biophysical Journal* **2008**, *95*, 2292–300, DOI: `10.1529/biophysj.108.133421`.

(107)   Groeneveld, M.; Slotboom, D. J. *Biochemistry* **2010**, *49*, 3511–3513, DOI: `10.1021/bi100430s`.

(108)   Holley, D. C.; Kavanaugh, M. P. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **2009**, *364*, 155–161, DOI: `10.1098/rstb.2008.0246`.

(109) Larsson, H. P.; Wang, X.; Lev, B.; Baconguis, I.; Caplan, D. A.; Vyleta, N. P.; Koch, H. P.; Diez-Sampedro, A.; Noskov, S. Y. *Proceedings of the National Academy of Sciences* **2010**, *107*, 13912–13917, DOI: 10.1073/pnas.1006289107.

(110) Tao, Z.; Rosental, N.; Kanner, B. I.; Gameiro, A.; Mwaura, J.; Grewer, C. *Journal of Biological Chemistry* **2010**, *285*, 17725–17733, DOI: 10.1074/jbc.M110.121798.

(111) Huang, Z.; Tajkhorshid, E. *Biophysical Journal* **2010**, *99*, 1416–25, DOI: 10.1016/j.bpj.2010.06.052.

(112) Tao, Z.; Zhang, Z.; Grewer, C. *Journal of Biological Chemistry* **2006**, *281*, 10263–10272, DOI: 10.1074/jbc.M510739200.

(113) Heinzelmann, G.; Bastug, T.; Kuyucak, S. *Journal of Physical Chemistry B* **2013**, *117*, 5486–5496, DOI: 10.1021/jp4010423.

(114) Lev, B.; Noskov, S. Y. *Physical Chemistry Chemical Physics* **2013**, *15*, 2397–2404, DOI: 10.1039/c2cp42860a.

(115) Venkatesan, S.; Saha, K.; Sohail, A.; Sandtner, W.; Freissmuth, M.; Ecker, G. F.; Sitte, H. H.; Stockner, T. *PLOS Computational Biology* **2015**, *11*, e1004551, DOI: 10.1371/journal.pcbi.1004551.

(116) Hänelt, I; Jensen, S; Wunnicke, D; Slotboom, D. J. *Journal of Biological Chemistry* **2015**, *290*, 15962–15972, DOI: 10.1074/jbc.M115.656876.

(117) Ewers, D; Becher, T; Machtens, J. P.; Weyand, I; Fahlke, C *Proceedings of the National Academy of Sciences* **2013**, *110*, 1–6, DOI: 10.1073/pnas.1300772110.

(118) Reyes, N; Oh, S; Boudker, O *Nature Structural and Molecular Biology* **2013**, *20*, 634–640, DOI: 10.1038/nsmb.2548.

(119) Crisman, T. J.; Qu, S.; Kanner, B. I.; Forrest, L. R. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 20752–20757, DOI: 10.1073/pnas.0908570106.

(120) Grazioso, G.; Limongelli, V.; Branduardi, D.; Novellino, E.; De Micheli, C.; Cavalli, A.; Parrinello, M. *Journal of the American Chemical Society* **2012**, *134*, 453–63, DOI: 10.1021/ja208485w.

(121) DeChancie, J.; Shrivastava, I. H.; Bahar, I. *Molecular BioSystems* **2011**, *7*, 832–842, DOI: 10.1039/c0mb00175a.

(122) Zomot, E.; Bahar, I. *Journal of Biological Chemistry* **2013**, *288*, 8231–8237, DOI: 10.1074/jbc.M112.438432.

(123) Gu, Y.; Shrivastava, I. H.; Amara, S. G.; Bahar, I. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 2589–2594, DOI: 10.1073/pnas.0812299106.

(124) Gur, M.; Zomot, E.; Bahar, I. *Journal of Chemical Physics* **2013**, *139*, DOI: 10.1063/1.4816375.

(125) Das, A.; Gur, M.; Cheng, M. H.; Jo, S.; Bahar, I.; Roux, B. *PLOS Computational Biology* **2014**, *10*, DOI: 10.1371/journal.pcbi.1003521.

(126) Erkens, G. B.; Hänelt, I.; Goudsmits, J. M. H.; Slotboom, D. J.; Van Oijen, A. M. *Nature* **2013**, *502*, 119–123, DOI: 10.1038/nature12538.

(127) Machtens, J.-P.; Kortzak, D.; Lansche, C.; Leinenweber, A.; Kilian, P.; Begemann, B.; Zachariae, U.; Ewers, D.; de Groot, B. L.; Rodolfo, B.; Fahlke, C. *Cell* **2015**, *160*, 542–553, DOI: 10.1016/j.cell.2014.12.035.

(128) Cater, R. J.; Vandenberg, R. J.; Ryan, R. M. *Biophysical Journal* **2014**, *107*, 621–9, DOI: 10.1016/j.bpj.2014.05.046.

(129) Borre, L; Kanner, B. I. *Journal of Biological Chemistry* **2004**, *279*, 2513–2519, DOI: 10.1074/jbc.M311446200.

(130) Canul-Tec, J. C.; Assal, R.; Cirri, E.; Legrand, P.; Brier, S.; Chamot-Rooke, J.; Reyes, N. *Nature* **2017**, *544*, 446–451, DOI: 10.1038/nature22064.

(131) Watzke, N; Rauen, T; Bamberg, E; Grewer, C *Journal of General Physiology* **2000**, *116*, 609–622, DOI: 10.1085/jgp.116.5.609.

(132) Grewer, C.; Watzke, N.; Rauen, T.; Bicho, A. *Journal of Biological Chemistry* **2003**, *278*, 2585–2592, DOI: 10.1074/jbc.M207956200.

(133) Heinzelmann, G.; Chen, P. C.; Kuyucak, S. *Journal of Physical Chemistry B* **2014**, *118*, 1813–1824, DOI: 10.1021/jp412195m.

(134) Teichman, S.; Qu, S.; Kanner, B. I. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 14297–14302, DOI: 10.1073/pnas.0904625106.

(135) Mwaura, J.; Tao, Z.; James, H.; Albers, T.; Schwartz, A.; Grewer, C. *ACS Chemical Neuroscience* **2012**, *3*, 1073–1083, DOI: dx.doi.org/10.1021/cn300163p.

(136) Jensen, S.; Guskov, A.; Rempel, S.; Hänelt, I.; Slotboom, D. J. *Nature Structural and Molecular Biology* **2013**, *20*, 1224–1226, DOI: 10.1038/nsmb.2663.

(137) Heinzelmann, G.; Kuyucak, S. *Biophysical Journal* **2014**, *106*, 2675–83, DOI: 10.1016/j.bpj.2014.05.010.

(138) Bahar, I. *Biophysical Journal* **2014**, *106*, 2547–2548, DOI: 10.1016/j.bpj.2014.05.011.

(139) Bröer, A.; Wagner, C.; Lang, F.; Bröer, S. *Biochemical Journal* **2015**, *346*, 705–710, DOI: 10.1042/bj3460705.

(140) Grewer, C.; Grabsch, E. *Journal of Physiology* **2004**, *557*, 747–759, DOI: 10.1113/jphysiol.2004.062521.

(141) Zander, C. B.; Albers, T.; Grewer, C. *Journal of general physiology* **2013**, *141*, 659–72, DOI: 10.1085/jgp.201210948.

(142) Scopelliti, A. J.; Heinzelmann, G.; Kuyucak, S.; Ryan, R. M.; Vandenberg, R. J. *Journal of Biological Chemistry* **2014**, *289*, 17468–79, DOI: 10.1074/jbc.M114.565242.

(143) Heinzelmann, G.; Kuyucak, S. *PLOS One* **2014**, *9*, e92089, DOI: 10.1371/journal.pone.0092089.

(144) Verdon, G.; Oh, S.; Serio, R.; Boudker, O. *eLife* **2014**, *2014*, 1–23, DOI: `10.7554/eLife.02283.001`.

(145) Ehrnstorfer, I. A.; Geertsma, E. R.; Pardon, E.; Steyaert, J.; Dutzler, R. *Nature Structural and Molecular Biology* **2014**, *21*, 990–996, DOI: `10.1038/nsmb.2904`.

(146) Bozzi, A. T.; Bane, L. B.; Weihofen, W. A.; McCabe, A. L.; Singharoy, A.; Chipot, C. J.; Schulten, K.; Gaudet, R. *Proceedings of the National Academy of Sciences* **2016**, *37*, 10310–10315, DOI: `10.1073/pnas.1607734113`.

(147) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14*, 33–38, DOI: `10.1016/0263-7855(96)00018-5`.

(148) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Laxmikant, K.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802, DOI: `10.1002/jcc.20289`.

(149) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *Journal of Computational Chemistry* **2010**, *31*, 671–690, DOI: `10.1002/jcc.21367`.

(150) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *Journal of Physical Chemistry B* **2003**, *107*, 9535–9551, DOI: `10.1021/jp0217839`.

(151) Carlsson, J.; Åqvist, J. *Journal of Physical Chemistry B* **2005**, *109*, 6448–6456, DOI: `10.1021/jp046022f`.

(152) Chipot, C; Pohorille, A, *Free Energy Calculations. Theory and Applications in Chemistry and Biology*, 1st ed.; Springer-Verlag Berlin Heidelberg: Berlin, 2007, DOI: `10.1007/978-3-540-38448-9`.

(153) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. *Journal of Physical Chemistry B* **2007**, *111*, 13052–13063, DOI: `10.1021/jp0735987`.

(154) Shannon, R. D. *Acta Crystallographica Section A* **1976**, *32*, 751–767, DOI: `10.1107/S0567739476001551`.

(155) Haynes, W. M., *CRC Handbook of Chemistry and Physics*; CRC Press: 2003.

(156) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. *Biochemistry* **1981**, *20*, 849–855, DOI: `10.1021/bi00507a030`.

(157) Deng, Y.; Roux, B. *Journal of Physical Chemistry B* **2004**, *108*, 16567–16576, DOI: `10.1021/jp048502c`.

(158) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *Journal of Chemical Physics* **2003**, *119*, 5740–5761, DOI: `10.1063/1.1587119`.

(159) Valley, C. C.; Cembran, A.; Perlmutter, J. D.; Lewis, A. K.; Labello, N. P.; Gao, J.; Sachs, J. N. *Journal of Biological Chemistry* **2012**, *287*, 34979–34991, DOI: `10.1074/jbc.M112.374504`.

(160) Rosental, N.; Kanner, B. I. *Journal of Biological Chemistry* **2010**, *285*, 21241–21248, DOI: `10.1074/jbc.M109.087163`.

(161) Setiadi, J.; Kuyucak, S. *Journal of Physical Chemistry B* **2017**, *121*, 9526–9531, DOI: `10.1021/acs.jpcb.7b07366`.

(162) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *Journal of Chemical Physics* **1983**, *79*, 926–935, DOI: 10.1063/1.445869.

(163) Neria, E.; Fischer, S.; Karplus, M. *Journal of Chemical Physics* **1996**, *105*, 1902–1921, DOI: 10.1063/1.472061.

(164) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. *Journal of Physical Chemistry B* **2012**, *114*, 7830–7843, DOI: 10.1021/jp101759q.

(165) Feller, S. E.; Pastor, R. W.; Rojnuckarin, A.; Bogusz, S.; Brooks, B. R. *Journal of Physical Chemistry* **1996**, *100*, 17011–17020, DOI: 10.1021/jp9614658.

(166) Cecchini, M.; Krivov, S. V.; Spichty, M.; Karplus, M. *Journal of Physical Chemistry B* **2009**, *113*, 9728–9740, DOI: 10.1021/jp9020646.

(167) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *Journal of Chemical Physics* **1994**, *100*, 9025–9031, DOI: 10.1063/1.466707.

(168) Arriza, J. L.; Eliasof, S; Kavanaugh, M. P.; Amara, S. G. *Proceedings of the National Academy of Sciences* **1997**, *94*, 4155–60, DOI: 10.1073/pnas.94.8.4155.

(169) Limongelli, V.; Bonomi, M.; Parrinello, M. *Proceedings of the National Academy of Sciences* **2013**, *110*, 6358–6363, DOI: 10.1073/pnas.1303186110.

(170) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophysical Journal* **1997**, *72*, 1568–1581, DOI: 10.1016/S0006-3495(97)78804-0.

(171) Szabo, A; Schulten, K; Schulten, Z *Journal of Chemical Physics* **1980**, *72*, 4350–4357, DOI: 10.1063/1.439715.

(172) Woolf, T. B.; Roux, B. *Journal of the American Chemical Society* **1994**, *116*, 5916–5926, DOI: 10.1021/ja00092a048.

(173) Hummer, G. *New Journal of Physics* **2005**, *7*, 34, DOI: 10.1088/1367-2630/7/1/034.

(174) Harned, H. S.; Owen, B. B., *The Physical Chemistry of Electrolytic Solutions*; Reinhold: New York, 1958.

(175) Grewer, C.; Watzke, N.; Wiessner, M.; Rauen, T. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 9706–9711, DOI: 10.1073/pnas.160170397.

(176) Gan, W.; Yang, S.; Roux, B. *Biophysical Journal* **2009**, *97*, 8–10, DOI: 10.1016/j.bpj.2009.06.016.

(177) Arseniev, A. S.; Barsukov, I. L.; Bystrov, V. F.; Lomize, A. L.; Ovchinnikov, Y. A. *FEBS Letters* **1985**, *186*, 168–174, DOI: 10.1016/0014-5793(85)80702-X.

(178) Ketchem, R. R.; Hu, W; Cross, T. A. *Science* **1993**, *261*, 1457–1460, DOI: 10.1126/science.7690158.

(179) Sarges, R.; Witkop, B. *Journal of the American Chemical Society* **1965**, *87*, 2011–2020, DOI: 10.1021/ja01087a027.

(180) Ramachandran G. N. Chandrasekaran, R *Progress in Peptide Research* **1972**, *2*, 195–215.

(181) Allen, T. W.; Andersen, O. S.; Roux, B. *Biophysical Journal* **2006**, *90*, 3447–3468, DOI: `10.1529/biophysj.105.077073`.

(182) Roux, B.; Karplus, M. *Annual Review of Biophysics and Biomolecular Structure* **1994**, *23*, DOI: `10.1146/annurev.bb.23.060194.003503`.

(183) Partenskii, M. B.; Jordan, P. C. *Quarterly Reviews of Biophysics* **1992**, *25*, 477–510, DOI: `10.1017/S0033583500004388`.

(184) Edwards, S.; Corry, B.; Kuyucak, S.; Chung, S.-H. *Biophysical Journal* **2002**, *83*, 1348–60, DOI: `10.1016/S0006-3495(02)73905-2`.

(185) Baştuğ, T.; Gray-Weale, A.; Patra, S. M.; Kuyucak, S. *Biophysical Journal* **2006**, *90*, 2285–96, DOI: `10.1529/biophysj.105.073205`.

(186) Baştuğ, T.; Kuyucak, S. *Chemical Physics Letters* **2007**, *436*, 383–387, DOI: `10.1016/j.cplett.2007.01.078`.

(187) Timko, J.; Kuyucak, S. *Journal of Chemical Physics* **2012**, *137*, 205106, DOI: `10.1063/1.4768247`.

(188) Van Meer, G.; Voelker, D. R.; Feigenson, G. W. *Nature Reviews Molecular Cell Biology* **2008**, *9*, 112, DOI: `10.1038/nrm2330`.

(189) Lee, S. Y.; Lee, A.; Chen, J.; MacKinnon, R. *Proceedings of the National Academy of Sciences* **2005**, *102*, 15441–15446, DOI: `10.1073/pnas.0507651102`.

(190) Jensen, M.; Mouritsen, O. G. *Biochimica et Biophysica Acta* **2004**, *1666*, 205–226, DOI: `10.1016/j.bbamem.2004.06.009`.

(191) Elmore, D. E.; Dougherty, D. A. *Biophysical Journal* **2003**, *85*, 1512–1524, DOI: `10.1016/S0006-3495(03)74584-6`.

(192) Valiyaveetil, F. I.; Zhou, Y.; MacKinnon, R. *Biochemistry* **2002**, *41*, 10771–10777, DOI: `10.1021/bi026215y`.

(193) Qin, Z.; Tepper, H. L.; Voth, G. A. *Journal of Physical Chemistry B* **2007**, *111*, 9931–9939, DOI: `10.1021/jp0708998`.

(194) Cukierman, S *Biophysical Journal* **2000**, *78*, 1825–34, DOI: `10.1016/S0006-3495(00)76732-4`.

(195) Dreyer, J.; Zhang, C.; Ippoliti, E.; Carloni, P. *Journal of Chemical Theory and Computation* **2013**, *9*, 3826–3831, DOI: `10.1021/ct400374n`.

(196) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *Journal of Cheminformatics* **2012**, *4*, 17, DOI: `10.1186/1758-2946-4-17`.

(197) Vanommeslaeghe, K.; MacKerell, A. D. *Journal of Chemical Information and Modeling* **2012**, *52*, 3144–3154, DOI: `10.1021/ci300363c`.

(198) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. *Journal of Chemical Information and Modeling* **2012**, *52*, 3155–3168, DOI: `10.1021/ci3003649`.

(199) Venable, R. M.; Sodt, A. J.; Rogaski, B.; Rui, H.; Hatcher, E.; MacKerell, A. D.; Pastor, R. W.; Klauda, J. B. *Biophysical Journal* **2014**, *107*, 134–145, DOI: `10.1016/j.bpj.2014.05.034`.

(200) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. *Journal of Computational Chemistry* **2013**, *34*, 2757–2770, DOI: `10.1002/jcc.23422`.

(201) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. Gaussian09, 2009.

(202) Giorgino, T. *Computer Physics Communications* **2014**, *185*, 317–322, DOI: `10.1016/j.cpc.2013.08.022`.

(203) Dutagaci, B; Becker-Baldus, J; Faraldo-Gómez, J. D.; Glaubitz, C *Biochimica et Biophysica Acta* **2014**, *1838*, 2511–2519, DOI: `10.1016/j.bbamem.2014.05.024`.

(204) Kučerka, N.; Nieh, M. P.; Katsaras, J. *Biochimica et Biophysica Acta* **2011**, *1808*, 2761–2771, DOI: `10.1016/j.bbamem.2011.07.022`.

(205) Carrer, D.; Schreier, S.; Patrito, M.; Maggio, B *Biophysical Journal* **2006**, *90*, 2394–2403, DOI: `10.1529/biophysj.105.074252`.

(206) Jurkiewicz, P.; Cwiklik, L.; Vojtíšková, A.; Jungwirth, P.; Hof, M. *Biochimica et Biophysica Acta* **2012**, *1818*, 609–616, DOI: `10.1016/j.bbamem.2011.11.033`.

(207) Paloncyéovaé Markeéta Vaévrovaé, K.; Sovovaé, Z.; DeVane, R.; Otyepka, M.; Berka, K. *Journal of Physical Chemistry B* **2015**, *119*, 9811–9819, DOI: `10.1021/acs.jpcb.5b05522`.

(208) Schmitt, U. W.; Voth, G. A. *Journal of Physical Chemistry B* **1998**, *102*, 5547–5551, DOI: `10.1021/jp9818131`.

(209) Kim, T.; Lee, K. I.; Morris, P.; Pastor, R. W.; Andersen, O. S.; Im, W. *Biophysical Journal* **2012**, *102*, 1551–1560, DOI: `10.1016/j.bpj.2012.03.014`.

(210) Ingólfsson, H. I.; Li, Y.; Vostrikov, V. V.; Gu, H.; Hinton, J. F.; Koeppe, R. E.; Roux, B.; Andersen, O. S. *Journal of Physical Chemistry B* **2011**, *115*, 7427–7432, DOI: `10.1021/jp200904d`.

(211) Hünenberger, P. H.; Reif, M. M., *Single-Ion Solvation: Experimental and Theoretical Approaches to Elusive Thermodynamic Quantities.* Theoretical and Computational Chemistry Series; Royal Society of Chemistry: Cambridge, 2011, DOI: `10.1039/9781849732222`.

(212) Marcus, Y., *Ion Sovation*; Wiley: New York, 1986.

(213) Brooks, B. et al. *Journal of Computational Chemistry* **2009**, *30*, 1545–1614, DOI: `10.1002/jcc.21287`.

(214) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174, DOI: `10.1002/jcc.20035`.

(215) Straatsma, T. P.; Berendsen, H. J. C. *Journal of Chemical Physics* **1988**, *89*, 5876–5886, DOI: `10.1063/1.455539`.

(216) Hummer, G.; Pratt, L. R.; García, A. E. *Journal of Physical Chemistry A* **1998**, *102*, 7885–7895, DOI: `10.1021/jp982195r`.

(217) Jackson, J. D., *Classical Electrodynamics*, 3rd ed; Wiley: New York, 1999.

(218) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. *Journal of Computational Chemistry* **2009**, *30*, 2157–2164, DOI: 10.1002/jcc.21224.

(219) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. *PLOS Computational Biology* **2017**, *13*, e1005659, DOI: 10.1371/journal.pcbi.1005659.

(220) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *Journal of Molecular Graphics and Modelling* **1999**, *16*, 213, DOI: 10.1016/S1093-3263(98)80006-5.

(221) Gomer, R; Tryson, G *Journal of Chemical Physics* **1977**, *66*, 4413–4424, DOI: 10.1063/1.433746.

(222) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Journal of Physical Chemistry B* **2006**, *110*, 16066–16081, DOI: 10.1021/jp063552y.

(223) Noyes, R. M. *Journal of the American Chemical Society* **1962**, *84*, 513–522, DOI: 10.1021/ja00863a002.

(224) Klots, C. E. *Journal of Physical Chemistry* **1981**, *85*, 3585–3588, DOI: 10.1021/j150624a013.

(225) Roux, B. *Biophysical Journal* **1996**, *71*, 3177–3185, DOI: 10.1016/S0006-3495(96)79511-5.

(226) Reif, M. M.; Hünenberger, P. H.; Oostenbrink, C. *Journal of Chemical Theory and Computation* **2012**, *8*, 3705–3723, DOI: 10.1021/ct300156h.

(227) Sitkoff, D.; Sharp, K. A.; Honig, B. *Journal of Physical Chemistry* **1994**, *98*, 1978–1988, DOI: 10.1021/j100058a043.

(228) Jensen, K. P. *Journal of Physical Chemistry B* **2008**, *112*, 1820–1827, DOI: 10.1021/jp077700b.

(229) Rashid, M. H.; Kuyucak, S. *Journal of Physical Chemistry B* **2014**, *118*, 707–716, DOI: 10.1021/jp410950h.

(230) Rashid, M. H.; Kuyucak, S. *Journal of Physical Chemistry B* **2012**, *116*, 4812–4822, DOI: 10.1021/jp300639x.

(231) Sultan, M. M.; Pande, V. S. *Journal of Chemical Physics* **2018**, *149*, DOI: 10.1063/1.5029972.

(232) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *Journal of Chemical Physics* **2006**, *125*, 1–15, DOI: 10.1063/1.2212942.

(233) Branduardi, D.; Gervasio, F. L.; Parrinello, M. *Journal of Chemical Physics* **2007**, *126*, DOI: 10.1063/1.2432340.

(234) Moradi, M.; Enkavi, G.; Tajkhorshid, E. *Nature Communications* **2015**, *6*, 1–11, DOI: 10.1038/ncomms9393.

(235) Comer, J.; Chipot, C.; González-Nilo, F. D. *Journal of Chemical Theory and Computation* **2013**, *9*, 876–882, DOI: 10.1021/ct300867e.

(236) Essex, J. W.; Jorgensen, W. L. *Journal of Computational Chemistry* **1995**, *16*, 951–972, DOI: `10.1002/jcc.540160805`.

(237) Venable, R. M.; Luo, Y.; Gawrisch, K.; Roux, B.; Pastor, R. W. *Journal of Physical Chemistry B* **2013**, *117*, 10183–10192, DOI: `10.1021/jp401512z`.

(238) Robinson, R. A.; Stokes, R. H., *Electrolyte solutions*; Courier Corporation: 2002.

(239) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *Journal of Computational Chemistry* **2011**, *32*, 2319–2327, DOI: `10.1002/jcc.21787`.

(240) Gowers, R. J.; Linke, M; Barnoud, J; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Dotson, D. L.; Domański, J; Buchoux, S; Kenney, I. M.; Beckstein, O *Proceedings of the 15th Python in Science Conference* **2016**, 98–105, DOI: `10.25080/majora-629e541a-00e`.

*"Nothing ventured, nothing gained."*