MOLECULAR INSIGHTS INTO
ALTERNATING ACCESS MECHANISM OF SECONDARY ACTIVE TRANSPORTERS
FROM MOLECULAR DYNAMICS SIMULATIONS

BY

JING LI

DISsertation

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biophysics and Computational Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Emad Tajkhorshid, Chair
Professor Klaus Schulten
Professor Deborah Leckband
Professor Claudio Grosman
Abstract

Membrane transporters are specialized molecular machinery for selective, regulable, efficient, and more importantly, active transport of diverse chemical species, e.g., nutrients, reaction precursors and products, and chemicals transmitting signals, across the cellular membrane. As a major class of membrane transporters, secondary active transporters couple vectorial translocation of one solute, typically Na\(^+\) or H\(^+\) ions, along its electrochemical gradient to uphill transport of their substrates. To fulfill its function, transporter operates via an alternating-access mechanism, in which it undergoes structural transitions between outward-facing (OF) and inward-facing (IF) states, to translocate substrate from one to the other side of the membrane. The alternating access mechanism has received substantial supports from recent structural studies. However, as an inherently dynamic process, several critical aspects of alternating access mechanism remain poorly understood solely on the basis of limited static snapshots of the transport cycle.

To address several key questions of alternating access model and to elucidate the underlying molecular mechanism at atomic level, we conducted computational studies mainly using molecular dynamics (MD) simulations on two secondary active transporters, i.e., bacterial Na\(^+\)-coupled glucose transporter (vSGLT), and the benzyl-hydantoin transporter (Mhp1). vSGLT presents the first IF structural state of a LeuT-fold transporter, and Mhp1 is the first secondary active transporter structurally resolved in both OF and IF states. Both of vSGLT and Mhp1 provided crucial information toward better characterization of the alternating-access mechanism. Based on the crystal structures of these two transporters, the computational studies elucidated several key functional molecular events in the transport cycle, and provided deeper insights of the underlying molecular
mechanism of alternating access model.

My early study on vSGLT identified the first ion-releasing state in the secondary active transporters. The crystal structure of vSGLT reports the transporter in its substrate-bound state, with a Na\(^{+}\) ion modeled in a binding site corresponding to that of a homologous protein, leucine transporter (LeuT). In repeated MD simulations, however, the Na\(^{+}\) ion is found instable, invariably and spontaneously diffusing out of the transporter through a pathway lined by D189, which appears to facilitate the diffusion of the ion toward the cytoplasm. Further analysis of the trajectories and close structural examination, in particular comparison of the Na\(^{+}\)-binding sites of vSGLT and LeuT, strongly indicates that the crystal structure of vSGLT actually represents an ion-releasing state of the transporter. Structural comparison of LeuT-fold transporters provides the first example in which we clearly see how global structural changes (tilting and shift of the helices) that take place during the transition between the IF and OF states, propagate into specific binding site of the ion (expansion of the site), thus, allowing the protein to release the ion into the solution. The observed dynamics of the Na\(^{+}\) ion, in contrast to the substrate, also suggests that the cytoplasmic release of the Na\(^{+}\) ion precedes that of the substrate, thus, shedding light on a key step in the transport cycle of this secondary transporter.

The second study on vSGLT is to characterize the next step after ion releasing, the substrate releases to cytoplasm. Employing both equilibrium and Steered MD (SMD) simulations, the pathway and mechanism of substrate unbinding from the IF state of the vSGLT have been investigated. During a 200-ns equilibrium simulation, repeated spontaneous unbinding events of the substrate from its binding site have been observed. In contrast to the previously proposed gating role of a tyrosine residue (Y263), the unbinding mechanism captured in our equilibrium simulation does not rely on the displacement and/or rotation of this side chain. Rather, the unbinding involves an initial lateral displacement of the substrate out of the binding site which allows the substrate to completely emerge from the region covered by the side chain of Y263 without any noticeable conformational changes of the latter. Starting with the snapshots taken from this equilibrium
simulation with the substrate outside the binding site, SMD simulations were then used
to probe the translocation of the substrate along the remaining of the release pathway
within the protein’s lumen and to characterize the nature of protein-substrate interactions
involved in the process. Combining the results of the equilibrium and SMD simulations,
a full translocation pathway is provided for the substrate release from the binding site
into the cytoplasm. The observed molecular events indicate that no gating is required for
the release of the substrate from the crystallographically captured IF structure of SGLT,
suggesting that this conformation represents an open, rather than occluded, state of the
transporter.

Although the alternating access mechanism successfully accounts for the efficient ex-
change of the primary substrate across the membrane, accruing evidence on significant
water transport and even uncoupled ion transport mediated by transporters has chal-
lenged the concept of perfect mechanical coupling and coordination of the gating mech-
anism in transporters, which might be expected from the alternating access model. I
and my colleagues performed a large set of extended equilibrium molecular dynamics
simulations on several classes of membrane transporters, e.g., vSGLT, Mhp1 and etc., in
different conformational states, to test the presence of the phenomenon in diverse trans-
porter classes and to investigate the underlying molecular mechanism of water transport
through membrane transporters. The simulations reveal spontaneous formation of tran-
sient water-conducting (channel-like) states allowing passive water diffusion through the
lumen of the transporters. These channel-like states are permeable to water but occluded
to substrate, thereby not hindering the uphill transport of the primary substrate, i.e.,
the alternating access model remains applicable to the substrate. The rise of such water-
conducting states during the large-scale structural transitions of the transporter protein is
indicative of imperfections in the coordinated closing and opening motions of the cytoplas-
mic and extracellular gates. We propose that the observed water-conducting states likely
represent a universal phenomenon in membrane transporters, which offers an expanded
understanding of alternating access mechanism.

At the heart of the transport mechanism of these secondary active transporters is the
coupling among species providing driving forces, substrate, and the conformational events during the whole transport cycle. To elucidate this crucial and long-term question, a serial of MD simulations were employed to study the impact of Na\(^+\)-binding on the structure and dynamics of Mhp1 in multiple functional states and on the transition between them. The results of microsecond-long equilibrium MD simulations suggest that Na\(^+\) binding stabilizes conformation favorable for the substrate binding in the OF state. Furthermore, the results of a special-protocol time-dependent biased simulation and subsequent free energy calculation for state transition, illustrate that Na\(^+\) binding increases the free energy barrier along the OF–IF transition. All the results suggest that Na\(^+\) binding reshapes the free-energy landscape of the ion/protein complex, thereby shifting the conformational preference toward a specific OF structure, which is favorable for substrate binding. The increased substrate affinity provided by Na\(^+\) binding facilitates uptaking of the substrate from its low-concentration environment by the transporter. The results, therefore, provide a deeper and more comprehensive understanding for the ion-coupling mechanism of secondary active transporters.
Acknowledgments

I would like to express my special appreciation to my adviser, Emad Tajkhorshid. Thank you for helping me to grow as a research scientist, and for your excellent guidance, caring, encouraging and supporting of my research. I would like to thank my collaborator as well as my committee member, Professor Deborah Leckband, for your valuable advice and insightful discussion about our project. I also appreciate the guidance of my other two committee members, Professor Claudio Grosman and Professor Klaus Schulten, and thank you for your brilliant comments and suggestions.

Many many thanks to all the current and past members of Emad Tajkhorshid lab, especially Giray Enkavi, Po-chao Wen, Mahmoud Moradi, Saher Shaikh and Zenmei Ohkubo, for your help, suggestions, discussions, and friendship during the last six years. My research would not have been possible without all of your help. Special thanks goes to Cindy Dodds, the administrative coordinator of our Center for Biophysics and Computational Biology. Thanks for your help, caring and patience from the first to the last day of my graduate study.

Many thanks to my friends in Champaign. Great friends make my life greater here. Finally, a thousand thanks to my family. Your great supporting was that sustained me thus far! Especially for my wife Ting. You always stood by me through the good times and the bad.

To my baby on the way, Yisu, you will be the best graduation gift from God for me.
# Table of Contents

## Chapter 1 Introduction to Membrane Transporters

1.1 Alternating Access Mechanism ........................................... 2
1.2 Secondary Active Transporters ........................................ 4
1.3 LeuT-fold Transporters .................................................. 5
1.4 Specific Membrane Transport Proteins Studied ........................ 7
  1.4.1 Na\(^+\)-coupled Glucose Transporter (vSGLT) ..................... 8
  1.4.2 Sodium-hydantoin Transporter (Mhp1) ............................. 9

## Chapter 2 Overview of the Methodology

2.1 Molecular Dynamics (MD) .............................................. 10
2.2 Nonequilibrium MD simulations ...................................... 13
2.3 Free Energy Calculations ............................................. 14
  2.3.1 Umbrella Sampling (US) ........................................... 15
  2.3.2 Bias-exchange Umbrella Sampling (BEUS) ........................ 15
  2.3.3 Reconstruction of the Potential of Mean Force (PMF) ........... 16
2.4 Preparation of Membrane Protein Systems .............................. 17

## Chapter 3 Ion Releasing from LeuT-fold Transporters

3.1 Introduction .................................................................. 19
3.2 Methods .................................................................... 21
  3.2.1 System Preparation .................................................. 21
  3.2.2 Simulation Protocols ................................................ 22
3.3 Results and Discussion .................................................. 23
Chapter 6  Ion-coupling Mechanism of Na\textsuperscript{+}-driven Transporter . . . . . 66

6.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

6.2 Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 69

6.2.1 System Preparation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 69

6.2.2 Conventional Simulation Protocol . . . . . . . . . . . . . . . . . . . . . 69

6.2.3 Exploration of OF↔IF Transition Pathway . . . . . . . . . . . . . . . . 70

6.2.4 Conformational Free Energy Calculations . . . . . . . . . . . . . . . . 71

6.3 Results and Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71

6.3.1 Na\textsuperscript{+} Binding Stabilizes OF State for Substrate Binding . . . . . 71

6.3.2 Na\textsuperscript{+} Binding Increases the Free Energy Barrier
along OF↔IF Transition . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 76

6.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 81

Bibliography . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82
Chapter 1

Introduction to Membrane Transporters *

All living cells rely on continuous exchange of diverse molecular species, e.g., nutrients, reaction precursors and products, and chemicals transmitting signals, across the cellular membrane for their proper function [4]. The phospholipid bilayer, the main constituent of the fluid-mosaic plasma membrane is only permeable to small and hydrophobic molecules. Continuous passage of a diverse set of polar and large molecular species, therefore, requires specialized membrane transport proteins [4,5]. Apart from small and/or highly hydrophobic molecules that can readily permeate lipid bilayers, membrane transport proteins are specialized molecular devices that provide the machinery for selective, regulable and efficient transport of most chemical species across the membrane. The transport proteins are classified into membrane channels and transporters, depending on their mechanism of transport.

Membrane Channels selectively facilitate passive diffusion of ions (ion channels), water

---

Membrane transporters are proteins that serve as active gate-keepers closely regulating the traffic of their substrates across the membrane. Similar to channels, membrane transporters provide selective pathways for permeation of their substrates across the membrane, but they are endowed the additional capability of actively pumping their substrates across the membrane, often against their electrochemical potential gradient \[5\]. These specialized molecular devices provide the machinery to intimately couple active transport to various forms of cellular energy. According to the source of cellular energy, membrane transporters are categorized into primary and secondary active transporters. Primary active transporters utilize light or the energy released by chemical reactions (mainly ATP hydrolysis), while secondary active transporters exploit the already established electrochemical gradient of specific solutes \[4\].

The fundamental role of membrane transporters in diverse biological and physiological processes has rendered them as important drug targets, further stimulating widespread interest in mechanistic studies of these proteins at a molecular level \[8\].

## 1.1 Alternating Access Mechanism

Central to efficient transport and to preventing dissipation of the substrate down the electrochemical gradient, is to ensure that the substrate-binding site will not be accessible to both side of the membrane at any given time, a widely accepted general mechanistic model for membrane transporters, termed the alternating access model \[9\]. It proposes that the transporter switches substrate accessibility from one side of the membrane to the other, through undergoing structural transitions between outward-facing (OF) and inward-facing (IF) states, temporarily residing in several possible intermediate states, and forming a transport cycle. The alternating access model supports the presence of two

(aquaporins) or other solutes through the membrane down their electrochemical gradient. Membrane channels, in general, fulfill their function with small-scale conformational change, e.g., gating motions \[5, 6\]. No requirement for large-scale conformational change generally allows channels to function at higher rates than transporters \[7\].
gates that open asynchronously, exposing the substrate binding site to only one side of the membrane at a time. This coordinated cycle in transporters contrasts with the general mechanism of channels, where usually opening of a single gate exposes the transmembrane pathway for substrate permeation to both sides of the membrane simultaneously. The mechanisms of transporters and channels are thus traditionally considered distinct from each other \[7\].

Accumulating evidence from biochemical kinetic, and structural studies have established that, various transporters operate via the alternating-access mechanism \[10\]-\[14\]. Especially recent rapidly-increasing crystal structures for several of these proposed states across various families of secondary active transporters, have provided substantial support for this mechanism \[14\]-\[17\].

**Figure 1.1:** The alternating access model of ion-coupled secondary active transporters. It includes two major states, outward facing (OF) and inward facing (IF) states, which can exist for both the apo and substrate-bound forms of the transporter. Substrate access from the two sides of the membrane is controlled by protein conformational changes.
1.2 Secondary Active Transporters

As a major class of membrane transporters, secondary active transporters use the energy stored in transmembrane electrochemical gradient of one solute to power translocation of their another, usually against the chemical gradient \([14, 18–21]\). A large number of secondary active transporters use the electrochemical gradients of ions across the membrane, most prominently \(\text{Na}^+\) or \(\text{H}^+\) ions for their function, and are, therefore, termed ion-coupled secondary active transporters \([15]\). The energy-coupling mechanism relies on a complex set of molecular processes within the transporter protein engaging many of its structural elements and its interactions with the transport species, rendering the dynamics of membrane transporters highly relevant to their function.

Figure 1.2: Two categories of membrane transport proteins: membrane channels and secondary active transporter. Two different transported solutes are respectively colored in red and orange.

At the heart of the transport mechanism of these secondary active transporters is the coupling among species providing driving forces, substrate, and the conformational events during the whole transport cycle \([14,17,22]\). Or in other words: how can the electrochemical potential of other ions facilitate the transport of substrate molecules against their concentration gradients \([17]\). Although in recent years, accumulating structural and biophysical studies have contributed substantial understanding to the structural basis of secondary active transporters, the question is still far from understood.
In spite of the fast growth of structural information of secondary active transporters, the alternating access mechanism, as an inherently dynamic process, also remains unclear solely on the basis of limited static snapshots of the transport cycle \[1]^{14}. The complete transport cycle involves many other sub–states of functional importance. For instance, it is known that both IF and OF states can exist at least in two other forms, namely the open or occluded sub–states. In the majority of structurally characterized transporters, only a few states could be experimentally captured. Sometimes the functional state represented by the structurally determined state is, however, not always easy to characterize. Specifically, it is sometimes difficult to judge whether the structure represents an open, closed, or semi-occluded state on the two ends (cytoplasmic and periplasmic/extracellular) of the transporter protein. In other words, it is not clear whether the substrate and/or the co-transported ions as well as water molecules have full, partial, or no access to their respective binding sites within the protein lumen.

Furthermore, although by combining the available crystal structures of different proteins in a family/superfamily we have achieved a uniquely well resolved description of the number of states involved in the transport cycle, the nature of the structural transitions involved, and how they underlie the transport mechanism, are still far from well understood \[14]. Another important mechanistic aspect regards the sequence of binding and unbinding events for the substrate and the co-transported ion(s) in the transport cycle which is poorly understood for the majority of secondary membrane transporters. Describing such mechanistic details based on static structures is rather challenging and calls for methodologies that can offer a dynamical treatment of the protein of interest.

### 1.3 LeuT-fold Transporters

Despite broad diversity in the transported substrates, a large number of secondary active transporters, surprisingly to be identified to bear significant architectural resemblance – two structural inverted repeats, each with a set of five transmembrane helices (TMs) oppositely oriented with respect to the membrane, and the first helix of each
repeat always unwound \([\text{14, 15, 17, 23}]\). These include leucine transporter (LeuT) \([\text{24–28}]\), bacterial \(\text{Na}^+\)–coupled glucose transporter (vSGLT) \([\text{29, 30}]\), the benzyl-hydantoin transporter (Mhp1) \([\text{31, 32}]\), betaine transporter (BetP) \([\text{33, 34}]\), carnitine transporter (CaiT) \([\text{35, 36}]\), dopamine transporter (DAT) \([\text{37}]\), \(\text{Na}^+\)–independent amino acid transporter (ApcT) \([\text{38}]\), arginine:agmatine antiporters (AdiC) \([\text{39–42}]\), and glutamate-GABA antiporter (GadC) \([\text{43}]\). The common architecture among these transporters has been termed the LeuT-fold architecture, due to the fact LeuT was the first among the group to be structurally characterized at high resolution \([\text{15}]\).

Figure 1.3: LeuT-fold transporters and their common architecture. (top panel) Topology diagram of LeuT-fold and view of the core structure parallel to the membrane. This image is from Krishnamurthy, H., et al., *Nature*, 459:347–355 \([\text{15}]\). Reuse by permission from Nature Publishing Group. Substrate and ions are depicted as yellow triangle and yellow circles, respectively. The large beige and blue triangles overlap the five-helix repeats related by the pseudo-two-fold axis of symmetry. (bottom panel) Various structurally resolved LeuT-fold transporters. Substrates and ions are shown in VDW representation, and proteins are shown in new-cartoon representation, with two five-helix repeats respectively colored in blue and purple.

Structural alignment of these transporters indicates a similar substrate–binding sites in LeuT, Mhp1, vSGLT and BetP, which in all cases, is located at the interface of the two inverted structural repeats \([\text{15, 23}]\). Sequence and structural comparisons of LeuT, Mhp1, vSGLT, BetP, DAT and ApcT also suggest a conserved cation–binding site which is named as the \(\text{Na}2\) site in LeuT \([\text{24}]\) and found to be critical for substrate binding and
symport 15, 22, 23, 38. All of these structural similarities suggest that the secondary active transporters with the LeuT-fold topology presents a growing superfamily that share a similar mechanism of transport. We refer to these transporters as “LeuT-fold transporters” hereafter.

As the prototype in this LeuT-fold transporters, LeuT is an amino acid transporter, and a bacterial homologue of the neurotransmitter sodium symporter (NSS) family 24. The NSS family includes transporters responsible for clearing the synaptic cleft of neurotransmitters, e.g., serotonin, dopamine, norepinephrine and GABA 44. This family thus includes important drug targets for treatment of neurological conditions such as depression, anxiety, and drug addiction 45–47. LeuT is the first to be structurally characterized in the NSS family, and also represents the first example of LeuT-fold transporters 24. LeuT thus serves as an important model to study the dynamics and mechanism not only for NSS members, but also for the whole superfamily of LeuT-fold transporters.

1.4 Specific Membrane Transport Proteins Studied

Elucidating the molecular mechanisms of alternating access model in secondary active transporters requires dynamical description of the transport process in atomistic detail. The wealthy structural information available in LeuT-fold transporters offers an unparalleled opportunity to establish a general transport mechanism for the entire LeuT-fold superfamily, and more importantly, to unlock the key molecular mechanisms of alternating access model using MD simulations.

In this work, our studies mainly focus on two LeuT-fold transporters: vSGLT 48, 50, and Mhp1 50. vSGLT presents the first IF structural state of LeuT-fold transporters 29, and Mhp1 is the first secondary active transporter structurally resolved in both OF and IF states 31, 32. In all of the studies based on these two transporters, we aimed to elucidate the conformational dynamics of the protein at different stages in transport cycle, and contribute deeper insights to the key aspects of the molecular mechanism of the alternating access model.
1.4.1 Na\(^{+}\)-coupled Glucose Transporter (vSGLT)

The solute sodium symporter (SSS) family consists of secondary active transporters that couple Na\(^{+}\) symport to the transport of a wide range of solutes, including sugars, amino acids, vitamins and inorganic ions [51]. SSS members play crucial roles in human health, and their malfunction is associated with various metabolic disorders [52]. Meanwhile, Na\(^{+}\)-coupled glucose transporter is also the best-documented case of a transporter exhibiting water cotransport, with several experimental reports discussing its characteristic leakiness to small species, such as water, and even ions [53–59].

vSGLT is the first structurally resolved SSS transporter in 2008 [29, 30]. Although it adopts the LeuT-fold, it is functionally divergent from LeuT. Currently two crystal structures of vSGLT are available. One is in an IF substrate-bound state, with a Na\(^{+}\) ion modeled in a binding site corresponding to that of a homologous protein, leucine transporter (LeuT) [29], and the other is crystallized in IF apo state two years later (2010) [30].

As the first IF structural state of a LeuT-fold transporters, the structure of vSGLT [29] provided crucial information toward better characterization of the alternating-access mechanism and the sequence of molecular events of the cytoplasmic-half transport cycle. Based on the first structure, MD studies of vSGLT have provided much insight into key aspects of the transport mechanism, i.e., its dynamics and transport mechanism [30, 48, 49, 60], the functional importance of structural element missing in crystal structure [61], as well as water permeation through vSGLT [50, 58, 59]. In MD studies, I have identified the first ion-releasing state of the transporter, and cytoplasmic release order for Na\(^{+}\) and substrate [48] (Chapter 3), characterization of the functional state of the crystallographically captured structure, and the substrate release mechanism [49] (Chapter 4), and water permeation through vSGLT as well as transient channel-like state in transporters [50] (Chapter 5).
1.4.2 Sodium-hydantoin Transporter (Mhp1)

Mhp1 is a Na\(^{+}\)-hydantoin transporter in the nucleobase-cation-symport-1 (NCS1) family, which mediates the uptake of nucleobase and related metabolites \(62\). Known substrates for the other NCS1 family transporters include allantoin, uracil, cytosine, purines, thiamine, pyridoxal-based compounds, and nicotinamide riboside \(62\). In terms of structure, Mhp1 belongs to the LeuT-fold transporters superfamily \(31,32\).

As the first secondary active transporter structurally resolved in both OF and IF states \(31,32\), Mhp1 becomes a key model for the secondary active transporters to study the alternating access mechanism, especially the OF–IF transition. Currently, Mhp1 has been structurally resolved in three major functional states, i.e., OF Na\(^{+}\)–bound state \(31\), OF Na\(^{+}\)/substrate–bound state \(31\), and IF apo state \(32\). Mhp1 bears significant architectural resemblance with other LeuT-fold transporters, including a similar substrate-binding site, and a conserved Na\(^{+}\)–binding site. This Na\(^{+}\)–binding site is named as the Na2 site in LeuT and found to be critical for substrate binding and symport in various LeuT-fold transporters \(15,23,38,63\). Different from the other Na\(^{+}\)–binding site, Na1 site in LeuT, the conserved Na2 site is further away from substrate-binding site (\(~6\) Å), less studied for the Na\(^{+}\)/substrate–coupling, but represent a more general coupling mechanism in LeuT-fold transporters \(22\). With the simplest Na\(^{+}\)/substrate stoichiometry (1:1) \(31\), Mhp1 presents a concise model for the coupling-mechanism studies.

In the MD studies, I have identified water permeation through Mhp1, and then proposed an expanded alternating access model \(50\) (Chapter \(3\)), and I used Mhp1 as a representative model to elucidate the general ion-coupling mechanism for secondary active transporters(Chapter \(6\)).
Chapter 2

Overview of the Methodology *

2.1 Molecular Dynamics (MD)

Molecular dynamics (MD) is a computational method to investigate the dynamics of molecules based on classical and statistical mechanics [64]. High spatial and temporal resolutions of MD simulations provide the ability to trace atomic motions of biological molecules [65] and have brought deep insights into molecular mechanisms that experimental approaches could not have achieved alone [6, 66–70].

The simulation system is composed of a collection of particles connected with bonds to reflect their chemical properties. MD simulations require a set of initial coordinates and velocities to start the integration of Newton’s equations of motion [64, 65, 71]. The initial velocities are generated randomly based on a Maxwell-Boltzmann distribution for a temperature, while initial coordinates for biomolecular simulations are either experimentally derived from crystallographic or NMR structures with atomic resolution, or obtained through a different modeling approach.

In an atomistic MD simulation, interactions are calculated between atoms using a set of parameters that define a potential energy function, representing a “force field”. The

*This section is partially based on a review article in Annual Reports in Computational Chemistry. Giray Enkavi, Jing Li, Po-Chao Wen, Sundarapandian Thangapandian, Mahmoud Moradi, Tao Jiang, Wei Han and Emad Tajkhorshid. “A Microscopic View of the Mechanisms of Active Transport Across the Cellular Membrane”. Annual Reports in Computational Chemistry, in press.
derivation of force field parameters is an iterative process, where initial parameters obtained from experimental data and quantum mechanical calculations \cite{72} are optimized to reproduce structure and vibrational modes, as well as thermodynamic properties of the molecular systems of interest. Various force fields are available for biomolecular simulations, with minor differences in their potential energy functions, and corresponding differences in parameters \cite{73-76}. A typical potential energy function for biomolecular simulations, $U$, includes terms that describe bonded (bonds, bond angles, dihedral angles, etc.) and non-bonded (van der Waals and electrostatic) interactions:

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{vdW} + U_{elec}$$

In non-bonded interactions electrostatic potential is calculated based on Coulomb’s Law

$$U_{elec} = k_e \frac{q_i q_j}{r},$$

where $k_e$ is Coulomb’s constant, $r$ is the distance between the two atoms with charges, $q_i$ and $q_j$. The van der Waals interactions are represented by the Lennard Jones potential:

$$V_{vdW} = \varepsilon \left[\left(\frac{R_{min}}{r}\right)^{12} - 2\left(\frac{R_{min}}{r}\right)^6\right],$$

where $\varepsilon$ is the well depth, $R_{min}$ is the distance at which the potential is at the minimum. Non-bonded interactions form the more expensive part of the calculations and often play a more important role than bonded interactions in describing interactions between different molecules, or even between different functional groups within the same molecule, e.g., side chain interaction within a protein. Thus non-bonded interactions can be separated into two parts, i.e., short-range and long-range interactions, and truncated with little loss of accuracy and great improvement in required computational time \cite{77}. In our simulations, short-range non-bonded interactions are always calculated using a cutoff distance of 12 Å, and long-range electrostatic interactions are usually calculated using the particle mesh Ewald (PME) method \cite{77}. 

11
The force, $\vec{F}$, acting on each atom is calculated as the negative gradient of the potential energy, $U$ with respect to its coordinates, $\vec{r}$:

$$\vec{F} = -\nabla U$$

The Newtonian equations of motion are then integrated \[64, 71\] to propagate positions, $\vec{r}$, and velocity, $\vec{v}$, of each atom in time, $t$:

$$\vec{F} = m\vec{a}; \quad \vec{a} = \frac{d\vec{v}}{dt}; \quad \vec{v} = \frac{d\vec{r}}{dt}$$

where $m$ is the mass and $\vec{a}$ is the acceleration. Integrating the forces forward in time results in a trajectory representing an ensemble of system configurations over time. Several thermodynamic and dynamic properties of the systems can be calculated from the ensemble of conformations obtained from MD simulations.

MD simulations require calculation of interactions between each pair of particles within the cut-off distance in the simulation system, making them highly dependent on developments of algorithms and computer technology for efficient performance \[78\]. Thus one of the major limits of MD for the biological studies is the accessible timescales. Many biological processes of interest take place in longer than millisecond timescales. However, simulations of large biomolecular systems are currently limited to microsecond timescales and most of the time even shorter. The timescale limitation usually results in inadequate sampling of the configuration space.

Besides, the use of simplified potential energy functions do not necessarily capture electronic properties, such as polarization effects. In order to solve this problem, “polarizable” force fields are being developed \[79–82\].

Despite these limitations, MD has been successfully employed in studying a wide range of biomolecular systems and phenomena, including membrane transport proteins and their mechanisms \[1, 6, 83–86\]. Furthermore, with continuous algorithmic improvement, availability of faster hardware, better force fields, and enhanced sampling techniques, the gap between simulations and experiments is rapidly closing, as evidenced by recent studies.
reporting simulations reaching timescales on the order of µs-sub-ms \[66–69,87\].

### 2.2 Nonequilibrium MD simulations

Many biologically relevant phenomena in biomolecular systems generally take place in timescales that are currently inaccessible to conventional equilibrium MD. To describe such long time-scale events, various nonequilibrium and biased simulations methods have been developed \[70,88–94\]. With regard to the large-scale conformational change between functional states, a novel approach has been developed by my colleagues recently \[70\] to investigate the transitions of membrane transporters using nonequilibrium work measurements as a search tool for practical biasing protocols \[70\]. The approach is primarily based on the application of nonequilibrium driven MD, a methodology that has been extensively employed in the past, most prominently in steered \[95\] and targeted \[91\] MD. In these nonequilibrium simulations, a time-dependent biasing potential is used to drive the system from an initial known state towards a final state along a reaction coordinate (collective variable), \(\zeta\), which could be, for instance, the center of mass of a group of atoms (steered MD), or the root mean square deviation (RMSD) from a target structure (targeted MD):

\[
U_{bias}(\zeta^t, t) = \frac{1}{2} k_{bias} (\zeta^t - \zeta(t))^2,
\]

where \(k_{bias}\) is the force constant, \(\zeta^t\) is the instantaneous value of the collective variable, and \(\zeta(t)\) is the scheduled target. The major distinction of our approach lies in the fact that we focus a major fraction of our computational effort on defining relevant collective variables, exploring the collective variable space, and optimizing the driving protocol. Instead of using simple reaction coordinates such as RMSD, we examine many combinations of various driving protocols, which are based on mechanistically relevant collective variables, for example an “orientation quaternion” \[70,96,97\] that describes a semi-rigid-body rotation in 3-dimensional space. The effectiveness of each transition protocol (pathway)
is assessed by the nonequilibrium work required to complete the transition:

\[ W(t) = \int_{t_0}^{t} \frac{\partial U_{bias}}{\partial t'} dt' = -k_{bias} \int_{t_0}^{t} (\zeta' - \zeta(t')) \dot{\zeta}(t') dt', \]

where \( W(t) \) is the work accumulated until \( t \), and \( \dot{\zeta}(t) \) is the velocity of the scheduled target. In this manner, we examine a large number of different driving protocols in short simulations repeated at least a few times. Convergence toward an optimal pathway is achieved by selecting a smaller number of the most promising pathways (protocols) that consistently result in lower nonequilibrium work and using them in longer driven simulations during the next iteration. After a few iterations, we arrive at an optimal protocol which is employed in extended simulations the following stage to generate a refined trajectory for the transition. As all of the initial simulations are nonequilibrium, it is important that they are repeated at least a few times. The optimal protocol and the associated refined trajectory can be then used to perform accurate free energy calculations.

The process of designing the driving protocols is mostly empirical and heavily relies on our knowledge of the system. Defining a set of reaction coordinates relevant to the transition of interest reduces the dimensionality of the configuration space and allows for a systematic optimization of the driving protocols. Although we cannot rule out the possibility of other transition pathways not sampled due to the use of particular biasing protocols, we can qualitatively assess the relative importance of the sampled transition pathways based on the nonequilibrium work measurements [70].

### 2.3 Free Energy Calculations

The calculation of free energy differences from MD is one of the main challenges in biomolecular simulations. Most of the time equilibrium MD simulations suffer from lack of sufficient sampling due to the system’s rugged energy landscape and the shortness of the time scale captured in MD.
2.3.1 Umbrella Sampling (US)

Umbrella sampling is an enhanced sampling method used to sample low probability events \[98, 99\] and calculate the free energy along a reaction coordinate, namely the potential of mean force (PMF):

\[
\Delta F(\zeta) = -k_B T \ln(\rho(\zeta))
\]

where \(\Delta F\) is the free energy, \(k_B\) is the Boltzmann constant, \(T\) is the temperature, and \(\rho(\zeta)\) is the probability distribution along a reaction coordinate of interest, \(\zeta\). In umbrella sampling, the reaction coordinate is divided into several windows. Each window is, then, simulated, independently with a different artificial biasing potential, often in harmonic form added to the system’s Hamiltonian:

\[
U_i(\zeta) = \frac{1}{2} k_i (\zeta - \zeta_i)^2
\]

where \(i\) is the index, \(k_i\) is the force constant, and \(\rho_i\) is the center of the window. The added bias is designed to make each window center a local minimum along the reaction coordinate, and consequently, improve sampling in a certain region where ergodicity is hindered by the form of the system’s energy landscape.

2.3.2 Bias-exchange Umbrella Sampling (BEUS)

Employing conventional umbrella sampling as described above to large-scale transitions is often challenging and may produce unreliable estimates for free energies, especially in large conformational changes. One of the methods to improve convergence of free energy estimations from the umbrella sampling simulations and to ensure the continuity of the phase space sampled is to employ a replica-exchange scheme \[100, 101\]. This scheme is referred as bias-exchange umbrella sampling (BEUS) (also as window-exchange or replica-exchange umbrella sampling \[101, 103\]). In bias exchange umbrella sampling, the individual windows, each with a different bias, are allowed to exchange potentials based on
a Metropolis criterion that preserves the detailed balance. The exchange probability is
\[ P_{ij}^{\text{exchange}} = \min(\exp(-\Delta E/k_BT), 1), \]
where \( \Delta E = (U_i(X^i_t) - U_i(X^j_t)) + (U_j(X^i_t) - U_j(X^j_t)) \),
i and \( j \) are indices of the windows, \( U_i(X) \) is the biasing potential for configuration \( X \) according to umbrella \( i \), and \( X^i_t \) and \( X^j_t \) represent two different configurations whose biases are attempted to be exchanged at time \( t \).

### 2.3.3 Reconstruction of the Potential of Mean Force (PMF)

To reconstruct the PMF, the biased probability densities calculated from each window can, then, be unbiased and combined using weighted histogram analysis method (WHAM) [104]. In WHAM, the following equations are solved iteratively to self-consistency:

\[
\begin{align*}
\rho(\zeta) &= \frac{\sum_{i=1}^{N_{\text{win}}} n_i(\zeta)}{\sum_{i=1}^{N_{\text{win}}} N_i e^{(U_i(\zeta))/k_BT}}, \\
f_i &= -k_BT \ln \left( \sum_{\text{bins}} \rho(\zeta)e^{-U_i(\zeta)/k_BT} \right),
\end{align*}
\]

where \( N_{\text{win}} \) is the total number of windows, \( n_i(\zeta) \) is the histogram counts, and \( U_i(\zeta) \) is the biasing potential in bin associated with \( \zeta \) and window \( i \), \( \rho(\zeta) \) is the estimate of the unbiased probability distribution.

One of the limits of the regular WHAM is that PMF can only be reconstructed as a function of the reaction coordinate biased in the umbrella sampling scheme. A generalization of both WHAM and Bennett acceptance ratio (BAR) [105], called Multistate Bennett Acceptance Ratio or generalized WHAM (GWHAM) [70], on the other hand, allows assignment of weights to individual configurations \( X^i_t \), thus, reconstruction of PMF as a function of any reaction coordinate [106, 107]. The weight of each configuration \( X^i_t \), \( p^i_t \), is determined by iteratively solving the following equations to self-consistency [106]:

\[
\begin{align*}
1/p^i_t &= \sum_j N_j f_j \exp(-\beta U_j(X^i_t)), \\
1/f_j &= \sum_i \sum_{t=1}^{N_i} p^i_t \exp(-\beta U_j(X^i_t))
\end{align*}
\]
in which \( U_i(X) \) is the biasing potential for configuration \( X \) according to umbrella \( i \). PMF
in terms of any reaction coordinate can, then, be constructed by measuring any property of
the configuration $X^i_t$ and reweighting it with $p^i_t$. GWHAM has been the preferred method
of analysis for BEUS, and regular WHAM for the conventional umbrella sampling.

## 2.4 Preparation of Membrane Protein Systems

In membrane protein simulation, such as what will be presented here, experimentally
solved atomistic protein structure is placed in explicit water, lipids (membrane), and
ions. The membrane, generated using the MEMBRANE BUILDER plugin of VMD [108], is
a patch of POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylethanolamine) for both
vSGLT and Mhp1 systems.

In order to mimic experimental conditions, MD simulations are often carried out under
NPT (constant-temperature (300–310 K) and constant-pressure (1 atm), constant number
of particles). For MD simulations using NAMD, constant temperature was maintained
by employing Langevin dynamics with a damping coefficient of 0.5 ps$^{-1}$. The Langevin
piston method [109,110] was employed to maintain a constant pressure of 1.0 atm with a
piston period of 100 fs. The CHARMM27 force field [76] were used for proteins, lipids,
ions, sugar molecule [111], and TIP3P model for explicit water [112]. All the simulations
were performed under periodic boundary conditions with a time step of 2 fs. Throughout
the simulations, bond distances involving hydrogen atoms were fixed using the SHAKE
algorithm [113].

The simulations involve a brief period of initial membrane equilibration (typically
1–5 ns), wherein the lipid tails are allowed to equilibrate while the lipid head groups
and the protein are constrained to their initial positions. This is then followed by an
unconstrained equilibration of the lipids and the protein for simulation times ranging
tens of nanoseconds to microseconds depending on the system and the problem. MD
simulations, presented in this work, were performed using NAMD [114] on supercomputers
from the Extreme Science and Engineering Discovery Environment (XSEDE), Blue Waters
and other campus clusters of University of Illinois, or Desmond on Anton [78]. The
Figure 2.1: MD simulation of a membrane protein system. In a typical MD simulation of membrane proteins, atomistic protein structure (ribbon representation) with bound substrate (VDW spheres) is placed in explicit water (transparent surface), membrane (POPE) (VDW spheres), and ions (VDW spheres).

Simulation trajectories were analyzed with VMD [108].
Chapter 3

Ion Releasing from LeuT-fold Transporters *

3.1 Introduction

Secondary membrane transporters couple electrochemical ionic gradients across the membrane to “uphill” translocation of their substrates. Despite diversity in sequence and substrate, accumulating evidence from biochemical, kinetic, and structural studies suggests that all ion-coupled transporters operate via “alternating-access” mechanism, in which the transporter alternates between two major conformational states, an IF and an OF one, thereby, switching the substrate access between the two sides of the membrane [10,11].

Due to the lack of high-resolution structures of the same transporter in multiple functional states, the nature of the structural transitions involved in the process and how they underlie the transport mechanism are largely unknown. Furthermore, only limited information is currently available on the sequence of binding and unbinding events during the transport cycle for most transporters [115][117]. Most of the crystal structures of Na⁺-coupled LeuT-fold transporter resolved before 2010, namely, LeuT [24][26], Mhp1 [31], and BetP [33], are either in the OF or in an intermediate “occluded” state. Capturing

the first IF conformation, the structure of a bacterial Na\(^+\)-coupled galactose transporter (vSGLT) takes a critical step towards a better characterization of the alternating-access mechanism in secondary active transporters \[29\].

![Figure 3.1: Overview of the structure of vSGLT and the simulation system.](image)

**Figure 3.1: Overview of the structure of vSGLT and the simulation system.** (A) The simulation system. vSGLT is shown in cartoon representation, with the bound substrate drawn in VDW. The POPE lipids and the ions in solution are also drawn in VDW, while the water is in surface representation. Some lipid molecules have been hidden from the view to provide a clear depiction of the protein. (B and C) The substrate binding site, viewed from within the membrane (B) and from the cytoplasm (C). Important side chains in the substrate binding site along with their individually colored TM helices are explicitly shown and labeled. (D) The proposed Na\(^+\)-binding site in crystal structure viewed from within the membrane. Residues in the Na\(^+\)-binding site are displayed as sticks.

Na\(^+\)/substrate stoichiometry in vSGLT has been identified as 1:1 \[118\]. However, in the reported crystal structure no Na\(^+\) binding site could be verified by cocrystallization of the protein with electron-dense ions \[29\]. Based on the structure of LeuT \[24\], another Na\(^+\)-coupled secondary transporters with significant architectural similarity to vSGLT, a Na\(^+\) ion was modeled in a “plausible” Na\(^+\) binding site at the intersection of TM1 and TM8 (Fig. 3.1D) \[29\]. The observation that the S365A mutation completely abrogates Na\(^+\)-dependent transport in vSGLT strongly supports the notion that the proposed bind-
ing site is indeed a Na\textsuperscript{+}-binding site \cite{29}. Moreover, several residues in this binding site, i.e., A62, I65, A361, S364 and S365, correspond to conserved residues in the Na\textsuperscript{+}-binding site of LeuT. However, whether this site is occupied by a Na\textsuperscript{+} ion in the IF conformation of vSGLT reported in the crystal structure remains an open question. The occupancy of this site is of high relevance to the sequence of events in the transport cycle of vSGLT. While the binding sequence of the substrate and ion(s) is characterized for vSGLT and hSGLT1 (human Na\textsuperscript{+}/glucose transporter) \cite{115-117}, the unbinding sequence of these species into the cytoplasm continues to be unknown.

### 3.2 Methods

#### 3.2.1 System Preparation

The simulation system was constructed by embedding monomeric vSGLT including the substrate galactose and one Na\textsuperscript{+} ion, taken from the RCSB Protein Data Bank (PDB ID 3DH4) \cite{29} into a lipid bilayer, as described in detail below. The titration states of ionizable amino acids (aspartate, glutamate, lysine, arginine, histidine, and tyrosine) were assigned based on pK\textsubscript{a} calculations using the PBEQ-Solver in CHARMM GUI \cite{119}. Coordinates for 6 missing residues (179 to 184) were built with the Psfgen plugin of VMD \cite{108} employing the CHARMM27 topology for proteins \cite{76,120,121}.

The first 46 N-terminal residues, including unassigned residues of helix (-1) (residues 3-19; helix numbering based on the recent convention used for inverted repeat symporters) and the unresolved loop (residues 20-46) were not included in the model. These residues do not belong to the core structure of the transporter, and, thus, are not expected to be essential for the mechanism. Water molecules were added using DOWSER \cite{122}, and the protein was solvated using the program Solvate \cite{123}. vSGLT was then aligned with the membrane normal using the OPM ( Orientations of Proteins in Membranes) database \cite{124}, and water molecules in the potential lipid protein interface were deleted. Solvated vSGLT was then inserted into a patch of POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylethanolamine) bilayer (100 × 100 Å\textsuperscript{2}) generated using the Membrane Builder.
plugin of VMD [108] with the membrane normal along the z-axis. The lipid molecules overlapping with the protein were deleted. Additional solvent was then added, and the system was neutralized with 100 mM NaCl using the Solvate and Autoionize plugins of VMD [108]. The final dimensions of the system are 100 × 98 × 95 Å³ including approximately 85,000 atoms.

### 3.2.2 Simulation Protocols

All simulations used TIP3P model for explicit water [112], and the CHARMM27 force field [76] for proteins (including the CMAP corrections [121]), lipids, ions, and the substrate - galactose [111]. All simulations were performed under periodic boundary conditions with a time step of 2 fs using NAMD 2.6 [114]. After an initial Conjugate Gradient minimization for 1000 iterations, lipid tails were melted in a 500-ps NVT simulation at 310 K in which all atoms except the lipid tails were fixed. The system was then equilibrated in an NPT (constant pressure/constant temperature) ensemble with a constant pressure of 1 atm for 500 ps during which all atoms of the protein, the substrate and the Na⁺ ion were constrained by harmonic potentials (k = 7.2 kcal/mol/Å²) to allow for packing of the lipids and water around the protein.

Following the 500-ps protein-constrained simulation, the system was simulated for additional 3 ns during which only heavy atoms of the protein, the substrate, and the Na⁺ ion were kept constrained. Then, another 3 ns of equilibration was performed with only the Cα atoms of the protein, the substrate, and the Na⁺ ion constrained. The final structure of this simulation was used to initiate six independent simulations (Na1-Na6) in which the system was simulated for additional 10 ns without any constraints.

Constant temperature was maintained by employing Langevin dynamics with a damping coefficient of 0.5 ps⁻¹. The Langevin piston method [109, 110] was employed to maintain a constant pressure of 1.0 atm with a piston period of 100 fs. Short-range non-bonded interactions were calculated using a cutoff distance of 12 Å, and long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) method [77].
3.3 Results and Discussion

3.3.1 Na\(^+\) Unbinding from vSGLT

In order to address the questions mentioned above, we performed several equilibrium simulations of a membrane-embedded model of vSGLT. Invariably, in all of the simulations, the Na\(^+\) ion unbinds spontaneously from the binding site within a few nanoseconds and diffuses into the cytoplasmic solution (Fig. 3.2). The reproducible, rapid unbinding of Na\(^+\) strongly suggests that the binding site does not provide an optimal configuration to retain the ion. In other words, the crystal structure appears to have captured an “open” Na\(^+\)–binding site.

![Figure 3.2: Spontaneous Na\(^+\) unbinding from its binding site in vSGLT.](image)

(A) Overview of the release trajectory of the Na\(^+\) ion. Residues in the Na\(^+\)-binding sites and D189 are displayed as sticks. TM1 (pink) and TM8 (blue) are the helices directly participating in the Na\(^+\)-binding site. (B) Displacement of the Na\(^+\) ion from its position in the crystal structure. The gray bar highlights the region in the vicinity of D189.

Interestingly, in all the simulations, after detaching from the binding site, the Na\(^+\) ion moves first to the region around D189, where it spends some time before completely leaving the protein (Fig. 3.2). It appears that the Na\(^+\) ion is somehow attracted by this residue, which is about 5 Å away from the Na\(^+\) ion in the crystal structure, and represents a highly conserved residue within the SSS family. The observed dynamics of the Na\(^+\) ion suggests that this residue might act as a “fishhook” above the Na\(^+\)–binding site facilitating the diffusion of the ion toward the cytoplasm in vSGLT. This residue, however, does not seem to constitute a binding site for the ion, as the ion only transiently interacts with it.
along its unbinding pathway.

The importance of D189 is supported by studies of homologous proteins, Na\(^+\)/proline transporter PutP and hSGLT1, where mutation of this residue results in complex effects ranging from significant decrease of the Na\(^+\)-coupled transport \[^{125}\], to even altered Na\(^+\)/H\(^+\) selectivity \[^{126}\].

### 3.3.2 Exploring Other Possible Na\(^+\)-binding Sites

In order to examine other possible Na\(^+\) binding sites that might be occupied in the crystal structure of vSGLT, a 10-ns simulation was done on a Na\(^+\)-free protein. Any site occupied with a Na\(^+\) ion under crystallization conditions would be expected to exhibit detectable, rapid local conformational changes in response to the absence of the ion in the simulation. Therefore, we should be able to identify potential Na\(^+\) binding sites through examination of fluctuations and structural deviation of individual side chains. The RMSD of the whole protein remains consistently below 2.5 Å in this simulation, indicating a stable conformation in the absence of Na\(^+\). The individual RMSD of the majority of residues in vSGLT is below 2.5 Å (Fig. S1).

The residues with higher RMSD values (F172, F195, F201, K285, F289, I408, L130, L137, L143, I438, F442 and the motif L103-P104-I105) are mostly facing outwards, e.g., facing the membrane (thus, irrelevant to specific Na\(^+\) binding) (Fig. 3.3). Furthermore, they are spatially too scattered to be able to cluster and form a Na\(^+\) binding site (Fig. 3.3). Note that coordination of a Na\(^+\) ion requires multiple ligands, usually provided by oxygen atoms from two or more residues in proteins. The substrate and Na\(^+\) have been suggested to be transported with 1:1 stoichiometry \[^{118}\] in vSGLT. Faham *et al.* report that no second Na\(^+\) binding site could be identified \[^{29}\], i.e., the site constituted by A62, I65, A361, S364 and S365 is probably the only ion binding site. In the Na\(^+\)-free simulation, these residues except I65 demonstrate RMSD values less than 1.5 Å, without any apparent major conformational changes, consistent with the notion that the site was likely not occupied by a Na\(^+\) ion in the crystal.
3.3.3 “Closed” and “Open” Na⁺–binding Sites

Our proposal that the Na⁺ binding site in vSGLT cannot stably bind Na⁺ in the IF conformation captured in the crystal structure is strongly supported by the comparison of the Na⁺ binding sites of vSGLT and LeuT. At the sequence level, the residues in the Na⁺ binding sites of vSGLT and LeuT are very similar. In LeuT [24], residues coordinating the ion, i.e., G20, V23, A351, T354 and S355, form a tight, square pyramidal arrangement around the Na⁺ ion, with Na⁺–O distances ranging between 2.1 and 2.4 Å (Tab. 3.1). All of the corresponding distances in the binding site of vSGLT, however, are above 3.1 Å (Tab. 3.1). In particular the distances between the Na⁺ ion and A62:O or S365:OG are more than 3.6 Å, with angles that are too irregular for an optimal Na⁺ coordination (Tab. 3.1).
Figure 3.4: Comparison of the “Closed” and “Open” Na\(^+\)-binding sites. (A) Comparison of the Na\(^+\)-binding sites of vSGLT and LeuT. (top) Superposition of the Na\(^+\)-binding sites of vSGLT and LeuT. Alignment was done using TM1 and TM8 helices from vSGLT and LeuT. Residues in the Na\(^+\)-binding site in vSGLT are drawn using darker colors. The TMs and the residue labels are blue for vSGLT and red for LeuT. The Na\(^+\) ion is yellow in vSGLT and orange in LeuT. (bottom) Comparison of the distances between the Na\(^+\) coordinating, oxygen atoms and the Na\(^+\) ion in the Na\(^+\) binding sites of vSGLT and LeuT. (B) Comparison of the Na\(^+\)-binding sites in OF and IF states of LeuT-fold transporters.

In such a configuration, stable ion binding is very difficult for the site in vSGLT to achieve. We note that for the Na\(^+\)-binding site of LeuT, Na\(^+\) was found to be completely stable in MD simulations of over 10s of nanoseconds \[127\]. Based on these results, we suggest that the Na\(^+\) binding sites of LeuT and vSGLT represent two distinct states, a “closed” and an “open” binding site, respectively. The configurational differences between these two binding site states seem to have originated from a shift in the interhelical angle of TM1 and TM8 in vSGLT (corresponding to TM1 and TM8 in LeuT), which becomes apparent by overlaying the structures of the two proteins (Fig. 3.4 A). The smaller interhelical angle of LeuT furnishes a closed binding site which binds Na\(^+\) tightly, whereas the larger angle in vSGLT results in the opening of the Na\(^+\) binding site. These conclusions are all reinforced by similar comparisons made with another transporter, Mhp1. The OF state of Mhp1 \[31\], similar to LeuT, exhibits a closed Na\(^+\)-binding site, while the IF state of Mhp1 \[32\] shows an open site as vSGLT (Fig. 3.4 B).
Table 3.1: Comparison of the bond distances and angles between the Na\(^+\) coordinating residues and the Na\(^+\) ion in the Na\(^+\) binding sites of vSGLT and LeuT.

<table>
<thead>
<tr>
<th>Bond</th>
<th>Distance (Å)</th>
<th>Bond</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G20(O)–Na</td>
<td>2.2</td>
<td>A62(O)–Na</td>
<td>3.6</td>
</tr>
<tr>
<td>V23(O)–Na</td>
<td>2.2</td>
<td>I65(O)–Na</td>
<td>3.3</td>
</tr>
<tr>
<td>A351(O)–Na</td>
<td>2.3</td>
<td>A361(O)–Na</td>
<td>3.2</td>
</tr>
<tr>
<td>T354(O(\gamma))–Na</td>
<td>2.3</td>
<td>S364(O(\gamma))–Na</td>
<td>3.1</td>
</tr>
<tr>
<td>S355(O(\gamma))–Na</td>
<td>2.4</td>
<td>S365(O(\gamma))–Na</td>
<td>3.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bond Angle</th>
<th>Angle (degree)</th>
<th>Bond Angle</th>
<th>Angle (degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V23(O)–Na–S355(O(\gamma))</td>
<td>99.3</td>
<td>I65(O)–Na–S365(O(\gamma))</td>
<td>60.5</td>
</tr>
<tr>
<td>V23(O)–Na–T354(O(\gamma))</td>
<td>112.8</td>
<td>I65(O)–Na–S364(O(\gamma))</td>
<td>133.0</td>
</tr>
<tr>
<td>T354(O(\gamma))–Na–S355(O(\gamma))</td>
<td>147.9</td>
<td>S364(O(\gamma))–Na–S365(O(\gamma))</td>
<td>73.5</td>
</tr>
<tr>
<td>G20(O)–Na–A351(O)</td>
<td>166.5</td>
<td>A62(O)–Na–A361(O)</td>
<td>150.6</td>
</tr>
</tbody>
</table>

3.4 Summary

Sequence comparison of LeuT, Mhp1, and vSGLT suggests that the Na\(^+\)-binding site proposed in the crystal structure of vSGLT is indeed a conserved ion binding site for divergent transporters and is critical for coupled substrate binding and symport [15]. Our study, however, strongly suggests that this site is in an ion-releasing state in the crystal structure. This finding has important ramifications in our structural view of the alternating access mechanism. Viewing the Na\(^+\) binding site of vSGLT as an “open” site, and comparing it with the corresponding “closed” sites in LeuT and Mhp1, we can deduce an important functional feature from the crystal structures of these proteins. Structural comparison of these transporters provides the first example in which we clearly see how global structural changes (tilting and shift of the helices) that take place during the transition between the IF and OF states, propagate into specific binding site of the ion (expansion of the site), thus, allowing the protein to release the ion into the solution. Furthermore, the rapid release of the Na\(^+\) ion within a few nanoseconds observed in all simulations, in sharp contrast to the substrate which did not exhibit any appreciable displacement from its binding pocket even after extension of one of the simulations to 200 ns, suggests that the cytoplasmic release of the Na\(^+\) ion precedes that of the substrate,
thus, shedding light on a key step in the transport cycle of this secondary transporter.

Figure 3.5: The substrate and ion binding and unbinding sequence in the alternating access model. It includes six major states. The Na\(^+\) ion precedes the substrate in both the periplasmic binding and the cytoplasmic release.
Chapter 4

A Gate-free Mechanism for Substrate Releasing

4.1 Introduction

Despite the fast progress over the last few years in structural biology of membrane proteins, the functional state represented by the structurally determined state is, however, not always easy to characterize. Specifically, it is sometimes difficult to judge whether the structure represents an open, closed, or semi-occluded state on the two ends (cytoplasmic and periplasmic/extracellular) of the transporter protein. In other words, it is not clear whether the substrate and/or the co-transported ions as well as water molecules have full, partial, or no access to their respective binding sites within the protein lumen. Furthermore, although by combining the available crystal structures of different proteins in the LeuT-fold family we have achieved a uniquely well resolved description of the number of states involved in the transport cycle of this family, the nature of the structural transitions involved, and how they underlie the transport mechanism, are still far from well understood. Describing such mechanistic details based on static structures is rather challenging and calls for methodologies that can offer a dynamical treatment of the protein


29
As the first IF structural state of LeuT-fold transporters, the structure of vSGLT \cite{29} provided crucial information toward better characterization of the alternating-access mechanism and the sequence of molecular events in these transporters. This functionally relevant feature turned out to be the first to be studied using MD simulations \cite{30,48,60}. In the first report, repeated equilibrium MD simulations together with detailed structural comparison of the Na\textsuperscript{+}–binding sites of vSGLT and LeuT were used to argue against the presence of an ion in the protein, proposing that the crystallographically captured structure represents an ion-free state \cite{48}. These results were consistently reproduced by simulations from other laboratories \cite{30,60} establishing the notion that the crystal structure is the state in which the ion binding site has sufficiently opened to allow the release of the Na\textsuperscript{+} ion \cite{16,23}, while the substrate is still bound to the transporter.

The next step along the transport cycle is the release of the substrate from its own binding site into the cytoplasmic solution. Based on the crystal structure, it is not easy to determine to what degree the substrate binding site is accessible from outside solution, especially after taking into account the effect of thermal fluctuation of the protein. In other words, it is not clear whether further protein conformational changes are required, and if so to what degree, for the release of the substrate into the cytoplasmic solution. In membrane transporters, the transition between major functional states is coupled to either global protein conformational changes, which usually involve several TM helices, e.g., transition between the IF and OF states of Mhp1 \cite{32}, or more localized gating motions of a few residues, e.g., the inward bending of the N-terminal half of TM10 in Mhp1 observed after substrate binding \cite{31}.

In the crystal structure \cite{29}, the substrate is bound about halfway across the membrane with its binding site flanked by hydrophobic residues on both the cytoplasmic and extracellular sides (Fig. 3.1 A). On the cytoplasmic side, Y263 from the broken helix TM6E stacks with the pyranose ring of the galactose (Fig. 3.1 B and C), a feature commonly found in a sugar-binding proteins \cite{128,129}. The position of this residue along a pathway that linearly connects the substrate to the cytoplasmic opening of the lumen triggered the
idea that Y263, together with the flanking residues Y262 and W264, might play the role of the cytoplasmic gate, and that the crystal structure represents a substrate–occluded state in which the exit of the substrate toward the cytoplasmic solution is blocked by Y263 [29]. Based on this mechanism, the release of the substrate from its binding site would rely on either large conformational changes of the TM region of the protein, or more likely, a gate–like, side–chain motion of Y263 and/or its neighboring residues that would further open the cytoplasmic lumen and increase the substrate accessibility. Two independent simulation studies in which the translocation of the substrate from its binding site toward the cytoplasmic solution was induced by either umbrella sampling or SMD have presented results supporting the notion that Y263 plays a gating role in the cytoplasmic lumen of vSGLT. In both simulations, the translocation of the substrate toward the cytoplasmic solution required was accompanied by side–chain rotation of Y263 [30,60].

Structural alignment of these transporters indicates a close similarity in the location of the substrate binding sites of LeuT, Mhp1, vSGLT and BetP, which in all cases, is formed at the interface of the two inverted structural repeats [15,23]. These structural similarities suggest that the galactose releasing from vSGLT might present a general mechanism of substrate releasing from LeuT-fold transporters.

4.2 Methods

4.2.1 System Preparation

The simulation system was constructed by embedding monomeric vSGLT including the bound substrate galactose, which was taken from the RCSB Protein Data Bank (PDB ID: 3DH4) [29] into a lipid bilayer, as described in detail below. The titration states of ionizable residues (aspartate, glutamate, lysine, arginine, histidine, and tyrosine) were assigned based on pK_\text{a} calculations performed using the H++ server [130], which resulted in a model in which all residues have their default titration states. Coordinates for the six missing residues (179–184) were constructed using the Psfgen plugin of VMD [108] employing the CHARMM27 topology file for proteins [76,120,121]. The first 46 N-terminal
residues, including unassigned residues (3–19) of helix –1 (helix numbering based on the recent convention used for inverted repeat symporters [15]) and the unresolved loop (residues 20-46) were not included in the model. These residues do not belong to the core structure of the transporter, and thus are not expected to be essential for the mechanism.

About 150 water molecules were added using DOWSER [122] to the protein. The first principal axis of the protein was aligned with the z axis using the OPM (Orientations of Proteins in Membranes) database [124]. Then, the system was inserted into a patch of POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylethanolamine) bilayer (100 × 100 Å²) generated using the Membrane Builder plugin of VMD [108] with the membrane normal along the z-axis. The lipid molecules overlapping with the protein were deleted. The system was then solvated using the program SOLVATE [123], and water molecules in the lipid-protein interface were deleted. The simulation system was then neutralized with 100 mM NaCl using the AUTOIONIZE plugins of VMD [108]. The final dimensions of the system before equilibration were 100 × 98 × 95 Å³ including ~85,000 atoms.

4.2.2 Simulation Protocols

All the simulations were performed using NAMD 2.6 [114], the CHARMM27 force field [76] for proteins, lipids, ions and the substrate galactose [111], and TIP3P model for explicit water [112]. All the simulations were performed under periodic boundary conditions with a time step of 2 fs. Throughout the simulations, bond distances involving hydrogen atoms were fixed using the SHAKE algorithm [113]. After an initial 1000 steps of minimization, lipid tails were melted in a 500–ps NVT (constant volume and temperature) simulation at 310 K, during which all atoms except the lipid tails were fixed. The system was then equilibrated in an NPT (constant pressure and temperature) ensemble with a constant pressure of 1 atm for 500 ps, during which all heavy atoms of the protein, the substrate and the bound Na⁺ ion were constrained by harmonic potentials (k = 7.2 kcal/mol/Å²) to allow for relaxation and packing of the lipid molecules and water around the protein. After a 500–ps NPT MD with heavy atoms constrained, and a 2.5–ns unconstrained NPT
simulation, the production run was performed for 200 ns in an NPₙT ensemble (constant area, temperature, and normal pressure) under equilibrium conditions.

Although two complete substrate unbinding events from its binding site were captured spontaneously during the equilibrium simulation, to probe the remainder of the exit pathway from the protein lumen into the cytoplasmic milieu, and to probe potential barrier regions along the unbinding pathway, four steered molecular dynamics (SMD) simulations were also performed, each starting from a different snapshot taken from the equilibrium trajectory (snapshots at 0, 89, 115, and 155 ns) representing typical different configurations of the substrate during the production run. We employed the protocol of constant–velocity (cv) SMD (cv-SMD [90,131]), with the arising force applied to the substrate’s center of mass toward the cytoplasmic side of the protein. The force was directed along the z axis (membrane normal) and was defined positive for pulling toward the cytoplasmic direction (Fig. 3.1 A). A force constant of \( k = 7 \text{ kcal/mol/Å}^2 \) and a pulling speed of \( \nu_z = 2 \text{ Å/ns} \) were used for the SMD simulations. The \( \text{C}_\alpha \) atoms of three remote residues from the translocation pathway (S313, I413, and F472) were fixed to prevent the overall translation of the system in response to the applied external force.

For all the MD and SMD simulations, constant temperature was maintained by employing Langevin dynamics with a damping coefficient of 0.5 ps⁻¹. The Langevin piston method [109,110] with a piston period of 100 fs was used to maintain the pressure at 1.0 atm. Short-range non-bonded interactions were calculated using a cutoff distance of 12 Å, and long-range electrostatic forces were described using the particle mesh Ewald (PME) method [77].

4.3 Results and Discussion

4.3.1 Capturing Spontaneous Substrate Unbinding

In order to investigate the dynamics of the IF state of vSGLT in its substrate–bound form, a 200–ns equilibrium MD simulation was performed on the membrane-bound model of the transporter. During the simulation, vSGLT maintains its IF state with the pro-
tein’s backbone RMSD invariably below 3 Å throughout the trajectory. We note that the initially bound Na\(^+\) ion reported in the crystal structure \(^{29}\) was released rapidly (within a few nanoseconds) from the protein into the cytoplasmic solution during the initial phase of the simulation. This feature has been uniformly observed in all previous simulations of vSGLT reported by our laboratory \(^{48}\), as well as those reported later by other groups \(^{30,60}\). As put forward earlier \(^{48}\), we associate this observation with the crystal structure of vSGLT representing a Na\(^+\)-free state, a hypothesis which is supported by structural comparison of the ion binding site in vSGLT and in other structurally known LeuT-fold transporters \(^{48}\), e.g., LeuT \(^{24}\) and Mhp1 \(^{31}\). Therefore, in the following discussion, we will refer to the state simulated in this study simply as the substrate-bound state.

In contrast to the protein, the substrate galactose shows significant fluctuations during the simulation, resulting in its occasional complete displacement from the binding site within the lumen of vSGLT, as monitored by the substrate’s heavy atom RMSD and the \(z\)-coordinate of its center of mass (Fig. 4.1A). Note that despite complete unbinding from its binding site, the substrate does not leave the protein lumen during the simulation, and after its apparently barrier-free, back and forth motion outside the binding site, it returns to the binding site region (Fig. 4.1 A and B). Nevertheless, the unbinding events from the binding site observed during the simulation clearly capture an initially curved (non-linear) exit pathway for the substrate. Of high relevance to the mechanism of transport in vSGLT is the fact that during the observed unbinding events only marginal conformational changes, even at the level of side chains, were observed to accompany the complete departure of the substrate from its binding site. This is particularly important with regard to Y263, which has been suggested to act as a cytoplasmic gate in vSGLT \(^{30,60}\). Based on the equilibrium unbinding events observed in our simulation, we propose that no gating motion is required for the release of the substrate from its binding site in the IF state of vSGLT as captured in the crystal structure \(^{29}\).

To better characterize the unbinding process, the dynamics of the substrate and its unbinding pathway during the two major unbinding events observed in the simulation
Figure 4.1: Spontaneous substrate unbinding in the equilibrium simulation. (A) Substrate’s heavy–atom RMSD (top) and its displacement along the membrane normal (bottom) during the 200 ns equilibrium simulation. The position of the substrate is shown using the maximum and minimum z-coordinate (gray solid lines) and the geometrical center (black solid line) of the substrate, and is compared to that of the geometrical center of the ring of Y263 (red solid line). (B) Snapshots showing the position and configuration of the substrate (VDW spheres) taken respectively at $t=0$ ns, 89 ns, 115 ns and 150 ns from the equilibrium simulations. The substrate is shown in VDW representation and residues in the substrate binding site are shown in overlaid stick and transparent surface representations. (C) Force-time profiles calculated for induced unbinding of the substrate from vSGLT in SMD simulations I–IV, seeded respectively from snapshots taken from the equilibrium simulation at $t=0$ ns, 89 ns, 115 ns and 150 ns. [49]
are described here in more detail. The first unbinding event (Unbinding Event I) was captured at $t = 81$ ns, at which point galactose adopts a position that is $\sim 7$ Å away from its original position towards the cytoplasmic side of the membrane, almost reaching the level of the plane of the aromatic ring of Y263 along the membrane normal ($z = \sim 4$ Å; Fig. 4.1A). The substrate remains in this position for about 15 ns (Fig. 4.1A).

Energetic analysis of the trajectory (Tab. 4.1) suggests that in the configuration achieved through Unbinding Event I, the substrate has the highest probability to exit the protein, since it experiences the lowest protein-substrate interaction energy and the highest water-substrate interaction energy. During Unbinding Event I (Fig. 4.1B) the substrate completely moves out of the original binding site, losing almost all of its original contacts with the binding site residues (Q69, E88, S91, N260, and K294), and instead establishes new contacts with several polar residues outside the binding site (N142, N267, T431), as indicated by the calculated contact frequency. During the substrate’s exit from the binding site, H–bonds with E68, T431, N142, and N267 appear to facilitate substrate unbinding. Most importantly, the substrate is no longer blocked along its path toward the cytoplasmic solution by Y263, which has been viewed as a “cytoplasmic plug” (Fig. 4.1B). Supporting the notion of the involvement of these residues in the exit pathway, previous studies have shown that the mutation of the residue corresponding to N142 in the homologous protein SGLT1 (K157A) impairs transport in oocytes [132]. The residue corresponding to T431 in vSGLT (T460 in SGLT1) was even suspected to be a substrate-binding residue since its mutation to cysteine altered sugar selectivity and decreased the affinity for glucose [133]. Interestingly, neither N142 nor T431 appears to contribute directly to the substrate binding site in the crystal structure [29]. The characterized unbinding pathway in our simulation, however, provides a molecular explanation for the importance of these residues and how they might assist the substrate with its unbinding and translocation toward the cytoplasmic side. Examination of the trajectory suggests that, the interaction of the polar side chains of these residues (N142 and T431) with the substrate might be important for facilitating the unbinding of the substrate and its translocation around the major obstacle of Y263. Mutation of either residue to a
nonpolar, or even less polar, side chain would therefore be expected to lower the chance of efficient substrate unbinding, as observed experimentally \cite{132,133}.

Table 4.1: Averaged interaction energies (kcal/mol) between the substrate and various subsets of the environment (protein, water, or individual residues) calculated for the unbound state obtained after Unbinding Event I (\(t = 81–96\) ns) and the Silent Phase (\(t = 130–200\) ns).

<table>
<thead>
<tr>
<th>Environment</th>
<th>Unbinding Event I</th>
<th>Silent Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein</td>
<td>(-31.90 \pm 3.51)</td>
<td>(-59.61 \pm 1.30)</td>
</tr>
<tr>
<td>water</td>
<td>(-32.96 \pm 5.26)</td>
<td>(-16.58 \pm 0.99)</td>
</tr>
<tr>
<td>N64</td>
<td>(-0.48 \pm 1.03)</td>
<td>(-1.62 \pm 1.20)</td>
</tr>
<tr>
<td>E68</td>
<td>(-15.48 \pm 4.41)</td>
<td>(-5.65 \pm 2.84)</td>
</tr>
<tr>
<td>Q69</td>
<td>(-0.16 \pm 0.47)</td>
<td>(-4.22 \pm 0.85)</td>
</tr>
<tr>
<td>E88</td>
<td>(1.28 \pm 0.32)</td>
<td>(-35.38 \pm 1.23)</td>
</tr>
<tr>
<td>S91</td>
<td>(-0.16 \pm 0.62)</td>
<td>(1.36 \pm 0.35)</td>
</tr>
<tr>
<td>N142</td>
<td>(-2.18 \pm 0.03)</td>
<td>(-0.75 \pm 0.01)</td>
</tr>
<tr>
<td>N260</td>
<td>(0.22 \pm 0.39)</td>
<td>(-1.67 \pm 1.14)</td>
</tr>
<tr>
<td>Y263</td>
<td>(-2.55 \pm 0.92)</td>
<td>(-1.81 \pm 0.65)</td>
</tr>
<tr>
<td>K294</td>
<td>(-1.18 \pm 0.31)</td>
<td>(2.27 \pm 1.56)</td>
</tr>
<tr>
<td>Q428</td>
<td>(-1.24 \pm 3.23)</td>
<td>(-1.38 \pm 0.06)</td>
</tr>
</tbody>
</table>

The displacement of the substrate out of its binding pocket involves a large lateral (parallel to the membrane plane) component, leading to an unbinding mechanism that is independent of the rotation of Y263. In other words, in the pathway captured in our simulation, the substrate goes around the side chain of Y263, an important detail that might be missed if the substrate is pulled linearly toward the cytoplasmic solution. As will be shown later by SMD simulations performed here as well as those reported in an a previous simulation study \cite{60}, the forced rotation of Y263 appears to be costly and can only be induced by applying rather large forces in the simulations (Tab. 4.2). In contrast, the unbinding mechanism captured in our simulation uses a pathway along which the substrate can circumvent the Y263 plug on its exit toward the cytoplasmic milieu.

A second major unbinding event (Unbinding Event II) is captured at \(t = 115\) ns, although in this case the substrate diffuses back into the binding site more rapidly (in \(\sim 1\) ns) than in Unbinding Event I. The displacement of the substrate from the binding site in Unbinding Event II is very similar to Unbinding Event I, as the substrate takes the same exact pathway out and reaches approximately the same level (\(z = \sim 4\) Å; Fig. 4.1A) within the protein lumen. The overall displacement of the substrate from its original
position in the binding site in Unbinding Event II measures to \( \sim 7 \) Å, and during the process it forms contacts with E68, N142, and N267. Similar to Unbinding Event I, the substrate is seen to have completely emerged out of the binding site and uncovered from Y263 along its exit pathway towards the cytoplasmic side (Fig. 4.1B). We note that the side chain of Y263 in either case does not undergo any conformational changes during the substrate unbinding (Fig. 4.3). In both unbinding events, E68 appears to play an important role in facilitating the exit of the substrate from its binding pocket. Based on the observed direct interaction with the substrate, we suggest that the mutation of E68 or corresponding residues in other SSS members, especially to an apolar side chain, will have a diminishing impact on the transport efficiency of SGLTs. This remains to be determined experimentally, since to our knowledge, no mutagenesis analysis of the effect of this residue on transport has been reported.

After the significant fluctuations and two complete unbinding events during the first 130 ns of the simulation, the substrate moves back into the binding site and stays there for the rest of the simulation (last 70 ns). The position and the pose of the substrate, however, are different from those observed in the crystal structure (Fig. 4.1B) [29]. In this state, which we refer to as the “Silent Phase”, the substrate is closer to the opening of the binding site (Fig. 4.1B) than in the crystal structure (Fig. 4.1B) and is stabilized mainly by N64, E68, Q69 and E88. We note that the term “silent” is used to only underline the reduced degree of fluctuation of the substrate during the last phase of the simulation. The transport cycle of SGLTs is approximated to be roughly on the order of tens of milliseconds [134, 135], providing ample time for such transient states, which might not necessarily be needed for the overall transport, to arise and disappear during the full cycle.

The charged side chains of E88 and E68, especially the former, have large contributions to the overall substrate–protein interaction energy in this phase, with E88 responsible for \( \sim -35 \) kcal/mol of the total interaction energy of \( \sim -60 \) kcal/mol (Tab. 4.1). E88 appears to be the main residue forming H–bonds with the substrate, and the number of H–bonds to this side chain shows a strong correlation with the stability of substrate in the trajectory.
Consistent with the major role of E88 in stabilizing the substrate in the binding site as an “anchor”, the mutation of this conserved residue has been shown to abolish Na$^+$–dependent galactose transport. Breaking from E88 is likely a high-barrier step in the unbinding of the substrate from the transporter. To exemplify the contrast between the bound state and the substrate’s poses achieved during the unbinding events, we note that after Unbinding Event I the E88–substrate interaction energy is dropped to below 1.5 kcal/mol (Tab. 4.1) clearly indicating the detachment of the substrate from the binding site.

4.3.2 Water Dynamics and Accessibility of the Substrate Binding Site

The accessibility of binding site to water is a prerequisite for successful substrate unbinding, and the degree of hydration and water dynamics can be viewed as strong measures for the degree of opening of the binding site. While Y263 along with its neighboring residues Y262 and W264 establish a plug-like structure that appears to prevent the linear displacement of the substrate from the binding site toward the cytoplasm, it does not occlude the binding site from water molecules, a finding more consistent with an open state of the binding site. During the simulation, water molecules readily reach the substrate binding site from the cytoplasmic side and form multiple H–bonds with the substrate. Water entrance into the binding site and hydration of the substrate facilitate the unbinding of the substrate by breaking protein–substrate H–bonds. The 80.0–81.5 ns segment of the trajectory provides a good example, during which the approaching water molecules compete with E88 for H–bonding to the substrate and gradually replace the original H–bonds between E88 and the substrate (Fig. 4.2). Since E88 is responsible for over than half of the protein–substrate interaction energy in the binding site, breaking of its H–bonds to the substrate is critical for freeing the latter from the binding site.
4.3.3 The Unbinding Pathway from the Binding Site to the Cytoplasm

Although the substrate exhibits full unbinding fully from its crystallographically determined binding pocket on several occasions during the equilibrium simulation, it does not completely leave the protein’s lumen during the simulated time scale. To explore the remainder of the unbinding pathway for substrate release, two independent SMD simulations were performed starting from the unbound states characterized during the equilibrium simulation, i.e., those obtained after Unbinding Event I and II (snapshots...
taken at $t = 89$ ns and 115 ns, respectively). In addition, for comparison of the forces required to induce the unbinding, we have also performed two SMD simulations starting from the fully bound state of the substrate, one starting from the initial state which is very close to the crystal structure ($t = 0$ ns), and the other starting from a snapshot representing the Silent Phase of the equilibrium MD simulation ($t = 150$ ns). During the SMD simulations, which are labeled as SMD–I, SMD–II, SMD–III, and SMD–IV, based on the time point of their starting configuration, the force is applied to the center of mass of the substrate along the $+z$-axis towards the cytoplasmic solution.

![Figure 4.3: Conformational dynamics of Y263 during the simulations.](image)

A major force peak ($\sim 1200$ pN) is obtained around $t = 2.8$ ns in the force-time profiles of both SMD–I and SMD–IV simulations (both starting from the bound states), which
can be clearly attributed to the blocking effect of the aromatic ring of Y263 along the
enforced displacement of the substrate linearly toward the cytoplasm (Fig. 4.1C). The
substrate passes the physical barrier of Y263 in SMD–I and SMD–IV in two different
ways; while in SMD–I, side chain rotation of Y263 allows the substrate to pass, in SMD–
IV substrate takes a pathway around Y263 without causing significant rotation of this side
chain (Fig. 4.3). After crossing the apparent barrier of Y263, no comparably large barriers
is observed during the remainder of the pulling simulations (Fig. 4.1C). These high-force
profiles have also been observed in a previously reported SMD simulation starting from
the bound state of the substrate in vSGLT [60]. A significantly smoother pathway and
lower force profiles are obtained for SMD–II and SMD–III simulations, since in the start-
ing configurations for these simulations the substrate has already completely unbound
from its binding site, and is no longer “blocked” by Y263 along the cytoplasmic translo-
cation pathway. No major barriers similar to those observed in SMD–I and SMD-IV are
observed for these simulations during the unbinding of the substrate (Fig. 4.1C). While
these results agree with the side chain of Y263 introducing a major obstacle along the
substrate exit pathway, they also clearly show that the substrate does not have to wait
for the conformational change of this residue to overcome the obstacle. During Unbinding
Events I or II, the substrate appears to have had time to explore the space available and to
have completely circumvented Y263 through lateral displacement, that is perpendicular
to the apparent exit pathway. As such, during the subsequent SMD-II and SMD-III simu-
lations, no conformational changes are observed in the rotameric state of Y263 (Fig. 4.3).
The overall structure of the protein remains intact during all of the SMD simulations,
and even in SMD–I and SMD–IV simulations where large forces are used to induce likely
artificial events, no major conformational changes of the whole protein is observed, sug-
gesting that the conformational state captured in the crystal structure [29] is very close
to, if not the same as, the IF open state.

Combining Unbinding Events I and II from the equilibrium simulation and the fol-
lowing smooth exit trajectories of SMD–II and SMD–III, we can describe the complete
release pathway for the substrate from the binding site to the cytoplasm (Fig. 4.4 and
Table 4.2: Comparison of the SMD simulations investigating the cytoplasmic substrate release pathway in vSGLT.

<table>
<thead>
<tr>
<th>SMD</th>
<th>equilibration time (ns)</th>
<th>SMD time (ns)</th>
<th>velocity (Å/ns)</th>
<th>direction $^a$</th>
<th>spring constant $^b$</th>
<th>Y263 blockade $^c$</th>
<th>maximum force (pN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0</td>
<td>10</td>
<td>2</td>
<td>$z$-axis</td>
<td>7</td>
<td>yes</td>
<td>$\sim$1300</td>
</tr>
<tr>
<td>II</td>
<td>89</td>
<td>10</td>
<td>2</td>
<td>$z$-axis</td>
<td>7</td>
<td>no</td>
<td>$\sim$600</td>
</tr>
<tr>
<td>III</td>
<td>115</td>
<td>10</td>
<td>2</td>
<td>$z$-axis</td>
<td>7</td>
<td>no</td>
<td>$\sim$600</td>
</tr>
<tr>
<td>IV</td>
<td>150</td>
<td>10</td>
<td>2</td>
<td>$z$-axis</td>
<td>7</td>
<td>yes</td>
<td>$\sim$1300</td>
</tr>
<tr>
<td>I $^d$</td>
<td>30</td>
<td>10</td>
<td>2</td>
<td>$z$-axis</td>
<td>7.14</td>
<td>yes</td>
<td>$\sim$1200</td>
</tr>
<tr>
<td>II $^d$</td>
<td>30</td>
<td>10</td>
<td>4</td>
<td>60° to $z$-axis</td>
<td>7.14</td>
<td>yes</td>
<td>$\sim$1100</td>
</tr>
</tbody>
</table>

$^a$ The $z$-axis is the membrane normal, with the positive direction being from the extracellular side to the cytoplasmic one as shown in Fig. 3.1 A.

$^b$ kcal/mol/Å$^2$.

$^c$ Whether or not the translocation takes place through the space occupied by the side chain of Y263.

$^d$ SMD simulations taken from Zomot’s report [60].

Figure 4.4: The cytoplasmic substrate release pathway and mechanism. Overview of the release trajectory of the substrate, obtained by combining the Unbinding Event I taken from the equilibrium trajectory (segment between $t = 80$ ns and $t = 83$ ns of the equilibrium simulation shown using pink spheres) with the unbinding trajectory of SMD–II (red spheres). The spheres specify the position of the C5 atom of the substrate in each frame. The protein is shown in tube representation with the residues lining the substrate’s release pathway displayed as sticks and labeled. [49]
Supplementary Movie). We note that the two exit pathways in SMD–II and SMD–III are essentially identical. During the equilibrium unbinding, residues E68, N142, T431, and N267 facilitate the initial lateral translocation of the substrate and its escape from the binding site. Subsequently, the substrate passes through a low-resistance open pathway toward the cytoplasm lined by S365, A63, N267, Y138, S368, S372, Y269, N371, R273, and T375, which are mostly located on TM6 and TM8 helices (Fig. 4.4). The majority of these polar or basic residues, most notably an array of equally spaced serine and threonine residues (S365, S368, S372, and T375) from TM8, form direct H–bonds with the substrate, thereby lubricating its passage along the exit pathway. Notably, several of these polar residues, namely, Y138, N267, R273, S368, N371, and S372, are highly conserved in Na\(^+\)–coupled sugar symporters of the SSS family, but replaced by nonpolar amino acids in the Proline:Na\(^+\) symporter from the same family (Fig. S5), supporting their importance in facilitating the translocation of sugar substrates. Finally, the characterized substrate exit pathway is through the putative Na\(^+\) binding site \[29\]. In other words, the lumen opening exploited by the substrate to exit the protein coincides with the region that has been proposed to constitute the putative Na\(^+\) binding site. As the substrate directly interacts with the residues of and around this putative Na\(^+\)–binding site, e.g., S365 and A63, it is conceivable that in the presence of a bound Na\(^+\) to this site, the pathway would be at least partially blocked. These observations are in line with our earlier conclusion that cytoplasmic release of Na\(^+\) precedes that of the substrate \[48\].

### 4.3.4 Mechanistic Implications

Characterizing the functional state of crystallographically solved structures of membrane transporters is not always straightforward. This is particularly important since significant effort and resources are invested in capturing the protein in different conformational states which might represent functional intermediates involved in the transport cycle. The currently available structures for membrane transporters are generally classified as either IF or OF states, and as open or occluded sub–states, primarily based on the degree of accessibility of the substrate binding site from the two sides of the membrane. A major difficulty
in this regard is that the degree of accessibility of cavities and openings within the protein mass and lumen cannot be always easily assessed based on static structures. A dynamical description of the structure that accounts for the ability of the lining residues to fluctuate and to form various types and degrees of interaction with incoming molecular species is necessary to know whether a particular conformation is representing a closed/occluded state, a semi-open state, or an open state.

Based on the position and intimate interaction of the substrate and Y263 in its binding site of the substrate-bound structure of IF vSGLT [29], a cytoplasmic gating role for this residue was proposed, and the structure was assigned to represent the IF–occluded state of the transporter. Starting from this structure, simulation studies employing SMD to induce the unbinding of the substrate [60], as well as a more recent study using equilibrium simulations [30] supported the gating role of Y263, since in both studies the exit of the substrate from its binding site was found to depend on the rotation of this residue.

In the equilibrium simulation presented here, we find a completely different mechanism and pathway for cytoplasmic unbinding of the substrate; the substrate main unbinding event is facilitated by its initial lateral displacement parallel to the membrane plane (lateral exit) and through a pathway that takes the substrate around Y263 rather than through it, resulting in an unbinding mechanism that is independent of the rotation of this side chain (Fig. 4.3). The observed unbinding pathway suggests that in the crystallographically captured structure [29], the substrate already has access to the space outside its binding site, that is, the binding site is in an open state. We note that an open state is in a much closer agreement with the very large fluctuation of the substrate in its binding site, which is consistently observed in all the reported simulations, including the previous two reports that support the gating role of Y263 [30, 60]. Furthermore, we argue that the recent crystal structure of IF vSGLT in its substrate–free form [30] is in closer agreement with a gate–free release of the substrate, since in this structure [30], the rotameric state of Y263 is unchanged compared to that in the substrate–bound state [29]. Taken together, the high degree of similarity of the binding site configurations in the two IF structures of vSGLT [29, 30], and the observed spontaneous unbinding of the substrate in
the absence of any conformational change in the residues lining the binding site observed in the simulation presented here, lead us to the conclusion that the crystal structure of the substrate-bound vSGLT \cite{29} represents an IF-open state rather than an IF-occluded state. We would like to note that the substrate might still take advantage of the space that might be occasionally and transiently vacated as a result of the rotation of the Y263 side chain, as observed in previous simulations \cite{30, 60}, but it does not rely on such a pathway and can take advantage of alternative, more permanently accessible routes for its initial unbinding from the binding pocket.

**Figure 4.5: The revised alternating access model.** It is not necessary to have either global conformational change or gating motion for the transition from substrate-bound to apo IF states.

### 4.4 Summary

Employing MD simulations, the pathway and mechanism of substrate unbinding from the IF state of the \(\text{Na}^+\)-coupled galactose transporter, vSGLT, have been investigated. During a 200–ns equilibrium simulation, repeated spontaneous unbinding events of the substrate from its binding site have been observed. In contrast to the previously proposed
gating role of a tyrosine residue (Y263), the unbinding mechanism captured in the present equilibrium simulation does not rely on the displacement and/or rotation of this side chain. Rather, the unbinding involves an initial lateral displacement of the substrate out of the binding site which allows the substrate to completely emerge from the region covered by the side chain of Y263 without any noticeable conformational changes of the latter. Starting with the snapshots taken from this equilibrium simulation with the substrate outside the binding site, SMD simulations were then used to probe the translocation of the substrate along the remaining of the release pathway within the protein’s lumen and to characterize the nature of protein-substrate interactions involved in the process. Combining the results of the equilibrium and SMD simulations, we provide a description of the full translocation pathway for the substrate release from the binding site into the cytoplasm. Residues E68, N142, T431, and N267 facilitate the initial substrate’s displacement out of the binding site, while the translocation of the substrate along the remainder of the exit pathway formed between TM6 and TM8 is facilitated by H–bond interactions between the substrate and a series of conserved, polar residues (Y138, N267, R273, S365, S368, N371, S372, and T375). The observed molecular events indicate that no gating is required for the release of the substrate from the crystallographically captured structure of the inward-facing state of SGLT, suggesting that this conformation might represent an open, rather than occluded, state of the transporter.
Chapter 5

Transient Formation of Water-conducting States in Membrane Transporters

5.1 Introduction

From a structural perspective of membrane transporters, the overall process of transport is furnished through transitions between two major conformational states, the OF and IF states, during which substrate accessibility is shifted from one side of the membrane to the other. Central to efficient transport and to preventing dissipation of the substrate down the electrochemical gradient, is to ensure that the substrate binding site will not be accessible simultaneously to both sides of the membrane, a mechanism best known as the alternating access model. This model relies on a high level of coordination between two gating mechanisms, a cytoplasmic and an extracellular one, exposing the substrate binding site to only one side of the membrane at a time. This key mechanistic aspect defines the main distinction of transporters from membrane channels, in which usually

opening of a single gate creates the transmembrane substrate permeation pathway to connect both sides of the membrane simultaneously [7]. The mechanisms of transporters and channels are thus traditionally considered distinct from each other [7].

Interestingly, various experimental studies have reported water transport and uncoupled ion transport for several membrane transporters [136, 138], a phenomenon that might appear to be contradicting the traditional understanding of the alternating access model. Water exchange mediated by membrane transporters has been attributed to two fundamentally different mechanisms: passive, osmosis-driven water exchange [55], and active, stoichiometrically related cotransport of water [136]. While the involvement of osmosis-driven mechanism has been well established [55], the role of active water cotransport is still a subject of debate and various models have been proposed to account for it [136]. These models include the carrier-mediated model, which relies on the uptake of a large number of water molecules along with the substrate into the binding region/lumen of the transporter and their simultaneous pumping to the other side of the membrane [54, 139, 140], and transport of water as a result of osmotic imbalance introduced by the local either inside the cavity or cytoplasm adjacent to the membrane, accumulation of the nonaqueous substrate [56, 141, 142].

A detailed investigation of the mechanism(s) for transporter-mediated water exchange across the membrane requires not only high resolution structures of membrane transporters, but also a dynamical description of the protein and transported species. Recent years have witnessed a sharp increase in the number of crystal structures of membrane transporters in different conformational states, thereby providing valuable insight into their mechanisms [12, 14]. While several of these structures are reported at resolutions high enough to gain information on the position of a few water molecules [24, 37, 143], none seems to offer a convincing mechanism for water transport. Thus it appears necessary to examine the dynamics of the protein as well as water to gain insight into the mechanistic details of the process. In this context, MD offers a powerful tool to capture conformational fluctuations of the transporter, which are closely coupled to formation of water permeation pathways, and to describe the dynamics of smaller species, such as
water, substrate, and ions during the structural breathing of the protein [6].

5.2 Methods

5.2.1 System Preparation

A summary of the studied transporters is provided in Table 5.1 [50]. All the transporter proteins were simulated in the explicit presence of membrane, water, and ions. The titration states of ionizable amino acids (aspartate, glutamate, lysine, arginine, histidine, and tyrosine) were assigned based on pK\textsubscript{a} calculations performed using the H++ server [130]. The protein was inserted into a patch of lipid bilayer generated using the MEMBRANE BUILDER plugin of VMD [108], with the membrane normal aligned along the z-axis. The lipid molecules overlapping with the protein were removed. The systems were then fully solvated using the program SOLVATE [123]. NaCl was used to neutralize the systems and to bring the ionic concentration to \( \sim 200 \text{mM} \).

<table>
<thead>
<tr>
<th>Transporter</th>
<th>Substrate</th>
<th>Family or State superfamily</th>
<th>Simulation time</th>
</tr>
</thead>
<tbody>
<tr>
<td>vSGLT</td>
<td>Na\textsuperscript{+}/glucose</td>
<td>SSS IF substrate-bound</td>
<td>1.15 ( \mu \text{s} )</td>
</tr>
<tr>
<td>Glt\textsubscript{ph}</td>
<td>Na\textsuperscript{+}/glutamate</td>
<td>SLC1 intermediate apo</td>
<td>160 ns</td>
</tr>
<tr>
<td>GlpT</td>
<td>glycerol-3-phosphate/P\textsubscript{i}</td>
<td>MFS IF substrate-bound</td>
<td>200 ns</td>
</tr>
<tr>
<td>Mhp1</td>
<td>Na\textsuperscript{+}/benzyl-hydantoin</td>
<td>NCS1 OF substrate-free</td>
<td>1.2 ( \mu \text{s} )</td>
</tr>
<tr>
<td>maltose transporter</td>
<td>maltose</td>
<td>ABC OF</td>
<td>100 ns</td>
</tr>
</tbody>
</table>

5.2.2 Simulation Protocols

Simulations were performed using NAMD 2.6 [114] or Desmond on Anton [78], using the CHARMM force field [76] for proteins, lipids, ions and substrate - galactose [111], and TIP3P model for explicit water [112]. All simulations were performed under periodic boundary conditions with a time step of 2 fs. Throughout the simulations, bond distances involving hydrogen atoms were constrained using the SHAKE algorithm [113]. After initial minimization of at least 1000 steps, all systems were simulated using the following
protocol: (1) 0.5 ns NVT simulation with all atoms constrained except for the acyl chains of the lipid molecules, in order to introduce a higher degree of disorder in the lipid tails; (2) simulation in an NPT ensemble with positional restraints applied to all protein and substrate atoms; and (3) equilibration in an NPT ensemble, without restraints. After the initial equilibration, the systems were subjected to production simulations in the NPT ensemble (Table 1). For MD simulations using NAMD, constant temperature was maintained by employing Langevin dynamics with a damping coefficient of 0.5 ps\(^{-1}\). The Langevin piston method \(^{109,110}\) was employed to maintain a constant pressure of 1.0 atm with a piston period of 100 fs. Short-range non-bonded interactions were calculated using a cutoff distance of 12 Å, and long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) method \(^{77}\). For MD simulations using Desmond on Anton \(^{78}\), the Berendson coupling scheme was employed to maintain a constant pressure of 1.0 atm, and long-range electrostatic interactions were computed using the k-space Gaussian split Ewald method, with a 64 × 64 × 64 grid.

### 5.2.3 Specific Analysis

Only full permeation events are counted, i.e., a water molecule has to traverse the entire membrane span in order to be counted as a permeation event. This is done by defining two planes parallel to the membrane and positioned at different points along the membrane normal, with one on the cytoplasmic side and the other on the periplasmic/extracellular side in a way that they cover the entire lumen region of the transporter. The planes are about 25-30 Å apart, depending on the transporter system. For a water permeation event to take place, a water molecule needs to have crossed the membrane and both planes, i.e., traveling from one side of the membrane all the way to the other side. The analysis is confined to a cylindrical region covering the lumen of the protein, and therefore excluding any possible water leak through the lipid bilayer. Repeating the analysis clearly shows that including half permeation events, i.e., those from the lumen region into one side of the membrane would only marginally affect the results. The reported numbers in the manuscript are, therefore, based on full permeation events. Since there is no osmotic
concentration gradient across the membrane in our simulation, we estimated the $P_f$ by using the theory developed by Zhu et al., in which $P_f$ is related to change in the square of the number of permeation events determined from an equilibrium simulation [144].

5.3 Results and Discussion

In this study, we have employed MD simulations to study a diverse set of membrane transporters (Table 5.1) with the primary goal of characterizing the mechanism of water transport. Extended equilibrium simulations of membrane-embedded structures are performed for transporters from the solute sodium symporter (SSS), solute carrier 1 (SLC1), major facilitator superfamily (MFS), nucleobase-cation-symport-1 (NCS1), and ATP binding cassette (ABC) families/superfamilies, namely, Na$^+$-coupled glucose transporter (vSGLT), glutamate transporter (Glt), glycerol-3-phosphate transporter (GlpT), benzyl hydantoin transporter (Mhp1), and maltose transporter, respectively. We note that for some of the transporter families studied here, the phenomenon of water transport has not yet been reported experimentally. We observe the formation of water-conducting states in all cases, describe their nature and mechanism of formation, and quantify the rate of water transport through these states. To our knowledge, this is the first report where the formation of water-conducting states has been captured in multiple transporter families, suggesting that the phenomenon might represent a universal aspect of membrane transporter function.

5.3.1 Water-conducting States of SGLT

The best-documented case of a water-conducting transporter is SGLT, with several studies reporting its characteristic leakiness to small species, such as water, and even ions [53][59]. Water transport has also been reported for Na$^+$-I$^-$ symporter (NIS) [140] and lactate transporter (MCT1) [145], both belonging to the same family (SSS) of transporters as SGLT. In addition, other LeuT-fold transporters, have also been reported to exhibit water transport [146][147]. The availability of a crystal structure for the substrate-bound IF
state of a bacterial SGLT (vSGLT) [29] provides the opportunity to study the molecular mechanism of water transport in a transporter family and structural topology that are well established experimentally to exhibit the phenomenon.

During a 1.15 $\mu$s MD simulation of vSGLT a large number of water permeation events were detected. On examining the origin and structural basis for water transport, we detected frequent formation of water-conducting states during the simulation (Fig. 5.1). A clear relationship exists between the degree of the opening of the water-conducting state at its narrowest point (the “constriction point”), and the number of permeating water molecules (Fig. 5.1C).

The water-conducting states are transient in nature, forming and disappearing frequently, thereby resulting in the fluctuation of the number of permeating water molecules (Fig. 5.1). The aqueous pathway in these states coincides with the putative substrate translocation pathway. This observation is in agreement with experimentally measured inhibitory action of phlorizin, a competitive inhibitor of substrate transport, on water flux, suggesting that these may share the same pathway [55]. Remarkably, the radius of the constriction point of the water pore formed within the lumen never reaches above 2.5 Å, which is larger than the radius of water ($\sim$1.15 Å), but smaller than that of the primary substrate ($\sim$4.0 Å), thus preventing it from leaking back down its chemical gradient.

During the simulation the protein maintains its IF state as in the crystal structure. This is indicated by the low structural deviation (RMSD < 2 Å) during the simulation (Fig. 5.1C) for the 10 transmembrane (TM) helices. Meanwhile, structural deviation of side chains lining the constriction point, namely Y87, F424, and Q428, and the resulting change in their distances, show a high correlation with the water transport (Fig. 5.1C). Y87 and F424 have been proposed to form the extracellular gate [29], and Q428 has been implicated in direct substrate binding [29]. Therefore, in addition to their role in controlling substrate access, these three residues appear to also act as the “gate” for water transport in the IF state of SGLT. The formation of the water-conducting states, in SGLT, therefore, is mainly a result of local conformational changes of a few side chains of these gating residues.
Figure 5.1: Water-conducting states and water transport in vSGLT. (A) Water-permeable state of vSGLT. The upper image shows the overall view of the water-conducting state and the continuous water channel formed within the transporter lumen. The protein is drawn in white cartoon, with some of the helices removed for clarity. Water molecules are drawn in both sticks and surface (red) representations, and the substrate in orange van der Waals (vdW). The three gating residues controlling water flow, namely Y87, F424, and Q428, are shown in sticks and colored by atom type. The lower image depicts a closeup view of the same conformation, highlighting the narrowest part of the water permeation pathway. Protein is shown in white surface, except for the three gating residues which are in cyan surface. (B) Water-impermeable conformation of vSGLT. The upper and lower views show the same as in (A) but for a water-impermeable state formed due to side chain motions of the gating residues (cyan surface) closing the narrowest point along the water permeation pathway. (C) Protein dynamics and water transport. The number of observed water permeation events in either direction (influx or efflux) is plotted against the simulation time in the top two panels. The third panel depicts the time series of the minimum radius of the aqueous pathway. The fourth panel shows the RMSD of gating residues and of the core transmembrane domain (10 TMs). The last panel shows the time evolution of the distances between the gating residues, i.e., Q428:N$_2$–F424:C$_2$, F424:C$_2$–Y87:C$_1$, and Q428:N$_2$–Y87:C$_1$. The vertical black dashed lines in C mark the snapshots shown in A and B, respectively, and the horizontal red dashed line is used to indicate the minimum radius needed to accommodate a water molecule. [50]

The osmotic permeability ($P_f$) of vSGLT estimated from these simulations is $4.75 \times 10^{-15} cm^3/s$, which is an order of magnitude higher than the experimentally measured value of $4.5 \times 10^{-16} cm^3/s$ for rSGLT [148]. It may be noted here that the $P_f$ values obtained from these simulations are only based on a partial view of the complete transport cycle, i.e., only the IF substrate-bound state, hence do not reproduce the average permeability of the complete cycle, reflected in the experimental $P_f$ value. It is expected that
increased sampling of more states would bring experiment and simulation results closer to each other. The larger gap between the $P_f$ values for vSGLT calculated from a shorter simulation (200 ns) and the experimental value is in line with this notion.

### 5.3.2 Water-conducting States in Other Transporter Families

Examining the dynamics of water in other transporter systems, we were surprised to find that the formation of water-conducting states is not unique to vSGLT. In fact, we observed water transport in four other transporter systems, namely, Glt$_{ph}$, GlpT, Mhp1, and maltose transporter (Figs. 5.2 and 5.3). Unlike vSGLT, water-conducting states in these transporters appear to be associated with the transition of the transporters from one major functional state to another. This was either due to the intermediate nature of the initial crystal structure or due to large structural changes induced during the simulations by mechanistically relevant molecular events as described below.

GlTs (also termed excitatory amino acid transporters (EAATs)) belong to the solute carrier 1 (SLC1) family of neurotransmitter transporters, which are responsible for the clearance of neurotransmitters from the synapse after their release from the presynaptic neurons. Several states of a bacterial GlT homologue (Glt$_{ph}$) have been structurally resolved, i.e., the OF, IF, and intermediate states. GlT is an experimentally well-documented case of a transporter mediating water exchange across the membrane during its function, and both osmotic-driven and cotransport mechanisms have been proposed for it.

Using the trimeric structure of Glt$_{ph}$ with one monomer in the intermediate state (and the other two monomers in the IF state), we have performed a 160 ns MD simulation of the apo (substrate/ion-free) state. These simulations have been performed under the exact same conditions used previously in our lab to simulate the OF and IF states of Glt$_{ph}$. During the simulation, numerous water permeation events occur along a continuous aqueous pathway formed at the interface between the transport domain (specifically helical hairpin 1; HP1) and the trimerization domain, but only in the monomer which is in the intermediate state (Fig. 5.2). The characterized pathway and
mechanism are consistent with a proposed mechanism for water transport in a human
GIT \[156\] and the crystal structure of Glt\textsubscript{ph} in the intermediate state \[153\], suggesting
that water transport through GIT is mediated via a putative Cl\textsuperscript{−} leak pathway of the
transporter. The radius of the constriction point in the intermediate Glt\textsubscript{ph} (\(\sim 1.5\) Å)
remains always smaller than that of the substrate (Fig. 5.2). Although the intermediate
nature of this monomer well accounts for the most part of the water permeation pathway,
进一步 conformational fluctuation of HP1 seems to be necessary for the formation of
the full water channel. No water exchange was observed in the other two monomers,
as well as in our previous simulations performed on trimeric states of either OF or IF
states in numerous different bound states. Altogether, these results strongly associate the
observation of the water-conducting state with the intermediate character of Glt\textsubscript{ph}.

The third studied membrane transporter is GlpT, one of the few structurally known
members of the MFS superfamily, the largest superfamily of secondary active trans-
porters \[159\]. Although no study implicating GlpT in water transport currently exists,
other MFS transporters, most prominently glucose transporters GLUT1 and GLUT2,
have been reported to exhibit water transport \[160,161\].

During a 200 ns simulation of substrate-bound, IF-state GlpT \[162,163\], we observed
copious water transport events (Fig. 5.2). Similar to vSGLT, water permeation in GlpT
occurs via an aqueous pathway running through the substrate binding site and the radius
of the constriction point remains smaller than that of the main substrate (Fig. 5.2). How-
ever, the scale of conformational changes preceding the formation of the water-conducting
state differs from vSGLT. It is important to note that water permeation in GlpT was only
observed following two functionally relevant events, namely, substrate binding, which
happened spontaneously in equilibrium simulations \[162,163\], and a subsequent proton
transfer reaction, which was modeled by the change of the protonation state of a conserved
histidine in the binding pocket. These molecular events are deemed to be necessary for
triggering/facilitating large-scale structural transitions of GlpT from the crystallographi-
cally captured state to other major states \[164,165\], the early stages of which seem to have
been captured in our simulation. These structural changes include the rearrangements of
Figure 5.2: Water-conducting states for Glt<sub>ph</sub> and GlpT. Glt<sub>ph</sub> (EAAT) of the SLC1 family (top panel), and GlpT from the MFS superfamily (bottom panel). In each panel, the leftmost figure depicts the topology of the protein using the crystal structure, highlighting in different colors functionally relevant domains and repeats. While Glt<sub>ph</sub> is simulated as a trimer, here only the monomer in the intermediate state is depicted, with the transport and trimerization domains colored in green and purple, respectively. The middle picture of each panel shows a representative water-conducting frame from the simulations. The charts on the right side of each panel depict time series for the number of water permeation events along the efflux and influx directions, radius of the narrowest part along of the aqueous lumen, and C<sub>α</sub>-RMSD of the TM region. The vertical black dashed lines represent the snapshots shown in the molecular images, and the horizontal red dashed lines show the minimal radius to accommodate a water molecule.

the two TM domains with respect to each other resulting in partial opening and closure in the extracellular and cytoplasmic halves, respectively, forming an intermediate state between the OF and IF states, exhibiting water transport. The results offer a putative mechanism of water transport for other MFS members as well, in particular for GLUT1 and GLUT2, where water transport has been demonstrated experimentally [160,161].

The next transporter studied here, Mhp1, is a Na<sup>+</</sup>-hydantoin transporter in the NCS1
family, which mediates the uptake of nucleobases and related metabolites \[62\]. Structurally \[31,32\], Mhp1 belongs to the LeuT-fold family of transporters. However, while other LeuT-fold transporters, e.g., SGLT1 and LeuT, have been experimentally characterized as water transporting proteins \[53,146\], until now, no water transport has been reported for the NCS1 family.

Transient water-conducting states were observed during a 1.2 µs simulation of the OF state of Mhp1 in its substrate-free form (Fig. 5.3). As in vSGLT and GlpT, the aqueous pathway overlaps with the putative substrate pathway, while the radius of the constriction point (\(~1.5\) Å) is consistently smaller than the substrate (\(~4.0\) Å) (Fig. 5.3), thereby preventing free diffusion of the substrate. Mhp1 is the another transporter discussed here for which the formation of water-conducting states is associated with large-scale structural changes of the protein, in contrast to more localized fluctuations which seem to be sufficient for the formation of water pores in vSGLT, even though it shares the same fold with Mhp1. A significant rearrangement of the cytoplasmic half of the protein appears to accompany the formation of the observed water-conducting state in Mhp1 (Fig. 5.3). This change is triggered by the spontaneous unbinding of the Na\(^+\) ion from its binding site during the equilibrium simulation. The presence of the Na\(^+\) ion has been demonstrated experimentally to directly affect the stability of the OF state in LeuT-fold transporters \[166,167\].

The last transporter examined for water transport is the \textit{E. coli} maltose transporter \[168\]. Distinct from the secondary active transporters described above, maltose transporter is a primary active transporter from the ABC transporter superfamily, with two TM domains forming the substrate translocation pathway (Fig. 5.3), whose conformational states are determined by ATP binding and hydrolysis within two cytoplasmic nucleotide-binding domains (NBDs) \[169,173\].

Starting from the crystal structure of maltose transporter in the OF state \[169\], a 100 ns simulation was performed after removing ATP from the two NBDs, as an efficient method to trigger the structural transition toward the IF state \[173,174\]. The removal of ATP leads to a rapid separation of the NBDs, which in turn translates into the opening of
Figure 5.3: Water-conducting states for Mhp1 and maltose transporter. Water-conducting states for Mhp1 of the NCS1 family (top panel), and maltose transporter of the ABC superfamily (bottom panel). The panels are arranged in the same manner as in Fig. 5.2, the crystal structure on the left, a representative water-conducting state in the middle, and time series for key events on the right. For clarity here only the TM domains and NBDs are depicted in maltose transporter.

the cytoplasmic ends of the closely coupled TM domains [173]. Since the periplasmic end of the transporter is already open in the crystal structure (OF state), the above-described molecular events result in the formation of a water-conducting state through which water permeation events were observed (t > 30 ns, Fig. 5.3). We note that in control simulations performed on the ATP-bound state, no large conformational changes or water permeation events were observed.

The maximal radius recorded at the constriction point of the aqueous pathway during the simulations is ~1.8 Å, which is sufficiently large to allow water exchange intermittently but not the translocation of the much larger transported substrate (maltose or maltodextrin). As was the case for most of transporters studied here, the aqueous pathway in maltose transporter runs through the substrate-binding site (Fig. 5.3). To our knowledge,
this study is the first to report water transport for a primary active transporter.

5.3.3 Mechanistic Features of Water Transport in Transporters

The results presented in this study illustrate the presence of a common molecular mechanism of water transport in diverse transporters. Earlier experimental studies on members of SSS, the neurotransmitter sodium symporter (NSS), SLC1, MFS, and other families have detected conduction for water and/or ions [55, 136–138, 146, 155, 160, 161, 175, 176]. Together with the results of the present study demonstrating the phenomenon in diverse transporter families, the accumulated evidence on leakiness of membrane transporters suggest that the coordination between opening and closing motions of the two gates in these molecular machines might not be as perfect as generally perceived. Water transport is mediated by well characterized water-conducting states (pores) most likely during the transition between major functional states. In all studied cases except one (GlT) the aqueous pathway forms along the putative substrate pathway. The observed formation of a water pore in a different region in Glt_{ph} (between the transport and trimerization domains) is an example highlighting the possibility that such permeation pathways can arise at any mechanically active interface due to various structural defects induced by large scale transitions of membrane transporters.

In the absence of well resolved structural and mechanistic descriptions for water and ion leak, the phenomenon might appear initially detrimental to the transporter function, since such leaky states could potentially dissipate substrate gradient across the membrane. Our results, however, clearly demonstrate how such water-conducting states can form frequently in response to a different structural fluctuations, without interfering with the tight control of the alternating access for the substrate. In all cases, the observed water-conducting states remain impermeable to the main substrate, thereby conforming with the required alternating access model, a principle that only needs to be obeyed by the “substrate”. While structural fluctuations of the protein, whether localized gating-like motions or more global transitions, mediate the formation of the water-conducting states, the constriction point within the transporter lumen is found to always remain
sufficiently small to prevent substrate from leaking. In other words, while the transient water-conducting states can be readily permeable to water, or even ions, they remain effectively occluded to the substrate, hence not expected to hinder the uphill transport function of the transporter.

The results presented here provide a structural basis only for a passive (osmosis-driven) mode of water transport, which is also the only mode of water transport consistent with the structures available for membrane transporters. These structures, particularly those captured in an occluded state, e.g., LeuT [143], vSGLT [29], Glt\textsubscript{ph} [153], and BetP [34], indicate that the occluded cavity is too small to accommodate more than a few water molecules to be transported along with the substrate in a carrier-mediated fashion. This low capacity differs by an order of magnitude from the number of water molecules transported per turnover in most transporters [136]. Nevertheless, since neither our simulations nor the available crystal structures can claim to have fully examined a transport cycle in its entirety, the involvement of an active (carrier-mediated) mode of transport cannot be discarded based on the currently available data.

Stoichiometric relation between the water flux and the substrate has been considered the primary evidence in support of the carrier-mediated model [136]. We note that the water-conducting states captured in the simulations can readily account for a high ratio of water/substrate transport. The amount of water transported through these leaky states is simply a function of their lifetime and average water permeability.

However, we note that these crystal structures or our simulations crystal structures do not represent the complete transport cycle, hence the possibility of the carrier-mediated model being applicable at a different stage of the cycle cannot be eliminated. Finally, we note that our simulations do not have any indication of a biased water flux across the membrane in either direction. The total amount of water efflux and influx are always almost the same as in all of the studied transporters (Figs. 5.1, 5.2, 5.3). This observation, which is expected given the absence of an osmotic gradient or hydrostatic pressure difference across the simulated membrane, also supports a completely passive mode of water transport in membrane transporters.
5.3.4 Universality of Water-conducting States in Transporters

One of the unique aspects of the present study is the demonstration of water-conducting states in diverse families and distinct functional states of membrane transporters: vSGLT and Mhp1 represent respectively a substrate-bound IF state and an OF state of LeuT-fold transporters, Glt\textsubscript{ph} and GlpT represent respectively an intermediate state of an SLC1 and an IF state of an MFS transporter, and maltose transporter is an OF state of ABC transporters. The simulations describe the phenomenon not only for several experimentally demonstrated cases \cite{55,136,138,146,155,160,161} but also for transporters for which water conduction is being proposed for the first time.

Although the formation of water-conducting states in vSGLT is mediated by mostly localized structural fluctuations, for the remainder of the studied transporters, the phenomenon appears to be correlated with more global structural changes, which are necessary for cycling through major functional states. This observation is in line with the notion that water-conducting states may generally arise as short-lived intermediates every time the protein undergoes a transition from one major state to another (Fig. 5.4).

Why would a transporter act like a channel? Membrane transporters face an extremely challenging task to accomplish in order to function effectively: a transporter needs to achieve highly coordinated gating motions at the cytoplasmic and extracellular gates, which are often spatially far from each other. Only a perfect coordination of the two gates can ensure the formation of a “truly” occluded state that would prevent even the leak of small species such as water. Given the soft mechanical properties of transporter proteins, it comes at no surprise to observe harmless imperfections in the overall gating motions, which manifest themselves in the formation of water-conducting states. It would, of course, be a concern if these channels were large enough to leak the substrate, and/or long-lived to allow very large amounts of smaller species to permeate across the membrane. Neither of these aspects appears to be the observed in our results, as the leaky states are only large enough for small species such as water. Furthermore, it appears that these states only transiently rise during the transport cycle, an attribute that might make them difficult to capture experimentally. Viewing the formation of these leaky
Figure 5.4: Formation of water-conducting states during the transport cycle. The core transport cycle of a membrane transporter (dark green region) relies on interconversion of the protein structure between major known functional states, namely, outward-facing open (OF-o), outward-facing occluded (OF-occ), inward-facing open (IF-o), and inward-facing occluded (IF-occ) states, which have been observed in various crystal structures [29,31,32,159,169]. An expanded view of the transport cycle (light green area) would involve a number of additional intermediates that arise during the transition between the major states, which due to their transient nature have not yet been structurally characterized by experiments, but can account for the uncoupled water and ion transport during the transport cycle.

states as a consequence of deviation of a transporter protein from a perfect, machine-like behavior, it is not surprising that such mechanical defects might not necessarily arise during each and every large-scale transition of every transporter. In fact, in several other transporter systems studied in our lab by comparably long simulations, we did not observe the formation of such water-conducting states. Furthermore, several experimental studies
have also illustrated the occurrence of water and ion leak only for specific states of the studied transporters [59, 136, 138, 177]. Depending on the protein architecture, degree of conformational changes involved in the cycle, and even the nature of the substrate, some transporters and some transitions might be more prone to the formation of water-conducting states.

5.4 Summary

Membrane transporters rely on highly coordinated structural transitions between major conformational states for their function, in such a perfect manner so as not to allow simultaneous access of the substrate binding site to both sides of the membrane, a mode of operation known as the alternating access model. Although this mechanism successfully accounts for the efficient exchange of the primary substrate across the membrane, accruing evidence on significant water transport and even uncoupled ion transport mediated by transporters has challenged the concept of perfect mechanical coupling and coordination of the gating mechanism in transporters which might be expected from the alternating access model. Here, we present a large set of extended equilibrium molecular dynamics simulations performed on several classes of membrane transporters in different conformational states, to test the presence of the phenomenon in diverse transporter classes and to investigate the underlying molecular mechanism of water transport through membrane transporters. The simulations reveal spontaneous formation of transient water-conducting (channel-like) states allowing passive water diffusion through the lumen of the transporters. These channel-like states are permeable to water but occluded to substrate, thereby not hindering the uphill transport of the primary substrate, i.e., the alternating access model remains applicable to the substrate. The rise of such water-conducting states during the large-scale structural transitions of the transporter protein is indicative of imperfections in the coordinated closing and opening motions of the cytoplasmic and extracellular gates. We propose that the observed water-conducting states likely represent a universal phenomenon in membrane transporters, which is consistent with their reliance
on large-scale motion for function.
Chapter 6

Ion-coupling Mechanism of Na\(^{+}\)-driven Transporter

6.1 Introduction

As a major class of membrane transporters, secondary active transporters use the energy stored in transmembrane electrochemical gradient of one solute to power uphill translocation of another \[14,18,21\]. A large number of secondary active transporters use the electrochemical gradients of various ions across the membrane, most prominently Na\(^{+}\) or H\(^{+}\) ions for their function, and are, therefore, termed ion-coupled secondary active transporters \[15\]. At the heart of the transport mechanism of these secondary active transporters is the coupling among driving forces provided by the ions, substrate, and the conformational events during the whole transport cycle \[14,17,22\]. Or in other words: how can the electrochemical potential of other ions facilitate the transport of substrate molecules against their concentration gradients \[17\].

In recent years the rapid increasing of high-resolution crystal structures of secondary active transporters provided considerable support to the alternating access mechanism, and improved our understanding substantially for the transporters at the atomic level \[13-\].

---

*This chapter is in preparation as a research article. Jing Li and Emad Tajkhorshid, “Mechanistic and Energetic Ion-coupling Mechanism of Na\(^{+}\)-driven Secondary Active Transporter”, in preparation.
In spite of the fast growth of structural information, our understanding of transport coupling mechanism is rather preliminary. For instance, there are a large number of structures crystallized for ion-coupled secondary active transporters in a superfamily, termed as LeuT-fold transporters, due to the fact leucine transporter (LeuT) was the first among the group to be structurally characterized. Currently, crystal structures are available for several transporters in this superfamily, i.e., LeuT, the Na+–hydantoin transporter (Mhp1), betaine transporter (BetP), in several major functional states. However, the coupling among ion, substrate and protein during the transport, is inherently a dynamic process, and is still far from understood solely on the basis of limited static snapshots of the transport cycle.

Meanwhile, various advanced biophysical approaches to study the dynamics of LeuT-fold transporters have offered valuable insights for the ion-coupling mechanism. Electron paramagnetic resonance (EPR) analysis suggests that Na+ binding increases accessibility of the extracellular vestibule and induces an OF conformation. Single-molecule fluorescence energy transfer (smFRET) studies show that higher Na+ concentration stabilizes OF state and decreases the overall frequency of transitions. Unfortunately, the underlying molecular mechanism is missing due to the resolution of these experimental approaches. At the same time, several computational studies using MD simulation also provided ideas from different perspectives for the coupling mechanism, i.e., secondary substrate binding allosterically triggers intracellular release of Na+ and substrate, substrate-binding affinity depends on the ion occupancy, ion-controlled conformational change from OF occluded to open state. Recently both smFRET and MD simulation studies estimated the free energy barrier along OF-IF transition, but there is a big discrepancy between them (80kJ/mol vs. 2-4k_BT). Despite all these, it remains unclear whether and how all of these points fit into the bigger picture, a clear, comprehensive, and convincing understanding for the ion-coupling transport mechanism.

To provide a deeper and more comprehensive understanding for the ion-coupling mechanism at atomic level, we employed several unconventional MD simulation approaches to
study the impact of Na\(^+\)-binding on the structure and dynamics of Mhp1 in multiple functional states and on the transition between them. First, microsecond-long equilibrium MD simulations were performed to investigate and compare the OF apo to OF Na\(^+\)-bound state, in order to reveal the underlying mechanism of how Na\(^+\) stabilizes OF states. Secondly, we would like to quantitatively characterize the free energy associated with the OF\(\leftrightarrow\)IF transition in the absence and presence of Na\(^+\), to elucidate the ion–coupling mechanism in terms of energetics. Characterization of the large-scale structural transitions of LeuT-fold transporters at an atomic level and associated free energy, is challenging both experimentally and computationally. In the present study, I used a novel computational approach developed in our group to search a reliable transition pathway \(^{70}\), and then performed free energy calculations, using umbrella sampling \(^{99}\) in conjunction with a replica-exchange scheme \(^{100}\), termed bias-exchange umbrella sampling (BEUS) \(^{70}\), to generate trustworthy free energy profiles.

In this study I would like to focus on Mhp1, a key model for the secondary active transporters with the LeuT-fold topology. As a Na\(^+\)-hydantoin symporter with the simplest Na\(^+\)/substrate stoichiometry (1:1) \(^{31}\), Mhp1 presents a concise case for studying the coupling-mechanism studies in LeuT-fold transporters. Currently, Mhp1 has been structurally resolved in three major functional states, i.e., OF Na\(^+\)-bound state \(^{31}\), OF Na\(^+\)/substrate–bound state \(^{31}\), and IF apo state \(^{32}\). Mhp1 bears significant architectural resemblance with other LeuT-fold transporters, including a similar substrate-binding site, and a conserved Na\(^+\)–binding site. This conserved Na\(^+\)–binding site is named as the Na2 site in LeuT \(^{24}\) and found to be critical for substrate binding and symport in various LeuT-fold transporters \(^{15\, 23\, 38\, 63}\). Different from the Na1 site in LeuT, the other Na\(^+\)–binding site in LeuT, the Na2 site, also the only Na\(^+\)–binding site in Mhp1, is further away from substrate-binding site (> 5 Å), less studied for the Na\(^+\)/substrate–coupling in LeuT, but represent a more general coupling mechanism in LeuT-fold transporters \(^{22}\).
6.2 Methods

6.2.1 System Preparation

The simulation systems were constructed by embedding Mhp1, which was taken from either substrate-free, Na\(^+\)–bound OF crystal structure (PDB ID: 2JLN), or apo IF crystal structure (PDB ID: 2X79), into a lipid bilayer, as described in detail below. The titration states of ionizable residues (aspartate, glutamate, lysine, arginine, histidine, and tyrosine) were assigned based on pK\(_a\) calculations performed using the H++ server [130], which resulted in a model in which all residues have their default titration states. The first principal axis of the protein was aligned with the z axis using the OPM (Orientations of Proteins in Membranes) database [124]. Then, the system was inserted into a patch of POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylethanolamine) bilayer (100 × 100 Å\(^2\)) generated using the MEMBRANE BUILDER plugin of VMD [108] with the membrane normal along the z-axis. The lipid molecules overlapping with the protein were deleted. The system was then solvated using the program SOLVATE [123], and water molecules in the lipid-protein interface were deleted. The simulation system was then neutralized with 100 mM NaCl using the AUTOIONIZE plugins of VMD [108]. The final dimensions of the system before equilibration were 98 × 98 × 100 Å\(^3\) including ~82,000 atoms.

6.2.2 Conventional Simulation Protocol

All the equilibrium simulations were performed using NAMD 2.6 [114] or Desmond on Anton [78], the CHARMM27 force field [76] for proteins, lipids, ions [111], and TIP3P model for explicit water [112]. All the simulations were performed under periodic boundary conditions with a time step of 2 fs. Throughout the simulations, bond distances involving hydrogen atoms were fixed using the SHAKE algorithm [113].

After initial minimization of at least 1000 steps, all systems were simulated using the following protocol: (1) 0.5 ns NVT simulation with all atoms constrained except for the acyl chains of the lipid molecules, in order to introduce a higher degree of disorder in the lipid tails; (2) simulation in an NPT ensemble with positional restraints applied to all pro-
tein and substrate atoms; and (3) equilibration in an NPT ensemble, without restraints. After the initial equilibration, the systems were subjected to production simulations in the NPT ensemble.

For MD simulations using NAMD, constant temperature was maintained by employing Langevin dynamics with a damping coefficient of $0.5 \text{ ps}^{-1}$. The Langevin piston method \cite{109,110} was employed to maintain a constant pressure of 1.0 atm with a piston period of 100 fs. Short-range non-bonded interactions were calculated using a cutoff distance of 12 Å, and long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) method \cite{77}. For MD simulations using Desmond on Anton \cite{78}, the Berendson coupling scheme was employed to maintain a constant pressure of 1.0 atm, and long-range electrostatic interactions were computed using the $k$-space Gaussian split Ewald method, with a $64 \times 64 \times 64$ grid \cite{78}.

Four microsecond-long equilibrium simulations were performed, in which three were in OF state, and the fourth was in IF state. The first one (Traj.1) was initiated from the substrate-free, Na$^+$–bound OF crystal structure of Mhp1 (PDB: 2JLN) \cite{31} without any constraint for Na$^+$ in the production simulation (1.2 µs). In the second trajectory (Traj.2) Na$^+$ was constrained in its binding site for all the 3 µs, and in the third one (Traj.3, 3 µs) the Na$^+$ ion was removed at the beginning. The fourth one (Traj.4) was initiated from the apo IF crystal structure of Mhp1 (PDB: 2X79) \cite{32}, and then performed for 1.2 µs.

6.2.3 Exploration of OF ↔ IF Transition Pathway

The exploration of the OF ↔ IF transition pathway used a novel approach elaborated in reference \cite{70}. The approach is primarily based on the application of nonequilibrium driven MD simulations. The mechanistically relevant collective variable space is explored by various time-dependent biasing protocols. The effectiveness of each protocol is assessed by the nonequilibrium work required to complete the transition. Select number of these nonequilibrium simulations were followed by restrained MD (RMD) simulations in which the system is subject to a time-independent biasing potential centered at the final target. Select number of the conformations resulted from the biased simulations were further
equilibrated with no bias to verify the complete transition. The optimized biasing protocol involves the collective variables that describe orientations of two transmembrane helices TMs 1 and 8 and the distance between them (Fig. 6.5 B). Specifically, we use the TM1–TM8 distance, defined as the distance between the center of mass (COM) of backbone atoms of the residues from A38 to I41 (TM1) and residues from A309 to T313 (TM8). As TM1 is unwound and there is a kink at the middle of TM8 (Fig. 6.5 B), the orientations of these two helices are described as four collective variables, i.e., the intracellular part of TM1 (TM1i, residues 30 to 40), the extracellular part of TM1 (TM1e, residues 44 to 54), the intracellular part of TM8 (TM8i, residues 319 to 329), the extracellular part of TM8 (TM8e, residues 302 to 312). This protocol then are used in multiple simulations for OF↔IF transition in both apo and Na\(^+\)-bound forms.

6.2.4 Conformational Free Energy Calculations

We also performed bias-exchange umbrella sampling (BEUS) (also known as window-exchange [102] or replica-exchange [101] umbrella sampling) simulations to quantify the free energies associated with OF↔IF transition. The centers and the force constants of the biases were optimized in short simulations to achieve \(~15\%\) exchange rate between replicas. For free energy calculations of both apo and Na\(^+\)-bound Mhp1, 14 replicas were used with initial conformations sampled from the extended nonequilibrium simulations for apo form and 18 were used for Na\(^+\)-bound form. Multistate Bennett Acceptance Ratio or generalized WHAM (GWHAM) [70, 106, 107], is used to reconstruct the PMF as a function of several coordinates.

6.3 Results and Discussion

6.3.1 Na\(^+\) Binding Stabilizes OF State for Substrate Binding

In order to investigate the impact of Na\(^+\)-binding on the conformation and dynamics of OF state, microsecond-long equilibrium simulations were performed initiated from the
substrate-free, Na\textsuperscript{+}–bound OF crystal structure of Mhp1 (PDB: 2JLN) \cite{31}. During a 1.2 µs equilibrium simulation of Mhp1 (hereafter referred to as Traj 1), the Na\textsuperscript{+} ion present in the starting crystal structure, spontaneously unbinds from its original binding site and diffuses into the extracellular side after 600 ns (Fig. 6.1). Note that the affinity of Na\textsuperscript{+} for Mhp1 in the substrate-free state, is not high ($K_d$ 1.15±0.28 mM) \cite{31}, spontaneous Na\textsuperscript{+}-unbinding can be expected to be captured in a simulation. Following Na\textsuperscript{+} unbinding, Mhp1 undergoes large-scale conformational changes for two Na\textsuperscript{+}–binding helices, i.e., TM1 and TM8, as well as the whole protein (Fig. 6.1). It reveals that the Na\textsuperscript{+}–binding plays a key role in stabilization of the conformation.

Figure 6.1: Dynamics and conformational changes for OF apo and Na\textsuperscript{+}–bound forms. The displacement of Na\textsuperscript{+} ion unbound from its binding site in Traj.1 is plotted against the simulation time in the top panel. The second and third panels respectively depict the time series of the RMSD of the 10 TMs (second panel) or TMs 1 and 8 (third panel) of the protein during three simulations. The fourth panel shows the distance between TM1 and TM8 in three simulations. It is measured as the TM1–TM8 distance defined in Method. The vertical black dashed line marks the snapshot that Na\textsuperscript{+} unbound from the Mhp1.

To compare the conformations and dynamical behaviors of the OF state of Mhp1 in
the presence and absence of Na\(^+\) ion, two additional simulations were performed, one with constraint of Na\(^+\) in its binding site all the time (Traj.2), the other with the Na\(^+\) ion removed at the beginning (Traj.3). Both of the simulations were performed for 3 \(\mu s\) to achieve enough sampling respectively for Na\(^+\)-bound and apo form for Mhp1 in its OF states. The comparison between Na\(^+\)-bound and apo forms shows that Na\(^+\)-binding holds two Na\(^+\)-binding helices, i.e., TM1 and TM8, together tightly, then stabilizes the core of the protein, 10 TMs (Fig. 6.1), in an OF conformation similar to the Na\(^+\)/substrate-bound structure (PDB ID:2JLO) \(^\text{31}\). In the absence of the Na\(^+\), the protein deviates away from the Na\(^+\)/substrate-bound structure and fluctuates significantly (Fig. 6.1).

The comparisons based on multiple coordinates confirm that Na\(^+\)-bound form is much more stable and similar to the substrate-bound structure. With both IF apo (PDB ID:2X79) \(^\text{32}\) and OF Na\(^+\)/substrate-bound (PDB ID:2JLO) \(^\text{31}\) crystal structures as references, the RMSD analysis for the core of the protein, 10TMs, clearly suggest that the ensemble of Na\(^+\)-bound form from Traj.3 is more localized in the configurational space similar to the OF Na\(^+\)/substrate structure, and with less fluctuation compared with the apo form from Traj.2 (Fig. 6.2). Upon Na\(^+\) binding, the helical angles between TM1e and TM8e, and TM3e and TM6e are also much stable and similar to the substrate-bound structure, indicating the Na\(^+\) binding could maintain the extracellular lumen in a conformation more favorable for substrate binding (Fig. 6.2). This is consistent with the EPR and smFRET data, suggesting that Na\(^+\) binding increases accessibility of the extracellular vestibule and stabilizes the OF states \(^\text{166,167,180}\).

More interestingly, Na\(^+\) binding also stabilizes the local conformation of the substrate-binding site, which is over 5 Å from the Na\(^+\)-binding site with no shared residues. In the presence of Na\(^+\) in its binding site, there is much less fluctuation of the substrate-binding residues than the apo form (Fig. 6.3 D and E). Na\(^+\) binding even maintains the side chains of the residues in the substrate-binding site similar to the substrate-bound structure (Fig. 6.3 B and C). As in the Na\(^+\)/substrate-bound crystal structure, the substrate is sandwiched by two tryptophans, i.e., W117 and W220, in the substrate-binding site, and forms pi-stacking interactions with both of the tryptophans, the distance between the
indole-rings of these two tryptophans is a good indicator of the binding-affinity of the substrate-binding site. Upon the Na\(^+\) binding, the distance is much more similar to that in the substrate–bound structure than the apo form (Fig. 6.3 D and E). All of these data indicate that Na\(^+\)–binding stabilizes an OF conformation with high affinity for substrate binding. This is consistent with the experiment that the affinity of benzyl-hydantoin to the protein is raised over 10-fold in the presence of saturated sodium [31].

Detailed analysis on our simulations shed light on the underlying molecular mecha-
Figure 6.3: Na\(^+\)-binding effect on the local conformation of substrate–binding site in OF state. (A). Conformation of substrate–binding site in OF Na\(^+\)/substrate–bound (PDB:2JLO) crystal structure. The substrate is shown in VDW representation, the substrate–binding residues are shown in stick, and the substrate–binding relevant helices are shown in cyan tube. (B) and (C). Conformational dynamics of substrate–binding site in Na\(^+\)-bound (B, blue) and apo (C, red) trajectories. The substrate–binding residues are shown in overlapped sticks from several snapshots for every 750 ns in respective trajectory. For Clarity only one typical conformation of the substrate–binding helices is shown in either blue (Na\(^+\)-bound) or red (apo) tube, with the Na\(^+\)/substrate–bound crystal structure shown in cyan. (D) and (E). Local Conformational fluctuations within Na\(^+\)-bound (D) and apo (E) trajectories. RMSDs of sidechains of substrate–binding residues are measured using OF Na\(^+\)/substrate–bound (PDB:2JLO) crystal structure as references. As an important indicator of the binding affinity, the distance between W117 and W220 is measured as that between the center of mass of two indole-rings. The distance in OF Na\(^+\)/substrate–bound (PDB:2JLO) crystal structure is shown as dashed line in (D) and (E).

nism of how Na\(^+\) binding stabilizes the protein and the substrate–binding site. The Na\(^+\) binding could lock the Na\(^+\)-binding site, and then restrain the conformation of TM 1
and 8 (Fig. 6.1). Dynamical network analysis [185] was also derived from MD simulations to describe the residue–residue dynamical correlation, and then to characterize the communication network within the protein. In the presence of Na\(^+\) ion in its binding site, there are much more allosteric interactions among TMs 1, 3, 6 and 8, holding these four substrate-binding helices together as a stable bundle (Fig. 6.4). In the absence of the Na\(^+\), there are much less connections among these helices, especially between TMs 3 and 6 (Fig. 6.4). Thus the presence of Na\(^+\) in its binding site not only locks the dynamics of TMs 1 and 8, but also holds the TMs 1, 3, 6, 8 and even the whole core part in a compact conformation to form more allosteric interactions.

![Figure 6.4: Dynamical network analysis for apo and Na\(^+\)--bound forms.](image)

**Figure 6.4: Dynamical network analysis for apo and Na\(^+\)--bound forms.** It is performed derived from Traj.2 and Traj.3 to describe the residue–residue dynamical correlation by using dynamical network analysis approach [185] for apo (left) and Na\(^+\)--bound (right) forms, and the allosteric interactions within network are shown as edges weighted by correlation data.

### 6.3.2 Na\(^+\) Binding Increases the Free Energy Barrier along OF↔IF Transition

In the Na\(^+\)--coupled symporter, Na\(^+\) should also play a key role during the OF↔IF transition. More importantly, as the driving force, how does the electrochemical Na\(^+\) gradients
affect the shape of the free energy profile associated with the transition is the central question to unlock the ion–coupling mechanism from the energetic perspective. Thus studying the Na\(^+\)–binding role during OF↔IF transition would be a critical step to achieve a comprehensive understanding of the ion–coupling mechanism. It not only requires full description of such structural transitions at atomic level, which is practically impossible in current experiments, but also needs continuous and enough sampling along the reliable transition pathway to characterize the free energy of IF↔OF transitions in different conditions. As large-scale conformational change, the OF↔IF transition is beyond the scope of conventional computational methods in terms of both timescale and complexity. Characterizations of both the transition pathway and free energy calculation continue to be extremely challenging computationally. In this study, we have used a novel computational approach [70], using extensive simulations to explore optimum biasing protocol for reliable transition pathway, and then performed BEUS [70] to accurately characterize the energetics of OF↔IF conformational changes.

To identify the optimum biasing protocol for characterizing the OF↔IF transition pathway, we have performed extensive MD simulations based on a wide range of mechanistically distinct collective variables, i.e., orientations, RMSDs and distances of different structural elements (as shown in Method). The optimum protocol is identified based on three criteria: (i) complete OF↔IF transition, (ii) nonequilibrium work relations, and (iii) functional relevancy. TMs 1 and 8 is the smallest structural element identified to trigger a complete OF↔IF transition (Fig. 6.6 A). Driving the local conformation for TMs 1 and 8 from OF to IF configuration, using either (i) RMSD or (ii) orientation of two helices and distance between them (Fig. 6.5 B) as collective variables, induces the global OF↔IF transition of the whole protein. Among all of the biasing protocols which could induce complete OF↔IF transition, one that consistently requires less work than others is using orientation of TMs 1 and 8 and distance between them (Fig. 6.6 B). More importantly, TMs 1 and 8 are the helices to form the binding site for Na\(^+\) ion, which provide the driving force for the functional transport cycle. In the equilibrium simulations for OF Na\(^+\)-bound state (Traj.2), OF apo state (Traj.3) and IF apo state (Traj.4), these three
Figure 6.5: Collective variables for non-equilibrium simulations and their functional relevance. Distribution of distance between TM1 and TM8 (A), angle between TM1e and TM8e (C), and angle between TM1i and TM8i (D) from three equilibrium ensembles: OF Na⁺–bound simulation (Traj.2), OF apo simulation (Traj.3), and an additional equilibrium simulation for IF apo state (1.2 µs). (B). Comparison of two Na⁺–binding helices, TMs 1 and 8, between OF Na⁺–bound (PDB:2JLN) and IF apo (PDB:2X79) crystal structures, to show the differences of the collective variables between OF and IF states.

collective variables present distinct distributions (Fig. 6.5 A, C and D). All of these data suggest that the local conformational change of TMs 1 and 8 is highly correlated with the global OF ↔ IF transition, and the biasing protocol based on the orientation of TMs 1 and 8 and distance between them could provide a reliable pathway to characterize the OF ↔ IF transition.

Then the same optimum biasing protocol is used to induce the OF ↔ IF transition respectively in apo and Na⁺–bound form. To compare the pattern and trend of work, rather than individual simulation profile, in these two conditions, we repeated the simulations several times to ensure the validity of our statements. In these simulations, the
Figure 6.6: The optimum biasing protocol for characterization the OF↔IF transition pathway. (A) Complete OF↔IF transition to reach a stable IF state. Evolution of backbone RMSDs using the optimum biasing protocol are measured for different structural elements by using IF apo crystal structure (PDB:2X79) (solid lines) or Na\textsuperscript{+}/substrate–bound crystal structure (PDB:2JLO) (dashed lines) as reference. This simulation includes three phases: 20 ns nonequilibrium simulation, 20 ns constrained simulation, and 10 ns without any constraint. (B) Nonequilibrium work analysis. Work profiles from nonequilibrium simulations which have induced complete transition in Na\textsuperscript{+}–bound form are shown in plot, and the work profile using the optimum biasing protocol is highlighted in red. (C) The comparison of nonequilibrium work for OF↔IF transition between apo (blue) and Na\textsuperscript{+}–bound (red) forms using the same optimum biasing protocol.

Non-equilibrium work for OF↔IF transition is reproducibly lower in apo than Na\textsuperscript{+}-bound form (Fig. 6.6 C). More importantly, based on the trend of work shown in Fig. 6.6 C, there is always a bump (between 5 and 7 ns) in non-equilibrium work profiles in Na\textsuperscript{+}–bound form, which results in more work than apo form. It is related with the breaking of the Na\textsuperscript{+}–binding site in the presence of Na\textsuperscript{+} ion. Thus non-equilibrium simulations by using the optimum biasing protocol provide an estimate of the free energy barrier along the OF↔IF transition, which suggest that Na\textsuperscript{+} binding increases the free energy for the OF↔IF transition.

To more accurately characterize the free energy for OF↔IF transition in apo and Na\textsuperscript{+}–bound forms, we performed bias-exchange umbrella sampling (BEUS) free energy calculation along the transition pathway identified from the nonequilibrium simulation using the optimum biasing protocol. Fig. 6.7 shows the reconstructed potential mean force (PMF) along three coordinates: distance between TM1 and TM8, angle between TM1\textsubscript{i} and TM8\textsubscript{i}, and angle between TM1\textsubscript{e} and TM8\textsubscript{e}. There exists a deepest minimum around 8.7 Å for the distance between TM1 and TM8, 67° for the angle between TM1\textsubscript{i} and TM8\textsubscript{i}, and ranging from 50° to 65° for the angle between TM1\textsubscript{e} and TM8\textsubscript{e} for the apo form (Fig. 6.7), which highly matches with the distribution of OF apo equilibrium state.
Figure 6.7: 2d PMF of apo and Na\textsuperscript{+}–bound forms along OF↔IF transition. The PMFs are obtained from BEUS MD simulations, respectively in apo (upper panel) or Na\textsuperscript{+}–bound form (lower panel), along three different coordinates: distance between TM1 and TM8, angle between TM1i and TM8i, and angle between TM1e and TM8e.

ensemble (Fig. 6.5 A, C and D). Another local minimum in apo form is around 10.5 Å for the distance between TM1 and TM8, 76° for the angle between TM1i and TM8i, which is also very similar to the IF apo state (Fig. 6.5 A and D). While in the Na\textsuperscript{+}–bound form, the deepest free energy basin is around 7.8 Å for the distance between TM1 and TM8, 70° for the angle between TM1i and TM8i, and 65° for the angle between TM1e and TM8e (Fig. 6.7), also highly similar to the OF Na\textsuperscript{+}–bound equilibrium ensemble (Fig. 6.5 A, C and D). As the free energy profiles are not converged yet, a better PMF to estimate the free energy barrier requires more sampling along the transition pathway.
6.4 Summary

In secondary active transporters, the electrochemical potential of ions across the membrane is used to fuel the “uphill” translocation of the substrate via the alternating access mechanism. The mechanism of this crucial coupling, however, remains unclear, despite significant recent experimental and computational studies. Mhp1, Na\(^+\)/Benzyl-hydantoin transporter, has become a key model for the secondary active transporters sharing the LeuT-fold topology. In the present study, we employed molecular dynamics (MD) simulations to study the impact of Na\(^+\)–binding on the structure and dynamics of Mhp1 in multiple functional states and on the transition between them. Microsecond-long equilibrium MD simulations suggest that Na\(^+\) binding stabilizes the substrate-binding conformation in the outward-facing (OF) state, thereby conferring high affinity for substrate binding. Furthermore, the results of a special-protocol time-dependent biased simulation and subsequent free energy calculation for state transition, illustrate that Na\(^+\) binding can increase the free energy barrier along the OF-IF transition. All the results suggest that cation binding reshapes the free-energy landscape of the ion/protein complex, thereby shifting the conformational preference toward a specific OF structure, which is favorable for substrate-binding. The increased substrate affinity provided by Na\(^+\) binding will facilitate capturing the substrate from its low-concentration environment by the transporter. The results, therefore, provide a deeper and more comprehensive understanding for the ion-coupling mechanism of secondary active transporters.

Figure 6.8: The hypothetical ion-coupling mechanism for Na\(^+\)-coupled symporters. The scheme shows how Na\(^+\) binding might change the free energy landscape along OF↔IF transition pathway, and provide an possible mechanism from energetic perspective.
Bibliography


