Scuola Internazionale Superiore di Studi Avanzati

Doctoral Thesis

Atomistic Study of Structural and Functional Properties of Membrane Proteins

Author:
Ina Bisha

Supervisors:
Alessandro Laio
Alessandra Magistrato

Physics and Chemistry of Biological Systems
October 16th, 2014

SISSA

Abstract

Physics Area
Molecular and Statistical Biophysics

Doctor of Philosophy

Atomistic Study of Structural and Functional Properties of Membrane Proteins

by Ina Bisha

Living cells exploit membrane proteins to carry out crucial functions like transport of nutrients, signal transduction, energy conversion, etc. Recently, the remarkable and continuous improvement of computational algorithms and power allowed simulating and investigating relevant aspects of the mechanisms of this important class of proteins.
In this thesis we focused on the study of two membrane proteins: a transporter and an ion channel. Firstly, we investigated the bacterial homologue of Sodium Galactose Transporter (SGLT), which plays an important role in the accumulation of sugars (i.e. glucose or galactose) inside cells, assuring a correct intestinal absorption and renal re-absorption. Using enhanced sampling techniques, we focused in understanding selected aspects of its transport mechanism. First, we identified a stable Na^+ ion binding site, which was not solved crystallographically. Second, based on the results of the first study, we investigated the mechanism of the binding/release of both ligands to/from the protein in the inward-facing conformation and their interplay during this process.
Finally, we also worked on another membrane protein: the Cyclic Nucleotide-Gated (CNG) channel. Using a chimera, the NaK2CNG mimic, we investigated the structural basis of the linkage among gating and permeation and of the voltage dependence shown by this channel. Large-scale molecular dynamics (MD) simulations, together with electrophysiology and X-ray crystallography, have been used to study the permeation mechanism of this mimic as a model system of CNG in presence of different alkali cations.

Contents

List of Figures

List of Tables

Abbreviations

MD	Molecular Dynamics
MTD	Metadynamics
BE-MTD	Bias Exchange Metadynamics
SGLT	Sodium-galactose transporter
vSGLT	Vibrio parahaemolyticus sodium-galactose transporter
hSGLT	human sodium-galactose transporter
Gal	galactose
COM	center of mass
SSS	Solute sodium symporter
LeuT	Leucine transporter
NIS	Na^+/I^- transporter
Mhp1	Benzyl-hydantoin transporter
BetP	Betaine transporter
CV	collective variable
FE	free energy
FES	free energy surface
VGCC	voltage-gated cation channels
cGMP	cyclic guanosine monophosphate
cAMP	cyclic adenine monophosphate
CNB	cyclic nucleotide binding

To my parents
Leta e Gori

Chapter 1

Overview

Membrane proteins are macromolecular machineries associated or embedded in cellular membranes playing crucial roles in biological processes like energy conversion, transport of ions and small molecules, signal recognition, transduction etc. [1]. A detailed structural and mechanistic characterization of this kind of proteins is nowadays much more difficult than that of the cytosolic ones, due to important experimental difficulties in their expression, solubilization, crystallization, etc. However, recent improvements in electrophysiological and crystallographic techniques, together with a remarkable and continuous improvement of computational tools, have greatly accelerated the progress in this field, allowing understanding of relevant aspects of the working mechanisms of these proteins.

In this thesis, we focused on two membrane proteins: the Sodium/Galactose Transporter (SGLT) and the Cyclic Nucleotide-Gated (CNG) ion channel, fundamental players in the transport of nutrients inside the cell and in the sensory transduction, respectively. Using enhanced sampling techniques (Metadynamis and Bias-Exchange Metadynamics) we investigated several important aspects of their mechanisms.

Sodium/Galactose Transporter

Sodium/galactose transporter (SGLT) is a secondary active symporter that accumulates sugars (glucose, galactose) inside the cells exploiting the electrochemical gradient of Na^+ ions across the membrane. This process is very important for the correct functioning of intestine and kidney and it has been hypothesized to occur via an alternating access mechanism [2, 3], in which the protein switches from an outward to an inward-facing conformation.

The crystal structure of a bacterial homologue of this transporter has been solved in 2008 [4]. This homologue shares a high sequence identity with the human protein (hSGLT) and translocates in the same direction one Na^+ ion and one galactose molecule (while hSGLT can use one or two ions depending on the isophorm). The structure revealed the symporter in the inward-facing occluded conformation with the galactose bound. Even if Na^+ ion was not detected by the X-ray experiments, a plausible binding site was proposed by a structural comparison with other transporters that share the same mechanism and fold [5]. However, recent molecular dynamics (MD) simulations studies suggested that the crystal structure represents an ion-releasing state of the transporter [6–8]. Indeed, after a few ns of MD simulation, the Na^+ spontaneously diffuses outside the protein, towards the intracellular space.

As a first goal of this thesis we focused on identifying a stable binding site for the ion in the inward-facing conformation, namely an 'ion-retaining' state.
Performing classical MD and metadynamics (MTD) [9, 10] simulations, we explored the free energy surface (FES) and the conformations involved in the binding and dissociation of the Na^+ ion. In Figure 1.1, we report the FES as a function of the coordination number of the ion with the five closest residues in the crystal binding site, and the distance between Na^+ ion and a residue seen to interact with the ion along its exit pathway, D189. Among the several binding sites detected, we identified the most stable

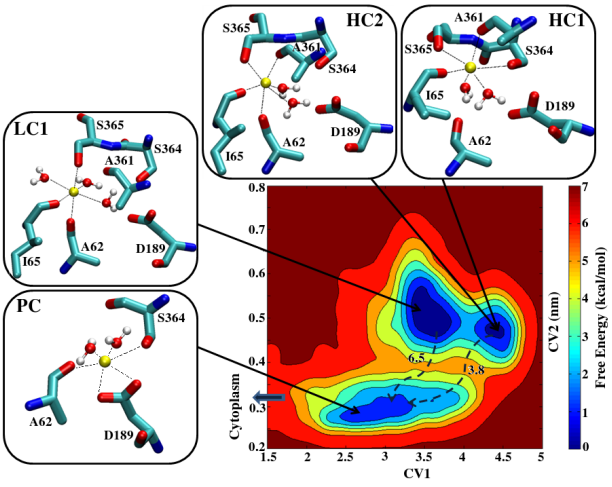

FIGURE 1.1: Free energy surface in the neighbourhood of the stable binding site of Na^+ discovered in this work. More detail in the text, Chapter 4.

one, called LC1. By long MD simulations we then showed that LC1 is indeed an ion-retaining binding site.

As a second goal of this thesis, we investigated the dissociation mechanism of the two ligands, Na^+ and galactose, from their binding sites into the cytoplasm. This topic

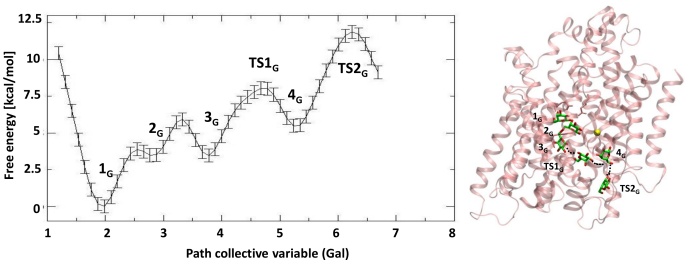

FIGURE 1.2: The free energy along the exit pathway of galactose. For more details, see Chapter 5.

has already been investigated by molecular simulations and experimental studies, but contradictory conclusions concerning the possible gating role of Y263 in the galactose release and the overall order of dissociation among the ligands were drawn [7, 8, 11–13]. Starting our simulations from the stable binding site identified in the first part of our study, we performed bias-exchange metadynamcs exploiting nine different collective variables (among which the path collective variable, coordination numbers, hydrogen bonds, distances, etc). Our results revealed that: 1) the minimum free energy exit pathway of the galactose does not involve any rotameric transition of the side-chain of Y263; 2) the main barriers characterizing the releasing mechanism are of the order of 11-12 kcal/mol for both the ion and the substrate; 3) the release mechanism of the galactose and the sodium ion is non-cooperative. Based on these results, we propose to extend the six-state kinetic model introduced by Wright and coworkers [12, 13] by adding one more state. Indeed, we suggest that, from the ligand bound inward-facing conformation, the transporter can follow independently two paths for Gal and Na^+ release.

Cyclic Nucleotide-Gated (CNG) ion channel

In the third part of this thesis we addressed the CNG channels, cation channels that mediate sensory transduction in the process of vision and olfaction. Despite belonging to the voltage-gated cation channels (VGCC) superfamily and, thus, harboring a voltage sensor domain, CNG channels respond to the binding of cyclic nucleotides to their intracellular domain. They are tetramers, with four subunits symmetrically organized around a central pore whose narrowest part, called selectivity filter, underlies the crucial role of selecting which type of ion can pass through the channel. Unfortunately, the only crystal structures available for this type of ion channels are those of a mimic based on the NaK channel, where the residues DGNFS of the selectivity filter were substituted by the ETTP which are characteristic of the CNG channels [14].

Recently, functional studies highlighted two remarkable features, previously overlooked, characterizing CNG channels: first, the permeation (the passage of specific/aspecific

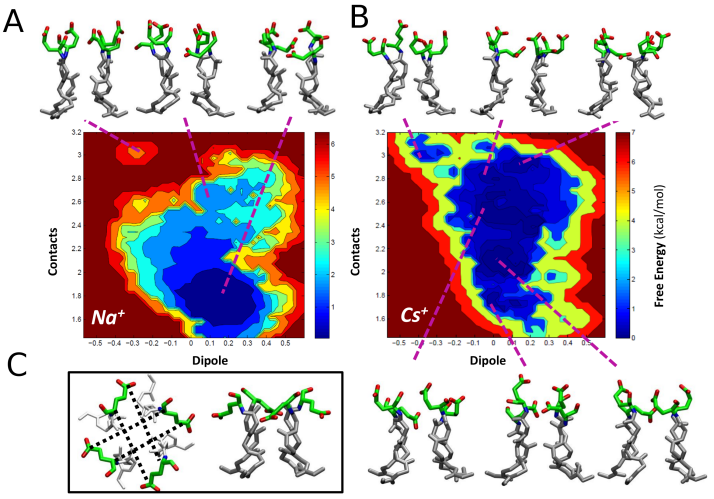

FIGURE 1.3: Free energy of Na^+ and Cs^+ systems. More details in Chapter 6.

ions) and gating (conformational changes that open and close their ion-permeable pores) are tightly coregulated at the selectivity filter; second, even if they were thought not to be voltage gated, a very recent work showed that CNG channels behave in a voltage dependent way in the presence of large cations (Rb^+, Cs^+) [15].

In order to investigate these issues we performed bias-exchange metadynamics using several collective variable (the dipole, distances, coordination number of ions, etc.). Our results can be summarized as follows: 1) the Cs^+ system is more flexible than the Na^+ system, and is characterized by a liquid-like behaviour in which E66 assumes a large number of conformations (Figure 1.3); 2) the conformation of P68 depends strongly on the voltage only if the channel is occupied by Cs^+. Consistently with this finding, this residue was previously shown to affect the selectivity of the channel [16]; 3) the movement of P68 and the permeation of the top ion in the Cs^+ system are strongly coupled. Based on our simulations, we then conclude that the pore flexibility plays a relevant role in the mechanism of the non-selective cation channels.

Chapter 2

Biological Background of Transporters and Ion Channels

2.1 Membrane Proteins

An essential condition for life is the compartmentalization of cells and organelles from their environment. This is made possible by the presence of biological membranes which, beyond this basic role as a barrier, promote many other functions, mainly determined by the type of proteins associated with or embedded in the bilayer. Membrane proteins are therefore key players in crucial processes like energy conversion, transport of ions and small molecules, signal recognition, transduction etc. [1]. Indeed, over 25% of all genes in most genomes encode for membrane proteins [17, 18]. Their important biological role is further highlighted by the fact that they represent the target of about 60% of approved drugs and their experimentally determined three-dimensional structures are eagerly sought to assist in structure-based drug design [19].
Unfortunately, a detailed structural and mechanistic characterization of this kind of proteins is much more difficult than that of the cytosolic ones, due to experimental difficulties (at expression, solubilization and crystallization levels). The first atomic-resolution structure of a membrane protein was obtained by Johann Deisenhofer and Hartmut Michel in 1985 [20]. They were able to determine the structure of a photosynthetic reaction center, contribution awarded with the Nobel Prize in 1988. The most important accomplishment was probably the demonstration that membrane proteins could be crystallized, thereby opening the way for atomic-resolution X-ray crystallography also for this kind of proteins. Nowadays, 29 years later, more than 494 atomic resolution structures (excluding the mutants or the same proteins in presence of different substrates)

are available in the PDB database, obtained mainly by X-ray crystallography, but also by nuclear magnetic resonance (NMR) and electron cryo-microscopy [21, 22].

In this thesis we focused on two main types of membrane proteins: transporters and ion channels, which play a key role in the transport of nutrients inside the cell and in the sensory transduction, respectively.

2.2 Transporters

Transporters are membrane proteins involved in the translocation of small organic molecules across the cellular membrane. While the primary active transporters work using directly the light or the energy released by chemical reactions (i.e. ATP hydrolysis), the secondary active transporters exploit the energy stored as transmembrane electrochemical gradient of ions (mostly Na^+ or H^+). These proteins can transport two or more molecules and ions in the same direction (symporters) or in opposite directions (antiporters). Sodium symporters, in particular, use this alkali metal ion to cotransport (in the same direction) a variety of substrates (sugars, amino acids, neurotransmitters, nucleobases) [3, 23] against their chemical concentration. They play a crucial role in the physiology of the brain, intestine, kidney, thyroid and skin, representing thus the target for therapeutic intervention in the treatment of depression, diabetes, obesity, etc. [3, 23]. Despite belonging to different gene families, the hypothesized transport cycle of these transporters occurs via a common alternating access mechanism [2, 3], in which the protein switches from an outward to an inward-facing conformation (Figures 2.1 and 2.2).

In recent years, the atomistic structures of several secondary symporters have been obtained by X-ray diffraction: the leucine transporter, LeuT, belonging to Neurotransmitter Sodium Symporter (NSS) family [5, 24], the benzyl-hydantoin transporter, Mhp1, belonging to Nucleobase Cation Symporter-1 (NCS1) family [25], the betaine transporter, BetP, belonging to Betaine Carnitine Choline Transporter (BCCT) family [26, 27], the glutamate transporter, GltPh, belonging to Excitatory Amino Acids Transporter (EAAT) [28] and the sodium/galactose transporter, SGLT, belongs to Sodium Solute Symporters (SSS) [4]. These structures provided insights into the molecular architecture of the transporters in the different stages of the transport cycle [3], unveiling their common characteristics. Indeed, despite their unrelated sequences (sequence identity 10-16%), they share a common structural core of 10 transmembrane (TM) helices composed by two five-helix inverted repeat (5HIR) motifs. In Figure 2.3, where the ligands' binding sites and the 10-TM core domains of four different symporters with a very low sequence identity are superposed, this striking structural similarity is highlighted.

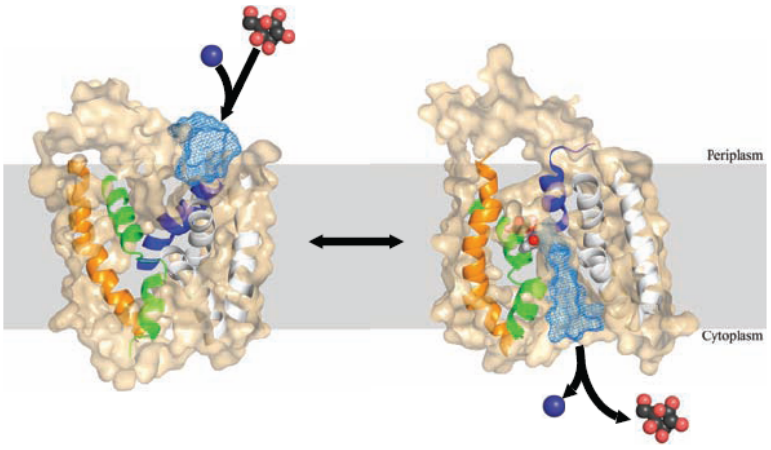

FIGURE 2.1: Conformational switch depicted here for the *Vibrio parahaemolyticus* bacterial homologue of SGLT (vSGLT). Slice of the protein surface is reported in the outward (left) and inward-facing (right) conformations, showing the hydrophilic cavity (blue mesh). Helices TM3, TM7 and TM11 are colored orange, green and blue, respectively. Galactose is shown as black and red spheres (for the C and O atoms), while the ion is colored as a blue sphere. Image from Ref. [4].

Thus, the study of the molecular mechanism of one of these proteins is important not only per se, but also because it can provide information on the general mechanism of this class of secondary active transporters, referred now on as 'LeuT-fold'.

In this work we focused on the sodium/galactose transporter (SGLT), which has an important role in accumulating sugars (i.e. glucose or galactose) inside the cells. This process is fundamental for the physiological intestinal absorption and renal reabsorption of sugars, and it is nowadays a promising field for the development of a new class of drugs for the treatment of type 2 diabetes [13]. Moreover, SGLT has a very important role in water absorption as it can promote the uptake of about 6 liters per day in normal adult intestine. The exact mechanism of water permeation through this transporter is still debated [29, 30]. The two main hypotheses are: (i) the active cotransport [29, 31], where water flux is coupled to the flux of ion/solute, and (ii) the passive permeation [32], for which a flux of water is induced as a response to the local osmotic gradient created by the accumulation of solutes near the intracellular side of the membrane during solute transport. However, despite this relevant function has still to be clarified in detail, it is already greatly exploited in the oral rehydration therapy (ORT) for the treatment of secretory diarrhea [13, 33].

2.2.1 Ion Binding Sites

The binding of sodium ions (one, two or even three ions, depending on the protein) represents a crucial event in many of secondary transporters. However, details of ion

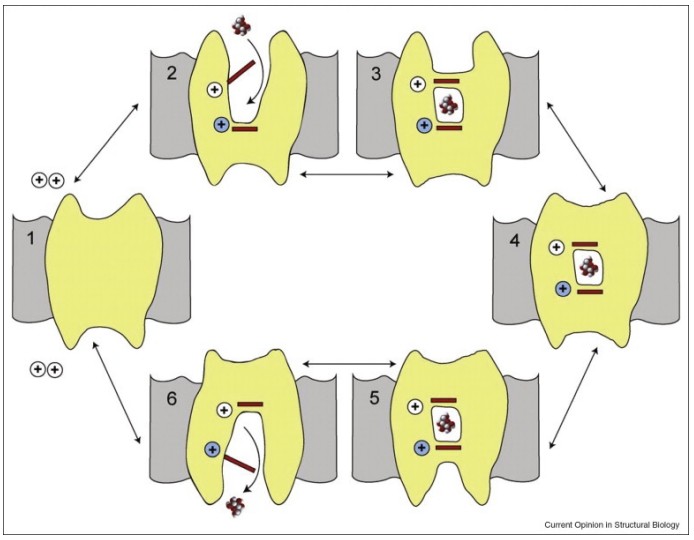

FIGURE 2.2: Hypothesized alternating access mechanism. Simplified model showing the hypothesized possible conformational states of the symporter and the role of gates in the transport. In state 1 no ligands are bound to the transporter. In the presence of extracellular sodium, once the ion(s) and the substrate bind, state 2, the external gate closes, state 3. The next step is the conformational switch, that leads to the closure of the external vestibule, state 4, and the opening of the internal vestibule, state 5. Opening of the internal gate permits substrate and sodium ion(s) to dissociate and exit at the intracellular face of the membrane, state 6. The cycle is completed by the closure of the internal vestibule and the return to state 1. Image from Ref. [3].

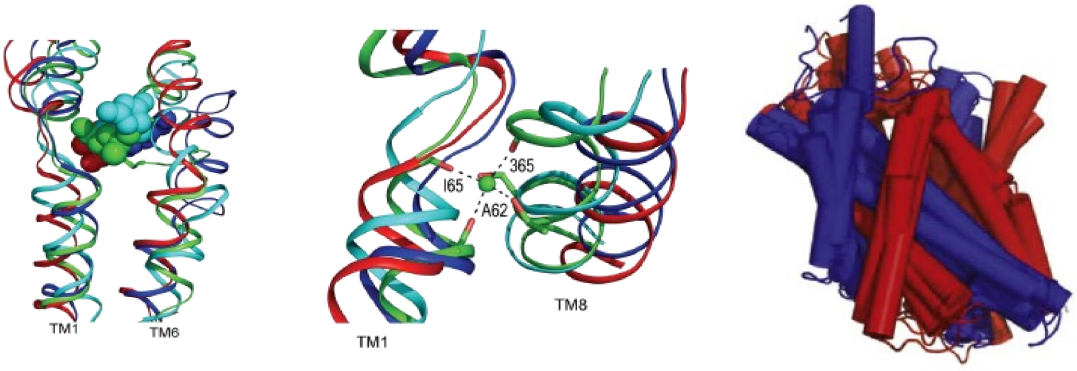

FIGURE 2.3: Common binding sites. Left: a structural alignment of TM1 and TM6 (TM2 and TM7 for vSGLT), and the substrate-binding site for vSGLT (red), BetP (blue), LeuT (green), and Mhp1 (cyan). Center: the same as before but for the sodium-binding site. Right: structural alignment of the core domains from vSGLT, BetP, LeuT, and Mhp1 yield an RMSD between 3.8 and 4.5 Å. The two 5-helix repeats are colored red and blue, respectively. Image from Ref. [3]

binding sites are poorly characterized in most cases, in part because of the challenges of membrane protein crystallography, the relatively low resolution of the data currently available, and, in part, because of the similarity of the scattering factors of sodium ions and oxygen atoms of water. Thus, the location of their biding site(s) is often unclear. It is interesting to note from the X-ray structures available nowadays for the different members of LeuT family that, while for the outward or occluded-facing conformations

the ions are tightly coordinated by protein residues, in the inward-facing conformations no ions are detected. Moreover, in case of suggested binding sites, the ions are loosely surrounded by residues. Indeed, for SGLT the crystal structure of the inward-facing conformation was solved by Faham et al. [4] from the bacterial homologue, *Vibrio parahaemolyticus* (vSGLT), with only galactose (Gal) bound inside the protein (Figure 2.4). The Na^+ ion was not detected and a plausible ion-binding site was proposed, on the basis of a structural comparison with the LeuT structure [5] and by mutational analysis. Molecular dynamics (MD) simulations studies, however, suggested that the crystal structure represents an ion-releasing state of the transporter [6–8] since the ion departs from the putative binding site after a few ns of MD.

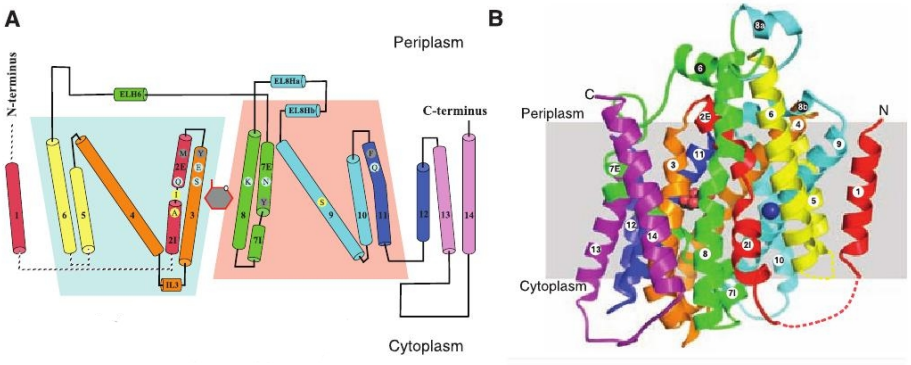

FIGURE 2.4: The vSGLT structure. A) The structure is colored as a rainbow from the N-terminus (red) to the C-terminus (purple). The blue and red trapeziums represent the inverted topology of TM2-6 and TM7-11. Galactose is represented as a gray hexagon with red trim. Residues involved in sugar recognition, gate residues, and a proposed ion site are shown (cyan, gray, yellow circles, respectively). B) The whole structure viewed in the membrane plane. Same coloring scheme and numbering of helices as in A). Galactose (black and red spheres) and proposed Na^+ (blue sphere) are shown. Image from Ref. [4]

Recent studies addressed several ion-related aspects in the 5HIR proteins. Khafizov et al. [34], for example, have recently employed MD simulations in parallel with experimental studies to characterize the Na^+-binding sites in BetP for which experimental evidence was lacking. In Ref. [35], Perez et al. studied the sequential formation of the substrate and sodium binding sites in the outward open state of BetP. Using a combination of structural and computational studies, they unraveled that the binding of the substrate leads to the occlusion of the sodium sites, especially the Na2 site, which is the one conserved among sodium coupled transporters, leading to a progressive dehydration and reshaping of the binding sites.

For LeuT, a close inspection of the role of the local hydration and of the ions in Na1 and Na2 binding sites has been addressed [36, 37].

Concerning GltPh transporter, free energy calculations revealed the presence of a third Na^+ binding site, not solved in the crystal structure [38].

2.2.2 Binding/Releasing Mechanism of the Ligands

The binding/dissociation mechanism of the ligands to/from their binding sites in the outward/inward-facing conformation of the symporter represents another relevant step in the study of the functional mechanism of this class of proteins. Indeed, many theoretical and experimental studies have been conducted in order to investigate these processes.
For vSGLT, galactose release has showed to happen spontaneously or by applying an external force [7, 8, 11, 39]. A more detailed discussion is reported on this topic in Chapter 5.
Performing MD simulations on LeuT, Zhao et al. hypothesized that the binding of the ion in Na2 site leads the protein to a more outward-open state, which facilitates the entrance of the substrate [37].
Computational studies on GltPh have described the time-resolved events that drive the recognition and binding of substrate at the outward-facing conformation, identifying a helical hairpin (HP2) as the extracellular gate that undergoes large motions driving the diffusion of the substrate towards the binding site [40]. An attempt to describe the energy profile along the exit pathway of the aspartate from GltPh using Steered MD simulations led to quite high barriers, around 15 kcal/mol [41]. The substrate uptake and release, together with the role of ions during these processes, have been addressed in Ref. [42] by metadynamics simulations. The authors defined different roles for the two ions: while the first ion with the substrate stabilize the closed conformation of HP2, the second Na^+ plays a fundamental role in locking GltPh in a conformation competent for substrate internalization.

2.2.3 Inward-to-Outward Switch

The study of the whole transport cycle of these symporters is the most challenging and difficult part. Several scientists have approached this issue for different proteins [24, 27, 43, 44].
In 2009 Forrest and Rudnick proposed a mechanism to explain how the ion-coupled solute flux works in symmetrical transporters: the so-called rocking bundle model [43]. Based on the outward-facing structure of LeuT, they defined a four-helix bundle, made by the first two helices of each repeat (see Figure 2.4), and a scaffold that surrounds the bundle protecting it from the lipids. The different tilt of the bundle with respect to the scaffold determines the inward or outward conformation of these transporters. According to this rocking-bundle mechanism, the binding sites would be located at the interface between the moving bundle and the stationary scaffold, providing a way for the ligand to control the orientation of the bundle through direct interactions. In 2012 the

resolution of the inward-facing conformation of LeuT by X-ray crystallography confirmed this hypothesis in a more complex picture where both local conformational changes and rigid body movements of groups of helices were associated with the transport [24].
The alternating access mechanism has been investigated also for BetP, based on the three major conformations solved crystallographycally: outward-facing, closed and inward-facing ones [27]. Here, a hybrid of rigid body movements and individual flexing of symmetry-related helices has been described, resembling changes seen for LeuT, but also showing differences attributed by the authors to the osmolytic nature of betaine [27].
The analysis of three structures together with molecular dynamics simulations have been useful to Shimamura et al. to study the transition from the outward through the occluded to the inward state of Mhp1 [44]. They found the mechanism of this symporter to be more consistent with the rocking bundle model.

2.3 Ion channels

The movements of ions across biological membranes is crucial to physiological processes like nerve excitation, muscle cell contraction and hormone secretion [45]. To this end, ion channels play a vital role by providing passages through pores within membranes, allowing specific ions to move down their electrochemical gradients. At difference with transporters, whose transport mechanism involves the movement of transmembrane segments leading thus to a relatively slow turnover rate (200-50000 molecules per second), an open ion channel, consisting basically in an aqueous pore, can permeate molecules with rates up to 10^8 per second, close to the diffusion limit [46]. Moreover, ion channels can be highly selective towards specific ionic species, and this property, together with the high conduction rates, leads to a very interesting paradox: how is possible to conciliate high selectivity (that implies high binding affinity) with the high turnover rate?

The modern history of ion channels begins in 1952 with the fundamental work of Hodgkin and Huxley on the definition and description of the action potential in the squid giant axon [47]. 46 years later, the atomic resolution of the first ion channel by X-ray crystallography represented another breakthrough in this field [48]. The structure of this bacterial K^+-selective channel (KcsA) revealed a tetramer with four identical subunits symmetrically organized around a central pore. Located at the narrowest part of the pore, the selectivity filter underlies the crucial role in determining which type of ion can pass through the channel.
In this thesis we studied the Cyclic Nucleotide-Gated (CNG) channels, non-selective cation channels expressed not only in vertebrate photoreceptors and olfactory sensory neurons, mediating sensory transduction in the process of vision and olfaction, but also in other neurons and non-neuronal tissues. In the visual transduction, CNG channels, present in the rod photoreceptors, open in dark conditions, activated by the binding of cyclic guanosine monophosphate (cGMP), and allow a steady cation current (*dark current*), keeping the membrane depolarized. Light triggers a cascade of enzymatic reactions that leads to the hydrolysis of cGMP and thus to the closure of CNG channels (Figure 2.5, left). During olfaction, volatile odorant molecules reach the olfactory epithelium in the nasal cavity and activate odorant receptors. This interaction triggers an enzymatic cascade that leads to an increase of intracellular concentration of cyclic adenine monophosphate (cAMP) and to the subsequent activation of CNGs, whose opening generates a current of Na^+ and Ca^{++} ions (Figure 2.5, right).
CNG channels belong to the voltage-gated cation channels (VGCC) superfamily, which also includes voltage gated K^+, Ca^{++} and Na^+ channels, believed to have evolved from a common ancestral predecessor [45]: a two-transmembrane domain channel probably similar to the bacterial K^+ channel solved by the MacKinnon group [48]. Despite belonging

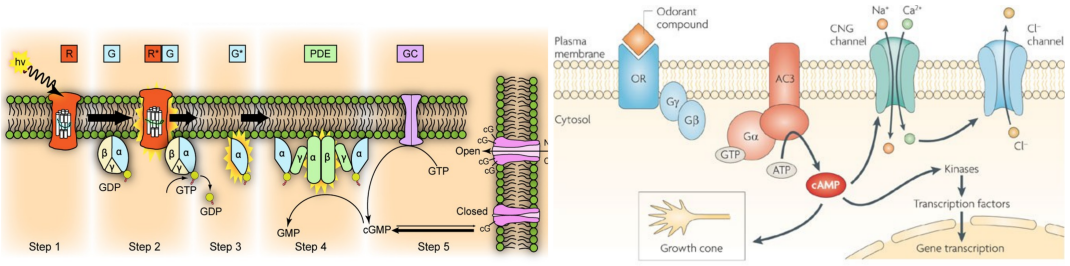

FIGURE 2.5: Vision and olfaction mechanisms. Left: representation of the molecular steps of photoactivation in the outer membrane disk of a rod. Activated proteins are labelled by $*$. The absorption of a photon activates the rhodopsin molecule (R), step 1. R^*, interacting with the transducin molecule (G), catalyzes its activation by exchanging the bound guanosine diphosphate (GDP) with the cytoplasmic guanosine triphosphate (GTP), step 2. The activated form of transducin, G^*, binds the inhibitory γ subunit of the phosphodiesterase (PDE), activating its α and β subunits, step 3. The activated PDE hydrolyzes cGMP, step 4, a process that causes the closure of the CNG channels, and thus a decrease of the intracellular concentration of Na^+ and Ca^{++}. Guanylyl cyclase (GC) is the enzyme that synthesizes cGMP from GTP, step 5. Right: binding of odorant compounds to an odorant receptor (OR) initiates a transduction cascade involving a G protein and activation of adenylate cyclase 3 (AC3), which in turn produces the second messenger cAMP. cAMP binds to a CNG channel and results in the influx of cations (Na^+ and Ca^{++}) which depolarizes the cell membrane. Ca^{++} can also activate a Ca^{++}-dependent Cl^- channel. Olfactory sensory neurons (OSNs) maintain a high intracellular Cl^- concentration, such that this channel supports an efflux of negatively charged Cl^-, producing a further depolarization of the cell membrane. Notably, the elevated levels of cAMP in the soma have a crucial role in regulating the phosphorylation of proteins and the transcription of genes important for growth and survival of the axons of OSNs. Images taken from Ref. [49, 50].

to the VGCC superfamily and, thus, harboring a voltage sensor domain (S1-S4 transmembrane helices), CNG channels respond to the binding of cyclic nucleotides (cAMP, cGMP) to their intracellular domain (Figure 2.6, left) [51]. Another important region is the pore domain, comprising two transmembrane helices (S5, S6), a short P-helix and the selectivity filter, lined by backbone carbonyl oxygens (Figure 2.6). Unfortunately, details of the selectivity and dynamic properties of CNG channels are not fully understood, since no high resolution structures are nowadays available. However, to reveal some insights on the structural mechanism and functional properties, Derebe et al. engineered and solved the crystal structures of a set of mimics of CNG [14]. The mimics, named NaK2CNG, were based on the NaK channel, a non-selective cation channel, where the residues DGNFS of the selectivity filter were substituted by the residues ETTP (and their mutants DTPP or NTPP), a characterizing and highly conserved motif in the CNG channels (Figure 2.7). The structures of all chimeras, in complex with various cations (Na^+, K^+ and Ca^{++} ions), were determined in an open conformation with a resolution < 2 Å. The selectivity filter adopts an architecture intermediate between NaK and KcsA. It contains three contiguous ion binding sites equivalent to sites 2-4 of K^+ channel and numbered in the same way for comparison. In Figure 2.8, the selectivity filters of K^+, NaK and NaK2CNG show a different number of sites for the three types of channels.

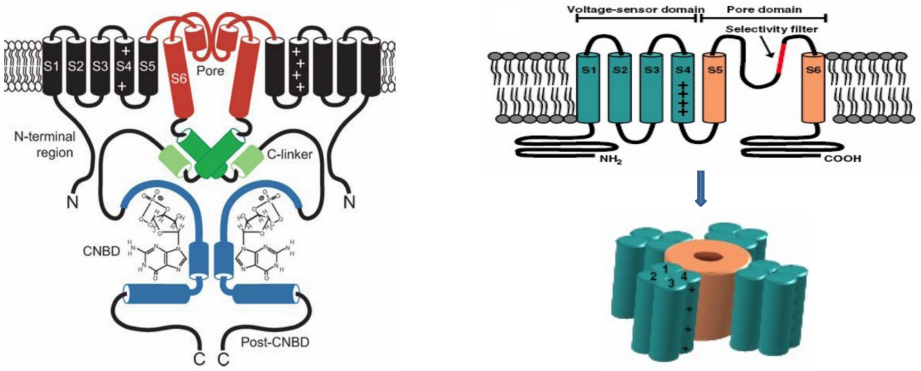

FIGURE 2.6: Topology of CNG channels. Left: two of the four subunits of a CNG channel are shown. The transmembrane segments (S1–S6) are shown in black, except for the pore region (S6 and pore loop), which is shown in red. The voltage sensor (S4) is indicated by positive charges. The C-terminal region contains the cyclic nucleotide binding domain (CNB, blue, shown with cGMP bound) and the C-linker (green), which connects the CNB to the pore. Right: the selectivity filter is highlighted in red and the 3D structural rearrangement at the bottom. Images taken from Ref. [52]

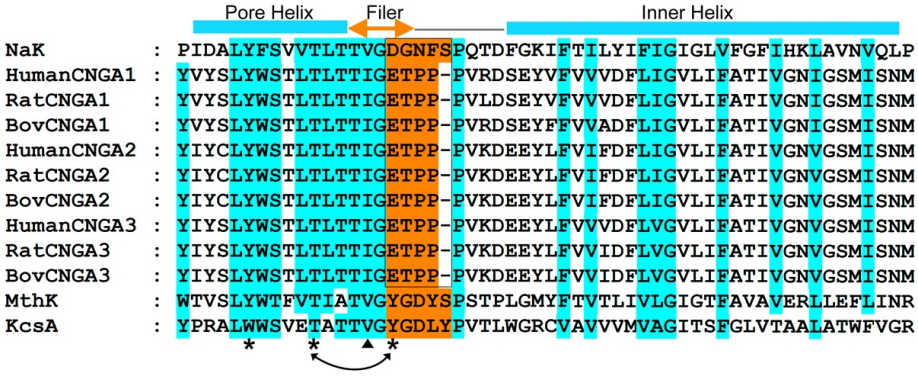

FIGURE 2.7: Sequence alignment of channels. While the pore of K^+ channels are characterized by TXGXG (where X denotes a hydrophobic residue), CNG channels contain an ETPP motif flanking the conserved TXG residues. Image taken from Ref. [14].

Ion channels undergo conformational changes that open and close their ion-permeable pores, a process referred to as *gating*. This process is regulated by different factors: changes of the transmembrane potential (voltage-gated channels), ligand binding, pH variations or mechanical forces. Once in the open form, ion channels allow the passage of specific/aspecific ions, in the so-called *permeation* process.

2.3.1 Selectivity

The ability to select specific ionic species, known as ion selectivity, is a fundamental property of ion channels. K^+ channels, for example, are able to transport 10^3-10^4 K^+ ions over one Na^+. In the last years, many attempts have been done in order to describe

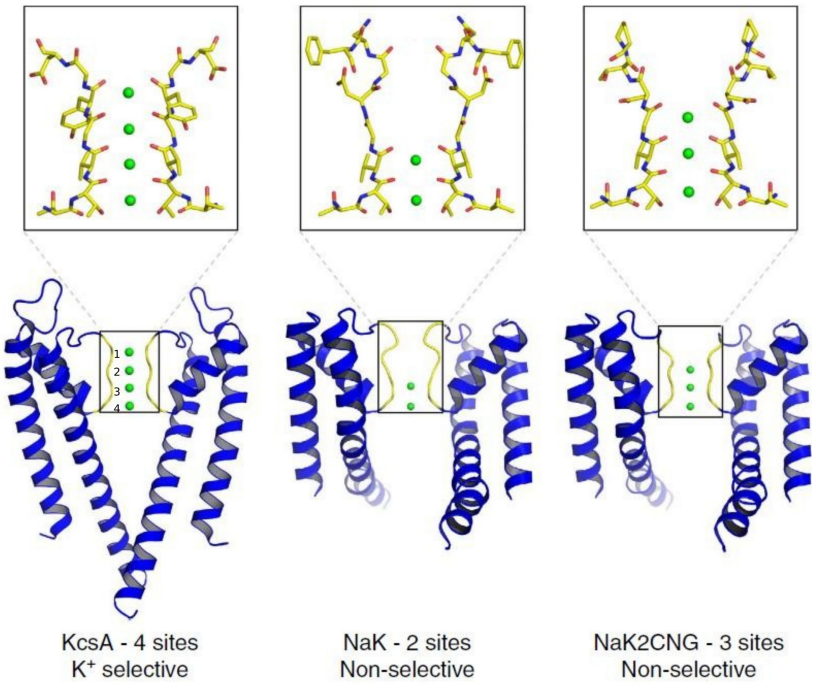

FIGURE 2.8: Selectivity filters of selected ion channels. The selectivity filter of KcsA contains four contiguous ion binding sites (S1-S4), on the left. While the NaK harbors only two sites (central), in the NaK2CNG we can see three sites, equivalent to S2-S4 (right) of KcsA. S1 in the chimera is replaced by a funnel-shaped vestibule. Image taken from Ref. [53].

the mechanisms underliying selectivity.

Two main hypotheses have been proposed for K^+ channels: (i) the mechanism of ionic selectivity by selective exclusion (a kinetic view) and (ii) the mechanism of selectivity by selective binding (a thermodynamic view). In the first mechanism, proposed in electrophysiological studies, the rates at which ions enter the pore are the key determinants [54], while in the second view, the presence of multiple binding sites with high affinities for permeant ions and low affinities for impermeant ions represents the key factor [55]. High resolution structures of KcsA appear consistent with the selective binding mechanism: the oxygen atoms in the selectivity filter coordinate K^+ to compensate for the energetic cost of dehydratation, while they will not optimally coordinate the smaller Na^+ [48, 56]. Using calorimetry and crystallography, Lockless et al. [57] showed for KcsA that the formation of the conductive structure involves not only the atoms of the selectivity filter that are in direct contact with bound ions, but also the surrounding atoms up to a distance of 15 Å from the ions, concluding that ion selectivity is a property of size-matched ion binding sites created by the protein structure rather than a property of charge density.

Also the hydration and the dynamics have been described as important for the selectivity in ion channels [58]. Indeed, computational studies showed that the S2 site of KcsA is the most selective binding site favouring K^+ over Na^+. Thus, the disruption of this

site in NaK (Figure 2.8) appears the major determinant to explain why this channel permits the conduction of both ions. Moreover, an increase in average hydration number strongly correlates with a loss of selectivity for K^+ in favor of Na^+ [58]. The ability of NaK to bind both ions may also arise from the ions' ability to utilize the existing environment in unique ways (different configurations, hydration, sites.) [59].
Sauer et al. [60] demonstrated that K^+ channel filter exists as an energetically strained structure and requires key residues Tyr and Asp to hold the filter in the precisely defined single-file four-sited configuration, essential for selective K^+ permeation. Free energy calculations of NaK2CNG suggested that the presence of the site S23 (namely, a site at the boundary between S2 and S3) for Na^+ ions is important in determining non-selectivity in channels [61].
A recent work, however, done using in concert electrophysiology, MD simulations and crystallography [62], is consistent with the kinetic hypothesis. Here, small cations have been observed to bind a separate site within KcsA, suggesting that selective permeation may primarily results from a large barrier blocking the entry for Na^+ in the presence of K^+, and not from a difference of binding affinity between ions.
Concluding, it is likely that the story is more complex than what appears to date and that multiple mechanisms are responsible for selectivity in ion channels.

2.3.2 Gating and Permeation

For several decades gating and permeation were considered two distinct and independent processes. Indeed, in Na^+ and K^+ channels they have different structural basis. Gating is controlled by the so-called *bundle crossing*, where S6 helices cross over and form a tight bundle towards the intracellular face of the channel [48, 63]. Permeation, instead, reflects ion-ion repulsion and ion-pore attraction (whose effects underlie the high turnover rate [64, 65]) within the selectivity filter at the extracellular side. Namely, the gate and selectivity filter are localized at the opposite sides of the protein.
In contrast to K^+ channels, in CNG channels the bundle crossing does not gate ion permeation. Indeed, Contreras et al. [66] showed that ion permeation in CNG channels is tightly regulated at the selectivity filter, and not at the intracellular end of the pore. By scanning the entire selectivity filter using small cysteine reagents (like cadmium and silver), they observed a state-dependent accessibility pattern consistent with gated access at the middle of the filter [66].
A model of the complex events initiated by the binding of the cyclic nucleotide to the cyclic nucleotide binding (CNB) domain has been proposed [67]: (1) upon ligand binding, the CNB domain and the C-linker rotate by 30° to 60°; (2) this rotation is transmitted to the pore region by a rotation of the S6 helices around their axis; and (3)

as a consequence of this rotation and of the tight coupling between S6 and the P-helix, the lumen of the pore widens and the channel opens. Thus, the selectivity filter acts also as a gate, namely, the same part of the structure controls ion permeation as well as gating [66, 67].

It is interesting to note that, if the primary gating of K^+ channels is located in the S6 helix, there is now increasing evidence that conformational changes occur also in the selectivity filter of K^+ and NaK channels [68–70]. Indeed, by free energy calculations, the occupancy of S2 and the strength of intersubunit interactions involving the GYG signature have been depicted as microscopic factors attributing a gating role to the filter of K^+ channel [68]. Chakrapani et al. identified E71 as having a key function in this process, by experimental means [69].

Potential of mean force for complete conduction events of Na^+ and K^+ ions through NaK showed that the filter of cation selective channels can adopt a conductive state (represented by the mutated NaK, namely NaK2CNG) and a closed one (WT-NaK) [70]. Hydrogen bond between D66 and N68 seems to be correlated to the permeation barriers.

2.3.3 Conduction

The paradox of ion channels mentioned before, namely, high conduction rates coupled to high selectivity, has been described in terms of concerted mechanisms in which ion-channel attraction and ion-ion repulsion have compensating effects, since several ions are moving simultaneously in single-file through the selectivity filter [48, 71]. Two main configurations - with ions at sites S1, S3, or S2, S4 - are at equilibrium. The entrance of a third ion from one side is coupled to the exit of an ion on the opposite side, recalling the 'knock-on' mechanism proposed in 1955 by Hodgkin and Keynes [72].

Bernèche and Roux [64] found in KcsA that the largest free energy barrier in this process was on the order of 2-3 kcal mol^{-1}, meaning that conduction is essentially barrierless and limited by diffusion.

Similarly to KcsA, also the pore of the voltage gated Na^+ channel, NavAb, preferentially occupied by two ions, can switch between different configurations by crossing low free energy barriers. However, in contrast to K^+ channels, the movement of the ions appear to be weakly coupled and, what is believed to confer selectivity, Na^+ ions are partially dehydrated and a selective site is present in the narrowest region of the filter, where a hydrated Na^+ is energetically stable [73].

2.3.4 Voltage Dependence of CNG Channels

Despite harbouring a voltage sensor domain, CNG channels were thought not to be voltage gated until a very recent work, showing that they behave in a voltage dependent manner in the presence of large permeating ions [15]. From Figure 2.9 it is evident the

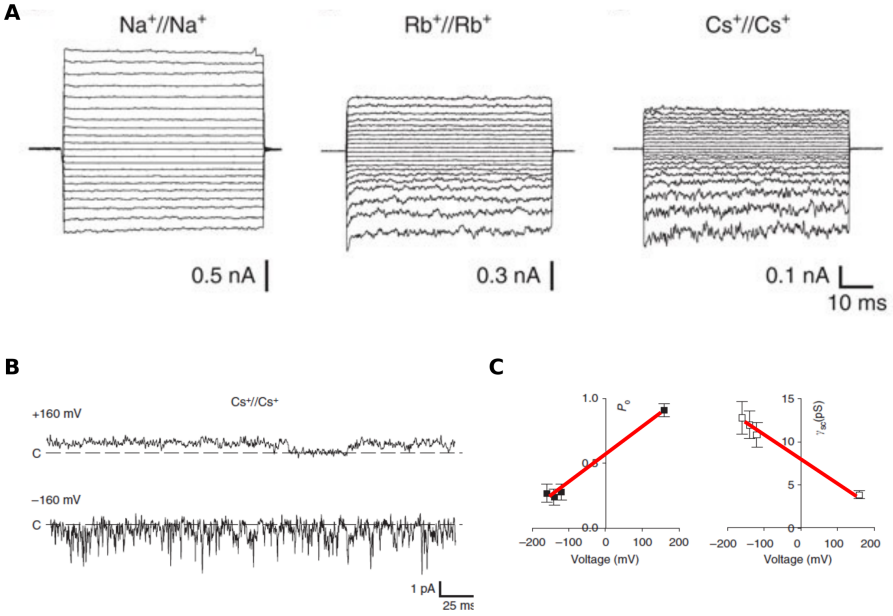

FIGURE 2.9: Current recordings of WT CNGA1 channel in symmetrical conditions. Panel A: Macroscopic currents recorded in Na^+, Rb^+ and Cs^+. Panel B: single-channel recordings at +160 mV and -160 mV in Cs^+. Dashed lines indicate the closed state (C). Panel C: dependence of open probability (P_O) and conductance (γ_{sc}) on voltage (V) for symmetrical Na^+ [15].

asymmetric behaviour of the channel at positive and negative voltages in the presence of cations larger than Na^+ ion. At single-channel level, it is clear that in presence of Rb^+ and Cs^+ both open probability (P_O) and channel conductance (γ_{sc}) depend on voltage (V), but in an opposite way: P_O increases with V, while γ_{sc} decreases. Moreover, at negative potentials channel openings are more flickering.

The authors hypothesize that, during evolution, these channels lose their voltage-sensing ability when Na^+ and K^+ permeate so that the vertebrate photoreceptor CNG channels are open at negative voltages, a necessary condition for phototransduction [15].

Chapter 3

Methods

3.1 Molecular Dynamics

Molecular Dynamics (MD) is a simulation technique that allows computing the equilibrium and transport properties of a many-particle system. This method calculates the time evolution of atoms and molecules by integrating the Newton's equations of motion. The main ideas at the basis of MD are: 1. set up of the system (defining the conditions of the simulation, like temperature, number and typology of particle/atoms; defining the time step and the initial positions and velocities of particle/atoms); 2. computation of the forces acting on all the particles; 3. integration of Newton's equations of motion; 4. calculation of the averages of measured observables. The two central steps represent the core of the simulation; indeed, they are repeated until we reach the desired length of time. In general, for biomolecules, the initial positions of the atoms are taken from PDB crystal structures, NMR data or built by homology modeling.

3.1.1 Molecular Mechanics Force Field

The forces acting on the atoms are generated by a potential that provides approximations of the physics and chemistry of the system fitting experimental or high level quantum chemical data into simple functional forms. The potential energy of the system under investigation can be written as a sum of intra- and inter-molecular forces:

$$U(r) = \sum_{bonds} U^{str} + \sum_{angles} U^{bend} + \sum_{torsions} U^{dihe} + \sum_{imp-tors} U^{im-dihe} + \sum_{i<j} (U^{LJ} + U^{Coul}) \tag{3.1}$$

where the first four terms include the bonded interactions, while the last summation includes the non-bonded ones. More in detail:

$$U(r) = \sum_{bonds} \frac{1}{2} k_r (r - r_{eq})^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_{eq})^2 + \sum_{torsions} \frac{1}{2} V_n (1 + \cos(n\phi - \phi_{eq})) + \sum_{imp-tors} k_\xi (\xi - \xi_{eq})^2 + \sum_{i<j} 4\epsilon k_{LJ} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) + \sum_{i<j} \frac{1}{4\pi\epsilon_0 \epsilon_r} \frac{q_i q_j}{r_{ij}} \quad (3.2)$$

where the first two terms model the bond stretching and angle bending by harmonic potentials that give an increase in energy as the bond with distance r or the angle θ among bonded atoms deviates from the equilibrium reference values, r_{eq} and θ_{eq}. k_r and k_θ are the force constants. The equilibrium values of these terms are usually derived from structural databases, while the force constants are derived from infrared spectroscopy [74] or quantum chemical calculations. In the third term, the torsion potential models how the energy changes as a bond rotates. V_n is the torsional barrier of a given torsional angle. Dihedral parameters are calibrated on small model compounds, comparing the energies with those obtained by quantum chemical calculations [46]. Improper dihedrals are added to preserve planarity in aromatic rings or to prevent molecules from flipping over their mirror images. The last two terms refer to non-bonded interactions between all pairs of atoms (i and j) separated by at least three bonds. The van der Waals (vdW) interactions are modelled using a Lennard-Jones potential containing an attractive and a repulsive term. In the formula, ϵ is the depth of the potential well, σ_{ij} is the finite distance at which the potential is zero, r_{ij} the distance between the two particles. The electrostatic interactions are modelled with Coulomb potential. The latter is evaluated using Coulomb's law for the partial atomic charges associated with each atom q_i and q_j. ϵ_0 ϵ_r are the dielectric constants in vacuum and in the material used.
The non-bonded interactions are by far the most expensive part of MD calculations. In practical applications, however, the number of interactions is limited by a pre-defined cutoff distance, introducing in this way a small truncation error ($< 10^{-4}$ of the total) in the vdW energy. Instead, this trick can not be used for Coulomb term, since the interaction between charges q_i and q_j decays much less rapidly with distance. In this case a different approach is applied (see 3.1.3).

3.1.2 Equations of Motion

Once the forces between the particles are computed (using $F_i = -\frac{\delta V}{\delta r_i}$), the Newton's equations are integrated in order to proceed along the trajectory. Several algorithms

have been designed to carry out this task. One of the simplest, and used here, is the so-called *Leap Frog* algorithm[75].

$$r(t+\delta t) = r(t) + \delta t v(t+\frac{1}{2}\delta t) \tag{3.3}$$

$$v(t+\frac{1}{2}\delta t) = v(t-\frac{1}{2}\delta t) + \frac{\delta t}{m}F(t) \tag{3.4}$$

It calculates positions and forces at time $t+\delta t$, while velocities at time $t+\frac{1}{2}\delta t$. As a consequence, kinetic and potential energy are also not defined at the same time, and hence it is not possible to directly calculate the total energy. However, it can still be computed using:

$$v(t) = \frac{1}{2}(v(t+\frac{1}{2}\delta t) + v(t-\frac{1}{2}\delta t)) \tag{3.5}$$

3.1.3 Periodic Boundary Conditions and Long Range Interactions

In order to minimize boundary effects and to mimic the presence of the bulk MD simulations are usually performed under periodic boundary conditions (PBC). In this approach, the system containing N atoms is replicated in all directions to yield an infinite periodic lattice of identical cells. When a particle moves in the central cell, its periodic image in every other cell moves accordingly. As one molecule leaves the central cell, its image enters from the opposite side without any kind of interactions with the cell boundary.

Non-bonded interactions. PBCs are taken into account calculating non-bonded interactions. Since vdW interactions are truncated above a certain cutoff distance, each of the interactions is calculated between atom i and the nearest periodic image of atom j. A natural consequence of this treatment is that the periodicity of the system in each direction must be higher than twice the cutoff distance, in order to avoid additional errors arising from double counting of interactions between atoms separated by $\sim 1/2$ the unit cell length. Electrostatic interactions are instead long ranged, and as such they are calculated between all the atoms of the system and all their periodic images: as a consequence, the total electrostatic energy is composed by an infinite number of terms:

$$U^{Coul} = \frac{1}{4\pi\epsilon_0\epsilon_r}\left(\sum_{\vec{n}}^{'}\sum_{i,j=1}^{N}\frac{q_i q_j}{|\vec{r}_i\vec{r}_j + \mathbf{U}\vec{n}|}\right) \tag{3.6}$$

where $\vec{n} = (n_x, n_y, n_z)$ and $\mathbf{U}\vec{n} = n_x\vec{u}_x, n_y\vec{u}_y, n_z\vec{u}_z$ and the prime on the summation indicates that we omit $\vec{n} = 0$ for $i = j$.
This energy contribution is usually calculated by Particle Mesh Ewald method [76, 77], which splits the sum in a short range term, comprising the interactions within a cutoff distance, and a long range term, in which the remaining interactions are calculated by Fourier transform.

3.1.4 Temperature and Pressure Control

In this work we performed the classical MD simulations at constant temperature, T, and constant pressure, P. This means that the equations of motion are modified, taking into account the temperature and pressure couplings.
The Nosé-Hoover thermostat is a deterministic algorithm originally developed by Nosé and improved in the formulation of Hoover. Briefly, the system is extended by introducing a thermal reservoir and a friction term in the equations of motion. The friction force is proportional to the product of each particle's velocity and a friction parameter, ξ. This friction parameter (or *heat bath*) is a fully dynamic quantity with its own momentum (p_ξ) and equation of motion. In this formulation, the particles' equations of motion are:

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} - \frac{p_\xi}{Q}\frac{dr_i}{dt} \tag{3.7}$$

where the equation of motion for the heat bath parameter ξ is:

$$\frac{dp_\xi}{dt} = T - T_0 \tag{3.8}$$

The heat bath parameter will have the effect of increasing or reducing frictional forces affecting in this way the velocities of the particles in order to control the current instantaneous temperature (T) of the system by the reference one (T_0). The strength of the coupling is determined by the constant Q (usually called the *mass parameter* of the reservoir) in combination with the reference temperature:

$$Q = \frac{\tau_T^2 T_0}{4\pi^2} \tag{3.9}$$

where τ_T is the period of the oscillations of kinetic energy between the system and the reservoir. This provides a much more intuitive way of selecting the Nosé-Hoover coupling strength (in addition τ_T is independent of system size and T_0).

However, this dynamics can be non-ergodic, meaning that only a fraction of phase space is sampled, even if the simulation is run for a infinitely long time. For this reason, the approach was improved and made ergotic adding a chain of thermostats [78].

In the Parrinello-Rahman barostat [79, 80], the box vectors as represented by the matrix b obey the matrix equation of motion:

$$\frac{db^2}{dt^2} = VW^{-1}b'^{-1}(P - P_{ref}) \tag{3.10}$$

where V is the volume of the box, W a matrix parameter determining the strength of the coupling. The matrices P and P_{ref} are the current and the reference pressures, respectively.
Since we work with membrane proteins, we used a semi isotropic ensemble firstly applying the surface tension coupling ($NP_\gamma T$ ensemble), then, after the desired Area Per Lipid (APL) is reached, we fixed the Area of the membrane, in a NAT ensemble.
The average surface tension $\gamma(t)$ can be calculated from the difference between the normal and the lateral pressure:

$$\gamma(t) = \frac{L_z}{n}\left(P_{zz}(t) - \frac{P_{xx}(t) + P_{yy}(t)}{2}\right) \tag{3.11}$$

where L_z is the height of the box and n is the number of surfaces (here it is equal to 2). P_{zz}, P_{xx} and P_{yy} are the pressure along the z, x and y-directions, respectively.

3.2 Advanced Techniques

Molecular dynamics simulation method is a very important computational tool used in several fields of science. However, most important processes we are interested in happen at much higher time scales we can simulate using MD with the present-day computational resources. Indeed, the trajectories usually got stuck in local minima, close to the starting point of the dynamics, missing an efficient exploration of the conformational space. A solution to this problem has been proposed with the metadynamics approach [9].

3.2.1 Metadynamics (MTD)

This method has been developed in order to escape from local minima in the free energy surface (FES) [9, 10]. It is based on the addition of an external history-dependent

potential, $V_G(s,t)$, acting on a small number of collective variables (CVs). The CVs can be any explicit function of the Cartesian coordinates r. The bias is constructed as a sum of gaussians:

$$V_G(S(r),t) = w \sum_{t'=\tau_G,2\tau_G,..} exp\left(-\sum_{\alpha=1}^{d} \frac{(S_\alpha(r) - s_\alpha(t'))^2}{2\delta s_\alpha^2}\right) \tag{3.12}$$

where $s(t) = S(r(t))$ is the value taken by the CV at time t, τ_G is the frequency at which the gaussians with height w and width δs are added. These three parameters influence the typical deviation of V_G from the free energy. If the gaussians are large, the deviations will be large. It has been empirically [81] and analytically [82] demonstrated that, in the limit in which the CVs evolve according to a Langevin dynamics, the average of the bias converges to the negative of the free energy, thus providing an optimal bias to enhance transition events:

$$-F(s) \sim \bar{V}_G(s) = \frac{1}{t_{tot} - t_{eq}} \int_{t_{eq}}^{t_{tot}} dt V_G(s,t') \tag{3.13}$$

where $\bar{V}_G(s)$ is the time average of the bias, t_{tot} the total simulation time, t_{eq} the time needed to fill all the relevant minima and $V_G^i(s)$ reaches a stationary state.
The time required to reconstruct a free energy surface of a given accuracy scales exponentially with the number of CVs. Thus, the performance of the algorithm rapidly deteriorates as the dimensionality of the CVs space increases, representing an important limitation for the study of complex biological processes, like protein folding, gating of ion channels or other events involving major conformational changes, and so several independent degrees of freedom. To overcome this constraint, some approaches combining metadynamics and replica exchange [83, 84] have been developed. Exchanges between replicas evolved by metadynamics at different temperatures are exploited in Ref. [85] to enhance convergence. In this approach, it is not necessary to know in advance all the important coordinates of the system, since the barriers are crossed due to diffusion in temperature. A different way to combine replica exchange and metadynamics is proposed in Ref. [86], which is the method used in this thesis.

3.2.2 Bias Exchange Metadynamics (BE-MTD)

This method allows reconstructing the free energy in a virtually unlimited number of variables [86]. It is based on the simulation of several replicas of the system at the same temperature but biased by time-dependent potentials acting on different variables. Exchanges of conformations among pairs of replica are attempted according to the replica

exchange method [84, 85]. The exchange move consists of swapping the atomic coordinates r^a r^b of two replicas a b selected at random. Since the two replicas are evolved under the action of two different potentials, the move is accepted with a probability p_{ab}:

$$p_{ab} = min(1, exp\{\beta[V_G^a(r^a,t) + V_G^b(r^b,t) - V_G^a(r^b,t) - V_G^b(r^b,t)]\}) \tag{3.14}$$

If the move is accepted, the CVs of replica a perform a jump from $s^a(r^a)$ to $s^a(r^b)$, and replica b from $s^b(r^b)$ to $s^b(r^a)$. Since the gaussians are time dependent, detailed balance is violated in BE-MTD, as in ordinary MTD. However, the jumps greatly increase the capability of each replica to diffuse in the CV space, and thus the accuracy of the free energy reconstruction.

The analysis of the BE-MTD simulation is performed as follows.

In the first step one groups together configurations in microstates sets of configurations in which the value of the CVs are similar. An important issue is how many and which CVs should be used for the analysis. It is not necessary to use all the CVs that have been biased in the replicas, as some of those might prove a posteriori to be less relevant for the process or strongly correlated with other variables.

The next step of the analysis is computing the free energy of the microstates. In MTD the history-dependent potential provides an estimate of the low dimensional projections of the free energy. In order to pass from low-dimensional projections to multi-dimensional microstates one uses the weighted-histogram technique (WHAM) [87]. Following the approach of Ref. [88], the equilibrium probability of microstate α is estimated as:

$$p_\alpha^i = n_\alpha^i exp[\beta(V_\alpha^i - f^i)] \tag{3.15}$$

where i is the replica index, f_i is a shift constant fixing the normalization, n_α^i is the number of time state α is observed in replica i and $V_\alpha^i = V^i(s_\alpha)$ is the bias potential evaluated on the microstate α. V^i is estimated from the time-average between the equilibration time of the bias potential t_{eq} and the total simulation time t_{tot}:

$$V^i(s) = \frac{1}{t_{tot} - t_{eq}} \int_{t_{eq}}^{t_{tot}} dt V_G^i(s,t) \tag{3.16}$$

where $V_G^i(s,t)$ is the metadynamics bias at time t, computed from the sum of gaussians. The p_α^i-s are then combined in a single estimate of the probability following the standard WHAM procedure. This leads to:

$$p_\alpha = C \frac{\sum_i \frac{1}{\gamma_\alpha^i} n_\alpha^i}{\sum_i \frac{1}{\gamma_\alpha^i} exp[\beta(f^i - V_\alpha^i)]} \tag{3.17}$$

where C is a normalization constant, γ_α^i is a constant that takes into account the correlation time [88].
Finally, the free energy of microstate α is given by the usual formula:

$$F_\alpha = -k_B T \log p_\alpha \tag{3.18}$$

3.2.2.1 Collective Variables

The choice of the CVs plays an essential role in determining the convergence and efficiency of the free energy calculation. In order to avoid hysteresis, caused by the insufficient exploration of important regions in the conformational space, the CVs should distinguish the different metastable states of the system. Although there is no a priori recipe for finding the correct set of collective variables, in the BE-MTD method the number of variables can be large at will, making the selection less critical than in other methods. We here list some of the collective variables used in this work.

Coordination number: counts the total number of contacts (e.g. hydrogen bonds, hydrophobic contacts) between the atoms of two groups:

$$CN = \sum_{i,j} \frac{1 - (\frac{r_{ij}}{r_0})^n}{1 - (\frac{r_{ij}}{r_0})^m} \tag{3.19}$$

where r_{ij} is the distance between atoms i and j belonging to two different groups and m, n are exponents that allow to tune the smoothness of the function.

Hydrogen bonds: similar to the coordination number, where i belongs to the group of hydrogen donors and j to the acceptors.

Dihedral similarity: measures the similarity of dihedral angles to a reference value ϕ_i^{Ref}.

$$s_{\alpha\beta} = \frac{1}{2} \sum_{i=1}^{N_D} (1 + cos(\phi_i - \phi_i^{Ref})) \tag{3.20}$$

where N_D represents the number of dihedrals used. This variable is useful, for example, when the torsions of some particular residues are relevant for the process studied.

Distance: measures the distance between the center of mass of two groups of atoms.

Gyration radius: it is defined as:

$$s_{Gyr} = \left(\frac{\sum_{i}^{n} \mid r_i - r_{COM} \mid^2}{\sum_{i}^{n} m_i} \right)^{1/2} \tag{3.21}$$

where m_i is the mass of atom i and the sums are over the n atoms in the group. The center of mass is:

$$r_{COM} = \frac{\sum_{i}^{n} r_i m_i}{\sum_{i}^{n} m_i} \tag{3.22}$$

In our case, the gyration radius is useful in estimating the changing of the volume and the stiffness of the binding sites.

Path collective variable (PCV) [89]: this is one of the most relevant CVs used in this work. It requires choosing a set of N reference conformations, R_i, that define a path in configuration space, from an initial state to a final state. The s variable (identified with the keyword S_PATH), measuring the position of the system along the path, is defined as:

$$s = Z^{-1} \sum_{i=1}^{N} i e^{-\lambda d(R_i, R(t))} \tag{3.23}$$

where $R(t)$ is the configuration of the system at any given time, d is a distance metric on χ, and $Z = \sum_{i=1}^{N} e^{-\lambda d(R_i, R(t))}$ is a normalization factor in which the prefactor, λ, should be chosen so as to have $\lambda d(R_i, R_{i\pm1}) \sim 2.3$ on average.

3.2.2.2 Choice of the Parameters

Here we discuss the choice of some parameters that are specific to a BE-MTD calculation. Part of this analysis is based on Ref. [90], where extensive benchmarks were performed using a coarse-grained force field, UNRES, comparing BE-MTD with standard REMD. The authors considered the PDB code 1E0G, a 48-residue-long alpha-beta protein that folds with UNRES into the experimental native state.

Frequency of exchange moves

Extensive simulations of 1E0G indicate that very frequent or very rare exchanges make BE-MTD marginally less efficient (see Figure 3.1). The optimal exchange time for this system is $\tau_{exch} = 120$ ps, which gives the system enough time to explore the local free-energy wells before a change in the direction of bias.

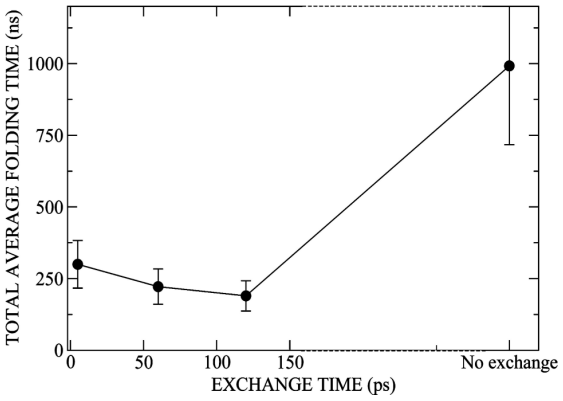

FIGURE 3.1: BE-MTD folding simulations on PDB 1E0G at $T = 280$ K with different τ_{exch}. The Gaussian height is held fixed $w = 0.012$ kcal/mol, and a set of 8 CVs has been used (adapted from Ref. [90]). The average folding times are shown.

Dimensionality of the bias potentials

The efficiency of standard metadynamics degrades with increasing dimensionality of the Gaussian hills, as it takes more time to fill with Gaussians a high dimensional space than a low-dimensional one. On the other hand, if the free-energy landscape of the system is intrinsically high dimensional and one uses only a one-dimensional bias, the metadynamics estimate of the free energy is affected by large errors. BE-MTD provides, at least in principal, a solution to this problem, as one can explore high dimensional free energy landscape by using several low-dimensional biases. Benchmarks indicate that using one-dimensional Gaussians increases the efficiency of the method by a factor of almost 2 compared with two-dimensional Gaussians [90]. The difference in performance might be due to the fact that the gain in the speed by which wells are filled with one-dimensional hills overwhelms the larger flexibility in finding complicated transition pathways with two-dimensional hills. The general validity of this assertion may however require to be checked on different systems.

Treatment of the boundaries: overcoming an artifact

When a metadynamics simulation reaches convergence, the history-dependent potential fluctuates around the negative of the free energy of the system, namely $V_G(s(r),t) \sim -F(s)$. The lack of stationary fluctuations may indicate serious convergence problems. One technical problem that has to be solved to allow convergence concerns the boundaries of the CV space. Metadynamics simulations typically take advantage of a finite and not too small width of Gaussian hills to fill the free energy surface quickly. However, finite width Gaussians can induce systematics errors at the boundaries of the CV space (whether natural or artificially imposed ones): these errors are due to the fact that a sum of finite-width Gaussian-shaped hills can not accurately reproduce discontinuities on the free energy profile. The presence of the discontinuities is common for several types of CVs that are intrinsically limited (e.g. to be non-negative, like coordination numbers, alpha/betarmsd, etc). At the beginning of the simulation these errors are small and are usually overlooked, but at long times they can become important, preventing the system from reaching stationary condition. A procedure to overcome this problem has been introduced in ref [91] that works only for one-dimensional biases, but that has the great advantage of being very robust and parameter-free. The solution is to change the metadynamics algorithm setting the bias force equal to zero beyond the boundary. For example, to compute the free energy as a function of a CV, s, for $s > s_w$ region, one still updates the history dependent potential according to Eq 3.12, but then sets

$$\frac{dV_G(s,t)}{ds} = 0, \qquad \forall s < s_w \tag{3.24}$$

Updating V_G implies that Gaussians are added also if $s < s_w$, as the tails of these Gaussians influence V_G in the region where it is relevant, $s > s_w$. In this way, the force on the system in region $s > s_w$ comes from both metadynamics and the force field, in the region $s < s_w$ only from the latter. This approach allows to reconstruct correctly the free energy. In Fig 3.2, an example of this treatment for a dialanine molecule is presented [91]. It is clear that the two estimates differ only on a tiny region close to the boundaries.

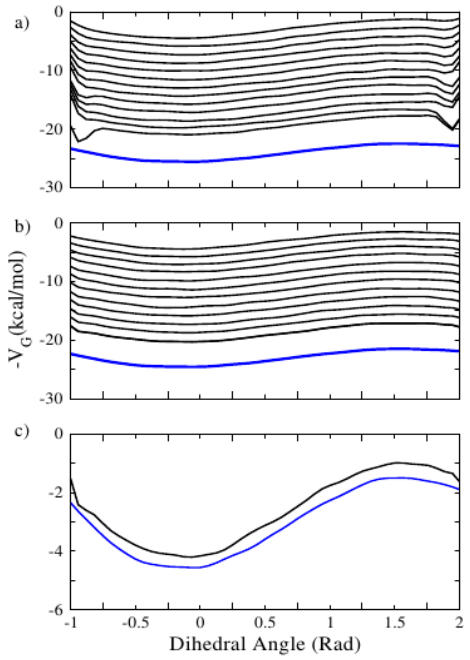

FIGURE 3.2: Time evolution of V_G for a metadynamics performed on a dialanine, using the dihedral ψ as a collective variable. The metadynamics parameters are $w = 0.1$, $\delta s = 0.2$, $\tau_G = 2$ ps. The blue line is the exact free energy of the system. a) Metadynamics is performed using parabolic wall with $K = 100000$ at $\psi = -1$ and $\psi = 2$ *rad*. Significant errors in the reconstruction appear close to the boundaries. b) Metadynamics is performed by setting the force equal to zero beyond $\psi = -1$ and $\psi = 2$ *rad*, according to the approach described in this section. The errors close to the boundaries are much reduced and do not become larger with simulation time. c) The average value of V_G in panel b is compared with the exact result (adapted from Ref. [91]).

Chapter 4

A Stable Ion-Binding Site

4.1 Overview

In this part of the thesis we addressed the issue of identifying a stable Na^+ binding site in vSGLT. Indeed, recent studies showed that the sodium ion spontaneously leaves its initial position diffusing outside the transporter, towards the intracellular space. This suggested that the crystal structure corresponds to an ion-releasing state of the transporters. Here, using metadynamics, we defined a more stable Na^+ binding site corresponding to an ion-retaining state of the transporter. In addition, our simulations, consistently with mutagenesis studies, highlight the importance of D189 that, without being one of the Na^+-coordinating residues, regulates its binding/release.

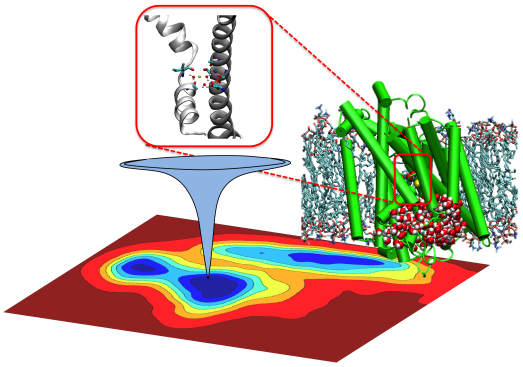

FIGURE 4.1

4.2 Introduction

As already discussed in Chapter 2, in 2008 Faham et al. [4] reported the crystallization and structural resolution of SGLT from *Vibrio parahaemolyticus* bacterium (that shares the 32% of its sequence with the human homologue), capturing the protein in the inward-facing galactose-bound conformation. Even if Na^+ ion was not detected by the X-ray experiments, a plausible binding site was proposed by a structural comparison with the LeuT structure [5], by considering the conservation of putative Na^+ binding residues among the SSS family and by mutational analysis [4]. However, recent molecular dynamics (MD) simulations studies suggested that the crystal structure represents an ion-releasing state of the transporter [6–8]. Indeed, starting from the binding site of the Na^+ ion proposed experimentally, after a few ns of MD simulation, the Na^+ spontaneously diffuses outside the protein, towards the intracellular space.
Here we focused on identifying of a stable binding site for the ion in the inward-facing conformation, corresponding to a putative 'ion-retaining' state by performing classical MD and metadynamics (MTD) [9, 10] simulations to explore the free energy surface and the conformations involved in the binding and dissociation of the Na^+ ion.

4.3 Computational Details

4.3.1 Building the Model

The model of vSGLT was built using the chain A of the ∼3.0 Å resolution crystal structure (PDB accession code 3DH4 [4]). The first helix, solved only for the backbone atoms of residues S3-Y19, was eliminated and it was then reconstructed taking these residues from a more recent crystal structure in which they were solved (PDB code 2XQ2 [8], 2.7 Å resolution). We did not choose this latter crystal as the starting structure of our simulations since it represents the inward-open ligands-free conformation of the transporter and, therefore, it was not useful for the aim of our study. In the whole work we preferred to use the original numbering of the helices of the protein, i.e. from helix 1 to helix 14 [4]. The missing atoms of side chains of residues K124, V185, R273, K454, K547 were reconstructed using SwissPDBViewer [92] application. All the Hys residues, far from the binding sites and oriented towards the solvent, were kept protonated in ϵ position. The Asp and Glu residues were considered in the deprotonated form, while the Arg and Lys residues in the protonated form. Residues S31 to L46, located between transmembrane helix (TM) 1 and TM2, and residues Y179 to A184, between TM5 and TM6, were constructed using Loopy [93] program. This latter loop is far from the ion binding site and does not affect our results. The final monomeric

structure contained 539 residues (S9 to K547). The protein was embedded in a pure, pre-equilibrated 1-palmitoyl-2-oleilphosphatidylcholine (POPC) lipid model (kindly supplied by T. A. Martinek) [94, 95] using the g_membed [96] tool of Gromacs and then it was oriented following OPM [97] database model. Afterwards, the system was neutralized and solvated with TIP3P model [98] water molecules (80969 total atoms in a box size of 97.6 x 96.7 x 85.1 $Å^3$). In Figure 4.2 is reported an overview of the whole simulated system.

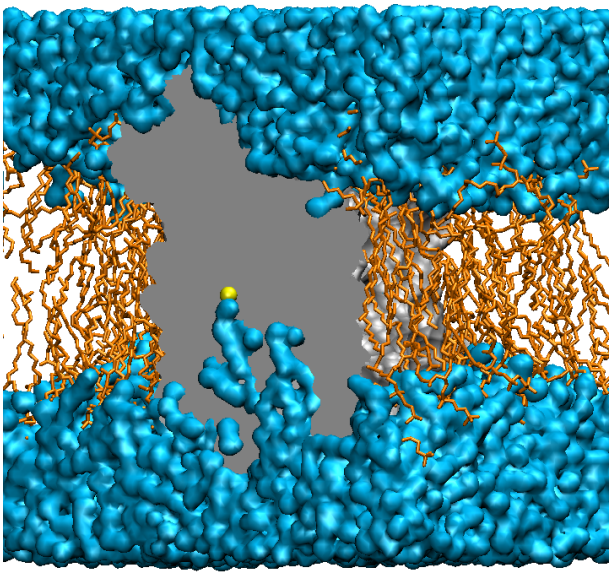

FIGURE 4.2: Overview of the simulated system. A section of vSGLT protein is shown in gray. From this view, the hydrophilic cavity of the protein, filled with many water molecules coming from the cytoplasmic side (the lower part of the figure with respect to the ion), is visible. The sodium ion in its binding site is colored in yellow. The POPC lipids are drawn in thick orange lines and water molecules in blue spheres.

4.3.2 Molecular Dynamics Simulations

Simulations were carried out with GROMACS4 [99] package using Amber03 [100] force field for protein and GAFF [101] for membrane and galactose; for this latter, RESP [102] charges were calculated fitting an electrostatic potential calculated using the 6-31G* basis set and B3LYP [103, 104] functional and the Gaussian03 [105] program. The simulations were done in periodic boundary conditions at 310 K using the Nose-Hoover thermostat [106] and Parrinello-Rahman barostat [79, 80], described in Chapter 3, with a semiisotropic pressure coupling type and a time step of 2 fs.

Position restraints of atoms were fixed with a force constant (K) equal to 1000 kJ mol^{-1} nm^{-2}. Distance restraints were imposed using Umbrella Sampling keyword as implemented in Plumed [107] and considering the distances measured in the crystal structure as reference values and a K=20000 kJ mol^{-1} nm^{-2}. The equilibration was performed in three stages: (1) the system was heated for 3 ns with protein backbone

and ligands completely fixed, while side chains were left free to move. (2) 5 ns were run using distance restraints between the ligands and a few close residues. These two stages were conducted in a NPγT ensemble with a surface tension equal to 600.0 bar*nm [94]. (3) Then the system was simulated for 26 ns where the membrane area was kept constant (Area Per Lipid around 64 Å^2).
15 MD production runs of different lengths (one of 125 ns, 2 runs of 70 ns, 12 runs of 12 ns) were simulated. During all the simulations, we took care in particular that important parameters of the membrane, such as the Area Per Lipid (APL) and thickness, were in agreement with experimental values [108].

4.3.3 Metadynamics Simulations

MTD was used in order to explore the FES and the conformations involved in the binding and dissociation of the Na^+ ion. The approach in described in detail in Section 3.2.1. Two different CVs were found to describe this process: the coordination number of the ion with the carbonyl oxygens of A62, I65, A361, and the hydroxyl oxygens (Oγ) of S364, S365, the five closest residues in the crystal binding site, and the distance between Na^+ ion and the carboxylate group of D189. This latter is the only charged residue in the vicinity of the putative binding site and it has been also previously observed to interact with the sodium ion along its exit pathway [7, 8]. The functional form of these CVs is provided in Section 3.2.2.1, where, for the coordination CV, we used $n = 6$, $m = 12$ and $r_0 = 0.4$ nm. We remark that we considered also different sets of CVs (i.e. the water coordination of the Na^+ and the distance between two charged residues, D189-R273, that could have moved closer the two helices, TM6-TM7, making a saltbridge). However, the set previously described is the only one able to correctly describe the binding/release of Na^+ in a relatively short simulation time. 10 MTD simulations, of about 10 ns each, were performed using PLUMED 1.2 [107] in NVT ensemble. Gaussians width (δs) was 0.1 for CV1 and 0.02 nm for CV2, and the height (w) equal to 0.4 kJ/mol for both CVs. The repulsive gaussian potentials were added every 5 ps. These values were found to be optimal for the correct FES reconstruction (see Section 3.2 and Ref. [10]). To exactly estimate the exit barriers from the two most important minima determined by MTD we followed the procedure described in Ref. [10]. Namely, we performed six independent MTD simulations starting from a configuration corresponding to different minima with initial velocities drawn at random from a Maxwell-Boltzmann distribution at 310 K. Then, we calculated the averages of the barriers and estimated the accuracy as the standard deviations of these values.

4.4 Results

4.4.1 Molecular Dynamics Simulations

After the setup of the system (see Section 4.3.1), we carried on an unbiased MD simulation aimed at elucidating the features of the experimental structure. We first analyzed the mechanism of water permeation [13, 32, 109]. As mentioned in Chapter 2, two different transport mechanisms have been proposed for this symporter: the water cotransport mechanism [110, 111], in which water is actively cotransported with Na^+ and glucose, and the local osmotic gradient hypothesis [112], in which water molecules are transported passively through permeation pathway. Nevertheless, the last theoretical and experimental studies [32] support the passive mechanism. Consistently with these studies [32], a visual inspection of our simulations shows that water molecules permeate relatively easily through the protein following a pathway [32] that presents a small gap of low water density localized in proximity of residues Y87 (on helix TM3), A259 (TM7), Q425 and Q428 (TM11), F479 and M483 (TM13).

We then verified whether the crystal structure of vSGLT represents an ion-releasing state of the transporter, consistently with the proposal of Li and Tajkhorshid [6]. Indeed, according to previous theoretical studies [6–8], our simulation shows that Na^+ leaves its binding site in a few ns. In particular, initially the sodium ion interacts with D189, I65 and A62, but is loses its coordination with I65 after 1 ns. Then, it starts increasing its hydration (up to 3-4 water molecules). As it has been already shown for LeuT [36], water molecules can penetrate from the cytoplasm into the protein core, entering through the hydrophilic cavity. This increases the local hydration of the Na^+ ion, probably facilitating its release from the Na2 site. In fact, after 6 ns, Na^+ exits the binding site. It moves towards the cytoplasm along TM9, passing close to the TM5-TM6 loop and finally exiting the protein after 12.5 ns. In order to evaluate the behavior of galactose we extended our simulation up to 125 ns. In spite of previous observations [8, 11], reporting spontaneous unbinding events of the substrate, in our case the galactose remained stably placed inside its binding site, conserving all its starting interactions. Indeed, the distances of the substrate from some of the residues important in keeping it in place through hydrogen bonds are (values in Å): Gal(O2)-E88(CD) 4.7 (±0.9); Gal(O4)-N260(CG) 4.6 (±0.7); Gal(O5)-Q428(CD) 4.0 (±0.5); Gal(O4)-Y87(O) 3.6 (±0.6). Also the Y263, considered the intracellular gate [4], stably interacts with N64, another galactose hydrogen-bonding residue (N64(ND)-Y263(OH) 3.1 (±0.2)).

4.4.2 Metadynamics Simulations

To investigate which residues form a stable sodium binding site and affect ion specificity in the ion-occluded state of vSGLT, we performed six independent MTD simulations. The free energy landscape explored is similar in all the independent metadynamics simulations (Figure 4.3). There are three distinguishable minima. Two minima correspond to high values of CVs and they are close to each other. The third one, corresponding to lower values of CVs, is located far from the first two and it is separated from them by a high barrier. By a cluster analysis, performed on all the MTD trajectories, in the regions corresponding to the three minima of the FES, we identified three putative binding states. All these states show the expected octahedral coordination for the Na^+ ion (Table 4.1). For sake of clarity, the different binding states were classified according to the coordinating residues. The first two states show a Na^+ ion coordinated by 4 residues and 2 water molecules. In particular, the first state (hereafter named Highly Coordinated 1 or HC1) involves residues I65 (on TM2), A361, S364 and S365 (on TM9); in the second state (hereafter named Highly Coordinated 2 or HC2) Na^+ interacts with I62 and A62 (on TM2), A361 and S365 (on TM9). The chosen CVs cannot distinguish between HC1 and HC2, which correspond to the same minimum of the FES. Except for the hydroxyl oxygens (Oγ) of the serines, the other residues coordinate the ion with the oxygens of their carbonylic moieties. The coordinating distances Na^+-O are in the range of 2.3-2.7 Å, in good agreement with the experimental data of ion binding sites of LeuT and other ion channels or transporters [5, 113] (see Table 4.2). Instead, the Na^+-O distances in the crystallografic structure of vSGLT (PDB code 3DH4) are significantly larger, in the range 3.1-3.6 Å (Table 4.2). This is consistent with the hypothesis that the experimental structure corresponds to an ion-releasing state. Another important minimum (hereafter Low Coordinated 1 or LC1) corresponds to a configuration where the ion is coordinated by 3 residues and 3 water molecules (3+3). The residues forming the coordination shell are A62, I65 and S365 and they bind to Na^+ with the carbonyl oxygens of A62 and I65 and hydroxyl oxygen of S365, as in HC1 and HC2. In this state the coordination distances Na^+-O are in the range of 2.3-2.6 Å, marginally smaller than in HC1-2. The three coordinated water molecules exchange with other waters during the simulation. Moreover, one of them makes simultaneously a H-bond with Na^+ and D189. The third minimum, visible in Figure 4.3, shows the Na^+ ion poorly coordinated by protein residues (A62 and S364 forming the binding site hypothesized on the base of the crystal structure, and also one or two Oδ of D189, close to the exiting pathway). This minimum (called hereafter Poorly Coordinated or PC) is quite close to the hydrophilic cavity opened towards the cytoplasm, formed from the intracellular portions of helices TM2, TM3, TM4, TM7, TM9 and TM11 (Figure 4.4). Therefore, it was not

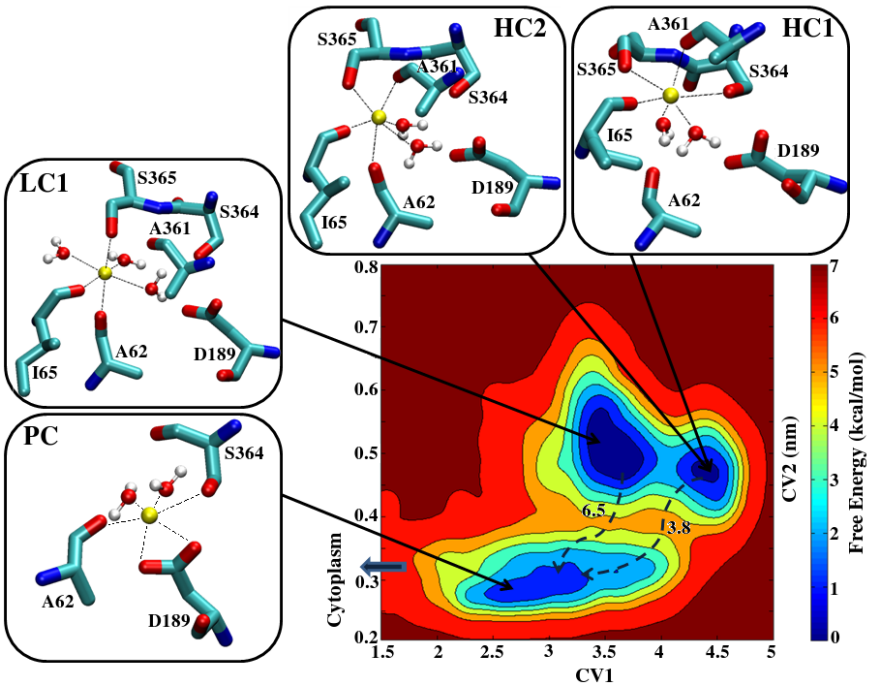

FIGURE 4.3: Free energy surface with the coordination states shown for the three minima. In the axes: CV1, the coordination number of the ion with the five closest residues in the crystal binding site (A62, I65, A361, S364, S365); CV2, the distance between Na^+ ion and D189. Anticlockwise boxes: HC1 state, in which the ion interacts with residues I65 (on TM2), A361, S364 and S365 (on TM9); HC2 state, with residues I62 and A62 (on TM2), A361 and S365 (on TM9); both these states involve also two coordinating water molecules. Then LC1 state, where the ion is coordinated by 3 residues (A62, I65 and S365) and 3 water molecules. The third minimum reports the PC state, coordinated by A62 and S364 from the crystal binding site, one or two Oδ of D189 and two water molecules. Except for the hydroxyl oxygens (Oγ) of the serines, the other residues coordinate the ion with carbonyl oxygens. The barriers on the path arrows are expressed in kcal/mol.

considered as a good ion-retaining candidate. Indeed, MD simulations started from this site unavoidably lead to the release of Na^+ in the cytoplasm.

4.4.3 Validation of Metadynamics' Results

This FES in Figure 4.3 has been obtained by metadynamics with a bias acting on two variables, ending the simulation as soon as the Na^+ ion reached the cytoplasm, well before the dynamics in CV space becomes diffusive [10]. In these conditions the FES obtained by metadynamics is accurate for what concerns the position of the free energy minima [9], but not for their relative depth and the height of the barriers. Thus, to determine exactly the barrier to exit the most important minima (HC1-2 and LC1) we followed the procedure described in Ref. [10]. Namely, we performed three MTD

States \ Residues	A62	I65	A361	S364	S365	D189	W
HC1		X	X	X	X		2
HC2	X	X	X		X		2
LC1	X	X			X		3
PC	X			X		X	2

TABLE 4.1: Protein residues included in the coordination sphere of Na^+ in the four different coordination states (HC1, HC2, LC1 and PC) identified analyzing the MTD simulations.

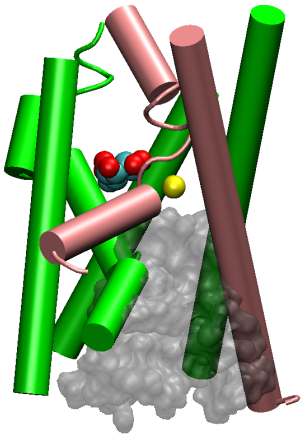

FIGURE 4.4: Part of the protein in cartoon, showing TM2 and TM9 helices forming sodium binding site, colored in pink. The ion is represented in yellow, while galactose in cyan and red. The hydrophilic cavity is shown in gray.

simulations starting both from HC1 and LC1, assigning different initial velocities (see Section 4.3).

The barrier that Na^+ has to overcome to leave the HC1 or HC2 sites towards the exiting PC site is of 3.8 (±0.4) kcal/mol; the exit pathway of the ion from these sites does not involve a passage from LC1. It involves a transient interaction with A62, S364, S365 and D189 and two water molecules (i.e. PC ligands plus S365), and it ends directly in PC site, after losing the coordination with S365. The exit barrier from LC1 is significantly higher: 6.5 (±0.4) kcal/mol. The exit pathway from LC1 also ends directly in PC, without passing through HC1-2. These results suggest that LC1, despite its lower Na^+-protein coordination, is marginally more stable than HC1-2, leading to the consideration that this site is a more stable ion-retaining state than the one proposed on the basis of the crystal structure and bioinformatics analysis. With our metadynamics simulations we performed only an exploration of the FES relative to small structural rearrangements of the residues surrounding the binding site. This does not rule out the existence of other Na^+ binding sites more stable than the one we identified in this work, which may require larger conformational changes with respect to the initial structure.

Symporter	PDB code	Cystallographic distances between Na^+ ion and coordinating ligands				
vSGLT		A62 (O)	I65 (O)	S365 (Oγ)	A361 (O)	S364 (Oγ)
	3DH4	3.72	3.33	3.64	3.15	3.10
vSGLT[1]		**A62 (O)**	**I65 (O)**	**S365 (Oγ)**		
	This work	**2.42 ($\pm$0.12)**	**2.35 ($\pm$0.09)**	**2.62 ($\pm$0.19)**		
LeuT [2]		A22 (O)	N27 (Oδ)	T254 (Oγ)	T254 (O)	N286 (Oδ)
(Na1)	2A65	2.11	2.18	2.28	2.39	2.54
LeuT		G20 (O)	V23 (O)	A351 (O)	T354 (Oγ)	S355 (Oγ)
(Na2)	2A65	2.21	2.11	2.25	2.21	2.32
GltPh		G306 (O)	N310 (O)	N401 (O)	D405 (Oδ1)	D405 (Oδ2)
(Na1)	2NWX	2.70	2.61	2.73	2.71	2.77
GltPh		T308 (O)	M311 (S)	S349 (O)	T352 (O)	
(Na2)	2NWX	2.75	3.43	1.96	2.48	
Mhp1		A38 (O)	I41 (O)	A309 (O)	S312 (Oγ)	T313 (Oγ)
(Na2)	2JLN (out)	2.62	2.66	2.09	2.53	2.77
	2JLO (occl)	2.84	2.79	2.56	2.62	2.65
	2X79 (inw)	residues far apart from each other, no ion				
BetP		A147 (O)	M150 (O)	F464 (O)	T467 (Oγ)	S468 (Oγ)
(Na2)	4AIN	2.17	2.48	2.33	2.31	2.36

[1] The three coordinating water molecules are not reported.
[2] The sixth coordinating ligand in LeuT Na1 site is represented by the substrate, Leu (2.52 Å).

TABLE 4.2: Distances (expressed in Å) for sodium binding sites in the crystal structures of several Na^+ symporters. Except for vSGLT, we report the residues identified in the crystal structures within 3 Å from the ion. For vSGLT we report (in bold) also the average distances and standard deviations of the coordinating residues.

In order to further validate this prediction, considering that multiple short MD simulations enable a better exploration of the conformational space with respect to a single long simulation [114], we performed 14 independent unbiased MD runs starting from different configurations and monitoring the conformational states explored by the dynamics. In seven simulations (4 performed starting from HC1 and 3 from HC2), the Na^+ ion, after spending few nanoseconds on HC2 state, moves to LC1 and remains in this configuration for the entire simulation time (12 ns). The variation of the CVs values during one of these simulations and their dependence on time have been reported in Figure 4.5. In two additional simulations, started from LC1, the ion remains in that configuration for 70 ns without observing any modification (Figure 4.6). The last five runs, two started from HC1 and the other three from HC2, quickly lead the ion to interact with D189 and to the loss of its original coordinating ligands in a few ns. These observations confirm that HC1 and HC2, despite their larger protein coordination ligands, are marginally less stable than LC1. The barrier we find to escape from this site are consistent with literature data of other symporters [32, 37, 38, 64, 73]. Indeed, the free energy barrier for the Na^+ dissociation from Na1 site in LeuT has been recently estimated to range between 14 and 6 kcal/mol [37], depending on the extent of outward open conformation of the transporter. Interestingly, the minimal value of this barrier is consistent with our results. The binding affinities calculated recently for the ion in the Na1 site of GltPh

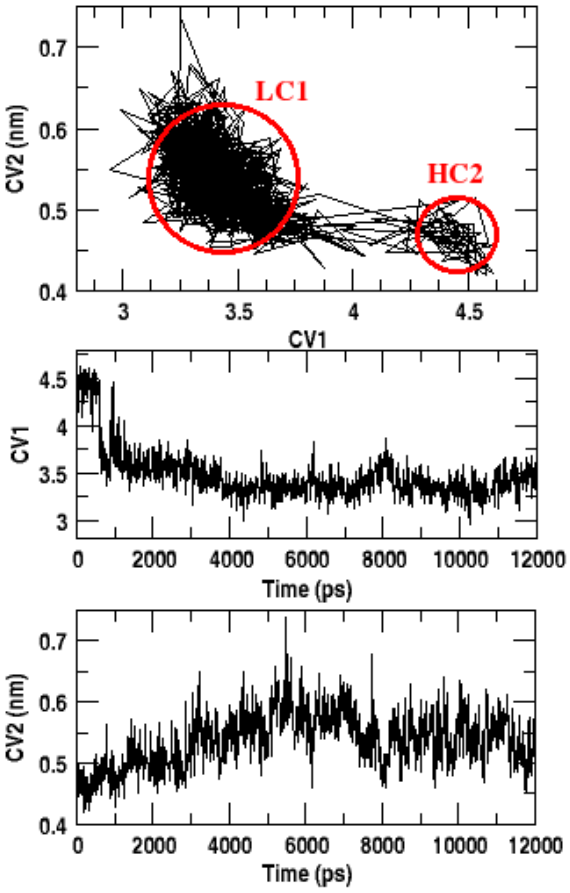

FIGURE 4.5: One of the 12 ns-long MD simulations started from HC2 site, reporting the values assumed by the two CVs (top), CV1 values as a function of time (center) and CV2 values as a function of time (bottom). After the first ns, sodium ion moves towards the LC1 site, remaining stably placed until the end of the run. On the top panel is visible that LC1 is a more populated state than HC2.

are also consistent with our hypothesis [38]. Moreover, our values are in line also with the barriers estimated for ion conduction via membrane proteins estimated in computational studies [64, 73]. Indeed, theoretical studies conducted on the X-ray structure of the KcsA K^+ channel showed that the largest free energy barrier for the process of ion conduction is on the order of 2-3 kcal/mol [64]. Similar values were found also by Furini et al. [73] in a theoretical work aimed at studying the conduction of Na^+ in bacterium sodium channel.

4.4.4 Final Remarks

The three binding states for the Na^+ ion identified in this work (HC1-2, LC1) share the interaction with the hydroxyl oxygen (Oγ) of S365 on TM9. Remarkably, this residue is conserved throughout the SSS family. Most importantly, mutation S365A leads to a

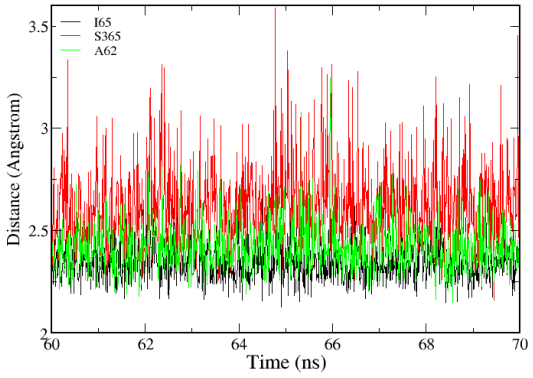

FIGURE 4.6: Distances (Å) between sodium ion and the oxygen of coordinating aminoacids (Oγ for serine, carbonyl oxygens for the other residues) in the last 10 ns of a 70 ns MD simulation.

complete abrogation of Na^+-dependent transport [4].

Comparing the crystal structure of vSGLT and the protein conformation assumed when the Na^+ ion is in LC1 state, we couldn't find any relevant difference in the orientation of the helices. However, we noted that, when Na^+ is in LC1, D189 is closer to the binding site with respect to the position occupied in the crystal structure (Figure 4.7). This conformation is stabilized by a hydrogen bond with S364 (Oγ). Moreover, a water bridge between Na^+ and D189 is present both in LC1 and HC1-2. In addition, in the simulations where the Na^+ was unstable we observed a rotation of D189 towards the hydrophilic cavity that seems to be related with the exit of the ion. This residue is highly conserved throughout the SSS family and experimental studies have highlighted its importance in a correct Na^+-galactose cotransport and in the cation selectivity [115, 116]. Therefore, we hypothesize that D189 plays an important role in stabilizing the ion-retaining LC1 state.

Finally, molecular mechanics simulations strongly depend on the force field parameters especially when metal ions are considered [117, 118]. To test the dependence of our results on the ion force field adopted [100], we repeated one MTD simulation considering the ion force field parameters developed by Cheatham and coworkers [119]. Also in this case, we found results (the same free energy picture with the three minima containing the binding states described previously) similar to the other MTD simulations. This excludes a strong influence of the force field parameters on the FES.

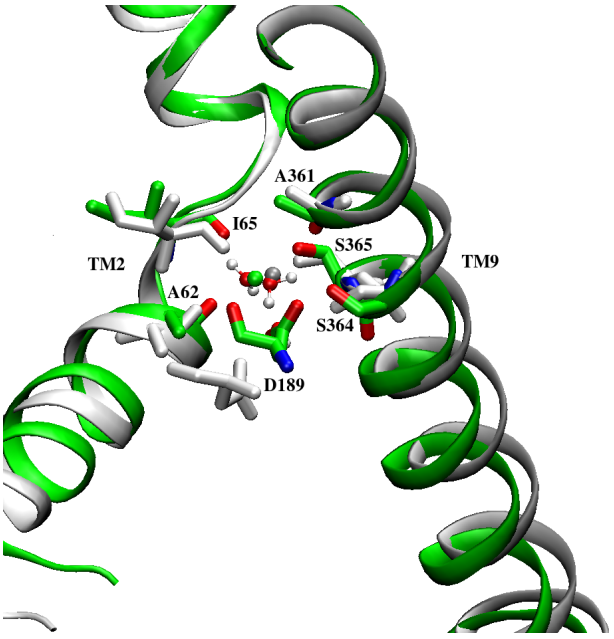

FIGURE 4.7: Overlap of crystal structure, colored in light gray, and LC1 conformation, in green. The most notable difference concerns the position of residue D189, much closer to the binding site in LC1.

4.4.5 Conclusions

In the last years several efforts have been done to investigate the ion binding sites of the LeuT-fold secondary active transporters [34, 36–38, 42]. In this study we identified a putative ion-retaining state of vSGLT [4, 8], which is more stable than the one proposed on the basis of the crystal structure and bioinformatics considerations and mutational studies [4, 8]. In fact, several theoretical works performed on this transporter showed a spontaneously-diffusing ion outside the protein, leading to the hypothesis that the crystal structure represents an ion-releasing state of vSGLT [6–8]. By performing force field based MD and MTD simulations we identified the LC1 binding site, a site in which the Na^+ ion has an octahedral coordination geometry with three water molecules and three residues A62, I65 (carbonyl oxygens) and S365 (hydroxyl oxygen) forming its coordination shell. This state turns out to be stable for long (70 ns) MD simulations and is characterized by a barrier of approximately 6 kcal/mol to leave the coordination site. The coordination distances (Table 4.2, Figure 4.6) and the calculated barrier that the Na^+ ion has to overcome to exit from this minimum are consistent with literature data [32, 37, 38, 64, 73]. The identification of this stable, ion-retaining binding site is very important not only for better understanding the mechanism of this symporter, but also for identifying sodium binding site in other 5HIR members for which crystallographic evidence is not available. Furthermore, a stable ion-binding site is fundamental for investigating the concerted binding/release mechanism of both ligands, the ion and the galactose. We characterized this part of the transport mechanism in the next Chapter.

Chapter 5

Releasing mechanism of Gal and Na^+ from vSGLT

5.1 Overview

In this part of the thesis we study some aspects of the transport mechanism of vSGLT symporter addressing, in particular, a detailed kinetic and thermodynamic characterization of the exit path of the two ligands. We use bias-exchange metadynamics, the approach described in Section 3.2.2, to characterize the free energy profile of the galactose and Na^+ release processes towards the intracellular side. Our study shows that (i) the Na^+/Gal interplay along the dissociation path is minimal and it is limited only to the initial displacement of both Na^+ and Gal from their binding sites; (ii) the dissociation of both Na^+ and Gal occurs with free energy barriers of about 11-12 kcal/mol, and the rate limiting step is associated to conformations in which Na^+ and Gal are more than 10 Å far from their binding sites; (iii) no gating role can be assigned to Y263. Simulation of the Y263F mutant reveals a rather significant change in the binding site of Gal, confirming that this residue has an important functional role, even if it does not act as a gate. On the basis of these results we propose a branched six-state alternating access mechanism, which, according to bioinformatic analysis, may be shared also by other members of the LeuT-fold transporters.

5.2 Introduction

The dissociation mechanism of galactose in vSGLT has been investigated by molecular simulations studies which showed that Gal release occurs either spontaneously or by applying an external force [7, 8, 11]. These studies lead to contradictory conclusions on a possible gating role of Y263, on the exact conformational state of the transporter (open or occluded) captured crystallographically and on the free energy profile of Gal release [7, 8, 11]. Namely, Zomot et al. showed that Gal exited the protein only by using steered molecular dynamics (SMD), after the rotameric transition of the side chain of Y263, which, according to this study, acts as a gate. A second gate represented by Y269 was also encountered later on along the path [7]. Consistently with these findings, Watanabe et al. showed that the sodium exit triggers the substrate release after the new rotameric conformation acquired by Y263 and that Gal has to overcome very small barriers ($\Delta G^{\#}$ $\sim$2 kcal/mol) along its exit pathway [8]. A different scenario was instead provided by Li and Tajkhorshid in 2011. By combining MD and SMD simulations, they identified a curved translocation pathway for Gal release. In this path Gal moves around Y263, requiring no gating event. This study pointed out that the crystal structure represents an open state of the transporter [11]. Unfortunately, experiments do not help solving the puzzle, as data on the order of dissociation of the two ligands are incomplete [120]. We here perform bias exchange metadynamics simulations (BE-MTD) [86, 88, 121], the approach described in Section 3.2.2, to establish the reciprocal influence of the Na^+ and Gal in their dissociation mechanism and at characterizing the kinetics and thermodynamics of the process.

5.3 Computational Details

5.3.1 System Setup

We used the same setup described in Section 4.3.1. In short, the model of vSGLT, built using the chain A of the 3 Å resolution crystal structure (PDB accession code 3DH4 [4]), was embedded in a pure, pre-equilibrated 1-palmitoyl-2-oleilphosphatidylcholine (POPC) lipid model (kindly supplied by T. A. Martinek) [94, 95] using the g_membed [96] tool of GROMACS4 [99] and then it was oriented following OPM [97] database model. Afterwards the system was neutralized and solvated with TIP3P [98] water molecules (80969 total atoms in a box size of 97.6 x 96.7 x 85.1 $Å^3$). Simulations were carried out with GROMACS4 package using Amber03 [100] force field for protein, GAFF [101] for galactose and for membrane the parameters supplied by T. A. Martinek [94].

5.3.2 Bias-Exchange Metadynamics Simulations

The starting point of this study was the structure of vSGLT in the ion-retaining state obtained from our previous MTD simulations, described in Chapter 4. In order to study the binding/dissociation of galactose and its coupling with the binding/dissociation of sodium ion by using BE-MTD, we exploited nine different collective variables (whose functional form is described in Section 3.2.2.1) using the Plumed plugin [107]. Seven of them were reserved to the Gal. In particular, to assess the controversial role of Y263 we used:

1) a combination of two dihedral angles of Y263 (C-Cα-Cβ-Cγ and Cα-Cβ-Cγ-Cδ) using the dihedral similarity CV;

2) the hydrogen bonds between Y263 and N64.

To focus on the dissociation of Gal from its binding site we used:

3) the distance of galactose from its binding site (represented by the center of mass (COM) of selected residues, see Table 5.1);

4) the hydrogen bonds between galactose and its binding site;

5) the hydrogen bonds between galactose and the likely exiting pathway [7, 8, 11];

6) the radius of gyration of a group of atoms belonging to the galactose binding site;

7) a path collective variable (where $\lambda = 53.5\ nm^{-2}$), that describes the progression of the galactose along its exit channel [89]. This variable, added after a spontaneous exit of the galactose was observed using the other variables, is defined by a set of 7 reference conformations.

Two additional variables were used to characterize the release mechanism of Na^+:

8) the distance between the sodium ion and its binding site (defined by the COM of selected residues, see Table 5.1);

9) the coordination number between sodium and four residues of the binding site (carbonyl oxygens of I65 and A361 and hydroxyl oxygens of S364 and S365), where $n = 10$, $m = 20$ and $r_0 = 0.4$ nm.

The BE-MTD calculations were conducted in the NVT ensemble. Moreover, some parabolic restraints (walls) (force constant, K, 5000 kJ mol^{-1} nm^{-2}) were introduced, in order to avoid the sampling of regions which were not relevant for the ligand dissociation process, see Table 5.1 In order to reduce the error close to the boundaries we applied the approach in Ref. [91], see Section 3.2.2.2. The gaussians were added every 5 ps with a height (W) equal to 0.08 kJ/mol for all the CVs. The exchanges between the bias potentials of the different variables were periodically attempted every 20 ps. We started the simulation with 8 walkers using all CVs except the path collective variable. After 740 ns we added other 4 walkers for the distance between Gal and the center of mass of its site; at 820 ns we introduced 4 more walkers biased on the path collective variable of the Gal.

CV	δs	Upper Wall	Interval
1. Dihedral angle Y263	0.025	-	0.1 - 1.9
2. H-Bonds Y263-N64	0.35	-	0.1 - 2
3. Distance Gal-site	0.05	3.5	0 - 3.5
4. H-Bonds Gal-site	1	-	1 - 30
5. H-Bonds Gal-path	1	-	1 - 31
6. Radius of Gyration Gal site	0.02	1.05	0 - 1.05
7. PCV	0.3	-	1.2 - 6.8
8. Distance Na^+-site	0.05	3.5	0 - 3.5
9. Coordination Na^+ site	0.1	-	0.1 -3.5

TABLE 5.1: CVs used in this study. The units of the CV3, CV6 and CV8 are expressed in nanometers, CV1 in radiants. The galactose binding site was defined by the Cα of residues Q69, N260, K294 and Q428. The ion binding site was defined by the Cα of A62, I65, S66, V363, S364 and S365.

In this way, we simulated a total of 1400 ns using at most 16 walkers. Parameters like gaussians width (ds), intervals, walls were added/changed and adapted for a better and faster convergence of the simulation (Table 5.1). Simulations of ligands dissociations were also performed in the absence of Na^+ in order to assess the role of the ion. Moreover, we have performed an additional simulation starting from the Y263F mutant to clarify the role of this residue in shaping the free energy landscape. At this scope we elongated the BE-MTD simulation mutating Y263F (for a total of 15 ns*16 CVs = 240 ns), maintaining all the CVs and their parameters used in the wild type (WT) simulation.

5.3.3 Analysis

All structural and free energy analyses were performed using METAGUI [122], a VMD [123] interface for analyzing metadynamics and molecular dynamics simulations.
The error of the free energy profiles was calculated as the standard deviation of two different time averages of the biased potential in the first and the second part of the converged interval of the simulation.
In order to assess the role of selected residues along the ligands dissociation paths we calculated the average interaction energies at the relevant minima and transition states. We considered the van der Waals and coulombic interactions. We remark that this analysis is qualitative and is meant only to provide a picture of the role of selected residues in the relative stabilization/destabilization of transition states and minima, as shown in other studies [124, 125].
H-bond analysis was performed using Plumed [107].
In order to understand the role of the residues that turned out to be involved in the exit pathway, we also performed a bioinformatic analysis. We conducted sequence alignment

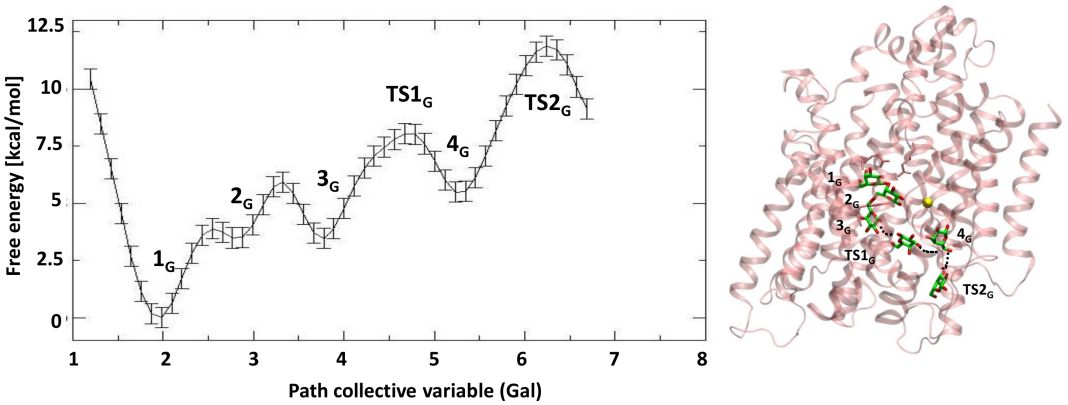

FIGURE 5.1: Exit pathway of galactose. On the left is shown the free energy profile (kcal/mol) along the path collective variable representing Gal exit. On the right it is reported the position of the Gal in the different minima along exit pathway on the top of the protein. This latter is depicted in pink cartoons, the substrate in licorice, while the Na^+ is depicted as a yellow sphere. The curved path is connected by dark dashed lines

of the Solute Sodium Symporters (SSS) family (namely, the human homolog hSGLT1, the Na^+/I^- transporter, NIS, and the Na^+/L-proline transporter, PutP) with ClustalW [126]. We also performed structural alignment of the vSGLT, Mhp1, BetP and LeuT transporters using the HHpred server [127]. Other alignments found in the literature were used to validate and complete our analysis [4, 34].

5.4 Results

5.4.1 Galactose Exit Pathway

We first investigated the dissociation path of Gal. The projection of the free energy along the path collective variable, representing the progression of Gal along its exit channel, is reported in Figure 5.1. For sake of clarity, we add a subscript $_G$ for the states relative to the Gal exit path, and a subscript $_{Na}$ for those relative to the ion path. Initially, Gal is in the deepest minimum (Min 1_G), which is the same binding site identified in the crystal structure (the RMSD of the heavy atoms of residues within 4 Å of Gal with respect to the conformation in the crystal structure is 1.0 Å ($\pm$ 0.1)). The substrate is stabilized by an extended H-bond network with E88, Q428, Q69, E68, N64 (Table 5.2). Residue Y263 (OH) interacts with N64 (HNδ). There are 2-3 water molecules in the binding site, interacting with the substrate. While Gal is in its binding site, Na^+ is coordinated by three water molecules and three residues (A62, I65 and S365). This corresponds to the ion-retaining binding site of the inward-facing conformation of vSGLT identified in Chapter 4, the LC1 site.

The next free energy minimum along the exit path of the substrate is Min 2_G, where,

HB	Mean (st.dev.)	% ≤ 3 Å	% ≤ 3.2 Å
	Min 1_G		
E68OE1-GalO1	2.70 (± 0.27)	94.6	95.0
E88OE2-GlaO3	2.62 (± 0.095	100.0	100.0
Q428NE2-GalO6	3.88 (± 1.14)	22.7	44.1
N64ND2-GalO2	3.28 (± 0.27)	13.2	45.5
N64ND2-Y263OH	3.09 (± 0.21)	38.2	75.0
Q69NE2-GalO2	3.03 (± 0.15)	50.0	86.4
	Min 2_G		
GalO6-T431OG	2.92 (± 0.26)	72.9	90.6
GalO2-N64O	2.92 (± 0.21)	68.8	92.7
S66N-E68OE2	2.78 (± 0.07)	100.0	100.0
S66OG-E68OE1	2.76 (± 0.33)	88.5	89.6
N64ND2-Y263OH	3.17 (± 0.26)	28.1	64.6
	Min 3_G		
GalO1-T431OG	3.40 (± 1.17)	50.6	74.7
S66N-E68OE1	3.10 (± 0.20)	35.6	71.3
S66N-E68OE2	2.81 (± 0.12)	95.4	97.7
S66OG-E68OE1	2.80 (± 0.42)	87.4	87.4
GalO2-N142OD	4.42 (± 1.94)	44.8	47.1
GalO3-N142OD	3.84 (± 1.41)	57.5	58.6
GalO5-Y262OH	3.62 (± 1.32)	36.8	54.0
	Min 4_G		
GalO1-D189OD2	2.76 (± 0.78)	90.4	97.9
GalO2-D189OD2	2.77 (± 0.58)	94.9	98.6
GalO4-S368OG	2.98 (± 0.55)	68.9	90.4
GalO6-A184O	3.06 (± 0.94)	74.4	87.7
	$TS2_G$		
GalO3-N371OD	3.54 (± 1.09)	52.1	64.6

TABLE 5.2: The average distances (Å), the standard deviation and the life time (%) of relevant H-bonds along the Gal dissociation path.

consistently with the suggestion of Li et al. [11], Gal undergoes a lateral displacement towards a position in which it is only partially shielded by the ring of the Y263. From this point, the substrate will find its way out by rotating about 90 degrees (assuming a conformation in which its ring is roughly parallel to the protein axis) and continues his progression along a curved path beyond Y263. This residue is at the edge of the hydrophilic cavity and the presence of water molecules confers flexibility to it, which hence is able to accomodate to the passage of Gal (Figure 5.2). In this new position, several water molecules enter the binding site, while the substrate (C6-O) interacts with T431 (HOγ) and, through a water bridge, with N142. In this minimum residue E68 assumes a new rotameric conformation. Indeed, its side chain, initially heading towards the Gal binding site, rotates towards the Na^+ binding site, making one or two H-bonds with S66, a conserved residue across the SSS (E68 (Oδ1) with S66 (HN) and E68 (Oδ2) with S66 (HOγ)), Figure 5.3. The interactions of N64 (O) with galactose (HO-C2) and

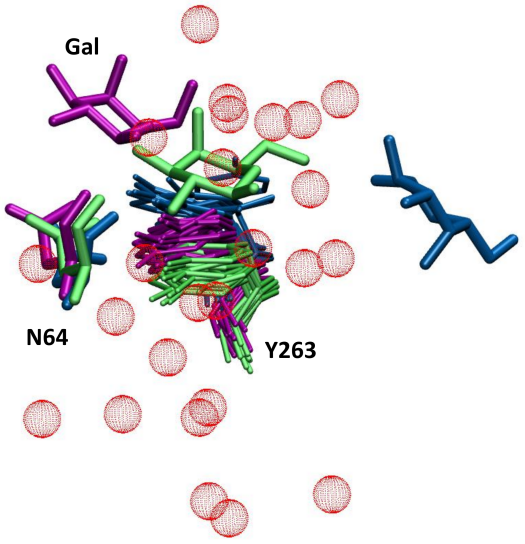

FIGURE 5.2: Flexibility of Y263. Min 1 (purple), Min 2 (green) and Min 3 (blue) of the Gal exit path are shown. Residues Y263, N64 and Gal are reported in licorice. In order to capture the flexibility of the Y263, several conformations of the same minimum are represented. Water molecules, freely coming from the hydrophilic cavity, are shown in red dotted spheres.

```
                                                    43                  66
vSGLT     1 MSNIEH----GLSFIDIMVFAIYVAIIIGVGLWVSRDKKGTQKSTEDYFLAGKSLPWWAVGASLIAANISAEQFIGMSGSGYSIGLAIASYEWMSAITL----- 95
hSGLT1   17 VETHEL----IRNAADISIIVIYFVVVMAVGLWAMFSTN--RGTVGGFFLAGRSMVWWPIGASLFASNIGSGHFVGLAGTGAASGIAIGGFEWNALVLV----- 109
NIS       1 MEAVETGERPTFGAWDYGVFALMLLVSTGIGLWVGLARGG-QRSAEDFFTGGRRLAALPVGLSLSASFMSAVQVLGVPSEAYRYGLKFLWMCLGQLLNS----- 98
PutP      1 ---------MAISTPMLVTFCVYIFGMILIGFIAW----RSTKNFDDYILGGRSLGPFVTALSAGASDMSGWLLMGLPGAVFLSGISESWIAIGLTLGAWINWK 91
                                                    *                   *
                                                                                                   181
VSGLT    96 IIVGKYFLPIFIEKGIYTIPEFVEKRFNKKL-K-TILAVFWISLYIFVNLTSVLYLGGLALETILGIPLMYSILGLALFALVYSIYGGLSAVVWTDVIQVFFLV 197
hSGLT1  110 VVLGWLFVPIYIKAGVVTMPEYLRKRFGGQRIQ-VYLSLLSLLLYIFTKISADIFSGAIFINLALGLNLYLAIFLLLAITALYTITGGLAAVIYTDTLQTVIML 202
NIS      99 VLTALLFMPVFYRLGLTSTYEYLEMRFSRAVRL-CGT--LQYIVATMLYTGIVIYAPALILNQVTGLDIWASLLSTGIICTFYTAVGGMKAVVWTDVFQVVVML 199
PutP     92 LVAGRLRVHTEYNNNALTLPDYFTGRFEDKSRILRIISALVILLFFTIYCASGIVAGARLFESTFGMSYETALWAGAAATILYTFIGGFLAVSWTDTVQASLMI 195
                                                                                                   *
                                                                                  267
vSGLT   198 LGGFMTTYMAVSFIGGTDGWFAGVSKMVDA----APGHFEMILD---------QSNPQYMNLPGIAVLIGGLWVANLYYWGFN----QYIIQRTLAAKSVSEAQ 284
hSGLT1  203 VGSLILTGFAFHEVGGYDAFMEKYMKAIPTIVSDGNTTFQEKCYTPRADSFHIFRDPLTGDLPW-PGFIFGMSILTLWYWCTD----QVIVQRCLSAKNMSHVK 311
NIS     200 SGFWVVLARGVMLVGGPRQVLTLAQN-----------------HSRINLMDFNPDPRSRYTFW--TFVVGGTLVWLSMYGVN----QAQVQRYVACRTEKQAK 279
PutP    196 FALILTPVIVIISVGGFGDSLEVIKQKSI-------------------ENVDMLK----G-LNF----V---AIISLMGWGLGYFGQPHILARFMAADSHHSIV 268
                                                                                        *
                                                                                                   368
vSGLT   285 KGIVFAAFLKLIVPFLVVL----PGIAAYVITSDPQLMASLGDIAATNLPSAANADKAYP-WLTQFLP--VGVKGVVFAALAAAIVSSLASMLNSTATIFTMDI 381
hSGLT1  312 GGCILCGYLKLMPMFIMVM----PGMISRILYTEKIACVVPSECEKYCGTKVGCTNIAYPTLVVELMP--NGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDI 409
NIS     280 LALLI----NQVGLFLIVSSAACCGIVMFVFY---------TDCDPLLLGRISAPDQYMPLLVLDIFEDLPGVPGLFLACAYSGTLSTASTSINAMAAVTVEDL 370
PutP    269 HA----RRISMTWMILCLAGAVAVGFFGIAYF---------NDHPALAGAVNQNAERVFIELAQILFN--PWIAGILLSAILAAVMSTLSCQLLVCSSAITEDL 357
                                                                                        *
                          396                                 427
vSGLT   382 YKEYISPDSGDHKLVNVGRTAAVVALIIACLIA--PMLGGIGQAFQYIQEYTGLYSPGILAVFLLGLFWKKTTSKGAIIGVVASIPFA-LFLKF---------- 472
hSGLT1  410 YAKVRKRASE-KELMIAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFWKRVNEPGAFWGLILGLLIG-ISRMITEFAYGTGSC 511
NIS     371 IKP-RLRSLAPRKLVIISKGLS-LIYGSACLTVAALSSLLGGGVLQGSFTVMGVISGPLLGAFILGMFLPACNTPGVLAGLGAGLALS-LWVALGATLYPPSEQ 471
PutP    358 YKAFLRKHASQKELVWVGRVMVLVVALVAIALAANPENRVLGL-VSYAWAG---FGAAFGPVVLFSVMWSRMTRNGALAGMIIGALTVIVWKQFGWLGL----- 452
                          *                                   *

vSGLT   473 -M---PLSMPFMDQMLYTLLFTMVVIAFTSLSTSINDDDP-KGISVTSSMFVTDRSFNI--------------------AAYGIMIVL--AVLYTLFW* 543
hSGLT1  512 ME---PSNCPTIICGVHYLYFAIILFAISFITIV----------------------VI--------------------SLLTKPIPD--VHLYRLCW* 561
NIS     472 TMRVLPSSA--ARCVALSVNA-SGLLDPALLPANDSSRAPSSGMDAS-RPALADSFYAISYLYYGALGTLTTVLCGALISCLTGPTKRSTLAPGLLWW* 565
PutP    453 -YEIIPGF----------IFGSIGIVVFSLLG-----KAPSAAMQK--RFAEADAHYHS----------------------APPSR-------L--* 499
```

FIGURE 5.3: Sequence alignment among SSS family using ClustalW. The residues conserved across the family and depicted in our work as relevant amino acids along the exit pathways are in bold character. The color of the boxes highlight the identical residues (yellow) or the similar ones, concerning the polar/apolar character of the aminoacid (magenta). The star and the numbers are put in order to simplify their identification.

N64 (HNδ) with Y263 (OH) are still present, even if characterized by large fluctuations. A qualitative analysis of the interaction energies between the substrate and selected residues shows that in the first two minima the van der Waals interactions regard the substrate and Y263, while the electrostatic interactions involve Gal-N64 (Table 5.3)
Afterwards, the substrate, hydrated by 4-5 water molecules, enters into a narrow cavity created by the residues N267, Q268, W134, T431, V434, transiently interacting with N142 and Y262 (Min 3_G). The H-bonds of E68 (Oδ1) with S66 (HN) and E68 (Oδ2) with S66 (HOγ) contribute to stabilize Gal in this minimum (Table 5.2). Y263 and N64 become very flexible as they can not form reciprocal H-bonds.
After overcoming a transition state ($TS1_G$), where the substrate is partially hydrated, Gal finds another minimum (Min 4_G). Here, it is almost fully hydrated, and surrounded by S368, V185, residues conserved in the SSS family and among LeuT-fold transporters (see Discussion), and the TM2-I, TM9, TM6, above loop TM5-6 and it is right below the sodium binding site, inside the hydrophilic cavity of the transporter. Gal (C1-OH and C2-OH) makes H-bonds with D189 (Oδ) and Gal (C6-OH) with A184 (carbonyl oxygen). Residue D189 is highly conserved throughout the SSS family (Figure 5.3) and it has been experimentally seen to play an important role for a correct Na^+-Gal cotransport and cation selectivity [115, 116]. We see here that it is also involved in the exit path of galactose, contributing to the stabilization of this minimum. Here, the aromatic ring of Y263 maintains an orientation similar to the crystal structure, while the side chain of N64 assumes a new conformation, pointing towards the carboxylic group of E68. The H-bond between E68-S66 is present also when Gal is in this minimum (Table 5.2).
In order to leave this site, moving deeper in the hydrophilic cavity, Gal has to overcome a transition state, $TS2_G$ ($\Delta F^{\#}$ ($TS2_G$) 11.9 $\pm$ 0.4 kcal/mol with respect to the minimum, which corresponds to the largest free energy barrier of the exit pathway. The breaking of the H-bonds between Gal and D189 contributes to the barrier, as suggested by the interaction energies among Gal and D189 along the path (Table 5.2 and 5.3).
At $TS2_G$, the substrate is at the protein surface and, although being hydrated, it is still interacting with a few surface residues forming a H-bond (Gal (C3-OH) with N371 (Oδ) and hydrophobic interactions with other residues (G181, L182 (on loop TM5-6), V396 (TM10), N371 and T375 (TM9) (Table 5.2). These latter residues are all conserved across the SSS family (Figure 5.3).
Remarkably, the free energy barriers associated to the exit path of Gal are significantly higher than those calculated by Watanabe et al. [8]. This is most probably due to the fact that our simulations start from a stable ion-retaining state of the transporter, and since a subtle cooperativity between Na^+ and Gal is observed at the very beginning of the path, simulations starting from the crystal structure, as those of Watanabe et al. [8], which corresponds to an ion-releasing state, may lead to simulate a less realistic dissociation process.

State	Coul Y263	LJ Y263	Coul N64	LJ N64	Coul D189	LJ D189
1_G	0.10 (±0.45)	-4.04 (±0.65)	-2.90 (±1.39)	-1.40 (±0.70)	0	0
2_G	-0.40 (±0.52)	-4.38 (±0.66)	-3.36 (±1.75)	-1.68 (±0.80)	0	0
3_G	0.05 (±0.15)	-0.61 (±0.41)	0	0	0	0
$TS1_G$	0.01 (±0.02)	-0.10 (±0.03)	-0.01 (±0.04)	-0.15 (±0.06)	-0.02 (±0.03)	-0.02 (±0.02)
4_G	0	0	-0.03 (±0.03)	-0.20 (±0.05)	-22.82 (±6.38)	2.39 (±2.22)
$TS2_G$	0	0	0	0	0	0

TABLE 5.3: Average electrostatic and van der Waals interaction energies (kcal/mol) between galactose and residues Y263, N64 and D189, at selected minima and transition states visited along the exit path of Gal.

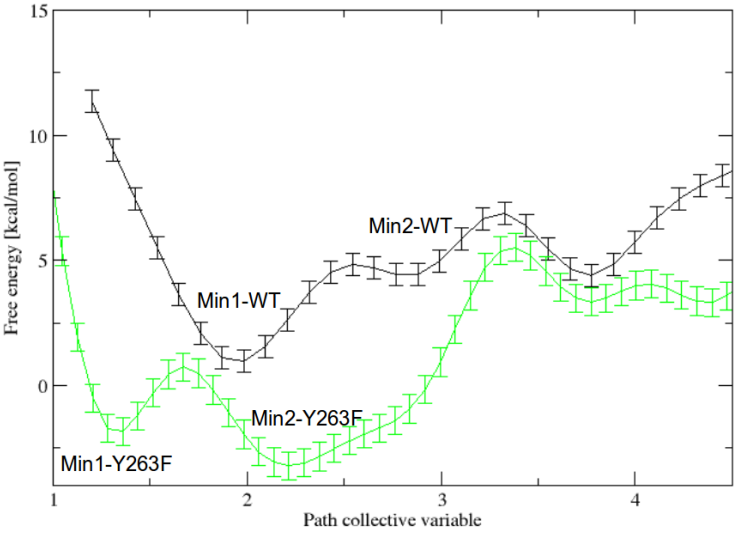

FIGURE 5.4: Free energy profile of wild type and the mutant. The free energy profile of wild type (WT) system along the path collective variable of Gal (black line) and that of the mutant Y263F (green line).

Since Y263F mutation has been observed to impair the transport mechanism, to further check the controversial role of Y263 in the dissociation of the substrate, we performed a BE-MTD simulation of the mutant using the same setup of the WT simulation. Looking at the Figure 5.4, we can clearly see a different profile, where the second minimum becomes the global one, more stable and broader than the first minimum. In short, Min1-Y263F corresponds to Min1-WT, while Min2-Y263F is broad and thus characterized by different configurations of Gal (containing among them the Min2-WT minimum). Their relative free energy has changed, meaning that Y263 decides in this transporter the relative stability of the minima characterizing the releasing path of the substrate. Thus, this mutation has a role in reshaping the free energy surface of Gal exit path. The stabilization of the other minimum does not seem to influence the barrier height significantly, but it may hamper Gal from assuming a position necessary to undergo the inward-outward conformational change affecting in turn the overall transport cycle, in

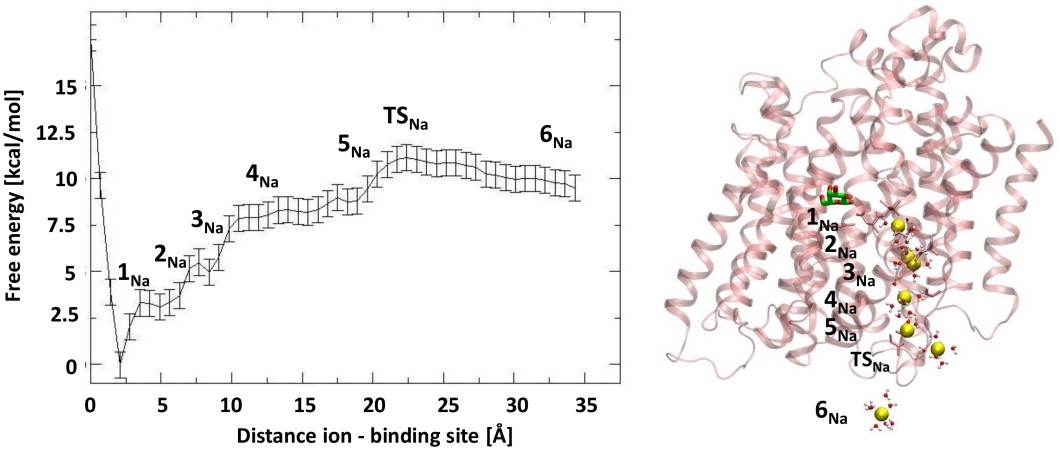

FIGURE 5.5: Exit pathway of sodium ion. Free energy profile (kcal/mol) projected along the distance ion - center of mass of its binding site (Å) is displayed. On the right the position of the different states along the exit path of Na^+ are shown on the top of the protein.

line with experimental findings [8]. We remark that the change in the free energy profile of the releasing of Gal caused by this mutation does not imply a gating role for this residue.

5.4.2 Sodium Ion Exit Pathway

The dissociation path of Na^+ is characterized by an overall free energy barrier of 11.1 (± 0.7) kcal/mol and by the presence of a few faint metastable states (see Figure 5.5). The most stable ion binding site is LC1 in which the ion is coordinated by three water molecules and three residues (A62, I65 and S365). As soon as the ion starts moving towards the cytoplasm, it loses its coordination with I65; then, it approaches the mouth of the binding site keeping the interaction with A62 and coordinating S364 and D189. This latter is often found to bind Na^+ along the exit. This state is referenced to as PC in Chapter 4.

In this configuration the side chain of E68 rotates from a configuration in which it heads towards the Gal binding site to a new conformation in which it forms hydrogen bonds with S66. In State 2_{Na} the ion, at almost 5 Å from its binding site, has overcome D189, moving deeper into the hydrophilic cavity, and it is fully hydrated. Then, it continues descending into the cavity coordinating G181 (loop 5-6) and S368 (on TM9) and four water molecules (State 3_{Na}). Remarkably, D189 interacts with the ion through water bridges, accompanying it from LC1 to State 3_{Na}, confirming its important role in the exit pathway of Na^+ [115, 116, 128]. After interacting with L182 (loop 5-6) (State 4_{Na}), it reaches State 5_{Na}, which is at approximately 1.8 Å from the binding site. Here, Na^+ is coordinated by a few residues of loop 1-2 and by 3-4 water molecules. It finally overcomes TS_{Na}, where it is still transiently coordinated by D43, R400, (both conserved residues in

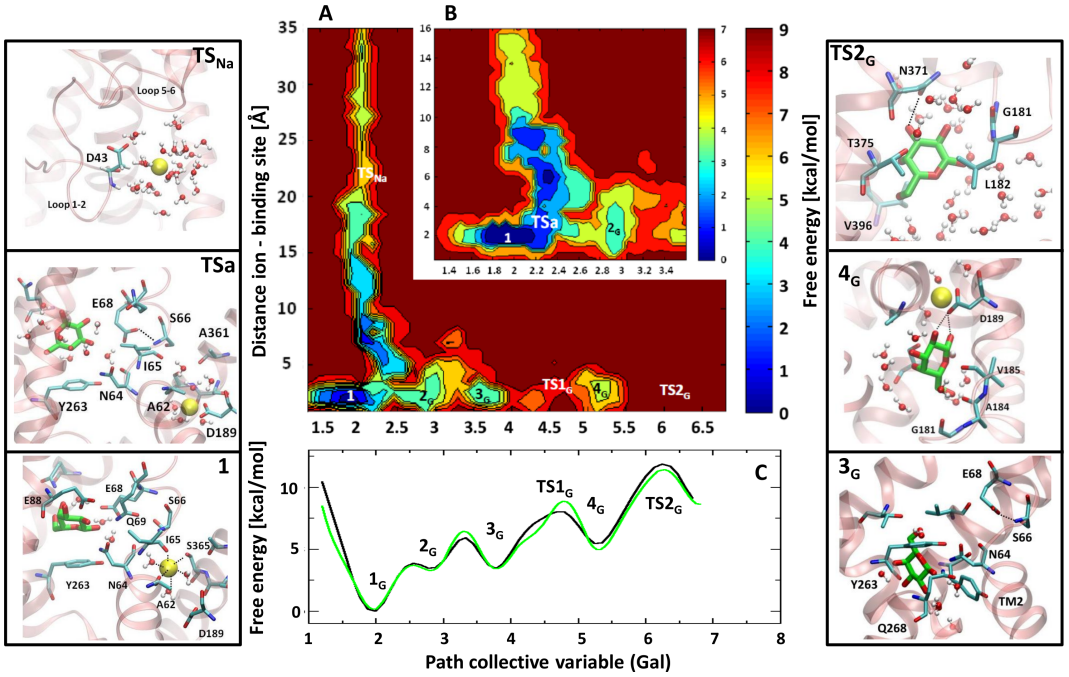

FIGURE 5.6: The cooperativity in the release mechanism. (A) Free energy surface (kcal/mol) depicted with respect to the distance between Na^+ and its binding site (Å), and the galactose exit path. Note that the deepest minima for both exit pathways are the same, i.e. $1_{Na} = 1_G = 1$. (B) Close view of the initial part of the FES. (C) The projection of the free energy (kcal/mol) along the path collective variable representing Gal exit. Black and green lines correspond to the simulations carried out in presence or in absence of Na^+. On the two sides of the image, the most relevant structures corresponding to minima and transitions states (TS) are depicted.

the SSS family, Figure 5.3), and even by a POPC molecule. Thus, the total free energy barrier is due to cumulative energy cost of small structural changes accompanying the Na^+ release without the formation of any stable intermediate. After the TS_{Na} Na^+ is quite delocalized, in a vestibule mainly formed by loops TM9-10, TM1-2 and loop TM5-6. It is important to note that the two ligands, starting from different binding sites, exit the protein through the same hydrophilic cavity (characterizing the inward-facing conformation) communicating to the cytoplasm.

5.4.3 Is the Release Mechanism of the Galactose and the Sodium Ion Cooperative?

In order to investigate possible cooperative effects in the release mechanism of Na^+ and Gal, in Figure 5.6 we report a projection of the FES as a function of two CVs, the distance between the ion and its site and the path variable of Gal. It is possible to note that the deepest minimum for both exit pathways is the same, i.e. $1_{Na} = 1_G$. Thus, it is labeled as Min 1. A zoom of the free energy landscape in the region close to the binding sites is also reported. The shape of the free energy landscape clearly suggests an

interplay between the two ligands. Indeed, upon displacement of Gal from its binding site to move towards Min 2_G, the Na^+ loses its coordination with I65 and moves towards a site with a reduced number of coordinating residues, the PC site (state TS_a). The residue linking the two binding sites is the E68. Indeed, upon Gal displacement from Min 1, E68, initially heading towards the Gal binding site, rotates towards the Na^+ binding site, establishing one H-bond with S66 (HN) (Table 5.2), a conserved residue across the SSS family (Figure 5.3). This functional rotation of E68 is also observed between the holo (PDB 3DH4) and the apo (PDB 2XQ2) forms of the vSGLT crystals. These results are in line with the previous hypothesis suggesting that the departure of Na^+ from its stable putative ion-retaining site, LC1, towards the PC site triggers the conformational changes at the basis of Gal displacement from the binding site, heading to the second metastable minimum of the path (Min 2_G). However, the free energy barrier associated to this initial displacement is very small and the highest barriers lay further along the dissociation path of Na^+ and Gal. This fact, along with the overall L-shape of the free energy for large values of the collective variables (see Figure 5.6), indicates that the rate limiting steps of the release of the ion and the substrate are independent. Indeed, the values of the two CVs (the path collective variable of Gal and the distance Na^+ - binding site) at the highest transition state ($TS2_G$) of Gal exit are 6.2 and 3 Å. While, those at the TS_{Na} of Na^+ exit are 2.2 and 21 Å. Namely, at the transition state of Gal, Na^+ is close to its binding site, and vice versa. In order to quantitatively verify this point, we computed the free energy of Gal exit in absence of Na^+. The free energy profile of Gal dissociation in the absence of Na^+ is practically identical to the profile in the presence of Na^+, confirming unambiguously this important result (Figure 5.6, panel C).

5.5 Discussion

In this Chapter we used BE-MTD to study the binding/dissociation mechanism of the two ligands of vSGLT symporter. We observed that the minimum free energy exit pathway of the galactose does not involve any rotameric transition of the side chain of Y263. Indeed, as already observed [11], Gal circumnavigates the so-called inner gate Y263 and proceeds along the hydrophilic cavity. However, our simulation of the mutant points to a possible functional role of Y263 in determining the relative stability of the minima observed along the Gal exit path. The global free energy minimum for the mutant and for the WT turns out to be different (Figure 5.4).

The main barriers characterizing the releasing mechanism are of the order of 11-12 kcal/-mol for both the ion and the substrate. These barriers are strikingly higher than the ones reported by Watanabe et al. [8]. This is probably due to the fact that the free energy space explored in our case includes an occluded state of the transporter, with

both ligands stable in their binding sites, while the work of Watanabe et al. [8], starting from a different structure, may simulate a different process, the departure of the ion from an ion-releasing state.

We underline that these barriers are in line with the kinetic experimental data. Indeed, the galactose turnover in vSGLT was estimated to be around 0.4 s^{-1} by Turk et al. [129], or on the order of tens of ms in other works [13, 130–132]. Lapointe and coworkers measured the turnover rate of hSGLT1, finding values of 8 s^{-1} or, being near V_{max} conditions, 13 s^{-1}. Moreover, a transition rate of 50-60 s^{-1} was found for hSGLT1 [120, 131]. These values correspond to a $\Delta G^{\#}$ in the range of 11-15 kcal/mol, consistently with our results. Our simulations suggest that the barrier associated with the dissociation of the ligands from the inward-facing conformation may be of the same order of magnitude of those associated with the inward-outward switch.

Importantly, our simulations provide for the first time direct insights on the possible cooperativity between Na^+ and Gal for their release mechanism towards the cytoplasm. A small interdependence is observed only at the very beginning of the ligands release process, with residue E68 playing a central role in the communication between the two binding sites. Remarkably, this intercommunication occurs far from the point of the free energy profile associated to the highest free energy barriers. Our simulations, carried out in the absence of Na^+, reveal that the whole free energy profile of Gal exit is essentially unaffected, Figure 5.6 (green line, panel C). The lack of a marked cooperativity in the release mechanism of Gal and Na^+ from the binding site is at first sight surprising. However, it is likely that the cooperativity might be associated to the first steps of the transport cycle, when the symporter, in the outward-facing conformation, binds the sodium ion and then the substrate, and their binding triggers the outward-to-inward facing conformational change, as observed in the LeuT-fold superfamily [27, 37, 133–135].

Due to this non-cooperativity in the Gal and Na^+ release mechanism from the inward-facing conformation of vSGLT and to the almost identical rate limiting free energy barriers, we propose to extend the six-state kinetic model introduced by Wright and coworkers [12, 13] by adding one more state, Figure 5.7 (blue region). Indeed, we suggest that, from the ligand bound inward-facing conformation, the transporter can follow independently two paths for Gal and Na^+ release. The very similar free energy barriers observed for the Na^+ and Gal release from the inward-facing conformation may be in part responsible for the difficulties encountered experimentally in providing a detailed and clear picture for this part of the transport path of hSGLT [12, 13, 120].

We remark that the binding site of Na^+ observed here is the same described in Chapter 4, by independent metadynamic simulations, demonstrating the reliability of our results. We also observe that the crystal structure of the inward-facing conformation in the apo form of (PDB code 2XQ2) [8] differs in the presence of a kink of a few degrees in TM2-I

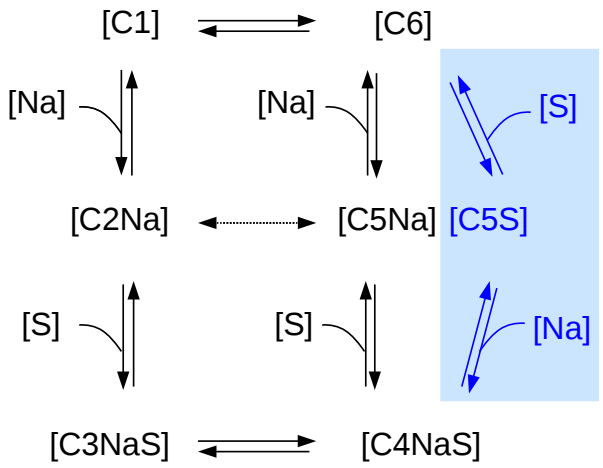

FIGURE 5.7: Six-state kinetic model for Na^+-galactose cotransporter updated to a branched six-state model. In the absence of ligands, the transporter can be in two states: outward or inward-facing conformation (C1 and C6, respectively). After the binding of Na^+ to the outward conformation (C2Na), the substrate enters the protein and finds its site (C3NaS). This step is then followed by the crucial event that sees the transporter switching to the inward-facing conformation (C4NaS). In case the transport follows the dashed line, namely passing from C2Na to C5Na, a uniport of Na^+ ion happens. From the ligands-loaded inward conformation (C4NaS) the protein can lose at first the substrate (C5Na), as suggested in Ref. [12, 13], or, based on our results (blue region), in an independent way characterized by similar barriers, the ion (C5S).

(intracellular half) and the side chain of E68 heading towards S66. Remarkably, these structural features are observed during the Gal release pathway, suggesting that starting from a Na^+ occluded and holo state of vSGLT (PDB code 3DH4) [4], we are able to visit these structural features of the apo state captured crystallographically (PDB code 2XQ2) [8].

An important process associated to this transporter is the permeation of water (as mentioned in Chapter 2), whose precise mechanism is still under debate [29, 30]. The two mechanisms considered more viable are the active cotransport [29, 31], where water flux is coupled to ion/solute flux, or the passive permeation [32], where the accumulation of the solutes near the intracellular side of the membrane during solute transport induces a flux of water as a response to the local osmotic gradient. A detailed analysis of this controversial issue is beyond the scope of our study. However, in line with the passive mechanism [39], we observe that: (i) the water molecules permeate easily through the whole protein during the releasing process; (ii) the water is free to enter from the cytoplasm into the hydrophilic cavity and then into the binding sites. Consistently with Ref. [36, 37], in our simulations water molecules help the breaking of the H-bonds that

keep the ligands bound to the protein, playing an important role in the whole releasing pathway. Their presence at the edge of the Gal binding site confers, indeed, flexibility to Y263, facilitating the initial displacement of the substrate (Figure 5.2).
Finally, a bioinformatic analysis reveals that several of the residues cited in describing the exit pathways of the ligands are conserved among the SSS family or, despite the low sequence similarity and the fact that different substrates are transported, even among different LeuT-fold transporters (LeuT, Mhp1 and BetP). This underlies their importance in the ligand release process (see Figure 5.3 and 5.8 for the alignments). For example, S66, with whom E68 makes H-bonds as soon as the system leaves the deepest minimum, is conserved across the whole SSS family. Mutation T460C in hSGLT1, corresponding to T431 of vSGLT, which in our simulations stabilizes Gal in Min 2_G through a H-bond, decreases glucose affinity in the human homolog [136]. A mutation of K157A of hSGLT1, corresponding to N142 of vSGLT, has been experimentally seen to impair the glucose transport in oocyte [137]. Moreover, Q268R mutation in the human homolog causes glucose/galactose malabsorption [138]. This residue surrounds Gal in Min 3_G and the substitution by a positively charged residue can further stabilize it, affecting the overall free energy profile and reducing in this manner the sugar transport rate. As we can see from Figure 5.3 and 5.8, E68 is not conserved by performing a standard sequence or structure alignment of vSGLT with the other symporters. So, the coupling mechanism between the ion and the substrate in the dissociation path may be different in other transporters. However, superimposing the structure of vSGLT with those of Mhp1 and LeuT, we found that, close to the location of the corresponding residue, other polar residues (Q42 in Mhp1, N27 in LeuT) are present and could play the same role of E68 in vSGLT.

Mutations regarding Y263, the hypothesized external gate, and N64 highlight their important role in the transport mechanism [8, 120]. Consistently with this picture, the simulation of the Y263F mutant reveals that this residue has an important role in reshaping the free energy profile of Gal exit path. This mutation may not influence the kinetic of Gal release, but it probably hampers Gal from assuming a position necessary to undergo the inward-outward conformational change. This may affect the overall transport cycle, in line with experimental findings. For both Na^+ and galactose exit pathways residue D189, a highly conserved residue in SSS family, plays an important role: stabilizing Gal in Min 4_G and participating in the highest barrier of its exit path, and accompanying the ion along its releasing mechanism [115, 116]. In short, a conservation of key residues involved in the Na^+ and Gal dissociation path has been observed certainly across the SSS family and, at times, even across the whole LeuT super family, both in terms of identity or similarity (for a graphical representation of the conserved residues along the exit pathways see Figure 5.9). This suggests that the mechanism of release of Na^+ and Gal elucidated here for vSGLT is likely to be common in the SSS

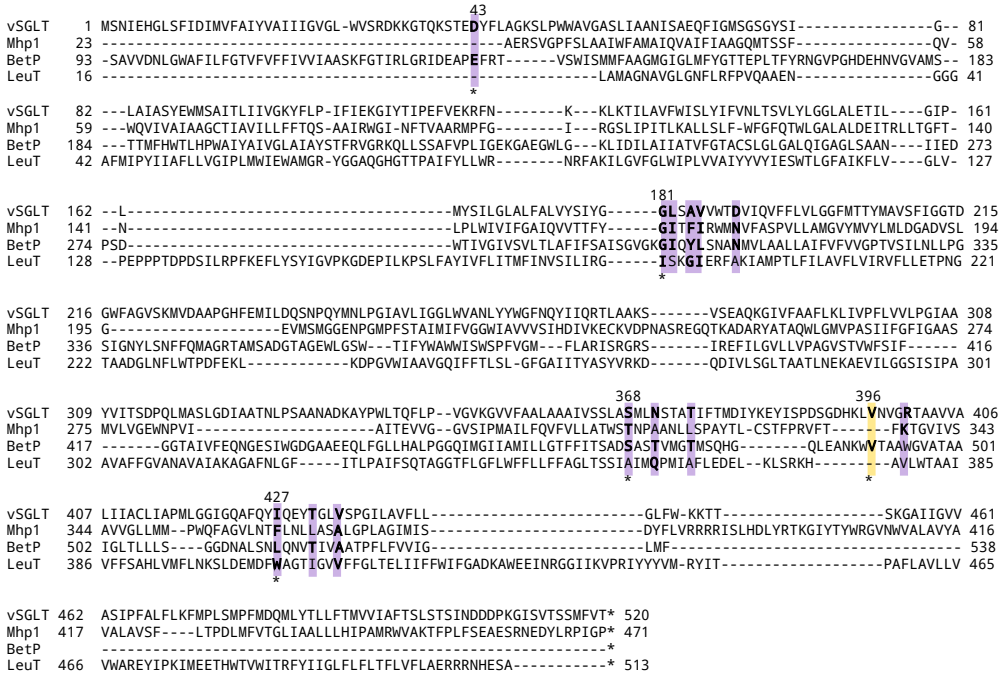

FIGURE 5.8: Structure alignment among LeuT-fold transporters using HHPred server. The residues conserved across the family and depicted in our work as relevant amino acids along the exit pathways are in bold character. The color of the boxes highlight the identical residues (yellow) or the similar ones, concerning the polar/apolar character of the aminoacid (magenta). The star and the numbers are put in order to simplify their identification.

family and to share remarkable mechanistic similarities with other LeuT transporters despite harboring different substrates and working with different stoichiometries ions/-substrates.

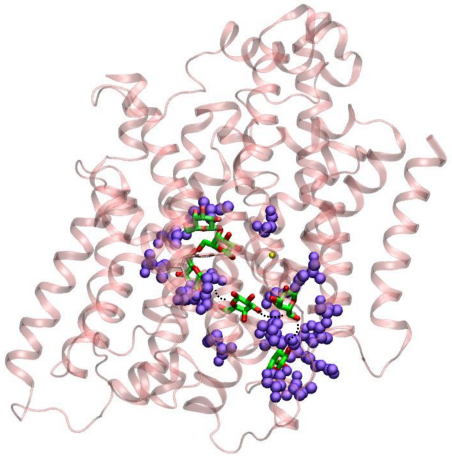

FIGURE 5.9: Conserved residues (violet spheres) along Gal and Na^+ dissociation path among the SSS family or LeuT-fold members.

Chapter 6

A Computational Study of CNG Channels

6.1 Overview

Permeation and gating in K^+ and Na^+ channels are considered as independent processes, while in cyclic nucleotide gated (CNG) channels a significant coupling occurs. In this part of the thesis we discuss the work done in collaboration with electrophysiologists and crystallographers to unravel the molecular basis of this coupling using a bacterial CNG mimics as model system. When the CNG mimics is crystallized in the presence of alkali monovalent cations, the side chain of E66 in the pore shows multiple conformations. MD simulations indicate that E66 and the prolines in the outer vestibule have a large mobility, modulated by the ionic species and the voltage. The pore of CNG channels is highly flexible with an almost "liquid-like" behavior. This allows rationalizing the anomalous coupling between gating and permeation observed in this channel.

6.2 Introduction

As discussed in Section 2.3, in K^+ selective channels, the opening and closing of the ion channel pore (gating) and the translocation of ions through the pore (permeation) are normally considered independent processes [45]. Gating is indeed controlled by the 'bundle crossing' at the intracellular side [48, 63], while permeation reflects ion-ion and ion-pore interactions within the selectivity filter. These experimental observations have led to the paradigm in which the 3D structure of the selectivity filter is rigid and that the mechanisms of ionic permeation can be deduced in essence from its crystal structure, as successfully done for K^+ channels [48, 65]. Cyclic nucleotide-gated (CNG) channels underlie sensory transduction in the retina and olfactory epithelium and share a high degree of homology with K^+ channels in the pore region. Despite of this strong homology, in contrast to K^+ channels, CNG channels gate at the selectivity filter, implying that the same protein region controls ion permeation as well as gating. Moreover, in CNG the ionic species present inside the channel influences channel gating [139–141], but how this coupling occurs is not well understood. In the presence of large cations, such as Rb^+ and Cs^+, this coupling is also controlled by membrane voltage [15] and current-voltage relationships activated by 1 mM cGMP depend on the radius of the permeating ion. Since structural information on CNG channels is still patchy and limited to a low-resolution electron-microscopy map [142], a chimeric channel where the CNG selectivity filter sequence is embedded into a NaK channel, was engineered and crystallized, creating a CNG mimic, the NaK2CNG [14] (Section 2.3). This CNG mimic provides a suitable model to understand the properties of the pore since it shares with the CNG channels the low ionic selectivity and the coupling between gating and permeation.

To investigate the structural basis of the linkage between gating and permeation, we combined our efforts with those of two experimental groups expert in electrophysiology and X-ray crystallography.
A) The electrophysiologists showed that the mimic recapitulates the voltage dependent coupling observed in CNGA1 channel. Indeed, with single channel recordings they observed that in the presence of Na^+ similar openings were visible at 100 and -100 mV, while in the presence of Cs^+ clear voltage-dependent asymmetries are observed: the channel displays flickering openings at -100 mV and long-lasting well-defined channel openings at +100 mV (Figure 6.1).

B) A structural analysis of the mimic in the presence of monovalent alkali cations (Li^+, Na^+, Rb^+, Cs^+, $NH4^+$, MA^+, DMA^+) complexes resulted in small but significant structural changes in the polypeptide backbone lining the pore lumen. An expansion of the latter has been detected with increasing cation size, and especially in the presence of the larger Cs^+ and DMA^+ (Figure 6.2). Smaller structural variations were observed for

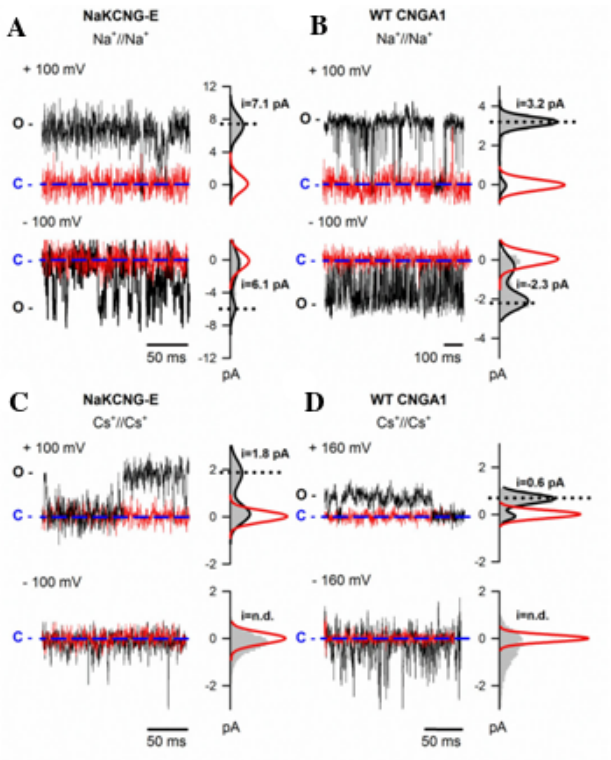

FIGURE 6.1: Electrophysiology of Na^+ and Cs^+ in NaK2CNG channel. A) Current recordings from a single CNG mimicking NaK chimera channel at +100 mV (upper panel) and -100 mV (lower panel) in the absence (black traces) and in the presence (red traces) of the blocker TPeA in symmetrical 110 mM Na^+ solution. B) current recordings from a single WT CNGA1 channel at +100 mV (upper panel) and -100 mV (lower panel) in the presence (black traces) and in the absence (red traces) of 1 mM cGMP in symmetrical 110 mM Na^+ solution. C) As in A but in symmetrical 110 mM Cs^+ solution. D) As in D but in symmetrical 110 mM Cs^+ solution and at + and -160 mV. Amplitude histograms for recordings in the absence of 30 μM TPeA and in the presence of 1 mM cGMP, for the mimic and for the WT channel respectively, are shown at the right (gray area). Black lines represent a two-terms Gaussian fit to the histograms. Red lines represent a one-term Gaussian fit to the histograms for recordings in the presence of TPeA and in the absence of 1 mM cGMP, for the mimic and for the WT channel respectively. Blue dashed lines indicate the 0 current level. C and O refer to the closed and open states, respectively.

P68 at the external vestibule. When Na^+ and Cs^+ structures were compared, a small displacement of the backbone polypeptide chain in this region was observed (Figure 6.2), resulting in a 0.7 Å widening of the funnel. Significant structural differences are also observed in the region surrounding the selectivity filter. Indeed, the structures showed a variety of conformations for the E66 side chain (Figure 6.2).

In order to rationalize these observations from an atomic perspective, we performed molecular dynamics (see Section 3.1) and bias-exchange metadynamics (see Section 3.2.2) simulations.

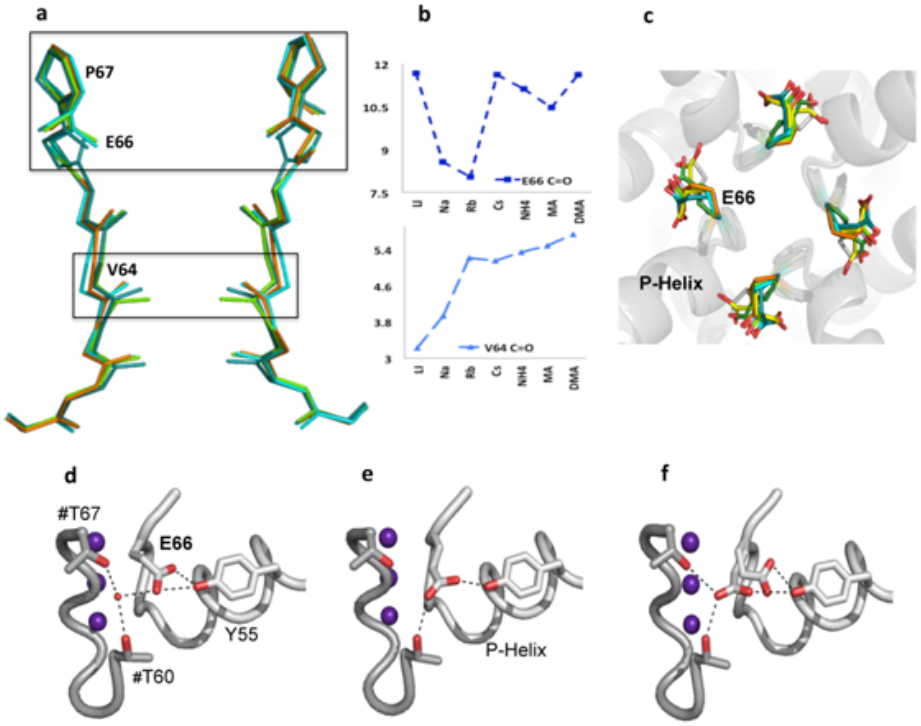

FIGURE 6.2: Crystallography of the NaK2CNG-E channel pore. (a) The tetrameric structures of Li^+, Na^+, Cs^+ and DMA^+ complexes have been overlapped and are viewed along the fourfold axis. The main chains of residues 62 to 68 are shown for two opposite subunits in the tetramer and are colour-coded as following: Li^+ (light green), Na^+ (teal), Cs^+ (cyan) and DMA^+ (orange). Changes in the pore size and the orientation of E66 carbonyls can be observed. (b) Top: the distances for the carbonyl oxygens of two V64 residues on opposite subunits are plotted against the various ions, in order of ionic radius. Bottom: the distances between opposite E66 carbonyl oxygens versus the different ions tested are plotted versus ion type. (d) An overview of the changes in the relative orientation of the E66 side chains as seen down the tetrameric channel axis, viewed from the extracellular side. (e, f, g) A sample of the different relative orientations and hydrogen bonding patterns of E66. Purple spheres represents sites 2, 3 and 4 (from top to bottom); the neighboring molecule is shown in darker green; side chains residues involved in the interactions are shown; residues coming from an adjacent subunit are labelled with a hash (#).

6.3 Methods

6.3.1 System Setup

The model of the chimera NaK2CNG-E was built using the chain B (residues from 19 to 113) of the 2 Å resolution crystal structure soaked with Na^+ ions (PDB accession code 3K0G [14]). Applying transformation matrices we generated the tetramer using VMD. The Hys residues, all exposed to the intracellular solvent, were kept protonated in ϵ position. The Asp and Glu residues were considered in the deprotonated form, while the Arg and Lys residues in the protonated form. The final tetrameric structure contained 95

residues (A2 to R96). The protein was embedded in a pure, pre-equilibrated 1-palmitoyl-2-oleilphosphatidylcholine (POPC) lipid model (kindly supplied by T. A. Martinek) [94, 95] using the g_membed [96] tool of Gromacs and then it was oriented following OPM [97] database model, similarly to the system of vSGLT (Chapter 4). Afterward the system was neutralized and solvated with TIP3P model [98] water molecules (76305 total atoms in a box size of 92.8 91.9 87.5 $Å^3$).

6.3.2 MD Simulations

Two systems were prepared with different permeant ions. Firstly, in the three Na^+ binding sites identified by the crystal structure were put two sodium ions alternated by a water molecule.
We followed a protocol similar to the one described in Section 4.3.2. The simulations were performed in periodic boundary conditions at 300 K using the Nose-Hoover thermostat [106] and Parrinello-Rahman barostat [79, 80] with a semiisotropic pressure coupling type and a time step of 2 fs. Position restraints of atoms were added with a force constant (K) equal to 1000 kJ mol^{-1} nm^{-2}. We used GROMACS [99] package with Amber03 [100] force field for protein and GAFF [101] for the membrane. The equilibration was performed in three stages: (1) the system was heated for 2.5 ns with protein backbone and ions fixed, while side chains were left free to move; (2) 5.2 ns were run using position restraints only for the selectivity filter and the ions. In the first stage we used the NPT ensemble, while in the second one a surface tension equal to 600.0 bar*nm [94] was added; (3) For the next 1 ns the membrane area was kept constant.
A configuration taken from this step was used as a starting point for a free molecular dynamics (MD) simulation of ∼100 ns. Moreover, from the same conformation the 2 Na^+ ions were substituted by two Cs^+, and the new system was then evolved for a ∼100 ns long simulation.

6.3.3 Bias-Exchange Metadynamics

In order to explore the free energy surface of the conformational states of the channel and better characterize its variability in the two systems, we performed two bias-exchange metadynamics (BE-MTD) simulations of 384 ns (48 ns * 8 walkers) (for the system with two Na^+ ions in the selectivity filter) and 320 ns (40 ns * 8 walkers) (for the system with two Cs^+ ions in the selectivity filter). The approach is described in Section 3.2.2. The collective variables (see Section 3.2.2.1) used here are:

1. the dipole of the selectivity filter, only heavy atoms (164 atoms, 6 residues for each monomer, two walkers), defined as:

$$D_t = \sum_i q_i(z_{it} - \bar{z}_t) \tag{6.1}$$

where for each atom i the charge q is moltiplied by the distance of its z coordinate at time t from the average of the z coordinates of all atoms at time t.

2. the distance between the Cα of E66 and the C of the carboxylic group (Cδ) of E66 in the opposite monomer (4 variables, see Figure 6.3, panel C);

3. the coordination number of the ions with the two oxygens of the carboxylic group of E66 (where $n = 6$ and $m = 12$ for both systems; $r_0 = 0.3$ nm for Na^+ and $r_0 = 0.43$ for Cs^+);

4. the coordination pairs of the native contacts between carbonylic oxygen (O) of E66 and the hydroxyl group of Y55 and carboxylic oxygens (Oδ) of E66 with hydroxyl group of T60 ($n = 6$, $m = 12$ and $r_0 = 0.25$).

In the simulation time the free energy converged for almost all CVs, both in the Na^+ and Cs^+ case. In particular, for the analysis of the Na^+ system we used the two walkers biased on the dipole CV and the four distances (for a total time of 108 ns (18 ns * 6 walkers)), while for the Cs^+ system we used one dipole, the four distances and also the coordination number of the ions (for a total time of 90 ns, namely 15 ns * 6 walkers).
In order to understand and quantify the differences of the two systems, and especially explain the voltage-dependence observed by electrophysiological experiments, we measured some observables and their linear dependence on the external field. The dependence of an observable O on the external voltage V has been calculated as:

$$O(V) = \frac{\sum_\alpha O_\alpha e^{-\frac{1}{kT}(F_\alpha - fVd_\alpha)}}{\sum_\alpha e^{-\frac{1}{kT}(F_\alpha - fVd_\alpha)}} \tag{6.2}$$

where F_α is the free energy of microstate α, d_α is its dipole, f the Faraday constant (96485.34 C/mol) and O_α is the value of the observable in the microstate α. We considered three observables:

1. the RMSD of the P68 of each monomer with respect to the crystal structure;

2. the distance of the center of mass of the four Cδ of E66 from the center of the pore (measured as the center of mass of the four G65) and

3. the position of the top ion, measured as the distance of the ion from the center of the mass of the heavy atom of the four G65, using only the z-direction.

With the four distances used as CVs in BE-MTD we built a new variable, called *Contacts*, describing approximately the location of the four E66. It is defined as the sum of 4 switching functions, each one going from 0 to 1. The higher the value of this variable, the closer to the crystallographic position of E66:

$$\text{Contacts} = \sum_{i=1}^{4} 1 - \frac{(1 - \frac{r_i^{\alpha\delta}}{0.8})^6}{(1 - \frac{r_i^{\alpha\delta}}{0.8})^{12}} \tag{6.3}$$

where $r_i^{\alpha\delta}$ is the distance between the Cα of E66 and the Cδ of E66 in the opposite monomer. The sum runs over the four pairs of Cα-Cδ.
The free energy as a function of the voltage is estimated as:

$$F(V) = -kTln\frac{\sum_\alpha e^{-\frac{1}{kT}(F_\alpha - fVd_\alpha)}}{N_\alpha} \tag{6.4}$$

where N_α is the number of configurations assigned to microstates α.

6.4 Results

Electrophysiological recordings strikingly reveal a significant voltage-dependence of the current and of the gating probability in the presence of Cs^+, but not in the presence of Na^+. As stated above, crystallography provides a hint that a key factor determining this difference might be the flexibility of the E66 chains. However, it does not provide a clear explanation of why gating and permeation are so tightly coupled only in the presence of Cs^+. In order to shed some light on these intriguing topics we first performed ~100 ns of MD at 300 K in the presence of either two Na^+ and or two Cs^+ ions within the selectivity filter. Visual inspection of these trajectories reveals a significantly higher flexibility for the system in the presence of the cesium ions. A quantitative measure of this difference was given by a recently developed clustering algorithm [143], capable of finding from a MD trajectory the metastable states of the system, namely the configurations which correspond to local free energy minima, separated by the other minima by barriers that are crossed rarely on the time scale of the simulation. Remarkably, we observed 38 metastable states for the Cs^+ case, and only 3 for the Na^+ system. The metastable states observed in the presence of Cs^+ differ in most of the cases for the position of the side chain of E66, that assumes a large array of conformations ranging from the crystallographic position to a re-orientation towards the external vestibule and finally towards the pore, where it often coordinates the ion. A similar mobility has been observed also in the corresponding residue, E177, in the Na^+ selective NavAb channel [144]. While these 100 ns of molecular dynamics are sufficient to reveal that the presence Na^+ or Cs^+ triggers a qualitatively different behavior, they do not provide

a quantitative picture of the free energy landscape, which for a simulation of this duration is not converged (for example, the population of the aforementioned metastable states is different when is computed on the first and the second half of the trajectory). In order to overcome this limitation, we used bias-exchange metadynamics (BE-MTD). Since the electrophysiological experiments revealed a strong voltage-dependence in the presence of Cs^+, the z-component of the dipole of the selectivity filter was selected as a CV. Since plain MD and structural data reveal an important mobility of the side chains of E66, we also included as CVs the four distances between the carboxylate and the carbonyl groups of E66 in opposite monomers. Finally, we included a CV measuring the coordination of the ions with the carboxylate oxygens and a CV measuring the presence of the crystallographic H-bonds between E66 and the two residues Y55 and T60 (see Section 6.3.3 for the exact definition of these CVs as explicit functions of the coordinates of the atoms). In Figure 6.3 we show the free energy as a function of the z-dipole of the selectivity filter and of a variable, called Contacts, counting the number of contacts between the carboxylate and carbonyl groups of E66 in opposite monomers. The higher the value of this variable, the closer is E66 to the crystallographic position. While in the presence of Na^+ (Figure 6.3, A) the landscape is characterized by a deep broad minimum, in the presence of Cs^+ (Figure 6.3, B) the landscape is much more flat, with several local minima characterized by different number of contacts and different dipoles within a range of $\sim$1 kcal/mol from the global free energy minimum. These local minima are separated by free energy barriers that are rather small, of 0.5 kcal/mol or less, but this is likely to be an effect of the two-dimensional projection, since plain MD reveals that these states are stable for 1 ns or more. This picture is in agreement with what observed by classical MD: the Cs^+ system is more flexible and characterized by several metastable states, differing significantly for the value of their dipole and for other structural features. This provides the first hint on the molecular mechanism at the basis of the differences observed by electrophysiology.

In order to investigate the difference in the voltage-dependence in the presence of Na^+ and of Cs^+, we then computed the two-dimensional projection of the free energy on the dipole and on several geometrical observables, describing the position of the ions in the selectivity filter, the position of the E66 residues, and the position of the P68 residue. Computing these projections allow predicting the dependence of any geometric observable on the external voltage V. In Figure 6.4 we show the dependence on V of the distance of the top ion (red lines), of the variable counting the number of contacts of E66 (blue lines) and the RMSD of P68 from the crystallographic positions (black lines). This latter variable has been chosen because P68 is the residue whose backbone changes conformation more significantly with V among those in the vicinity of the selectivity filter. The top panel shows the dependence of these variables for the channel in the

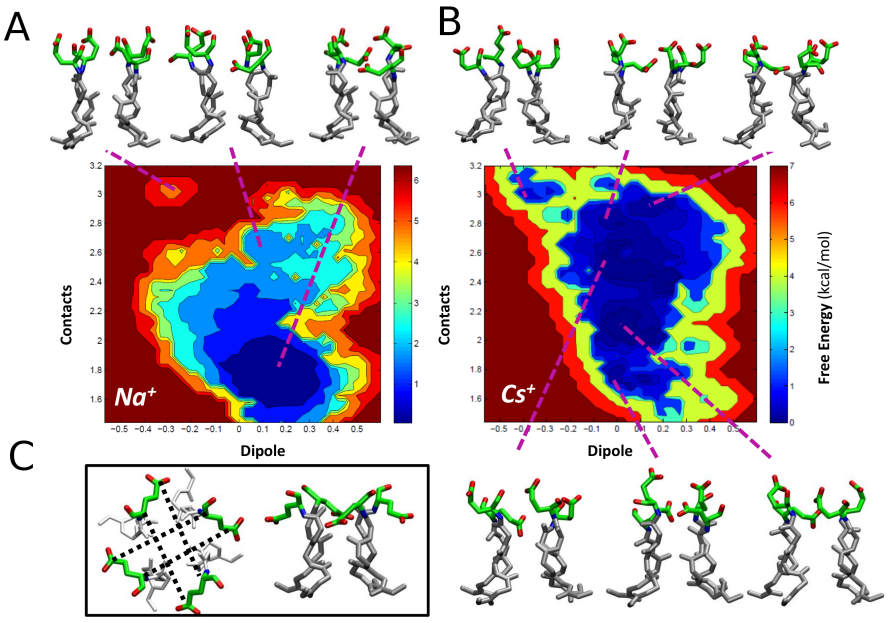

FIGURE 6.3: Free energy of Na^+ and Cs^+ systems. (A, B). The free energy as a function of the z-dipole and of the contacts variable for the Na^+ (A) and Cs^+ system (B). The crystallographic structure used as initial model (PDB ID: 3K0G) falls at the coordinates (0.69; 3.62). (C). The selectivity filter in the crystallographic structure is shown from a top and a side view (E66 in green licorice).

presence of Na^+, the bottom panel in the presence of Cs^+. The difference between the two cases is striking. The average position of the top Na^+ ion changes by only 0.2 Å when the external potential is changed between -0.2 and 0.2 V, while the top Cs^+ moves by more than 2 Å. The Contacts variable shows a net change in its trend in the Cs^+ system. The RMSD of P68 shows a strong voltage dependence in the presence of Cs^+, and not in the presence of Na^+, suggesting that this residue might be related to the anomalous coupling between gating and permeation observed in electrophysiology. Interestingly, previous electrical recordings from CNGA1 channels pointed out the involvement of this proline in Cs^+ permeation [16]. Indeed, when P365, equivalent to P68 in the mimic, was replaced with a threonine, Cs^+ permeation was almost abolished (Figure 6.5). Notably, the presence of intracellular Cs^+ appears to slow down the deactivation kinetics as large Na^+ transients are observed at hyperpolarized membrane potentials only in the presence of Cs^+ (Figure 6.5). Therefore, while Cs^+ ions could still interact or possibly enter the selectivity filter and affect channel gating, the absence of Cs^+ currents at depolarized potentials suggest the existence of a high energy barrier within the selectivity filter preventing Cs^+ translocation.

As a next step, we investigated the difference between the transport mechanism in the

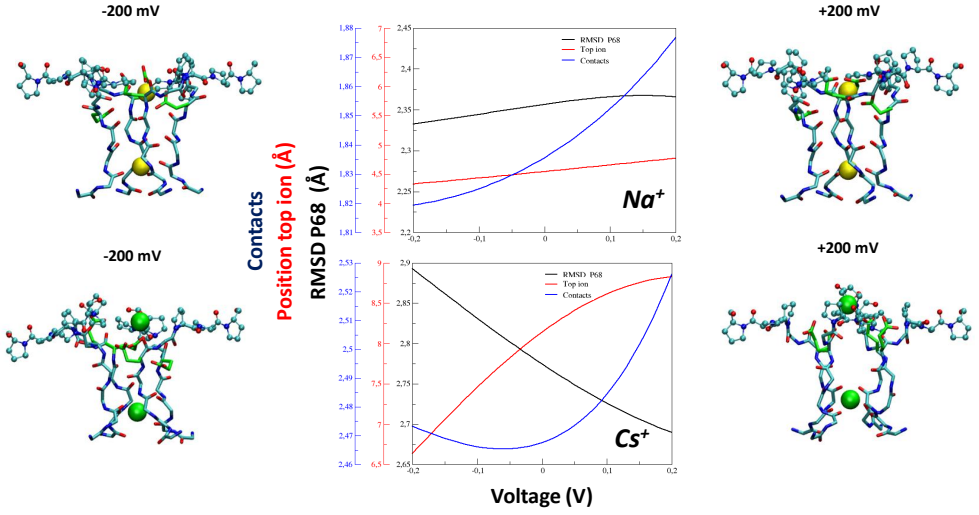

FIGURE 6.4: Dependence on the external potential of the RMSD of P68 with respect to the crystal structure (black line), of the position of the top ion (measured as the distance of the ion from the center of the mass of the heavy atom of the four G65, using only the z-direction) (red line), and of the variable Contacts (blue line), for Na^+ (upper panel) and Cs^+ system (lower panel). Structures of the selectivity filter (licorice), ions (yellow and green spheres, for Na^+ and Cs^+, respectively) and the P68-70 (stick and balls) are reported on the two sides of the graphs, corresponding to different voltages.

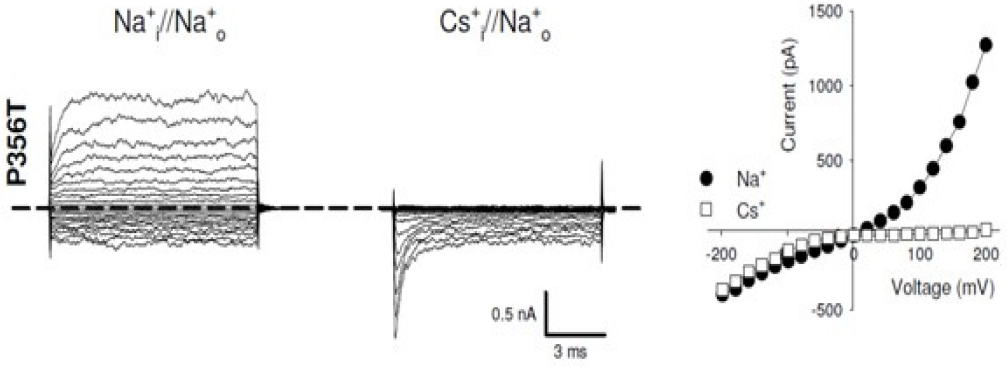

FIGURE 6.5: Macroscopic currents recorded from mutant P365T CNGA1 channel in the presence of 110 mM intracellular Na^+ (left) and Cs^+ (center) with 1 mM cGMP on the cytoplasmic side. Patch pipette was filled with 110 mmM Na^+. Voltage steps from -200 to +200 mV ($\Delta V = 20$ mV). Right: current versus voltage relationship for the recordings in Na^+ (filled circles) and Cs^+ (open squares).

presence of Na^+ and Cs^+, and we searched for structural signatures at the transition states that could explain the observed coupling between permeation and gating. To this end, we projected the free energy along the vertical position of the two ions in the selectivity filter (Figure 6.6). Different minima in this landscape correspond to states in which the position of at least one of the two ions has changed. Therefore, transitions between these minima are associated with transport. While this landscape does not provide a full description of the permeation process, it reveals differences between the Na^+ and the Cs^+ case that are in striking agreement with experimental evidence. In the presence of Na^+, the bottom ion can move between the two local minima without influencing the location of the top ion. Instead, in the presence of Cs^+ the position of the two ions is strongly coupled and, when the bottom ion moves upward, the top ion must move towards the outer vestibule. In the Na^+ case the free energy between the two minima does not change with the external voltage. In the case of Cs^+ the minimum in which both ions are in the selectivity filter is occupied with the same probability of the other minimum only for positive external voltage. For V=0, the first minimum is disfavored by 0.70 kcal/mol and for negative V it is disfavored by 1.29 kcal/mol. This indicates that V influences this step of the transport mechanism only in the Cs^+ case.

6.5 Discussion

By combining electrophysiology, molecular dynamics simulations and X-ray crystallography we have found a remarkable flexibility of the pore region of the CNG mimic. This flexibility is intimately linked to the ion residing within the pore, influences ionic conduction and is in our opinion at the basis of the coupling between gating and permeation. A simple model of the selectivity of ionic channels [145], based on general physical principles of the thermodynamics of ion hydration and geometrical factors, suggested indeed that the poor selectivity of CNG channels could be explained by a significant larger flexibility. These simple physical considerations indicated that the high selectivity of K^+ channels required a very low flexibility of the pore of K^+ channels. In agreement with this earliest notion, crystallographic data of the structure of the CNG mimics collected in the presence of different ionic species inside the pore show highly flexible residues at the pore region. The distance between the opposite carbonyl of V64, located at the narrowest section of the channel pore, increases with the size of the ion accommodated in the pore: for the smallest alkali cation Li^+ is 3 Å and increases up to 5.5 Å for the largest alkali cation, i.e. Cs^+, Figure 6.2. Therefore the pore adapts its shape to the ion it houses. MD simulations and crystallographic data demonstrate a large flexibility of side chains of E66, best seen in Figure 6.3, and suggest that the flexibility of P68 in the outer vestibule could play an important role in the permeation of large alkali cations,

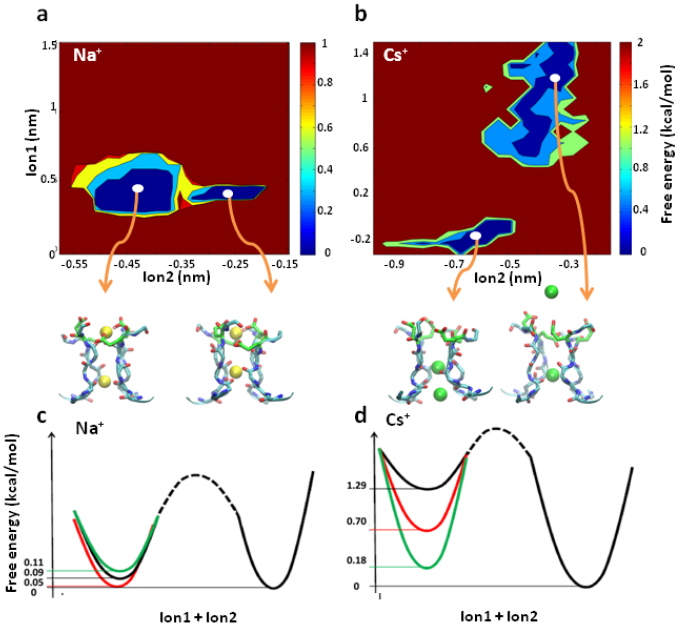

FIGURE 6.6: (a, b) Projection of the free energy (Kcal/mol) along the positions of the top (Ion1) and bottom (Ion2) ion (measured as distances of the ions from the center of the mass of the heavy atom of the four G65s, using only the z-direction). On the x-axes is reported the lower ion, on the y-axes the upper ion. Representative structures of the minima are shown: the selectivity filter drawn in cyan licorice, E66 in green licorice, while the ions as yellow (Na^+) or green (Cs^+) spheres. (c,d) Schematic view of the dependence of the free energy of the two minima on the external field (-200 mV, green line; 0 mV, red line; +200 mV black line).

such as Cs^+ (see Figure 6.4 and 6.5). This conduction mechanism is reminiscent of what has been recently suggested to occur in Ca^{++} and Na^+ channels, whose selectivity filter is formed also by the side chains of several aminoacid in the pore region [144, 146, 147]. Ca^{++} and Na^+ channels are less selective than K^+ channels and have been proposed to be rather flexible. Multi-ion occupancy in these channels is believed to arise from the mobility of a ring of four glutamates residues, known as the EEEE locus, lining the extracellular end of the selectivity filter. Indeed, MD simulations in a bacterial Na^+ channels indicate that the binding of Na^+ is intimately coupled to conformational isomerization of the four glutamate side chains resulting in a highly degenerate, liquid-like energy landscape propitious to Na^+ diffusion [144]. The side walls of the selectivity filter of all K^+ channels as well as the CNG mimics is formed by the carbonyl groups of the aminoacids lining the pore [48]. However, while K^+ channel selectivity filter is viewed as a relatively static and rigid structure our data demonstrates that CNG channel filter is rather flexible and dynamic. What are the structural basis of this strikingly different behavior? In the K^+ channels two key protein interactions involving the tyrosine residue

in the selectivity filter and the charged aspartate, immediately following the signature sequence, play a major role in holding the conductive conformation of the filter and are necessary for K^+ selectivity [60]. The ring of bulky tyrosine residues, known as the aromatic girdle [48, 56], is conserved in most K^+ channels and is thought to provide the putative rigidity and the network of hydrogen bonds necessary to stabilize the structure of the selectivity filter [48, 60, 68]. Moreover, the aspartate immediately following the filter has been shown to be part of a multipoint H-bond network defining the energetic profile of C-type inactivation in KcsA channels [60, 148]. The aromatic girdle is replaced by the ring of E363 in CNG channels (E66 in the mimic), while the aspartate immediately following the filter is substituted by P365 (P68 in the mimic). Therefore these two key structural pillar anchoring the K^+ channels selectivity filter in its conductive configuration are lost in CNG channels. It is therefore not entirely surprising that the results of our MD simulations and our crystallographic data identify the mobility of just E66 and P68 as the major molecular players underlying the coupling between gating and permeation best seen in the presence of symmetrical Na^+ and Cs^+ conditions. Based on our results, we propose that the selectivity filter "liquid-like" behavior herein described underscores the poor ionic selectivity of CNG channels and represents a conduction mode very different from that of classical K^+ channels, which are highly selective and with a fairly rigid molecular structure. This flexibility is likely shared by others non-selective ionic channels, such as TRP and chloride channels, where a significant coupling among gating and permeation is also known to occur.

Chapter 7

Conclusions

Membrane proteins are an important class of biological molecules playing fundamental functions in several aspects of cells' life. Their working mechanisms are less explored with respect to globular proteins as they are more difficult to handle experimentally. For this reason simulations techniques can be of help in assessing structural and functional aspects of this class of proteins, allowing in some cases interpreting and rationalizing puzzling experimental data. In this thesis we focused on two membrane proteins: the Sodium/Galactose Transporter (SGLT) and the Cyclic Nucleotide-Gated (CNG) ion channel, fundamental players in the transport of nutrients inside the cell and in the sensory transduction, respectively. Using enhanced sampling techniques (metadynamis and bias-exchange metadynamics), we elucidated several important aspects of their mechanisms. Concerning the *Vibrio parahaemolyticus* SGLT, we first identified a stable Na^+ ion binding site, which was not discovered crystallographically. Then, we investigated the mechanism of the binding/release of both ligands to/from the protein in the inward-facing conformation, focusing in understanding their interplay during the dissociation process. We were able to simulate this complex process, directly addressing the relative influence of Na^+ and galatose during their dissociation towards the cytoplasm. Our results, for the first time, enlighten a non-cooperative mechanism of Gal and Na^+ ion in their dissociation process from the vSGLT symporter, remanding the cooperativity of the two ligands at the beginning of the transport mechanism (namely the binding of the ligands in the outward-facing conformation or in their conformational switch). Moreover, they point out that the dissociation free energy barriers of Na^+ and galactose are very similar and of the same order of magnitude of the rate limiting step of the whole transport cycle. With these studies we obtained important results on issues that are difficult to address by experimental techniques. Indeed, concerning the direct identification of a stable ion binding site, crystallographic structures need to be solved at a very high resolution, which can be quite difficult for this class of proteins. In a

similar manner, the exit order of the ligands from this secondary active transporter was unknown before our work. Our results offer a rationalization to the difficulties encountered experimentally to characterize this part of the transport mechanism. Indeed, we propose that the experiments can be rationalized with a branched transport mechanism, which is more complex that what originally believed.

In the last part of this work we used the same methods to investigate the mechanistic aspects of the CNG ion channel. For this channel, electrophysiological experiments revealed a linkage permeation-gating and voltage dependence in the presence of large cations. However, an atomistic understanding of these two aspects was missing. Our simulations of a CNG mimic immediately revealed the key role of two residues: E66, in the selectivity filter, and P68, at the external vestibule, and how this flexibility is correlated with the type of cation transported and depends on the external voltage. Together with the crystallographic data, our simulations demonstrated the relevant role of the pore flexibility for this class of channels.

An important part in unraveling the transport mechanism of SGLT that is missing from our study would be the description of the whole inward-to-outward conformational transition. We have recently started to approach this aspect, modeling the outward-facing conformation and starting another BE-MTD simulation in order to capture the whole structural and energetic aspects of this step of the transport mechanism. In the same manner, to do a further step in the mechanistic characterization CNG channels, a very interesting and fulfilling study would be the investigation of the whole permeation pathway of Na^+ and Cs^+ ions along the CNG mimic. This would provide a detailed picture and highlight the precise role of the residues involved in this process.

In short, in this thesis we have provided examples of how molecular simulations, thanks to recent development in computational methods, in the accuracy of force fields and to the ever increasing computational power available, can give an important contribution in complementing experimental techniques to characterize the mechanism of complex biological systems and in particular of membrane proteins.

Acknowledgments

It is quite difficult to thank all the people that made these four years so beautiful and happily lived. Where should I start from?...let's try from the beginning of this special era of my life.
I want to thank Alessio Lodola that suggested to me to do the entrance exam at SISSA. Then, I warmly thank the Professors/Doctors/Researchers that were at my entrance exam and that allowed me to start this adventure. SISSA, such an amazing School, inspiring nature all around, high-level professors, very nice and good people from all the world, in love with Science and here to share their company, thoughts and optimism with the others.
Next thanks goes to my SUPER supervisors: Alessandra Magistrato and Alessandro Laio, that taught to me so much about how to do research, what to look for, how to critical think with an honest and open mind, but in a firm way. Not only, they also made me think about how important is to enjoy life and to give time to our precious dears.
All this journey started with my good friends Marco dS, Francesco dP and Francesca M. Francesco, how can I ever forget our splendid experience in San Francisco? and the mess following 'che bel volante'? Francesca, our playing music together, the conversations. Marco dS, my so dear friend, our jogging at Barcola, swimming at San Giovanni's pool, eating so many pizzas.
My sector: how could have I imagined to receive such a warm welcome in the breast of my sector!? Such good friends, gentle people: Luca T, Emmanuela (from whom I inherited the Secret Friend game!), Zhaleh, Colizzi, Rolando and Trang, Giuseppe F (our conversations, our biking trips), Luca I (biking in via del Pucino), Daniele and Giorgia, Gianpaolo, Duvan, Pilar. A special thanks to my good Fahimeh, deep conversations on not easy topics, great help at work in difficult moments, her smiles and hugs that so easily improved my days, sharing with her for a year our beautiful 329 office. To all these friends, now around the world, I wish a bright and happy life.
And then Angelo, Gdad (sorry, I could not resist), Maria the elegant dancer, Sandro, Richard, Ivan, Jessica.

A grateful and warm thanks to Jacopo and Alex, that helped me so much with my work! And thanks to Giovanni, Cristian and all the rest of SBP sector!
I would like to thank also my friends of the other sectors: Lorenzo, Marco M, David, Marko, Ambra, Gianluca, Danoe, Lee, Gcrl, il Morandotti, Riccardo (Cristofer), Arin, Alessio, Shima, Milad (always check for the right wagon, ok?).
Antonio (Sciá), thank you for being a good friend, and for giving to me such great opportunities to resuscitate my pianist side. For organizing and taking part to such interesting and cultural events. Massimo, Arsenii and Francesca S, so happy to have shared my music time, even if not for too long, with you. Francesca, thanks for letting me play four-hands piano for the first time in my life.
Vanni, so funny (ops, I didn't do it on purpose!), thanks for the laughs we shared together. Marilena, my favourite psicologist, thanks for all our meaningful conversations.
A special thanks to Jamir. So many parties, events, happy moments happend because of you. Without you, probably it would have taken longer (who knows how much) to meet my beloved Marco. Jamir and Gaia, thank you also for the beautiful weekend at Innsbruck. I wish you all the joy of a soon-to-be-merried couple.
And so we came to my Marco: thank you for making the last three years so profoundly serene and glowing.
A dear thanks to Mary and Toni, that hosted my winter and summer holidays with warm hearts and cards on the table. Thanks also to Gimi, for his eternal support.
As all the beautiful and blooming trees have healthy and solid roots, my roots are my parents and my sweet sister. My parents gave everything they had to grant me and my sister greater opportunities in life. I hope to have given back to them at least a small part of what they offered to me.

Ina

ps: a great place like SISSA wouldn't be so perfect without the sunny faces of the people that feed us everyday. So thanks Lucia, Patrizia, Eleonora, Silvia, Lucio, Jack..
The whole picture is marvellously framed by the beautiful Trieste.

Bibliography

[1] C. Kandt, W. L. Ash, and D. Peter Tieleman, "Setting up and running molecular dynamics simulations of membrane proteins," *Methods*, vol. 41, no. 4, pp. 475–488, 2007.

[2] O. Jardetzky, "Simple allosteric model for membrane pumps," *Nature*, vol. 211, pp. 969–970, Aug 1966.

[3] J. Abramson and E. M. Wright, "Structure and function of Na^+-symporters with inverted repeats," *Curr. Opin. Struct. Biol.*, vol. 19, pp. 425–432, Aug 2009.

[4] S. Faham, A. Watanabe, G. M. Besserer, D. Cascio, A. Specht, B. A. Hirayama, E. M. Wright, and J. Abramson, "The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport," *Science*, vol. 321, pp. 810–814, Aug 2008.

[5] A. Yamashita, S. K. Singh, T. Kawate, Y. Jin, and E. Gouaux, "Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters," *Nature*, vol. 437, pp. 215–223, Sep 2005.

[6] J. Li and E. Tajkhorshid, "Ion-releasing state of a secondary membrane transporter," *Biophys. J.*, vol. 97, pp. 29–31, Dec 2009.

[7] E. Zomot and I. Bahar, "The sodium/galactose symporter crystal structure is a dynamic, not so occluded state," *Mol. Biosyst.*, vol. 6, pp. 1040–1046, Jun 2010.

[8] A. Watanabe, S. Choe, V. Chaptal, J. M. Rosenberg, E. M. Wright, M. Grabe, and J. Abramson, "The mechanism of sodium and substrate release from the binding pocket of vSGLT," *Nature*, vol. 468, pp. 988–991, Dec 2010.

[9] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, no. 20, pp. 12562–12566, 2002.

[10] A. Laio and F. Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Rep. Prog. Phys.*, vol. 71, p. 126601, 2008.

[11] J. Li and E. Tajkhorshid, "A gate-free pathway for substrate release from the inward-facing state of the Na^+-galactose transporter," *Biochim. Biophys. Acta*, vol. 1818, pp. 263–271, Feb 2012.

[12] S. Eskandari, E. Wright, and D. Loo, "Kinetics of the reverse mode of the Na^+/glucose cotransporter," *J. Membrane Biol.*, vol. 204, no. 1, pp. 23–32, 2005.

[13] E. M. Wright, D. D. Loo, and B. A. Hirayama, "Biology of human sodium glucose transporters," *Physiol. Rev.*, vol. 91, pp. 733–794, Apr 2011.

[14] M. G. Derebe, W. Zeng, Y. Li, A. Alam, and Y. Jiang, "Structural studies of ion permeation and Ca^{++} blockage of a bacterial channel mimicking the cyclic nucleotide-gated channel pore," *Proc. Natl. Acad. Sci.*, vol. 108, no. 2, pp. 592–597, 2011.

[15] A. Marchesi, M. Mazzolini, and V. Torre, "Gating of cyclic nucleotide-gated channels is voltage dependent," *Nat. Commun.*, vol. 3, p. 973, 2012.

[16] A. Becchetti, K. Gamel, and V. Torre, "Cyclic nucleotide-gated channels pore topology studied through the accessibility of reporter cysteines," *J. Gen. Physiol.*, vol. 114, no. 3, pp. 377–392, 1999.

[17] A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden markov model: application to complete genomes," *J. Mol. Biol.*, vol. 305, no. 3, pp. 567 – 580, 2001.

[18] M. S. Almén, K. J. Nordström, R. Fredriksson, and H. B. Schiöth, "Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin," *BMC biology*, vol. 7, no. 1, p. 50, 2009.

[19] M. A. Yıldırım, K.-I. Goh, M. E. Cusick, A.-L. Barabási, and M. Vidal, "Drug—target network," *Nat. Biotechnol.*, vol. 25, no. 10, pp. 1119–1126, 2007.

[20] J. Deisenhofer, O. Epp, K. Miki, R. Huber, and H. Michel, "Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3 angstrom resolution," *Nature*.

[21] S. H. White, "Biophysical dissection of membrane proteins," *Nature*, vol. 459, no. 7245, pp. 344–346, 2009.

[22] "http://blanco.biomol.uci.edu/mpstruc/."

[23] H. Krishnamurthy, C. L. Piscitelli, and E. Gouaux, "Unlocking the molecular secrets of sodium-coupled transporters," *Nature*, vol. 459, no. 7245, pp. 347–355, 2009.

[24] H. Krishnamurthy and E. Gouaux, "X-ray structures of LeuT in substrate-free outward-open and apo inward-open states," *Nature*, vol. 481, no. 7382, pp. 469–474, 2012.

[25] S. Weyand, T. Shimamura, S. Yajima, S. Suzuki, O. Mirza, K. Krusong, E. Carpenter, N. Rutherford, J. Hadden, J. O'Reilly, and others., "Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter," *Science*, vol. 322, no. 5902, p. 709, 2008.

[26] S. Ressl, A. van Scheltinga, C. Vonrhein, V. Ott, and C. Ziegler, "Molecular basis of transport and regulation in the Na^+/betaine symporter BetP," *Nature*, vol. 458, no. 7234, pp. 47–52, 2009.

[27] C. Perez, C. Koshy, Ö. Yildiz, and C. Ziegler, "Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter BetP," *Nature*, vol. 490, no. 7418, pp. 126–130, 2012.

[28] D. Yernool, O. Boudker, Y. Jin, and E. Gouaux, "Structure of a glutamate transporter homologue from pyrococcus horikoshii," *Nature*, vol. 431, no. 7010, pp. 811–818, 2004.

[29] T. Zeuthen and N. MacAulay, "Transport of water against its concentration gradient: fact or fiction?," *Wiley Interdiscip. Rev.: Membrane Transport and Signaling*, vol. 1, no. 4, pp. 373–381, 2012.

[30] F. Zhu, "How does water pass through a sugar transporter?," *Biophys. J.*, vol. 106, no. 6, pp. 1229–1230, 2014.

[31] T. Zeuthen, A. Meinild, D. Loo, E. Wright, and D. Klaerke, "Isotonic transport by the Na^+-glucose cotransporter SGLT1 from humans and rabbit," *J. Physiol.*, vol. 531, no. 3, pp. 631–644, 2004.

[32] L. Sasseville, J. Cuervo, J. Lapointe, and S. Noskov, "The structural pathway for water permeation through sodium-glucose cotransporters," *Biophys. J.*, vol. 101, no. 8, pp. 1887–1895, 2011.

[33] E. Wright, B. Hirayama, and D. Loo, "Active sugar transport in health and disease," *J. Intern. Med.*, vol. 261, no. 1, pp. 32–43, 2007.

[34] K. Khafizov, C. Perez, C. Koshy, M. Quick, K. Fendler, C. Ziegler, and L. Forrest, "Investigation of the sodium-binding sites in the sodium-coupled betaine transporter BetP," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 44, pp. E3035–E3044, 2012.

[35] C. Perez, B. Faust, A. R. Mehdipour, K. A. Francesconi, L. R. Forrest, and C. Ziegler, "Substrate-bound outward-open state of the betaine transporter BetP provides insights into Na^+ coupling," *Nat. Commun.*, vol. 5, 2014.

[36] C. Zhao and S. Noskov, "The role of local hydration and hydrogen-bonding dynamics in ion and solute release from ion-coupled secondary transporters," *Biochemistry*, vol. 50, no. 11, pp. 1848–1856, 2011.

[37] C. Zhao, S. Stolzenberg, L. Gracia, H. Weinstein, S. Noskov, and L. Shi, "Ion-controlled conformational dynamics in the outward-open transition from an occluded state of LeuT," *Biophys. J.*, vol. 103, no. 5, pp. 878–888, 2012.

[38] T. Bastug, G. Heinzelmann, S. Kuyucak, M. Salim, R. Vandenberg, and R. Ryan, "Position of the third Na^+ site in the aspartate transporter GltPh and the human glutamate transporter, EAAT1," *PloS one*, vol. 7, no. 3, p. e33058, 2012.

[39] J. L. Adelman, Y. Sheng, S. Choe, J. Abramson, E. M. Wright, J. M. Rosenberg, and M. Grabe, "Structural determinants of water permeation through the sodium-galactose transporter vSGLT," *Biophys. J.*, vol. 106, no. 6, pp. 1280–1289, 2014.

[40] I. H. Shrivastava, J. Jiang, S. G. Amara, and I. Bahar, "Time-resolved mechanism of extracellular gate opening and substrate binding in a glutamate transporter," *J. Biol. Chem.*, vol. 283, no. 42, pp. 28680–28690, 2008.

[41] Y. Gu, I. Shrivastava, S. Amara, and I. Bahar, "Molecular simulations elucidate the substrate translocation pathway in a glutamate transporter," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 8, pp. 2589–2594, 2009.

[42] G. Grazioso, V. Limongelli, D. Branduardi, E. Novellino, C. De Micheli, A. Cavalli, and M. Parrinello, "Investigating the mechanism of substrate uptake and release in the glutamate transporter homologue GltPh through metadynamics simulations," *J. Am. Chem. Soc.*, vol. 134, no. 1, pp. 453–463, 2011.

[43] L. Forrest and G. Rudnick, "The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters," *Physiology*, vol. 24, no. 6, pp. 377–386, 2009.

[44] T. Shimamura, S. Weyand, O. Beckstein, N. Rutherford, J. Hadden, D. Sharples, M. Sansom, S. Iwata, P. Henderson, and A. Cameron, "Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1," *Science*, vol. 328, no. 5977, pp. 470–473, 2010.

[45] B. Hille, *Ion channels of excitable membranes*, vol. 507. Sinauer Sunderland, MA, 2001.

[46] Y. Shi, "Common folds and transport mechanisms of secondary active transporters," *Annu. Rev. Biophys.*, vol. 42, pp. 51–72, 2013.

[47] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J. Physiol.*, vol. 117, no. 4, p. 500, 1952.

[48] D. A. Doyle, J. M. Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon, "The structure of the potassium channel: molecular basis of K^+ conduction and selectivity," *Science*, vol. 280, no. 5360, pp. 69–77, 1998.

[49] I. B. Leskov, V. A. Klenchin, J. W. Handy, G. G. Whitlock, V. I. Govardovskii, M. D. Bownds, T. D. Lamb, E. N. Pugh Jr, and V. Y. Arshavsky, "The gain of rod phototransduction: reconciliation of biochemical and electrophysiological measurements," *Neuron*, vol. 27, no. 3, pp. 525–537, 2000.

[50] D.-J. Zou, A. Chesler, and S. Firestein, "How the olfactory bulb got its glomeruli: a just so story?," *Nat. Rev. Neurosci.*, vol. 10, no. 8, pp. 611–618, 2009.

[51] L. Stryer, "Cyclic GMP cascade of vision," *Annu. Rev. Neurosci.*, vol. 9, no. 1, pp. 87–119, 1986.

[52] K. B. Craven and W. N. Zagotta, "CNG and HCN channels: two peas, one pod," *Annu. Rev. Physiol.*, vol. 68, pp. 375–401, 2006.

[53] D. B. Sauer, W. Zeng, J. Canty, Y. Lam, and Y. Jiang, "Sodium and potassium competition in potassium-selective and non-selective channels," *Nat. Commun.*, vol. 4, 2013.

[54] F. Bezanilla and C. M. Armstrong, "Negative conductance caused by entry of sodium and cesium ions into the potassium channels of squid axons," *J. Gen. Physiol.*, vol. 60, no. 5, pp. 588–608, 1972.

[55] J. Neyton and C. Miller, "Discrete Ba^{++} block as a probe of ion occupancy and pore structure in the high-conductance Ca^{++}-activated K^+ channel.," *J. Gen. Physiol.*, vol. 92, no. 5, pp. 569–586, 1988.

[56] Y. Zhou, J. H. Morais-Cabral, A. Kaufman, and R. MacKinnon, "Chemistry of ion coordination and hydration revealed by a K^+ channel–fab complex at 2.0 å resolution," *Nature*, vol. 414, no. 6859, pp. 43–48, 2001.

[57] S. W. Lockless, M. Zhou, and R. MacKinnon, "Structural and thermodynamic properties of selective ion binding in a K^+ channel," *PLoS biology*, vol. 5, no. 5, p. e121, 2007.

[58] S. Y. Noskov and B. Roux, "Importance of hydration and dynamics on the selectivity of the KcsA and NaK channels," *J. Gen. Physiol.*, vol. 129, no. 2, pp. 135–143, 2007.

[59] A. Alam and Y. Jiang, "Structural analysis of ion selectivity in the NaK channel," *Nat. Struct. Mol. Biol.*, vol. 16, no. 1, pp. 35–41, 2008.

[60] D. B. Sauer, W. Zeng, S. Raghunathan, and Y. Jiang, "Protein interactions central to stabilizing the K^+ channel selectivity filter in a four-sited configuration for selective K^+ permeation," *Proc. Natl. Acad. Sci.*, vol. 108, no. 40, pp. 16634–16639, 2011.

[61] S. Furini and C. Domene, "Nonselective conduction in a mutated NaK channel with three cation-binding sites," *Biophys. J.*, vol. 103, no. 10, pp. 2106–2114, 2012.

[62] A. N. Thompson, I. Kim, T. D. Panosian, T. M. Iverson, T. W. Allen, and C. M. Nimigean, "Mechanism of potassium-channel selectivity revealed by Na^+ and Li^+ binding sites within the KcsA pore," *Nat. Struct. Mol. Biol.*, vol. 16, no. 12, pp. 1317–1324, 2009.

[63] R. Blunck, J. F. Cordero-Morales, L. G. Cuello, E. Perozo, and F. Bezanilla, "Detection of the opening of the bundle crossing in KcsA with fluorescence lifetime spectroscopy reveals the existence of two gates for ion conduction," *J. Gen. Physiol.*, vol. 128, no. 5, pp. 569–581, 2006.

[64] S. Berneche and B. Roux, "Energetics of ion conduction through the K^+ channel," *Nature*, vol. 414, no. 6859, pp. 73–76, 2001.

[65] J. H. Morais-Cabral, Y. Zhou, and R. MacKinnon, "Energetic optimization of ion conduction rate by the K^+ selectivity filter," *Nature*, vol. 414, no. 6859, pp. 37–42, 2001.

[66] J. E. Contreras, D. Srikumar, and M. Holmgren, "Gating at the selectivity filter in cyclic nucleotide-gated channels," *Proc. Natl. Acad. Sci.*, vol. 105, no. 9, pp. 3310–3314, 2008.

[67] M. Mazzolini, A. Marchesi, A. Giorgetti, and V. Torre, "Gating in CNGA1 channels," *Pflug. Arch. Eur. J. Phy.*, vol. 459, no. 4, pp. 547–555, 2010.

[68] S. Bernèche and B. Roux, "A gate in the selectivity filter of potassium channels," *Structure*, vol. 13, no. 4, pp. 591–600, 2005.

[69] S. Chakrapani, J. F. Cordero-Morales, V. Jogini, A. C. Pan, D. M. Cortes, B. Roux, and E. Perozo, "On the structural basis of modal gating behavior in K^+ channels," *Nat. Struct. Mol. Biol.*, vol. 18, no. 1, pp. 67–74, 2011.

[70] S. Furini and C. Domene, "Gating at the selectivity filter of ion channels that conduct Na^+ and K^+ ions," *Biophys. J.*, vol. 101, no. 7, pp. 1623–1631, 2011.

[71] B. Hille and W. Schwarz, "Potassium channels as multi-ion single-file pores.," *J. Gen. Physiol.*, vol. 72, no. 4, pp. 409–442, 1978.

[72] A. L. Hodgkin and R. Keynes, "The potassium permeability of a giant nerve fibre," *J. Physiol.*, vol. 128, no. 1, pp. 61–88, 1955.

[73] S. Furini and C. Domene, "On conduction in a bacterial sodium channel," *PLoS Comput. Biol.*, vol. 8, no. 4, p. e1002476, 2012.

[74] A. R. Leach, *Molecular modelling: principles and applications*. Pearson Education, 2001.

[75] W. Van Gunsteren and H. Berendsen, "A leap-frog algorithm for stochastic dynamics," *Mol. Simulat.*, vol. 1, no. 3, pp. 173–185, 1988.

[76] P. P. Ewald, "Die berechnung optischer und elektrostatischer gitterpotentiale," *Annalen der Physik*, vol. 369, no. 3, pp. 253–287, 1921.

[77] T. Darden, D. York, and L. Pedersen, "Particle mesh ewald: An $nlog(n)$ method for ewald sums in large systems," *J. Chem. Phys.*, vol. 98, no. 12, pp. 10089–10092, 1993.

[78] G. J. Martyna, M. L. Klein, and M. Tuckerman, "Nosé–hoover chains: the canonical ensemble via continuous dynamics," *J. Chem. Phys.*, vol. 97, no. 4, pp. 2635–2643, 1992.

[79] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *J. Appl. Phys.*, vol. 52, no. 12, pp. 7182–7190, 1981.

[80] S. Nose and M. Klein, "Constant pressure molecular dynamics for molecular systems," *Mol. Phys.*, vol. 50, no. 5, pp. 1055–1076, 1983.

[81] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello, "Assessing the accuracy of metadynamics," *J. Phys. Chem. B.*, vol. 109, pp. 6714–6721, 2005.

[82] G. Bussi, A. Laio, and M. Parrinello, "Equilibrium free energies from nonequilibrium metadynamics," *Phys. Rev. Lett.*, vol. 96, p. 090601, MAR 10 2006.

[83] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chem. Phys. Lett.*, vol. 314, no. 1, pp. 141–151, 1999.

[84] Y. Sugita, A. Kitao, and Y. Okamoto, "Multidimensional replica−exchange method for free-energy calculations," *J. Chem. Phys.*, vol. 113, no. 15, pp. 6042–6051, 2000.

[85] G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello, "Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics," *J. Am. Chem. Soc.*, vol. 128, pp. 13435–13441, OCT 18 2006.

[86] S. Piana and A. Laio, "A bias-exchange approach to protein folding," *J. Phys. Chem. B*, vol. 111, no. 17, pp. 4553–4559, 2007.

[87] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method," *J. Comput. Chem.*, vol. 13, no. 8, pp. 1011–1021, 1992.

[88] F. Marinelli, F. Pietrucci, A. Laio, and S. Piana, "A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations," *PLoS Comput. Biol.*, vol. 5, no. 8, p. e1000452, 2009.

[89] D. Branduardi, F. L. Gervasio, and M. Parrinello, "From a to b in free energy space," *J. Chem. Phys.*, vol. 126, no. 5, p. 054103, 2007.

[90] P. Cossio, F. Marinelli, A. Laio, and F. Pietrucci, "Optimizing the performance of bias-exchange metadynamics: folding a 48-residue LysM domain using a coarse-grained model," *J. Phys. Chem. B*, vol. 114, no. 9, pp. 3259–3265, 2010.

[91] F. Baftizadeh, P. Cossio, F. Pietrucci, and A. Laio, "Protein folding and ligand-enzyme binding from bias-exchange metadynamics simulations," *Curr. Phys. Chem.*, vol. 2, pp. 79–91, 2012.

[92] N. Guex and M. C. Peitsch, "Swiss-model and the swiss-pdbviewer: an environment for comparative protein modeling," *Electrophoresis*, vol. 18, pp. 2714–2723, Dec 1997.

[93] Z. Xiang, C. S. Soto, and B. Honig, "Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, pp. 7432–7437, May 2002.

[94] B. Jójárt and T. A. Martinek, "Performance of the general amber force field in modeling aqueous popc membrane bilayers," *J. Comput. Chem.*, vol. 28, pp. 2051–2058, Sep 2007.

[95] J. Sgrignani and A. Magistrato, "Influence of the membrane lipophilic environment on the structure and on the substrate access/egress routes of the human aromatase enzyme. a computational study," *J. Chem. Inf. Model.*, vol. 52, no. 6, pp. 1595–1606, 2012.

[96] M. G. Wolf, M. Hoefling, C. Aponte-Santamaría, H. Grubmüller, and G. Groenhof, "g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation," *J. Comput. Chem.*, vol. 31, pp. 2169–2174, Aug 2010.

[97] M. A. Lomize, A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, "Opm: orientations of proteins in membranes database," *Bioinformatics*, vol. 22, pp. 623–625, Mar 2006.

[98] M. Mahoney and W. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *J. Chem. Phys.*, vol. 112, p. 8910, 2000.

[99] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, "Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *J. Chem. Theory Comp.*, vol. 4, no. 3, pp. 435–447, 2008.

[100] Y. Duan, C. Wu, S. Chowdhury, M. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, and others., "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations," *J. Comput. Chem.*, vol. 24, no. 16, pp. 1999–2012, 2003.

[101] J. Wang, R. Wolf, J. Caldwell, P. Kollman, and D. Case, "Development and testing of a general amber force field," *J. Comput. Chem.*, vol. 25, no. 9, pp. 1157–1174, 2004.

[102] C. Bayly, P. Cieplak, W. Cornell, and P. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model," *J. Phys. Chem.*, vol. 97, no. 40, pp. 10269–10280, 1993.

[103] C. Lee, W. Yang, and R. Parr, "Development of the colle-salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, vol. 37, no. 2, p. 785, 1988.

[104] P. Stephens, F. Devlin, C. Chabalowski, and M. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields," *J. Phys. Chem.*, vol. 98, no. 45, pp. 11623–11627, 1994.

[105] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, J. Montgomery Jr, T. Vreven, K. Kudin, J. Burant, and others., "Gaussian 03, rev. c. 02. gaussian," *Inc., Wallingford, CT*, 2004.

[106] D. Evans and B. Holian, "The nose–hoover thermostat," *J. Chem. Phys.*, vol. 83, p. 4069, 1985.

[107] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. Broglia, and others., "Plumed: A portable plugin for free-energy calculations with molecular dynamics," *Comput. Phys. Commun.*, vol. 180, no. 10, pp. 1961–1972, 2009.

[108] N. Kučerka, S. Tristram-Nagle, and J. Nagle, "Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains," *J. Membrane Biol.*, vol. 208, no. 3, pp. 193–202, 2006.

[109] S. Choe, J. Rosenberg, J. Abramson, E. Wright, and M. Grabe, "Water permeation through the sodium-dependent galactose cotransporter vSGLT," *Biophys. J.*, vol. 99, no. 7, pp. L56–L58, 2010.

[110] D. Loo, E. Wright, and T. Zeuthen, "Water pumps," *J. Physiol.*, vol. 542, no. 1, pp. 53–60, 2004.

[111] T. Zeuthen, A. Meinild, D. Loo, E. Wright, and D. Klaerke, "Isotonic transport by the Na^+-glucose cotransporter SGLT1 from humans and rabbit," *J. Physiol.*, vol. 531, no. 3, pp. 631–644, 2004.

[112] P. Duquette, P. Bissonnette, and J. Lapointe, "Local osmotic gradients drive the water flux associated with Na^+/glucose cotransport," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, no. 7, pp. 3796–3801, 2001.

[113] M. Harding, "Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium," *Acta Crystallogr. D*, vol. 58, no. 5, pp. 872–874, 2002.

[114] J. Evanseck, L. Caves, and M. Karplus, "Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin," *Protein Sci.*, vol. 7, no. 3, pp. 649–666, 1998.

[115] M. Quick, D. Loo, and E. Wright, "Neutralization of a conserved amino acid residue in the human Na^+/glucose transporter (hSGLT1) generates a glucose-gated H^+ channel," *J. Biol. Chem.*, vol. 276, no. 3, pp. 1728–1734, 2001.

[116] M. Quick and H. Jung, "A conserved aspartate residue, asp187, is important for Na^+-dependent proline binding and transport by the Na^+/proline transporter of escherichia coli," *Biochemistry*, vol. 37, no. 39, pp. 13800–13806, 1998.

[117] M. Zimmer, "Are classical molecular mechanics calculations still useful in bioinorganic simulations?," *Coordin. Chem. Rev.*, vol. 253, no. 5-6, pp. 817–826, 2009.

[118] J. Sgrignani and R. Pierattelli, "Nuclear magnetic resonance signal chemical shifts and molecular simulations: a multidisciplinary approach to modeling copper protein structures," *J. Biol. Inorg. Chem.*, vol. 17, no. 1, pp. 71–79, 2012.

[119] I. S. Joung and T. E. Cheatham, "Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations," *J. Phys. Chem. B*, vol. 112, pp. 9020–9041, Jul 2008.

[120] M. Sala-Rabanal, B. A. Hirayama, D. D. Loo, V. Chaptal, J. Abramson, and E. M. Wright, "Bridging the gap between structure and kinetics of human SGLT1," *Am. J. Physiol.-Cell Ph.*, vol. 302, no. 9, p. C1293, 2012.

[121] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, pp. 12562–12566, Oct 2002.

[122] X. Biarnés, F. Pietrucci, F. Marinelli, and A. Laio, "Metagui. a vmd interface for analyzing metadynamics and molecular dynamics simulations," *Comput. Phys. Commun.*, vol. 183, no. 1, pp. 203–211, 2012.

[123] W. Humphrey, A. Dalke, and K. Schulten, "Vmd: visual molecular dynamics," *J. Mol. Graphics*, vol. 14, no. 1, pp. 33–38, 1996.

[124] D. Franco, A. V. Vargiu, and A. Magistrato, "Ru [(bpy) 2 (dppz)] 2+ and rh [(bpy) 2 (chrysi)] 3+ targeting double strand dna: The shape of the intercalating ligand tunes the free energy landscape of deintercalation," *Inorg. Chem.*, vol. 53, no. 15, pp. 7999–8008, 2014.

[125] A. V. Vargiu, P. Ruggerone, A. Magistrato, and P. Carloni, "Dissociation of minor groove binders from DNA: insights from metadynamics simulations," *Nucleic Acids Res.*, vol. 36, no. 18, pp. 5910–5921, 2008.

[126] "http://www.ebi.ac.uk/."

[127] J. Söding, A. Biegert, and A. N. Lupas, "The hhpred interactive server for protein homology detection and structure prediction," *Nucleic Acids Res.*, vol. 33, no. suppl 2, pp. W244–W248, 2005.

[128] I. Bisha, A. Laio, A. Magistrato, A. Giorgetti, and J. Sgrignani, "A candidate ion-retaining state in the inward-facing conformation of sodium/galactose symporter: Clues from atomistic simulations," *J. Chem. Theory Comp.*, vol. 9, no. 2, pp. 1240–1246, 2013.

[129] E. Turk, O. Kim, J. le Coutre, J. P. Whitelegge, S. Eskandari, J. T. Lam, M. Kreman, G. Zampighi, K. F. Faull, and E. M. Wright, "Molecular characterization of Vibrio parahaemolyticus vSGLT a model for sodium-coupled sugar cotransporters," *J. Biol. Chem.*, vol. 275, no. 33, pp. 25711–25716, 2000.

[130] D. D. Loo, B. A. Hirayama, A. Cha, F. Bezanilla, and E. M. Wright, "Perturbation analysis of the voltage-sensitive conformational changes of the Na^+/glucose cotransporter," *J. Gen. Physiol.*, vol. 125, no. 1, pp. 13–36, 2005.

[131] D. D. Loo, B. A. Hirayama, M. H. Karakossian, A.-K. Meinild, and E. M. Wright, "Conformational dynamics of hSGLT1 during Na^+/glucose cotransport," *J. Gen. Physiol.*, vol. 128, no. 6, pp. 701–720, 2006.

[132] J.-P. Longpré and J.-Y. Lapointe, "Determination of the Na^+/glucose cotransporter (SGLT1) turnover rate using the ion-trap technique," *Biophys. J.*, vol. 100, no. 1, pp. 52–59, 2011.

[133] L. Shi, M. Quick, Y. Zhao, H. Weinstein, and J. A. Javitch, "The mechanism of a neurotransmitter: Sodium symporter—inward release of Na^+ and substrate is triggered by substrate in a second binding site," *Mol. Cell*, vol. 30, no. 6, pp. 667–677, 2008.

[134] C. Zhao and S. Y. Noskov, "The molecular mechanism of ion-dependent gating in secondary transporters," *PLoS Comput. Biol.*, vol. 9, no. 10, p. e1003296, 2013.

[135] M. H. Cheng and I. Bahar, "Coupled global and local changes direct substrate translocation by neurotransmitter-sodium symporter ortholog LeuT," *Biophys. J.*, vol. 105, no. 3, pp. 630–639, 2013.

[136] B. A. Hirayama, D. D. Loo, A. Díez-Sampedro, D. W. Leung, A.-K. Meinild, M. Lai-Bing, E. Turk, and E. M. Wright, "Sodium-dependent reorganization of the sugar-binding site of SGLT1," *Biochemistry*, vol. 46, no. 46, pp. 13391–13406, 2007.

[137] E. Turk, C. J. Kerner, M. P. Lostao, and E. M. Wright, "Membrane topology of the human Na^+/glucose cotransporter SGLT1," *J. Biol. Chem.*, vol. 271, no. 4, pp. 1925–1934, 1996.

[138] M. Raja and R. K. Kinne, "Structural insights into genetic variants of Na^+/glucose cotransporter SGLT1 causing glucose–galactose malabsorption: vSGLT as a model structure," *Cell Biochem. Biophys.*, vol. 63, no. 2, pp. 151–158, 2012.

[139] M. Ruiz and J. W. Karpen, "Single cyclic nucleotide-gated channels locked in different ligand-bound states," *Nature*, vol. 389, no. 6649, pp. 389–392, 1997.

[140] D. H. Hackos and J. I. Korenbrot, "Divalent cation selectivity is a function of gating in native and recombinant cyclic nucleotide–gated ion channels from retinal photoreceptors," *J. Gen. Physiol.*, vol. 113, no. 6, pp. 799–818, 1999.

[141] K. Gamel and V. Torre, "The interaction of Na^+ and K^+ in the pore of cyclic nucleotide-gated channels," *Biophys. J.*, vol. 79, no. 5, pp. 2475–2493, 2000.

[142] M. K. Higgins, D. Weitz, T. Warne, G. F. Schertler, and U. B. Kaupp, "Molecular architecture of a retinal cGMP-gated channel: the arrangement of the cytoplasmic domains," *The EMBO journal*, vol. 21, no. 9, pp. 2087–2094, 2002.

[143] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[144] N. Chakrabarti, C. Ing, J. Payandeh, N. Zheng, W. A. Catterall, and R. Pomès, "Catalysis of Na^+ permeation in the bacterial sodium channel NaVAb," *Proc. Natl. Acad. Sci.*, vol. 110, no. 28, pp. 11331–11336, 2013.

[145] A. Laio and V. Torre, "Physical origin of selectivity in ionic channels of biological membranes," *Biophys. J.*, vol. 76, no. 1, pp. 129–148, 1999.

[146] L. Tang, T. M. G. El-Din, J. Payandeh, G. Q. Martinez, T. M. Heard, T. Scheuer, N. Zheng, and W. A. Catterall, "Structural basis for Ca^{++} selectivity of a voltage-gated calcium channel," *Nature*, vol. 505, no. 7481, pp. 56–61, 2014.

[147] J. Payandeh, T. Scheuer, N. Zheng, and W. A. Catterall, "The crystal structure of a voltage-gated sodium channel," *Nature*, vol. 475, no. 7356, pp. 353–358, 2011.

[148] J. F. Cordero-Morales, V. Jogini, A. Lewis, V. Vásquez, D. M. Cortes, B. Roux, and E. Perozo, "Molecular driving forces determining potassium channel slow inactivation," *Nat. Struct. Mol. Biol.*, vol. 14, no. 11, pp. 1062–1069, 2007.