MOLECULAR MECHANISMS OF GATING AND SELECTIVITY IN TRANSPORT CHANNELS

BY

RAMYA BHARGAVI GAMINI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biophysics and Computational Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Klaus Schulten, Chair
Associate Professor Emad Tajkhorshid
Associate Professor Claudio F. Grosman
Assistant Professor Charles M. Schroeder

Abstract

Transport channels and pores are of fundamental importance for translocation of molecules across the otherwise-impermeable biological membrane. As a conduit for the passage of material, specific membrane channels must accomplish the task of selective transport of molecules, while preventing unnecessary loss of other cellular material. The nuclear pore complex (NPC), and the mechanosensitive channel of small conductance, MscS are the focus of the research presented in this thesis.

NPCs are sole gateways for passage of material across the nuclear envelope of eukaryotic cells. Several unstructured proteins that are rich in phenylaline- glycine motifs (FG-nups) form the central transport channel. Small molecules passively diffuse through the channel but, larger molecules are selectively transported via transport factors (TFs), which apparently interact with the FG-repeats of the nups in the channel. To understand how nups are assembled in the interior of the NPC, assemblies of one kind of nup, starting from different initial states, are investigated. Results suggest nups form different structures in different regions of the central channel. While only molecules of size < 9nm can penetrate through, a limit known for passive diffusion, the resulting structure posed a selectivity barrier for larger molecules that can be penetrated only via TF interactions.

Mechanosensitive (MS) channels, bacterial inner-membrane proteins, open and close in response to mechanical stimuli such as changes in membrane tension during osmotic stress. These channels act as safety valves preventing cell lysis upon hypoosmotic cell swelling: the channels open under membrane tension to release osmolytes along with water. MscS, consists, beside the transmembrane channel, of a large cytoplasmic domain (CD) that features a balloon-like, water filled chamber opening to the cytoplasm through seven side pores and a small distal pore. The CD is apparently a molecular sieve covering the channel, that optimizes loss of osmolytes during osmoadaptation. Diffusion theory and MD simulations are employed to explore the transport kinetics of Glu^- and K^+ as representative osmolytes. A role of a filter is suggested for the CD such that it balances passage of Glu^- and K^+ osmolytes, to yield a largely neutral efflux, thereby, reducing cell depolarization in the open state that also conserves to a large degree the essential metabolite Glu^-.

I lovingly dedicate this to my parents and Srinu, for their endless love.

Acknowledgements

This dissertation is a milestone in my academic carrier. I am grateful to a number of people who have helped me in achieving this goal.

I would like to thank first my adviser, Klaus Schulten, without whom none of this work would have been accomplished. I am grateful to him for his guidance, encouragement and help throughout the past five and half years. Conversations with him have always been insightful and motivating and his frank remarks encouraging to strive and perform better. Although, he leads a large group, he is never tired and keeps in the best interests for the group. To me, he has been exceptionally supportive and understanding especially during the last phase of my Ph.D. while I was also adapting as a young mother. I shall cherish this support forever.

I would also like to thank the other members of my committee, Emad Tajkhorshid, Claudio Grosman and Charles Schroeder, for all the support and for providing guidance to me outside of their role as committee members. Emad Tajkhorshid with whom I rotated during my first year in graduate school, and Claudio Grosman and Charles Schroeder advised me during tutorials. I have had a great learning experience interacting with all of them.

To my family. To mom and dad, for their love, unconditional support and for having faith in me. To them, I owe everything. I would also like to extend my thanks to my in-laws, for welcoming into their family as their own and for their encouragement to continue on my goal. To my dear sisters, Soujanya and Divya, and brother-in-law, Vijay, who are always there for me. And to my little nephew Abhinav, who makes me happy. And of course, my husband, Srinu, who remains willing to be there for me, for his support and endurance that allowed me to finish this journey. To you, I promise you will see the wife-y me more often. Finally, to my own little baby, Vaibhav, for his priceless smile is probably the only comfort in the last months while preparing for my final defense.

Special thanks to all members (past and present) of the TCB group. They were not merely labmates but also invaluable friends to me. To Xueqing, my wonderful office-mate and friend, for her admirable patience listening to me during my downtime and doubling my happiness when happy.

To Wei Han, who replaced Xueqing not only as an office-mate but in every way when she left the group to join her husband. Wei also helped me tremendously with my research. I am particularly thankful to former group member Peter Fedrellino for his great patience in answering my endless questions when I started as a first-year graduate student. To Marcos Sotomayor, also former group member, who helped me get started with research in the group. To Eduard Schriener, with whom I shared many events in and outside Lab, for instilling in me confidence, and for being a very good friend. And to Eric Lee, for those special thanksgiving dinners, especially warm during my early years in the US when I had just left my family far back in India. I'd also like to thank the good company of Danielle Chandler, Chaitanya Sathe, Wen Ma, Kirby Vandivort, Barry Isralewitz, David Hardy, Leonardo Trabuco, JC, Yi Wang, Ying Yin, Anton Arkhipov, Aruna Rajan, David Tanner.

To Jim Phillips, John Stone, for their outstanding technical support and for answering all my questions helping me learn NAMD and VMD. To Haley Mui, for keeping our computers functioning and providing me infinite disk space.

My sincere gratitude to Cindy Dodds, Nancy Mallon and Jo Miller for taking care of the academic paperwork, publication procedures, which helped me concentrate on my research goal. To Nancy and Jo also for bringing joy and providing moral support whenever I needed it.

Outside lab, thanks to some of my friends who kept me happy with their good company Jeff, Lance, Leonardo, Gemma, Taras, Seher, Girai, Po-chao and Vita Solov'yov.

This work was supported by grants R01-GM067887 and P41-GM104601 from the National Institutes of Health. Supercomputer time was provided by the National Center for Supercomputing Applications, Blue waters (NSF OCI 07-25070) and Extreme Science and Engineering Discovery Environment (XSEDE) via XSEDE Resources Allocation Committee grant MCA93S028.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABF	Adaptive biasing force.
NPC	Nuclear pore complex.
CG	Coarse-grained.
FG	Phenylalanine-Glycine motif in nuclear pore proteins.
NUP	Nucleoporins or nuclear pore proteins.
CHARMM	Chemistry at Harvard macromolecular mechanics software and force-field.
CMAP	Extension to the CHARMM22 protein force-field correcting backbone dynamics.
MscS	Mechanosensitive channel of small conductance.
MD	Molecular dynamics.
NAMD	Software for nanoscale molecular dynamics.
NPT	Simulation conditions of constant number of particles, pressure and temperature.
NVT	Simulation conditions of constant number of particles, volume and temperature.
PDB	Protein data bank.
PME	Particle mesh ewald, method to calculate electrostatics.
PMF	Potential of mean force.
TFs	Transport factors; carrier proteins for cargo transport.
VMD	Software for visualization of molecular dynamics.

Chapter 1

Introduction

Exchange of materials across biological membranes is one of the most fundamental and highly regulated processes in biology of all living organisms. While small hydrophobic molecules can readily permeate lipid bilayers, specialized membrane proteins have been evolved to facilitate, efficiently the transport of almost every chemical species. Two such protein channels, the nuclear pore complex (NPC) and the mechanosensitive channel for small conductance (MscS) are the focus of the research presented in this thesis.

Below, a brief description of the two channels, as well as the key questions addressed in our research is provided. Following which is a description for the different computational methods employed. Lastly, an outline of the remaining thesis is provided.

1.1 Gating by the nuclear pore complex

Most of the genetic material and the required protein transcription machinery of the eukaryotic cell is stored and organized inside its membrane-bound nucleus. The double membrane of the nucleus, the nuclear envelop (NE), is highly impermeable to molecules and spatially separates the cytoplasm from the nucleus. While the NE protects the cell's genome and transcription machinery inside the nucleus, the nucleus also needs to communicate with the rest of the cell, importing proteins from the cytoplasm across the NE and exporting products, such as RNAs. As the only pathway connecting the cellular cytoplasm with the nucleoplasm, nuclear pore complexes (NPCs), forming large pores through the NE, enable and control the required import and export. Small molecules up to a limit of 20-40 KDa can diffuse freely through NPCs [1], while transport of large molecules (>40 KDa) are excluded unless facilitated and carried by transport factors (TFs), proteins that selectively pass through NPCs despite their large size. Thus, as the only conduit for nucleocytoplasmic transport, the selective passage through the NPC is essential for cell growth and function; however, the mechanism underlying the selection is still unknown.

Proteins building up the NPC are generally called nucleoporins/nups. NPC nups can be broadly partitioned into three structural and functional groups [2], i.e., transmembrane, central scaffold and transport. The transport group mainly contains FG-nups [2], proteins including intrinsically unstructured FG-repeat domains rich in phenylalanine and glycine (FG) repeating sequences (FG-repeats) [3–6]. Anchored on the central pore surface by their coiled-coil domains [2], FG-nups effectively fill the NPC central pore [3, 7, 8] and directly helps to mediate the NPC transport [9–16]. The interaction between FG-nups and the transport receptors is key to the selective transport of NPC. Both experiment and simulation have demonstrated that transport receptors bind to FG-nups through special binding spots [10, 12–14, 17–20].

Two questions appear to be central to this subject: 1) what is the structure adopted by the unstructured nups in the interior of NPC central channel; and 2) how does this structure play a role in gating of molecules in and out of the nucleus wherein, only small molecules, < 9m are allowed to passively diffuse through the channel and transport of large molecules is selectively carried out via TFs interacting that bind to nups in the channel through FG-motifs. This is addressed in chapter 2.

1.2 Filter function of the MscS' cytoplasmic domain

MscS, acts as a safety valve protecting a bacterial cell from bursting under osmotic pressure [21–24]. Such pressure can arise when the liquid around a bacterium changes suddenly, for example when salt water is replaced by pure water. The pressure in the cell is released when tension in the cell wall widens the channel such that osmolytes, positively and negatively charged ions, along with water, can leave the cell reducing the pressure [22, 25–27].

Since the inside of a bacterium, like all living cells, has a negative potential relative to its outside, mainly negatively charged osmolytes would leave the cell through the channel, keeping the cell from bursting, but abolishing the electrical cell potential and draining the cell of a crucial energy source. So the cell would not die from rupture, but from starvation. The protein channel prevents this, but the mechanism is unclear.

The transmembrane channel on the cytoplasmic side, extends to form a large balloon-like, cytoplasmic domain with openings to the cytoplasm. The CD is apparently a molecular sieve covering the transmembrane channel, that optimizes loss of osmolytes during osmoadaptation. In chapter 3, we address the role of the cytoplasmic domain and its filtering mechanism.

To find out how these channels function, a super-microscope that shows every atom of the channel and its action is required. Experiments such as X-ray crystallography, cryoelectron microscopy, and NMR spectroscopy, provide a very good snapshots of proteins, for instance, a single static picture of the channel protein, but they do not take videos. However, molecules in reality are, highly dynamic, and their motions are critical to understand their function. Thus, what is needed is a super-high resolution video. Apparently, such videos can actually be taken today through computer simulations. The simulations take into account all what is known from experiment and simulate the channels in action.

1.3 Molecular dynamics

The present work focuses on molecular dynamics (MD) simulations of biomolecules. In such simulations, biomolecules and their surroundings– for example, a protein and water surrounding it–are represented with a set of point masses (one per atom or, in certain cases, multiple atoms), with the interactions between those masses governed by an effective potential energy function (or force field) which has been parametrized based on quantum mechanical calculations and experimental data. The dynamics of the system can then be simulated by numerically integrating Newton's equations of motion over discrete timesteps, generally on the order of 1-2 fs (the timestep must be significantly shorter than the highest frequency oscillation in the system to avoid numerical instability [28]).

In a simplified level, molecular dynamics is a method in which one iteratively solves Newton's second law for N particles in a potential $U(\vec{r_1}, \vec{r_2}, ..., \vec{r_N})$ [28], i.e.,

$$m_i \ddot{r}_i = -\frac{\partial}{\partial r_i} U(\vec{r_1}, \vec{r_2}, ..., \vec{r_N}) \qquad i = 1..N. \tag{1.1}$$

1.3.1 The potential energy function

The potential function commonly used for biomolecular simulations express the total force on an atom as a sum of three components: *(a)* bonded force, which involves interactions between small group of atoms connected by one or more covalent bonds; *(b)* van der Waals forces, which involves interactions between all pairs of atoms in the system but which fall off quickly with distance and are generally evaluated only for nearby pairs of atoms; and *(c)* electrostatic forces, which involves

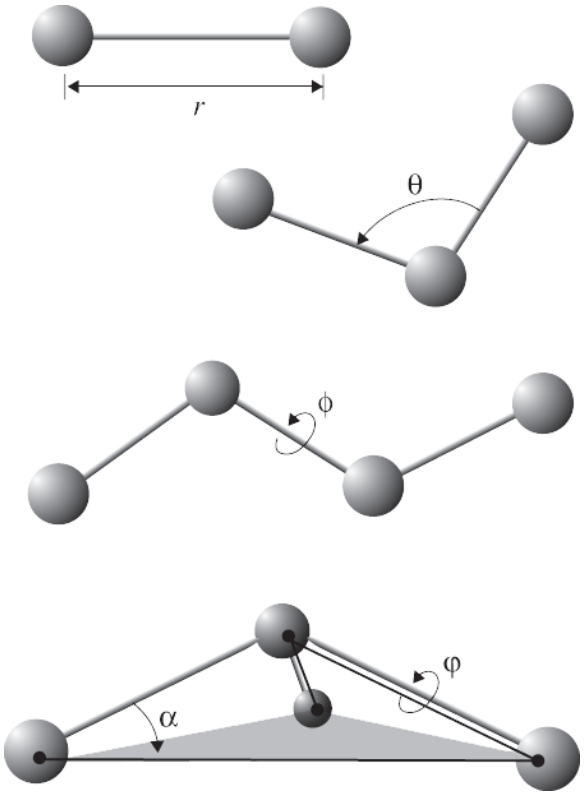

Figure 1.1: Contributions to the bonded potential function. Internal coordinates for bonded interactions: r governs bond stretching; θ represents the bond angle term; ϕ gives the dihedral angle; the small out-of-plane angle α is governed by the so-called "improper" dihedral angle φ. Reproduced with permission from [28].

interactions between all pairs of atoms and fall off slowly with distance. Electrostatic interactions are computed explicitly between nearby pairs of atoms, whereas long- range electrostatic interactions are typically handles via one of several approximate methods that are more efficient than explicitly computing interactions between all distant pairs of atoms.

There are four terms representing the covalently bonded nature of biological molecules, defined for sets of two, three, and four contiguous linked atoms, illustrated in Fig. 1.1. The first of these four terms is for the bonds themselves, given as a harmonic approximation for each pair of bonded atoms:

$$U_{bond} = k^{bond}(r - r_0)^2. \tag{1.2}$$

The force constant, k^{bond}, and equilibrium distance, r_0, are taken from a variety of experimental and computational sources, including IR spectra, high resolution crystal structures, and ab initio quantum chemical calculations [29]. The next term in the potential is for angles between three atoms, defined as

$$U_{angle} = k^{angle}(\theta - \theta_0)^2. \tag{1.3}$$

The last two terms, dihedrals and impropers, are both defined for four linked atoms. Dihedral terms represent rotation around a covalent bond, defined by

$$U_{dihe} = k^{dihe}(1 + \cos(n\phi + \delta)) \tag{1.4}$$

where n represents the symmetry of the rotation and δ is the phase shift. Finally, improper terms are a further necessary addition used to maintain the planarity of three atoms centered on a fourth and take the same form as that for dihedrals, given in Eq. 1.4.

The remaining two terms of the potential energy function U_{MD} govern non-bonded interactions between atoms, and are applied to all pairs of atoms in the system, subject to certain exclusions. The first term represents van der Waals interactions, i.e., those that govern the deviation of the system from an ideal gas of non-interacting, point-like particles. In MD, this term takes the form of the Lennard-Jones potential,

$$U_{vdW} = 4\epsilon_{ij}(\frac{\sigma_{ij}}{r_{ij}^{12}} - \frac{\sigma_{ij}}{r_{ij}^{6}}) \tag{1.5}$$

where ϵ_{ij} and σ_{ij} are parameters based on experiment. The r^{12} term in Eq. 1.5 approximates the exponential hard-sphere repulsion between atoms coming very close to each other, while the r^6 term represents the attractive London dispersion forces (i.e., induced dipole/induced dipole) between atoms. Finally, the remaining term in the potential U_{MD} arises from the coulombic electrostatic interactions between charged atoms:

$$U_{coulomb} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{1.6}$$

where q_i and q_j are the atomic partial charges. The total potential U_{MD} is thus

$$U_{MD}(\vec{r_1}, \vec{r_2}, ..., \vec{r_N}) = \sum_{bonds} U_{bond} + \sum_{angles} U_{angle} + \sum_{dihe} U_{dihe} + \sum_{impr} U_{impr} + \sum_{i \neq j} U_{vdW} + U_{coulomb}. \tag{1.7}$$

MD simulations are computationally expensive. Force calculation, Eq. 1.1, for each step requires substantial computation. For instance, one billion arithmetic operations for a system of 100, 000

atoms. Also, the force calculations must be repeated many times. Individual steps are limited to a few femtoseconds by fast atomic vibrations, so simulating a millisecond of physical time requires one trillion time steps. Using a single high-end processor core for such a simulation would take thousands of years to complete!

1.4 Accessing longer timescales

Classical MD suffers a major challenge, in its applicability to biological systems, which is the accessible timescales to which the systems can be simulated. Molecular dynamics simulations are routinely carried out on timescales of 10-100 ns. In contrast, the timescales required for functionally relevant motions of biomolecules range from nanoseconds for local relaxation and hundreds of nanoseconds for relative motions of protein domains to microseconds, milliseconds, or longer for key processes such as protein folding, drug binding, membrane transport.

Longer timescales can be reached, for example, through improvements in hardware, software, and algorithms: enhanced sampling techniques which bias the simulation to pass energetic barriers (*e.g.*, Umberella sampling, Adaptive biasing force, metadynamics) [30–32], coarse graining methods for simulating simplified system representations (*e.g.*, Residue-based coarse-graining, Shape-based coarse- graining, implicit solvent) [33–37]), and generalized ensemble methods [38].

Chapter 2, makes use of residue based coarse-grained simulations (RBCG) to address the system of size 15 million CG beads (150 million in all-atoms) and the timescales which both exceed the normal limitations of MD. Chapter 3 makes use of sampling technique, adaptive biasing force (ABF), that provides access to longtime scale event–overcoming an energy barrier for solute transport–unattainable with conventional MD.

The two methods are briefly described below. The details of the specific simulation procedures used are omitted here and the reader is encouraged to consult the methods sections of the appropriate chapters.

1.4.1 Residue-based coarse-grained molecular dynamics

In residue-based coarse-grained (RBCG) model for biological systems comprising proteins and or lipids, several atoms are grouped together in a "virtual" bead that interacts through an effective potential. For example, each amino acid residue and 4 water molecules are represented by 2-5 beads and 1 bead respectively (Fig. 1.2). This replacement already yields a reduction in system size by

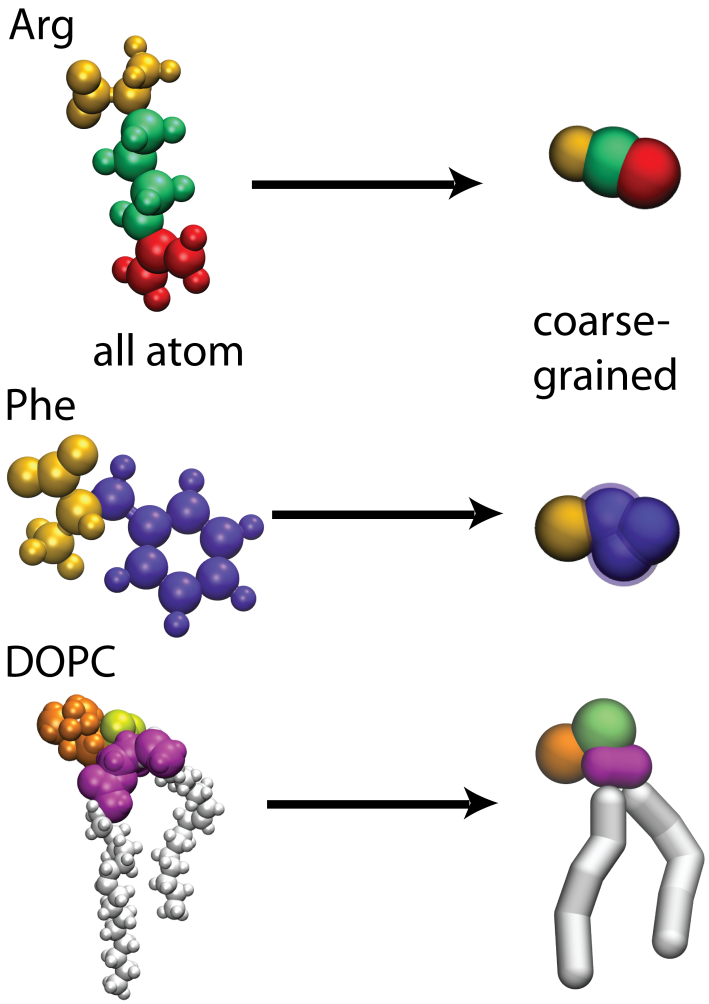

Figure 1.2: Amino acid and lipid structure in all-atom (left) and residue-based CG (right) representations.

a factor of ∼10. The coarser system also permits a longer simulation timestep (from 15 to 50 fs) and typically does not require long-range electrostatics calculations. The reduction of the number of degrees of freedom and the use of shorter-range potential functions makes the model computationally very efficient, allowing an increase in the base time-step and thus a reduction of the simulation time by 2 - 3 orders of magnitude compared to the traditional atomistic models [33–36].

The coarse-grained models employ a similar form of the total potential energy as the above all-atom force fields, yet, with different parameters. The all-atom force field parameters are generally obtained through a combination of empirical techniques and quantum mechanical calculations [39–42]; bulk water and experimentally well-characterized small molecules are then simulated using

these parameters to test whether their structural, dynamic, and thermodynamic properties can be correctly reproduced. On the other hand, the coarse-grained force field parameters are generated based on all-atom force field parameters, and are then tested for fidelity in reproducing the all-atom simulation results [35].

As noted above, the interactions between the beads are governed by a potential similar to that in Eq. 1.7, with a few differences. Bond and dihedral potentials are identical to those in all-atom MD, given in Eqs. 1.2 and 1.4, respectively, although for all dihedrals other than those for the protein backbone, k^{dihe} is set to 0 [35]. The angle potential is defined differently and takes the form

$$U_{angle} = k^{angle}(\cos(\theta) - \cos(\theta_0))^2 \tag{1.8}$$

where the cosine is used to "soften" the potential [33]. The nonbonded potential is identical to that given in Eqs. 1.5 and 1.6. Beads are divided in four types, polar, non-polar, apolar, and charged; the van der Waals interaction energy constant, ϵ_{ij}, is then assigned to be one of five levels based on the combination of types for beads i and j [33]. In Eq. 1.6, the dielectric constant in CG MD is $15\epsilon_0$, intead of ϵ_0 as in all-atom MD [35].

1.4.2 Adaptive biasing force

The adaptive biasing force (ABF) scheme is a molecular-dynamics based method for overcoming barriers of the free-energy landscape. Integration of the mean force measured along a chosen reaction coordinate (RC) yields the so-called potential of mean force (PMF). The RC is a coarse-grained description of the transition mechanism. The mean force is estimated by accruing and averaging the instantaneous force exerted on the system. The PMF is then used to bias the standard dynamics of the system in order to improve sampling in the RC [31, 43–45].

The connection between the derivative of the free energy with respect to the reaction coordinate, $\frac{dA}{d\xi}$, and the forces exerted along the latter may be written as

$$\frac{dA}{d\xi} = \left\langle \frac{\partial V(x)}{d\xi} - \frac{1}{\beta}\frac{\partial \ln |J|}{\partial \xi} \right\rangle_{\xi} = -\left\langle F_{\xi} \right\rangle_{\xi} \tag{1.9}$$

where $|J|$ is the determinant of the Jacobian for the transformation from generalized to Cartesian coordinates. The first term of the ensemble average corresponds to the Cartesian forces exerted on the system, derived from the potential energy function, $V(x)$. The second contribution is a geometric correction arising from the change in the metric of the phase space due to the use of generalized

coordinates. It is worth noting, that, contrary to its instantaneous component, F_ξ , only the average force, $\left\langle F_\xi \right\rangle_\xi$, is physically meaningful.

In the framework of the average biasing force (ABF) approach, F_ξ is accumulated in small windows or bins of finite size, $\delta\xi$, thereby providing an estimate of the derivative $\dfrac{dA}{d\xi}$ defined in equation 1.9. The force applied along the reaction coordinate, ξ , to overcome free energy barriers is defined by:

$$F^{ABF} = \nabla_x \widetilde{A} = -\left\langle F_\xi \right\rangle_\xi \nabla_x \xi \tag{1.10}$$

where $\widetilde{A}$ denotes the current estimate of the free energy and $\left\langle F_\xi \right\rangle_\xi$, the current average of F_ξ.

As sampling of the phase space proceeds, the estimate $\nabla_x \widetilde{A}$ is progressively refined. The biasing force, F^{ABF} , introduced in the equations of motion guarantees that in the bin centered about ξ, the force acting along the reaction coordinate averages to zero over time. Evolution of the system along ξ is, therefore, governed mainly by its self-diffusion properties.

A particular feature of the instantaneous force, F_ξ , is its tendency to fluctuate significantly. As a result, in the beginning of an ABF simulation, the accumulated average in each bin will generally take large, inaccurate values. Under these circumstances, applying the biasing force along according to equation 1.10 may severely perturb the dynamics of the system, thereby biasing artificially the accrued average, and, thus, impede convergence. To avoid such undesirable effects, no biasing force is applied in a bin centered about ξ until a reasonable number of force samples has been collected. When the user-defined minimum number of samples is reached, the biasing force is introduced progressively in the form of a linear ramp. For optimal efficiency, this minimal number of samples should be adjusted on a system-dependent basis.

In addition, to alleviate the deleterious effects caused by abrupt variations of the force, the corresponding fluctuations are smoothed out, using a weighted running average over a preset number of adjacent bins, in lieu of the average of the current bin itself. It is, however, crucial to ascertain that the free energy profile varies regularly in the ξ-interval, over which the average is performed.

To obtain an adequate sampling in reasonable simulation times, it is recommended to split long reaction pathways into consecutive ranges of ξ. In contrast with probability-based methods, ABF does not require that these windows overlap by virtue of the continuity of the force across the the reaction pathway.

1.5 Thesis outline

In order to shed light on the gating mechanism of NPCs and on the filtering mechanism of MscS, one has to go beyond static pictures as supplied, for example, through experimental structure analysis. MD simulations offer an opportunity here as they predict dynamic behavior of macromolecular systems. Unfortunately, these methods can describe only short time scales, but forfeiting atomic resolution long time scales can be addressed by means of coarse graining. If the issue is conformational sampling, rather than dynamics, enhanced sampling methodologies permit one to account for a wide range of conformations otherwise visited only in impossibly long simulations. In this thesis the mentioned approaches are applied to access long timescale events revealing motions in NPC and MscS involved in selectivity and transportation processes. In chapter 2, we examine the barrier structures formed by the unstructured proteins that protrude into the interior of the NPC central channel, concluding with a model for gating. In chapter 3, we investigate the role of the large cytoplasmic domain of MscS, to suggest a filter function. Chapter 4 summarizes the impact of the results and presents directions for future research. Chapters 2 and 3 in this volume are based on previously published or submitted work.

Chapter 2

Assembly of Nsp1 nucleoporins provides insights into nuclear pore complex gating

2.1 Introduction

Nuclear pore complexes (NPCs) are large protein assemblies embedded in the nuclear membrane that provide the only conduit for exchange of molecules between cytoplasm and nucleoplasm, a process defined as nucleocytoplasmic transport (NCT). While the membrane bound nucleus of a eukaryotic cell protectively envelopes the genetic material, separating it from the cytoplasm, processes such as DNA transcription require access to the genetic material in the interior of the nucleus by a myriad of molecules, including large molecules such as proteins . Two modes of transport exist for the NCT: passive diffusion of molecules, smaller than 40 kDa or 9-10 nm in diameter [1], and selective transport into and out of the nucleus for larger molecules such as proteins, RNA, ribosomal subunits, the latter being recognized by transport factors (TFs), carrier proteins, which shuttle back and forth ferrying their cargo through the NPC.

NPCs are giant complexes having molecular masses ranging from ∼65 MDa (in yeast) to ∼125 MDa (in higher eukaryotes) that are made of copies of 30 different proteins termed nucleoporins (nups) [3, 46–48]. Recently, a detailed 3D model for the position and abundance of each nup in the Saccharomyces cerevisiae NPC structure was proposed based on experimental data obtained from molecular, biochemical and structural studies of the NPCs and their components [7, 8]. In the modeled structure, the scaffold of the NPC is formed by two protein subcomplexes that, through linker proteins, anchor a set of FG-containing nups [8]. The complex has 8-fold symmetry about its central axis and 2-fold symmetry about the equatorial plane such that each nup is repeated 8-, 16-, 32-, or 48-fold. Altogether, the NPC forms a multiprotein complex of nearly ∼450 proteins [3, 7, 8, 48].

About one-third of all pore proteins constitute the barrier proteins also known as FG-nups (F and G represent amino-acids phenylalanine and glycine) . FG-nups are intrinsically disordered, rich in FG-repeat motifs [3, 4, 6, 48, 49] and form the central transport channel of the NPC, extending however also into the cytoplasmic and nucleoplasmic space. While the FG-nups extending toward

cytoplasm and nucleoplasm sides are assymetric in distribution, the FG-nups (eg., Nsp1) of the central region have copies distributed symmetrically around the scaffold of the NPC and are critical for bidirectional NCT. The FG-nups are present in different lengths (a few to several hundred amino-acids long) and exhibit different amino acid composition and repeat motifs; their different spatial localizations render the central channel heterogeneous. These nups play a central role in selective transport of the NPC and this selectivity is attained by bearing specific binding sites for the TFs that undergo multiple, low- affinity interactions with nups as they permeate their way through the channel [10, 12–14, 17–19].

Given that the central transport channel is filled with heterogenous, natively unstructured FG-nups that are not susceptible to crystallization for structure determination, the important question how the unstructured FG nups are assembled in the interior to promote efficient, selective gating could not be answered unequivocally.

Several models have been proposed so far to explain the structure of the interior of the NPC. According to the virtual gate model [3, 50], the tethered, unstructured FG-repeat proteins form an entropic barrier that repels non-specific cargo, preventing it from reaching the interior of the channel, whereas TFs that interact with the FG-repeats of nups in the NPC interior channel have higher probability of entering and permeating the channel.

An alternative model, the polymer brush model [5, 51], is based on the above model suggesting that the FG-nups form extended brush-like polymers with inter-linked bristles that reversibly collapse upon binding of TFs. TFs pass the channel by repetitive binding and unbinding to the nups' FG repeats until they reach the exit of the pore [52].

In contrast, the selective phase/hydrogel model [53] proposes that FG nups in the central channel form a sieve-like meshwork through weak inter-repeat FG-FG hydrophobic interactions to form a hydrogel within the central channel [54]. This meshwork prevents non-specific TFs from binding to the FG repeats via low-affinity interactions which could transiently melt these FG-FG crosslinks, allowing thus only TFs to permeate the channel. Further, non-specific cargo that cannot bind would not permeate and thus would be filtered out [54, 55].

Another model, the reduction of dimensionality (ROD) model, proposes that the TFs act as ferries, carrying cargo and slide on the surface of phenylalanine glycine (FG) motifs by interacting with the FGs according to a 2D random walk rather than a 3D diffusion [56].

The 'forest' model, later modified to the tube gate model [57, 58], proposes the interior of the NPC to be a hydrogel and the periphery to be brush-like, featuring two separate zones of traffic

with distinct physicochemical properties. In their study, based on the hydrodynamical radius of individual nups localized in different regions of the NPC, the respective authors classify nups as (a) 'shrubs' that form collapsed coil, low-charge content, cohesive domains; (b)'extended coil' that forms high-charge content, non-cohesive domains; (c) 'trees' having features of both (a) and (b). Thus, a 'forest-like' structure of FG-nups is formed in the npc. In this model, the collapsed-coil domains are hypothesized to form a transport zone 1 in the central pore and the extended-coil domains form a peripheral zone 2. The forest model is based on the observations of individual nup conformations, not considering interactions between nups that drives assemblies. Therefore, the model cannot describe the collective behavior of the nup assembly in an environment as crowded as found in the interior of the channel.

In a recent study on the role of nups in NPC gating, Tagliazucchi et. al [59] deduced a potential of mean force from the amino acid (positive, negative, hydrophobic) distribution and claimed that this potential can explain the use of TFs for traffic across the nuclear envelope. The claim is based, however, on the assumption that nups are individually random and do not form any structures in the channel interior as suggested by the earlier models of the NPC interior. The authors also do not attribute an explicit role to TFs interacting with nups through FG-motifs.

Most models propose an FG network as the main mechanism for pore selectivity and address to a much smaller degree a role for the actual molecular structures formed by the disordered nups of the NPC. Likely, the qualitative differences of the models can be reconciled with the heterogeneity found along entire transport pathways through the NPC that furnishes different environments on the cytoplasmic side, in the interior and on the nucleoplasm side of the NPC.

To identify the structure formed by the nups in the interior of the NPC pore, Miao and Schulten [51] carried out the first MD simulation of nup polymers, using a coarse-grained model. Unlike the other models, their model included details of individual amino acids and, more importantly, included the interactions between nup polymer chains, end-grafted to a flat surface in a 5×5 array. The MD simulations suggest that the nups have a strong tendency to form bundles of typically 2 - 6 proteins and that the bundles are inter-linked in a mesh-like manner as single nups cross frequently from one bundle to another bundle. From this behavior results a brush structure that is made of bundle-like, interlinked bristles. This structure can explain to some degree the gating of NPC pores: the structure can be passed by small biomolecules readily while larger biomolecules need to alter the structure to pass through. This role, namely of cutting through the mesh-like structure, could be played by transport factors.

While the structure model of Miao and Schulten is highly suggestive, the study had obvious limitations imposed at the time by limitations in technologically available computer power. One limitation is the short length (100-aa) of nups employed that is in contrast to the actual, much longer (about 609-aa) lengths of many nups. A second limitation is that only a single type of nup tethering was investigated, namely tethering to a flat surface. A third limitation is that the simulated ensemble started with all nups initially in a fully-extended, straight chain conformation.

Enjoying today tremendously more computer power than was available to Miao and Schulten we have performed for the present study microsecond-long simulations on various conformational ensembles of 609-aa-long FG domains of Nsp1, a key nup in the NPC central channel. Three different ensembles were investigated, namely of untethered Nsp1s, Nsp1s tethered to a flat surface as in the Miao and Schulten study, and Nsp1s tethered to a ring-like surface that is qualitatively similar to the geometry of the NPC. Most essential, however, is that the present study explores Nsp1 ensembles in different initial protein conformations, namely fully-extended conformations as assumed by Miao and Schulten as well as random polymer-like conformations. Nevertheless, the new MD simulations support the structural model as seen in the simulations of Miao and Schulten, in which Nsp1 form bundles linked through single proteins crossing between bundles. An important new finding, however, is that the structure formed as described depends little on FG side groups as qualitatively the same structure comes out when FG side groups are replaced by alanines. The new, more extensive simulations demonstrate also that the actual cross-linked bundle system depends to a significant degree on the actual tethering of nups' ends as well as on the initial nups conformation.

2.2 Results

To elucidate how the NPC poses a barrier for transport of molecules into and out of the nucleus, we characterized through simulations the general structural features in the NPC central channel, arising from the assembly of disordered proteins. We investigated the assembly using only one type of FG-nups, namely, Nsp1. Nsp1 is a FXFG-rich nup present in the central channel and has been studied as a model protein in previous studies [51, 55, 57]. The nup is present in 32 copies in each of the upper and lower half of the NPC equatorial plane; its spatical location is $\sim$ 30 nm radius from the center of the NPC [60]. The large number of copies and its spatial location inside the central channel renders Nsp1 critical for the bi-directional NCT. The 609-aa-long FG-domain of this protein was described through coarse-grained simulations that investigated the structures formed by three

Name	Simulated System	Time (ns)	CG-beads	System Size (Å^3)
wild-type_ring	120 fully extended, wild-type nsp1s grafted on a gold nano-pore	1000	15,453,214	1244 Å× 1244 Å× 1627 Å
mutant_ring	120 fully extended, mutant (FG→AA) nsp1s grafted on a gold nano-pore	1000	16,019,434	1244 Å× 1244 Å× 1627 Å
random_array	5 × 5 array of worm-like chain nsp1s grafted on a gold substrate	1000	1,097,433	290 Å× 290 Å× 1075 Å
random_bath	120 worm-like chain nsp1s, freely floating in a water bath	1000	3,091,910	755 Å× 755 Å× 755 Å

Table 2.1: Systems simulated. In the left column, conventional names of the simulations, as used throughout the paper, are defined. In column "CG Beads", the total number of coarse-grained beads simulated is given, adding up beads describing protein, gold substrate and solvent (water and ions).

different ensembles of Nsp1s. In the first two simulations, the proteins were tethered to a gold ring surface (simulations `wild-type_ring` and `mutant_ring`). For the initial protein conformations we assumed fully-extended Nsp1s as was assumed in the earlier study by Miao et al [51]. However, in the present study, we use the full-length (609-aa-long) FG-domain. The two simulations differ in that the first employs wild-type Nsp1 whereas the second employs a mutant Nsp1. In the third simulation, Nsp1s were tethered to a flat gold surface (simulation `random_array`). In this case the proteins were assumed initially in a random polymer conformation. In a fourth simulation the effect of end-tethering was investigated by simulating Nsp1s untethered and initially homogeneously distributed in bulk solvent in random polymer conformations (simulation `random_bath`). All ensembles contain the protein at a concentration characteristic for the interior of the NPC channel. The simulations carried out are listed in Table 2.1.

For simulation `wild-type_ring`, the full length FG-domain of Nsp1 (1-609 amino acids) were grafted to a gold nano-ring that matches the pore dimension of the pore system investigated by an experimental group [52, 60] engineered to mimic the NPC geometry. The system was simulated for 1 μs after which time the initially fully streched Nsp1s formed brush-like bundle structures (see Fig. 2.1), the same structure as reported in an earlier study of 100-aa-long nups [51]. Such structures can clearly function as a selective barrier to transport, supporting the virtual-gate model of nuclear pore gating [1, 51]. Key for the gating mechanism is that the bundles are interconnected via single Nsp1 chains cross-linking adjacent bundles (see Fig. 2.1).

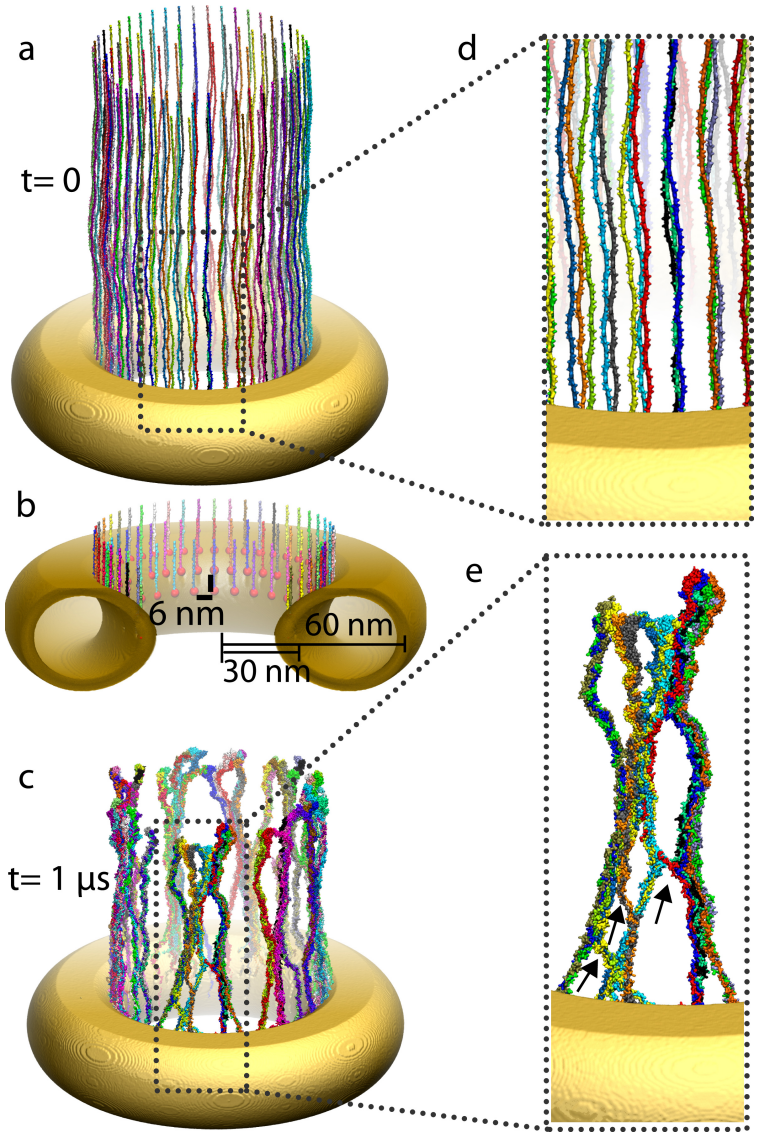

Figure 2.1: Initial and final configuration of simulated wild-type Nsp1s grafted to a gold ring (simulation `wild-type_ring`). *(a)* Initial configuration. Shown are fully extended, wild-type Nsp1s grafted on the ring, the geometry of which matches that of an experimentally constructed nanodevice mimicking an NPC as reported in [52, 60]. Colors distinguish 120 wild-type Nsp1s grafted on the ring in three concentric rows. *(b)* Close-up view of grafted ends of the Nsp1 chains. The gold nano-ring is cut open to expose the C-termini, shown as red spheres, fixed to the gold ring, as well as the terminal parts of the Nsp1 chains. *(c)* Snapshot of the (1 μs) end of simulation `wild-type_ring`. One can recognize that the Nsp1 chains, shown in surface representation, have formed brush-like bundles. *(d)* Close-up view of the structure in *(a)*. Shown is a region as marked. The close-up view reveals the initially straight conformation of the Nsp1 chains; bumps in the surface of the individual chains correspond to amino acid side groups. *(e)* Close-up view of a segment of *(c)*. The view reveals the brush-like bundles formed by the Nsp1 chains. Arrows point to cross-links between bundles formed when single Nsp1 chains cross from one bundle to another bundle. As a result of such cross links the bundles form a mesh of thick (bundles made of several Nsp1 chains) and thin (cross links made of single Nsp1 chains) segments.

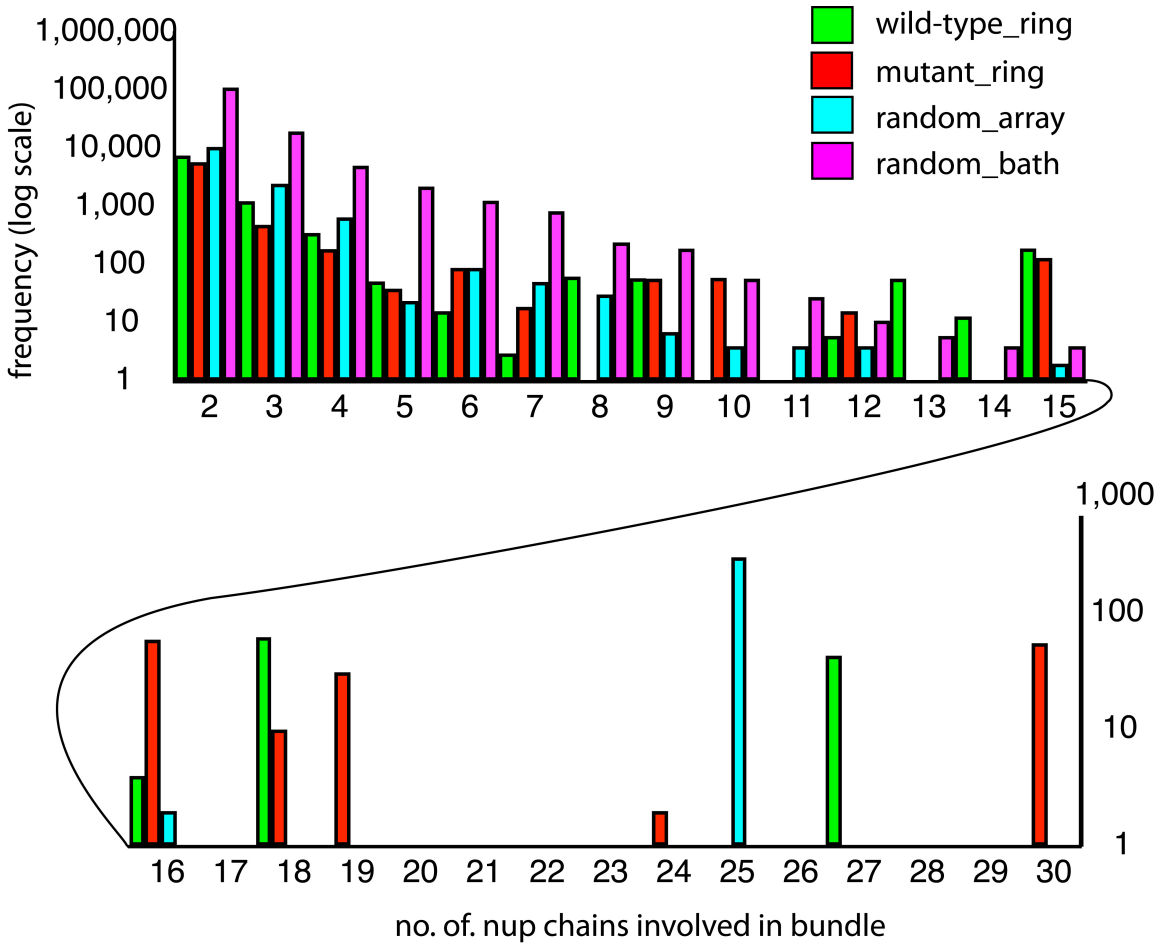

Figure 2.2: Bundle thickness distribution. Bundle thickness is determined by the number of Nsp1 chains involved in a bundle. Shown is here the distribution of these numbers for the simulations carried out. The frequencies with which chain numbers arise were averaged for the last 30 ns of the four 1 μs simulations `wild-type_ring`, `mutant_ring`, `random_array`, and `random_bath`. Bundles with fewer than ten Nsp1 chains favor a mesh-like structure, namely bundles with frequent cross-links, whereas bundles with more than ten Nsp1 chains exhibit brush-like structures with relatively few cross-links. Green represents the frequency distribution of Nsp1s for simulation `wild-type_ring`, red for `mutant_ring`, cyan for `random_array`, and purple for `random_bath`.

We analyzed these structures through computation of the bundle-thickness distribution. For the brush-like structure seen in Fig. 2.1 the distribution has a peak at more than fifteen different chains (Fig. 2.2). Thus, the brush-like structure is characterized by thick and strong bundles.

To test which role FG-repeats play in the formation of the structure in Fig. 2.1 we simulated a mutant system replacing all FGs in Nsp1 by alanines (simulation `mutant_ring`). The assembly structure of mutant Nsp1s in the nanopore model resulting from the 1 μs simulations is very similar to that seen for wild-type (WT) Nsp1s. The similarity is also reflected in the bundle distribution that shows for the structural assembly of mutant Nsp1s thick bundles.

The initially fully- extended, straight chain Nsp1s, simulated in [51] and in the present simulations, `wild-type_ring` and `mutant_ring`, are highly ordered at the outset and may lead to an unrealistic quasi-equilibrium, i.e., trapped, state calling the assembly structure shown in Fig. 2.1 into question. Therefore, we modeled in simulation `random_array` end-tethered Nsp1s with a grid spacing of 6 nm, similar to that in the gold ring geometry, but assuming for the initial state random, rather than straight Nsp1 conformations. The simulation involved also full length (609-aa) Nsp1. From the present simulation resulted, nevertheless a brush-like structure of bundles with the bundle distribution shown in Fig. 2.2. However, the bundles seen exhibit more cross-links than seen in the earlier simulation [51], (see Fig. 2.3).

The unstructures nups in the central channel are fixated at their base by other proteins that link with scaffold nups in the nuclear envelop. Brush-like structures, a feature distinctive to end-tethered and initially conformation ally random Nsp1s, are likely representative for the multi-protein structures formed inside the NPC. The structures exhibited by a single untethered Nsp1 protein monitored in a 1.1 μs CG MD simulation showed very different features from the structures resulting from the ensembles of closely spaced Nsp1 proteins in simulations `wild-type_ring`, `mutant_ring` and `random_array`. An initially fully extended, untethered single Nsp1, is found to coil up into a globule-like structure, with its radius of gyration (R_g) decreasing from over 300 Å (corresponding to the initially fully extended form) to 65 Å (corresponding to the final, coiled form) as shown in Fig. 2.4. Similar globule-like structures were also observed in earlier studies of dynamics of individual, end-tethered short fragments of Nsp1s [51].

The brush-like structure resulting for ensembles of closely spaced Nsp1s can similarly be characterized by brush height. For the end-tethered Nsp1s in simulations `wild-type_ring`, `mutant_ring` and `random_array` we monitored the average brush height, namely the end-to-end distance of the main-chain C_α beads of the Nsp1 chains in the z-direction. The average was taken over all Nsp1 chains

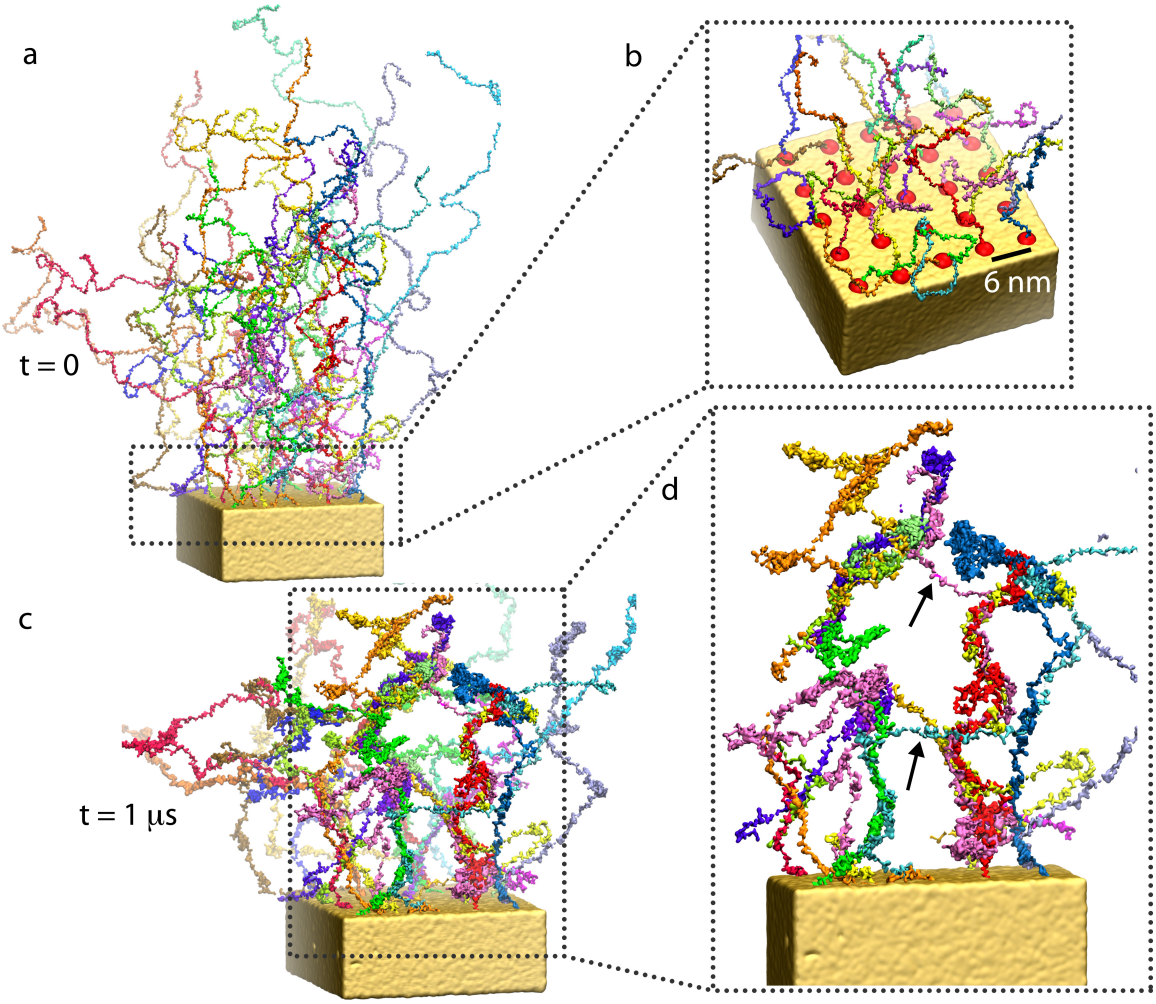

Figure 2.3: Initial and final configuration of a simulated array of wild-type Nsp1s grafted to a gold substrate (simulation `random_array`). *(a)* Initial configuration. Shown is the 5×5 array of the wild-type Nsp1s grafted with their C-terminal ends to a gold substrate. The proteins are placed initially in random polymer-like conformations obtained computationally through a description of non-overlapping worm-like chains. Colors distinguish the 25 grafted wild-type Nsp1s. *(b)* Close-up view of grafted ends of the Nsp1 chains in an array. Shown as red spheres are C-terminal ends of the Nsp1 chains fixed to the gold substrate, as well as the terminal segments of the Nsp1 chains. *(c)* Snapshot of the (1 μs) end of simulation `random_array`. One can see that the end-tethered, randomly placed (matching a worm-like chain model) chain Nsp1s, shown in surface representations, form brush-like structures as in case of simulation `wild-type_ring`, but with a higher density of cross-links compared to the gold ring case shown in Fig. 2.1. *(d)* Close-up view of a segment of *(c)*. The view reveals cross-linked Nsp1 bundles. Arrows point to cross-links between bundles formed when Nsp1 chains cross from one bundle to another bundle.

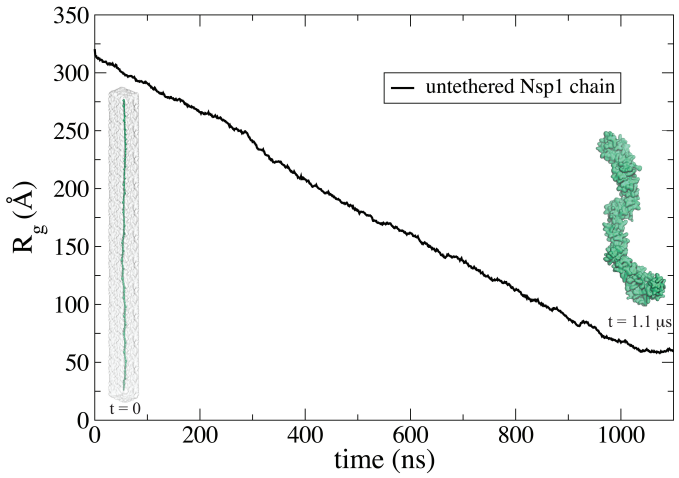

Figure 2.4: Coiling of an initially fully extended, Nsp1 segment. The time evolution of the radius of gyration (R_g) is shown for 1.1 μs CG simulation.

in the given simulation system. For the Nsp1s in simulations `wild-type_ring` and `mutant_ring`, the average height exhibited in both cases a very similar time dependence, namely a decay from an initial height of 115 nm to a height of 80 nm (70% of initial height) in 1 μs. The Nsp1 chains in `random_array` exhibited an average height that decayed to 65% of the initial height (60 nm) namely by 37.5 nm in 1 μs.

In order to investigate how the final structure of closely spaced Nsp1s is affected by end-tethering, we carried out simulation of initially fully disordered Nsp1s freely floating in a bath (`random_bath`) at a protein density similar to that in `random_array`. These Nsp1s formed again bundles, but in this case with many more links between bundles (formed by single Nsp1s crossing between bundles) than seen in case of simulations `wild-type_ring`, `mutant_ring` and `random_array` (see Fig. 2.6). This mesh-like structure can be clearly distinguished from the brush-like structures of `wild-type_ring`, `mutant_ring` and `random_array` through the bundle thickness distribution shown in (Fig. 2.2) as the bundles arising in simulation `random_bath` exhibit fewer than ten chains per bundle.

Given the different structures formed by Nsp1 assemblies as they result from the present simulations and represent likely the interior of the NPC, one wonders if any specific interactions, in particular, hydrophobic FG-FG interactions, favor the structures seen. We determined, therefore, for the different types of amino acids involved in the formation of bundles how often particular amino acids arise in bundles. For simulation `wild-type_ring` FGs clearly are not critical for the formation

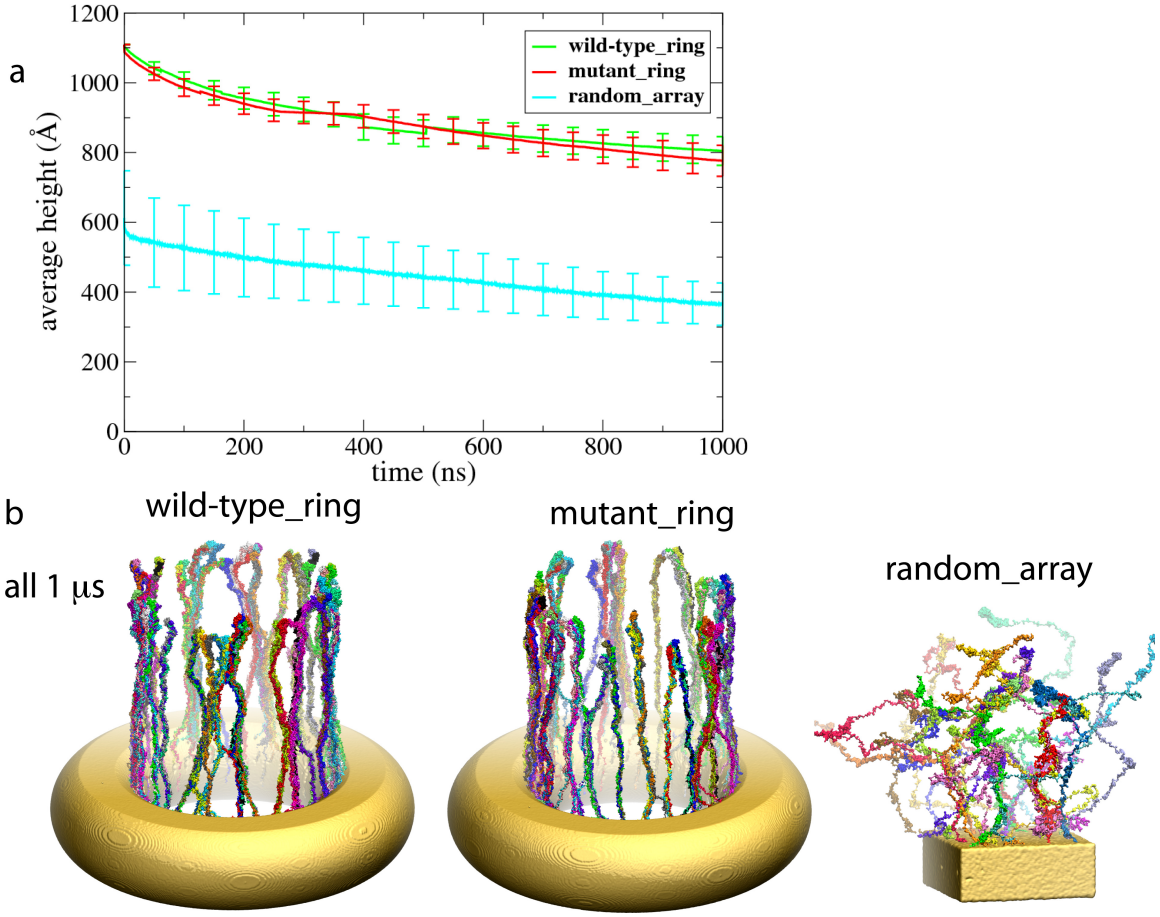

Figure 2.5: Height of Nsp1 chains. *(a)* Time evolution of the average height. The height is shown for 1 μs simulations for the end-tethered Nsp1s in simulations `wild-type_ring` (green line), `mutant_ring` (red line), and `random_array` (cyan line). The heights are calculated as the average end-to-end distance in the z-direction, the average being taken over all chains in a given simulation system. *(b)* Snapshot of the (1 μs) end of simulation `wild-type_ring` (see also Fig. 2.1*(c)*), `mutant_ring`, and `random_array` (see also Fig. 2.3*(c)*). The mutated Nsp1 chains (FG-to- AA) arising in simulation `mutant_ring` form brush-like bundles with similar average brush height as seen to arise for wild-type Nsp1s in simulation `wild-type_ring`.

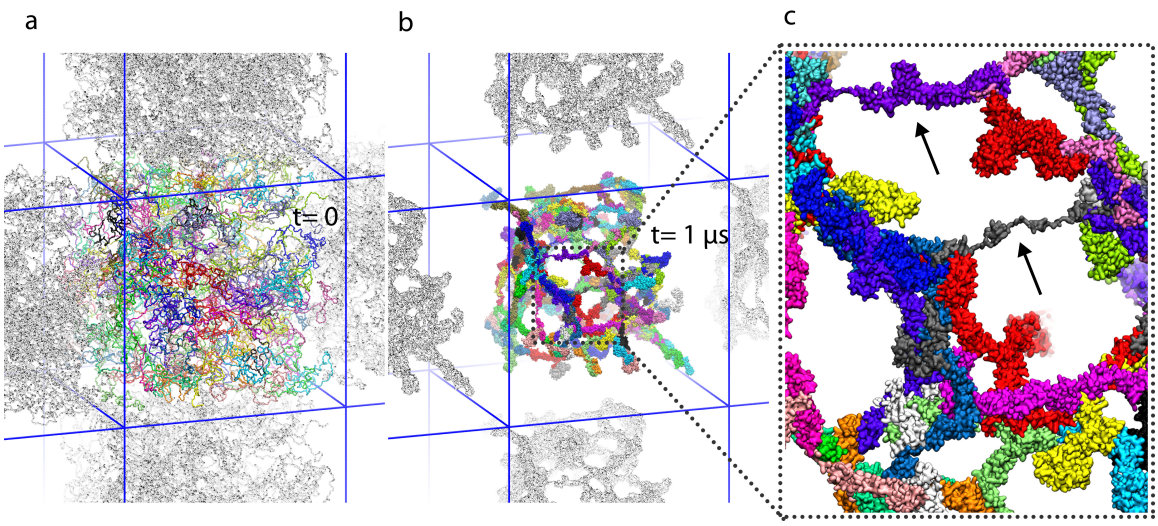

Figure 2.6: Initial and final configuration of simulated wild-type Nsp1s in a solvent bath (simulation `random_bath`). *(a)* Initial configuration. Shown is the periodic (see simulation conditions as described in Methods) system of wild-type Nsp1s, freely floating in a solvent bath (water and ions). The initial random conformations match a polymer melt modeled from worm-like chains. Colors distinguish 120 freely floating Nsp1 chains. Neighboring boxes in x- and y- directions are shown with the Nsp1s colored in grey. *(b)* Snapshot of the (1 μs) end of simulation `random_bath`. The Nsp1 chains, shown in surface representations, are seen to form a porous mesh of cross-linked Nsp1 bundles. *(c)* Close-up view of the structure in *(b)*. Shown is a region as marked. The view reveals a system of short bundles that are frequently cross-linked. Arrows point to the cross-links between bundles.

of these structures (Fig. 2.7). This is also supported by simulation `mutant_ring`, in which the FG motif was replaced by alanines, and by an earlier study reported in [51]. Moreover, all amino-acids are equally favorable for the brush-like structures of the `wild-type_ring`. Actually, no structures arising from `wild-type_ring`, `mutant_ring`, `random_array` or `random_bath` exhibit particular favorable amino acid preference. Therefore, the observed bundle structures are not sequence-specific and likely come about through the fraction of non-polar amino acids in the Nsp1 sequence.

Name	Average pore size ± std.dev Å
wild-type_ring	77.32 ± 0.92
mutant_ring	77.35 ± 0.94
random_array	50.38 ± 1.40
random_bath	43.33 ± 1.33

Table 2.2: Average pore size. The pore size is defined through the radius of the largest spherical cargo capable of passing through the final structure of a simulated system of nsp1s. The radius was determine according to the algorithm presented in Methods and is illustrated in Fig. 7. The average was taken over the last 30 ns of the 1 μs simulation for each of the systems `wild-type_ring`, `mutant_ring`, `random_array`, and `random_bath`.

In order to determine in how far the structures of Nsp1 assemblies described above provide a

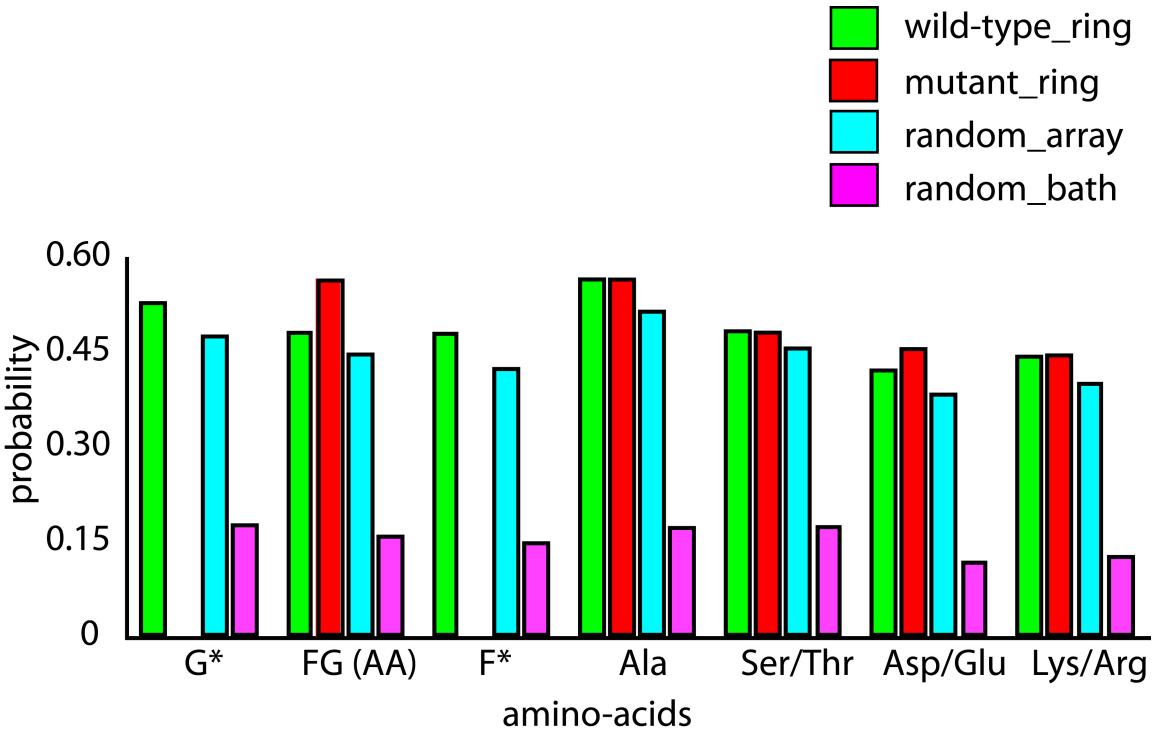

Figure 2.7: Propensity for certain amino-acids to be involved in the formation of bundles. The probability for different kinds of amino acids to be involved in the formation of bundles as determined from an average over the last 30 ns of the 1 μs simulations `wild-type_ring` (green), `mutant_ring` (red), `random_array` (cyan), and `random_bath` (purple). G* refers to glycines and F* refers to the phenylalanines that are not included in FGs of FG-repeat motifs.

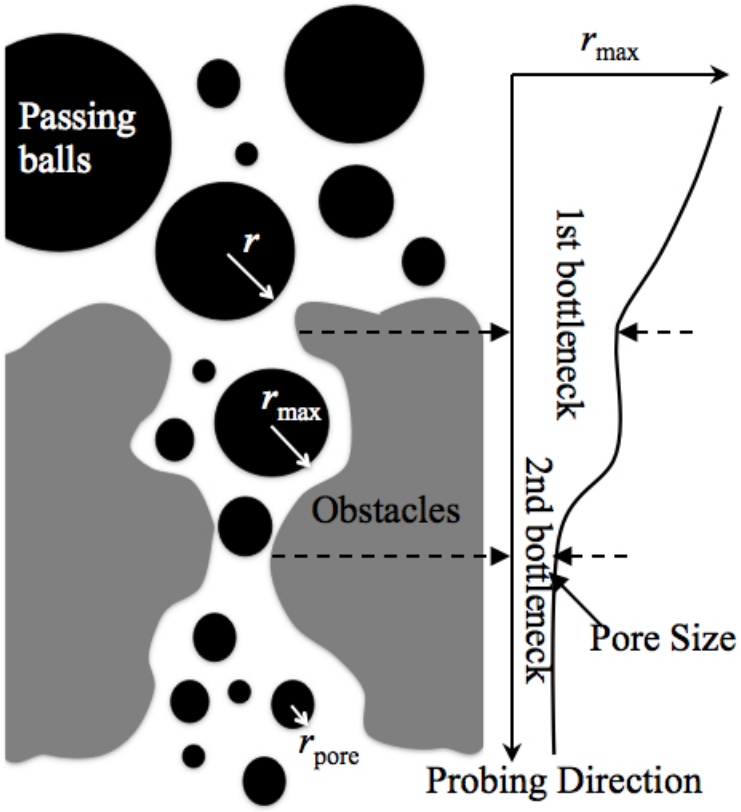

Figure 2.8: Schematic algorithm for calculating pore sizes. The figure shows a schematic depiction of the algorithm employed in calculating the pore sizes listed in Table 2. The pore size is defined through the radius of the largest spherical cargo capable of passing through the final 1 μs of simulation structure of a system of Nsp1s. A search for the largest cargo starts on one side of the system, the latter shown in grey. The cargo (black sphere) of a certain size moves towards the other side while the algorithm probes if the cargo can pass. The panel at right illustrates what is measured by the algorithm, namely the radius of the largest cargo that could pass through along a probing direction. As shown in this panel, there are two bottlenecks, the second one of which determines how large a ball can pass through the obstacles and, therefore, characterizes the pore size of obstacles.

barrier against the diffusion of molecules through them, we computed the size of passing molecules as described in Methods and Fig. 2.8. Table 2.2 lists for the structures resulting from simulations `wild-type_ring`, `mutant_ring`, `random_array`, and `random_bath`, the average radius for the cargo molecules that can diffuse through the structures. Large pores are available for diffusive passage when Nsp1s are assembled into bundle structures with few cross-links; in this case molecules with a radius as large as ∼77 Å can pass, whereas mesh-like bundle structures with many cross-links furnish only relatively small pores for molecules to pass through, namely only ones with radii smaller than 43-50 Å.

2.3 Discussion

The structure of the NPC central channel interior, made up of unstructured nups with FG repeats, governs transport through the NPC. The structure formed by the assembly of tethered nups determines the size of small molecules that can passively diffuse through. The structure also interacts, likely through the FG repeats, with the transport factors that carry various cargoes through the NPC.

To characterize the structure of the channel interior, we investigated the assembly using one type of FG-nups, namely, Nsp1, that is present in the central channel. We studied, using CG MD simulations, the dynamics of system of Nsp1s starting from three different initial states: tethered, straight-chain conformations; tethered, random chain conformations; and untethered, random chain conformations, simulating each system for 1 μs. We observe in all cases the formation of brush-like bundles linked through single Nsp1s crossing between bundles. This assembly structure is very similar to one seen in prior simulations of is initially fully extended, Nsp1s fragments (100-aa-long), that were closely grafted (2.6 nm apart compared to 6 nm-spacing in the present study) in an array like arrangement [51].

Our interpretation, based on the structural features observed for tethered, unstructured FG-nups of the transport channel agrees with the other views [5, 20, 50–52, 54–58] in that the nups form characteristic quasi-stable, i.e., slowly varying, structures bearing FG spots, with which TFs are known to interact. Thus, interaction between TF surfaces and nups is structured, and cannot solely be described properly through a potential of mean force as suggested by Tagliazucchi et. al [59]. Experimental studies using high-resolution single-molecule studies that deduced spatial density plots for TF-FG-repeat interactions have also shown that the interaction sites are not evenly distributed

in the NPC; instead, they form spatial clusters inside and outside the central nuclear pore [61].

A new finding of the current study is that frequency of crossings between bundles and bundle thickness depends on the protein length, on the geometry of the simulated volume, on the degree of the tethering, and, particularly, on the initial conformations of the proteins.

An interesting observation is that FGs do not drive the formation of bundles as observed in our study. Given our present mutant study (simulation `mutant_ring`) and that reported in [51], basically the same structures of brush-like bundles with crosslinking nups are formed as are formed with wild-type Nsp1s, suggesting the structures to be independent of FG-interaction. This finding is also supported by our analysis that shows that all kinds of amino acids have similar propensity to arise in the bundles formed. This observation is in contrast to the cross-linking arising in a hydrogel-like structure suggested for the selective phase model: in such structure low affinity inter-FG hydrophobic interactions are responsible for cross-linking [54, 55]. We suggest that FG repeats mainly provide binding spots for transport factors [10, 12–14, 17–19]. The observation that FGs are not mainly involved in bundling implies that they are readily available for TF binding. This supports the reduction-of-dimensionality model [56] that postulates that transport factors pass through the NPC by sliding along the FG surface. A recent study showed that the presence of unbound, thus freely available, FG-repeat motifs helps binding all possible sites of the transport factor and is required for efficient transportion of cargo [62].

ur goal is to model the structure of the entire transport channel of the NPC. In our study, we did not simulate the entire region of the channel interior as it is computationally very expensive. Instead, we simulated systems of Nsp1s starting from different initial states, expecting that the differences represent to some degree the heterogeneity of the NPC pore interior. The salient features of nups in a real NPC channel arise in our models through (i) end-tethering of nups, (ii) protein density, (iii) random conformations for initial states. We interpret our results then to imply that the nups form different structures in different regions of the central channel (Fig. 2.9). The NPC inner diameter is about 30 nm and unstructured nups that are several hundred amino-acids long span this volume of an open pore. Since tethering effects should be minimal in the central region due to the large distance from the NPC wall we expect that the structure found here is similar to that seen in simulation `random_bath`, exhibiting bundles with a high degree of crosslinking forming a mesh. The region in the periphery of the central channel, away from the center, should be represented by a system of tethered nups as in simulation `random_array`, where brush-like bundles with less cross-linking arise.

A recent model [58] for the structure inside a NPC transport channel proposed that the central

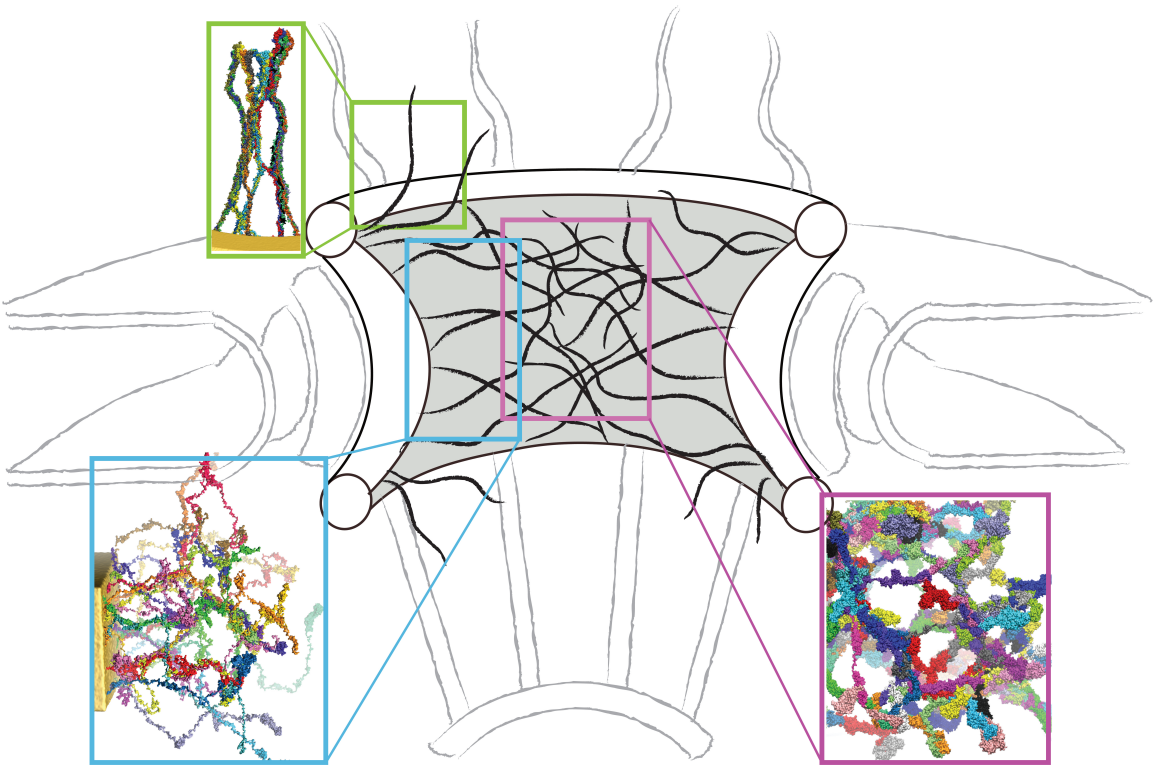

Figure 2.9: Schematic model for the structural assembly of nups in the NPC channel. The strands in black represents bundles of two or more chains. The frequency of cross-linking is higher in the central region and can be identified with the assembly of nups formed in simluation `random_bath` (inset in purple) into a sieve-like mesh; in the periphery brush-like bundles with less cross-linking arise and the respective structure can be identified with with the ones developed in simulations `random_array` (inset in blue) and `wild-type_ring` (inset in green).

region is gel-ike and the periphery is brush-like. The authors postulate two zones, a central and a peripheral zone, for trafficking of molecules in and out of the NPC. According to our model, the central region is a sieve-like mesh arising due to severe cross-linking between bundles protruding into the volume interior.

The major unsolved problem regarding our still limited knowledge of the NPC and NPT is how specific large cargoes with the assistance of TFs manage to pass the NPC. Clearly the NPC being one of the largest and at the same time one of the most dynamic macromolecular systems in eukaryotic cells still holds great secrets and offers opportunity for great discoveries.

Our suggestion here of the principal assembly structure of disordered nups in the NPC, even if completely true, does not imply yet how transport factors can melt the assembly structure for passage. Straight forward MD simulations, even when simplified through coarse groaning, cannot bring about answers as the systems and processes that need to be simulated are much too large and much too slow, respectively. This calamity is actually a bonus as the combination of theory, experiment, and simulation needed is intellectually more rewarding than a straightforward tour de force strategy. But in pursuing the role of transport factors one needs to be open minded about the possibility that yet unsuspected mechanisms play a major role.

2.4 Methods

2.4.1 Fully extended Nsp1s on a gold nanopore

The FG-repeat domain of wild-type Nsp1 was built from Nsp1 sequence 1-609 (Swiss-Prot P14907) by using the 2004.03 release of Chemical Computing Group's Molecular Operating Environment (MOE) software. The backbone dihedral angles (phi, psi) were set to (180 °, 180 °) so that an unstructured straight Nsp1 was obtained. For the model simulation `wild-type_ring`, we tested this model through comparison to experiments reported in [52] that have engineered, using nanotechnology, an artificial pore channel employing nuclear pore proteins inside a gold ring. The system dimension and protein density were chosen to imitate volume-wise the interior of the NPC. We reproduced the gold ring dimensions in [60] and covered the ring with full-length proteins (609-aa-long). As shown in Fig. 2.1, 120 fully extended wild-type Nsp1 chains were grafted to the ring in three concentric rows with 6 nm spacing between adjacent Nsp1s [60]. The C-terminus of the end-tethered Nsp1s was modified by adding five cystine residues that remained fixed to the gold surface throughout the simulation. In [51, 52, 63], these cystine residues formed thiol linkages with

the gold substrate. With the C-terminus of each chain attached to the gold-ring and the rest of the chain fully extended in the direction shown in Fig. 2.1, the whole system was coarse-grained and solvated with CG water in a box large enough to prevent proteins from interacting with their periodic images. A total of 100 mM NaCl was added to the water box, adjusting the relative concentrations of Na^+ and Cl^- to render the whole system neutral. The resulting sytem simulation `wild-type_ring` has 15,453,214 CG beads. For the simulation `mutant_ring` we constructed the ring system as above and replaced all phenylalanines (F) and glycines (G) of the Nsp1s by alanine (A). The resulting system described in simulation `mutant_ring` has 16,019,434 CG beads. Both systems were simulated for 1 μs using coarse-grained molecular dynamics simulations as described below.

2.4.2 Generating random conformations for Nsp1s

To introduce disorder in Nsp1 chains, initial Nsp1 conformations needed in simulations `random_array` and `random_bath` were modeled according to the widely used worm-like chain model [64]. The assignment of the random conformations proceeded in two steps. In a first step, the main-chain C_α beads of Nsp1 constituting the protein backbone were modeled as random homo-polymers constructing the backbone from a self-avoiding walk (SAW) procedure [65]. In this procedure, the SAW is directed under two local geometric restraints, namely keeping a fixed distance of 3.7 Å between neighboring C_α beads and keeping a fixed angle of 127 ° for three adjacent C_α beads. The stated values are used in the MARTINI force field for polypeptide chains with coil conformations. Any two C_α beads were considered to be in close contact if their distance is shorter than 8 Å. We discarded any conformations of backbone chains with beads in close contact within one chain (intra) or between different chains (inter). In the second step, CG side-chains of amino-acids were grafted on the resulting backbone chains. The geometries of the side-chains and gold nano-particles were modeled with standard parameters in the MARTINI force field [35, 66].

2.4.3 Random Nsp1s end-tethered on a 2D array

For modeling end-tethered Nsp1s in simulation `random_array`, we employed again the worm-like chain model described above and chose the last five C_α beads of residues 610-614 as the starting points for the SAW. In the experimental systems [52] nupss are tethered to the gold substrate by means of thiol bonds to cystine residues added to the C-terminus. In order to be consistent with the description adopted for the ring-like geometry, the five C_α beads of each Nsp1 chain were stretched

toward the z-direction and placed on a 5×5 grid in the x, y-plane, with a grid spacing of 6 nm. The full-length worm-like chain was then modeled as described above, such that, 25 Nsp1s were placed as shown in Fig. 2.3. The whole system was coarse grained and solvated with CG water. The system was then ionized with 100 mM NaCl, adjusting again the relative concentrations of Na^+ and Cl^- to render the whole system neutral. The resulting simulation `random_array` involves 1,097,433 CG beads. The system was simulated for 1 μs using coarse-grained molecular dynamics simulations as described below.

2.4.4 Freely floating random Nsp1s in a bath

For simulation `random_bath`, the first three C_α beads of each Nsp1 chain, treated as a rigid body, were chosen as starting points. They were randomly placed in simulation boxes with random orientations. The full-length random conformation Nsp1s were then modeled as worm-like chains as described above, with 120 self-avoiding Nsp1s being placed in box of volume 725 Å$\times$ 725 Å$\times$ 725 Å to match the concentration of Nsp1s as in the simulation `wild-type_ring` (see Fig. 2.6). The whole system was coarse grained and solvated in a CG water box. 100 mM NaCl was then added to the sytem, adjusting the relative concentrations of Na^+ and Cl^- to render the whole system neutral. The resulting simulation `random_bath` involved 3,091,910 CG beads. The system `random_bath` was simulated for 1 μs using coarse-grained molecular dynamics simulations as described below.

2.4.5 Simulation protocol

Simulations were performed using coarse-grained (CG) molecular dynamics based on Marrink's et al model for proteins [35] in NAMD 2.9 [28]. For the use of MARTINI parameters for CG modeling, we adapted the GROMACS switiching function for the LJ potential and a shifting function for the Colomb potential. Non-bonded interactions were cut off at 12 Å, with shifting throughout the interaction range for electrostatic interactions and beginning at 9 Å for vdW interactions, implementing a smooth cut-off. Simulations were performed using a 10 fs timestep. Pair lists were updated at least once every ten steps, with a 14 Å pair list cut-off. In all cases we performed Langevin dynamics with a damping coefficient of 5 ps^{-1}. A constant pressure of 1 atm was maintained with a Nosé-Hoover Langevin piston [67], using a piston period of 2000 fs and a decay time of 1000 fs. All systems were allowed to equilibrate as follows: first, the system was energy minimised for 5000 steps and molecular dynamics was performed for 2 ns in an NVT ensemble (T=300 K). The resulting system was then simulated for 1 μs assuming an NPT ensemble.

2.4.6 Analysis of bundles

As described below in Results, the simulated Nsp1s formed strands that we refer to as bundles. We define a bundle as a linearly arranged cluster of multiple Nsp1 chains. In such a cluster, every amino-acid in a segment of one Nsp1 chain is in contact with at least one amino-acid of another Nsp1 chain. In order to provide a quantitative characterization of these bundles we applied graph theory to identify bundle segments in a given configuration of the Nsp1 chains. In this approach, each amino-acid is represented by a node in a graph. If the distance of two amino-acids A and B from different chains is shorter than 6 Å, the corresponding nodes in the graph are connected with an edge. In addition, the nodes for amino-acids adjacent in sequence to the two amino acids A and B forming an edge are considered to be connected to the nodes for amino-acids A and B. Therefore, bundle segments can be identified by examining connected components in a graph as constructed above. For this purpose we employed the breadth-first search algorithm [68].

Thickness of a bundle is determined as the number of different Nsp1 chains that belong to the bundle. Another bundle characteristic determined here is the fraction of each type of amino-acids involved in a bundle.

2.4.7 Pore characterization

Nsp1 polymers self-assemble, as shown in Results, into network structures. In order to assess how these structures are related to gating in the NPC we determine the size limit of spherical particles being capable geometrically to pass through the structures. This size limit defines the pore size, namely as the largest possible radius of passing particles. Fig. 2.8 illustrates how the pore size is determined algorithmically. Starting from one side of the network structure, spheres of various sizes are moved towards the other side. The analysis was applied to the final structures resulting from simulations `random_array`, `random_bath`, `wild-type_ring`, and `mutant_ring`; in the case of the latter two simulations structures are formed with Nsp1s arranged in a ring-like arrangement. In these cases, the spheres were moved radially from the inside of the Nsp1 ring structure to its outside.

We implemented the above analysis through an algorithm in which the simulation box with the network structure was mapped into a cubic lattice. We then determined for each grid point (x) in the lattice the radius $r(x)$ of the largest sphere that could be placed on this grid point without sterical

clash with any Nsp1 chain; this radius property was calculated as the shortest distance between the grid point and protein beads. A sphere is considered capable of moving in the lattice along a given pathway, if the size of the sphere does not exceed the bottleneck of the pathway, i.e., the smallest $r(x)$ of the grid points on the pathway. To characterize the largest sphere that could permeate the network structures one searches all possible pathways, from one side of the network structure to the other side, to find the pathway that has the largest bottleneck. The pathway search was performed using Dijkstra's algorithm [69].

Chapter 3

Cytoplasmic domain filter function in the mechanosensitive channel of small conductance

[1]

3.1 Introduction

MscS, the mechanosensitive channel (MS) of small conductance, belongs to a family of pressure sensitive channels that play a major role in osmoadaptation of bacterial cells [23, 24]. Bacteria experience changes in membrane tension as a consequence of water influx and efflux induced by an imbalance of osmolytes between intra- and extracellular spaces. The channels perceive membrane tension resulting from swelling and function as safety valves to alleviate cell turgor by releasing osmolytes along with water and, thereby, prevent cell lysis [25]. The MscS channel opens at slight pressure differences of tens of mmHg, assuming then a conductance of 1 nS.

Bacteria feature three functionally similar MS channels: MscM (M for mini), MscS (S for small), and MscL (L for large). These channels respond in a graded manner to relieve the cell from osmotic stress [21, 22]. MscM opens most readily in response to membrane tension, assuming then a conductance of 0.1-0.3 nS. MscL opens in response to pressure changes approaching the cell lytic limit, namely $\Delta p < -40$ mmHg (-40 torr), assuming a conductance of 3 nS [71]; MscL opens only as a last resort, i.e., it functions as an emergency valve to release all contents, even proteins, to help the cell escape imminent cell lysis [22, 25–27]. Experiments have provided a great deal of phenotypic evidence to suggest the physiological role of these channels in osmosensing and adaption during osmotic shock.

MscS is a homo-heptameric complex, each monomer (P1 to P7) of the *Escherichia (E.) coli* channel consisting of 286 amino acids. Two crystal structures of *E. coli* MscS have been determined, one of the wild-type (Pdb: 2OAU) with MscS in a putative closed state [72, 73] and the other of a

[1] This chapter appeared as a research article in Biophysical Journal [70]. R. Gamini *et al.*, "Cytoplasmic domain filter function in the mechanosensitive channel of small conductance". *Biophysical Journal*, 101(1):80–89, **2011**. doi: 10.1016/j.bpj.2011.05.042

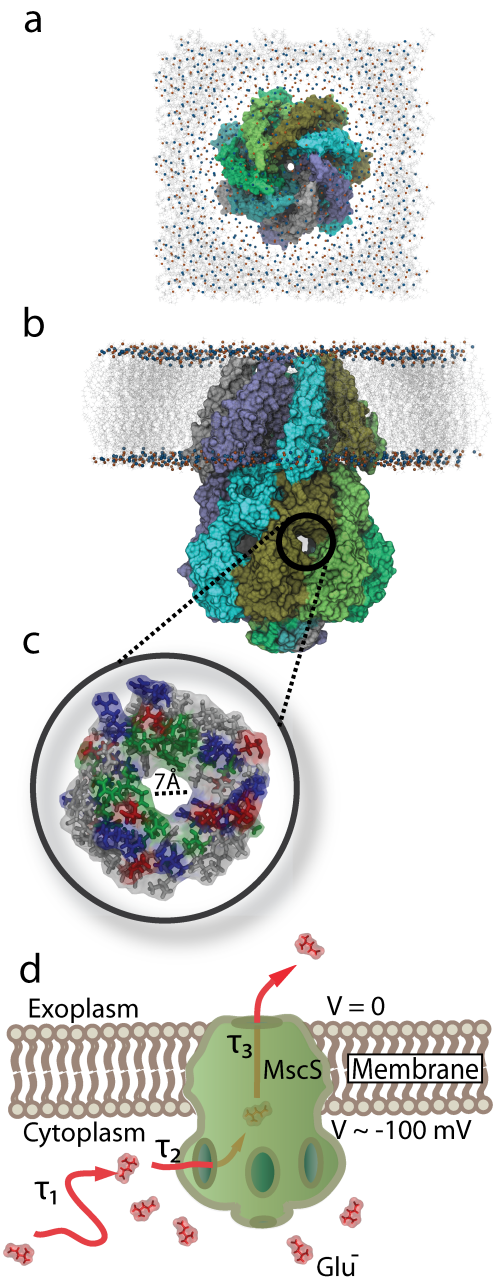

Figure 3.1: Placement of MscS in the cell membrane. Placement of MscS in the cell membrane. *(a-b)* Top and side views of the homoheptameric MscS (Pdb:2OAU), respectively. The latter view distinguishes clearly the transmembrane and the cytoplasmic domains. The cytoplasmic domain contains seven equal openings on its side, one being clearly visible and enlarged, exhibiting the 7 Å opening in *(c)*. Colors distinguish the seven MscS subunits. *(c)* View of the side opening. The circular insert shows MscS as a transparent surface with side groups lining the opening shown in licorice representation, colors grey, green, red, and blue denoting non-polar, polar, negative, and positive side groups, respectively. *(d)* Schematic illustration of MscS and the passage of cellular material from the cytoplasm to the extracellular space upon opening of the channel due to an osmotic pressure difference. A negative intracellular voltage drives negative ions, e.g., the osmolyte Glu^-, out of the cell. MscS combines a transmembrane domain with a mechanosensitive channel and a large cytoplasmic domain of unclear function. Here the possible function of the cytoplasmic domain as a filter is investigated. An intracellular solute would leave the cell in three steps: (1) diffusion towards the cytoplasmic domain; (2) translocation through openings in this domain; (3) translocation through the transmembrane channel. τ_1, τ_2 and τ_3 denote the durations of the respective steps.

mutant (A106V, Pdb: 2VV5) with MscS trapped in an open state [74]. The channel subunits have two distinct domains, a transmembrane (TM) domain and a cytoplasmic domain (CD) (Figs. 3.1a, b). Each TM domain comprises three α-helices labeled TM1, TM2 and TM3. The seven subunits form a central TM pore made of 21 helices that sense tension and open and close the channel as discussed in [72, 73, 75–83].

The two crystal structures of MscS [72–74], although representing different putative states of the TM domain, depict the same architecture for MscS' large cytoplasmic domain (Fig. 3.1b). The CD comprises $\sim$65% of the total mass of the protein and is a conserved structural feature in the MscS family whose function remains unclear. The CD is mainly composed of β-sheets that form a balloon-shaped, water-filled chamber with a diameter of $\sim$80 Å. The CD has seven distinct, roughly hourglass-shaped side openings, each $\sim$7 Å wide in the middle, formed between adjacent MscS subunits and has a single narrow axial opening distal to the pore. The side openings (Fig. 3.1c) are lined by side chains that are basic, acidic, polar and non-polar, thus suggesting only weak selectivity, if any at all, in regard to solutes allowed to pass through. In accordance with this feature, both K^+ and Cl^- have been observed in simulations to pass through the openings [76, 78]. The narrow distal opening has its interior lined with hydrophobic residues which prevent easy passage of hydrated solutes, such that this opening appears to be effectively closed.

The CD side openings are the only gateway for water and cytoplasmic solutes to reach the cell exterior. Translocation of solutes involves three steps as shown in Fig. 3.1d: (1) diffusion of solutes towards the CD requiring a time τ_1 to find a side opening; (2) passage of solutes through an opening requiring a time τ_2; and finally (3) translocation through the TM pore to the cell exterior requiring a time τ_3.

The present study focusses on function and mechanism of the MscS CD. Various investigations suggest the CD to play a role in channel gating [84–88] and MscS conductivity [80]. Deletion studies suggest the domain to be required for gating and stability of the channel [85, 89]. Here we consider, however, another likely function of the CD, namely that of a sieve that regulates the efflux of solutes valuable to the cell. Such function is consistent with the observation that MscS with largely deleted CDs remains physiologically functional [89], the respective cells being, however, less viable under high salt conditions [80]. MscS without its sieve could still protect the cell from osmotic shock by opening and closing its TM domain, but would do this at a higher cost to the cell as it would spill too much of its valuable content during osmoadaptation. Furthermore, the CD appears to be specifically designed to prevent dominant efflux of negative ions, e.g., Glu^-, which is favored by the

cellular potential, and instead appears to maintain, by mixing equal amounts of Glu^- and K^+ ions, neutral efflux that conserves the cellular potential.

In our study of MscS' filter and mixing function we select two osmolytes for particular attention, namely, Glu^- and K^+. We consider Glu^- and K^+ as examplary cytoplasmic solutes that shed light on MscS' function in general due to their high concentration in *E. coli* cells. Our approach involves mathematical and computational modeling. We present a mathematical description of the diffusion-controlled approach of Glu^- and K^+ to the MscS CD and to the side opening, obtaining an estimate for the time τ_1 (see Fig. 3.1d). We then describe computationally the transport of Glu^- through a CD opening providing an estimate for the time τ_2 and through an account of the transmembrane potential we estimate the channel passage time τ_3. τ_1, τ_2 and τ_3 are also estimated for K^+. The calculations lead us to the overall conclusion that the CD of MscS balances passage of Glu^-, K^+ and other osmolytes. The key finding is that MscS achieves its functionality seemingly without its side openings getting clogged and with minimizing loss of cell polarization.

3.2 Results

MscS furnishes a cell's initial defense against cell lysis, being gated at low osmotic pressure differences, namely, of a few tens mmHg. At this low pressure point, i.e., without threat of imminent lysis, the cell may protect itself while preventing cell depolarization and loss of particularly valuable osmolytes, employing for this purpose the cytoplasmic domain (CD) as a filter.

One such osmolyte should be the Glu^- ion, the main negative counter charge for the cell's positive ions like K^+ and also an essential metabolite. In this section we investigate, hence, the role of the CD as a filter that controls Glu^- and K^+ efflux. This can be achieved by the CD acting as what we like to refer to as an entropic filter, that prevents Glu^- from reaching the CD pores, but once Glu^- enters a pore, lets it pass unhindered due to favorable enthalpic interactions to prevent unclogging of the pores. We note that the cell potential of about -100 mV tends to expel Glu^- from the cell and retain K^+. The CD can act by slowing efflux of Glu^-, while attracting K^+, achieving thereby a neutral osmolyte current and conserving partially the cellular potential along with retaining as much as possible the valuable Glu^- ion.

In the following we demonstrate theoretically and computationally CD's action as such entropic filter. We first describe mathematically the diffusion controlled approach of Glu^- to the CD and its side opening entrance (c.f. Fig. 3.1d). We then demonstrate that Glu^- can readily pass the pores

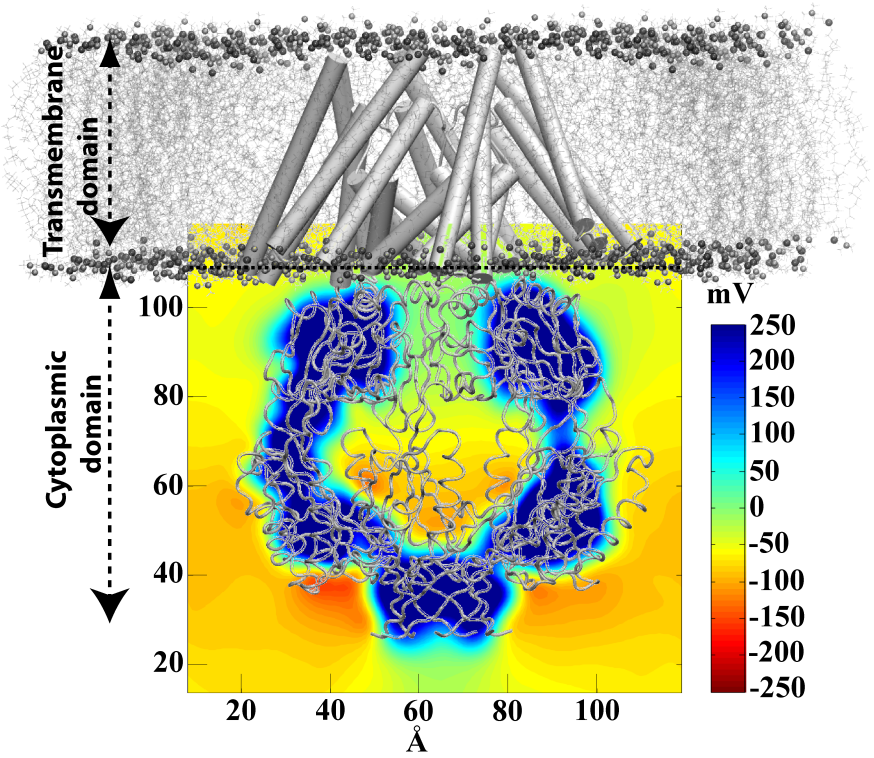

Figure 3.2: Time-averaged electrostatic potential due to the MscS cytoplasmic domain. Time-averaged electrostatic potential due to the MscS cytoplasmic domain. The potential shown has been determined through averaging over our all-atom MD simulations as described in [90]. The transmembrane domain of MscS, shown in gray as an overlay over a standard membrane, has not been included in the calculation. The potential has been obtained under zero biasing potential. The structure of the cytoplasmic MscS domain is superimposed to facilitate the interpretation of the electrostatic features shown.

by computing Glu^-'s potential of mean force in the pores and then calculating on this basis the mean time of passage through the pore (c.f., Fig. 3.1d). Our conclusions regarding CD's function can be based on solely a qualitative treatment of the overall translocation behavior of Glu^- as the filter effect of the CD turns out to be quite drastic.

3.2.1 Electrostatic potential

The first step in Glu^- translocation involves the diffusive approach to the CD and Glu^- entering one of CD's pores, in the following referred to as the "openings". As Glu^- carries a $-e$ charge it experiences a repulsive, long-range interaction with the $-7e$ overall charge of the CD, screened by the water dielectric property. Fig. 3.2 shows the electrostatic potential inside and outside of the CD averaged over an MD trajectory as described in Methods. The potential is weakly repulsive for Glu^-

outside and inside the CD, in particular in the vicinity of the side openings while the inaccessible protein part is attractive.

3.2.2 Diffusion of Glu^- and K^+ to cytoplasmic domain and its openings

.

A mathematical description of the diffusive approach for the realistic geometry and potential of the CD shown in Fig. 3.2 requires a numerical approach which would not be very illuminating at this point, but should be furnished eventually. For the sake of qualitative argument a simplification to a radially symmetric geometry and potential, that permits an analytical treatment, is sufficient and will be pursued now. For this purpose we employ the stationary rate constant of the diffusion controlled binding to the CD opening, $k(R_o, w)$, as derived in Methods for the case of a Coulomb potential (see Eq. (3.33)). In case of the potential $U(r) = q_{\rm CD} q_{\rm Glu}/\epsilon\, r$ with $q_{\rm CD} = -7e$, $q_{\rm Glu} = -e$, and $\epsilon = 80$ (approximate relative dielectric constant of water) one determines for $R_{\rm L} = 7\,\beta\, e^2/\epsilon$ the value $R_{\rm L} = 49$ Å. The radius of the CD at the level of the side openings is $R_o = 43$ Å(as measured using VMD). We assume for the diffusion constant an r-independent value of $D_{\rm Glu} = 0.75 \times 10^{11}$ Å^2/s [91]. To determine $k(R_o, w)$ one needs to specify the reaction constant w accounting for the probability, once Glu^- has encountered the CD surface, that the ion actually enters an opening.

In order to estimate w for the case of a CD surface dotted with seven openings we envision the following process: when Glu^- has approached the CD surface it diffuses around until it hits the mouth of a CD opening (c.f., Fig. 3.1c) and then enters the opening. The probability of this to happen is the quantum yield $Q(R, R, w)$ of a diffusion process starting at the surface of the CD and being absorbed by an opening, i.e., for $R = R_o$. Since the electrostatic potential near the surface of the CD is largely negative, as seen in Fig. 3.2, we assume that openings are entered only for more or less direct hits, such that we choose w through the condition

$$Q(R, R, w) \;=\; q_o \tag{3.1}$$

where $q_o = 7\, r_o^2/4R_o^2$ is the ratio of the combined area of seven openings, i.e., of $7 \times \pi r_o^2$, to the total CD surface, i.e., $4\pi R_o^2$. Using $r_o = 4.5$ Å (this value will be rationalized below) and $R_o = 43$ Å, we estimate $q_o \approx 0.02$. We derived in Methods the expression Eq. (3.36) for $Q(R, R, w)$ which can be

written

$$Q(R, R, w) \;=\; 1 \;-\; \frac{D}{w\,R \;+\; D}\;. \tag{3.2}$$

From this follows the desired expression for w, namely,

$$w \;=\; \frac{q_o}{1 \,-\, q_o}\,\frac{D}{R}\;. \tag{3.3}$$

Rewriting the last result as $D/wR = (1 - q_o)/q_o$ one can express Eq. (3.33)

$$k(R_o, w) \;=\; 4\,\pi\,D\,R_{\rm L}\left[\frac{R_{\rm L}}{R_o}\,\frac{1 \,-\, q_o}{q_o}\,e^{R_{\rm L}/R_o} \;+\; e^{R_{\rm L}/R_o} \;-\; 1\right]^{-1} \tag{3.4}$$

$$k(R_o, w) \;=\; 4\,\pi\,D\,R_o\,\frac{q_o}{1 \,-\, q_o}\,e^{-R_{\rm L}/R_o}\;. \tag{3.5}$$

Approximating $\exp(-R_{\rm L}/R_o) \approx 1$ and $1 - q_o \approx 1$ one obtains the rate constant $k = 4\pi D R_o q_o$, i.e., the diffusion-controlled rate constant reduced by the factor q_o [c.f., Eq. (3.30)]. This is the expected result in the case of weak Coulomb repulsion and a small reaction probability.

We conclude from the mathematical description provided that the approach of Glu^- to the CD is governed roughly by the rate constant stated in Eq. (3.5). The factor $4\pi D R_o$ describes the rate constant for diffusion towards the CD (radius R_o), the factor $q_o/(1 - q_o)$ accounts for the time delay due to the small size of the CD openings, and the Boltzmann factor $\exp(-R_{\rm L}/R_o)$ includes the effect of the electrostatic repulsion between Glu^- and the CD. Employing the already stated numerical values for R_o and $R_{\rm L}$ along with the stated diffusion coefficient $D_{\rm Glu}$ and a concentration $c_{\rm Glu} = 0.2$ mol/liter [92], one obtains for the time τ_1 of diffusion controlled entering of the CD openings by Glu^-

$$\tau_1 \;=\; 1\,/\,k(R_o, w)\,c_{\rm Glu} \;\approx\; 32\,\mathrm{ns}\;, \tag{3.6}$$

i.e., on average about one Glu^- per 32 nanoseconds enters a CD opening.

If the entire CD would be porous for Glu^-, the average time τ_1 of reaching the pores would be 0.6 ns, i.e., 50 times shorter than without the filter function of the CD openings. Clearly, the CD acts as an entropic filter, turning randomly a large fraction of approaching Glu^- away, preventing, thereby, a quick loss of Glu^-.

For K^+ the electrostatic potential around the CD is attractive. One can assume that, once the ion has reached the CD surface, the quantum yield of finding a pore opening is close to one as K^+ remains attracted to the surface long enough to eventually find an opening. The rate of K^+ entering

the CD openings is then given by Eqs. (3.32, 3.33) for $w \to \infty$ and $q_s = e$, i.e., $R_L = -49$ Å i.e., is $4\pi D_K |R_L| c_K / (1 - \exp(-|R_L/R_o|)$; here we assume that the search time of K^+ on the CD surface is negligible. With $D_K = 2 \times 10^{11}$ Å^2/s [93], $c_K = 0.2$ mol/liter, and $|R_L/R_o| = 1.139$ one obtains $\tau_1 = 0.045$ ns. To this time should be added the search time, τ_{search} of a few nanoseconds. Even though τ_{search} is larger than τ_1, the added contribution does not matter as we will see below.

3.2.3 Free energy of Glu^- and K^+ translocation

The entropic filter mechanism of the CD is based on the premise that CD wall openings do not get clogged by substances like Glu^-, i.e., that the passage (through the CD openings) time of Glu^-, τ_2, is very short compared to the "open"-time of MscS. τ_2 is estimated computationally in the next section.

The free energy profile characterizing Glu^- translocation through the side openings, calculated as described in Methods, is shown in Fig. 3.3. The profile shows Glu^-, initially at a distance 51 Å away from the center of the CD, moving toward the entrance of the opening subject to a slightly attractive potential due to interaction with positively charged residues, lysines and arginines, that line the walls of the openings. The openings establish a barrier of only $\sim$ 2 kcal/mol for the translocation of Glu^- which is sufficiently low to prevent Glu^- from clogging the opening. It is important to note that throughout the 0.38 μs of ABF simulations, the opening remained quite stable with only minute fluctuations in structure detected. The distance RMSD computed for the backbone of residues lining the opening never exceeded 1.45 Å. We note that the harmonic restraints applied to the C_α atoms of residues 274 and 120 in subunits P1 through P7 (see Methods) should not affect the stated RMSD values as residues 274 and 120 are far away from the side openings. Once Glu^- enters the CD interior it experiences mainly unfavorable Coulomb interactions of 0.5 kcal/mol; this slightly unfavourable interaction is also discernable in Fig. 3.3 as well as in the electrostatic potential Fig. 3.2.

We note from Fig. 3.3 that from the cytoplasmic direction a first barrier arises inside the side opening at $\zeta = 20$ Å; Fig. 3.3d shows that at this ζ-value the opening has a radius of 4.5 Å, which is why we assumed in our calculation of $k(R_o, w)$ above this as the r_o-value.

The small barrier for Glu^- ion suggests τ_2 for Glu^- ion to be small such that these ions may readily pass through the openings. This is consistent with spontaneous diffusion of Glu^- ions passing through the side openings of CD; such diffusion was observed during our simulations. Indeed,

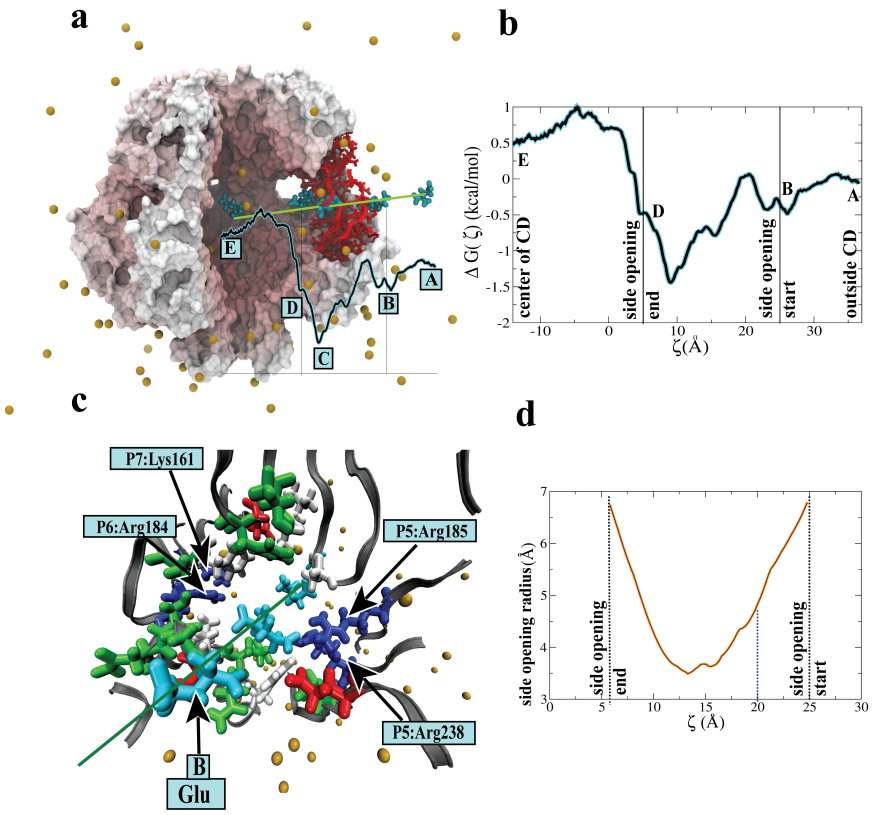

Figure 3.3: Free energy profile for Glu^-, a key cellular osmolyte, translocating a side opening in the MscS cytoplasmic domain. Free energy profile for Glu^-, a key cellular osmolyte, translocating a side opening in the MscS cytoplasmic domain. *(a)* View of the outside and inside of the cytoplasmic domain, cut open. The protein is shown in surface representation. On the right, an opening is presented with the protein shown in red in ribbon representation with external side groups in licorice representation. Ions (K^+) are shown as gold spheres, water is not shown. Glu^- is shown in five stages of translocation, labeled A, B, C, D, E. The potential of mean force, accounting for the energy of Glu^- during translocation and obtained from an equilibrium ABF simulation totaling 0.38 μs of sampling, is shown below the Glu^- snapshots; the potential has been determined along the line connecting the Glu^- snapshots which are also denoted in the potential. *(b)* Potential of mean force for translocating Glu^-. The potential of mean force in *(a)* is shown again with fully labeled axes. The domain side opening is about 20 Å long; the distance from the inside end of the opening and the domain center is about 20 Å with the center of CD corresponding to -14 Å on the ζ-axis; the potential of mean force varies over a range of 2 kcal/mol, i.e., only by 4 k_BT. *(c)* Close-up view of MscS amino acid side groups lining the domain opening. The residues coming within 4 Å of Glu^- along the translocation pathway are depicted in licorice representation in white, green, red, and blue for non-polar, polar, negative, and positive amino side groups, respectively; the protein distant from the opening is shown in dark grey ribbon representation; K^+ ions are shown in gold. The Glu^- ion, shown at locations B, C, D (c.f. *(a, b)*), is colored in light blue. The green line shown passing through the Glu^- snapshots corresponds to the ζ-axis in *(b)*. Glu^- position C corresponds to an enthalpic energy minimum where the ion is stablized by interactions with Lys161 of subunit P7, Arg184 of subunit P6, and with Arg238 and Arg185 of subunit P5. *(d)* Size of the openings in the MscS cytoplasmic domain. The graph shows the radius profile of the side opening as determined by the program HOLE [94]; the radius is smallest at the center of the opening (c.f., *(b)*); the opening stretches from 5 Å to 25 Å.

during 0.38 μs, four events of spontaneous diffusion of Glu^- ions were observed. In one of the four cases, a Glu^- ion transited from the bulk to the CD interior, whereas, in other cases, the ions moved out of the CD. We also observed spontaneous transit of a K^+ ion moving from bulk to the interior of the CD. Spontaneous diffusion of K^+ and Cl^-ions through the side opening were also reported in earlier studies [76, 78]. These observations suggest that the free energy barrier for K^+ is as small as for the Glu^- ion, allowing unhindered permeation of K^+ ions via the openings. Using free-energy perturbation (FEP) (see Appendix A), replacing Glu^- by a K^+, at positons A, C and E (Fig. 3.3a and 3.3b) along the translocation pathway, the barrier at the opening was estimated to be $\sim$ 3.1 kcal/mol. These observations also suggest that the passage times τ_2 for Glu^- and K^+ ions are comparable.

3.2.4 Mean passage time τ_2 of Glu^- and K^+ passing through a CD opening

To further illustrate that Glu^- can readily pass through with only a short resident time inside the CD openings, we evaluate the mean passage time, τ_2, from the obtained PMF, $U(x)$, according to [95, 96]

$$\tau_2(x_0) \;=\; \frac{1}{D}\int_{x_1}^{x_0} dx\, e^{\beta U(x)} \int_{x}^{x_0} dy\, e^{-\beta U(y)}\;. \tag{3.7}$$

Here, we assume the same diffusion coefficient of glutamate in water as above, i.e., $D_{\mathrm{Glu}} = 0.75 \times 10^{11}\,\text{Å}^2/\text{s}$. The integration limits x_1 and x_0 correspond to Glu^- set at an initial position outside the CD, x_1, and reaching the center of the CD, x_0 (see Fig. 3.3b). Numerical integration of the above expression yields $\tau_2 \approx 40$ ns. MscS channel open times from patch clamp measurements are about 100 ms [75]; τ_2 is extremely short in comparison such that one can conclude the CD openings to remain unclogged.

To obtain the corresponding value of τ_2 for K^+, the above equation (3.7) was numerically integrated after splining over $U(x)$ obtained at three positions viz., A, C and E (Figs. 3.3a and 3.3b) that was estimated from FEP calculations as described in methods (also see Appendix A). Here, the same integration limits x_1 and x_0 were used as for Glu^-. Assuming for the K^+ ion a diffusion coefficient $D_{\mathrm{K}} = 2.0 \times 10^{11}\,\text{Å}^2/\text{s}$, τ_2 for K^+ ion is $\approx$ 20 ns. The stated τ_2 values for Glu^- and K^+ suggest that both ions permeate the CD openings quickly.

ion	τ_1 (ns)	τ_2 (ns)	τ_3 (ns)	$\tau_3{}'$ (ns)	$\tau = \tau_1 + \tau_2 + \tau_3$(ns)
Glu^-	32	40	10	260	82
K^+	0.045	20	35	375	55

Table 3.1: Computed mean passage times τ_1, τ_2, τ_3 and τ_3' for Glu^- and K^+ ions. τ_3 is the mean first passage time through the actual transmembrane channel (see text); τ_3' is the mean first passage time from the center of the CD through the transmembrane channel.

3.2.5 Mixing of ions

The simulated CD contains about 2,600 water molecules. At the standard water molar concentration of 56 mol/liter the water molecules occupy about $8{\times}10^4$ Å^3 and in equilibrium with Glu^- and K^+ concentrations of 0.2 mol/liter should harbor about 10 ions of each type delivered with a time constant of $\tau' = \tau_1 + \tau_2$ where τ_1 and τ_2, already discussed are given in Table 3.1. We derive in Appendix B that the ions in a spherical cavity prefer energetically an equal number due to Coulomb interaction; for a cavity the size of the CD a well mixed state is being favored by several k_BT. We conclude from our description that the CD maintains on average a state of about ten Glu^- and ten K^+ ions. The ions spread diffusively throughout the CD interior such that there are always some ions near the membrane channel exit. The scenario suggested depends critically on the entrance times τ' for the ions being short compared to their exit times τ_3 being evaluated now. From Table 3.1 we find that the entrance times are 72 ns for Glu^- and 20 ns for K^+.

3.2.6 Times τ_3 of Glu^- and K^+ exiting through transmembrane channel

The exit times τ_3 can be estimated as mean first passage times of the ions permeating from inside the CD to the extracelluar space, assuming the open state conformation of MscS (PDB:2VV5) [74]. τ_3 can be evaluated assuming that the passage is governed by an effective potential, $\tilde{U}(x)$, and then is given by the expression (see Appendix C)

$$\tau_3 \;=\; \int_{x_{\rm center}}^{x_{\rm out}} dx D^{-1}(x) \exp[\beta\tilde{U}(x)] \int_{x_{\rm apex}}^{x} dx' \exp[-\beta\tilde{U}(x')] \tag{3.8}$$

In the above expression, $\tilde{U}(x) = U(x) - k_{\rm B}T \ln Z(x)$ combines enthalpic $[U(x)]$ and entropic $[-k_{\rm B}T \ln Z(x)]$ contributions. $U(x)$ is the transmembrane potential characterized through the resting potential of the *E. coli* cell namely, $-V_o = -100$ mV [97, 98]; $Z(x)$ accounts for the CD geometry along the conduction path CD $\rightarrow$ channel and is $Z(x) = \pi(r^2(x)/r_0^2)$, where $r(x)$ is the radius of the cross section along the symmetry axis, x, of the MscS and r_0 is a reference radius, the value

of which drops out of the evaluation. The HOLE program [94] was used to determine the radius profile of MscS and, thus, $Z(x)$. In Eq. C.2, x_{apex}, x_{center} and x_{out} are positions along the CD axis located at the CD apex, CD center, and end point of the transmembrane channel. We note that the integration limit x_{apex} in Eq. C.2 accounts for the possibility that an ion diffuses in the direction away from the channel mouth before it exits the CD.

$Z(x)$ accounts for an effective entropic barrier arising from the need that the diffusing ions has to find the channel mouth. In case of well mixed ions of sufficient density one can assume that there are always Glu^- and K^+ ions near the mouth such that little time may need to be spent to find it. To estimate the time to exit the CD one can determine τ_3 for two scenarios, (i) starting near the channel mouth (in which case x_{apex} and x_{center} in Eq. C.2 are replaced by x_{exit}), (ii) starting from the center with $Z(x)$ included. Table 3.1 lists for the two scenarios the times 10 ns and 260 ns in case of Glu^- and 35 ns and 375 ns for K^+. Comparing these times with the entrance times ,τ', discussed above we conclude that the times to search and exit the CD, τ_3', are significantly longer than the entrance times while the pure exit times, τ_3, are shorter in case of Glu^- and longer in case of K^+. We conclude then that the CD, in mixing the ions, keeps always some Glu^- and K^+ ions ready near the channel mouth to exit faster than a new ion could that just permeated a CD side pore.

We compare finally the scenario (i) exit times, τ_3, with the ion conductances measured for the respective ions. The 10 ns value for Glu^- corresponds to a current of 100×10^6 ions per second or a channel conductance of ≈ 230 pS. Since an open MscS channel can accommodate two Glu^- ions at any one time, the channel conductance for Glu^- increases by a factor of two and should then be $\approx$ 500 pS which is in agreement with the experimentally measured channel conductance for Glu^- [24]. This agreement is an argument in favor of the scenario described for the overall ion conduction of MscS in which the CD provides an ion buffer that effectively accelerates ion permeation as mixed ions search for the channel mouth in parallel, rather than sequentially.

Taken together, the total passage time (see Table 3.1), $\tau = \tau_1 + \tau_2 + \tau_3$, is nearly the same for the two ions, irrespective of the time τ_{search} (discussed above) which should only be a few nanoseconds. The values suggests that the CD serves to maintain an overall balance between positively and negatively charged osmolytes leaving the cell, precluding loss of cell polarization. The overall balance of osmolytes is also enforced by electrostatic interaction between osmolytes inside the CD.

3.3 Discussion

The cytoplasmic domain (CD) is a conspicuous component of the mechanosensitive channel, MscS. It is not directly involved in the protein's TM channel. Yet, it makes up the major body of the protein. Given the CD's sieve-like and mixing bowl-like architecture along with its cytoplasmic position, a filter and mixer function suggests itself and has been investigated in this study.

The filter function is more complex than one realizes right away, the key problem arising from the need to avoid clogging the filter. A naive solution regarding the filter mechanism would be to involve in the filter openings an enthalpic barrier to solutes. However, this would be the worst solution as it would promote clogging. In fact, the opposite solution is more advisable, namely, that solutes pass openings quickly, but that they do not reach the openings in the first place. We call this an entropic filter mechanism.

In our study we have identified indeed such property for the CD openings of MscS. We focussed first on Glu^-, a key osmolyte and metabolite in *E. coli*, present at high concentration. We find that Glu^- passes the CD openings readily without significant enthalpic barrier which suggests that the CD acts as an entropic filter. The average electrostatic field around the CD supports this notion as the CD surface potential is repulsive in case of the negative Glu^- osmolyte. We provide in our study a formal description of the entropic filter function employing the theory of diffusion-controlled reaction processes. Clogging of the CD openings is investigated through calculation of the mean first passage time τ_2 of Glu^- making it through the channel, which was determined from a potential of mean force obtained through simulation: τ_2 turned out to be much shorter than typical "open" times of MscS which implies that the CD openings remain basically unclogged as regards Glu^-. We then calculated the mean passage time for Glu^- to leave the CD and reach the extracellular space. The total time $\tau = \tau_1 + \tau_2 + \tau_3$ is estimated to be about 80 ns.

We then focussed on K^+, also present at high concentration in *E. coli* cells. In this case, the overall exit time τ is about 55 ns, i.e., close to the 80 ns Glu^- value, but τ_1 is shorter and τ_3 is longer than for Glu^-. Based on these times for Glu? and K+, we conclude the filter function of the CD shown in 3.4. MscS, through the combination of CD and channel, manages an amazing feat, namely even though the cell potential by itself clearly favors expulsion of Glu^-, compensation through a short K^+ τ_1 makes the overall exit times of Glu^- and K^+ about the same. This impressive balance is likely stabilized through electrostatic interactions in the CD. The size of the CD with an inside opening of volume 8×10^4 Å^3 permits that at any one time about ten positive and ten negative ions are being harbored, given the cytoplasmic ion concentration of 0.2 mol/liter. The lowest electrostatic

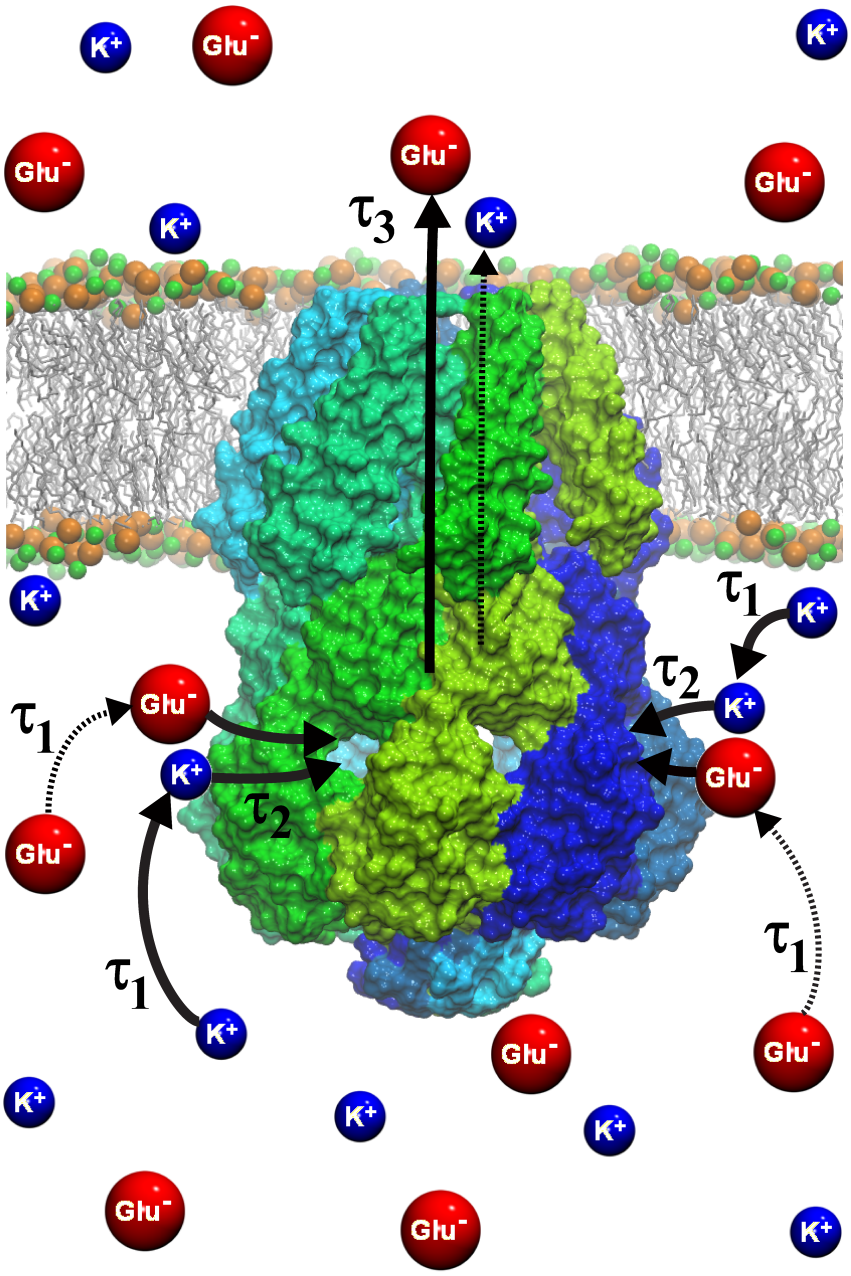

Figure 3.4: Diffusion of Glu^- and K^+ ions through the cytoplasmic domain of MscS. Diffusion of Glu^- and K^+ ions through the cytoplasmic domain of MscS. Due to the topology of the electrostatic potential in conjunction with the intrinsic characteristics of the osmolytes, K^+ ions diffuse faster than Glu^- ions towards the CD (τ_1), permeate through the pores at comparable times (τ_2), but exit the channel slower than their anionic counterpart (τ_3); (fast translocation shown as thick, solid arrows and slow one as thin, dashed arrows)

energy is assumed when the CD harbors the same number of positive and negative ions, thereby favoring a neutral osmolyte current through MscS in the open state.

The present study is but a first step into the filter and mixing function of the MscS CD. Given the many types of solutes that need to be considered in this regard, the study needs to be extended, though for a first investigation it seemed wise to concentrate on just two ions, Glu^- and K^+. Future studies should involve a multi-scale approach that combines transport theory with MD simulations, finite element method of the bulk cytosol with the detailed geometry of MscS CD and membrane surface, and covers the entire solute translocation, i.e., from cytoplasm to the extracellular space, as a non-stationary process. A particularly exciting aspect of future studies should be ion-ion interaction effects in the CD that favor neutral efflux and preserve cell polarity.

3.4 Methods

3.4.1 Systems modeled

The MscS cytoplasmic domain (CD) was modeled starting from the full atomic structure obtained at a resolution of 3.9 Å from X-ray crystallography (PDB 2OAU) [72, 73]. We note here that the crystal structure lacks a 6 amino-acid segment at the C-terminus which has zero net charge; we do not model this segment in our system. The CD of the wild-type protein modeled here is essentially similar to that found in the putative open mutant A106V [74]. The CD construct was obtained by removing the N-terminal transmembrane residues 27-112; the resulting N-terminus was modified with N-acetyl using the `PSFGen` plugin of VMD [99] to have an uncharged N-terminus. The plugin also placed the missing hydrogen atoms into the structure. The CD was solvated using the `Solvate` plugin of VMD to add 13 Å of water padding in each direction. In order to compute the intrinsic electrostatics of the CD, the solvated system of the MscS CD, which has an overall charge of -7.0 e, was neutralized by placing into the simulated volume seven K^+ counter ions using the `cIonize` plugin of VMD. The resulting system of the CD in a water box, `SimA`, had 128,390 atoms (in a 10.8 nm×11.1 nm×10.3 nm volume). System `SimB` was modeled adding K^+ and Glu^- ions to the solvated system at random positions using the VMD `ionize` plugin such that the neutralized system contained 200 mM KGlu. This concentration is within the typical KGlu range found in bacterial cells [92]. We modeled free glutamate molecules as zwitter ions carrying a net -1.0 e charge due to its side-chain. The resulting setup had 123,462 atoms (in a 10.8 nm×10.8 nm×10.1 nm volume). In all simulations, constraints were imposed on the C_α atoms of residues 274 and 120 of all subunits,

P1 through P7, to prevent the domain from drifting during simulations; the constraint assumed for this purpose was harmonic with a spring constant of 1.0 kcal mol^{-1} Å^{-2}. Molecular dynamics simulations were performed as described below.

3.4.2 Molecular dynamics simulations

All molecular dynamics simulations were performed using NAMD 2.6 [28] with CHARMM27 parameters [100] and CMAP corrections [41]. The TIP3P water model [101] was employed for water. The simulations were carried out in an NVT ensemble with the temperature maintained at 300 K using Langevin dynamics. During initial equilibration, NPT ensemble simulations were employed as noted, with a pressure maintained at 1 atm using a Nosé-Hoover Langevin piston [28]. The simulations used multiple timestepping, with a base timestep of 1 fs, short-range interactions calculated every step, and long-range electrostatics every 2 steps. Electrostatic forces were evaluated through the particle-mesh Ewald method [28] with a grid density of 1.0 Å^{-3}. Periodic boundary conditions were assumed for all simulations. All MD trajectories were saved once per ps. Initial equilibration of `SimA` in the NPT ensemble covered 300 ps and of `SimB` 2000 ps. The equilibrium NVT ensemble simulations used a base timestep of 2 fs reaching a total sampling time of 380 ns for PMF calculations and of 14.3 ns for calculation of the electrostatic map.

3.4.3 Electrostatic potential map

In order to evalutate the average instrinsic electrostatic map of the CD, 14.3 ns equilibrium simulations of the `SimA` set up in the NVT ensemble were performed at zero bias potential. The VMD `PMEpot` plugin [90, 102], which uses the particle-mesh Ewald method [28], was employed to compute the electrostatic potential map for every frame (2 ps) . This electrostatic potential was evaluated using a three-dimensional grid of 108 × 112 × 108 points (which ensured at least one grid point per Å in each direction), and was averaged over entire trajectories including all atoms of the system.

3.4.4 Potential of mean force

The potential of mean force (PMF) of glutamate (Glu^-) translocating through the MscS side opening was evaluated from equilibrium adaptive biasing force (ABF) simulations [31, 43–45, 103]. The ABF

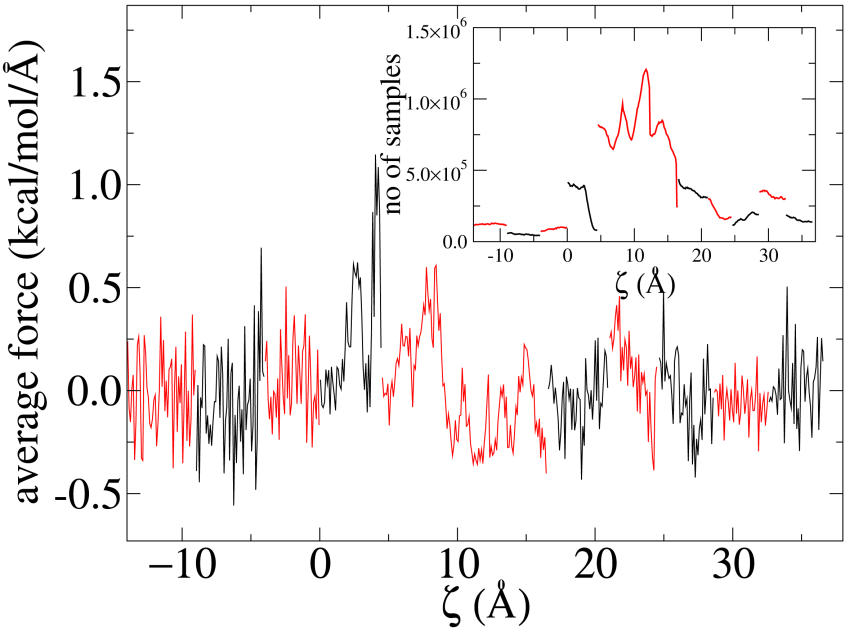

Figure 3.5: Convergence of ABF simulation. Average forces across each of the 10 windows are shown to be continuous. The graph presents in its inset the sampling distribution in each window along the ζ-axis.

method has been widely used in conjuction with molecular dynamics simulation for the evaluation of the PMF of many biologically and chemically interesting systems [45, 104–106]. In ABF simulations, the system is subject to a continuously updated biasing force designed to eventually remove all energetic barriers along a chosen coordinate and, thus, allow free diffusion of the system along that coordinate. Once the calculation converges, a PMF along the chosen coordinate is obtained through thermodynamic integration [44]. To simulate the translocation of Glu^- through one of the side openings of the MscS CD, a Glu^- ion was initially positioned on the cytoplasmic side at a distance 51 Å away from the center of the CD along the line joining the center of the CD and the center of the side opening. For the purpose of evaluating the PMF of Glu^- along the translocation axis, the reaction coordinate, ζ, is defined as the distance of Glu^- from the center of the CD. Residues 158, 143, 243 and 247 of subunits P1 through P7 were chosen to define the center of the CD as the center of mass of these residues and residues 158, 143, 243 and 247 of subunit P6, lining the side opening, to define similarly the center of the side opening. In computing the reaction coordinate we chose only the carbonyl C atoms of the residues mentioned above to avoid contamination of the measured force due to holonomic constraints [44]. To keep the Glu^- of interest from moving away from the side opening, an additional, cylindrical restraint was imposed along the translocation axis by attaching a virtual harmonic spring with spring constant 1.0 kcal mol^{-1} $Å^{-2}$ to the backbone

carbonyl C atom of this Glu^-, the constraint forces pointing towards the translocation axis passing through the middle of the CD opening. This restraint still allows the Glu^- ion to explore the whole volume of the opening.

ABF calculations were performed using 12 non-overlapping windows along the translocation axis, each of width 4-5 Å in which several ns of MD trajectory were generated. In addition, in cases where the average force was discontinuous across a bin barrier, further ABF calculations were run to convergence using the union of adjacent bins (across the three bins spanning the entire stretch of the side opening, from 5 Å to 20 Å). All PMF calculations were performed with the `SimB` set-up in the NVT ensemble assuming a Langevin damping coefficient of 1.0 ps^{-1}. Uniform sampling and continuity of average forces across the windows that demonstrate convergence of our ABF simulation is illustrated in Fig. 3.5.

3.4.5 Rate of diffusion-controlled reaction

Here we derive, following [107], the rate of diffusion controlled approach of Glu^- in the cellular cytoplasm to the CD followed by successful entry into one of the side openings of the CD. The derivation is provided for the sake of completeness as the one in [107] is extremely brief. The theory of diffusion controlled reactions is due to Debye [108] and Eigen [109].

As not every diffusive encounter between solute and CD leads to entering of an opening, the diffusive encounter is characterized as a reaction with a finite, in fact, small, reaction probability, controlled through the parameter w employed below, $w = 0$ corresponding to no reaction and $w \to \infty$ corresponding to every encounter leading to a reaction. We approximate the CD as a sphere as we are only interested in a rough estimate of the diffusion rate.

The diffusion-reaction process (reaction = entrance into an opening), described through the probability distribution $p(\vec{r}, t)$ for finding Glu^- at position $\vec{r}$ at time t near the CD, is governed by the Smoluchowski equation [110]

$$\partial_t \, p(\vec{r}, t) \;=\; \nabla \cdot D(\vec{r}) \, [\nabla \;-\; \beta \, \vec{F}(\vec{r})] \, p(\vec{r}, t) \; . \tag{3.9}$$

Here $\vec{F}(\vec{r})$ is the force acting on the solute and $D(\vec{r})$ is the local diffusion coefficient. The solution of Eq. (3.9) is subject to the boundary condition

$$\hat{n}(\vec{r}) \cdot D(\vec{r}) \, [\nabla \;-\; \beta \vec{F}(\vec{r})] \, p(\vec{r}, t) \;=\; w \, p(\vec{r}, t) \qquad \text{at } |\vec{r}| \;=\; R_o \tag{3.10}$$

for some constant w that has been introduced above [$\hat{n}(\vec{r})$ is the normal to the surface]. Assuming that the concentration, $c_{\rm S}$, of $\rm Glu^-$ in the cytoplasm at a far distance from the CD remains constant, i.e., each translocated $\rm Glu^-$ is immediately replaced at a random location in the cytoplasm, one can assume that the solution of Eq. (3.9) is time-independent, such that the Smoluchowski equation becomes

$$\nabla \cdot D(\vec{r}\,) \left[\nabla \, - \, \beta \, \vec{F}(\vec{r}) \right] p(\vec{r}\,) \; = \; 0 \; , \tag{3.11}$$

with the additional (Eq. (3.10) also holds) boundary condition

$$\lim_{|\vec{r}| \to \infty} \; p(\vec{r}\,) = \; c_{\rm S} \; . \tag{3.12}$$

This approach describes well the initial phase of $\rm Glu^-$ depletion, when $c_{\rm S}$ is still at a maximum. The later phase of depletion of $c_{\rm S}$ around the CD would have to be described through a more cumbersome, time-dependent solution of Eq. (3.9). The occurrence of an entrance of the solute into a CD opening implies that a stationary current develops which describes the diffusive approach of $\rm Glu^-$. We consider in the following the case that the interaction between CD and $\rm Glu^-$ is due to a Coulomb potential $U(r) \; = \; q_{\rm CD} q_{\rm S} / \epsilon r$ ($\epsilon =$ dielectric constant), which depends solely on the distance $|\vec{r}|$ of $\rm Glu^-$ from the CD center. We also assume that the diffusion coefficient D depends solely on $|\vec{r}|$. The stationary Smoluchowski equation (3.11) is then angle-independent and reads

$$\partial_r D(r) [\, \partial_r \, - \, \beta \, F(r) \,] \, p(r) \; = \; 0 \tag{3.13}$$

to which is associated the radial current

$$J_{\rm tot}(r) \; = \; 4 \pi r^2 D(r) [\, \partial_r \, - \, \beta \, F(r) \,] \, p(r) \tag{3.14}$$

where we have integrated over all directions $\hat{r}$ obtaining the total current, $J_{\rm tot}$, at radius r. Employing $F(r) \; = \; - \, \partial_r U(r)$, one can express Eq. (3.14) as:

$$J_{\rm tot}(r) \; = \; 4 \pi r^2 D(r) \, \exp[-\beta U(r)] \, \partial_r \, \exp[\beta U(r)] \, p(r) \; . \tag{3.15}$$

However, $J_{\rm tot}(r)$ must be the same at all r as otherwise $p(r)$ would change in time, in contrast to

the assumption that the distribution is stationary. It must hold, in particular,

$$J_{\rm tot}(R_o) \;=\; J_{\rm tot}(r)\;. \tag{3.16}$$

The boundary condition (3.10), together with (3.15), yields

$$4\pi R_o^2\, w\, p(R_o) \;=\; 4\pi r^2\, D(r)\, \exp[-\beta U(r)]\, \partial_r\, \exp[\beta U(r)]\, p(r)\;. \tag{3.17}$$

This relationship, a first order differential equation, allows one to determine $p(r)$.

For the evaluation of $p(r)$ we write (3.17)

$$\partial_r \left[e^{\beta U(r)}\, p(r) \right] \;=\; \frac{R_o^2 w}{r^2 D(r)}\, p(R_o)\, e^{\beta U(r)}\;. \tag{3.18}$$

Integration $\int_r^\infty dr \cdots$ yields

$$p(\infty)\, e^{\beta U(\infty)} \;-\; p(r)\, e^{\beta U(r)} \;=\; R_o^2 w\, p(R_o) \int_r^\infty dr' \frac{e^{\beta U(r')}}{r'^2\, D(r')} \tag{3.19}$$

or, using (3.12) and $U(\infty) = 0$,

$$p(r)\, e^{\beta U(r)} \;= c_{\rm S} \;-\; R_o^2 w\, p(R_o) \int_r^\infty dr' \frac{e^{\beta U(r')}}{r'^2\, D(r')}\;. \tag{3.20}$$

Evaluating this at $r = R_o$ and solving for $p(R_o)$ leads to

$$p(R_o) \;=\; \frac{c_{\rm S}\, e^{-\beta U(R_o)}}{1 \;+\; R_o^2 w\, e^{-\beta U(R_o)} \int_{R_o}^\infty dr\, e^{\beta U(r)}/r^2 D(r)}\;. \tag{3.21}$$

We are presently interested in the rate at which the diffusive approach to the CD and entering into the CD openings proceeds. This rate is given by $J_{\rm tot}(R_o) = 4\pi R_o^2 w\, p(R_o)$. Hence, we can state

$$\text{Rate} \;=\; \frac{4\pi R_o^2 w\, c_{\rm S}\, e^{-\beta U(R_o)}}{1 \;+\; R_o^2 w\, e^{-\beta U(R_o)} \int_{R_o}^\infty dr\, e^{\beta U(r)}/r^2 D(r)}\;. \tag{3.22}$$

This expression is proportional to $c_{\rm S}$, a dependence expected for the monomolecular reaction between solute and CD.

We define a corresponding, $c_{\rm S}$-independent rate constant k as follows

$$\text{Rate} \;=\; k(R_o, w)\, c_{\rm S} \tag{3.23}$$

where we state explicitly the dependence of k on R_o and w. This rate constant is then, in the present case,

$$k(R_o, w) \;=\; \frac{4\pi}{\int_{R_o}^{\infty} dr\, \mathcal{R}(r) \;+\; e^{\beta U(R_o)}/R_o^2 w}\;. \tag{3.24}$$

Here, we defined

$$\mathcal{R}(r) \;=\; e^{\beta U(r)}/r^2 D(r)\;. \tag{3.25}$$

We consider first the case of very ineffective reactions described by small w values. In this case the time required for the diffusive encounter of solute and CD can become significantly shorter than the time for the local reaction to proceed, if it proceeds at all. In this case it may hold

$$\frac{e^{\beta U(R_o)}}{R_o^2 w} \;>>\; \int_{R_o}^{\infty} dr\, \mathcal{R}(r) \tag{3.26}$$

and the reaction rate (3.24) becomes

$$k(R_o, w) \;=\; 4\pi R_o^2\, w\, e^{-\beta U(R_o)}\;. \tag{3.27}$$

This expression conforms to the well-known Arrhenius law.

We want to apply (3.24) and (3.25) to two cases, free diffusion ($U(r) \equiv 0$) and diffusion in a Coulomb potential ($U(r) = q_1 q_2/\epsilon r$). We assume in both cases a distance-independent diffusion constant. In case of free diffusion holds $\mathcal{R}(r) = D^{-1} r^{-2}$ and, hence,

$$\int_{R_o}^{\infty} dr\, \mathcal{R}(r) \;=\; 1/DR_o\;. \tag{3.28}$$

From this results

$$k(R_o, w) \;=\; \frac{4\pi D R_o}{1 \;+\; D/R_o w}\;. \tag{3.29}$$

In case of very effective reactions, i.e., for very large w, this becomes

$$k(R_o, w) \;=\; 4\pi D R_o \tag{3.30}$$

which is the well-known rate constant for diffusion-controlled reaction processes. In case of a Coulomb interaction between Glu^- and CD one obtains

$$\int_{R_o}^{\infty} dr\, \mathcal{R}(r) \;=\; \frac{1}{D} \int_{R_o}^{\infty} dr\, \frac{1}{r^2}\, \exp\left[\frac{\beta q_1 q_2}{\epsilon r}\right] \;=\; \frac{1}{R_L D}\left(e^{R_L/R_o} \;-\; 1\right) \tag{3.31}$$

where

$$R_L \;=\; \beta q_{\rm CD} q_{\rm S} \,/\, \epsilon \tag{3.32}$$

defines the so-called Onsager radius. Note that R_L can be positive or negative, depending on the sign of $q_1 q_2$, but that the integral over $\mathcal{R}(r)$ (3.31) is always positive. The rate constant (3.24) can then be written

$$k(R_o, w) \;=\; \frac{4\pi D R_L}{(R_L D / R_o^2 w)\, e^{R_L/R_o} \;+\; e^{R_L/R_o} \;-\; 1}\,. \tag{3.33}$$

3.4.6 Quantum yield of diffusion controlled reaction

One can express the rate constants derived above in the form

$$k(R, w) \;=\; k(R', \infty)\; \phi(R', R, w) \tag{3.34}$$

where, according to the definition above, $k(R', \infty)$ is the rate constant for the case that every encounter at R' leads to a reaction and where $\phi(R', R, w)$ is then the probability, referred to as the quantum yield, that for a finite w the reaction actually occurs at R when diffusion started at R'. From (3.24) follows

$$\phi(R', R, w) \;=\; \frac{\int_{R'}^{\infty} dr\, \mathcal{R}(r)}{e^{\beta U(R)} / R^2 w \;+\; \int_{R}^{\infty} dr\, \mathcal{R}(r)} \tag{3.35}$$

In the case of free diffusion and when $R' \;=\; R$ holds, using (3.28), we obtain

$$\phi(R' = R, R, w) \;=\; \frac{1}{D\,/Rw \;+\; 1}\,. \tag{3.36}$$

Chapter 4

Conclusions and Outlook

The simulation results presented in this thesis have provided unique insights into 1) the structure of the interior of the central channel of NPC that facilitates gating for selective transport of molecules in and out of the nucleus, and 2) the filter function of MscS' cytoplasmic domain to preserve membrane potential in bacterium during osmotic shocks. In this chapter, I will summarize the results of these studies in the context of the field as a whole and also describe future directions of research.

4.1 Gating by the nuclear pore complex

The unstructured, FG-rich nups that makes the central transport channel, apparently plays a role in gating of the NPC. We sought to understand how these unstructured FG-nups are arranged in the interior of the NPC such that large complexes are efficiently and selecively transported in and out of the nucleus.

We used coarse-grained molecular dynamics to study assemblies of one kind of nup, namely, Nsp1 (an FXFG rich nup present in the central channel; x is any amino acid) starting from different initial conformations and geometrical arrangements. Three different geomentries of the Nsp1 systems were studied: 1) initially fully-extended, straight chain conformations for Nsp1s, end-tethered in a ring like geometry; 2) initially random conformations for Nsp1s in an array-like arrangement; and 3) initially random conformations but freely floating Nsp1s in a bath of water. Nsp1s collectively formed brush-like structures with bristles made of bundles of few to several nups. The bundles are cross-linked through single nups leaving one bundle and joining a nearby one. The cross-linked bundle structure is independent of the initial confomations and spatial arrangement of nups. However, the degree of cross-linking varies for different initial nup conformations and arrangement. The structures described are FG-FG interaction independent suggesting that FGs are readily available and that act mainly as binding sites for transport factors.

We proposed a model for gating by the NPC based on the resulting structures. The observations

support the central channel featuring brush-like bundles near the periphery and a sieve-like mesh structure arising in the central region, as the free ends of unstructured nups, protruding to the volume interior may have higher degree of cross-linking due to crowding. The increased cross-linking between bundles also sets the size limit of <9 nm known for passive diffusion of molecules through the channel. Larger biomolecules need to be bound by transport factor to pass through nuclear pores; the transport factor must have the ability to melt the cross-linking between nups bundles creating wide enough space for their own and their cargo to pass.

How the TFs melt the structural assembly to penetrate through, this mechanism remains unanswered and is particularly exciting aspect of future studies.

4.2 Filter function of the MscS' cytoplasmic domain

MscS' transmembrane channel is covered on the inside of the cell by a balloon-like filter (80 Å in diameter), the cytoplasmic domain (CD) that makes up actually the major part of the whole channel in terms of atom count (65% of the total mass of the protein) . Given that the balloon is such a large piece of the channel, it must be important. In this study, we focus on the function and mechanism of the baloon-like cytoplasmic domain.

We noticed that the balloon has seven holes through which osmolytes and water can leave the cell, passing first through the balloon and then through the transmembrane channel coming behind it. We first thought that the side openings in the balloon-like filter, select some osmolytes over others, but that was not so: the free energy profile of Glu- and K+ making it through the openings having a low barrier showed in fact the openings let anything small enough pass through in no time as the resulting mean passage time through the openings turned out to be much shorter than typical open times of MscS.

These openings are very small (7 Å) and osmolytes need to find them. Using a mathematical description, we deduced an estimate for the time for the diffusion controlled approach of Glu- and K+ to the MscS CD and to the side opening and observed that the solutes donot reach the openings in the first place. The positive ions adhere to the balloon surface and always find sooner or later a hole to pass through. The negative ions are repelled by the balloon surface and have a hard time finding a hole to pass through.

We found that the balloon counteracts the cell electrical potential. The transmembrane potential (-100 mv) favors negative osmolytes to leave the cell as is also observed by channel exit times is

faster for negative Glu- than for the K+, but the balloon favors positive ions.

From the total translocation times described from the diffusion of solutes toward the CD requiring a time to find a side opening, passage of solute through the openings and finally through the transmembrane channel, for Glu- and K+, we conclude the filter function of the CD. So on balance, positive and negative ions leave the cell at the same overall speed, that means, they leave the cell together.

We suggest, therefore, the large cytoplasmic domain is a filter which is required by the cell as the cell wants to make really sure that positive and negative osmolytes leave the cell together. It uses the big balloon as a mixing bowl for positive and negative osmolytes such that they are in 1:1 proportion. We evaluated the penalty paid in terms of electrostatic energy which is a few kbT for every additional type of ion present. And, the large ballon can hold about twenty pair such that when they leave the balloon, they leave together in equal numbers of positives and negatives.

The balloon thus solves two problems at once: it lets osmolytes leave the cell to reduce pressure, but lets only an equal number of positive and negative osmolytes leave, so that the cell does not drain its electrical battery.

A particularly exciting aspect of future studies should be ion-ion interaction effects in the CD that favor neutral efflux and preserve cell polarity.

Appendix A

Free-energy perturbation calculation

To examine the diffusion of K^+ from the bulk aqueous environment to the interior of the CD, a PMF could be generated in the spirit of the ABF calculation endeavored for Glu^-. However, this route has proven (for the case of Glu^-) to be computationally costly, in particular when, in the case of potassium, one is interested only at milestones along the reaction pathway — viz. points A, C and E, rather than the entire free-energy landscape. Accordingly, we resorted to an alternative numerical scheme for the K^+ PMF, namely, to calculate the free-energy change incurred in the replacement of Glu^- by K^+. Free-energy perturbation [111] (FEP) is particularly well suited to tackle this problem, wherein the anion located along the reaction pathway used in the computation of the PMF is mutated reversibly into its cationic counterpart, as depicted in Figure SA.1.
In the thermodynamic cycle depicted in Figure SA.1, in the left vertical transition Glu^- is transformed alchemically into K^+, both in the free state, i.e., in a 200 mM potassium glutamate bath in the absence of the CD. In case of the right side vertical transition the alchemical transformation occurs in association with CD. Our computational strategy involved calculating simultaneously the energy differences of the two vertical transitions, namely, $\Delta G^2_{\text{mut.}} - \Delta G^2_{\text{mut.}}$ and to calculate from this, through the identity $\Delta G^1_{\text{assoc.}} - \Delta G^2_{\text{assoc.}} = \Delta G^1_{\text{mut.}} - \Delta G^2_{\text{mut.}}$, the energy $\Delta G^1_{\text{assoc.}} - \Delta G^2_{\text{assoc.}}$ that yields the desired energy difference between Glu^- and K^+ in the side opening pathway. To increase the cost-effectiveness of the calculation of $\Delta G^2_{\text{mut.}} - \Delta G^2_{\text{mut.}}$, the two point mutations were performed concomitantly, the glutamate ion localized on the reaction pathway being replaced by its cationic counterpart, whilst symmetrically, a potassium ion lying sufficiently far away from the CD was transmuted into the corresponding anionic species. Aside from obviating the need to run separate simulations, concurrent alchemical transformation of a potassium and a glutamate ion in the same periodic cell circumvents size-consistency issues rooted in an incomplete treatment of long-range charge-dipole interactions [112], size-dependence corrections being mutually compensated here.
The position of the glutamate ion was restrained harmonically at points A and E as it was

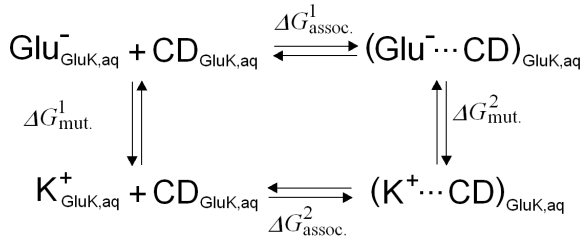

Figure A.1: Thermodynamic cycle utilized to measure the free-energy change upon replacement by potassium of a glutamate ion. The horizontal legs of the cycle delineate the association of the ion, either glutamate or potassium, with the CD. The association free energy, $\Delta G^1_{\text{assoc.}}$, characterized by the upper, horizontal leg can be inferred from the PMF by integration of the latter [113]. The vertical legs correspond to the point mutation of the anion in the free and in the bound states, so that $\Delta G^1_{\text{assoc.}} - \Delta G^2_{\text{assoc.}} = \Delta G^1_{\text{mut.}} - \Delta G^2_{\text{mut.}}$.

point	ΔG (kcal/mol)	number of strata	time per stratum (ns)	total time (ns)
A	+0.2±0.3	100	0.096	9.60
C	+4.6±0.4	120	0.128	15.36
E	+0.2±0.4	100	0.096	9.60

Table A.1: Stratification strategy and net free-energy change for the alchemical transformations into potassium of a glutamate ion located at three milestone positions of the potential of mean force characterizing the diffusion of the anion in and out of the CD. A potassium ion lying sufficiently far away from the CD is mutated concurrently into glutamate following the thermodynamic cycle of Figure A.1.

progressively transformed alchemically into the potassium ion. Conversely, at point C, the anion was free to diffuse radially in a thin, circular section, the radius of which was that of the pore. The width of this section is equal to that of the bin used in the ABF calculation to accrue the instantaneous force acting along the reaction coordinate. Point mutations of the glutamate ion into the potassium ion was carried out in the framework of the dual-topology paradigm [114], wherein the reference and the target states of the transformation coexist without interacting mutually. The reaction pathway connecting these two states was stratified by means of 100 to 120 intermediate states, or strata, wherein up to 0.128 ns of sampling was performed, as summarized in Table A.1. On account of the additional configurational entropy, convergence of the ensemble average required appreciably longer simulation time at point C than at points A and E. In each stratum, the first fourth of the sampling consisted of a thermalization of the molecular system prior to data collection. To alleviate singularities in the Lennard-Jones potential upon creation of the ions, either glutamate or potassium, a separation-shifted scaling scheme was employed [115], and van der Waals interactions were turned on prior to their electrostatic counterpart. The statistical error reflecting the precision of the free-energy calculation [116] was determined from the variance of the ensemble average. [103, 117] The correlation length of the time series was measured using the Flyvbjerg-Petersen renormalization group blocking method [118].

Appendix B

Electrostatic energy of a "gas" of N_+ positive and N_- negative charges confined in a sphere at constant density

In the following we will demonstrate that the CD favors harboring an equal number of positive and negative osmolytes. We assume that N_+ positive and N_- negative ions are held inside the cytoplasmic domain at constant density, i.e., without correlation due to the mutual Coulomb attraction and repulsion or due to excluded volume effects. We define $N_+ = N + n$ and $N_- = N - n$ and we seek to determine the total Coulomb energy in case of a spherical cavity of radius R_o

$$E(N, n, R_o) \;=\; (\rho_{++} \;+\; \rho_{--} \;-\; \rho_{+-})\; I \tag{B.1}$$

where

$$\rho_{++} \;=\; \frac{N_+\,(N_+ \,-\, 1)}{2}\,\rho_o^2 \tag{B.2}$$

$$\rho_{--} \;=\; \frac{N_-\,(N_- \,-\, 1)}{2}\,\rho_o^2 \tag{B.3}$$

$$\rho_{+-} \;=\; N_+\,N_-\,\rho_o^2 \tag{B.4}$$

$$\rho_o \;=\; \frac{3\,e}{4\,\pi\,R_o^3} \tag{B.5}$$

and

$$I \;=\; \int d\vec{r}_1 \int d\vec{r}_2 \frac{1}{|\vec{r}_1 \,-\, \vec{r}_2|} \tag{B.6}$$

I can be written $I = I_> + I_< + I_=$ where

$$I_> \;=\; \iint_{r_1 > r_2} \frac{1}{|\vec{r}_1 \,-\, \vec{r}_2|}\, d\vec{r}_1\, d\vec{r}_2 \tag{B.7}$$

and similarly for $I_<$. $I_=$ is the sum over all Coulomb energies for $r_1 = r_2$. One can expand, following well known procedures [119],

$$I_> \;=\; \iint_{r_1>r_2} d\vec{r}_1\, d\vec{r}_2\; \frac{1}{r_1} \sum_\ell \left(\frac{r_2}{r_1}\right)^\ell P_\ell(\cos\gamma) \tag{B.8}$$

and

$$I_= \;=\; \iint_{r_1=r_2=r} d\vec{r}_1\, d\vec{r}_2\; \frac{1}{r_1} \sum_\ell \left(\frac{r_2}{r_1}\right)^\ell P_\ell(\cos\gamma) \tag{B.9}$$

where P_ℓ denotes the ℓ-th Legendre polynomial and γ the angle between $\vec{r}_1$ and $\vec{r}_2$. According to the addition theorem of spherical harmonics, $Y_{\ell m}$, one can write

$$P_\ell(\cos\gamma) \;=\; \frac{4\pi}{2\ell+1} \sum_m Y^*_{\ell m}(\theta_1,\phi_1)\, Y_{\ell m}(\theta_2,\phi_2)\;; \tag{B.10}$$

here angles θ_j, ϕ_j define the direction of $\vec{r}_j$. The orthogonality relationships of the spherical harmonics lead to all terms in the sums to vanish, except the term with $\ell = 0$ and $m = 0$, such that one obtains, noting $Y_{00} = 1/\sqrt{4\pi}$,

$$I_> \;=\; 4\pi \iint_{r_1>r_2} r_1^2 dr_1\; r_2^2 dr_2\; \frac{1}{r_1}\,. \tag{B.11}$$

$$I_= \;=\; 4\pi \iint_{r_1=r_2=r} r_1^2 dr_1\; r_2^2 dr_2\; \frac{1}{r_1}\,. \tag{B.12}$$

One can readily derive

$$I_> \;=\; 4\pi R_o^5/15 \tag{B.13}$$

$I_=$ is negligible compared to $I_>$ when R_o is large, and also $I_> = I_<$. Altogether, one obtains finally

$$E(N,n,R_o) \;=\; (2\,n^2 \,-\, N)\,\frac{3\,e^2}{10\,\pi\,R_o}\,. \tag{B.14}$$

The expression shows that the lowest energy arises for an equal number of positive and negative ions. The actual energy increase experienced in case of changing n from zero to one is $3e^2/5\,\pi\,R_o$, but would be scaled down by the effective dielectric constant, ϵ, in the cytoplasmic domain. For $\epsilon = 1$ and $R_o = 43$ Å, the energy would be a few k_BT. In reality, it should be less due to dielectric screening; however, the volume inside the CD available to ions might be characterized by a smaller R_o value; also the electrostatic energy in the CD would increase quadratically with n (excess number of positive over negative ions) and, hence, the interaction would likely favor

significantly a balance of positive and negative ions in the cytoplasmic domain. Such balance would also be strengthened through correlation effects, e.g., through Glu^{-}-K^{+} pairing.

Appendix C

Calculation of the transmembrane conduction time τ_3

In the following we will explain the mathematical expression stated in Eq. (36) of the main text. τ_3 describes how long it takes ions to leave the cytoplasmic domain through the open transmembrane channel to reach the extracellular space. Ion motion is controlled to a large degree by the cellular potential $U(x)$ which assumes a value V_o inside ($x < a$) the cell, zero outside ($x > a$) the cell and is close to constant in the cell interior. The potential varies from V_o to zero across the inner cellular membrane with a center position $x = a$. We assume here a sigmoidal variation of $U(x)$ inside the membrane with width σ. $U(x)$ is described through the function

$$U(x) \;=\; \frac{V_o}{2} \;-\; \frac{V_o}{2}\,\tanh[(x\,-\,a)/\sigma] \tag{C.1}$$

We assume also that the x-axis points along the center of the MscS channel in which case holds a = -16.86 Å and σ = 6 Å, the values having been determined through matching of expression (C.1) to the cellular potential calculated in [80]. For the cell potential we assume V_o = -100 mV. We note that in applying $U(x)$ to Glu^- or K^+, one needs to multiply the potential with the sign of the ion charge, i.e., Glu^- experiences the potential $-U(x)$ as defined above, while K^+ experiences the potential $+U(x)$.

To determine τ_3 we will employ the mean first passage time description, as adopted also for calculation of τ_2. However, we need to account here for the narrowing of the cytoplasmic domain towards the transmembrane channel erecting an entropy barrier. To account for this barrier, at least qualitatively, we assume that diffusion orthogonal to the x-axis is more rapid than along the x-axis such that we can assume the distribution function for an ion, $p(x, y, z, t)$ to assume the form $\tilde{p}(x, t)\bar{p}_o(x|y, z)$ where $p_o(x|y, z)$ is the time-independent equilibrium distribution along y and z for fixed x. Integrating $\bar{p}(x, y, z, t)$ over the y- and z-coordinates yields $\bar{p}(x, t) \;=\; \tilde{p}(x, t)\, Z(x)$, where $Z(x)$ is the partition function in the y, z-plane. We assume here also that $\tilde{p}(x, t)$ corresponds to Brownian motion in the potential $U(x)$ stated above, i.e., at equilibrium $\tilde{p}(x, t) \;\propto\; \exp[-\beta U(x)]$. It

should then hold $p(x, t \to \infty) \sim \exp[-\beta U(x)]\, Z(x) = \exp[-\beta \tilde{U}(x)]$ where

$\tilde{U}(x) = U(x) - k_{\mathrm{B}}\mathrm{T} \ln Z(x)$.

The mean first passage time for the ions moving in the free energy $\tilde{U}(x)$ is then, according to Szabo et al. [95],

$$\tau_3 = \int_{x_{\mathrm{center}}}^{x_{\mathrm{out}}} dx D^{-1}(x) \exp[\beta \tilde{U}(x)] \int_{x_{\mathrm{apex}}}^{x} dx' \exp[-\beta \tilde{U}(x')] \tag{C.2}$$

Here x_{center} denotes the center of the cytoplasmic domain, x_{out} a position just outside of the transmembrane channel, i.e., just in the extracellular space, and x_{apex} the position at the apex of the cytoplasmic domain, i.e., the position inside the cytoplasmic domain widest from the transmembrane channel. Expression (C.2) can be employed to determine τ_3 once $Z(x)$ is specified. In the present case, $Z(x)$ is just the area of the cross section of the cytoplasmic domain and channel at point x along the x-axis, i.e., $Z(x) = \pi(r^2(x)/r_0^2)$, where $r(x)$ is the radius of the cross section along the symmetry axis, x, of the MscS and r_0 is a reference radius, the value of which drops out of the evaluation. We assume here that there is no obstacle impeding diffusion away or towards the x-axis.

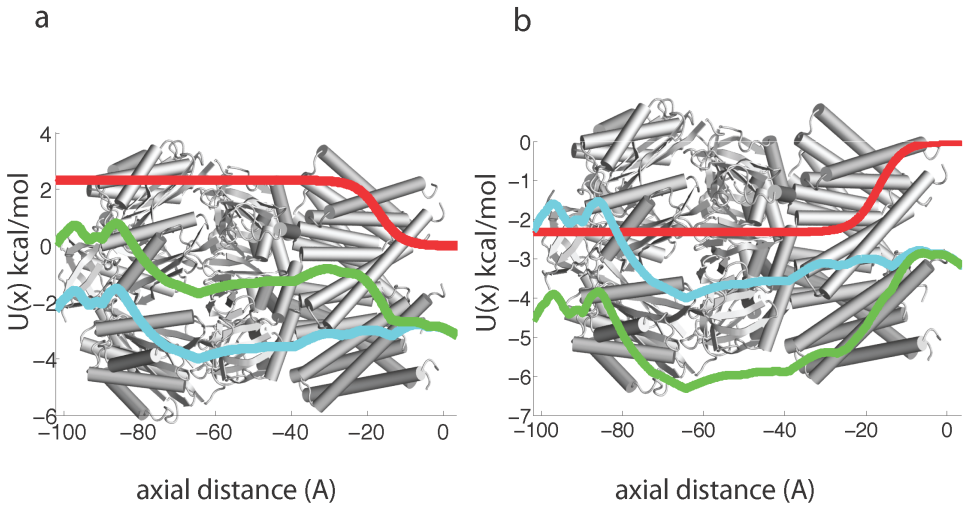

Figure C.1: Comparision between the enthalpic and total potential for ions permeating along the symmetry axis of MscS. *(a)* Shown in red is the enthalpic potential $U(x)$, in blue $-k_{\mathrm{B}}\mathrm{T} \ln Z(x)$, and in green $\tilde{U}(x)$ for Glu^- permeating the transmembrane pore along the symmetry axis. *(b)* The corresponding potentials for K^+. MscS in its open conformation (PDB:2VV5) is shown in gray in both *(a)* and *(b)*.

Figure C.1 compares the potentials $U(x)$, $-k_{\mathrm{B}}\mathrm{T} \ln Z(x)$ and $\tilde{U}(x)$ for Glu^- and K^+, showing that due to the presence of the spherical cytoplasmic domain the effective potential $\tilde{U}(x)$ develops a

minimum at the domain center with a depth of about 4.0 k_{B}T. The origin of the minimum is the wider y, z-space available near the CD center than near the CD apex or the transmembrane channel.

References

[1] Rout, M. P. & Blobel, G. (1993). Isolation of the yeast nuclear pore complex. *J. Cell Biol.*, **123**, 771–783.

[2] Devos, D., Dokudovskaya, S., Williams, R., Alber, F., Eswar, N., Chait, B., Rout, M. & Sali, A. (2006). Simple fold composition and modular architecture of the nuclear pore complex. *Proc. Natl. Acad. Sci. USA*, **103**, 2172–7.

[3] Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y. & Chait, B. T. (2000). The yeast nuclear pore complex: Composition, architecture, and transport mechanism. *J. Cell Biol.*, **148**, 635–651.

[4] Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L. & Rexach, M. (2003). Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl. Acad. Sci. USA*, **100**, 2450–2455.

[5] Lim, R., Huang, N., Köser, J., Deng, J., Lau, K., Schwarz-Herion, K., Fahrenkrog, B. & Aebi, U. (2006). Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc. Natl. Acad. Sci. USA*, **103**, 9512–7.

[6] Denning, D. & Rexach, M. (2007). Rapid evolution exposes the boundaries of domain structure and function in natively unfolded FG nucleoporins. *Mol. Cell Proteomics*, **6**, 272–282.

[7] Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Sali, A. & Rout, M. P. (2007). The molecular architecture of the nuclear pore complex. *Nature*, **450**, 695–701.

[8] Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Rout, M. P. & Sali, A. (2007). Determining the architectures of macromolecular assemblies. *Nature*, **450**, 683–694.

[9] Rexach, M. & Blobel, G. (1995). Protein import into nuclei: Association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. *Cell*, **83**, 683–692.

[10] Liu, S. M. & Stewart, M. (2005). Structural basis for the high-affinity binding of nucleoporin Nup1p to the *Saccharomyces cerevisiae* importin-β homologue, Kap95p. *J. Mol. Biol.*, **349**, 515–525.

[11] Cushman, I., Bowman, B., Sowa, M., Lichtarge, O., Quiocho, F. & Moore, M. (2004). Computational and biochemical identification of a nuclear pore complex binding site on the nuclear transport carrier NTF2. *J. Mol. Biol.*, **344**, 303–10.

[12] Morrison, J., Yang, J., Stewart, M. & Neuhaus, D. (2003). Solution NMR study of the interaction between NTF2 and nucleoporin FxFG repeats. *J. Mol. Biol.*, **333**, 587–603.

[13] Bednenko, J., Cingolani, G. & Gerace, L. (2003). Importin-β contains a COOH-terminal nucleoporin binding region important for nuclear transport. *J. Cell Biol.*, **162**, 391–401.

[14] Bayliss, R., Littlewood, T., Strawn, L. A., Wente, S. R. & Stewart, M. (2002). GLFG and FxFG nucleoporins bind to overlapping sites on importin-β. *J. Biol. Chem.*, **277**, 50597–50606.

[15] Bayliss, R., Leung, S. W., Baker, R. P., Quimby, B. B., Corbett, A. H. & Stewart, M. (2002). Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats. *EMBO J.*, **21**, 2843–2853.

[16] Strawn, L. A., Shen, T. & Wente, S. R. (2001). The GLFG regions of Nup116p and Nup100p serve as binding sites for both Kap95p and Mex67p at the nuclear pore complex. *J. Biol. Chem.*, **276**, 6445–6452.

[17] Isgro, T. A. & Schulten, K. (2005). Binding dynamics of isolated nucleoporin repeat regions to importin-β. *Structure*, **13**, 1869–1879.

[18] Isgro, T. A. & Schulten, K. (2007). Association of nuclear pore FG-repeat domains to NTF2 import and export complexes. *J. Mol. Biol.*, **366**, 330–345.

[19] Isgro, T. A. & Schulten, K. (2007). Cse1p binding dynamics reveal a novel binding pattern for FG-repeat nucleoporins on transport receptors. *Structure*, **15**, 977–991.

[20] Miao, L. & Schulten, K. (2010). Probing a structural model of the nuclear pore complex channel through molecular dynamics. *Biophys. J.*, **98**, 1658–1667.

[21] Berrier, C., Coulombe, A., Houssin, C. & Ghazi, A. (1989). A patch-clamp study of ion channels of inner and outer membranes and of contact zones of *E. coli*, fused into giant liposomes. *FEBS LETTERS*, **259**, 27–32.

[22] Berrier, C., Besnard, M., Ajouz, B. & Ghazi, A. (1996). Multiple mechanosensitive ion channels from *E. coli*, activated at different thresholds of applied pressure. *J. Membr. Biol.*, **151**, 175–187.

[23] Hamill, O. P. & Martinac, B. (2001). Molecular basis of mechanotransduction in living cells. *Physiol. Rev.*, **81**, 685–740.

[24] Martinac, B., Buechner, M., Delcour, A. H., Adler, J. & Kung, C. (1987). Pressure-sensitive ion channel in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **84**, 2297–2301.

[25] Levina, N., Totemeyer, S., Stokes, N. R., Louis, P., Jones, M. A. & Booth, I. (1999). Protection of *Escherichia coli* cells against extreme turgor by activation of MscS and MscL mechanosensitive channels: identification of genes required for MscS activity. *EMBO J.*, **18**, 1730–1737.

[26] Ajouz, B., Berrier, C., Garrigues, A., Besnard, M. & Ghazi, A. (1998). Release of thioredoxin via the mechanosensitive channel MscL during osmotic downshock of *Escherichia coli* cells. *J. Biol. Chem.*, **273**, 26670–26674.

[27] Berrier, C., Garrigues, A., Richarme, G. & Ghazi, A. (2000). Elongation factor tu and dnak are transferred from the cytoplasm to the periplasm of *Escherichia coli* during osmotic downshock presumably via the mechanosensitive channel mscl. *J. Bacteriol.*, **182**, 248–251.

[28] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J. Comp. Chem.*, **26**, 1781–1802.

[29] MacKerell, Jr., A. D., Bashford, D., Bellott, M., Dunbrack, Jr., R. L., Evanseck, J., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, I. W. E., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.

[30] Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comp. Chem.*, **13**, 1011–1021.

[31] Darve, E. & Pohorille, A. (2001). Calculating free energies using average force. *J. Chem. Phys.*, **115**, 9169–9183.

[32] Laio, A. & Parrinello, M. (2002). Escaping free energy minima. *Proc. Natl. Acad. Sci. USA*, **99**, 12562–12566.

[33] Marrink, S. J., de Vries, A. H. & Mark, A. E. (2004). Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B*, **108**, 750–760.

[34] Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. (2007). The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, **111**, 7812–7824.

[35] Monticelli, L., Kandasamy, S., Periole, X. & Marrink, R. L. D. T. S. (2008). The MARTINI coarse grained forcefield: Extension to proteins. *J. Chem. Theor. Comp.*, **4**, 819–834.

[36] Arkhipov, A., Freddolino, P. L. & Schulten, K. (2006). Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure*, **14**, 1767–1777.

[37] Roux, B. & Simonson, T. (1999). Implicit solvent models. *Biophys. Chem.*, **78**, 1–20.

[38] Sugita, Y. & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, **314**, 141–151.

[39] MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, J. R. L., Evanseck, J., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1992). Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. *FASEB J.*, **6**, A143–A143.

[40] MacKerell, Jr., A. D., Bashford, D., Bellott, M., Dunbrack, Jr., R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, I. W. E., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.

[41] MacKerell, Jr., A. D., Feig, M. & Brooks, III, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comp. Chem.*, **25**, 1400–1415.

[42] Weiner, S. P., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, J. & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Appl. Cryst.*, **106**, 765–784.

[43] Darve, E., Wilson, D. & Pohorille, A. (2002). Calculating free energies using a scaled-force molecular dynamics algorithm. *Mol. Sim.*, **28**, 113–144.

[44] Hénin, J. & Chipot, C. (2004). Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.*, **121**, 2904–2914.

[45] Henin, J., Schulten, K. & Chipot, C. (2006). Conformational equilibrium in alanine-rich peptides probed by reversible stretching simulations. *J. Phys. Chem. B*, **110**, 16718–16723.

[46] Reichelt, R., Holzenburg, A., E. L. Buhle, J., Jarnik, M., Engel, A. & Aebi, U. (1990). Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components. *J. Cell Biol.*, **110**, 883–894.

[47] Yang, Q., Rout, M. P. & Akey, C. W. (1998). Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol Cell*, **1**, 223–34.

[48] Cronshaw, J., Krutchinsky, A., Zhang, W., Chait, B. & Matunis, M. (2002). Proteomic analysis of the mammalian nuclear pore complex. *J. Cell Biol.*, **158**, 915–927.

[49] Macara, I. G. (2001). Transport into and out of the nucleus. *Microbiol. Mol. Biol. Rev.*, **65**, 570–594.

[50] Rout, M. P., Aitchison, J. D., Magnasco, M. O. & Chait, B. T. (2003). Virtual gating and nuclear transport: the hole picture. *Trends Cell. Biol.*, **13**, 622–628.

[51] Miao, L. & Schulten, K. (2009). Transport-related structures and processes of the nuclear pore complex studied through molecular dynamics. *Structure*, **17**, 449–459.

[52] Lim, R. Y. H., Fahrenkrog, B., Koser, J., Schwarz-Herion, K., Deng, J. & Aebi, U. (2007). Nanomechanical basis of selective gating by the nuclear pore complex. *Science*, **318**, 640–643.

[53] Ribbeck, K. & Görlich, D. (2002). The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. *EMBO J.*, **21**, 2664–2671.

[54] Frey, S., Richter, R. & Görlich, D. (2006). FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science*, **314**, 815–7.

[55] Frey, S. & Görlich, D. (2007). A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell*, **130**, 512–523.

[56] Peters, R. (2005). Translocation through the nuclear pore complex: Selectivity and speed by reduction-of-dimensionality. *Traffic*, **6**, 421–427.

[57] Patel, S. S., Belmont, B. J., Sante, J. M. & Rexach, M. F. (2007). Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell*, **129**, 83–96.

[58] Yamada, J., Phillips, J. L., Patel, S., Goldfien, G., Calestagne-Morelli, A., Huang, H., Reza, R., Acheson, J., Krishnan, V. V., Newsam, S., Gopinathan, A., Lau, E. Y., Colvin, M. E., Uversky, V. N. & Rexach, M. F. (2010). A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in fg nucleoporins. *Mol. Cell Proteomics*, **9**, 2205–22204.

[59] Tagliazucchi, M., Peleg, O., Kroger, M., Rabin, Y. & Massimo, I. S. (2013). Effect of charge, hydrophobicity, and sequence of nucleoporins on the translocation of model particles through the nuclear pore complex. *Proc. Natl. Acad. Sci. USA*, **110**, 3363–3368.

[60] Peleg, O. & Lim, R. Y. H. (2010). Converging on the function of intrinsically disordered nucleoporins in the nuclear pore complex. *Biol. Chem.*, **391**, 719–730.

[61] Yang, W. (2011). 'natively unfloded' nucleoporins in nucleocytoplasmic transport: Clustered or evenly distributed ? *Nucleus*, **2**, 10–16.

[62] Moussavi-Baygi, R., Jamali, Y., Karimi, R. & Mofrad, M. R. K. (2011). Brownian dynamics simulation of nucleocytoplasmic transport: A coarse-grained model for the functional state of the nuclear pore complex. *PLoS Comput. Biol.*, **7**, e1002049.

[63] Schoch, R. L., Kapinos, L. E. & Lim, R. (2012). Nuclear transport receptor binding avidity triggers a self-healing collapse transition in fg-nucleoporin molecular brushes. *Proc. Natl. Acad. Sci. USA*, **109**, 16911–16.

[64] Doi, M. & Edwards, S. F. (1999). *The Theory of Polymer Dynamics*. Oxford University Press, New York.

[65] Flory, P. J. (1953). *Principles of polymer chemistry*. Cornell University Press.

[66] Kyrychenkoa, A., Karpushinaa, G. V., Bogatyrenkoa, S. I., Kryshtala, A. P. & Doroshenkoa, A. O. (2011). Preparation, structure, and a coarse-grained molecular dynamics model for dodecanethiol-stabilized gold nanoparticles. *Comput. Theor. Chem.*, **977**, 34–39.

[67] Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.*, **103**, 4613–4621.

[68] Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C., eds. (2001). *Introduction to Algorithms, 2nd edition*. MIT Press, Cambridge, MA.

[69] Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, **1**, 269–271.

[70] Gamini, R., Sotomayor, M., Chipot, C. & Schulten, K. (2011). Cytoplasmic domain filter function in the mechanosensitive channel of small conductance. *Biophys. J.*, **101**, 80–89.

[71] Sukharev, S. I., Sigurdson, W. J., Kung, C. & Sachs, F. (1999). Energetics and spatial parameters for gating of the bacterial large conductance mechanosensitive channel, MscL. *J. Gen. Physiol.*, **113**, 525–539.

[72] Bass, R. B., Strop, P., Barclay, M. & Rees, D. C. (2002). Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science*, **298**, 1582–1587.

[73] Steinbacher, S., Bass, R. B., Strop, P. & Rees, D. C. (2007). Structures of the prokaryotic mechanosensitive channels MscL and MscS. In O. P. Hamill, ed., *Current Topics in Membranes in Mechanosensitive Ion Channels, Part A*, volume 58, 1–24. Elsevier Inc., New York, NY.

[74] Wang, W., Black, S. S., Edwards, M. D., Miller, S., Morrison, E. L., Bartlett, W., Dong, C., Naismith, J. H. & Booth, I. R. (2008). The structure of an open form of an *E. coli* mechanosensitive channel at 3.45 Å resolution. *Science*, **321**, 1179–1183.

[75] Edwards, M. D., Li, Y., Miller, S., Barlett, W., Black, S., Dennison, S., Iscla, I., Blount, P., Bowie, J. U. & Booth, I. R. (2005). Pivotal role of the glycine-rich TM3 helix in gating the MscS mechanosensitive channel. *Nat. Struct. Mol. Biol.*, **12**, 113–119.

[76] Sotomayor, M. & Schulten, K. (2004). Molecular dynamics study of gating in the mechanosensitive channel of small conductance MscS. *Biophys. J.*, **87**, 3050–3065.

[77] Sotomayor, M., van der Straaten, T. A., Ravaioli, U. & Schulten, K. (2006). Electrostatic properties of the mechanosensitive channel of small conductance MscS. *Biophys. J.*, **90**, 3496–3510.

[78] Vora, T., Corry, B. & Chung, S.-H. (2006). Brownian dynamics investigation into the conductance state of the MscS channel crystal structure. *Biochim. Biophys. Acta Biomembr.*, **1758**, 730–737.

[79] Spronk, S., Elmore, D. E. & Dougherty, D. A. (2006). Voltage-dependent hydration and conduction properties of the hydrophobic pore of the mechanosensitive channel of small conductance. *Biophys. J.*, **90**, 3555–3569.

[80] Sotomayor, M., Vasquez, V., Perozo, E. & Schulten, K. (2007). Ion conduction through MscS as determined by electrophysiology and simulation. *Biophys. J.*, **92**, 886–902.

[81] Akitake, B., Anishkin, A. & Sukharev, S. (2007). Straightening and sequential buckling of the pore-lining helices define the gating cycle of MscS. *Nat. Struct. Mol. Biol.*, **14**, 1141–1149.

[82] Vasquez, V., Sotomayor, M., Cortes, D. M., Roux, B., Schulten, K. & Perozo, E. (2008). Three dimensional architecture of membrane-embedded MscS in the closed conformation. *J. Mol. Biol.*, **378**, 55–70.

[83] Vasquez, V., Sotomayor, M., Cordero-Morales, J., Schulten, K. & Perozo, E. (2008). A structural mechanism for MscS gating in lipid bilayers. *Science*, **321**, 1210–1214.

[84] Koprowski, P. & Kubalski, A. (2003). C termini of the *Escherichia coli* mechanosensitive ion channel (MscS) move apart upon the channel opening. *J. Biol. Chem.*, **278**, 11237–11245.

[85] Miller, S., Bartlett, W., Chandrasekaran, S., Simpson, S., Edwards, M. & Booth, I. R. (2003). Domain organization of the MscS mechanosensitive channel of *Escherichia coli*. *EMBO J.*, **22**, 36–46.

[86] Miller, S., Edwards, M. D., Ozdemir, C. & Booth, I. R. (2003). The closed structure of the MscS mechanosensitive channel. *The Journal of Biological Chemistry*, **278**, 32246–32250.

[87] Nomura, T., Sokabe, M. & Yoshimura, K. (2008). Interaction between the cytoplasmic and transmembrane domains of the mechanosensitive channel MscS. *Biophys. J.*, **94**, 1638–1645.

[88] Machiyama, H., Tatsumi, H. & Sokabe, M. (2009). Structural changes in the cytoplasmic domain of the mechanosensitive channel mscs during opening. *Biophys. J.*, **97**, 1048–1057.

[89] Schumann, U., Edwards, M., Li, C. & Booth, I. R. (2004). The conserved carboxy-terminus of the mscs mechanosensitive channel is not essential but increases stability and activity. *FEBS Lett.*, **572**, 233–237.

[90] Aksimentiev, A. & Schulten, K. (2005). Imaging alpha-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability and the electrostatic potential map. *Biophys. J.*, **88**, 3745–3761.

[91] Zheng, K., Scimemi, A. & Rusakov, D. A. (2008). Receptor actions of synaptically released glutamate: The role of transporters on the scale from nanometers to microns. *Biophys. J.*, **95**, 4584–4596.

[92] Leirmo, S., Harrison, C., Cayley, D. S., Burgess, R. R. & Jr., M. T. R. (1987). Replacement of potassium chloride by potassium glutamate dramatically enhances protein-dna interactions in vitro. *Biochemistry*, **26**, 2095–2101.

[93] Hille, B. (1992). *Ionic channels of excitable membranes*. Sinauer Associates, Sunderland, MA, 2nd edition.

[94] Smart, O. S., Goodfellow, J. M. & Wallace, B. A. (1993). The pore dimensions of Gramicidin A. *Biophys. J.*, **65**, 2455–2460.

[95] Szabo, A., Schulten, K. & Schulten, Z. (1980). First passage time approach to diffusion controlled reactions. *J. Chem. Phys.*, **72**, 4350–4357.

[96] Schulten, K., Schulten, Z. & Szabo, A. (1981). Dynamics of reactions involving diffusive barrier crossing. *J. Chem. Phys.*, **74**, 4426–4432.

[97] Foster, J. W. (2004). *Escherichia Coli* acid resistance: Tales of an amateur acidophile. *Nature Reviews*, **2**, 898–907.

[98] Felle, H., Porter, J. S., Slayman, C. L. & Kaback, H. R. (1980). Quantitative measurements of membrane potential in *escherichia coli*. *Biochemistry*, **19**, 3585–3590.

[99] Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD – Visual Molecular Dynamics. *J. Mol. Graphics*, **14**, 33–38.

[100] MacKerell Jr., A. D., Brooks, B., Brooks, III, C. L., Nilsson, L., Roux, B., Won, Y. & Karplus, M. (1998). CHARMM: The energy function and its parameterization with an overview of the program. In P. Schleyer et al., eds., *The Encyclopedia of Computational Chemistry*, 271–277. John Wiley & Sons, Chichester.

[101] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

[102] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.

[103] Chipot, C. & Pohorille, A. (2007). *Free Energy Calculations: Theory and Applications in Chemistry and Biology*. Springer.

[104] Henin, J., Pohorille, A. & Chipot, C. (2005). Insights into the recognition and association of transmembrane alpha-helices. the free energy of alpha-helix dimerization in glycophorin A. *J. Am. Chem. Soc.*, **127**, 8478–84.

[105] Henin, J., Tajkhorshid, E., Schulten, K. & Chipot, C. (2008). Diffusion of glycerol through *Escherichia coli* aquaglyceroporin GlpF. *Biophys. J.*, **94**, 832–839.

[106] Chipot, C. & Schulten, K. (2008). Understanding structure and function of membrane proteins using free energy calculations. In E. Pebay-Peyroula, ed., *Biophysical analysis of membrane proteins. Investigating structure and function*, 187–211. Wiley, Weinheim.

[107] Schulten, Z. & Schulten, K. (1977). The generation, diffusion, spin motion, and recombination of radical pairs in solution in the nanosecond time domain. *J. Chem. Phys.*, **66**, 4616–4634.

[108] Debye, P. (1942). Reaction rates in ionic solutions. *Trans. Electrochem. Soc.*, **82**, 265–272.

[109] Eigen, M. (1954). Kinetics of high-speed ion reactions in aqueous solution. *Z Physik.Chem.*, **1**, 176.

[110] Gardiner, C. W. (1983). *Handbook of Stochastic Methods*. Springer, New York.

[111] Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.*, **22**, 1420–1426.

[112] Hummer, G., Pratt, L. & Garcia, A. E. (1996). Free energy of ionic hydration. *J. Phys. Chem.*, **100**, 1206–1215.

[113] Shoup, D. & Szabo, A. (1982). Role of diffusion in ligand binding to macromolecules and cell–bound receptors. *Biophys. J.*, **40**, 33–39.

[114] Gao, J., Kuczera, K., Tidor, B. & Karplus, M. (1989). Hidden thermodynamics of mutant proteins: A molecular dynamics analysis. *Science*, **244**, 1069–1072.

[115] Zacharias, M., Straatsma, T. P. & McCammon, J. A. (1994). Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.*, **100**, 9025–9031.

[116] Kofke, D. A. & Cummings, P. T. (1998). Precision and accuracy of staged free-energy perturbation methods for computing the chemical potential by molecular simulation. *Fluid Phase Equil.*, **150**, 41–49.

[117] Pohorille, A., Jarzynski, C. & Chipot, C. (2010). Good practices in free-energy calculations. *J. Phys. Chem. B*, **114**, 10235–10253.

[118] Flyvbjerg, H. & Petersen, H. G. (1989). Error estimates on averages of correlated data. *J. Chem. Phys.*, **91**, 461–466.

[119] Jackson, J. D. (1999). *Classical Electrodynamics 3rd ed.)*. John Wiley & Sons, Inc, New York.