

C O M P U T A T I O N A L S T U D Y O F T H E
M O L E C U L A R D E T A I L S O F I O N
P E R M E A T I O N A C R O S S T H E
F O R M A T E - N I T R I T E T R A N S P O R T E R S

D i s s e r t a t i o n

for the award of the degree

“Doctor rerum naturalium”

of the Georg-August-Universität Göttingen

within the doctoral program

Physics of Biological and Complex Systems

of the Georg-August University School of Science (GAUSS)

submitted by

Kalina Atkovska

from Skopje

Göttingen, April 29th, 2016

Members of the Thesis Committee:

Dr. Jochen S. Hub (Supervisor, 1st Reviewer)
Institute for Microbiology and Genetics
Faculty of Biology and Psychology

Prof. Dr. Kai Tittmann (2nd Reviewer)
Albrecht von Haller Institute
Faculty of Biology and Psychology

Prof. Dr. Claudia Steinem
Institute of Organic and Biomolecular Chemistry
Faculty of Chemistry

Additional Members of the Examination Committee:

Prof. Bert de Groot, Ph.D.
Max Planck Institute for Biophysical Chemistry

Prof. Dr. Michael Meinecke
European Neuroscience Institute

Prof. Dr. Ricardo Mata
Institute of Physical Chemistry
Faculty of Chemistry

Date of the oral examination: June 13th, 2016

I hereby declare that this doctoral thesis entitled “Computational study of the molecular details of ion permeation across the formate-nitrite transporters ” has been written independently with no other sources and aids than those quoted.

Kalina Atkovska
April 29th, 2016
Göttingen, Germany

When a few hundred microseconds last three and a half years...

ACKNOWLEDGMENTS

First, I am grateful to my supervisor Dr. Jochen Hub, for patiently guiding me through every step of my studies and showing continuous support and enthusiasm that truly helped me complete this work.

I would also like to thank Prof. Dr. Kai Tittmann and Prof. Dr. Claudia Steinem for the valuable advice and suggestions during our TAC meetings.

I am grateful to my current and former colleagues from the Computational Biophysics Molecular group: Poker Chen, Neha Awasthi, Igor Ariz, Levin Brinkmann, Chris Knight, Roman Shevchuk, and Milos Ivanovic, for being kind, supportive, and happy to share ideas, making my experience in the group unforgettable.

I am indebted to the IMPRS-PBCS graduate school for the financial support during most of the duration of my studies.

Computational time at the GWDG and HLRN supercomputing centers is gratefully acknowledged.

I am grateful to Dr. Thomas Ullmann for providing the scripts for the GMCT calculations and the help with the setup.

I am grateful to Christian Hoffman, Dragomir Milovanovic, and Mitja Platen, for reading the thesis and providing helpful comments and suggestions that made it better, and for the friendship that enriched my PhD experience.

Finally, I dedicate this work to my parents, Lile and Lale, to my sister Irena, and to my Mitja.

ABSTRACT

The selective flux of molecules across biological membranes is essential for the normal function of any cell, and is governed by diverse membrane transport proteins. Historically, the focus in the membrane transport field was placed on cation channels of excitable tissues. Although this picture has changed in recent decades, there is still a world of non-mammalian membrane transport proteins, and proteins for transport of anions or neutral solutes, that is largely unexplored. The goal of this work is to quantitatively describe the molecular details of permeation across the formate-nitrite transporters (FNTs). FNTs transport a range of small monovalent anions across the membrane in bacteria, archaea, fungi, parasites, and green algae. Structurally, they share a fold with the aquaporin water channel, however, they have no sequence homology, and exhibit a different geometry of the permeation pore. Despite their almost ubiquitous nature, many aspects of their permeation mechanism are still unknown. It is not yet entirely clear whether they function as channels or transporters, or whether the family contains proteins of both types. A proton coupling to the permeation process has been strongly suggested, however, its exact nature has remained elusive. The role of a highly-conserved histidine residue in the center of the permeation pore has been debated. Finally, it is not certain whether all FNTs fit under the umbrella of one unifying fundamental mechanism of function.

In this thesis, atomistic molecular dynamics (MD) simulations were employed to study the permeation mechanism across all FNT subfamilies with known structure: the nitrite channel NirC, the formate channel FocA, and the hydrosulfide channel HSC. The free energy profiles for permeation of multiple substrates across the FNTs were obtained by potential of mean force (PMF) calculations. The possibility of a “knock-on” permeation mechanism was studied using computational electrophysiology simulations. The role of the central histidine residue was thoroughly investigated by accounting for different protonation states, and by studying the details of its protonation using both, microscopic and macroscopic methods. Finally, the substrate protonation during permeation was studied using combined quantum mechanics/molecular mechanics (QM/MM) simulations.

These calculations led to a general picture of the permeation across the FNTs, revealing that anions are not able to completely traverse the pore, and need to be therefore protonated in order to complete the permeation. This process was studied with most details in NirC. The permeation seems to occur in the following way: first, the central histidine protonation and anion binding in the pore occur in a coupled manner, after which the central histidine protonates the permeating anion, thereby weakening its binding and enabling its release from the pore. In FocA, an additional level of complexity is present in the permeation process, represented by flexible regions in the cytoplasmic portion of the protein. Such a general mechanism allows for high adaptability depending on the metabolic context and current needs of the cell, since in principle,

it has the capacity for both, export and import of substrates, with or without proton co-transport.

CONTENTS

I	INTRODUCTION	1
1	MOTIVATION	3
1.1	Proteins govern transport across biological membranes	3
1.2	Proteins move and this is functionally relevant	4
1.3	Computational methods for membrane protein studies	5
2	FORMATE-NITRITE TRANSPORTERS	7
2.1	History of classification of the FNT family	7
2.2	Physiological roles of the FNTs	8
2.2.1	FocA - formate channel	9
2.2.2	NirC - nitrite channel	10
2.2.3	HSC - hydrosulfide channel	11
2.3	Structural properties of the FNTs	12
2.3.1	Common structural features	12
2.3.2	Variable structural features	14
2.4	Permeation mechanism hypotheses	17
2.5	Aim of this study	18
II	METHODS	21
3	GENERAL THEORY	23
3.1	Molecular dynamics simulations	23
3.1.1	Force fields	24
3.1.2	MD algorithms	25
3.2	Potentials of mean force	26
3.3	Thermodynamic integration	27
3.4	Combined quantum mechanics/molecular mechanics simulations	28
3.5	Generalized Monte Carlo titration method	30
4	SIMULATION DETAILS	33
4.1	Structure preparation	33
4.2	Equilibrium MD simulations	33
4.3	Thermodynamics of transmembrane transfer	34
4.4	Permeation PMFs	35
4.4.1	Umbrella simulations	36
4.4.2	Construction of the PMFs	37
4.5	Substrate molecules parameters	38
4.5.1	Parameterization	38
4.5.2	Validation	38
4.6	Computational electrophysiology	39
4.7	Histidine protonation by thermodynamic integration	40
4.7.1	POPC membrane simulations	41

4.7.2	Charged membrane simulations	41
4.8	GMCT calculations	43
4.9	PMFs for simultaneous proton/anion internalization	43
4.9.1	Validation	44
4.9.2	PMFs for simultaneous internalization of formate and hydronium ions in NirC	45
4.10	QM/MM calculations	46
4.10.1	Optimization of the QM parameters	46
4.10.2	Protein QM/MM simulations	48
III	RESULTS	51
5	ENERGETICS OF PERMEATION ACROSS THE FNTS	53
5.1	Molecular dynamics of the FNTs	53
5.2	Thermodynamics of transmembrane transfer	57
5.3	Potentials of mean force for permeation across the FNTs	61
5.3.1	Substrate parameters validation	61
5.3.2	Optimization of the umbrella sampling protocol	64
5.3.3	PMFs for full permeation across the FNTs	68
6	COMPUTATIONAL ELECTROPHYSIOLOGY OF THE FNTS	75
6.1	Permeation of chloride and iodide	75
6.2	Permeation of formate in a quasi-electrophysiology setup	81
6.3	Permeation of nitrite	81
7	PROTONATION OF THE CENTRAL HISTIDINE	85
7.1	Free energy of protonation by thermodynamic integration	85
7.2	Effect of the lipid content of the membrane	87
7.3	Free energy of protonation by GMCT	89
7.4	Proton internalization in the pore	92
7.4.1	Validation of the “sum of distances” reaction coordinate	92
7.4.2	PMFs of simultaneous internalization of formate and hydronium ions	93
8	CENTRAL HISTIDINE-ANION PROTON TRANSFER	97
8.1	Selection of QM parameters	97
8.2	Proton transfer in NirC and FocA	100
IV	DISCUSSION AND CONCLUSIONS	105
9	DISCUSSION	107
9.1	Permeation mechanism in NirC	110
9.2	Permeation mechanism in HSC	115
9.3	Permeation mechanism in FocA	117
9.4	FNTs among other membrane transport proteins	120
10	CONCLUSIONS AND OUTLOOK	123
10.1	Outlook	123
V	APPENDIX	125
11	RVFV PROTEIN-LIPID INTERACTIONS	127

11.1 Introduction	127
11.2 Materials and methods	127
11.3 Results and discussion	129
12 SMALL MOLECULE PARAMETERS	133
BIBLIOGRAPHY	137
Curriculum Vitae	153

LIST OF FIGURES

Figure 2.1	Role of FocA in formate metabolism	10
Figure 2.2	Role of NirC in nitrogen metabolism	11
Figure 2.3	Role of HSC in hydrosulfide metabolism	12
Figure 2.4	FNT pentamer	13
Figure 2.5	Structural alignment of FNT and aquaporin monomers	13
Figure 2.6	FNT monomer	14
Figure 2.7	Structure of the FNT pore	15
Figure 2.8	Structure of the N-termini	16
Figure 2.9	Structure of the Ω -loop	16
Figure 2.10	Surface electrostatic potential	17
Figure 2.11	Proton relay hypothesis	19
Figure 4.1	Typical MD simulation system	34
Figure 4.2	Placement of substrates in the umbrella simulations	36
Figure 5.1	Protein backbone RMSD in free simulation	54
Figure 5.2	Average residue fluctuations in free simulation	54
Figure 5.3	RMSF of pore-lining residues	55
Figure 5.4	Flexibility of the Ω -loop	56
Figure 5.5	Flexibility in <i>StFocA</i>	56
Figure 5.6	Free energy of transfer for the export of formate/formic acid	59
Figure 5.7	Free energy of transfer for the import of nitrite/nitrous acid	60
Figure 5.8	FNT permeating substrates	61
Figure 5.9	Permeation PMF reaction coordinate	65
Figure 5.10	Optimization of the umbrella parameters	66
Figure 5.11	Single-chain PMF convergence	67
Figure 5.12	Overview of PMFs for permeation across the FNTs	69
Figure 5.13	Permeation of anions across HIS0 FNTs	70
Figure 5.14	Comparison of the permeation barriers for cations and anions across HIS0 FNTs	70
Figure 5.15	Permeation of neutral substrates across HIS0 FNTs	71
Figure 5.16	Example snapshots from the umbrella simulations	73
Figure 5.17	Permeation of anions across HIS+ FNTs	74
Figure 6.1	Computational electrophysiology setup	76
Figure 6.2	CompEl protein stability and membrane potential	78
Figure 6.3	Typical snapshots from the NirC/Cl and NirC/I CompEl simu- lations	80
Figure 6.4	Typical snapshots from the formate and nitrite CompEl simu- lations	82
Figure 7.1	Free energy of the central histidine protonation by TI	86
Figure 7.2	Free energy of protonation in FocA	87

Figure 7.3	Electrostatic potential across neutral and charged lipid bilayers	89
Figure 7.4	Dependence of the free energy of protonation of the central histidine on the proton motive force	91
Figure 7.5	Protonation probability of the central histidine in NirC in <i>EcFocA</i>	91
Figure 7.6	Scheme of the “sum of distances” reaction coordinate ξ	92
Figure 7.7	Test PMFs of the “sum of distances” reaction coordinate	93
Figure 7.8	PMFs for simultaneous internalization of formate and hydronium ions into the central binding site in NirC	94
Figure 7.9	Comparison of the free energy for separate and simultaneous internalization of formate and hydronium ions into NirC	95
Figure 8.1	Optimization of the QM/MM parameters	99
Figure 8.2	Long-range electrostatics in the protein QM/MM simulations .	100
Figure 8.3	Proton jumps in NirC	102
Figure 8.4	Proton jumps in FocA	103
Figure 9.1	Permeation mechanism across the FNTs	110
Figure 9.2	Combined PMFs for permeation across NirC	112
Figure 9.3	Combined PMFs for permeation across HSC	116
Figure 9.4	Combined PMFs for permeation across FocA	119
Figure 11.1	Thermodynamic cycle for calculation of $\Delta\Delta G_{\text{binding}}$	128
Figure 11.2	Effect of the cholesterol content on the RVFV fusion protein binding	129
Figure 11.3	RVFV simulation snapshots	130
Figure 11.4	Cholesterol accumulation in the region below the protein	131

LIST OF TABLES

Table 5.1	Hydration free energies of neutral substrates	62
Table 5.2	Box size effect on the hydration free energy	63
Table 5.3	Hydration free energies of standard ions	64
Table 5.4	Hydration free energies of ionic substrates	64
Table 6.1	Chloride and iodide full permeations across the FNTs	79
Table 6.2	Formate full permeations across NirC and <i>VcFocA</i>	79
Table 6.3	Nitrite full permeations across NirC and <i>VcFocA</i>	80
Table 7.1	Effect of the lipid content on the central histidine ΔG_{prot}	88
Table 7.2	Free energy of protonation of the central histidine by GMCT . .	90

ACRONYMS

ATP	Adenosine triphosphate
nirBCD	Cytoplasmic nitrite reductase
CPU	Central processing unit
CompEl	Computational electrophysiology
DFT	Density functional theory
DOPC	1,2-Dioleoyl-sn-glycero-3-phosphocholine
DOPS	1,2-dioleoyl-sn-glycero-3-phospho-L-serine
DTI	Discrete thermodynamic integration
FDH	Formate dehydrogenase
FHL	Formate hydrogenlyase
FNT	Formate-nitrite transporter
FocA	Formate channel A
GAFF	General Amber force field
GMCT	Generalized Monte Carlo titration
HF	Hartree Fock
HIS+	Positively-charged central histidine residue
HIS0	Neutral central histidine residue
HSC	Hydrosulfide channel
LT	Linear transit
MC	Monte Carlo
MD	Molecular dynamics
MIP	Major intrinsic protein
MM	Molecular mechanics
MP2	Second order Møller-Plesset perturbation theory
NADPH	Nicotinamide adenine dinucleotide phosphate
NMR	Nuclear magnetic resonance
NPT	Isothermal-isobaric conditions
NarK/U	Nitrate transporter K/U
NirC	Nitrite channel C
PDB	Protein data bank
PFL	Pyruvate formate-lyase
PLB	Planar lipid bilayers

PME	Particle mesh Ewald
PMF	Potential of mean force
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol)
QM	Quantum mechanics
RESP	Restrained electrostatic potential
RSS	Relaxed surface scan
RVFV	Rift Valley fever virus
SPR	Surface plasmon resonance
SSM	Solid-supported membrane
TCDB	Transporter classification database
TI	Thermodynamic integration
TM	Transmembrane helix
WHAM	Weighted histogram analysis method

Part I

INTRODUCTION

MOTIVATION

THE selective flux of molecules across biological membranes is essential for the normal function of any cell. This process is governed by diverse membrane channels and transporters. Detailed understanding of their function is not only of interest to fundamental science, but it also holds a potential for practical applications, such as design of molecular filters or drugs. A crucial aspect on the road to fully understand a certain membrane transport protein, is elucidating the molecular details of its permeation mechanism. Computational methods represent a versatile toolbox which allows to directly follow the structural dynamics of biomolecules in time, with atomic resolution. Moreover, this is achieved in a quantitative manner, enabling for the calculation of thermodynamic and kinetic properties of biological processes. This thesis studies the structural dynamics of formate-nitrite transporters using computational methods, in order to elucidate the molecular details of permeation across this protein family.

1.1 PROTEINS GOVERN TRANSPORT ACROSS BIOLOGICAL MEMBRANES

Biological membranes function as barriers that enable the maintenance of relatively fixed content inside cells and cell compartments. To be able to play this role on one hand, and allow normal cell function on the other, the flux across the membrane needs to be highly selective and regulated. The lipid bilayer is permeable only to hydrophobic and small polar molecules, therefore the necessary transport of other substrates, such as ions, is governed by membrane transport proteins [1]. These proteins are essential to the cell, as they are involved in numerous biological processes, among which nutrient ingestion, waste excretion, osmolarity regulation, establishing of transmembrane potentials, or signaling.

Proteins involved in membrane transport typically span the bilayer, thereby communicating with both membrane-separated compartments. Based on several criteria, transport proteins can be divided into two big groups: channels and transporters (carriers) [1], although it seems a clear boundary between these groups does not exist [2]. Usually, the distinction is made based on their turnover rates, and on the extent of conformational change they exhibit on permeation. Channels represent continuous pores across the membrane, which allow for substrates to passively move down their electrochemical gradients [3]. As such, channels have high, near-diffusion permeation rates. Most channels have some sort of a gate, which allows for its regulation mediated by ligand binding, voltage, pH, or mechanical signals. Transporters on the other

hand, undergo conformational cycles on permeation, and are therefore orders of magnitude slower than channels [1]. Substrate binding triggers a conformational shift that closes the transporter in an “occluded state”. Next, the transporter opens on the other side of the membrane, and releases the substrate, after which regeneration of the initial conformation follows. Transporters usually carry (pump) their substrate against its electrochemical gradient, and must therefore use some energy source, such as ATP, or coupling to the downhill transport of another substrate [1]. The presence of an occluded state prohibits opening towards both membrane-separated compartments simultaneously, and thereby prevents passive diffusion of the substrate of interest.

This work focuses on the formate-nitrite transporters, whose role is to transport certain anions across the membrane in many organisms. They represent a typical case of membrane transport proteins that exist on the border between channels and transporters, as it will be further elaborated in the following chapters.

1.2 PROTEINS MOVE AND THIS IS FUNCTIONALLY RELEVANT

The function of a protein is closely connected to its structure, and we therefore observe a rich spectrum of protein structures in nature, to match the specific tasks they perform in the cell. The early views of proteins as being majorly rigid, was soon replaced by a more dynamic picture in which proteins can exhibit different levels of flexibility [4]. Among others, nuclear magnetic resonance (NMR) spectroscopy and molecular dynamics (MD) simulations contributed significantly to establish this dynamic view. Various NMR spectroscopy methods enable the discovery and quantification of internal protein motions of different scales [5]. MD simulations on the other hand, produce trajectories of the motions of particles in time, providing a direct insight into the molecular details of protein flexibility [6].

Protein flexibility is manifested on different length and time scales, ranging from fast bond and angle vibrations, to side chain rotamer flexibility, and slower correlated local and global conformational motions [7, 8]. Today, it is not uncommon to think about proteins in terms of conformational ensembles [9, 10], which are determined by the protein’s energy landscape [11]. Such landscapes capture the thermodynamic and kinetic properties of proteins, encoded in the magnitude and shapes of the energy wells and barriers, and also help define states and categorize protein motions.

The notion that proteins are inherently dynamic, naturally led to expansion of the “structure-determines-function” paradigm, in a way that protein function is now thought to be determined rather by the interplay between its structure and dynamics [12]. Protein fluctuations and transitions between (sub)states are known to be relevant for many biological processes, such as protein folding, enzyme catalysis, protein-ligand interactions, signaling, and allostery [12–14]. The inherent conformational variability of proteins, has been also suggested to be responsible for their ability to evolve, thereby reconciling the need for both, protein specificity and protein adaptability [15]. Finally, the discovery of proteins that do not clearly follow Anfinsen’s hypothesis of protein folding [16], such as prions [17], or intrinsically disordered proteins and protein regions [18], speak for the complex structure-dynamics-function relationships that exist in nature.

In view of everything above, it becomes clear that precise description of the protein dynamics, ideally with molecular/atomistic detail, is necessary to understand the underlying mechanisms of protein function.

1.3 COMPUTATIONAL METHODS FOR MEMBRANE PROTEIN STUDIES

Many experimental methods have been developed to probe protein structure and dynamics, including X-ray crystallography, NMR spectroscopy, cryo-electron microscopy, small- and wide-angle X-ray scattering, and fluorescence single-molecule methods. Each of these methods produces data of certain resolution, and is appropriate for studying protein motions of certain time scales, where gain in dynamic information often leads to decrease in resolution [8]. In contrast, computational methods, such as MD, are in principle capable to describe the full protein dynamics, as they produce trajectories of atoms in time, thereby providing good spatial, as well as time resolution. However, two main limitations prevent MD of reaching its full capacity, namely 1) the approximations to the used potential energy function (force field) for computing the positions of the particles, and 2) limited sampling [19]. Significant progress has been made on both of these fronts, since the first MD simulation of a biological molecule (the bovine pancreatic trypsin inhibitor in vacuum) was published almost 40 years ago [20]. Continuous improvement of force field parameters has yielded state-of-art protein force fields that well reproduce multiple experimental observables and are able to describe the overall protein dynamics [21–23]. Additionally, the advance of high-performance computers [24] and the increase of the efficiency of the MD software [25, 26], allow for simulation lengths in the microsecond to millisecond range, while the development of enhanced sampling techniques enables for studying even slower processes [27]. In this way, MD simulations developed into a commonly used complementary method to experiment, providing molecular detail to macroscopic findings, and moreover, revealing previously not known functionally relevant features [6, 13].

Historically, computational studies on membrane proteins have developed slower in comparison to soluble proteins. The main reason was the scarcity of good-quality high-resolution protein structures, which are typically the starting point for MD simulations. However, technological advances in areas such as protein engineering and X-ray diffraction, have significantly contributed to the growing number of crystal structures of membrane proteins [28]. In addition, the development of lipid force fields is slowly closing the gap behind protein force fields [29], thereby enabling an adequate description of the protein environment. Today, there are even multiple computational tools designed specifically for membrane and membrane protein simulations, such as the MemGen lipid bilayer builder [30], the `g_membed` protein embedding tool [31], or the computational electrophysiology software for protein channel studies [32]. All these advances have resulted with remarkable insight into the molecular mechanisms of membrane proteins, prominent examples including the permeation, selectivity, and proton exclusion mechanisms in aquaporins [33–35], or the permeation mechanism and voltage gating in potassium channels [36–39].

 FORMATE-NITRITE TRANSPORTERS

FORMATE-NITRITE TRANSPORTERS (FNTs) constitute an ancient family of membrane proteins involved in the translocation of monovalent anions across biological membranes [40, 41]. They are widely distributed among all domains of life and have so far been identified in bacteria, archaea, fungi, parasites and algae [42]. The FNTs are known to carry out diverse cellular functions in the context of various metabolic pathways, through their role in transport of formate [43–45], nitrite [46–48], hydrosulphide [49], lactate [50], acetate [51], or bicarbonate [52]. These proteins have struck a significant interest in the past years due to multiple reasons. Generally speaking, the study of anion channels has gained momentum only in the past two decades, since historically, the field of membrane transport has been dominated by cation channels of excitable tissues [53]. In this sense, a widespread and diverse family such as the FNTs, for which fairly abundant structural information is available, could enrich our knowledge of anion transport across membranes. Further, the exact mechanism of permeation across these proteins has been a matter of debate [40, 41], with puzzling experimental data accumulating over time. Finally, keeping in mind their wide distribution in pathogens, and the lack of a homologue in humans, a pharmacological relevance of the FNTs was hinted in two recent studies. Specifically, the nitrite channel NirC has been implicated in the pathogenesis of *Salmonella typhimurium* [54], while the *Plasmodium falciparum* lactate transporter PfFNT has been suggested as a potential malaria drug target [50]. Such an application however, necessitates a detailed understanding of the permeation mechanism across these proteins.

This chapter summarizes the current knowledge on the biochemical, structural and functional properties of the FNTs, with a special focus on three FNT members: the formate channel FocA, the nitrite channel NirC, and the hydrosulfide channel HSC. So far, these are the only FNTs with a known crystal structure, which was the basis for all computational analyses performed on the FNTs in this work.

2.1 HISTORY OF CLASSIFICATION OF THE FNT FAMILY

In the search for clues about the mechanism(s) of formate transport in *E.coli*, Suppmann and Sawers in 1994 identified and characterized a gene of the *pfl* (pyruvate formate-lyase) operon, which they named *focA* (denoting “formate channel”) [43]. They suggested its product, FocA, probably functions as a specific bidirectional formate channel, without however, excluding the possibility for formate/H⁺ symport. Furthermore, via sequence analysis the authors related *focA* to two previously identified, but func-

tionally not characterized genes, *fdhC* of the *fdhCAB* (formate dehydrogenase) operon [44] and *nirC* of the *nirBCD* (NADH-dependent nitrite reductase) operon [55]. Based on their genetic context, and on the sequence similarity of their products to FocA, they proposed FocA, FdhC, and NirC might represent a new class of transport proteins for short-chained acids.

The term “formate-nitrite transporters” appeared for the first time as a phylogenetic category representing this group of proteins, in an early classification of transport proteins derived from genome analyses [56]. In this work, the FNTs were suggested to function as secondary carriers (hence “transporters” in the name), based on the functional characterization of a proposed yeast acetate/ H^+ symporter from this family [51]. The prokaryotic members were suggested to be involved in the transport of nitrite and formate.

Since then, a number of new FNT members have been identified in highly diverse organisms, continuously increasing the variety of discovered transported substrates. In prokaryotes, the FNTs are represented with at least three subfamilies. Specifically, in addition to FocA, involved in formate transport [43], and NirC in nitrite transport [46], a third prokaryotic FNT was recently discovered in *Clostridium difficile* [49], and suggested to be involved in the transport of hydrosulfide ions. In archaea, the abovementioned FdhC was identified in *Methanobacterium formicicum* [44, 45] and is a putative formate uptake transporter. In fungi, additional to the putative acetate/ H^+ symporter identified in *Saccharomyces cerevisiae* [51], a nitrite specific transporter NitA was identified and characterized in *Aspergillus nidulans* [48, 57]. A similar nitrite transport protein, Nar1, was also identified in green algae, specifically *Chlamydomonas reinhardtii* [58]. A study of the *Nar1* gene family revealed that certain members are also carbon regulated (in addition to nitrogen), which suggested they might be involved in bicarbonate transport [59]. Indeed, NAR1.2, also known as LCIA, was associated to bicarbonate uptake across the chloroplast envelope in *C. reinhardtii* [52, 60]. Finally, a lactate transporter was identified and characterized in the *P. falciparum* parasite [50, 61]. So far, no FNTs were found in higher plants or animals.

An unexpected revelation was brought by the first crystal structure solved for an FNT protein, namely FocA from *E.coli* [42]. The structure revealed a shared fold with aquaporins, despite the lack of sequence homology. This suggested a relation of the FNTs to the major intrinsic protein (MIP) family, and more importantly, promoted the view that the FNTs might function in a channel-like manner. These new developments led to reclassification of the FNT family in the transporter classification database (TCDB) [62–65] from the porter subclass (old number: 2.A.44) to the α -type channel subclass (new number: 1.A.16). Here, they are grouped together with the MIP family in a superfamily of major intrinsic proteins. However, the opinion that the FNTs represent a superfamily on their own is still present in the field [66, 67].

2.2 PHYSIOLOGICAL ROLES OF THE FNTS

As previously mentioned, FNTs are involved in the transport of monovalent anions across biological membranes. Being an ancient protein family, they perform this func-

tion in a very diverse range of metabolic pathways in a variety of living organisms. The largest portion of the knowledge regarding the FNTs originates from the known prokaryotic members: FocA, NirC, and HSC, and their wider metabolic context is described in more details in this section.

2.2.1 *FocA* - formate channel

Formate is an important metabolite in the energetic metabolism of bacteria, with especially prominent role during mixed-acid fermentation. The intracellular formate pool, as well as the formate transport across the bacterial inner membrane, require tight regulation. This is because on one hand, formate is toxic for the cell above certain levels and leads to overacidification, and on the other, it is a highly reducing compound that can be utilized by the cell. Due to its pK_a of ~ 3.8 , at physiological pH formate is mostly present in anionic form, necessitating a membrane transport facilitator. Formate metabolism in enteric bacteria has been exhaustively reviewed [68–71], on which the largest part of this section is based.

Pyruvate formate-lyase (PFL) is a central enzyme of the mixed-acid fermentation, expressed in anaerobic or microanaerobic conditions. In such conditions, PFL substitutes or complements the role of the aerobic pyruvate dehydrogenase, by cleaving pyruvate into formate and acetyl-CoA. Since a third of the glucose carbon is converted to formate in this way during mixed-acid fermentation, formate has to be exported in order to prevent overacidification of the cytoplasm and decoupling of the proton gradient. More importantly, when exported, formate can be used as a substrate for the anaerobic formate dehydrogenase FDH-N, a membrane protein with an active site located on the periplasmic side. This enables the cell to efficiently harness the energy stored in formate when there is available nitrate as an electron acceptor. The entire process leads to generation of three protons in the extracellular space for each exported formate molecule. Alternatively, when the periplasmic pH drops, and there is abundance of protons in the periplasmic space, formate is re-imported into the cell and indirectly activates the “last resort” metabolic pathway for formate consumption, involving the formate hydrogenlyase (FHL) complex. In these conditions, another formate dehydrogenase, FDH-H, is activated, which is located on the cytoplasmic side of the membrane. The final product of the FHL complex are CO_2 and H_2 gas in 1:1 ratio, and the main role of this pathway is to maintain the pH homeostasis.

The *focA* gene shares an operon with the gene for the formate-producing enzyme PFL, and as previously mentioned, was identified as a formate transport system in *E. coli* [43]. *In vivo*, FocA was confirmed: 1) to export mainly formate to the periplasm, where it can be utilized as an electron donor by FDH-N, and in minor portions to also export acetate, succinate, and lactate, and 2) to import exclusively formate, when the pH of the medium falls below 6.5, directing it into the FHL pathway [43, 72]. A scheme of the role of FocA in formate metabolism is shown in figure 2.1.

Inherent polyspecificity of FocA was demonstrated in an electrophysiological study in planar lipid bilayers [73], which revealed that it is able to permeate also other products of the mixed-acid fermentation with comparable conductances to formate. However,

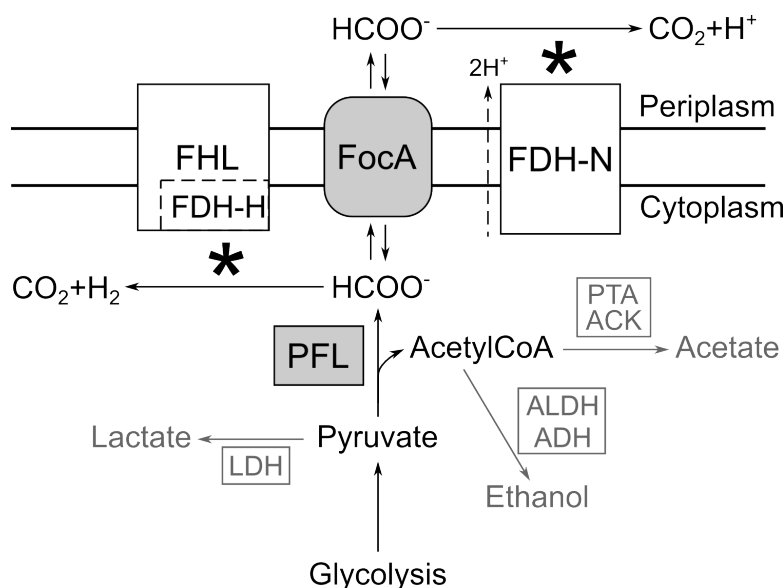


Figure 2.1: Role of FocA in formate metabolism and mixed acid fermentation. PFL (pyruvate formate-lyase) is genetically linked to FocA. Location of active sites of membrane proteins marked with star. FHL: formate hydrogenlyase; FDH: formate dehydrogenase; PFL: pyruvate formate-lyase; PTA: phosphotransacetylase; ACK: acetate kinase; ALDH: aldehyde dehydrogenase; ADH: alcohol dehydrogenase; LDH: lactate dehydrogenase. Based on [73].

Beyer et al. argue that this polyspecificity is not relevant *in vivo*, where formate is the most significant translocated ion [72]. A clue for the rationalization of the discrepancy between *in vivo* and *in vitro* data on FocA permeability came with the discovery that PFL itself binds to FocA and could regulate formate translocation, possibly by funneling formate directly into the channel [66].

2.2.2 NirC - nitrite channel

Nitrogen is necessary for the biosynthesis of proteins and nucleic acids. It exists in several oxidation states connected in a biogeochemical cycle, which is evolving in a tight relation to the enzymatic systems of the nitrogen metabolism in living organisms, wherein microorganisms have the most prominent role [74]. In the nitrogen cycle, nitrite represents a central branching point, which among other, connects nitrate (available in the environment) to ammonia (the only form of nitrogen that can be assimilated into biomass) [75]. In *E. coli* and related enteric bacteria living in electron-rich environments, the two-step nitrate reduction to ammonia via nitrite is coupled to the respiratory system and is overall an energy-conserving process. It occurs in anaerobic conditions via the activity of the periplasmic or cytoplasmic nitrate and nitrite reductases [76]. The cytoplasmic branch of the respiratory reduction of nitrate involves a nitrate reductase of the Nar family, which is activated by high nitrate concentrations, and the nitrite reductase NirBD, which is activated by both, nitrate and nitrite. A simplified scheme is shown in figure 2.2.

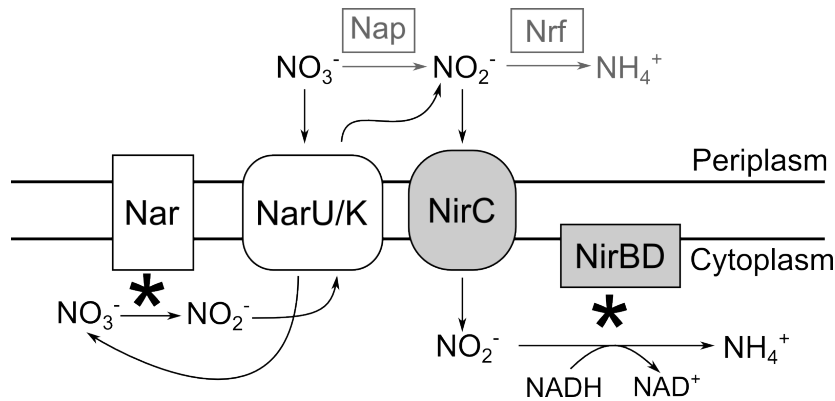


Figure 2.2: Role of NirC in nitrogen metabolism. NirBD (cytoplasmic nitrite reductase) is genetically linked to NirC. Active sites of Nar and NirBD marked with star. Nap: periplasmic nitrite reductase; Nrf: periplasmic nitrite reductase; Nar: cytoplasmic nitrate reductase; NarU/NarK: nitrate/nitrite antiporters. Based on [77].

Clearly, the transport systems for the different nitrogen species, as well as their regulation, play an important role in the described processes. Moreover, nitrite concentration in the cell must be regulated, since it can be converted to the cytotoxic nitric oxide (NO), if accumulated [77]. There are three transport systems which are known to translocate nitrite: NarK, NarU, and NirC. While the primary function of NarK and NarU is import of nitrate, NirC is able to transport only nitrite [46]. NirC is expressed together with the cytoplasmic nitrite reductase NirBD, therefore, it is expected to be mainly involved in nitrite import, supplying in this way NirBD with its substrate [78]. Indeed, in absence of NirC, nitrite is rather exported, and furthermore, in cells grown in a glucose-free medium, which hinders the function of NirBD, nitrite also accumulates in the medium [79, 80]. Currently, it is believed that most of the nitrite produced by the cytoplasmic nitrate reduction is first expelled into the periplasm through the nitrate/nitrite antiport activity of NarK or NarU, and then uptaken by NirC for further reduction by NirBD [77, 80].

Apart from the role in nitrogen assimilation, uptake of nitrite via NirC has been also shown to have a direct function in the virulence of *S. typhimurium* [54]. Namely, the internalization of nitrite by NirC and its detoxification by NirBD leads to reduced NO content in the *Salmonella*-containing vacuole in active macrophages, which in turn, via a cascade of regulation factors, leads to reduced production of NO by the inducible NO synthase. In this manner, intracellular *Salmonella* avoid killing by NO in active macrophages. NirBD is found also in multiple other pathogens resistant to NO, such as *Mycobacterium tuberculosis* and *Neisseria gonorrhoeae* [81].

2.2.3 HSC - hydrosulfide channel

Hydrogen sulfide is a major product of the assimilatory and dissimilatory sulfite reduction during anaerobic bacterial growth [82]. It has been shown to inhibit growth of sulfate-reducing bacteria [83], and its toxicity above certain concentration is probably

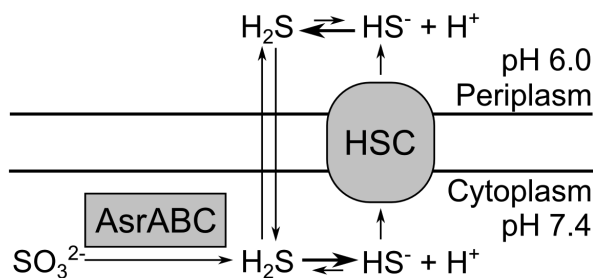


Figure 2.3: Role of HSC in hydrosulfide metabolism. AsrABC (anaerobic sulfate reductase) is genetically linked to HSC. Based on [49].

due to its binding to the metal centers of enzymes [84]. Therefore, hydrogen sulfide has to be efficiently exported from the cell. The first $\text{p}K_a$ of hydrogen sulfide is 6.8, raising the question whether passive diffusion across the membrane in a neutral form is sufficient for detoxification purposes. However, it is thought that due to the intracellular-extracellular pH difference, it is the transport of hydrosulfide ions across the membrane that would mostly contribute to the export from the cell, which in turn would require a specific protein transport system [49] (fig. 2.3).

Czyzewski et al. [49] identified an FNT-related gene (known as *HSC*, *FNT3*, or *asrD*) in the operon of the anaerobic sulfate reductase *asrABC*. This reductase catalyzes the process of sulfite reduction to sulfide, which under physiological conditions is quickly reduced to hydrogen sulfide. In a concentrative uptake assay these authors showed that HSC is able to transport formate, and that the formate uptake was inhibited by nitrite and hydrosulfide in a concentration-dependent manner. Due to its genetic context, the export of hydrosulfide ions is believed to be the primary physiological role of HSC.

2.3 STRUCTURAL PROPERTIES OF THE FNTS

So far, the crystal structure of five FNT members has been solved: FocA from *E. coli* [42], *Vibrio cholerae* [85], and *S. typhimurium* [86], NirC from *S. typhimurium* [78], and HSC from *C. difficile* [49] (hereafter *EcFocA*, *VcFocA*, *StFocA*, NirC, and HSC, respectively).

2.3.1 Common structural features

FNTs have a homopentameric architecture with a five-fold symmetry axis perpendicular to the membrane, and approximate thickness and diameter of 50-60 Å and 80 Å, respectively (fig. 2.4). The FNT monomer has a significant structural similarity to an aquaporin monomer (fig. 2.5), with slight differences due to the fact that FNTs accommodate an extra monomer in their quaternary structure, compared to a typically tetrameric aquaporin. The center of the pentamer is represented by a central cavity of various sizes, which in the crystal structures is filled with electron density resembling

detergent lipid tails, and is therefore believed to be plugged by 2-3 lipid molecules when it is in the membrane [42, 78, 85].

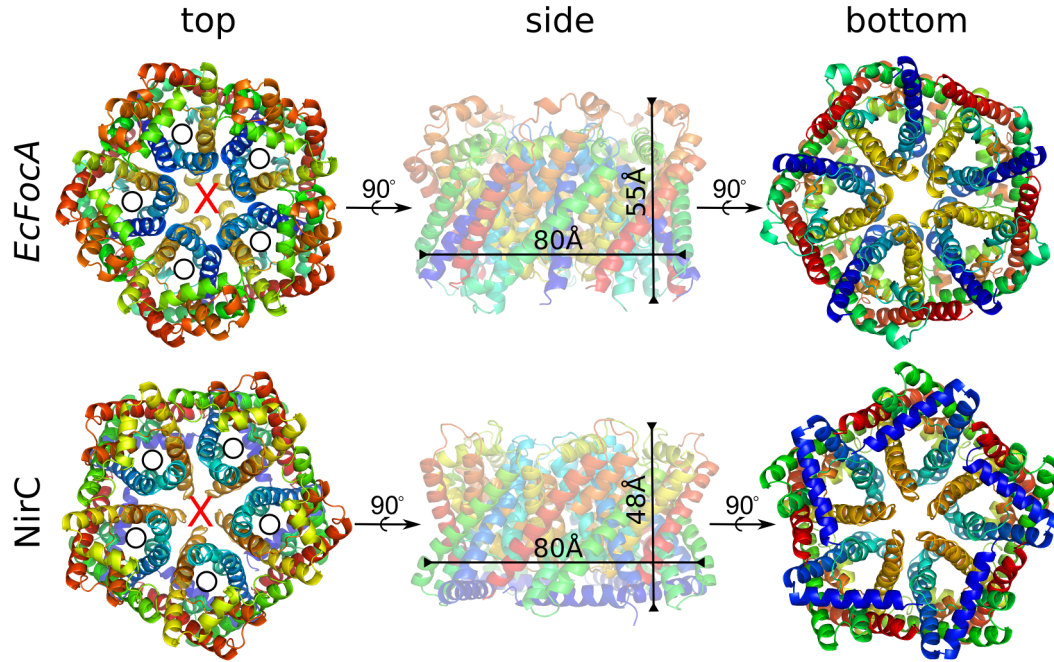


Figure 2.4: FNT homopentameric structure. The structures of *EcFocA* (PDB ID: 3KCU) and *NirC* (PDB ID: 4FC4) are shown as ribbon diagrams. The “top” view corresponds to the periplasmic, and the “bottom” view to the cytoplasmic side of the proteins.

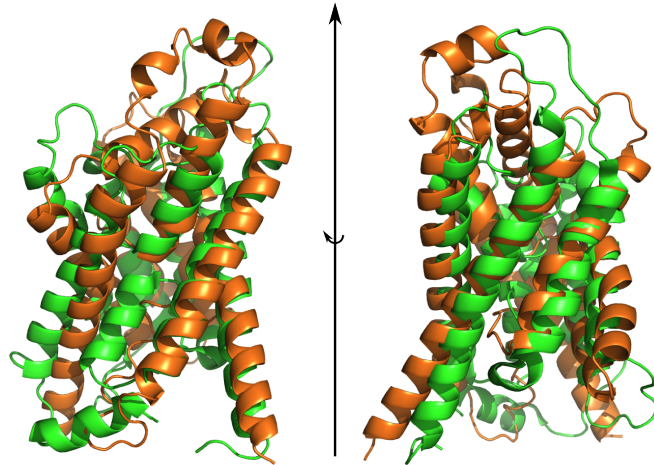


Figure 2.5: Structural alignment of an FNT (PDB ID: 3KLY, orange) and aquaporin (PDB ID: 1J4N, green) monomer.

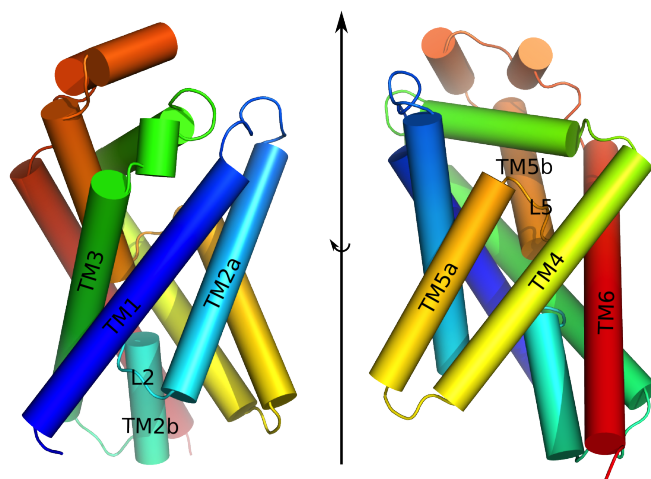


Figure 2.6: Structure of the FNT monomer. Chain A from *EcFocA* (PDB ID: 3KCU) shown as a ribbon diagram with helices depicted as cylinders. The transmembrane helices and internal loops are labeled according to their topology.

Each of the five monomers represents a right-handed twisted bundle of six transmembrane helices exhibiting a two-fold pseudosymmetry (fig. 2.6). Transmembrane helices 2 and 5 (TM2, TM5) are broken into two distinct subhelices by loops 2 and 5 (L2, L5). Both termini are located on the cytoplasmic face of the protein.

Each of the monomers is traversed by a continuous pore, open to both sides of the membrane, which is thought to be the substrate permeation pathway [42, 49, 78, 85, 86] (fig. 2.7). It is made of a central narrow pore, approx. 15-20 Å long, and two funnel-shaped vestibules through which the pore communicates with the cytoplasmic and periplasmic space. Two constriction sites are found in the narrow pore and delimit a central chamber (sometimes termed the “central vestibule” [78]). Both constrictions have on average a diameter smaller than 2.5 Å and are formed by one valine and two leucine residues on the cytoplasmic side, and two phenylalanine residues on the periplasmic side. This makes the permeating pore rather narrow and hydrophobic. A highly-conserved histidine residue is located on L5, almost directly in the middle of the pore (hereafter “the central histidine”).

2.3.2 Variable structural features

While the constrictions and central histidine are well conserved among the FNTs, certain variability can be observed mainly in the cytoplasmic half of the pore. Specifically, the N-termini in FocA were either truncated (*EcFocA* structure [42]) or not resolved (*VcFocA* structure [85]), when crystallized at neutral pH. At pH = 4 (*StFocA* structure), the N-terminus was not resolved in one of the monomers (termed “open” conformation), and showed two different orientations in the rest of the monomers (termed “intermediate” and “closed” conformations) [86]. In contrast, the N-termini in NirC and HSC exhibit

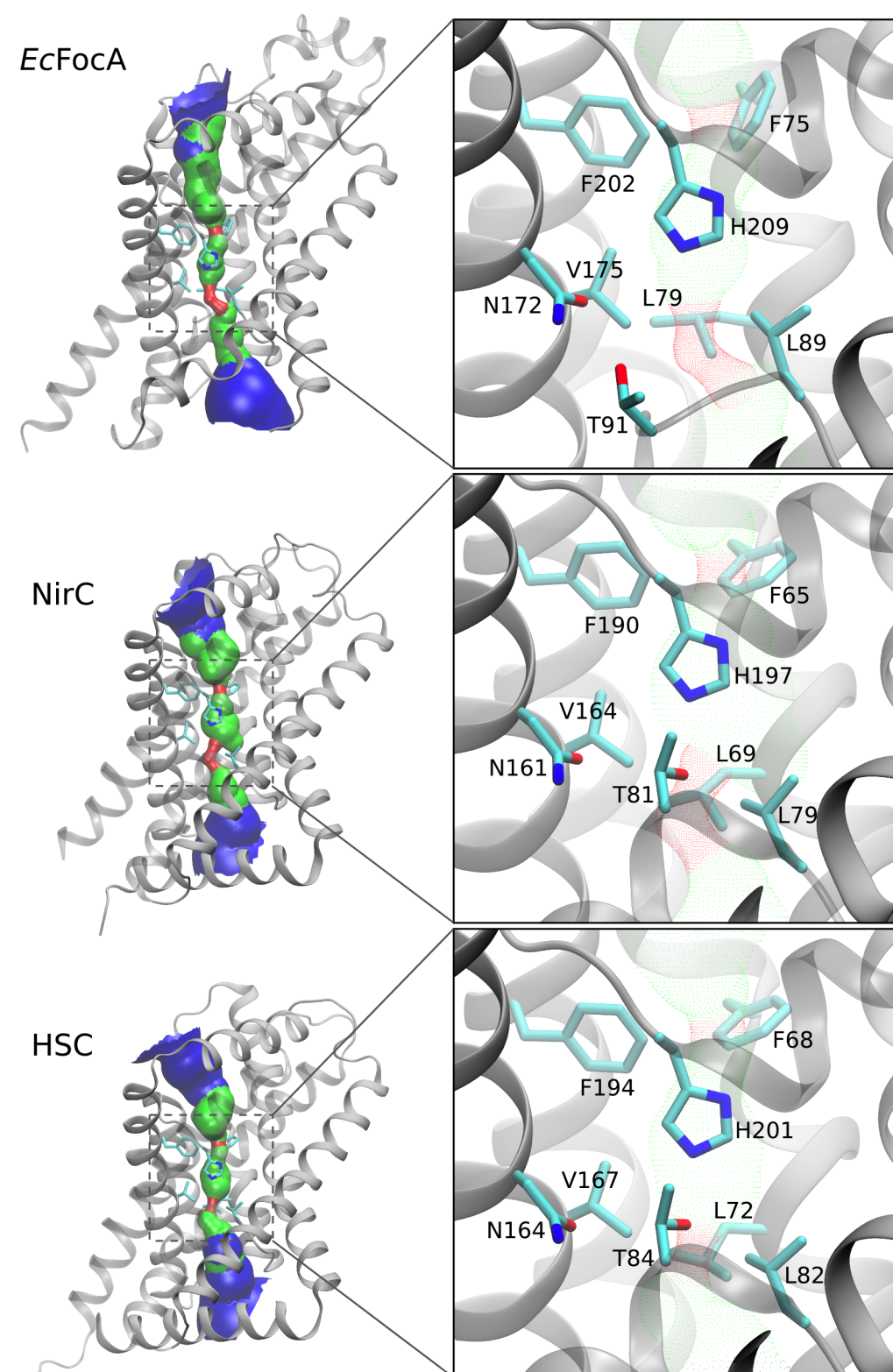


Figure 2.7: Structure of the FNT pore. Monomer shown as a ribbon diagram, the side chains of the pore-lining residues in the region of the constriction sites shown as sticks, pore representation by HOLE [87], color coded by pore radius: $< 1 \text{ \AA}$ red, $1 \text{ \AA} < 2.5 \text{ \AA}$ green, $> 2.5 \text{ \AA}$ blue. TM6 removed for clarity in the zoomed representation on the right. PDB IDs: *EcFocA* 3KCU, *NirC* 4FC4, *HSC* 3TDO.

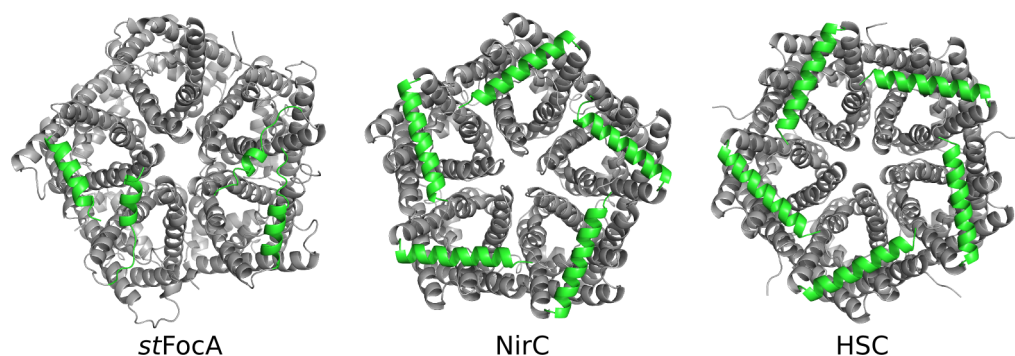


Figure 2.8: Structure of the N-termini. Protein shown as a ribbon diagram, N-termini colored green. In *StFocA*, the N-termini are resolved in only four monomers and exhibit two different orientations. In *NirC* and *HSC*, all termini are well structured and have the same orientation. PDB IDs: *StFocA* 3Q7K, *NirC* 4FC4, *HSC* 3TDO.

well defined helical structure in all monomers and at all crystallization conditions [49, 78] (fig. 2.8).

Similarly, the region connecting the TM2a subhelix to the TM3 helix, termed the “ Ω -loop”, was resolved in two different orientations in the *VcFocA* structure, which were suggested to correspond to an open and closed orientation [85]. In *NirC* and *HSC*, this region exhibits virtually no variability and has a defined helical structure with an identical orientation in all monomers and hydrogen bonding to residues from the N-terminus [49, 78] (fig. 2.9).

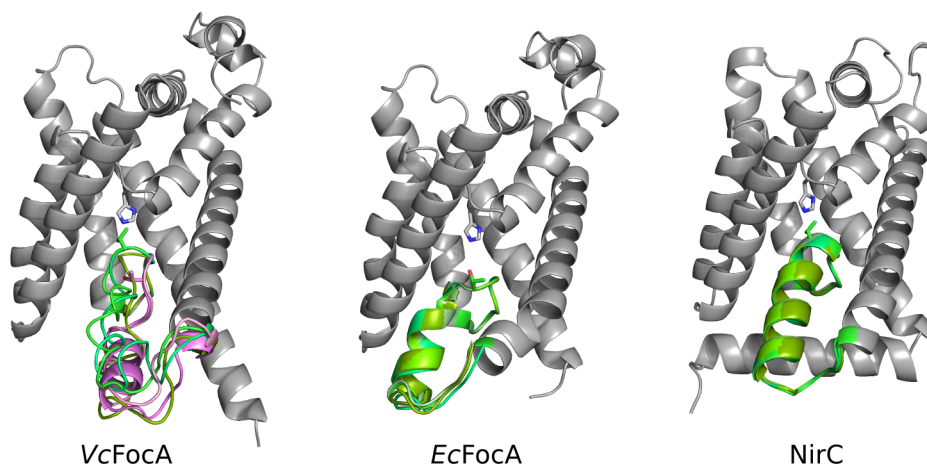


Figure 2.9: Structure of the Ω -loop. Monomer shown in gray, overlay of the Ω -loop from all five monomers shown in different shades of green. In *VcFocA*, the Ω -loop from chains A, B, and C (“closed”) is shown in different shades of green, and from chain D and E (“open”) in different shades of violet. TM6 removed for clarity. PDB IDs: *VcFocA* 3KLY, *EcFocA* 3KCU, *NirC* 4FC4.

Finally, the electrostatic potential of the cytoplasmic surface is positive in all FNT structures, in accordance to the “positive-inside” rule. In contrast, the potential of the periplasmic surfaces varies among the FNTs, with NirC and HSC exhibiting rather positive, while FocA rather negative electrostatic potential on the periplasmic side (fig. 2.10).

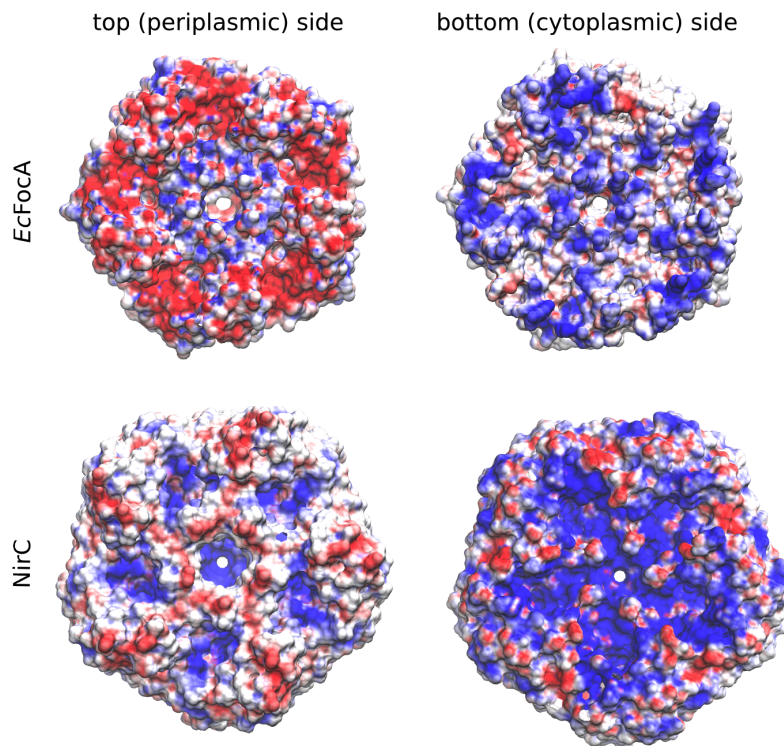


Figure 2.10: Surface electrostatic potential in FocA and NirC, calculated with APBS [88] and colored from blue at $-5 k_B T$ to red at $+5 k_B T$. PDB IDs: *EcFocA* 3KCU, *NirC* 4FC4.

2.4 PERMEATION MECHANISM HYPOTHESES

The permeation mechanism of the FNTs has been a matter of discussion since early on, as it is already hinted in the nomenclature inconsistencies over two decades of FNT studies. Although the entire family was named “transporters” when it was first classified [56], the first FNT for which a function could be assigned was named FocA, denoting “formate channel” [43, 71, 72]. The yeast FNT, Yha8, was characterized as a transport system for short-chained acids (primarily acetate), functioning in an electroneutral proton-symport manner [51, 56]. The discovery of the aquaporin-similar structure of the FNTs shifted the view in the field towards a channel-like permeation mechanism [42]. This view was additionally supported by two electrophysiological studies of *StFocA* [73] and *NirC* [78] in planar lipid bilayers (PLB), demonstrating voltage-independent electrogenic ion transport. The measured conductances of ~ 25 pS in these studies place them in the range of moderately fast channels. These studies also revealed that FocA

and NirC lack strict substrate specificity, but are rather polyspecific to a range of monovalent anions, such as formate and nitrite. Furthermore, chloride seems to only bind FocA without being efficiently translocated [73]. It is worth noting that FocA exhibited pH-gating with currents ceasing at $\text{pH} < 5.6$ [73]. Such gating was absent in NirC, however, change in pH did produce an effect on the reversal potential [78]. Another *in vitro* study involving SSM-based electrophysiology, suggested nitrite/ H^+ antiport activity for NirC [89]. The latest addition to the FNT family, *PfFNT*, was demonstrated to co-transport lactate together with a proton in a symport mechanism [50, 61].

Taken together, current structural, functional, and biochemical data do not paint a unified picture regarding the selectivity and permeation mechanism of the FNT family. The constriction sites have been suggested as obvious selectivity filters, and the central histidine has been suggested a crucial role in the permeation process [40–42, 78, 85, 90]. Mutating this residue resulted with no measurable current in *StFocA* [73]. The N-termini have been proposed to have a role in the pH-gating of FocA, while the surface electrostatics have been suggested to influence the directionality of transport [41, 73, 86]. A substrate-gated mechanism has been proposed for FocA, based on the different orientations of the Ω -loop in the *VcFocA* structure [40, 85]. Here, the hydrogen bond between a threonine residue from this loop and the central histidine is present in the “closed”, but absent in the “open” orientation. The authors propose that such mechanism could involve a competition between the permeating substrate and the threonine side chain for hydrogen bonding with the N_ϵ -atom from the central histidine.

In their review [40], Waight et al. describe the selectivity of the FNTs as a combination of “surface electrostatics, geometric constraints, and interaction with the central histidine”. The authors recognize the prevailing evidence for FNTs as channels, while still acknowledging some may be transporters, drawing a parallel to the ClC chloride channel/transporter family. On the other hand, Lü et al. [41, 78], propose a “proton relay” permeation mechanism, in which the central histidine cycles between its neutral and positively-charged (imidazolium) form, while transiently protonating the permeating ion, enabling it an easier passage through the hydrophobic constrictions (fig. 2.11). This mechanism would enable FNTs to work optionally as channels or proton-coupled transporters, though not all possibilities need to be physiologically relevant [91].

2.5 AIM OF THIS STUDY

The goal of the current study is to offer a first comprehensive and comparative analysis of the energetics of permeation across all FNT subfamilies with known structure. The character of permeation is not yet clear, as experimental evidence for both, channel and porter activity of the FNTs exists. This work explores whether this duality implies functional segregation within the FNT family, or alternatively, whether the experimental observations can be integrated by a single fundamental mechanism. Further, a proton involvement of some sort in the permeation process has repeatedly been suggested. Therefore, this work aims to elucidate the nature of proton coupling to the permeation.

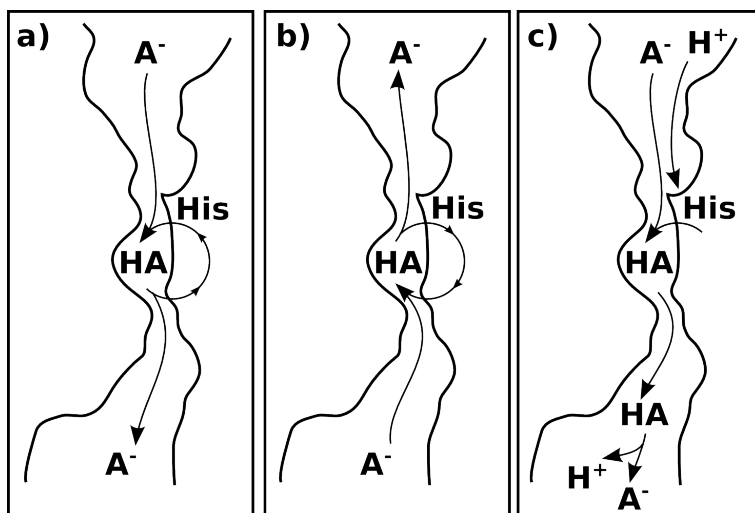


Figure 2.11: Proton relay hypothesis (adapted from [41]). Import (a) and export (b) of the substrate in an anionic form (A^-) by transient substrate protonation (HA) via proton relay that includes the central histidine. (c) Symport of anion and proton in case the central histidine is re protonated before the substrate returns the proton to the relay.

Finally, by considering all subfamilies with known structure, possible adaptations among the FNT members corresponding to their physiological roles are examined.

For this purpose, molecular dynamics simulations were employed to study the free energy barriers and molecular details of the permeation across the FNTs, by calculation of potentials of mean force for permeation, and performing computational electrophysiology [32] simulations. The role of the highly-conserved central histidine residue in the permeation was thoroughly investigated by considering different protonation states and examining the thermodynamic and kinetic details of its protonation. Mixed quantum mechanics / molecular mechanics simulations were performed to investigate the bound state of the substrate in the channel. These data resulted in a general picture of the permeation mechanism across the FNTs, involving substrate protonation by the central histidine.

Part II

METHODS

GENERAL THEORY

THIS chapter describes in general the theoretical basis for most methods used in this thesis. These include: molecular dynamics (MD) simulations, potential of mean force (PMF) calculations, free energy calculations by thermodynamic integration (TI), combined quantum mechanics/molecular mechanics (QM/MM) simulations, and free energy calculations with the generalized Monte Carlo titration (GMCT) method.

3.1 MOLECULAR DYNAMICS SIMULATIONS

MD simulations are an irreplaceable tool for analyzing and understanding the underlying physical principles of structure and functions of biological macromolecules [6]. They are used to calculate the dynamic behavior of a system by solving Newton's equation of motion for the particles in the system as a function of time [19]. As such, conventional MD simulations involve three approximations: 1) decoupling of the motion of nuclei and electrons (Born-Oppenheimer approximation), 2) classical description of the motion of the nuclei, and 3) approximation of the potential energy surface by an empirical energy function (force field) [92]. In short, the time-evolution (trajectory) of a system with n number of particles and their positions $\mathbf{R} = \{\mathbf{R}_1, \dots, \mathbf{R}_n\}$ is calculated at each time step by:

$$m_i \frac{\partial^2 \mathbf{R}_i}{\partial t^2} = -\nabla_{\mathbf{R}_i} V(\mathbf{R}), i = 1, \dots, n \quad (3.1)$$

which corresponds to the second Newton's law:

$$m_i \mathbf{a}_i = \mathbf{F}_i \quad (3.2)$$

where m_i , \mathbf{R}_i , \mathbf{a}_i and \mathbf{F}_i denote the mass, position, acceleration and force on particle i respectively, and $V(\mathbf{R})$ is the potential energy of the system. This classical description of the motions of particles in a system is often sufficient for the study of biological macromolecules, since they typically operate at non-extreme conditions, and moreover, proteins in solvent were found to behave classically [93]. Out of the scope of classical MD simulations using standard force fields, are excited electron states, charge transfer processes, or any chemical reactions involving formation or breaking of covalent bonds.

3.1.1 Force fields

The forces that act on the particles in the system and dictate their motion are calculated from the potential energy, which is in turn approximated by an empirical potential energy function, known as the “force field”. Typically, the potential energy is represented by a sum of expressions for the bonded and non-bonded interactions between the particles in the system [19, 94, 95], such as the one given in equation 3.3.

$$\begin{aligned}
 V(\mathbf{R}) &= V_b + V_a + V_{\text{dih}} + V_{\text{imp.dih}} + V_{\text{LJ}} + V_{\text{coul}} \\
 &= \sum_{\text{bonds } i} \frac{k_i}{2} (l_i - l_{i,0})^2 \\
 &\quad + \sum_{\text{angles } i} \frac{f_i}{2} (\varphi_i - \varphi_{i,0})^2 \\
 &\quad + \sum_{\text{dihedrals } i} \frac{V_i}{2} [1 + \cos(n\phi_i - \phi_{i,0})] \\
 &\quad + \sum_{\text{imp.dih. } i} \kappa_i (\xi_i - \xi_{i,0})^2 \\
 &\quad + \sum_{\text{pairs } i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}}
 \end{aligned} \tag{3.3}$$

Here, the first four terms represent the bonded interactions: the bond potential V_b , the bond angle potential V_a , and the improper dihedral potential $V_{\text{imp.dih}}$ (describing out-of-plane distortions), which are modeled as harmonic potentials, and the proper dihedral potential V_{dih} , which is modeled by a sinusoidal term. The last two terms describe the pair-wise non-bonded interactions. The short-range repulsive and attractive dispersion interactions are typically represented by a Lennard-Jones (LJ) potential, where the parameters σ_{ij} and ε_{ij} shape the width and strength of the potential. The electrostatic interactions are represented by the Coulomb term, where q_i denotes the partial charge of particle i , and the relative dielectric constant ε_r is typically set to 1.

There are several families of protein force fields that are currently widely used: AMBER [96], CHARMM [97], GROMOS [98], and OPLS [99]. GAFF and CGenFF are general force fields that can be used also for parameterizing other small molecules in addition to the “standard” biomolecules [100, 101]. The lipid force fields are more heterogeneous than protein force fields, as the lipid universe shows more diversity in comparison to the twenty building blocks of proteins. Some of the popular lipid force fields are: Berger-lipids [102], CHARMM36 [103], GAFF-lipids [104], Stockholm lipids (S-lipids) [105], and different adaptations of Gromos lipid parameters [106, 107]. When not stated otherwise, in this work the Amber-99SB*-ILDN parameters [108–111] were used for the proteins, in combination with the Berger lipid parameters [102, 112–114]. Specialized force fields for nucleic acids and sugars also exist, but are not of interest for this thesis, as such molecules were not a subject for simulation.

3.1.2 MD algorithms

Gromacs [26, 115–118] was used to perform all MD simulations in this thesis. Apart from the force field parameters, there are certain MD parameters that dictate the features of the simulation and have to be properly treated to enable physical description of the system. Among others, these parameters include: the integration scheme, temperature and pressure coupling, and the treatment of long-range interactions.

Time integration in Gromacs is typically performed using the leap-frog algorithm [119], in which the velocities and positions of the particles at each time step are computed as:

$$\begin{aligned}\mathbf{v}(t + \Delta t/2) &= \mathbf{v}(t - \Delta t/2) + \mathbf{F}(t)\Delta t/m \\ \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}(t + \Delta t/2)\Delta t.\end{aligned}\tag{3.4}$$

The time step Δt needs to be smaller than the fastest motions in the system, in order to prevent integration errors. However, not all vibrations need to be explicitly modeled to achieve a realistic description of the system, which enables the usage of a larger time step and renders the computations more efficient. Namely, bond vibrations are in their quantum ground state and are therefore better represented by a constraint, rather than a harmonic potential [120]. Constraining bond lengths allows increase of the time step to 2 fs. Often used constraint algorithms are SETTLE [121] (for the water molecules) and LINCS [122] (for the rest of the system). The next fastest oscillations are given by the bond angles of hydrogen atoms. This degree of freedom can be removed by modeling hydrogen atoms as virtual interactions sites [123], thereby allowing a time step of 4 fs [120].

In order to reproduce typical experimental conditions, MD simulations of biomolecules are often performed in an NPT ensemble (isobaric-isothermal conditions). This necessitates coupling of the simulation system to a temperature and pressure bath. Temperature coupling prevents heating of the system (or parts of it) due to e.g. numerical inaccuracies or use of cut-offs. Pressure coupling is necessary to avoid high pressure that might arise due to the low compressibility of the condensed phase in a fixed volume. In this work, mainly the velocity-rescale scheme [124] was used for temperature coupling, which is based on the Berendsen thermostat [125], but provides the correct kinetic energy distribution. For pressure coupling, the Berendsen barostat [126] was usually used for equilibration, as it leads to faster convergence, while the Parrinello-Rahman barostat was usually used for the production runs, as it results with a correct NPT ensemble [127, 128].

Finally, the treatment of the long-range non-bonded interactions has a tremendous effect on the computational cost of the simulation. As previously mentioned, the calculation of non-bonded interactions requires a sum of pairs of atoms, meaning they scale quadratically with the number of particles N in the system. To avoid that the computational cost scales with N^2 , LJ interactions are usually cut off beyond a distance of 1.0–1.4 nm [129]. Coulomb interactions on the other hand, cannot simply be cut off, due the long-range nature of the Coulomb potential that decays slowly, with only r^{-1} (see equation 3.3). Therefore, long-range electrostatic interactions in this work

were treated with the particle-mesh Ewald (PME) method [130, 131], which allows for $N \log N$, instead of N^2 scaling with the number of particles N .

3.2 POTENTIALS OF MEAN FORCE

The potential of mean force (PMF) is a concept in statistical mechanics which was introduced in 1935 by Kirkwood [132], and has been commonly used for describing the free energy profile of transitions in various, including biomolecular, systems [133]. As its name says, the PMF $\mathcal{W}(\xi)$ results from integrating the mean force acting on the particle(s), along some reaction coordinate ξ . It is calculated from the average distribution function $\langle \rho(\xi) \rangle$ as [134]:

$$\mathcal{W}(\xi) = \mathcal{W}(\xi^*) - k_b T \ln \left[\frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right]. \quad (3.5)$$

Here, k_b and T denote the Boltzmann constant and the temperature, respectively, ξ^* corresponds to an arbitrary reference, and the brackets $\langle \bullet \rangle$ denote the ensemble average. When $\mathcal{W}(\xi^*)$ is set to zero, $\mathcal{W}(\xi)$ represents the relative free energy change with respect to ξ^* . The average distribution function along ξ is in turn calculated from a Boltzmann weighted average as [134]:

$$\langle \rho(\xi) \rangle = \frac{\int d\mathbf{R} \delta(\xi'[\mathbf{R}] - \xi) e^{-U(\mathbf{R})/k_b T}}{\int d\mathbf{R} e^{-U(\mathbf{R})/k_b T}}, \quad (3.6)$$

where $U(\mathbf{R})$ is the total energy of the system as a function of its coordinates \mathbf{R} , and $\xi'[\mathbf{R}]$ is the function that maps the system to the reaction coordinate, and usually depends on a few degrees of freedom in the system. To give it a slightly more intuitive description, the PMF represents the free energy whose Boltzmann factor gives the probability distribution along the respective reaction coordinate [135]. A proof that the PMF $\mathcal{W}(\xi)$ is analogous to the free energy $G(\xi)$ is given in reference [136]. The reaction coordinates used in this thesis are defined later.

Proper sampling along the reaction coordinate is essential for computing the PMF accurately, and can be especially challenging when the reaction pathway is characterized with high barriers, since they will not be frequently crossed (or at all) in equilibrium simulations. Multiple methods have been developed to address these challenges [137], and one such frequently used method is umbrella sampling [138, 139]. The general idea is to split the reaction coordinate ξ into multiple regions (“umbrella windows”), and restrain the system in each of the windows by applying appropriate biasing potential. The biasing potential $\omega_i(\xi)$ is typically a harmonic potential centered on ξ_i , with a force constant K [133, 134, 140]:

$$\omega_i(\xi) = \frac{1}{2} K (\xi - \xi_i)^2 \quad (3.7)$$

Separate simulations are performed for each of the umbrella windows, where the biasing potential determines the sampled region of the reaction coordinate in each simulation. When a sufficient number of windows with sufficient overlap are employed, good

sampling along the coordinate is achieved. The resulting biased distribution function obtained from window i is given by [134]:

$$\langle \rho(\xi) \rangle_{(i)} = e^{-\omega_i(\xi)/k_b T} \langle \rho(\xi) \rangle \langle e^{-\omega_i(\xi)/k_b T} \rangle^{-1}, \quad (3.8)$$

and the unbiased PMF for this window is [134]:

$$\mathcal{W}_i(\xi) = \mathcal{W}(\xi^*) - k_b T \ln \left[\frac{\langle \rho(\xi) \rangle_{(i)}}{\langle \rho(\xi^*) \rangle} \right] - \omega_i(\xi) + F_i. \quad (3.9)$$

As $\mathcal{W}(\xi^*)$, $\langle \rho(\xi^*) \rangle$, $\langle \rho(\xi) \rangle_{(i)}$, and $\omega_i(\xi)$, are all known, the only undetermined constant required for unbiasing the results from the umbrella simulations is F_i , which represents the free energy difference for introducing the umbrella potential and is defined as [134]:

$$e^{-F_i/k_b T} = \langle e^{-\omega_i(\xi)/k_b T} \rangle. \quad (3.10)$$

Over the years, there have been significant efforts to develop accurate and robust methods for calculation of unbiased PMFs from biased umbrella simulations [138, 139, 141–143]. In this thesis, the weighted histogram analysis method (WHAM) [142] was used for this purpose. WHAM is probably one of the most popular methods for unbiasing the distributions obtained from umbrella simulations [133, 134]. The unbiased average distribution function $\langle \rho(\xi) \rangle$ can be calculated by iteratively solving the WHAM equations [133, 134]:

$$\begin{aligned} \langle \rho(\xi) \rangle &= \frac{\sum_{i=1}^{N_w} n_i \langle \rho(\xi) \rangle_i}{\sum_{j=1}^{N_w} n_j e^{-[\omega_j(\xi) - F_j]/k_b T}} \\ e^{-F_j/k_b T} &= \int d\xi e^{-\omega_j(\xi)/k_b T} \langle \rho(\xi) \rangle. \end{aligned} \quad (3.11)$$

Here, N_w denotes the number of umbrella windows, and $n_{i/j}$ the number of independent data points used for constructing the individual biased distribution functions $\langle \rho(\xi) \rangle_{(i)}$ (this term ensures that windows with longer autocorrelation times will be weighted less). As both equations have two unknowns ($\langle \rho(\xi) \rangle$ and F_i), they have to be solved self-consistently, where hundreds to ten thousands of iterations are typically necessary to achieve converged value for $\langle \rho(\xi) \rangle$, which can be related to the PMF by equation 3.5. All WHAM calculations in this thesis were performed with the `g_wham` module [133], as implemented in Gromacs 4 [26, 117, 118].

3.3 THERMODYNAMIC INTEGRATION

Free energy differences are frequently calculated by connecting the states of interest via non-physical transformations, which are in many cases easier to simulate in comparison to the physical process [144]. The fact that the free energy is a state variable (meaning path-independent), allows for incorporating these non-physical, but computationally convenient pathways into various thermodynamic cycles, in order to calculate quantities such as relative binding affinities, solvation energies, pK_a values, etc. The

methods, basic elements, and limitations of free energy calculations have been extensively reviewed [144–147]. The free energy in this section refers to the Gibbs free energy, as simulations were typically performed in an NPT ensemble.

The Gibbs free energy difference ΔG_{AB} between states A and B is defined as [146, 147]:

$$\Delta G_{AB} = G_B - G_A = -k_b T \ln \frac{\mathcal{Q}_B}{\mathcal{Q}_A}. \quad (3.12)$$

Here, $\mathcal{Q} = \mathcal{Q}(N, P, T)$ represents the partition function in an NPT ensemble, which is defined as [145–147]:

$$\mathcal{Q}(N, P, T) = \frac{1}{h^{3N} N!} \int \int e^{\frac{U(\mathbf{r}) + PV}{k_b T}} dV d\mathbf{r}, \quad (3.13)$$

where N , P , and V represent the number of particles, pressure, and container volume, respectively, h denotes the Planck constant, and $U(\mathbf{r})$ is the potential energy as a function of the coordinates \mathbf{r} . Strictly speaking, the exponent in equation 3.13 should also include a term for the kinetic energy, however, as the end states A and B usually have equal masses, the kinetic energy is excluded for simplicity [146].

Typical methods for free energy calculations use equilibrium simulations, although, non-equilibrium methods have recently gained popularity [147]. Equilibrium methods have their roots in the free energy perturbation method introduced by Zwanzig [148]. Currently used methods are expanded to include intermediate states between the end states of interest, in order to ensure a good phase-space overlap and hence better convergence [146]. The alchemical path that connects the end states via (non-physical) intermediates is usually described by the λ variable, where $\lambda = 0$ and $\lambda = 1$ correspond to the end states. With the thermodynamic integration (TI) method [132], the free energy difference is calculated by integrating the average force exerted on the system along λ [146, 147]:

$$\Delta G_{AB} = \int_0^1 \left\langle \frac{\partial U(\lambda, (r))}{\partial \lambda} \right\rangle_\lambda d\lambda. \quad (3.14)$$

There are two main flavors of TI: slow-growth, and discrete TI (DTI). In slow-growth TI, the system is slowly transformed from $\lambda = 0$ to $\lambda = 1$ in simulation, which may result with inaccurate estimates due to the non-equilibrium conditions in the simulation. Therefore, DTI was used in this thesis, in which the average $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda_i}$ is evaluated from multiple simulations, each conducted at discrete value of λ_i . The resulting derivatives are integrated by some numerical integration scheme, typically using the trapezoid rule. When sufficient intermediates are simulated for a sufficient time, this results with a robust estimate of the free energy difference [146].

3.4 COMBINED QUANTUM MECHANICS/MOLECULAR MECHANICS SIMULATIONS

As previously mentioned, classical MD simulations are not able to describe chemical reactions, such as charge transfer processes, which limits the scope of their application

in many biological systems. Such non-classical processes have to be treated by using the computationally-expensive quantum-mechanical methods, which is usually not feasible for systems containing more than several tens of atoms. A combined QM/MM approach enables to investigate chemical reactions in big systems, while maintaining the computational feasibility. This is achieved by defining a (typically small) QM region in the system, which is treated quantum-mechanically, while the interactions in the rest of the system (MM) are described by the classical force field.

In contrast to MD, where the potential energy surface is approximated by an empirical energy function, most QM methods compute approximate solutions of the time-independent electronic Schrödinger equation, based on the Born-Oppenheimer approximation [149]. Two approaches are commonly used for this purpose: wave-function based approaches, and approaches based on density functional theory [150]. The former represent the electronic wave-function as Slater determinants whose parameters are numerically optimized, with Hartree-Fock (HF) [151–153] and the second order Møller-Plesset perturbation theory (MP2) [154] being some of the most popular methods of this kind. Density functional theory (DFT) [155–157] methods represent the total energy of the system as a functional of the electron density, thereby significantly reducing the computational cost, while continuously improving the accuracy (by development of new functionals) [150]. The choice of an appropriate QM level of theory is essential for accurate description of the process of interest, and was therefore systematically investigated for the needs of the QM/MM calculations in this thesis, as it will be later described in details.

One of the critical elements in combined QM/MM simulations is the coupling of the QM and MM regions in the system [149, 158, 159]. The most simple approach is given by the subtractive scheme, where the potential energy of the entire system $V(\text{MM} + \text{QM})$ is evaluated on MM level, and the potential energy of the QM region $V(\text{QM})$ is evaluated on both, QM and MM levels, and these terms are combined as:

$$V_{\text{QM/MM}} = V_{(\text{MM})}(\text{MM}+\text{QM}) + V_{(\text{QM})}(\text{QM}) - V_{(\text{MM})}(\text{QM}), \quad (3.15)$$

where the subscripts denote the level of theory, and the term in the parenthesis denotes the respective (QM or MM) region of the system [159]. While this scheme is simple and easy to implement, it lacks the very essential ability of the MM region to polarize the QM region, and can therefore not be used for processes where a significant influence of the protein environment is expected. Another approach to couple the QM and MM regions is given by the additive scheme, in which the interactions between these regions are treated explicitly [159]:

$$V_{\text{QM/MM}} = V_{(\text{MM})}(\text{MM}) + V_{(\text{QM})}(\text{QM}) + V_{(\text{QM-MM})}(\text{QM}+\text{MM}), \quad (3.16)$$

where the first and second terms represent the potential energy of the MM region calculated on MM level, and of the QM region calculated on QM level, respectively, while the third term represents the potential energy from the interactions between the QM and MM atoms. The bonded and LJ interactions between the QM and MM atoms are usually treated with the force field description, as shown in equation 3.3. The electrostatic interactions are treated with a Coulomb potential in the “mechanical

embedding” scheme, where the partial charges of the QM atoms can be taken from the force field, or recalculated each step. The mechanical embedding has the same disadvantage as the subtractive scheme, as it does not allow polarization of the QM atoms by the MM region. Therefore, the “electrostatic embedding” scheme was used in this thesis, where the charges of the MM atoms explicitly enter the QM energy function as one-electron operators [159]:

$$h_i^{\text{QM-MM}} = h_i^{\text{QM}} - \sum_J^M \frac{e^2 q_J}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_J|}, \quad (3.17)$$

where h_i^{QM} is the original one-electron operator for the energy of electron i , \mathbf{r}_i and \mathbf{R}_J are the positions of electron i and MM atom J with partial charge q_J , and M is the number of MM atoms. As with the other coupling schemes, this type of QM/MM coupling also carries some concerns, such as the compatibility of the QM and MM electrostatics, or the risk from over-polarization at the QM/MM boundary [149]. However, it is essential for accurate description of processes which are influenced or mediated by the protein environment.

3.5 GENERALIZED MONTE CARLO TITRATION METHOD

GMCT represents a program suite for studying thermodynamic properties of binding, based on a microstate description of the system and formalism in terms of electrochemical potentials. When the ligand is a proton, and the receptor is a protein with titratable sites, then this program can be used to study the thermodynamics of protonation. Detailed information on the theory and implementation can be found in references [160–162], on which this short description is based. Some expressions were simplified corresponding to the actual application of GMCT in this work.

In the system of interest, the receptor (in this case FNT protein), and the ligand species (in this case protons) are treated explicitly through their chemical potentials, while all other components are treated implicitly through their influence on the electrochemical potentials of the receptor and the ligands. The electrochemical potential $\bar{\mu}_i$ of a substrate i is given as [163]:

$$\bar{\mu}_i = \mu_i^\circ + k_b T \ln a_i + z_i F \Theta \Delta \Psi, \quad (3.18)$$

where μ_i° , a_i , and z_i denote the standard chemical potential, activity, and formal charge of the substrate, respectively, $\Delta \Psi$ denotes the electric transmembrane potential, F is the Faraday constant, and Θ the Heaviside step function with value 0 in the extracellular phase, and value 1 in the intracellular phase, by convention. The energy of the system in microstate \mathbf{n} is related to the electrochemical potentials of its components and their respective stoichiometric coefficients ν_i by:

$$E_{\mathbf{n}}^{\text{micro}} = \sum_i \nu_i \bar{\mu}_i. \quad (3.19)$$

The receptor is further divided into multiple components: 1) the binding sites, defined as the parts of the protein that can bind the ligand (in this case, all titratable protein

side chains), and 2) the background, defined as the parts of the receptor that do not contain binding sites. Each binding site can adopt multiple “instances” k , depending on the number of ligands bound to the site. In this way, the microstate energy can be written as:

$$E_{\mathbf{n}}^{\text{micro}} = E_c^{\text{conf}} + \sum_{i=1}^{N^{\text{sites}}} \left(E_{c,i,k}^{\text{int,r}} - \nu_{c,i,k} \bar{\mu} \right) + \sum_{i=1}^{N^{\text{sites}}} \sum_{j=1}^{j < i} W_{c,i,k,j,l}. \quad (3.20)$$

Here, the first term on the right side represents the energy of the global conformation c of the receptor. In this work, only one global conformation was always considered, since the FNTs do not undergo major conformational changes, therefore, this term was identical in all microstates. The second term on the right side runs over all N^{sites} binding sites i , in their respective instance k (in this case, all titratable protein side chains represent sites, and the different protonation forms of these side chains represent instances). Here, $E_{c,i,k}^{\text{int,r}}$ corresponds to the so-called “intrinsic” energy of instance k of site i in the receptor environment, which accounts for the interaction with the background parts of the receptor and the receptor environment. The term $\nu_{c,i,k} \bar{\mu}$, where $\bar{\mu}$ represents the electrochemical potential of the ligand in bulk solution, and $\nu_{c,i,k}$ denotes the number of ligands bound to site i in instance k , accounts for the energy cost of removing the bound ligands from bulk solution. The third term on the right side runs over all pairs of sites and their instances, where $W_{c,i,k,j,l}$ represents the interaction between instance k of site i and instance l of site j . Given this microstate energy function, the partition function of a system with N number of microstates \mathbf{n} is given by:

$$Z = \sum_{\mathbf{n}}^N e^{-E_{\mathbf{n}}^{\text{micro}}/k_b T}. \quad (3.21)$$

One can then define substate a as a group of microstates that fulfill a certain condition (e.g. all microstates where the central histidine is in its doubly-protonated form), and the partition function of this substate is given by:

$$Z_a = \sum_{\mathbf{n}}^N \delta_{\mathbf{n},a} e^{-E_{\mathbf{n}}^{\text{micro}}/k_b T}, \quad (3.22)$$

where $\delta_{\mathbf{n},a}$ is equal to 1 if the microstate \mathbf{n} belongs to substate a , or equal to 0 if it does not. The probability of substate a and the free energy difference between substates a and b can then be respectively calculated as:

$$p_a = \frac{Z_a}{Z} \quad (3.23)$$

$$\Delta G_{a \rightarrow b} = -k_b T \ln \frac{Z_b}{Z_a}.$$

Since analytical solution of these equations is impractical for big systems with many sites (such as protein receptors), GMCT performs Monte Carlo (MC) simulations in order to compute the thermodynamic properties of interest. The energy terms used in equation 3.20 need to be calculated prior to the GMCT calculation, using the GCEM module of the extended MEAD program suite. The GCEM calculations are based on

a combined continuum electrostatics/molecular mechanics model, in which the protein receptor and bound ligands are described in atomic detail, while the protein environment and unbound ligands are treated implicitly. The calculation of the energy terms by GCEM is described in detail in reference [161], and is based on the CHARMM force field for the molecular mechanics part, and on the linearized Poisson-Boltzmann equation for the continuum electrostatics part.

SIMULATION DETAILS

4.1 STRUCTURE PREPARATION

The crystal structures of the FNT proteins were obtained from the Protein Data Bank [164] (www.rcsb.org), with the following PDB IDs: 4FC4 (NirC, chains A-E), 3TDO (HSC), 3KCU (*EcFocA*), 3KLY (*VcFocA*), and 3Q7K (*StFocA*, chains A-E). All detergent fragments were removed, while water molecules were kept. The missing Glu23 side chain in the *VcFocA* structure was added using the WhatIf web server [165–167]. The missing loop in the *EcFocA* structure (residues 102–110 in chain C) was not modeled, only its terminal residues were capped with N-methyl and acetyl groups. Three missing loops in the *StFocA* structures: residues 99–109 in chain B, 100–103 in chain D, and 100–107 in chain E, were modeled using the Rosetta loop modeler. Here, the kinematic closure protocol [168, 169] was used, and the lowest energy model was selected for further usage in the simulations. The missing Met 1 residue in chain D of the NirC structure was added in PyMol [170]. For most of the simulations, the incomplete N- and C-termini in all structures were capped with an acetyl or N-methyl group, respectively, in order to avoid artificially charged termini.

4.2 EQUILIBRIUM MD SIMULATIONS

The simulation setup procedure was same for all equilibrium simulations of the investigated FNTs. The protein was embedded in a POPC lipid patch with the `g_membed` tool [31]. The simulation box was solvated with explicit water molecules (TIP3P water model [171]), and the appropriate number of counterions (Cl^-) was added. A rhombic dodecahedron with hexagonal orientation and an approximate volume of 1400 nm^3 was used as a simulation box. This procedure resulted with 253 lipid molecules and ~ 31000 water molecules in each system. Typical simulation box is shown in figure 4.1.

The system was minimized using the steepest descent algorithm. The water and lipid molecules were equilibrated for minimum 20 ns, while restraining the positions of all heavy atoms of the protein with reference to the crystal structure. The restraining potential $V_{pr}(\mathbf{r}_i) = 0.5k_{pr}|\mathbf{r}_i - \mathbf{R}_i|^2$ on atom i with coordinates \mathbf{r}_i , with respect to the reference position \mathbf{R}_i , was applied with a force constant k_{pr} of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. In the next step, free unrestrained MD simulations of different duration ($> 100 \text{ ns}$) were conducted and used for structural analysis, or for generating starting structures for the umbrella and TI simulations. In both, restrained and free simulations, all bonds were constrained by using the SETTLE algorithm [121] (water molecules) or LINCS

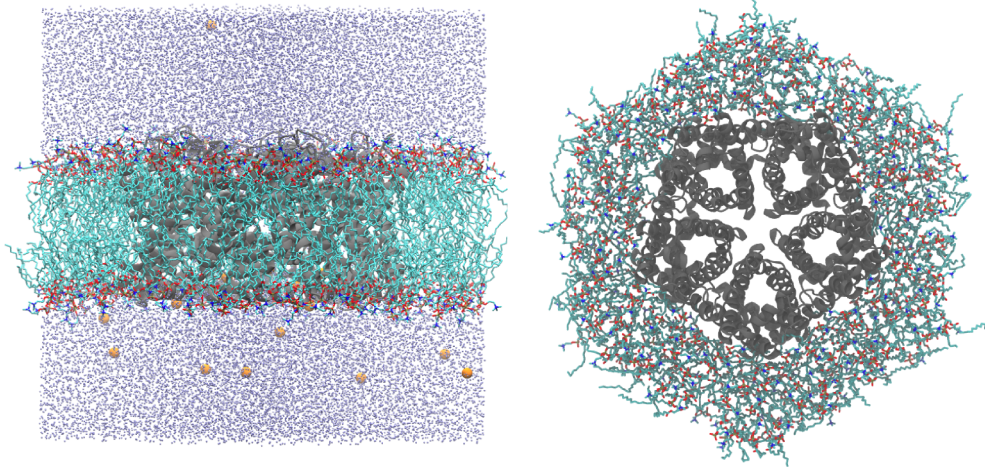


Figure 4.1: Side (left) and top (right) view of a typical MD simulation system. Protein shown as a ribbon diagram, lipid molecules shown as sticks, water molecules shown as blue dots, and counterions shown as spheres. In the right figure water molecules and counterions are not shown for clarity.

algorithm [122] (rest of the system). This, together with modeling the hydrogen atoms as virtual sites, allowed for an integration step of 4 fs. LJ and Coulomb interactions were cut off beyond distance of 1.0 nm, with neighbor searching frequency of 20 fs. Long-range electrostatics were treated with the PME scheme [130, 131] using grid spacing of 0.12 nm and interpolation order of 3. The temperature was kept 300 K using the velocity rescale thermostat [124] with time constant of 2.5 ps. The pressure was kept constant at 1 bar by semi-isotropic coupling to a Berendsen barostat [126] with time constant of 2 ps in the restrained simulations, and to a Parrinello-Rahman barostat [127, 128] with a time constant of 5 ps in the free simulations. The average root mean square fluctuations (rmsf) per residue were calculated with the `g_rmsf` module after fitting the protein to the initial structure, and converted into B-factors by $B = \frac{8\pi^2}{3}\text{rmsf}^2$.

4.3 THERMODYNAMICS OF TRANSMEMBRANE TRANSFER

The transfer energies $\Delta G_{\text{transfer}}$ of the substrates of interest, between two membrane-separated compartments, was calculated as the difference in their electrochemical potentials $\bar{\mu}$ in each compartment [163]:

$$\Delta G_{\text{transfer}} = \bar{\mu}^{\text{B}} - \bar{\mu}^{\text{A}}. \quad (4.1)$$

When “A” refers to the cytoplasmic, and “B” to the periplasmic compartment, the resulting free energy corresponds to substrate export, while in the opposite case, it corresponds to substrate import. These calculations exclude any mechanistic details of the transport across the membrane. The equations for calculation of the electrochemical potentials presented in the following are based on reference [163], where a detailed derivation can be found.

The electrochemical potential $\bar{\mu}_i$ of a substrate i is given in equation 3.18. In this work, the transfer of two weak acids of interest (formic and nitrous acid) were considered, with protonation equilibrium given as:



where HA and A^- denote the neutral and anionic form of the substrate, respectively. As the total electrochemical potential is equal for each side of the equation, the electrochemical potentials of both protonation forms are related as:

$$\bar{\mu}_{HA} = \bar{\mu}_{A^-} + \bar{\mu}_{H^+}. \quad (4.3)$$

Analogous to the definition of pH ($a_{H^+} \equiv 10^{-\text{pH}}$), a “pX” value can be defined, related to the substrate concentration:

$$a_{HA} + a_{A^-} = 10^{-\text{pX}}. \quad (4.4)$$

When the substrate is formate/formic acid, the notation pF is used, while in the case of nitrite/nitrous acid, the pX value is denoted as pN. For small substrate concentrations and far from the pH extremes, the activity coefficients of all protonation forms are assumed to be close to 1, which allows for expressing pX via the substrate concentration as:

$$\text{pX} \approx -\log_{10} \frac{c_{HA} + c_{A^-}}{c^\circ}, \quad (4.5)$$

where c° is the standard concentration of 1 mol L^{-1} . The activity of the anionic form can then be written as a function of pX [163]:

$$a_{A^-} = \frac{10^{-\text{pX}}}{1 + e^{-(\mu_{HA}^\circ - \mu_{A^-}^\circ - \mu_{H^+})/k_b T}}. \quad (4.6)$$

This can be used for calculating the electrochemical potential of the anionic form via equation 3.18, which can then be used for calculation of the electrochemical potential of the protonated form via equation 4.3.

Using this formalism, $\Delta G_{\text{transfer}}$ of the aforementioned substrates was calculated as function of the proton motive force and the intracellular pH^{in} at a fixed substrate concentration ($c_{HA} + c_{A^-}$) in both compartments of 20 mM ($\text{pX}^{in} = \text{pX}^{out} = 1.7$), or as a function of the proton motive force and the concentration gradient ($\Delta \text{pX} = \text{pX}^{in} - \text{pX}^{out}$), at $\text{pH}^{in} = 7$ and $\text{pX}^{in} = 1.7$. The standard chemical potentials of the substrates were taken from reference [172].

4.4 PERMEATION PMFS

The reaction coordinate for the permeation PMFs was defined as the z coordinate (where the z -axis is normal to the membrane) of the center of mass of the permeating substrate, with respect to the center of mass of the transmembrane part of the pentamer (fig. 5.9). This reaction coordinate was sampled using the umbrella sampling method, whereby the substrates were restrained at i number of windows by a harmonic umbrella potential $\omega_i(z) = \frac{1}{2}K(z - z_i)^2$.

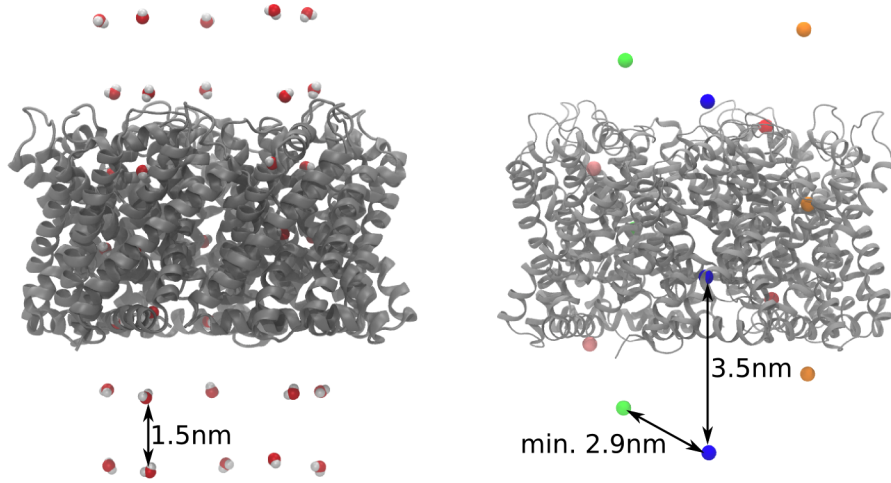


Figure 4.2: Placement of neutral substrates (left) and ions (right) in the umbrella simulations. Protein shown as a ribbon diagram, water molecules (left) and chloride ions (right) shown as spheres. The neutral substrates are placed equally in each chain, such that the distance between substrates in one chain is 1.5 nm. The chloride ions are color-coded by their association to different protein chains, and they are placed such that the distance between two ions in one chain is 3.5 nm, and the minimum distance between ions from neighboring chains is 2.9 nm.

4.4.1 Umbrella simulations

The starting structures for the umbrella simulations were taken from equilibrium simulations performed as described in section 4.2, typically from 10-20 ns long intervals beyond the 100th nanosecond in the trajectory. Multiple umbrella windows were simulated per simulation in order to enhance sampling without increase of the computational cost. The permeating substrates were inserted into the pore at the respective umbrella centers in a different manner for the neutral and ionic substrates. The neutral substrates were inserted equally in each monomer, such that the distance between two windows in a pore was 1.5 nm (fig. 4.2, left). For the ionic substrates, a substrate was inserted in only one monomer per window, in a manner such that the distance between two windows in one monomer was 3.5 nm, and the minimum distance between any two permeating ions is 2.9 nm (fig. 4.2, right).

The overlapping water molecules with the inserted substrates were removed, and in the case of adding ionic substrates, the system was neutralized as necessary. After short energy minimization, umbrella simulations of different duration were ran. The simulation parameters as described for the free simulations in section 4.2 were applied, with the following differences: 1) to keep the temperature at 300 K, stochastic dynamics temperature coupling with coupling constant of 0.5 ps was used [173], and 2) the compressibility of the box in the z direction was turned off, in order to avoid artifacts in the umbrella histograms. Additional to the umbrella potential, the substrates were also restrained to a cylinder with radius $r_c = 7\text{\AA}$ with axis centered along the pore.

This was achieved by applying a flat-bottom quadratic potential in the xy plane, with resulting additional force on the particle of $F(r) = -k_c(r - r_c)H(r - r_c)$ pointing towards the cylinder axis. Here, r represents the substrate distance from the cylinder axis, $k_c = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ is the force constant, and H is the Heaviside step function. The optimization of the umbrella parameters (force constant, window density, and simulation duration) is described in details in section 5.3.2. For most of the cases, force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ and $4000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ was used in the umbrella simulations involving neutral and ionic substrates, respectively, with distance between neighboring umbrella windows of 0.1-0.2 Å, and umbrella duration of 5-10 ns. This resulted on average with ~ 900 and ~ 550 windows per chain for the neutral and ionic substrates, respectively, and five times more per five-chain average PMF.

4.4.2 Construction of the PMFs

The first 1 ns was removed from the umbrella simulations with duration of 5 ns, and the first 3 ns were removed from the simulations with duration of 10 ns, after which the umbrella histograms were extracted. The single-chain PMFs were calculated using the periodic WHAM procedure as implemented in Gromacs 4.6. The periodicity was imposed in order to prevent a non-physical off-set of the PMFs in the bulk phases. The integrated autocorrelation times were calculated and implemented in the WHAM procedure, with prior smoothing along the reaction coordinate with a Gaussian filter with width of 2 Å. The statistical error was estimated by applying the Bayesian bootstrap procedure using 50 bootstraps. The used WHAM implementation and options are described in detail in reference [133].

With the cylinder restraint, the umbrella simulations yield an energy profile that corresponds to a channel density of one channel per cylinder cross-section. To obtain a profile that corresponds to a channel density of one channel per membrane cross-section occupied by one monomer, a trapezoidal correction was applied to the PMF in the pore entrances, which reads: $\Delta G_{\text{corr}} = k_b T \ln(A_{\text{mono}}/A_C)$ [174]. Here A_{mono} and A_C represent the cross-sections of the monomer and the cylinder, respectively. A_{mono} was estimated to be $\sim 10.6 \text{ nm}^2$ for an FNT monomer. A_C was estimated from the radius r_c and force constant k_c of the flat-bottom potential, as: $A_C = \pi(r_c + 2\sigma_c)^2 = 2.01 \text{ nm}^2$, where $\sigma_c = (k_b T/k_c)^{1/2}$ is the width of the Gaussian-shaped substrate distribution at the edge of the potential. This resulted with $\Delta G_{\text{corr}} = 4.1 \text{ kJ mol}^{-1}$.

Finally, the average five-chain PMF ΔG_{avg} was calculated from the single-chain PMFs ΔG_j as:

$$e^{-\frac{\Delta G_{\text{avg}}(z)}{k_b T}} = \frac{1}{5} \sum_{j=1}^5 e^{-\frac{\Delta G_j(z)}{k_b T}}. \quad (4.7)$$

Here, the permeation across the single-chains was considered to be independent. The average statistical error s_{avg} was calculated with error propagation of the single-chain error estimates s_j from the bootstrapping procedure as:

$$s_{\text{avg}} = \sqrt{\sum_{j=1}^5 \left(\frac{e^{-\Delta G_j / k_b T}}{\sum_{j=1}^5 e^{-\Delta G_j / k_b T}} \right)^2} s_j^2. \quad (4.8)$$

4.5 SUBSTRATE MOLECULES PARAMETERS

The following small molecules were parameterized for use in the permeation PMF calculations: formic acid, formate ion, nitrous acid, nitrite ion, hydrogen sulfide, and hydrosulfide ion.

4.5.1 *Parameterization*

The parameterization was performed using the Antechamber module [175] from the AmberTools (release 1.4) package. The general Amber force field (GAFF) [100] was used for the atom types (therefore also for the bonded and LJ parameters), since it is entirely compatible with the used Amber protein force field. HF/6-31* RESP charges were used, consistent with all GAFF force field parameterizations. Here, geometry optimization and calculation of the electrostatic potential was performed with the Gaussian software [176] using HF/6-31* level of theory, and the RESP charge fit was performed by Antechamber. The topology files were created by the tleap module in AmberTools, and converted into Gromacs format by ACPYPE [177]. All parameters can be found in chapter 12.

4.5.2 *Validation*

Validation of the parameters of the small molecules was performed by calculation of their hydration free energies ΔG_{hydr} . Additionally, ΔG_{hydr} was also calculated for several “standard” ions: Li^+ , Na^+ , K^+ , F^- , Cl^- , Br^- , and I^- , in order to determine the necessary correction of the calculated values, as it is discussed in section 5.3.1. For this purpose, the TI method was employed to calculate the free energy of transferring the solute from vacuum to a water box of certain size, by setting the coupling parameter λ to represent the interactions between the solute and solvent molecules. In order to avoid simulation artifacts, the solute-solvent LJ and Coulomb interactions were treated separately, such that in one set of TI simulations, λ was related to turning on the LJ interactions (while the partial charges of the solute atoms were set to 0), and in another set, λ was related to turning on the Coulomb interactions (while keeping the LJ interactions on).

In both cases, the system was simulated at multiple equidistant points along λ for 500 ps. The Coulomb interactions were turned on linearly, therefore 5 simulations, each at different λ value, were sufficient. When turning on the LJ interactions, a soft-core

potential was used to modify the LJ potential, in order to avoid singularities that might arise due to appearance or disappearance of atoms. In this case, the system was simulated at 21 values of λ . This resulted with 26 simulations per substrate. In all simulations, a stochastic dynamics integrator with integration step of 2 fs and coupling constant of 0.1 ps was used. The reference temperature was set to 300 K. The pressure was kept constant at 1 bar by coupling to the Berendsen barostat with time constant of 5 ps. Dispersion correction for the energy and pressure was applied. The LJ interactions were cut off at 1.4 nm, while the Coulomb interactions at 0.9 nm. The long-range electrostatics were treated with the PME algorithm, using grid spacing of 0.12 nm and interpolation order of 4. The TIP3P water model was used. A water box of 50^3 \AA^3 was used, however, different sizes were tested (see section 5.3.1). Finally, the `g_analyze` module was used to average the resulting $\partial H / \partial \lambda$ values from each simulation (discarding the first 50 ps for equilibration) and estimate the error using block averaging, and to subsequently integrate the averages over λ using the trapezoid rule.

4.6 COMPUTATIONAL ELECTROPHYSIOLOGY

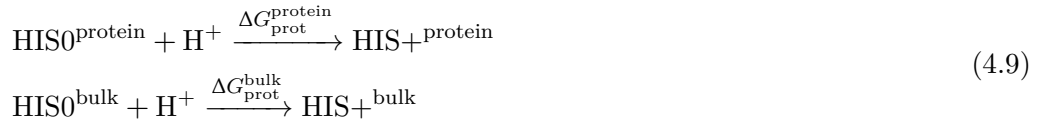
The computational electrophysiology (CompEl) code [32] was used as implemented in Gromacs 4.6 for the halide ion simulations, and Gromacs 5 for the nitrite ion simulations. The main idea of this method is to reproduce an electrophysiology setup in which channel conductance can be investigated in the presence of concentration gradient and/or transmembrane voltage. When using periodic boundary conditions (as in the case of all simulations in this work), such voltage can not be simply established across the membrane, as there is only one bulk compartment. Therefore, a double membrane system needs to be used, as the one shown in fig. 6.1b. Membrane voltage can then be established by introducing charge imbalance between the two bulk compartments. To prevent dissipation of the charge difference on ion permeation across the channel, the CompEl method exchanges ion/water pairs between buffer regions in both compartments, thereby maintaining the transmembrane voltage. The quasi-electrophysiology setup uses the same double membrane system, however, without the particle-exchange method, meaning it is not able to maintain the initially introduced voltage on ion permeation.

The systems for all CompEl or quasi-CompEl simulations were set up using the same procedure. Initially, a single system was set up, as described in section 4.2 up to and including the energy minimization, with two differences. First, the appropriate number of ions was added to reach the desired salt concentration (1M NaCl, NaI, HCOONa, or NaNO₂, or 200mM NaNO₂). Second, as suggested by the electron density features identified in the central cavity between the five monomers in the crystal structures [78, 85], three (NirC, HSC) or two (VcFocA) lipid molecules were placed in this region, in order to prevent possible leak permeation events. The single system was duplicated, and the duplicate was translated along the *z*-axis, and stitched to the single system to yield a double membrane system as shown in fig. 6.1. The double system was minimized, and equilibrated for minimum 50 ns with position restraints on all heavy atoms in the proteins and no charge imbalance. Afterwards, multiple unrestrained simulations with

different duration and different charge imbalance between the compartments ($4 e - 16 e$) were ran (fig. 6.2 and tables 6.1, 6.2, and 6.3). The simulation parameters for both, the restrained and free simulations were the same as described in section 4.2. In the simulations involving NaCl and NaI salts, the parameters for the sodium, chloride, and iodide ions were taken from reference [178], since they did not exhibit formation of salt crystals at high concentrations. The transmembrane potential was calculated by the `g_potential` module in Gromacs. The number of permeations had to be manually counted, as ions were very often found in the middle of the pores, and were therefore sometimes crossing the compartment border without in fact performing a full permeation.

4.7 HISTIDINE PROTONATION BY THERMODYNAMIC INTEGRATION

The free energy of protonation of the central histidine in the protein, with respect to histidine in water was calculated using the TI method. First we consider the following protonation reactions:



where HIS0 and HIS+ represent the singly- and doubly-protonated forms of histidine, ΔG_{prot} denotes the protonation free energy, and the superscripts correspond to the respective environments. Their subtraction and rearrangement yields:

$$\text{HIS0}^{\text{protein}} + \text{HIS}^+{}^{\text{bulk}} \xrightarrow{\Delta \Delta G_{\text{prot}} = \Delta G_{\text{prot}}^{\text{protein}} - \Delta G_{\text{prot}}^{\text{bulk}}} \text{HIS}^+{}^{\text{protein}} + \text{HIS0}^{\text{bulk}} \quad (4.10)$$

In this way, $\Delta \Delta G_{\text{prot}}$ can be calculated by TI, where λ is related to the process of transferring a proton between the central histidine and a histidine residue in the bulk solvent, i.e. the protonation of one histidine residue, and deprotonation of the other. Simulating the opposite transformations of the histidine residue in the protein and the one in bulk in the same simulation box, enables to obtain $\Delta \Delta G_{\text{prot}}$ from a single set of TI simulations, and moreover, provides a way to maintain the neutrality of the system at all values of λ . $\Delta G_{\text{prot}}^{\text{protein}}$ can then be calculated by $\Delta G_{\text{prot}}^{\text{protein}} = \Delta \Delta G_{\text{prot}} + \Delta G_{\text{prot}}^{\text{bulk}}$, where $\Delta G_{\text{prot}}^{\text{bulk}}$ is related to the known $\text{p}K_a$ of histidine in bulk and the pH via the Henderson–Hasselbalch equation:

$$\begin{aligned} \text{p}K_a &= \text{pH} - \log_{10} \frac{[\text{HIS0}]}{[\text{HIS}^+]} \\ \text{p}K_a &= \text{pH} - \log_{10} \left(\frac{1}{e^{-\Delta G_{\text{prot}}/k_b T}} \right) \\ \Delta G_{\text{prot}} &= k_b T \ln 10^{(\text{pH} - \text{p}K_a)}. \end{aligned} \quad (4.11)$$

Considering $\text{p}K_a \approx 6$ and $\text{pH} = 7$, $\Delta G_{\text{prot}}^{\text{bulk}} \approx 5.7 \text{ kJ mol}^{-1}$. The ΔG_{prot} values that were calculated with the TI method are reported in kJ mol^{-1} as $\Delta G_{\text{prot}} = \Delta \Delta G_{\text{prot}} + 5.7 \text{ kJ mol}^{-1}$.

4.7.1 POPC membrane simulations

Starting structures were taken from equilibrium simulations performed as described in section 4.2, after equilibration with various numbers of monomers with doubly-protonated central histidine and for a different time. A histidine residue capped with N-methyl and acetyl groups was added in the bulk water (hereafter the “bulk histidine”), and the overlapping water molecules were removed. Afterwards, the system was minimized, equilibrated for a few hundred picoseconds with position restraints on all heavy protein and bulk histidine atoms, and for another few nanoseconds with no other restraints, apart from a very weak restraint (force constant of $10 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) on the position of the C_α atom from the bulk histidine in z -direction. This restraint was used in order to prevent interactions of the bulk histidine with the lipids or the protein (and was therefore also present in the TI simulations). The simulation parameters for the equilibration simulations were identical to the ones described in section 4.2.

The transfer of the positive charge between the central histidine in the protein and the bulk histidine was achieved by λ transformation of the system, taking advantage of the possibility to define the end states ($\lambda = 0$ and $\lambda = 1$) in a single topology file. The system was simulated at 11 equidistant points of λ for minimum of 10 ns. A stochastic dynamics integrator with an integration step of 4 fs was used, enabled by the use of constraints for all bonds (SETTLE for water molecules, and LINCS for the rest of the system) and modeling the hydrogen atoms as virtual sites. The LJ interactions were cut off beyond 1.2 nm, and the short-range Coulomb interactions beyond 1.0 nm. Long-range electrostatics were treated with the PME algorithm with grid spacing of 0.12 nm and interpolation order of 4. Dispersion correction was applied to the energy and pressure. The temperature was kept at 300 K by stochastic dynamics temperature coupling with a coupling constant of 0.1 ps, while the pressure was semiisotropically kept at 1 bar with the Parrinello-Rahman barostat with coupling constant of 1 ps. In all TI simulations, the non-perturbed central histidine residues were in HIS0 state, with exclusion of the *EcFocA* “cooperativity” analyses, where different combinations of monomers with charged and neutral central histidine residues were investigated (fig. 7.2). $\Delta\Delta G_{\text{prot}}$ was calculated using the `g_analyze` module, by averaging the resulting $\partial H / \partial \lambda$ values from each simulation (discarding the first 1 ns for equilibration, and estimating the error using block averaging), and subsequent integration over λ using the trapezoid rule.

4.7.2 Charged membrane simulations

Initially, a system with artificially charged lipid bilayer was created from the NirC / POPC system. The starting structure was taken after 200 ns equilibration as described in section 4.2. In 64 lipid molecules ($\sim 25\%$ of the total) the choline group was transformed as such: one N-attached methyl group was replaced by a dummy atom, and the partial charges of the remaining two methyl groups and the N-atoms were adjusted to give 0 net charge. In this way, a “mutated” POPC molecule with net negative charge was created. The system was neutralized with Na^+ or Cs^+ counterions with different

parameters, equilibrated for 10 ns with position restraints on all heavy protein atoms, and for another 100 ns without any restraints. The simulation parameters for the restrained and free equilibration simulations were identical as described in section 4.2. The last frame of the free simulation was further used for the TI calculations. The preparation of the system (adding bulk histidine, short equilibration), as well as the TI simulations of 10 ns duration and the subsequent analysis, were performed as described in section 4.7.1.

For the simulations with a mixed POPE/POPG (3:1) lipid bilayer, GROMOS-CKP parameters for POPE and POPG were used [107]. The parameter files were acquired from the Lipidbook repository [179] and modified as necessary to be used in combination with the Amber-99SB*-ILDN protein force field. Hydrogen atoms were modeled as virtual sites in order to allow for an integration step of 4 fs. A solvated lipid patch with 96 POPE and 32 POPG lipid molecules equilibrated for 100 ns, was downloaded from <http://www.softsimu.net/downloads.shtml> [180, 181]. The simulation box was increased for ~ 2 nm in the z -direction and additional water molecules were added respectively. The system was minimized, equilibrated for 5 ns with position restraints on all heavy lipid atoms, and another 5 ns without any restraints. All simulation parameters were identical to the ones described for the restrained and free simulations in section 4.2, except for the cut-off for the LJ interactions, which was here set to 1.4 nm. Afterwards, two sets of simulations were performed, one set using the TIP3P water model and another using the SPC water model [182]. In each set, simulations with different counterions (Na^+ and Cs^+) and their parameters [178, 183, 184] were performed. For each water model/counterion combination, the system was first equilibrated for minimum 200 ns, and then two production simulations of 100-500 ns were performed, which were finally concatenated and used for analysis of the electrostatic potential. The simulation parameters were the same as described for the free simulations in section 4.2, except for the cut-off for the LJ interactions (which was here set to 1.4 nm), and the pressure coupling parameters. Namely, the Berendsen barostat was used in the equilibration simulations, and the Parrinello-Rahman barostat was used in the production simulations, in all cases with a coupling constant of 2 ps. The electrostatic potential across the box was calculated using the `g_potential` module.

The system of NirC in a POPE/POPG bilayer was constructed in the following way. The resulting lipid patch from the above-described simulations from the TIP3P/ Na^+ [178] combination was used to build three times bigger patch (295 POPE and 88 POPG lipid molecules) by duplicating and stitching the small patch. The system was equilibrated with position restraints on all heavy lipid atoms for 1 ns, and without any restraints for 50 ns. The simulation parameters were same as the one used for the equilibration and production simulations of the small patch, respectively. The protein was embedded into the big lipid patch using `g_membed`. The system was neutralized by removing the appropriate number of Na^+ ions, minimized, and equilibrated for 30 ns with position restraints on all heavy protein atoms, and for another 300 ns without any restraints. The simulation parameters were the same as for the restrained and free simulations in section 4.2. All central histidine residues were doubly-protonated in the equilibration simulations. The TI calculations were performed same as for the protein/POPC systems, described in section 4.7.1. The TI simulations were ran for

10 ns. As before, the non-perturbed central histidine residues were singly protonated. $\Delta\Delta G_{\text{prot}}$ was calculated using the `g_analyze` module, by averaging the resulting $\partial H / \partial \lambda$ values from each simulation (discarding the first 1 ns for equilibration, and estimating the error using block averaging), and subsequent integration over λ using the trapezoid rule.

4.8 GMCT CALCULATIONS

Preparation and GCEM pre-calculations were same for both, ΔG_{prot} and protonation probabilities calculations. For all GMCT data shown in this thesis, the crystal structure of the proteins was used. Hydrogen atoms, atomic radii, and atomic charges were added according to the CHARMM22 force field. All histidine, arginine, lysine, glutamate, aspartate, cysteine, and tyrosine residues were considered protonable in the protonation probability (titration) calculations. For the free energy calculations, the same residues plus all threonine and serine residues were considered protonable.

MEAD/GCEM parameters: The dielectric constant of the solvent was set to 80. The solvent-accessible volume in the protein was determined with a spherical probe with 1.4 Å radius, and the dielectric constant in these protein cavities was also set to 80. The membrane region was divided in three layers: membrane core with dielectric constant of 2, and polar lipid-head regions with thickness of 5 Å and dielectric constant of 20. The dielectric constant of the protein (except the cavities) was set to 4. The ionic strength of the solution was set to 150 mM, the temperature to 298.15 K, and the ion exclusion layer to 2 Å.

GMCT parameters: All GMCT calculations were performed with the Metropolis MC method, with temperature set to 298.15 K and interaction energy limits of 1 kcal mol⁻¹ for pair moves, and 2 kcal mol⁻¹ for triplet moves. For the free energy calculations, the free energy perturbation method combined with the Bennet-Pande method [185, 186] was used, including statistical error tolerance of 0.02 kcal mol⁻¹, a staging procedure with two chimeric intermediates, and multiple simulations according to the multi-move simulation scheme. Each simulation consisted of 1000 MC scans for equilibration, and another 1000 for production. For the protonation probability calculations, 5000 MC scans were used for equilibration, and another 10000 scans for production. The proton motive force was calculated from the transmembrane potential and the pH gradient according to equation 5.2.

4.9 PMFS FOR SIMULTANEOUS PROTON/ANION INTERNALIZATION

The free energy profile for simultaneous internalization of anion and hydronium ion into the pore was calculated using the umbrella sampling method along a “sum of distances” reaction coordinate that I implemented into Gromacs. The reaction coordinate ξ was defined as the sum of distances D of two particles a and b with coordinates $\vec{r}_a = (x_a, y_a, z_a)$ and $\vec{r}_b = (x_b, y_b, z_b)$ to a reference particle with coordinates $\vec{R} = (x, y, z)$:

$$\xi = D_a + D_b = \sqrt{(\vec{r}_a - \vec{R})^2} + \sqrt{(\vec{r}_b - \vec{R})^2}. \quad (4.12)$$

A scheme of the reaction coordinate is found in the Results part, under figure 7.6. Given the definition of umbrella potential: $\omega_i = 0.5K(\xi - \xi_i)^2$, the resulting force $\vec{\mathbf{F}}_a = (F_x, F_y, F_z)$ on particle a is derived as following:

$$\vec{\mathbf{F}}_a = -\frac{\partial \omega_i}{\partial \vec{\mathbf{r}}_a} = -\frac{\partial \omega_i}{\partial \xi} \frac{\partial \xi}{\partial \vec{\mathbf{r}}_a}, \quad (4.13)$$

where evaluating both derivatives on the right side finally yields:

$$\vec{\mathbf{F}}_a = -K(\xi - \xi_i) \frac{\vec{\mathbf{r}}_a - \vec{\mathbf{R}}}{|\vec{\mathbf{r}}_a - \vec{\mathbf{R}}|}. \quad (4.14)$$

Here, ω_i represents the umbrella potential centered at ξ_i . The force acting on particle b can be derived analogously. Similarly, one-dimensional reaction coordinate can be defined, in cases when only the distance along one direction is of interest. The above equations are in that case simplified as follows:

$$\begin{aligned} \xi &= D_a + D_b = \sqrt{(x_a - x)^2} + \sqrt{(x_b - x)^2} \\ F_{x_a} &= -K(\xi - \xi_i) \frac{x_a - x}{\sqrt{(x_a - x)^2}}, F_{y_a} = F_{z_a} = 0. \end{aligned} \quad (4.15)$$

Then, the force related to the umbrella potential acts on the particle in only one direction. In this work however, only the 3D implementation of the “sum of distances” coordinate was used.

The distance pull code in Gromacs 4.6 was modified to accommodate the “sum of distances” reaction coordinate in one and three dimensions. Moreover, it was combined with the dynamic reference group code (which had to be modified) in order to enable simulating multiple pairs of particles with respect to multiple reference groups.

4.9.1 Validation

The implementation was tested by employing a system of non-interacting particles. A scheme of this system can be found in the Results part under figure 7.6. The freedom of movement of such particles is then dictated only by entropy, to the extent allowed by the umbrella potential. In such case we have:

$$\Delta G_{\text{pmf}} = -T\Delta S, \quad (4.16)$$

where the entropy $S(\xi)$ of a state with $W(\xi)$ microstates can be calculated with respect to the reference state ξ_0 as:

$$S(\xi) = S(\xi_0) + k_b \ln \frac{W(\xi)}{W(\xi_0)}. \quad (4.17)$$

The free energy difference between states ξ and ξ_0 is given by:

$$\Delta G_{\text{pmf}} = -k_b T \ln \frac{W(\xi)}{W(\xi_0)}. \quad (4.18)$$

Taking into account the definition of ξ as given by equation 4.12, then we can write the following for the 3D case:

$$W(\xi) = \int_0^\xi 4\pi D_a^2 4\pi (\xi - D_a)^2 dD_a = \frac{75\xi^5}{19}. \quad (4.19)$$

By substitution in equation 4.18, we finally get:

$$\Delta G_{\text{pmf}} = -k_b T \ln \frac{\xi^5}{\xi_0^5}. \quad (4.20)$$

Analogously, the function of the PMF for the 1D case can be obtained, where $W(\xi) = \xi$, leading to $\Delta G_{\text{pmf}} = -k_b T \ln \frac{\xi}{\xi_0}$.

4.9.2 PMFs for simultaneous internalization of formate and hydronium ions in NirC

PMFs for simultaneous internalization of formate and hydronium ions into the NirC central binding site were calculated using umbrella sampling along the reaction coordinate described above (3D implementation). The starting structures for the umbrella simulations were taken from equilibrium MD simulations performed as described in section 4.2, from 10 ns intervals after minimum of 140 ns of equilibration. Two sets of starting structures were considered: one set was taken from a NirC HIS0 simulation, and another from a NirC HIS+ simulation. In the latter set, the central histidine residues were morphed from HIS+ to HIS0 form prior to the umbrella simulations. The parameters for the hydronium ion were taken from reference [187], slightly modified by inclusion of virtual sites, in order to allow for a 4 fs integration step (the used parameters can be found in chapter 12). For each umbrella window i , the substrates were inserted in the pore from the opposite or same side of the central histidine along the z -axis, such that the condition $\xi_i = D_a + D_b$ is fulfilled. Here, D_a and D_b are the distances of the centers of mass of the ions to the C_ϵ atom from the central histidine in the respective pore, which was also used as the reference atom for the distances during the umbrella simulations. As $\xi_i = D_a + D_b$ has many solutions, the ions were inserted at distances $(\xi_i/2) \pm d$, where d is a random distance between 0 and 5 Å. A step of 0.4 Å between umbrella windows, and a force constant of 8000 kJ mol⁻¹ nm⁻² was used. Only one umbrella window in each of the monomers was simulated per umbrella simulation. Similarly to the single-substrate umbrella simulations for the PMFs for permeation, a flat-bottom potential restraining the ions to a cylinder with radius of 7 Å with a force constant of 4000 kJ mol⁻¹ nm⁻² was applied. Per PMF, 150 umbrella simulations were ran for 5 ns each.

The PMFs were calculated using the Gromacs WHAM implementation `g_wham`, similarly to the PMFs for permeation. The first 1 ns was removed from the umbrella simulations, after which the umbrella histograms were extracted. No periodicity was imposed to the PMFs. The integrated autocorrelation times were calculated and implemented in the WHAM procedure, with prior smoothing along the reaction coordinate with a Gaussian filter with width of 2 Å. The statistical error was estimated by applying the Bayesian bootstrap procedure using 50 bootstraps. The average PMFs over all

monomers were also calculated by g_wham, by taking into account the simulation data from all monomers.

The single-ion permeation PMFs for formate and hydronium ion were calculated in an identical procedure as described in section 4.4. The starting structures for the umbrella simulations were taken from the second set used for the “sum of distances” umbrella simulations. To compare to the case where the ions enter the pore from opposite sides, the sum of the single-ion PMFs was calculated as $\Delta G(\xi) = \Delta G(z)_{\text{formate}} + \Delta G(-z)_{\text{hydronium}}$, where $\Delta G(z)_{\text{formate}}$ and $\Delta G(-z)_{\text{hydronium}}$ are the PMFs for permeation of formate and hydronium ion, respectively, at z and $-z$ points along the reaction coordinate. Similarly, to compare to the case where the ions enter the pore from the periplasmic side, the sum of the single-ion PMFs was calculated as $\Delta G(\xi) = \Delta G(z)_{\text{formate}} + \Delta G(z)_{\text{hydronium}}$. In both cases, $\xi = |2z| + 0.5$, where the small 0.5 nm shift is necessary due to the fact that z is one-dimensional, while ξ is three-dimensional coordinate. The shift is minor, since the movement of a substrate across the pore is mainly along z . Prior to the summation, the single-ion PMF for formate was transformed such that $\Delta G(z) = \Delta G(z = -1.26)$, when $z < -1.26$, and $\Delta G(z) = \Delta G(z = 1.18)$, when $z > 1.18$, in order to account for the fact that formate always entered the pore first in the “sum of distances” umbrella simulations, even at ξ_{max} .

4.10 QM/MM CALCULATIONS

The QM and QM/MM calculations related to the imidazolium/anion proton transfer were performed with the quantum chemistry software Orca [188] and its combination with Gromacs 4.6, respectively. Here, only the technical details of the calculations are presented. For a detailed description of the methods, test simulations and results, please refer to chapter 8.

4.10.1 Optimization of the QM parameters

Test QM calculations were performed on a minimal system consisted of 4-methylimidazolium (4-MEI) and a formate or nitrite anion (fig. 8.1a). The initial screening for accuracy/speed of multiple combinations of QM methods and basis sets was performed using a rough reaction coordinate defined as the position of the proton on the line connecting the N_δ atom from the imidazolium ring and one oxygen atom in the formate ion. Single point energies without geometry optimization were calculated at 11 equidistant points along this coordinate. The calculations were performed in implicit water, using the COSMO(water) solvation model as implemented in Orca.

A relaxed surface scan (RSS) enables a more precise calculation of the energies along a reaction coordinate, by restraining the degree of freedom of interest, and relaxing (optimizing) all other degrees of freedom in the system. For the RSS calculations involving the minimal 4-MEI/formate or 4-MEI/nitrite systems, the distance between the 4-MEI N_δ and H_δ atoms was fixed, while the rest of the system was relaxed, and the potential energy was calculated. The RSS in vacuum was performed by sampling 25 equidistant points along the reaction coordinate from 0.8 Å to 2.1 Å. For the RSS in

water calculations, 35 or 50 equidistant points from 0.9 Å to 2.0 Å were sampled, using the COSMO(water) model. All RSS calculations were performed using DFT with the B3LYP functional and various basis sets.

For the test QM/MM calculations, a system of capped doubly-protonated histidine residue and a formate ion in vacuum (fig. 8.1b) or TIP3P water was constructed. To ensure a similar orientation of the constituents as in the protein, the central histidine and the bound formate ion were taken from the final structure of a NirC HIS+/formate umbrella simulation, from an umbrella window at the minimum of the calculated PMF for permeation. The rest of the NirC/membrane system was removed. The histidine residue was capped with N-methyl and acetyl groups. The QM region consisted of the histidine side chain and the formate ion. The QM/MM boundary was represented by a dummy “link atom” which was constructed as a virtual site by linear combination of the C_α and C_β atoms from the histidine side chain. As such, the link atom was located on the line between these two atoms, and it is was considered as a hydrogen atom in the QM calculation, thereby closing the electron shell of the QM region. Simulations of this system in vacuum and water were performed.

The calculation of the potential energies in vacuum was performed by employing the linear transit (LT) setup. LT is analogous to RSS in a sense that the potential energy along a reaction coordinate is calculated by restraining the system at multiple points of the coordinate, and minimizing the remaining degrees of freedom. As before, the reaction coordinate was defined as the distance between the N_δ and H_δ atoms from the histidine residue. The coordinate was sampled at 101 equidistant points from 0.8 Å to 3.0 Å, by constraining the N_δ - H_δ bond, and minimizing the system. The final structure from each step was used as the starting structure for the next one. The energy was minimized using the steepest descent algorithm for 1000 steps, or until the maximum force was below $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. The LJ and Coulomb interactions were cut off beyond 1.6 nm. Long-range electrostatics were calculated using the reaction field method. The QM calculations were performed using the DFT method with B3LYP functional and the SV(P) basis set. The QM region was embedded in the MM region by employing the ONIOM scheme.

Test QM/MM calculations in water were performed on the same capped-histidine/formate system, in this case solvated with 2163 TIP3P water molecules. Here, the TI method was employed to scan the N_δ - H_δ reaction coordinate, by using the coupling constant λ to transform the N_δ - H_δ distance. Since λ in this case has units of length, integration of $\partial H / \partial \lambda$ over λ yields the PMF along this reaction coordinate. The system was first equilibrated in pure MM simulation as follows: first, a short steepest descent minimization was performed, after which the system was equilibrated for 10 ns. Weak position restraints on the N_δ atom from the histidine, and one of the oxygen atoms from the formate ion was applied in this simulation, in order to prevent dissociation of the formate. The following MD parameters were used: integration step of 2 fs, cut-off for the LJ and Coulomb interactions of 1.2 nm, PME treatment of the long-range electrostatics, with grid spacing of 0.12 nm and interpolation order of 4, weak temperature coupling at 300 K with the velocity rescale scheme with coupling constant of 2.5 ps, and weak pressure coupling at 1 bar with the Parrinello-Rahman barostat and coupling constant of 5 ps. After equilibration, the energy of the system was minimized again with

the conjugate gradient algorithm. Afterwards, a set of QM/MM minimizations were performed: first the system was minimized without any constraints using the steepest descent algorithm, after which 26 energy minimization simulations were spawned, in which the $N_\delta-H_\delta$ bond was constrained at equidistant values from 0.9 to 2.1 Å. For all QM/MM minimizations, the QM software was used only to calculate the energy gradients of the atoms in the QM region, while the optimization was performed by Gromacs. Furthermore, the B3LYP method with the SVP basis set was used, with the electrostatic embedding scheme. The minimized structures were finally used to start three sets of TI calculations. For each PMF total of 26 TI simulations each with duration of 15 ps was performed. The stochastic dynamics integrator was used with integration step of 2 fs and coupling constant of 0.1 ps. The temperature was set to 300 K. In the MM region, all bonds were constrained using the SETTLE algorithm (water molecules) or the LINCS algorithm (the rest of the system). In one set of TI simulations, no cut-offs were used for the LJ and Coulomb interactions, and no periodic boundary or pressure coupling was applied. In the other two sets, the LJ and Coulomb interactions were cut off beyond 1.6 nm, where in one of the sets the long-range electrostatic interactions were treated by the reaction field method with dielectric constant of 78. In these two sets, the pressure was set to 1 bar by coupling to the Berendsen barostat with coupling constant of 1 ps. In all TI simulations, the QM region was treated with B3LYP/SVP level of theory, and the electrostatic embedding of the QM region was used. The resulting $\partial H/\partial \lambda$ values from each simulation were averaged by `g_analyze` (discarding the first 2 ps for equilibration, and estimating the error using block averaging), and ΔG was obtained by subsequent integration.

4.10.2 Protein QM/MM simulations

The starting structures for the NirC/POPC and VcFocA/POPC QM/MM calculations were taken from the end frames of the permeation umbrella simulations (corresponding to the energy minima identified in the PMF curves). For each of the NirC/formate, NirC/nitrite, and VcFocA/formate combinations, two different starting structures were used for the production QM/MM simulations.

Test simulations to estimate the effect of the MD and QM parameters were performed only for the NirC/formate combination. Prior to running the QM/MM simulations, the system was shortly minimized and equilibrated for few hundred picoseconds in a pure MM simulation. The following MD parameters were used: integration step of 2 or 4 fs, cut-off for the LJ and Coulomb interactions of 1 nm, PME treatment of the long-range electrostatics, with grid spacing of 0.12 nm and interpolation order of 4, weak temperature coupling at 300 K with the velocity rescale scheme with coupling constant of 2.5 ps, and weak pressure coupling at 1 bar with the Parrinello-Rahman barostat and coupling constant of 5 ps. Afterwards, the system was optionally minimized in an MM and/or QM/MM setup, using B3LYP/SVP level of theory for the QM region. Test QM/MM free simulations were ran for several picoseconds, employing the following parameters: a molecular dynamics or stochastic dynamics integrator with integration step of 1 or 2 fs was used, the temperature was set to 300 K by applying the velocity-rescale ther-

mostat with constant of 2.5 ps, or the stochastic dynamics thermostat with constant of 0.1 ps, the pressure was set to 1 bar by semiisotropic coupling to the Parrinello-Rahman barostat with constant of 2 or 5 ps, the LJ and Coulomb interactions were cut off beyond 2 nm, and the long-range electrostatics were treated with PME or reaction field. In the thermostat/electrostatics tests, the QM region was described with B3LYP/aug-cc-pVDZ level of theory, while in the bond length tests additionally the B3LYP/SVP level of theory was used. The QM region was always embedded in the MM region using the electrostatic embedding scheme.

The production QM/MM simulations were performed according to a similar procedure. The starting structure originating from an umbrella simulations was first minimized and shortly equilibrated in pure MM simulations, using the same parameters as for the test simulations described above. Optionally, the system was minimized in a QM/MM simulation using the steepest descent algorithm, in which the QM calculations were performed only to obtain the energy gradients of the atoms in the QM region (B3LYP/aug-cc-pVDZ), while the optimization step was performed by Gromacs. Afterwards, multiple replicas of production QM/MM simulation of various duration were performed, with the following parameters: stochastic dynamics integrator with integration step of 1 fs was used and with coupling constant of 0.1 ps, the temperature was set to 300 K, and the pressure was set to 1 bar by semiisotropic coupling to the Parrinello-Rahman barostat with constant of 2 ps, the LJ and Coulomb interactions were cut off beyond 2 nm, dispersion correction for the energy and pressure was applied, and all bonds in the MM region were constrained using the SETTLE algorithm (water molecules) or the LINCS algorithm (the rest of the system). For the QM region the B3LYP/aug-cc-pVDZ level of theory was used, and the electrostatic embedding scheme was applied.

Part III

RESULTS

ENERGETICS OF PERMEATION ACROSS THE FNTS

USING currently available structural data on the FNTs as a starting point, the determinants of their permeation mechanism were studied by employing atomistic MD simulations. Initially, a simulation system was set up for all FNTs of known structure and long equilibrium simulations were performed. Certain dynamical features could be extracted from the resulting data. Further, the free energy barriers for permeation of physiologically relevant substrates across the FNTs were estimated by calculation of potentials of mean force (PMFs) for full permeation events. The resulting free energy profiles give an overview of the energetics of permeation of different substrates across the FNTs.

5.1 MOLECULAR DYNAMICS OF THE FNTS

For each of the five FNTs with known structure, two simulation systems representing the protein embedded in a POPC membrane patch were set up, one in which all five central histidine residues were singly protonated (HIS0), and one in which they were all doubly protonated (HIS+). Two replicas of 500 ns were simulated for each of the ten systems. With exception of the *StFocA* simulations, the system seems to equilibrate within 100 ns, which is evident from the protein backbone root mean square deviation (rmsd) convergence (fig. 5.1). The lowest rmsd compared to the crystal structure is observed for NirC and HSC (~ 1 - 1.5 Å), regardless of the protonation state of the central histidine. Within the FocA group, the *EcFocA* and *VcFocA* structures are more stable (equilibrium rmsd of ~ 1.5 - 2 Å) than the *StFocA* structure, which does not seem to equilibrate in 500 ns (rmsd > 2.5 Å).

The protein root mean square fluctuation (rmsf) during the simulation, decomposed per residue, provides a general overview of the flexibility of separate regions in the protein. In figure 5.2, the crystal structures are color coded by average residue B-factors calculated from the fluctuations observed in simulation. As anticipated, the largest fluctuations in NirC, HSC, and *EcFocA* are observed in loop and terminal regions, as well as in the surface residues. The other two FocA structures exhibit largest flexibility overall, where additional to the loop/surface regions, also the Ω -loop (*VcFocA* and *StFocA*) and the N-termini (*StFocA*) are highly flexible.

The rmsf of the central histidine and the residues forming the periplasmic and cytoplasmic constrictions are shown in figure 5.3 (mean and standard deviation over all monomers). For illustration, the rmsf values of all residues in the FNTs typically range from minimum 0.38 Å, to more than 1.5 Å in high-flexibility regions, such as loops

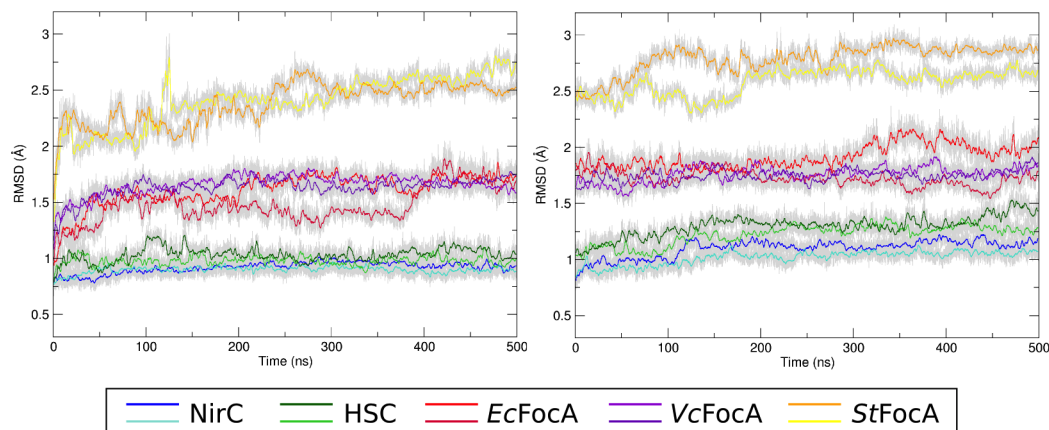


Figure 5.1: Root mean square deviation of the protein backbone from the crystal structure. Running averages shown with colored lines (see legend), and raw data shown as gray background. Left: HIS0 simulations. Right: HIS+ simulations. (Note: the starting structure for all HIS+ simulations, with exception of NirC, was taken after certain equilibration in a HIS0 system).

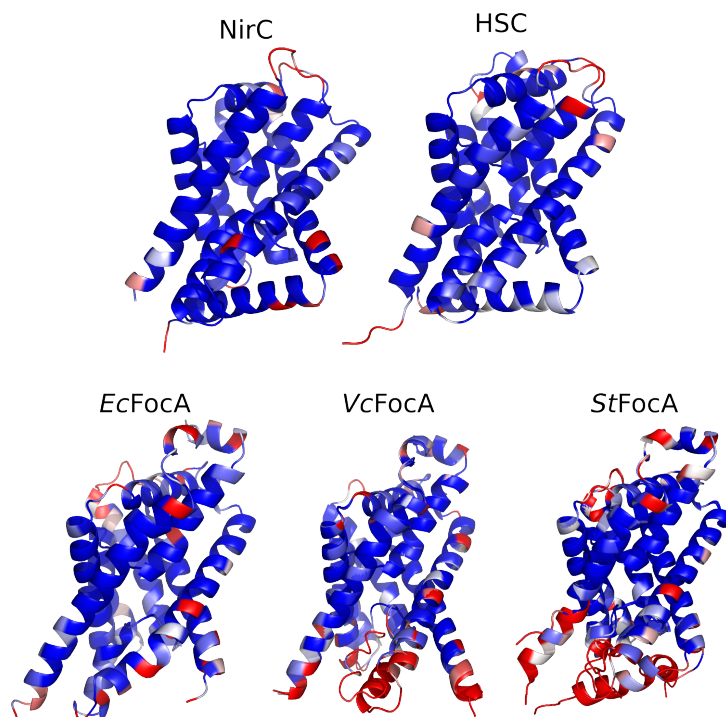


Figure 5.2: Average residue fluctuations in free simulation. The monomer crystal structure is shown as a ribbon diagram, colored by B-factors, from blue at 15 \AA^2 to red at 50 \AA^2 . One representative monomer is shown for NirC, HSC, and EcFocA, and overlay of two monomers are shown for VcFocA and StFocA.

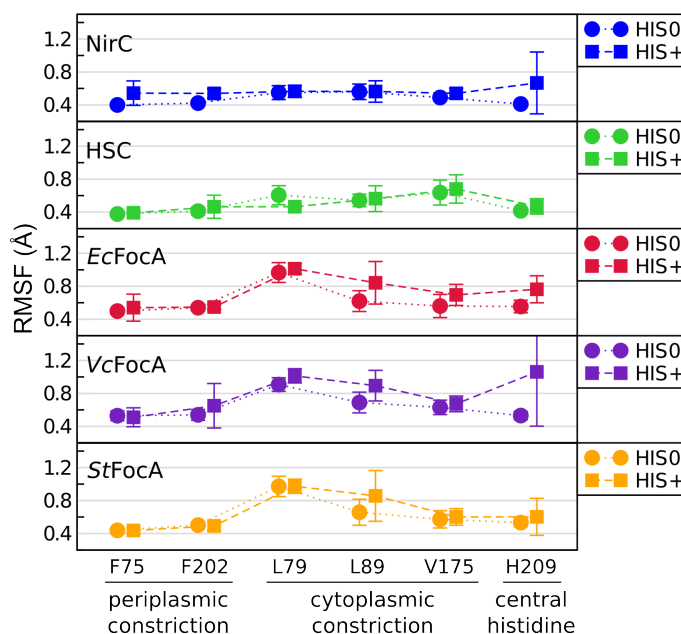


Figure 5.3: Root mean square fluctuations of the residues forming the constriction sites, and the central histidine. Each point represents the mean and standard deviation over all monomers from two simulation replicas. The residue numbering is consistent with *EcFocA*.

and termini. In all FNTs, the periplasmic constriction exhibits minimum fluctuations in equilibrium conditions. In FocA, the cytoplasmic constriction seems slightly more flexible, with Leu78/79 exhibiting the highest fluctuations, to a similar extent in all monomers. Curiously, Leu88/89 seems to fluctuate less than Leu78/79, even though it is located on the Ω -loop. When it is positively charged, the central histidine exhibits higher fluctuations, which is especially pronounced in *VcFocA* and *NirC*. The fluctuations in this residue are probably indicative of changes in orientation that occur in the range of tens to hundreds of nanoseconds, and therefore do not occur in all monomers within simulation time, leading to the high standard deviation. For example, in the *NirC* simulations, in only one HIS+ monomer from one of the simulation replicas, the central histidine assumes a flipped orientation, with the side chain pointing towards the periplasm.

An evident region with high flexibility in FocA is the Ω -loop connecting the trans-membrane helices TM2a and TM3. Notably, this region shows less variability in *EcFocA*, compared to the other FocA structures (fig. 5.4). In *VcFocA*, the Ω -loop exhibits larger fluctuations and assumes different orientations within the pore. However, the putative “open” and “closed” states from the crystal structure (nomenclature as in [85]) are not maintained, suggesting this region is more dynamic than anticipated, which could in turn play a role in permeation.

As previously mentioned, the *StFocA* structure exhibited the biggest deviation from the one determined experimentally, and did not converge within 500 ns. The regions of highest flexibility in these simulations were the N-termini and the wider Ω -loops (parts of which had to be modeled due to their absence in the crystal structure). As evident

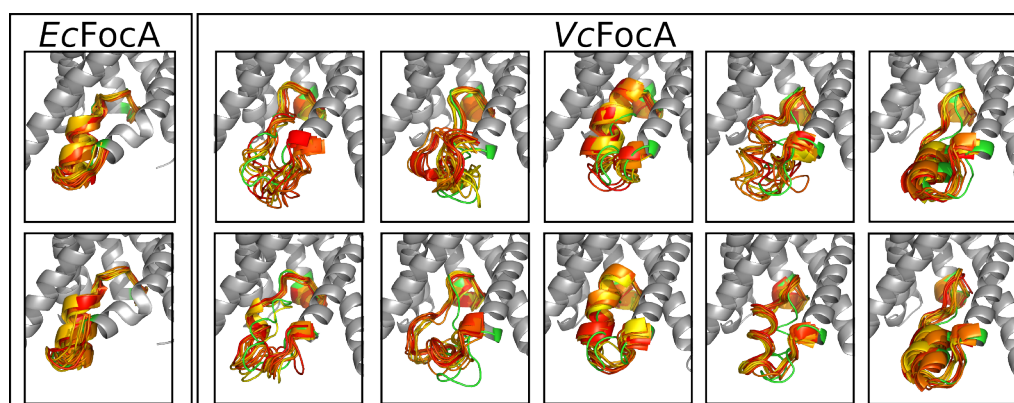


Figure 5.4: Flexibility of the Ω -loop in FocA. Protein shown in gray. Ω -loop shown as an overlay of the crystal structure (green), and simulation snapshots from 40 ns intervals (yellow to red). Each image represents one monomer. For *EcFocA* only two monomers from one HIS0 simulation are shown. For *VcFocA* all monomers from both HIS0 simulation replicas are shown.

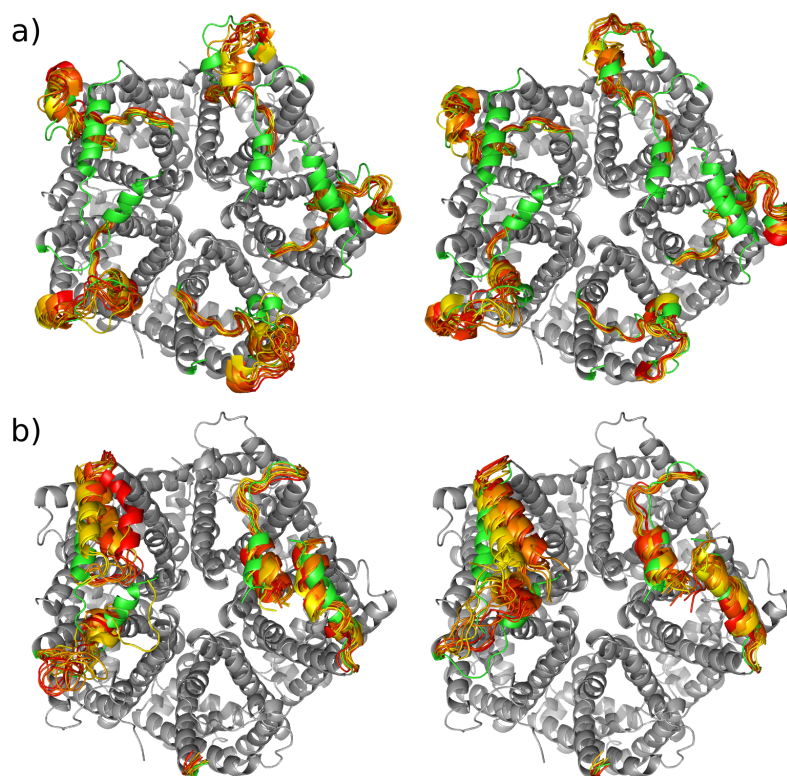


Figure 5.5: Flexibility in *StFocA*. Bottom view of the protein shown in gray. a) Ω -loops and b) N-termini shown as an overlay of simulation snapshots from 40 ns intervals (yellow to red). The Ω -loop (a) and N-termini (a,b) from the crystal structure are colored green. Each column represents one HIS0 simulation replica.

from figure 5.5b, the N-termini do not strictly retain the “closed” and “intermediate” orientations from the crystal structure (nomenclature as in [86]). The N-termini from two neighboring monomers do interact with each other, however, the overall conformation deviates significantly from the crystal structure in two out of four “closed-intermediate” pairs of monomers from two simulations. The N-termini occasionally form transient interactions with the peripheral portion of the Ω -loop, which are more stable in cases when the N-terminus adopts an orientation close to the “intermediate” from the crystal structure.

Taken together, the dynamics of the FNTs in equilibrium and in absence of permeating substrate, reveal that the FocA channels exhibit higher flexibility than NirC and HSC. This is mainly due to the fluctuations of the Ω -loop and the N-termini, on the scale of tens to hundreds of nanoseconds. The crystal structure of these regions is generally not retained in simulation, suggesting they are inherently flexible and their orientation may be highly dependent on the protein environment or other factors. The N-termini form contacts among each other and with the peripheral portions of the Ω -loops. As the Ω -loop represents a major part of the cytoplasmic vestibule, this may play a role in the permeation of substrates across FocA, as it will be discussed later. At this point, it is difficult to speculate whether certain correlated motions of the N-termini and the Ω -loop might lead to a stable “open” FocA conformation, since the N-termini are missing in two of the FocA crystal structures, and moreover, such motions could occur on longer time scales than those accessible in simulation.

5.2 THERMODYNAMICS OF TRANSMEMBRANE TRANSFER

As previously mentioned, in a minimal system, such as the one used in the PLB electrophysiology studies of FocA and NirC, these FNTs were found to translocate anions in an electrogenic manner [73, 78]. Additionally, FocA was proposed to function as a proton symporter at certain external pH [43, 72]. To account for the fact that not all physiologically relevant cases are treated in experiment, the following calculations aim to estimate whether proton co-transport is thermodynamically necessary for anion transport across the membrane in a range of physiological conditions. This was achieved by calculation of the driving force for anion transport, with or without proton co-transport.

For this purpose, the thermodynamics of transfer of a substrate molecule from one to the other side of the membrane were investigated, while excluding any assumptions about the mechanistic details of the permeation process across the membrane. The free energy difference associated with such an alchemical process in which a substrate molecule disappears on one, and appears on the other side of the membrane, can be calculated as a difference between the electrochemical potentials of the substrate in the two respective compartments [163]. For substrates that assume different protonation states, such as formate and nitrite, the electrochemical potential will depend on the total concentration and the concentration ratio of both protonation forms, and on the transmembrane potential (for details see section 4.3).

Figures 5.6 and 5.7 show the free energy of transfer of formate/formic acid from the cytoplasmic to the periplasmic space, and nitrite/nitrous acid in the opposite direction. The free energy is plotted as a function of the proton motive force and the intracellular pH, at a fixed substrate concentration equal in both compartments (figures 5.6a and 5.7a), and as a function of the proton motive force and the substrate concentration difference between the two compartments, at a fixed intracellular pH of 7 (figures 5.6b and 5.7b). Here, the term “substrate concentration” denotes the total concentration including both protonation forms of the substrate. The proton motive force can be conveniently separated into a chemical and electrical component [163, 189, 190] as:

$$\text{proton motive force} = -\frac{k_b T \ln 10}{F} \Delta \text{pH} + \Delta \Psi, \quad (5.1)$$

where ΔpH and $\Delta \Psi$ correspond to the pH difference and the electrostatic potential difference between the cytoplasmic and the periplasmic space, and k_b , F , and T represent the Boltzmann and Faraday constants, and the temperature, respectively. This allows for a separation of the influence of the transmembrane potential (left columns) and the transmembrane pH difference (right columns) on the transfer free energy.

Figure 5.6a (left column) shows that the transmembrane potential does not affect the transfer of the uncharged formic acid, while physiological (negative) values of this potential thermodynamically favor the export of formate. The effect of the transmembrane pH difference is displayed in the right column of the same figure, where it is evident that at physiological conditions (negative proton motive force, higher abundance of protons in the periplasmic space), the export of formic acid is disfavored, while the export of formate ions is favored. Similarly, figure 5.6b shows the effect of the transmembrane potential and pH difference on the transfer free energy, this time taking into account a concentration gradient across the membrane (negative $\Delta \mu$ denotes higher substrate concentration inside the cell). It is evident that at any concentration gradient, both components favor the export of formate more than the export of formic acid. Taken all together, these simple calculations confirm that in physiological conditions, formate would be the preferred exported counterpart from the formate/formic acid pair. In contrast, in case of import of the substrate, the transfer of formic acid would be thermodynamically favored over the transfer of formate.

Equivalent calculations were performed for the free energy of transfer of nitrite/nitrous acid, but in the opposite direction: from the periplasmic to the cytoplasmic space (in compliance to the physiological role of NirC). From figure 5.7a it is evident that in the physiological range, both the transmembrane potential and pH difference disfavor import of nitrite, when no concentration gradient is present. The import of nitrous acid is not affected by the transmembrane potential, and favored by the pH gradient in the physiological range. When a concentration gradient is introduced in fig. 5.7b, the voltage and pH gradients favor the import of nitrous acid in the physiological range. The import of nitrite is indeed thermodynamically allowed, but at (high) inwards-driving concentration gradients. All in all, when not accounting for the mechanistic details of permeation across the membrane, the import of nitrous acid (or a nitrite plus proton) is thermodynamically favored over the import of nitrite, in the physiological ranges for the transmembrane voltage, pH and concentration gradients.

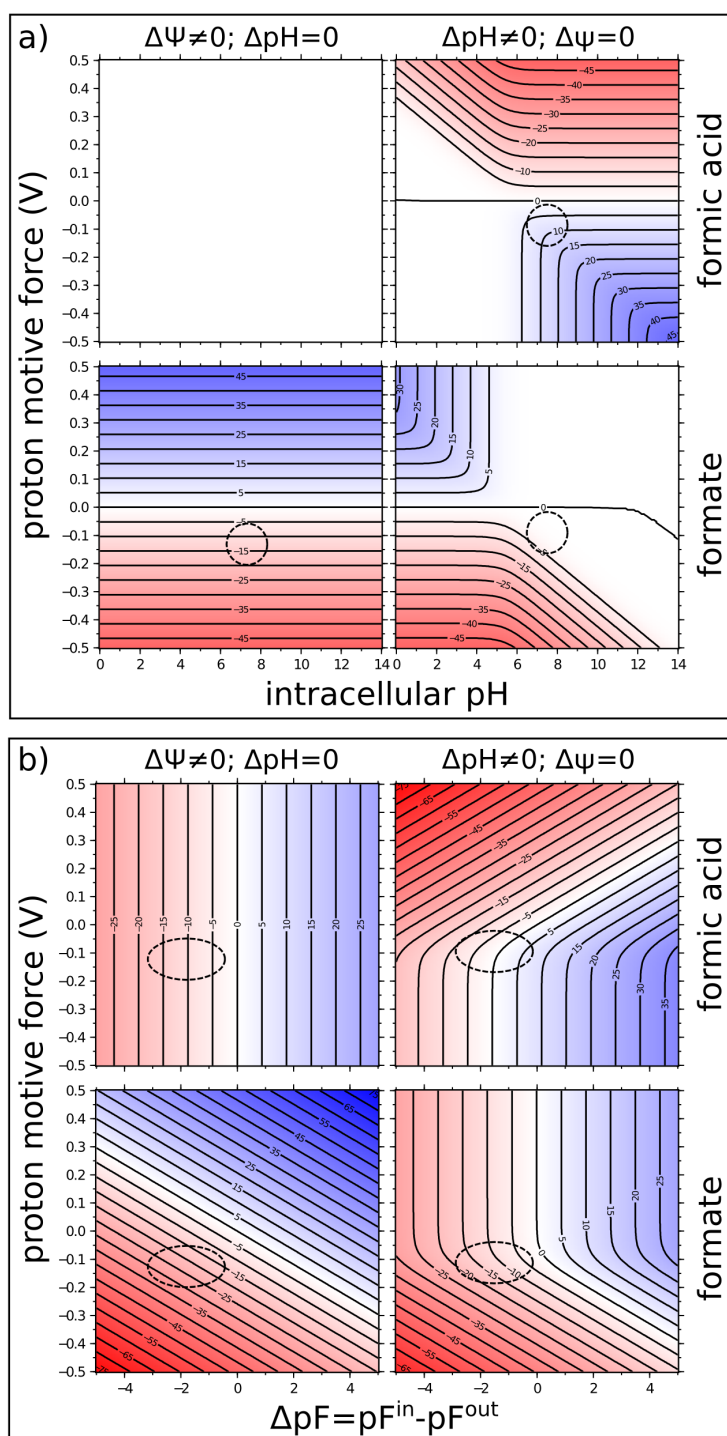


Figure 5.6: Free energy of transfer for the export of formic acid (upper rows) and formate (lower rows), shown as isocontours drawn in intervals of 5 kJ mol^{-1} , as a function of the proton motive force and a) the intracellular pH, or b) the concentration gradient $\Delta pF = pF^{in} - pF^{out}$. pF has an analogous relationship to the substrate concentration, as pH to the proton concentration (for details see chapter 4). Left column: the proton motive force consists entirely of the electric transmembrane potential $\Delta\Psi = \Psi^{in} - \Psi^{out}$ (no pH gradient); right column: the proton motive force consists entirely of the transmembrane difference $\Delta pH = pH^{in} - pH^{out}$. The dashed lines enclose a wider physiologically relevant region.

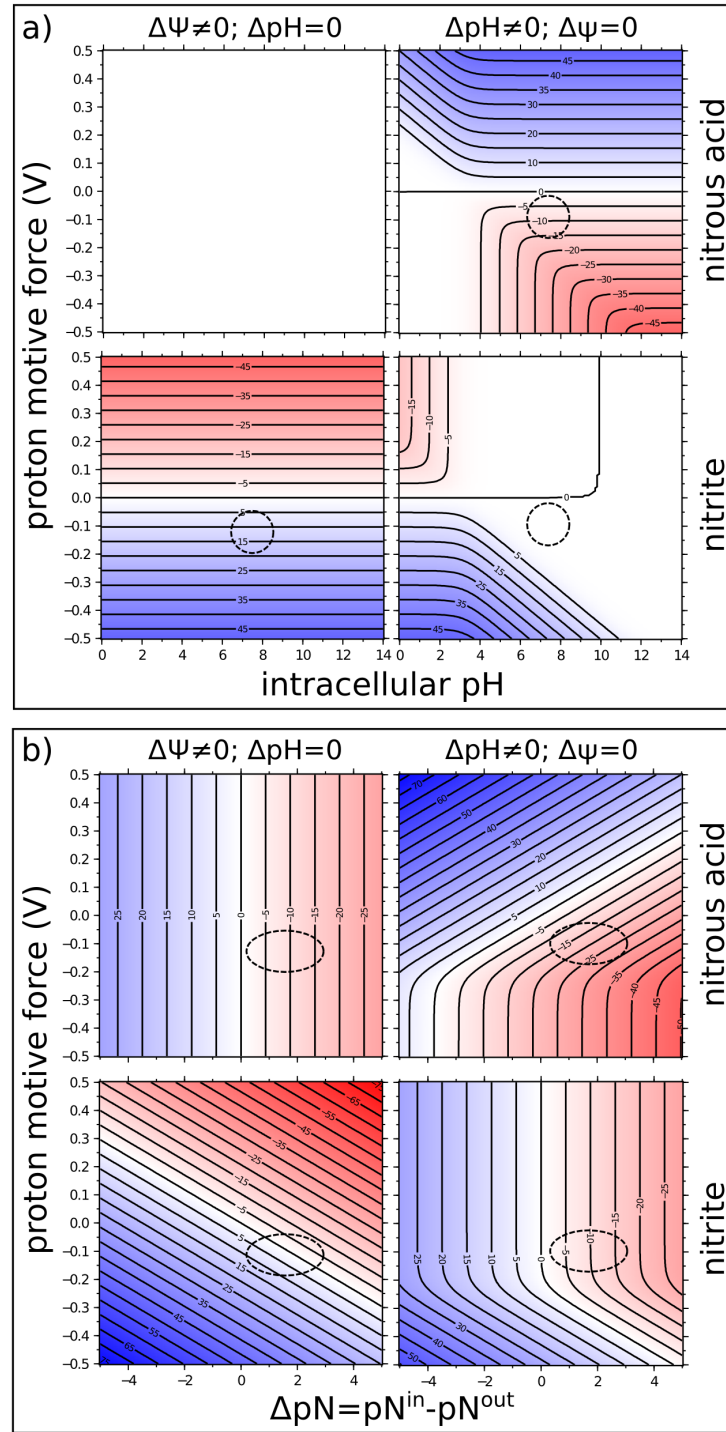


Figure 5.7: Free energy of transfer for the import of nitrous acid (upper rows) and nitrite (lower rows), shown as isocontours drawn in intervals of 5 kJ mol^{-1} , as a function of the proton motive force and a) the intracellular pH, or b) the concentration gradient $\Delta pN = pN^{in} - pN^{out}$. pN has an analogous relationship to the substrate concentration, as pH to the proton concentration (for details see chapter 4). Left column: the proton motive force consists entirely of the electric transmembrane potential $\Delta\Psi = \Psi^{in} - \Psi^{out}$ (no pH gradient); right column: the proton motive force consists entirely of the transmembrane difference $\Delta pH = pH^{in} - pH^{out}$. The dashed lines enclose a wider physiologically relevant region.

These findings show that for the export function of FocA, co-transport of proton is not only unnecessary, but also thermodynamically disfavored at physiological conditions. A different situation is valid for NirC, whose main function is nitrite import. At most physiological conditions, import of nitrite is allowed with or without proton co-transport, when sufficient concentration gradient is present, but proton co-transport is thermodynamically more favorable.

5.3 POTENTIALS OF MEAN FORCE FOR PERMEATION ACROSS THE FNTS

In order to comparatively and systematically study the energetics of permeation across the FNTs, PMFs for full permeation events of various substrates were calculated by using the umbrella sampling technique. A “full permeation event” entails moving a permeating substrate from the bulk solvent on one side of the membrane, across the channel, to the bulk solvent on the other side of the membrane. To describe this process, the reaction coordinate was defined as the position of the permeating substrate on the axis normal to the membrane (z). The physiologically relevant permeating substrates of the FNTs do not belong to the “standard” molecules included in common force-field packages, therefore, they had to be parameterized prior to the PMF calculations. Furthermore, due to the high computational cost of umbrella sampling, the simulation parameters needed to be optimized in order to achieve the best accuracy within reasonable computational time.

5.3.1 Substrate parameters validation

The following substrates were parameterized using the GAFF force field with HF/6-31* RESP charges (for details see section 4.5): formic acid, formate ion, nitrous acid, nitrite ion, hydrogen sulfide, and hydrosulfide ion (figure 5.8). The parameters were validated by calculation of the free energy of hydration (ΔG_{hydr}) using the thermodynamic integration method and comparing to experimental data.

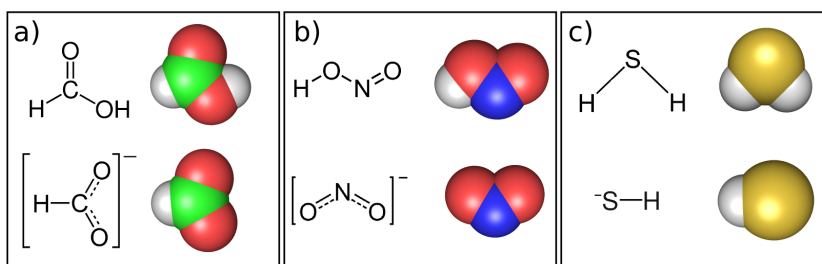


Figure 5.8: Chemical formulas and sphere representation of the FNT permeating substrates. a) formic acid (top) and formate ion (bottom), b) nitrous acid (top) and nitrite ion (bottom), c) hydrogen sulfide (top) and hydrosulfide ion (bottom).

Table 5.1: Hydration free energies of neutral substrates, decomposed by Coulomb and Lennard-Jones (LJ) contributions. Values are in kJ mol^{-1} . ^aReference [191]. ^bReference [192].

ΔG_{hydr} kJ mol^{-1}	Coulomb	LJ	Total	Exp.	Exp.-Total
formic acid	-31.64	3.93	-27.71	-29.3 ^a	-1.59
nitrous acid	-17.85	1.93	-15.92	-9.6 ^b	6.32
hydrogen sulfide	-8.18	8.21	0.03	-2.93 ^a	-2.96

Experimentally, ΔG_{hydr} of neutral solutes can be estimated directly from partition coefficient measurements, with typically low uncertainty ($< 1 \text{ kJ mol}^{-1}$) [191]. The calculated hydration free energies of the neutral solutes parameterized in this work are in good agreement with experiment, where the largest deviation ($\sim 6 \text{ kJ mol}^{-1}$) is observed for nitrous acid (table 5.1). Formic acid and hydrogen sulfide exhibit slightly less favorable ΔG_{hydr} , while nitrous acid slightly more favorable ΔG_{hydr} than experimentally measured.

Validation of the calculated parameters for the ionic substrates was far less trivial than for the neutral substrates, mainly due to two aspects: 1) limitations arising from available experimental data, and 2) technical considerations regarding the hydration free energy calculations. Regarding the former, collections of single-ion hydration free energies [191, 193–197] are limited, suffer from poor standardization, and contain values that inherently carry high uncertainties. Thermodynamic properties of single-ion solvation are not directly accessible in experiment, due to the near electroneutrality of bulk systems. Therefore, the relative contribution of a single ion in reference to another ion is typically determined in calorimetric or electrochemical measurements, using appropriate thermodynamic cycles [198, 199]. The proton is usually taken as the reference ion, with ΔG_{hydr} set to zero by convention [191]. These “conventional” free energies of hydration are however, not directly comparable to calculated quantities in simulation, which correspond to “absolute” (intrinsic) free energies of hydration. The estimation of the absolute ΔG_{hydr} values based on experimental data requires the knowledge of ΔG_{hydr} of the proton, a quantity that has long been a matter of debate, with experimental values spanning an interval of 230 kJ mol^{-1} [183, 200]. Currently, the value of $-1112.5 \text{ kJ mol}^{-1}$ [201] (corresponding to standard states of 1 mol L^{-1} in both, the gas and liquid phase) is most widely accepted, however, with uncertainty of no less than 8.4 kJ mol^{-1} [191].

The other challenge in ion parameter validation involves certain technical aspects of the method for calculation of their hydration free energies. Typically, the system is “alchemically” perturbed, such that the solute-solvent interactions are gradually decoupled (for details see section 4.5). Here, the calculated energy of charging of the solute is especially sensitive to simulation details, mainly due to finite-size effects and the approximate treatment of long-range electrostatics. In a series of publications [202–204], Hünenberger et al. systematically studied these effects and proposed four types of correction terms that should be applied to calculated ΔG_{hydr} values, in order to obtain methodology-independent values, which are comparable to experiment. Three of those corrections (B, C, and D) were applicable to the calculations done in this thesis.

Table 5.2: Effect of the box size on the hydration free energy ΔG_{hydr} . The total energy is decomposed by Coulomb and Lennard-Jones (LJ) contributions. Values are in kJ mol^{-1} .

ΔG_{hydr} kJ mol^{-1}	Na^+				Cl^-			
Box side (\AA)	30	40	50	60	30	40	50	60
Coulomb	-366.5	-365.6	-364	-363.9	-378.7	-377.9	-377.5	-377.8
LJ	13	12.8	12.8	12.9	18.9	18.8	18.3	18.6
Total	-353.5	-352.8	-351.2	-351	-359.8	-359.1	-359.2	-359.2

Correction B arises from the difference in solvent polarization in the periodic simulation system as compared to an ideal macroscopic system. This effect depends essentially on the size of the simulation box and was tested for two standard ions. Increasing the box size from a 30^3\AA^3 -cube to a 60^3\AA^3 -cube did not significantly affect the calculated charging free energy (table 5.2), implying the finite-size effect is minimal and type B correction is not necessary. Type D correction accounts for inaccuracies in the relative permittivity of the water model and was found to be only $\sim 1 \text{ kJ mol}^{-1}$ for the used TIP3P model.

Type C correction accounts for the deviations of the electrostatic potential at the ion site from the “real” potential, which arise mainly due to the usage of particle-based instead of molecule-based summation scheme for the contribution of individual atomic charges to this potential. This correction was found to have a significant effect on the calculated charging energy and was therefore considered in the calculations. It depends mainly on the solvent properties and solute charge, it is negative for cations and positive for anions, and it was estimated to be $\pm 74 \text{ kJ mol}^{-1}$ (analytical solution) and $\pm 52 \text{ kJ mol}^{-1}$ (numerical solution) for the CHARMM-TIP3P water model [205], and $\pm 79 \text{ kJ mol}^{-1}$ for the SPC water model [183].

Instead of calculating the correction C explicitly, I calculated the hydration free energies of standard ions. If we consider that their parameters have been validated, then the difference between the calculated and experimental ΔG_{hydr} values gives the correction that arises due to the used setup for ΔG_{hydr} calculation. For this purpose, the parameters for the standard ions were taken from reference [183], since they were optimized to reproduce experimental values based on the latest proton solvation energy, using an optimization procedure that accounted for the abovementioned corrections. The resulting ΔG_{hydr} values are offset from the experimental values for $\pm 69 \text{ kJ mol}^{-1}$ on average (table 5.3), which was finally used as the correction that needs to be applied to the raw calculated ΔG_{hydr} values, in order to be able to compare to experiment.

In this way calculated and corrected ΔG_{hydr} values of formate, nitrite and hydrosulfide ions are shown in table 5.4. Both, formate and nitrite exhibit $\sim 15 \text{ kJ mol}^{-1}$ more favorable ΔG_{hydr} values compared to experiment. For nitrite, unfortunately only an old experimental hydration energy could be found, which was based on an older value for the proton ΔG_{hydr} . The “real” value is probably less negative, meaning that the deviation for this ion could be slightly higher. Only the hydrosulfide ion parameters result

Table 5.3: Hydration free energies ΔG_{hydr} of standard ions [183], decomposed by Coulomb and Lennard-Jones (LJ) contributions. Values are in kJ mol^{-1} . Experimental values are taken from [191].

ΔG_{hyd} kJ mol^{-1}	Coulomb	LJ	Total	Exp.	Exp.-Total
F^-	-514	11	-503	-437	-66
Cl^-	-395	14	-381	-311	-70
Br^-	-367	15	-352	-286	-66
I^-	-330	17	-313	-251	-62
Li^+	-470	8	-462	-538	76
Na^+	-371	12	-359	-432	73
K^+	-299	13	-286	-358	72
Average					± 69

Table 5.4: Hydration free energies ΔG_{hydr} of ionic substrates, decomposed by Coulomb and Lennard-Jones (LJ) contributions. Values are in kJ mol^{-1} . The correction of the total hydration free energy is 69 kJ mol^{-1} . ^aReference [191]. ^bReference [195].

ΔG_{hyd} kJ mol^{-1}	Coulomb	LJ	Total	Corrected	Exp.	Exp.-Corr.
formate	-406	3	-403	-334	-319 ^a	15
nitrite	-391	1	-390	-321	-305 ^b	16
hydrosulfide	-453	8	-445	-376	-302 ^a	74

in a ΔG_{hydr} significantly off from the experimental value, therefore, this substrate was exempt from the PMF calculations.

5.3.2 Optimization of the umbrella sampling protocol

The umbrella sampling method [138] in combination with atomistic MD simulations is an accurate tool for PMF calculations and has been widely used to study permeation barriers in channels [34, 174, 206, 207]. In this sense, the umbrella technique provides good sampling along the reaction coordinate, by introducing (known) bias in order to restrain the system at a certain point, while the atomistic approach provides an explicit and physical description of the studied system. However, for a good accuracy, an appropriate number of points along the reaction coordinate (umbrella windows) need to be sampled for a sufficient time, and the choice of these parameters highly depends on the studied system and process, as well as on the reaction coordinate. For big membrane protein systems involving slow transitions, this can lead to very high computational cost of the method. To reduce this cost to an affordable level, and still achieve good accuracy, multiple umbrella sampling parameters had to be optimized. They are schematically represented together with the reaction coordinate in figure 5.9.

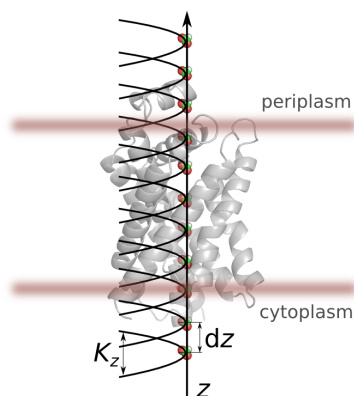


Figure 5.9: Scheme of the reaction coordinate z for the PMFs for permeation. The reaction coordinate is defined as the position of the substrate (shown as spheres) along the z axis normal to the membrane plane, with respect to the transmembrane parts of the pentamer. For clarity, only one monomer is shown. The substrate is restrained with harmonic potential with force constant K_z at multiple points along z with interval dz .

First, the force constant K_z of the biasing harmonic potential needs to be strong enough to enable sampling over the barrier, but also not too strong, since sufficient overlap between the histograms from neighboring windows is necessary for WHAM. While force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ was already acceptable for neutral substrates, a much higher value of $4000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ had to be applied to ions. Ions tend to interact stronger and therefore need a stronger restraining potential, especially in regions of unfavorable free energy (fig. 5.10a and 5.10b).

Two umbrella sampling parameters determine the overall CPU time per PMF: a) the density of umbrella windows along the coordinate (number of simulations), and b) the duration of each simulation. The reaction coordinate in question has units of length, therefore the windows density is quantified by the distance between two windows (“step-size” dz). At lower simulation times, the PMF curves lie within $\sim 3 \text{ kJ mol}^{-1}$ from each other when the step-size is below 0.2 \AA (fig. 5.10c), suggesting a step of 0.1 – 0.2 \AA is appropriate. To be on the safe end of this range, a step-size of 0.1 \AA was used for the neutral substrates, which resulted in total of 150 umbrella simulations per PMF (while using the benefit of multiple windows per simulation, since at a sufficient distance the interactions between the substrates from different windows become negligible). For charged solutes, the distance between the windows in a single simulation has to be greater, requiring also a special placement of the solutes in the system, which leads to a minimum of 200 umbrella simulations per PMF in order to achieve a step-size of 0.1 – 0.2 \AA . With such high number of simulations per PMF, the upper limit for the simulation time is in the range of 10 ns per simulation. The PMF curves vary only little ($< 5 \text{ kJ mol}^{-1}$) for simulation duration between 5 ns and 10 ns (fig. 5.10d), therefore, the umbrella simulations of the neutral substrates were ran for 5 ns, while most of those of the ions for 10 ns, in order to account for possible slower sampling of the ion motions.

As described in details in section 4.4, the entire pentamer is simulated in the umbrella simulations, and therefore, umbrella windows are considered in each of the five

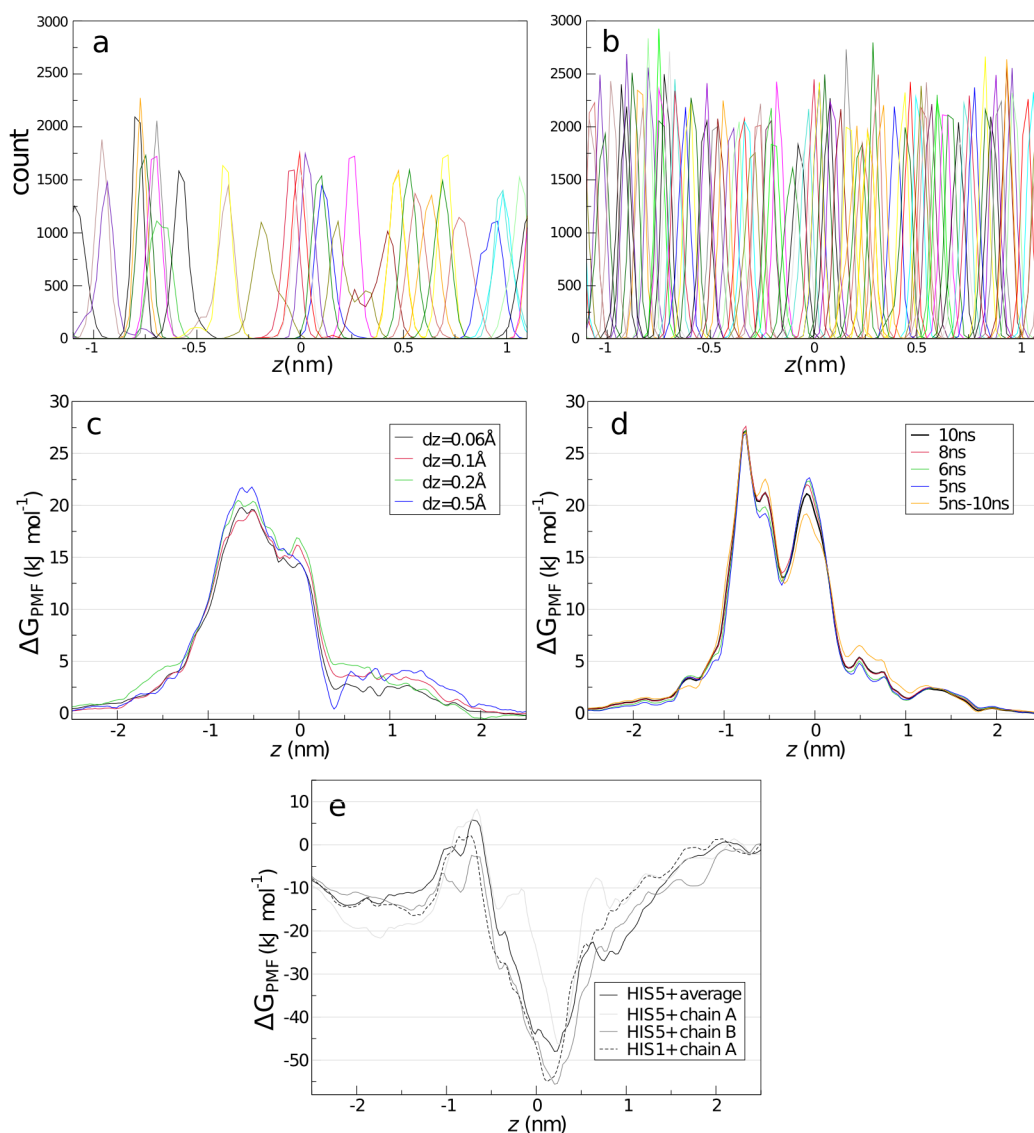


Figure 5.10: Optimization of the umbrella parameters. a,b) Force constant K_z . Probability distributions from the umbrella windows in the $-1 < z < 1$ region (substrate: chloride ion, protein: *VcFocA* HIS0, simulation duration: 0.5 ns). a) $K_z = 1000 \text{ kJ mol}^{-1}$, $dz = 0.1 \text{ \AA}$, b) $K_z = 4000 \text{ kJ mol}^{-1}$, $dz = 0.04 \text{ \AA}$. c,d) Step-size and duration. PMFs for permeation of water shown as function of c) the step-size dz , d) the duration of the umbrella simulation. c) Protein: *VcFocA* HIS0, simulation duration: 0.5 ns, five-chain average shown. d) Protein: *EcFocA* HIS0, $dz = 0.1 \text{ \AA}$, single-chain PMF of chain A shown. e) Central histidine protonation. Protein: NirC, substrate nitrite, simulation duration 10 ns. HIS5+: All five central histidine residues were doubly protonated in simulation. HIS1+: Only one central histidine was doubly protonated in simulation. Note: All PMFs are without a trapezoid correction.

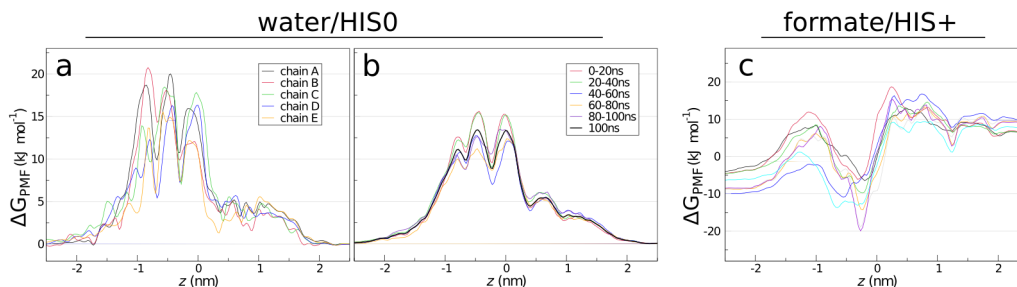


Figure 5.11: Single-chain PMF convergence. a) Single-chain PMFs. b) Five-chain average PMFs computed from different intervals from umbrella simulations with duration of 100 ns. c) Five-chain average PMFs computed from different sets of starting structures and different parameters for the umbrella simulations. Simulation parameters: a) substrate: water, protein: VcFocA HIS0, $dz=0.1$ Å and simulation duration: 5 ns; b) substrate: water, protein: VcFocA HIS0, $dz=0.5$ Å and different simulation duration (see legend); c) substrate: formate ion, protein: VcFocA HIS+, $dz=0.175$ Å, simulation duration 5-10 ns. Note: All PMFs are without a trapezoid correction.

chains. This results with sampling the reaction coordinate five times in each set of umbrella simulations. The resulting single-chain PMFs are then averaged, and a five-chain average PMF is reported. The single-chain PMFs show certain degree of variability, especially in the FocA systems (fig. 5.11a). This is reflective of the side-chain flexibility in the pore, which might sometimes not be properly sampled per single chain. Two strategies were employed to test whether increased sampling is necessary to reduce the variability between single-chain PMFs, and thereby obtain more converged average PMF: 1) significantly increasing the umbrella simulation time (100 ns), or, 2) taking multiple sets of starting structures for the umbrella simulations from different-sized intervals from equilibrium simulations (10- 450 ns). The PMFs shown in figures 5.11b,c demonstrate the averaging over the single chains produces reasonably accurate PMFs already within shorter simulation times. Namely, the average PMFs calculated from 20 ns intervals from the long umbrella simulations are all within 5 kJ mol^{-1} (fig. 5.11b). Furthermore, the PMFs calculated from multiple different sets of starting structures are all within $\sim 15 \text{ kJ mol}^{-1}$ (fig. 5.11c). In conclusion, the pore flexibility and side-chain re-orientations may indeed introduce uncertainty in the calculated single-chain PMFs, however, the average PMFs result with an overall consensus free energy profile.

Finally, for the PMF calculations in the systems with doubly protonated central histidine, the possible inter-monomer influence was investigated. It was found that the PMF does not change drastically when only one, or all five monomers have a doubly protonated central histidine (fig 5.10e). Therefore, to maximally increase sampling, as discussed above, all HIS+ PMFs were calculated from a system in which all five central histidine residues were doubly protonated.

In summary, the optimization procedure led to the following protocol: 1) for neutral substrates, 150 umbrella simulations of 5 ns with inter-window step-size of 0.1 Å and force constant for the umbrella potential of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ were performed per PMF calculation, and 2) for ionic substrates, 200 umbrella simulations of 5- 10 ns with

inter-window step-size of 0.1-0.2 Å and force constant for the umbrella potential of 4000 kJ mol⁻¹ nm⁻² were performed. This protocol leads to PMFs with typically low statistical error (~ 1 -2 kJ mol⁻¹, as calculated by bootstrapping), and uncertainties which come mainly from: 1) limited sampling, and 2) uncertainties in the substrate parameters (see section 5.3.1), and are estimated to be < 5 kJ mol⁻¹ for the PMFs for permeation of neutral substrates, and < 20 kJ mol⁻¹ for the PMFs for permeation of ions.

5.3.3 PMFs for full permeation across the FNTs

An overview of the PMFs for permeation of formate, nitrite, chloride, and water across NirC, HSC, EcFocA, and VcFocA are shown in figure 5.12 as a function of the pore coordinate z , where $z = 0$ corresponds to the center of mass of the transmembrane parts of the pentamer. The PMF profiles are not only illustrative of prominent structural features along the permeation path, but also indicative of the kinetics of permeation, and can be directly used to compare permeabilities of different substrates and among different FNTs. Using concepts from transition state theory, the maximum energy barrier, ΔG_{\max} can be related to the single-channel permeability as $P = \omega e^{-\Delta G_{\max}/k_b T}$, where the pre-exponential factor ω is related to the attempt frequency for crossing the barrier, k_b denotes the Boltzmann factor, and T denotes the temperature [206]. This represents only an approximate estimation of the permeability, due to assumptions regarding the kinetics of the permeation reaction and the value of the prefactor. Nonetheless, it is instructive to compare to experimentally measured quantities. The permeation of anions across FocA and NirC has been quantified in electrophysiology experiments [73, 78] in terms of the conductance $g = I/V$, defined as the measured current I at a membrane potential V . Since the PMFs reflect permeation at equilibrium conditions, the expected equilibrium flux Φ_0 must be first estimated from the experimentally measured conductances of FocA and NirC. At symmetric concentrations of the substrate of $c = 20$ mM and voltage $V = 150$ mV, a current $I = 2.6$ pA was measured for formate across FocA [73], and $I = 3.6$ pA for nitrite across NirC [78], resulting with conductance g of 17.3 pS and 24 pS, respectively. Using rate theory, the flux $\Phi = I/q$ (where q denotes the elementary charge) can be approximated to $\Phi \approx \Phi_0 \sinh(qV/2k_b T)$. Considering near equilibrium conditions ($qV \ll k_b T$), and using $I = gV$, gives a rough estimate for the equilibrium flux $\Phi_0 \approx 2k_b T g q^{-2}$ [208, 209] of $5.6 \mu\text{s}^{-1}$ (permeation events per μs) and $7.7 \mu\text{s}^{-1}$ for formate across FocA and nitrite across NirC, respectively. The equilibrium flux can be used to calculate the channel permeability via $P = \Phi_0/C$, where C is the number concentration of the permeable substrate. This results with an estimate for the expected maximum barrier ΔG_{\max} of 6.3 kJ mol⁻¹ for permeation of nitrite across NirC and 7.7 kJ mol⁻¹ for permeation of formate across FocA (when the prefactor is approximated to the water attempt frequency of $\omega \approx 10^{-14} \text{ dm}^{-3} \text{ s}^{-1}$).

All plausible combinations for the protonation states of the central histidine and the permeating substrates were taken into account: 1) anions permeating across FNTs with neutral central histidine (fig. 5.12 middle row), 2) neutral substrates permeating across

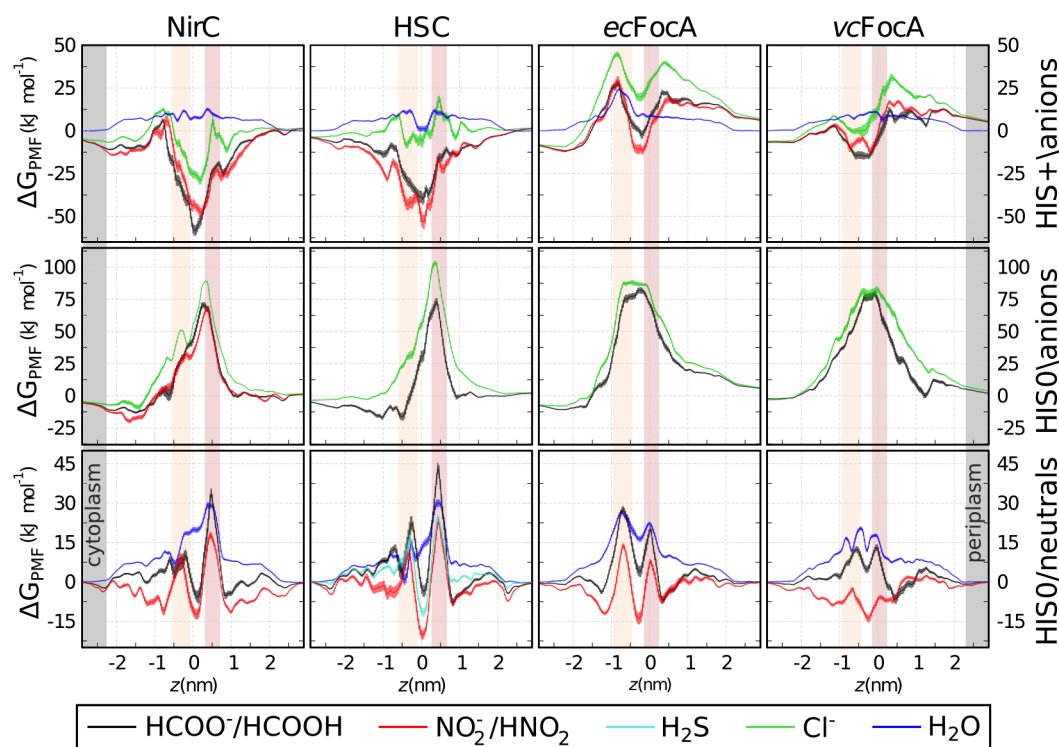


Figure 5.12: Overview of PMFs for permeation across the FNTs. PMFs from different FNTs are separated by columns (see labels on top). PMFs from different combinations of the protonations states of the central histidine and the substrate are separated by row (see labels on the right). The regions of the constriction sites are marked with orange (cytoplasmic constriction) and pink (periplasmic constriction) bars. The orientation with respect to the membrane compartments is marked with gray bars and is equal for all PMFs.

FNTs with neutral central histidine (fig. 5.12 bottom row), and 3) anions permeating across FNTs with positively-charged central histidine (fig. 5.12 top row).

5.3.3.1 Case A: Neutral central histidine vs. anionic substrates

All tested anions encounter a very high permeation barrier of 70-100 kJ mol⁻¹ when the central histidine is in its neutral form (fig. 5.13). The barrier heights are similar in all studied FNTs. The unfavorable region overlaps with the narrow pore in the channel, and seems to be shorter in NirC and HSC (~2 nm), in comparison to the FocA channels. In all FNTs, the highest barrier is encountered by the chloride anions. For illustration, the free energy barrier for cation permeation is even higher in comparison to the anions (fig. 5.14).

The small diameter and hydrophobic nature of the permeation pore cause the permeating ion to lose the hydration shell during permeation, therefore, the free energy barrier could be mainly reflective of the ion desolvation penalty. However, it is evident that formate and nitrite anions can partly compensate for this penalty, probably by favorable interactions with the pore-lining residues, keeping in mind that their solva-

tion energy is by $\sim 40 \text{ kJ mol}^{-1}$ more negative in comparison to chloride (tables 5.3 and 5.4), while the permeation barrier is typically lower. In any event, the observed barriers for all anions would practically result in no permeation across the FNTs. This is consistent with a permeation mechanism in which a proton involvement is necessary for a successful permeation.

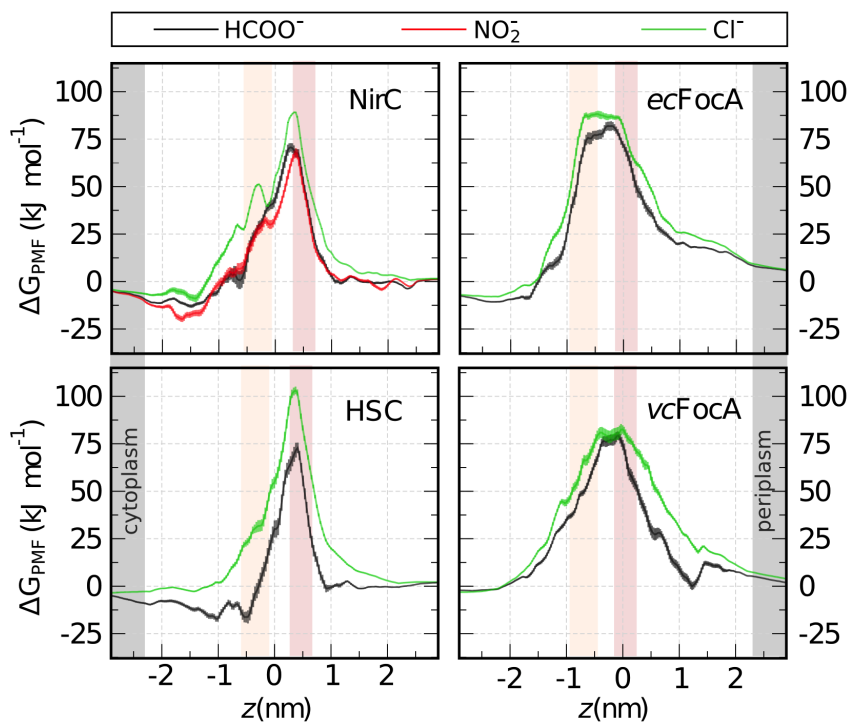


Figure 5.13: PMFs for permeation of anions across FNTs with neutral central histidine. All labels are same as in fig. 5.12.

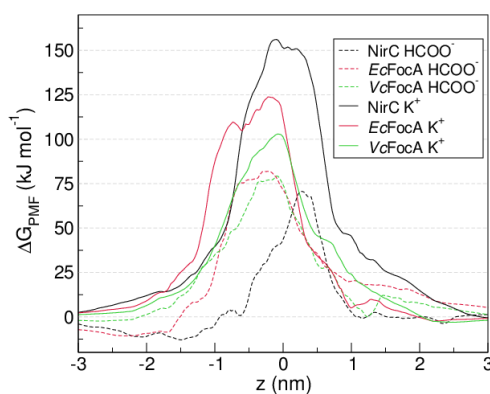


Figure 5.14: Comparison of the permeation barriers for cations and anions across FNTs with neutral central histidine. The labels in the legend correspond to the FNT/substrate combination.

5.3.3.2 Case B: Neutral central histidine vs. neutral substrates

The permeation barrier is significantly lower when the permeating substrate is protonated to its neutral form (fig. 5.15). For the permeation of the acids, the PMFs in most cases exhibit two peaks, corresponding to the two constriction sites along the permeation pathway, with an in-between well, which corresponds to the region of the central histidine. Although in this region the substrate is dynamic and does not assume any stable orientation, it frequently forms contacts with the central histidine, the only polar residue here. Example snapshots from the umbrella windows corresponding to the minima in the PMF curves are shown in fig. 5.16.

In NirC and HSC, the periplasmic constriction site seems to be the bottleneck, with ΔG_{\max} for permeation of formic and nitrous acid of 35 kJ mol^{-1} and 16 kJ mol^{-1} in NirC, and 45 kJ mol^{-1} and 22 kJ mol^{-1} in HSC, respectively. The situation is opposite in EcFocA, where ΔG_{\max} is related rather to the cytoplasmic constriction, with values of 30 kJ mol^{-1} for permeation of formic, and 15 kJ mol^{-1} for permeation of nitrous acid. The lowest energy barriers are observed in VcFocA, with ΔG_{\max} of 15 kJ mol^{-1} for formic acid, and virtually no barrier for permeation of nitrous acid. In all FNTs, the barrier for nitrous acid is consistently lower in comparison to formic acid, which might simply reflect steric and/or desolvation effects. Nitrous acid is smaller (fig. 5.8) and has a lower desolvation free energy (table 5.1) in comparison to formic acid.

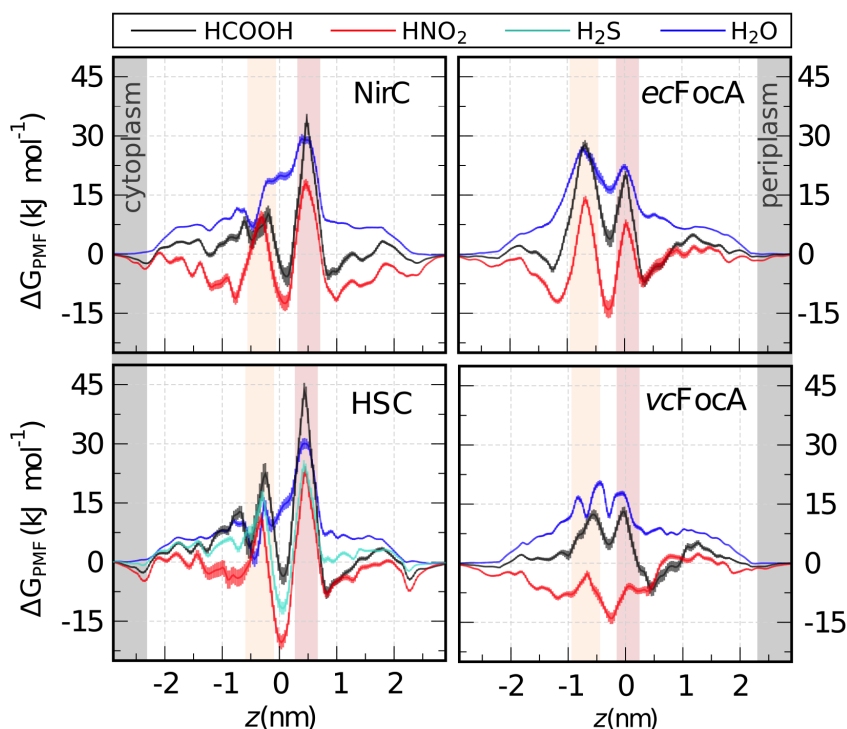


Figure 5.15: PMFs for permeation of neutral substrates across FNTs with neutral central histidine. All labels are same as in fig. 5.12.

These data show that a neutral substrate would in theory permeate easier than its anionic counterpart across a HIS0 FNT in equilibrium conditions. The observed maximum free energy barriers are however still too high to account for the permeation rates observed in experiment, especially when the pK_a of these substrates is taken into account. With respective pK_a values of ~ 3.8 and ~ 3.4 , formic and nitrous acid are mainly found in deprotonated form at physiological pH [71], therefore, the translocation of the neutral species would not significantly contribute to the permeation. Moreover, a permeation of a neutral solute across the channel would generate no signal in electrophysiological studies. In summary, this group of PMFs provide insight into the molecular details of possible protein-substrate interactions, but by no means do they account for a full permeation mechanism across the FNTs.

Water permeation is characterized with an energy barrier of around 30 kJ mol^{-1} , with the exception of *VcFocA*, where ΔG_{max} for water permeation is around 20 kJ mol^{-1} . This corresponds to water permeability of $3.3 \times 10^{-18} \text{ dm}^3 \text{ s}^{-1}$ for *VcFocA*, which is almost an order of magnitude lower than the typical water permeability of aquaporins of $\sim 5 \times 10^{-17} \text{ dm}^3 \text{ s}^{-1}$ [210]. This is to be expected if we take into account that water permeation barrier in aquaporins has been estimated to be only 13 kJ mol^{-1} [211]. Hence, in a state with a neutral central histidine, FNTs are not efficient water channels. This is also consistent with the equilibrium MD simulations of the FNT HIS0 systems, where the permeating pore was almost never occupied by water.

5.3.3.3 Case C: Positively-charged central histidine vs. anionic solutes.

The permeation picture entirely changes when the extra proton resides on the central histidine, rendering it positively charged (fig. 5.17). The anions experience a relatively strong binding to the imidazolium ring, with ΔG_{min} for formate and nitrite in the range of -60 to -30 kJ mol^{-1} in NirC and HSC, and -15 kJ mol^{-1} in *VcFocA*. Moreover, this binding site seems to act as a trap for anions, especially in the case of NirC and HSC, virtually eliminating any barrier for their internalization. This is not the case in *EcFocA*, where the first peak at approx. $z = -1$ shows that the cytoplasmic constriction still poses a significant barrier for anion permeation, suggesting a necessity of a flexible Ω -loop for a successful permeation across the FocA channels. Chloride ions also bind to the central binding site in all FNTs, albeit significantly weaker than formate and nitrite. Similarly to case A, this implies that FNTs might indeed be specialized to form more favorable interactions with these anions. Example snapshots from the umbrella windows corresponding to the minima in the PMF curves are shown in fig. 5.16, where formate and nitrite are shown to interact with the $N_\delta\text{-H}_\delta$ part of the central histidine. With exception of *EcFocA*, water seems to permeate all FNTs with a positively-charged central histidine, with permeation barriers comparable to aquaporins [211].

To summarize, the permeation barrier for anions across the FNTs is canceled or significantly decreased when the central histidine is positively-charged, mainly due to 1) the strong favorable interaction of the permeating anion to the central histidine (hereinafter “the central binding site”), and 2) the hydration of the permeating pore, which might assist the internalization of the anions. However, it is not clear alone from

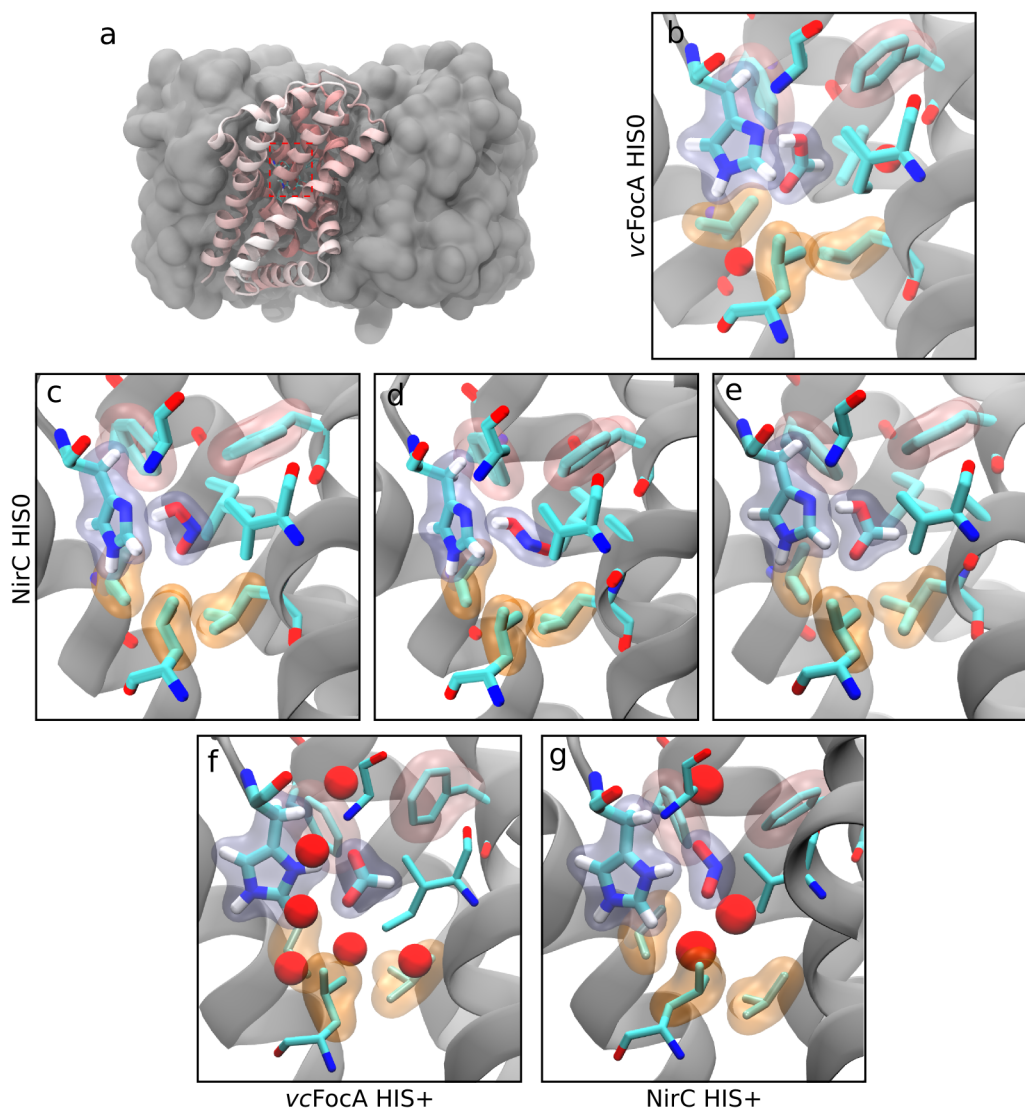


Figure 5.16: Example snapshots from the umbrella windows in the region of the central histidine. a) Global view of an FNT protein (NirC). The red dashed square shows the approximate region shown in figures b-g. b-g) Enlarged view of the central histidine region. The permeating substrate (b,e formic acid; c,d nitrous acid; f formate; g nitrite), the residues forming the constriction sites, and all residues and water molecules with atoms within 4 Å from the permeating substrate are shown as sticks (protein residues) or spheres (water molecules). Hydrogen atoms not shown, except for the permeating substrate and the central histidine. Shown as a simplified surface are the side chains of: the central histidine and the permeating substrate (purple), the cytoplasmic constriction (orange), and the periplasmic constriction (pink). The protein is shown as a gray ribbon diagram, front helices are removed for clarity. b) *VcFocA* HIS0, $z=-0.25$ nm; c,d,e) NirC HIS0, $z=0.06$ nm; f) *VcFocA* HIS+, $z=-0.25$ nm; g) NirC HIS+, $z=0.23$ nm.

the PMFs whether the anion binding is too strong to allow for full permeation across these channels.

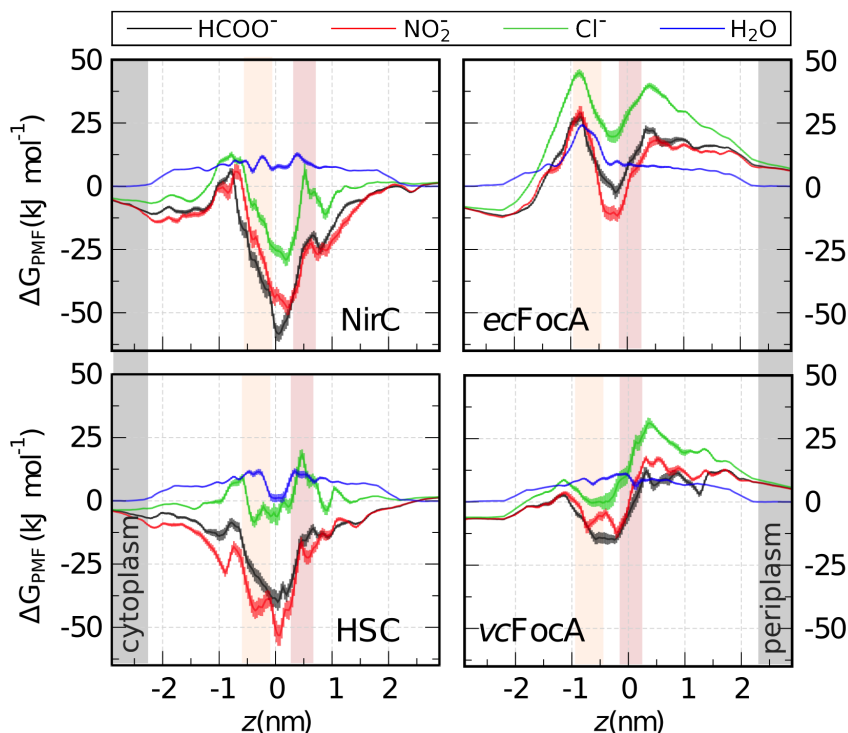


Figure 5.17: PMFs for permeation of anions across FNTs with positively-charged central histidine. All labels are same as in fig. 5.12.

Recap. The PMF calculations suggest that anions are probably not permeable across any of the FNTs. When the central histidine is neutral, anions experience a high barrier on permeation, and when it is positively charged, it strongly binds the permeating anion, especially in NirC and HSC. The permeation barriers are lower for neutral substrates, and the binding to the histidine is weaker. Permeation of neutral substrates however, would not contribute significantly to the flux at physiological conditions, in which the substrate is mostly deprotonated. Therefore, for an anion to completely permeate the HIS⁺ pore, it needs to be either expelled from the binding site in a knock-on mechanism, or alternatively, it needs to be protonated on permeation.

COMPUTATIONAL ELECTROPHYSIOLOGY OF THE FNTS

THE energy profiles for permeation across the four tested FNTs demonstrate that anions are practically not able to cross the channel when the central histidine is neutral (fig. 5.13). They also show that in a situation where this histidine is positively charged, anions are easily internalized into NirC and HSC, and into FocA when it is allowed by the Ω -loop (fig. 5.17). However, the central binding site has to be overcome for a successful permeation. One possible solution could be a “knock-on” mechanism, in which a second anion within the pore could destabilize the binding of the first anion, aiding its release from the channel. To test this possibility, the computational electrophysiology (CompEl) [32] setup was employed (for details see section 4.6). At the time of performing these calculations the electrophysiology code was only available for single-particle ions, therefore the initial tests were done with chloride and iodide ions. In order to confirm the observed phenomena with a physiologically relevant solute, a quasi-electrophysiology setup was simulated for formate ions. Only recently, the electrophysiology code became available also for multi-particle ions in a test version, therefore, the simulations were repeated for nitrite permeation across NirC and VcFocA at two membrane potentials. The central histidine was kept positively charged in all CompEl simulations.

6.1 PERMEATION OF CHLORIDE AND IODIDE

Even though in physiological conditions chloride ions are probably not permeable for the FNTs, it is still possible to extract information on the conductance by applying high (non-physiological) electrochemical gradients, which could increase the probability of observing full permeation events. In electrophysiology experiments, both NirC and FocA showed Ohmic behavior and no voltage-gating [73, 78]. This allows for usage of artificially high transmembrane potentials in simulation, in order to boost permeation. Iodide has significantly lower desolvation energy than chloride, which might further facilitate permeation. In order to achieve high electrochemical gradients, the salt concentration (NaCl or NaI) in the bulk was set to 1 M, with a charge difference between the compartments of 14 e or 16 e , which resulted with a membrane potential of 620-1080 mV for different systems. A scheme and an example image of the double system is shown in fig. 6.1. Permeation across protein A corresponds to a permeation from the periplasmic to the cytoplasmic space (hereinafter “direction A”), while permeation across protein B corresponds to the opposite direction (“direction B”). Five simulation

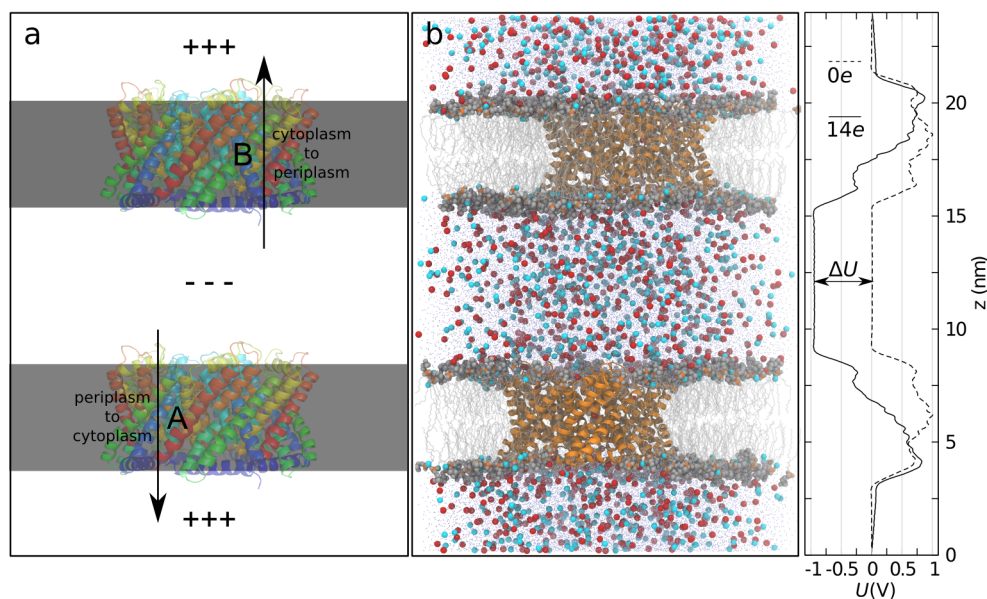


Figure 6.1: Computational electrophysiology setup. Scheme (a) and example simulation system (b) of the double membrane setup (NirC). a) The protein (ribbon diagram) has the same orientation in both membranes. “+” and “-” illustrate the relative potential of one compartment with respect to the other. Arrows indicate the direction of the anion gradient. b) Protein shown as a ribbon diagram, ions (Na^+ blue, Cl^- red), and lipids heads shown as spheres, lipid tails shown as lines. Water not shown. The potential energy of systems with and without compartment charge difference shown on the right for illustration, where ΔU corresponds to the transmembrane potential.

replicas were conducted for three FNTs (NirC, HSC, and *VcFocA*) in combination with both salts. An overview of the number of observed permeations is given in table 6.1.

Permeation across NirC. For the NirC system, the described protocol resulted with total of $2.23 \mu\text{s}$ simulation time for each salt. The charge difference between the compartments was set to $14e$, which resulted with a membrane potential of ± 900 - 1080 mV (fig. 6.2a, right). Despite the high potential, the membrane was stable and no pore formation was observed in any of the simulation replicas. The protein structure was also stable, with backbone rmsd converging to similar values as in the conventional free simulations (fig. 6.2a, left). In total of $4.6 \mu\text{s}$, only 4 full permeation events of chloride and iodide were observed, which is orders of magnitude lower than experimentally measured conductance (single-channel flux of $\sim 22 \mu\text{s}^{-1}$ for nitrite, at a membrane potential of 150 mV and symmetrical nitrite concentration of 20 mM [78]).

In direction A (periplasm to cytoplasm, fig. 6.1a, lower membrane), chloride ions enter the central binding site within 20 ns in most of the pores, and remain tightly bound for the rest of the simulation (fig. 6.3a,b,c). Similar observation can be made for iodide ions, with the difference that they usually take tens to hundreds of nanoseconds to enter the pore, possibly due to their larger size. Very often an anion is found in the periplasmic vestibule, few Å above the periplasmic constriction, interacting with Lys145 and sometimes the central histidine (fig. 6.3a,b). This binding seems to be dynamic and

relatively weak, and might correspond to the slight well at $z \approx 1$, observed in the PMF for permeation of chloride across NirC (fig. 5.17). However, the frequent presence of an ion in the periplasmic vestibule does not sufficiently destabilize the ion bound to the central histidine to enable its escape from the pore. Finally, the central histidine itself seems to be perturbed by the transmembrane potential. In this direction, the potential probably pushes the positive charge on this residue “upwards”, which on several occasions causes for the central histidine to change its orientation (“semi-flip”, fig. 6.3c) or entirely flip out (fig. 6.4d) of the binding site.

In direction B (cytoplasm to periplasm, fig. 6.1a, lower membrane), the anions generally enter the central binding site after tens to hundreds of nanoseconds, and remain bound until the end of the simulation, interacting with the $N_\delta-H_\delta$ group from the central histidine side chain. In contrast to direction A, chloride ions enter the binding site less frequently than iodide ions. Both anions sometimes enter from the periplasmic side and occupy the abovementioned “periplasmic site”, however less frequently and much shorter, since they are expelled by the electrochemical gradient.

Permeation across HSC. For the HSC system, a total of $2.23 \mu s$ and $2.25 \mu s$ were simulated with chloride and iodide salt respectively, employing a charge difference of $14 e$, which resulted with a membrane potential of $\pm 620 - 880 mV$ (fig. 6.2g,h). As before, the protein and membrane were stable and no pore formation was observed (fig. 6.2g,h). During this time, a total of 22 chloride and 16 iodide permeation events were observed, resulting in single-channel flux of $0.99 \mu s^{-1}$ and $0.71 \mu s^{-1}$, respectively (the single channels were considered bidirectional and independent, since the observed permeation events occurred in both directions and through different monomers). So far, the conductance of HSC has not been directly measured in an electrophysiology experiment. However, it was shown in a concentrative uptake assay that formate permeates across HSC with comparable, albeit lower magnitude in comparison to FocA [49]. Since the experimentally measured flux of FocA is in the order of $10 \mu s^{-1}$ [73], the observed flux in the simulations is estimated to be too low to account for experimental observations, especially when keeping in mind the very high electrochemical gradient. The fact that more permeation events of the halide ions are observed in HSC in comparison to NirC, is reflected also in the respective PMF curves. Namely, chloride ions bind much weaker in HSC than in NirC, and the observed permeation barriers can apparently be overcome at such high transmembrane potential. The physiologically relevant ions however, bind stronger to the central binding site than chloride, and are therefore in practice not expected to permeate across HSC in a “knock-on” manner.

In direction A, similar observations could be made as for NirC. In general, anions enter in the binding site in most of the monomers (chloride ions faster and more frequent than iodide ions), and there is usually a second ion loosely interacting with Lys148 in the periplasmic vestibule. The central histidine seems perturbed by the potential and is often in “flipped” (fig. 6.3d) or “semi-flipped” orientation. In direction B, the halide ions enter the central binding site in all monomers within 20-30 ns, which is faster than this direction in NirC. Occasionally, a second ion enters the same site, which can lead to a knock-on permeation, or more frequently, to a long-lived ($> 100 ns$) state of the pore with two bound ions.

Table 6.1: Number of full permeations of chloride and iodide ions across the FNTs. Direction A: periplasm to cytoplasm. Direction B: cytoplasm to periplasm. Simulation time given in μs .

N° permeations	Cl ⁻ 1 M				I ⁻ 1 M			
direction, sim. time, voltage	A	B	μs	± mV	A	B	μs	± mV
NirC	0	1	2.23	900 - 1080	0	3	2.23	820 - 1100
HSC	8	14	2.23	620 - 820	2	14	2.25	800 - 890
VcFocA	11	38	2.21	680 - 830	2	55	2.25	630 - 800

Table 6.2: Number of full permeations of formate ions across NirC and VcFocA. Direction A: periplasm to cytoplasm. Direction B: cytoplasm to periplasm. Simulation time given in μs .

N° permeations	HCOO ⁻ 1 M			
direction, sim. time, voltage	A	B	μ s	± mV
NirC	0	0	1.9	580 - 660
VcFocA	0	4	1.9	230 - 450

Permeation across FocA. From the FocA channels, the VcFocA structure was used for this analysis, since it has a flexible Ω -loop which is necessary for permeation, as evidenced in the respective PMFs (fig. 5.17). A total of 2.21 μs and 2.25 μs were simulated with chloride and iodide salt, respectively, employing a charge difference of 16 e , which resulted with a membrane potential of \pm 630-820 mV (fig. 6.2b, right). As before, the protein and membrane were stable and no pore formation was observed (fig. 6.2b, left). During this time, a total of 49 chloride and 57 iodide permeation events were observed, resulting in single-channel flux of 2.22 μs^{-1} and 2.53 μs^{-1} , respectively. Most of the permeations occurred in direction B (cytoplasm to periplasm), resulting in a chloride single-channel flux of 3.44 μs^{-1} and iodide single-channel flux of 4.89 μs^{-1} in this direction. This is at least five times lower than the experimentally measured single-channel flux of formate across FocA, which in saturation conditions, and at a much lower membrane potential of 150 mV, was found to be $\sim 25 \mu\text{s}^{-1}$ [73].

In direction A, both anions rarely enter the binding site, and when they do, they usually quickly ($< 1 \text{ ns}$) complete the permeation. This is consistent with the PMF for permeation of chloride across VcFocA, where it is evident that the binding site is more accessible from the cytoplasmic, than the periplasmic side (fig. 5.17). As in NirC and HSC, there are frequently ions in the periplasmic vestibule, interacting with Lys63 and/or Lys155, and this binding is short-lived and dynamic. The central histidine also seems to be perturbed by the membrane potential and is often in “flipped” or “semi-flipped” orientation, as shown for NirC in fig. 6.3c,d. Most of the permeations occur on a nanosecond time scale. In direction B, the binding site is occupied more frequently, and the ions stay bound usually for tens to hundreds of nanoseconds. Iodide ions on average stay bound longer than chloride ions. There are no anions in the periplasmic vestibule. The observed permeations occur on different time-scales and in a different manner.

Table 6.3: Number of full permeations of nitrite ions across NirC and *VcFocA*. Direction A: periplasm to cytoplasm. Direction B: cytoplasm to periplasm. Simulation time given in μs .

N° permeations	NO_2^- 1 M				NO_2^- 0.2 M			
	A	B	μs	\pm mV	A	B	μs	\pm mV
NirC	0	0	0.8	780 - 830	0	0	1.04	170 - 330
<i>VcFocA</i>	3	1	0.73	600 - 650	0	0	1.15	130 - 410

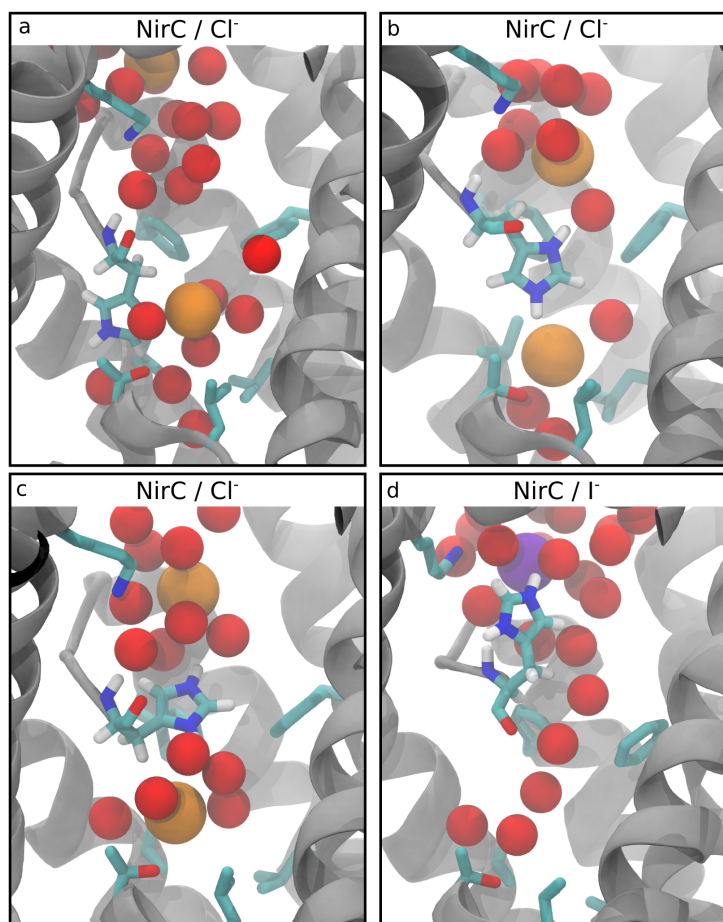


Figure 6.3: Typical snapshots of the central binding site from the CompEl simulations involving NirC and chloride or iodide ions. Protein shown as a ribbon diagram, front helices are removed for clarity. The central histidine residue, and the side chains of the residues of both constriction sites and Lys145 are shown as sticks. Only the hydrogen atoms from the central histidine are shown. Water oxygen atoms shown as red spheres, chloride and iodide ions shown as orange and purple spheres, respectively. An ion is present in the central binding site in a,b and c, and in the “periplasmic binding site” in b and c. The central histidine is semi-flipped in c, and flipped in d.

6.2 PERMEATION OF FORMATE IN A QUASI-ELECTROPHYSIOLOGY SETUP

The electrophysiology simulations of the halide ions suggest that even when the central histidine is doubly protonated, and there is a high electrochemical gradient present, the anions get internalized in the channel relatively easily, but are unlikely to undergo a full permeation. To test whether this is the case also for physiological ions, the permeation of formate across NirC and VcFocA was studied in a quasi-electrophysiological setup. Equally as in the CompEl setup, a double membrane system was used and charge difference between the compartments was introduced, however, without the ability to maintain the membrane potential in case of frequent permeation events. Two simulation replicas of $1\ \mu\text{s}$ were simulated for each of the tested FNTs, with initial charge difference of $16\ e$.

Permeation across NirC. In NirC, the membrane potential eventually equilibrated to approx. $\pm 580\ \text{mV}$ and $\pm 660\ \text{mV}$ in the two replicas respectively (fig. 6.2c). Similarly to the chloride and iodide ions, in $2\ \mu\text{s}$ no full permeation events of formate across NirC were observed. In direction A (periplasm to cytoplasm), the formate ions enter in the central binding site in all monomers within $10\ \text{ns}$ and remain bound for the rest of the simulation. They usually interact with the $\text{N}_\delta\text{-H}_\delta$ group of the central histidine (fig. 6.4a). A second ion is frequently found in the pore in the periplasmic vestibule, interacting mainly with Lys145 (fig. 6.4a). On three occasions, second ion enters the central binding site. This however does not destabilize the binding of the first ion sufficiently for it to complete the permeation, and they both stay bound for the rest of the simulation. In direction B (cytoplasm to periplasm), the central binding site in all monomers is occupied with a formate ion within $230\ \text{ns}$, which stays bound till the end of the simulation.

Permeation across FocA. Similarly, only few permeation events were observed in VcFocA (total of 4 in $2\ \mu\text{s}$), which is far below the experimentally measured flux of ~ 25 per μs per single channel in saturation conditions and at $150\ \text{mV}$. All permeations occurred in direction B within the first $100\ \text{ns}$ of the simulation, suggesting they might be result of equilibration effects. The membrane potential varied between $\pm 330\ \text{mV}$ and $\pm 450\ \text{mV}$ in one replica, and $\pm 220\ \text{mV}$ and $\pm 280\ \text{mV}$ in the other. Similarly as for the halide ions, formate ions entered the binding site easier from the cytoplasmic than the periplasmic side. A typical bound formate in the central binding site is shown in figure 6.4c.

6.3 PERMEATION OF NITRITE

Fortunately, the CompEl code with the option of using multi-particle ions has recently become available in a test version, and it was used to investigate the permeation of the physiologically relevant nitrite ion across NirC and VcFocA. Moreover, in this set of calculations also systems with lower membrane potential and salt concentration were included, in order to eliminate any effects of non-physiological gradients.

Permeation across NirC. For NirC, $0.8\ \mu\text{s}$ were simulated in a system with $1\ \text{M}$ sodium nitrite and $12\ e$ charge difference resulting in a membrane potential between

± 780 mV and ± 830 mV; and $1.04 \mu\text{s}$ were simulated in a system with 0.2 M sodium nitrite and 4 e charge difference resulting in membrane potential between ± 170 mV and ± 330 mV (fig. 6.2e). No permeation of nitrite was observed in any of the simulations. In direction A, nitrite ions enter in the central binding site in almost all monomers within 30 ns in the replicas with higher gradient, and within 80 ns in the replicas with lower gradient. Typical binding is shown in fig. 6.4b. A second ion is frequently found in the periplasmic vestibule, interacting mainly with Lys145. In the high gradient simulations, there are two ions in the central binding site on two occasions, this however does not seem to facilitate permeation. In direction B, the central binding site is very rarely occupied, and only in the higher gradient simulations, while the central histidine is flipped in few monomers.

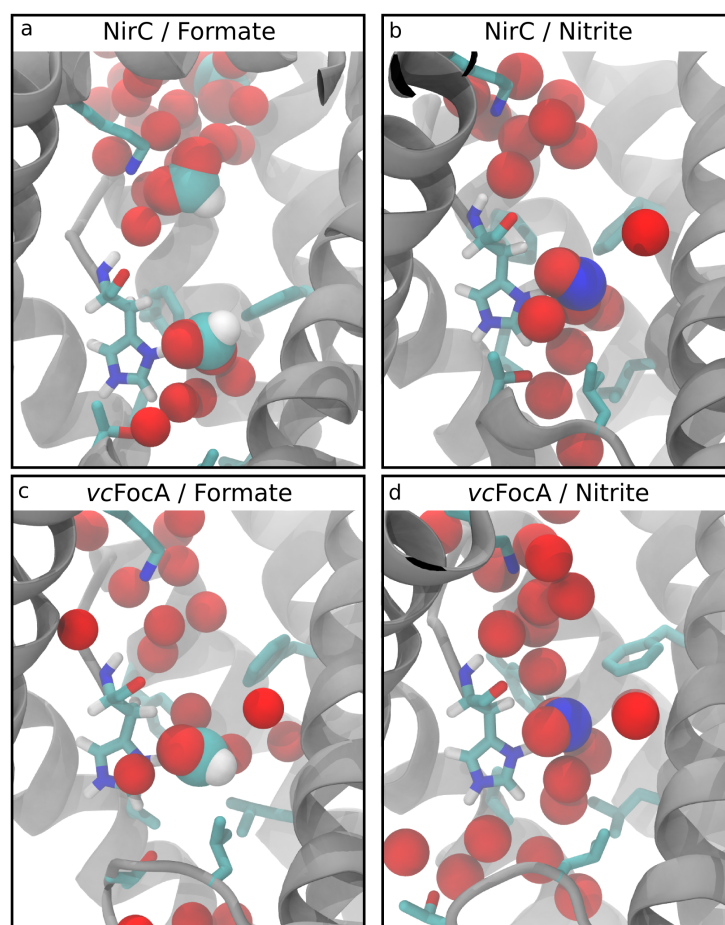


Figure 6.4: Typical snapshots of the central binding site from the CompEl simulations involving formate and nitrite ions. Protein shown as a ribbon diagram, front helices are removed for clarity. The central histidine residue, and the side chains of the residues of both constriction sites and Lys145 (NirC) or Lys155 (VcFocA) are shown as sticks. Only the hydrogen atoms from the central histidine are shown. Water oxygen atoms (red), formate and nitrite ions are shown as spheres. An ion is present in the central binding site in all figures, and in the “periplasmic binding site” in a.

Permeation across FocA. For *VcFocA*, $0.73\ \mu\text{s}$ were simulated in a system with 1 M sodium nitrite and 12 e charge difference resulting in membrane potential between $\pm 600\text{ mV}$ and $\pm 650\text{ mV}$; and $1.15\ \mu\text{s}$ were simulated in a system with 0.2 M sodium nitrite and 4 e charge difference resulting in membrane potential between $\pm 130\text{ mV}$ and $\pm 410\text{ mV}$ (fig. 6.2f). Only 4 permeation events were observed in the higher gradient simulations, and none in the ones with the lower gradient. Similarly to formate, nitrite enters the central binding site easier from the cytoplasmic than the periplasmic side. On several occasions at lower gradients, nitrite even enters the pore from the cytoplasmic side against the gradient. Less monomers are occupied by nitrite in direction A, as compared to direction B. In the lower gradient simulations, the presence of an ion in the periplasmic vestibule is significantly less frequent than in the higher gradient simulations. Overall, regardless of the direction of the gradient, more anions accumulate on the cytoplasmic than on the periplasmic side. Typical binding of nitrite in the central binding site is shown in fig. 6.4d.

Recap. The CompEl simulations demonstrate that anions are too tightly bound to the central binding site in order to achieve efficient permeation. In NirC, almost no permeations of any tested anion at any voltage were observed, while the flux observed in HSC and *VcFocA* was consistently lower even than the experimentally measured flux at much lower concentration and membrane voltage (tables 6.1, 6.2, and 6.3). In contrast to the rare full permeations, anions frequently entered the pore and bound to the central binding site. In NirC, anions entered the pore more frequently from the periplasmic side. This situation was opposite in *VcFocA* and HSC. This observation might be related to the roles the FNTs play in their physiological context (NirC - import, *VcFocA* and HSC - export). In all tested FNTs, a “periplasmic binding site” was identified, where anions were interacting with a lysine residue in the periplasmic vestibule, several Å above the constriction. The binding to this site was however much weaker, and more dynamic in comparison to the central binding site, particularly in the low gradient simulations of *VcFocA*. In certain occasions, the central histidine exhibited a flipped or semi-flipped orientation. At this point, it cannot be concluded whether such an orientation has some role in permeation, as it might be an artifact from the strong transmembrane potential.

The CompEl simulations involving physiologically relevant anions, together with the permeation PMF calculations, finally confirm that a simple mechanism of anion permeating across the FNTs is unlikely, regardless of the protonation state of the central histidine. This is consistent with a mechanism in which the anion has to undergo protonation while permeating, in order to successfully complete the permeation.

PROTONATION OF THE CENTRAL HISTIDINE

THE central histidine is generally thought to be crucial for permeation across the FNTs [40, 42, 85], and more specifically, it was suggested to have a key role in the proton-coupling to the permeation process [41, 78, 90]. To be able to achieve this task, it needs to be (at least transiently) positively-charged. This assumption is not straightforward, given this residue is located deep in the middle of the protein, within a rather hydrophobic pore. Histidine is easily protonated in bulk solution at physiological pH, however, it is well known that the pK_a values of ionizable amino acid residues can be strongly influenced by the protein microenvironment [212]. Therefore, the details of the central histidine protonation were investigated.

Calculation of pK_a of ionizable protein side chains, especially ones burried in the protein core, is known to be challenging and methodology-dependent [213]. Hence, the free energy of protonation ΔG_{prot} of the central histidine was calculated using multiple methods that employ different schemes for the description of electrostatic interactions. Computational methods used for pK_a calculations are generally divided into microscopic and macroscopic methods. Microscopic methods explicitly account for the protein flexibility, as well as for the interactions with the solvent and the protein environment. However, they can be hindered by finite sampling, particularly in proteins with high flexibility and/or slow dielectric relaxation. Macroscopic methods on the other hand, employ a continuum description of the system, thereby implicitly accounting for the protein relaxation. In this way, they do not face sampling issues, however they lose the atomic resolution. Finally, there are also combined molecular mechanics / continuum electrostatics methods, in which parts of the system can be treated explicitly, while the rest is represented by a continuum description. In this work, ΔG_{prot} of the central histidine was calculated using 1) a fully microscopic method, by employing atomistic MD simulations in combination with the thermodynamic integration (TI) method for free energy calculations, and 2) a combined molecular mechanics / continuum electrostatics method, by employing the generalized Monte Carlo titration (GMCT) method. Moreover, the effect of the membrane content on the free energy of protonation was investigated. Finally, the pathway of proton internalization in the protein was studied using another set of umbrella simulations.

7.1 FREE ENERGY OF PROTONATION BY THERMODYNAMIC INTEGRATION

The pK_a shift of the central histidine with respect to the bulk environment was calculated by alchemically transferring a proton between the central histidine in the channel

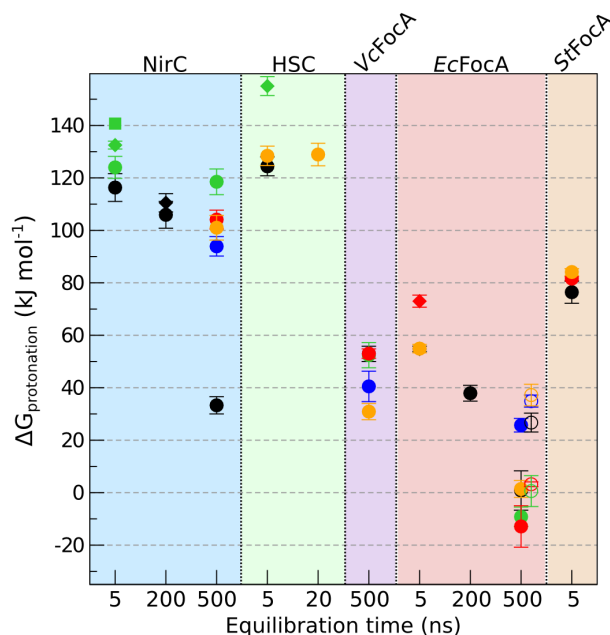


Figure 7.1: Free energy of protonation of the central histidine (ΔG_{prot}) calculated by TI. Each point on the x-axis represents a different starting structure (protein and HIS+ equilibration time) for the TI calculations. The results for different monomers of the same structure are distinguished by color. Symbols denote the duration of the TI simulations: (●) 10 ns, (◆) 20 ns, (■) 100 ns. Empty circles correspond to a second starting structure from the same protein and with the same equilibration time.

and a histidine residue in the bulk, in a set of TI simulations (for details see section 4.7). An overview of the resulting ΔG_{prot} values for all FNTs are shown in figure 7.1, which reveal a highly positive ΔG_{prot} in most cases. This suggests that the pK_a of the central histidine is shifted towards lower values, therefore, it is less probable for it to be protonated in comparison to bulk histidine. Moreover, the energy cost for protonating the central histidine tends to be higher in NirC and HSC, compared to the FocA channels.

However, these calculations alone do not allow for more quantitative conclusions. The calculated ΔG_{prot} strongly depends on the starting structure for the TI simulations, as can be observed in the calculated values from structures with different pre-equilibration time (HIS+ equilibration), or originating from different simulation replicas. The free energy varies tens of kJ mol^{-1} even between monomers from a single structure, especially in the FocA channels. These observations indicate that 1) the protonation state of the central histidine strongly depends on the orientation of the flexible side chains in the protein, and 2) the FNTs are clearly associated with long dielectric relaxation times, probably significantly longer than the longest pre-equilibration time (500 ns) or TI simulation duration (100 ns) employed in this work (which are already at the upper limit of computational feasibility).

The possibility of inter-monomer influence on the central histidine pK_a , in terms of cooperativity (or anti-cooperativity) between the protonation of this residue in discrete

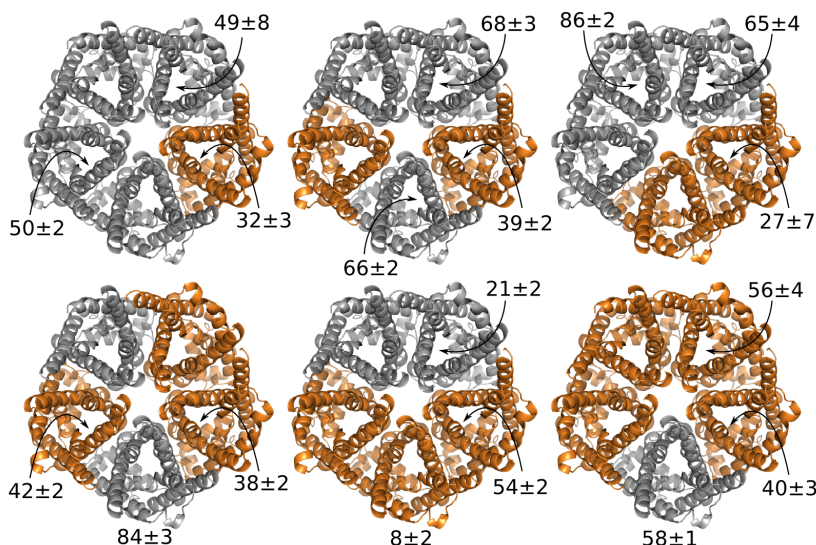


Figure 7.2: Dependence of ΔG_{prot} of the central histidine in *EcFocA* on the number of monomers with positively-charged central histidine (HIS+). The HIS+ monomers are colored orange. ΔG_{prot} of the central histidine from the monomer indicated by arrows given in kJ mol^{-1} . (Pre-equilibration: min. 200 ns. TI simulations duration: 10 ns.)

monomers of one protein, was investigated in *EcFocA*. Namely, ΔG_{prot} was calculated in different monomers in structures with different number of doubly-protonated central histidine residues (fig. 7.2). The starting structures for the TI calculations were equilibrated for minimum 200 ns with the corresponding number of HIS+ residues. However, the aforementioned relaxation effects probably outweigh any effects from the monomers interplay (if such do exist), since no consistent trends in the protonation energies could be observed.

7.2 EFFECT OF THE LIPID CONTENT OF THE MEMBRANE

Despite their shortfalls, the TI calculations provide a rough idea of the protonation probability of the central histidine in NirC, whereby the calculated ΔG_{prot} was above 80 kJ mol^{-1} in majority of the cases. This finding means that this residue is practically always in its singly-protonated state. However, additional to the protein dielectric environment, the membrane could also potentially affect the protonability of protein side chains. To account for this effect, it was necessary to exchange the model lipid bilayer of pure zwitterionic lipids (POPC), with a patch that more closely reproduces the natural environment of NirC.

The bacterial inner membrane is slightly negatively charged, therefore, a system with a 25% “artificially-charged” lipids was initially constructed from the POPC/NirC system by neutralizing the charge of the choline group, in order to avoid the necessity of the time-consuming equilibration steps. Due to known deficiencies in the description of the ion-membrane interactions [214], the simulations were repeated with different

Table 7.1: Free energy of protonation ΔG_{prot} of the central histidine in membranes containing 25% lipids with net negative charge, given in kJ mol^{-1} . Ion parameters from reference [184] (amber-ff99SB-ILDN*) ^(a), and reference [178] ^(b).

ΔG_{prot} (kJ mol^{-1})	“Artificial” system			POPE/POPG
counter-ions	$\text{Na}^{+(a)}$	$\text{Na}^{+(b)}$	$\text{Cs}^{+(b)}$	$\text{Na}^{+(b)}$
Chain A	105 ± 1	61 ± 4	99 ± 7	64 ± 3
Chain B				100 ± 5
Chain C	120 ± 5	114 ± 4	107 ± 4	62 ± 4
Chain D				109 ± 5
Chain E				82 ± 5

counter-ions and their parameters. As outlined in table 7.1, charging of the membrane did not have a detectable effect on the calculated ΔG_{prot} . Similarly to the conventional TI calculations, the variance in the calculated ΔG_{prot} values is probably due to its strong dependence on the side-chain orientations and the slow protein relaxation, now with the added component of the ion-membrane interactions, which typically need a long time to equilibrate. In order to strongly prevent binding of ions to the membrane and therefore test the upper limit of the effect a charged membrane could have on ΔG_{prot} , large cesium counterions were used in one simulation set. However, no significant difference in the calculated ΔG_{prot} was observed (table 7.1).

In order to confirm these findings in a less artificial model membrane, a system with POPE:POPG (3:1) lipid bilayer was set up. A good estimate of the effect of membrane charge on charged groups within the hydrophobic layer of the membrane can be extracted from the electrostatic potential across the membrane with respect to bulk solvent. In this regard, the electrostatic potential of a pure POPC membrane patch was compared to that of the mixed POPE/POPG patch. Long 300 ns equilibrium simulations of the solvated model membranes were performed with two different water models, SPC and TIP3P, since the former is consistent with the lipid force-field, while the latter with the protein force-field (and was therefore used in all TI calculations). As before, different ions (and ion parameters) were used to neutralize the system, among which also the larger cesium ions. Figures 7.3a and b depict the electrostatic potentials of the simulated systems with TIP3P and SPC water models, respectively, as function of the z coordinate normal to the membrane. The central histidine is located almost perfectly in the middle of the protein, so the relevant difference in the potential between the POPC and POPG/POPE lipid patches is located in the region around $z \approx 4$, and ranges from $\sim 0.1 \text{ V}$ when TIP3P water is used, to $\sim 0.2 \text{ V}$ when SPC water is used. Disregarding the effect of the protein environment, when translated to free energy difference ($\Delta G = FV$, where F is the Faraday’s constant), this potential difference results with the estimate that a positive charge would be stabilized by approx. 10-20 kJ mol^{-1} in the middle of the charged bilayer, with respect to the middle of the POPC patch.

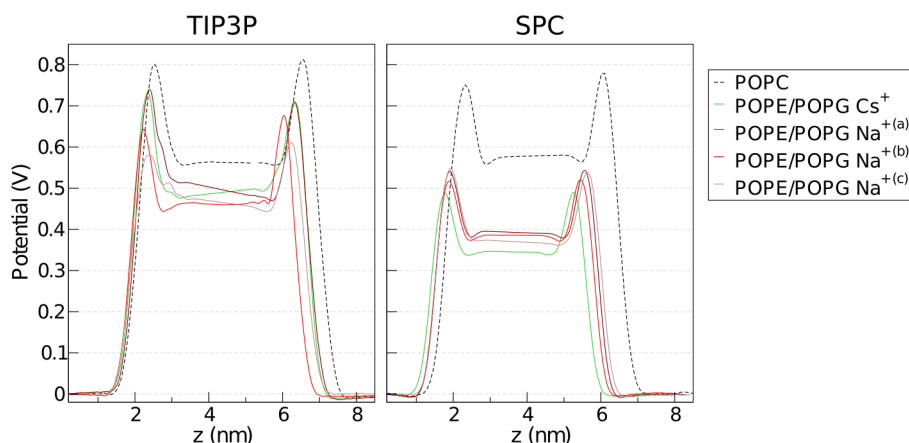


Figure 7.3: Electrostatic potential across neutral (POPC) and charged (POPE:POPG(3:1)) lipid bilayers. The charged membrane simulations were repeated with different counter-ions. Ion parameters: reference [184] (amber-ff99SB-ILDN*) ^(a), reference [183] (gromos54a7) ^(b), and reference [178] ^(c).

Finally, a set of TI calculations was performed for NirC in the POPE/POPG membrane (table 7.1). As before, there is a big variance in the calculated ΔG_{prot} values for different monomers. It is tempting to interpret the slightly lower ΔG_{prot} values in two of the monomers as a stabilizing effect of the membrane charge on the doubly protonated central histidine, with magnitude as predicted by the electrostatic potential calculations. However, this result is probably purely circumstantial, since as before, these TI calculations are suffering from issues related to the limited sampling and the slow relaxation of the protein.

7.3 FREE ENERGY OF PROTONATION BY GMCT

The TI calculations clearly demonstrate that the FNTs are associated with long dielectric relaxation times, which makes these calculations inappropriate to quantitatively estimate the thermodynamics of protonation of the central histidine. In such cases, using continuum methods is advantageous over purely microscopic methods, since they implicitly capture the contributions of the protein dielectric relaxation and are therefore not subjected to issues arising from limited sampling [215]. In this work, the generalized Monte Carlo titration method [160, 162] was used for calculation of ΔG_{prot} of the central histidine. GMCT is used for computing thermodynamic properties of binding (in this work binding of protons), based on microstate description of the system and formalism in terms of electrochemical potentials. The energy terms are calculated with a combined molecular mechanics / continuum electrostatics approach, where the receptor protein and the bound ligands (in this case protons) are described with atomic detail, while all other components are treated implicitly through their influence on the electrochemical potentials of the receptor and the ligands. The microstate description of the system accounts for the interaction between all titratable sites in the protein, the

Table 7.2: Free energy of protonation (ΔG_{prot}) of the central histidine in equilibrium conditions (no pH gradient or transmembrane potential), calculated with GMCT. The values for each monomer (A-E) and the five-monomer average with standard deviation are given in kJ mol^{-1} . The calculated Born (desolvation) energy is given in parenthesis for illustration.

ΔG_{prot} (kJ mol^{-1})	A	B	C	D	E	mean \pm s.d.
NirC	76 (47)	69 (46)	68 (47)	67 (47)	71 (46)	70.2 \pm 3.6
HSC	74 (46)	71 (46)	71 (46)	72 (46)	70 (46)	71.6 \pm 1.5
<i>EcFocA</i>	67 (47)	71 (51)	70 (54)	75 (62)	71 (49)	70.8 \pm 2.9
<i>VcFocA</i>	65 (46)	56 (45)	51 (44)	87 (67)	85 (67)	68.8 \pm 16.5

interaction with the protein “background” (all non-titratable parts of the protein), as well as the pH and transmembrane potential. It also allows for separating regions with different dielectric properties (solvent, protein, protein cavities, membrane core, and membrane polar-head regions), thereby enabling for a good description of the protein environment. For more details, please refer to section 3.5.

GMCT was used to calculate ΔG_{prot} of the central histidine in four FNTs (7.2), as well as protonation probabilities of this histidine in NirC and *EcFocA* (fig. 7.5). The results shown in table 7.2 were computed using the protein crystal structure, and correspond to equilibrium conditions, indicating absence of pH gradient or membrane voltage. As anticipated, in such conditions the energetic cost for protonating this residue is high in all tested FNTs. A large portion of this cost is evidently due to high unfavorable Born (desolvation) energy. However, in physiological conditions, as well as electrophysiological experiments, there is usually a certain proton motive force present, which could potentially shift the protonation probability of the protein side chains. To quantify this effect in NirC, ΔG_{prot} was calculated in the presence of a proton motive force (fig. 7.4), which was further decomposed into a chemical and electrical component (analogously to section 5.2). At intracellular pH of 7, ΔG_{prot} is reduced by 10-15 kJ mol^{-1} in the physiological range (proton motive force of approx. -170 mV). Such high protonation free energy cost would in practice result with an extremely low protonation probability. This is demonstrated in fig. 7.5, where the protonation probability of the central histidine in NirC and *EcFocA* is shown as a function of the intracellular pH and the proton motive force. It is evident that for the majority of pH and proton motive force values, the protonation probability of the central histidine is near zero.

Recap. Taken together, these data provide a general idea of the range of the energetic cost of protonation of the central histidine, and suggest that, in absence of other ions in the channel, this residue would probably not be positively charged. Moreover, these calculations reveal that the histidine protonability is strongly influenced by the protein microenvironment, and that FNTs have slow dielectric relaxation. The PMFs for permeation (see section 5.3.3.3) show a strong favorable interaction of an anion with the positively-charged central histidine. In NirC, ΔG_{prot} of the central histidine has a strikingly similar absolute value to the ΔG of binding as calculated from the umbrella simulations ($\sim 60 \text{ kJ mol}^{-1}$). Therefore, it is likely that the processes of histi-

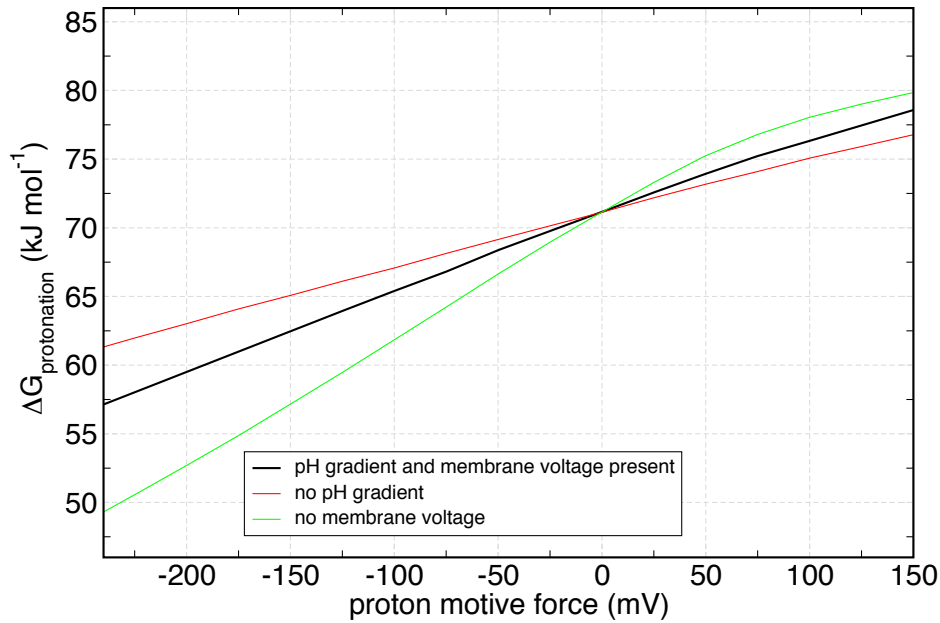


Figure 7.4: Dependence of the free energy of protonation of the central histidine in NirC on the proton motive force. The proton motive force is defined as in equation 5.2. Black: Both, the chemical (pH gradient) and electrical (membrane voltage) components are present, in ratio 1:2. Red: only membrane voltage is present. Green: only pH gradient is present.

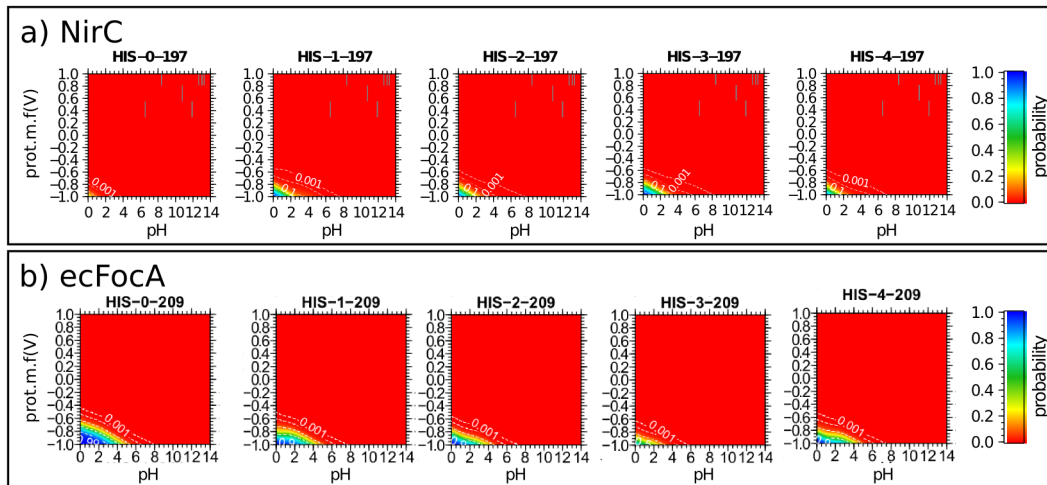


Figure 7.5: Protonation probability of the central histidine in each monomer in NirC and *EcFocA*, plotted as a function of the proton motive force and the intracellular pH. Both, the chemical and electrical components of the proton motive force are present, in ratio 1:2.

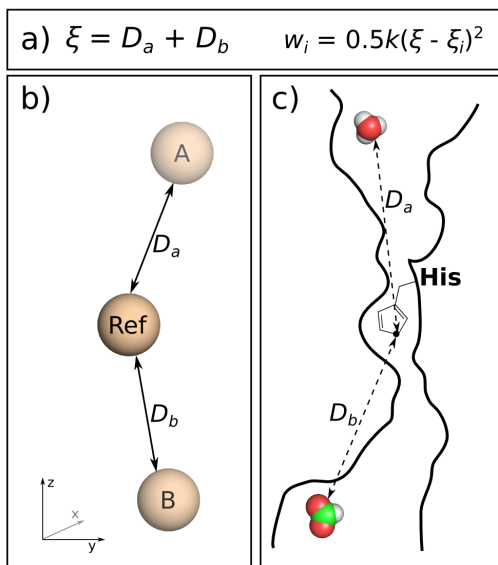


Figure 7.6: Scheme of the “sum of distances” reaction coordinate ξ . a,b,c) The reaction coordinate ξ is defined as the sum of the distances D_a and D_b of molecules A and B to the reference atom. The umbrella potential w_i harmonically restrains the sum of the distances to the reference value ξ_i . A scheme of (a) the test system, and (b) an example simulation system is shown.

dine protonation and anion binding occur in a coupled manner, resulting overall in a thermodynamically feasible state.

7.4 PROTON INTERNALIZATION IN THE PORE

To investigate whether the aforementioned state representing HIS+ protein with bound anion in the central binding site is also kinetically available, PMF of simultaneously bringing a proton and an anion into this site was calculated. The reaction coordinate ξ was defined as the sum of distances of an anion and hydronium cation to the central histidine (fig. 7.6), and, as before, the umbrella sampling technique was applied (see chapter 4 for details). This reaction coordinate had to be implemented into Gromacs, therefore, extensive initial tests were necessary prior to the production simulations.

7.4.1 Validation of the “sum of distances” reaction coordinate

The manually implemented reaction coordinate was validated on test systems containing only non-interacting particles (fig. 7.6b). One particle was the reference particle for the distances of the other particles, whose sum was kept restrained at a certain value ξ_i by a harmonic umbrella potential w_i . In this manner, the freedom of movement of the restrained particles is dictated only by the sum of their distances to the reference particle. Theoretically, this results with a free energy difference with respect to a reference state ξ_0 as: $\Delta G_{\text{pmf}} = -k_b T \ln(\xi/\xi_0)$ and $\Delta G_{\text{pmf}} = -k_b T \ln(\xi^5/\xi_0^5)$ in

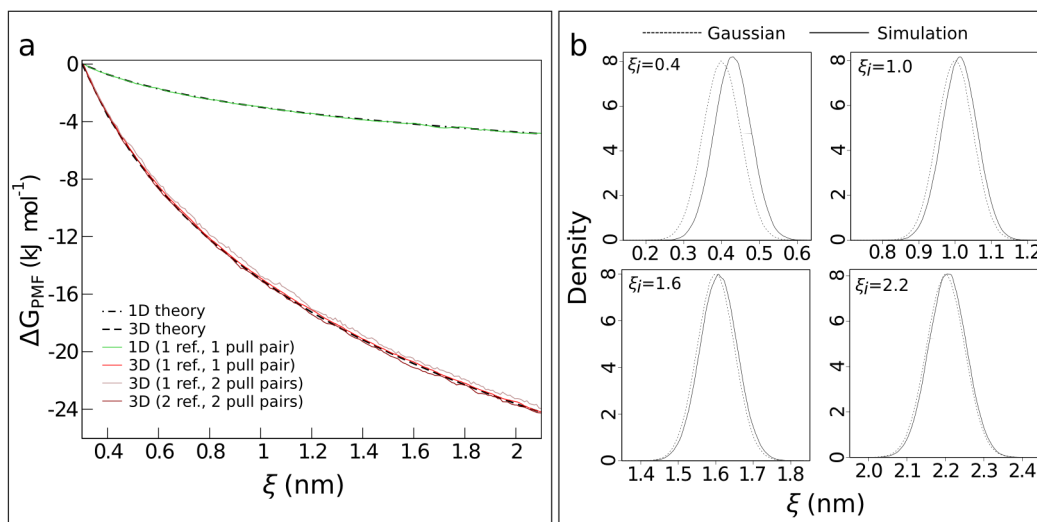


Figure 7.7: a) Test PMFs of the “sum of distances” coordinate $\xi = D_a + D_b$ in one and three dimensions, calculated on a system of non-interacting particles. b) Examples of the distribution of ξ from several umbrella windows, compared to a Gaussian distribution with corresponding center and width to the umbrella potential (simulation data corresponds to the red PMF curve).

one and three dimensions respectively (for full derivation see section 4.9). Figure 7.7a depicts the perfect overlap of the calculated PMFs with the theoretical model, which confirms the successful and accurate implementation of the reaction coordinate. This is also evident from the Gaussian distributions of ξ in each umbrella simulation (fig. 7.7b). To enhance sampling as in the umbrella simulations from the permeation PMFs, the possibility of multiple reference points was also implemented, enabling simultaneous simulation of each umbrella window in all five monomers. Therefore, test PMFs were calculated also for systems with multiple pairs of pull groups and multiple reference particles (fig. 7.7a), which in both cases match well with theory, and any fluctuations are simply due to finite sampling.

7.4.2 PMFs of simultaneous internalization of formate and hydronium ions

The newly implemented coordinate was employed to calculate the PMF of simultaneous internalization of a formate and hydronium ion into the central binding site in NirC. This was achieved for setup in which both ions approach the channel from the same, or the opposite side of the membrane. The resulting single-chain and average PMFs are shown in fig. 7.8 as a function of the sum of the distances of both ions to the C_ϵ atom of the central histidine in the respective chain. In general, the process of bringing the two ions into the central binding site as defined with this reaction coordinate is energetically unfavorable. This is especially the case when the ions approach the channel from the opposite side of the membrane, where bringing the formate ion from the periplasmic, and the hydronium ion from the cytoplasmic side, is the least favorable. In contrast,

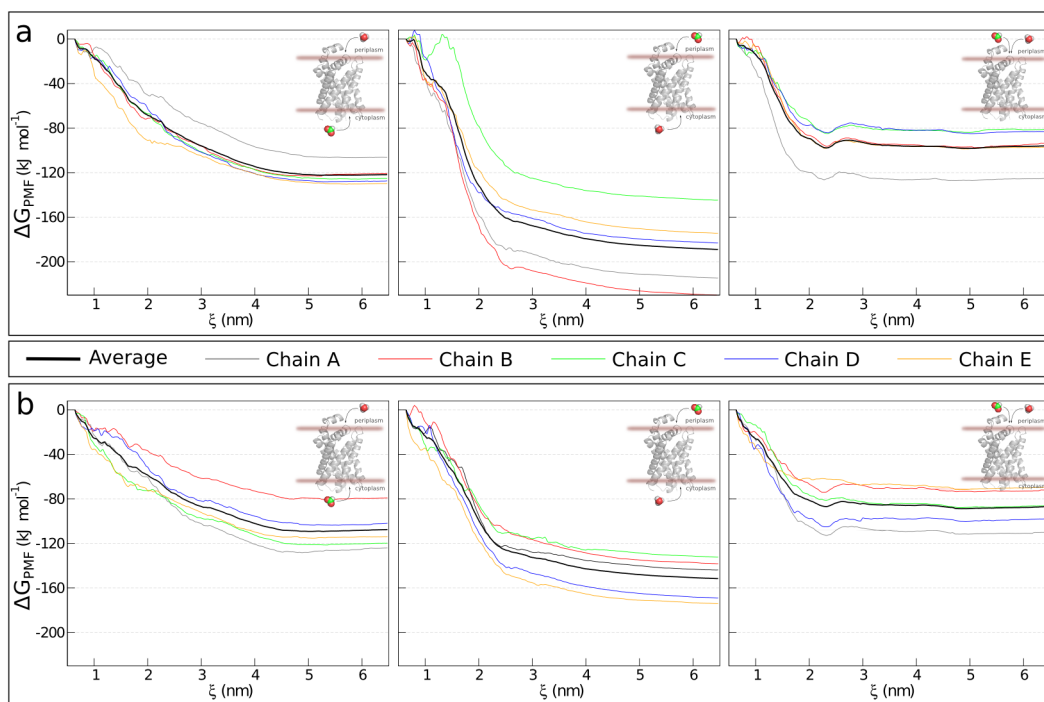


Figure 7.8: PMFs for simultaneous internalization of formate and hydronium ions into the central binding site in NirC. Colored curves represent the single-chain PMFs, while the bold black curve the five-chain average. The direction of origin of each ion is labeled in the top right image of each graph. The starting structures for the umbrella simulations originate from HIS0 (top row) or HIS+ (bottom row) equilibrium simulations. In all umbrella simulations the central histidine was neutral.

bringing both ions from the same side of the membrane seems to be significantly easier. As can be expected, the calculated PMFs are also dependent on the level of hydration in the pore. While in all umbrella simulations the central histidine was neutral, the starting structures for these simulations were taken from equilibrium simulations in which the central histidine was neutral (fig. 7.8, top row), or positively-charged (fig. 7.8, bottom row). As previously mentioned, HIS+ pores are significantly more hydrated than HIS0 pores, and this seems to facilitate the internalization of the ions by $\sim 10\text{--}30\text{ kJ mol}^{-1}$.

It is useful to compare to the single-ion PMFs for permeation of formate and hydronium ions (fig. 7.9b), in order to estimate whether the simultaneous internalization of anion and hydronium ion is of any advantage. As evident from figure 7.9a, bringing both oppositely charged ions into the pore simultaneously, indeed facilitates the process of ion internalization, since the “sum of distances” PMFs do not add up to the sum of the single-ion PMFs, regardless of the direction of origin of the ions.

However, the exact interpretation of these PMFs is not straightforward for multiple reasons. First, these calculations suffer from poor convergence, which is evident from the great variance in the single-chain PMF curves. Furthermore, a direct comparison between the PMFs with a different direction of origin of the ions is not possible, since a

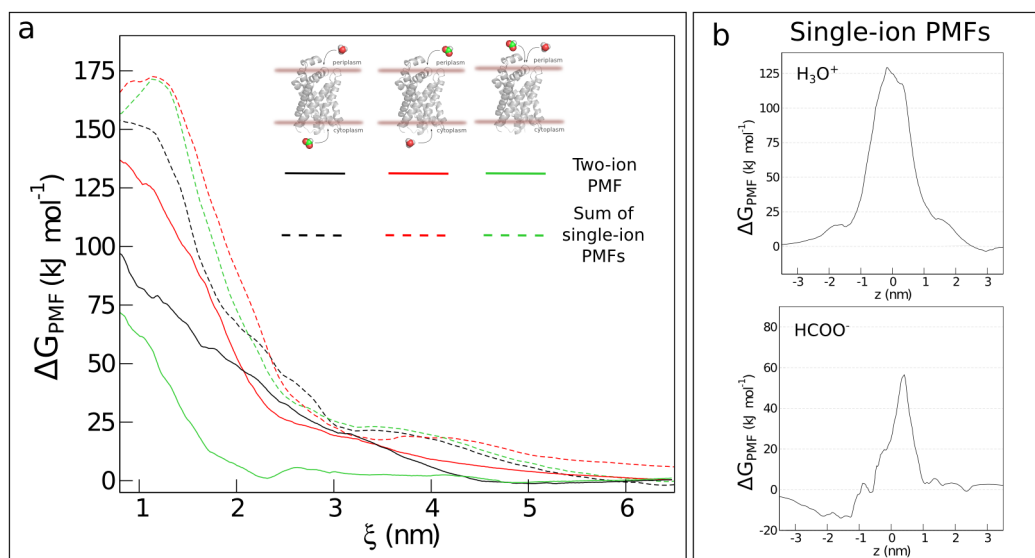


Figure 7.9: Comparison of the free energy of separate and simultaneous internalization of formate and hydronium ions into the central binding site of NirC. a) The curves are color coded by the direction of origin of each ion (illustrated in the top schemes). The bold lines represent the “sum of distances” PMFs, and the dashed lines represent the sum of the formate and hydronium single-ion PMFs shown in (b) (while taking into account the respective direction of entrance of each ion). The starting structures for all umbrella simulations originate from HIS+ equilibrium simulations.

“bulk” state is not achieved. Namely, the formate ion is usually attracted more to the protein than the hydronium ion (possibly due the positive electrostatic potential of the protein surfaces (fig. 2.10)), and binds in the cytoplasmic and periplasmic entrances even at the highest ξ values. Moreover, in the cases where both ions enter from the same side, the oppositely charged ions have a stabilizing effect on each other during the umbrella simulations. Additionally, the state at ξ_{\min} does not strictly represent the target state of anion bound to a positively-charged histidine, therefore, the free energy of transfer of the proton to the histidine needs to be additionally considered. Finally, since the central histidine residue has a certain flexibility on one hand, and is used as a reference for the distances of the moving ions on the other, the ξ_{\min} state might not be well defined, and might differ even among single-chains from one set of umbrella simulations.

Recap. This set of PMF calculations reveals that the entrance of a hydronium ion into the central binding site is indeed facilitated by simultaneous internalization with a permeation anion. As anions are more attracted to the pore than the hydronium cation, in practice, the anion would enter one of the vestibules first, from where it would assist the protonation of the central histidine. However, the exact sequence of events of how both ions reach the central binding site, and at which point the histidine protonation occurs, remains unclear.

CENTRAL HISTIDINE-ANION PROTON TRANSFER

BASED on multiple methods, the obtained results in this thesis suggest that anions are not able to perform a full permeation across the FNTs, regardless of the protonation state of the central histidine. As elaborated in previous sections, this finding is either due to a high permeation barrier, or due to a very strong binding to the central binding site. One possible exit strategy for an anion bound to this site, is a proton transfer from the positively-charged central histidine to the anion, enabling it an escape from the channel in a neutral form, as was previously suggested [41, 90]. To test the probability of proton transfer between the central histidine and the permeating anion, combined QM/MM calculations were performed, where the side chain of the central histidine and a formate or nitrite ion in the center of the pore were described on a QM level, while the rest of the system was treated with a classical force-field description. Initially, a suitable level of QM theory needed to be chosen, one that accurately describes the studied process, but is computationally feasible. Once a protocol was established, free QM/MM simulations of NirC and VcFocA with a bound anion in the central binding site were performed.

8.1 SELECTION OF QM PARAMETERS

QM calculations are typically computationally expensive, since they involve solving the electronic Schrödinger equation for the system of interest [150]. Therefore, a common practice is to find a reasonable compromise between the accuracy and the cost of the level of theory used for description of the system. In this respect, two major components need to be considered: the method used for approximating the electronic equation, and the set of basis functions used for building the molecular orbitals. In this work, the process of interest is the proton transfer between a histidine side chain and a formate/nitrite anion, therefore, a minimal system containing 4-methylimidazolium and the anion of interest was set up (fig. 8.1a). Initially, a rough “reaction coordinate” for the proton transfer was used, defined as the position of the proton on the line connecting the N_δ atom from the imidazolium ring and one oxygen atom in the formate ion, in order to get a quick estimate of the efficiency of multiple method/basis set combinations, without performing any geometry optimizations (fig. 8.1c). Except for the Hartree-Fock calculations and the DFT/6-311⁺G* combination, the rest of the method-basis set combinations fall within few kJ mol⁻¹. Since the much faster DFT method produced very similar results to the higher level MP2 method, DFT (in combination with the B3LYP functional) was chosen to be used in all further calculations.

To test the performance of different basis sets with this method, as well as its capability to accurately describe the process of interest, a relaxed surface scan (RSS) of the N_δ - H_δ bond was performed in vacuum (fig. 8.1d) and in implicit solvent (fig. 8.1e). RSS enables for scanning a reaction coordinate by fixing a certain degree of freedom of interest (in this case the N_δ - H_δ distance), and optimizing all other degrees of freedom. As evident from fig. (fig. 8.1e), the minimal SVP and SV(P) basis sets deviate the most from the rest of the calculations, suggesting the more expensive, correlation consistent polarized basis sets, including diffuse functions, need to be used. From those, there is no significant difference when double- or triple-zeta functions are used, therefore, the cheaper aug-cc-pVDZ basis set was selected for the production simulations of the protein system. On a closer inspection of the shape of the calculated potential energy curves, only one minimum is observed in vacuum (fig. 8.1d), corresponding to neutral imidazole ring and neutral formic acid, since a state with two ions is unstable in vacuum. In water, two minima are observed, corresponding to the proton residing on the imidazole ring, or on any of the formate's oxygen atoms (fig. 8.1e). The observed potential energy difference between these two states is only few kJ mol^{-1} , with an energy barrier of the same magnitude. This difference might seem low, if the pK_a value of 7.5 for 4-methylimidazolium and 3.8 for formate are considered, which suggest that the formate ion would not be easily protonated by the imidazolium ring. However, it has to be noted that the resulting energies from the relaxed surface scan do not represent free energies, and are therefore not directly comparable to the pK_a . Moreover, the potential energy surface for proton transfer depends strongly on the donor-acceptor distance, and a surface scan of the proton position offers only qualitative trends for the energy barriers [216]. The only quantity which could truly be compared to experiment is the proton affinity, defined as the enthalpic contribution of the gas-phase protonation reaction of a species [217]. The performance of different QM methods in reproducing proton affinities has been extensively studied [218], where B3LYP was found to slightly overestimate the proton affinity of 4-methylimidazolium, and slightly underestimate the one of formic acid, however, with deviations comparable to experimental errors.

The QM/MM setup was tested on a minimal system consisted of a capped histidine residue and a formate ion either in vacuum, or solvated in explicit water. In this system, only the histidine side chain and the anion were described with QM level of theory. In vacuum, the QM RSS calculations were imitated using the linear transit method (for details see section 4.10), and the resulting potential energy curve matches closely the pure-QM results (fig. 8.1d). In water, the free energy of proton transfer was calculated using the TI method and the abovementioned reaction coordinate. In such a setup, the TI calculations effectively result in the PMF for the proton transfer reaction, offering information for the energy barrier between the two end states, in addition to the free energy difference. As seen in fig. 8.1f, the minima and shape of the curve match reasonably well with the RSS curves, but should not be expected to perfectly reproduce the results from the pure QM calculations, due to differences in the studied system, the description of the water model, and the calculated quantity. Notably, the method used for description of long-range electrostatics seems to influence the PMF the most, therefore, this effect was investigated also in the protein system.

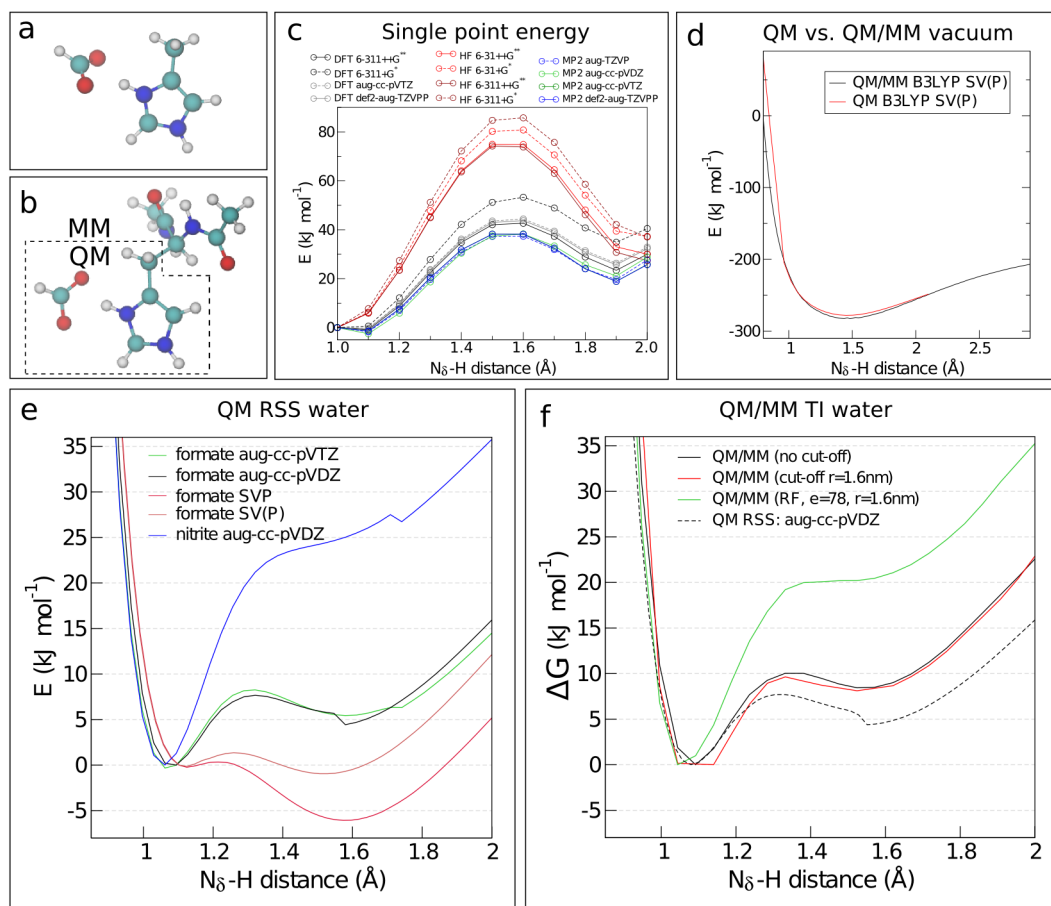


Figure 8.1: Optimization of the QM/MM parameters. a) Minimal system for optimization of the QM parameters. b) Minimal system for testing the QM/MM setup. c) Single point energies of the QM system shown in (a). Each point corresponds to a geometry with different position of the proton on the line connecting N_δ from the imidazolium ring and one oxygen atom from the formate ion. d) QM relaxed surface scan (red) and QM/MM potential energy calculated by linear transit (black) in vacuum. Reaction coordinate: distance between N_δ from the imidazolium ring and its bound proton. e, f) QM relaxed surface scan (e) and QM/MM free energy calculated by thermodynamic integration (f) in water. Reaction coordinate as in (d).

Overall, the described QM/MM setup in combination with B3LYP/aug-cc-pVDZ level of theory, should enable semi-quantitative description of the proton transfer reaction in the protein.

8.2 PROTON TRANSFER IN NIRC AND FOCA

The feasibility of proton transfer between the central histidine and the permeating anion was investigated in three protein-anion systems: NirC with bound formate or nitrite, and VcFocA with bound formate in the central binding site. As in the test simulations, the histidine side chain and the bound anion were described quantum mechanically, while the rest of the system was described classically. Starting structures were taken from the permeation umbrella simulations (corresponding to the energy minima identified in the PMF curves), and used for free QM/MM simulations.

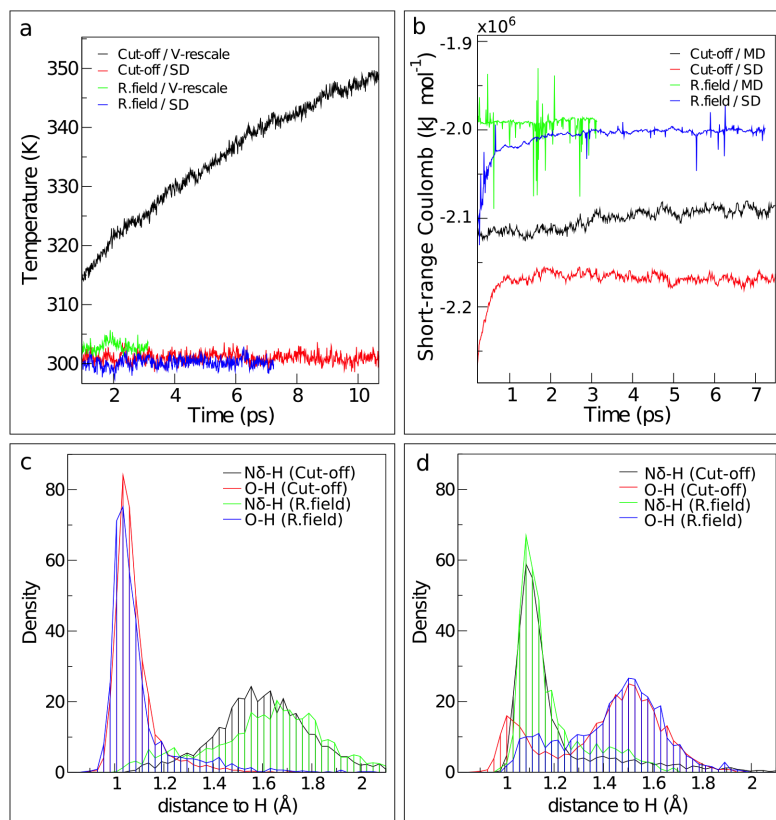


Figure 8.2: Effect of the long-range electrostatics method in the protein QM/MM simulations. a,b) Evolution of the system temperature (a) and short-range Coulomb interactions (b) in simulations employing different electrostatic schemes (cut-off/reaction field), thermostats (V-rescale/stochastic dynamics) or integrator (molecular/stochastic dynamics). c,d) Distribution of distances between the proton and the N_δ atom from the central histidine (N_δ-H_δ), and the proton and formate oxygen atom (O-H), considering different electrostatic schemes (cut-off/reaction field) and basis sets: (c) SVP, (d) aug-cc-pVDZ.

As mentioned in the previous section, the method used for treatment of long-range electrostatic interactions could potentially have an effect on the process of interest in the protein/anion system. To investigate the magnitude of this effect, two long-range electrostatic schemes compatible with the QM/MM setup were employed: cut-off scheme and the reaction field method. Although simple cut-off of the electrostatic interactions is usually not recommended due to the long-range nature of the Coulomb potential, it is still commonly used in QM/MM calculations. This is due to the fact that 1) most typical QM/MM implementations do not yet offer any more advanced electrostatics schemes (e.g. PME), 2) large enough cut-off distances might offer a good enough approximation in many cases. Due to truncation effects, the cut-off scheme leads to gradual heating of the system (fig. 8.2a), but this issue can be circumvented by using a stochastic dynamics thermostat. The reaction-field scheme, in which beyond certain cutoff the system is treated as a dielectric medium with a certain permittivity, does not lead to such magnitudes of the heating issue. However, it was found to result overall in less stable simulations, exhibiting high, non-physical spikes in the short-range Coulomb interactions (fig. 8.2b). This is probably due to Gromacs / Orca implementation issues, as the QM/MM interface was originally implemented for use with the cut-off scheme. Finally, the choice of the electrostatics scheme does not seem to have a significant effect on the observed protein transfer process in the protein, which is demonstrated in figures 8.2c and d. Here, the distribution of the distances between 1) the proton and the N_δ atom from the central histidine, and 2) the proton and the formate oxygen atom closest to the central histidine are shown, considering different basis sets and electrostatic schemes. It is clear that the effect of the electrostatics scheme is much smaller than that of the selected basis set for the QM subsystem. Keeping that in mind, and taking into account the fact that the Orca/Gromacs QM/MM interface used in the current work was originally implemented to function with the cut-off scheme, for the production simulations this electrostatics scheme was employed in combination with a stochastic dynamics thermostat.

For each of the three systems (NirC/formate, NirC/nitrite, and VcFocA/formate), several simulation replicas were performed, using different starting structures, with and without energy minimization prior to the simulation. The distances of the proton to the donor atom (the N_δ atom from the central histidine) and the acceptor atoms (any of the oxygen atoms in formate and nitrite) are shown as function of time (fig. 8.3 and 8.4). In the NirC system, frequent jumps of the proton between the histidine and both, the formate and nitrite ions can be observed, occurring on a picosecond time scale. Overall, the proton resides more often and longer on the histidine side chain, however, in certain cases it moves “permanently” to the permeating ion (fig. 8.3a, 8.3c, 8.3g) which can then move away from the histidine (fig. 8.3a).

In VcFocA (fig. 8.4), it is evident that the central histidine and formate ion communicate very closely, and there are events when the proton is actually closer to the formate, but these are much less frequent and much shorter than in NirC.

Recap. Taking all into account, it is evident that the proton transfer between the central histidine and the bound anion can occur in the protein microenvironment, and in NirC, it does so on a picosecond time-scale. In this way, the permeating anion is enabled to escape from the central binding site and is “free” to complete the permeation in a

neutral form. In FocA, this process maybe occurs on longer time-scales, and/or requires certain orientation of the Ω -loop.

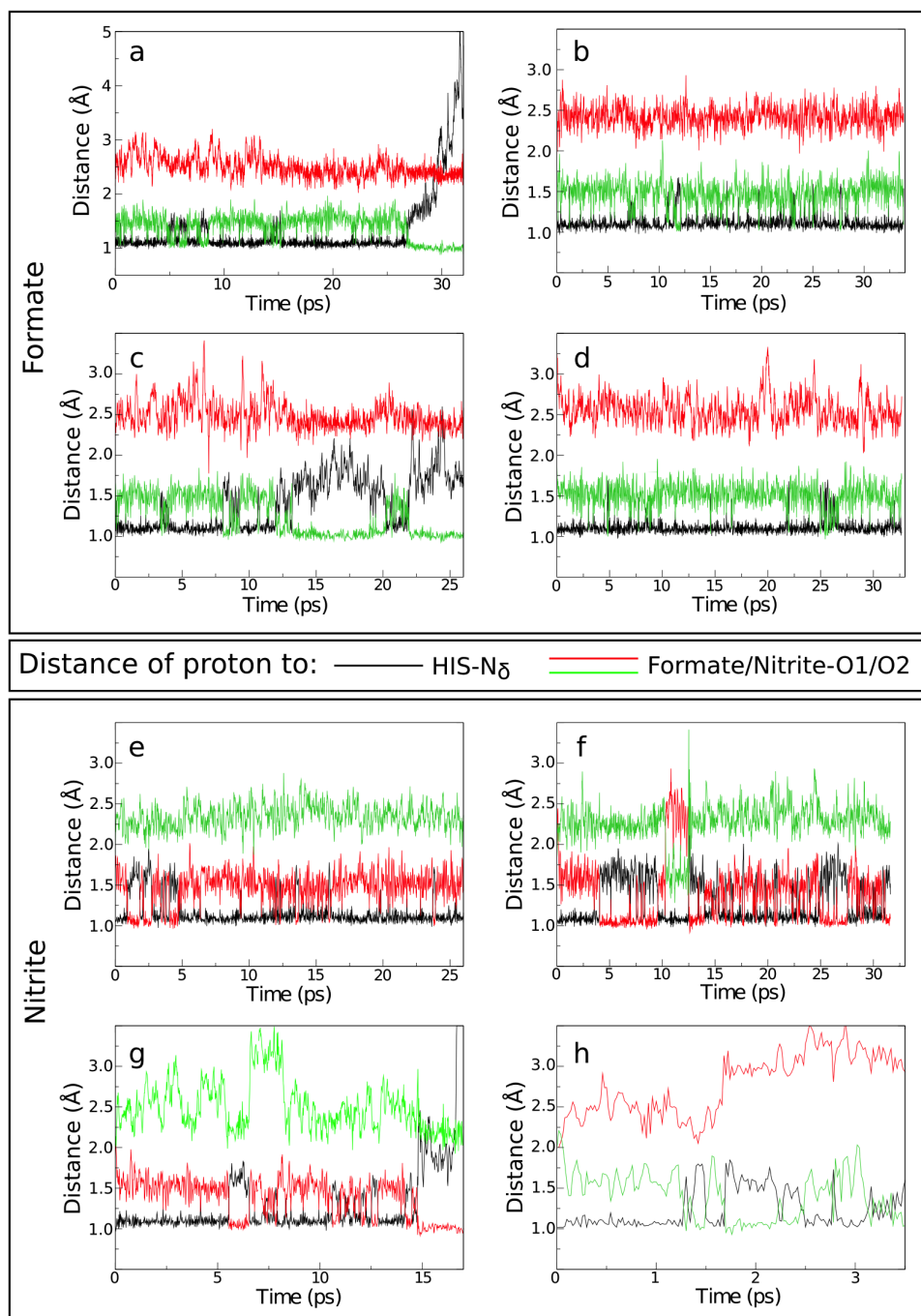


Figure 8.3: Proton jumps in NirC. (a-h) Distances of the proton to the donor atom (N_δ atom from the central histidine, black) and the acceptor atoms (any of the oxygen atoms in formate (a-d) or nitrite (e-h), red, green) from different simulation replicas are shown. Left column: with prior energy minimization, right column: no prior energy minimization. Each row corresponds to simulations originating from a different starting structure.

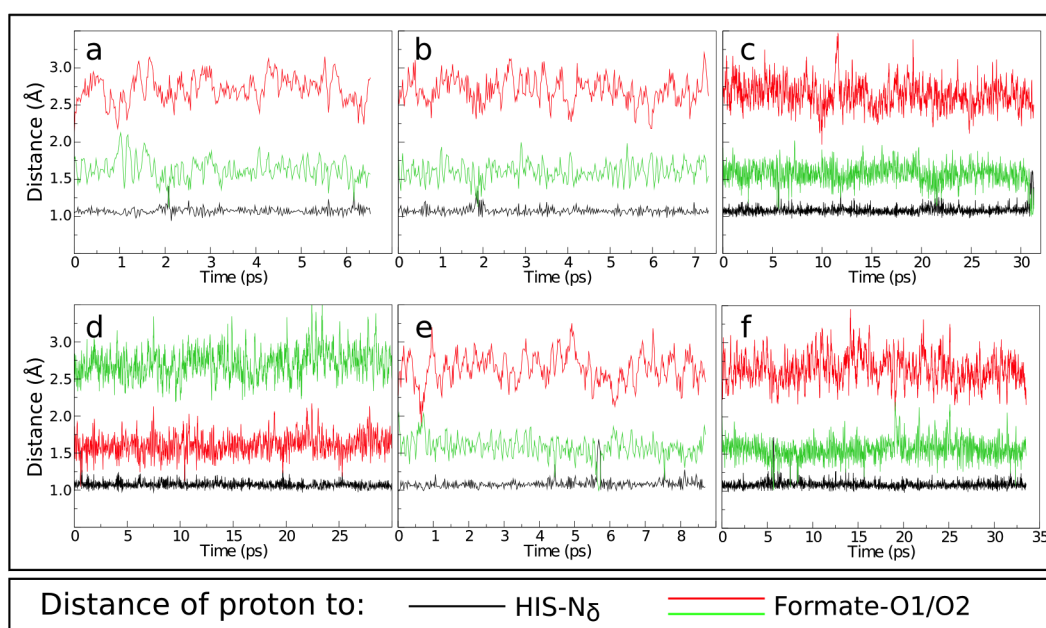


Figure 8.4: Proton jumps in FocA. (a-f) Distances of the proton to the donor atom (N δ atom from the central histidine, black) and the acceptor atoms (any of the formate oxygen atoms, red, green) from different simulation replicas are shown. (a,d) with prior energy minimization, (b,c,e,f) no prior energy minimization. Each row corresponds to simulations originating from a different starting structure.

Part IV

DISCUSSION AND CONCLUSIONS

DISCUSSION

The FNT family of membrane transport proteins was studied using multiple computational methods, in order to elucidate the molecular details of their permeation mechanism. Among other, this work aimed to reveal the character of permeation, as well as the nature of proton coupling. Furthermore, the question whether a general mechanism describes the permeation in all FNTs, or significant differences are in place in order to accommodate the versatile functions of this family, was investigated. All FNT subfamilies with a known structure (NirC, HSC, and FocA) were included in the analyses. To the best of my knowledge, there are no *in silico* experiments performed for NirC and HSC so far. The free energy barriers for permeation of different substrates across the FNTs were estimated by calculation of PMFs for full permeation events. The possibility for a knock-on permeation mechanism was eliminated by extensive computational electrophysiology simulations. In all calculations, the central histidine was given a special focus, in order to specifically pinpoint its role in permeation. The details of protonation of this residue were thoroughly studied. Finally, the proton transfer between the central histidine and the permeating anion was demonstrated in a set of QM/MM calculations. These efforts led to a comprehensive picture of the permeation across the FNT channels. Generally, similar trends are observed in all investigated FNTs. Certain variations could be recognized between NirC and HSC on one, and FocA on the other hand, consistent with their phylogenetic relationship and structural similarity.

Role of the central histidine. As previously mentioned, the central histidine residue had to be treated with special care, due to its confirmed, but so far not described role in permeation. This residue is highly conserved in the FNTs, and is located almost perfectly in the center of the permeating pore, in a small chamber between two constriction sites (see for example fig. 2.7). Its importance was already anticipated following the discovery of the FNT structure, where it was hypothesized that its side-chain orientation could affect the FocA channel activity [42], it could bind formate and thus be a part of a substrate-gated mechanism in FocA [85], or it could protonate the substrate in an elaborate proton-relay mechanism [78, 90] (fig. 2.11). Indeed, mutation of this residue to phenylalanine resulted with no measurable current in FocA in PLB electrophysiology experiments [73]. As hydrogen atoms are typically not determined in crystal structures, two major questions emerge: What is the protonation state of the central histidine in the protein, and more importantly, does it influence the permeation of substrates? This was investigated in details in this thesis, where both protonation states of the central histidine: imidazole (HIS0), and imidazolium (HIS+), were taken into account.

Permeation mechanism across the FNTs. Generally speaking, the clear channel-like structure of the FNTs points to a permeation that does not involve considerable conformational changes inherent to a typical transporter. The permeation rates across open channels (up to 10^7 - 10^8 s $^{-1}$) are orders of magnitude higher than typical rates across transporters (10^1 - 10^4 s $^{-1}$), which usually undergo “conformation cycles” including an occluded state [219, 220]. Indeed, the measured single-channel flux of NirC and FocA of $\sim 10^7$ s $^{-1}$ at near-physiological conditions places them in the range of moderately-fast to fast channels. Permeant substrates of FNTs are commonly weak acids, introducing an additional level of complexity to the permeation mechanism, as they could in principle adopt multiple protonation states. The presence of detectable current in electrophysiology experiments indicates that FNTs must be able of translocating net charge across the membrane (at least *in vitro*). Assuming an open channel, and disregarding for the moment any kinetic or thermodynamic details of the protonation of the central histidine, we can conceive a simplistic view of the possibilities for permeation across the FNTs, which would account for these experimental observations:

1. Substrate permeates the channel as an anion, while the central histidine is neutral;
2. Substrate permeates the channel as an anion, while the central histidine is positively charged;
3. Substrate changes its protonation state during permeation, possibly with involvement of the central histidine.

Anion vs. HIS0 pore. The first of these options represents the simplest imaginable pathway of an anion across an FNT, and corresponds to the classical description of a protein channel, which entails a continuous pathway through the membrane with a selectivity filter for certain ions [3]. The current work demonstrates that this pathway is not feasible in any of the studied FNTs. As clearly evident from the PMFs for permeation (fig. 5.13), the energy barrier for anions of $\Delta G^\ddagger > 70$ kJ mol $^{-1}$, is too high to allow any detectable flux across the channel. Moreover, this observation is valid in all studied FNT/anion combinations, including the putative “open” conformations in VcFocA. The observed energy barrier can be mainly attributed to the high dehydration penalty the anion experiences on permeation. The permeating pore with average radius smaller than 3 Å in a region approx. 20 Å long, is too narrow to allow the anion to take its hydration shell past the vestibules. Moreover, the central portion of the pore including the two constriction sites is only very rarely hydrated. This is reflected in the PMFs for water permeation (fig. 5.15) that exhibit approximately two-times higher barrier in comparison to aquaporins [211]. In protein channels, dehydration offers in a way the basis of ion selectivity, as highly-selective channels provide very specific substitute solvation for a certain type of ions by a special arrangement of their pore-lining residues [221]. However, HIS0 FNTs clearly do not offer sufficient compensating solvation for the permeating anions to overcome the dehydration penalty. Interestingly, the permeation barrier for the functionally relevant substrates, formate and nitrite, is consistently lower in comparison to chloride, despite their higher dehydration energies. This suggests that formate and nitrite can form favorable interactions with the pore, but not to the extent

that would permit efficient permeation across the channel. In summary, the described case does not represent the permeation mechanism in the studied FNTs, and most probably also neither in the entire FNT family, since the pore-lining residues, including the constrictions and the central histidine, are well conserved among the family.

Anion vs. HIS+ pore. The second option for permeation across the FNTs that was considered and is consistent with electrogenic transport, is similar to the one previously described, with the difference of the protonation state of the central histidine (HIS+). Disregarding any details of the process of histidine protonation, a HIS+ FNT pore offers a strong binding site to the permeating anion in all studied FNTs, especially for formate and nitrite ($\Delta G_{min} = -60 \text{ kJ mol}^{-1}$ to -15 kJ mol^{-1} , fig. 5.17). As previously mentioned, a good selective ion channel needs to recognize and bind the permeating substrate. However, in order to allow the high permeation rates necessary for function, this binding should be relatively loose and is typically in the range of 10-100 mM [222], which is not the case for the FNTs. In some channels, the Coulomb repulsion from a second (or more) ions in the pore, destabilizes the bound ion and pushes it outside in a so-called “knock-on” mechanism [36]. In the FNTs however, such a mechanism was shown unfeasible by the extensive CompEl simulations. Certain number of permeation events could be observed (tables 6.1, 6.2, and 6.3), but only at very high non-physiological electrochemical gradients, and never to the extent to match experimental data. Moreover, these permeations did not follow any particular pattern or mechanism, and occurred on very different time-scales. At near-physiological conditions, no permeation events could be observed, nonetheless, the permeating anions were found to be able to enter the channels and bind to the central binding site.

Permeation by substrate protonation. Taken together, the data presented in this thesis suggest that anions are not capable of fully permeating across the FNTs, regardless of the protonation state of the central histidine, while neutral substrates experience significantly lower permeation barriers (fig. 5.15). This points towards a mechanism in which the permeating anion needs to be protonated at some point in the channel, in order to complete the permeation. A good candidate for the protonation role is the central histidine, and one of the reasons for this is its central location in the pore. Since FNTs are thought to be bidirectional [41, 43, 80], and are shown to be physiologically relevant for import (NirC), export (HSC), or import and export (FocA) of their respective substrates, a titratable residue with a central location, which is practically equally accessible from both sides of the membrane, is compatible with a permeation mechanism that involves substrate protonation. On one hand, protonating the central histidine is thermodynamically unfavorable ($\Delta G_{prot} \approx 70 \text{ kJ mol}^{-1}$ in NirC, table 7.2), and on the other, binding of an anion to a positively charged central histidine is a strong favorable interaction ($\Delta G_{min} \approx -60 \text{ kJ mol}^{-1}$ in NirC, fig. 5.17). Considering everything above, an anion would almost never enter the central binding site of a HIS0 pore, and in turn, the central histidine would almost never be in a HIS+ state in absence of a bound anion. However, combining both processes - histidine protonation and anion binding - results in an overall thermodynamically feasible state (fig. 9.2). Therefore, the following permeation mechanism for weak acids across the FNTs is hereby proposed:

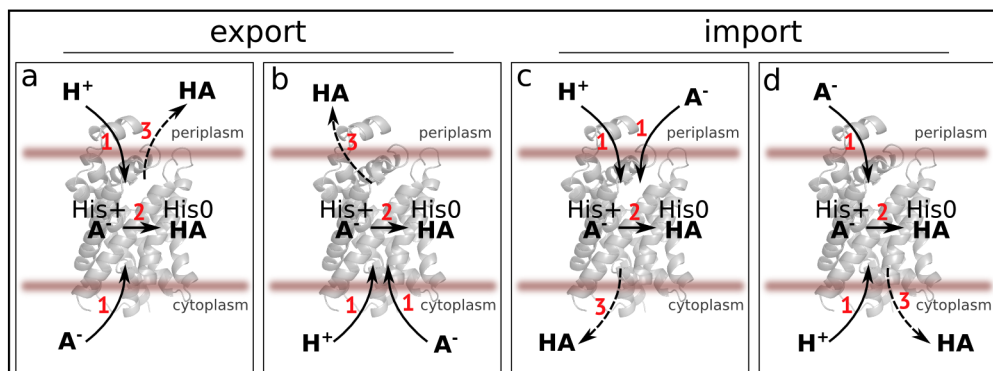


Figure 9.1: Scheme of the proposed mechanism of permeation across the FNTs. All possible combinations of proton origin and direction of transport are shown. His0: neutral central histidine, His+: positively-charged central histidine, HA: neutral substrate, A^- : anionic substrate. The order of the reaction is labeled with numbers. The exit of the substrate is marked with dashed line.

1. Protonation of the central histidine and translocation of an anion into the pore occur in a coupled manner,
2. The anion picks up a proton from the doubly-protonated central histidine, and
3. The now neutral substrate experiences much weaker binding, and is therefore capable to exit the pore.

This results with an electrogenic or proton symport transport, depending on the compartment of origin of the proton and anion. If they enter the pore from opposite sides, net charge is transported across the membrane, while in the case they both enter from the same side, the anion will be co-transported with a proton. In principle, all options schematically shown in fig. 9.1 are possible in experiment, depending on the anion and proton electrochemical gradients, however they might not all be physiologically relevant. As previously proposed, a mechanism involving substrate protonation would account for the fact chloride ions bind, but do not efficiently permeate the FNTs [41]. The details of the proposed permeation mechanism were studied separately in all FNT subfamilies with known structure, and will be presented per FNT protein in the following sections.

9.1 PERMEATION MECHANISM IN NIRC

In this work, the nitrite channel NirC was studied in most detail, and used as the model FNT in an attempt to describe an unifying picture of the permeation across this protein family. The overall structure of NirC was found to be stable in simulation for several hundreds of nanoseconds. As anticipated from the crystal structure, the N-termini, together with the region corresponding to the Ω -loop, stayed well-structured without large-scale conformational changes. Due to the vicinity of this loop to the cytoplasmic constriction, the NirC crystal structure was suggested to represent a closed channel [40], based on the fact that FocA adopts multiple orientations of the Ω -loop in the crystal

structure [85]. However, Lü et al. suggested that side-chain re-orientations allowed by the protein’s inherent flexibility could be sufficient for passage of cargo, without the need for a lasting “open state” per se [78]. The current work demonstrates that other factors, such as the protonation state of the central histidine and the permeating substrate, dictate the aspects of permeation across NirC. Moreover, the PMFs for permeation across NirC clearly establish the periplasmic constriction as posing the higher barrier to both, ions and neutral substrates, when compared to the cytoplasmic constriction (fig. 5.13 and 5.15). This suggests that even an unlikely major conformational shift in the region corresponding to the Ω -loop occurs, which could affect the size of the cytoplasmic constriction, it would not effectively create an “open” state and facilitate permeation.

As already established for all FNTs, anions are not able to completely permeate NirC. The reason for this is the high permeation barrier ($\Delta G^\ddagger \approx 70 \text{ kJ mol}^{-1}$) when the central histidine is neutral, or the strong binding to the central binding site (ΔG_{min} of -60 kJ mol^{-1} to -50 kJ mol^{-1}) when the central histidine is positively charged. However, the performed QM/MM simulations demonstrate that in NirC, proton transfers between the central histidine and the bound formate or nitrite ions occur on a picosecond time-scale (fig. 8.3), strongly suggesting a permeation mechanism involving substrate protonation.

Protonation of the central histidine. To be able to protonate the permeating anion, the central histidine must itself first be doubly-protonated. In this work, ΔG_{prot} of the central histidine was calculated using both, fully microscopic and a macroscopic method. The former has the advantage of accounting explicitly for the protein flexibility and the interactions with the solvent or the charged residues and lipids, however, it is limited by sampling issues. A macroscopic continuum electrostatics method on the other hand, implicitly captures the effects of the protein dielectric environment, which in cases of slow protein relaxation may offer overall a better description of the system [215]. This turned out to be the case for the FNTs, since the pK_a of the central histidine was found to be strongly dependent on the protein side-chain orientations, and the FNTs were found to have longer dielectric relaxation times than the TI or equilibration simulations. In any event, both, TI and GMCT calculations (table 7.2 and fig. 7.1) suggest that protonating this residue is not an easy task in any of the FNTs, and particularly not in NirC and HSC. This is not surprising given its location in the middle of a narrow hydrophobic pore that is only very rarely hydrated, leading to a highly unfavorable desolvation energy (table 7.2) on introduction of positive charge. That the pore is “cation-unfriendly” is also reflected in the PMFs for permeation of cations which are characterized by significantly higher permeation barriers in comparison for anions ($>100 \text{ kJ mol}^{-1}$, fig. 5.14). With the GMCT method, ΔG_{prot} of the central histidine in NirC was estimated to be $\sim 70 \text{ kJ mol}^{-1}$, with a possible decrease of 10-15 kJ mol^{-1} in the presence of a proton motive force in the physiological range. This suggests that in practice, the central histidine is not protonated in the absence of other ions in the pore.

Permeation across NirC. As previously mentioned, coupling the histidine protonation to the binding of an anion in the central binding site could provide a compensation for the high energy cost of protonation, and means to internalize the anion into the pore. This coupled process is the first step in the proposed mechanism for permeation of

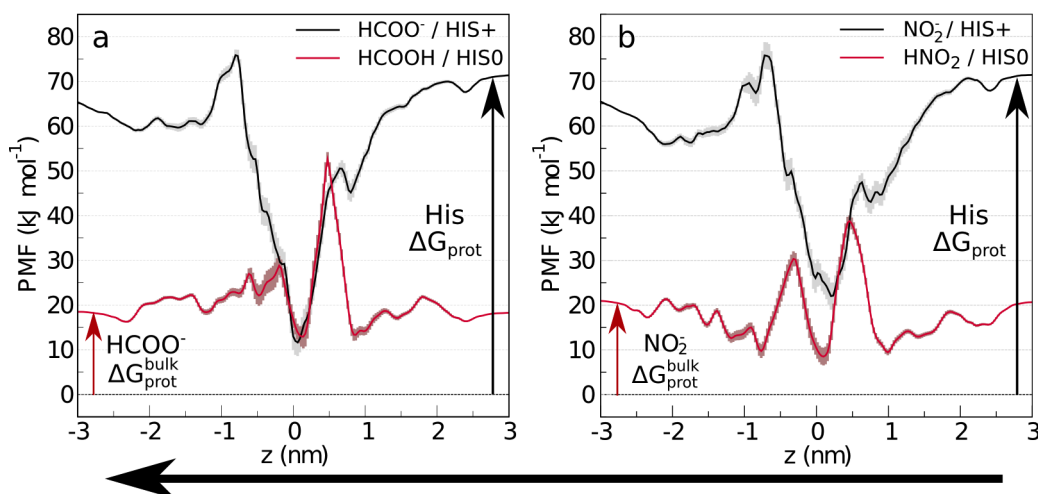


Figure 9.2: Combined PMFs for permeation across NirC. The PMFs for permeation of formate/formic acid (a) or nitrite/nitrous acid (b) across HIS+ or HIS0 pore are shown corrected by the respective free energy of protonation of the central histidine in the protein or the anion in bulk (PMF shift is indicated by arrows). The arrow on the bottom indicates the physiological direction of permeation (left: cytoplasm; right: periplasm).

formate and nitrite across NirC, as schematically shown in fig. 9.1 and 9.2. In the latter figure, the PMFs for permeation of formate/nitrite across a HIS+ pore, and formic/nitrous acid across a HIS0 pore, are shown corrected for the energy cost of protonating the central histidine, or protonating the anion in bulk, respectively (considering $\text{pH} = 7$ and absence of pH gradient or membrane voltage). Once the anion is bound to the central histidine ($z \approx 0$ at the black curve), the QM/MM calculations predict it will be very quickly protonated by the imidazolium ring, which in fig. 9.2 is represented by moving from the black to the red PMF curve at $z \approx 0$. That the states corresponding to bound anion, or bound neutral substrate (in HIS+ or HIS0 pore, respectively) are energetically close, is well reflected in the fact that the minima in both curves are almost identical for formate, and within $\sim 10 \text{ kJ mol}^{-1}$ for nitrite (which is below the uncertainty of the anion PMFs). Typical binding poses of the anion and the neutral acid from the umbrella and CompEl simulations are shown in fig. 5.16 and 6.4. Once the substrate is protonated, it experiences weaker binding to the central histidine, and a lower permeation barrier. In NirC, the cytoplasmic constriction presents a lower barrier than the periplasmic constriction, therefore, the net movement of the neutral substrate is in the direction towards the cytoplasm, which is in accordance with the physiological role of NirC (nitrite import). However, bidirectionality is in principle allowed.

Coupling of the histidine protonation and anion binding. One important aspect in such a permeation mechanism, which is not yet entirely clear, is the pathway of simultaneous proton and anion internalization into the central binding site of the pore, in order to reach the state of a HIS+ pore with a bound anion. In physiological conditions, both the proton and the anion would likely tend to enter the pore from the periplasmic side, following their respective electrochemical gradients (assuming high

NirBD activity in the cytoplasm, that would quickly process the imported nitrite ions). In PLB electrophysiological experiments, the proton and anion gradients are always opposite in case of symmetrical proton and substrate concentrations in both compartments, and mostly opposite in case of asymmetric substrate concentrations. The permeation PMFs show that the periplasmic vestibule is more accessible to formate and nitrite ions, regardless of the protonation state of the central histidine, although in principle they could also enter the cytoplasmic vestibule (fig. 5.13 and 5.17). In the HIS0 state, formate and nitrite enter deep in the periplasmic vestibule almost barrier-free, to reach the “periplasmic binding site” where they loosely interact with Lys145 (also exhibited as a small well at $z = 1$ in the HIS+/anion PMF curve). From the cytoplasmic side, the anions seem to have stronger favorable interactions with the pore, but located more peripherally. These observations might indicate structural optimizations in NirC for a preferable transport in the import direction.

The PMFs for simultaneous internalization of a hydronium and formate ion into the central binding site in NirC provide some insight into this coupled process (fig. 7.8). As expected, at large values of the reaction coordinate ξ (equal to the sum of the distances of each ion to the central histidine), the system equilibrates to a state where the formate ion is much closer to the protein than the hydronium ion, regardless of the direction of its approach. This suggests that in a first step, an anion enters the periplasmic or cytoplasmic vestibule (direction depending on their gradient), and its presence then assists the protonation of the central histidine. When compared to the sum of the single-ion PMFs, it is evident that the presence of an anion in the pore does indeed aid the entrance of a hydronium ion. Namely, the energetic cost for internalization of both ions is reduced by ~ 30 - 80 kJ mol^{-1} , where the smallest effect is observed in the case when the hydronium ion enters from the cytoplasmic side, and the largest when both, hydronium and formate ions enter from the periplasmic side (fig. 7.9). Notably, the largest effect coincides with the physiological direction of movement of the ions, while the smallest effect with the direction of proton movement opposite of physiological pH gradients.

Unfortunately, the kinetic details of this process could not be precisely quantified from PMF calculations for simultaneous internalization of anion and hydronium ion, mainly due to 1) poor convergence, and 2) oversimplification of the reaction coordinate. Regarding the latter, the first large approximation is the definition of the final state ξ_{\min} as a HIS0 pore with hydronium and formate ions in the central binding site, which does not strictly represent the target state of an anion bound to a positively charged histidine. Further, apart from the distance of the ions to the central histidine, this process is probably influenced by other, orthogonal degrees of freedom, which are therefore not visible in the resulting PMFs. Such degrees of freedom could include side-chain orientations of the pore-lining residues, or existence of a “proton ladder” which might assist the proton internalization. In summary, although these PMFs do not reveal the exact mechanism of the coupled process of histidine protonation and anion binding, they demonstrate that an anion and hydronium ion on average assist each others entrance into the central binding site.

Proton fate on substrate exit. Finally, the question regarding the fate of the proton once the permeating substrate crosses the constriction remains. In principle,

the described permeation mechanism by substrate protonation accounts for different combinations of electrogenic or electroneutral transport. Namely, when the substrate completes the permeation in a neutral form, the overall flux would be electrogenic if the anion and proton originate from different sides of the membrane, and conversely, electroneutral when the anion and proton enter the pore from the same side (scheme 9.1). In the PLB electrophysiology experiments, the clear Ohmic and Nernstian behavior of the current suggests electrogenic transport [78]. Since at symmetrical pH and substrate concentrations the proton and substrate gradients are opposite, this would be consistent with situation as depicted in fig. 9.1a or 9.1b. In physiological conditions, where both, nitrite and proton would probably enter from the periplasmic side, such a permeation would result in an anion/ H^+ symport. This is thermodynamically consistent, since the transfer of nitrous acid (or nitrite plus proton) from the periplasm to the cytoplasm, is favored more than the transfer of nitrite ion (which is also thermodynamically allowed, when sufficient concentration gradient is present, see section 5.2). However, such a mechanism would result with dissipation of the pH gradient, since it would co-transport a proton down its gradient for every transported nitrite ion. The complex nitrogen metabolism in bacteria with (only seemingly) redundant pathways makes it challenging to estimate whether such energy loss is justified. Some clues can be found by examining the origin and fate of the transported substrate across NirC. As previously mentioned, NirC is genetically linked to the cytoplasmic nitrite reductase NirBD, whose primary role is nitrogen assimilation by nitrite reduction to ammonia [80]. While nitrogen assimilation is usually an energy-consuming process in nature, certain bacteria, such as *E.coli* and *Salmonella*, developed a strategy to conserve energy in anaerobic conditions by coupling the participating nitrate and nitrite reductases to the respiratory chain [76]. NirBD seems to be an exception to this rule, since it uses NADPH to reduce nitrite. However, it is believed that this process indirectly results in ATP production, via promoting the acetate-producing branch of the mixed-acid fermentation in order to maintain redox balance [76, 223]. Therefore, this metabolic network could possibly compensate the energy loss by nitrite/proton co-transport. The origin of the permeating nitrite across NirC is even more curious. It is now believed that the majority ($\sim 70\%$) of the nitrite produced by the energy-conserving cytoplasmic nitrate reductases (Nar), is first expelled from the cell by NarK or NarU, (possibly in antiport with nitrate), after which it is re-imported by NirC [46, 77, 79, 80] (fig. 2.2). For each reduced nitrate by Nar, two protons are pumped across the membrane. If nitrate/nitrite antiport in ratio of 1:1 is assumed, the nitrite/proton symport by NirC would use up approx. 35 % of the energy conserved by Nar. This is only a very rough estimate, since the transport across NarK and NarU is not fully understood (proton involvement and stoichiometry), and moreover, it is not clear why nitrite is first expelled, and then re-imported into the cell [76].

An alternative possibility is that the substrate is deprotonated after passing the final constriction, via the proposed proton relay [41] or some other mechanism, and exits the channel in an anionic form, thereby leaving the extra proton in the pore, able to re-protonate the central histidine. In this case, the transport would be always electrogenic. The primary protonation of the central histidine could be then considered as channel gating, since it converts the channel into its permeation-competent form. As previously

mentioned, in physiological conditions the transfer of nitrous acid from the periplasm to the cytoplasm is thermodynamically more favorable than the transfer of nitrite, however, transfer of nitrite is indeed allowed when there is sufficient concentration gradient present, as in the case of active cytoplasmic NirBD. The slow gating (in the range of seconds) observed in electrophysiology experiments was proposed to represent central histidine protonation with a proton from the bulk, after an eventual loss of the extra proton in one of the permeation events [41]. Proton re-cycling would also account for the relatively high permeation rate observed in electrophysiology. Since the PMF calculations of the simultaneous hydronium and anion internalization did not result with a quantitative estimation of the kinetic barrier of this process, this suggestion cannot be confirmed or disproved at this point.

9.2 PERMEATION MECHANISM IN HSC

The hydrosulfide channel HSC is structurally more similar to NirC than to the FocA channels, which was also reflected in the data presented in this work. Namely, similarly to NirC, the N-termini and the region of the Ω -loop in HSC had a stable helical structure during the simulation without major conformational changes. The PMFs for permeation across HSC show the same trends as in NirC. The HIS0 pore presents high permeation barrier, and the HIS+ pore strong binding site for anions. Chloride ions bind weaker to the central binding site than formate and nitrite ions, and this difference is more pronounced than in NirC. This is also the reason for the few permeation events across HSC that were observed in the CompEl simulations, in the presence of very high electrochemical gradients. However, as in NirC, anions are probably not able to complete a permeation across the HIS+ pore at physiological gradients, especially if we keep in mind that typical FNT substrates (formate and nitrite) bind much stronger than chloride ions (fig. 5.17), and hydrosulfide ions have been shown to bind even stronger [49].

The crystal structure of HSC has been shown to be almost identical in a range of pH values, as well as in the presence of certain single-point mutations in the pore-lining residues [49]. This led the authors to suggest that significant conformational shifts in the N-terminus and TM2b (corresponding to the region of the Ω -loop) helices are necessary for opening the channel. However, as in NirC, the periplasmic constriction presents a bigger obstacle in comparison to the cytoplasmic (fig. 5.15 and 5.13), suggesting that such unlikely conformational change would not necessarily create an “open” state. Moreover, anions can enter barrier-free quite deep into the cytoplasmic vestibule, in fact, deeper than in NirC. This is not only reflected in the permeation PMFs, but was also observed in the CompEl simulations, which demonstrate that anions enter the central binding site faster and more frequently in the direction from the cytoplasmic side, in comparison to NirC. A better accessibility to the pore from this side is in accordance with the export function of HSC in the cell.

Taking all into account, the permeation mechanism by substrate protonation described for NirC can be conceived also for HSC (fig. 9.3). Since this mechanism does not assume directionality, it can accommodate the physiological role of export of hydro-

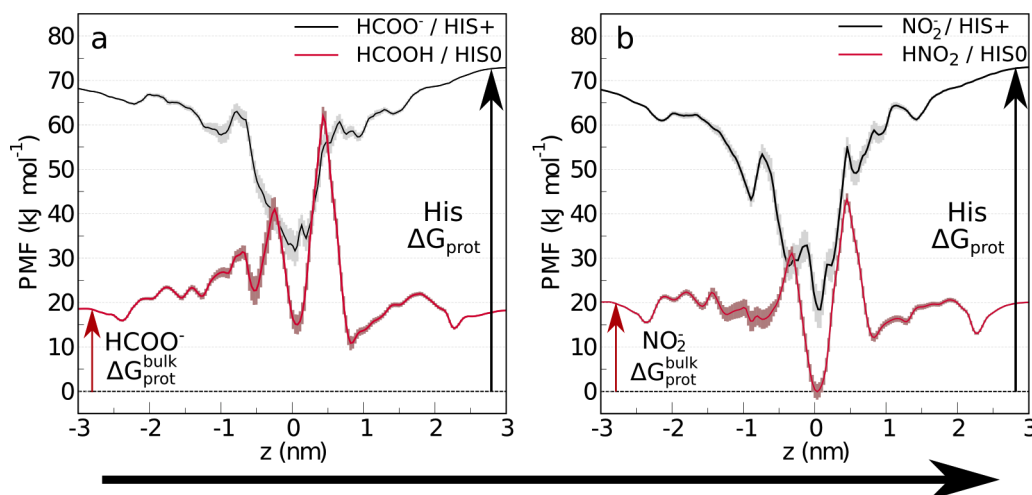


Figure 9.3: Combined PMFs for permeation across HSC. The PMFs for permeation of formate/formic acid (a) or nitrite/nitrous acid (b) across HIS+ or HIS0 pore are shown corrected by the respective free energy of protonation of the central histidine in the protein or the anion in bulk (PMF shift is indicated by arrows). The arrow on the bottom indicates the physiological direction of permeation (left: cytoplasm; right: periplasm).

sulfide ions. At physiological pH gradients, the outward flux is believed to be predominated by the ionic form of this substrate, despite its pK_a of 6.8 [49]. Therefore, an anion would first enter the cytoplasmic vestibule, where it would assist the protonation of the central histidine. Once it is bound in the central binding site, it is itself protonated and can leave the pore in a neutral form (or as an anion in the case of proton recycling in the pore). In the event the proton enters from the periplasmic side, this process would be electrogenic, and in the opposite event, it would result with anion/proton symport. As demonstrated in fig. 9.3 (equivalent to fig. 9.2 for NirC), the free energy of binding of formate or nitrite ion in the central binding site, does not entirely account for the energetic cost for protonation of the central histidine ($\sim 70 \text{ kJ mol}^{-1}$). However, this might not be an issue for the functionally relevant hydrosulfide ion, which was shown in thermostability experiments to bind stronger than nitrite and formate [49], thereby possibly allowing for an appropriate compensation of the energy cost of central histidine protonation. Unfortunately, this could not be investigated in the current work, due to lack of appropriate parameters for the hydrosulfide ion.

Notably, in contrast to the NirC protein, the physiological role of HSC is export of anions, meaning that once the substrate is protonated by the central histidine, it needs to cross the periplasmic constriction to complete the permeation. Even though this constriction poses lower barrier to the neutral substrate than to an anion, it is still higher than the barrier posed by the cytoplasmic constriction by approx. 25% for hydrogen sulfide, suggesting that the neutral substrate would move towards the periplasm only once every ~ 100 times in equilibrium conditions. This suggests a very low conductance for HSC. In a concentrative uptake assay, HSC was shown to be less active in comparison to FocA [49]. More importantly, the notion that hydrogen sulfide

will move frequently towards the cytoplasm, suggests that the proton involved in the permeation would likely originate from the cytoplasm. Otherwise, this mechanism would lead to unnecessary dissipation of the pH gradient without efficient substrate transport across the membrane. Proton expulsion together with HS^- ion transport might be beneficial for the cell, as it prevents proton accumulation in the cytoplasm, which is though to possibly occur due to passive diffusion of H_2S via the membrane [49].

In summary, the described permeation mechanism depicted in scheme 9.1 is a viable scenario also for HSC. The friendlier environment for anions in the cytoplasmic vestibule in comparison to NirC probably reflects an adjustment to the functionally relevant direction of anion transport across HSC, while the difference between the barriers posed by the cytoplasmic and periplasmic constriction to the neutral substrate might account for the lower conductance of this channel in comparison to other FNTs.

9.3 PERMEATION MECHANISM IN FOCA

As previously mentioned, the FocA channels follow the same general trends as observed for the other FNTs, including certain specific properties that separate them from the NirC/HSC group, arising from structural and functional adaptations. Most notably, the flexibility of the N-termini and the region of the Ω -loop add a layer of complexity to the permeation mechanism in FocA. Already in the free MD simulations it became evident that these regions show much higher orientational variability in comparison to the other investigated FNTs (fig. 5.4). Curiously, the *EcFocA* structure, in which the N-termini were truncated prior to crystallization, exhibited well-structured Ω -loop, which is more stable than in the other two FocA structures. This suggests that the N-termini, which are shown to form contacts with the peripheral portion of the Ω -loop (fig. 5.5), might be necessary for its flexibility. Moreover, a flexible Ω -loop is probably necessary for permeation, which is evident from the permeation PMFs across *EcFocA* and *VcFocA* (fig. 5.12). For both, anions and neutral substrates, the permeation barrier in the region of the cytoplasmic constriction in *EcFocA* is significantly higher than the one in the corresponding region in *VcFocA*, as well as than the periplasmic barrier. Therefore, the Ω -loop in the *EcFocA* structure might represent either a “closed” conformation of the protein, or a crystallographic artifact arising from the truncation of the protein.

Similarly to all FNTs, anions experience a very high barrier on permeation across an HIS0 pore, and relatively strong binding on permeation across a HIS+ pore (fig. 5.12). The PMFs qualitatively agree with the ones previously calculated for *VcFocA* using the adaptive biasing force method [224]. Precise quantitative comparison is not possible due to different definitions of the PMF, whereby the PMFs in this thesis correspond to channel density of one channel per membrane area occupied by a single monomer (for details see section 4.4). Notably, much higher sampling was achieved in the current work, by using all monomers (as opposed to only one in [224]), longer equilibration and umbrella simulation times, and tighter sampling along the reaction coordinate. In comparison to NirC and HSC, a longer portion of the HIS0 pore seems to be unfavorable for anion permeation, suggesting that entering of anions in the cytoplasmic or periplasmic vestibule in equilibrium conditions might occur less frequently than

in the other investigated FNTs. In a HIS+ pore, the binding of anion to the central binding site is weaker than in NirC and HSC, however, sufficiently strong to prevent complete permeation of the anion. This is evident both from the permeation PMFs, as well as the CompEl simulations, where no permeation events were observed at lower, near-physiological gradients (table 6.2). Typical binding poses of formate in the HIS+ pore and formic acid in the HIS0 pore are shown in fig. 5.16.

The free energy of protonation of the central histidine in FocA is also generally lower than in NirC and HSC. Here however, a precise value is even more challenging to achieve, since the orientation of the Ω -loop apparently affects the pK_a of this residue, additional to the dielectric relaxation of the protein (table 7.2 and fig. 7.1, 7.2). In any event, these calculations predict a high free energy cost to protonate the central histidine, consistent with the rest of the FNTs. The pK_a of this residue has been previously estimated to have value of 8.26 (for the *StFocA* structure) [225], and 5 (for the *VcFocA* structure) [224], using two other continuum electrostatics methods [226, 227], respectively. The finding in the former study is particularly curious, since it suggests the central histidine in the protein is protonated easier than a bulk histidine residue. However, it is difficult to comment the data from these studies, since 1) results for a single monomer from a single structure are reported, and 2) the used parameters, such as dielectric constants, are not stated. Moreover, it is not stated whether the membrane environment was taken into account. In this thesis, the free energy of protonation of the central histidine was extensively studied in all FNTs, taking into account multiple structures and all monomers, applying multiple methods, and considering the effect of the membrane environment, pH gradient, and membrane voltage. As previously mentioned, FocA presented a bigger challenge in this sense, due to the increased flexibility of the pore-lining residues and the Ω -loop. However, the general tendency reveals that the presence of a doubly-protonated central histidine is on average thermodynamically unfavorable. Given that FNTs are water permeable only when the central histidine is protonated, keeping a low probability of its protonation could be a mechanism to prevent proton leakage across the channel via the Grotthuss mechanism, in situations when an anionic substrate is not present.

Finally, the QM/MM simulations revealed that in FocA, proton transfers between the central histidine and the bound anion do not occur regularly on a picosecond time scale, in contrast to NirC. Similar results were observed by Lv et al. [224] in longer, albeit semi-empirical simulations (as opposed to the *ab initio* calculations presented in this thesis). This suggests that there is probably a larger barrier for the proton transfer process in FocA. Additionally, it could also be affected by the orientation of the Ω -loop, which unfortunately could not be sampled in the computationally-costly QM/MM simulations.

Overall, the same observations as for all FNTs apply: an anion would not enter the pore when the central histidine is neutral, and the central histidine would not be doubly protonated without the presence of an anion in the binding site. Moreover, the generally lower ΔG_{prot} of this residue is consistent with the weaker HIS+/anion binding, although it does not seem to be perfectly compensated by it (fig. 9.4). Therefore, the previously described mechanism for all FNTs is probably also viable for FocA, but requires additionally a certain re-orientation in the Ω -loop. Strictly thermodynamically, the export

of formate is favored over the export of formic acid at physiological conditions. In such conditions, the permeating anion would enter the pore from the cytoplasmic side, while the central histidine would be protonated from the periplasmic side, resulting in an electrogenic process that does not use the energy stored in the pH gradient. As for NirC, in a PLB electrophysiology setup, the proton and anion gradients are opposite at most membrane voltages, also resulting with an electrogenic transport. Such mechanism would also fit the FocA import function at conditions of low extracellular pH and accumulation of formate in the periplasmic space, in which case the permeating anion and the extra proton would originate from the same membrane side, resulting in formate/proton symport.

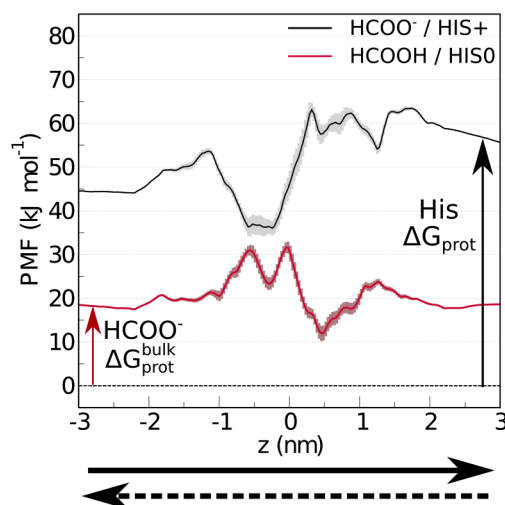


Figure 9.4: Combined PMFs for permeation across FocA. The PMFs for permeation of formate/formic acid across HIS+ or HIS0 pore are shown corrected by the respective free energy of protonation of the central histidine in the protein or the anion in bulk (PMF shift is indicated by arrows). The arrows on the bottom indicates the physiological direction of permeation (left: cytoplasm; right: periplasm). The lowest calculated ΔG_{prot} of the central histidine was used for correction of the HIS+ PMF.

This picture seems to be more complex *in vivo*, since the pyruvate formate-lyase (PflB) was found to bind to the cytoplasmic surface of FocA, and the N-termini are crucial in this interaction [66]. This can have two major influences on the permeation mechanism as conceived in this thesis: 1) If the binding promotes certain orientation of the N-termini, and thereby affect the orientation of the Ω -loop, it could in principle stabilize it in a permeation-competent orientation, and 2) The close vicinity of the formate-producing enzyme to FocA might significantly influence the substrate gradient by accumulating formate near the channel entrance.

Notably, the PMFs for water permeation revealed that the HIS0 pore does not permeate water efficiently, in contrast to the HIS+ pore, which shows water permeability similar to aquaporins (fig. 5.12). Therefore, the FNTs would be capable of water permeation only in the presence of anions which bind the central binding site, since this is a prerequisite for the central histidine protonation. This might account for the observa-

tion that stop-flow experiments, performed in the absence of such ions, show no water permeation across FocA [42], while water molecules were found in the pore in the FocA crystal structure, when also formate was present in the buffer [85].

9.4 FNTS AMONG OTHER MEMBRANE TRANSPORT PROTEINS

The binary view of membrane transport proteins in terms of “channels vs. transporters”, has since long been enriched to include a broad spectrum of proteins offering a variety of strategies to facilitate transport of substrates across biological membranes [2]. The investigation of such proteins, existing on the border between these two standard groups, has led to the realization that certain fundamental characteristics are probably shared among channels and transporters [220]. For example, the excitatory amino acid transporter (EEAT), whose primary function is glutamate uphill transport, coupled to Na^+ , H^+ , and K^+ gradients, was found to also harbor a chloride channel responsible for Cl^- ion leak [228–230]. Another interesting case represents the Na^+, K^+ -ATPase pump, which is effectively converted into a passive cation channel on binding palytoxin, whereby its internal and external gates are decoupled [231, 232]. Probably the most studied protein family that exhibits both, channel and transporter properties, is the ClC family. Here, the same protein structure can apparently accommodate two distinct functions, since some ClC proteins are passive chloride channels, while others are Cl^-/H^+ exchangers, in which the gradient of one ion is used for uphill transport of the second one [53, 233, 234]. Interestingly, a mutation of two gating residues in a Cl^-/H^+ exchanger led to a passive chloride transport, with rates faster than any known transporter, but slower than the wild-type ClC channels [235]. Therefore, this designed transport protein earned the name “chansporter” [220].

Given this continuum of membrane transport proteins, with no well-defined intrinsic borders, where do the FNTs, characterized by permeation involving substrate protonation by the central histidine residue, fit? The FNTs are certainly not consistent with the classical definition of transporters, since 1) they exhibit permeation rates comparable to passive channels, 2) they do not undergo major conformation shifts involving an occluded state, in order to avoid opening towards both membrane compartments simultaneously, and 3) many transport their substrates down their electrochemical gradients. On the other hand, FNTs do not quite fit typical ion channels either, in the sense of ion diffusion through a continuous pore with a certain selectivity filter. Since they have strikingly similar structure to aquaporins, it is natural to ask whether they share any functional features. The protein fold shared by aquaporins and FNTs is apparently conserved in nature, as it provides a narrow pore for efficient, but selective transport of substrates. The multimerization into pentamers (FNTs) or tetramers (aquaporins) is thought to stabilize the tertiary structure in the membrane, providing the necessary rigidity to maintain the selectivity filter [40, 85]. However, this is where the resemblance to aquaporins stops, as they do not share sequence homology with the FNTs [42, 85], and, as shown in this thesis, they are not efficient water channels in HIS0 state. Intriguingly, certain aquaporins are known to transport nitrate and halide anions (AQP6), or monocarboxylic acids (AQP9 and the nodulin intrinsic proteins NIPs) [90], which

demonstrates that this protein fold holds the capacity to accommodate the permeation of typical FNT substrates. Nonetheless, the molecular details of permeation across FNTs are distinct from aquaporins, as they exhibit entirely different constellation of pore-lining residues resulting in a different geometry of the permeation pore [42, 85].

Role of the pore flexibility. Three elements seem to be essential for anion transport across the FNTs: 1) protonation of the central histidine, 2) substrate protonation on permeation, and 3) pore (side-chain) flexibility. Regarding the latter, the data presented in this thesis demonstrate that orientation of the pore-lining residues probably plays a role in all steps of the permeation process. The results from both, PMF and TI calculations, exhibited single-chain variability and dependence on the starting structures for simulation. Such inherent side-chain flexibility, occurring on the scale of tens to hundreds of nanoseconds, confers the FNTs with capacity to influence both, the substrate movement across the pore, and the protonation reactions involved in the permeation. In this regard, the growingly-popular term “chansporter” might be appropriate [220, 236], implying involvement of slight side-chain movements during permeation, and excluding major conformational changes, as was previously suggested for NirC [78].

Channel opening by histidine protonation and selectivity by substrate protonation. To be in a permeation competent form, the central histidine residue needs to be doubly-protonated. However, the histidine protonation step does not occur in the absence of an anion in the pore. Such “gating” that implies channel opening only in the presence of substrate, has already been observed in nature. In the ClC family of chloride channels and Cl^-/H^+ exchangers, the protonation of a glutamate side-chain is known to be involved in gating [237], and it was shown that this process occurs only in the presence of chloride, which likely influences the $\text{p}K_a$ of the carboxylate group from the gating glutamate [238]. Hence, the necessity of anion present in the pore for channel opening, enables tight regulation of the permeation across the FNTs and prevents collateral leakage. The substrate protonation step on the other hand, combined with geometric and electrostatic features, grants the selectivity of the FNT channels, as only monovalent anions that are strong bases, are amenable to permeation.

Roles of the proton coupling. As schematically shown in fig. 9.1, the permeation mechanism across the FNTs can lead to anion/proton symport in cases the proton used for substrate protonation originates from the same membrane compartment as the permeating substrate. This raises the question whether the proton gradient is used to pump anions across the membrane. Looking at NirC, where the import of nitrous acid is thermodynamically favored over the import of nitrite (although import of both substrates is allowed), the proton gradient might indeed thermodynamically assist the transport of nitrite ions. However, this is not the primary role of the proton coupling to the permeation process. Instead, the proton involvement in the permeation, by subsequent histidine and substrate protonation, is more of a “kinetic” nature, namely to first reduce the permeation barrier for the substrate, and then weaken the binding and enable substrate release. This is in accordance to the observation from uptake assays and electrophysiology experiments, where it was shown that FocA, NirC and HSC only required substrate gradient for permeation [49, 73, 78, 85]. It would also account for fact that pH did have an effect on the reversal potential in NirC, even though pH gradient was not necessary for permeation [78]. However, the effect of proton concentration and

its gradient on permeation across the FNTs is complex and difficult to decompose, since 1) they affect the ratio of the different protonation forms of the permeating substrate, thereby affecting their gradients, and 2) they influence both, the thermodynamics and kinetics of protonation of the central histidine, thereby affecting the opening probability.

Additional to the “gating” role of the proton, the FNTs seem to exhibit different types of relationship between the substrate and proton transport, which can all be accounted for by the mechanism shown in scheme 9.1. For example, FocA in its export mode does not transport protons across the membrane, since the proton involved in permeation exits the same side from where it entered. The putative acetate/proton symporter in yeast was suggested to use the proton gradient to thermodynamically enable the import of acetate [51]. The opposite situation is present in the plasmodial *Pf*FNT, where the protons are the ones moving against their gradient, using the downhill lactate transport [50, 239]. And finally, proton co-transport with the substrate might result in energy dissipation, when both are transported down their gradients, as it is probably the case in NirC.

Additionally, the fate of the proton on substrate exit remains unclear. The possibility that the substrate is deprotonated before it leaves the pore (as envisioned in the proton relay mechanism [41, 78]) can not be excluded. Such a mechanism would account for the observed high conductances, and would in principle prevent dissipating the energy stored as pH gradient, in cases of anion/proton symport when both substrates travel downhill. Alternatively to the proton relay, adaptations in the protein which would favor proton entrance from the opposite side with respect to the anion, regardless of the pH gradient, would also solve this issue. As it is the case with the origin of the proton, its fate in different FNTs probably also depends on the interplay between substrate and pH gradients, as well as the wider metabolic context of the channel. Finally, it should not be forgotten that *in vivo*, the FNTs might be involved in complex regulatory networks, exhibited among other, by direct protein-protein interactions, as it was shown to be the case in FocA [66].

CONCLUSIONS AND OUTLOOK

This work establishes the view of the formate-nitrite transporters as transport proteins of mixed nature, combining channel and carrier properties into a remarkable permeation mechanism. Moreover, it confirms the pivotal role of the central histidine in the permeation process. Regardless of the protonation state of the central histidine, anions are not able to fully permeate the pore. This is due to the very high free energy barrier for permeation across the narrow hydrophobic central portion, in the case of a neutral central histidine, or due to the strong binding to the central binding site, in the case of a positively-charged central histidine. Therefore, a permeation mechanism of at least two protonation events arose, in order to overcome these challenges for anion permeation. In one protonation event the central histidine is protonated, and in another the permeating anion is protonated by the central histidine. With the first protonation event, the permeation barrier is annihilated, allowing for anion internalization into the central binding site. Here, the histidine protonation is coupled to the anion binding, since it is thermodynamically unfavorable in the absence of anions. The second protonation enables release of the substrate from the strong binding, and completion of the permeation. In this way, an efficient pathway for the substrate across the pore is achieved, enabling to reach channel-like permeation rates. The primary role of the proton involvement is to modify the free energy profile along the pore, in order to enable such an efficient anion transport. Additionally, the proton can be used to also thermodynamically aid the anion transport across the membrane, if this is necessary. The coupling of the histidine protonation to the anion binding ensures that the channel opens only when the substrate is present, and prevents proton leakage via the Grotthuss mechanism. The substrate protonation in turn, gives the channel the selectivity for weak acids, and prevents leak of other anions, such as chloride. This mechanism was shown in full details for the nitrite channel NirC. The permeation mechanism across the formate channel FocA involves the additional component of the Ω -loop, whose flexibility seems to be necessary for allowing anions to enter the pore, and probably also affects the protonation events.

10.1 OUTLOOK

The details of the coupled process of histidine protonation and anion binding remain elusive. This work demonstrates that it is energetically more favorable to bring formate and hydronium ions into the central binding site simultaneously, rather than separately. However, it does not conclusively reveal the exact sequence of events that leads to an

open channel. The presence of an anion in one of the vestibules can have effect on the protonability of the central histidine (or other side chains in the vestibule) on one hand, and can aid the hydration of the pore by dragging water molecules on entering. In this way, the proton maybe reaches the central histidine by a complex network including amino acid side chains and water molecules.

Further, the fate of the proton after the substrate crosses the constriction site is unclear. The high conductances measured in electrophysiology experiments suggest that the substrate returns the proton into the pore, thereby regenerating the channel open state faster in comparison to reprotonation by a bulk proton. On the other hand certain FNTs have been reported to function as anion/proton symporters. In principle, both options are allowed depending on the substrate and pH gradients.

An insight into these remaining questions could be obtained by a careful decomposition of the pH effect on permeation rates on one hand, and opening probabilities on the other. This should ideally be achieved in minimal systems in which the directionality of the protein embedding in the bilayer can be controlled.

Finally, the recent finding that FocA binds the formate-producing enzyme, emphasizes the fact that *in vivo*, the permeation across each FNT needs to be considered in the wider network of processes where it belongs. Interaction partners may trigger some structural rearrangements and/or have a major influence on the substrate gradient, thereby assisting the transport in a certain direction. Further interaction studies, supported with structural data and models from computational studies are necessary to clarify these questions.

In summary, this computational study provides the molecular details of the permeation of anions across the FNTs. Combining multiple computational tools enabled for a comprehensive study of various aspects of this process, resulting with a detailed picture of the permeation mechanism. The story of the FNTs however, is probably just at its beginning. As they are found in numerous pathogens, and are not found in humans, FNTs may turn into a rich source for novel drug targets. Moreover, they expand our knowledge about the range of strategies that arose in nature, in order to facilitate substrate transport across biological membranes. This can be witnessed for example in the profound effect a single side-chain protonation has on the free energy profile for permeation. The FNTs research field is surely entering an exciting phase, which could ultimately lead to practical application of these fascinating proteins.

Part V

APPENDIX

RVFFV PROTEIN-LIPID INTERACTIONS

11.1 INTRODUCTION

Membrane fusion is a universal biological process involved among other, in the infection of host cells by viruses that have lipid envelopes [240]. The fusion of the virus and host lipid bilayers is mediated by viral membrane fusion proteins, which act as catalyzers of the fusion process by reducing the kinetic barrier [241]. The Rift Valley fever virus (RVFFV) uses a class II fusion protein to attack and inhabit the host cell. The crystal structure of this protein was solved by our collaborators in its trimeric organization. Importantly, each of the monomers was co-crystallized with a fragment of a short acyl-chain-phosphocholine, revealing a specific binding pocket. Here, conserved arginine (R776) and aspartate (D961) residues interact with the phosphate and choline groups respectively. Moreover, stronger binding of RVFFV to the membrane was observed with increased cholesterol content. Finally, they also investigated the D961K mutation, and revealed no binding of the mutant to the zwitterionic DOPC membrane. However, adding the negatively-charged lipid DOPS to the membrane, managed to rescue the binding of the mutant. In this work, I investigated the molecular details of the RVFFV fusion protein-lipid interactions from three aspects: 1) the influence of the cholesterol content on the binding, 2) the stability of the lipid head binding in the protein binding pocket, and 3) the effect of the D961K mutation on the binding.

11.2 MATERIALS AND METHODS

Atomistic MD simulations were performed with Gromacs 4.6. The Amber-99SB*-ILDN protein force field was used in combination with the Slipids lipid force field. All simulations were performed in explicit solvent (TIP3P water model) and in neutral environment achieved by counterions. To reduce the bias, the lipid fragments from the crystal structure were not taken into account. The temperature and pressure were kept constant at 310K (except in the equilibration phase) and 1bar respectively, using the velocity rescale thermostat and the Berendsen barostat (for equilibration runs) or the Parrinello-Rahman barostat (for production runs). Long-range electrostatic interactions were calculated using PME with a real space cut-off of 1 nm, while the LJ interactions were cut off at 1.5 nm with a switching function from 1.4 nm. The SETTLE algorithm was used to constrain bond lengths and angles of the water molecules, while all other bonds were constrained using LINCS. An integration step of 2 fs was used.

Six sets of simulations (each using a bilayer with a different cholesterol concentration) were performed according to the following procedure. First, the fully restrained protein was docked to an equilibrated membrane patch of pure DOPC or DOPC / cholesterol mixture (10, 20, 30, 40 or 50 mol% cholesterol) with an area of $\sim 65 \text{ nm}^2$, by continuously reducing the protein-bilayer distance with a rate of $1 - 2 \times 10^{-4} \text{ nm ps}^{-1}$, using a harmonic potential with a constant of $2000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ for $\sim 15 \text{ ns}$. In order to achieve better sampling of the lateral diffusion of the lipids, the system was subsequently equilibrated at higher temperature (373 K, or 403 K for the 50% cholesterol system), while keeping the protein-bilayer distance restrained. In a subsequent equilibration step of 5 ns, the distance restraint was released. Finally, the position restraints on the protein were released (with exception of the C_α atoms from the upper third of the protein, which were always held restrained in x - and y -direction in order to prevent protein tilting, while allowing for full freedom of movement in the z -direction). A minimum of three production runs of minimum 550 ns each were performed for each cholesterol concentration, and the analyses were performed on the last 350 ns of each run.

Free energy calculations. For calculation of the difference in the free energy of binding of the wild type protein and the D961K mutant to the membrane, the thermodynamic cycle shown in figure 11.1 was constructed.

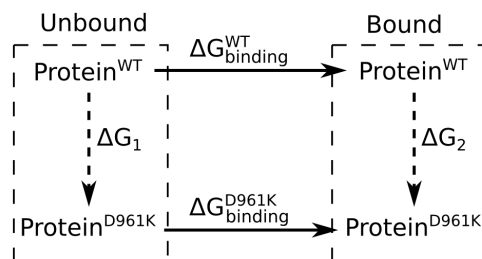


Figure 11.1: Thermodynamic cycle for calculation of $\Delta\Delta G_{\text{binding}}$. The dashed arrows denote the simulated branches, where ΔG_1 and ΔG_2 correspond to the wild type (WT) to mutant (D961K) transformation of the unbound and bound protein, respectively. The difference between the free energies of binding of the wild type and mutant protein is calculated as: $\Delta\Delta G_{\text{binding}} = \Delta G_{\text{binding}}^{\text{D961K}} - \Delta G_{\text{binding}}^{\text{WT}} = \Delta G_2 - \Delta G_1$.

The alchemical perturbation was performed for the two branches (membrane-bound protein and unbound protein in water) using the discrete TI method. The starting structure for the “bound” branch was obtained from a 200 ns production run, following the same procedure from the previous paragraph, with the difference that the lipid coordinates from the crystal structure were retained, and full DOPC molecules were reconstructed upon them. A membrane patch with 30% cholesterol was used. For the “unbound” branch, the crystal structure of the protein was equilibrated for 250 ps fully restrained, followed by 25 ns production run in which the same xy position restraints were used as for all other production simulations. The hybrid structures and topologies were generated using the pmx topology generator [242]. For both branches, the system was equilibrated for 1 ns with the hybrid topology, after which a quick, continuous

perturbation from state A to B was performed in 1 ns. This trajectory was used to spawn the starting structures for the discrete TI steps. Coulomb and LJ parameters were perturbed simultaneously employing soft core potential [243] for all non-bonded interactions. For each branch, 34 TI steps were simulated for 100 ns each.

11.3 RESULTS AND DISCUSSION

The RVFV fusion protein has two hydrophobic residues (phenylalanine and tryptophan) on each chain, which serve as anchors to the membrane (see for example fig. 11.3e). It is evident that with increased cholesterol content, the protein is able to sink deeper into the membrane (fig. 11.2, top). The potential energy due to the protein-lipid interactions (as defined by the force field) is also more favorable as the cholesterol content grows (fig. 11.2, middle). This quantity is indicative of the number of contacts that the protein can form with the membrane lipids. Moreover, fig. 11.4 demonstrates that the cholesterol accumulates in the membrane region directly below the protein. This is visible at all cholesterol concentrations above 10%. The accumulation of cholesterol is also evident from the cholesterol enrichment plot (fig. 11.2, bottom), in which the ratio of the cholesterol densities in the area below the protein and in the rest of the bilayer is shown. This value is greater than 1 for all cholesterol concentrations above 10%, further speaking for the cholesterol accumulation below the protein. Visual inspection of the

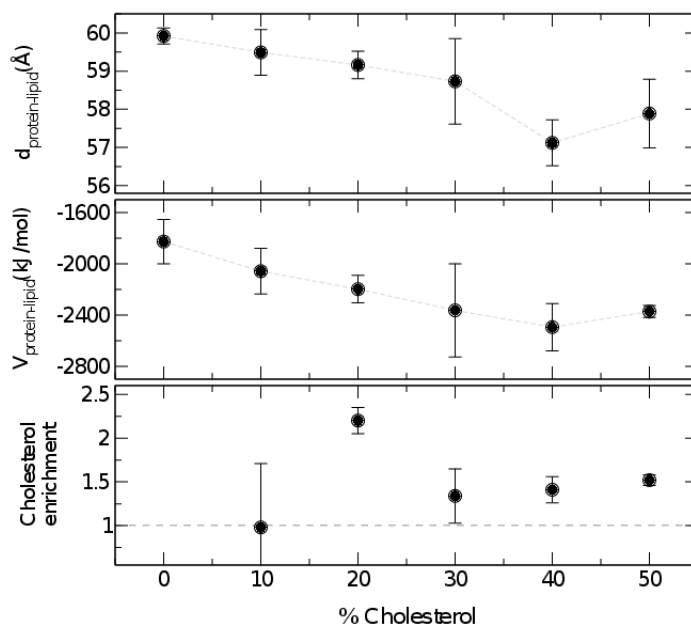


Figure 11.2: Effect of the cholesterol content on the RVFV fusion protein binding. Top: Depth of the protein in the membrane, calculated as the distance in z between the center of mass of the protein and the center of mass of the DOPC phosphorus atoms in the upper leaflet. Middle: Potential energy from the interactions between the protein and the lipids, as calculated from the force field. Bottom: Ratio of the cholesterol densities in the area below the protein and in the rest of the bilayer.

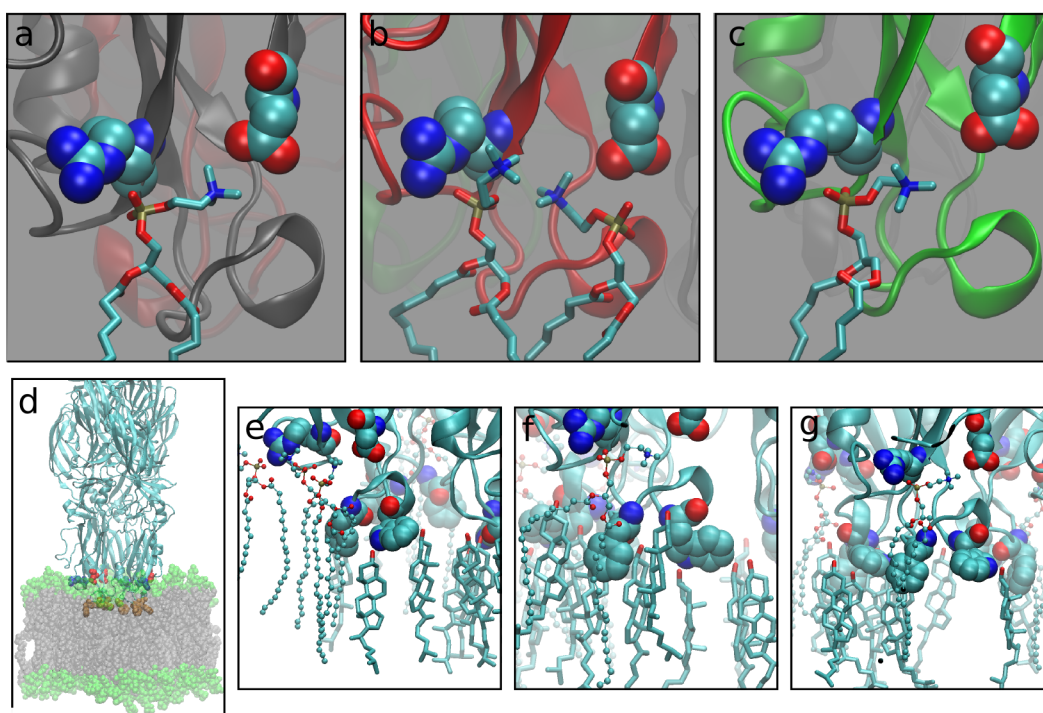


Figure 11.3: RVFV snapshots from simulation of the system with 40% cholesterol. a,b,c) Typical binding of DOPC in the binding pockets. Protein shown as a ribbon diagram, R776 and D961 side chains shown as spheres, and DOPC molecules shown as sticks. d) Side view of a typical simulation system. Protein shown as a ribbon diagram, anchor residues and residues in the binding pocket shown as orange or red/blue spheres, respectively. Lipid bilayer shown as spheres, green: lipid heads, gray: lipid tails. e,f,g) Orientation of the anchor residues in the bilayer. Protein shown as a ribbon diagram, anchor residues and residues in the binding pocket shown as spheres, bound DOPC molecules shown as balls and sticks, and cholesterol molecules near the anchor residues shown as sticks.

MD trajectories reveal that the protein-cholesterol interactions are generally not specific (fig. 11.3e,f,g). The amino group from the anchoring tryptophan often forms a hydrogen bond with the cholesterol hydroxyl group, which is probably a result of the hydrophobic effect of the membrane core environment. The accumulation of cholesterol seems to provide the necessary space in the bilayer to accommodate the binding parts of the protein among the DOPC lipids. In summary, these data confirm that the RVFV fusion protein binds stronger to cholesterol-rich membranes, and it reveals that this is not due to specific interactions of the anchor residues with the cholesterol, but due to overall better docking of the protein to the membrane that results with more favorable interactions.

DOPC molecules usually bound to the binding pocket already during the equilibration simulations. This binding seems to be more dynamic than anticipated, and multiple variations were observed in simulation (fig. 11.3a,b,c). The choline-aspartate interaction is particularly transient, in comparison to the phosphate-arginine salt bridge. On multiple occasions, several DOPC molecules were found to interact with the binding site.

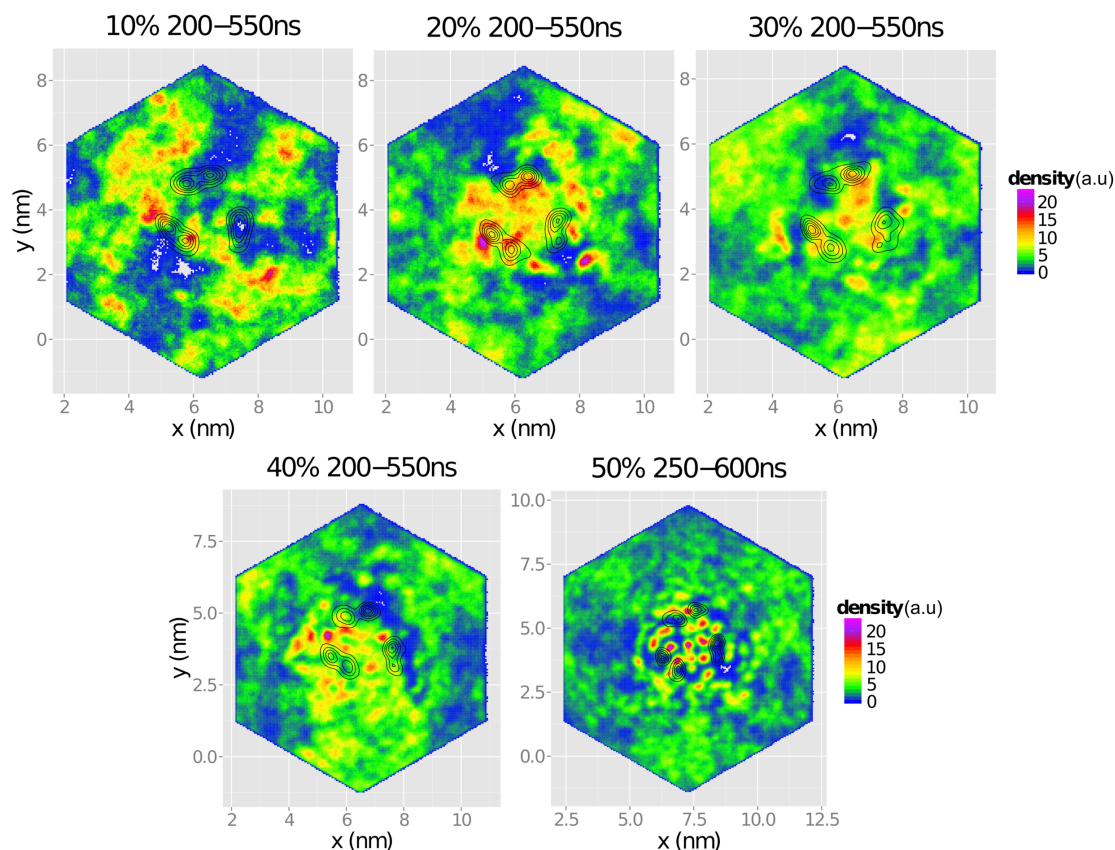


Figure 11.4: Cholesterol accumulation in the region below the protein. The cholesterol density in the upper leaflet is shown as a function of the x - and y -axes in the system. The region of the anchoring residues shown as contours. A.u.: arbitrary units. The cholesterol percentage, and the interval used from each simulation replica for analysis is labeled on top of each figure.

The D961K mutant was found to lose the ability to bind to the membrane in experiment. To study the molecular details of this observation, the difference between the free energy binding between the wild type and mutant protein, $\Delta\Delta G_{\text{binding}} = \Delta G_{\text{binding}}^{\text{D961K}} - \Delta G_{\text{binding}}^{\text{WT}}$, was calculated in a set of TI calculations. Even with very long (100 ns) TI simulations, this value could not be accurately estimated, since it was found to be smaller than the statistical uncertainty of the method. This might be due to the fact that the starting structures for the bound branch were taken from a wild type protein equilibration. However, it might also suggest that the inability of the mutant to bind to a DOPC/cholesterol membrane is rather of a kinetic nature, possibly due to electrostatic repulsion. This is also in accordance to the fact that adding a negatively-charged lipid to the membrane, rescued the binding of the D961K mutant in experiment.

SMALL MOLECULE PARAMETERS

The following topologies were derived for the small molecules parameterized in this work according to the procedure described in section 5.3.1. The hydrogen virtual sites are included. The hydronium ion parameters were taken from reference [187] and modified to be able to use bond constraints and hydrogen virtual sites.

Formic acid

```
[ atomtypes ]
;name      bond_type      mass      charge      ptype      sigma      epsilon      Amb
h5         h5              0.00000    0.00000    A          2.42146e-01  6.27600e-02 ; 1.36  0.0150
c          c              0.00000    0.00000    A          3.39967e-01  3.59824e-01 ; 1.91  0.0860
o          o              0.00000    0.00000    A          2.95992e-01  8.78640e-01 ; 1.66  0.2100
oh         oh              0.00000    0.00000    A          3.06647e-01  8.80314e-01 ; 1.72  0.2104
ho         ho              0.00000    0.00000    A          0.00000e+00  0.00000e+00 ; 0.00  0.0000

[ moleculetype ]
;name      nrexcl
FMA        3

[ atoms ]
;      nr      type      resi      res      atom      cgnr      charge      mass      ; qtot      bond_type
;      1      h5        1      FMA      H1        1      0.045999    0.000000 ; qtot 0.046
;      2      c         1      FMA      C1        2      0.710562    13.01800 ; qtot 0.757
;      3      o         1      FMA      O1        3      -0.578108    16.00000 ; qtot 0.178
;      4      oh        1      FMA      O2        4      -0.650276    16.00000 ; qtot -0.472
;      5      ho        1      FMA      H2        5      0.471822     1.00800 ; qtot 0.000

[ bonds ]
;      ai      aj      funct      r      k
;      1      2      1      1.1053e-01  2.6727e+05 ;      H1 - C1
;      2      3      1      1.2140e-01  5.4225e+05 ;      C1 - O1
;      2      4      1      1.3060e-01  3.9028e+05 ;      C1 - O2
;      4      5      1      9.7400e-02  3.0928e+05 ;      O2 - H2

[ constraints ]
;      ai      aj      funct      c0      c1
;      2      5      2      1.8477e-01

[ pairs ]
;      ai      aj      funct
;      1      5      1 ;      H1 - H2
;      3      5      1 ;      O1 - H2

[ angles ]
;      ai      aj      ak      funct      theta      cth
;      1      2      3      1      1.2326e+02  4.5095e+02 ;      H1 - C1      - O1
;      1      2      4      1      1.1622e+02  4.3790e+02 ;      H1 - C1      - O2
;      2      4      5      1      1.0737e+02  4.2836e+02 ;      C1 - O2      - H2
;      3      2      4      1      1.2288e+02  6.4752e+02 ;      O1 - C1      - O2

[ dihedrals ] ; props
; treated as RBs in GROMACS to use combine multiple AMBER torsions per quartet
;      i      j      k      l      func      C0      C1      C2      C3      C4
;      C5
;      1      2      4      5      3      19.24640    0.00000    -19.24640    0.00000    0.00000
;      0.00000 ;      H1-      C1-      O2-      H2

[ dihedrals ] ; impropers
; treated as props in GROMACS to use correct AMBER analytical function
;      i      j      k      l      func      phase      kd      pn
;      1      3      2      4      1      180.00    43.93200    2 ;      H1-      O1-      C1-      O2

[ virtual_sites3 ]
;      ai      aj      ak      al      funct      c0      c1
```

```

1      2      3      4      2

Formate ion

[ atomtypes ]
;name      bond_type      mass      charge      ptype      sigma      epsilon      Amb
c          c              0.00000    0.00000    A          3.39967e-01  3.59824e-01 ; 1.91  0.0860
o          o              0.00000    0.00000    A          2.95992e-01  8.78640e-01 ; 1.66  0.2100
h5         h5             0.00000    0.00000    A          2.42146e-01  6.27600e-02 ; 1.36  0.0150

[ moleculetype ]
;name      nrexcl
FMT        3

[ atoms ]
;  nr  type  resi  res  atom  cgnr      charge      mass      ; qtot  bond_type
    1    c    1    FMT   C1    1       0.849269    13.01800 ; qtot  0.849
    2    o    1    FMT   O1    2       -0.825141   16.00000 ; qtot  0.024
    3    o    1    FMT   O2    3       -0.825141   16.00000 ; qtot -0.801
    4   h5    1    FMT   H1    4       -0.198987    0.00000 ; qtot -1.000

[ bonds ]
;  ai  aj  funct  r      k      C1 - O1
    1    2    1    1.2140e-01  5.4225e+05 ; C1 - O1
    1    3    1    1.2140e-01  5.4225e+05 ; C1 - O2
    1    4    1    1.1053e-01  2.6727e+05 ; C1 - H1

[ angles ]
;  ai  aj  ak  funct  theta  cth      O1 - C1  - O2
    2    1    3    1    1.3038e+02  6.5413e+02 ; O1 - C1  - O2
    2    1    4    1    1.2326e+02  4.5095e+02 ; O1 - C1  - H1
    3    1    4    1    1.2326e+02  4.5095e+02 ; O2 - C1  - H1

[ dihedrals ] ; impropers
; treated as propers in GROMACS to use correct AMBER analytical function
;  i  j  k  l  func  phase  kd  pn      O2-  C1-  O1-  H1
    3    1    2    4    1   180.00  4.60240  2 ; O2-  C1-  O1-  H1

[ virtual_sites3 ]
;  ai  aj  ak  al  funct  c0      c1
    4    1    2    3    2

Nitrous acid

[ atomtypes ]
;name      bond_type      mass      charge      ptype      sigma      epsilon      Amb
o          o              0.00000    0.00000    A          2.95992e-01  8.78640e-01 ; 1.66  0.2100
n2         n2             0.00000    0.00000    A          3.25000e-01  7.11280e-01 ; 1.82  0.1700
oh         oh              0.00000    0.00000    A          3.06647e-01  8.80314e-01 ; 1.72  0.2104
ho         ho              0.00000    0.00000    A          0.00000e+00  0.00000e+00 ; 0.00  0.0000

[ moleculetype ]
;name      nrexcl
NIA        3

[ atoms ]
;  nr  type  resi  res  atom  cgnr      charge      mass      ; qtot  bond_type
    1    o    1    NIA   O1    1      -0.132386    16.00000 ; qtot -0.132
    2   n2    1    NIA   N1    2       0.094006    14.01000 ; qtot -0.038
    3    oh    1    NIA   O2    3      -0.346016    16.00000 ; qtot -0.384
    4    ho    1    NIA   H1    4       0.384396     1.00800 ; qtot  0.000

[ bonds ]
;  ai  aj  funct  r      k      O1 - N1
    1    2    1    1.2090e-01  6.6099e+05 ; O1 - N1
    2    3    1    1.3940e-01  3.4828e+05 ; N1 - O2
    3    4    1    9.7400e-02  3.0928e+05 ; O2 - H1

[ constraints ]
;  ai  aj  funct  c0      c1
    2    4    2      1.8684e-01

[ pairs ]
;  ai  aj  funct  O1 - H1
    1    4    1 ; O1 - H1

[ angles ]
;  ai  aj  ak  funct  theta  cth      O1 - N1  - O2
    1    2    3    1    1.1215e+02  6.3229e+02 ; O1 - N1  - O2
    2    3    4    1    1.0274e+02  4.2317e+02 ; N1 - O2  - H1

[ dihedrals ] ; propers
; treated as RBs in GROMACS to use combine multiple AMBER torsions per quartet
;  i  j  k  l  func  C0      C1      C2      C3      C4
    C5

```

```

1      2      3      4      3      26.77760      0.00000      -26.77760      0.00000      0.00000
      0.00000 ;      O1-      N1-      O2-      H1

```

Nitrite ion

```

[ atomtypes ]
;name      bond_type      mass      charge      ptype      sigma      epsilon      Amb
o          o              0.00000      0.00000      A          2.95992e-01      8.78640e-01 ; 1.66      0.2100
n2         n2              0.00000      0.00000      A          3.25000e-01      7.11280e-01 ; 1.82      0.1700

[ moleculetype ]
;name      nrexcl
NIT        3

[ atoms ]
;      nr      type      resi      res      atom      cgnr      charge      mass      ; qtot      bond_type
      1      o          1      NIT      O1          1      -0.475567      16.00000 ; qtot -0.476
      2      n2         1      NIT      N1          2      -0.048866      14.01000 ; qtot -0.524
      3      o          1      NIT      O2          3      -0.475567      16.00000 ; qtot -1.000

[ bonds ]
;      ai      aj      funct      r      k
      1      2      1      1.2090e-01      6.6099e+05 ;      O1 - N1
      2      3      1      1.2090e-01      6.6099e+05 ;      N1 - O2

[ angles ]
;      ai      aj      ak      funct      theta      cth
      1      2      3      1      1.1537e+02      6.7798e+02 ;      O1 - N1      - O2

```

Hydrogen sulfide

```

[ atomtypes ]
;name      bond_type      mass      charge      ptype      sigma      epsilon      Amb
sh         sh              0.00000      0.00000      A          3.56359e-01      1.04600e+00 ; 2.00      0.2500
hs         hs              0.00000      0.00000      A          1.06908e-01      6.56888e-02 ; 0.60      0.0157

[ moleculetype ]
;name      nrexcl
HSH        3

[ atoms ]
;      nr      type      resi      res      atom      cgnr      charge      mass      ; qtot      bond_type
      1      sh          1      HSH      S1          1      -0.357670      32.06000 ; qtot -0.358
      2      hs          1      HSH      H1          2      0.178835      1.00800 ; qtot -0.179
      3      hs          1      HSH      H2          3      0.178835      1.00800 ; qtot 0.000

[ bonds ]
;      ai      aj      funct      r      k
      1      2      1      1.3370e-01      2.5288e+05 ;      S1 - H1
      1      3      1      1.3370e-01      2.5288e+05 ;      S1 - H2

[ constraints ]
;      ai      aj      funct      c0      c1
      2      3      2      1.9510e-01

[ angles ]
;      ai      aj      ak      funct      theta      cth
      2      1      3      1      9.3720e+01      1.9606e+02 ;      H1 - S1      - H2

```

Hydronium ion

```

[ atomtypes ]
;name      bond_type      mass      charge      ptype      sigma      epsilon      Amb
OW+        OW+              0.00000      0.00000      A          3.22000e-01      6.430808e-01

[ moleculetype ]
;molname      nrexcl
H3O           2

[ atoms ]
; id      at      type      res      nr      res      name      at      name      cg      nr      charge      mass
1      OW+          1      H3O      OW          1      -0.590      16.00000
2      HW          1      H3O      HW1         1      0.530      1.51200
3      HW          1      H3O      HW2         1      0.530      1.51200
4      HW          1      H3O      HW3         1      0.530      0.00000

[ constraints ]
1      2      1      0.10200
1      3      1      0.10200
;1      4      1      0.10200
2      3      1      0.169124
;2      4      1      0.169124
;3      4      1      0.169124

[ virtual_sites3 ]

```

```
; Site      from      funct      a      b      c
4          1 2 3      4      -0.59899 -0.59899 -7.85056

[ exclusions ]
1          2          3          4
2          1          3          4
3          1          2          4
4          1          2          3
```

BIBLIOGRAPHY

1. Alberts, B. *et al.* *Molecular Biology of the Cell* 4th edn. (Garland Science, 2002).
2. Ashcroft, F., Gadsby, D. & Miller, C. Introduction. The blurred boundary between channels and transporters. *Philos Trans R Soc Lond B Biol Sci* **364**, 145–7 (2009).
3. Hille, B. *et al.* *Ion channels of excitable membranes* (Sinauer Sunderland, MA, 2001).
4. Karplus, M & McCammon, J. Dynamics of proteins: elements and function. *Annual review of biochemistry* **52**, 263–300 (1983).
5. Mittermaier, A. & Kay, L. E. New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224–8 (2006).
6. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* **9**, 646–52 (2002).
7. McCammon, J. & Karplus, M. Simulation of protein dynamics. *Annual Review of Physical Chemistry* **31**, 29–45 (1980).
8. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–72 (2007).
9. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**, 789–96 (2009).
10. Orozco, M. A theoretical view of protein dynamics. *Chem Soc Rev* **43**, 5051–66 (2014).
11. Frauenfelder, H, Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–603 (1991).
12. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat Chem Biol* **5**, 808–14 (2009).
13. Karplus, M & Kuriyan, J. Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* **102**, 6679–85 (2005).
14. Bu, Z. & Callaway, D. J. E. Proteins move! Protein dynamics and long-range allostery in cell signaling. *Adv Protein Chem Struct Biol* **83**, 163–221 (2011).
15. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203–7 (2009).
16. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–30 (1973).
17. Diaz-Espinoza, R. & Soto, C. High-resolution structure of infectious prion protein: the final frontier. *Nat Struct Mol Biol* **19**, 370–7 (2012).

18. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197–208 (2005).
19. Karplus, M & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347**, 631–9 (1990).
20. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585–90 (1977).
21. Lindorff-Larsen, K. *et al.* Systematic validation of protein force fields against experimental data. *PLoS One* **7**, e32131 (2012).
22. Beauchamp, K. A., Lin, Y.-S., Das, R. & Pande, V. S. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *Journal of chemical theory and computation* **8**, 1409–1414 (2012).
23. Martín-García, F., Papaleo, E., Gomez-Puertas, P., Boomsma, W. & Lindorff-Larsen, K. Comparing molecular dynamics force fields in the essential subspace. *PLoS One* **10**, e0121114 (2015).
24. Shaw, D. E. *et al.* Millisecond-scale molecular dynamics simulations on Anton in High Performance Computing Networking, Storage and Analysis, *Proceedings of the Conference on* (2009), 1–11.
25. Sanbonmatsu, K. Y. & Tung, C.-S. High performance computing in biology: multimillion atom simulations of nanoscale systems. *J Struct Biol* **157**, 470–80 (2007).
26. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
27. Mitsutake, A., Mori, Y. & Okamoto, Y. Enhanced sampling algorithms. *Biomolecular Simulations: Methods and Protocols*, 153–195 (2013).
28. Bill, R. M. *et al.* Overcoming barriers to membrane protein structure determination. *Nat Biotechnol* **29**, 335–40 (2011).
29. Piggot, T. J., Piñeiro, Á. & Khalid, S. Molecular Dynamics Simulations of Phosphatidylcholine Membranes: A Comparative Force Field Study. *J Chem Theory Comput* **8**, 4593–609 (2012).
30. Knight, C. J. & Hub, J. S. MemGen: a general web server for the setup of lipid membrane simulation systems. *Bioinformatics* **31**, 2897–9 (2015).
31. Wolf, M. G., Hoefling, M., Aponte-Santamaría, C., Grubmüller, H. & Groenhof, G. g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *J Comput Chem* **31**, 2169–74 (2010).
32. Kutzner, C., Grubmüller, H., de Groot, B. L. & Zachariae, U. Computational electrophysiology: the molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophys J* **101**, 809–17 (2011).
33. De Groot, B. L. & Grubmüller, H. Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* **294**, 2353–7 (2001).

34. Hub, J. S. & de Groot, B. L. Mechanism of selectivity in aquaporins and aquaglyceroporins. *Proc Natl Acad Sci U S A* **105**, 1198–203 (2008).
35. De Groot, B. L., Frigato, T., Helms, V. & Grubmüller, H. The mechanism of proton exclusion in the aquaporin-1 water channel. *J Mol Biol* **333**, 279–93 (2003).
36. Köpfer, D. A. *et al.* Ion permeation in K⁺ channels occurs by direct Coulomb knock-on. *Science* **346**, 352–355 (2014).
37. Medovoy, D., Perozo, E. & Roux, B. Multi-ion free energy landscapes underscore the microscopic mechanism of ion selectivity in the KcsA channel. *Biochim Biophys Acta*. doi:10.1016/j.bbamem.2016.02.019 (2016).
38. Chanda, B., Asamoah, O. K., Blunck, R., Roux, B. & Bezannila, F. Gating charge displacement in voltage-gated ion channels involves limited transmembrane movement. *Nature* **436**, 852–6 (2005).
39. Jensen, M. Ø. *et al.* Mechanism of voltage gating in potassium channels. *Science* **336**, 229–33 (2012).
40. Waight, A. B., Czyzewski, B. K. & Wang, D.-N. Ion selectivity and gating mechanisms of FNT channels. *Curr Opin Struct Biol* **23**, 499–506 (2013).
41. Lü, W. *et al.* The formate/nitrite transporter family of anion channels. *Biol Chem* **394**, 715–27 (2013).
42. Wang, Y. *et al.* Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature* **462**, 467–72 (2009).
43. Suppmann, B. & Sawers, G. Isolation and characterization of hypophosphite-resistant mutants of *Escherichia coli*: identification of the FocA protein, encoded by the pfl operon, as a putative formate transporter. *Mol Microbiol* **11**, 965–82 (1994).
44. White, W. B. & Ferry, J. G. Identification of formate dehydrogenase-specific mRNA species and nucleotide sequence of the fdhC gene of *Methanobacterium formicicum*. *J Bacteriol* **174**, 4997–5004 (1992).
45. Nölling, J. & Reeve, J. N. Growth- and substrate-dependent transcription of the formate dehydrogenase (fdhCAB) operon in *Methanobacterium thermoformicicum* Z-245. *J Bacteriol* **179**, 899–908 (1997).
46. Clegg, S., Yu, F., Griffiths, L. & Cole, J. A. The roles of the polytopic membrane proteins NarK, NarU and NirC in *Escherichia coli* K-12: two nitrate and three nitrite transporters. *Mol Microbiol* **44**, 143–55 (2002).
47. Galván, A., Rexach, J., Mariscal, V. & Fernández, E. Nitrite transport to the chloroplast in *Chlamydomonas reinhardtii*: molecular evidence for a regulated process. *J Exp Bot* **53**, 845–53 (2002).
48. Unkles, S. E. *et al.* Physiological and biochemical characterization of AnNitA, the *Aspergillus nidulans* high-affinity nitrite transporter. *Eukaryot Cell* **10**, 1724–32 (2011).

49. Czyzewski, B. K. & Wang, D.-N. Identification and characterization of a bacterial hydrosulphide ion channel. *Nature* **483**, 494–7 (2012).
50. Wu, B. *et al.* Identity of a Plasmodium lactate/H(+) symporter structurally unrelated to human transporters. *Nat Commun* **6**, 6284 (2015).
51. Casal, M., Cardoso, H & Leão, C. Mechanisms regulating the transport of acetic acid in *Saccharomyces cerevisiae*. *Microbiology* **142** (Pt 6), 1385–90 (1996).
52. Yamano, T., Sato, E., Iguchi, H., Fukuda, Y. & Fukuzawa, H. Characterization of cooperative bicarbonate uptake into chloroplast stroma in the green alga *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci U S A* **112**, 7315–20 (2015).
53. Miller, C. ClC chloride channels viewed through a transporter lens. *Nature* **440**, 484–9 (2006).
54. Das, P., Lahiri, A., Lahiri, A. & Chakravorty, D. Novel role of the nitrite transporter NirC in *Salmonella* pathogenesis: SPI2-dependent suppression of inducible nitric oxide synthase in activated macrophages. *Microbiology* **155**, 2476–89 (2009).
55. Peakman, T *et al.* Nucleotide sequence, organisation and structural analysis of the products of genes in the nirB-cysG region of the *Escherichia coli* K-12 chromosome. *Eur J Biochem* **191**, 315–23 (1990).
56. Saier, M. H. Jr *et al.* Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochim Biophys Acta* **1422**, 1–56 (1999).
57. Wang, Y. *et al.* Nitrite transport is mediated by the nitrite-specific high-affinity NitA transporter and by nitrate transporters NrtA, NrtB in *Aspergillus nidulans*. *Fungal Genet Biol* **45**, 94–102 (2008).
58. Rexach, J, Fernández, E & Galván, A. The *Chlamydomonas reinhardtii* Nar1 gene encodes a chloroplast membrane protein involved in nitrite transport. *Plant Cell* **12**, 1441–53 (2000).
59. Mariscal, V. *et al.* Differential regulation of the *Chlamydomonas* Nar1 gene family by carbon and nitrogen. *Protist* **157**, 421–33 (2006).
60. Wang, Y. & Spalding, M. H. Acclimation to very low CO₂: contribution of limiting CO₂ inducible proteins, LCIB and LCIA, to inorganic carbon uptake in *Chlamydomonas reinhardtii*. *Plant Physiol* **166**, 2040–50 (2014).
61. Marchetti, R. V. *et al.* A lactate and formate transporter in the intraerythrocytic malaria parasite, *Plasmodium falciparum*. *Nat Commun* **6**, 6721 (2015).
62. Saier, M. H. Jr, Tran, C. V. & Barabote, R. D. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* **34**, D181–6 (2006).
63. Saier, M. H. Jr, Yen, M. R., Noto, K., Tamang, D. G. & Elkan, C. The Transporter Classification Database: recent advances. *Nucleic Acids Res* **37**, D274–8 (2009).
64. Saier, M. H. Jr, Reddy, V. S., Tamang, D. G. & Västermark, A. The transporter classification database. *Nucleic Acids Res* **42**, D251–8 (2014).

65. Saier, M. H. Jr *et al.* The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res* **44**, D372–9 (2016).
66. Doberenz, C. *et al.* Pyruvate formate-lyase interacts directly with the formate channel FocA to regulate formate translocation. *J Mol Biol* **426**, 2827–39 (2014).
67. Hunger, D., Doberenz, C. & Sawers, R. G. Identification of key residues in the formate channel FocA that control import and export of formate. *Biol Chem* **395**, 813–25 (2014).
68. Leonhartsberger, S., Korsa, I. & Böck, A. The molecular biology of formate metabolism in enterobacteria. *J Mol Microbiol Biotechnol* **4**, 269–76 (2002).
69. Sawers, R. G. & Clark, D. P. Fermentative Pyruvate and Acetyl-Coenzyme A Metabolism. *EcoSal Plus* **1**. doi:10.1128/ecosalplus.3.5.3 (2004).
70. Sawers, R. G., Blokesch, M. & Böck, A. Anaerobic Formate and Hydrogen Metabolism. *EcoSal Plus* **1**. doi:10.1128/ecosalplus.3.5.4 (2004).
71. Sawers, R. G. Formate and its role in hydrogen production in *Escherichia coli*. *Biochem Soc Trans* **33**, 42–6 (2005).
72. Beyer, L. *et al.* Coordination of FocA and pyruvate formate-lyase synthesis in *Escherichia coli* demonstrates preferential translocation of formate over other mixed-acid fermentation products. *J Bacteriol* **195**, 1428–35 (2013).
73. Lü, W. *et al.* The formate channel FocA exports the products of mixed-acid fermentation. *Proc Natl Acad Sci U S A* **109**, 13254–9 (2012).
74. Canfield, D. E., Glazer, A. N. & Falkowski, P. G. The evolution and future of Earth's nitrogen cycle. *Science* **330**, 192–6 (2010).
75. Einsle, O. & Kroneck, P. M. H. Structural basis of denitrification. *Biol Chem* **385**, 875–83 (2004).
76. Cole, J. A. & Richardson, D. J. Respiration of Nitrate and Nitrite. *EcoSal Plus* **3**. doi:10.1128/ecosal.3.2.5 (2008).
77. Andrade, S. L. A. & Einsle, O. The tricky task of nitrate/nitrite antiport. *Angew Chem Int Ed Engl* **52**, 10422–4 (2013).
78. Lü, W. *et al.* Structural and functional characterization of the nitrite channel NirC from *Salmonella typhimurium*. *Proc Natl Acad Sci U S A* **109**, 18395–400 (2012).
79. Jia, W. & Cole, J. A. Nitrate and nitrite transport in *Escherichia coli*. *Biochem Soc Trans* **33**, 159–61 (2005).
80. Jia, W., Tovell, N., Clegg, S., Trimmer, M. & Cole, J. A single channel for nitrate uptake, nitrite export and nitrite uptake by *Escherichia coli* NarU and a role for NirC in nitrite export and uptake. *Biochem J* **417**, 297–304 (2009).
81. Sparacino-Watkins, C., Stolz, J. F. & Basu, P. Nitrate and periplasmic nitrate reductases. *Chemical Society Reviews* **43**, 676–706 (2014).
82. Dhillon, A., Goswami, S., Riley, M., Teske, A. & Sogin, M. Domain evolution and functional diversification of sulfite reductases. *Astrobiology* **5**, 18–29 (2005).

83. Reis, M. A., Almeida, J. S., Lemos, P. C. & Carrondo, M. J. Effect of hydrogen sulfide on growth of sulfate reducing bacteria. *Biotechnol Bioeng* **40**, 593–600 (1992).
84. Goffredi, S. K., Childress, J. J., Desaulniers, N. T. & Lallier, F. J. Sulfide acquisition by the vent worm *Riftia pachyptila* appears to be via uptake of HS⁻, rather than H₂S. *J Exp Biol* **200**, 2609–16 (1997).
85. Waight, A. B., Love, J. & Wang, D.-N. Structure and mechanism of a pentameric formate channel. *Nat Struct Mol Biol* **17**, 31–7 (2010).
86. Lü, W. *et al.* pH-dependent gating in a FocA formate channel. *Science* **332**, 352–354 (2011).
87. Smart, O. S., Neduvilil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J Mol Graph* **14**, 354–60, 376 (1996).
88. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037–41 (2001).
89. Rycovska, A., Hatahet, L., Fendler, K. & Michel, H. The nitrite transport protein NirC from *Salmonella typhimurium* is a nitrite proton antiporter. *Biochim Biophys Acta* **1818**, 1342–50 (2012).
90. Rambow, J., Wu, B., Rönfeldt, D. & Beitz, E. Aquaporins with anion/monocarboxylate permeability: mechanisms, relevance for pathogen-host interactions. *Front Pharmacol* **5**, 199 (2014).
91. Rycovska-Blume, A., Lü, W., Andrade, S., Fendler, K. & Einsle, O. Structural and Functional Studies of NirC from *Salmonella typhimurium*. *Methods Enzymol* **556**, 475–97 (2015).
92. Berendsen, H. J. *Simulating the physical world: hierarchical modeling from quantum mechanics to fluid dynamics* (Cambridge University Press, 2007).
93. Kiefer, C. Emergence of a classical Universe from quantum gravity and cosmology. *Philos Trans A Math Phys Eng Sci* **370**, 4566–75 (2012).
94. Mackerell, A. D. Jr. Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem* **25**, 1584–604 (2004).
95. Monticelli, L. & Tieleman, D. P. Force fields for classical molecular dynamics. *Methods Mol Biol* **924**, 197–213 (2013).
96. Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Adv Protein Chem* **66**, 27–85 (2003).
97. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586–616 (1998).
98. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* **25**, 1656–76 (2004).

99. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **118**, 11225–11236 (1996).
100. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157–74 (2004).
101. Vanommeslaeghe, K *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **31**, 671–90 (2010).
102. Berger, O, Edholm, O & Jähnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys J* **72**, 2002–13 (1997).
103. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry B* **114**, 7830–7843 (2010).
104. Dickson, C. J., Rosso, L., Betz, R. M., Walker, R. C. & Gould, I. R. GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter* **8**, 9617–9627 (2012).
105. Jämbeck, J. P. M. & Lyubartsev, A. P. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *J Phys Chem B* **116**, 3164–79 (2012).
106. Kukol, A. Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins. *J Chem Theory Comput* **5**, 615–26 (2009).
107. Piggot, T. J., Holdbrook, D. A. & Khalid, S. Electroporation of the E. coli and S. aureus membranes: molecular dynamics simulations of complex bacterial membranes. *The Journal of Physical Chemistry B* **115**, 13381–13388 (2011).
108. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–25 (2006).
109. Best, R. B. & Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* **113**, 9004–15 (2009).
110. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–8 (2010).
111. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys J* **100**, L47–9 (2011).
112. Tieleman, D. P., Berendsen, H. J. & Sansom, M. S. An alamethicin channel in a lipid bilayer: molecular dynamics simulations. *Biophysical journal* **76**, 1757–1769 (1999).

113. Bachar, M., Brunelle, P., Tieleman, D. P. & Rauk, A. Molecular dynamics simulation of a polyunsaturated lipid bilayer susceptible to lipid peroxidation. *The Journal of Physical Chemistry B* **108**, 7170–7179 (2004).
114. Cordero, A., Caltabiano, G. & Pardo, L. Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters. *Journal of Chemical Theory and Computation* **8**, 948–958 (2012).
115. Berendsen, H. J., van der Spoel, D. & van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, 43–56 (1995).
116. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J Comput Chem* **26**, 1701–18 (2005).
117. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **4**, 435–47 (2008).
118. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
119. Hockney, R., Goel, S. & Eastwood, J. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics* **14**, 148–158 (1974).
120. Van der Spoel, D, Lindahl, E, Hess, B & the GROMACS development team. *GROMACS User Manual version 4.6* www.gromacs.org (2012).
121. Miyamoto, S. & Kollman, P. A. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* **13**, 952–962 (1992).
122. Hess, B., Bekker, H., Berendsen, H. J., Fraaije, J. G., *et al.* LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **18**, 1463–1472 (1997).
123. Feenstra, K. A., Hess, B. & Berendsen, H. J. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem* **20**, 786–798 (1999).
124. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101 (2007).
125. Berendsen, H. in *Computer Simulation in Materials Science* 139–155 (Springer, 1991).
126. Berendsen, H. J., Postma, J. v., van Gunsteren, W. F., DiNola, A. & Haak, J. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **81**, 3684–3690 (1984).
127. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **52**, 7182–7190 (1981).
128. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **50**, 1055–1076 (1983).

129. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry* **4**, 187–217 (1983).
130. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics* **98**, 10089–10092 (1993).
131. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of chemical physics* **103**, 8577–8593 (1995).
132. Kirkwood, J. G. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics* **3**, 300–313 (1935).
133. Hub, J. S., De Groot, B. L. & Van Der Spoel, D. g_wham - A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of Chemical Theory and Computation* **6**, 3713–3720 (2010).
134. Roux, B. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* **91**, 275–282 (1995).
135. Zuckerman, D. M. *Statistical Physics of Biomolecules: An Introduction* (CRC Press, 2010).
136. Hub, J. S. *Selectivity, Regulation, and Inhibition of Aquaporin Channels. A Molecular Dynamics Study* PhD thesis (Georg-August-Universität Göttingen, 2008).
137. Kosztin, I., Barz, B. & Janosi, L. Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality. *J Chem Phys* **124**, 64106 (2006).
138. Torrie, G. M. & Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard-Jones fluid. *Chemical Physics Letters* **28**, 578–581 (1974).
139. Valleau, J. & Torrie, G. in *Statistical Mechanics* 169–194 (Springer, 1977).
140. Kästner, J. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 932–942 (2011).
141. Haydock, C, Sharp, J. C. & Prendergast, F. G. Tryptophan-47 rotational isomerization in variant-3 scorpion neurotoxin. A combination thermodynamic perturbation and umbrella sampling study. *Biophys J* **57**, 1269–79 (1990).
142. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **13**, 1011–1021 (1992).
143. Kästner, J. & Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration". *J Chem Phys* **123**, 144104 (2005).
144. Shirts, M. R., Mobley, D. L. & Chodera, J. D. Alchemical free energy calculations: ready for prime time. *Ann Rep Comput Chem* **3**, 41–59 (2007).

145. Christ, C. D., Mark, A. E. & van Gunsteren, W. F. Basic ingredients of free energy calculations: a review. *J Comput Chem* **31**, 1569–82 (2010).
146. Shirts, M. R. & Mobley, D. L. An introduction to best practices in free energy calculations. *Methods Mol Biol* **924**, 271–311 (2013).
147. Gapsys, V., Michielssens, S., Peters, J. H., de Groot, B. L. & Leonov, H. Calculation of binding free energies. *Methods Mol Biol* **1215**, 173–209 (2015).
148. Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. nonpolar gases. *The Journal of Chemical Physics* **22**, 1420–1426 (1954).
149. Brunk, E. & Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem Rev* **115**, 6217–63 (2015).
150. Friesner, R. A. Ab initio quantum chemistry: methodology and applications. *Proc Natl Acad Sci U S A* **102**, 6648–53 (2005).
151. Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik* **61**, 126–148 (1930).
152. Hartree, D. Cambridge Phil. Soc. 1928, 24, 89–110. The wave mechanics of an atom with a non-coulomb central field. I. Theory and methods.(b) Fock, V. *Z. Physik* **61**, 126–148 (1930).
153. Slater, J. C. Note on Hartree’s method. *Physical Review* **35**, 210 (1930).
154. Møller, C. & Plesset, M. S. Note on an approximation treatment for many-electron systems. *Physical Review* **46**, 618 (1934).
155. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Physical review* **136**, B864 (1964).
156. Sham, L. & Kohn, W. One-particle properties of an inhomogeneous interacting electron gas. *Physical Review* **145**, 561 (1966).
157. Kohn, W., Becke, A. D. & Parr, R. G. Density functional theory of electronic structure. *The Journal of Physical Chemistry* **100**, 12974–12980 (1996).
158. Senn, H. M. & Thiel, W. QM/MM methods for biomolecular systems. *Angew Chem Int Ed Engl* **48**, 1198–229 (2009).
159. Groenhof, G. in *Biomolecular Simulations* 43–66 (Springer, 2013).
160. Ullmann, R. T. & Ullmann, G. M. GMCT : a Monte Carlo simulation package for macromolecular receptors. *J Comput Chem* **33**, 887–900 (2012).
161. Ullmann, R. T. *Monte Carlo Simulation Methods for Studying the Thermodynamics of Ligand Binding Transfer Processes in Biomolecules* PhD thesis (University of Bayreuth, 2012).
162. developers, T. G. *GMCT Manual* version 1.2.3. Structural Biology / Bioinformatics Group, University of Bayreuth ().
163. Ullmann, R. T., Andrade, S. L. & Ullmann, G. M. Thermodynamics of transport through the ammonium transporter Amt-1 investigated with free energy calculations. *The Journal of Physical Chemistry B* **116**, 9690–9703 (2012).

164. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–42 (2000).
165. Rodriguez, R, Chinae, G, Lopez, N, Pons, T & Vriend, G. Homology modeling, model and software evaluation: three related resources. *Bioinformatics* **14**, 523–8 (1998).
166. Hekkelman, M. L. *et al.* WIWS: a protein structure bioinformatics Web service collection. *Nucleic Acids Res* **38**, W719–23 (2010).
167. Chinae, G, Padron, G, Hooft, R. W., Sander, C & Vriend, G. The use of position-specific rotamers in model building by homology. *Proteins* **23**, 415–21 (1995).
168. Mandell, D. J., Coutsiadis, E. A. & Kortemme, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat Methods* **6**, 551–2 (2009).
169. Coutsiadis, E. A., Seok, C., Wester, M. J. & Dill, K. A. Resultants and loop closure. *International Journal of Quantum Chemistry* **106**, 176–189 (2006).
170. Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 1.7* 2015.
171. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926–935 (1983).
172. Wiberg, E. *Die chemische Affinität: eine erste Einführung in die Lehre von der Triebkraft chemischer Reaktionen* (Walter de Gruyter, 1972).
173. Berendsen, H. & Van Gunsteren, W. Practical algorithms for dynamic simulations. *Molecular-dynamics simulation of statistical-mechanical systems*, 43–65 (1986).
174. Allen, T. W., Andersen, O. S. & Roux, B. Molecular dynamics - potential of mean force calculations as a tool for understanding ion permeation and selectivity in narrow channels. *Biophys Chem* **124**, 251–67 (2006).
175. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* **25**, 247–60 (2006).
176. Frisch, M. J. *et al.* *Gaussian09 Revision E.01* Gaussian Inc. Wallingford CT 2009.
177. Sousa da Silva, A. W. & Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interface. *BMC Res Notes* **5**, 367 (2012).
178. Joung, I. S. & Cheatham, T. E. 3rd. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B* **112**, 9020–41 (2008).
179. Domański, J., Stansfeld, P. J., Sansom, M. S. P. & Beckstein, O. Lipidbook: a public repository for force-field parameters used in membrane simulations. *J Membr Biol* **236**, 255–8 (2010).

180. Zhao, W., Róg, T., Gurtovenko, A. A., Vattulainen, I. & Karttunen, M. Atomic-scale structure and electrostatics of anionic palmitoyloleoylphosphatidylglycerol lipid bilayers with Na⁺ counterions. *Biophys J* **92**, 1114–24 (2007).
181. Zhao, W., Róg, T., Gurtovenko, A. A., Vattulainen, I. & Karttunen, M. Role of phosphatidylglycerols in the stability of bacterial membranes. *Biochimie* **90**, 930–8 (2008).
182. Berendsen, H. J., Postma, J. P., van Gunsteren, W. F. & Hermans, J. in *Intermolecular forces* 331–342 (Springer, 1981).
183. Reif, M. M. & Hünenberger, P. H. Computation of methodology-independent single-ion solvation properties from molecular simulations. IV. Optimized Lennard-Jones interaction parameter sets for the alkali and halide ions in water. *J Chem Phys* **134**, 144104 (2011).
184. Åqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *Journal of Physical Chemistry* **94**, 8021–8024 (1990).
185. Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **22**, 245–268 (1976).
186. Shirts, M. R., Bair, E., Hooker, G. & Pande, V. S. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys Rev Lett* **91**, 140601 (2003).
187. Wolf, M. G. & Groenhof, G. Explicit proton transfer in classical molecular dynamics simulations. *J Comput Chem* **35**, 657–71 (2014).
188. Neese, F. The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2**, 73–78 (2012).
189. Azzone, G. *et al.* Transmembrane measurements across bioenergetic membranes. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1183**, 1–3 (1993).
190. Mitchell, P. Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. 1966. *Biochim Biophys Acta* **1807**, 1507–38 (2011).
191. Kelly, C. P., Cramer, C. J. & Truhlar, D. G. Aqueous solvation free energies of ions and ion-water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. *J Phys Chem B* **110**, 16066–81 (2006).
192. Becker, K. H., Kleffmann, J., Kurtenbach, R. & Wiesen, P. Solubility of nitrous acid (HONO) in sulfuric acid solutions. *The Journal of Physical Chemistry* **100**, 14984–14990 (1996).
193. Pearson, R. G. Ionization potentials and electron affinities in aqueous solution. *Journal of the American Chemical Society* **108**, 6109–6114 (1986).
194. Marcus, Y. Thermodynamics of solvation of ions. Part 5.—Gibbs free energy of hydration at 298.15 K. *Journal of the Chemical Society, Faraday Transactions* **87**, 2995–2999 (1991).

195. Florián, J. & Warshel, A. Langevin dipoles model for ab initio calculations of chemical processes in solution: Parametrization and application to hydration free energies of neutral and ionic solutes and conformational analysis in aqueous solution. *The Journal of Physical Chemistry B* **101**, 5583–5595 (1997).
196. Pliego, J. R. & Riveros, J. M. New values for the absolute solvation free energy of univalent ions in aqueous solution. *Chemical Physics Letters* **332**, 597–602 (2000).
197. Kelly, C. P., Cramer, C. J. & Truhlar, D. G. SM6: a density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters. *Journal of Chemical Theory and Computation* **1**, 1133–1152 (2005).
198. Friedman, H. & Krishnan, C. in *Aqueous Solutions of Simple Electrolytes* 1–118 (Springer, 1973).
199. Donald, W. A. & Williams, E. R. Gas-phase electrochemistry: Measuring absolute potentials and investigating ion and electron hydration. *Pure and Applied Chemistry* **83**, 2129–2151 (2011).
200. Alongi, K. S. & Shields, G. C. Theoretical calculations of acid dissociation constants: a review article. *Annual reports in computational chemistry* **6**, 113–138 (2010).
201. Tissandier, M. D. *et al.* The proton's absolute aqueous enthalpy and Gibbs free energy of solvation from cluster-ion solvation data. *The Journal of Physical Chemistry A* **102**, 7787–7794 (1998).
202. Kastenholz, M. A. & Hünenberger, P. H. Computation of methodology-independent ionic solvation free energies from molecular simulations. I. The electrostatic potential in molecular liquids. *J Chem Phys* **124**, 124106 (2006).
203. Kastenholz, M. A. & Hünenberger, P. H. Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. *J Chem Phys* **124**, 224501 (2006).
204. Reif, M. M. & Hünenberger, P. H. Computation of methodology-independent single-ion solvation properties from molecular simulations. III. Correction terms for the solvation free energies, enthalpies, entropies, heat capacities, volumes, compressibilities, and expansivities of solvated ions. *J Chem Phys* **134**, 144103 (2011).
205. Lu, X. & Cui, Q. Charging free energy calculations using the Generalized Solvent Boundary Potential (GSBP) and periodic boundary condition: a comparative analysis using ion solvation and oxidation free energy in proteins. *J Phys Chem B* **117**, 2005–18 (2013).
206. Hub, J. S., Winkler, F. K., Merrick, M. & de Groot, B. L. Potentials of mean force and permeabilities for carbon dioxide, ammonia, and water flux across a Rhesus protein channel and lipid membranes. *J Am Chem Soc* **132**, 13251–63 (2010).

207. Zhu, F. & Hummer, G. Theory and simulation of ion conduction in the pentameric GLIC channel. *J Chem Theory Comput* **8**, 3759–3768 (2012).
208. Beckstein, O. & Sansom, M. S. P. The influence of geometry, surface character, and flexibility on the permeation of ions and water through biological pores. *Phys Biol* **1**, 42–52 (2004).
209. Beckstein, O. & Sansom, M. S. A hydrophobic gate in an ion channel: the closed state of the nicotinic acetylcholine receptor. *Physical biology* **3**, 147 (2006).
210. Walz, T, Smith, B. L., Zeidel, M. L., Engel, A & Agre, P. Biologically active two-dimensional crystals of aquaporin CHIP. *J Biol Chem* **269**, 1583–6 (1994).
211. Hub, J. S. & de Groot, B. L. Does CO₂ permeate through aquaporin-1? *Biophys J* **91**, 842–8 (2006).
212. Bashford, D & Karplus, M. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **29**, 10219–25 (1990).
213. Alexov, E. *et al.* Progress in the prediction of pK_a values in proteins. *Proteins* **79**, 3260–75 (2011).
214. Knecht, V. & Klasczyk, B. Specific binding of chloride ions to lipid vesicles and implications at molecular scale. *Biophys J* **104**, 818–24 (2013).
215. Fitch, C. A. *et al.* Experimental pK(a) values of buried residues: analysis with continuum methods and role of water penetration. *Biophys J* **82**, 3289–304 (2002).
216. Li, G.-S., Maigret, B., Rinaldi, D. & Ruiz-López, M. F. Influence of environment on proton-transfer mechanisms in model triads from theoretical calculations. *Journal of computational chemistry* **19**, 1675–1688 (1998).
217. McNaught, A. D. & Wilkinson, A. *IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book")* ISBN 0-9678550-9-8 (Blackwell Scientific Publications, Oxford, 1997).
218. Moser, A., Range, K. & York, D. M. Accurate proton affinity and gas-phase basicity values for molecules important in biocatalysis. *J Phys Chem B* **114**, 13911–21 (2010).
219. Compton, E. L. R. & Mindell, J. A. Bacterial Ion Channels. *EcoSal Plus* **4**. doi:10.1128/ecosalplus.3.3.2 (2010).
220. Csanády, L. & Mindell, J. A. The twain shall meet: channels, transporters and things between. Meeting on Membrane Transport in Flux: the Ambiguous Interface Between Channels and Pumps. *EMBO Rep* **9**, 960–5 (2008).
221. Mullins, L. Ion selectivity of carriers and channels. *Biophysical journal* **15**, 921 (1975).
222. Kaczmarek, L. K. & Levitan, I. B. *Neuromodulation: the biochemical control of neuronal excitability* (Oxford University Press, USA, 1987).
223. Cole, J. & Brown, C. Nitrite reduction to ammonia by fermentative bacteria: a short circuit in the biological nitrogen cycle. *FEMS Microbiol. Lett* **7**, 65–72 (1980).

224. Lv, X., Liu, H., Ke, M. & Gong, H. Exploring the pH-dependent substrate transport mechanism of FocA using molecular dynamics simulation. *Biophys J* **105**, 2714–23 (2013).
225. Feng, Z., Hou, T. & Li, Y. Concerted movement in pH-dependent gating of FocA from molecular dynamics simulations. *J Chem Inf Model* **52**, 2119–31 (2012).
226. Spassov, V. Z. & Yan, L. A fast and accurate computational approach to protein ionization. *Protein Sci* **17**, 1955–70 (2008).
227. Gordon, J. C. *et al.* H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **33**, W368–71 (2005).
228. Fairman, W. A., Vandenberg, R. J., Arriza, J. L., Kavanaugh, M. P. & Amara, S. G. An excitatory amino-acid transporter with properties of a ligand-gated chloride channel. *Nature* **375**, 599–603 (1995).
229. Tzingounis, A. V. & Wadiche, J. I. Glutamate transporters: confining runaway excitation by shaping synaptic transmission. *Nat Rev Neurosci* **8**, 935–47 (2007).
230. Holley, D. C. & Kavanaugh, M. P. Interactions of alkali cations with glutamate transporters. *Philos Trans R Soc Lond B Biol Sci* **364**, 155–61 (2009).
231. Artigas, P. & Gadsby, D. C. Na⁺/K⁺-pump ligands modulate gating of palytoxin-induced ion channels. *Proc Natl Acad Sci U S A* **100**, 501–5 (2003).
232. Gadsby, D. C., Takeuchi, A., Artigas, P. & Reyes, N. Review. Peering into an ATPase ion pump with single-channel recordings. *Philos Trans R Soc Lond B Biol Sci* **364**, 229–38 (2009).
233. Lísal, J. & Maduke, M. Review. Proton-coupled gating in chloride channels. *Philos Trans R Soc Lond B Biol Sci* **364**, 181–7 (2009).
234. Accardi, A. & Picollo, A. CLC channels and transporters: proteins with borderline personalities. *Biochim Biophys Acta* **1798**, 1457–64 (2010).
235. Jayaram, H., Accardi, A., Wu, F., Williams, C. & Miller, C. Ion permeation through a Cl[−]-selective channel designed from a CLC Cl[−]/H⁺ exchanger. *Proc Natl Acad Sci U S A* **105**, 11194–9 (2008).
236. Stockbridge, R. B. *et al.* Crystal structures of a double-barrelled fluoride ion channel. *Nature* (2015).
237. Zifarelli, G. & Pusch, M. The role of protons in fast and slow gating of the Torpedo chloride channel ClC-0. *Eur Biophys J* **39**, 869–75 (2010).
238. Feng, L., Campbell, E. B. & MacKinnon, R. Molecular mechanism of proton transport in CLC Cl[−]/H⁺ exchange transporters. *Proc Natl Acad Sci U S A* **109**, 11699–704 (2012).
239. Ginsburg, H. Abundant proton pumping in *Plasmodium falciparum*, but why? *Trends Parasitol* **18**, 483–6 (2002).
240. Jahn, R., Lang, T. & Südhof, T. C. Membrane fusion. *Cell* **112**, 519–33 (2003).
241. Harrison, S. C. Viral membrane fusion. *Nat Struct Mol Biol* **15**, 690–8 (2008).

- 242. Gapsys, V., Michielssens, S., Seeliger, D. & de Groot, B. L. pmx: Automated protein structure and topology generation for alchemical perturbations. *J Comput Chem* **36**, 348–54 (2015).
- 243. Gapsys, V., Seeliger, D. & de Groot, B. L. New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy Calculations. *J Chem Theory Comput* **8**, 2373–82 (2012).