TECHNISCHE UNIVERSITÄT MÜNCHEN

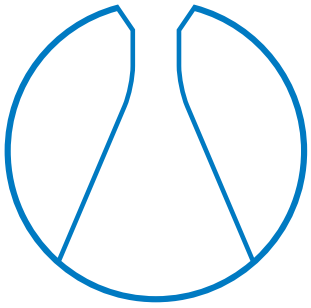
Fakultät für Chemie

Lehrstuhl für Computergestützte Biokatalyse

D I S S E R T A T I O N

Computational Studies of Membrane-bound Proteins
Na^+/H^+-Antiporter and Respiratory Complex I

Judith Warnau



Garching b. München, 2019

TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Chemie

Lehrstuhl für Computergestützte Biokatalyse

Computational Studies of Membrane-bound Proteins Na^+/H^+-Antiporter and Respiratory Complex I

Judith Warnau

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender : Prof. Dr. Michael Groll

Prüfer der Dissertation:

1. Prof. Dr. Ville R. I. Kaila
2. Prof. Dr. Gerhard Hummer

Die Dissertation wurde am 08.04.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 25.06.2019 angenommen.

Contents

List of Abbreviations

1D	1-dimensional
3D	3-dimensional
ABC	ATP binding cassette
ADP	adenosine diphosphate
AH	amphiphilic helix
ATP	adenosine triphosphate
b	backward
BAR	Bennett acceptance ratio
BPTI	*bovine* pancreatic trypsin inhibitor
ccc	cross-correlation coefficient
COM	center of mass
COSMO	conductor-like screening model
CPA	cation/proton antiporter
cryo-EM	cryo-electron microscopy
cyc *c*	cytochrom *c*
D	diffusion coefficient
DFT	density-functional theory
DNA	deoxyribonucleic acid
EcNhaA	NHA from *Escherichia coli*
EM	electron microscopy
EPR	electron paramagnetic resonance
ETC	electron transport chain
EXP	exponential averaging
f	forward
FEP	free energy perturbation
FeS	iron-sulfur (clusters)
FMN	flavin mononucleotide
GGA	generalized-gradient approximation
IN	inward-open conformational state
In	inward-open conformational state
LDA	local density approximation
LYP	Lee, Yang, and Parr

mBAR	multistate Bennett acceptance ratio
MD	molecular dynamics
MDFF	molecular dynamics flexible fitting
MjNhaP1	NHA from *Methanocaldococcus jannaschii*
MLE	maximum likelihood estimate
MM	outer MM-region in QM/MM simulation
MtrA-H	methyl-transferase subunits A-H
NADH	nicotinamide adenine dinucleotide
NHA	Na^+/H^+ antiporter
NHE	Na^+/H^+ exchanger
OUT	outward-open conformational state
Out	outward-open conformational state
PaNhaP	NHA from *Pyrococcus abyssi*
PDB	protein data base
PME	particle-mesh Ewald
PMF	potential of mean force
pmf	proton motive force
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholin
POPE	1-palmitoyl-2-oleoylphosphatidylethanol-amine
PSI/TM	position specific iterative/transmembrane
Q	ubiquinone/ol
Q_H	high redox-potential side of ubiquinone
Q_L	low redox-potential side of ubiquinone
Q_{ox}	ubiquinone
QH_2	ubiquinol
QM	inner QM-region in QM/MM simulation
QM/MM	quantum mechanics/molecular mechanics
RMSD	root-mean-square deviation
ROS	radical oxygen species
SEM	standard error of the mean
SMD	steered molecular dynamics
Sol	in bulk water
SQ_{Nf}	fast-relaxing signal of a semiquinone radical
SQ_{Ns}	slow-relaxing signal of a semiquinone radical
SQ_{Nvs}	very-slow relaxing signal of a semiquinone radical
std	standard derivation
TMH	transmembrane helix
TMP	transmembrane protein

TtNapA	NHA from *Thermus thermophilus*
vdW	van der Waals
VMD	visualization molecular dynamics
WHAM	weighted histogram analysis method
WT	wildtype

Zusammenfassung

In biologischen Systemen wird ein aktiver Stofftransport über die Zellmembran von Transmembranproteinen bewerkstelligt. In dieser Dissertation werden zwei Transmembranproteine, Komplex I der Atmungskette von *Thermus thermophilus* und der Na^+/H^+-Antiporter von *Methanocaldococcus jannaschii* (MjNhaP1) untersucht, die Ionen aktiv über die Zellmembran transportieren. In beiden Systemen führen Änderungen des Ladungszustandes, durch Ionenbindung bzw. Protonierung in MjNhaP1 oder durch Reduktion des Ubichinons in Komplex I, gekoppelt mit konformationellen Änderungen des Proteins zu einem aktiven Ionentransport.

Der Na^+/H^+-Antiporter MjNhaP1 ist ein sekundär aktives Transportprotein, der durch die Kontrolle des pH-Wertes und der Na^+-Konzentration zur Selbstregulation der Zelle beiträgt. Eine Fehlfunktion von menschlichen Na^+/H^+-Austauschern (NHE) wird mit der Entwicklung verschiedener Krankheiten in Verbindung gebracht, wie Bluthochdruck, Herzinsuffizienz, Autismus und Epilepsie [1]. MjNhaP1 ist ein funktionelles Homolog des menschlichen NHE1, welcher ein wichtiges Wirkstoffziel ist [2]. Na^+/H^+-Antiporter tauschen Na^+- gegen H^+-Ionen über die Zellmembran, nach dem „alternate access mechanism“ [3], [4] aus. Der molekulare Kopplungsmechanismus zwischen Ionenbindung und konformationellen Änderungen des gesamten Proteins, der zu einem aktiven Ionentransport führt, ist noch nicht vollständig aufgelöst worden. Im ersten Teil der Arbeit wird die Ionenbindung und Ionenselektivität in MjNhaP1 mit verschiedenen Methoden untersucht: Sequenziellem und strukturellem Vergleich mit homologen Antiporter-Proteinen, Quantenchemischen Berechnungen, Molekulardynamik (MD) und Quantenmechanischen/Molekularmechanischen Simulationen und Berechnungen der freien Energie. Wir haben gezeigt, dass MjNhaP1, insbesondere in dem nach innen offenen konformationellen Zustand, eine hohe Selektivität für Na^+ aufweist. Wir konnten außerdem die Beiträge von einzelnen Aminosäureresten zu der Selektivität der Ionenbindung bestimmen. Damit eröffnen unsere Ergebnisse ein Verständnis der Ionenselektivität von Kation/Proton-Austauschern auf molekularer Ebene [5], [6].

Komplex I der Atmungskette ist eine durch Redoxreaktionen angetriebene Protonenpumpe, die zum Aufbau des Protonengradienten über die Zellmembran beiträgt. Komplex I ist einer der kompliziertesten und am wenigsten verstandenen biologischen Energieumwandlungsmaschinen. Der molekulare Mechanismus, wie die Redoxreaktionen an der Bindungsstelle für Ubichinon die Protonenpumpen katalysiert, die *ca.* 200 Å voneinander entfernt sind, muss noch weiter erforscht werden. Im zweiten Teil dieser Arbeit wird das dynamische Verhalten von Ubichinon in dem Substratbindungstunnel von Komplex I unter der

Verwendung von MD-Simulationen und der Berechnung von freien Energieprofilen und Diffusionsprofilen untersucht. Dadurch konnten wir eine zweite Bindungsstelle für Ubichinon strukturell charakterisieren, deren mechanistische Bedeutung für den Kopplungsmechanismus zu den Protonenpumpen noch geklärt werden muss [7].

Summary

In biological systems, transmembrane proteins are responsible for an active transport across cell membranes. In this dissertation, two transmembrane proteins were investigated that catalyze an active ion transport across the cell membrane, namely the respiratory Complex I from *Thermus thermophilus* and the Na^+/H^+ antiporter from *Methanocaldococcus jannaschii* (MjNhaP1). Common in both systems is a coupling between changes in the charge state, through either ion binding/protonation in MjNhaP1 or reduction of ubiquinone in complex I, and conformational changes of the protein, which lead to active ion transport.

The Na^+/H^+ antiporter MjNhaP1 is a secondary active transporter. Its biological function is to maintain cell homeostasis, by controlling the pH and Na^+ concentration in the cell. Dysfunction of the human Na^+/H^+ exchanger (NHE) is linked to the pathogenesis of several diseases including hypertension, heart failure, autism, and epilepsy [1]. MjNhaP1 is a functional homolog of the human NHE1, which is an important drug target [2]. Na^+/H^+ antiporter exchanges Na^+ against H^+ by employing an alternate access state mechanism [3], [4]. However, the exact molecular coupling mechanism between ion binding and conformational changes remain elusive. In the first part of this thesis, I studied the ion binding and selectivity of MjNhaP1 by applying multiple methods: sequence and structural comparison with antiporter homologs, quantum chemical cluster calculation, molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) simulation and free energy estimations. We showed that the MjNhaP1 is highly selective towards Na^+, particularly in the inward-open conformational state. We determined the contribution of individual amino acid residues to the selectivity of the ion binding. Our findings thus established the molecular basis for understanding ion selectivity in cation/proton exchangers [5], [6].

Respiratory complex I is one the most intricate and least understood types of biological energy conversion machinery. The molecular mechanism of how the redox reaction at the ubiquinone binding site catalyzes the proton-pump up to *ca.* 200 Å away, remains elusive. In the second part of this work, the dynamical behavior of ubiquinone and ubiquinol in the substrate binding tunnel of complex I was studied by performing MD simulations and calculating free energy and diffusion profiles. We were able to characterize a second ubiquinone binding site in the ubiquinone binding tunnel, which could have a mechanistic impact on the coupling mechanism to activate the proton pumps [7].

Chapter 1

Introduction

Central to all life are biological membranes [8]. The biological membrane consists of a lipid bilayer, which defines the boundaries of all living cells. As a hydrophobic barrier through which most molecules cannot pass, the lipid bilayer prevents substrates generated inside the cell from leaking out and unwanted molecules from diffusing in [9]. Lipid membranes contain abundant transmembrane proteins, which perform specific substrate exchanges with the environment. Such transport systems give the biological membrane the property of selective permeability [9] and serve therefore several functions crucial for life, such as information transduction and energy storage in form of electrochemical gradients. For example a proton electrochemical gradient across the membrane established by the electron transport chain (ETC) is used for adenosine triphosphate (ATP) synthesis.

In all living organisms, membranes play a key role in energy conversion as the location of ETCs, where an electron is transferred to acceptors of an increasingly oxidizing nature. In the case of aerobic cellular respiration (section 1.1), food molecules provide high-energy electrons which, move through the ETC, eventually reducing dioxygen to water and using the free energy released by this process to generate an ion gradient. This gradient is used for multiple processes across the cell, including membrane transport (section 1.2). Substrate transport can be divided into passive and active transport across the membrane (section 1.3). The common mechanism of secondary active transporters is the so-called alternating access state mechanism [4], [3] (section 1.4), which Na^+/H^+ antiporters (NHAs) (section 1.5) also follows. In this thesis two transmembrane proteins are studied, NHA from *M. jannaschii* (MjNhaP1) (section 1.6) and respiratory complex I from *T. thermophilus* (section 1.7).

1.1 Cellular Respiration

In cellular respiration, chemical energy from nutrients, *e.g.*, sugars, fatty acids, or amino acids is converted using O_2 as oxidizing agent, into energy stored as ATP via a branch of metabolic reactions and processes in the cell. A series of exothermic redox reactions take place in many biochemical steps. Respiration thus releases chemical energy to fuel cellular processes. The energy stored as ATP is made available throughout the cell for energy-requiring processes like biosynthesis, cellular transport, or locomotion. Respiration is realized in two ways: under aerobic conditions with O_2, and under anaerobic conditions without O_2.

The focus of this thesis is on aerobic respiration. In the first part of cellular respiration, a glucose molecule is gradually oxidized and broken down by multiple biochemical reactions into CO_2 and nicotinamide adenine dinucleotide (NADH), a cellular electron carrier.

NADH subsequently provides electrons to the ETC in the inner mitochondrial membrane in eukaryotes. In bacteria, the ETC is located in the cell membrane.

Oxidative phosphorylation comprises the ETC shown in Fig. 1.1, and leads to transfer of electrons from an electron donor, NADH, in multiple redox reactions to an electron acceptor, O_2, which is finally reduced to water. A small percentage of the electrons are not properly transferred through the ETC, but leak directly to oxygen resulting in formation of radical oxygen species (ROS). ROS formation contributes to oxidative stress and has been implicated in several diseases and aging [10–13]. Coupled to the redox processes, protons are pumped from the mitochondrial matrix into the intermembrane space. The proton gradient thus generated across the membrane (proton motive force; pmf) drives the F_oF_1-ATPase to produce the major part of ATP in cellular respiration [14]. In the chemiosmotic theory [15], Peter Mitchell proposed that this coupling element between the redox reactions in the electron transfer chain and the ATP-synthase is the electrochemical-proton gradient across the membrane, providing the energy source for multiple cellular processes.

The electron transfer chain starts with the oxidation of NADH by complex I (NADH: ubiquinone oxidoreductase), which transfers the electrons to ubiquinone. Its reduction to ubiquinol is coupled to active proton translocation across the membrane. Complex II reduces ubiquinone to ubiquinol by oxidizing succinate to fumarate without contributing to the proton gradient. Ubiquinol passes electrons to the cytochrome bc_1 (complex III). Complex III catalyzes the so-called Q-cycle, in which two electrons are passed to cytochrome c, a water-soluble electron carrier located in the intermembrane space. Complex III takes up two protons from the matrix and releases four protons into the intermembrane space. Cytochrome c is oxidized by complex IV, which employs the electrons together with eight protons taken up from the matrix, to reduce oxygen to water and to pump four protons across the membrane. Complex IV contributes four pumped protons to the pmf. The proton gradient is finally used to generate ATP. Protons flow back through the membrane-embedded F_o domain of ATP-synthase, driving the rotation of the central stalk within the matrix domain where mechanical rotary energy is converted to chemical energy in the form of ATP.

In addition to ATP synthesis, the transmembrane gradient is also used for a number of other energetically costly processes, for example by highly specialized transporter proteins to catalyze active substrate transport across the membrane.

1.2 Membrane Transport

Small inorganic ions, such as Na^+, K^+, Ca^{2+}, Mg^{2+}, HCO_3^-, and PO_4^{3-}, play an important role in physiological processes including cell signaling, neural impulse conduction, muscle contraction, and pH and osmotic pressure balance. Maintaining cellular homeostasis requires that the concentration of ions is carefully regulated in each cellular compartment. This is achieved by active transport and exchange of ions. The plasma membrane, consisting of a lipid bilayer, works as physiological barrier to prevent passive substrate flux. For example in humans, K^+ is highly concentrated in the cytosol, whereas Na^+ and Cl^- are more abundant in the extracellular fluid. Effective and highly specific substrate transport is achieved using

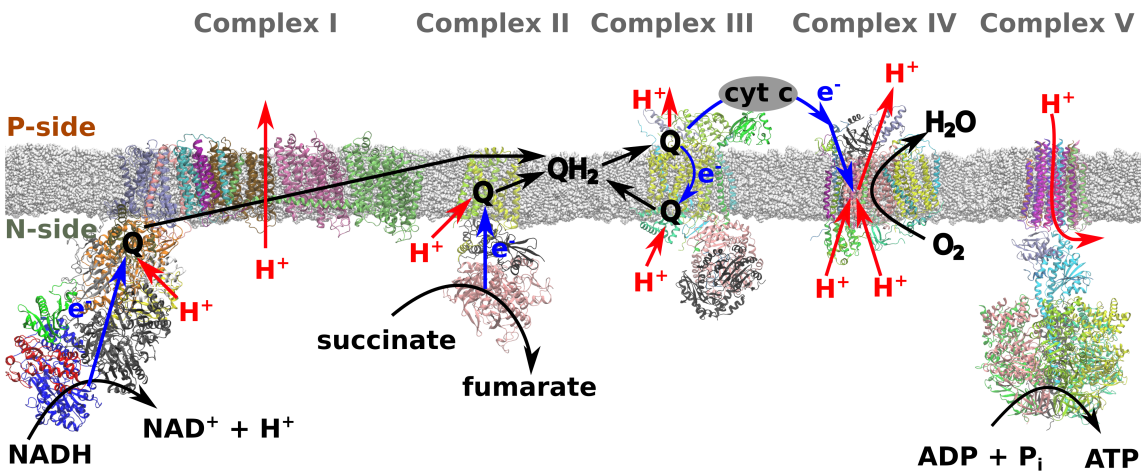


Figure 1.1: Electron transport chain in aerobic cellular respiration. Oxidized ubiquinone (Q) is reduced in complex I and complex II to ubiquinol (QH_2). In complex III, the Q-cycle donates electrons (blue) to cytochrome c (cyt c), which shuttles the electrons to complex IV, where they are used to reduce oxygen (O_2) to water. A proton gradient is generated by pumping protons across the membrane from the N- to the P-side by complexes I, III, and IV. ATP synthesis is powered by down-gradient flow of protons in complex V. Structures of transmembrane complexes are taken from protein data base (PDB) ID: 4HEA [16], 1ZOY [17], 1BCC [18], 1V54 [19], and 2XOK [20].

transmembrane proteins such channels and active transporters. Small inorganic ions maybe transported alone or co-transported with larger organic substrates like glucose or amino acids. Understanding and controlling ion transport is vital among many processes and also they are linked to the parthenogenesis of several diseases such as diabetes, epilepsy, cystic fibrosis, and osteoporosis [21].

In theoretical biophysics, we aim to achieve a molecular understanding of the function of the protein complex [22]. A central point of ion transporters is an ion-selective site, which is often composed of highly conserved residues. These ion binding sites usually have a high selectivity and are effective transporters at the same time. The ion coordination and selectivity mechanisms rely on coupling processes in ion binding and protein dynamics, which can be explored using compositional methods.

1.3 Passive and Active Transport Across the Membrane

Molecules can cross the lipid membrane by diffusion, passive transport or active transport. Only hydrophobic compounds are free to pass the lipid bilayer by simple diffusion, whereas hydrophilic, polar and charged molecules have to use highly specialized transporter systems to overcome the hydrophobic lipid barrier (Fig. 1.2). Passive transport is facilitated by channels or carrier proteins. Ion channels are highly selective and achieve a very high transport rate near the diffusion limit (*e.g.* K^+ channel KcsA [23], [24]). In passive transport, the transport direction always follows the electrochemical gradient.

Active transport processes, by contrast, transport substrates against a concentration gradient. The energy source for active transport is obtained either from the hydrolysis of ATP in so-called primary active transporters, or by coupling to a favorable transport

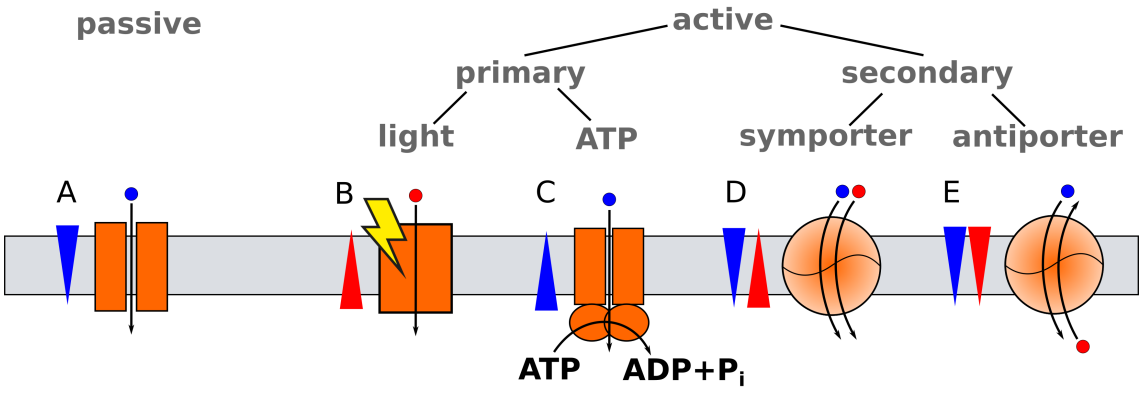


Figure 1.2: Active and passive transport across the membrane. A) Passive transporter proteins, such as ion channels, accomplish substrate transport with a concentration gradient. Primary active transport against a concentration gradient is driven by (B) light or (C) ATP hydrolysis. Secondary active transporters are divided into (D) symporter, in which both substrates are transported in the same direction, and (E) antiporter or exchangers, where two substrates are transported in opposite directions.

process in order to use the energy of an electrochemical gradient across the membrane in secondary active transporters. The central role of primary active transporters is to build up a specific ion gradient, which is then used by secondary active transporters [25].

ATP binding cassette (ABC) transporters belong to the primary active transporters. For example, ABCB10 [26] actively transports 6-20 amino acid-long peptides, using energy from ATP hydrolysis. Active transport can also be driven by light energy, as in bacteriorhodopsin.

Secondary active transporters combine passive and active transport. The active substrate transport is facilitated by a passive co-transport of another substrate. Depending on the relative transport direction, the secondary active transporters can be subdivided into symporter, transporting both substrates in same direction, or antiporter, transporting them in opposite directions (Fig. 1.2). The Na^+/H^+-antiporters (NHA) including NhaP1 from *M. jannaschii*, studied in this work, are examples of secondary active transporters.

1.4 Alternating Access State Mechanism

The underlying molecular-transport mechanism of secondary active transporters is the alternating access state mechanism. This mechanism was first proposed primarily by Mitchell in 1957 [3], and formulated in its final version by Jardetzky in 1966 [4]. A growing number of available structures of transporter proteins in the last years has improved our understanding of the mechanism. The alternating access state mechanism explains how the substrate binding site of the transporter protein, located in the middle of the membrane, is accessible to only one side of the membrane at a given time. Conformational changes of the protein are needed to make the binding site accessible to the other side of the membrane. The key aspect of the mechanism is the tight coupling between substrate binding and conformational transition, which occur in a sequential order of a series of fully reversible conformational changes, which ensures an effective transport cycle [27].

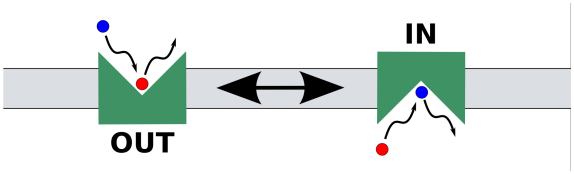


Figure 1.3: Schematic representation of the alternate access mechanism for an antiporter. In the inward-open conformational state (IN), the binding site is accessible from the inside of the cell and in the outward-open conformational state (OUT), the binding site is accessible from the outside.

1.5 Structure and Function of Na^+/H^+ Antiporters

Na^+/H^+ antiporters belong to the family of cation/proton antiporter (CPA), which are present in all kingdoms of life. Their biological function is to maintain cell homeostasis, by controlling pH and Na^+ concentration in the cell. The CPA super-family can be divided into CPA1 and CPA2 subfamilies, depending on the evolutionary origin [1]. CPAs can be further distinguished as electroneutral or electrogenic, according to their transport stoichometry. Electroneutral antiporters transport one H^+ against one Na^+, while electrogenic antiporters exchange two H^+ for one Na^+ ion. The human NHE are linked to the pathogenesis of several diseases including hypertension, heart failure, autism and epilepsy, and they are well-established drug targets [1].

Na^+/H^+ antiporters function by the alternate access mechanism. Depending on the dominant gradient employed, the NHAs transport their substrate in either direction. In *Escherichia coli*, the NhaA uses a H^+ gradient to actively pump Na^+ out of the cell, whereas the human NHE uses the Na^+ gradient between extra- and intracelullar fluids to actively alkalize the inside of the cell. Similarly, the archaeal organism *M. jannaschii* uses the Na^+ gradient between its saline environment and the cytosol with low Na^+ concentration to pump H^+ out of the cell, protecting the cell from acidification. Na^+/H^+ antiporters are one of the fastest known transporters, reaching transport rates of up to 89000 ions min^{-1} (*ca.* 1500 ions sec^{-1}) [28]. This rate implies that an ion requires on average *ca.* 700 μs to cross the membrane.

The Na^+/H^+ antiporter from *E. coli* (EcNhaA) is the best studied antiporter. The first high-resolution X-ray structure of an NHA was EcNhaA, resolved 3.45 Å at pH 4 by Hunte *et al.* in 2005 [29]. The monomeric structure with 12 transmembrane helices was captured in an inactive acidic-pH-locked conformational state [29]. EcNhaA belongs to the CPA2 branch and is electrogenic, transporting two H^+ along the gradient into the cell and one Na^+ actively out of the cell. This exchange activity enables the organism to resist high salinity and alkaline stress [29], [30]. In EcNhaA, the subsequent aspartate residues Asp-163 and Asp-164 form the conserved DD motif [31], [29]. Computational studies, showed the mechanistic relevance of the protonation state of these two aspartic acids [30]. Na^+ binding takes place when Asp-164 is deprotonated, while Asp-163 works as a conformational switch, controlling the connectivity of the binding site to one or the other side of the membrane. The binding site is water accessible from the cytosol when Asp-163 is protonated and accessible from the periplasm when Asp-163 is deprotonated [30]. The ion selectivity of EcNhaA was estimated with free energy calculations; Na^+ binding in the binding site of EcNhaA is

3.4 kcal mol^{-1} more favorable than K^+ binding. This corroborates a clear selectivity of EcNhaA for Na^+ over K^+ transport [30].

Electrophysiology measurements [32] demonstrated that the pH dependence of EcNhaA is central to the substrate transport mechanism, and excluded that pH regulation could be established by allosteric effects. This implies that substrate transport is intrinsically downregulated in the acidic and alkaline pH range. Na^+ and H^+ compete for the same binding site, and therefore the protein becomes inactive under acidic conditions (< pH 5) when both aspartic acids in the binding site are protonated and no Na^+ binding or transport can occur. Under alkaline conditions (> pH 8) EcNhaA loses its function due to substrate depletion [32].

A second Na^+/H^+ antiporter structure, from *T. thermophilus* NapA [33] at pH 7.8, was obtained at 3 Å resolution. Under these mild conditions, a dimeric structure in an outward-open conformational state was obtained [33]. TtNapA is an electrogenic antiporter and belongs to the CPA2 branch. Similar to EcNhaA, it transports two H^+ against one Na^+ ion. In comparison to the inward-open structure of EcNhaA, the six-helix bundle domain is translated as a rigid body by 10 Å towards the outside, and rotated by 20° with respect to the dimerization interface. The suggested large conformational rearrangements for the conformational transformation from the inward- to outward-open state is known as the two-domain rocking bundle model [33].

The structures of electroneutral CPA1-type Na^+/H^+ antiporters NhaP from *Pyrococcus abyssi* (PaNhaP) and NhaP1 from *M. jannaschii* (MjNhaP1), were recently resolved by Paulino *et al.* [34] and Wöhlert *et al.* [35]. Two dimeric structures of PaNhaP in an inward-open state were resolved at pH 4 and pH 8. The crystal structure of PaNhaP at pH 8 was resolved at 3.2 Å with a bound Tl^+ ion in the binding site, the first time it was possible to resolve a Na^+/H^+ antiporter with an ion bound in the intact binding site. The bound Tl^+ ion is surrounded by three carboxylate groups: Asp-159 (part of the characteristic ND motif of CPA1 antiporter [34]), Asp-130 (corresponding to Asp-133 in EcNhaA) and an additional Glu-73 that is not conserved. PaNhaP is maximally active at pH 5. Extrapolating the transport rate to 100°C gives an estimated turnover of 5000 ions sec^{-1}. PaNhaP shows cooperative ion binding of the two chains at pH 6 but not at pH 5. Driven by the Na^+ gradient, PaNhaP is likely to actively pump H^+ out of the cell, which is supported by the high activity of PaNhaP under acidic conditions. PaNhaP is, like MjNhaP1, a functional homolog of human Na^+/H^+ exchanger NHE1, which is an important drug target [2].

1.6 Structure and Function of MjNhaP1

MjNhaP1 has 20% sequence identity of functionally important regions with the mammalian NHEs [36], [37]. Both MjNhaP1 and the mammalian NHEs are likely to utilize the Na^+ gradient, to actively expel H^+ in order to maintain the intracellular pH.

The crystal structure of MjNhaP1 was determined at 3.5 Å resolution at pH 8 in a homodimeric state. Each monomer compromises 13 transmembrane helix (TMH) arranged into two major structural elements: a 6-helix bundle and an interface domain consisting of 7 TMHs.

A Na^+ ion bound in the binding site is not resolved in the X-ray structure of MjNhaP1 [34], but given the structural homology to PaNhaP, it is reasonable to assume that the binding site of MjNhaP1 is in a comparable position. Paulino *et al.* [34] suggested the following residues to be involved in ion coordination: Ser-157, Asn-160, and Asp-161 in TMH-6, Asp-132, and the backbone carbonyl of Thr-131 in the unwound stretch of the TMH-5 (the middle part of the disconnected TMH-5, which is not ordered as helix). Asn-160 and Asp-161 are form the characteristic ND motif of CPA1 antiporter [34]. There is no equivalent ion-binding residue in the binding site of MjNhaP1 for the ion-coordinating Glu-73 of PaNhaP. Based on the comparison with the Tl^+ bound structure of PaNhaP, it is unlikely that Asn-160 directly coordinates the ion. However, based on mutational experiments [34], Asn-160 seems to be important for the antiporter function. A mutation of Asn-160 to alanine (N160A) renders MjNhaP1 inactive, while mutation to aspartic acid (N160D) prodruces a mutant with a slower transport rate, but does not transform MjNhaP1 into an electrogenic transporter [34].

MjNhaP1 has an even higher ion transport rate than PaNhaP, because of this the turnover rate of over 5000 ions sec^{-1} at physiological temperatures (85°C) is extrapolated. Unlike its homologue PaNhaP, there is no evidence for a cooperative ion binding of the two monomers in MjNhaP1.

A structure of MjNhaP1 in an alternative state was resolved with electron cryo-crystallography at pH 4 without NaCl. A 3D map was obtained with an in-plane resolution of 6 Å [34]. In this thesis, computer simulations are used to study the ion binding and selectivity in the inward- and outward-open state of MjNhaP1.

1.7 Structure and Function of Respiratory Complex I

Complex I has a central role in cellular respiration and energy production. It functions as the electron entry point into the ETC by the oxidation of NADH, to reduce ubiquinone to ubiquinol. This redox reaction is coupled to proton translocation across the membrane, contributing four protons to the pmf [38], [39], [40].

Mutations in the central subunits of complex I give rise to several human neurodegenerative diseases [41]. Complex I is the major source of radical oxygen species in mitochondria, which lead to mitochondrial deoxyribonucleic acid (DNA)-damage and are implicated in Parkinson's disease and aging [11], [12], [13].

Complex I is one of the largest known membrane proteins. In the minimal prokaryotic version it comprises 14 core subunits ($\sim$ 550 kDa), which are highly conserved through all organisms from bacteria to human, suggesting that the overall mechanism is likely to be conserved. Mitochondrial complex I is around 1 MDa with up to 45 subunits. Around 30 accessory subunits are present around the core domains, forming a protective shell. The accessory subunits are important for complex I assembly and stabilization, and some of them have specific functions [42].

Complex I assembly takes place in multiple distinct intermediate steps, and is aided by multiple assembly factors [43]. One step requires a concerted insertion of preformed iron-sulfur (clusters) (FeS) clusters [44]. Pathogenic mutations were identified not only

in the central subunits, but also in the accessory subunits and assembly factors. The accessory subunit NUMM from *Yarrowia lipolytica* is important for complex I assembly and ensures the correct insertion of FeS cluster N4. Chromosomal deletion of the NUMM gene or mutations in the Zn-binding motif block a late step of complex I assembly [42]. In α-proteobacteria a homolog of the accessory subunit NUMM was previously found with a conserved Zn-binding motif of three cycteines and one histidine [45]. It is also reported that a Zn^{2+} is bound in bovine complex I structure but its functional relevance and position is unknown [46], [47]. The orthologous subunit in human is NDUFS6, where a mutation of one of the cysteines of the Zn-binding motif is fatal [48], [49]. In a mouse model, knock-down of the corresponding gene leads to complex I-linked diseases [50].

The 14 core subunits which accomplish the bioenergetic functions are assembled in an L-shaped structure. Half of the subunits are inserted into the lipid bilayer building the membrane arm, including the H^+ pumping antiporter-like subunits Nqo12, Nqo13, Nqo14, and probably Nqo7/Nqo8/Nqo10; *T. thermophilus* naming). The hydrophilic domain extends into the matrix or cytosol in mitochondrial and bacterial complex I, respectively. The end of the hydrophilic domain harbors a flavin mononucleotide (FMN) binding site, at which the oxidation of NADH takes place. The electrons released from this oxidation are transferred via FeS clusters to the ubiquinone headgroup. The ubiquinone binding site is at the interface region of the hydrophilic domain and the membrane arm, surrounded by subunits Nqo4, Nqo6, and Nqo8. The ubiquinone binding site and the FMN are connected by 9 FeS clusters, which conduct electrons from one site to the other.

Complex I is the most recenly resolved membrane protein complex of the ETC. The first full structure containing all core subunits was resolved from *T. thermophilus* at 3.3 Å resolution in 2013 [16]. Serveral structures of complexes from *E. coli* [51, 52], *Y. lipolytica* [53], [54], *Ovis aries* [55], *mouse* [56], *bos taurus* [57], [58] species, have been published since.

A ubiquinone-bound form of complex I has never been structurally resolved. Biochemical studies [59], [60] and unresolved densities in the *T. thermophilus* X-ray data set [16] suggest a binding site for ubiquinone in complex I in which the ubiquinone headgroup is located 20-30 Å above the membrane plane [61], [16]. The ubiqunone binding tunnel is a *ca.* 30 Å-long narrow chamber surrounded by subunits Nqo4 and Nqo6 of the hydrophilic domain and the transmembrane subunit Nqo8 (Fig. 1.4). The entrance to the tunnel is a triangle formed by three helices of subunit Nqo8 at the protein-membrane interface. The triangular entry is formed by two strongly tilted long transmembrane helices, TMH-1 and TMH-6, and a short amphiphilic helix (AH) AH-1 located parallel to the membrane surface (Fig. 1.4). The ubiquinone binding (Q) tunnel ends deep in the protein close to the last FeS cluster N2, where the ubiquinone headgroup binds. The long tail of 10 isoprenyl units is elongated in the tunnel towards the membrane exit. Several charged residues, mostly arginines [62] of the Nqo8 subunit line the Q tunnel. Mutation of conserved residues in Nqo8 (Leu-15, Glu-35, Arg-36, Arg-45, Asn-49, Ala-63, Glu-70, Glu-163, Glu-235, Tyr-236, Thr-324) [41] causes inhibition of complex I activity, and are implicated in mitochondrial disease (Glu-35, Arg-36, Arg-45, Asn-49, Glu-70, Ala-75, Gly-151, Glu-163, Iso-208, Arg-216, Glu-235, Tyr-

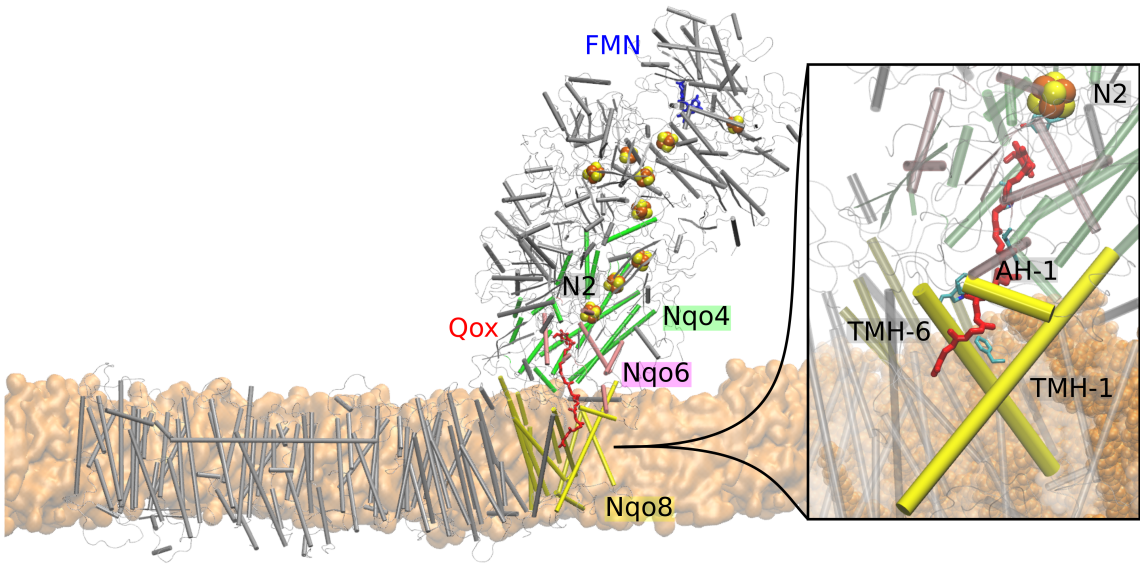


Figure 1.4: Respiratory complex I. Snapshot of MD simulation of the membrane embedded complex I structure from *T. thermophilus* (PDB ID: 4HEA). Ubiquinone is bound in the primary binding site close to FeS cluster N2. Close-up to the entrance of the Q-binding tunnel from the membrane side, framed by TMH-1, TMH-6, and the amphiphilic helix AH-1.

236, Iso-293) [63–70]. It is suggested that the ubiquinone headgroup in proximity to the N2 cluster, which would make an effective electron transfer from N2 to ubiquinone headgroup feasible, according to biological electron transfer theories [71]. The ubiquinone headgroup is within hydrogen bonding distance of Tyr-87$_{\mathrm{Nqo4}}$ and His-38$_{\mathrm{Nqo4}}$ [16], which donate two H^+ to the reduced ubiquinone [72], [73].

A key question is how the redox reaction of NADH and ubiquinone is coupled to the spatially separated H^+ translocation in the membrane domain. The distance between the ubiquinone binding site and the most distant antiporter-like subunit is around 200 Å.

A widely-held theory postulates that upon ubiquinone reduction, a signal in the form of electrostatic/conformational changes percolates through the protein toward the membrane arm to activate the H^+ pumps [74]. Many recent studies investigate this idea [16, 53, 75–79]. MD simulations studies [72] suggest that the reduction of ubiquinone is coupled to a local proton transfer reaction from the residues Tyr-87$_{\mathrm{Nqo4}}$ and the ion pair His-38$_{\mathrm{Nqo4}}$/Asp-139$_{\mathrm{Nqo4}}$ to the reduced ubiquinone headgroup. This leads to the reorientation of the ion pair, with Asp-139$_{\mathrm{Nqo4}}$ pointing towards the membrane domain.

In the standard single ubiquinone model an oxidized ubiquinone enters the Q tunnel from the membrane and moves to the primary binding site at 12 Å distance from the N2 cluster. After reduction and local protonation from the surrounding residues, Tyr-87$_{\mathrm{Nqo4}}$ and His-38$_{\mathrm{Nqo4}}$, QH_2 migrates back towards the membrane exit of the tunnel and exchanges with an oxidized ubiquinone entering the Q tunnel for the next catalytic turnover.

The redox potential of the ubiquinone ubiquinone radical couple bound in complex I is $<$ -300 mV [80], [81], [73], whereas ubiquinone in the membrane has a much higher redox-potential of +90 mV. The redox potential of NADH/NAD^+ is -320 mV [80]. The nearly equal redox potentials of NADH and ubiquinone bound in complex I, implies that the free

energy that drives the H^+ pumps in the membrane arm must come from a step after the reduction of ubiquinone to semiquinone [39]. The very low redox potential of ubiquinone bound in complex I might have some significant mechanistic implications, as outlined by DeVault 1971 [82]. In the coupling between oxidoreduction and proton translocation in the membrane arm, ubiquinone could play the role of a redox-coupled energy transducer. Following this idea, ubiquinone could exist in two states with a low (Q_L) and an high (Q_H) redox potential. The low-potential site would be the primary binding site found in the crystal structure [16]. The location of the high-potential site is unknown. There would be a change of free energy corresponding to the jump in redox potential when a ubiquinone molecule moves from the low- to the high-potential site. In order to drive endergonic proton pumping, the transition of ubiquinone from one site to the other is expected to be exergonic. Depending on the details of the mechanistic model, transitions from Q_L to Q_H and/or from Q_H to Q_L are exergonic providing one or two power strokes [39].

The existence of multiple ubiquinone binding sites in complex I is indirectly supported by experimental data [83–87]. A second ubiquinone binding site is consistent with findings by affinity labeling studies with acetogenin from Miyoshi and co-workers [88], [89]. The crystal structure of complex I from *T. thermophilus* does not seem to accommodate a second ubiquinone molecule [16]. In EPR studies, Ohnishi and coworkers [90], [59] measured two signals of semiquinone radicals: a fast-relaxing signal (SQ_{Nf}) 8-13 Å from the N2 cluster, which matches with the primary binding site of ubiquinone, and a slow-relaxing signal (SQ_{Ns}) belong to a ubiquinone bound $\geq$ 30 Å away from the N2 cluster. It is possible that both signals arise from a single ubiquinone molecule at two different binding sites in the Q tunnel, because both signals represent less than one ubiquinone per FMN [90], [59]. Recent EPR experiments [91] found a third signal due to a very-slow relaxing semiquinone (SQ_{Nvs}), at larger distances from N2 center than the slow-relaxing signal SQ_{Ns}. This signal was suggested to be due to a membrane bound ubiquinone at the membrane-protein interface [91].

Several mechanistic suggestions exist of how to incorporate two ubiquinone binding sites into one catalytic turnover of complex I. The ubiquinone could work as a piston to transfer energy to the proton pumping mechanism. Shuttling between these two sites, ubiquinone would work as a dynamic redox transducer [39], [77]. This hypothesis has some similarities to the two-stroke stabilizing-change mechanism of Brandt *et al.* [75]. This mechanistic model proposes that stabilization of negatively charged ubiquinone intermediates drives a conformational change, thereby transmitting energy to drive the H^+ pumps in the membrane arm [53].

A central point in the coupling mechanism of ubiquinone reduction and the activation of the proton pumps, is to determine the location of the putative second ubiquinone binding site, and resolve the migration profile of ubiquinone in the Q tunnel, depending on its oxidation state. Elucidating the location of the second binding site would give new perspectives that could reveal the coupling mechanism, and would advance our understanding of complex I function.

Chapter 2

Theoretical Basis of Computational Biophysics

2.1 Quantum Chemical Calculations

In quantum chemistry, we describe the properties of atoms and molecules by so-called wave functions Ψ, which satisfy the time-independent Schrödinger equation

$$\hat{H}\Psi = E\Psi. \tag{2.1}$$

Mathematically, Eqn. 2.1, can be seen as an eigenvalue equation, where the Ψ are eigenfunctions of the Hamiltonian operator $\hat{H}$ with the corresponding eigenvalues E, which represent the energies of the states.

The Born-Oppenheimer approximation separates the motion of the nuclei and electrons, so that we can reduce Eqn. 2.1 to an electronic part, which defines chemical properties of the system. The electronic Schrödinger equation,

$$\hat{H}_{\mathrm{el}}\Psi_{\mathrm{el}} = E_{\mathrm{el}}\Psi_{\mathrm{el}} \tag{2.2}$$

with the electronic Hamiltonian operator[1]

$$\hat{H}_{\mathrm{el}} = \hat{T}_{\mathrm{el}} + \hat{V}_{\mathrm{Ke}} + \hat{V}_{\mathrm{ee}} = -\frac{1}{2}\sum_{i=1}^{n}\nabla_i^2 - \sum_{i=1}^{n}\sum_{I=1}^{N}\frac{Z_I}{r_{iI}} + \sum_{i<j}^{n}\frac{1}{r_{ij}}, \tag{2.3}$$

where $\hat{T}_{\mathrm{el}}$ and $\hat{V}_{\mathrm{ee}}$ are the kinetic and potential energy operators of the electrons. $\hat{V}_{\mathrm{Ke}}$ is the operator describing the attractive force between each electron and the nucleus, n and N are the numbers of electrons and nuclei, respectively. Z_I is the charge of the nucleus I, and r_{iI} is the distance between the nucleus I and electron i, r_{ij} is the distance between electrons i and j. Based on the Born-Oppenheimer approximation, the positions of the nuclei only enters the Hamiltonian as parameters.

Due to the complex electron-electron interaction $\hat{V}_{\mathrm{ee}}$, it is not possible to solve the Schrödinger equation analytically for more than two particles, *i.e.*, an hydrogen atom. For that reason one has to apply approximations, which covers at the same time all physical laws of the fermionic electrons and satisfy the Pauli principle.

To fulfill these criteria for a description of a wave function, the Slater determinant is

[1] All operators are given in atomic units.

introduced,

$$\Psi^{\mathrm{SD}}(1,\ldots,n) = \frac{1}{\sqrt{n!}}\sum_{k=1}^{n!}(-1)^{p_k}\hat{P}_k\prod_{i=1}^{n}\psi_i(i) = \frac{1}{\sqrt{n!}}\begin{vmatrix}\psi_1(1) & \psi_2(1) & \ldots & \psi_n(1)\\ \psi_1(2) & \psi_2(2) & \ldots & \psi_n(2)\\ \vdots & \vdots & \ddots & \vdots\\ \psi_1(n) & \psi_2(n) & \ldots & \psi_n(n)\end{vmatrix}. \quad (2.4)$$

The Slater determinant is an antisymmetrized product of spin orbitals ψ_n, which consists of spatial orbitals and spin functions. It converts an n-electron problem into n one-electron problems. In Eqn. 2.4, the factor $\frac{1}{\sqrt{n!}}$ normalizes the Slater determinant, and p_k is the parity of the k-s permutation, that controls the sign of the wave function. $\hat{P}_k$ is the permutation operator, and $\prod_{i=1}^{n}\psi_i(i)$ is the Hartree product. The variational principle is applied by optimizing the spin orbitals ψ_i to find the Slater determinant with the lowest energy.

2.1.1 Hartree-Fock Formalism

The Hartree-Fock method forms the basis of several methods in quantum chemistry, that are based on the mean-field approximation. In the Hartree-Fock formalism the wave function is approximated using the Slater determinant,

$$E\left[\Psi^{\mathrm{SD}}\right] = \frac{\left\langle\Psi^{\mathrm{SD}}|\hat{H}|\Psi^{\mathrm{SD}}\right\rangle}{\langle\Psi^{\mathrm{SD}}|\Psi^{\mathrm{SD}}\rangle}. \quad (2.5)$$

By applying the Slater-Condon rules to Eqn. 2.5, we get an expression for the energy

$$E_{\mathrm{HF}} = \left\langle\Psi^{\mathrm{SD}}|\hat{H}|\Psi^{\mathrm{SD}}\right\rangle = \sum_{i}^{n}\left\langle\phi_i|\hat{h}|\phi_i\right\rangle + \frac{1}{2}\sum_{ij}^{n}\langle\phi_i\phi_j||\phi_i\phi_j\rangle\,, \quad (2.6)$$

where $\hat{h}$ is the one-electron Hamiltonian operator, and ϕ_i are the spin orbitals of electron i. The last term contains the exchange and correlation operators $\hat{J}$ and $\hat{K}$, which will be explained later. This equation can be reformulated into the Hartree-Fock integro-differential equation,

$$\hat{h}_i(1)\phi_i(1) + \sum_{j\neq i}^{n}\left[\int dx_2|\phi_j(2)|^2\frac{1}{r_{12}}\right]\phi_i(1) - \sum_{j\neq i}^{n}\left[\int dx_2\phi_j^*(2)\phi_i(2)\frac{1}{r_{12}}\right]\phi_j(1) = \epsilon_i\phi_i(1). \quad (2.7)$$

The solution to this eigenvalue problem gives the spin orbitals $\phi_i(1)$, which are associated with the lowest Hartree-Fock-energy. $\hat{h}_i(1)$ is the one-electron Hamiltonian operator, which covers the kinetic energy and the attractive force between electrons and nuclei. The orbital energy ϵ_i in Eqn. 2.7 is a good approximation for the negative energy needed to remove the electron (1) from the orbital $\phi_i(1)$, also known as Koopmans' theorem.

The last two terms on the left-hand side of Eqn. 2.7 are the two-electron terms, which are describing the mean field electron-electron interactions in an average way, without correlation effects. The first of the two-electron terms is the Coulomb operator $\hat{J}_j$ defined

as

$$\hat{J}_j(1) = \int dx_2 |\phi_j(2)|^2 \frac{1}{r_{12}}, \tag{2.8}$$

and represents the potential that an electron 1 experiences due to the average charge distribution of another electron in spin orbital ϕ_j. $|\phi_j(2)|^2 dx_2$ is the probability that the second electron 2 is in the volume element dx_2. The Coulomb potential describes the effect of an average Coulomb term $1/r_{12}$ of $(N-1)$-electrons, integrated over the entire space $\int dx_2$, weighted with the probability $|\phi_j(2)|^2$, acting on the first electron in $\phi_i(1)$. The Coulomb operator is a local operator because the application of $\hat{J}_j(1)$ on a spin orbital $\phi_j(1)$ depends only on the ϕ_j at the position of electron 1.

The second two-electron term in Eqn. 2.7 is the exchange operator $\hat{K}_j$. The exchange operator has only a quantum mechanical interpretation, defining the exchange of two electrons. The definition of $\hat{K}_j$ when it acts on a spin function is,

$$\hat{K}_j(1)\phi_i(1) = \int dx_2 \phi_j^*(1)\phi_i(2)\frac{1}{r_{12}}\phi_j(1). \tag{2.9}$$

The exchange operator describes the correlation effect of two electrons with the same spin, that accounts for their decreased repulsion energy, due to the Pauli principle. The expectation values of the Coulomb and exchange operators acting on a spin orbital are given by,

$$\begin{aligned} J_{ij} &= \left\langle \phi_i | \hat{J}_j | \phi_i \right\rangle = \int\int dx_1 dx_2 \phi_i^*(1)\phi_j^*(2)\tfrac{1}{r_{12}}\phi_i(1)\phi_j(2), \\ K_{ij} &= \left\langle \phi_i | \hat{K}_j | \phi_i \right\rangle = \int\int dx_1 dx_2 \phi_i^*(1)\phi_j^*(2)\tfrac{1}{r_{12}}\phi_i(2)\phi_j(1). \end{aligned} \tag{2.10}$$

Using the definitions in 2.8 and 2.9 for the Coulomb and exchange operator, respectively, allows us to rewrite Eqn. 2.7 as,

$$\left[\hat{h}(1) + \sum_{j\neq i}^{n} \hat{J}_i(1) - \sum_{j\neq i}^{n} \hat{K}_j(1)\right] \phi_i(1) = \left[\hat{h}(1) + \hat{v}^{\text{eff}}(1)\right] \phi_i(1) = \hat{f}_i\phi_i(1) = \epsilon_i\phi_i(1), \tag{2.11}$$

where $\hat{f}$ is the Fock operator. This equation can be solved in an iterative scheme. In a self-consistent field (SCF) approach, the eigenvalue problem in Eqn. 2.11 is solved by varying the spin orbitals until the orbital energy is not changing anymore.

The spin orbitals are represented as linear combination of basis functions χ_i

$$\phi_i = \sum_{k=1}^{k_{\max}} c_{ki}\chi_k, \tag{2.12}$$

where the c_{ki} are expansion coefficients, which determine the mixture of the individual basis functions. The expansion coefficients are obtained by applying the Hartree-Fock procedure. The basis functions are one-electron functions and also referred to as atomic orbitals. Typical choices of basis functions include Gaussian- or Slater-type functions. Molecular orbitals are constructed in a linear combination of atomic orbitals (LCAO). The exact Hartree-Fock energy is obtained at the HF-limit with a complete basis set ($k_{\max} = \infty$).

Introducing basis functions to the Hartree-Fock formalism reduces the problem to calculating a set of expansion coefficients c_{ki} and gives rise to the Roothaan-Hall equation, here formulated as a matrix,

$$\underline{\underline{F}} \quad \underline{\underline{C}} = \underline{\underline{S}} \quad \underline{\underline{C}} \quad \underline{\epsilon}. \tag{2.13}$$

Here, $\underline{\underline{F}}$ is the Fock matrix with matrix elements $F_{kl} = \left\langle \chi_k | \hat{f} | \chi_l \right\rangle$. $\underline{\underline{C}}$ is the sought-after coefficient matrix. $\underline{\underline{S}}$ is the overlap matrix with elements $S_{kl} = \langle \chi_k | \chi_l \rangle$. The diagonal matrix $\underline{\epsilon}$ contains the spin-orbital energies. After orthonormalization of the basis functions, obtained by the transformation $\underline{\underline{X}}^{-1} \quad \underline{\underline{S}} \quad \underline{\underline{X}} = \underline{1}$, Eqn. 2.13 reduces to a common eigenvalue problem,

$$\underline{\underline{\tilde{F}}} \quad \underline{\underline{\tilde{C}}} = \underline{\underline{\tilde{C}}} \quad \underline{\epsilon}. \tag{2.14}$$

We have thus reduced the functional eigenvalue problem Eqn. 2.7 to a matrix eigenvalue problem.

2.2 Density-Functional Theory

The Hartree-Fock method is based on the wave function to minimize the total energy. Density-functional theory (DFT) provides an alternative approach, in which the electron density gives all the information we need to construct an Hamiltonian of the system. Using the electron density as a fundamental quantity leads to a dimensional reduction in comparison to wave-function-based methods. A many-body wave function is a function of $4 \times N$ dimensions (three spatial coordinates and one spin information per atom), whereas the electron density only has three spatial coordinates (x, y, z). The electron density is also an experimental observable, *e.g.* in X-ray crystallography.

The electron density $\rho(\vec{r})$ is defined as an integral over the square of the wave function Ψ, namely

$$\rho(\vec{r}_1) = n \int |\Psi|^2 ds_1 d\vec{q}_2 d\vec{q}_3 \dots d\vec{q}_n, \tag{2.15}$$

where s_1 is the spin coordinate of electron 1, $\vec{q}_i = (x_i, y_i, z_i, s_i)$ contains the spatial and spin coordinate of the i^{th} electron. The electron density vanishes in the limit of large distances $\lim_{|\vec{r}|\to\infty} \rho(\vec{r}) = 0$, and it is normalized, meaning that integration over the probability $\rho d\vec{r}$ gives the total number of electrons n,

$$\int \rho(\vec{r}) dr = n. \tag{2.16}$$

$\rho(\vec{r})$ has maxima, and cusps, at the positions $\vec{R}_I$ of the nuclei.

The idea to construct the Hamiltonian based on the electron density came originally from Thomas and Fermi [92–94] in 1927. The Thomas-Fermi model is based on a quantum statistical model of electrons that takes into account the kinetic energy of the electrons, and the attractive potential between the nuclei and the electrons. The electron-electron interactions are treated in a completely classical way. The expression for the kinetic energy

is based on the uniform electron gas. The Thomas-Fermi model is only of limited use, as it cannot describe the energy of an atom, due to very crude approximations to the kinetic energy, and the fact that exchange and correlation effects are neglected.

2.2.1 The Hohenberg-Kohn Theorems

The proof that there is only one electron density, which defines the ground-state energy was shown by Hohenberg and Kohn [95]. In their first theorem they showed that the electron density uniquely determines the Hamiltonian and thus all properties of the system. For a non degenerated ground state, let us assume there exist two external potentials V_{ext} and V'_{ext}, that both describe the same ground state $\rho(\vec{r})$. The two external potentials are part of two Hamiltonians, $\hat{H} = \hat{T} + \hat{V}_{\text{ee}} + \hat{V}_{\text{ext}}$ and $\hat{H}' = \hat{T} + \hat{V}_{\text{ee}} + \hat{V}'_{\text{ext}}$, which only differ in their external potentials. The two Hamiltonians $\hat{H}$ and $\hat{H}'$ give rise to two different ground-state wave functions Ψ and Ψ' with energies E_0 and E'_0, with $E_0 \neq E'_0$. Let us further assume that both wave functions Ψ and Ψ' are different and belong to the same electron density. For Ψ' as a trial wave function for $\hat{H}$ we obtain,

$$E_0 < \left\langle \Psi' | \hat{H} | \Psi' \right\rangle = \left\langle \Psi' | \hat{H}' | \Psi' \right\rangle + \left\langle \Psi' | \hat{H} - \hat{H}' | \Psi' \right\rangle. \tag{2.17}$$

Due to the assumption that the two Hamiltonians only differ in their external potential we can reformulate Eqn. 2.17 as follows,

$$E_0 < E'_0 + \left\langle \Psi' | \hat{T} + \hat{V}_{\text{ee}} + \hat{V}_{\text{ext}} - \hat{T} - \hat{V}_{\text{ee}} - \hat{V}'_{\text{ext}} | \Psi' \right\rangle \tag{2.18}$$

which gives,

$$E_0 < E'_0 + \int \rho(\vec{r}) \left\{ V_{\text{ext}} - V'_{\text{ext}} \right\} d\vec{r}. \tag{2.19}$$

By interchanging the primed and unprimed quantities, and repeating the steps in Eqn. 2.17, 2.18, and Eqn. 2.19, we obtain a corresponding relation,

$$E'_0 < E_0 + \int \rho(\vec{r}) \left\{ V_{\text{ext}} - V'_{\text{ext}} \right\} d\vec{r}. \tag{2.20}$$

Adding Eqn. 2.19 and 2.20 leads to a clear contradiction,

$$E_0 + E'_0 < E'_0 + E_0. \tag{2.21}$$

This proof shows that there cannot be two different external potentials that yield the same ground-state electron density. This means that the electron density uniquely specifies the external potential V_{ext}. To summarize what we know so far about $\rho(\vec{r})$, we can write the following relation,

$$\rho \Rightarrow n, Z_I, R_I \Rightarrow \hat{H} \Rightarrow \Psi_0 \Rightarrow E_0. \tag{2.22}$$

In the Hohenberg-Kohn formalism, the ground state energy E_0 is define as,

$$E_0[\rho_0] = T[\rho_0] + E_{\text{ee}}[\rho_0] + E_{\text{Ke}}[\rho_0] = \int \rho_0(\vec{r}) V_{\text{Ke}} d\vec{r} + F_{\text{HK}}[\rho_0]. \tag{2.23}$$

Where $T[\rho_0]$ is the kinetic energy, $E_{\mathrm{Ke}}[\rho_0]$ denotes the potential energy between electrons and nuclei, and the electron-electron interactions are described by $E_{\mathrm{ee}}[\rho_0]$. In Eqn. 2.23, the only system dependent term is $\int \rho_0(\vec{r}) V_{\mathrm{Ke}} d\vec{r}$, whereas all other terms are system independent and contained in the Hohenberg-Kohn functional $F_{\mathrm{HK}}[\rho_0]$. However, this functional is unknown, and therefore Eqn. 2.23 can only be solved by employing approximations.

The Hohenberg and Kohn theorem states that the ground state density is sufficient to obtain all properties of interest. However, the question remains: how do we know that the density really belongs to the ground state? Hohenberg and Kohn suggest in their second theorem to tackle this problem by applying the following variational principle

$$E_0 \leq E[\tilde{\rho}] = T[\tilde{\rho}] + E_{\mathrm{ee}}[\tilde{\rho}] + E_{\mathrm{Ke}}[\tilde{\rho}]. \tag{2.24}$$

The energy of the functional of the electron density $\rho(\vec{r})$ in Eqn. 2.23, gives the lowest energy that defines the ground state and the test electron density $\tilde{\rho}$ is varied to find the lowest energy.

2.2.2 Kohn-Sham Formalism

Kohn and Sham [96] derived a method to approximate the energy as functional of the electron density as

$$E[\rho] = T_{\mathrm{s}} + J + E_{\mathrm{Ke}} + E_{\mathrm{XC}}. \tag{2.25}$$

The energy functional in the Kohn-Sham formalism is composed of the kinetic energy T_{s}, the Coulomb interaction J, the electron-nuclear attraction E_{Ke}, and the exchange-correlation term E_{XC}. Kohn and Sham defined this in terms of a non-interacting system with the same density as the real system. They expressed the wave function as a Slater determinant with so-called Kohn-Sham orbitals ϕ_i, which are obtained from solving the following eigenvalue problem,

$$\hat{f}^{\mathrm{ks}} \phi_i = \epsilon_i \phi_i, \tag{2.26}$$

where $\hat{f}^{\mathrm{ks}}$ is an effective one-electron Kohn-Sham operator, given by

$$\hat{f}^{\mathrm{ks}} = -\frac{1}{2}\Delta + V_s(\vec{r}), \tag{2.27}$$

where the Laplace operator $\Delta = \nabla^2$ and $V_s(\vec{r})$ is the Kohn-Sham potential. The sum over the square of the Kohn-Sham orbitals gives the electron density,

$$\sum_{i}^{n} \sum_{s} |\phi_i(\vec{r}, s)|^2 = \rho_0(\vec{r}), \tag{2.28}$$

while the kinetic energy T_{s} constructed with these orbitals reads

$$T_{\mathrm{s}} = -\frac{1}{2} \sum_{i}^{n} \langle \phi_i | \Delta | \phi_i \rangle \,. \tag{2.29}$$

With the definitions 2.27, 2.25, and 2.29 the Kohn-Sham equation 2.26 is formulated as

$$\left(-\frac{1}{2}\Delta + \left[\int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 - \sum_I^N \frac{Z_I}{r_{1I}} + V_{\mathrm{XC}}\right]\right)\phi_i = \left(-\frac{1}{2}\Delta + V_{\mathrm{eff}}(\vec{r}_1)\right)\phi_i = \epsilon_i \phi_i. \quad (2.30)$$

The only unknown is the exchange-correlation potential V_{XC}, which corresponds to the exchange-correlation functional $E_{\mathrm{XC}}[\rho(\vec{r})]$ in Eqn. 2.25. Since it cannot be derived rigorously from first principles, it must be approximated and there exist various ansatzes to do so, as we shall see in the next section.

2.2.3 Exchange-Correlation Functional

The exchange-correlation functional is split into two parts: the first part includes the exchange energy $E_{\mathrm{X}}[\rho]$, and the second part covers the correlation energy $E_{\mathrm{C}}[\rho]$.

There are different ways to approximate the exchange-correlation functional. The local density approximation (LDA) approximates E_{XC} as a functional of the electron density $\rho(\vec{r})$ at the position $\vec{r}$,

$$E_{\mathrm{XC}}^{\mathrm{LDA}}[\rho] = \int \rho(\vec{r})\epsilon_{\mathrm{XC}}(\rho) d\vec{r} \quad \text{and} \quad \epsilon_{\mathrm{XC}}(\rho) = \epsilon_{\mathrm{X}}(\rho) + \epsilon_{\mathrm{C}}(\rho). \quad (2.31)$$

Here, $\epsilon_{\mathrm{XC}}(\rho)$ is the exchange-correlation energy per particle of a homogeneous electron gas with density $\rho(\vec{r})$. The exchange part $\epsilon_{\mathrm{X}}(\rho)$ is calculated using the so-called Slater exchange,

$$\epsilon_X = -\frac{3}{4}\sqrt[3]{\frac{3\rho(\vec{r})}{\pi}}. \quad (2.32)$$

Unlike $\epsilon_{\mathrm{X}}(\rho)$, there exists no analytical expression for the correlation part of the energy $\epsilon_{\mathrm{C}}(\rho)$. Approximate expressions for $\epsilon_C(\rho)$ have been obtained from quantum Monte-Carlo simulations of a homogeneous electron gas [97], [98]. The LDA approach is known to fail at the surface of atoms, because of rapid changes of the electron density over small volumes.

To improve the LDA, the exchange-correlation functional can be estimated with the generalized-gradient approximation (GGA). In this method, the gradient of the electron density is also taken into account. The exchange-correlation functional is expanded in a Taylor series with respect to the electron density, and truncated at the linear term, which depends on the gradient,

$$E_{\mathrm{XC}}^{\mathrm{GGA}} = \int \rho(\vec{r})\epsilon_{\mathrm{xc}}\left[\rho(\vec{r}), \nabla\rho(\vec{r})\right] d\vec{r}. \quad (2.33)$$

There are many GGA functionals which contain parameters calibrated against reference values rather than derived from first principles. Popular examples for GGA functionals are Becke88 [99], or the functional from Lee, Yang, and Parr (LYP) [100]. The LYP correlation functional is often used in combination with the Becke88 exchange functional in the BLYP functional.

In *meta*-GGA functionals, the approximation behind Eqn. 2.33 is further improved by including a second order term. The general expression of a *meta*-GGA is,

$$E_{\mathrm{XC}}^{\text{meta-GGA}} = \int F\left(\rho(\vec{r}), \nabla\rho(\vec{r}), \nabla^2\rho(\vec{r})\right) d\rho(\vec{r}). \quad (2.34)$$

TPSS is an example of a *meta*-GGA functional [101].

In hybrid functionals, fitted parameters from experimental data are included, as well as a certain amount of exact Hartree-Fock exchange. A popular hybrid functional is the B3-LYP functional [102]. It consists of an exchange part from Becke with three parameters, in combination with the correlation part from the LYP functional,

$$E_{\mathrm{XC}}^{\mathrm{B3\text{-}LYP}} = (1-a)E_{\mathrm{X}}^{\mathrm{LDA}} + aE_{\mathrm{X}}^{\mathrm{HF}} + b\Delta E_{\mathrm{X}}^{\mathrm{B}} + (1-c)E_{\mathrm{C}}^{\mathrm{LDA}} + cE_{\mathrm{C}}^{\mathrm{LYP}}, \tag{2.35}$$

where $a = 0.2$, $b = 0.72$, and $c = 0.81$. The B3-LYP functional was used in all quantum chemical calculations in this work.

2.3 Molecular-Dynamics Simulation

In MD simulations the motion of atoms is treated with tools of classical physics. The evolution in time is calculated by numerically solving Newton's equations of motion. The history of MD simulations goes back to 1957, when Alder and Wainwright simulated the phase transition for a hard sphere system [103]. In 1964, Rahman used a Lennard-Jones potential in an MD simulation to describe the interactions of liquid argon [104]. The first MD simulation of a protein was performed by McCammon in 1977 on the *bovine* pancreatic trypsin inhibitor (BPTI) [105]. With increasing computational power and more available protein structures, MD simulations are now a central tool in biophysics and molecular biology, acting as a "molecular microscope". In the following I will briefly review the fundamentals of classical mechanics, statistical mechanics and free energy calculations [106], [107], which form the foundations of MD simulations.

2.3.1 Classical Mechanics

The equations of motion used in MD simulations are derived from Newton's second law,

$$\vec{F}_i = m_i \frac{d^2\vec{r}_i(t)}{dt^2}, \tag{2.36}$$

where m_i is the mass and $\vec{r}_i$ the time dependent coordinates of the i^{th} atom. For conservative systems, the force $\vec{F}_i$ is the negative gradient of the potential U,

$$F_i = -\nabla_i U. \tag{2.37}$$

The differential Eqn. 2.36, and 2.37 are solved using numerical algorithms (*e.g.* the Verlet algorithm) to obtain a trajectories $\vec{r}_i$ of the atoms.

2.3.2 Statistical Mechanics

MD simulations deal with microscopic observables. A microscopic state is defined by the phase space variable $\Gamma = (\vec{q}_1, \ldots, \vec{q}_N, \vec{p}_1, \ldots, \vec{p}_N)$, with three cartesian coordinates $\vec{q}_i$ for every atom i of an N-particle system and three associated momentum coordinates $\vec{p}_i$. An MD trajectory can be seen as time series where various microstates of the phase space are sampled. A macrostate is a collection of microstates, which share the same macro-

scopic quantities measured in experiments, *e.g.*, particle number, temperature, pressure, and volume. Statistical mechanics provide a framework to bridge the two worlds of microscopic and macroscopic quantities. The collection of microstates build an ensemble, from which average macroscopic quantities can be derived. The central point in the analysis of an MD trajectory is to formulate an equilibrium ensemble average of a observable A as,

$$\langle A \rangle = \frac{\int A(\Gamma) p(\Gamma) d\Gamma}{\int p(\Gamma) d\Gamma}, \tag{2.38}$$

where $A(\Gamma)$ denotes the observed property as function of the phase space variables, $p(\Gamma)d\Gamma$ is the probability of a microstate, which is defined by Γ and $d\Gamma = \int d\vec{p}^N d\vec{q}^N$ is the volume element in phase space over N coordinates and N momenta.

To extract macroscopic quantities from an MD simulation, we have to build a time average over the trajectory. The ergodic hypothesis states that under equilibrium conditions, as $t \to \infty$ the time average and the ensemble average coincide,

$$\lim_{t\to\infty} \frac{1}{\tau} \int_0^\tau dt A(t) = \langle A \rangle_{eq}. \tag{2.39}$$

This allows us, in case of long simulation times, to use the time average extracted from MD simulation to estimate ensemble averages and extract macroscopic quantities like the free energy.

One can perform MD simulations in various ensembles, which are distinguished by what is kept constant, *e.g.*, NVE ensemble is performed with constant particle number N, constant volume V, and energy E. The ensemble mainly used in this work is the isothermal-isobaric (NpT) ensemble, where the temperature T and pressure p are kept constant next to the particle number N. In a simulation performed at constant temperature of a system with discrete energy levels, the ensemble average in Eqn. 2.38 assume the form,

$$\langle A \rangle = \sum_i^M A_i p_i. \tag{2.40}$$

M is the total number of microstates, and p_i is the Boltzmann distribution given by

$$p_i = \frac{g_i e^{-\beta \mathcal{E}_i}}{Q}, \tag{2.41}$$

where $\beta = 1/(k_\mathrm{B}T)$, k_B is the Boltzmann constant, and T is the absolute temperature of the system. g_i is the degeneracy of the state i with energy $\mathcal{E}_i$. The denominator Q is the partition function that normalizes the distribution. The Boltzmann distribution is the probability of occupying a microstate i. At equilibrium the relative probability of two microstates is determined by their energy difference in the exponential. The Boltzmann factor is the ratio of probabilities of occupying the states i and j, respectively,

$$\frac{p_i}{p_j} = e^{\beta(\mathcal{E}_j - \mathcal{E}_i)}, \tag{2.42}$$

where $\mathcal{E}_i$ and $\mathcal{E}_j$ denote the energies of the microstates. The Boltzmann factor tells us how much more likely it is to occupy one microstate over another one.

The partition function Q, in Eqn. 2.41, is defined in systems with discrete energy states, as the sum of the exponential energies over all microstates M,

$$Q = \sum_{i=1}^{M} g_i e^{-\beta \mathcal{E}_i}, \tag{2.43}$$

The partition function is a central quantity in statistical physics. Most thermodynamic observables of a system, such as free energy, entropy, and pressure, can be expressed in terms of the partition function. For example, the free energy G is given by

$$G = -k_{\mathrm{B}} T \ln(Q). \tag{2.44}$$

2.3.3 Force Fields

In MD simulations, the interactions between atoms are described by force fields. The most generic functional form of a force field, is given by

$$\begin{aligned} U_{\mathrm{EM}} = & \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi - \zeta)] \\ + & \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_u (u - u_0)^2 \\ + & \sum_{\text{nonbonded}} \varepsilon \left[\left(\frac{R_{\min(i,j)}}{r_{i,j}} \right)^{12} - \left(\frac{R_{\min(i,j)}}{r_{i,j}} \right)^{6} \right] + \frac{1}{4\pi\epsilon_0 \epsilon_r} \frac{q_i q_j}{r_{i,j}}. \end{aligned} \tag{2.45}$$

The first term of Eqn. 2.45 describes covalent bonds by a harmonic potential with the force constant k_b and $(b - b_0)$ are displacements from equilibrium. The second term accounts for the bond angles, k_θ is the angle force constant, and $(\theta - \theta_0)$ is the angle deviating from the equilibrium configuration. The third term describes the dihedral angle of four covalently bound atoms with a force constant k_ϕ, and dihedral angle ϕ, the multiplicity n, and the phase shift ζ of the cosinus function. The fourth term is the so-called improper term. It conserves specific chemical properties, for example it prevents out-of-plain bending with k_ω as the improper force constant, and ω as the out-of-plane angle. The Urey-Bradley term is an additional term that describes angle bending of three atoms (labeled 1, 2, and 3), it is defined between the nonbonded atoms 1 and 3. k_u is the respective force constant of a harmonic potential acting along the distance u between atoms 1 and 3. The two last terms represent the nonbonded interactions between atoms i and j, which are separated by three or more bonds. The van-der-Waals interactions are modeled using a standard 12-6 Lennard-Jones potential and the electrostatic interactions are given by a Coulombic potential.

2.3.4 Langevin Dynamics

Langevin dynamics is an extension of Newton's dynamics where the thermal excitations of a fictitious surrounding heat bath, acting as a thermostat, are taken in to account. The

Langevin equation is

$$m_i \frac{d^2\vec{r}_i}{dt^2} = -\nabla U(\vec{r}_i) - \gamma \frac{d\vec{r}_i}{dt} + \sqrt{2\gamma k_\mathrm{B} T} R(t), \tag{2.46}$$

with masses m_i and coordinates $\vec{r}_i$ of the i^{th} atom. $U(\vec{r}_i)$ is the interaction potential between the particles. The negative second term on the right hand side describes the viscosity of the solvent via Stockes friction, γ is a friction constant multiplied by the velocity. Setting $\gamma = 0$, reduces the dynamics to Newton's law (Eqn. 2.36). The last term adds random forces and in combination with the friction therm it represents the coupling to a heat bath, where $R(t)$ is a Gaussian distributed random variable with mean zero and unit standard derivation.

2.3.5 Molecular Dynamics Flexible Fitting

The molecular molecular dynamics flexible fitting (MDFF) [108] can be used to fit atomistic structures into electron microscopy (EM) maps using MD simulations. It drives the selected atoms via a force that biases the structure towards the high-density areas of the electron map. The total potential energy function of an MDFF simulation,

$$U_\mathrm{total} = U_\mathrm{MD} + U_\mathrm{EM} + U_\mathrm{SecStruc}. \tag{2.47}$$

is composed of the all-atom potential energy function U_MD (Eqn. 2.45), and the potential energy U_EM of the electron density map into a potential, which acts additionally on the selected atoms. The stereochemical properties (*cis/trans* chirality of the peptide bond) of the structure, and the integrity of the secondary structure is preserved in MDFF simulations with help of a third potential, U_SecStruc. The potential energy function, which represents the electron density map is defined as,

$$U_\mathrm{EM}(R) = \sum_i w_i V_\mathrm{EM}(\vec{r}_i), \tag{2.48}$$

where w_j are weights, that depend on the atomic masses. The potential $V_\mathrm{EM}(\vec{r}_i)$ acts on the atomic coordinates $\vec{r}_i$ of the selected atoms, and is defined as

$$V_\mathrm{EM}(r) = \begin{cases} \zeta \frac{\Phi(r) - \Phi_{thr}}{\Phi_\mathrm{max} - \Phi_\mathrm{thr}} & \text{if} \quad \Phi(r) \geq \Phi_\mathrm{thr}, \\ \zeta & \text{if} \quad \Phi(r) < \Phi_\mathrm{thr}. \end{cases} \tag{2.49}$$

The electron density map is converted into a Coulomb-potential $\Phi(r)$, the maximum value of all voxels is Φ_max. Φ_thr is a threshold value to neglect densities, which do not belong to the selected atoms, and $\zeta < 0$ is an arbitrary negative scaling factor.

2.3.6 Steered Molecular Dynamics

In steered molecular dynamics simulations (SMD), selected atoms are steered in a certain direction by introducing a biasing force. The SMD potential acting on the selected atoms

additionally to the all-atom potential energy function (Eqn. 2.45) can be written as,

$$U(\vec{r}_1, \vec{r}_2, ..., t) = \frac{1}{2}k\Big[vt - (\vec{R}(t) - \vec{R}_0)\vec{n}\Big]^2. \tag{2.50}$$

The potential acts on the coordinates $\vec{r}_1, \vec{r}_2, ...,$ of the selected atoms at every time step t. $\vec{R}(t)$ and $\vec{R}_0$ are the current and initial positions of the center of mass of the selected atoms, respectively. The center of mass is harmonically restrained with the force constant k and is moved along $\vec{n}$ with the velocity v. In SMD simulations, the center of mass of the selected atoms is pulled along the restrained direction, this allows the selected molecule to adapt to the environment it is pulled through.

2.4 Hybrid Quantum Mechanical/Molecular Mechanical Approach

The hybrid approach quantum QM/MM combines the strength of classical molecular mechanics and quantum mechanical methods in a molecular simulation. The QM accuracy is required to describe chemical reactions, because chemical reactions cannot be described using classical force fields. DFT-based QM methods are normally restricted to a few hundreds of atoms. To simulate a large biomolecular system we need to capture hundreds of thousands of atoms on a microsecond time scale, which require MD simulations based on classical force fields. The QM/MM approach was introduced in 1976 by Warshel and Levitt [109], and together with Karplus, their findings were honored with the Nobel prize in 2013 for "the development of multi scale models for complex chemical systems". Nowadays QM/MM method has become popular when investigating chemical reactions in the condensed phase. In this section the basic fundamentals of different QM/MM schemes are described (see also [110], a current review article).

The QM/MM system is partitioned into an inner (QM) and an outer (MM) region. The MM region is described by a classical force field (Eqn. 2.45). For the QM region, a quantum chemical method like DFT, as discussed in section (2.2), is applied. There are two main ways to calculate the total energy of the QM/MM system. In the subtractive scheme the total energy is determined via,

$$E^{\text{sub}}_{\text{QM/MM}} = E_{\text{MM}}(S) + E_{\text{QM}}(I + L) - E_{\text{MM}}(I + L). \tag{2.51}$$

Here, the first term on the right-hand side accounts for the entire system S and is evaluated using MM (force fields). In the second term, a QM calculation is performed for the inner system I with additional link atoms L (see below). Finally, we subtract the energy of the inner region $(I+L)$ evaluated on a MM level. The advantage of this method is its simplicity and that no QM/MM coupling terms have to be evaluated. The method requires classical parameters for the QM region, which is problematic in cases where the atoms are polarized. Another disadvantage of the subtractive scheme in Eqn. 2.51 is that the coupling between the regions is handled on the MM level. There are, however, also substractive schemes with electrostatic embedding [111].

The QM/MM energy can also be computed using an additive scheme,

$$E^{\text{add}}_{\text{QM/MM}} = E_{\text{MM}}(O) + E_{\text{QM}}(I+L) + E_{\text{QM/MM}}(I,O), \tag{2.52}$$

where the MM energy evaluation in the first term is only applied to the outer region O. The second term is equivalent to the second term in Eqn. 2.51. The third term is the QM/MM coupling term. There are different ways to handle this term, the simplest of which is mechanical embedding. In this approach, the interactions are treated on a classical level, where a charge model of MM is applied to the QM region. The mechanical embedding has several limitations. The charge of the outer region does not interact with the QM electron density, and there is no direct influence to the electrostatic environment. Another disadvantage is the fact that application of appropriate point charges to the QM atoms is non-trivial, especially when their electrostatic character changes during the reaction. Additionally, the MM charges do not necessary reproduce the true charge distribution. This can lead to drastic errors, since the accurate description of the electrostatic interactions between the outer and inner regions is essential for the realistic modeling of biomolecules.

The electrostatic embedding scheme overcomes the major shortcomings of the mechanical embedding scheme, by incorporating the MM point charges in the form of one-electron Hamiltonian terms,

$$\hat{H}^{\text{el}}_{\text{QM-MM}} = -\sum_{i}^{N}\sum_{J\in O}^{L}\frac{q_J}{|\vec{r}_i - \vec{R}_J|} + \sum_{\alpha\in I+L}^{M}\sum_{J\in O}^{L}\frac{q_J Q_\alpha}{|\vec{R}_\alpha - \vec{R}_J|}, \tag{2.53}$$

in the QM Hamiltonian. Here, q_J denotes the MM partial charges at the positions $\vec{R}_J$, $\vec{r}_i$ are the electron positions, and Q_α and $\vec{R}_\alpha$ are the respective charges and positions of the nuclei. The indices i, J, and α run over the N electrons, L point charges of the outer region, and M nuclei in the QM region, respectively. $\hat{H}^{\text{el}}_{\text{QM-MM}}$ allows the electronic structure of the inner region to adapt to changes in the charge distribution of the environment. The QM-MM electrostatic interaction allows for the QM region to become polarized by the MM region. Furthermore, there is no need to assign point charges to the QM atoms that would otherwise require additional parametrization.

The electrostatic coupling between the QM and MM region can be further improved via a polarized embedding scheme, which requires a flexible charge model that is polarized by the QM charge density and thus introduces mutual polarization. However, polarizable biomolecular force fields are still under development.

A possible computational problem can arise when covalent bonds, crossing the QM/MM boundary. One approach to solve this problem is to introduce link atoms that saturate the free valences of QM region atoms [112], [113]. This is often achieved by introducing a hydrogen atom between the last QM atom and the first MM atom. Seen from the QM region, it behaves as a hydrogen atom, but its structural degrees of freedom remain undetected by the MM region.

2.5 Free Energy Perturbation

The free energy is a thermodynamic quantity that determines the behavior of a system at equilibrium. Processes like ligand binding to a macromolecule, partitioning of drugs across cell membranes, or conformational changes of proteins are governed by favorable changes in the free energy. To understand and make reliable predictions for such processes, we have to know the associated free energy change. Free energy calculations are widely used and have become powerful tools due to the increase computational resources. The following review of the theory behind free energy perturbation calculations is largely based on the discussion found in [114].

Free energy perturbation (FEP) is one of the oldest, most frequently used, and well-established methods. For a rigorous derivation of FEP, see Lev Landau's textbook from 1938 [115]. In 1954 Zwanzig developed the FEP theory for statistical mechanics of condensed phases [116].

In order to explain the basic concepts of FEP, we will start with the Hamiltonian,

$$H(\Gamma, \lambda) = \sum_{i=1}^{N} \frac{\vec{p}_i^2}{2m_i} + U(\vec{q}_1, \ldots, \vec{q}_N; \lambda). \tag{2.54}$$

$\Gamma = \vec{p}^N \vec{q}^N$ is a point in phase space of a N-particle system. λ is an external parameter, and U is the potential energy, which depends for a conservative system only on the spatial coordinates. The Boltzmann distribution at equilibrium with fixed temperature T and in case of continuous energy levels, is given by

$$p_{\mathrm{eq}}(\Gamma, \lambda) = \frac{e^{-\beta H(\Gamma,\lambda)}}{\int d\Gamma' e^{-\beta H(\Gamma',\lambda)}} = \frac{1}{Q_\lambda} e^{-\beta H(\Gamma,\lambda)}, \tag{2.55}$$

where Q_λ is the partition function (defined in Eqn. 2.43). The free energy difference $\Delta G_{0,1}$ of the states 0 and 1, witch are characterized by the parameter values $\lambda = 0$ and $\lambda = 1$, respectively, is directly related to the ratio of the corresponding partition functions,

$$\Delta G_{0,1} = G_1 - G_0 = -k_{\mathrm{B}} T \ln\Big(\frac{Q_1}{Q_0}\Big) \tag{2.56}$$

In many cases, λ is used as a coupling parameter. Especially in "computational alchemy", this parameter varies the Hamiltonian, resulting in an interpolation between two different systems. This modifies the interatomic interactions, while the system passes through various (even non-physical) intermediate states. Expressing the partition functions of Eqn. 2.56 with integrals and expanding the numerator via the identity $e^{\beta H_0} e^{-\beta H_0} = 1$, gives

$$\begin{aligned} \Delta G_{0,1} &= -k_{\mathrm{B}} T \ln\Big(\frac{\int d\Gamma e^{-\beta H_1} e^{\beta H_0} e^{-\beta H_0}}{\int d\Gamma e^{-\beta H_0}}\Big) \\ &= -k_{\mathrm{B}} T \ln\Big(\frac{\int d\Gamma e^{-\beta (H_1 - H_0)} e^{-\beta H_0}}{\int d\Gamma e^{-\beta H_0}}\Big) = -k_{\mathrm{B}} T \ln \Big\langle e^{-\beta (H_1 - H_0)} \Big\rangle_0 . \end{aligned} \tag{2.57}$$

We thus end up with an ensemble average over the state $\lambda = 0$. Equivalently, we can

formulate an ensemble average over the state $\lambda = 1$,

$$\Delta G_{1,0} = -k_{\mathrm{B}}T \ln \left\langle e^{-\beta(H_0 - H_1)} \right\rangle_1 . \tag{2.58}$$

Accurate free energy estimation is only possible when there is a sufficient overlap in phase space of the two states for which the free energy difference is computed. If this important condition is not fulfilled, one has to introduce M intermediate states, which are not necessarily physically meaningful. The free energy is evaluated for two neighboring intermediate states, and the sum of all intermediate steps gives the total free energy difference between the two thermodynamic states of interest,

$$\Delta G_{0,1}^{\mathrm{total}} = \sum_{n}^{M} \Delta G_{\lambda_n \to \lambda_{n+1}}. \tag{2.59}$$

In this work only free energy differences with constant temperature are considered. There exist a broad selection of techniques to evaluate the free-energy differences. The ones relevant for this thesis are discussed in more details below.

2.5.1 Exponential Averaging

The exponential averaging (EXP) method, first derived by Robert Zwanzig [116], is one of the earliest free energy calculation methods. The derivation of free energy expressions for the EXP method makes use of our result from Eqn. 2.57, which can be written out as follows,

$$\Delta G_{0,1} = -k_{\mathrm{B}}T \ln\Big(\frac{Q_1}{Q_0}\Big) = -k_{\mathrm{B}}T \ln\Big(\langle e^{-\beta \Delta U(\Gamma,\lambda)} \rangle_0\Big) = -k_{\mathrm{B}}T \ln\Big(\frac{1}{N}\sum_{i=1}^{N} e^{-\beta \Delta U(\Gamma_i,\lambda)}\Big) \tag{2.60}$$

where the last expression gives the relation that the free energy difference of two states can be calculated directly by averaging over the exponential of potential-energy differences $\Delta U(\Gamma, \lambda) = U(\Gamma, 1) - U(\Gamma, 0)$ between the reference state $\lambda = 0$ and the target state $\lambda = 1$. We are not interested in kinetic energies, so only the potential-energy differences are considered at each sampled point in phase space with the external parameter λ.

The EXP identity is exact, but the method has a slow convergence in case of small phase space overlaps. The EXP is an unidirectional method, and it is based on simulations of only one state at equilibrium, which makes the method inefficient as compared to bidirectional methods, which take the simulations of both the initial and final state into account for one free energy difference estimation. In the next subsections the bidirectional BAR method and the multidirectional mBAR method, will be introduced.

2.5.2 Bennett Acceptance Ratio

The Bennett acceptance ratio method (BAR) is named after Charles H. Bennett, who developed the method in 1976 [117]. Bennett demonstrated that an estimation of the free energy differences between two states, using input from simulations in both states, is significantly better than an estimation based on a simulation in only one state, as in

the EXP method. Let us consider a system that exist in two states A and B, and we want to calculate the free energy difference between them. With a Metropolis Monte-Carlo algorithm, we sample the energy landscape of the two states,

$$p(\mathrm{A} \to \mathrm{B}) = \min(e^{-\beta \Delta U}, 1) = M(\beta \Delta U), \tag{2.61}$$

where $\Delta U = U(\mathrm{B}) - U(\mathrm{A})$ is the potential energy difference between two states. $M(x) = \min(e^{-x}, 1)$ is the Metropolis function, which serves as a decision criteria to determine which state is more probable. The states are sampled according to their relative occupation probabilities given by their respective Boltzmann distributions at temperature T. We further assume that both states A and B have common conformational states. To calculate the free energy difference based on simulations in states A and B, we make the ansatz

$$e^{-\beta(\Delta G - C)} = \frac{\langle f(\beta(U_B - U_A - C))\rangle_A}{\langle f(\beta(U_A - U_B - C))\rangle_B}, \tag{2.62}$$

where ΔG is the free energy difference we want to estimate, f is an arbitrary function, and C is a constant. Bennett found that the most optimal choice for the function f in Eqn. 2.62 is a Fermi function, $f(x) = (1 + e^x)^{-1}$, and the constant C in Eqn. 2.62 is set to $C = \ln\left(\frac{Q_0}{Q_1}\right) + C'$[2]. The choice of f and C gives the best asymptotically unbiased estimator possible for the acceptance ratio method [118].

Shirts and Pande [118] derived the BAR Eqn. 2.62 via maximum likelihood estimation. The derivation starts from the Crooks relation [119],

$$\ln\left[\frac{P_{\mathrm{F}}(W)}{P_{\mathrm{R}}(-W)}\right] = \beta(W - \Delta G), \tag{2.63}$$

where $P_{\mathrm{F}}(W)$ and $P_{\mathrm{R}}(-W)$ are the conditional probabilities of observing the non-equilibrium work value W in the forward (F) and $-W$ in a reverse (R) measurement. Equation 2.63 can be reformulated by interpreting the arguments of the logarithm as Bayesian posteriors, and expressing them as products of priors and likelihoods, *e.g.* $P(\mathrm{W}|\mathrm{F}) = P(\mathrm{F}|\mathrm{W})P(\mathrm{W})/P(\mathrm{F})$. Finally, with the help of the normalization condition $P(\mathrm{F}|\mathrm{W}) + P(\mathrm{R}|\mathrm{W}) = 1$, we arrive at the following expression,

$$\frac{P(\mathrm{W}|\mathrm{F})}{P(\mathrm{W}|\mathrm{R})} = \frac{P(\mathrm{F}|\mathrm{W})P(\mathrm{R})}{P(\mathrm{R}|\mathrm{W})P(\mathrm{F})} = \frac{P(\mathrm{F}|\mathrm{W})P(\mathrm{R})}{(1 - P(\mathrm{F}|\mathrm{W}))P(\mathrm{F})}. \tag{2.64}$$

The ratio $P(\mathrm{R})/P(\mathrm{F})$ is equal to the occurrence number ratio $n_{\mathrm{R}}/n_{\mathrm{F}}$ of forward and reverse measurements. With Eqn. 2.64 the Crooks relation in Eqn. 2.63 can be reformulated as,

$$\ln\left(\frac{P(\mathrm{F}|\mathrm{W})}{1 - P(\mathrm{F}|\mathrm{W})}\right) = \beta(M + W - \Delta G), \tag{2.65}$$

with the constant $M = k_{\mathrm{B}}T \ln(n_{\mathrm{F}}/n_{\mathrm{R}})$, which is related to $C' = \ln\left(\frac{n_1}{n_0}\right)$ in Eqn. 2.62.

The probability of a single measurement $P(\mathrm{F}|\mathrm{W}_i)$ can be formulated as

$$P(\mathrm{F}|\mathrm{W}_i) = \frac{1}{1 + e^{-\beta(M + W_i - \Delta G)}} \tag{2.66}$$

[2] C' will be defined later.

by rearranging the terms in Eqn. 2.65. Furthermore, we have $P(\mathrm{F}|\mathrm{W}) + P(\mathrm{R}|\mathrm{W}) = 1$ and therefore

$$P(\mathrm{R}|\mathrm{W}_i) = \frac{1}{1 + e^{\beta(M+W_i-\Delta G)}}. \tag{2.67}$$

The overall likelihood of obtaining a given measurement, is given by

$$L(\Delta G) = \prod_{i=1}^{n_\mathrm{F}} P(\mathrm{F}|\mathrm{W}_i) \prod_{j=1}^{n_\mathrm{R}} P(\mathrm{R}|\mathrm{W}_j), \tag{2.68}$$

and is maximized by the most likely value of ΔG. In order to find the free energy difference which maximizes the likelihood L. Or, equivalently, the logarithm of L, we take its derivative with respect to ΔG and set it to zero, giving

$$\frac{\partial \ln L(\Delta G)}{\partial \Delta G} = \sum_{i=1}^{n_\mathrm{F}} \frac{\beta}{1 + e^{\beta[M+W_i-\Delta G]}} - \sum_{j=1}^{n_\mathrm{R}} \frac{\beta}{1 + e^{-\beta[M+W_j-\Delta G]}} \stackrel{!}{=} 0. \tag{2.69}$$

Solving this equation for ΔG results in an expression equivalent to Eqn. 2.62. The asymptotic behavior of the two terms in Eqn. 2.69 guarantees the uniqueness of the root, from which we extract ΔG, *i.e.* the maximum likelihood estimate (MLE) of the measured data. The Bennett acceptance ratio yields the free energy difference which, given a series of work measurements in the forward and reverse direction, maximizes the likelihood that these work values would be observed [118].

2.5.3 Multiple State Bennett Acceptance Ratio

The multiple state Bennett Acceptance Ratio method (mBAR) is an extension of the previously introduced BAR method. Unlike BAR, it is not restricted to only two states, but allows instead for the use of data from multiple states. mBAR reduces to BAR whenever only two states are sampled. In the large sampling limit, the mBAR free energy estimation is accurate, meaning it is an unbiased estimator, and it achieves the highest precision by having the lowest possible variance for the available data sampled in equilibrium in multiple states [120]. Shirts and Chodera used results from the field of statistical inference to construct a statistically optimal estimator for computing free energy differences. In the following, the mBAR equation is derived along the lines of [120].

Let us assume we have N_i uncorrelated equilibrium samples from each of the K thermodynamic states within, an NpT ensemble. We furthermore assumed that the states i differ by their potential energy function $\Delta U_i(x)$, where x is a molecular configuration determined by a point in phase space. For sake of simplicity, the reduced potential $u_i(x) = \beta U_i(x)$ is defined.

The configurations $\{x_{i,n}\}_{n=1}^{N_i}$ from state i are sampled from the Boltzmann distribution $p_i(x)$ (see also Eqn. 2.41),

$$p_i(x) = \frac{q_i(x)}{c_i}, \quad c_i = \int_\Gamma dx q_i(x), \tag{2.70}$$

where $q_i(x)$ is the Boltzmann weight $q_i(x) = \exp[-u_i(x)]$. c_i is the unknown partition

function that serves as a normalization constant. We want to construct a estimator for the free energy difference, which is related to the ratio of partition functions (Eqn. 2.56). First we note the identity, for an arbitrary choice of function $\alpha_{ij}(x)$,

$$\begin{aligned} c_i\langle\alpha_{ij}q_j\rangle_i &= \left[\int_\Gamma dx q_i(x)\right]\frac{\int_\Gamma dx q_i(x)\alpha_{ij}(x)q_j(x)}{\int_\Gamma dx q_i(x)} \\ &= \int_\Gamma dx q_i(x)\alpha_{ij}(x)q_j(x) \\ &= \left[\int_\Gamma dx q_j(x)\right]\frac{\int_\Gamma dx q_j(x)\alpha_{ij}(x)q_i(x)}{\int_\Gamma dx q_j(x)} \\ &= c_j\langle\alpha_{ij}q_i\rangle_j. \end{aligned} \tag{2.71}$$

By summing over the index j, and substituting $\langle g\rangle_i$ with the empirical estimator $N_i^{-1}\sum_{n=1}^{N_i} g(x_{in})$ for the expectation value of the observable g, one can reformulate Eqn. 2.71 and obtain a set of K estimating equation.

$$\sum_{j=1}^{K}\frac{\hat{c}_i}{N_i}\sum_{n=1}^{N_i}\alpha_{ij}q_j(x_{in}) = \sum_{j=1}^{K}\frac{\hat{c}_j}{N_j}\sum_{n=1}^{N_j}\alpha_{ij}q_i(x_{jn}) \tag{2.72}$$

The index i runs from 1 to K and gives with Eqn. 2.72 a set of equations, which solutions for all $\hat{c}_i$ yields estimates for the associated partition function c_i. The function $\alpha_{ij}(x)$ is chosen as

$$\alpha_{ij}(x) = \frac{\frac{N_j}{\hat{c}_j}}{\sum_{k=1}^{K}\frac{N_k q_k(x)}{\hat{c}_k}}, \tag{2.73}$$

which is optimal, as it minimizes the variance of the estimator.

Combining Eqns. 2.72 and 2.73, one obtains an expression for the dimensionless free energies,

$$\hat{g}_i = -\ln\sum_{j=1}^{K}\sum_{n=1}^{N_j}\frac{e^{-u_i(x_{jn})}}{\sum_{k=1}^{K}N_k e^{\hat{g}_k - u_k(x_{jn})}}. \tag{2.74}$$

Equation 2.74 must be solved self-consistently. The estimated free energies $\hat{g}_i$ are determined uniquely up to an additive constant, so only differences $\Delta\hat{g}_{ij} = \hat{g}_j - \hat{g}_i$ are meaningful. The mBAR Eqn. 2.74 is consictens with the weighted histogram analysis method (WHAM) in the limit of zero bin width [121].

2.6 One-Dimensional Diffusion Model

The 1-dimensional (1D) diffusion model derived by Hummer in 2005 [122] estimates self-consistently free energies and diffusion coefficients based on MD simulation via Bayesian inference. The central idea is to compare the observed motion along a proposed reaction coordinate Q in an MD simulation with those expected from diffusion dynamics. It is possible to extract kinetic constants from a Markov state model and relate those to the equilibrium probability distributions and free energy differences.

The comparison is done by using Bayesian analysis of the simulated data. Using the

Bayesian theorem, a posterior probability $P(\text{parameter}|\text{data})$ can be written as a likelihood function $P(\text{data}|\text{parameter})$ times a prior $P(\text{parameter})$,

$$P(\text{parameter}|\text{data}) \propto P(\text{data}|\text{parameter})P(\text{parameter}). \tag{2.75}$$

In the context of the Bayesian formalism, a likelihood function $P(\text{data}|\text{parameter})$ entering Eqn. 2.75 can be constructed that gives the probability of observing the motion along Q seen in the simulations. The prior $P(\text{parameter})$ assigns the parameter values certain weights, irrespective of the data. The prior is assumed to be uniform, which reduces the problem to a maximum likelihood estimation and is used to smoothen the transition between neighboring cells. We are therefore able to construct a coarse master equation, characterized by a rate matrix, which contains all necessary information about $G(Q)$ and $D(Q)$.

In the 1D diffusion approach, the master equation is constructed based on Zwanzig's generalized master equation for Newtonian dynamics in configuration space [123]. For a configuration space that is divided into N unit cells, and under the assumption that the dynamics of the system becomes Markovian at sufficiently long times, one can approximate the time evolution of the probability $p_i(t)$ of being in cell i at time t by the Markovian rate equation,

$$\dot{p}_i(t) = \sum_j R_{ij} p_j(t). \tag{2.76}$$

R_{ij} is an element in the i^{th} row and j^{th} column of a rate matrix $\mathbf{R}$ with $R_{ij} \geq 0$ for $i \neq j$, $R_{ii} \leq 0$, and $\sum_i R_{ij} = 0$. Equation 2.76 can be solved in terms of a matrix exponential,

$$p_i(t) = \sum_j \left(e^{tR}\right)_{ij} p_j(0). \tag{2.77}$$

The propagators of the Markovian model are given by

$$p(i,t|j,0) = \left(e^{tR}\right)_{ij} \tag{2.78}$$

and represent the conditional probability to be located in the i^{th} cell at the time t, under the premise of being previously in the cell j at the time 0.

The Smoluchowski equation describes the evolution in time of the probability density along a reaction coordinate Q for a diffusive process. Biscout and Szabo [124] discretized the Smoluchowski diffusion equation in space, and obtained the following rate equations,

$$\dot{p}_i(t) = R_{i,i-1}p_{i-1}(t) - (R_{i-1,i} + R_{i+1,i})p_i(t) + R_{i,i+1}p_{i+1}(t) \quad \text{for} \quad 1 \leq i \leq N, \tag{2.79}$$

where $\dot{p}_i(t)$ is the time derivative of the probability of being in cell i around Q_i at time t.

The free energy surface $G(Q_i)$ can be estimated from the equilibrium probabilities P_i,

$$G(Q_i) \approx -k_{\text{B}}T \ln \frac{P_i}{\Delta Q}, \tag{2.80}$$

where $\Delta Q = |Q_{i+1} - Q_i|$ is the bin width, and $\mathbf{P} = (P_1, \ldots, P_N)^T$ is the vector of equilibrium probabilities, which is an eigenvector of $\mathbf{R}$ with eigenvalue zero, *i.e.* $\mathbf{RP} = 0$. The position-

dependent diffusion coefficients are related to the rate matrix and its equilibrium probability according to

$$D_{i+1/2} \approx \Delta Q^2 R_{i+1,i} \left(\frac{P_i}{P_{i+1}} \right)^{1/2}. \tag{2.81}$$

The diffusion coefficient is calculated at the bin center $D_{i+1/2} = D[(Q_i + Q_{i+1})/2]$, where i and $i+1$ are two neighboring cells. The rate matrix satisfies detailed balance since Eqn. 2.81 is symmetric in respect to exchange of i and $i+1$, $R_{i+1,i}/R_{i,i+1} = P_{i+1}/P_i = e^{-\beta[G(Q_{i+1})-G(Q_i)]}$. In order to solve for $G(Q)$ and $D(Q)$ in Eqns. 2.80 and 2.81, one has to estimate the rate coefficients for transitions between adjacent cells along Q.

In the context of Bayesian inference, we can use the following likelihood function to estimate the rate matrix $\mathbf{R}$ from short MD simulations,

$$L = \prod_{\alpha} p(i_\alpha, t_\alpha | j_\alpha, 0) = \prod_{\alpha} \left(e^{t_\alpha \mathbf{R}} \right)_{i_\alpha j_\alpha}. \tag{2.82}$$

It uses as input a series of observations i_α, separated by equispaced time steps t_α.

The detailed balance condition halves the number of rate constants entering $\mathbf{R}$. Furthermore, there are only $N-1$ independent equilibrium probabilities P_i, because of the normalization condition $\sum_i P_i = 1$, which leaves us with $N(N-1)/2$ free rate parameters in $\mathbf{R}$. The rate matrix $\mathbf{R}$ thus has the structure

$$R_{ij} = \begin{cases} R_{ij} & \text{if} \quad i > j \\ -\sum_{l(\neq)i} R_{li} & \text{if} \quad i = j \\ R_{ji} \frac{P_i}{P_j} & \text{if} \quad i < j. \end{cases} \tag{2.83}$$

Maximizing the likelihood $L(\text{parameter}|\text{data})$ via Monte Carlo simulations in parameter space, gives the most likely equilibrium free energies $g_i = -\ln P_i$, and rate coefficients R_{ij}.

Chapter 3

Aims of the Study

The crystal structure of the Na^+/H^+ antiporter MjNhaP1 from archaeal *M. jannaschii* was recently resolved by Paulino *et al.* in 2014 [34]. Using biochemical characterization, the pH activity profile and transport rates were measured for the wildtype and for specific single mutations, to elucidate the physiological meaning of individual residues [34]. Electrophysiological measurements confirm the alternating access mechanism and determined that H^+ and Na^+ are competing for the same binding site [32, 125].

However, the molecular mechanism of ion binding and ion selectivity, remains unclear. In this thesis the protein dynamics was studied to understand the mechanism of Na^+ binding by applying multiple theoretical approaches, in particular MD and QM/MM simulation, QM cluster calculation, and free energy calculation. These simulations contributed to the understanding of the transport mechanism of MjNhaP1 on the molecular level.

The aims of this work was to:

- Identify the ion binding site in comparison to the Tl^+ bound PaNhaP homologe structure [35].
- Optimize the geometry of the binding site with Na^+ bound.
- Model the outward-open conformation of MjNhaP1 based on the low resolution (6 Å) cryo-electron microscopy (cryo-EM) map of MjNhaP1 [34].
- Probe the ion binding and conformational stability and protein dynamics of the inward- and outward-open states of MjNhaP1 in MD simulation and QM/MM simulation.
- Estimate the free energy of ion binding of MjNhaP1 by applying alchemical transformation simulation.
- Determine the influence of individual residues close to the binding site to the Na^+ selectivity.
- Compare the sequence of Na^+/H^+ antiporters, which are selective for Na^+ with those that are selective for K^+.
- Based on the simulations, suggest mutations to manipulate the ion selectivity of MjNhaP1 towards K^+.

Complex I is the last resolved structure among the respiratory complex systems (Baradaran *et al.* in 2013 [16]). In recent years, many experimental and theoretical studies focused on complex I to reveal its mechanism [16,53,54,56,74–80,126]. However, a detailed molecular mechanism of the coupling processes between reduced ubiquinol and the proton pumps is still unknown. The work of respiratory complex I, presented here, focuses on studies of how the ubiquinone substrate binds to the enzyme. In these simulations, the dynamics of ubiquinone/ol (Q) in the Q tunnel was studied using extensive atomistic MD simulation of Q_{ox} and QH_2 in the *ca.* 40 Å long Q tunnel. The key questions for this thesis can be outlined as: are there multiple binding sites in the Q tunnel? On what timescales does Q migrate through the binding-channel? Is a single ubiquinone sufficient for an efficient complex I function?

These questions were studied by using the following approaches:

- To obtain a rough estimate of the binding properties and interactions of the Q in the Q tunnel, the bound ubiquinone was pulled out of the binding pocked by applying steered-molecular dynamics simulation.
- Based on multiple short equilibrium MD simulations at different positions of Q headgroup in the Q tunnel, the dynamics of Q in the Q tunnel was explored.
- To identify preferred binding positions and stabilizing interactions of Q with the protein environment, the MD simulations were in detail analyzed.
- Extract free energy and distance dependent diffusion profiles by applying the one-dimensional diffusion model.
- Estimate the forward and backward round-trip time of Q migrating through the entire Q tunnel based on the calculated diffusion coefficients.

A subtopic in the context of complex I arose from a collaboration with Prof. Zickermann[1]. By applying theoretical simulation (MD and QM/MM) and QM calculation, the structural stability and function of accessory subunit NUMM in *Y. lipolytica* was studied.

[1]Structural Bioenergetics Group, Institute of Biochemistry II, Medical School, Goethe University, 60438 Frankfurt am Main, Germany

Chapter 4

Computational Details

The challenge for an accurate theoretical description of transmembrane protein transporters is their large system size of several 100,000 atoms. In such models, we aim to capture the biologically relevant surrounding, including the lipid bilayer and the water with physiological ion concentrations. Due to the large system size we are limited to highly efficient force fields to cover at least some parts of the protein dynamics. On the other side, we are interested in having a close look at the binding site with high accurate methods. To cover the crucial electronic effects of the bound ion and the binding site environment, we need to apply quantum chemical methods. Another challenge in describing biological systems is to appropriately model protonation states and dynamics, which is essential for Na^+/H^+ antiporters and complex I. This requires a good understanding of the performance of various computational methods when applied to ion transport.

4.1 MD Simulations of MjNhaP1

To construct a MD simulation box, the crystal structure of MjNhaP1 (PDB ID: 4CZB) [34] resolved as a homodimer at pH 8 was embedded in 1-palmitoyl-2-oleoylphosphatidylethanolamine (POPE) lipid bilayer and solvated with 150 mM NaCl solvation to mimic physiological conditions. The initial box size was 156 Å × 126 Å × 105 Å. The protonation state of titrable amino acides was assigend based on electrostatic calculations using PROPKA [127], [128]. According to these calculations, the buried acidic residues Asp-93, Glu-344, and Glu-401, were protonated in both monomers. The MD simulations were performed with NAMD2.9 - 2.11 [129], in combination with CHARMM36 force field [130]. The force field parameters for the Na^+, K^+, and Cl^- ions, were taken from [130], [131]. The MD simulation was performed with Langevin dynamics using a 2 fs time step, in an NpT ensemble, with 186383 particles, at 310 K, and at 1.01325 bar. For pressure coupling a Langevin piston was used, with flexible cell setting. The particle-mesh Ewald (PME) approach [132] was applied to describe the long-range electrostatics. 1 μs of equilibrium MD simulation of the inward-open conformational state was performed based on the crystal structure.

To model an outward-open state, MDFF was applied to the crystal structure as starting structure and the cryo-EM map (PDB ID: 4D0A) [34] resolved with 6 Å resolution at pH 4 as target. In the first step the dimer crystal structure of MjNhaP1 (PDB ID: 4CZB) [34] was structurally aligned to the outward-open structure modelled previously on the cryo-EM map (PDB ID: 4D0A) [34] with MultiSeq module of visualization molecular dynamics (VMD) [133]. In the flexible fit a scaling factor $\zeta = 0.3$ kcal mol^{-1} was used. To preserve

the secondary structure, harmonic restrains were applied to ψ and ϕ angels for amino acid residues in helices and in sheets with a spring constant $k_\mu = 200$ kcal mol^{-1}rad^{-2}, as well as restrains for hydrogen bonds of residues on these secondary structure elements with a spring constant $k_\mu = 200$ kcal mol^{-1}Å^{-2}. The plugins “cispeptide" and “chirality" were used to restrain the cis and trans peptide bonds and restrain chiral centers, respectively, ensuring proper peptide stereochemisty. 2 ns of flexible fitting were followed by 200 steps of minimization. The MD trajectories were analyzed and visualized in VMD [133]. The final structure of the MDFF simulation was used as initial structure for MD simulations of an outward-open state of MjNhaP1. The MD simulation were performed with NAMD2.9-2.11 [129] and the CHARMM36 force field [130]. The dimer structure was embedded into a POPE lipid bilayer and solvated with 150 mM NaCl. The initial box size was 161 Å × 130 Å × 98 Å. Langevin dynamics with 2 fs time steps was used to simulate a NpT ensemble with constant temperature at 310 K, constant pressure at 1.01325 bar, and constant total number of atoms in the simulation box with 177581 particles. Langevin piston was applied for temperature control and flexible cell setting was used. To cover long-range electrostatic interactions, the PME approach [132] was used. The prepared system was equilibrated for 68 ns with restraints to the final structure of the MDFF simulation, which was removed in the subsequent 90 ns free MD simulations. The outward-open state of MjNhaP1 was simulated for 1 μs.

The dimer structure in the inward- and outward-open state of MjNhaP1 was also simulated with a bound K^+ in the binding sites. After 100 ns of equilibrium MD simulation of MjNhaP1, only the bound Na^+ in the binding sites were substituted by K^+ in both monomers. Subsequently 1 μs of equilibrium MD simulation was performed in the inward-open state and 500 ns in the outward-open state with bound K^+ in the binding site.

4.2 Studies of the Ion Binding Site in Na^+/H^+ Antiporters

4.2.1 Comparison of the Ion Binding Site from Homologous Proteins

For a comparison of the four homologs of known crystal structures, a sequence and structural alignment of MjNhaP1 (PDB ID: 4CZB), PaNhaP (PDB ID: 4CZ8), EcNhaA (PDB ID: 4ZCD), and TtNapA (PDB ID: 4BWZ) was performed. For the structural alignment the MultiSeq module of VMD [133] was used. The sequence alignment was performed with the membrane protein specific version of multi sequence alignment web server position specific iterative/transmembrane (PSI/TM)-Coffee [134,135], using the homology extension UniRef100. The sequence alignment was visualized using jalview [136].

4.2.2 QM Cluster Calculations and QM/MM Simulations of MjNhaP1

The ion binding site of monomeric model of MjNhaP1 was also studied by applying hybrid quantum mechanic/classical mechanic (QM/MM) simulation. The QM/MM simulations were initiated based on 100 ns of equilibrium MD simulations of MjNhaP1 in the inward- and outward-open state with bound Na^+ and bound K^+ in the binding site. A reduced model with *ca.* 10100 atoms of an MjNhaP1 was build consisting a monomer and a surrounding layer of water and ions of 5 Å. The QM region comprises the ion binding site, the bound

ion (Na^+ or K^+), the surrounding residues (Thr-131, Asp-132, Ser-157, Asn-160, and Asp-161), and three distance wise closest water molecules to the bound ion. The QM region was truncated using link atoms (Fig. 4.1). The basis sets def2-TZVP and def2-SVP [137] was used for the alkali-metal ion and for all other atoms in the QM region, respectively, in combination with the hybrid functional B3-LYP [138], [139]. The rest of the reduced model is described classically (MM), using CHARMM27 force field [140], [130]. In the QM/MM simulations the temperature was set at 310 K using a 1 fs time step. In total, four QM/MM simulations were performed, each 7 ps long, of the inward- and outward-open state with bound Na^+ and K^+ in the binding site, respectively. For the QM/MM simulation TURBOMOLE-CHARMM [141] interface was used.

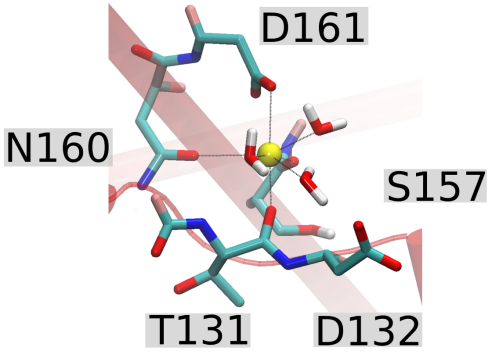


Figure 4.1: QM region in QM/MM simulation of the inward-open state chain A with bound Na^+. The atoms included in the QM region are represented in licorice. The link atoms are shown in pink.

QM cluster calculations were performed of the ion binding site of the crystal structures of four homologous Na^+/H^+ antiporter: MjNhaP1 (PDB ID: 4CZB), PaNhaP (PDB ID: 4CZ8), EcNhaA (PDB ID: 4ZCD), and TtNapA (PDB ID: 4BWZ). The QM cluster models comprise the ion in the binding site with coordinating residues Thr-131, Asp-132, Ser-157, Asn-160, and Asp-161, for MjNhaP1; Glu-73, Thr-129, Asp-130, Ser-155, Asn-158, and Asp-159, for PaNhaP; Thr-132, Asp-133, Ala-160, Asp-163, and Asp-164, for EcNhaA; and Thr-126, Ser-127, Ala-153, Asp-156, and Asp-157, for TtNapA. The QM cluster model was saturated with hydrogen atoms. To mimic the protein structure, the backbone atoms were kept fix. The geometry optimizations were performed with the functional B3-LYP [138], [139] in combination with def2-TZVP [137] basis set. The QM cluster calculations were performed using TURBOMOLE [142].

4.3 Calculations of Ion-Binding Free Energies

To estimate the free energy of Na^+ and K^+ binding to the inward- and outward-open state of MjNhaP1, alchemical free energy perturbation (FEP) calculations were performed. The alchemical transformation between Na^+ and K^+ were performed using the single-topology approach. The Lennard-Jones potential of single positively charged fictitious particle is linearly scaled by the parameter λ in 41 steps from Na^+ (ϵ = -0.0469 kcal mol^{-1} and σ = 1.41075 Å, corresponds to $\lambda = 0$) to K^+ (ϵ = -0.0870 kcal mol^{-1} and σ = 1.76375 Å, corresponds to $\lambda = 1$) [131, 143]. In this way the system is minimally perturbed, and there

is no need to restrain the emerging or vanishing particle unlike the dual-topology approach. Each of the 41 λ windows was equilibrated for 0.5 ns, followed by 1 ns for data collection. The alchemical transformation in forward and subsequently in backward direction, meaning transforming Na^+ to K^+ and subsequently K^+ back to Na^+, followed the same protocol. Ten independent forward and backward alchemical-transformation simulations were initiated based on extracted snapshots from an equilibrium 1 μs MD simulation of bound Na^+ in the inward- and outward-open state. For each λ window 5000 configurations at intervals of 0.2 ps were saved, and the energy differences of each configuration to the neighboring λ window were determined. The free energy difference between Na^+ and K^+ is estimated with exponential averaging (EXP), Bennett acceptance ratio (BAR) and multistate Bennett acceptance ratio (mBAR) methods. To probe possible hysteresis effects, the free energies were also calculated using data only from the forward and backward paths. FEP calculations were performed on the B chain of dimeric MjNhaP1, where the ions are coordinated fully in both access states, whereas for the inward-open state with bound Na^+, the FEP calculations were performed on chain A. The statistical uncertainty was estimated with the standard error of the mean, based on 10 independent forward and backward calculations. The same FEP protocol was used to calculate the difference in ion-solvation free energy in bulk aqueous solution. A single ion was introduced in a water box of the size 40 Å $\times$ 40 Å $\times$ 40 Å. Since the charge state is not changed during the transformation between Na^+ and K^+, finite-size effects are kept at a minimum [144], except for possible dielectric effects that could arise from the vicinity to the low-dielectric protein interior and the membrane environment [145, 146].

4.4 Ubiquinone/ol Dynamics

4.4.1 MD Simulations

The full atomistic and equilibrated model system of complex I for an MD simulation was build and delivered as starting point for my work, from AG Kaila. The system setup was described and published earlier [72]. The simulations were based on the crystal structure from *T. thermophilus* (PDB ID: 4HEA), set it into an 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholin (POPC) lipid membrane and added water with physiological NaCl-salt concentration. The full atomistic model comprises 809413 atoms. The force field parameters for ubiquinone/ubiquinol, and FeS cluster were derived from quantum chemical calculation on the level of DFT as described in previous work [72], [73]. The MD simulation were performed with NAMD2 [129] with the force field CHARMM27 [140], [130]. In an NpT ensemble the temperature and pressure were kept constant at 310 K, and at 1 atm. The time step in the MD simultaion was set to 2 fs and the sharkH algorithm was used to restrain all hydrogen bonds. To describe long range electrostatic interaction the PME method was used. The simulated trajectories were analyzed using VMD [133].

4.4.2 Steered MD Simulations

Two independent steered molecular dynamics (SMD) simulations were performed for with ubiquinone and ubiquinol bound in the Q-binding tunnel of complex I. The SMD simu-

lations were initiated based on 100 ns of equilibrium MD simulation of the full atomistic system setup (described above) in which a ubiquinone was bound in the primary binding site in hydrogen bond distance to Tyr-87_{Nqo4}. SMD simulation were applied to pull the bound ubiquinone out into the membrane. The SMD simulation with bound ubiquinol was prepared based on the same starting structure, but the bound ubiquinone was substituted by an ubiquinol. Tyr-87_{Nqo4} and His-38_{Nqo4} were deprotonated, in order to model the proton transfer from these residues to the reduced ubiquinone [72]. The ubiquinone/ol was pulled out of the protein binding tunnel in direction to the membrane exit with constant pulling force of 0.5 Å/ns and a force constant of k = 100 kcal mol^{-1}Å^{-2}. The force was applied to the center of mass of the ubiquinone/ol molecule. The 100 C_α atoms of the transmembrane helices of subunit Nqo10 were fixed in their position with a force constant k = 2 kcal mol^{-1}Å^{-2} to stabilize the protein and lipid membrane in the box. The SMD were performed with NAMD2 [129] with CHARMM27 [140], [130] force field without application of temperature and pressure control.

4.4.3 Dataset of Short MD Simulations of Ubiquinone/ol

Snapshots were extracted every 2 ns from the SMD simulations, in order to initiate short (11 ns) equilibrium MD simulations, with different ubiquinone/ol headgroup positions in the Q tunnel. For ubiquinone, initially 38 × 11-ns MD simulations were started. To improve sampling in rarely visited regions of the Q tunnel 23 additional 11-ns MD simulations were performed. And finally, 4 × 11-ns MD simulations were performed in which the ubiquinone headgroup is in a hydrogen bond distance to His-38_{Nqo4} and Tyr-87_{Nqo4}, using a structure based on previous work [72]. The dataset of ubiquinone comprises in total 65 × 11 ns MD simulations.

For ubiquinol, 38 snapshots were extracted directly from the SMD simulation with ubiquinol, due to a 2 Å spacing of the ubiquinol headgroup position in the Q tunnel. Based on these snapshots, 38 × 11 ns MD simulations were initiated. 14 additional 11-ns MD simulations were performed, to fill the sampling gaps in rarely visited regions in the Q tunnel. Finally, 5 × 11-ns MD simulations were added, based on the same snapshots as in the case of ubiquinone, in which the ubiquinone headgroup is initially in a hydrogen bond distance to His-38_{Nqo4} and Tyr-87_{Nqo4} but with substituted ubiquinone to ubiquinol and accordingly deprotonated His-38_{Nqo4} and Tyr-87_{Nqo4}. The complete dataset of ubiquinol comprises 57 × 11-ns MD simulations. The last 10 ns of all 11-ns MD simulations were used for further analysis.

4.4.4 Calculations of PMF and Diffusion Coefficients Profiles of Ubiquinone/ol

Free energy (potential of mean force, PMF) profiles and diffusion profiles were calculated of ubiquinone and ubiquinol bound in the Q tunnel of complex I. An in house implementation of the one-dimensional diffusion model [122] was applied based on 65 × 11 ns MD simulations of Q_{ox} and 54 or 57 × 11-ns MD simulations of QH_2.

4.4.5 The Definition of Two Reaction Coordinates

To test the sensitivity of the reaction coordinate, the potential of mean force (PMF) with respect to different reaction coordinates was calculated. The first reaction coordinate is defined as, the Tyr-87$_{\mathrm{Nqo4}}$-Q distance between the hydroxyl oxygen of Tyr-87$_{\mathrm{Nqo4}}$(OH) and the distance-wise closest carbonyl/hydroxyl oxygen of Q_{ox}/QH_2 headgroup. The second reaction coordinate is defined as, the projected Tyr-87$_{\mathrm{Nqo4}}$-Q distance between the C_α of Tyr-87$_{\mathrm{Nqo4}}$ and the distance-wise closest carbonyl/hydroxyl oxygen of Q_{ox}/QH_2 headgroup, projected on the SMD pulling vector.

4.5 QM and QM/MM Simulations of Accessory Subunit NUMM of *Y. lipolytica*

Due to the absence of structural data of NUMM subunit, a homology model was generated as starting structure of the quantum chemical calculations. The C-terminal domain of the NUMM subunit (L69-H137) was constructed based on an NMR structure of the RHOS4_26430 protein (PDB ID: 2JVM) using the MODELLER program [147]. The sequence identity is 27.5%. The homology structure of NUMM was provided by Prof. Zickermann[1].

The QM region consists of the residues Cys-97, Cys-125, Cys-128, and His-110 truncated at the C_α position and an added Zn^{2+}-ion. Link atoms were introduced between the C_α and C_β atoms for each residue. The hybrid functional B3-LYP [138], [139] in combination with the basis set def2-TZVP [137], for the Zn, and the basis set def2-SVP [137] for other atoms in the QM region was used. Identical to the QM region, a QM cluster model was build comprising the Zn^{2+} ion and the four coordinating residues (Cys-97, Cys-125, Cys-128, and His-110) cut at the C_α position, and saturated with hydrogen atoms. Apart from the QM region, the rest of the NUMM subunit was treated classically (MM), using the force field CHARMM27 [140], [130]. The QM/MM simulation was performed with temperature and pressure coupling at 310 K and 1 atm using a 1 fs time steps. Two independent QM/MM simulations were performed with and without weak harmonic restrains of 1 kcal $\mathrm{mol}^{-1}\mathrm{\AA}^{-2}$ on the C_α atoms. For QM/MM simulation TURBOMOLE-CHARMM [141] was used. Two geometry optimizations of the QM cluster model were performed with fixed positions of C_α atoms. One geometry optimization was performed in vacuum. The second geometry optimization was performed with the conductor-like screening model (COSMO) [148], in order to model the protein environment as polarizable low-dielectric medium with $\epsilon = 4$. The QM cluster calculations were performed using TURBOMOLE [142].

[1]Structural Bioenergetics Group, Institute of Biochemistry II, Medical School, Goethe University, 60438 Frankfurt am Main, Germany

Chapter 5

Ion Binding and Selectivity in Na^+/H^+ Antiporters

5.1 The Outward-Open State of MjNhaP1

An outward-open structure of MjNhaP1 was modeled by applying the MDFF method based on the inward-open crystal structure and a medium resolution cryo-EM map of the outward-open state [34] (Fig. 5.1). The inward-open crystal structure fits into the cryo-EM map by small changes. The root-mean-square deviation (RMSD) obtained during the 2 ns MDFF simulation, (Fig. 5.2 C) reveals that the MDFF simulation is converged by reaching a plateau. The structural changes during the simulation are small (RMSD < 3 Å) (Fig. 5.2 A, B).

The Pearson's correlation coefficient or cross-correlation coefficient (*ccc*) between the structure (S), converted to a map of the same resolution, and the experimental map (E) was evaluated to measure the similarity between the data and the atomistic model. The *ccc* is given by,

$$ccc = \frac{\langle (S - \langle S \rangle)(E - \langle E \rangle) \rangle}{\sigma_S \sigma_E}, \tag{5.1}$$

where $\langle S \rangle$ and $\langle E \rangle$ correspond to the average voxel values of the simulated and experimental maps, respectively, and σ_S and σ_E are their standard derivations. The *ccc* is normalized from -1 to 1.

The crystal structure has a *ccc* value of 0.32, and after the MDFF simulation, the *ccc* value increases to 0.79. A structural alignment of the crystal structure and the final fitted model reveals that the largest changes of *ca.* 3 Å RMSD take place at the binding domains (Fig. 5.2 A, B).

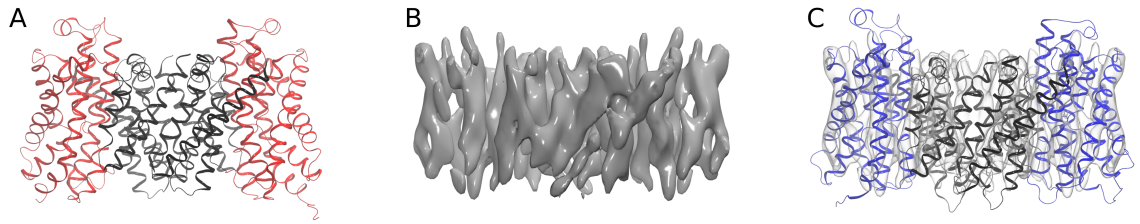


Figure 5.1: A) Crystal structure (PDB ID: 4CZB), B) cryo-EM map of 6 Å resolution represented at a sigma value of 110, and C) the fitted outward-open structure of MjNhaP1 into the cryo-EM map.

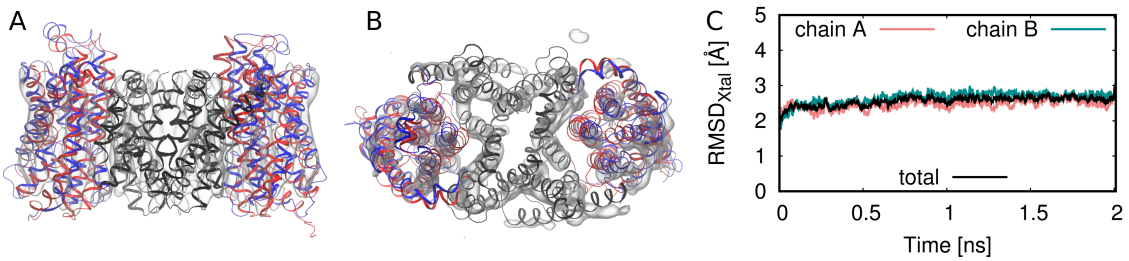


Figure 5.2: A) Structural alignment of the fitted model (blue) with crystal structure (red), and the cryo-EM map (gray transparent) in side view, and (B) top view. C) RMSD in Å relative to the crystal structure calculated for the 2 ns MDFF simulation for chain A (light coral), chain B (dark cyan), and for both monomers ("total"; black). Only the backbone atoms of the helices were taken into account in the RMSD calculation.

Figure 5.3: Sequence alignment of four homologs of MjNhaP1. From top to bottom: MjNhaP1, PaNhaP, EcNhaA, and TtNapA. The sequence alignment was performed with the membrane protein specific version of multiple sequence alignment web server PSI/TM-Coffee [134, 135], using the homology extension UniRef100. The sequence alignment was visualized using jalview [136]. The residues are colored based on their conservation (light blue = 50 % conservation, blue = 75 %, and dark blue = 100 %). Conserved residues of the ion binding site are marked with red boxes.

5.2 The Ion Binding Site

5.2.1 Structural Comparison of Homologous Proteins

In literature, there are several suggestions about the exact location of the ion binding site in Na^+/H^+ antiporters [29, 30, 33–35], but a bound Na^+ ion has not been resolved in any available crystal structure. A strong proof for the putative ion binding site was obtained by the Tl^+-bound structure of PaNhaP [35]. To deduce the Na^+ binding site based on the Tl^+-bound structure of PaNhaP in the homologous protein, a sequence and a structural alignment of the four antiporter crystal structures was performed, shown in Fig. 5.3 and Fig. 5.4.

The sequence alignment (Fig. 5.3) reveals that the residues, which are mainly responsible for the ion binding in MjNhaP1 (Thr-131, Asp-132, Asp-161), are conserved. The residue Ser-157 is partly conserved in MjNhaP1 and PaNhaP, whereas the electrogenic antiporters EcNhaA and TtNapA, have an alanine at this position.

In the structural alignment (Fig. 5.4), the protein structures of the four homologs antiporters aligned well. The residues of the binding sites are superimpose nearly on top of each other, with an exception of the binding site in TtNapA. The TtNapA structure was resolved in the outward-open state, and the binding site residues were shifted towards the outside [33].

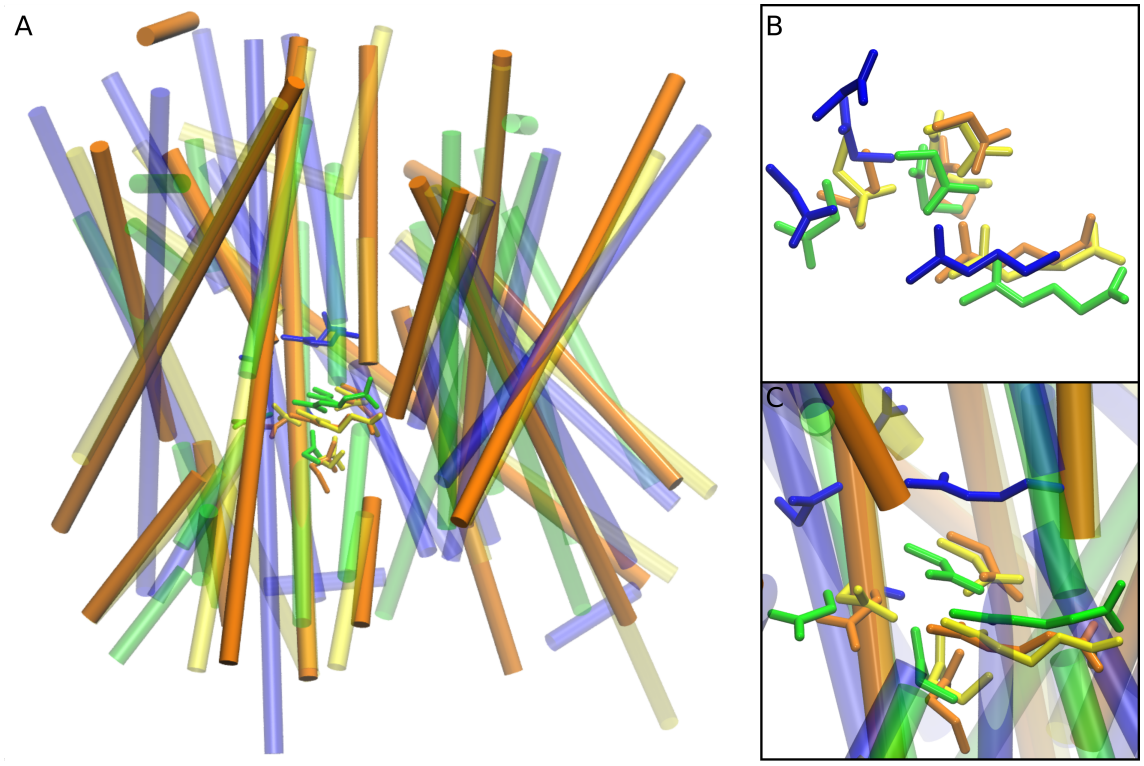


Figure 5.4: Structural alignment of four homologs of MjNhaP1 with MultiSeq alignment tool of VMD [149]. The secondary structure elements of MjNhaP1 (orange), PaNhaP (yellow), EcNhaA (green), and TtNapA (blue), are shown in the transparent cartoon representation. The residues of the binding site (Thr-131, Asp-132, Ser-157, Asn-160, and Asp-161) are represented as licorice in the respective color. A) Alignment of one monomer from four homologs in side view. B) Close-up of the binding site without protein surrounding in top view. C) Close-up of the binding site with protein helices in background in side view.

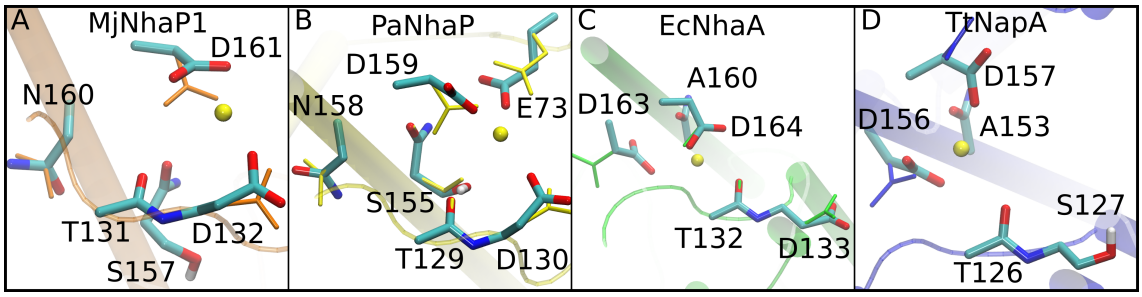


Figure 5.5: Geometry optimized binding sites with fixed protein backbone atoms represented in licorice. The protein crystal structures are shown in background. A) MjNhaP1 in orange, B) PaNhaP in yellow, C) EcNhaA in green, and D) TtNapA in blue. The starting crystal structure conformation for the optimization are shown in thin lines in the respective color.

Due to the high similarity of the binding site region of the four Na^+/H^+-antiporter homologs in the sequence and structural alignments, we can assume that the Tl^+ binding site mimics the ion-binding site in the other Na^+/H^+-antiporters.

5.2.2 Geometry Optimization of the Ion Binding Sites in Homologous Proteins

To probe the stability of Na^+ binding in the putative ion binding site, quantum-chemical geometry optimizations were performed. A Na^+ was placed in the binding sites of the four crystal structures of MjNhaP1, PaNhaP, EcNahA, and TtNapA, between Thr-131 and Asp-161 (MjNhaP1 nomenclature), in a position comparable to that of Tl^+ in PaNhaP. Geometry optimization at the DFT level of the Na^+ loaded binding sites in the conformation of the crystal structures of the four homologs was performed. The geometry optimizations converged in all four cases with minimal rearrangements of the ligands (Fig. 5.5). This finding supports the idea that the putative binding sites can stably bind a Na^+ ion. To assess the ability of MjNhaP1 to bind an ion, MD simulations were performed with bound Na^+ and K^+ in the inward- and outward-open states of MjNhaP1.

5.2.3 Characterization of the Ion Binding Site

Equilibrium MD simulations of the inward- and outward-open states

In a 1 μs MD simulation of the inward-open conformation of MjNhaP1, a continous water and ion accessibility of the binding site to the inside was observed (Fig. 5.6 A). The RMSD of the protein structure during MD simulations relative to the crystal structure (Fig. 5.7) indicates that the inward-open conformation remains stable during the entire simulation. The ion binding takes place as suggested previously [34], or in the related Na^+/H^+-antiporter [30]. The ion binding residues predicted by the MD simulations are Thr-131, Ser-157, Asn-160, and Asp-161. We observe a stable and continuous Na^+ binding to the intact binding sites in both monomers (Fig. 5.8, 5.9). In the 1 μs MD simulation, which was started from the crystal structure, several features were captured for a physiologically meaningful inward-open state.

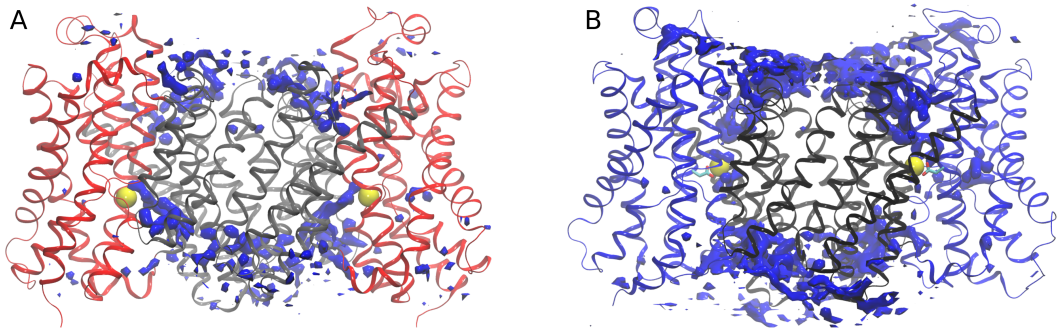


Figure 5.6: Water occupancy in 1 μs MD simulation in (A) the inward-open state and (B) the outward-open state of MjNhaP1. The blue surface represents the fractional occupancy level of 0.5 and 0.45 for the inward- and outward-open state, respectively, indicating the grid points at which at least 50 % or 45 % of the time a water molecule reside, averaged over 1 μs MD trajectory.

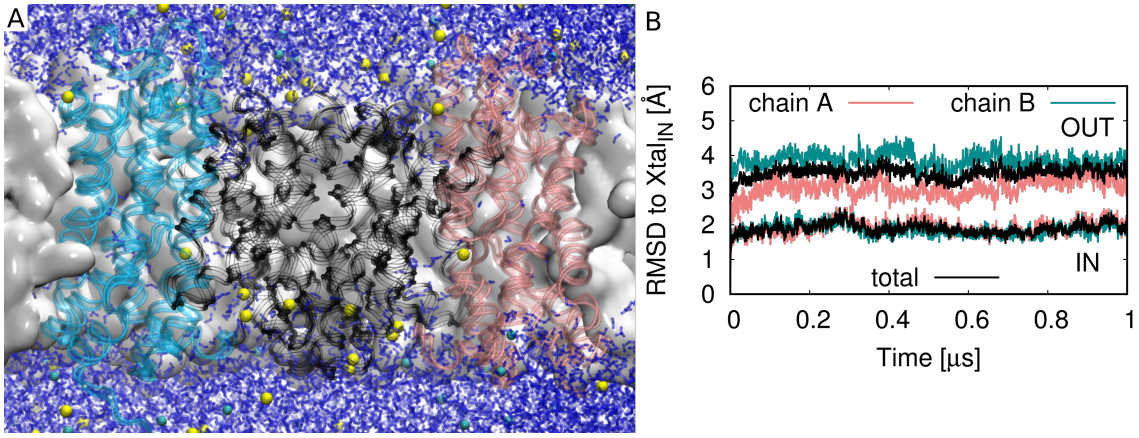


Figure 5.7: Atomistic MD simulation setup of MjNhaP1 embedded in a POPE lipid membrane. A) Snapshot of inward-open conformational state after 1 μs of MD simulation. The binding site domains of chains A and B are shown in light coral and dark cyan, respectively. The interface domains are shown in black. The membrane is represented as a white surface and the water and ions are drawn as sticks and spheres, respectively (Na^+: yellow; Cl^-: cyan). B) RMSD in Å calculated for the 1 μs MD simulations of the inward-open and outward-open conformational states in reference to the inward-open crystal structure. Results are shown for chain A (light coral), chain B (dark cyan), and for both monomers ("total"; black). Only the backbone atoms of the helices were taken into account in the RMSD calculation.

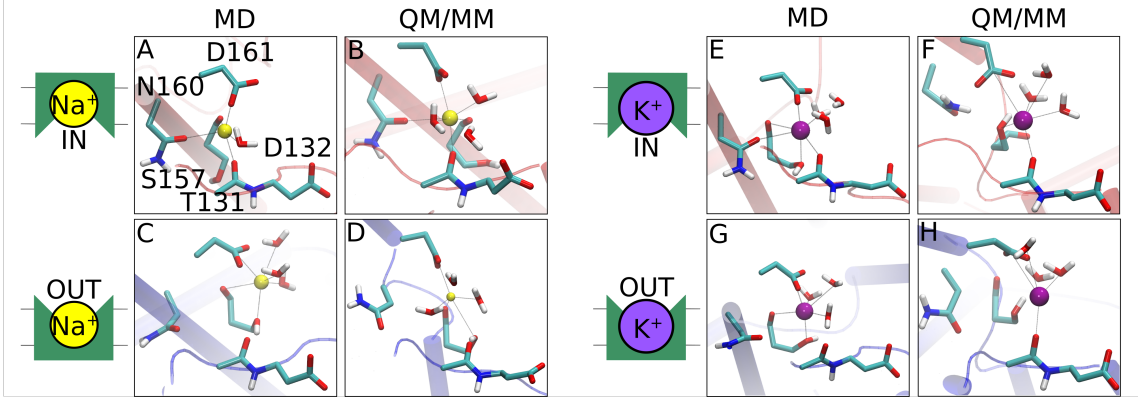


Figure 5.8: Comparison of the ion binding sites from MD simulations (left) and QM/MM simulations (right) with bound Na^+ and K^+ in inward- and outward-open conformational states. Binding site residues Thr-131, Asp-132, Ser-157, Asn-160, and Asp-161 are represented in licorice. (A, B) Chain A in the inward-open state with bound Na^+. (C, D) Chain B in the outward-open state with bound Na^+. (E, F) Chain B in the inward-open state with bound K^+. (G, H) Chain B in the outward-open state with bound K^+.

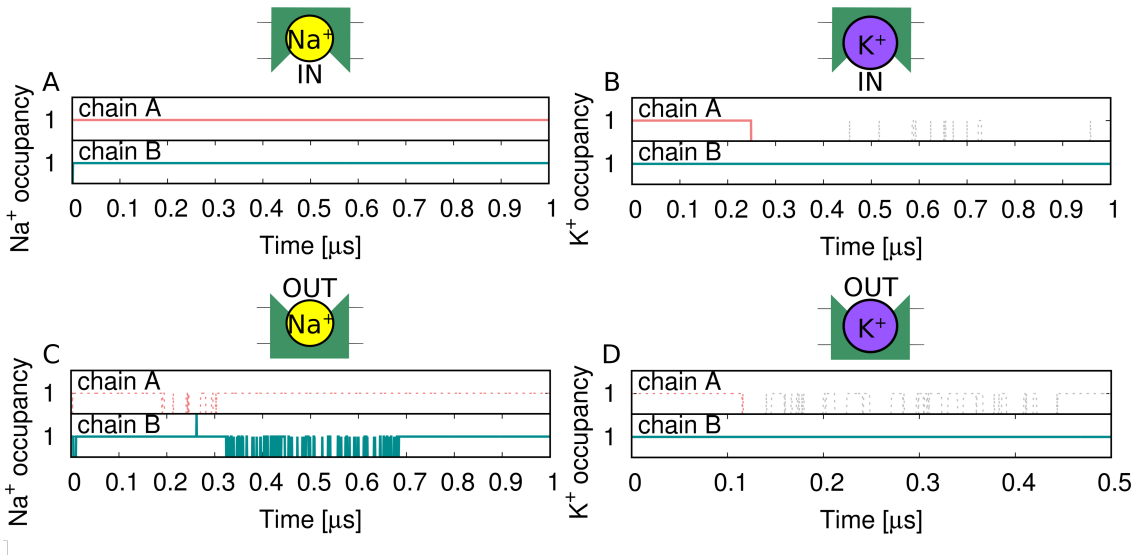


Figure 5.9: Ion occupancy in the binding sites of chain A and chain B of MjNhaP1 in light coral and dark cyan, respectively. A) Ion occupancy in the inward-open state and (C) the outward-open state with bound Na^+. B) Ion occupancy in the inward-open state and (D) the outward-open state with bound K^+. Dashed lines indicate that the ion has only contact to Asp-161 at the entry of the ion binding site and not deeply bound in the binding pocket 5.13. (B, D) Gray dashed lines indicate that Na^+ comes close to Asp-161 of chain A in inward- and outward-open simulations in which originally a K^+ ion was occupying the binding site. Ions are considered to be bound within 4.5 Å distance to Asp-161.

Table 5.1: Average (avg) and standard derivation (std) of the distance between the residues (Asp-161, Thr-131, Ser-157(O: carbonyl oxygen), Ser-157(OH: side chain hydroxyl oxygen), and Asn-160) and the bound Na^+, bound K^+ to the inward-open state in 1 μs MD simulations and bound Na^+ in 7 ps QM/MM simulation.

	MD	MD	QM/MM
Residue	Na^+ avg [Å] ± std	K^+ avg [Å] ± std	Na^+ avg [Å] ± std
Asp-161	2.28 ± 0.09	2.57 ± 0.11	2.35 ± 0.15
Thr-131	2.26 ± 0.09	2.61 ± 0.12	2.31 ± 0.11
Ser-157(O)	2.26 ± 0.10	2.65 ± 0.14	2.38 ± 0.16
Ser-157(OH)	5.18 ± 0.10	3.17 ± 1.00	5.23 ± 0.35
Asn-160	2.45 ± 0.24	3.10 ± 0.46	3.46 ± 0.29

The MD simulation of the outward-open state of MjNhaP1 was started from the flexible fitted structure into the cryo-EM map [34] (see above). After a careful equilibration phase of more than 60 ns long, a 1 μs MD simulation was performed of the outward-open state of MjNhaP1. The RMSD measurement of the 1 μs MD trajectory indicates a stable outward-open conformational state (Fig. 5.7). In the entire trajectory, the binding site is accessible for water and ions from the outside (Fig. 5.6). The ion binding to the binding sites of the two chains takes place in a different way. Na^+ forms a stable interaction with the binding site of chain B (Fig. 5.8), whereas several Na^+ binding and unbinding events take place to the binding site in chain A (Fig. 5.9). The modeled outward-open state exhibits several features of a physiological meaningful outward-open state.

To gain a better understanding of the Na^+ selectivity of MjNhaP1 and its disability to transport K^+, equilibrium MD simulations were performed with bound K^+ to the binding site in the inward- and outward-open state of MjNhaP1. Comparing both Na^+ and K^+ ions bound to the binding site, we found that the ion binding sites in the outward-open state are shallower than those of the inward-open state (Fig. 5.8). Weaker ion-protein interactions in the outward-open state are compensated with increased number of water ligands coordinating the bound ion (Fig. 5.10). We observe larger bound distances to K^+, which indicates that the coordinating ligands have to adapt to the lager size of the K^+ by conformational changes (Fig. 5.11, Table 5.1). Another difference between Na^+ and K^+ binding in the binding sites is the coordination of the hydroxyl side chain of Ser-157. It coordinates the bound K^+ in the binding site, whereas it points away from the bound Na^+ most of the time (Fig. 5.11). The binding modes observed in classical MD and QM/MM simulations are highly similar (Fig. 5.8, Table 5.1).

Ion occupancy of binding sites

The ion occupancy of the binding sites was characterized and compared in the inward- and outward-open state of MjNhaP1 with bound Na^+ and K^+. In the inward-open state, Na^+ binds stable in the binding sites of both protein chains A and B (Fig. 5.9). By contrast, in the outward-open state, several Na^+ binding and unbinding events were observed with different Na^+ ions. In chain B, the Na^+ ions binds in a stable conformation with multiple direct interactions with different residues (carboxylate group of Asp-161, carbonyl backbone of Thr-131 and Ser-157, and the hydroxyl group of Ser-157) (Fig. 5.8, 5.12). In chain A,

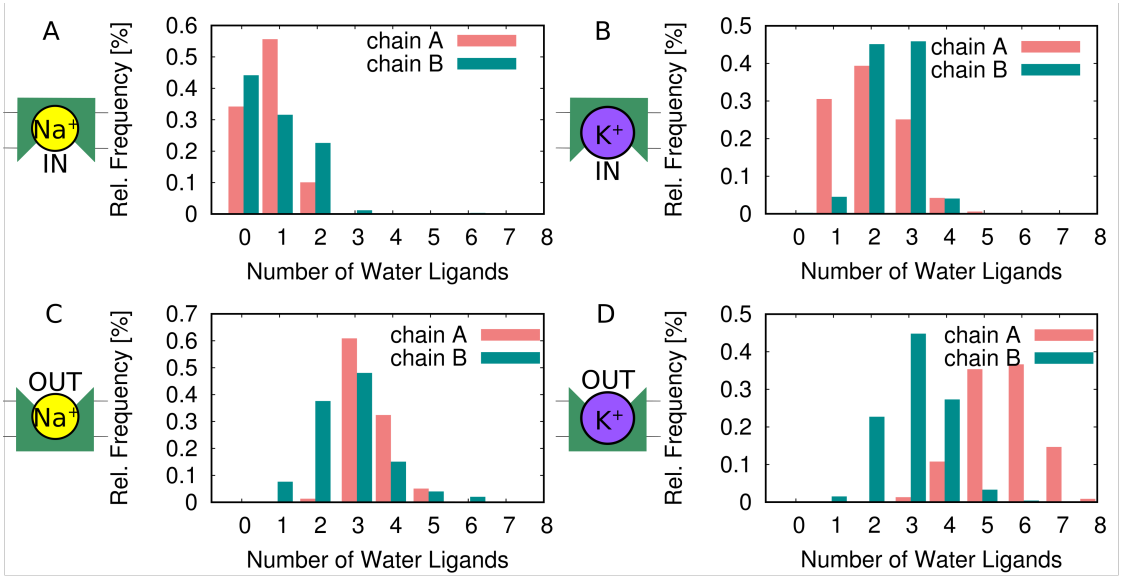


Figure 5.10: Quantitative measure of water ligands in the binding site, which additionally coordinate the bound ion, based on MD simulations. Histograms show the relative frequencies of the number of water ligands coordinating the bound ion in chain A and B (light coral and dark cyan, respectively), A) with Na^+ bound in the inward-open state and (C) in the outward-open state, B) with K^+ bound in the inward-open state and (D) in the outward-open state. Water molecules are considered as ligands if their oxygen atom is within 2.8 Å and 3.5 Å distance from Na^+ and K^+, respectively.

the Na^+ ions have prominent interactions only with Asp-161, but not with Thr-131 and Ser-157 as in the other chain (Fig. 5.9, 5.13).

A stable and continuous K^+ binding was observed to the binding site in chain B in the inward and outward-open state, whereas in chain A, the K^+ escape into the bulk in both access states, and subsequently several Na^+ ions from the bulk formed contacts with Asp-161 (Fig. 5.9, 5.13). In the outward-open state the K^+ forms direct contact only with Asp-161 and not with the other residues of the binding site (Thr-131, Ser-157). As a consequence of the solvents exposure, K^+ leaves the binding site. K^+ also leaves the binding site of chain A in the inward-open state, although K^+ binds deep in the binding site initially coordinating with Asp-161, Ser-157, and Thr-131.

Hydration of bound ions

Next the water molecules coordinating the ion bound in the binding sites were quantified. In the inward- and outward-open state, 1 ± 1 and 3 ± 1 water ligands coordinate the bound Na^+ ions, respectively (Fig. 5.10) in the MD simulations, suggesting that in the inward-open state the Na^+ is nearly saturated by the protein residues, whereas additional ligands are required to compensate the weaker binding mode in the outward-open state.

K^+ ions bound to chain B in the inward- and outward-open state follow a similar trend as the Na^+. K^+ coordinates 2-3 and 3 ± 1 water ligands in the inward- and outward-open state, respectively. In the inward-open chain A, K^+ coordinates 2 ± 1 water ligands, suggesting that K^+ is well coordinated by the binding site residues. By contrast, the K^+ in

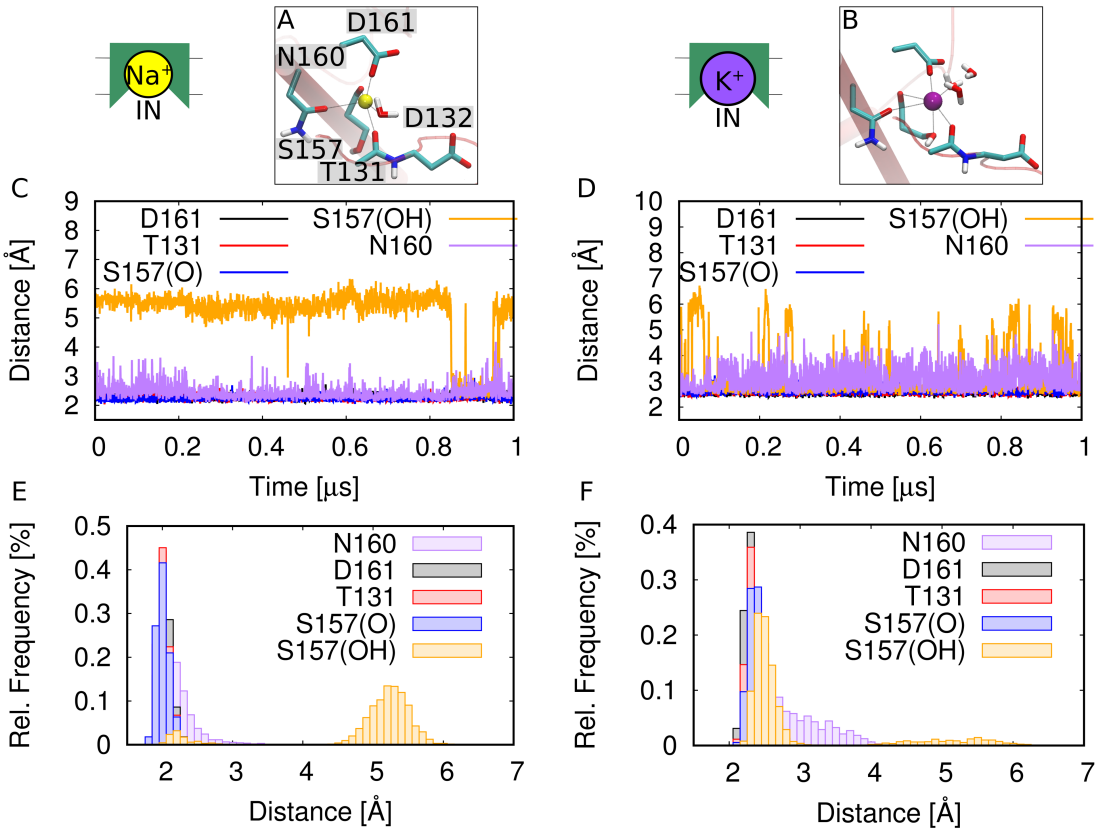


Figure 5.11: Comparison of Na^+ and K^+ bound to the inward-open state, in 1 μs MD simulations each. A) Close-up view to the binding site in chain A with bound Na^+ and (B) in chain B with bound K^+. C) Distance measurements of the coordinating residues (Asp-161, Thr-131, Ser-157(O: carbonyl oxygen), Ser-157(OH: side chain hydroxyl oxygen), and Asn-160) to Na^+ and (D) to K^+. E) Histograms of the distances to Na^+ and (F) to K^+.

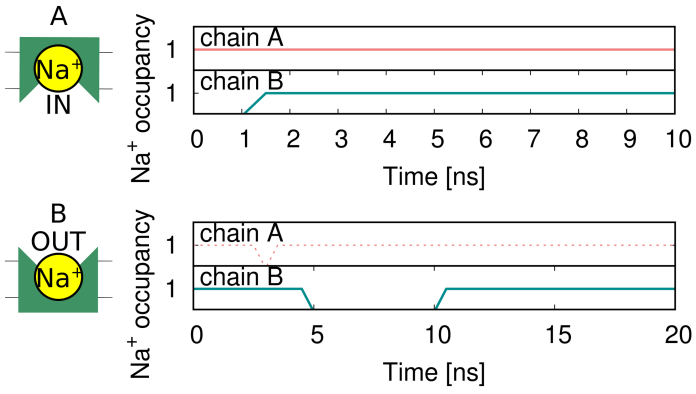


Figure 5.12: Ion binding events of Na^+ in the first 10-20 ns of 1 μs MD simulations (shown in full length in Fig. 5.9). A) First 10 ns of equilibrium MD simulation of MjNhaP1 in the inward-open state. Chain A (light coral) starts with bound Na^+ to the binding site and empty binding site in chain B (dark cyan). Na^+ binding takes place to the empty binding site of chain B in the first ns. B) First 20 ns of equilibrium MD simulation of MjNhaP1 in the outward-open state. The initial Na^+ binding event in both chains happens already in an early equilibration step. In the first 20 ns of the equilibrium MD simulation spontaneous ion unbinding and binding events take place. Dashed line, in the outward-open chain A, indicates that the Na^+ has only contact to Asp-161 at the entry of the ion binding site and is not deeply bound in the binding pocket (Fig. 5.13). Ions are considered to be bound within 4.5 Å distance to Asp-161.

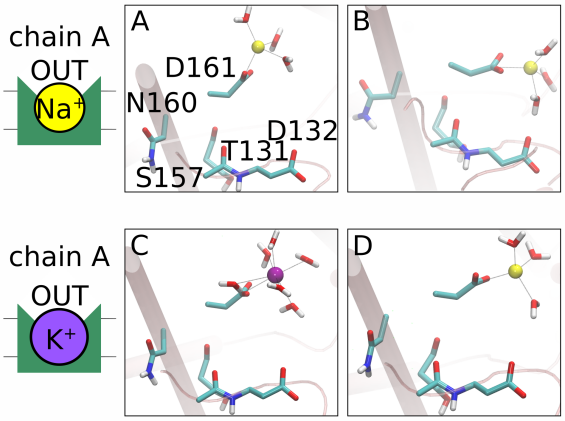


Figure 5.13: Snapshots from 1 μs MD simulations of the ion binding site of chain A in the outward-open state. The ions are not bound deep in the binding site pocket of chain A, they are just in contact with Asp-161 and exposed to the solvents. Snapshots at (A) *ca.* 100 ns and (B) *ca.* 900 ns of the MDsimulation of the outward-open state with Na^+. Snapshots at (C) *ca.* 100 ns and (D) *ca.* 500 ns of the MD simulation of the outward-open state with K^+ and Na^+, respectively. Initially the binding site is loaded with K^+, after K^+ escape from the binding site several Na^+ ions from the bulk come in close proximity to Asp-161 (Fig. 5.9).

the outward-open state forms contact only with Asp-161 and therefore it coordinates 5-6 water ligands. The high number of water ligands reflects a relatively weak interaction with the protein and the exposure to the solvent.

5.3 Ion Selectivity and Specificity

5.3.1 Free Energy of Ion Binding

In order to quantify the ion binding selectivity, the free-energy difference of bound Na^+ and K^+ was calculated in the inward- and outward-open state of MjNhaP1. The thermodynamic cycle in Fig. 5.14 is constructed with the transformation between Na^+ and K^+ in aqueous solution and bound to the binding site of MjNhaP1. The zero-sum rule of a thermodynamic cycle enables us to define the free-energy difference of ion binding $\Delta\Delta G^{\text{binding}}$ (vertical transformations) as difference between the free energies of the ions bound to the protein and solvated in bulk water (horizontal transformations). The direct comparison of $\Delta\Delta G^{\text{binding}}_{\text{In}}$ (in the inward-open state) and $\Delta\Delta G^{\text{binding}}_{\text{Out}}$ (in the outward-open state) gives the possible conformation-dependent ion binding selectivity of MjNhaP1.

The horizontal free-energy differences in Fig. 5.14 between Na^+ and K^+ are estimated based on classical alchemical transformations of Na^+ to K^+ in a single-topology approach. One positively charged fictitious particle is defined, in which the Lennard-Jones potential is scaled in 41 steps from Na^+ (ϵ = -0.0469 kcal mol^{-1}; σ = 1.41075 Å) to K^+ (ϵ = -0.0870 kcal mol^{-1}; σ = 1.76375 Å) [131, 143]. This method has the advantage that it minimally perturbs the system. A restraint on the emerging or vanishing particles is not needed unlike in the dual-topology approach.

The free-energy difference of Na^+ and K^+ was estimated in aqueous solution, and bound to the binding site in the inward- and outward-open state of MjNhaP1, based on

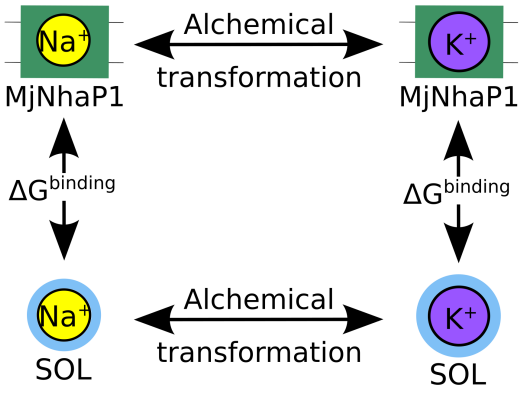


Figure 5.14: Thermodynamic cycle to estimate free energies of ion binding to MjNhaP1. Horizontal paths are implemented as alchemical transformations of the ions. The vertical paths are the free energies of ion binding to MjNhaP1 in reference to ion solvation free energies.

Table 5.2: Free-energy difference between Na^+ and K^+ estimated with EXP method in water (sol) and bound to the inward- (In) and outward-open (Out) state of MjNhaP1. Free energies in kcal mol^{-1} are estimated based on 10 independent forward (f) and backwards (b) trajectories. The mean and the standard error of the mean (sem) from 10 independent free-energy estimations are given in kcal mol^{-1}.

Run	ΔG_f^{sol}	ΔG_b^{sol}	ΔG_f^{In}	ΔG_b^{In}	ΔG_f^{Out}	ΔG_b^{Out}
1	18.4761	18.4592	23.59	22.21	18.51	18.53
2	18.4686	18.4459	25.23	24.84	19.08	19.30
3	18.4841	18.4515	24.45	24.77	18.60	18.49
4	18.4522	18.4717	24.01	22.31	18.90	19.11
5	18.4780	18.4623	24.16	24.42	18.93	19.13
6	18.4712	18.4510	22.48	22.81	18.06	18.53
7	18.4615	18.4760	22.03	21.68	19.01	18.92
8	18.4554	18.4808	24.47	23.47	18.84	18.72
9	18.4522	18.4580	24.36	24.32	19.16	19.19
10	18.4472	18.4585	24.54	24.39	19.42	19.24
mean ± sem	18.4646 ± 0.0040	18.4615 ± 0.0036	23.93 ± 0.31	23.52 ± 0.37	18.85 ± 0.12	18.92 ± 0.10

the alchemical-transformation trajectories with three different methods (EXP, BAR, and mBAR). The results of these calculations are shown in Tables 5.2, 5.3, 5.4, and in Fig. 5.15, 5.16.

The free energies estimated with EXP, BAR and mBAR methods are very similar. As an example a comparison of free energies in the inward-open state is demonstrated as follows. The difference between estimated free energies of the three methods is ten times smaller (0.01 kcal mol^{-1}) than the difference within each method (standard error of the mean (sem): 0.31 and 0.38 for forward and backward, respectively). There are minimal variations in the free-energy estimations, suggesting that the free-energy calculations are converged. For further analysis the free energies estimated with mBAR method, combining all available data, are taken and summarized in Fig. 5.17.

The estimated solvation free-energy differences $\Delta G^{sol}(Na^+, K^+)$ is 18.5 kcal mol^{-1} [1], which is close to the experimental estimate of 17.45 kcal mol^{-1} [150]. The free-energy difference between Na^+ and K^+ bound to the binding site in the inward-open state ($\Delta G^{In}(Na^+, K^+)$) was estimated to 23.7 kcal mol^{-1}, and in the outward-open state

[1] Noskov *et al.* [143] calculated $\Delta G_{bulk}(Na^+ \text{ to } K^+) = 18.12$ kcal mol^{-1}

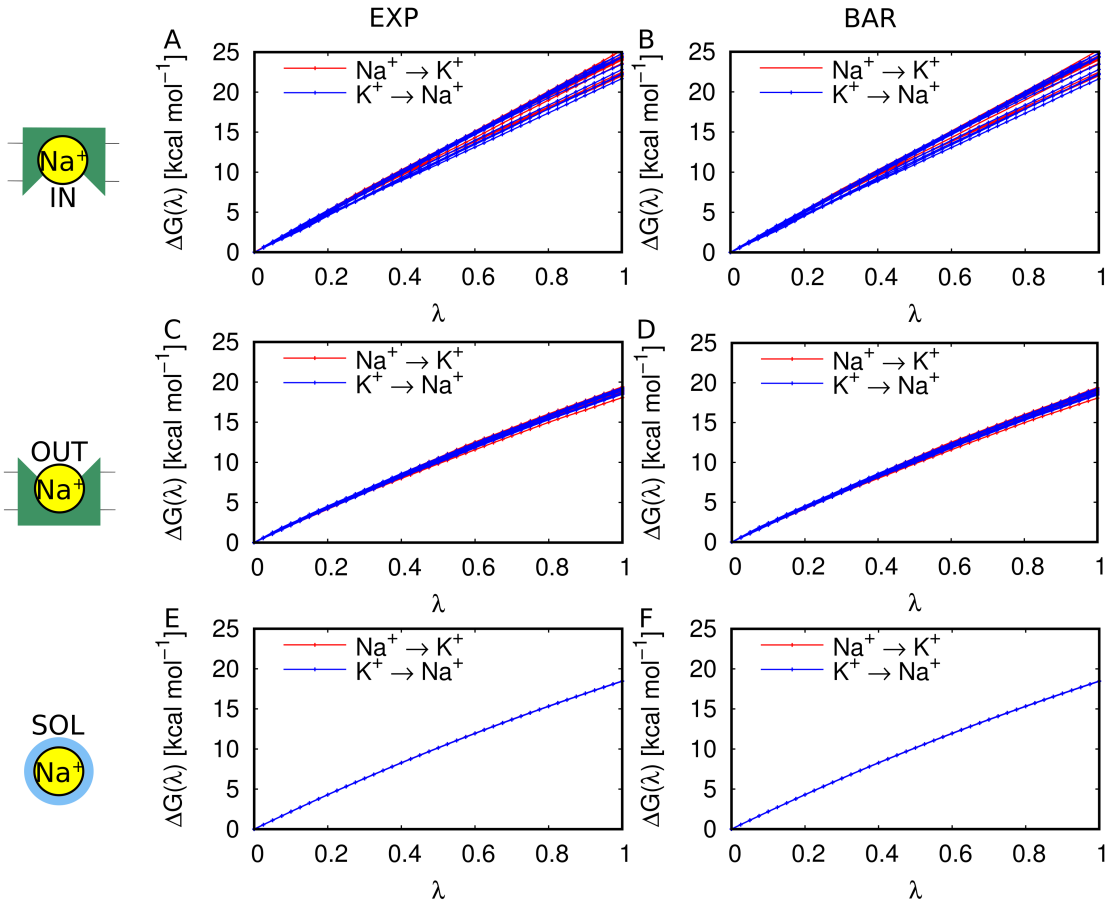


Figure 5.15: Free-energy profiles of alchemical transformations as a function of coupling parameter λ interpolating between Na^+ ($\lambda = 0$) and K^+ ($\lambda = 1$). Free energies are evaluated with EXP method (A) the inward- and (C) outward-open state of MjNhaP1 and (E) in aqueous solution. Free energies are evaluated with BAR method (B) the inward- and (D) outward-open state of MjNhaP1 and (F) in aqueous solution.

Table 5.3: Free-energy difference between Na^+ and K^+ estimated with BAR method in water (sol) and bound to the inward- (In) and outward-open (Out) state of MjNhaP1. Free energies in kcal mol^{-1} are estimated based on 10 independent forward (f) and backwards (b) trajectories. The mean and the standard error of the mean (sem) from 10 independent free-energy estimations are given in kcal mol^{-1}.

Run	ΔG_f^{sol}	ΔG_b^{sol}	ΔG_f^{In}	ΔG_b^{In}	ΔG_f^{Out}	ΔG_b^{Out}
1	18.4779	18.4608	23.58	22.16	18.50	18.51
2	18.4701	18.4468	25.23	24.84	19.07	19.28
3	18.4857	18.4531	24.44	24.76	18.59	18.46
4	18.4539	18.4730	24.00	22.28	18.88	19.09
5	18.4795	18.4636	24.15	24.40	18.93	19.12
6	18.4724	18.4527	22.47	22.80	18.06	18.53
7	18.4628	18.4781	22.02	21.62	18.95	18.88
8	18.4567	18.4825	24.45	23.46	18.76	18.72
9	18.4533	18.4596	24.35	24.32	19.17	19.18
10	18.4487	18.4600	24.53	24.39	19.40	19.24
mean ± sem	18.4661 ± 0.0040	18.4630 ± 0.0037	23.92 ± 0.31	23.50 ± 0.38	18.83 ± 0.12	18.90 ± 0.10

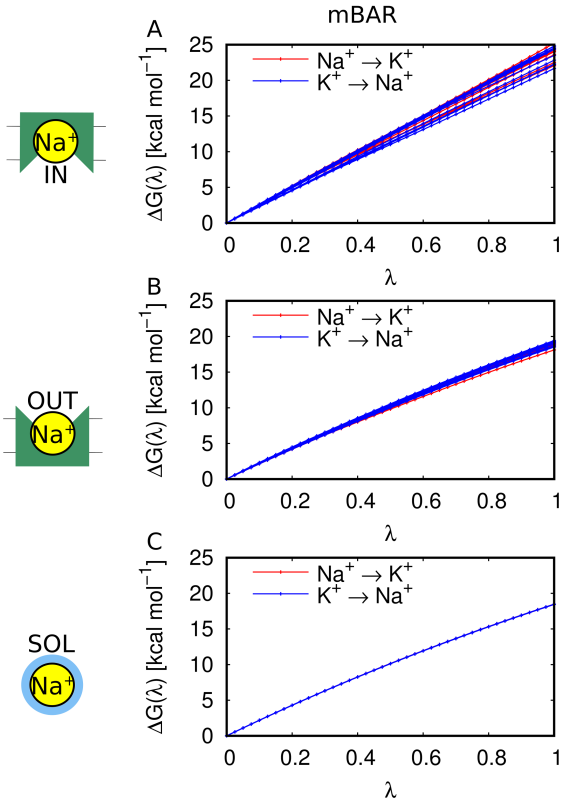


Figure 5.16: Free-energy profiles of alchemical transformations as a function of coupling parameter λ interpolating between Na^+ ($\lambda = 0$) and K^+ ($\lambda = 1$). Free energies are evaluated with mBAR method for (A) the inward- and (B) the outward-open state of MjNhaP1 and (C) in aqueous solution.

Table 5.4: Free-energy difference between Na^+ and K^+ estimated with mBAR method in water (sol) and bound to the inward- (In) and outward-open (Out) state of MjNhaP1. Free energies in kcal mol^{-1} are estimated based on 10 independent forward (f) and backwards (b) trajectories. The mean and the standard error of the mean (sem) from 10 independent free-energy estimations are given in kcal mol^{-1}.

Run	ΔG_f^{sol}	ΔG_b^{sol}	ΔG_f^{In}	ΔG_b^{In}	ΔG_f^{Out}	ΔG_b^{Out}
1	18.4714	18.4564	23.61	22.21	18.55	18.55
2	18.4654	18.4429	25.22	24.84	19.09	19.33
3	18.4805	18.4468	24.45	24.78	18.62	18.54
4	18.4503	18.4682	24.01	22.30	18.93	19.16
5	18.4735	18.4593	24.15	24.41	18.96	19.13
6	18.4688	18.4492	22.52	22.83	18.12	18.52
7	18.4561	18.4690	22.07	21.67	19.00	18.95
8	18.4493	18.4759	24.46	23.50	18.84	18.74
9	18.4473	18.4565	24.37	24.34	19.18	19.20
10	18.4416	18.4541	24.56	24.42	19.44	19.26
mean ± sem	18.4604 ± 0.0042	18.4578 ± 0.0033	23.94 ± 0.31	23.53 ± 0.38	18.87 ± 0.12	18.94 ± 0.10

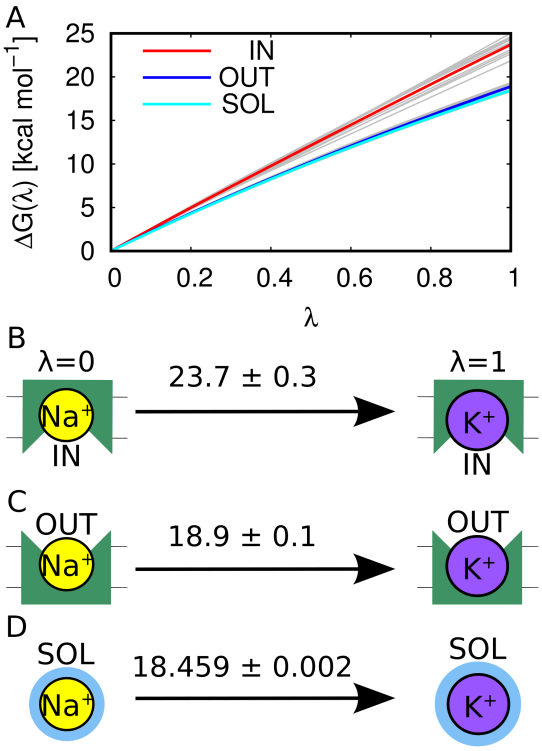


Figure 5.17: Free-energy difference between Na^+ and K^+ in water and bound to the inward- and outward-open state of MjNhaP1 estimated with mBAR method. (A) Free-energy profiles of alchemical transformations as a function of coupling parameter lambda λ interpolating between Na^+ ($\lambda = 0$) and K^+ ($\lambda = 1$). Free energies are evaluated with mBAR method for the inward- (red) and outward-open (blue) state of MjNhaP1 and the solvation free energy (cyan). Free energies are estimated from 10 independent forward and backward runs. The free energies estimated from one pair of forward and backward run are shown in gray. Overall free-energy difference and the standard error of the mean from 10 independent free-energy estimations in kcal mol^{-1} of the alchemical transformation from Na^+ to K^+ in (B) the inward-open state, (C) the outward-open state, (D) and in aqueous solution.

($\Delta G^{\mathrm{Out}}(Na^+, K^+)$) to 18.9 kcal mol^{-1}. This means that the inward-open state has a clear preference for Na^+ over K^+ with a $\Delta\Delta G_{\mathrm{In}}(Na^+, K^+) = 5.2$ kcal mol^{-1}, compared to bulk water. However, in the outward-open state a negligible preference for Na^+ over K^+ was obtained $\Delta\Delta G_{\mathrm{Out}}(Na^+, K^+) = 0.4$ kcal mol^{-1} compared to the bulk water. Overall we can conclude from the free-energy calculations that Na^+ is preferred over K^+.

5.3.2 Controlling the Ion Selectivity

A central point in understanding the ion selectivity is determining its structural origin. The contributions of different amino acids to ion binding selectivity were estimated with free-energy calculations. Based on a sequence alignment of MjNhaP1 with related homologs, with known ion selectivity (Fig. 5.18), residues that could have potential influence on the ion selectivity, were identified by sequence alignment of ten homologs antiporter (Fig. 5.18). These ten homologs are divided into three groups based on their ion selectivity; (1) Four homologs with known structure are Na^+ selective (MjNhaP1, PaNhaP, EcNhaA, TtNapA) [29,33,34]; (2) Three homologs proteins known to transport both ions (VcNhaP1, ScNHX1, AtNHX1) [1,151]; and (3) Three antiporters that are selective to K^+ (VcNhaP2, TaNHX2, EcNhaP2) [152–154]. The first and second group of Na^+ selective and non-selective antiporters are classified to the first clade of NhaP-I/NHE in the phylogenetic tree of Masrati *et al.* [155]. The third group of K^+ selective antiporters are classified in the third clade of NhaP-II K^+-specific antiporters.

The residues directly coordinating the ion are conserved (Thr-131 and Asp-161; MjNhaP1 nomenclature). Four residues were selected in a larger region around the binding site (Ala-130, Pro-133, Ser-157, and Pro-162; MjNhaP1 nomenclature). Based on the K^+ selective group, the following single mutations of MjNhaP1 in the inward-open state were suggested: A130S, P133S, S157G, and P162A. For instance, Ala-130 is conserved in the group of Na^+ selective antiporters, whereas in the group of K^+ selective antiporters there is mostly (2 out of 3) a serine at this position (Fig. 5.18). Additionally the mutants D132N and D161N were chosen to check the effect of reducing the charge density of the binding site.

Alchemical transformations were performed by transforming Na^+ into a K^+ ion bound to the binding site in all in-silico mutated systems and free-energy differences were estimated between the Na^+ and the K^+ bound state. To probe the influence of the individual residues on the ion binding selectivity, the calculated free-energy differences of the mutants were compared with the wildtype (WT) (Table 5.5, Fig. 5.19).

The largest effects were predicted for A130S and P162A towards a K^+ selectivity among the single mutants. A130S is 2 kcal mol^{-1} and P162A is 0.8 kcal mol^{-1} less favored for Na^+ relative to K^+, as compared to the WT (Fig. 5.19 C, Table 5.5). Based on these findings, the A130S/P162A double mutant was constructed. The double mutant A130S/P162A achieve a nearly additive effect of 2.5 kcal mol^{-1}. The reason for the enhanced K^+ favoring effect of A130S is that the hydroxyl group of the serine establishes a hydrogen bond to the ion when the ion becomes more similar to K^+ during the alchemical transformation (Fig. 5.19 B, D). A further possible explanation for the stabilizing effect of K^+ in P162A mutation is a small

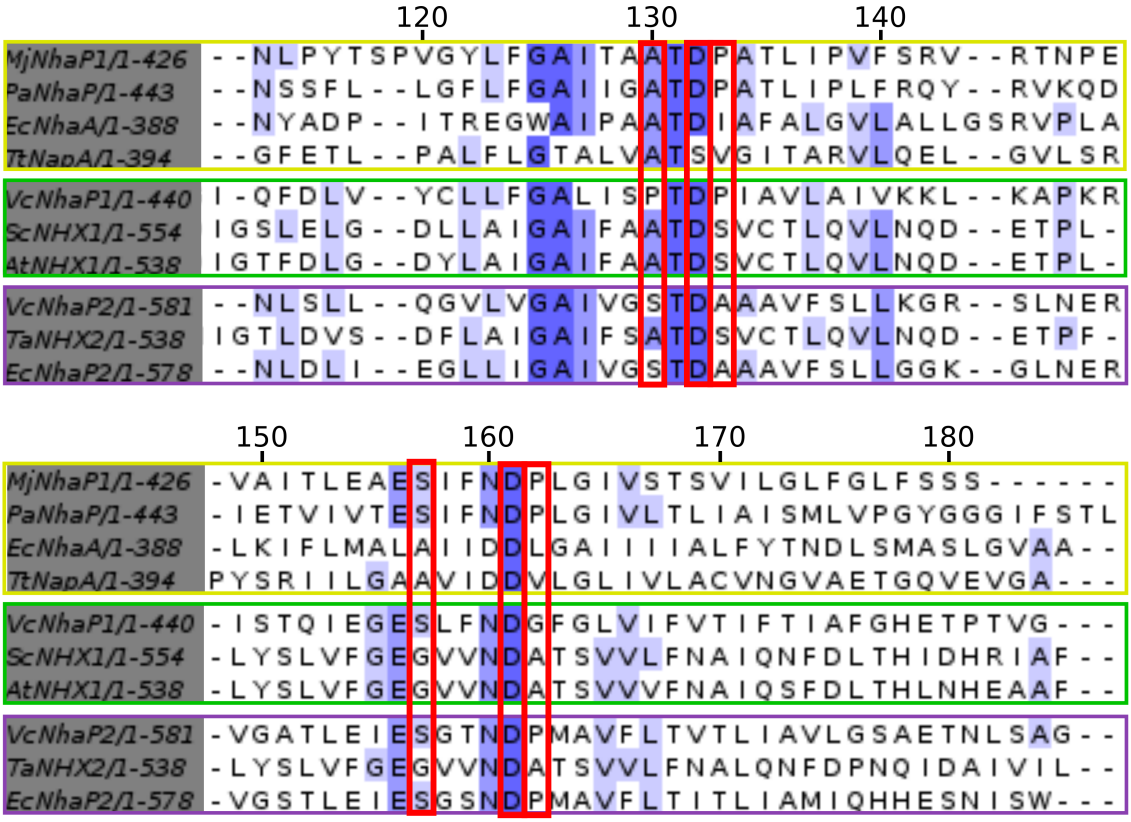


Figure 5.18: Sequence alignment of MjNhaP1 homologs with known ion selectivity. Group 1: (yellow frame) Na^+ selective antiporters. Group 2: (green frame) antiporters which transport both Na^+ and K^+. Group 3: (purple frame) K^+ selective antiporters. The red boxes marking the residues picked for mutations (A130S, S132N, P133S, S157G, D161N, and P162A). The sequence alignment was performed unsing the multiple sequence alignment web server PSI/TM-Coffee [134, 135] with "transmembrane" option and the homology extension option of UniRef100. The representation of the sequence alignment is created using jalview [136].

Table 5.5: Free-energy differences of Na^+ and K^+ bound to the binding site of the mutants relative to the WT, estimated with mBAR method and given in kcal mol^{-1}.

Mutant	ΔG	$\Delta\Delta G^{WT}$
WT	23.72	0.00
A130S/P162A	21.19	-2.53
A130S	21.70	-2.01
P162A	22.94	-0.77
S157G	23.50	-0.21
D161N	23.94	0.23
P133S	24.24	0.52
D132N	24.43	0.71

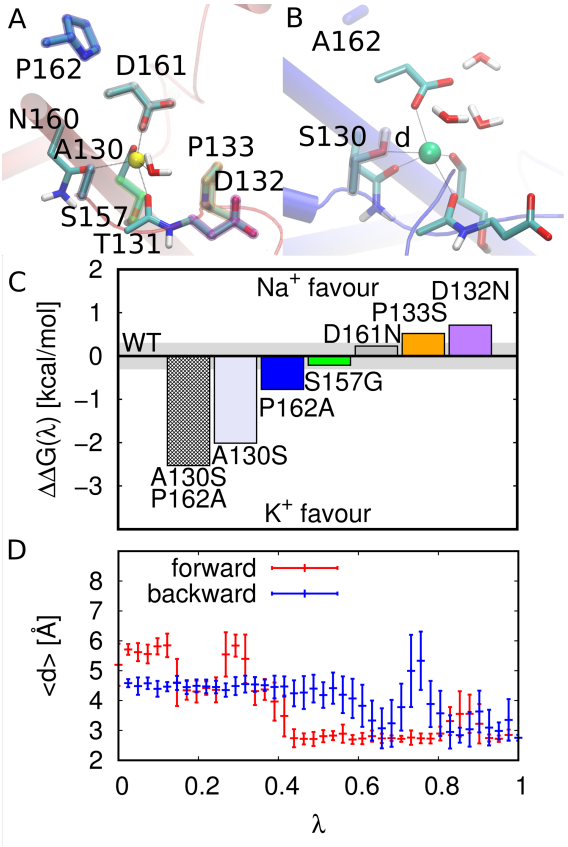


Figure 5.19: Effect of mutations on the ion binding selectivity of MjNhaP1 in the inward-open state. A) Binding site of chain A in the inward-open state indicating the mutated residues by outlined colors (A130S, D132N, P133S, S157G, D161N, P162A). B) A snapshot of the binding site during the alchemical transformation of Na^+ to K^+ in the double mutant A130S/P162A. C) Difference in free energy $\Delta G(\lambda = 0, \lambda = 1)$ in kcal mol^{-1} from alchemical transformation of the different mutants in reference to the free-energy difference of Na^+ and K^+ in the WT. A negative $\Delta\Delta G$ means that the mutant favors K^+ binding more than the WT. D) The average and standard deviation of the distance in Å in each λ window between the oxygen atom of the hydroxyl group of the Ser-130 in the (A130S/P162A) mutant and the bound ion (red: forward transformation; blue: backward transformation).

increase in the number of water ligands coordinating the bound ion as a result of a small opening of the binding site. The finding that A130S/P162A shows a significant lower Na^+ selectivity compared to the WT, is consistent with the sequence signatures associated with Na^+ and K^+ selectivity (Fig. 5.3, 5.18).

To access the impact of the mutations on the ion binding affinities, one can convert the free-energy differences to a concentration ratio between the two ions. In the WT, a concentration ratio between $[Na^+]$ and $[K^+]$ of 1 to about 4600 is required to achieve equal affinity for both ions. In the double mutant (A130S/P162A) the concentration ratio is reduce to about 80 to achieve equal affinity, which is thus still rather large.

Experimental measurements also confirmed the extreme Na^+ selective character of MjNhaP1. Dr. Wöhlert[2] performed transport measurements with WT and the double mutant (A130S/P162A) of MjNhaP1 complementary to the computational description of the ion binding and selectivity in MjNhaP1. It was first tested whether the addition of Na^+ and K^+ is able to shift an already established transport equilibrium. The WT and the double mutant of MjNhaP1 showed a clear shift of the equilibrium in case of Na^+ (Fig. 5.20 A, Fig. 5.21 A), whereas upon K^+ addition no effect was detected. Another experiment was performed to ensure a 80-fold express of K^+, in which proteoliposomes were loaded with either 200 mM Na^+ or 200 mM K^+ (Fig. 5.20 B, Fig. 5.21 B). Only in the case of Na^+-loaded liposomes, the double mutant and the WT showed a diminishing fluorescence over time, and no effect was detected with K^+-loaded liposomes. This indicates that no protons entering the liposomes via an K^+/H^+ exchange. The Na^+/H^+ exchange of the double mutant A130S/P162A is reduced compared to the WT MjNhaP1. Despite the changes in the binding site to enhance K^+ selectivity, both experiments demonstrate that K^+ is transported neither by WT MjNhaP1 nor the double mutant A130S/P162A.

5.4 Discussion

M. jannaschii lives in deep sea water close to hydrothermal vents [156]. The abundant Na^+ and a weakly alkalized environment causes a huge Na^+ gradient from outside to inside of the cell and a negligible H^+ gradient across the membrane. *M. jannaschii* is methanogenic archaea without cytochromes, and it therefore gains energy from methanogenesis to maintain its essential electrochemical gradient to sustain cellular function [157]. A central step in its bioenergetic cascade of reducing CO_2 to CH_4 is the methyl-transferase MtrA-H, which works as an active outside Na^+ pump. It can be assumed that the ATP synthase of *M. jannaschii* is Na^+ driven[3].

In our experiments MjNhaP1, was shown to transport Na^+ but not K^+ (Fig. 5.20). Consistent with the experimental outcome, the theoretical results show a strong favor for Na^+. However, the contribution of transition states on the pathway, connecting inward- and outward-open state, to the ion selectivity were not included in this work but certainly have a significant role. Na^+/H^+ antiporters are able to work in both directions, driven by

[2]Department of Structural Biology, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany.

[3]The UniProt entry of Q58623 describes a "putative cation driven ATPase". In related organisms the ATP synthase is Na^+ dependent [158–160]

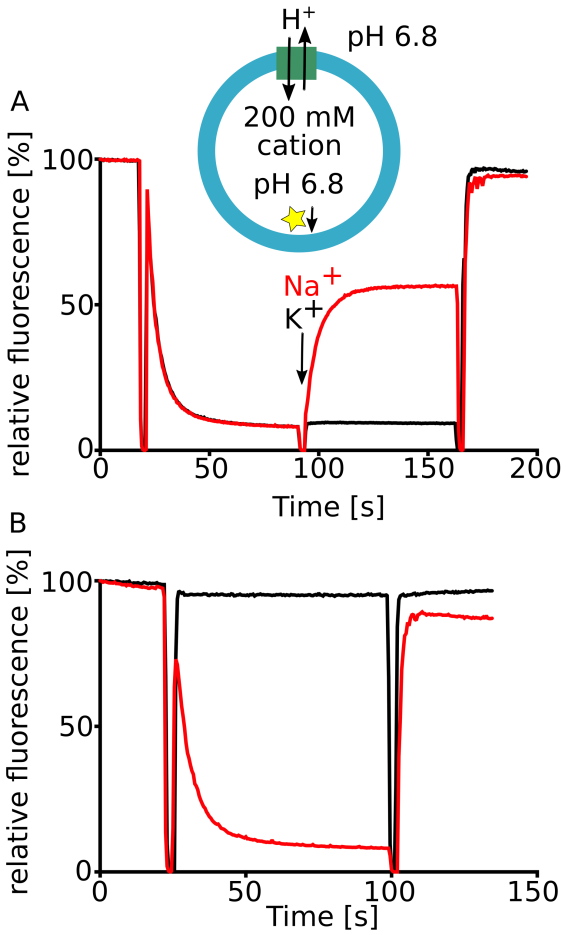


Figure 5.20: Experimental determination of MjNhaP1 ion selectivity (experiments performed by Dr. David Wöhlert). A) Na^+-loaded proteoliposomes are diluted in reaction buffer, resulting in a drop of fluorescence due to proton influx catalyzed by MjNhaP1. After initial equilibration, addition of Na^+ (red) to a concentration of 5 mM results in reverse transport of protons and an increase of fluorescence. Addition of K^+ to the same concentration (black) does not shift the transport equilibrium. B) Proteoliposmes were loaded with either 200 mM Na^+ (red) or K^+ (black) and diluted into reaction buffer. While a Δ pH is built up in Na^+-loaded liposomes, the fluorescence stays constant for the K^+-loaded liposomes. Thus, even at higher concentrations, K^+ driven proton influx is no detectable for WT MjNhaP1.

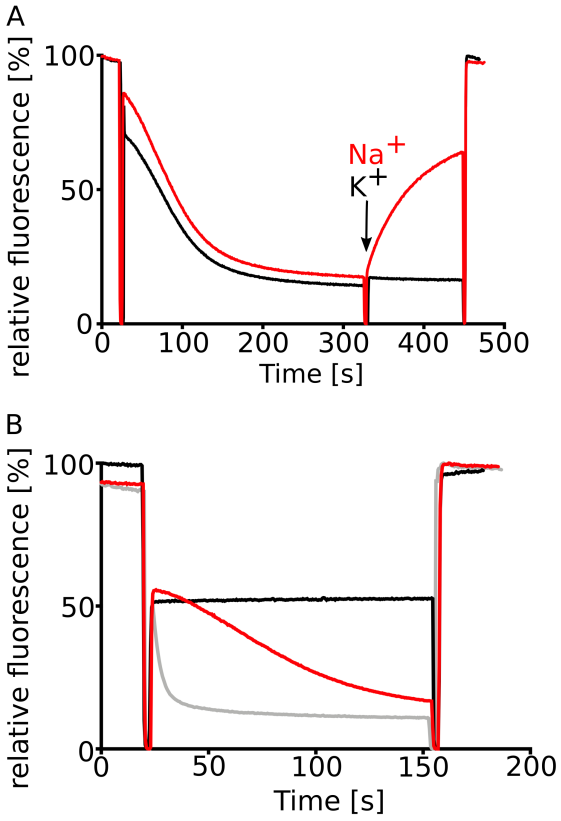


Figure 5.21: Transport measurements of the double mutant A130S/P162A (experiments performed by Dr. David Wöhlert). A) A shift of the transport equilibrium, previously established by Na^+ efflux is only observed for the addition of external Na^+ (red) to a concentration of 5 mM. The balance is uninfluenced by the addition of the same concentration of K^+ (black). B) Proteolipsomes were loaded with 200 mM of either potassium ion (black) or sodium ion (red). Sodium ion efflux transport activity of WT MjNhaP1 (gray) is shown for reference.

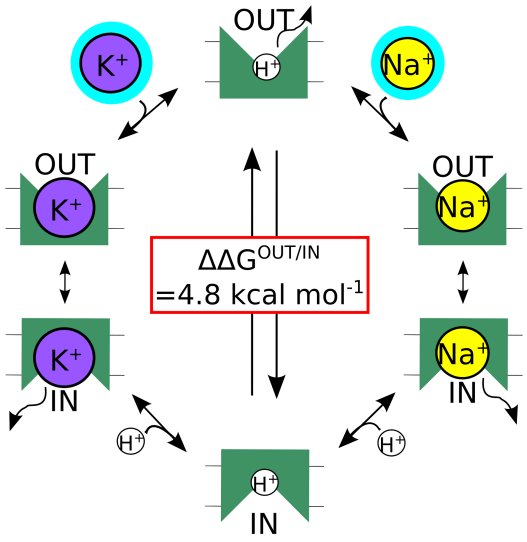


Figure 5.22: Transport cycle of MjNhaP1 constructed with K^+ transport on the left side and Na^+ transport on the right side. The discrimination between Na^+ and K^+ binding takes place in the inward-open state in reference to the solvation free energies. The inward-open state with bound Na^+ is by 4.8 kcal mol^{-1} more favorite than with bound K^+, in reverence to the free-energy difference between Na^+ and K^+ bound to the outward-open state in MjNhaP1.

the dominant gradient across the membrane. Based on what is so far known from the *M. jannaschii* and related organisms [34, 35], the MjNhaP1 is a Na^+ driven H^+ pump that ensures pH homeostasis in the cell.

Based on free-energy calculations a clear preference for Na^+ over K^+ binding was found in the inward-open state ($\Delta\Delta G_{\mathrm{In}}(Na^+, K^+) = 5.2$ kcal mol^{-1}), whereas ion binding to the outward-open state is nearly equally favorable for both ions ($\Delta\Delta G_{\mathrm{Out}}(Na^+, K^+) = 0.4$ kcal mol^{-1}). This inequality between the conformational states is consistent with the presumed functional requirements on MjNhaP1. Due to the abundant Na^+ presence in sea water at the outside of a *M. jannaschii* cell, there is no need for a highly selectivity of Na^+ in the outward-open conformational state. However, in the inside of the cell, the K^+ concentration is about 10 times higher than the Na^+ concentration. In order to avoid a K^+ efflux, MjNhaP1 has to be highly specific for Na^+ in the inward-open state. The strong preference for Na^+ is consistent with the free-energy calculations performed in this work.

Based on the results of this work, a transport cycle of MjNhaP1 can be drawn shuttling Na^+ to the inside of the cell and pump H^+ to the outside. In Fig. 5.22, free energies of Na^+ bound states are compared with K^+ bound state in a hypothetical K^+ transport. The main finding of this work is that due to a $\Delta\Delta G^{\mathrm{OUT/IN}}(Na^+, K^+) = 4.8$ kcal mol^{-1}, the cycle can only be completed with Na^+, but not with K^+. For a complete energetic description of the transport cycle, the proton affinity of the ion binding site in the two conformational states needed to be compared with the binding free energies of the ions.

The contribution of individual residues on the binding selectivity was estimated by applying free-energy calculations. In combination with a sequence alignment with partial and exclusive K^+ selective homologs, it was possible to identify residues responsible for the selectivity of ion binding. The residues Ala-130 and Pro-162 have the largest effect

on the preference for Na^+ binding when they are mutated to A130S and P162A. These findings supports the phylogeny-based hypothesis of Masrati *et al.* [155] in which a polar residue (A130S) close to the binding site at the transmembrane helix 4 is important for ion selectivity. The double mutant A130S/P162A reduce the Na^+ selectivity by 2.5 kcal mol^{-1} and with a preference of 2.7 kcal mol^{-1} for Na^+. The calculated ion selectivity of the WT and the double mutant A130S/P162A are consistent with experimental measurements. The calculated preference for Na^+ binding in the WT is consistent with the absence of a detected K^+ transport. For the double mutant A130S/P162A, the theoretical results predict a concentration ratio of $[K^+]/[Na^+]$ of *ca.* 80/1 to achieve equal affinity. The calculated equilibrium affinity can be converted into a difference in transport rates, resulting in a 80 times slower transport of K^+ than Na^+, which is not detectable with the presented experimental setup. Despite the improved K^+ in the binding site in the double mutant, the factors, such as the contribution of the transition state connecting inward- and outward-open state or the accessibility of the ion to the binding site might prevent K^+ selectivity.

The present findings and sequence conservation, provide a basis for introducing additional substitutions in order to change the ion selectivity.

Chapter 6

Bioenergetic Function of Respiratory Complex I

6.1 Multiple Binding Sites in the Ubiquinone Binding Tunnel

6.1.1 SMD Simulations of Ubiquinone/ol

A bound ubiquinone (Q_{ox}) and ubiquinol (QH_2) in the primary binding site of complex I was pulled out of the Q-binding tunnel, by applying SMD simulation technique (Fig. 6.1 A). The Q_{ox} tail in the extended conformation within the Q tunnel was used to define the SMD pulling vector (Fig. 6.1). Profiles of the Q_{ox} and QH_2 headgroup motion through the Q tunnel, measured from Tyr-87_{Nqo4}, are shown in Fig. 6.1 (B). The distance plot over time (Fig. 6.1 B) seems to evolve in discrete steps, indicating that the Q species (*i.e.* Q_{ox} and QH_2) experience stabilizing interactions at specific regions in the Q-binding tunnel. In the Q_{ox} migration profile, a plateau is observed at 4 Å, 10 Å, 25 Å, and 30 Å distance from Tyr-87_{Nqo4}, where the Q_{ox} headgroup remains during a few nanoseconds in the SMD simulation, despite the constant pulling force acting on the ubiquinone (Fig. 6.1 B). For QH_2, the plateau regions are observed at 5 Å, 12 Å, 22 Å, and 27 Å (Fig. 6.1 B).

The stabilizing interactions between the Q headgroups and protein residues were identified from detailed analysis of the SMD simulations. Q_{ox} mainly interacts with the aromatic side chains of Tyr-87_{Nqo4}, His-38_{Nqo4}, Phe-28_{Nqo6}, Trp-37_{Nqo6}, Trp-241_{Nqo8}, and Phe-28_{Nqo8}, affecting the dynamics of the ubiquinone (Fig. 6.2 A). The first plateau is observed when Q_{ox} forms a hydrogen bond with Tyr-87_{Nqo4}. After Q_{ox} leaves the primary binding site, it forms a stable interaction with Phe-28_{Nqo6} at 10 Å distance from Tyr-87_{Nqo4} (Fig. 6.2 A). The next significant stabilizing interaction is formed when Q_{ox} is between the aromatic side chains of Trp-37_{Nqo6} and Trp-241_{Nqo8} at *ca.* 25 Å distance from Tyr-87_{Nqo4} (Fig. 6.2 C-E). These interactions are stable, and Q_{ox} headgroup forms contacts with these residue for *ca.* 20 ns (Fig. 6.2 A). After the Q_{ox} passes this region, the Q_{ox} headgroup remains bound to Trp-241_{Nqo8} on the other side at *ca.* 30 Å distance from Tyr-87_{Nqo4} (Fig. 6.2 F). The final interaction before the headgroup leaves the protein binding tunnel is formed with Phe-28_{Nqo8} at 30 Å distance from Tyr-87_{Nqo4} (Fig. 6.2 E).

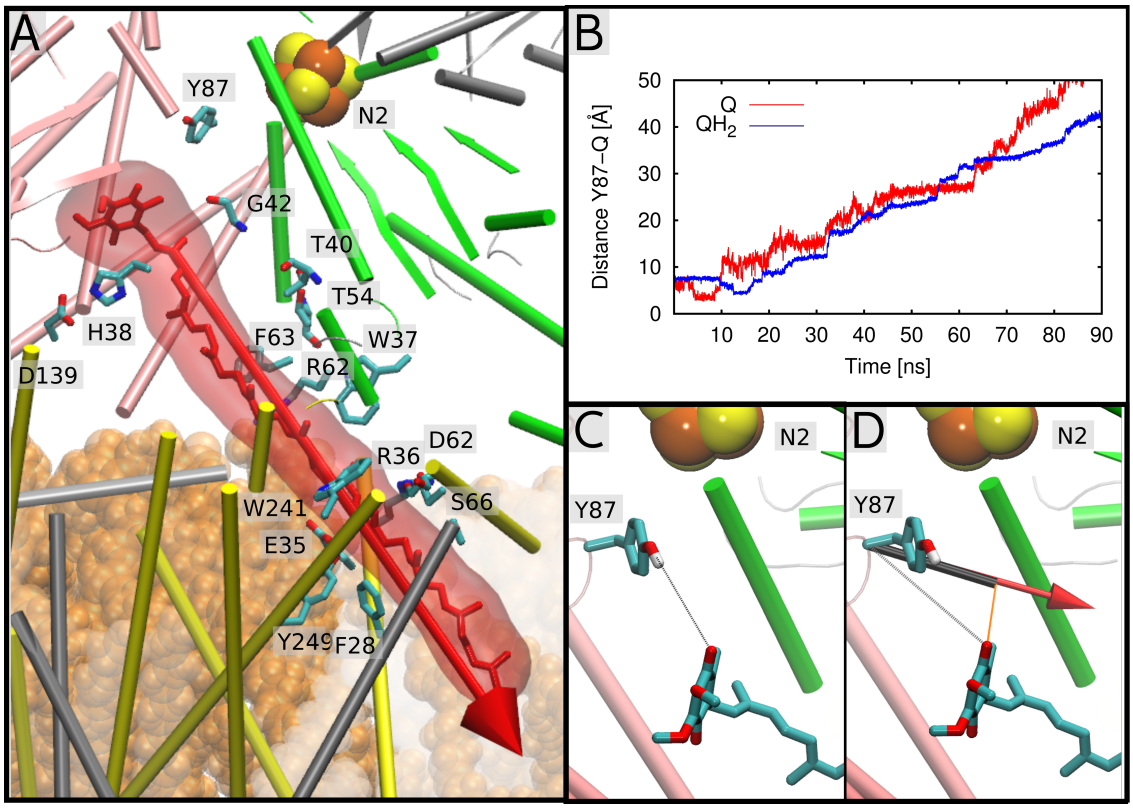


Figure 6.1: SMD simulations of Q_{ox} and QH_2. A) Oxidized ubiquinone (in red) bound in the primary binding site 1. The force acting on the complete Q_{ox}/QH_2 molecule during the SMD simulations is shown with a red arrow. B) The SMD pulling distance as a function of simulation time, showing the distance between the hydroxyl oxygen of Tyr-87_{Nqo4} and the carbonyl/hydroxyl oxygen (closest to Tyr-87_{Nqo4}) of the Q_{ox}/QH_2 headgroup. In the SMD simulations, the Q_{ox}/QH_2 molecule is pulled towards the membrane exit of its binding cavity. Illustration of the reaction coordinates C) the minimum distance between Tyr-87(OH) and a carbonyl oxygen of Q_{ox} headgroup, D) the projected minimum distance (in black) between Tyr-87_{Nqo4}(C_α) and the carbonyl oxygen of the Q_{ox} headgroup on the SMD pulling vector (in red). Figure adopted from [7], license: CC BY-NC-ND 4.0.

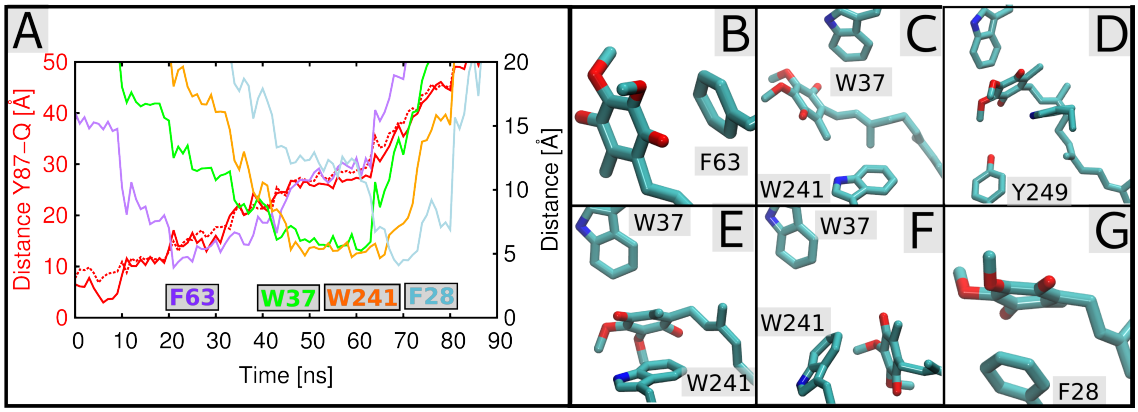


Figure 6.2: SMD simulation of Q_{ox}. A) Pulling distance between the Q_{ox} headgroup (carbonyl oxygen O2, in red, and O5, in dashed red) and the hydroxyl oxygen of Tyr-87$_{Nqo4}$ is shown as a function of the SMD simulation time. In the same graph the distances are plot to the second y-axis between the center of mass of QH_2 headgroup and the center of mass of stabilizing residues depending on the SMD simulation time. The figure highlights interactions with residues Phe-63$_{Nqo6}$ (in purple), Trp-37$_{Nqo6}$ (in green), Trp-241$_{Nqo8}$ (in orange), and Phe-28$_{Nqo8}$ (in light blue). The distance is measured between the Q_{ox} headgroup (carbonyl oxygen O2, in red, and O5, in dashed red) and the hydroxyl oxygen of Tyr-87$_{Nqo4}$. B-G) Snapshots from the SMD simulation when the Q_{ox} headgroup is stabilized by specific interactions with the selected amino acids. The snapshots are numbered in time sequence of the SMD trajectory. Snapshot of Q_{ox} interacting with (B) Phe-63$_{Nqo6}$ at 24 ns, (C) with Trp-37$_{Nqo6}$ at 44 ns, (D) with Tyr-249$_{Nqo8}$ at 56 ns, (E) with Trp-241$_{Nqo8}$ at 63 ns, which comprises the second binding site (see main text). (F) Snapshot of the Q_{ox} headgroup interacting with Tyr-241$_{Nqo8}$ at 64 ns after passing the second binding site, and (G) with Q_{ox} in site 2′, forming interactions with Phe-28$_{Nqo8}$ at 69 ns. Figure adopted from [7], license: CC BY-NC-ND 4.0.

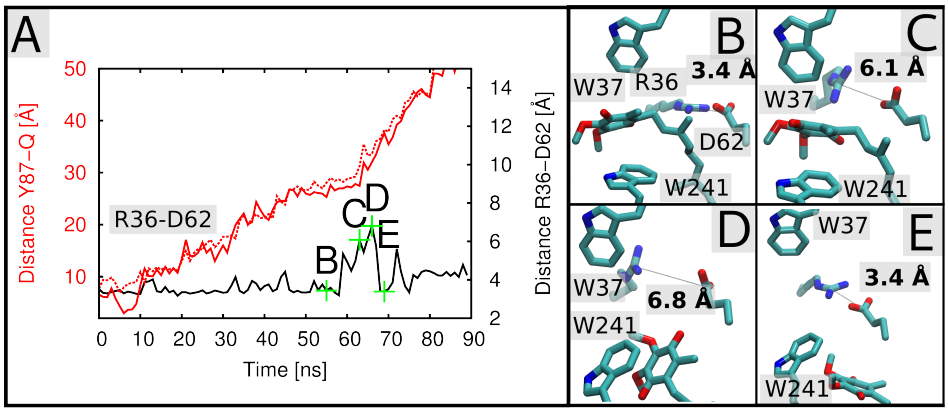


Figure 6.3: Correlation between the Tyr-87$_{Nqo4}$(OH)-Q_{ox} distance and the opening of the conserved salt bridge Arg-36$_{Nqo8}$/Asp-62$_{Nqo8}$ obtained from a SMD simulation. Experiments suggest that the Arg-36$_{Nqo8}$/Asp-62$_{Nqo8}$ interaction is functionally important in complex I. Selected structural snapshots along the SMD trajectory are indicated with green crosses and the corresponding labels in panel A. B-E) Snapshots of Q_{ox} and the Arg-36$_{Nqo8}$/Asp-62$_{Nqo8}$ ion-pair along the SMD trajectory. B) Q_{ox} in the second binding site, forming an interaction with Trp-241$_{Nqo8}$ at 55 ns, and (C) at 63 ns, upon dissociation of the Arg-36$_{Nqo8}$/Asp-62$_{Nqo8}$ ion-pair. D) Q_{ox} interacting with Trp-241$_{Nqo8}$ at 66 ns. The Arg-36$_{Nqo8}$/Asp-62$_{Nqo8}$ ion-pair is re-established after Q_{ox} passes this binding site. Figure adopted from [7], license: CC BY-NC-ND 4.0.

The salt bridge formed by Arg-36$_{Nqo8}$ and Asp-62$_{Nqo8}$ shows an interesting behavior in the SMD simulation with Q_{ox}. These residues are highly conserved and are presumed to play an important role in complex I assembly [68]. In all MD simulations performed, the side chains of Arg-36$_{Nqo8}$ and Asp-62$_{Nqo8}$ remain tightly closed, with an exception of a few frames in the Q_{ox} SMD simulation. Here the side chain distance correlates with the Q_{ox} position in the Q-binding tunnel. As shown in Fig. 6.3, the Arg-36$_{Nqo8}$/Asp-62$_{Nqo8}$ salt bridge has to open up allowing the Q_{ox} headgroup to pass through the bottleneck (around Trp-241$_{Nqo8}$), leading to the dissociation from the plateau region at *ca.* 25 Å distance from Tyr-87$_{Nqo4}$ (Fig. 6.1 A, B; 6.3 A, D). The Arg-36$_{Nqo8}$/Asp-62$_{Nqo8}$ salt bridge is re-established after Q_{ox} headgroup leaves the region (Fig. 6.3 E).

In the SMD simulation with bound ubiquinol, the QH_2 headgroup forms stabilizing interactions with the side chains of the residues Arg-62$_{Nqo6}$, Tyr-249$_{Nqo6}$, and Ser-66$_{Nqo8}$. Plateau regions are observed at distances of 5 Å, 12 Å, 17 Å, 22 Å, and 27 Å, between QH_2 headgroup and Tyr-87$_{Nqo4}$, respectively (Fig. 6.1 B). After QH_2 headgroup leaves the primary binding site (at 5 Å), the first important interaction is formed with Arg-62$_{Nqo6}$ at *ca.* 35 ns with the life time of *ca.* 20 ns in SMD simulation. After passing the Arg-62$_{Nqo6}$ region, the QH_2 headgroup moves further along the binding cavity, until a new interaction is established at *ca.* 22 Å from Tyr-87$_{Nqo4}$. The tight interaction with Arg-62$_{Nqo6}$ is replaced by a new interaction with Tyr-249$_{Nqo6}$ (Fig. 6.4 A, C). The final contact of the QH_2 headgroup is formed with Ser-66$_{Nqo8}$ beyond Trp-241$_{Nqo8}$, keeping it at 33-35 Å distance from Tyr-87$_{Nqo4}$. Both hydroxyl groups of QH_2 headgroup form subsequently interactions with Ser-66$_{Nqo8}$ at this site (Fig. 6.4 A, D, E).

6.1.2 Sampling Ubiquinone/ol Dynamics in the Q-Binding Tunnel

Based on the SMD simulations with Q_{ox} and QH_2 in the Q tunnel, multiple short (11 ns) equilibrium MD simulations were initiated, where the Q headgroup was distributed over the entire Q-binding tunnel. These simulations allowed us to extract redox-state dependent free energy profiles based on a Bayesian model [122]. This requires an appropriate sampling of Q_{ox} and QH_2 in the entire Q-binding tunnel.

Two reaction coordinates are introduced to understand the behavior of Q in complex I, and to test that the findings are independent of a specific definition of the reaction coordinate. The first reaction coordinate is defined as the distance between the hydroxyl oxygen of Tyr-87$_{Nqo4}$ and the closest carbonyl/hydroxyl oxygen of the Q_{ox}/QH_2 headgroup (Fig. 6.1 C). The second reaction coordinate is defined as the distance between the C_α carbon of Tyr-87$_{Nqo4}$ and the closest carbonyl/hydroxyl oxygen of the Q_{ox}/QH_2 headgroup, projected on the SMD pulling vector (Fig. 6.1 D). The second projected reaction coordinate has the advantage that it is independent of the side chain movements of Tyr-87$_{Nqo4}$. The 11 ns equilibrium MD simulations are represented along the Tyr-87$_{Nqo4}$(OH)-Q distance reaction coordinate in Fig. 6.5 and along the projected Tyr-87$_{Nqo4}$(C_α)-Q distance reaction coordinate in Fig. 6.6. The necessity of defining the projected distance reaction coordinate becomes clear from Fig. 6.6. Some trajectories of QH_2 stay in the primary binding site and sample regions that are somewhat deeper in the Q tunnel at 2-5 Å along the projected reac-

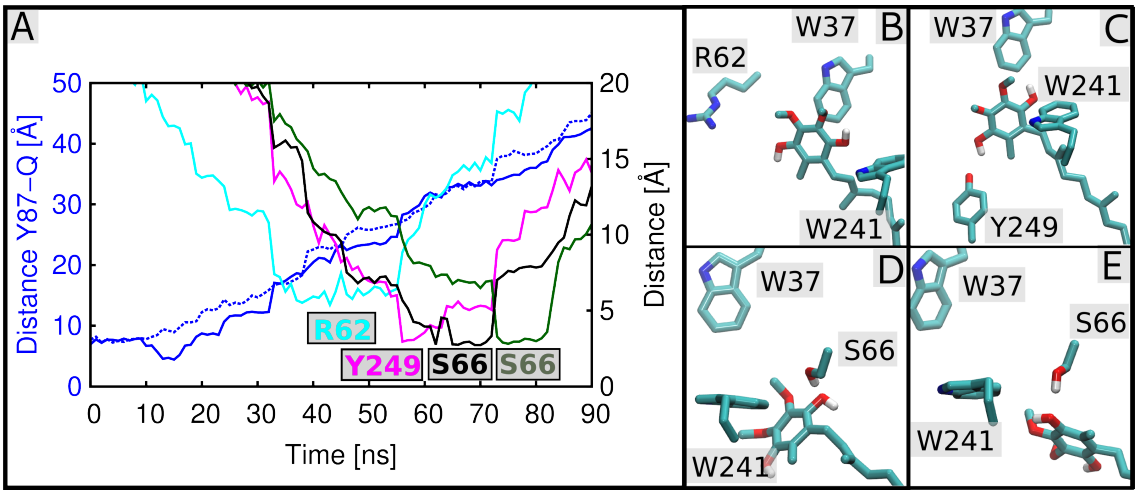


Figure 6.4: SMD simulations of reduced QH_2. A) Pulling distance as a function of the SMD simulation time between the QH_2 headgroup (hydroxyl oxygen O2, in blue, and O5, in dashed blue) and the hydroxyl oxygen of Tyr-87$_{Nqo4}$. In the same graph the distances are plot to the second y-axis between the center of mass of QH_2 headgroup and the center of mass of stabilizing residues depending on the SMD simulation time. The figure highlights interactions with residues Arg-62$_{Nqo6}$ (in cyan), Tyr-249$_{Nqo6}$ (in magenta), and Ser-66$_{Nqo8}$ (black/gray). B-E) Snapshots from the SMD simulation when the QH_2 headgroup is stabilized by specific interactions with the selected amino acids. The snapshots are numbered in time sequence of the SMD trajectory. Snapshot of QH_2 headgroup interacting with (B) Arg-62$_{Nqo6}$ at 55 ns, (C) Tyr-249$_{Nqo8}$ at 56 ns, (D) Ser-66$_{Nqo8}$ at 66 ns, and (E) at 77 ns. Figure adopted from [7], license: CC BY-NC-ND 4.0.

tion coordinate (Fig. 6.6). Using the distance-based reaction coordinate, these trajectories would have a value around 6-8 Å and therefore would coincide with the other trajectories moving out of the Q tunnel. This suggests that the distance-based reaction coordinate is not unique in this region.

The time vs. reaction coordinate plots in Fig. 6.5 and 6.6 indicate that with extensive sampling, apart from a few gaps, nearly the entire range of the Q-binding tunnel could be captured. The projected reaction coordinate (Fig. 6.6) reveals sampling gaps, indicating regions of rare or nearly no sampling. The sampling gaps for Q_{ox} are at 8-10 Å and 18-20 Å, and for QH_2 at 5-7 Å and 33-35 Å of the projected reaction coordinate (Fig. 6.6).

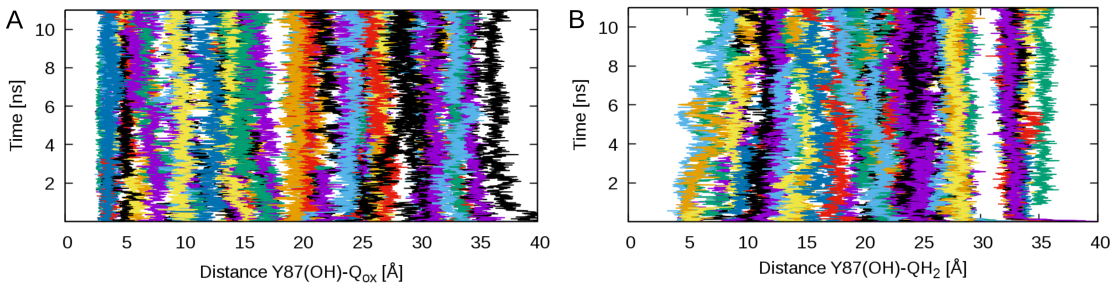


Figure 6.5: Sampling of the minimum Tyr-87$_{Nqo4}$(OH)-Q headgroup distance reaction coordinate (A) in 65 × 11 ns MD simulations that were employed for constructing diffusion model for Q_{ox}, and (B) in 54 × 11 ns MD simulations of QH_2.

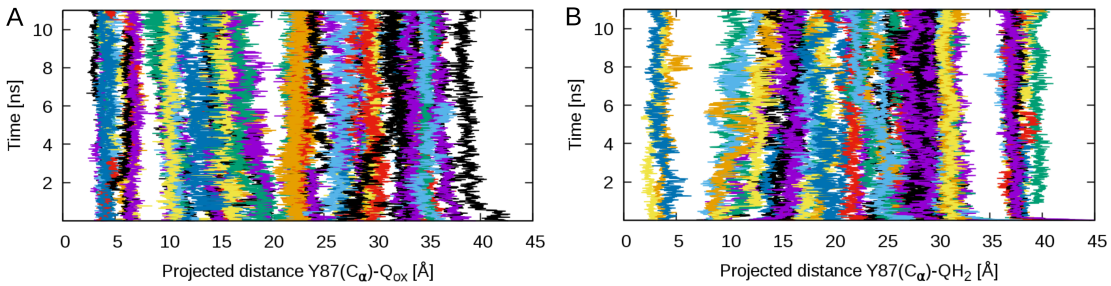


Figure 6.6: Sampling of the projected minimum Tyr-87$_{Nqo4}$(C_α)-Q headgroup distance reaction coordinate (see Method, Fig. 6.8) in (A) 65 × 11 ns MD simulations that were employed for constructing diffusion model for Q_{ox}, and (B) in 57 × 11 ns MD simulations of QH_2.

Attempts to fill the sampling gaps were unsuccessful by initiating additional trajectories from configurations close to these regions, suggesting that these regions have large uncertainties in the free energy profiles. The regions of enriched sampling were analyzed and representative configurations of Q headgroup with stabilizing interactions to protein side chains were extracted and are shown in Fig. 6.7. There are four preferred positions for Q_{ox} (Fig. 6.7 A, B) and QH_2 (Fig. 6.7 A, C), which coincide with the stabilizing interactions observed in the SMD simulations discussed above. The four preferred binding positions of Q can be summarized into two binding sites. The first binding site has two preferred positions (1, 1′), that are close to the N2 cluster in the upper part of the Q-binding tunnel. A second binding site with two preferred positions (2, 2′) are located close to the exit of the tunnel. The overview of the ubiquinone binding tunnel in Fig. 6.7 (A) also demonstrates that the exact location of the preferred positions is different for Q_{ox} and QH_2.

In binding site 1, the Q_{ox} headgroup forms a hydrogen bond to Tyr-87$_{Nqo4}$ (see 1 in Fig. 6.7 B). The next preferred position for Q_{ox} headgroup is at *ca.* 12 Å distance from Tyr-87$_{Nqo4}$ forming a stabilizing interaction with Phe-63$_{Nqo6}$ (see 1′ in Fig. 6.7 B), similar to what was observed in the SMD simulations. The second binding site has two preferred positions (2 and 2′) of Q_{ox} headgroup at 25-35 Å distance from Tyr-87$_{Nqo4}$. In binding site 2, the Q_{ox} headgroup forms interaction with the aromatic side chain of Trp-37$_{Nqo6}$ and Trp-241$_{Nqo8}$ (see 2 in Fig. 6.7 B). The fourth preferred position is the binding site 2′ when Q_{ox} headgroup interacts with Trp-241$_{Nqo8}$ from the opposite side compared to the binding site 2 (see 2′ in Fig. 6.7 B).

The QH_2 trajectories are dense and span a range of 10-32 Å, in projected reaction coordinate (Fig. 6.6), suggesting a proper sampling of this region. This region spans between two binding sites 1′ and 2. The majority of the QH_2 trajectories initiated from the primary binding site of Q_{ox} end in the binding site 1′ forming one or multiple hydrogen bonds with Gly-42$_{Nqo6}$, Thr-40$_{Nqo6}$, and Thr-54$_{Nqo6}$ at 10-15 Å distance from Tyr-87$_{Nqo4}$ (see 1′ in Fig. 6.7 C). A preferred binding position of QH_2 headgroup was observed at *ca.* 30 Å of the projected reaction coordinate, which is defined as binding site 2. At this position, QH_2 is surrounded by Trp-37$_{Nqo6}$, Trp-241$_{Nqo8}$, Tyr-249$_{Nqo8}$ and also interacts with two salt bridges, Arg-36$_{Nqo8}$/Asp-62$_{Nqo8}$ and Arg-62$_{Nqo6}$/Glu-35$_{Nqo8}$ (see 2 in Fig. 6.7 C). Two other

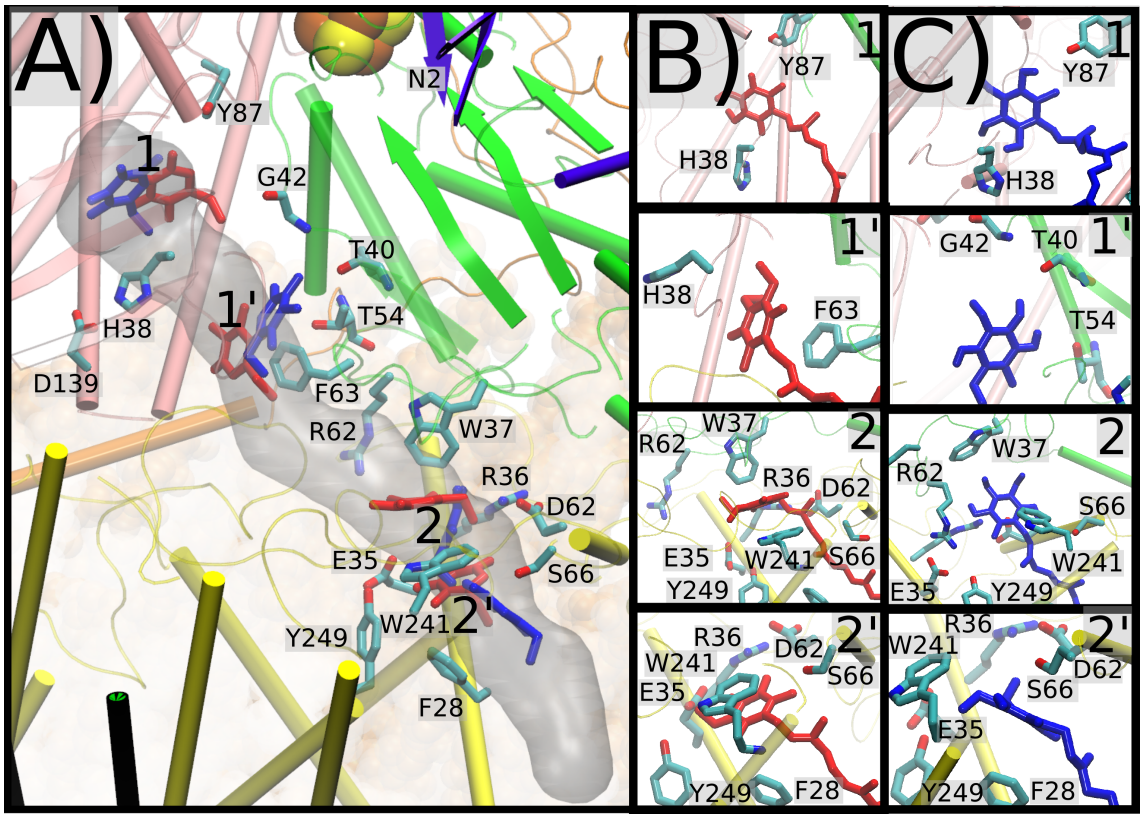


Figure 6.7: Structure of the Q-tunnel, with Q_{ox} (in red) and QH_2 (in blue) at transient binding sites 1, 1′, 2, and 2′. B) and C) close-ups of the Q-binding sites. Sites 1/1′ and 2/2′ correspond to the site close to iron-sulfur cluster N2 and the sites at the entrance of the Q tunnel, respectively. Figure adopted from [7], license: CC BY-NC-ND 4.0.

preferred positions of QH_2 were found, clearly separated from the properly sampled region described above. QH_2 makes hydrogen bond contact with Tyr-87_{Nqo4} in site 1 at 2-5 Å of the projected reaction coordinate. This interaction was observed in only a few of the 11-ns trajectories initiated from the primary binding site (Fig. 6.6 B). At binding site 2′ at 35-40 Å of the projected reaction coordinate (Fig. 6.6 B), the QH_2 headgroup interacts with Trp-241_{Nqo8} on the opposite side compared to the binding site 2, and with the side chain of Phe-28_{Nqo8}. At this site, QH_2 is additionally stabilized by a hydrogen bond interaction with the backbone of Ser-66_{Nqo8} (see 2′ in Fig. 6.7 C) similar to what was observed in the SMD simulation. The separation of the trajectories at distance reaction coordinate of 31 Å, indicates that binding site 2 and 2′ are separated by significant free-energy barrier, which leads to a sampling gap between 28-32 Å distances from Tyr-87_{Nqo4}.

6.1.3 PMF Profiles and Diffusion Coefficients of a Bound Ubiquinone/ol

Time series of reaction coordinate from all equilibrium MD trajectories were used to extract a PMF and profiles of the diffusion coefficients (D) under the assumption that the dynamics along the reaction coordinate is diffusive. The free energy and diffusion profiles are determined by matching the observed and predicted time evolution of the

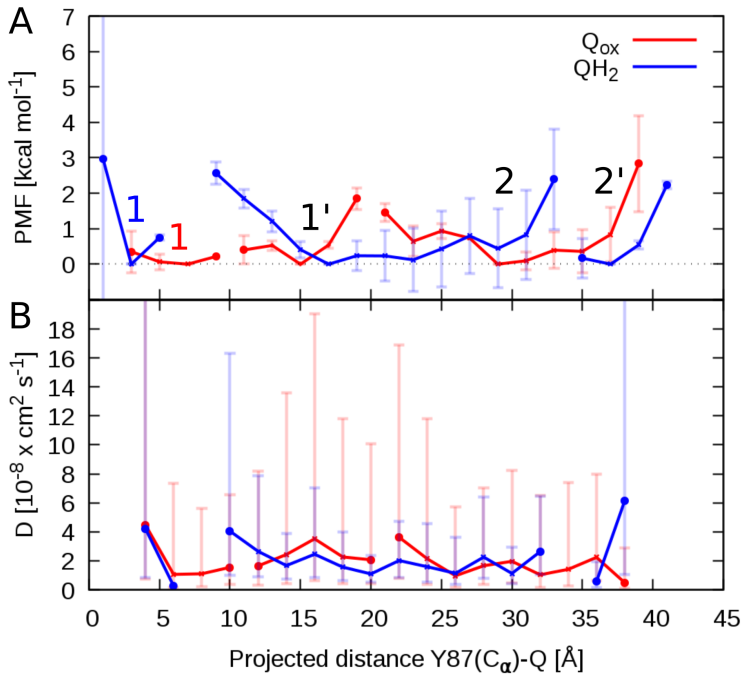


Figure 6.8: Free energy profile (PMF) and diffusion coefficient of Q_{ox} (red) and QH_2 (blue) in the ubiquinone tunnel. A) Free energy profiles as a function of the projected Tyr-87$_{Nqo4}$(C_α)-Q headgroup distance reaction coordinate, and (B) corresponding position-dependent diffusion coefficient profiles. The two PMF profiles are shifted vertically with their global minimum set to zero. 1, 1′, 2, and 2′ in panel A indicate local minima in the Q-binding cavity for the Q_{ox} and QH_2 headgroup. Marked gaps in the profiles indicate unresolved free energy differences in rarely sampled areas of the reaction coordinate (Fig. 6.6). Error bars in the PMF indicate standard errors of the mean, which were estimated by block averaging (see Methods).

probability distributions along the reaction coordinate [122]. The dynamics of a diffusion model are matched with the observed simulation data to find the best free energy and diffusion coefficients profile. In this section mainly the PMF and D profiles based on the projected Tyr-87$_{Nqo4}$(C_α)-Q distance reaction coordinate are discussed, because the sampling gaps are more clear and therefore one can easier judge the confidence of the PMFs. Additionally one is able to include the first binding site of QH_2 close to Tyr-87$_{Nqo4}$, which is only representable with the projected reaction coordinate.

The PMF and D profiles for Q_{ox} and QH_2 are shown in Fig. 6.8. The PMF profiles are flat with small energy amplitudes of less than 3 kcal mol^{-1}. The four previously discussed favorable binding positions of both Q species are also visible and coincide with the local minima in the PMF profiles. Overall, the Q tunnel can be divided into a part connecting sites 1, 1′, and 2 and separated by cluster of aromatic residues including Trp-241$_{Nqo8}$ at the tunnel entrance around site 2′. The PMF profile of Q_{ox} in the first binding site has local minima at 4 Å (1), and 14 Å (1′) of the projected reaction coordinate, and local minima at 28 Å (2), and 34 Å (2′) in the second binding site. The PMF is interrupted at two positions due to sampling gaps between 8-10 Å and 18-20 Å. The four local minima in the PMF profile of QH_2 of the first and second binding site are at 2 Å (1), 16 Å (1′), 28 Å (2), and 36 Å (2′) along the projected reaction coordinate. In site 1, where QH_2 forms a contact with Tyr-87$_{Nqo4}$ (see 1 in Fig. 6.7 C) is only resolved in the projected reaction coordinate and

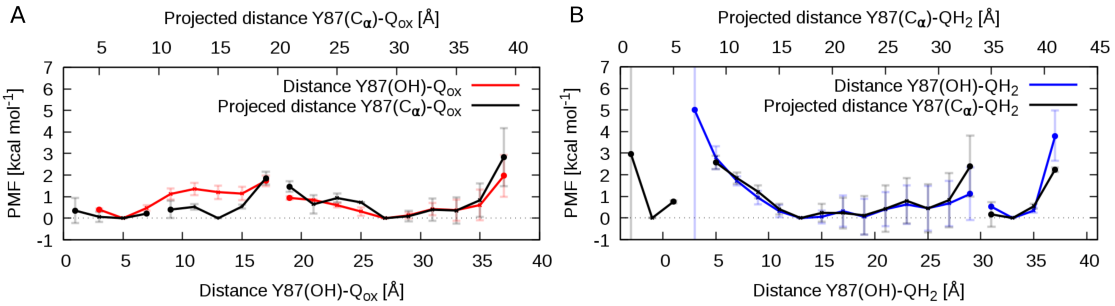


Figure 6.9: PMF profiles of Q_{ox} (A) and QH_2 (B) in two reaction coordinates. A) PMF profiles of Q_{ox} using the projected Tyr-87$_{Nqo4}$(C_α)-Q_{ox} headgroup distance (black) and the Tyr-87$_{Nqo4}$(OH)-Q_{ox} headgroup distance (red) as reaction coordinate. B) PMF profiles of reduced ubiquinol using the projected Tyr-87$_{Nqo4}$(C_α)-QH_2 headgroup distance (black) and the Tyr-87$_{Nqo4}$(OH)-QH_2 headgroup distance (blue) as reaction coordinate. Error bars in the PMFs indicate standard errors of the mean, which were estimated by block averaging (see Methods).

is not connected to the rest of the PMF profile. Due to clear sampling gaps at the regions 4-8 Å and 32-34 Å, the QH_2 PMF profile is interrupted. A main finding of the QH_2 PMF profile is that the local minimum 1′ at 16 Å of the projected reaction coordinate relates to 12 Å distance to Tyr-87$_{Nqo4}$ and coincides with a position found for a bound ubiquinone analog in *Y. lipolytica* structure from Zickermann *et al.* [53]. The distinct sampling gap at the second binding site at 32-34 Å indicates that minimum 2 and 2′ are separated by significant free-energy barrier, which leads to a sampling gap between 28 and 32 Å distance from Tyr-87$_{Nqo4}$ (Fig. 6.5).

The distance-dependent diffusion coefficients for Q_{ox} and QH_2 are shown in Fig. 6.8 (B). The values obtained for the diffusion of Q in the complex I Q binding tunnel are in the range of D = 1-4 $\times$ 10^{-8} cm^2 s^{-1}. Despite large statistical uncertainties, these values are about one order of magnitude smaller than the diffusion coefficients of a lipid in a typical membrane environment and of ubiquinone in vesicles [161]. Therefore, in the absence of a significant barrier the roughly estimated exit time for Q would be $\tau_{exit} \sim (30\ Å)^2/D \sim 0.01$ ms. Based on a Arrhenius correction, a barrier of 3 kcal mol^{-1} would slow down the exit time to about 1 ms, which is within the range of the experimental turnover rate measured of complex I [39].

To check the independence of the PMF profiles from a specific choice of the reaction coordinate, the PMFs were calculated and compared in both reaction coordinates (Fig. 6.9). The superposition of the PMFs with two reaction coordinate in Fig. 6.9, shows an overall agreement of PMF profiles. It reveals the two binding sites with for minimum preferred binding positions (1, 1′, 2, 2′) for Q_{ox} and QH_2, which are consistent in both reaction coordinates, with an exception of binding site 1 for QH_2, which could only be resolved using the projected reaction coordinate.

In summary, the main findings from the MD simulations of Q species in the Q binding cavity of complex I are: (1) that QH_2 moves away from Tyr-87$_{Nqo4}$ to the site 1′, occupied by ubiquinone analog in a structure *Y. lipolytica* complex I [53]; (2) A second binding site emerges close to the opening of the Q tunnel to the membrane, where the Q headgroup

forms tight interactions with a cluster of aromatic and charged amino acids; and (3) the motion of Q species along the tunnel is quite easy and smooth running, with a lower diffusion coefficient by a factor ten in comparison with free diffusion in a membrane with relatively modest free energy barrier.

6.1.4 Salt Bridges Analysis

A salt bridge analysis was performed for 183 ion pairs in complex I subunit Nqo4 and Nqo6-Nqo10. These salt bridges probably have a functional relevance. To probe the coupling between the Q dynamics and the conformational state of ion pairs, seven ion pairs were selected, which are in close proximity to either the Q binding cavity or close to the antiporter-like subunits in the membrane domain of complex I. The seven ion pairs (Asp-49_{Nqo7}/Lys-146_{Nqo8}, Asp-392_{Nqo4}/Arg-217_{Nqo4}, Glu-213_{Nqo8}/Arg-294_{Nqo8}, Glu-213_{Nqo8}/Arg-216_{Nqo8}, Asp-139_{Nqo4}/His-38_{Nqo4}, Asp-55_{Nqo6}/Arg-62_{Nqo6}, and Asp-62_{Nqo8}/Arg-36_{Nqo8}) are shown in simulations of Q_{ox} at sites 1, 1′, and 2 in the Q binding tunnel in Fig. 6.10 (A). The conformational state of these ion pairs depends on the Q position. For example the ion pair Asp-55_{Nqo6}/Arg-62_{Nqo6} is broken when Q_{ox} headgroup is in the binding site 1 and is formed when Q_{ox} headgroup is close to this ion pair in binding site 1′ and beyond in binding site 2 and 2′. Nearly all of the studied ion pairs remain stable salt bridges in simulations with Q_{ox}, whereas in the simulations with QH_2 they break with the exception of ion pair Asp-62_{Nqo8}/Arg-36_{Nqo8}, which is most of the time close. This effect was observed with ion pair Asp-139_{Nqo4}/His-38_{Nqo4} (consistent with earlier findings [72]), Asp-49_{Nqo7}/Lys-146_{Nqo8}, Asp-392_{Nqo4}/Arg-217_{Nqo4}, Asp-55_{Nqo6}/Arg-62_{Nqo6}.

6.2 Discussion

Experimental results [81] suggest that there is one tightly bound Q in complex I of *E. coli* with a ratio of 1.3 (± 0.1) per FMN molecule. However, due to the absence of a structurally resolved Q molecule bound to complex I, it remains unknown where the Q binds in the *ca.* 40 Å-long cavity. The calculated PMFs in this work suggest that Q could bind either at the primary binding site near the N2 FeS center, or at a second binding site close to the entrance of the Q tunnel. The second binding site is *ca.* 30 Å away from Tyr-87_{Nqo4} and it is surrounded by aromatic and charged groups stabilizing the Q headgroup.

In the standard single-Q model, a Q from the membrane pool enters the Q tunnel at the second binding site and the headgroup migrates through the *ca.* 40-Å long tunnel. Arriving at the primary binding site close to N2 cluster, the Q gets reduced and travels back to binding site 2, and finally diffuses out into the membrane. The diffusion coefficient of Q species inside the tunnel is ten times slower than that of free diffusion of Q. The exit time under most ideal circumstances (large Q pool, completely unrestrained motion of Q in the tunnel, and barriers in the PMF of maximal 3 kcal mol^{-1}) is estimated around 1 ms. This

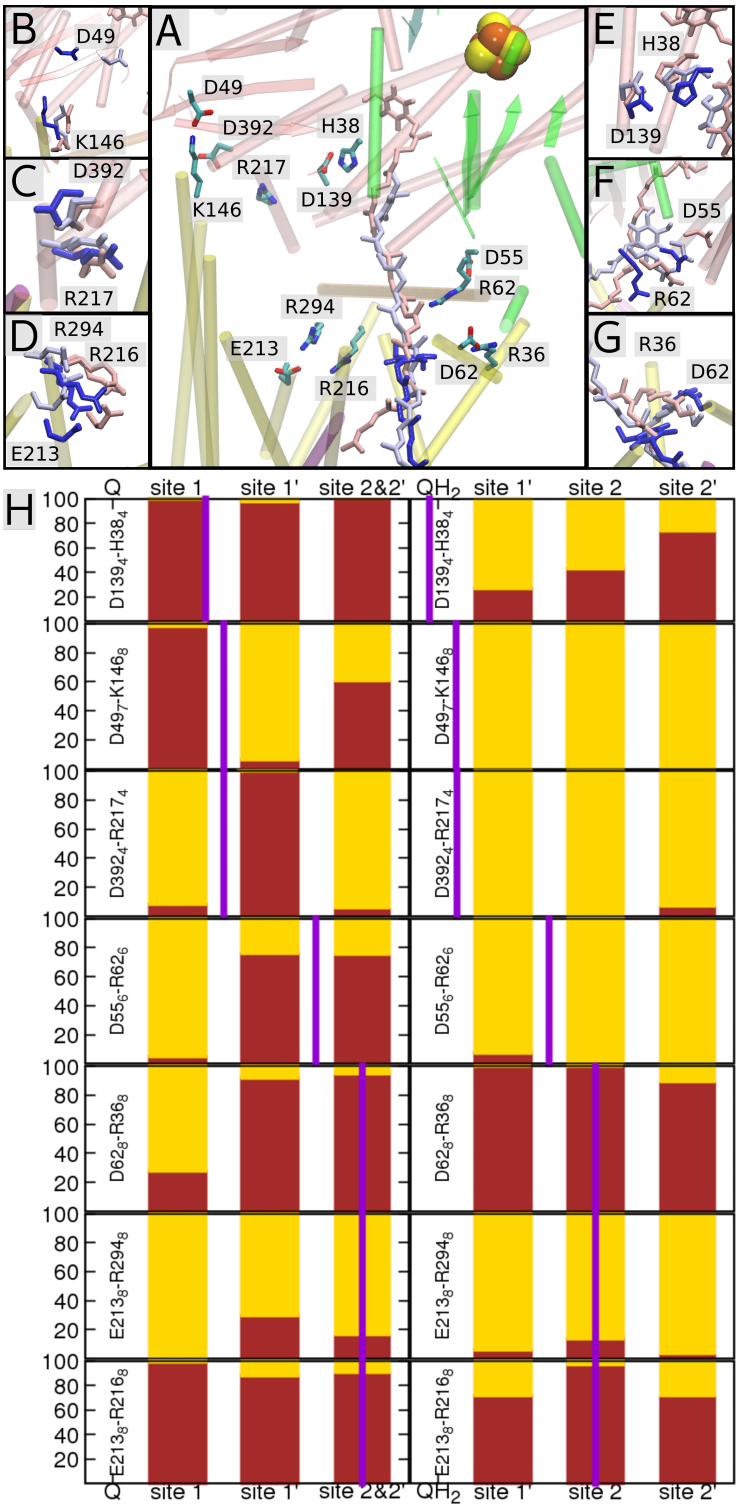


Figure 6.10: Salt bridge analysis of ion pairs in subunit Nqo4, and Nqo6-Nqo10 based on equilibrium MD simulations of Q_{ox} and QH_2 at different positions of the Q tunnel. A) Overview of Q-binding tunnel with three ubiquinone molecules are superimposed, bound to binding site 1 (pink), 1′ (ice blue), and 2 (blue), and seven salt bridges around the Q-binding sites. B)-G) The closeup view shows three superimposed snapshots (pink, ice blue, blue), in which Q_{ox} headgroup is bound to site 1 (pink), 1′ (ice-blue), and 2 (blue), for all the seven salt bridges. B) Asp-49$_{Nqo7}$/Lys-146$_{Nqo8}$, C) Asp-392$_{Nqo4}$/Arg-217$_{Nqo4}$, D) Glu-213$_{Nqo8}$/Arg-294$_{Nqo8}$ and Glu-213$_{Nqo8}$/Arg-216$_{Nqo8}$, E) Asp-139$_{Nqo4}$/His-38$_{Nqo4}$, F) Asp-55$_{Nqo6}$/Arg-62$_{Nqo6}$, and G) Asp-62$_{Nqo8}$/Arg-36$_{Nqo8}$. (H) Analysis of salt bridges formation (at distances $\leq$ 5 Å in yellow) and disruption (at distances $>$ 5 Å in brown) for each of the seven residue pairs, when Q_{ox} and QH_2 headgroup is in binding site 1, 1′, 2, and 2′ (see Fig. 6.8). The purple lines indicate the relative position of the ion pair to the Q binding sites. Figure adopted from [7], license: CC BY-NC-ND 4.0.

round-trip time is comparable to the experimental predicted turnover rate of complex I [39]. Any further increase in the barrier height would thus limit the overall turnover[1].

As an alternative to the standard single-Q model, it is also possible that a second Q binds to complex I to accomplish the bio-energetic function more efficiently [39], [77], [162]. Interpreting the results in this context, a single Q would be persistently bound in the Q binding tunnel and continuously shuttle between binding site 1 and 2. From the second binding site (2 or 2′) the electrons would have to be transferred to a second Q, which is in the membrane or close to binding site 2′.

The assumed high barrier in the QH_2 PMF profile between binding site 2 and 2′ supports the fact that the QH_2 might not exit the Q-binding tunnel. QH_2 stuck at the second binding site would need another membrane bound Q within 15 Å distance, to ensure an efficient electron transfer [71], for its oxidation. A putative Q bound at the surface of subunit Nqo8 would fulfill this criteria. Further support for the double-Q piston model could come from the conformational fluctuations of the E quartet [72], or conformational changes of Glu-218$_{\mathrm{Nqo8}}$ in the two crystal structures [16], [53]. Labeling studies [89] also support the hypothesis that a third Q binding site exist.

EPR measurements directly supports the possibility that there are multiple Q binding sites exist. Ohnishi *et al.* [59] found a fast relaxing (SQ_{Nf}) at *ca.* 10 Å and a slow relaxing (SQ_{Ns}) at 35 Å semiquinone distance to N2, corresponding to 4-10 Å, and 25-35 Å distance from Tyr-87$_{\mathrm{Nqo4}}$, respectively. More recent EPR data [91] reveal in addition a third signal of a very-slow-relaxing semiquinone (SQ_{Nvs}), which would be in a distant position and was suggested to originate from a Q bound at the membrane-protein interface [91]. In the context of the double-Q piston model one could interpret that SQ_{Nf} and SQ_{Ns} sit in the first and second binding sites.

Overall, the main findings of this work are: (1) the molecular structure of a putative second Q binding site is described. (2) The simulations of Q_{ox} and QH_2 suggest that a single Q shuttles between the first and the second binding sites, within the tight Q tunnel. (3) The diffusion process of Q in the 40 Å long Q tunnel is likely important for a strong coupling of the redox processes with Q and the proton pumps at the membrane domain, which are spatially separated by a large distance from each other.

[1]Note here that caution has to be taken against over interpretation of the estimated values, which may reflect in part the difficulties in equilibrating and sampling the motions of such a large and complex protein system.

Chapter 7

Assembly of Respiratory Complex I

7.1 Accessory Subunit NUMM of *Y. lipolytica* harbors a Zn^{2+} binding site

Complex I assembly takes place in multiple distinct intermediate steps, in which multiple assembly factors and accessory subunits are involved [43], however the detailed mechanism of complex I assembly is not completely understood. The functional role of the accessory subunit NUMM from *Y. lipolytica* in complex I assembly was elucidated by Kmita and coworkers [42]. I contributed to the understanding of the structural stability and function of accessory subunit NUMM in *Y. lipolytica* by applying MD and QM/MM simulation and QM calculation.

The accessory subunit NUMM is important for a proper assembly of complex I. Experimentally it was found, that deletion of the NUMM gene from the genome of *Y. lipolytica* (nummΔ) leads to disassembly of the accessory subunit N7BM [42]. Another effect is a significant reduction of the N4 FeS cluster signal was measured using EPR. The presence of NUMM is a prerequisite for a proper insertion of FeS clusters into complex I for reasons that remain unclear. In the nummΔ strain, the assembly factor N7BML, remains attached to complex I, suggesting that NUMM is required for the detachment of the assembly factor N7BML. The accessory subunit NUMM harbors, as the only subunit of complex I, a Zn^{2+} ion. The integrity of the Zn^{2+} binding site is needed for proper assembly of complex I. Site-directed mutations of the Zn^{2+} binding site reveal that mutation of Cys-128 lead to complete loss of subunit NUMM. Moreover, mutation of the homologs residue of Cys-115 in NDUFS6 in the human enzyme causes fatal neonatal lactic acidosis [49]. The relative position of Zn^{2+} in the *Y. lipolytica* complex I structure was determined, based on assignment of the Zn^{2+} electron density. The Zn^{2+} ion of NUMM binds in close distance to the N4 FeS cluster (16.4 Å) and the FeS cluster center of the 75 kDa subunit (20.2 Å).

A structural model of NUMM subunit could not be build based on the X-ray electron density. Instead a homology model of the C-terminal part of NUMM was constructed based on the structure of a bacterial homologs of NUMM RHOS4_26430 protein (PDB ID: 2JVM). This homology model was build by Prof. Zickermann[1] and provided for further theoretical studies of the structural stability of it.

The structural model of NUMM (Fig. 7.1) features a two stranded β-sheet (β2 and β3) and a three stranded β-sheet (β1, β4, and β5). The Zn^{2+} is coordinated by three cysteines

[1]Structural Bioenergetics Group, Institute of Biochemistry II, Medical School, Goethe University, 60438 Frankfurt am Main, Germany.

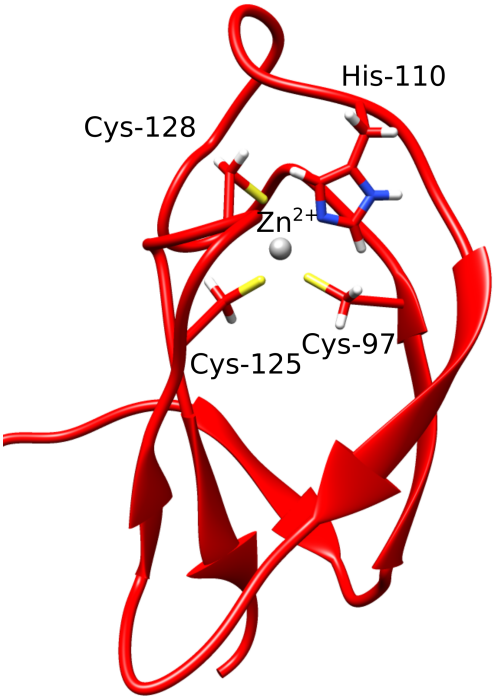


Figure 7.1: Snapshot at the end of a 5 ps QM/MM simulations of the structural model of the C-terminal domain of subunit NUMM constructed based on the homologous RHOS4_26430 protein (PDB ID: 2JVM).

(Cys-97, Cys-125, and Cys-128) and one histidine (His-110). Cys-97 is part of the β3 strand, Cys-125, and Cys-128 are located on the loop connecting β4 and β5 strand. His-110 is placed at the long β2-β3 loop.

Geometry optimizations at the DFT level of the Zn^{2+} binding site and QM/MM simulations of the complete NUMM subunit model were performed to test the stability of the model. To setup the calculation I added a Zn^{2+} ion to the homology model in the putative Zn-binding motif ($CX_8HX_{14}CX_2C$). The QM cluster calculations were performed in vacuum and with the implicit solvent model COSMO, mimicking the protein environment with ϵ=4. The NUMM structural model remains stable during 5 ps of QM/MM simulations performed with and without harmonic restraints on the C_α atoms. The changes in RMSD are small and converge to a plateau with and without harmonic restrains (Fig. 7.2). The bond distances and bond angles in the Zn^{2+} binding site were analyzes in QM/MM simulations and QM cluster calculations. Both geometry optimization and QM/MM (each performed in two different setups) predict consistently that the binding distances of the coordinating residues to Zn^{2+} are about 2.3-2.4 Å (Table 7.1) and that the angles between the Zn^{2+} and the coordinating residues are around 96-124°, suggesting that the Zn^{2+} ion remains tetrahedrally coordinated (Fig. 7.1, 7.3, Table 7.1). The simulations also reveal that the His-110 coordinating the Zn^{2+} is stabilized by a strong persistent hydrogen bond to the backbone of Pro-111 (Fig. 7.4).

Table 7.1: Structure of the zinc binding sites in subunit NUMM. Bond distances and angles calculated based on 5 ps restrained and unrestrained QM/MM simulations (mean and standard derivation), and from the optimized QM cluster models with COSMO (ϵ=4), and in vacuum. The QM systems were treated at the B3LYP-D/def2-SVP (for the C, H, N, and O atoms) and def2-TZVP (for the Zn) level.

Zn-binding site	QM/MM (restrained)	QM/MM	QM cluster (ϵ=4)	QM (vacuum)
Distances [Å]				
Zn-His110	2.35 ± 0.13	2.31 ± 0.15	2.15	2.22
Zn-Cys97	2.44 ± 0.08	2.44 ± 0.08	2.38	2.33
Zn-Cys125	2.36 ± 0.06	2.35 ± 0.06	2.35	2.36
Zn-Cys128	2.38 ± 0.07	2.35 ± 0.06	2.37	2.35
Angles [deg]				
His110-Zn-Cys128	93.6 ± 5.1	96.2 ± 6.4	100.3	96.1
His110-Zn-Cys125	105.3 ± 5.4	105.2 ± 5.2	100.4	98.5
His110-Zn-Cys97	99.5 ± 3.9	97.4 ± 5.6	99.7	106.4
Cys97-Zn-Cys128	107.1 ± 4.4	104.6 ± 6.6	117.9	116.0
Cys97-Zn-Cys125	116.0 ± 5.2	120.0 ± 6.8	108.2	115.4
Cys125-Zn-Cys128	127.8 ± 5.3	125.4 ± 6.1	124.5	119.0

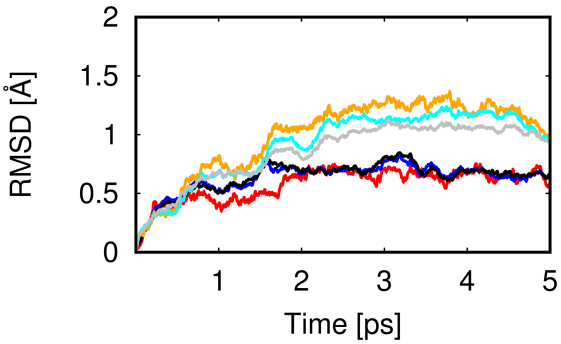


Figure 7.2: The root-mean square deviation (RMSD) of C_α atoms near the zinc-binding site calculated from 5 ps restrained (in red, blue, and black) and unrestrained (in orange, cyan, and gray) QM/MM molecular dynamics simulations (see Methods). The RMSD compared to the initial structure is shown in orange and red for the direct Zn^{2+} ligands, and for the C_α atoms of their flanking residues in sequence and within *ca.* 7 Å and 8.8 Å distance from the Zn^{2+}, shown in cyan/blue, and gray/black, respectively. The formation of tight interactions between Zn^{2+} and its four ligands, and the small RMSDs of 1-1.5 Å reached after 1 ps, are consistent with a stable zinc binding site.

The structural model of NUMM was used for a search in Brookhaven Protein Data Base for structurally similar proteins. The best candidate was COX4, an accessory subunit of cytochrome *c* oxidase from *Saccharomyces cerevisiae*, which also harbors a Zn^{2+} binding site with three cysteines and one histidine. Similar to NUMM, COX4 is essential for the correct assembly of cytochrome *c* oxidase [163]. In the structural alignment of NUMM and COX4, the structures superimpose well (Fig. 7.5). Both structures share the same overall secondary structural elements of two and three stranded β sheet. The Zn^{2+}-binding sites are superimpose very well, whereas in the loops away from the Zn-binding site they differ more.

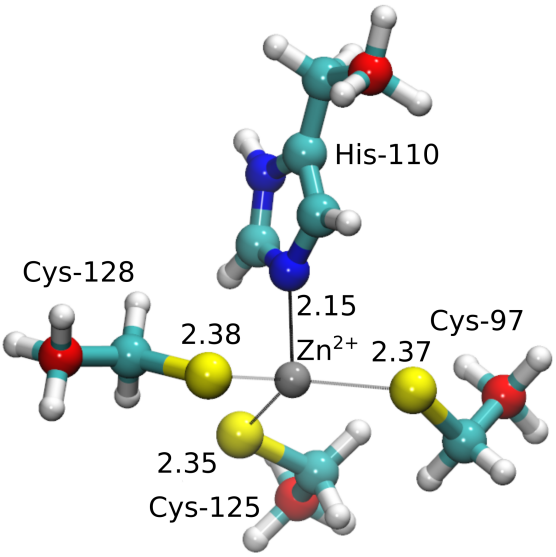


Figure 7.3: QM optimized geometry of the Zn^{2+}-binding site with fixed C_α atoms shown in red. The QM cluster model was constructed by modeling the protein surroundings using the conductor-like screening model (COSMO) with ϵ=4.

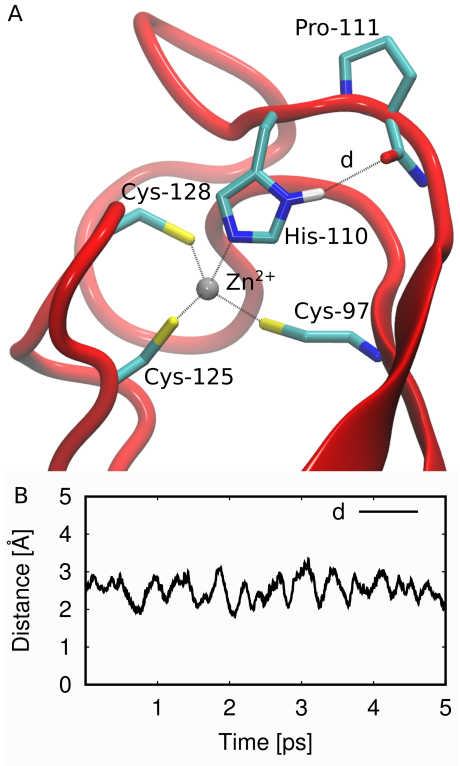


Figure 7.4: A) Snapshot from the QM/MM simulation. His-110 of the zinc binding site forms a hydrogen bond with the backbone of Pro-111. B) Distance of hydrogen bond between the backbone carbonyl of Pro-111 and the hydrogen of the delta nitrogen of His-110 in 5 ps QM/MM simulation.

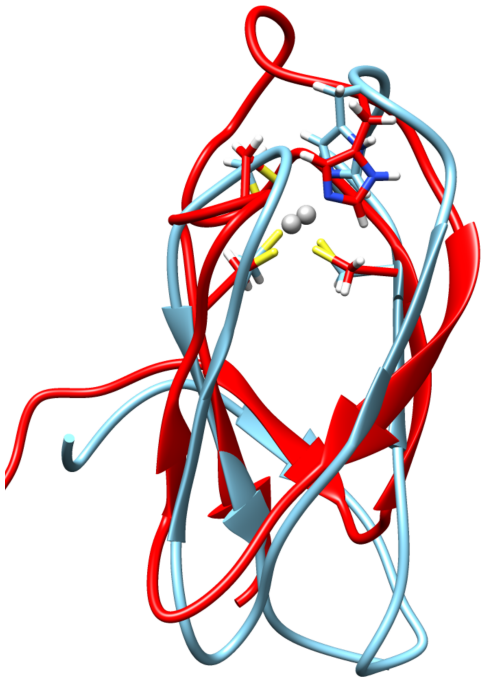


Figure 7.5: Structural alignment of the NUMM model with the cytochrome *c* oxidase subunit COX from *S. cerevisiae* (in light blue, PDB ID: 2ODX). The figure was prepared using chimera [164].

7.2 Discussion

The accessory subunit NUMM is required for a proper assembly of complex I in *Y. lipolytica*, as it is involved in a late assembly intermediate step of complex I. NUMM harbors a unique Zn^{2+} ion-binding site, the integrity of which is crucial for the subunit stability and complex I biogenesis. As it was not able to resolve the structure of NUMM experimentally, a homology model based on the structure of a bacterial homolog was built. We showed that the structural model of NUMM and also the Zn^{2+} binding is stable in both QM cluster calculations and in both QM/MM simulations.

The most similar structure to NUMM is COX4 from cytochrome *c* oxidase from *S. cerevisiae* [163]. COX4 binds a Zn^{2+} with the same set of ligands like NUMM. The Zn-binding site is also essential for the correct protein assembly. The results of the mutagenesis studies of NUMM are consistent with those found in COX4. Mutations of the cysteine residues in COX4 had a drastic effect on assembly, whereas mutations of the histidine residue are better tolerated [163]. A comparison of NUMM and COX4 suggesting that there is not only a structural similarity between them more over that they are also have similar functional role.

Upon deletion of NUMM, the assembly factor N7BML remains attached to the nearly complete assembled complex I. In the absence of NUMM and with bound N7BML, the N4 FeS site might become solvent exposed, which makes the complex I structure fragile and causes the loss of the FeS clusters. The NUMM center might have a possible role in the biogenesis of complex I by substituting N7BML after FeS cluster insertion, which would

lock the cluster entry site.

Interestingly, the same Zn-binding motif of three cysteines and one histidine as NUMM are also present in IscU [165] and mito-NEET [166], which are involved in the FeS transfer machinery of apo-ferredoxin. IscU and mito-NEET are also able to bind Zn^{2+} under specific conditions [167], [168]. Based on the similarity of these proteins to NUMM, it is possible that the subunit is also involved in FeS cluster transfer and insertion machinery. A possible way to further study this scenario would be to probe the stability of a Fe_2S_2 cluster in the Zn binding site of NUMM with QM and QM/MM simulation.

Chapter 8

Conclusion and Outlook

Na^+/H^+ antiporter are secondary active transporters. Depending on the organism they can work in both directions across the membrane. In human and probably in *M. jannaschii*, NHE works as Na^+ driven proton pump. These antiporters are well established drug targets in human [1]. Despite decades of research in multiple disciplines and with different methods, such as biochemical, structural, phylogenetic, biophysical, and theoretical studies, the molecular mechanism of how the ion transport couples to ion binding remains elusive.

This thesis contributes to a detailed mechanistical understanding of the Na^+/H^+ antiporter MjNhaP1. By applying complementary theoretical approaches (MD simulation, QM/MM simulation, QM cluster calculation and free energy calculation), I investigated the location and stability of the Na^+ binding site in MjNhaP1. The resolved structure of MjNhaP1 [34] is indeed a physiological meaningful inward-open conformational state. I successfully modeled a stable outward-open state of MjNhaP1 based on cryo-EM map (6 Å resolution) [34]. The Na^+ and K^+ binding behavior in the binding sites of MjNhaP1 in MD simulation and QM/MM simulation was characterized. The observed water molecules contribute to the coordination and to the stabilization of the bound ion. The number of additional water ligands depends on the bound ion (Na^+ or K^+) and on the conformational state of MjNhaP1 (inward- or outward-open). By applying free energy calculation, it was found that the inward-open MjNhaP1 is a Na^+ selective state (by 5.2 kcal mol^{-1} compared to K^+ binding), whereas the outward-open state has a negligible preference for Na^+ over K^+ (0.4 kcal mol^{-1}). Based on a sequence alignment of Na^+/H^+ antiporters with different ion selectivities, several mutations were suggested to change the ion selectivity of MjNhaP1. I estimated the individual contribution of residues to the selectivity of ion binding with the help of free energy calculation. The largest destabilizing effect was observed of Na^+ binding towards a K^+ selectivity in a double mutant A130S/P162A (by 2.5 kcal mol^{-1}).

Results of this work can provide valuable input for further mechanistic studies of MjNhaP1, with a final goal to reveal a complete molecular mechanism of ion transport in MjNhaP1. As an important next step to reach this goal, a transition path connecting inward- and outward-open state has to be found, with Na^+ and H^+ loaded in the binding site, by applying transition path sampling technique. Initially, a proper reaction coordinate has to be defined that separates the inward- and outward-open state. After an extensive sampling along the reaction coordinate, one could identify the transition state, the point at which half of the trajectories converge to the inward-open state and half of the trajectories converge to the outward-open state. Identification of the transition state is crucial for understanding the underlying mechanism, as the barrier height defines the time scale of

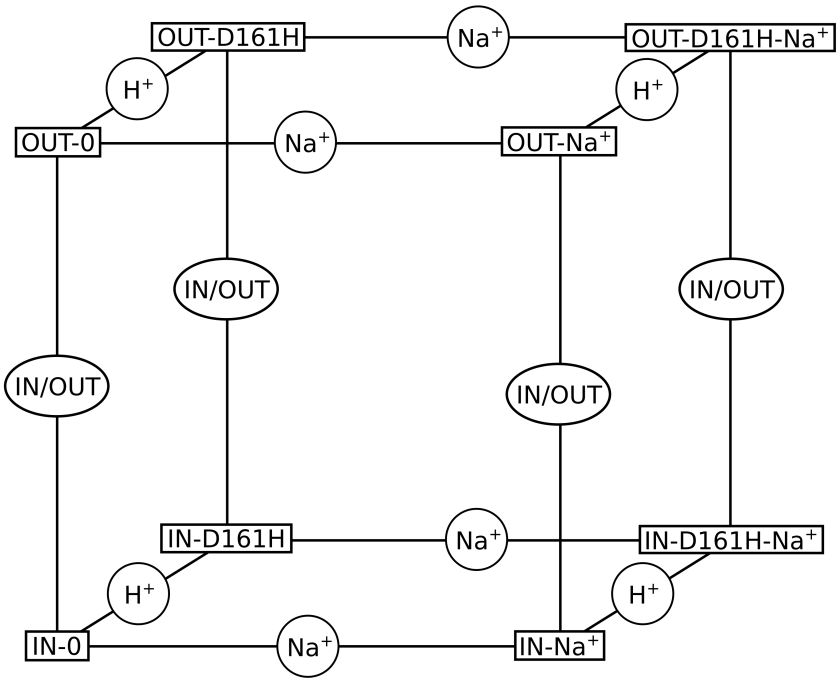


Figure 8.1: Thermodynamic cycle of ion transport in MjNhaP1. The horizontal paths refer to Na^+ binding and unbinding events. In the vertical paths the conformational inward/outward transition takes place. The third dimension covers H^+ binding and unbinding to Asp-161. IN-0: substrate unloaded inward-open state, OUT-0: substrate unloaded outward-open state, IN-Na^+: Na^+ bound inward-open state, OUT-Na^+: Na^+ bound outward-open state, IN-D161H: protonated Asp-161 in the inward-open state, OUT-D161H: protonated Asp-161 in the outward-open state, IN-D161H-Na^+: protonated Asp-161 in the Na^+ bound inward-open state, and OUT-D161H-Na^+: protonated Asp-161 in the Na^+ bound outward-open state.

the process in exponential law (Eyring or Kramers). By applying free energy perturbation calculation of the bound ion on the transition state, one can estimate the possible additional selectivity inferred from the different free energy of the transition state. Furthermore one can estimate the contribution of individual residues to the ion selectivity on the transition state, by free energy perturbation calculation of single mutants, extending the approach developed in this work for the inward-open state.

To complete the full transport cycle it is necessary to couple the energetics of the H^+ transport mechanism as well. The next step would be to perform pK_a calculations of the titrable residues in MjNhaP1. Of special interest is pK_a of Asp-161, which is mainly responsible for H^+ and Na^+ binding. In addition it is essential to investigate the influence of the protein conformational state and the presence of Na^+ ion on the pK_a value. In order to compare the preference of MjNhaP1 for Na^+, K^+, and H^+ binding, it is needed to estimate the protonation free energy of Asp-161, *e.g.*, by applying free energy perturbation calculation.

In a complete thermodynamic cycle in three dimensions covers Na^+ uptake, protonation of Asp-161, and the conformational transition of MjNhaP1 in Fig. 8.1. The aim would be to estimate the free energy differences between the eight states (IN-0, IN-Na^+, IN-D161H, IN-D161H-Na^+, OUT-0, OUT-Na^+, OUT-D161H, OUT-D161H-Na^+). This would allow

to identify the sequential ordering of every step in the transport mechanism, and would answer the following questions: Is the proton uptake spontaneous with bound Na^+ in one of the conformational states? or, Is the release of Na^+ necessary for H^+ binding?

Respiratory complex I is a redox-driven proton pump. It is the entry point for electrons into the electron transport chain originating from the oxidation of NADH. These electrons are transferred to ubiquinone, which is a central cofactor in the respiratory chain. The motion of Q in the Q tunnel is probably important to establish strong coupling between spatially separated electron transfer and proton transport processes in complex I. A widely accepted theory about the coupling mechanism between redox reaction and the proton transport machinery, is that upon ubiquinone reduction energy is transferred in form of conformational/electrostatic signal to activate the proton pumps in the membrane domain [74], [77]. This coupling mechanism was a focal point in the last few years by many groups with different approaches and techniques (structural strudies, EPR, time resolved spectroscopy, theoretical studies, biochemical experiments, *etc.*) [16, 53, 75–79]. However, the detailed mechanism on a molecular basis remains to be clarified.

I explored the dynamics of Q bound in the Q tunnel, as oxidized species Q_{ox} before the electron transfer from N2 and as reduced species QH_2 after the reduction and proton acceptance from the local environment (His-38$_{Nqo4}$/Asp-139$_{Nqo4}$ and Tyr-87$_{Nqo4}$). By an extensive sampling along the entire 40 Å long Q tunnel, four preferred binding sites were identified (1, 1′, 2, and 2′) of the Q_{ox} and QH_2 headgroup belonging to two binding sites. In combination with complementary approaches, we described for the first time the molecular structure of the second binding site of Q in the Q tunnel. The second site was identified at 25-35 Å away from N2 cluster and close to the membrane exit of the Q tunnel. Based on the MD simulation data extracted, free energy and difffusion profiles of Q_{ox} and QH_2 along the Q tunnel were calculated by applying the one-dimensional diffusion model [122]. Apart from a few gaps, the energy profiles are flat and predict that Q_{ox} and QH_2 can diffuse freely in most parts of the Q tunnel. Notably, the separation of the QH_2 trajectories at region 28-32 Å distance to Tyr-87$_{Nqo4}$ (Fig. 1.5 in chapter: The Bioenergetic Function of Respiratory Complex I) indicates that QH_2 headgroup may not switch between binding position 2 and 2′. This could imply that a reduced ubiquinol coming from the primary binding site might get trapped at the 2nd binding site, and be unable to leave the Q tunnel. Based on the distance dependent diffusion profiles the exit time of ubiquinone was estimated to be around 1 ms in the absence of a significant free-energy barrier. Therefore the forward and backward round-trip time for a single ubiquinone would be close to the catalytic turnover rate of complex I [39], and any further increase of the energy barrier would make it difficult for a single ubiquinone to fulfill the bioenergetic function in an appropriate time. These findings support the idea of a double Q-piston model [39], [7].

In further studies it would be interesting to probe the influence of individual residues on the free energy profile of Q, by applying the same procedure to extract PMF profiles based on short MD simulations with mutated residues in the Q tunnel (*e.g.* Trp-37$_{Nqo6}$, Trp-241$_{Nqo8}$, Arg-62$_{Nqo6}$, Glu-35$_{Nqo8}$, Arg-36$_{Nqo8}$, Asp-62$_{Nqo8}$, *ect.*). In this way one could also calculate PMF profiles of Q bound to complex I structures available from other species

(*e.g. Y. lipolytica*, *Bovine*). The overall goal, which would clarify the uncertainties about the Q binding sites, is high resolution complex I structures with resolved ubiquinone species in the primary and second binding sites.

The identification of the 2nd ubiquinone binding site is important in order to understand the coupling between the ubiquinone binding site and the proton pumps in the membrane domain, and strengthens the experimental evidence for multiple Q binding sites. The discovery of the second binding site opens new perspectives for complex I mechanism and raises at the same time new questions which are remained to be answered, for example:

- Where does a second ubiquinone bind if the substrate ubiquinol molecule occupies the second binding site?
- If the ubiquinol molecules do not exchange, how does the oxidoreduction process at the second binding site take place?
- What are possible electron and proton acceptors or transmitters?
- How is a conformational/electrostatical signal transferred from the second binding site to the proton pumps in the membrane arm?

Bibliography

[1] C. L. Brett, M. Donowitz, and R. Rao. Evolutionary origins of eukaryotic sodium/-proton exchangers. Am. J. Physiol. Cell Physiol. **288**, C223–C239 (2005).

[2] M. Karmazyn, X. T. Gan, R. A. Humphreys, H. Yoshida, and K. Kusumoto. The Myocardial Na^+/H^+ Exchange. Circ. Res. **85**, 777–786 (1999).

[3] P. Mitchell. A General Theory of Membrane Transport From Studies of Bacteria. Nature **180**, 134–136 (1957).

[4] O. Jardetzky. Simple Allosteric Model for Membrane Pumps. Nature **211**, 969–970 (1966).

[5] J. Warnau, D. Wöhlert, K. Okazaki, Ö. Yildiz, A. P. Gamiz-Hernandez, V. R. I. Kaila, W. Kühlbrandt, and G. Hummer. Ion Binding and Selectivity of the Na^+/H^+ Antiporter MjNhaP1 (in preparation).

[6] K. Okazaki, D. Wöhlert, J. Warnau, H. Jung, Ö. Yildiz, W. Kühlbrandt, and G. Hummer. Mechanism of the electroneutral sodium/proton antiporter PaNhaP from transition-path shooting. Nat. Commun. (in press).

[7] J. Warnau, V. Sharma, A. P. Gamiz-Hernandez, A. Di Luca, O. Haapanen, I. Vattulainen, M. Wikström, G. Hummer, and V. R. I. Kaila. Redox-coupled quinone dynamics in the respiratory complex I. Proc. Natl. Acad. Sci. U.S.A. **115**, E8413–E8420 (2018).

[8] D. S. Goodsell. *The Machinary of Life; 2nd Ed.* Springer, (2009).

[9] J. M. Berg, J. L. Tymoczko, and L. Stryer. *Biochemistry; 7th International ed.* W.H. Freeman, (2012).

[10] J. Hirst, M. S. King, and K. R. Pryde. The production of reactive oxygen species by complex I. Biochem. Soc. Trans. **36**, 976–980 (2008).

[11] M. P. Murphy. How mitochondria produce reactive oxygen species. J. Biochem. **417**, 1–13 (2009).

[12] T. M. Dawson and V. L. Dawson. Molecular Pathways of Neurodegeneration in Parkinson's Disease. Science **302**, 819–823 (2003).

[13] R. S. Balaban, S. Nemoto, T. Finkel, and C. Branch. Mitochondria, Oxidants, and Aging Review. Cell **120**, 483–495 (2005).

[14] A. Knopp, S. Thierfelder, B. Doepner, and K. Benndorf. Mitochondria are the main ATP source for a cytosolic pool controlling the activity of ATP-sensitive K^+ channels in mouse cardiac myocytes. Cardiovasc. Res. **52**, 236–245 (2001).

[15] P. Mitchell. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. Nature **191**, 144–148 (1961).

[16] R. Baradaran, J. M. Berrisford, G. S. Minhas, and L. A. Sazanov. Crystal structure of the entire respiratory complex I. Nature **494**, 443–448 (2013).

[17] F. Sun, X. Huo, Y. Zhai, A. Wang, J. Xu, D. Su, M. Bartlam, and Z. Rao. Crystal Structure of Mitochondrial Respiratory Membrane Protein Complex II. Cell **121**, 1043–1057 (2005).

[18] Z. Zhang, H. Lishar, V. M. Shulmeister, Y.-I. Chi, K. K. Kim, L.-W. Hung, A. Crofts, E. A. Berry, and S.-H. Kim. Electron transfer by domain movement in cytochrome. Nature **392**, 677–684 (1998).

[19] T. Tsukihara, K. Shimokata, Y. Katayama, H. Shimada, K. Muramoto, H. Aoyama, M. Mochizuki, K. Shinzawa-itoh, E. Yamashita, M. Yao, Y. Ishimura, and S. Yoshikawa. The low-spin heme of cytochrome c oxidase as the driving element of the proton-pumping process. Proc. Natl. Acad. Sci. U.S.A. **100**, 15304–15309 (2003).

[20] D. Stock, A. G. W. Leslie, and J. E. Walke. Molecular Architecture of the Rotary Motor in ATP Synthase. Science **286**, 1700–1706 (1999).

[21] J. Kim. Channelopathies. Pediatr. **57**, 1–18 (2014).

[22] H. R. Zhekova, V. Ngo, D. Salahub, and S. Noskov. Selective ion binding and transport by membrane proteins – A computational perspective. Coord. Chem. Rev. **345**, 108–136 (2017).

[23] Y. Zhou, J. H. Morais-Cabral, A. Kaufman, and R. Mackinnon. Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 $\AA$ resolution. Nature **414**, 43–48 (2001).

[24] J. H. Morais-Cabral, Y. Zhou, and R. Mackinnon. Energetic optimization of ion conduction rate by the K^+ selectivity filter. Nature **414**, 37–42 (2001).

[25] D. M. Scott. Sodium Cotransport Systems: Cellular, Molecular and degulatory Aspects. BioEssays **7**, 71–78 (1987).

[26] C. A. Shintre, A. C. W. Pike, Q. Li, J.-i. Kim, A. J. Barr, S. Goubin, L. Shrestha, J. Yang, G. Berridge, J. Ross, P. J. Stansfeld, M. S. P. Sansom, A. M. Edwards, C. Bountra, B. D. Marsden, A. N. Bullock, O. Gileadi, N. A. Burgess-brown, and E. P. Carpenter. Structures of ABCB10, a human ATP-binding cassette transporter in apo- and nucleotide-bound states. Proc. Natl. Acad. Sci. U.S.A. **110**, 1–6 (2013).

[27] L. R. Forrest, R. Krämer, and C. Ziegler. The structural basis of secondary active transport mechanisms. BBA - Bioenergetics **1807**, 167–188 (2011).

[28] D. Taglicht, E. Padan, and S. Schuldiner. Overproduction and Purification of a Functional Na^+/H^+ Antiporter Coded by nhaA (ant) from *Escherichia coli*. J. Biol. Chem. **266**, 11289–11294 (1991).

[29] C. Hunte, E. Screpanti, M. Venturi, A. Rimon, E. Padan, and H. Michel. Structure of a Na^+/H^+ antiporter and insights into mechanism of action and regulation by pH. Nature **435**, 1197–1202 (2005).

[30] I. T. Arkin, H. Xu, M. Jensen, E. Arbely, E. R. Bennett, K. J. Bowers, E. Chow, R. O. Dror, M. P. Eastwood, R. Flitman-Tene, B. A. Gregersen, J. L. Klepeis, I. Kolossváry, Y. Shan, and D. E. Shaw. Mechanism of Na^+/H^+ antiporting. Science **317**, 799–803 (2007).

[31] M. Xiang, M. Feng, S. Muend, and R. Rao. A human Na^+/H^+ antiporter sharing evolutionary origins with bacterial NhaA may be a candidate gene for essential hypertension. Proc. Natl. Acad. Sci. U.S.A. **104**, 18677–18681 (2007).

[32] T. Mager, A. Rimon, E. Padan, and K. Fendler. Transport Mechanism and pH Regulation of the Na^+/H^+ Antiporter NhaA from *Escherichia coli*. J. Biol. Chem. **286**, 23570–23581 (2011).

[33] C. Lee, H. J. Kang, C. Von Ballmoos, S. Newstead, P. Uzdavinys, D. L. Dotson, S. Iwata, O. Beckstein, A. D. Cameron, and D. Drew. A two-domain elevator mechanism for sodium/proton antiport. Nature **501**, 573–577 (2013).

[34] C. Paulino, D. Wöhlert, E. Kapotova, Ö. Yildiz, and W. Kühlbrandt. Structure and transport mechanism of the sodium/proton antiporter MjNhaP1. eLife **3**, e03583 (2014).

[35] D. Wöhlert, W. Kühlbrandt, and Ö. Yildiz. Structure and substrate ion binding in the sodium/proton antiporter PaNhaP. eLife **3**, e03579 (2014).

[36] P. Goswami, C. Paulino, D. Hizlan, J. Vonck, Ö. Yildiz, and W. Kühlbrandt. Structure of the archaeal Na^+/H^+ antiporter NhaP1 and functional role of transmembrane helix 1. EMBO J. **30**, 439–449 (2011).

[37] J. Hellmer, A. Teubner, and C. Zeilinger. Conserved arginine and aspartate residues are critical for function of MjNhaP1, a Na^+/H^+ antiporter of *M. jannaschii*. FEBS Lett. **542**, 32–36 (2003).

[38] U. Brandt. Energy converting NADH:quinone oxidoreductase (complex I). Annu. Rev. Biochem. **75**, 69–92 (2006).

[39] M. Wikström, V. Sharma, V. R. I. Kaila, J. P. Hosler, and G. Hummer. New perspectives on proton pumping in cellular respiration. Chem. Rev. **115**, 2196–2221 (2015).

[40] J. Hirst. Mitochondrial Complex I. Annu. Rev. Biochem. **82**, 551–575 (2013).

[41] A. H. V. Schapira. Human complex I defects in neurodegenerative diseases. BBA - Bioenergetics **1364**, 261–270 (1998).

[42] K. Kmita, C. Wirth, J. Warnau, S. Guerrero-Castillo, C. Hunte, G. Hummer, V. R. I. Kaila, K. Zwicker, U. Brandt, and V. Zickermann. Accessory NUMM (NDUFS6) subunit harbors a Zn-binding site and is essential for biogenesis of mitochondrial complex I. Proc. Natl. Acad. Sci. U.S.A. **112**, 5685–5690 (2015).

[43] M. Mimaki, X. Wang, M. Mckenzie, D. R. Thorburn, and M. T. Ryan. Understanding mitochondrial complex I assembly in health and disease. BBA - Bioenergetics **1817**, 851–862 (2012).

[44] R. Lill, B. Hoffmann, S. Molik, A. J. Pierik, N. Rietzschel, O. Stehling, M. A. Uzarska, H. Webert, C. Wilbrecht, and U. Mühlenhoff. The role of mitochondria in cellular iron–sulfur protein biogenesis and iron metabolism. BBA - Mol. Cell Res. **1823**, 1491–1508 (2012).

[45] C. Yip, M. E. Harbour, K. Jayawardena, I. M. Fearnley, and L. A. Sazanov. Evolution of Respiratory Complex I. J. Biol. Chem. **286**, 5023–5033 (2011).

[46] L. Giachini, F. Francia, F. Boscherini, C. Pacelli, T. Cocco, S. Papa, and G. Venturoli. EXAFS reveals a structural zinc binding site in the bovine NADH-Q oxidoreductase. FEBS Lett. **581**, 5645–5648 (2007).

[47] K. Shinzawa-itoh, J. Seiyama, H. Terada, R. Nakatsubo, K. Naoki, Y. Nakashima, and S. Yoshikawa. Bovine Heart NADH - Ubiquinone Oxidoreductase Contains One Molecule of Ubiquinone with ten Isoprene Units as One of the Cofactors. Biochemistry **49**, 487–492 (2010).

[48] D. M. Kirby, R. Salemi, C. Sugiana, A. Ohtake, L. Parry, K. M. Bell, E. P. Kirk, A. Boneh, R. W. Taylor, H.-h. M. Dahl, M. T. Ryan, and D. R. Thorburn. NDUFS6 mutations are a novel cause of lethal neonatal mitochondrial complex I deficiency. J. Clin. Invest. **114**, 837–845 (2004).

[49] R. Spiegel, A. Shaag, H. Mandel, D. Reich, M. Penyakov, Y. Hujeirat, A. Saada, O. Elpeleg, and S. A. Shalev. Mutated NDUFS6 is the cause of fatal neonatal lactic acidemia in Caucasus Jews. Eur. J. Hum. Genet. **17**, 1200–1203 (2009).

[50] B. Ke, S. Pepe, D. R. Grubb, J. C. Komen, A. Laskowski, F. A. Rodda, B. M. Hardman, J. J. Pitt, M. T. Ryan, M. Lazarou, J. Koleff, M. M. H. Cheung, J. J. Smolich, and D. R. Thorburn. Tissue-specific splicing of an Ndufs6 gene-trap insertion generates a mitochondrial complex I deficiency-specific cardiomyopathy. Proc. Natl. Acad. Sci. U.S.A. **109**, 6165–6170 (2012).

[51] J. M. Berrisford and L. A. Sazanov. Structural basis for the mechanism of respiratory complex I. Journal of Biological Chemistry **284**, 29773–29783 (2009).

[52] R. G. Efremov and L. A. Sazanov. Structure of the membrane domain of respiratory complex I. Nature **476**, 414–421 (2011).

[53] V. Zickermann, C. Wirth, H. Nasiri, K. Siegmund, H. Schwalbe, C. Hunte, and U. Brandt. Mechanistic insight from the crystal structure of mitochondrial complex I. Science **347**, 44–49 (2015).

[54] K. Parey, U. Brandt, H. Xie, D. J. Mills, K. Siegmund, J. Vonck, W. Kühlbrandt, and V. Zickermann. Cryo-EM structure of respiratory complex I at work. eLife **7**, e39213 (2018).

[55] K. Fiedorczuk, J. A. Letts, G. Degliesposti, K. Kaszuba, M. Skehel, and L. A. Sazanov. Atomic structure of the entire mammalian mitochondrial complex I. Nature **538**, 406–410 (2016).

[56] A. Agip, J. N. Blaza, H. R. Bridges, C. Viscomi, S. Rawson, S. P. Muench, and J. Hirst. Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states. Nat. Struct. Mol. Biol. **25**, 548–556 (2018).

[57] J. Zhu, K. R. Vinothkumar, and J. Hirst. Structure of mammalian respiratory complex I. Nature **536**, 354–358 (2016).

[58] J. N. Blaza, K. R. Vinothkumar, and J. Hirst. Structure of the Deactive State of Mammalian Respiratory Complex I. Structure **26**, 312–319 (2018).

[59] T. Yano, W. R. Dunham, and T. Ohnishi. Characterization of the $\Delta\mu H^+$-Sensitive Ubisemiquinone Species (SQNf) and the Interaction with Cluster N2: New Insight into the Energy-Coupled Electron Transfer in Complex I. Biochemistry **44**, 1744–1754 (2005).

[60] M. A. Tocilescu, V. Zickermann, K. Zwicker, and U. Brandt. Quinone binding and reduction by respiratory complex I. BBA **1797**, 1883–1890 (2010).

[61] C. Hunte, V. Zickermann, and U. Brandt. Functional modules and structural basis of conformational coupling in mitochondrial complex I. Science (New York, N.Y.) **329**, 448–451 (2010).

[62] J. G. Fedor, A. J. Y. Jones, A. D. Luca, V. R. I. Kaila, and J. Hirst. Correlating kinetic and structural data on ubiquinone binding and reduction by respiratory complex I. Proc. Natl. Acad. Sci. U.S.A. **114**, 12737–12742 (2017).

[63] K. Huoponen, J. Vilkki, P. Aula, E. K. Nikoskelainen, and M.-L. Savontaus. A New mtDNA Mutation Associated with Leber Hereditary Optic Neuroretinopathy. Am. J. Hum. Genet. **48**, 1147–1153 (1991).

[64] D. A. Mackey, R. J. Oostra, T. Rosenberg, E. Nikoskelainen, J. Bronte-Stewart, J. Poulton, A. E. Harding, G. Govan, P. A. Bolhuis, and S. Norby. Primary pathogenic mtDNA mutations in multigeneration pedigrees with Leber hereditary optic neuropathy. Am. J. Hum. Genet. **59**, 481–485 (1996).

[65] C. Piccoli, M. Ripoli, R. Scrima, P. Stanziale, M. Di Ianni, L. Moretti, B. Biscottini, M. Carella, D. Boffoli, A. Tabilio, and N. Capitanio. Mutation associated with mitochondrial dysfunction in megakaryoblastic leukaemic cells. Leukemia **22**, 1938–1972 (2008).

[66] E. L. Blakely, R. D. Silva, A. King, V. Schwarzer, T. Harrower, G. Dawidek, D. M. Turnbull, and R. W. Taylor. LHON/MELAS overlap syndrome associated with a mitochondrial MTND1 gene mutation. Eur. J. Hum. Genet. **13**, 623–627 (2005).

[67] J. Pätsi, P. Maliniemi, S. Pakanen, R. Hinttala, J. Uusimaa, K. Majamaa, T. Nyström, M. Kervinen, and I. E. Hassinen. LHON/MELAS overlap mutation in ND1 subunit of mitochondrial complex I affects ubiquinone binding as revealed by modeling in *Escherichia coli* NDH-1. BBA - Bioenergetics **1817**, 312–318 (2012).

[68] P. K. Sinha, J. Torres-bacete, E. Nakamaru-ogiso, N. Castro-guerrero, A. Matsuno-yagi, and T. Yagi. Critical Roles of Subunit NuoH (ND1) in the Assembly of Peripheral Subunits with the Membrane Domain of. J. Biol. Chem. **284**, 9814–9823 (2009).

[69] V. Zickermann, B. Barquera, and M. Finel. Analysis of the Pathogenic Human Mitochondrial Mutation ND1/3460, and Mutations of Strictly Conserved Residues in Its Vicinity, Using the Bacterium Paracoccus denitrificans. Biochemistry **37**, 11792–11796 (1998).

[70] A. Majander, K. Huoponen, M.-L. Savontaus, E. Nikoskelainen, and M. Wikström. Electron transfer properties of NADH:ubiquinone reductase in the ND 1/3460 and the ND4/11778 mutations of the Leber hereditary optic neuroretinopathy (LHON). FEBS Lett. **292**, 289–292 (1991).

[71] C. C. Page, C. C. Moser, X. Chen, and P. L. Dutton. Natural engineering principles of electron tunnelling in biological oxidation–reduction. Nature **402**, 47–52 (1999).

[72] V. Sharma, G. Belevich, A. P. Gamiz-Hernandez, T. Róg, I. Vattulainen, M. L. Verkhovskaya, M. Wikström, G. Hummer, and V. R. I. Kaila. Redox-induced activation of the proton pump in the respiratory complex I. Proc. Natl. Acad. Sci. U.S.A. **112**, 11571–11576 (2015).

[73] A. P. Gamiz-Hernandez, A. Jussupow, M. P. Johansson, and V. R. I. Kaila. Terminal Electron − Proton Transfer Dynamics in the Quinone Reduction of Respiratory Complex I. JACS **20**, 16282–16288 (2017).

[74] L. Euro, G. Belevich, M. I. Verkhovsky, M. Wikström, and M. Verkhovskaya. Conserved lysine residues of the membrane subunit NuoM are involved in energy conversion by the proton-pumping NADH:ubiquinone oxidoreductase (Complex I). BBA **1777**, 1166–1172 (2008).

[75] U. Brandt. A two-state stabilization-change mechanism for proton-pumping complex I. BBA - Bioenergetics **1807**, 1364–1369 (2011).

[76] M. Verkhovskaya and D. A. Bloch. Energy-converting respiratory Complex I: On the way to the molecular mechanism of the proton pump. Int. J. Biochem. Cell Biol. **45**, 491–511 (2013).

[77] V. R. I. Kaila. Long-range proton-coupled electron transfer in biological energy conversion: towards mechanistic understanding of respiratory complex I. J. Royal Soc. Interface **15**, 1–16 (2018).

[78] M. Wikström and G. Hummer. Stoichiometry of proton translocation by respiratory complex I and its mechanistic implications. Proc. Natl. Acad. Sci. U.S.A. **109**, 4431–4436 (2012).

[79] G. Hummer and M. Wikström. Molecular simulation and modeling of complex I. BBA **1857**, 915–921 (2016).

[80] M. L. Verkhovskaya, N. Belevich, L. Euro, M. Wikström, and M. I. Verkhovsky. Real-time electron transfer in respiratory complex I. Proc. Natl. Acad. Sci. U.S.A. **105**, 3763–3767 (2008).

[81] M. Verkhovskaya and M. Wikström. Oxidoreduction properties of bound ubiquinone in Complex I from *Escherichia coli*. BBA - Bioenergetics **1837**, 246–250 (2014).

[82] D. DeVault. Energy transduction in electron transport. BBA **225**, 193–199 (1971).

[83] S. T. Ohnishi, J. C. Salerno, and T. Ohnishi. Possible Roles of Two Quinone Molecules in Direct and Indirect Proton Pumps of Bovine Heart NADH-quinone Oxidoreductase (Complex I). BBA **1797**, 1891–1893 (2011).

[84] F. Schuler and J. E. Casida. Functional coupling of PSST and ND1 subunits in NADH:ubiquinone oxidoreductase established by photoaffnity labeling. BBA **1506**, 79–87 (2001).

[85] M. Kervinen, J. Pa, M. Finel, and I. E. Hassinen. A Pair of Membrane-Embedded Acidic Residues in the NuoK Subunit of *Escherichia coli* NDH-1, a Counterpart of the ND4L Subunit of the Mitochondrial Complex I, Are Required for High Ubiquinone Reductase Activity. Biochemistry **43**, 773–781 (2004).

[86] M. Kao, S. D. Bernardo, E. Nakamaru-ogiso, H. Miyoshi, A. Matsuno-yagi, and T. Yagi. Characterization of the Membrane Domain Subunit NuoJ (ND6) of the NADH-Quinone Oxidoreductase from *Escherichia coli* by Chromosomal DNA Manipulation. Biochemistry **44**, 3562–3571 (2005).

[87] M. Kao, E. Nakamaru-ogiso, A. Matsuno-yagi, and T. Yagi. Characterization of the Membrane Domain Subunit NuoK (ND4L) of the NADH-Quinone Oxidoreductase from *Escherichia coli*. Biochemistry **44**, 9545–9554 (2005).

[88] S. Nakanishi, M. Abe, S. Yamamoto, M. Murai, and H. Miyoshi. Bis-THF motif of acetogenin binds to the third matrix-side loop of ND1 subunit in mitochondrial NADH-ubiquinone oxidoreductase. BBA **1807**, 1170–1176 (2011).

[89] M. Murai, Y. Mashimo, J. Hirst, and H. Miyoshi. Exploring Interactions between the 49 kDa and ND1 Subunits in Mitochondrial NADH-Ubiquinone Oxidoreductase (Complex I) by Photoaffinity Labeling. Biochemistry **50**, 6901–6908 (2011).

[90] T. Ohnishi. Iron–sulfur clusters/semiquinones in Complex I. BBA - Bioenergetics **1364**, 186–206 (1998).

[91] M. Narayanan, S. A. Leung, Y. Inaba, M. M. Elguindy, and E. Nakamaru-Ogiso. Semiquinone intermediates are involved in the energy coupling mechanism of *E. coli* complex I. BBA **1847**, 681–689 (2015).

[92] L. H. Thomas. The calculation of atomic fields. Math. Proc. Camb. Philos. Soc. **23**, 542–548 (1927).

[93] E. Fermi. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente. Z. Phys. **48**, 73–79 (1928).

[94] P. A. M. Dirac. Note on exchange phenomena in the thomas atom. Math. Proc. Camb. Philos. Soc. **26**, 376–385 (1930).

[95] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. Phys. Rev. **136**, B864–B871 (1964).

[96] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. Phys. Rev. **140**, 1133–1138 (1965).

[97] D. M. Ceperley. Ground State of the Electron Gas by a Strochastic Method. Phys. Rev. Lett. **45**, 566–569 (1980).

[98] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, and G. E. Scuseria. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. J. Chem. Phys. **123**, 1–9 (2005).

[99] A. D. Becke. Density-functional exchange-energy approximation with correct asymtotic behavior. Phys. Rev. A **38**, 3098–3100 (1988).

[100] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B **37**, 785–789 (1988).

[101] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. Phys. Rev. Lett. **91**, 1–4 (2003).

[102] P. J. Stephens, C. Chabalowski, and M. J. Frisch. Initio Calculation of Vibrational Absorption and Circular Dichronism Spectra Using Density Functional Force Fields. J. Phys. Chem. **98**, 11623–11627 (1994).

[103] B. J. Alder and T. E. Wainwright. Phase Transition for a Hard Sphere System. J. Chem. Phys. **27**, 1208–1209 (1957).

[104] A. Rahman. Correlations in te Motion of Atoms in Liquid Argon. Phys. Rev. **136**, A405–A411 (1964).

[105] J. A. McCammon, B. R. Gelin, and K. Martin. Dynamics of folded proteins. Nature **267**, 585–590 (1977).

[106] D. M. Zuckerman. *Statistical Physics of Biomolecules: An Introduction.* CRC Press/Taylor and Francis, (2010).

[107] D. Frankel and B. Smith. *Understanding Molecular Simulation: From Algorithms to Applications.* Academic Press, (2001).

[108] L. G. Trabuco, E. Villa, M. Kakoli, J. Frank, and K. Schulten. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure **5**, 673–683 (2008).

[109] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J. Mol. Biol. **103**, 227–249 (1976).

[110] H. M. Senn and W. Thiel. QM/MM Methods for Biomolecular Sytems. Angew. Chem. Int. Ed. **48**, 1198–1229 (2009).

[111] L. Cao and U. Ryde. On the Difference Between Additive and Subtractive QM/MM Calculations. Front. Chem. **6**, 1–15 (2018).

[112] M. J. Field, P. A. Bash, and M. Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. J. Comput. Chem. **11**, 700–733 (1990).

[113] J. Sauer and M. Sierka. Combining Quantum Mechanics and Interatomic Potential Functions in *Ab Initio* Studies of Extended Systems. J. Comput. Chem. **21**, 1470–1493 (2000).

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

[115] L. Landau. *Statistical physics.* The Clarendon Press: Oxford, (1938).

[116] R. W. Zwanzig. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. J. Chem. Phys. **22**, 1420–1426 (1954).

[117] C. H. Bennett. Efficient estimation of free energy differences from Monte Carlo data. J. Comput. Phys. **22**, 245–268 (1976).

[118] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. Phys. Rev. Lett. **91**, 1–4 (2003).

[119] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys. Rev. E **60**, 2721–2726 (1999).

[120] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. J. Chem. Phys. **129**, 1–10 (2008).

[121] E. Rosta, M. Nowotny, W. Yang, and G. Hummer. Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations. J. Am. Chem. Soc. **133**, 8934–8941 (2011).

[122] G. Hummer. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. New J. Phys. **7**, 1–14 (2005).

[123] R. Zwanzig. From Classical Dynamics to Continuous Time Random Walks. J. Stat. Phys. **30**, 255–262 (1983).

[124] D. J. Bicout and A. Szabo. Electron transfer reaction dynamics in non-Debye solvents. J. Chem. Phys. **109**, 2325–2338 (1998).

[125] O. Călinescu, C. Paulino, W. Kühlbrandt, and K. Fendler. Keeping it simple, transport mechanism and pH regulation in Na +/H+ exchangers. J. Biol. Chem. **289**, 13168–13176 (2014).

[126] A. Di Luca, A. P. Gamiz-Hernandez, and V. R. I. Kaila. Symmetry-related proton transfer pathways in respiratory complex I. Proc. Natl. Acad. Sci. U.S.A. **114**, E6214–E6321 (2017).

[127] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. J. Chem. Theory Comput. **7**, 525–537 (2011).

[128] C. R. Søndergaard, M. H. Olsson, M. Rostkowski, and J. H. Jensen. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values. J. Chem. Theory Comput. **7**, 2284–2295 (2011).

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

[130] B. R. Brooks, C. L. B. III, J. A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. I. Hodoscek, and M. Karplus. CHARMM: The Biomolecular Simulation Program B. J. Comput. Chem. **30**, 1545–1614 (2009).

[131] D. Beglov and B. Roux. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. J. Chem. Phys. **100**, 9050–9063 (1994).

[132] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. **98**, 10089–10092 (1993).

[133] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. J. Mol. Graph. **14**, 33–38 (1996).

[134] C. Notredame, D. G. Higgins, and J. Heringa. T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. J. Mol. Biol. **302**, 205–2017 (2000).

[135] E. W. Floden, P. D. Tommaso, M. Chatzou, C. Magis, C. Notredame, and J.-M. Chang. PSI/TM-Coffee: a web server for fast and accurate multiple sequence alignments of regular and transmembrane proteins using homology extension on reduced databases. Nucleic Acids Res. **44**, 339–343 (2016).

[136] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. Barton. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics Applications Note **25**, 1189–1191 (2009).

[137] F. Weigend and R. Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. Phys. Chem. Chem. Phys. **7**, 3297–3305 (2005).

[138] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. **98**, 5648–5652 (1993).

[139] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B **37**, 785–789 (1988).

[140] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J. Phys. Chem. B **102**, 3586–3616 (1998).

[141] S. Riahi and C. N. Rowley. The CHARMM-TURBOMOLE interface for efficient and accurate QM/MM molecular dynamics, free energies, and excited state properties. J. Comput. Chem. **35**, 2076–2086 (2014).

[142] TURBOMOLE V6.2 2010, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; `http://www.turbomole.com`.

[143] S. Y. Noskov and B. Roux. Control of Ion Selectivity in LeuT: Two Na^+ Binding Sites with Two Different Mechanisms. J. Mol. Biol. **377**, 804–818 (2008).

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

[145] G. Hummer, L. R. Pratt, and A. E. Garcia. Ion sizes and finite-size corrections for ionic-solvation free energies. J. Chem. Phys. **107**, 9275–9277 (1997).

[146] P. H. Hünenberger and J. A. Mccammon. Ewald artifacts in computer simulations of ionic solvation and ion-ion interaction: A continuum electrostatics study. J. Chem. Phys. **110**, 1856–1872 (1999).

[147] A. Šali and T. L. Blundell. Comparative Protein Modelling by Satisfaction of Spatial Restraints. J. Mol. Biol **234**, 779–815 (1993).

[148] A. Schäfer, A. Klamt, D. Sattel, J. C. W. Lohrenz, and F. Eckert. COSMO Implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems. Phys. Chem. Chem. Phys. **2**, 2187–2193 (2000).

[149] E. Roberts, J. Eargle, D. Wright, and Z. Luthey-Schulten. MultiSeq: unifying sequence and structure data for evolutionary analysis. BMC Bioinformatics **7**, 1–11 (2006).

[150] B. E. Conway. The evaluation and use of properties of individual ions in slution. J. Solution Chem. **7**, 721–770 (1978).

[151] M. Gierth and P. Mäser. Potassium transporters in plants - Involvement in K^+ acquisition, redistribution and homeostasis. FEBS Lett. **581**, 2348–2356 (2007).

[152] C. T. Resch, J. L. Winogrodzki, C. T. Patterson, E. J. Lind, M. J. Quinn, P. Dibrov, and C. C. Häse. Putative Na^+/H^+ antiporter of Vibrio cholerae, Vc-NhaP2, mediates the specific K^+/H^+ exchange in vivo. Biochemistry **49**, 2520–2528 (2010).

[153] Y. Xu, Y. Zhou, S. Hong, Z. Xia, D. Cui, J. Guo, H. Xu, and X. Jiang. Functional characterization of a wheat NHX antiporter gene TaNHX2 that encodes a K^+/H^+ exchanger. PLoS ONE **8** (2013).

[154] M. V. Radchenko, K. Tanaka, R. Waditee, S. Oshimi, Y. Matsuzaki, M. Fukuhara, H. Kobayashi, T. Takabe, and T. Nakamura. Potassium/proton antiport system of Escherichia coli. J. Biol. Chem. **281**, 19822–19829 (2006).

[155] G. Masrati, M. Dwivedi, A. Rimon, Y. Gluck-Margolin, A. Kessel, H. Ashkenazy, I. Mayrose, E. Padan, and N. Ben-Tal. Broad phylogenetic analysis of cation/proton antiporters reveals transport determinants. Nat. Commun. **9**, 1–14 (2018).

[156] W. J. Jones, J. A. Leigh, F. Mayer, C. R. Woese, and R. S. Wolfe. Methanococcus jannaschii sp. nov., and extremely thermophilic methanogen from a submarine hydrothermal vent. Arch Microbiol **136**, 254–261 (1983).

[157] R. K. Thauer, A. K. Kaster, H. Seedorf, W. Buckel, and R. Hedderich. Methanogenic archaea: Ecologically relevant differences in energy conservation. Nat. Rev. Microbiol. **6**, 579–591 (2008).

[158] D. G. McMillan, S. A. Ferguson, D. Dey, K. Schröder, H. L. Aung, V. Carbone, G. T. Attwood, R. S. Ronimus, T. Meier, P. H. Janssen, and G. M. Cook. A_1A_O-ATP synthase of Methanobrevibacter ruminantium couples sodium ions for ATP synthesis under physiological conditions. J. Biol. Chem. **286**, 39882–39892 (2011).

[159] F. Mayer and V. Müller. Adaptations of anaerobic archaea to life under extreme energy limitation. FEMS Microbiol. Rev. **38**, 449–472 (2014).

[160] K. Y. Pisa, H. Huber, M. Thomm, and V. Müller. A sodium ion-dependent A_1A_O ATP synthase from the hyperthermophilic archaeon Pyrococcus furiosus. FEBS J. **274**, 3928–3938 (2007).

[161] A. Filippov, G. Orädd, and G. Lndblom. Influence of Cholesterol and Water Content on Phospholipid Lateral Diffusion in Bilayers. Langmuir **19**, 6397–6400 (2003).

[162] O. Haapanen and V. Sharma. A modeling and simulation perspective on the mechanism and function of respiratory complex I. BBA - Bioenergetics **1859**, 510–523 (2018).

[163] H. J. Coyne III, S. Ciofi-baffoni, L. Banci, I. Bertini, L. Zhang, G. N. George, and D. R. Winge. The Characterization and Role of Zinc Binding in Yeast Cox4. J. Biol. Chem. **282**, 8926–8934 (2007).

[164] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. J. Comput. Chem. **25**, 1605–1612 (2004).

[165] J. L. Markley, J. H. Kim, Z. Dai, J. Bothe, K. Cai, R. O. Frederick, and M. Tonelli. Metamorphic protein IscU alternates conformations in the course of its role as the scaffold protein for iron-sulfur cluster biosynthesis and delivery. FEBS Lett. **587**, 1172–1179 (2013).

[166] J. A. Zuris, Y. Harir, A. R. Conlan, M. Shvartsman, D. Michaeli, S. Tamir, M. L. Paddock, J. N. Onuchi, R. Mittler, Z. I. Cabantchik, P. A. Jennings, and R. Nechushtai. Facile transfer of [2Fe-2S] clusters from the diabetes drug target mitoNEET to an apo-acceptor protein. Proc. Natl. Acad. Sci. U.S.A. **108**, 13047–13052 (2011).

[167] G. Tan, A. P. Landry, R. Dai, L. Wang, J. Lu, and H. Ding. Competition of zinc ion for the [2Fe-2S] cluster binding site in the diabetes drug target protein mitoNEET. BioMetals **25**, 1–12 (2012).

[168] T. A. Ramelot, J. R. Cort, S. Goldsmith-fischman, G. J. Kornhaber, R. Xiao, R. Shastry, T. B. Acton, B. Honig, G. T. Montelione, and M. A. Kennedy. Solution NMR Structure of the Iron–Sulfur Cluster Assembly Protein U (IscU) with Zinc Bound at the Active Site. J. Mol. Biol. **344**, 567–583 (2004).

Acknowledgement

This work was performed at the Max Planck Institute of Biophysics in the Department of Theoretical Biophysics, Frankfurt am Main, and at the Department of Theoretical Chemistry (Lehrstuhl für Biokatalyse), Technical University of Munich (2013-2019). This work was financially supported by the Max Planck Society and the CRC807 "Transport and Communication across Biological Membranes". The computer resources from the Max Planck Society and Department of Chemistry at the Technical University Munich were used.

First, I want to thank my PhD adviser, Prof. Dr. Gerhard Hummer for his scientific guidance especially in free energy computations. I would also like to thank him for providing me the opportunity to pursue my post graduate studies in a shared supervision with Prof. Dr. Ville R. I. Kaila. He also provided me with all resources I required for my work. I enjoyed absolute freedom to plan the work related trips to TUM, workshops, and other scientific related activities during my PhD studentship. I thank him for his trust and support.

Next, I want to thank my second supervisor Prof. Dr. Ville R. I. Kaila for his supervision, especially in quantum chemical methods and insightful comments on this PhD thesis. I am also grateful to Dr. Ana P. Gamiz-Hernandez for introducing me to QM/MM approaches and her technical support and for sharing her knowledge on complex I. I also thank, the entire Kaila group for their hospitality. I felt always welcome and I had a wonderful time in Munich.

I would also like to thank all present and past members of the Department Theoretical Biophysics group for a supportive and encouraging atmosphere filled with enlightening discussions. Especially, I want to acknowledge Prof. Dr. Kei-ichi Okazaki for a tight collaboration in the early-stages of my studies on antiporter. I specially acknowledge Dr. Ahmadreza Mehdipour for his reliable support and great help, whenever I needed. I want to also thank Drs. Jakob Bullerjahn, Amir Bahrami, Bonnie Murphy, David Wöhlert, Ramachandra Bhaskara, Roberto Covino, Lukas Stelzl, and Hendrik Jung, and Laura Schulz for helpful suggestions and carefully reviewing this thesis. Many thanks for that.

I want to acknowledge Dr. David Wöhlert for performing the biochemical experiments in our collaborative work on MjNhaP1. Additionally, I want to acknowledge all co-authors of my publications for their contribution. Especially, Prof. Dr. Werner Kühlbrandt, Dr. Özkan Yildiz, Dr. Vivek Sharma, Prof. Dr. Volker Zickermann, Dr. Katarzyna Kmita. I wish to acknowledge the special PhD programs the CRC 807 integrated research training group, TRAM (TRansport Across Membranes), and the international Max Planck research school (IMPReS) for structure and function of biological membrane. I benefited from their additional talks, lectures, workshops, and retreats. I also acknowledge the ProCareer.Doc program from Mentoring Hessen.

Finally, a big thanks to my family for their wonderful support; my parents and my in-laws. I would like to thank my beloved husband, Thomas, for being a dedicated and

loving father, and taking care of family commitments. I greatly appreciate your support in all my endeavors and providing me with the freedom to pursue my PhD. Without you, it would not have been possible.