

Fakultät für Chemie der Technische Universität München

Multi-scale modeling and computational design of biological proton/ion-transport modules

Max Emanuel Mühlbauer

Vollständiger Abdruck der von der

Fakultät für Chemie

der Technischen Universität München zur Erlangung des akademischen Grades

eines Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitz: Prof. Dr. Michael Sattler

Prüfende/-r der Dissertation:

1. Prof. Dr. Ville R. I. Kaila

2. Prof. Dr. Michael Groll

Die Dissertation wurde am 14.03.2022 bei der Technischen Universität München eingereicht und durch die

Fakultät für Chemie am 04.07.2022 angenommen.

dedicated to my family

Contents

Abstract	iii
Zusammenfassung	iv
List of publications	v
Author contributions.....	vi
Abbreviations	vii
1 Introduction.....	1
1.1 The respiratory chain.....	1
1.2 Complex I as the initial proton pump in the respiratory chain	4
1.3 The role of membrane bound hydrogenase in generating a sodium motive force as an alternative in archaeal respiratory chains	6
1.4 Computational biochemistry	11
2 Aims of the study	13
3 Theory: Computational multi-scale modeling	14
4 Methods	26
4.1 General methods	26
4.2 Complex I from <i>Thermus thermophilus</i>	28
4.3 Complex I from Escherichia Coli	30
4.4 Mitochondrial complex I from <i>Mus musculus</i>	30
4.5 Membrane bound hydrogenase from <i>Pyrococcus furiosus</i>	32
4.6 Artificial 4-helix-bundle constructs.....	33
5 Results and Discussion	36
5.1 Hydration dynamics in the membrane domain of bacterial complex I (Articles 1 & 2)	36
5.2 Hydration dynamics in the membrane domain of mammalian complex I (Article 3)	41
5.3 Functional dynamics of a membrane bound hydrogenase (Article 5)	44
5.4 Design of an artificial protein with an ion-pair (Article 4).....	48
6 Conclusions and outlook	53
Bibliography	54
Acknowledgements	61
Appendix.....	I

Abstract

Adenosine triphosphate (ATP) is the universal energy currency of all organisms. It is synthesized by the ATP-synthase, a molecular turbine driven by the proton motive force (*pmf*). This transmembrane potential is generated by the complexes I, III and IV of the respiratory chain which pump protons across mitochondrial or cytoplasmatic membranes to build up a chemiosmotic potential in most species. Many variations on the common theme of respiratory chains have evolved over time. For instance, in some archaea, the ATP-synthase is not powered by a *pmf* but a sodium motive force (*smf*), which is generated by e.g. a membrane bound hydrogenase (Mbh). Molecular structures for most complexes of the respiratory chain have been solved in recent years, due to advances in structural biology, but many open questions remain concerning the mechanisms of these intricate enzymes. The mechanism by which the energy from the redox chemistry they perform, e.g., quinone-reduction or hydrogen gas production from protons and electrons, is converted into an electrochemical gradient is of particular interest for understanding key principles in bioenergetics. In this study, computational simulations in the form of classical molecular dynamics simulations and quantum chemical calculations have been used to elucidate the conformational dynamics of these enzymes. Here, we focus on Complex I from *T. thermophilus* and *M. musculus* as well as an Mbh from *P. furiosus*, which function as initial electron acceptors in respiratory chains. We demonstrate conserved coupling mechanisms in these evolutionarily related proteins, describe the central role of hydration dynamics in the ion transfer processes they catalyze, and apply protein design procedures and synthetic biology to build small model systems containing key electrostatic switches found in the complex I superfamily.

Zusammenfassung

Adenosintriphosphat (ATP) ist als universelle Energiewährung des Stoffwechsels allen Lebewesen gemein. In aeroben Organismen wird ATP durch die ATP-Synthase einer molekularen Turbine, die durch einen Protonengradienten angetrieben wird. Dieses Transmembranpotential wird in den meisten Organismen durch die Komplexe I, III und IV der Atmungskette aufgebaut, die Protonen über mitochondriale oder zytoplasmatische Membranen pumpen, um ein chemiosmotisches Potential zu erzeugen. Vielerlei Varianten der Atmungskette haben sich über die Zeit durch Evolution herausgebildet. Zum Beispiel wird die ATP-Synthase in manchen Archaeen nicht durch Protonengradienten sondern durch Natriumgradienten angetrieben, die zum Beispiel durch die Membrangebundene Hydrogenase (Mbh) erzeugt werden. Obwohl in der jüngeren Vergangenheit die molekularen Strukturen aller Komplexe der Atmungskette durch Fortschritte in der Strukturbiologie gelöst wurden, bleiben viele offene Fragen betreffs der Mechanismen dieser komplexen Enzyme bestehen. Insbesondere die Frage nach der Art, wie die Umwandlung der Energie aus deren Redoxreaktionen, wie zum Beispiel der Quinonreduktion oder der Wasserstoffsynthese, in einen elektrochemischen Gradienten umgewandelt wird, ist von großer Bedeutung. In dieser Arbeit wurden computergestützte Simulationen in der Form von klassischen Molekulardynamiksimulationen und quantenchemischen Berechnungen verwendet um die konformationellen Dynamiken dieser Enzyme herauszustellen. Wir richten unser Hauptaugenmerk auf die initialen Elektronenakzeptoren der Atmungskette: Komplex I aus *T. thermophilus* und *M. musculus*, sowie Mbh aus *P. furiosus*. Wir zeigen konservierte Kopplungsmechanismen in diesen evolutionär verwandten Proteinen auf, beleuchten den Einfluss von Wasserdynamik auf die Ionentransferprozesse und verwenden Proteindesignmethoden und synthetische Biologie um artifizielle Modellsysteme mit wichtigen elektrostatischen Schalterelementen aus der Komplex-I-Superfamilie herzustellen.

List of publications

Andrea Di Luca*, **Max E. Mühlbauer***, Patricia Saura and Ville R. I. Kaila (2018) How inter-subunit contacts in the membrane domain of complex I affect proton transfer energetics. *BBA – Bioenergetics*. **1859**, 734–741; <https://doi.org/10.1016/j.bbabbio.2018.06.001>.

Max E. Mühlbauer*, Patricia Saura*, Franziska Nuber, Andrea Di Luca, Thorsten Friedrich and Ville R. I. Kaila (2020) Water-Gated Proton Transfer Dynamics in Respiratory Complex I. *J. Am. Chem. Soc.* **142**, 13718–13728; <https://doi.org/10.1021/jacs.0c02789>.

Mona Baumgart, Michael Röpke*, **Max E. Mühlbauer***, Sam Asami, Sophie L. Mader, Kai Fredriksson, Michael Groll, Ana P. Gamiz-Hernandez and and Ville R. I. Kaila (2021) Design of buried charged networks in artificial proteins. *Nat. Commun.* **12**, 1895; <https://doi.org/10.1038/s41467-021-21909-7>.

Michael Röpke, Daniel Riepl*, Patricia Saura*, Andrea Di Luca*, **Max E. Mühlbauer**, Alexander Jussupow, Ana P. Gamiz-Hernandez, Ville R. I. Kaila (2021) Deactivation blocks proton pathways in the mitochondrial complex I. *Proc. Natl. Acad. Sci. USA* **118**, 29; <https://doi.org/10.1073/pnas.2019498118>.

Max E. Mühlbauer, Ana P. Gamiz-Hernandez, Ville R. I. Kaila (2021) Functional dynamics of an ancient membrane-bound hydrogenase. *J. Am. Chem. Soc.* **143**, 20873–20883; <https://doi.org/10.1021/jacs.1c09356>.

Other contributions

Jagna Witek, **Max Mühlbauer**, Bettina G. Keller, Markus Blatter, Axel Meissner, Trixie Wagner and Sereina Riniker (2017) Interconversion Rates between Conformational States as Rationale for the Membrane Permeability of Cyclosporines. *ChemPhysChem* **18**, 3309–3314; <https://doi.org/10.1002/cphc.201700995>.

Hamed Kooshapur, Nila Roy Choudhury, Bernd Simon, **Max Mühlbauer**, Alexander Jussupow, Noemi Fernandez, Alisha N. Jones, Andre Dallmann, Frank Gabel, Carlo Camilloni, Gracjan Michlewski, Javier F. Caceres and Michael Sattler (2018) Structural basis for terminal loop recognition and stimulation of pri-miRNA-18a processing by hnRNP A1. *Nat. Commun.* **9**, 2479; <https://doi.org/10.1038/s41467-018-04871-9>.

Kai Xue, **Max Mühlbauer**, Salvatore Mamone, Riddhiman Sarkar, and Bernd Reif (2019) Accurate Determination of ¹H-¹⁵N Dipolar Couplings Using Inaccurate Settings of the Magic Angle in Solid-State NMR Spectroscopy. *Angew. Chemie Int. Ed.* **58**, 4286–4290; <https://doi.org/10.1002/anie.201814314>.

* contributed equally

Author contributions

BBA Bioenergetics 2018

M.E.M., A.D.L., and V.R.I.K. designed research. **M.E.M.**, A.D.L., P.S., and V.R.I.K. performed research. **M.E.M.**, A.D.L., and P.S. analyzed data. **M.E.M.**, A.D.L., P.S., and V.R.I.K. wrote the paper.

JACS 2020

M.E.M. performed classical MD simulations, P.S. performed quantum chemical calculations and hybrid QM/MM simulations, **M.E.M.** and P.S. developed analytic computational tools and analyzed results. A.D.L. developed and analyzed the kinetic model. F.N. and T.F. expressed, purified, and characterized the complex I variants and performed biophysical characterization. V.R.I.K. designed and directed the project and wrote the manuscript with input from all authors.

Nature Communications 2021

M.B. expressed, purified, and characterized artificial proteins and performed computational analysis. M.B., **M.E.M.**, M.R., S.L.M., A.P.G.H., and V.R.I.K. performed molecular simulations and designed artificial proteins. **M.E.M.**, M.R., S.L.M., and V.R.I.K. developed tools for computational analysis. M.B. and S.A. conducted NMR experiments and refined NMR structure ensembles. M.B., M.G., and V.R.I.K. resolved the x-ray crystal structures. A.P.G.H. built the initial computational model. M.B., **M.E.M.**, M.R., S.A., S.L.M., K.F., M.G., A.P.G.H., and V.R.I.K. analyzed results. V.R.I.K. designed and directed the project, conducted computational design and biophysical characterization, and wrote the manuscript with input from all authors.

PNAS 2021

M.R., A.D.L. and **M.E.M.** built initial models of complex I. M.R., D.R., P.S., A.D.L., **M.E.M.**, A.J. and A.P.G. conducted simulations and performed computational analysis. V.R.I.K. directed the project and wrote the manuscript with input from all authors.

JACS 2021

M.E.M. performed classical molecular simulations, developed analytic computational tools and analyzed results. A.P.G. performed computational calculations and analysis. A.P.G. and V.R.I.K. performed quantum chemical calculations. V.R.I.K. directed the project and wrote the manuscript with input from all authors.

Abbreviations

ACMA	9-amino-6-chloro-2-methoxyacridine
ACS	acetyl-CoA synthase
ADP	adenosine diphosphate
ALS	antiporter like subunit(s)
AMP	adenosine monophosphate
APBS	adaptive Poisson-Boltzmann solver
ATP	adenosine triphosphate
BBK	Brünger-Brooks-Karplus method
BIRCH	balanced-iterative reducing and clustering using hierarchies
BMRB	biological magnetic resonance data bank
CD	circular dichroism (spectroscopy)
CGMD	coarse grained molecular dynamics
CI	respiratory complex I
CoA	coenzyme A
COSMO	conductor like screening model
CPU	central processing unit
Cyt c	cytochrome c
DHAMed	dynamic histogram analysis method
DFT	density functional theory
EM	electron microscopy
EPR	electron paramagnetic resonance (spectroscopy)
eT	electron transfer
FAD	flavine adenine dinucleotide (oxidized)
FADH ₂	flavine adenine dinucleotide (reduced)
Fd	ferredoxin
FMN	flavine mononucleotide
GAPOR	glyceraldehyde-3-phosphate ferredoxin oxidoreductase
Gdn	guanidinium
GPU	graphical processing unit
ISC	iron sulfur center
HISQC	heteronuclear in-phase single quantum coherence spectroscopy
HL	horizontal long (helix)

LCAO	linear combination of atomic orbitals
Mbh	membrane bound hydrogenase
Mbs	membrane bound sulfur reductase
MC	Monte-Carlo
MD	molecular dynamics
MDFF	molecular dynamics flexible fitting
MM	molecular mechanics
MO	molecular (spin) orbital
Mrp	multiple-resistance and pH adaptation
NADH	nicotinamide adenine dinucleotide (reduced)
NAD ⁺	nicotinamide adenine dinucleotide (oxidized)
NDH	NADH dehydrogenase, also known as photosynthetic complex I
NMR	nuclear magnetic resonance (spectroscopy)
N-side	negatively charged side (of a membrane)
<i>pmf</i>	proton motive force ¹
pmf	potential of mean force
PBC	periodic boundary conditions
PBE	Poisson-Boltzmann equation
PCET	proton-coupled electron transfer
PDB	protein database
P _i	inorganic phosphate
PME	Particle-Mesh-Ewald (method)
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POR	pyruvate ferredoxin oxidoreductase
P-side	positively charged side (of a membrane)
pT	proton transfer
PYPI	1-palmitoyl-2-palmitoeyl-sn-glycero-3-phosphoinositol
Q	quinone
QH ₂	quinol
QH ⁻	semiquinone
QM	quantum mechanics, quantum mechanical
QM/MM	quantum mechanics/molecular mechanics

¹ italics are used throughout the text to differentiate proton motive force (*pmf*) and potential of mean force (pmf)

<i>rc</i>	reaction coordinate
REUS	replica exchange umbrella sampling
RMSD	root mean square deviation
RMSF	root mean square fluctuation
ROS	reactive oxygen species
SHE	standard hydrogen electrode
SMD	steered molecular dynamics
<i>smf</i>	sodium motive force
TM	transmembrane (helix)
UQ	ubiquinone
WHAM	weighted histogram analysis method
WT	wild type

I would rather have questions that can't be answered than answers that can't be questioned.
Richard P. Feynman

1 Introduction

Modern bioenergetics arguably began with Peter Mitchell's seminal paper on the chemiosmotic theory of energy transduction in 1961.⁽¹⁾ Therein, Mitchell suggested that oxidative phosphorylation, i.e., the synthesis of adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and inorganic phosphate (P_i), is driven by the transfer of elementary particles across biological membranes.^(1, 2) 50 years later, the key ideas of this hypothesis still hold and have spawned a wide field of bioenergetic research, which has undergone a recent reinvention with the advent of high-resolution spatial structures for all members of the (canonical) electron transfer chain⁽³⁾, driven in part by the so-called resolution-revolution in cryo-EM microscopy.⁽⁴⁾ The availability of high-quality structures in combination with the wider availability of high performance computing resources has allowed for computational studies on the various respiratory complexes^(5–11).

In this section, the main features of the respiratory chain will be briefly summarized. Although this work is focused on the members and the evolution of the complex I superfamily, it is important to keep the biochemical framework in which these enzymes operate in mind. Further on, the architecture and mechanism of complex I will be discussed in more detail before alternative modes of action in archaeobacterial respiratory chains are highlighted. Lastly, the role of computational chemistry in current bioenergetics research will be discussed briefly.

1.1 The respiratory chain

Aerobic respiratory chains, or electron transport chains, catalyze the reaction between protons, electrons and oxygen to form water (Figure 1).⁽¹²⁾ This highly exothermic reaction, catalyzed by complexes I, III and IV, is coupled to the pumping of several protons from the negatively charged side (N-side) of a biological membrane (the inner mitochondrial membrane in Eukarya, the cytoplasmic membrane in Eubacteria and Archaea) to the positively charged side (P-side).⁽¹³⁾ The resulting electrochemical gradient, or proton motive force (pmf), is subsequently utilized by complex V, i.e. the ATP synthase, to synthesize the universal energy currency ATP from ADP and P_i .⁽¹⁴⁾ Alternatively, the pmf can also be used directly for active transport across biological membranes.⁽¹⁾ The nature of the pmf depends on the organism in question as well as on the ion- and proton-concentration on both sides of the membrane: In mitochondria, the pmf is mostly determined by its electrical component via the membrane potential $\Delta\Psi$, while in chloroplasts and some bacteria there may be a significant difference in proton concentration on both sides of the membrane, adding a stronger ΔpH component.⁽¹⁵⁾

The reduced forms of nicotinamide adenine dinucleotide (NADH) and succinate are the (by)product of many metabolic pathways, and electrons are usually funneled into the respiratory chain via these metabolites (via a flavin adenine dinucleotide ($FADH_2$) cofactor in the case of succinate).⁽¹⁶⁾ The redox pairs formed by their reduced and oxidized forms at pH 7 have low midpoint potentials (compared to the standard hydrogen

electrode (SHE) E_{m7}), of -320 mV for NADH/NAD⁺ and +31 mV for FADH₂/FAD⁺. Compared to the O₂/H₂O pair with an E_{m7} of +815 mV, this yields a redox potential of up to 1135 mV for the total reaction.(17)

Complex I and complex II provide two alternative entry points for electrons into the respiratory chain.(3) While both catalyze the reduction of quinone (Q) to quinol (QH₂), they differ in their initial redox pair: Complex I accepts electrons from NADH via a FMN cofactor, while complex II has a bound FAD, which is reduced to FADH₂ during the succinate to fumarate reaction performed by this enzyme, also known as succinate dehydrogenase, during the Krebs cycle.(18) The higher potential between the NADH/NAD⁺ redox couple and the QH₂/Q redox couple ($E_{m7} = +90$ mV in the membrane for ubiquinone)(13) compared to the potential between FAD/FADH₂ and QH₂/Q, allows complex I to not only form QH₂ but also to contribute to the *pmf* by pumping four protons across the membrane.(19, 20) Since complex II does not contribute directly to the *pmf*, it is often considered to be associated with the Krebs cycle rather than to the respiratory chain.(21)

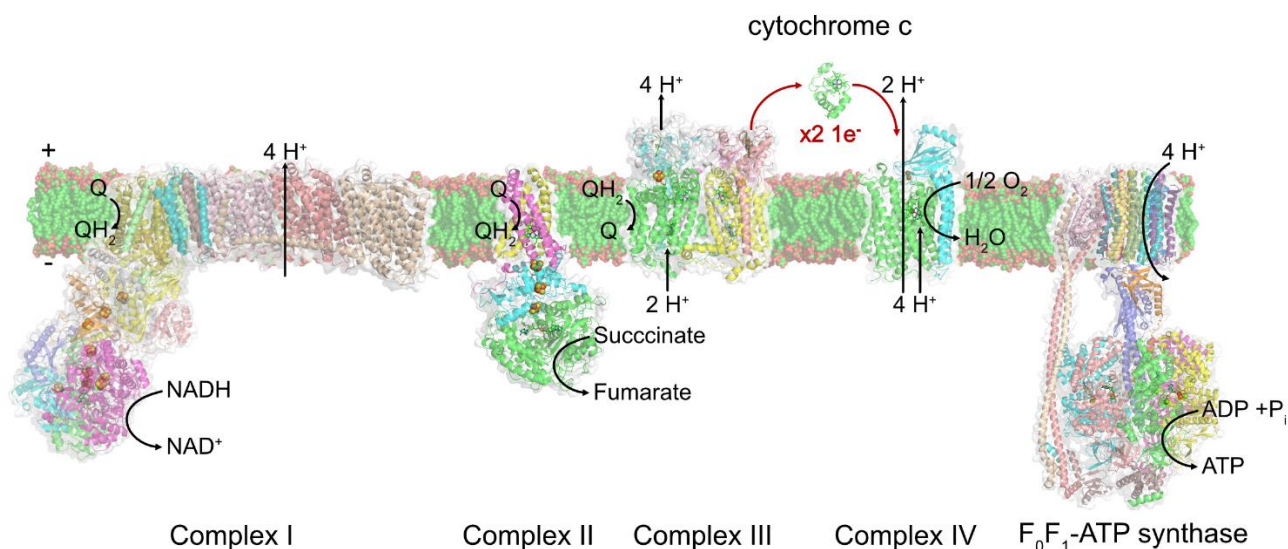


Figure 1: Example of a respiratory chain with bacterial complexes I to V and their interaction with the Q-pool as well as the *pmf*. PDB IDs of depicted protein structures: Complex I - 4HEA(22), complex II - 1NEN(23), complex III - 2YIU(24), cytochrome *c* - 2B4Z(25), complex IV aa₃ - 3HB3(26), complex V - 6OQR.(27) Note that complex II only contributes to the QH₂ pool but not to the *pmf*. Complexes I, III and IV directly contribute to the *pmf* by pumping protons, while complex V uses the energy provided by the *pmf* for ATP synthesis.(3)

While complexes I and II build up a pool of reduced quinol, complex III is responsible for transferring the electrons from QH₂ to cytochrome *c*, a mobile heme protein, while replenishing the Q-pool in the process and contributing to the *pmf*.(28) Notably, complex III is not a proton pump but generates *pmf* by a so-called Mitchellian redox loop mechanism(29): The charge transfer is based on electron transfer coupled to proton uptake and release processes in the four Q-sites of complex III. Two of these sites are located at the N-side and P-side each and the Q-cycle deployed leads to the release of four protons to the P-side. Two additional protons

are taken up from the N-side, while two QH₂ molecules are oxidized, and one Q is reduced to QH₂, using the protons from the N-side. This complex shuttling mechanism is catalyzed by a Rieske type 2Fe-2S iron-sulfur-center (ISC) in combination with one type *c* cytochrome and two type *b* cytochromes with different potentials.(30)

The cytochrome *c* reduced by complex III (with a E_{m7} of +250 mV)(17, 31) is utilized by complex IV to reduce oxygen to water (E_{m7} of +850 mV),(12) with four cytochrome *c* single electron carriers needed per reduction of one O₂ molecule. The stepwise increase in redox potential from NADH via QH₂ and reduced cytochrome *c* to water is important for preventing the dissipation of the ca. 23 kcal/mol of free energy.(3) Complex IV, or cytochrome *c* oxidase, contributes to the *pmf* by the reduction of dioxygen to 2 H₂O while consuming eight protons from the N-side. Four of these protons are pumped across the membrane while another four protons are transferred to the active site.(32) Two cycles of complex III are necessary to provide the reduced cytochrome *c* for the reduction of one molecule dioxygen.

The first four complexes of the respiratory chain are thus used to generate the *pmf* directly (I, III and IV) or indirectly (II). Their net reactions and proton pumping activity are summarized in Table 1.

Table 1: Net reactions of respiratory complexes I to IV and the number of protons pumped from the N-side to the P-side in addition to the net reaction. Note that the four protons in the complex IV reaction are removed from the N-side and also contribute to the *pmf*.

Complex	Net reaction	# of protons pumped
<i>I</i>	$NADH + H^+ + Q \rightarrow NAD^+ + QH_2$	4
<i>II</i>	$succinate + Q \rightarrow fumarate + QH_2$	0
<i>III</i>	$QH_2 + 2 \text{ cyt } c_{ox} \rightarrow Q + 2 \text{ cyt } c_{red}$	2
<i>IV</i>	$4 \text{ cyt } c_{red} + O_2 + 4H^+ \rightarrow 4 \text{ cyt } c_{ox} + 2 H_2O$	4

Complex V, the ATP synthase, finally uses this built-up *pmf* to synthesize ATP from ADP and P_i. The rotary mechanism of this enzyme has been studied extensively in the past,(14, 33–36) resulting in the Nobel prize in chemistry for Boyle, Walker and Skou in 1997(37), nineteen years after Peter Mitchell received his Nobel prize in chemistry for the chemiosmotic theory.(38)

1.2 Complex I as the initial proton pump in the respiratory chain

In most aerobic organisms, respiratory complex I is the main entry point for electrons to the respiratory chain.(39) This enormous L-shaped multi-domain protein shares a core of 14 subunits between most species with additional supernumerary subunits (of thus far debated function) common in higher organisms.(40) The core subunits can be grouped into two main domains (Figure 2): The hydrophilic domain, which drives the redox chemistry, and the membrane domain, which pumps protons.(39, 41) Molecular details of the intricate molecular machinery of this enzyme have started to emerge since its first (full) structure was solved by Baradaran et al. in 2013 (Complex I from *T. thermophilus*) closely followed by a second structure from *Y. lipolytica* in 2015(42), but its molecular mechanism has remained elusive. This has become a highly active field of research and new complex I structures are solved every year.(42–45)

Proton pumping in the membrane domain of complex I occurs in the three so-called antiporter like subunits (ALSs) Nqo12, Nqo13 and Nqo14 (*T. thermophilus* nomenclature) and possibly in a fourth pathway between Nqo14 and Nqo8, which is less established.(11, 41) Current hypotheses attribute one proton pumped to each of these sites and conserved S-shaped water pathways have been found in computational studies in each ALS.(11) The S-shape of these pathways reflects the internal symmetry of each ALS, which feature two symmetrical five-helix bundles with one broken helix element per bundle (typically transmembrane helices (TM) 7 and 12) each.(46) The ALSs are lined up sequentially in the membrane and clamped together by a horizontal long (HL-) helix of Nqo12, which has previously been implicated in so-called piston mechanisms considered largely disproven in current thinking, based on biophysical and site-directed mutagenesis studies.(47) Each ALS contains one conserved ion-pair facing its previous neighboring subunit in the structure. Together with the residues implicated in lateral proton transfer, these residues form a hydrophilic axis at the center of the membrane spanning the whole length of the membrane domain.(48) Considering the desolvation penalty associated with burying so many charged and polar residues in the hydrophobic protein core,(49, 50) they were implicated as key functional elements even before the first structure of complex I had been solved.(51) Prior computational studies showed that the ion-pairs can flip open, switching to an alternative ion-pair, formed with the terminal proton acceptor in the preceding subunit.(11) As such, they could be electrostatic switches coordinating the ALSs with each other. The first part of the results section in this work will reflect in more detail on their mechanistic role.

The hydrophilic domain catalyzes the electron transfer from NADH to Q via a bound flavin mononucleotide (FMN) and a number of ISCs (9 in *T. thermophilus*). (52) The number of ISCs varies with the species (for instance N7 in *T. thermophilus* is replaced by a zinc site in higher organisms)(45) and not all of them are likely on pathway.(53–55) Electrons are transferred down the quasi-linear chain of ISCs towards Q at a rate of ca. 90 μ s (derived from EPR measurements) with N2 being the final ISC acceptor.(56) It should be noted that while NADH donates two electrons and Q accepts two electrons, the ISCs themselves are one electron acceptors. The

chain of ISCs can roughly be divided into two parts: an upper part consisting of N3, N1b, N4 and N5 (with N7 and potentially N1a off pathway) and a lower part made up of N6a, N6b and N2. The two parts are separated by a gap of roughly 14 Å between N5 and N6a, close to the maximum distance allowing for rapid electron transfer according to Marcus theory.(52, 57, 58) It is no coincidence that the largest distance can be found between these two clusters, as will be discussed in section I.4.

It is currently debated, when and how Q obtains the two protons necessary for reduction to QH₂. Since Q/QH₂ is very flexible(59) in its cavity formed by Nqo4, Nqo6 and Nqo8, it has been challenging to resolve it experimentally.(39, 60) Studies on bound Q-analogs and inhibitors, particularly piericidin computational studies followed by biochemical validation have revealed two possible binding sites, one close to the N2 ISC with the head group bound to two conserved tyrosine and histidine residues of subunits Nqo4 and Nqo6 respectively (implicated in proton transfer to Q) and one closer towards the exit site of the cavity towards the membrane.(9, 60–62) This second binding site is close to a sequence of conserved acidic residues forming the so-called E-funnel, which leads towards the ALS.(39)

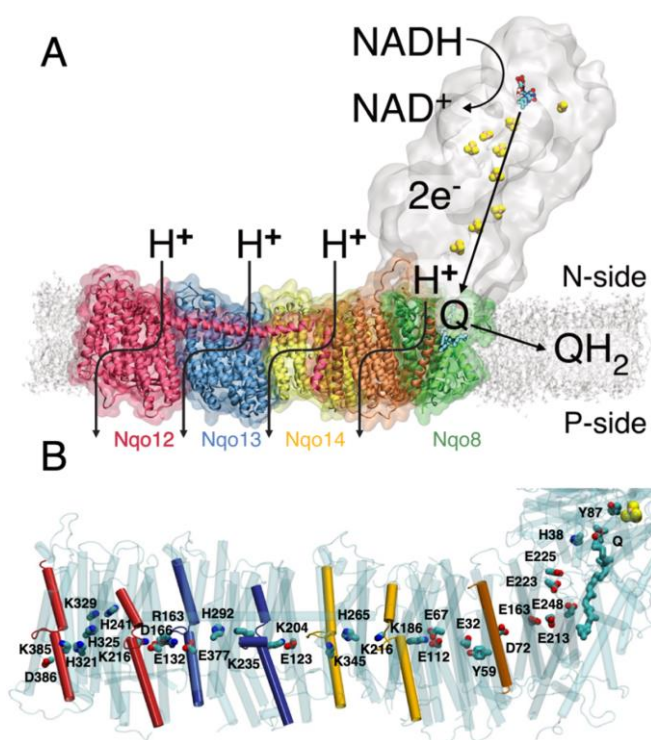


Figure 2: The architecture of respiratory complex I A) The hydrophilic domain (grey) catalyzes electron transfer from NADH to Q via FMN (sticks) and a chain of ISCs (spheres), the membrane domain with its antiporter-like-subunits (red, blue and orange) features S-shaped hydration pathways, which are used for proton transfer. A putative fourth pathway can be found between Nqo8 (green) and the ALS via subunits Nqo7/10/11. B) A hydrophilic axis (sticks) consisting of charged switches and proton transfer chains spans the membrane domain. Each antiporter-like-subunit features two rotary-symmetrical five-helix bundles with one broken helix element (cylinders) each. Q is shown in the first binding site. Figure adapted from article II.

The first binding site had already been described in biochemical studies(63) and is located close to three conserved loops: the TM1-TM2 loop in Nqo7, the β1-β2 loop in Nqo4 and the TM5-TM6 loop of Nqo8. These loops have attracted considerable interest since cryo-EM structures revealed different degrees of disorder associated with different levels of activity.(43, 45, 64, 65) Crosslinking experiments showed that locking the

TM1-TM2 loop of Nqo7, which is particularly long, in one conformation by a disulfide bridge decoupled oxidoreductase activity and proton pumping reversibly.(66) While this indicates functional relevance of the loop dynamics, possibly for coupling redox reactions and proton transfer reactions, their role is poorly understood its structure has been challenging to resolve due to the high flexibility.

The coupling between the oxidoreductase activity/ Q-chemistry and the proton pumping activity in the ALSs has been of major interest in complex I research in recent years.(67–69) Q-dynamics, changes in electrostatics during Q-reduction and even electrostatic effects by charged quinone species during turnover have all been discussed,(39) but a clear mechanistic model is still missing. Furthermore, the questions of how the energy from redox reactions is passed along the chain of ALSs and of how the proton transfer within the ALSs are coordinated is under debate, especially since Complex I operates under full microscopic reversibility.(70)

In higher organism, Complex I can switch into this reverse operation mode under substrate-deprived and high *pmf* conditions.(71) This leads to reverse electron transfer, a functional mode observed in ischemia, where it can lead to tissue damage and the formation reactive oxygen species (ROS) through uncontrolled release of electrons.(72) Mammalian isoforms of complex I can transition into a so-called 'deactive' form, which can be induced *in vitro*, e.g., by withholding substrate.(73) It takes a significant amount of time for complex I to revert to its active form and during this transition, neither forward nor reverse activity can be measured, suggesting that the deactive form blocks reverse electron transfer.(71, 74, 75) This deactivation is not observed in simpler organisms, which has sparked a discussion, whether it is an extreme form of an on-pathway conformational change during regular function or a separate, off-pathway transition.(43, 44, 66) Deactivation has been linked to increased rigidity of the conserved loops around the 1st Q binding site (see above) and an α -helix to π -bulge transition in TM3 of subunit Nqo6 (see part 2 of the results section for details). However, it is still unclear whether physiological deactivation is associated with a specific deactive structure.

1.3 The role of membrane bound hydrogenase in generating a sodium motive force as an alternative in archaeal respiratory chains

The membrane bound hydrogenase (Mbh) from *Pyrococcus furiosus* was found by chance at the turn of the millennium, when hydrogen production activity was observed in washed membrane preparations.(76) Closer inspection and operon analysis revealed a 14-subunit enzyme with significant homology to hydrogenases implicated in respiration from other organisms (Ech and Hyc, both proteins named after their operon)(77, 78), which sparked interest since *P. furiosus*, as a strictly anaerobic organism, had been previously considered a purely fermenting organism.

A follow-up study suggested a role for Mbh in the metabolism of the hyperthermophilic archaeon.(79) In the proposed pathway, primary metabolic reactions during glycolysis are used to reduce the single electron carrier ferredoxin (Fd), a small FeS protein with a low redox potential ($E_{m7} = -368$ mV for the *P. furiosus* variant).(80)

The oxidation of glyceraldehyde 3-phosphate to 3-phospho-glycerate catalyzed by glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) and the decarboxylation of pyruvate to acetyl-CoA catalyzed by pyruvate ferredoxin oxidoreductase (POR) are the two main electron-contributing reactions. Two Fds would then donate one electron each to Mbh, which in turn would use the energy released from hydrogen gas production ($2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2$) to generate a *pmf* by pumping protons across the membrane. The *pmf* is used, like in other respiratory chains, to drive ATP synthesis (Figure 3). Indeed, the same study found that Mbh activity does not only contribute to the $\Delta\Psi$, but also the ΔpH component of the membrane potential. This activity, monitored by changes in the fluorescence of acridine orange, could be induced by addition of Fd and selectively inhibited by copper ions (a known inhibitor of hydrogenases). It should be noted that the selective activation by addition of Fd prevents artifacts due to the orientation of Mbh in the vesicles (hydrophilic domain inside or outside). ATP synthase was present in the vesicles, could be driven by Mbh after Fd addition and selectively inhibited. These results suggested that Mbh functions as a redox-driven proton pump.

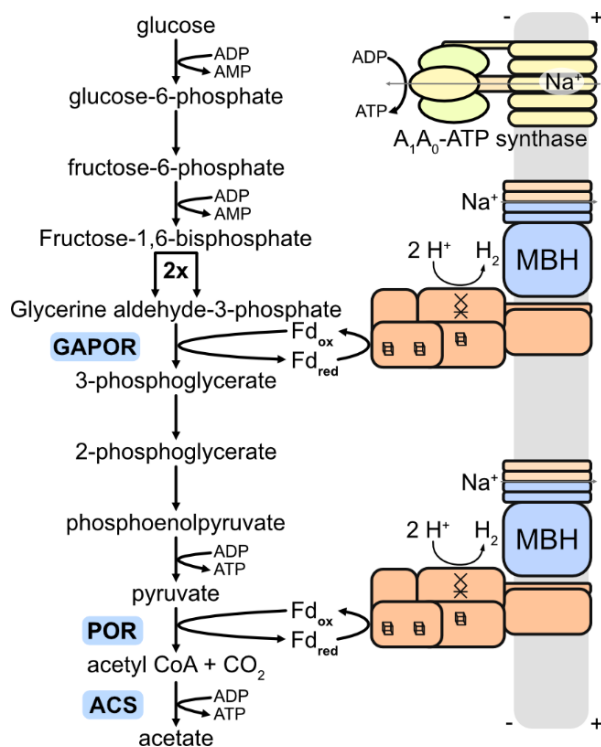


Figure 3: Metabolic role of Mbh proposed by Saprà et al. 2003(79) Four ferredoxin are reduced during one round of glycolysis before they are re-oxidized by Mbh. GAPOR: GAP ferredoxin oxidoreductase; POR: pyruvate ferredoxin oxidoreductase; ACS, acetyl-CoA synthase. Note that Mbh was originally suggested to be a proton transporter. The schematic function suggested in this figure has been updated to reflect current results, which show that the A_1A_0 -ATP synthase in *P. furiosus* is powered by sodium ions instead of protons(33) and Mbh is therefore a possible sodium transporter. Proton transport by Mbh has been omitted for simplicity.

This model was challenged in 2007, when it was found that the ATP synthase in *P. furiosus* is not proton but sodium driven, i.e., it utilizes a sodium motive force (*smf*). (33) The prior results show that Mbh was able to drive the A_1A_0 -ATPase in *in vitro* experiments and showed its contribution to ΔpH . These findings suggest that MBH must transport both sodium ions and protons in the same direction. This seems counterintuitive, especially considering the low thermodynamic driving force of hydrogen production: The Fd/ H_2 redox reaction yields only

a driving force of $\Delta E_{m7} = +86$ mV, about an order of magnitude smaller than the one provided by complex I.(17) At a membrane potential of -130 mV(81), this would allow for transduction of only one ion per two electrons, i.e. per H₂ formed.(17).

Mbh belongs to the class 4 NiFe-hydrogenases, and the sequence homology around the active site shows that it is closely related to the complex I superfamily rather than to soluble hydrogenases.(82) However, considering the unique nature of the active site metal cluster, it is expected that core parts of the reactive cycle are shared between NiFe-hydrogenases.

The active site of Mbh is a bimetallic center consisting of a nickel and an iron atom, bridged by sulfur atoms of two cysteine side chains. While the coordination sphere of the nickel is completed by two additional cysteine residues, the iron features three small molecules in its octahedral coordination: two cyanide ligands and one carbon monoxide ligand. As an oxygen-sensitive NiFe-hydrogenase, the active site can be reversibly inhibited by either binding of a CO to the nickel or binding of a OH⁻ between the nickel and iron atoms.(83–87)

NiFe-hydrogenases have been extensively studied by spectroscopic methods, especially electron paramagnetic resonance (EPR)(84), and computational methods, particularly density functional theory (DFT) based cluster models, as well as by hybrid quantum-chemicals/molecular mechanics (QM/MM) methods.(85, 86)

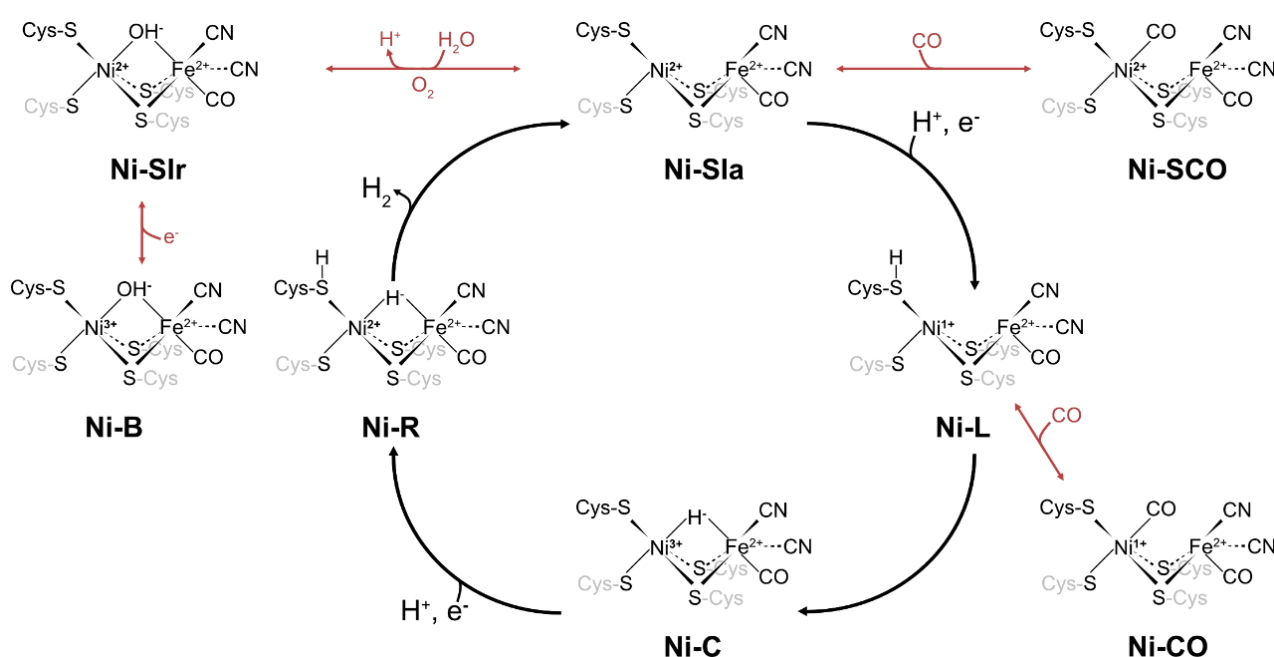


Figure 4: Reactivity of O₂ sensitive NiFe-hydrogenases as described in Ogata et al. 2016.(83) The experimentally observed states of the catalytic cycle are Ni-R, Ni-Sla and Ni-RC. The latter is the only EPR active state of the three. Inactivation reactions by O₂ and CO are indicated by red arrows.

The catalytic cycle features three experimentally detectable states, Ni-R, Ni-C and Ni-Sla, of which Ni-C is the only EPR active state with an oxidation state of Ni(III) Fe(II). It is also the resting state of these enzymes, while

the Ni-L state has not been observed experimentally due to its transient nature.(84–86) Starting from the Ni-S_{1a} state (Ni(II) Fe(II)), which has not yet received electrons or protons, the reaction cycle could proceed by (Figure 4):(83) 1) The Ni-L state (Ni(I) Fe(II)) is achieved by reducing Ni(II) by e⁻T from Fd via the chain of ISCs and PCET to the sulfur atom of the side chain of one of the terminal cysteine ligands (probably Cys374_L in Mbh). 2) The Ni-C state is reached by oxidizing Ni(I) to Ni(III), coupled to reduction of the bound proton to a hydride species, which binds between nickel and iron 3) a second e⁻T yields Ni-R, a Ni(II) Fe(II) state, still with the bound hydride and a second proton is added (possibly via the same cysteine side chain as the first). 3) Combining the proton with the hydride and releasing the H₂ product completes the cycle. The donors for the protons to the cysteine side chain are still debated.(87–90) One possible proton donor is a conserved glutamate on the β1β2-loop of MbhL (see below). A chain of histidine residues within MbhL toward the ‘top’ of the enzyme could also take part in pT. The involvement of the β1β2-loop is supported by the the importance of this site in complex I (see above) and further supports the close relation between the two enzymes.

The homology between Mbh and complex I is not restricted to the active site. Of its fourteen subunits, all except MbhA, MbhB, MbhC and MbhF are homologs to complex I subunits. These four subunits, on the other hand, can be found in the Multi-resistance and pH adaptation (Mrp) sodium-proton-antiporter. Homologs to MbhD, MbhE, MbhG, MbhH and MbhI can be found in all three proteins and ALSs (MbhH/MrpD/Nqo14) are core elements in all three classes of proteins, although their functionality differs significantly:(91–95) While Mbh and complex I build up a *pmf/smf*, Mrp is a proton sodium antiporter.(17, 82, 95, 96) Together with biophysical characterization, bioinformatic analysis have revealed new insight into the function of Mbh throughout the early 2000s.(82, 91–93, 95, 97, 98) The homology to complex I indicates that it could function as a redox-driven proton pump, while similarities to Mrp indicate antiporter activity. In a related respiratory chain, the two functionalities are achieved by an antiporter and a separate proton pump.(82)

With this in mind, one model trying to make sense of the seemingly contradictory experimental data proposed Mbh as a secondary sodium transporter. In this model, a proton would be first pumped against the gradient, then a proton would return across the membrane with the gradient and this in turn would be coupled with sodium transport against the membrane potential, effectively establishing a secondary sodium gradient. The delay between proton transport sodium-proton-exchange, which could only happen after the primary proton gradient has achieved sufficient magnitude, would account for the macroscopically observed ΔpH.(99) Although there is a basis for assuming proton transport followed by sodium-proton-exchange in Mbh’s evolutionary relations, this model has been criticized for failing to clearly address how such a proton back-and-forth pumping would conserve the small amount of driving force that Mbh commands.

The game for Mbh research changed in 2018, when Yu et al. managed to solve the cryo-EM structure of Mbh at a resolution of 3.7 Å (PDB ID 6cfw, see Figure 5)(100) (the first structure of Mrp would be solved a few years later)(101). The structure revealed strong homology to complex I not only on a sequence but also a structural

level: The characteristic L-shape is reproduced, the three ISCs form a chain virtually unchanged from the arrangement of N2, N6a and N6b in complex I (compare figures 2A (lowest three ISCs) and 5), and the three conserved loops (in Mbh, all three were resolved, although the MbhI/Nqo7 TM1-TM2 loop could only be backbone traced) are equally well reproduced. Surprisingly, even the Q-cavity is clearly present, although the enzyme does not catalyze any known Q reactions and the entrance to the Q cavity from the membrane is blocked by bulky aromatic amino acids.

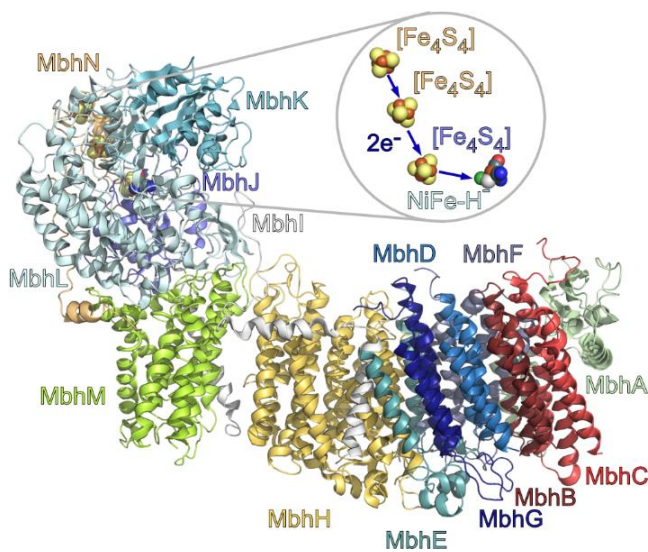


Figure 5: Structure of the Membrane bound hydrogenase from *P. furiosus* (PDB ID 6cfw)(100) Note the cleft between the ALS MbhH and the hinge region (MbhM). MbhG, MbhD, MbhB and MbhC form a symmetrical 12-helix bundle that constitutes the core of the putative sodium transport module. Inset: A chain of three iron-sulfur-centers accepts electrons from ferredoxin, which are combined with two protons in the characteristic NiFe active site (shown here with a H⁺ ligand in the Ni-C state, which is the resting state of NiFe-hydrogenases).

There are several noteworthy deviations between Mbh and respiratory complex I, though: In complex I, the 'hinge' subunit Nqo8 is separated from the first ALS (Nqo14/MbhH) by three small subunits (Nqo7/MbhI_{nt}, Nqo10/MbhD+MbhE and Nqo11/MbhG). This sequence of subunits (ALS + three small subunits, called the conserved core from here on) is rotated by 180° as a whole in the membrane plane of Mbh so that the ALS faces the hinge subunit and the three small subunits face the terminal edge of the membrane domain. Furthermore, a large, putatively lipid-filled, cleft in the electron density map separated hinge and ALS. Two of the three small subunits (MbhD and MbhG) form a symmetrical twelve-helix-bundle with MbhB and MbhC, which is at the center of the putative sodium-transport domain, thus bridging ALS and sodium-transport module.

The modularity of the complex I family throughout its evolution is one of its more fascinating features (Figure 6): A conserved core is present in quite different classes of proteins (incidentally also in the membrane bound sulfane reductase (Mbs)).(102) It can be used to couple different modes of energy accumulation and consumption. Energy can be provided as redox energy (complex I, Mbh, Mbs) or probably by proton transport along a gradient (MrpA/Nqo12 in Mrp). This energy can then be used for either active proton transport in additional ALS (complex I, Mbs) or active sodium transport (Mbh, Mrp: whether additional proton transport processes are involved here is an open question). This plug-and-play functionality can also be observed in the

hydrophilic domain: Additional subunits can be added or removed to tune the preferred electron donor: The shorter hydrophilic domains of Mbh and the so-called photosynthetic complex I (Ndh)(103) accept electrons from Fd with a smaller magnitude of the redox potential, while additional subunits and ISCs therein can be added to utilize the greater potential magnitude of NADH. Variations are also observed in the terminal electron acceptors: Various forms of Q (ubiquinone, plastoquinone, menaquinone, etc.) with different redox potentials are utilized by complex I isoforms, while Mbh and Mbs deploy a strikingly similar architecture featuring the conserved Q-cavity without actually having Q as a substrate. Notably, adding or removing subunits does not seem to require major changes in the topology of the single subunits.(82, 104)

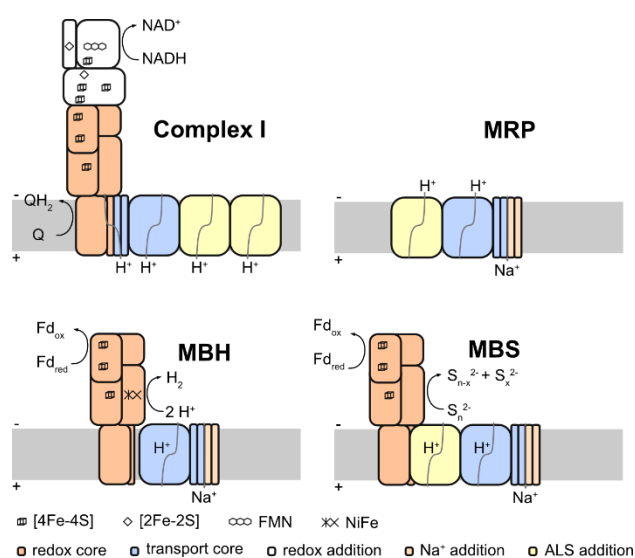


Figure 6: Modular evolution of the complex I family. Advances in structural biology have revealed conserved building blocks for protein complexes with different functionalities, which are shared between complexes catalyzing quite different reactions and transport processes.(104) The addition of the NiFe-center enables hydrogen chemistry(100), while an open coordination in the N2 cluster in Mbs enables sulfane reduction.(102) ALSs can be added and even rotated (e.g. subunit MbhH in comparison to Nqo14) to adjust for the way transport is powered or powering (figure inspired by Calisto et al.).(17)

Recently, structures for several relatives of Mbh have been solved at varying resolutions.(100–102, 105) These new structures and their strong homology to Mbh have prompted, again, speculations about Mbh functionality based on homology considerations. Different hypotheses, especially for the proton/ion-pathways have emerged, e.g., placing sodium transport in the ALS or suggesting sodium pathways throughout the sodium transport module.(101, 102, 105) However, at the time this thesis is written, new experimental evidence since the release of the first structure has been scarce and the question of the mechanism of Mbh remains wide open, a question we will try to address in this work.

1.4 Computational biochemistry

Traditionally, computational chemistry has been used as a means to interpret and explain experimental results. However with the steep rise of computing power (cf. Moore's law(106)) the computational predictive power was enhanced. *In silico* experiments offer access to a set of observables on timescales and length-scales complementary to those provided by *in vitro* and *in vivo* experiments. To a certain degree, the borders between

computation and experiment have blurred: Although cryo-EM is an experimental technique, the computing power invested into refining the raw data into molecular structure rivals those of dedicated simulations. On the other hand, synthetic biology and protein engineering routinely use computational predictions to inform experiments, shrink the pool of, e.g., drug candidates for *in vitro* experiments by weeding out putative constructs with undesirable qualities and generally easing the experimental burden.(107–109) Nowadays, even classical biochemical techniques such as protein purification by high pressure liquid chromatography are modeled *in silico* to predict optimal buffer compositions and resins for the columns.(110, 111) The rise of artificial intelligence and deep learning will probably tighten the bond between experimental and computational work.(112)

Yet, pure computational biochemistry itself has also evolved rapidly. In 1977, the very first protein molecular dynamics simulation featured bovine pancreatic trypsin inhibitor with 58 amino acid residues simulated for 8.8 picoseconds.(113) Contemporary simulations can leverage the power of modern GPU clusters, and supercomputers follow the dynamics of several millions of atoms for microseconds (or even longer if dedicated hardware is utilized).(114–116)

In this work, we use computational chemistry in all its facets: We show how it can be used as a molecular microscope to elucidate dynamics with atomic detail, follow thousands of water molecules at femtosecond time-resolution to understand how ions and protons are transported, calculate free energies for electrostatic switches and chemical reactions and finally use our insights to create new proteins based on motifs we have observed in nature.

2 Aims of the study

In this study, we used multi-scale modeling techniques to investigate the functional role of hydration and electrostatic switching principles in the regulation and control of biological proton and ion transporters. Key questions addressed in this work are:

- How can individual charge transfer processes be coupled by electrostatic interactions?
- What is the role of hydration dynamics in the regulation of charge transfer processes?
- Are ion-pairs and charged networks stabilized in the hydrophobic core of a protein?
- How is the free energy from redox reactions used to drive charge transfer processes?
- Can a charged network be engineered in a small artificial protein?

To address these questions, we used computational methods in both an analytical and a synthetic manner: On the one hand, we performed large-scale simulations of several members of the complex I family, particularly the bacterial complex I from *T. thermophilus* (Articles I & II), the mitochondrial complex I *M. musculus* (Article III) and the membrane bound hydrogenase from *P. furiosus* (Article V). On the other hand, we used *in silico* methods to guide the design of small *de novo* four-helix bundles incorporating charged networks forming important coupling elements in the complex I family (Article IV).

3 Theory: Computational multi-scale modeling

Biological processes span multiple time and length scales and are often intertwined between multiple scales.(117) Over the years, a plethora of computational methods with different capabilities and target scales has emerged allowing the modern computational chemist to address processes from the femtoseconds to the seconds scale (Figure 7).(118) The studied system size is usually correlated with the modeled time scale: While on the one end of the spectrum *ab initio* quantum chemical calculations are useful for describing reactions between a handful of atoms for femtoseconds to picoseconds, coarse grained methods sacrifice spatial and temporal resolution to address millions of atoms for milliseconds. Between these two extremes, classical MD simulations are often used to model systems on microsecond timescales. There is always a trade-off between accuracy and achievable sampling and it is crucial that the method used is tailored to the specific question:(119) Highly accurate methods may be able to yield precise predictions for specific processes, but the limitation in sampling can actually limit the accuracy if conformational sampling is impaired too much, and rare events may not be captured at all. Free energy methods provide a work-around by specifically enriching the sampling of the process in question but require a precise knowledge of the process in the first place to be able to choose a reaction coordinate for describing the reaction and steering the free energy calculation. On the other hand, it can also be challenging to achieve chemical accuracy ($< 1 \text{ kcal mol}^{-1}$) using classical force-field based models. (120)

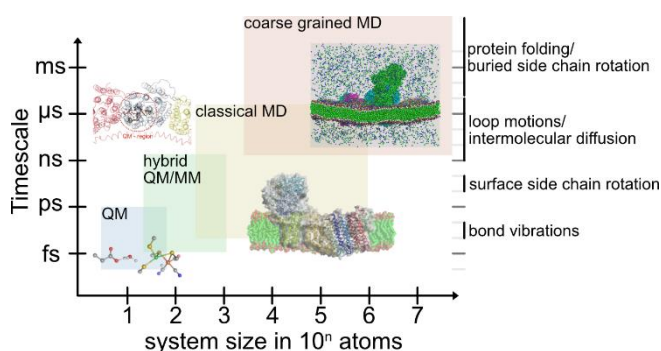


Figure 7: Overview over biological and computational length and time scales. Different levels of theory are used for probing different scale regimes. Examples of timescales, commonly involved in biological processes, are given on the right hand side based on Zwier et al.(121) Renderings are partially taken from Articles III and V.

Here, we used a combination of several levels of theory to address the scientific questions. While atomistic MD simulations on the μs scale were used for describing conformational transitions of proteins, we employed quantum chemical and hybrid QM/MM methods to describe chemical reactions, which cannot be addressed with classical methods. In this section, the theory behind the methods utilized in this study are briefly reviewed, while an overview of the specific simulations performed is given in chapter 4.

3.1 Quantum mechanics

The laws of classical physics are insufficient to accurately describe electronic structure and motion that form the basis of chemical reactivity, so it is necessary to turn to the laws of quantum mechanics. The theoretical basis of the simulations in this study relies on solving approximations to the time independent Schrödinger eigenvalue equation

$$\hat{H}\Psi = E\Psi \quad (3.1)$$

with the Hamiltonian operator \hat{H} describing the interactions within the system of interest with the energy E as its eigenvalue and the wavefunction Ψ associated with the system. By assuming time-independence of the Hamiltonian, we exclude interaction with any non-stationary external fields, e.g. photons. Apart from neglecting relativistic effects, we will further apply the Born-Oppenheimer approximation, neglecting the coupling between electronic and nuclear motion by using nuclear coordinates as parameters for solving electronic motion. This assumption is often valid because of the large difference (ca. four orders of magnitude) between the timescales associated with electronic and nuclear motion, and it allows us to describe the total wavefunction as a product of an electronic wavefunction Ψ_{el} and a nuclear wavefunction Ψ_n . The electronic Hamiltonian operator \hat{H}_{el} , associated with the former by the electronic time-independent Schrödinger equation

$$\hat{H}_{el}\Psi_{el} = E_{el}\Psi_{el} \quad (3.2)$$

is then

$$\hat{H}_{el} = \hat{T}_{el} + \hat{V}_{ne} + \hat{V}_{ee} \quad (3.3)$$

with the kinetic energy operator of the electrons \hat{T}_{el} , the potential energy operator for electron-nuclear interactions \hat{V}_{ne} and the potential energy operator for electron-electron interactions \hat{V}_{ee} . We can expand these operators in atomic units to

$$\hat{H}_{el} = -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \sum_{i=1}^n \sum_{I=1}^N \frac{Z_I}{r_{iI}} + \sum_{i<j}^n \frac{1}{r_{ij}} \quad (3.4)$$

with the Laplace operator for the kinetic energy of electron i ∇_i^2 , the nuclear charge of nucleus I Z_I , the distance between nucleus I electron i r_{iI} and the distance between electrons i and j r_{ij} .

3.1.1 Hartree-Fock theory

Unfortunately, analytical solutions to the electronic Schrödinger equation can only be obtained for one-electron systems (like the H_2^+ molecule). Since biological and chemical reactions feature many-electron systems, it is necessary to rely on numerical solutions. Hartree-Fock theory is a wavefunction based theory for solving the electronic structure of molecular systems and forms the basis for most higher levels of theory. The Hartree-Fock theory is based on the mean-field approximation, where we neglect electron correlations, and subjects each electron only to the average potential generated by all electrons. The wavefunction is modeled as a Slater determinant Φ_{SD} , which can be written for N electrons and N spin orbitals as follows:

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{vmatrix} \quad \text{with} \quad \langle \phi_i | \phi_j \rangle = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \quad (3.5)$$

While the columns of this determinant contain single-electron wavefunctions, the electron coordinates are in the rows. Entries in the Slater determinant are one-electron wavefunctions, which are spin orbitals. The Slater determinant inherently satisfies the anti-symmetry requirement, stating that interchange of any two electron coordinates in the total electronic wavefunction must change its sign (since electrons are fermions) and in extension adheres to the Pauli principle.

The energy $E[\Psi]$ associated with the single-determinant trial wavefunction Φ is optimized by the variational principle, which ensures that the true ground state energy E_0 is a lower bound for $E[\Psi]$.

$$E[\Phi] = \frac{\langle \Phi | \hat{H} | \Phi \rangle}{\langle \Phi | \Phi \rangle} \geq E_0 \quad (3.6)$$

Therefore, finding the trial wavefunction with the lowest energy corresponds to finding the one which is the best approximation of the true wavefunction. With the introduction of the molecular spin orbitals ϕ_i (MO), we can formulate the Fock pseudo-eigenequation as

$$\hat{F}_i \phi_i = \varepsilon_i \phi_i \quad (3.7)$$

with the Fock operator \hat{F}_i and its associated pseudo-eigenvalue ε_i , the orbital energy which can be interpreted as the ionization energy within the frozen molecular orbital approximation according to Koopman's theorem. The reason why this is only a pseudo-eigenvalue equation lies in the definition of the Fock operator, which can be expressed as

$$\hat{F}_i = \hat{h}_i + \sum_j^N (\hat{J}_j - \hat{K}_j) \quad (3.8)$$

in which the one-electron operator \hat{h}_i describes the movement of the electrons in the field of the nuclei ($\hat{T}_{el} + \hat{V}_{ne}$ in equation 3.3), the Coulomb integral J_{ij} describes the classical repulsion between electrons i and j and the exchange integral K_{ij} expands on this subject but has no classical interpretation. The Fock operator for a single electron depends on the interaction with all other electrons, i.e., occupied orbitals via the Coulomb and exchange integrals. This implies that to know one specific molecular orbital, one has to know all molecular orbitals, which is why self-consistent optimization methods have to be applied.

To solve this, the basis set expansion approach is used, where each unknown MO is expanded as a set of known basis functions χ , commonly known as atomic orbitals:

$$\phi_i = \sum_a^{M_{basis}} c_{ai} \chi_a \quad (3.9)$$

with c_{ai} the coefficients associated with each basis set function χ_a . This approach is known as linear combination of atomic orbitals (LCAO), and the Hartree-Fock equations in terms of the atomic basis are known as the Roothaan-Hall equations (for closed shell systems):

$$\mathbf{FC} = \mathbf{SC}\epsilon \quad \text{with } F_{ab} = \langle \chi_a | F | \chi_b \rangle \text{ and } S_{ab} = \langle \chi_a | \chi_b \rangle \quad (3.10)$$

with the Fock matrix F , the overlap matrix S , the coefficient matrix C and the diagonal matrix of the molecular orbital energies ϵ . The matrix eigenvalue problem in equation 3.10 can be solved iteratively from an initial guess (extended Hückel theory is commonly used) until self-consistency of the molecular orbitals is reached.

If electron correlation is important for the processes to be studied (e.g. excitation processes), so-called post-Hartree-Fock methods expressing the wavefunction using multiple determinants have to be used. However, the increased accuracy of such methods (e.g. configuration interaction(122, 123) or coupled cluster(124–126)) comes with a significant increase in computational cost.

3.1.2 Density functional theory

Wavefunction methods, like the Hartree-Fock theory described in the previous section, approximate the high-dimensional wavefunction in order to solve the Schrödinger equation for a system. Density functional theory (DFT) in contrast is formulated based on the electron density, a three-dimensional object with a clear interpretation and commonly obtained by experimental methods like X-ray crystallography. This dimensionality reduction from N electronic wavefunctions to three spatial coordinated (plus one spin coordinate) simplifies the interpretation considerably. DFT emerged when Hohenberg and Kohn showed that the electron density ρ is sufficient to completely determine the ground-state electronic energy and therewith the associated Hamiltonian, and further applied the variational principle to derive the electron density.

The key problem is that while DFT exactly maps the density as a function of electron coordinates to the energy, the exact relationship between the two is unknown. Although orbital-free variants of DFT exist, they are generally outperformed by methods based on the works of Kohn and Sham, who reintroduced orbitals to improve the poor representation of kinetic energy in orbital-free methods. The Kohn-Sham model solves the problem of calculating the precise kinetic energies by splitting the free energy in two parts: One part is calculated for a hypothetical system of non-interacting orbitals with the same density as the ‘real’ system, which is described by the Kohn-Sham-orbitals ϕ_i

$$\rho_{approx} = \sum_i^{N_{el}} |\phi_i|^2 \quad (3.11)$$

as a sum over all occupied orbitals. This corresponds to the kinetic energy of a single-determinant description and accounts for ca. 99% of the actual kinetic energy. The remaining part of the kinetic energy is packed into the so-called exchange-correlation term, which gives a functional of the form

$$E_{DFT}[\rho] = T_s[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \quad (3.12)$$

The first three terms on the right-hand side of equation 3.12 are exactly the same as in the Hartree-Fock formalism (see above). The exchange correlation term can be defined as the deviation of the non-interacting system from the exact solution

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (E_{ee}[\rho] - J[\rho]) \quad (3.13)$$

The first subtraction describes the kinetic correlation energy, while the second contains a mixture of potential correlation and exchange correlation. Approximations for the unknown exchange correlation functional have been a major focus of research since the 1990s and the functional form for this functional is the main difference between DFT methods.

3.2 Classical molecular dynamics simulations

Classical molecular dynamics simulations describe the system by a potential energy function $V(x(t))$ of atomic coordinates. Solving the Newtonian equations of motion by calculating the gradient ∇ of this potential energy function with respect to the time t , allows the calculation of forces $F(x(t))$, accelerations $a(t)$, velocities $v(t)$ and positions $x(t)$

$$-\nabla V(x(t)) = F(x(t)) = ma = m \frac{dv}{dt} = m \frac{d^2x}{dt^2} \quad (3.14)$$

Thus, the potential energy and Newton's second law suffice to determine the motion of a system over time from a given starting configuration, usually prepared from experimentally resolved structures. A plethora of methods exists to numerically solve these equations in practice (see section 3.2.2).

3.2.1 Force fields for molecular dynamics simulations

Electronic structure can be approximated with a good degree of accuracy for many condensed phase processes by assuming a model of atoms as points connected by bonds. For a chemist, this kind of model has the advantage that it exactly maps to common ways of depicting the molecular geometry of chemical compounds, e.g. by Lewis electron dot structures(127), which can be interpreted intuitively.

To ensure that such a model describes physical reality with a high degree of accuracy, the potential energy of the system is described as a parametric function of atomic, i.e., nuclear, coordinates and charges, and a global optimization of all parameters is performed for agreement with experimental observables or observables from higher levels of theory. A general way of formulating such an potential energy V as function of the nuclear coordinates R is

$$V(R) = V_{bonded} + V_{nonbonded} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{van-der-Waals} + V_{electrostatics} \quad (3.15)$$

Bonded terms describe the geometric constraints imposed by the electronic structure of the molecule, while nonbonded terms describe interactions between all atoms through space. Terms describing bond stretching and angle bending are usually given as the first terms of a Taylor expansion (although Morse potentials have been used, too), while dihedral terms describing rotation around bonds (i.e. 1-3 interactions) are derived by Fourier-series. These terms are commonly known as improper dihedrals or out-of-plane terms. The functional forms for the bonded terms used in the CHARMM36m force field applied in this work are as follows:

$$V_{bonds} = \sum_i^N k_b (b - b_0)^2 \quad V_{angles} = \sum_i^N k_\theta (\theta - \theta_0)^2 \quad V_{dihedrals} = \sum_i^N k_\phi (1 + \cos(n\phi - \delta)) \quad (3.16)$$

With k_b, k_θ and k_ϕ being equilibrium force constants, b_0 and θ_0 being equilibrium bond lengths and bond angles respectively, and n and δ being the periodicity and phase shift for dihedral angles. The sums run over all bonds between two covalently bound atoms, all angles between three covalently bound atoms and all dihedrals between four covalently bound atoms. As such, these terms scale linearly with the numbers of bonds angles and dihedrals respectively and are computationally cheap.

In contrast, non-bonded terms are calculated between all pairs, regardless of connectivity and thus scale with the square of the number of atoms simulated, although the nonbonded terms for pairs of atoms sharing a bond, angle or dihedral (i.e. neighbors up to $i+3$) are commonly excluded to limit cross-dependency effects between bonded and nonbonded terms. The functional forms for the nonbonded terms used in the CHARMM36m force field applied in this work are as follows:

$$V_{van-der-Waals} = \sum_{i,j \neq i} \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad V_{electrostatics} = \sum_{i,j \neq i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (3.17)$$

ϵ_{ij} and σ_{ij} in the van-der-Waals term for dispersive interactions describes the well-depth and position of the minimum in the Lennard-Jones potential with the n^6 term modeling the attractive regime radii r_{ij} greater than the potential minimum and the n^{12} term describing the strong repulsion experienced by atoms forced closer than the potential minimum. The electrostatics term is a direct pairwise implementation of Coulombs law for the interaction between two atoms i and j with charges q_i and q_j at a distance of r_{ij} from each other with ϵ_0 the permittivity of vacuum.

In principle, both these terms scale with $O(N^2)$. However, since the van-der-Waals term decays very fast with $1/r^6$ it can be treated efficiently with fairly short-range cut-off schemes without significant loss of accuracy. In contrast, the electrostatics term, which decays only with $1/r$ has significant magnitude even at higher distances and cannot simply be cut off. Solutions to the problem of long-range-electrostatics are discussed below.

Many force-fields include additional terms to improve the description of molecular systems. In the case of the CHARMM force field, three additional terms are included: The Urey-Bradley term, an additional correction for the 1-3 interaction between covalently bound atoms, the improper dihedral term, a four-residue term with a harmonic form enforcing the planarity of e.g. double-bonds and the grid based CMAP correction, which improves the agreement with experimentally determined dihedral angles e.g. in proteins.

3.2.2 Boundary conditions

Each simulations necessitates the treatment of boundaries, and the dominance of boundary effects is a phenomenon unique to molecular simulations. In reality only a minuscule part of a system lies at the boundary of the system, i.e. within a few nanometers of the interface with another phase. For experimental equilibrium observables, the parts of a system at the boundaries are largely irrelevant, since the part of the system that is not at the interfaces will dominate the ensemble average. In contrast, in simulations, most of the simulated

system is in the boundary region (depending on the system size), and boundary effects can have outsized influence.

The easiest boundary, which can be applied to a biological system, e.g. a protein, is no boundaries at all, i.e. a simulation in vacuum. This introduces severe artifacts: nonbonded interactions within the protein have outsized influence and electrostatics are not screened, since there is no solvent at all to counteract them. Furthermore, the systems tend to be overly compact since they try to maximize self-interaction and surface tension effects will arise. There are two trivial strategies to tackle these effects: First, one can apply an implicit solvent model.⁽¹²⁸⁾ These models solve the problem of electrostatic screening by solving the Poisson-Boltzmann equations with a low-dielectric mean field within the solvent-exclusion volume of the protein (often with epsilon around 4) and a field with a water-like dielectric, i.e. 80, outside. While this type of model is rather good at treating long-range electrostatics, the local effects at the dielectric boundary are not well described. Second, a solvent box or droplet can be constructed around the protein with a fixed boundary. This passes the boundary problem on to the solvent, which is usually of lesser interest, but still contains a hard boundary between vacuum and solvent by construction. Local effects between protein and explicit solvent are modeled well, but long range electrostatics beyond the reach of explicit solvation are described poorly, which may be counteracted by simulating a very large droplet. However, this is computationally expensive and impractical. The most common choice of boundaries is the application of periodic boundary conditions (PBC) in combination with lattice sum algorithms like the Particle-Mesh-Ewald (PME) method for the treatment of long-range electrostatics. In this framework, the protein is placed in an explicit solvent box, which is formally surrounded by periodic copies of itself. With a large enough simulation box, the interaction with artificial periodic images can be kept minimal, while treatment of electrostatics beyond a cut-off, often realized in practice by a smooth switching function, in Fourier-space keeps the computational burden tractable. This approach is used for nearly all biological classical MD simulations.

Another important choice for MD simulations is the choice of the ensemble. It is advisable to perform simulations in such a way that *in silico* observables can be easily compared to experimental measurements. Ensembles are determined by the conservation laws enforced during simulation. The most common conservation enforced is conservation of the number of particles N , which may be abandoned if conservation of chemical potential in the grand-canonical ensemble is desired. Temperature control, i.e., conservation of kinetic energy, is equally common but may be omitted if total energy conservation (NVE or microcanonical ensemble) is more important for a specific observable to be measured. For direct comparison with most biochemical experiments, NVT and NPT ensembles are preferred. In the former, the volume of the simulated system, the temperature and the number of particles in the system are conserved. This canonical ensemble corresponds to a closed test tube and is enforced by one of the many types of thermostats available for molecular simulations. The latter isothermal-isobaric ensemble applies a barostat in addition to the thermostat, which sacrifices

constant volume to keep the pressure in the system constant. This corresponds to an experiment in an open test tube and is preferred for membrane simulations. Earlier versions of lipid force-fields also required a constant area of the simulation box (in the xy -plane, so called $NPTa$ ensemble), but improved force-fields are able to reproduce correct extensions and dynamics of membranes without this constraint.

In this work, for temperature control in the NVT ensemble, the (stochastic) Langevin approach is generally used, which achieves a constant temperature by stochastic coupling to a reservoir based on the general Langevin equation

$$M\dot{v} = F(x) - \gamma v - \sqrt{\frac{2\gamma k_B T}{M}} R(t) \quad (3.18)$$

which uses a friction coefficient γ and a univariate Gaussian random process $R(t)$ to couple the system to a temperature reservoir by adding dissipative forces $-\gamma v$ and fluctuation $\sqrt{\frac{2\gamma k_B T}{M}} R(t)$ to the Newtonian equations of motion. This also necessitates a change to the integrator, which has to incorporate the coupling terms. In this work, we use the Brünger-Brooks-Karplus (BBK) method,(129) which is an extension of the Verlet integrator. Its position recurrence relation is given by

$$x_{n+1} = x_n + \frac{1 - \frac{\gamma\Delta t}{2}}{1 + \frac{\gamma\Delta t}{2}}(x_n - x_{n-1}) + \frac{1}{1 + \frac{\gamma\Delta t}{2}} \Delta t^2 \left[\frac{F(x_n)}{M} + \sqrt{\frac{2\gamma k_B T}{\Delta M}} Z^n \right] \quad (3.19)$$

which introduces a set of Gaussian random variables Z^n , which have a mean of zero and a variance of one. Therewith, time evolution of velocities and positions is described by the time evolution of the corresponding probability distribution, which is governed by the Focker-Planck equation. A similar formalism based on the Langevin-piston method and Hoover's method can be extended for temperature control in the NPT ensemble.

3.3 Hybrid quantum mechanics/ molecular mechanics simulations

As described in the previous section, boundary conditions play an important role in describing a system accurately. However, the PBC-PME approach, which is favored for classical MD-simulations, is not particularly suited for QM-simulations. Even including a solvent box is usually prohibitively expensive on a purely QM level due to its high cost and at least cubic scaling with the number of particles of QM-algorithms. While implicit solvent models and *in vacuo* simulations are commonly used for cluster models, there is a more elegant solution: Most biological systems have a more or less clearly defined active site, in which chemical reactions occur. For the rest of the system, a classical treatment it is often adequate. This allows to embed an QM-active region, treated, e.g., on the DFT level, in a classical surroundings (usually applying finite boundary conditions to that part). In this case the classical region is in principle the boundary treatment for the QM-region. Chemical bonds can be cut across this boundary by introducing so called link atoms, often hydrogen-like, which simulate a saturated end of the bond to each of the two sub-parts of the system. Link atoms are commonly introduced at

the C α -C β -bond in proteins, so that the side chain of a reactive amino acid can be treated quantum mechanically, while the backbone remains in the classical region.

There are multiple embedding schemes to treat the interaction between classical and QM regions of such a multiscale simulation. In a subtractive scheme, the energy of the QM-region is calculated both on a classical and a QM-level. The total energy of the system is then the sum of QM-level energy and the classical energy of the whole system minus the classical energy of the QM-region. In contrast, the additive scheme calculates the energy for the classical part classically, the QM region quantum chemically and introduces an additional coupling term that reflects the dependence of the two parts on each other. The coupling term can be calculated in a mechanical embedding scheme, where a classical charge model is developed for the QM-part and the coupling is treated as electrostatic interaction between these fixed charges and the charges of the classical region. Alternatively, the charges of the MD region can be incorporated as point charges, which polarize the QM-region in the electronic embedding scheme. However, this can lead to over-polarization of the QM-region by the MD point charges, so it is often necessary to re-distribute or neglect the classical charges close to the QM-region. Generally, it is important to check for convergence of calculated observables with the size and shape of the QM-region to ensure that boundary artifacts are minimized.

3.4 Free energy calculations

The free energy (either the Gibbs free energy (G) in an NPT ensemble or the Helmholtz free energy (F) in an NVT ensemble) is among the most commonly applied thermodynamic concepts since the free energy associated with a state i can be related to the probability of this state. Applied to a e.g. a chemical reaction or a conformational change, the associated free energy difference ΔG describes whether this process will occur spontaneously ($\Delta G < 0$ at constant pressure and temperature without non-PV work occurring), while the height of the rate-limiting free energy barrier along the process can be related to its associated timescale. In contrast to the potential energy V directly obtained from the MD energy function, the free energy does also include entropic contributions, which can be non-negligible for many processes.

Thus, calculating a free energy profile along a reaction coordinate is a common task in MD simulations and often carried out by umbrella sampling. The general idea is to perform several independent simulations (windows) with harmonic restraining potentials at different points along the reaction coordinate so that the potential of mean force (pmf) along the process described by ζ is fully sampled. The addition of the biasing potential ensures that points along the reaction barrier with a low probability of being sampled, most prominently transition state regions, are nevertheless well represented in the ensemble derived from the enhanced sampling simulations, and that the windows are spaced in a way that there are no gaps in sampling along the reaction coordinate. This allows to model processes occurring on timescales far beyond the reach of plain MD. However, the distribution of states derived from such biased simulations is *a priori* not the

distribution associated with an unbiased sampling along the reaction coordinate. Thus, free energies derived from such a distribution must be corrected for the biases applied. Traditionally, this is achieved with the weighted histogram analysis method (WHAM).(130, 131) Despite its widespread success, a plethora of modern successor methods has been developed, aiming to correct e.g. systematic errors arising from imperfect equilibrium sampling in single windows. One of these methods, the dynamic histogram analysis method extended to detailed balance (DHAMed)(132) was used in this work and will be discussed here. For a brief discussion of the WHAM formalism see e.g. Kumar et al.(131).

In the canonical ensemble, we will derive a free energy G_i , which is associated with an equilibrium distribution p_i of states i .

$$p_i \equiv e^{-G_i/k_B T} = \frac{\int_i e^{-U(x)/k_B T} dx}{\int e^{-U(x)/k_B T} dx} \quad (3.20)$$

for points in configuration space x with associated potential energy $U(x)$ at an absolute temperature T and with Boltzmann's constant k_B . The numerator extends over configurations associated with state i , while the denominator spans the whole configuration space. In umbrella sampling, the reaction coordinate typically is a collective variable $q(x)$, and the process of interest, for which a pmf should be derived, is defined as a state as the region between q_i and $q_i + dq$ allowing p_i to be described as:

$$p_i = \frac{\int dx \delta[q_i - q(x)] e^{-U(x)/k_B T}}{\int e^{-U(x)/k_B T} dx} \quad (3.21)$$

with the Dirac delta function δ . However, in umbrella sampling, we do not sample the unbiased potential energy surface $U(x)$ but usually simulate a set of A runs (defined by an iterator $\alpha=1,2,3,\dots,A$) biased by a potential $u^\alpha(x)$ each and associated with a biased potential energy surface $U^\alpha(x)$ each.

$$U^\alpha(x) = U(x) + u^\alpha(x) \quad (3.22)$$

It should be noted that this also applies for the common method of temperature replica exchange simulations, since a change in temperature can be expressed as an effective bias potential in the canonical ensemble. For umbrella sampling, it is typically valid that (harmonic) biasing potentials are constant within a state i , if configuration space is divided in a set of states $i=1,2,\dots,M$. We assume that the unbiased relative population P_i of states i changes over time t according to a master equation

$$\frac{dP_i(t)}{dt} = \sum_{i \neq j} [k_{ij} P_j(t) - k_{ji} P_i(t)] \quad (3.23)$$

with the first order rate coefficient for a transition from i to j k_{ij} . The equation formally constitutes a time-continuous Markov state model. For this model, we assume microscopic reversibility, detailed balance and equilibrium populations normalized to 1. It is valid to assume that the time evolution of the same states with an additional bias follows a similar master equation

$$\frac{dP_i^\alpha(t)}{dt} = \sum_{i \neq j} [k_{ij}^\alpha P_j^\alpha(t) - k_{ji}^\alpha P_i^\alpha(t)] \quad (3.24)$$

with the same assumptions, except that equilibrium populations under their bias are proportional to the product of the unbiased populations and the bias factors, which are known in umbrella sampling.

$$p_i^\alpha \propto p_i e^{-u_i^\alpha/k_B T} \quad (3.25)$$

Under the assumption of the master equation model defined by eq. 3.24, equilibrium populations can be obtained by likelihood-maximization. During each biased run α , we will measure both the time t_i^α spent in a state/bin i as well as the number of transitions from each state j to state i N_{ij}^α . Both observables are easily obtainable from time-discrete MD simulations with timestep Δt , for which

$$t_i^\alpha = \sum_j N_{ij}^\alpha \Delta t \quad (3.26)$$

assuming no transitions were missed due to time-discretization. Detailed balance can be enforced on the model by constructing the rate coefficients k_{ij}^α from equilibrium populations p_i^α with the help of coefficients x_{ij}^α that are positive and symmetric with respect to exchange of the state indices, i.e. $x_{ij}^\alpha = x_{ji}^\alpha$:

$$k_{ij}^\alpha = \frac{x_{ij}^\alpha}{p_j e^{-u_j^\alpha/k_B T}} \quad (3.27)$$

Maximizing the log-likelihood of the set of biased trajectories under this model with respect to the unknown parameters x_{ij}^α and p_j yields a closed set of expressions for the equilibrium populations:

$$p_j = \frac{\sum_{\alpha=1}^A \sum_{i \neq j} (N_{ij}^\alpha + N_{ji}^\alpha) / \left(\frac{1}{p_j} + \frac{1}{p_i} \frac{t_i e^{u_i^\alpha/k_B T}}{t_j e^{u_j^\alpha/k_B T}} \right)}{\sum_{\alpha=1}^A \sum_{i \neq j} N_{ij}^\alpha} \quad (3.28)$$

from which free energies can be calculated as

$$G_j = -k_B T \ln(p_j) \quad (3.29)$$

This solution is unique except for a scaling factor, which can be found by iterating eq. 3.28 from the right as:

$$p'_j = \frac{\sum_{\alpha=1}^A \sum_{i \neq j} (N_{ij}^\alpha + N_{ji}^\alpha) / \left(\frac{1}{p_j^{(m)}} + \frac{1}{p_i^{(m)}} \frac{t_i e^{u_i^\alpha/k_B T}}{t_j e^{u_j^\alpha/k_B T}} \right)}{\sum_{\alpha=1}^A \sum_{i \neq j} N_{ij}^\alpha} \quad \text{with } j = 1, 2, \dots, M \quad (3.30)$$

and with

$$p_j^{(m+1)} = p'_j / \sum_{i=1}^M p'_i \quad \text{with } j = 1, 2, \dots, M \quad (3.31)$$

Iteration can be started, e.g., from $p_j^{(1)} = 1/M$. It should be noted that states, which have fewer than two transitions into the state or fewer than two transition out of the state are excluded to ensure proper equilibrium estimation. Asymptotic convergence of the populations estimated from the DHAMed model to the correct limit can be shown from microscopic time reversibility in the case of imperfect sampling.

This methodology is also compatible with replica exchanged methods, where exchange between the biased runs α is periodically allowed so that the bias-factors $e^{-u_i^\alpha/k_B T}$ between runs α and α' can be swapped every $n \times \Delta t$ steps. Exchanging bias factors is usually accepted or declined based on energetic considerations e.g. via a

Metropolis-Monte-Carlo criterion. After the runs have been completed, they are sorted according to their initial bias factor so that continuous trajectories are regenerated and bias correction can be applied to generate the associated pmf.

4 Methods

This section will provide a high-level overview of the experiments performed and computational methods used in this work. For the sake of completeness and reproducibility, the experimental details for quantum chemical calculations and hybrid QM/MM simulations, which were performed by P.S.M, A.P.G. and V.R.I.K., are also summarized (see author contributions detailed on pg. vi). M.R. performed MDFF refinement (see section 4.4.1). For additional details, particularly on individual simulations, please refer to the supporting information of articles I to V. The description here focuses on the computational experiments contributing to the results presented in section 5.

4.1 General methods

In this section, an overview of standard methods that have been applied to multiple systems is given. In case the treatment of a specific system strays from the workflow described here, the deviations will be given in detail in the respective section below the general overview. If no further information is given, the workflow adheres to the one given in this section. Methods that were only applied to one specific system are listed in the corresponding section.

4.1.1 Classical molecular dynamics simulations

Generally, setups for MD simulations were prepared using the *psfgen* utility of VMD(133) together with CHARMM-GUI(134) to prepare initial conformations for most of the model membranes (exceptions are noted below). The modified TIP3P water model(135) was used for modeling solvent explicitly, since the charmm36(136) and charmm36m(137) forcefields, which were used in all cases, were parameterized with this specific water model in mind. Additional force field parameters for metal centres and cofactors were taken from our previous work(11), where we had obtained them based on DFT calculations at the B3LYP-D3/def2-SVP/TZVP(138–140) level, or, in the case of the NiFe active site of Mbh, had adapted them from existing literature.(141) ISCs were generally modeled in their oxidized states, except for N2, which was sometimes modeled in a reduced state (see below).

All classical simulations setups were built as boxes to enable periodic boundary conditions and treatment of long-range electrostatics by the Particle-Mesh-Ewald (PME) formalism(142) during the simulations.

Protonation states of titratable groups at pH 7 were estimated based on the PBE/MC formalism using the adaptive Poisson-Boltzmann-solver (APBS) in combination with Karlsberg+ (143–145) at an ion concentration of 100 mM KCl, a dielectric constant $\epsilon = 80$ and a probe radius of 1.4 Å for the solvent and a dielectric constant $\epsilon = 4$ for the protein, which was described as the point charges of its constituting atoms. If perturbations of

protonation states or point mutations were tested, they were introduced by decomposing the fully built system into segments, applying the perturbations to the respective segments and reassembling the system.

Simulations were performed with NAMD2 in the *NPT* ensemble at a temperature of 310 K and a pressure of 1 bar. These conditions were enforced using a Langevin thermostat(146) and a Nose-Hoover-Langevin barostat.(147) Long-range electrostatics were treated in Fourier space by applying the PME method(142) at a grid size of 1 Å with a Lennard-Jones cut-off of 12 Å. All bonds to a hydrogen atom were kept rigid with the ShakeH algorithm(148) to enable a timestep of 2 fs.

4.1.2 Quantum chemical density functional theory models

Starting structures for DFT models were generated by extracting coordinates from equilibrated classical MD simulations. Amino acids were usually cut at the $C\alpha$ - $C\beta$ bond and saturated by adding a single hydrogen atom. The position of $C\beta$ -atoms was fixed during structure optimization to mimic the protein framework and implicit solvation was included in the calculations via the conductor like screening model (COSMO)(149) at $\epsilon = 4$, appropriate for a model of the protein interior. The B3LYP-D3(138, 139) functional was used with a def2-SVP(150) basis set for geometry optimization (with def2-TZVP for all metals), while the basis was expanded to def2-TZVP(150) for all atoms to calculate single point energies. Zero-point vibrational energies and entropic effects at 310 K were calculated using the smaller basis set. Calculations were performed with TURBOMOLE in versions from 6.6 and 7.2 to 7.5 (see below).(151) A chain-of-state type optimization was used to predict minimum energy reaction pathways.(152, 153)

4.1.3 Hybrid QM/MM calculations

Starting structures for hybrid QM/MM simulations, which were performed using the CHARMM/TURBOMOLE interface,(154) were extracted from equilibrated MD simulations. The link atoms connecting the MM region and the QM region were introduced between the $C\alpha$ - $C\beta$ atoms of protein residues. While the QM region was simulated at the B3LYP-D3/def2-SVP level(138–140), the CHARMM36m force-field(137) was applied to the classical region with the respective in-house parameters described above. Simulations were performed at 310 K with a timestep of 1 fs.

4.1.4 Hydrogen-bonds graph analysis

H-bonding connectivity is central to the Grotthuss-type mechanism of proton transfer that is also present in the complex I family. (155, 156) Therefore, analysis of the H-bonding connectivity was performed for several members of this family, namely complex I from *E. coli* and *T. thermophilus* and Mbh from *P. furiosus*. The networkx(157) and mdtraj(158) python libraries were used to implement this algorithm. Briefly, side chains of polar residues and water molecules were considered as vertices in a weighted graph, while hydrogen bonds were considered as edges. Hydrogen bonds were considered an edge, if the distance between the two heavy atoms

involved in the H-bond was $< 3 \text{ \AA}$ and the angle formed by the two heavy atoms and the hydrogen atom in between was $> 120^\circ$. An artificial edge was introduced between the two nitrogen atoms in the sidechain of histidine since we considered histidine a whole residue as one ‘Grotthuss-entity’. The weight of each edge was set to the Euclidean distance between the two nodes it connected. Shortest path within the graph were calculated using Dijkstra’s algorithm.(159) Connectivity between two residues was defined as the fraction of frames in the MD trajectory, in which an unbroken path between them existed.

4.2 Complex I from *Thermus thermophilus*

4.2.1 Classical MD simulations

The crystal structure of the entire bacterial complex I from *T. thermophilus* (PDB ID 4HEA)(22) was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane, a well-established bacterial model membrane, solvated and adjusted with NaCl to an ion concentration of 150 mM. The position of the Q, which was modeled in a oxidized state, was taken from our previous work,(11) as was the model for the loop missing in the crystal structure of Nqo8. All ISCs were modeled in an oxidized state except for N2. 200 ns of classical MD simulations were performed to prepare starting conformations for free energy calculations.

The H292A_{NqoB} point mutation was introduced to the first frame of the fully hydrated state at an *rc* of -8 \AA (see section 4.2.2) and simulated for 100 ns, as was the wild type of the same frame with and without the proton transferred from K235_{NqoB} to E377_{NqoB}.

4.2.2 Classical free energy calculations

Classical free energies for the switching of E123_{NqoB} from an *intra*-subunit ion pair with K204_{NqoB} to an *inter*-subunit ion-pair with K345_{Nqo14} were calculated with replica exchange umbrella sampling (REUS) in different hydration states of the membrane domain. The reaction coordinate *rc* was defined as

$$rc = r_2 - r_1 = r(K345 - E123) - r(K256 - E123)$$

With *r* being the Euclidean distance between the centre of mass of the head group atoms of the respective amino acids, i.e., the amino group for lysine residues and the carboxy group for the glutamate residue. The reaction coordinate was scanned between -10 \AA (*intra*) and $+8 \text{ \AA}$ (*inter*). Starting conformations were taken from classical MD simulations after 100 ns (dry state) and 200 ns (medium hydrated state). For each of the two starting conformations, a 10 ns long steered MD simulation was performed with a harmonic bias with a force constant of $10 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ on the *rc*. From these trajectories, 20 starting structures evenly spacing the *rc* were extracted and equilibrated for an additional 5 ns before the start of the REUS. The free energy calculations themselves were performed until convergence of the free energy profile, which took ca. 36 ns per window at a harmonic potential with a force constant of $2.5 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ for each window and exchange attempts with neighboring windows every 10 ps. Additionally, a previously published free energy profile calculated for an even more

hydrated state ('fully hydrated state')(11), which was determined with only r_2 as the reaction coordinate, was re-mapped onto the r_c used here by re-calculating the r_c for each frame sampled in the calculation and sorting to the respective windows. This was possible because of a high correlation between the two reaction coordinates. The energy profiles were calculated from the trajectories using the python implementation of the dynamic histogram analysis method (pyDHAMed)(132), which is more robust than the classically used weighted histogram analysis method (WHAM)(131) at the edges of the profile. A bin width of 0.01 Å was applied for binning and rates were calculated from the free energy profiles with transition state theory at 310 K with a unity reflection coefficient and a standard pre-exponential factor of 0.16/ps.

4.2.3 Hybrid QM/MM calculations

For hybrid QM/MM calculations, structures were extracted from the classical MD simulations of the states described in section 4.2.1 and trimmed to only include subunits Nqo12, Nqo13 and Nqo14 as well as water, ions and lipid molecules in their direct vicinity. The QM region was made up of residues T232, K235, S291, H292, S318, Y321, T322, E377, S402 and Y405 in subunit Nqo13 as well as 10 surrounding water molecules. Before QM/MM dynamics were initiated, all atoms within 10 Å of the QM region were relaxed using the adopted Newton-Raphson algorithm to adapt the system to the new Hamiltonian. 7 ps of unbiased QM/MM-MD were performed for the fully hydrated state and the dry state with the K204-E123 ion-pair in an open and closed conformation each. QM/MM free energy profiles were obtained in the fully hydrated state with ion pair closed and open for the proton transfer from K235 to E377. This reaction was described by a coordinate constructed from the linear combination of all hydrogen bonds broken and formed during the process as extracted from the unbiased QM/MM dynamics. 58 evenly spaced windows along this reaction coordinate from -7 Å to +7 Å were simulated for 1.6 ps per profile with the respective value of the reaction coordinate enforced with a force constant of 100 kcal mol⁻¹ Å⁻¹. Profiles were calculated from the trajectories using WHAM.(131)

4.2.4 Quantum chemical DFT cluster model calculations

For DFT calculations, the QM region was made up of residues T232, K235, S291, H292/A292, S318, Y321, T322, E377, S402 and Y405 in subunit Nqo13 as well as 10/13 (for the WT/H292A_{Nqo13} systems respectively) surrounding water molecules extracted from the fully hydrated classical MD simulations. Potential energy profiles along the minimum energy path were calculated as described above.

Additionally, cluster models comprising residues K235, H292, E377 of subunit Nqo13 or the corresponding residues K216, H265 and K345 of subunit Nqo14 and in both cases three water molecules in between the residues. Here, the presence or absence of the spatially close lysine featured in the ion pair of each subunit (K204_{Nqo13} or K186_{Nqo14} respectively) was mimicked by a fixed α -amino group 10.5 Å from the proton-donor lysine, which was cut, saturated and fixed at the C δ -atom and not the C β -atom in contrast to the general protocol described above. Transition states and intermediates were optimized as described above.

4.2.5 Hydrogen bond graph analysis

The graph analysis algorithm was used to find all shortest paths within subunits NqoB of the WT and the H292A mutant. The dataset comprised 100 ns of MD simulations each. Shortest paths between K35 and E377 were extracted and analyzed with regard to their length and persistency.

4.3 Complex I from *Escherichia coli*

4.3.1 Classical MD simulations

The crystal structure of the membrane domain of complex I from *E. coli* (PDB ID 3RKO) was placed in a POPC membrane at an ionic strength of the solvated system of 150 mM NaCl. The following systems with point mutations were also created and simulated for 100 ns using the procedure described above: H322A_{NuoM}, H348A_{NuoM} and the H322A_{NuoM} + H348A_{NuoM} double mutant.

4.3.2 Quantum chemical DFT cluster model calculations

DFT cluster models were created for all the variants described in section 4.3.1. The DFT models comprised residues A260, L264, K265, T318, S319, S321, H/A322, M323, Q344, H/A348, S351, A352, T395, L396, N403, E407, L429, A432 and Y435 of subunit NuoM. L396 and L429 were not included in the model of the H322A variant, and L396 and A260 were not included in the model of the H322A/H348A variant, while 10 to 13 water molecules in contact with residues of the cluster model were included for all variants. Potential energies were calculated for optimized transition states and intermediates of the minimum energy reaction pathway obtained as described above.

4.3.3 Hydrogen bond graph analysis

The graph analysis algorithm was used to find all shortest paths within subunits NuoM for all variants described in section 4.3.2. The dataset comprised 100 ns of MD simulations each. Shortest paths between K365 and E407 were extracted and analyzed with regard to their length and persistency.

4.4 Mitochondrial complex I from *Mus musculus*

4.4.1 Classical MD simulations

The cryo-EM structure of the active state of mitochondrial complex I from *M. musculus* (PDB ID 6ZTQ)(59) was inserted into a 2:2:1 POPC:POPE:cardiolipin membrane, solvated and ionized with NaCl to an ion concentration of 150 mM NaCl. The piericidin molecules resolved in this structure were discarded but resolved lipids were retained in the model. The position of Q, which was modeled both in its oxidized and reduced state in the first as well as the second binding site, was taken from previous work.(9) Since this protein was too large to estimate

protonation states for the whole protein at the same time by our method of choice, we divided it into three overlapping parts.

To prepare the deactive-state model, for which no resolved structure of sufficient quality for MD studies existed, we drove the active-state structure (see above) to consistency with the cryo-EM density map of the deactive state (EMDB: 4356) using MDFF(160), an approach which converts the cryo-EM density map into an additional restraining potential. MDFF minimization was followed by 5 ns MDFF dynamics in implicit solvent, while harmonic restraints were used to preserve secondary structure elements.

Parameters for the ISCs and cofactors were taken from our previous work(59) and the N2-ISC was modeled in its reduced state in most simulations. The protein contains two unusual arginine residues: Arg87_{NDUFS7}, which was modeled in its γ -hydroxylated form, and Arg85_{NDUFS2}, which was modeled as a symmetric dimethylarginine. In total, 8.5 μ s of classical MD simulation were performed for the active and deactive state with Q in different binding sites and for complex I in its apo form.

4.4.2 Quantum chemical DFT cluster models

DFT cluster models were used to probe the proton transfer at the ND3/ND4/ND6 interface. To account for the large size of the region involved in pT, models were constructed in two parts: The first one contained Asp66_{ND3}, Glu34_{ND4L}, Tyr59_{ND6}, Phe62_{ND3}, Leu63_{ND3}, Leu67_{ND3}, Leu32_{ND6}, Leu57_{ND6}, Ile58_{ND6}, Met63_{ND6}, Leu30_{ND4L}, Leu31_{ND4L}, Val37_{ND4L}, Leu38_{ND4L}, Ala71_{ND4L}, Leu75_{ND4L}, Leu106_{ND2} and 8 water molecules, while the second consisted of Glu34_{ND4L}, Glu70_{ND4L}, Glu34_{ND2}, Lys105_{ND2}, Tyr59_{ND6}, Leu33_{ND4L}, Val37_{ND4L}, Ala67_{ND4L}, Ala71_{ND4L}, Leu37_{ND2}, Leu38_{ND2}, Gln63_{ND2}, Ala64_{ND2}, Ser67_{ND2}, Met68_{ND2} and 7 water molecules. After separate optimization of the models and the reaction path as described above, the models were combined to calculate the final energetics.

4.4.3 Hybrid QM/MM simulations

The same approach as for the DFT models was also utilized for QM/MM-models of the same pT reaction. However, the QM regions were slightly smaller than those in the cluster models, comprising Asp66_{ND3}, Glu34_{ND4L}, Tyr59_{ND6}, Phe62_{ND3}, Leu63_{ND3}, Leu32_{ND6}, Leu57_{ND6}, Ile58_{ND6}, Val37_{ND4L}, Ala71_{ND4L} and six water molecules, and Glu34_{ND4L}, Glu70_{ND4L}, Tyr59_{ND6}, Ala71_{ND4L}, Val37_{ND4L}, Ala67_{ND4L}, Met68_{ND2}, Ala64_{ND2}, Ser67_{ND2} and 3 water molecules respectively. The reactions were sampled independently for both transfer steps by QM/MM-US with 25 and 19 windows respectively with a sampling of 5 ps with a similar reaction coordinate as used for the *T. thermophilus* QM/MM-US and free energy profiles were calculated as described above.

Additionally, the proton transfer at the upper end of the E-funnel from Glu202_{ND1} to Glu 227_{ND1} via one water molecule was also probed by QM/MM-MD. The QM-region comprised Glu24_{ND1} (protonated), Arg195_{ND1}, Asp199_{ND1}, Glu202_{ND1}, Glu204_{ND1}, Ser209_{ND1}, Arg274_{ND1}, Phe224_{ND1}, Glu227_{ND1}, Tyr228_{ND1}, Arg279_{ND1}, Gln92_{NDUFS7}, Arg87_{NDUFS7}, QH₂/QH⁻ and 9 water molecules and was constructed from equilibrated MD simulations with Q in the second binding site in the usual way (see above). Proton transfer occurred within 2 ps of sampling.

4.5 Membrane bound hydrogenase from *Pyrococcus furiosus*

4.5.1 Classical MD simulations

The cryo-EM structure of membrane bound hydrogenase from *Pyrococcus furiosus* (PDB ID 6cfw)(100) was embedded in a 1-palmitoyl-2-palmitoeyl-sn-glycero-3-phosphoinositol (PYPI) membrane. This lipid type was chosen since phosphoinositols are abundant in the membrane of *P. furiosus*(161) but it should be noted that this is a bacterial lipid model, which differs from archaeal ones in terms of the linkage between lipid tails and headgroup (phosphoester linkage in bacteria versus phosphoether linkage in archaea) and the sterical properties of the tails themselves: the tails of the precise PI-lipids abundant in *P. furiosus* are methylated at every second carbon atom of the lipid tails.(161) The cryo-EM map showed weak density in the large cleft between subunits MbhH and MbhM, which was tentatively assigned to two lipids.(100) Therefore, the membrane was prepared in a way that two lipid molecules remained in that cleft, which was possible minor adjustment of the tail conformations using VMD's *molefactory* functionality.(133)

The ionic strength was adjusted to 250 mM NaCl and parameters for the metal centres were taken from existing literature.(141) In total, 8 μ s of MD simulation in different protonation states were performed. In six of these simulations, sodium ions were placed in putative binding sites and in one of the simulations, the TM3 of subunit MbhD was biased to be alpha-helical.

4.5.2 Quantum chemical DFT cluster models

DFT cluster models were used to model the reaction of the NiFe-cluster at the active site of Mbh. The cluster was in the NIC state, between the transfer of the second electron and the second proton to form the H₂ product. The model included the cluster with its four cysteine ligands (Cys68_{MbhL}, Cys71_{MbhL}, Cys374_{MbhL} and Cys377_{MbhL}), residues in direct contact (Glu21_{MbhL}, Asn36_{MbhL}, Ile70_{MbhL}, His75_{MbhL}, Arg320_{MbhL}, Thr345_{MbhL}, Asp372_{MbhL}, Ser376_{MbhL}) as well as three water molecules. The proton transfer from Glu21_{MbhL} via one water molecules to Cys374_{MbhL} was optimized as described above as was the subsequent flip of the proton around the CY374-Sy-atom and its transfer to the bound hydride to form H₂.

4.5.3 Water inlet cluster analysis

To trace possible proton and ion pathways, we traced the path of water molecules interacting with buried sites in Mbh via the Aquaduct program in version 1.0.11.(162) Our sites of interest were defined by their residues as follows: The possible proton donor residue at the active site (Glu21_{MbhL}), the members of the hydrophilic axis in the ALS MbhH (Lys409, His350, Lys354 and Lys256), the putative N-side sodium binding site (Asp35_{MbhB}), the putative P-side sodium binding site (Asap59_{MbhB}) and the O-site resolved in Mrp (Glu43_{MbhH}, Lys225_{MbhH}, Asp37_{MbhG}, His41_{MbhG}, Glu69_{MbhD}). For the active site and the hydrophilic axis, we tracked water molecules

within 4 Å of the defining residues, while for the sodium sites, solvent molecules within 6 Å were tracked. All water molecules entering these radii around the sites of interest were tracked and constituted the so-called object. The algorithm also recognizes when tracked molecules enter or exit the solvent exclusion volume of the protein, which we approximated by a convex hull around all residues involved in secondary structure elements ('scope'). Analysis was performed separately for each site and the scope was adjusted for the specific sites: For the sodium binding sites and the ALS, the scope was only taken as the solvent exclusion volume of subunits within the membrane, while for the active site, the convex hull was constructed around subunits MbhJ to MbhN plus the part of MbhI, which is homologous to Nqo7. Our dataset consisted of all performed MD simulations for the sodium sites but was reduced to a subset of simulations for the ALS and the active site, since they are far more hydrated, and good sampling can be achieved with a smaller dataset. The time-resolution of all MD data included in our analysis was 10 ps. The paths of all tracked water molecules were clustered according to the Barber algorithm with a cut-off of 1.4 Å according to the point, where they intersect the convex hull. These inlet clusters were then inspected visually with regard to how the paths continue within the protein behind the inlet cluster. If one inlet cluster corresponded to several distinct paths within the protein, it was subdivided by the balanced-iterative reducing and clustering using hierarchies (BIRCH) algorithm. This was only needed at very high hydration levels in the hydrophilic domain. Clusters with less than 10 paths were assigned to the outlier cluster. Path trimming with the autobarber algorithm was only performed for the sodium site analyses since the open cavities found for the ALS analysis and the active site analysis are not well treatable with this algorithm. The Q-site as a large, buried water filled cavity seems to be particularly problematic.

4.5.4 Hydrogen bond graph analysis

Here, graph analysis was used to determine, whether pT through the lipid cleft from MbhH to MbhM is feasible by calculating shortest paths between K409_{MbhM} and E141_{MbhH} with the latter both in a protonated and ionic state.

4.6 Artificial 4-helix-bundle constructs

4.6.1 Classical MD simulations

Three 4-helix bundles were designed and modeled computationally. For all three, simulations were performed in a water box at an ionic strength of 150 NaCl. For Maquette 2 and Maquette 3, experimental structures were determined by solution NMR spectroscopy (BMRB id 34518) and X-ray crystallography (PDB ID 6ZOC) respectively. The lowest-energy NMR structure and chain A in the unit cell of the crystal structure were each placed in the water box directly after prediction of protonation states. In the case of Maquette 3, the structure could only be solved by introducing a stabilizing element (PDB ID 5VJT). Therefore, the wild type was reconstructed from the stabilized variant by side chain replacement using the *mutate* functionality of psfgen.

Stabilizing mutants were tested for all three Maquette series using the same functionality. Maquette 1, for which no experimental structure could be resolved, was modeled in silico prior to MD simulations. Helix elements were predicted by PsiPred(163) for residues 1–27, 36–62, 71–97, and 106–130 and this prediction was used as basis for model building using established principles.(164, 165)

MD simulations on the μ s scale were performed in duplicates for all designed constructs to assess the stability of the structure as a whole and the buried ion-pair motifs in particular.

4.6.2 Computational design algorithm

We developed a computational design algorithm to predict stabilizing elements for motifs in proteins. The general idea is to introduce point mutations in the protein and equilibrate them by a sequence of short energy minimizations and MD simulations. After 5000 steps of initial equilibration, 10 ps of MD simulation are performed before the system is minimized for 5000 steps again. These simulations can be performed in explicit solvent, implicit solvent or even vacuum. During this process additional restraints on parts of the protein may be used to make the placement more efficient. It should be noted that the protein backbone is usually kept restrained during the whole process since the short duration of the simulations is only capable of capturing local effects. After equilibration, the new structure can be evaluated by arbitrary scoring functions. The algorithm was implemented in python in a way that new modules for scoring functions can be easily added without changing the core code. In this work, we focused on the simplest case of using energies calculated by a molecular dynamics single point after equilibration. We included a liner combination of three terms in our scoring function: The interaction between the residues constituting the motif of interest, the interaction between all other protein residues and the ones of the motif and the total potential energy of the protein as a measure of stability. The last term is particularly useful for filtering out clashes.

The score is then compared with the score of the last accepted structure (or the starting structure) and assessed according to a Metropolis-Monte-Carlo criterion:

$$e^{-\Delta score/k_B T} > R \in [0,1]$$

with Boltzmann's constant k_B and the temperature T (an input parameter for acceptance tuning). R is a random number within the interval 0 to 1. $\Delta score$ is the score difference between the last accepted structure and the candidate structure. While the score needs not necessarily be a physical energy (we used packing density as a measure, for instance), it may be useful to convert it to a measure leading to a decent acceptance rate (20 % acceptance seems to work well). Comparison with a random number allows the algorithm to accept substitutions, which make the structure worse, with a finite probability. This Monte-Carlo-tunnelling is useful for escaping local score minima. Since the score is usually based on protein structure, it tends to be equally degenerate. If the criterion is met, the candidate structure is accepted and used as basis for the next circle. If

not, the last accepted structure is used. The algorithm runs until either a certain number of tries has been performed or until a certain number of substitutions has been accepted. It is not advisable to go far beyond three accepted substitutions, since a high number of mutations is likely to alter the global structure of the protein, which is not captured by our algorithm and may lead to unreliable results.

5 Results and Discussion

In this section, we will give a brief overview of key results obtained in this work. The sections are grouped by the biological systems studied. The full articles can be found in appendix I.

5.1 Hydration dynamics in the membrane domain of bacterial complex I (Articles 1 & 2)

One of the most striking qualities of respiratory complex I is its complete reversibility and extremely low dissipation of energy while transducing chemical energy into an electrochemical gradient established by pumping protons. Not only do these features imply a high efficiency of the electron and proton transfer during the quinone reduction, but also the existence of conserved and highly effective coupling elements between these processes. Furthermore, complex I has been shown to pump four protons in four distinct subunits spanning nearly 200 Å.^(19, 20, 39) Three antiporter-like subunits (ALS) contain one of these proton paths each, while an additional proton pathway is thought to reside around the hinge region connecting membrane and hydrophilic domains. The internal symmetry of the ALS gives the proton paths a characteristic S-shape. (Figure 8C) with two vertical parts conducting protons to and from the hydrophilic axis and a horizontal part for lateral proton transfer.

Even before the first residue-resolving crystal structure of respiratory complex I had been solved (PDB IDs 3M9C/3M9S),⁽⁴⁶⁾ mutational studies showed that the conserved ion-pairs are central for proton pumping.^(51, 166, 167) The crystal structure revealed that these conserved ion-pairs are located at the interface between ALSs and form, together with residues implicated in proton transfer, a hydrophilic axis spanning the entire membrane domain. Each ion-pair faces the putative terminal proton acceptor of the preceding ALS, or in the case of subunit Nqo14 (*T. thermophilus* nomenclature will be used throughout this chapter) and its homologs, the proposed proton acceptor of the Q-site adjacent part of the membrane domain (see section 5.2 for details on this part of the complex I mechanism). This arrangement, together with the probably high desolvation penalty involved in burying this many charged and polar residues in the membrane plane, has sparked much interest in these ion-pairs as possible coupling elements, which could help regulate the timing/coupling between adjacent ALSs. Prior computational studies⁽¹¹⁾ have established that a switch from a ‘closed’ intra-subunit ion-pair to an ‘open’ ion-pair between the acidic residue of the ion-pair and the terminal proton acceptor of the preceding ALS is energetically feasible on the sub-microsecond timescale (Figure 9A).

The precise topology of the ion-pairs varies in complexity between subunits and organisms, from the plain lysine-glutamate ion pair found in subunits Nqo13 and Nqo14 to the more involved arrangement in Nqo12, which features a lysine-aspartate ion pair supplemented by an arginine-glutamate ion pair. However, all known ion pair arrangements feature a positively charged lysine residue in spatial vicinity to the conserved ‘middle lysine’

residue implicated as primary proton acceptor of the hydrophilic axis in each ALS. This close vicinity of two positively charged residues (in their standard protonation state at least) opens new venues of possible interaction and regulation of the proton transfer in the ALS.(11, 67)

In the publications presented here, we studied the dynamics of the conserved ion-pairs as well as their interplay with hydration and their effect on the proton transfer reactions in the ALS. We calculated (free) energy profiles for many of the key processes, using both classical and quantum chemical levels, and validated key predictions by biochemical assays in collaboration with the Friedrich group. Kinetic models and principal component analysis allowed us to tie the diverse processes together and develop a functional model of how proton transfer is affected and regulated in the ALS. For these studies, we focused on subunit Nqo13 (Figure 8A), the middle of the three ALSs, since this subunit has been characterized most thoroughly in the past and its position surrounded by other ALS could reduce boundary effects.(11, 51, 166, 167)

To establish, whether the positive charge of the lysine in the ion-pair actually influences the energetics of proton transfer in the ALS, we built DFT cluster models including the key residues implicated in the lateral part of the pT reaction in Nqo13, namely K235 (middle lysine), H292 (bridging histidine) and E377 (terminal acceptor) and intermediate water molecules. (Figure 10A) The lysine residue in the ion-pair was mimicked as a point charge and placed at a distance of 10 Å from the middle lysine, which is the mean of the distance distributions obtained from classical MD simulations for an open state of the ion-pair (Figure 9B). When modelling the lysine in its charged state (+1), i.e., mimicking an open state of the ion-pair, lowered the energies of all intermediates, transition states and products of the pT reaction by 2.4 to 3.2 kcal/mol, indicating that ion-pair opening facilitates the pT. This result could be reproduced for the homologous setup involving the corresponding residues in subunit Nqo14.

Having shown the effect of the ion-pair on the pT reaction, we looked at the effect of proton transfer on the hydration of the ALS. Since the proton transfer requires a stable hydrogen bonding connectivity between the residues involved, the hydration of the ALS could be an important regulatory element, allowing proton transfer by managing how many water molecules are present in the buried hydrophilic axis to connect key residues. We found in classical MD simulations that having a proton on the middle lysine, i.e., being in a possibly pT competent state, leads to hydration of the hydrophilic axis on the 100 ns timescale, whereas transferring the proton onward and thereby neutralizing the middle lysine leads to rapid drying within 50 ns. (Figure 8E, F) This effect is accompanied by a tilting of the conserved broken helix element in TM7 of Nqo13, in which the middle lysine is situated. The broken helix element, which is found twice in symmetry-related positions in all ALS is a common motif in biological transporters for enabling flexibility and regulating access to central sites. (Figure 8D)

Further PC analysis, in which we compared the correlation between the three ALS in different hydration states, once in a hydrated states and once in a drier state, revealed that the inter-residue correlation between subunits

strongly depends on the connection of the hydrophilic axis to the N-side of the membrane. Drying of these channels, e.g., observed after pT, lowers the correlation between subunits considerably pointing to a global regulatory effect of these local processes within each ALS (Figure 8B).

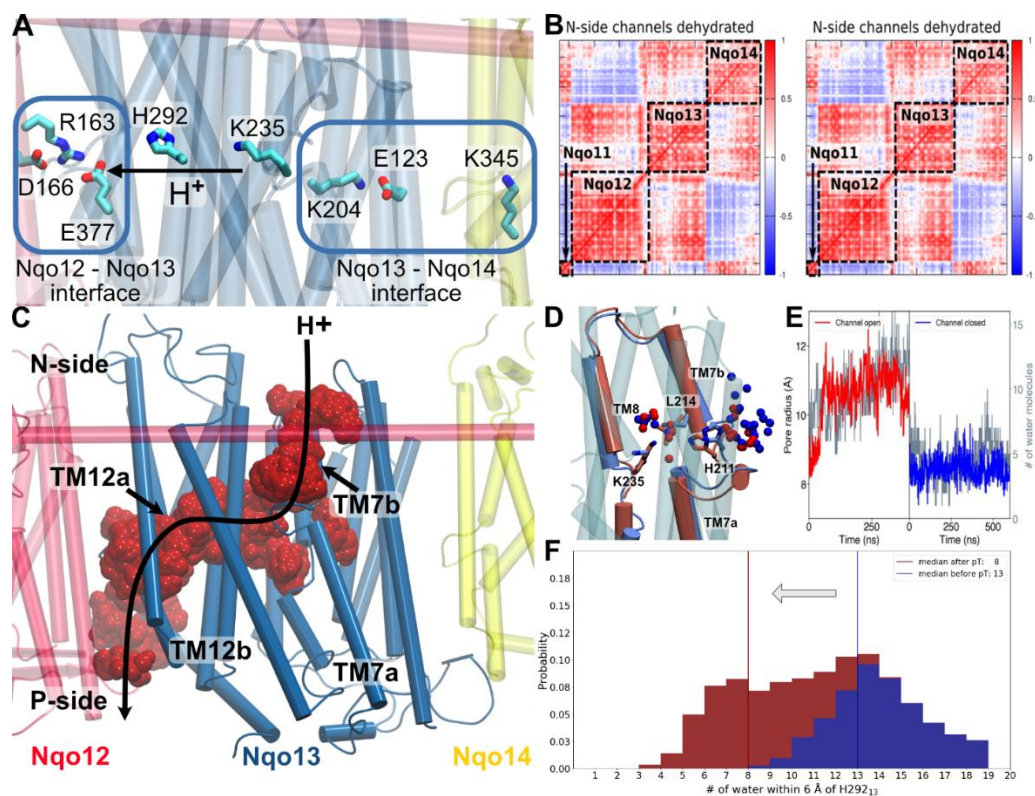


Figure 8: Hydration dynamics in subunit Nqo3 of complex I from *T. thermophilus*. **A)** The hydrophilic chain in the center of subunit Nqo3 can be divided into three overlapping parts: The residues involved in lateral proton transfer and the two interfaces to the preceding and succeeding subunits Nqo14 and Nqo12. **B)** Correlated motion between these subunits depends on the hydration state of the ALS from the N-side. **C)** Topology of the S-shaped water wire serving as a proton path in Nqo3 **D)** The N-terminal part of the conserved broken helix element on TM7 of Nqo3 serves as a gate, which regulates hydration of the subunit from the N-side. **E)** Opening and closing of that gate strongly correlates with subunit hydration **F)** Proton transfer within the subunit from K235 to E377 is coupled to the hydration dynamics. Panels adapted from articles I and II.

Having shown the interdependence of proton transfer and subunit hydration and their effect on crosstalk between subunits, we went on to establish the effect of hydration on the dynamics of the conserved ion-pairs. Since the coupling between subunits is affected by hydration changes and we hypothesized these ion-pairs to be central for the coupling between ALSs, we predicted them to be strongly affected by hydration dynamics as well. To this end, we calculated classical free energy profiles for changing the ion-pair conformation from ‘closed’ to ‘open’ under different levels of subunit hydration by REUS (Figure 9C). We found that, as predicted, subunit hydration strongly influences the energetic cost of opening the ion-pair: At very low hydration levels, a

free energy barrier of > 25 kcal/mol makes opening the ion pair unfeasible on biological timescales. Increased hydration levels abolish this barrier entirely and lower the energy of the open state relative to the open state by > 15 kcal/mol.

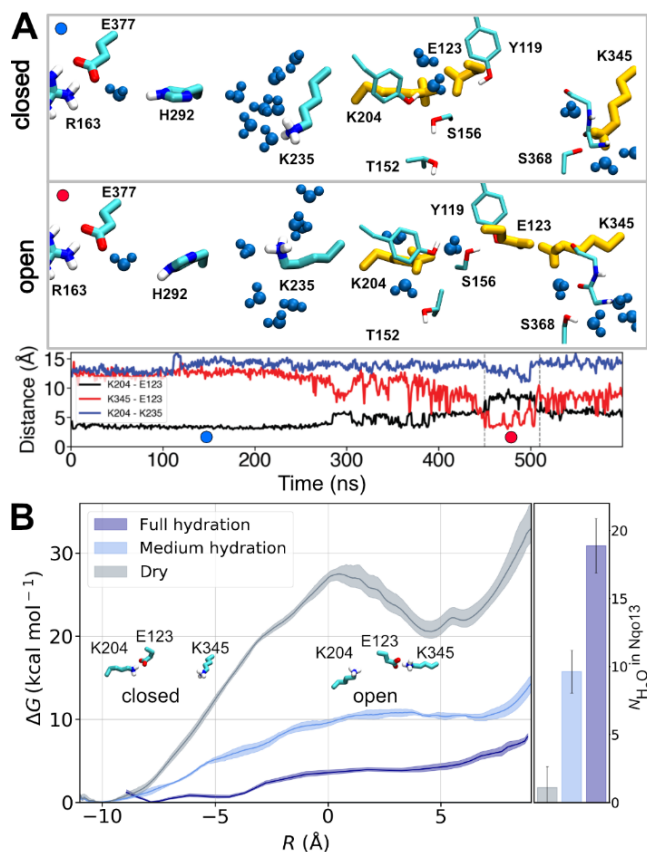


Figure 9: Dynamics of the conserved K204-E377 ion-pair at the interface between Nqo13 and Nqo14 in complex I from *T. thermophilus*. A) The buried charged residues of the hydrophilic axis are stabilized by a group of conserved residues and bound water molecules in both the closed state (top panel) and the open state (middle panel) of the interface ion-pair, whose dynamics can be observed in unbiased MD simulations at the hundreds of ns timescale (bottom panel). The states that the structural snapshots in the top panel belong to are marked in the traces in the bottom panel by circles. Note that opening brings K204 in closer contact with K235 involved in proton transfer B) Opening of the ion-pair is only possible under high hydration levels as indicated by free energies from REUS. Panels adapted from articles I and II.

Interestingly, it has previously been found in classical free energy calculations that the open state only shows as a minimum of the free energy profile if pT has already happened in the ALS(II), and our DFT cluster models corroborate the interdependence of pT and ion-pair dynamics. Therefore, we calculated a free energy profile by hybrid QM/MM umbrella sampling in a fully hydrated state with the ion-pair in both the open and closed conformation (Figure 10B, D). Our QM/MM free-energy calculations confirm the result we obtained earlier from cluster models that pT is made more feasible energetically by opening the ion-pair but the detailed free energy profile additionally revealed that the pT transfer is actually endergonic by ca. 4 kcal/mol, if the ion pair is closed, while opening renders it exergonic by ca. 2 kcal/mol. The barrier of the process is also reduced by more than a factor of two from ca. 10 kcal/mol to ca. 4 kcal/mol, but both these barriers could be mechanistically feasible on the turnover timescale of complex I. However, the change from an endergonic to an exergonic process could be mechanistically important. This change can also be observed in hybrid QM/MM MD simulations, which show the proton spontaneously progressing to the terminal acceptor within 2 ps if the

subunit is fully hydrated and the ion-pair open (Figure 10C, D). Closing the ion-pair stalls the reaction midway with the proton staying on the bridging histidine, while reducing hydration stops pT altogether on the probed timescales of 8 ps.

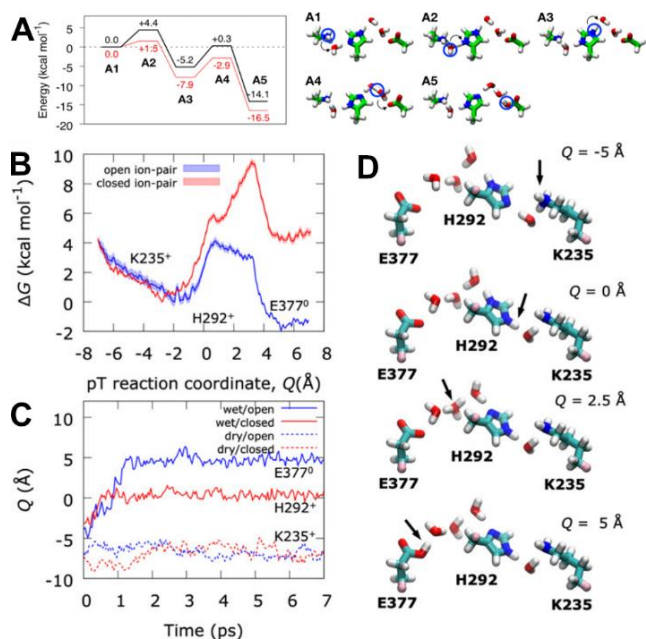


Figure 10: Energetics for the proton transfer from K235 to E377 in subunit Nqo13 of complex I from *T. thermophilus*. A) potential energies from QM-cluster models with the charge of K204 of the conserved ion-pair modeled as in the open position (red levels) or in the closed position (black levels). The reaction progress is depicted on the right. B) Free energies from hybrid QM/MM simulations of the same reaction show that the full reaction is only exergonic if the ion pair is open (blue curve) C) The same effect is observed in unbiased QM/MM-dynamics, where full hydration of the subunit and an open ion-pair make the reaction happen on the ps timescale. D) Progress along the reaction coordinate depicted in C) and D). Panel adapted from Articles I and II.

Our results, both from hybrid QM/MM calculations and cluster models highlighted the bridging histidine as an important intermediate facilitating protonic connectivity between the middle lysine and the terminal acceptor. Hence, we tested, both by *in vitro* and *in silico* mutagenesis experiments, how mutating this residue to an alanine, which as a small amino acid without side chain H-bonding activity, would abolish the intermediate without obstructing the quasi ID-channel between the middle lysine and the terminal glutamate. To enable the *in vitro* mutations and pertaining proton pumping experiments, we switched our organism from *T. thermophilus* to *E. coli*, which is much easier to grow and mutate in the laboratory. All subsequent *in silico* mutations were performed for both organisms to provide comparability. While NuoM, the *E. coli* homolog to Nqo13 is structurally extremely homologous, it features two bridging histidine residues instead of the single one found in *T. thermophilus* (Figure 11A). To account for this, mutation experiments were performed for both histidine residues being substituted by alanine separately as well as for the double substitution.

Proton pumping was monitored by both ACMA and oxonol VI quenching, while complex I oxidoreductase activity was monitored separately for all three mutant constructs (Figure 11B). All mutants had significantly reduced activity in comparison to the wild type, with the double mutant being more impaired than both single mutants, indicating that each of the two histidine residues could partially compensate for the absence of the other (Figure 11C). Our classical MD simulations provided a rationale for the experimental observations: In all

three cases, as well as in the homologous H292A mutation in *T. thermophilus*, substitution of the histidine led to considerably longer and more transient H-bonding wires between the middle lysine and the terminal acceptor, which are expected to make Grothuss-type transfer between the two more challenging (Figure IID).

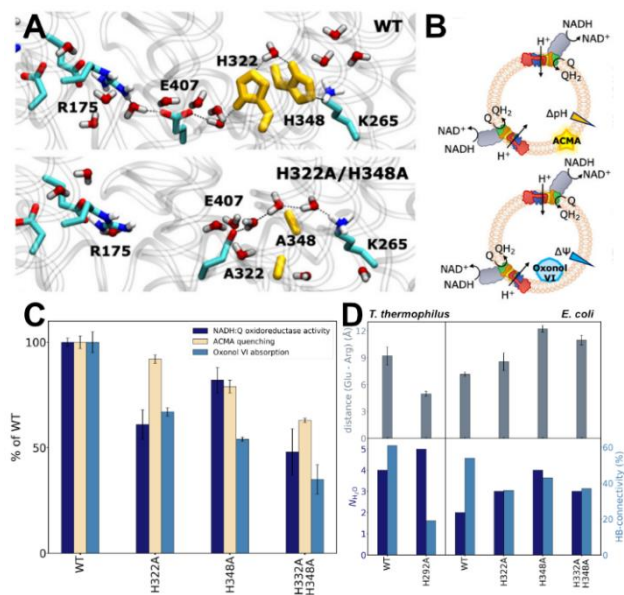


Figure II: Mutations of the bridging histidine(s) effect proton pumping both in *E. coli* and *T. thermophilus*. A) The lateral proton transfer chain in *E. coli* NuoM (homolog to Nqo13) features an additional histidine. Substituting the two histidines by alanine changes the water connectivity in the subunit. B) ACMA and oxonol VI can be used to monitor the $\Delta\Psi$ and ΔpH components of the membrane potential *in vitro* C) Replacing one or both histidines in *E. coli* not only affects proton pumping but also oxidoreductase affinity D) *In silico* experiments show a possible explanation by increased distances to the next subunit and longer water wires along the proton path. Figure adapted from Article II.

Hybrid QM/MM calculations showed that pT does not occur on the same timescale as for the WT, even under full hydration with the ion-pair opened.

In summary, we established the interplay between hydration levels in the ALS, dynamics of the conserved interface ion-pairs, which act as a coupling element, and pT within the ALS by deriving multi-scale free energy profiles and identifying important intermediates

5.2 Hydration dynamics in the membrane domain of mammalian complex I (Article 3)

Whereas the articles in section 5.1 focused on the coupling between ALSs, this work concerns the initial coupling of proton transfer to Q reduction. Complex I in higher organisms can switch into a so-called deactive state, a resting state that can regain activity after a longer reactivation period. This form has been found *in vitro* under lack-of-substrate conditions. The molecular-level details of the deactivation process, have nevertheless remained unclear. While its biological function remains debated, current thinking links it to a prevention of reverse electron transfer upon recovery from ischemic conditions and resulting tissue damages by uncontrolled

release of electrons. Recently, the structure of complex I from *M. musculus* has been resolved in putative active and deactive states.(45, 65)

While the active state structure at 3.3 Å resolution (PDB ID 6G2J) was suitable as a starting structure for MD simulations, the deactive structure at 3.9 Å resolution (PDB ID 6G72) was challenging not only because of the comparatively low resolution but also because of several poorly or unresolved regions. This was not unexpected since it has already been found for the bovine form of complex I that deactivation seems to be associated with increasing disorder of the three conserved loops flanking the first Q binding site. To address these issues and create a suitable model of deactive complex I, we created a refined deactive state model by targeting the active state structure to the density of the deactive state using MDFF (Figure 12A). Our simulations reproduced an intriguing feature of the experimental *M. musculus* deactive structure: TM3 of ND6 (Nqo10 in *T. thermophilus* nomenclature), which adopts a fully α -helical conformation in the active state, changes to a π -bulged structure in the deactive state. This change is accompanied by a rotation of the helix's top part.

To probe the influence of these changes on Q binding and proton transfer from the Q-sites towards the first ALS (ND2, Nqo14 in *T. thermophilus* nomenclature), we accumulated about 15 μ s of classical MD simulations of both states supplemented by QM-calculations and hybrid QM/MM calculations based on structures from the classical simulations. Our simulations revealed water connectivity between the Q-binding regions down the E-funnel and further towards the conserved ion-pair of the first ALS in the active state. However, in the deactive state, the rotation of TM3a in ND6 rotates residue Met63_{ND6} straight into the path of the water wire, thereby blocking connectivity (Figure 12B). This sterical constriction of the water pathway (Figure 12C) would not allow for proton transfer according to a Grotthuss mechanism and would thereby potentially explain the lack of proton pumping activity by complex I in the deactive state. In the active state, proton transfer along this pathway modulates the water wire itself through the interaction between water molecules and the charge shift during pt. Modeling different protonation states along the path with classical MD simulations revealed that while the water wire is also not fully continuous in the active state, the proton shift leads to every path of the tunnel being hydrated at one time during the reaction (Figure 12D). Introducing a charge in the dry part of the pathway in the deactive state does not pull water in as observed in the active state, suggesting that the steric seal of the tunnel is indeed tight.

Both, hybrid QM/MM and QM cluster model calculations show that the pT reaction through this gating region (i.e. from Asp66_{ND3} to Glu70_{ND4L}, see Figure 12E) is feasible in the active state on the timescales that complex I operates on with an overall barrier of about 15 kcal/mol. The entire reaction is endergonic by about 5 kcal/mol. Free energies from QM/MM-US and potential energies from optimized DFT-models are quite similar for most of the reaction and in reasonable agreement with higher levels of theory. As already observed for proton transfer within ALSs (see section 5.1), the conserved ion-pair in ND2 (Lys105_{ND2} / Glu34_{ND2}) modulates the barrier of the

final step of the proton transfer from Glu34^{ND4L} to Glu70^{ND4L}, while the first part-reaction from Asp66^{ND3} to Glu34^{ND4L} is unaffected, which corroborates the importance of the ion-pairs as regulatory elements (Figure 12F).

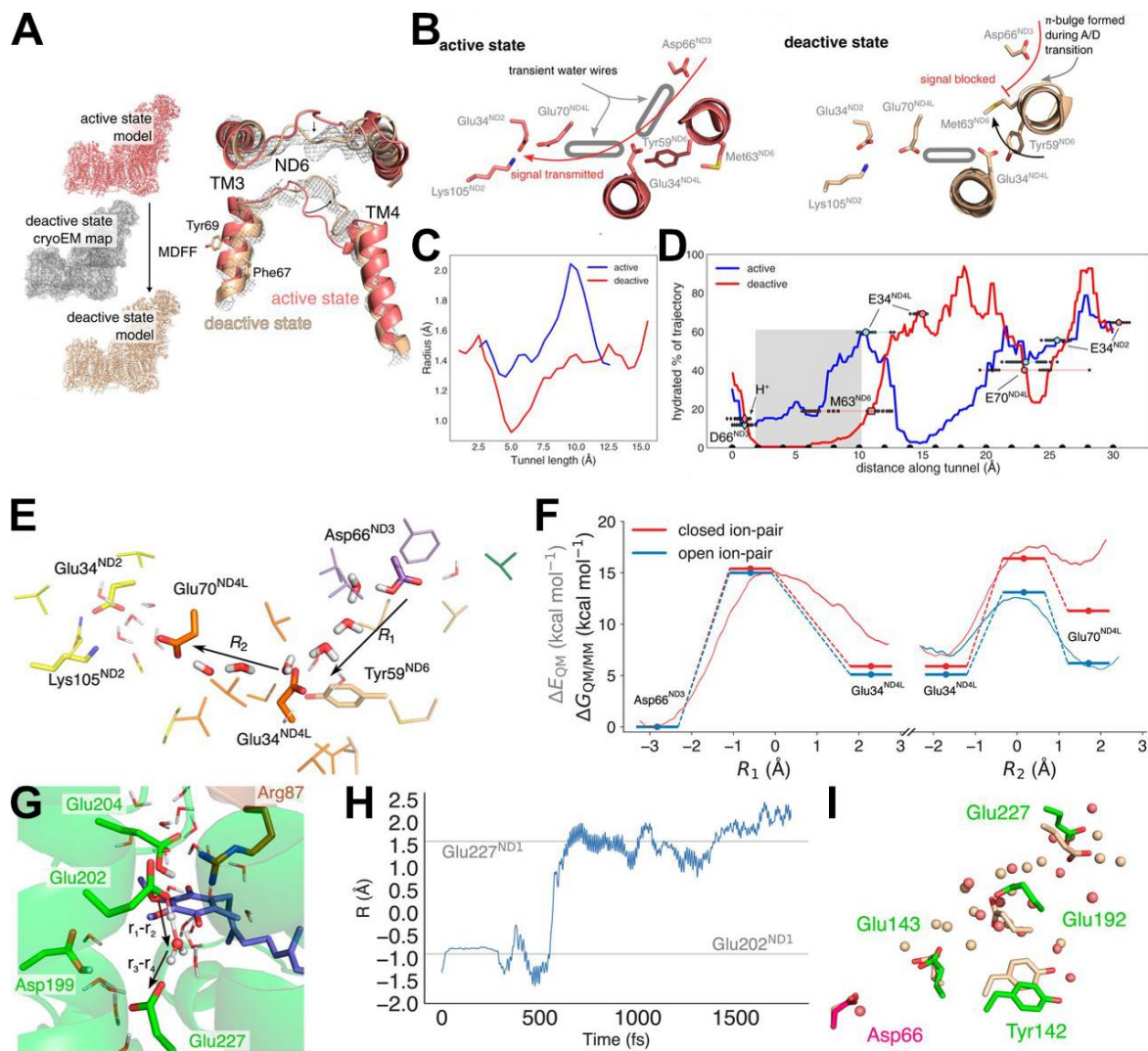


Figure 12: Deactivation of complex I from *M. musculus*. **A**) The deactive state, which was prepared by targeting the resolved active state structure to the deactive state density by MDFF, shows a pi-bulge in TM3 of ND6. **B**) Formation of the bulge and accompanying rotation of TM3 moves M63 into the course of the water wire following the hydrophilic axis and abolishes connectivity. This can be seen from **C**) decrease of the tunnel radius and **D**) breakdown of hydration at that position. **E**) The water wire, interrupted in the deactive state, facilitates proton transfer in the active state. **F**) This occurs at physiological timescales, which is affected by the conformation of the conserved ion-pair in the ALS ND2, as probed by QM and QM/MM calculations **G**) Protons enter the water wire from the E-funnel. The feasibility of injecting a proton into the E-funnel, while Q is bound in its 2nd binding site is shown by QM/MM-MD. **H**) Proton transfer from E202^{ND1} to E227^{ND1} is realized on the 0.5 ps timescale. The reaction coordinate used to describe this transfer is depicted in panel **G**). **I**) Hydration in this region also decreases during deactivation (red spheres) compared to the active state (tan spheres). Figure adapted from article III.

For the lateral proton transfer towards the first ALS described in the paragraph above to take place, a proton would have to be injected into the E-funnel and be transferred down the funnel from Glu202_{ND1} to Glu227_{ND1} via several other acidic residues (hence the name E-funnel) to Asp66_{ND3}. To ascertain the feasibility of such a process and probe its interaction with Q-dynamics, we performed QM/MM-MD simulations of the pT from Glu202_{ND1} to Glu227_{ND1} with Q bound in the second binding site, which is directly adjacent to both residues (Figure 12G). Our simulations show transfer via one bridging water molecule within less than one ps, highlighting that injecting a proton into the channel from the second Q binding site seems realistic (Figure 12H). However, this pathway is also abolished in the deactive state, where the region remains dry (Figure 12I). The dynamics of the Q itself and the aforementioned dynamic loops enclosing the first Q binding site are also affected by the active-to-deactive transition in our simulations. Our active state simulations show a more ordered conformation of the loops than the deactive state simulations and reproduce a partial alpha-helical conformation in the TM5-TM6 loop of ND1, which was previously found experimentally. This increased conformational dynamics might be behind the partial dissociation from the membrane facing binding site (2nd binding site) we find for Q in our deactive state simulations.

In summary, we characterized hydration dynamics and proton transfer from Q through an important gating region and provided a possible mechanism for the lack of proton transfer activity of complex I in its deactive state.

5.3 Functional dynamics of a membrane bound hydrogenase – the evolution of the complex I family (Article 5)

The recent cryo-EM structure of the membrane bound hydrogenase (Mbh) from *P. furiosus* (PDB ID 6CFW) added a new piece to the puzzle that the mechanism of the complex I superfamily remains. Mbh is an ancient precursor of modern respiratory enzymes and shares many of their characteristics. It combines a central ALS with a putative sodium-transport module and a redox module closely homologous to the redox machinery of complex I. However, several key differences between Mbh and complex I can be observed: The ALS (homolog to Nqo14) is rotated by 180° in the membrane plane, so the conserved ion pair is not facing the redox module but the sodium transport module. Furthermore, the ALS is separated from the redox module by a conspicuous cleft in the structure, for which the electron density shows weak density, which is consistent with lipids binding to the cleft. As a hydrogenase, Mbh does not use Q as a substrate but combines two protons with two electrons to form hydrogen. Incidentally, this is also a key difference to canonical NiFe-hydrogenases, which usually catalyze the reverse reaction. This explains the most striking feature of the Mbh structure: The Ni-Fe-cluster, which acts as its active site and replaces the N2 FeS-cluster found in complex I. However, the surrounding motifs of the active site are near-identical to the complex I setup. The three ISC ferrying electrons to the active sites can be found in complex I as can be the four conserved loops, which have also been mentioned in section 5.2.

Even the Q-cavity is already present in Mbh, although Q is no known substrate of the enzyme, but its entry to the membrane is sealed by bulky aliphatic residues. These tantalizingly similar dissimilarities to complex I led us to probe Mbh by classical MD simulations in combination with DFT cluster models to find mechanistic features hidden in their evolutionary relation. Again, we focused on hydration dynamics as an indicator for possible proton/ion pathways.

We modeled the active site in the NiC state with a hydride bound between Ni(III) and Fe(II). For the next step of the reaction to take place, another electron and a second proton must be transported to the reaction center. Since the path of the electrons is quite well defined by the chain of ISCs, we focused on potential proton paths. Our simulations showed that the conserved Glu21_{MbhL} residue forms a contact, which is mediated by a single water molecule with the S γ -atom of Cys374_{MbhL}, one of the four cysteine residues ligating the active site, if the acidic residue is modeled in a protonated state (as predicted by PBE calculations, Figure 13A). However, if modeled in a charged state, the residue and the conserved β 1 β 2-loop of MbhL (NDUFS2 in *M. musculus* complex I, see above) on which it resides, flip away from the active site by about 10 Å (Figure 13D). The water bridge between the cysteine and the glutamate residues is remarkably stable with a median occurrence of > 25% of time in all simulations, which can go up to > 80% under certain protonation states (notably a protonated state of His75_{MbhL}, which forms an H-bond with

Cys71_{MbhL}, another of the cysteine residues ligating the active site). This fits well with the heterolytic mechanism of hydrogen cleavage by canonical NiFe-hydrogenases, where the proton transfer between precisely these two residues has been postulated as an important mechanistic step. To probe, whether this mechanism is also applicable to the reverse reaction in Mbh, we built QM-cluster models based on structures from our classical simulations and followed the proton transfer from Glu21_{MbhL} to Cys374_{MbhL} and onwards to the hydride. Protonation of the cysteine is followed by a rotation of the proton around the residue to face the bound hydride. Injecting the second electron to achieve the Ni(II)/Fe(II) state allows further pT to the hydride and formation of the H₂ product (Figure 13B).

Having thus established Glu21_{MbhL} as a feasible proton donor for H₂ production in Mbh, we traced the paths via which the proton could be transported to the active site by following the path of every single water molecule coming near Glu21_{MbhL} during our MD simulations (at a 10 ps time resolution) and determining inlets on the protein surface by clustering the water paths. We found four distinct water clusters frequented by several thousand water molecules in total per simulation, which are well reproduced in independent simulations, although the extent to which the four different clusters are utilized vary over the simulations (Figure 13C). Two of the four inlets at the interface between MbhL and MbhM have been described in studies on complex I before, while two novel pathways were revealed: One originating in the cleft region between MbhM and MbhH, which follows the E-funnel upwards to the active site and one around the long loop in MbhI (i.e. the TM1-TM2 loop in ND3 of complex I from *M. musculus*). This last path is the only one of the three, whose location can shift slightly

based on the dynamics of the long TM1-TM2 loop of MbhI, but it is most often found at the interface of MbhI, MbhK and MbhL. Judging from the median contribution over all simulations, the four inlet clusters roughly are of equal importance for water access to the active site and in extension also for proton access to the site.

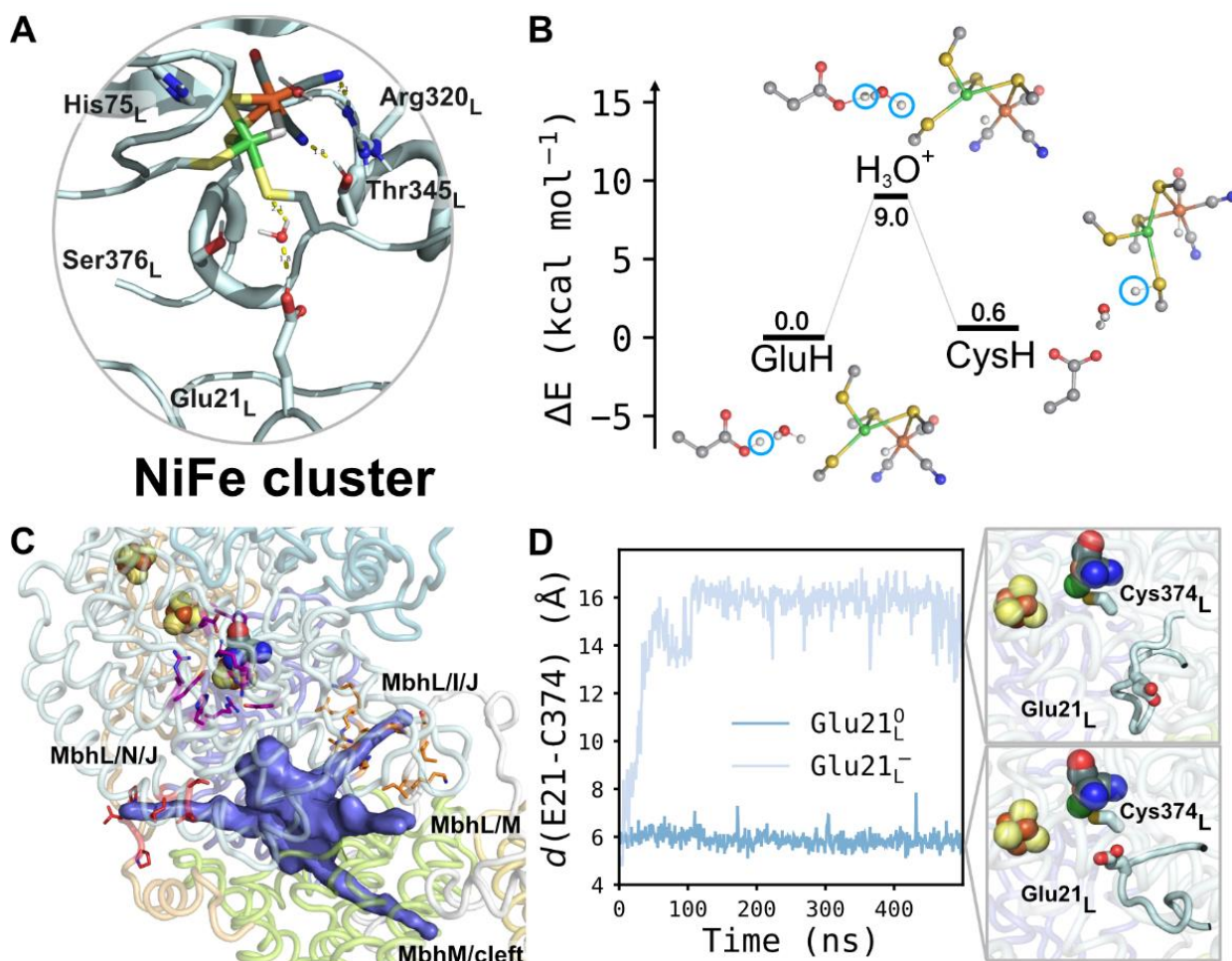


Figure 13: The active site of Mbh from *P. furiosus* catalyzes hydrogen formation from electrons and protons. **A)** The conserved Glu21_{MbhL}, a possible proton donor, is bridged to the H-NiFe-cluster (NIC state) via a water. **B)** QM-cluster models at DFT level show proton transfer from Glu21_{MbhL} to the active site is feasible **C)** Water pathways towards the cluster of conserved loops at the center of the hydrophilic domain show possible proton pathways to Glu21_{MbhL} on the conserved β 1 β 2-loop of MbhL. **D)** proton transfer from Glu21_{MbhL} to the active side leads to a flip of that conserved loop, which could be a possible coupling mechanism.

The cleft region is not only important for water access to the active site. We were able to shed some light on the dynamics of this region by modeling the two lipid residues predicted by the cryo-EM density map into the cleft region. We chose bacterial phosphoester PI lipids instead of archaeal phosphoether PI lipids, since there currently is no well-established model for the latter. Developing such a model and implementing it for further

studies of Mbh would be of significant benefit for understanding this protein. However, the bacterial lipids, which apart from the chemical linkage between headgroup and tail, only differ in the lack of tail methylation and slightly in tail length, fit into the cleft nicely and do not lead to significant rearrangements of the protein. They restrict water access to the cleft region from the P-side of the membrane but allow for water access from the N-side through a gap between MbhH, MbhM and MbhI.

The hydrophilic axis in Mbh's ALS features the well-known structural motif common to the complex I family: A middle lysine K256_{MbhH} is connected by bridging histidines to a putative terminal acceptor, K409_{MbhH}, to the one side and flanked by a conserved ion pair on the other side. Here, however, the ion-pairs acidic residue does not point to another putative proton acceptor but to a possible sodium binding site, which has been resolved in the cryoEM structure of the homologous Mrp from *Anoxybacillus flavithermus* (PDB ID 6Z16). Thereby, the ion-pair could maintain its role as a regulatory/coupling element in Mbh not by coupling two ALSs but by linking the proton transport within the ALS to sodium transport processes.

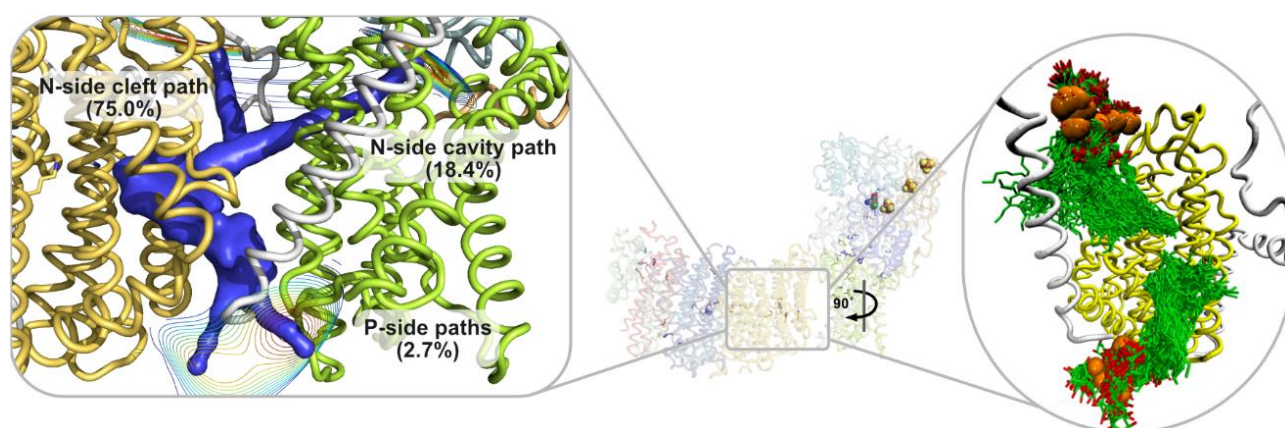


Figure 14: Water connectivity and lipid binding to the cleft region. The hydrophilic axis in the ALS MbhH (yellow) is predominantly hydrated from the cleft region (left inset). Secondary water paths lead down the E-funnel equivalent in MbhM (green) and via the broken helix element at TM12 of MbhH from the P-side. Water accessibility is regulated by lipids binding from the N-side and the P-side to the cleft region (right inset).

Although the structure of Mbh is quite recent, several mechanistic suggestions for sodium transfer have already been made based on the enzyme's structure and the recently resolved structures of the related enzymes Mrp and Mbs. In nearly all our simulations, we observe sodium binding to Mbh from the P-side of the membrane to a shallowly buried cavity around Asp59_{MbhB}, corroborating the hypothesis by Yu et al based on the Mbh structure (Figure 15A). While we do not find sodium binding from the bulk to the proposed binding site from the N-side around Asp35_{MbhB} (Figure 15A), hydration path cluster analysis, as performed for the active site, reveals that there is a strong water flow from the N-side through the proposed N-side binding site (N-site) to the

experimentally found occluded site (O-site) around His41_{MbhG}, pointing to a functional connection between the two sites (Figure 15A, B, C). No similar flow is found between the P-side binding site and the occluded site.

The water connectivity between N-site and O-site leads between TM2 and TM3 of subunit MbhD right beside the conserved Glu69_{MbhD}. It should be noted that MbhD is a structural homolog of ND6 in complex I from *M. musculus*, which shows the α -helix to π -bulge change during the active-to-deactive transition (see section 5.2). In Mbh, TM3 has been resolved in a π -bulged form, while an α -helical form has been found for the homologous subunit of Mrp. We hypothesized that a structurally similar transition to the active-to-deactive transition could be relevant for the sodium pumping processes in Mrp and Mbh. Therefore, we prepared an alpha-helical form of the Mbh structure via steered molecular dynamics (SMD, Figure 15A, right inset). The helical structure is completely stable on the half-microsecond timescales assessed here and leads to a pronounced increase in hydration of the O-site, indicating that this transition could be related to higher exchange between the two sites (Figure 15B).

To probe exchange between O-site and N-site directly, we prepared structures where sodium ions have been placed a) in the O-site as resolved in Mrp (two sodium ions) b) in the N-site directly as Asp35_{MbhB} as predicted from the Mbh structure and c) right in the middle of the O-site and N-site, i.e. in the geometrical center of the twelve-helix bundle comprising the core of the sodium transfer domain. Sodium transfer could be observed within half a microsecond from the center of the symmetrical element to the O-site. Interaction with Glu69_{MbhD}, which can reach in between TM2 and TM3 of subunit MbhD, seems crucial for this transfer through the proposed MbhD gate. Transport from the O-site to the N-site does not happen under predicted protonation states, but could be stimulated within 50 ns if an additional proton were injected into the site (Figure 15D). Proton transfer capabilities for this site has been proposed by Yu *et al.* before. Our water inlet cluster analysis suggests that a potential proton path could exist between subunits MbhG and MbhD from the N-side of the membrane. A cluster of conserved lysine residues at the start of this inlet makes sodium transfer unlikely, but proton transfer through a lysine cluster is a well-established functional motif in the complex I family.

In summary, we establish the first computational model for the membrane bound hydrogenase of *P. furiosus*, which enabled us to find proton paths, both previously described for complex I and novel, sodium binding sites and pathways and suggest a mechanistic model for the active NiFe-site. Additionally, we corroborated lipid-binding to the cleft between redox and transport modules.

5.4 Design of an artificial protein with an ion-pair – towards a *de novo* model system for complex I (Article 4)

Our work on members of the complex I family showed the importance of buried ion-pairs as regulatory elements of coupling and transport processes. However, the intricate nature of these giant biological machines makes direct investigation of these elements computationally expensive and technically challenging. Therefore, we

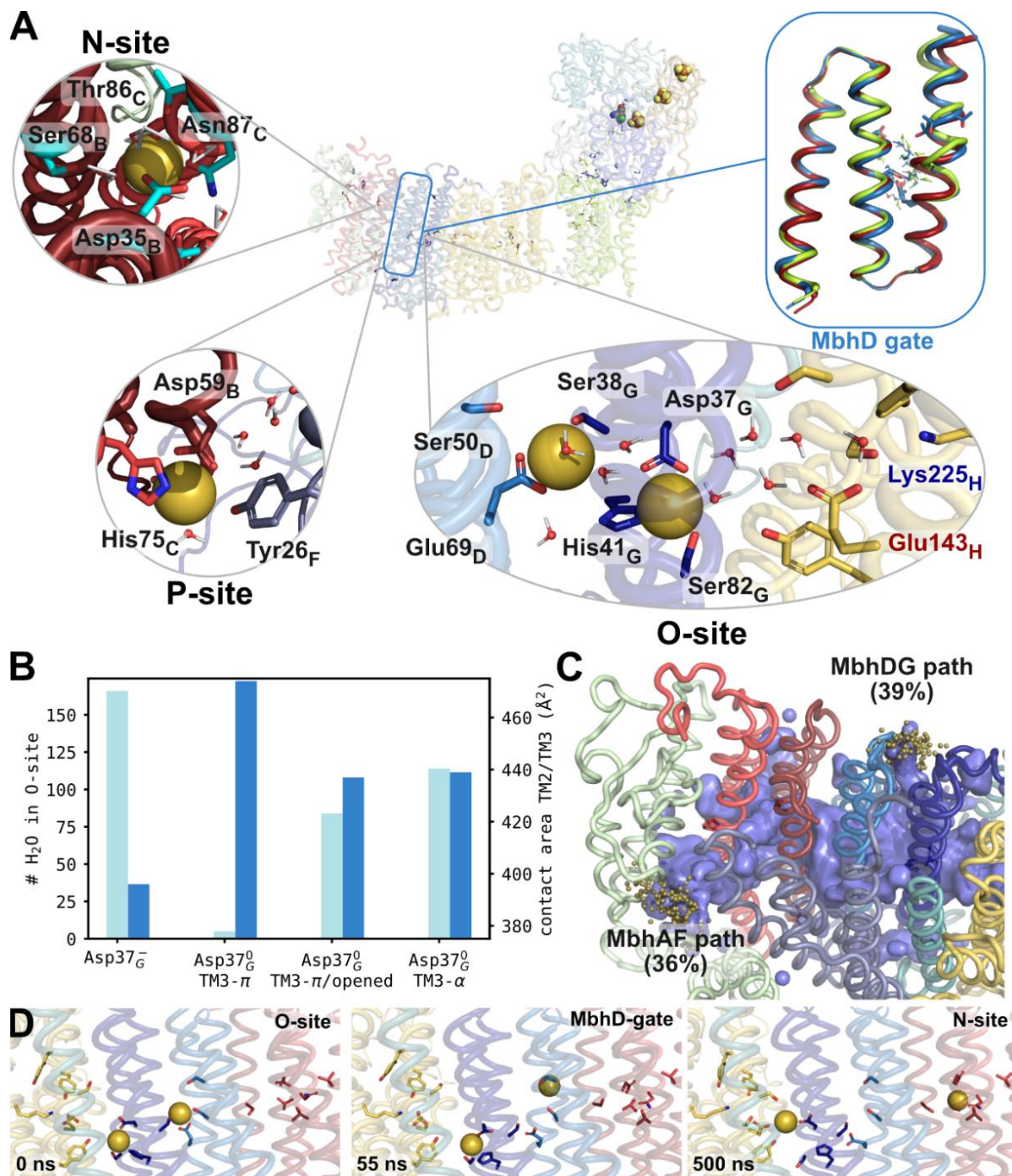


Figure 15: Ion transport in the putative sodium transport module. **A)** Three putative sodium binding sites in Mbh. N-site and P-site have been suggested by Yu et al.(100) from the cryo-EM structure. Spontaneous sodium binding to the P-site has been observed in nearly all MD-simulations. The N-site is open to hydration and a sodium can stably bind there, if placed. The same holds for two sodium ions in the O-site, which have been experimentally resolved in a structure of the homologous Mrp.(101) **B)** Hydration of the O-site occurs to a large degree from MbhA via the so-called MbhD-gate (inset panel A), which shows similar features to ND6 of complex I implicated in the active/deactive-transition. The hydration pathways are depicted in panel C). **D)** Injecting protons into the O-site leads to sodium transport to the N-site.

decided to design a *de novo* artificial protein featuring buried ion-pairs, to investigate such a molecular switch in a model system. As basis for our designed protein, we chose a four-helix bundle topology, which is a well-established and well-described scaffold in protein design. Knowledge of the geometric constraints on coiled-coil motifs as basis for four-helix bundle proteins makes the *novo design* feasible and reduces the sequence space considerably. As a basis for our computational design, we chose a previously suggested sequence suggested to form a four-helix bundle, although no structure had been resolved, yet. We predicted a 3D-structure of the sequence using established modeling methods, based on secondary structure predictions, and refined the structure with molecular dynamics simulations. The resulting artificial protein was dubbed Maquette 1. Biochemical characterization by circular dichroism (CD)-spectroscopy revealed a strongly alpha-helical fold, as predicted from *in-silico* design and high thermostability as well as stability in urea-unfolding assays. Introduction of the ion-pair significantly lowered melting temperature and the GdnHCl-concentration leading to unfolding, destabilizing the protein by ca. 5 kcal/mol. Different topologies, achieved by varying the loops connecting the 4 helices of the bundle showed comparable properties.

Although both *in vitro*-characterization and MD simulations of up to 1 μ s runtime indicated a high stability of Maquette with and without the buried ion-pair (which remained formed in MD simulations), we were unable to solve its structure either by x-ray crystallography or NMR-spectroscopy.

Therefore, we switched scaffolds from the Maquette 1 sequence, to the one of a recently designed porphyrin-binding four-helix bundle, which could be resolved by NMR-spectroscopy in both its apo- (PDB ID 5TGW) and its holo-state (PDB ID 5TGY).(109) Maquette 2 was modeled based on the structure of the apo-state of the protein, which shows an interesting design feature: The upper half was designed to be more polar and solvent-exposed in its core to be able to bind its ligand, while the lower half was designed as a proper hydrophobic core to provide stability. We were able to resolve a structural ensemble of Marquette 2 with an introduced ion-pair at positions Glu17/Lys72 (Figure 18A), a position we chose based on sterical considerations and since it showed great stability in MD simulations (Figure 16B).

However, the NMR-ensemble showed a mixture of open and closed conformations of the ion-pair, which had not been predicted computationally, and the protein stability decreased by ca. 5 kcal/mol (Figure 16C, D).

To address this issue and stabilize our introduced motif, we predicted stabilizing mutations with our newly developed design algorithm (see methods and Figure 17A). Furthermore, we conducted a survey of stabilizing elements for buried ion-pairs in membrane proteins, whose structure had been deposited in the PDB.

Our design algorithm was able to predict a plethora of stabilizing substitutions within few iterations (Figure 17B) and the distribution of predicted substitutions (Figure 17C) agrees relatively well with the one found in natural proteins (Figure 17D). Both predict enrichment of charged and polar residues to lower the desolvation penalty of the buried charged motif. Based on these results, we introduced Asn69 as a stabilizing element, which

resulted in both slightly better thermostability and a higher proportion of closed ion-pair conformations as assessed by NH₃-selective HISQC-NMR(168, 169) experiments.

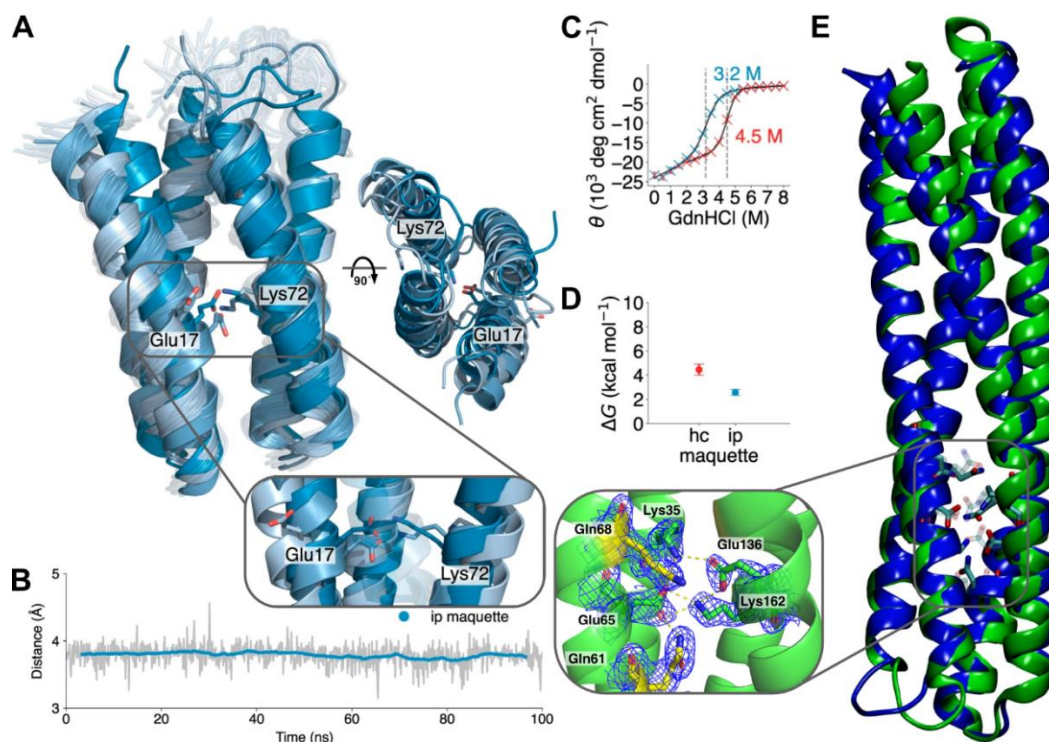


Figure 16: Designed four helix bundles with buried ion-pairs, which could be resolved experimentally. A) Maquette 2 was resolved by NMR-spectroscopy. The ensemble contained the buried ion-pair in both an open and closed conformation (see inset) B) During MD simulations of the designed construct, the ion pair stayed closed. C) The designed construct shows high stability against denaturing agents. D) Introduction of the ion pair costs about 5 kcal/mol. E) In Maquette 3, which was resolved by X-ray-crystallography to a resolution of 1.85 Å, we stabilized the ion-pair by adding a second ion-pair and polar residues as suggested by our design predictions. Figure adapted from article IV.

Guided by the prediction that addition of further ion-pairs to the one we inserted could increase the stability, we tried to build a buried charged network. However, we judged that the hydrophobic core of Maquette 2 was not extensive enough to support the necessary number of mutations. Therefore, we used a significantly longer Zn²⁺ binding four-helix bundle scaffold (PDB ID 5VJT) to introduce two ion-pairs flanked by two additional Gln residues. We guided the design of this network by computational modeling, which indicated good stability and feasibility of the designed conformation. Notwithstanding this large subtraction from the hydrophobic core of the protein, we were able to resolve a crystal structure of the construct with the buried charged network to a resolution of 1.85 Å, demonstrating the feasibility of extensive charged networks similar to those deployed by respiratory enzymes even in small proteins, if a sufficient polar stabilizing frame is built.

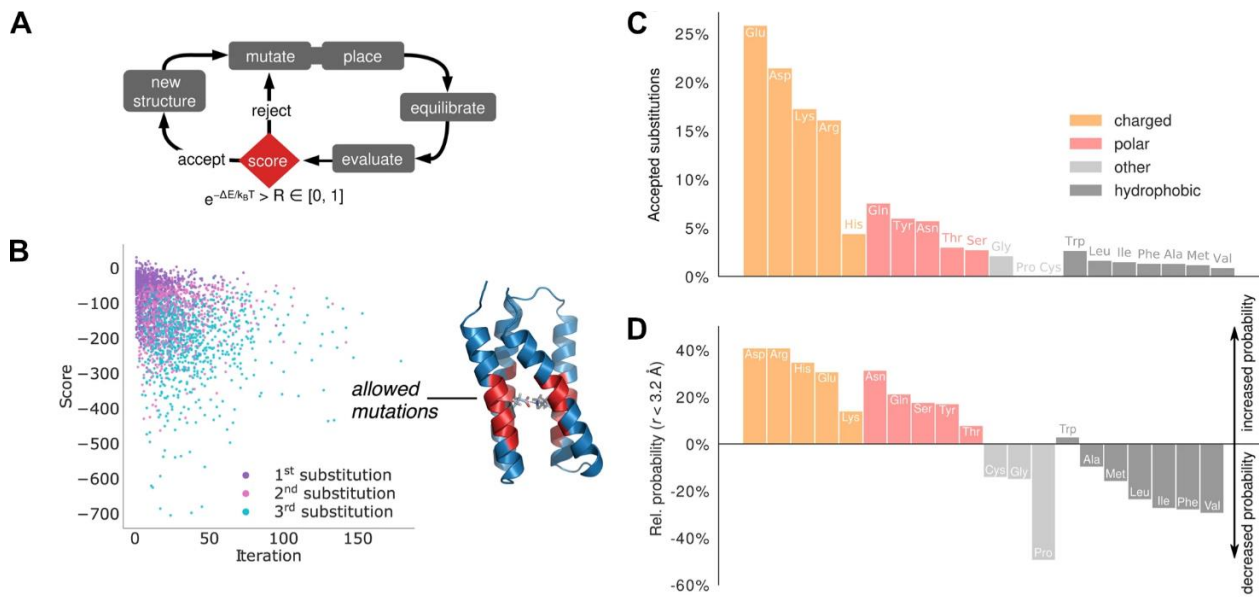


Figure 17: Protein design algorithm used to predict stabilizing elements for buried ion-pairs. **A)** general scheme of the algorithm based on MD simulations combined with an Metropolis-Monte-Carlo-acceptance criterion. **B)** Energy score as a function of the MC iteration step. The energy score was calculated as a combination of total energy, ion-pair stability and interaction between protein and ion-pair. The inset on the right shows the region of Maquette 2, in which mutations were allowed **C)** Frequency of predicted mutations by amino acid type. **D)** Analysis of natural stabilizing elements for buried ion-pairs in the PDB shows similar trends as obtained by the design algorithm.

These constructs are first steps on the way to build small model systems of respiratory complexes, which would allow detailed analysis of the charged switches used by such proteins.

6 Conclusions and outlook

In this work, we studied molecular principles by which the members of the complex I family transport proton or sodium ions across biological membranes to build up an electrochemical gradient. We revealed how the hydration dynamics not only enable proton transfer, but also has a direct influence on the electrostatic coupling between multiple antiporter like subunits, thus ensuring regulation of coupling between multiple ALSs. Our quantum chemical simulations of the proton transfer processes demonstrated important intermediates, like the conserved bridging histidine residues, and stressed the importance of conformational changes in ion-pairs providing a driving force for proton transfer lowering energy barriers. We also compared the dynamics of active and deactive forms of mitochondrial complex I. Our work reveals changes in the hydration structure that could disrupt the proton wire between the Q sites and the ALS providing a possible rationale for the inactivity of the deactive form. Scrutinizing the dynamics of Mbh, an ancient member of the complex I family, suggested how many of the features, which are conserved also in the modern complex I, can be used to generate a sodium motive force instead of a proton motive force and how the modularity of this family of enzymes allows for adaptations to different redox reaction and transport substrates. Lastly, we developed model systems for studying the above-mentioned conserved electrostatics switching principles in a simple artificial protein environment. Despite the insights gained in understanding principles of charge transfer in the complex I family, many mechanistic questions remain. For example, the proton release from the ALS remains as poorly understood as the precise mechanism of coupling the redox reaction to proton transfer. Moreover, while it is clear that the conserved loops at the Q cavity are important, it is still not understood how their dynamics or lack thereof (de)couple proton pumping and redox chemistry. We could also show that simple model systems, such as the 4-helical bundles, can provide insight into the mechanistic principles of complex biological machines, but a significant amount of work is still needed to build artificial proton pumps resembling those of the complex I family.

Bibliography

1. P. Mitchell, Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144–148 (1961).
2. P. Mitchell, Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *Biol. Rev. Camb. Philos. Soc.* **41**, 445–502 (1966).
3. V. R. I. Kaila, M. Wikström, Architecture of bacterial respiratory chains. *Nat. Rev. Microbiol.* **19**, 319–330 (2021).
4. W. Kühlbrandt, Cryo-EM enters a new era. *Elife* **3**, e03678 (2014).
5. M. Röpke, P. Saura, D. Riepl, M. C. Pöverlein, V. R. I. Kaila, Functional Water Wires Catalyze Long-Range Proton Pumping in the Mammalian Respiratory Complex I. *J. Am. Chem. Soc.* **142**, 21758–21766 (2020).
6. P. Saura, V. R. I. Kaila, Molecular dynamics and structural models of the cyanobacterial NDH-1 complex. *Biochim. Biophys. Acta - Bioenerg.* **1860**, 201–208 (2019).
7. V. Sharma, *et al.*, Redox-induced activation of the proton pump in the respiratory complex I. *Proc. Natl. Acad. Sci. USA* **112**, 11571–11576 (2015).
8. O. Haapanen, V. Sharma, Role of water and protein dynamics in proton pumping by respiratory complex I. *Sci. Rep.* **7**, 7747 (2017).
9. J. Warnau, *et al.*, Redox-coupled quinone dynamics in the respiratory complex I. *Proc. Natl. Acad. Sci. USA* **115**, E8413–E8420 (2018).
10. V. R. I. Kaila, M. Wikström, G. Hummer, Electrostatics, hydration, and proton transfer dynamics in the membrane domain of respiratory complex I. *Proc. Natl. Acad. Sci. USA* **111**, 6988–6993 (2014).
11. A. Di Luca, A. P. Gamiz-Hernandez, V. R. I. Kaila, Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl. Acad. Sci. USA* **114**, E6314–E6321 (2017).
12. V. R. I. Kaila, M. I. Verkhovsky, M. Wikström, Proton-coupled electron transfer in cytochrome oxidase. *Chem. Rev.* **110**, 7062–7081 (2010).
13. U. Brandt, Energy converting NADH:quinone oxidoreductase (complex I). *Annu. Rev. Biochem.* **75**, 69–92 (2006).
14. M. Yoshida, E. Muneyuki, T. Hisabori, ATP synthase - A marvellous rotary engine of the cell. *Nat. Rev. Mol. Cell Biol.* **2**, 669–677 (2001).
15. B. J. Berry, A. J. Trewin, A. M. Amitrano, M. Kim, A. P. Wojtovich, Use the Protonmotive Force: Mitochondrial Uncoupling and Reactive Oxygen Species. *J. Mol. Biol.* **430**, 3873–3891 (2018).
16. I. Vercellino, L. A. Sazanov, The assembly, regulation and function of the mitochondrial respiratory chain. *Nat. Rev. Mol. Cell Biol.* **23**, 141–161 (2022).
17. F. Calisto, F. M. Sousa, F. V. Sena, P. N. Refojo, M. M. Pereira, Mechanisms of Energy Transduction by Charge Translocating Membrane Proteins. *Chem. Rev.* **121**, 1804–1844 (2021).
18. G. Cecchini, Function and structure of complex II of the respiratory chain. *Annu. Rev. Biochem.* **72**, 77–109 (2003).
19. M. Wikström, Two protons are pumped from the mitochondrial matrix per electron transferred between NADH and ubiquinone. *FEBS Lett.* **169**, 300–304 (1984).
20. A. Galkin, S. Dröse, U. Brandt, The proton pumping stoichiometry of purified mitochondrial complex I reconstituted into proteoliposomes. *Biochim. Biophys. Acta - Bioenerg.* **1757**, 1575–1581 (2006).
21. M. Wikström, V. Sharma, V. R. I. Kaila, J. P. Hosler, G. Hummer, New perspectives on proton pumping in cellular respiration. *Chem. Rev.* **115**, 2196–2221 (2015).
22. R. Baradaran, J. M. Berrisford, G. S. Minhas, L. A. Sazanov, Crystal structure of the entire respiratory complex I. *Nature* **494**, 443–448 (2013).
23. V. Yankovskaya, *et al.*, Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science* **299**, 700–704 (2003).
24. T. Kleinschroth, *et al.*, X-ray structure of the dimeric cytochrome bc 1 complex from the soil bacterium *Paracoccus denitrificans* at 2.7-Å resolution. *Biochim. Biophys. Acta - Bioenerg.* **1807**, 1606–1615 (2011).
25. N. Mirkin, J. Jaconcic, V. Stojanoff, A. Moreno, High resolution X-ray crystallographic structure of bovine heart cytochrome c and its application to the design of an electron transfer biosensor. *Proteins Struct. Funct. Genet.* **70**, 83–92 (2008).
26. J. Koepke, *et al.*, High resolution crystal structure of *Paracoccus denitrificans* cytochrome c oxidase: New insights into the active site and the proton transfer pathways. *Biochim. Biophys. Acta -*

- Bioenerg.* **1787**, 635–645 (2009).
27. M. Sobti, *et al.*, Cryo-EM structures provide insight into how E. coli F1Fo ATP synthase accommodates symmetry mismatch. *Nat. Commun.* **11**, 1–10 (2020).
 28. C. Hunte, S. Solmaz, H. Palsdóttir, T. Wenz, A structural perspective on mechanism and function of the cytochrome bc1 complex. *Results Probl. Cell Differ.* **45**, 253–278 (2008).
 29. A. R. Crofts, The cytochrome bc1 complex: Function in the context of structure. *Annu. Rev. Physiol.* **66**, 689–733 (2004).
 30. H. Michel, J. Behr, A. Harrenga, A. Kannt, Cytochrome c oxidase: Structure and spectroscopy. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 329–356 (1998).
 31. I. Bertini, G. Cavallaro, A. Rosato, Cytochrome c: Occurrence and functions. *Chem. Rev.* **106**, 90–115 (2006).
 32. M. Wikström, Proton pump coupled to cytochrome c oxidase in mitochondria. *Nature* **266**, 271–273 (1977).
 33. K. Y. Pisa, H. Huber, M. Thomm, V. Müller, A sodium ion-dependent A1AO ATP synthase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *FEBS J.* **274**, 3928–3938 (2007).
 34. E. Morales-Rios, M. G. Montgomery, A. G. W. Leslie, J. E. Walker, Structure of ATP synthase from *Paracoccus denitrificans* determined by X-ray crystallography at 4.0 Å resolution. *Proc. Natl. Acad. Sci. USA* **112**, 13231–13236 (2015).
 35. J. E. Walker, The ATP synthase: The understood, the uncertain and the unknown. *Biochem. Soc. Trans.* **41**, 1–16 (2013).
 36. D. Stock, A. G. W. Leslie, J. E. Walker, Molecular architecture of the rotary motor in ATP synthase. *Science* **286**, 1700–1705 (1999).
 37. , ATP Synthesis by Rotary Catalysis (Nobel lecture) - Walker - 1998 - Angewandte Chemie International Edition - Wiley Online Library (February 15, 2022).
 38. P. Mitchell, Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *Biochim. Biophys. Acta - Bioenerg.* **1807**, 1507–1538 (2011).
 39. J. Hirst, Mitochondrial complex I. *Annu. Rev. Biochem.* **82**, 551–575 (2013).
 40. J. Hirst, J. Carroll, I. M. Fearnley, R. J. Shannon, J. E. Walker, The nuclear encoded subunits of complex I from bovine heart mitochondria. *Biochim. Biophys. Acta - Bioenerg.* **1604**, 135–150 (2003).
 41. L. A. Sazanov, A giant molecular proton pump: Structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **16**, 375–388 (2015).
 42. V. Zickermann, *et al.*, Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* **347**, 44–49 (2015).
 43. D. Kampjut, L. A. Sazanov, The coupling mechanism of mammalian respiratory complex I. *Science* **370** (2020).
 44. K. Parey, *et al.*, Cryo-EM structure of respiratory complex I at work. *Elife* **7**, e39213 (2018).
 45. A. N. A. Agip, *et al.*, Cryo-em structures of complex I from mouse heart mitochondria in two biochemically defined states. *Nat. Struct. Mol. Biol.* **25**, 548–556 (2018).
 46. R. G. Efremov, R. Baradaran, L. A. Sazanov, The architecture of respiratory complex I. *Nature* **465**, 441–445 (2010).
 47. G. Belevich, J. Knuuti, M. I. Verkhovskiy, M. Wikström, M. Verkhovskaya, Probing the mechanistic role of the long α -helix in subunit L of respiratory Complex I from *Escherichia coli* by site-directed mutagenesis. *Mol. Microbiol.* **82**, 1086–1095 (2011).
 48. R. G. Efremov, L. A. Sazanov, Structure of the membrane domain of respiratory complex I. *Nature* **476**, 414–421 (2011).
 49. D. G. Isom, C. A. Castañeda, B. R. Cannon, P. D. Velu, B. García-Moreno E, Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci. USA* **107**, 16096–16100 (2010).
 50. A. C. Robinson, C. A. Castañeda, J. L. Schlessman, E. Bertrand García-Moreno, Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. *Proc. Natl. Acad. Sci. USA* **111**, 11685–11690 (2014).
 51. L. Euro, G. Belevich, M. I. Verkhovskiy, M. Wikström, M. Verkhovskaya, Conserved lysine residues of the membrane subunit NuoM are involved in energy conversion by the proton-pumping NADH:ubiquinone oxidoreductase (Complex I). *Biochim. Biophys. Acta - Bioenerg.* **1777**, 1166–1172 (2008).
 52. L. A. Sazanov, P. Hinchliffe, Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. *Science* **311**, 1430–1436 (2006).
 53. P. Saura, V. R. I. Kaila, Energetics and Dynamics of Proton-Coupled Electron Transfer in the

- NADH/FMN Site of Respiratory Complex I. *J. Am. Chem. Soc.* **141**, 5710–5719 (2019).
54. E. Gnannt, J. Schimpf, C. Harter, J. Hoese, T. Friedrich, Reduction of the off-pathway iron-sulphur cluster N1a of Escherichia coli respiratory complex I restrains NAD⁺ dissociation. *Sci. Rep.* **7**, 1–10 (2017).
 55. T. Pohl, *et al.*, Iron-sulfur cluster N7 of the NADH:Ubiquinone oxidoreductase (complex I) is essential for stability but not involved in electron transfer. *Biochemistry* **46**, 6588–6596 (2007).
 56. M. L. Verkhovskaya, N. Belevich, L. Euro, M. Wikström, M. I. Verkhovsky, Real-time electron transfer in respiratory complex I. *Proc. Natl. Acad. Sci. USA* **105**, 3763–3767 (2008).
 57. R. A. Marcus, On the theory of oxidation-reduction reactions involving electron transfer. I. *J. Chem. Phys.* **24**, 966–978 (1956).
 58. R. A. Marcus, Electrostatic free energy and other properties of states having nonequilibrium polarization. I. *J. Chem. Phys.* **24**, 979–989 (1956).
 59. H. R. Bridges, *et al.*, Structure of inhibitor-bound mammalian complex I. *Nat. Commun.* **11**, 1–11 (2020).
 60. J. G. Fedor, A. J. Y. Jones, A. Di Luca, V. R. I. Kaila, J. Hirst, Correlating kinetic and structural data on ubiquinone binding and reduction by respiratory complex I. *Proc. Natl. Acad. Sci. USA* **114**, 12737–12742 (2017).
 61. J. Gutiérrez-Fernández, *et al.*, Key role of quinone in the mechanism of respiratory complex I. *Nat. Commun.* **11**, 1–17 (2020).
 62. A. Jussupow, A. Di Luca, V. R. I. Kaila, How cardiolipin modulates the dynamics of respiratory complex I. *Sci. Adv.* **5** (2019).
 63. M. A. Tocilescu, U. Fendel, K. Zwicker, S. Kerscher, U. Brandt, Exploring the ubiquinone binding cavity of respiratory complex. *J. Biol. Chem.* **282**, 29514–29520 (2007).
 64. P. G. Roberts, J. Hirst, The deactive form of respiratory complex I from mammalian mitochondria is a Na⁺/H⁺ antiporter. *J. Biol. Chem.* **287**, 34743–34751 (2012).
 65. J. N. Blaza, K. R. Vinothkumar, J. Hirst, Structure of the Deactive State of Mammalian Respiratory Complex I. *Structure* **26**, 312-319.e3 (2018).
 66. A. Cabrera-Orefice, *et al.*, Locking loop movement in the ubiquinone pocket of complex I disengages the proton pumps. *Nat. Commun.* **9**, 4500 (2018).
 67. V. R. I. Kaila, Long-range proton-coupled electron transfer in biological energy conversion: Towards mechanistic understanding of respiratory complex I. *J. R. Soc. Interface* **15**, 20170916 (2018).
 68. L. A. Sazanov, The mechanism of coupling between electron transfer and proton translocation in respiratory complex I. *J. Bioenerg. Biomembr.* **46**, 247–253 (2014).
 69. R. J. Rodenburg, Mitochondrial complex I-linked disease. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 938–945 (2016).
 70. A. D. Vinogradov, Catalytic properties of the mitochondrial NADH-ubiquinone oxidoreductase (Complex I) and the pseudo-reversible active/inactive enzyme transition. *Biochim. Biophys. Acta - Bioenerg.* **1364**, 169–185 (1998).
 71. K. R. Pryde, J. Hirst, Superoxide is produced by the reduced flavin in mitochondrial complex I: A single, unified mechanism that applies during both forward and reverse electron transfer. *J. Biol. Chem.* **286**, 18056–18065 (2011).
 72. A. Stepanova, *et al.*, Deactivation of mitochondrial complex I after hypoxia–ischemia in the immature brain. *J. Cereb. Blood Flow Metab.* **39**, 1790–1802 (2019).
 73. V. G. Grivennikova, A. N. Kapustin, A. D. Vinogradov, Catalytic activity of NADH-ubiquinone oxidoreductase (Complex I) in intact mitochondria: Evidence for the slow active/inactive transition. *J. Biol. Chem.* **276**, 9038–9044 (2001).
 74. E. L. Robb, *et al.*, Control of mitochondrial superoxide production by reverse electron transport at complex I. *J. Biol. Chem.* **293**, 9869–9879 (2018).
 75. A. J. Lambert, M. D. Brand, Superoxide production by NADH:ubiquinone oxidoreductase (complex I) depends on the pH gradient across the mitochondrial inner membrane. *Biochem. J.* **382**, 511–517 (2004).
 76. R. Saprà, M. F. J. M. Verhagen, M. W. W. Adams, Purification and characterization of a membrane-bound hydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus. *J. Bacteriol.* **182**, 3423–3428 (2000).
 77. M. C. Schoelmerich, V. Müller, Energy conservation by a hydrogenase-dependent chemiosmotic mechanism in an ancient metabolic pathway. *Proc. Natl. Acad. Sci. USA* **116**, 6329–6334 (2019).
 78. J. S. McDowall, *et al.*, Bacterial formate hydrogenlyase complex. *Proc. Natl. Acad. Sci. USA* **111**,

- E3948–E3956 (2014).
79. R. Sapra, K. Bagramyan, M. W. W. Adams, A simple energy-conserving system: Proton reduction coupled to proton translocation. *Proc. Natl. Acad. Sci. USA* **100**, 7545–7550 (2003).
 80. P. S. Brereton, M. F. J. M. Verhagen, Z. H. Zhou, M. W. W. Adams, Effect of iron-sulfur cluster environment in modulating the thermodynamic properties and biological function of ferredoxin from *Pyrococcus furiosus*. *Biochemistry* **37**, 7351–7362 (1998).
 81. M. Blaut, G. Gottschalk, Coupling of ATP synthesis and methane formation from methanol and molecular hydrogen in *Methanosarcina barkeri*. *Eur. J. Biochem.* **141**, 217–222 (1984).
 82. G. J. Schut, E. S. Boyd, J. W. Peters, M. W. W. Adams, The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiol. Rev.* **37**, 182–203 (2013).
 83. H. Ogata, W. Lubitz, Y. Higuchi, Structure and function of [NiFe] hydrogenases. *J. Biochem.* **160**, 251–258 (2016).
 84. M. E. Pandelia, H. Ogata, W. Lubitz, Intermediates in the catalytic cycle of [NiFe] hydrogenase: Functional spectroscopy of the active site. *ChemPhysChem* **11**, 1127–1140 (2010).
 85. P. E. M. Siegbahn, R. Z. Liao, The Energetics of Hydrogen Molecule Oxidation in NiFe-hydrogenase. *ACS Catal.* **10**, 5603–5613 (2020).
 86. P. E. M. Siegbahn, J. W. Tye, M. B. Hall, Computational studies of [NiFe] and [FeFe] hydrogenases. *Chem. Rev.* **107**, 4414–4435 (2007).
 87. J. Kalms, *et al.*, Tracking the route of molecular oxygen in O₂-tolerant membrane-bound [NiFe] hydrogenase. *Proc. Natl. Acad. Sci. USA* **115**, E2229–E2237 (2018).
 88. F. Oteri, M. Baaden, E. Lojou, S. Sacquin-Mora, Multiscale simulations give insight into the hydrogen in and out pathways of [NiFe]-hydrogenases from *aquifex aeolicus* and *desulfovibrio fructosovorans*. *J. Phys. Chem. B* **118**, 13800–13811 (2014).
 89. S. Zacarias, *et al.*, A Hydrophilic Channel Is Involved in Oxidative Inactivation of a [NiFeSe] Hydrogenase. *ACS Catal.* **9**, 8509–8519 (2019).
 90. D. N. Grba, J. Hirst, Mitochondrial complex I structure reveals ordered water molecules for catalysis and proton translocation. *Nat. Struct. Mol. Biol.* **27**, 892–900 (2020).
 91. V. K. Moparthi, B. Kumar, C. Mathiesen, C. Hägerhäll, Homologous protein subunits from *Escherichia coli* NADH:quinone oxidoreductase can functionally replace MrpA and MrpD in *Bacillus subtilis*. *Biochim. Biophys. Acta - Bioenerg.* **1807**, 427–436 (2011).
 92. V. K. Moparthi, *et al.*, Functional role of the MrpA- and MrpD-homologous protein subunits in enzyme complexes evolutionary related to respiratory chain complex I. *Biochim. Biophys. Acta - Bioenerg.* **1837**, 178–185 (2014).
 93. E. Virzintiene, *et al.*, Structure and function of the C-terminal domain of MrpA in the *Bacillus subtilis* Mrp-antiporter complex - The evolutionary progenitor of the long horizontal helix in complex I. *FEBS Lett.* **587**, 3341–3347 (2013).
 94. E. Sperling, K. Górecki, T. Drakenberg, C. Hägerhäll, Functional differentiation of antiporter-like polypeptides in complex I; a site-directed mutagenesis study of residues conserved in MrpA and NuoL but not in MrpD, NuoM, and NuoN. *PLoS One* **11** (2016).
 95. V. K. Moparthi, C. Hägerhäll, The evolution of respiratory chain complex I from a smaller last common ancestor consisting of 11 protein subunits. *J. Mol. Evol.* **72**, 484–497 (2011).
 96. F. Mayer, V. Müller, Adaptations of anaerobic archaea to life under extreme energy limitation. *FEMS Microbiol. Rev.* **38**, 449–472 (2014).
 97. G. J. Schut, *et al.*, The role of geochemistry and energetics in the evolution of modern respiratory complexes from a proton-reducing ancestor. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 958–970 (2016).
 98. C. Mathiesen, C. Hägerhäll, The “antiporter module” of respiratory chain Complex I includes the MrpC/NuoK subunit - A revision of the modular evolution scheme. *FEBS Lett.* **549**, 7–13 (2003).
 99. J. K. Lim, F. Mayer, S. G. Kang, V. Müller, Energy conservation by oxidation of formate to carbon dioxide and hydrogen via a sodium ion current in a hyperthermophilic archaeon. *Proc. Natl. Acad. Sci. USA* **111**, 11497–11502 (2014).
 100. H. Yu, *et al.*, Structure of an Ancient Respiratory System. *Cell* **173**, 1636-1649.e16 (2018).
 101. J. Steiner, L. Sazanov, Structure and mechanism of the MRP complex, an ancient cation/proton antiporter. *Elife* **9**, 1–42 (2020).
 102. H. Yu, *et al.*, Structure of the respiratory MBS complex reveals iron-sulfur cluster catalyzed sulfane sulfur reduction in ancient life. *Nat. Commun.* **11**, 1–13 (2020).

103. J. M. Schuller, *et al.*, Structural adaptations of photosynthetic complex I enable ferredoxin-dependent electron transfer. *Science* **363**, 257–260 (2019).
104. U. Brandt, Adaptations of an ancient modular machine Mechanism of energy conversion is conserved in the complex I superfamily. *Science* **363**, 230–231 (2019).
105. B. Li, *et al.*, Structure of the Dietzia Mrp complex reveals molecular mechanism of this giant bacterial sodium proton pump. *Proc. Natl. Acad. Sci. USA* **117**, 31166–31176 (2020).
106. G. E. Moore, Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. *IEEE Solid-State Circuits Soc. Newsl.* **11**, 33–35 (2009).
107. E. H. Kellogg, A. Leaver-Fay, D. Baker, Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins Struct. Funct. Bioinforma.* **79**, 830–838 (2011).
108. Z. Chen, *et al.*, Programmable design of orthogonal protein heterodimers. *Nature* **565**, 106–111 (2019).
109. N. F. Polizzi, *et al.*, De novo design of a hyperstable non-natural protein-ligand complex with sub-Å accuracy. *Nat. Chem.* **9**, 1157–1164 (2017).
110. Y. N. Sun, *et al.*, Model-based process development and evaluation of twin-column continuous capture processes with Protein A affinity resin. *J. Chromatogr. A* **1625**, 461300 (2020).
111. F. Ianni, *et al.*, Integrating experimental and computational techniques to study chromatographic enantioresolutions of chiral tetrahydroindazole derivatives. *J. Chromatogr. A* **1625**, 461310 (2020).
112. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
113. J. A. McCammon, B. R. Gelin, M. Karplus, Dynamics of folded proteins. *Nature* **267**, 585–590 (1977).
114. D. E. Shaw, *et al.*, Anton 3: Twenty Microseconds of Molecular Dynamics Simulation before Lunch in *International Conference for High Performance Computing, Networking, Storage and Analysis, SC*, (IEEE Computer Society, 2021) <https://doi.org/10.1145/3458817.3487397>.
115. R. B. Cooley, D. J. Arp, P. A. Karplus, Evolutionary Origin of a Secondary Structure: α -Helices as Cryptic but Widespread Insertional Variations of α -Helices That Enhance Protein Functionality. *J. Mol. Biol.* **404**, 232–246 (2010).
116. J. C. Phillips, *et al.*, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
117. M. Shamir, Y. Bar-On, R. Phillips, R. Milo, SnapShot: Timescales in Cell Biology. *Cell* **164**, 1302-1302.e1 (2016).
118. W. Thiel, G. Hummer, Nobel 2013 Chemistry: Methods for computational chemistry. *Nature* **504**, 96–97 (2013).
119. W. F. van Gunsteren, *et al.*, Validation of Molecular Simulation: An Overview of Issues. *Angew. Chemie - Int. Ed.* **57**, 884–902 (2018).
120. C. D. Christ, A. E. Mark, W. F. Van Gunsteren, Basic ingredients of free energy calculations: A review. *J. Comput. Chem.* **31**, 1569–1582 (2010).
121. M. C. Zwier, L. T. Chong, Reaching biological timescales with all-atom molecular dynamics simulations (Elsevier, 2010).
122. M. Head-Gordon, R. J. Rico, M. Oumi, T. J. Lee, A doubles correction to electronic excited states from configuration interaction in the space of single substitutions. *Chem. Phys. Lett.* **219**, 21–29 (1994).
123. D. Maurice, M. Head-Gordon, Analytical second derivatives for excited electronic states using the single excitation configuration interaction method: Theory and application to benzo[a]pyrene and chalcone. *Mol. Phys.* **96**, 1533–1541 (1999).
124. G. D. Purvis, R. J. Bartlett, A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **76**, 1910–1918 (1982).
125. K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **157**, 479–483 (1989).
126. T. Van Voorhis, M. Head-Gordon, Two-body coupled cluster expansions. *J. Chem. Phys.* **115**, 5033–5040 (2001).
127. G. N. Lewis, The atom and the molecule. *J. Am. Chem. Soc.* **38**, 762–785 (1916).
128. B. Roux, T. Simonson, Implicit solvent models in *Biophysical Chemistry*, (Elsevier, 1999), pp. 1–20.
129. A. Brünger, C. L. Brooks, M. Karplus, Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.* **105**, 495–500 (1984).

130. F. Zhu, G. Hummer, Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **33**, 453–465 (2012).
131. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
132. L. S. Stelzl, A. Kells, E. Rosta, G. Hummer, Dynamic Histogram Analysis To Determine Free Energies and Rates from Biased Simulations. *J. Chem. Theory Comput.* **13**, 6328–6342 (2017).
133. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
134. S. Jo, J. B. Lim, J. B. Klauda, W. Im, CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **97**, 50–58 (2009).
135. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
136. R. B. Best, *et al.*, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
137. J. Huang, *et al.*, CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2016).
138. A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
139. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104-1–19 (2010).
140. F. Weigend, 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).
141. D. M. A. Smith, Y. Xiong, T. P. Straatsma, K. M. Rosso, T. C. Squier, Force-field development and molecular dynamics of [NiFe] hydrogenase. *J. Chem. Theory Comput.* **8**, 2103–2114 (2012).
142. U. Essmann, *et al.*, A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
143. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041 (2001).
144. T. Meyer, E. W. Knapp, pKa Values in Proteins Determined by Electrostatics Applied to Molecular Dynamics Trajectories. *J. Chem. Theory Comput.* **11**, 2827–2840 (2015).
145. G. Kieseritzky, E. W. Knapp, Optimizing pKa computation in proteins with pH adapted conformations. *Proteins Struct. Funct. Genet.* **71**, 1335–1348 (2008).
146. T. Schneider, E. Stoll, Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **17**, 1302–1322 (1978).
147. S. E. Feller, Y. Zhang, R. W. Pastor, B. R. Brooks, Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).
148. R. Elber, A. P. Ruymgaart, B. Hess, SHAKE parallelization. *Eur. Phys. J. Spec. Top.* **200**, 211–223 (2011).
149. A. Klamt, G. Schüürmann, COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2* **12**, 799–805 (1993).
150. A. Schäfer, H. Horn, R. Ahlrichs, Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **97**, 2571–2577 (1992).
151. R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **162**, 165–169 (1989).
152. P. Plessow, Reaction path optimization without NEB springs or interpolation algorithms. *J. Chem. Theory Comput.* **9**, 1305–1310 (2013).
153. W. E. Ren, E. Vanden-Eijnden, String method for the study of rare events. *Phys. Rev. B - Condens. Matter Mater. Phys.* **66**, 523011–523014 (2002).
154. S. Riahi, 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).
155. N. Agmon, The Grotthuss mechanism. *Chem. Phys. Lett.* **244**, 456–462 (1995).
156. S. Cukierman, Et tu, Grotthuss! and other unfinished stories. *Biochim. Biophys. Acta - Bioenerg.*

- 1757**, 876–885 (2006).
157. A. A. Hagberg, D. A. Schult, P. J. Swart, Exploring network structure, dynamics, and function using NetworkX in *7th Python in Science Conference (SciPy 2008)*, (2008), pp. 11–15.
 158. R. T. McGibbon, *et al.*, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–1532 (2015).
 159. D. Turk, MAIN software for density averaging, model building, structure refinement and validation. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1342–1357 (2013).
 160. L. G. Trabuco, E. Villa, K. Mitra, J. Frank, K. Schulten, Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure* **16**, 673–683 (2008).
 161. S. Lobasso, *et al.*, Coupled TLC and MALDI-TOF/MS analyses of the lipid extract of the hyperthermophilic archaeon *pyrococcus furiosus*. *Archaea* **2012** (2012).
 162. K. Mitusińska, A. Raczyńska, P. Wojsa, M. Bzówka, A. Góra, AQUA-DUCT: Analysis of Molecular Dynamics Simulations of Macromolecules with the use of Molecular Probes [Article v1.0]. *Living J. Comput. Mol. Sci.* **2**, 1–36 (2020).
 163. D. W. A. Buchan, D. T. Jones, The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
 164. D. M. Popović, S. D. Zarić, B. Rabenstein, E. W. Knapp, Artificial cytochrome b: Computer modeling and evaluation of redox potentials. *J. Am. Chem. Soc.* **123**, 6040–6053 (2001).
 165. A. P. Gámiz-Hernández, G. Kieseritzky, A. S. Galstyan, O. Demir-Kavuk, E. W. Knapp, Understanding properties of cofactors in proteins: Redox potentials of synthetic cytochromes b. *ChemPhysChem* **11**, 1196–1206 (2010).
 166. J. Torres-Bacete, P. K. Sinha, N. Castro-Guerrero, A. Matsuno-Yagi, T. Yagi, Features of subunit NuoM (ND4) in *Escherichia coli* NDH-1. Topology and implication of conserved GLU144 for coupling site 1. *J. Biol. Chem.* **284**, 33062–33069 (2009).
 167. J. Torres-Bacete, E. Nakamaru-Ogiso, A. Matsuno-Yagi, T. Yagi, Characterization of the NuoM (ND4) subunit in *Escherichia coli* NDH-1: Conserved charged residues essential for energy-coupled activities. *J. Biol. Chem.* **282**, 36914–36922 (2007).
 168. T. Yuwen, N. R. Skrynnikov, CP-HISQC: A better version of HSQC experiment for intrinsically disordered proteins under physiological conditions. *J. Biomol. NMR* **58**, 175–192 (2014).
 169. J. Iwahara, Y. S. Jung, G. M. Clore, Heteronuclear NMR spectroscopy for lysine NH₃ groups in proteins: Unique effect of water exchange on ¹⁵N transverse relaxation. *J. Am. Chem. Soc.* **129**, 2971–2980 (2007).

Acknowledgements

There are many people in my life, without who this thesis would not have been written in the first place and I would like to thank all of them. Here is an incomplete list of people who deserve my deepest thanks:

First of all, I would like to thank my supervisor Ville Rolf Ilmari Kaila: You contributed to this thesis in a large number of ways (starting with hiring me) but I would like to thank you especially for lots of patience with all my idiosyncrasies (like using fancy words and labelling graphs with “cornflower blue”), years of advice, productive scientific discussion and generally giving me the opportunity to grow as a scientists over the last few years.

I would like to thank my colleagues for their camaraderie, help and a lot of fun throughout the years, in no particular order: Alex Jussupow, Sophie Mader, Mona Baumgart, Daniel Riepl, Hyunho Kim, Rike Algöwer, Max Pöverlein, Sven Klumpe, Patricia Saura, Ana Gamiz-Hernandez, Grant Kemp, Ina Bisha and Mikko Muuronen. Special thanks to Michi Röpke for being my long-term colleague (go Hallenoffice!), dear friend, climbing partner and best man. If I would try to add a list of all the things making my live richer you introduced me to, it would be far too long to be printed here. To Andrea Di Luca for, among many other things (being a mentor when I started in this group for instance or becoming a great friend), teaching me the Banana song (the good people of Helsinki are probably still asking themselves what was going on that night). To Toffer Hültenschmidt for a ton of friendship, lots of (tactical) advice and much insanity in RT, DH and all the rest (plus of course being there for me during the last decade (Yes, it’s been a decade. Yes, we are getting old). Also for introducing me to Jordan, Sanderson and many others.

For enabling most of the work collected in this thesis, I would like to thank my co-authors and collaborators. For granting the computing time that enabled the experiments in this thesis I would like to thank the Leibniz Rechenzentrum (projects pr27xu and pn34he) and PRACE for granting time at the Barcelona Supercomputing centre (project prlejk). I would also like to thank the ERC (Grant Agreement 715311, to V.R.I.K.) and the DFG (TRR 235) for providing the funding that paid my bills for most of these last four years.

Lastly, I want to thank my family: My parents, Gaby and Sigi for raising me to be the man I am (you did well in my (biased) opinion), always supporting me in whatever crackbrained dream I was chasing and being role models for how to raise my own son, Hagen (who made the last few days of writing this thesis so much more precious). To my beloved wife, Clauds, I would like to give the very last thanks: You are the love of my life, just being with you makes the worst days bearable and the best days special and sunsets will truly never be the same. Looking forward to see the rest of the world with you and share my cake. (also: lubb!)

Appendix: Published articles

This section contains all relevant published articles. Permission for printing these articles in this thesis within the scope of their respective licenses has been granted by all relevant publishing agencies. The articles were originally published in the following journals:

Article I: Andrea Di Luca, Max E. Mühlbauer, Patricia Saura and Ville R. I. Kaila (2018) How inter-subunit contacts in the membrane domain of complex I affect proton transfer energetics. *BBA – Bioenergetics*. 1859, 734–741.

Article II: Max E. Mühlbauer, Patricia Saura, Franziska Nuber, Andrea Di Luca, Thorsten Friedrich and Ville R. I. Kaila (2020) Water-Gated Proton Transfer Dynamics in Respiratory Complex I. *J. Am. Chem. Soc.* 142, 13718–13728.

Article III: Michael Röpke, Daniel Riepl, Patricia Saura, Andrea Di Luca, Max E. Mühlbauer, Alexander Jussupow, Ana P. Gamiz-Hernandez, Ville R. I. Kaila (2021) Deactivation blocks proton pathways in the mitochondrial complex I. *Proc. Natl. Acad. Sci. USA* 118, 29.

Article IV: Mona Baumgart, Michael Röpke, Max E. Mühlbauer, Sam Asami, Sophie L. Mader, Kai Fredriksson, Michael Groll, Ana P. Gamiz-Hernandez and and Ville R. I. Kaila (2021) Design of buried charged networks in artificial proteins. *Nat. Commun.* 12, 1895.

Article V: Max E. Mühlbauer, Ana P. Gamiz-Hernandez, Ville R. I. Kaila (2021) Functional dynamics of an ancient membrane-bound hydrogenase. *J. Am. Chem. Soc.* 143, 20873–20883.

Article I



How inter-subunit contacts in the membrane domain of complex I affect proton transfer energetics

Andrea Di Luca¹, Max E. Mühlbauer¹, Patricia Saura, Ville R.I. Kaila*

Department Chemie, Technische Universität München, Lichtenbergstr. 4, Garching, D-85747, Germany



ARTICLE INFO

Keywords:

Bioenergetics
Proton transfer
NADH:ubiquinone oxidoreductase
Enzyme dynamics

ABSTRACT

The respiratory complex I is a redox-driven proton pump that employs the free energy released from quinone reduction to pump protons across its complete ca. 200 Å wide membrane domain. Despite recently resolved structures and molecular simulations, the exact mechanism for the proton transport process remains unclear. Here we combine large-scale molecular simulations with quantum chemical density functional theory (DFT) models to study how contacts between neighboring antiporter-like subunits in the membrane domain of complex I affect the proton transfer energetics. Our combined results suggest that opening of conserved Lys/Glu ion pairs within each antiporter-like subunit modulates the barrier for the lateral proton transfer reactions. Our work provides a mechanistic suggestion for key coupling effects in the long-range force propagation process of complex I.

1. Introduction

The respiratory complex I (NADH:ubiquinone oxidoreductase) is a redox-driven proton pump that serves as an initial electron entry point in prokaryotic and eukaryotic respiratory chains [1–4]. By reducing quinone (Q) to quinol (QH₂), complex I transports four protons across a biological membrane [1,5,6] and establishes a proton motive force (μ_{mf}) that is employed for active transport and synthesis of adenosine triphosphate (ATP) [7,8]. Complex I is by far the largest and most intricate member of the respiratory chain, and despite recently resolved structures [9–12], the molecular mechanism by which it pumps protons still remains unclear.

Complex I is a 0.5–1 MDa L-shaped enzyme that comprises up to 45 subunits in eukaryotes [13], and is organized into a hydrophilic domain and a membrane domain. The electron transfer process takes place in the ca. 100 Å long hydrophilic domain, whereas the 200 Å long membrane domain is responsible for the proton pumping function (Fig. 1). The 14 conserved core subunits of the enzyme constitute the machinery needed to catalyze the long-range proton-coupled electron transfer (PCET) process. The remaining 31 supernumerary subunits in the eukaryotic enzyme are organized around the core subunits [14–16] and are possibly involved in the regulation of enzyme functions, e.g., by the active-to-deactive transition [11,12,17–20], which modulates complex I activity.

The electron transport process is initiated by the oxidation of nicotinamide adenine dinucleotide (NADH), which transfers its two electrons via a non-covalently bound flavin mononucleotide (FMN) cofactor to a chain of 8–9 iron sulfur centers (ISC) and further to the Q-binding site, located ca. 30 Å above the membrane surface (Fig. 1) [9,21,22]. The electron transfer (eT) between NADH and the terminal N2 iron-sulfur center takes place on ca. 90 μs timescales [23,24], which is fast relative to the millisecond turnover of complex I [2], and thus not rate-limiting for the proton pumping process [2,25].

The Q pocket is located at the interface between the three subunits Nqo4, Nqo6, and Nqo8 (*T. thermophilus* nomenclature), and extends ca. 40 Å toward the membrane domain. There are to date no experimentally resolved structures of complex I with bound Q, but computational studies [26,27] suggest that Tyr₄-87 and His₄-38 stabilize the Q head-group and function as local proton donors in the Q-reduction process. These findings are also supported by site-directed mutagenesis studies [28,29]. The Q cavity has a non-uniform polarity, and it comprises a kink region with many polar and charged amino acids [9,30]. From this kink, a chain of conserved charged/polar residues extends in the middle of the membrane domain toward the terminal Nqo12 subunit [9,31].

Of the seven membrane domain subunits, three antiporter-like subunits, Nqo12, Nqo13, and Nqo14 are evolutionary related to each other and to multi-resistance and pH adaptation (Mrp) Na⁺/H⁺ antiporters. The antiporter-like subunits share an internal pseudo-

* Corresponding author.

E-mail address: ville.kaila@ch.tum.de (V.R.I. Kaila).

¹ Contributed equally to this work.

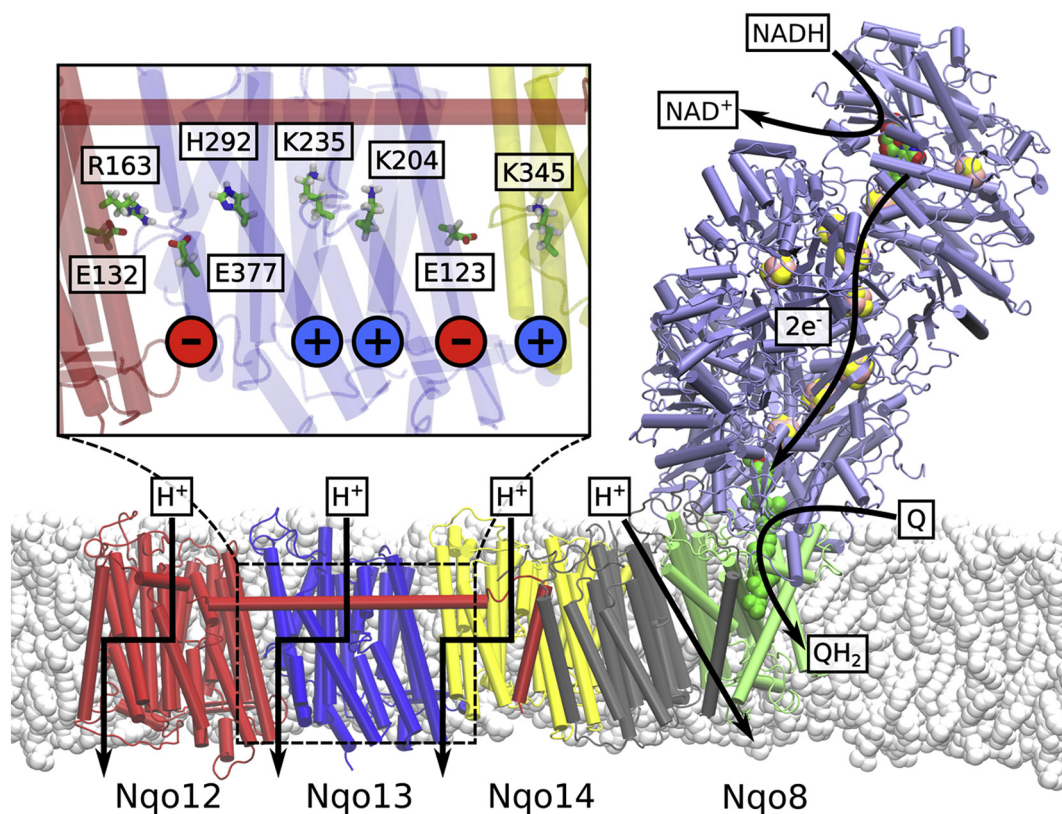


Fig. 1. Structure of the bacterial complex I from *T. thermophilus* (PDB ID: 4HEA). Electron transfer takes place in the hydrophilic domain (in purple) between NADH and Q. Energy released from the Q reduction process is employed to transfer four protons across the membrane. The antiporter-like subunits Nqo12 (in red), Nqo13 (in blue) and Nqo14 (in yellow) are likely to transfer one proton each, whereas the location of the fourth proton pathway is still not fully clear (but cf. [33,34]). *Inset:* Chain of conserved charged and polar residues in Nqo13. From right to left: terminal charged residue of Nqo14, the Glu/Lys ion pair, the central Lys, the bridging His, the terminal charged residue of Nqo13 (Glu), and the interface to Nqo12.

symmetry, with two trans-membrane (TM) helix bundles, TM4-8 and TM9-13, that contain a broken-helix element. A similar five-helical bundle, TM2-6, is also present in the Nqo8 subunits, and it comprises a part of the Q channel. The proton pumping is likely to occur in the Nqo12, Nqo13, and Nqo14 subunits, pumping one proton each. The location of the fourth proton pathway is still under debate, but a possible location is the region between Nqo8 and Nqo11 [9,10]. Recent simulations [32–34] show that the proton channels are likely to form at the broken-helix segment, similar as in other transporters [35,36].

It is possible to identify conserved repeated residue motifs in each antiporter-like subunit. These include a Lys/Glu ion pair (Arg/Glu in Nqo12), a central Lys residue, one or more bridging His residues, and a terminal charged Lys or Glu residue (Fig. 1, inset). Site-directed mutagenesis experiments [37–42] and molecular simulations suggest that these residues are crucial for the proton pumping activity. Moreover, the Q reduction activity is also affected by mutations of these residues, suggesting that the electron and proton transfer processes are tightly coupled in complex I. Importantly, to achieve such tight coupling between the Q reaction and the terminal proton transfer in Nqo12, the “Q-reduction signal” needs to propagate through the complete membrane domain.

Biochemical, structural, and computational studies have probed possible proton pumping mechanisms ([25,33,42–46], cf. [44] and refs. therein). Although the overall pumping process takes place on milliseconds timescales, individual transitions that couple to the pumping process may take place on much shorter timescales once a rate-limiting step has been overcome. Therefore, relaxation of such “non-equilibrium” state created here, e.g., by protonation changes, can be employed to obtain mechanistic information of rare events using molecular dynamics simulations that are shorter than the overall turnover

timescale.

Recently, we suggested a molecular mechanism where conformational changes in the Lys/Glu ion pairs are involved in the long-range force propagation process and transmit the signal between neighboring antiporter-like subunits. This mechanism involves sequential Lys/Glu ion-pair dissociation and lateral proton transfer processes, propagating the signal from Nqo14 to Nqo12. We found that the energetics of ion-pair dissociations depends on the protonation state of the central Lys residues, making also the reverse effect possible, i.e., that the ion-pair dynamics modulate the pK_a of the neighboring amino acids. The protonic connectivity to the two membrane sides (N- and P-side) was further suggested to be regulated by the hydration state of *input* and *output* channels, which in turn is controlled by the state of buried charged residues. The water channels “open” and “close” on the sub- μ s timescale, which may provide a rate-limiting element in the proton pumping process. It was suggested that the proton N-side input and P-side output channels are located at symmetry-related positions. More specifically, the channels form at the interface between subunits and the 5-TM helices bundles, sharing the same symmetry and connecting the buried central Lys and terminal charged residues within each subunit to the N- and P-sides of the membrane [33]. The lateral proton transfer takes place between the central Lys residue and the charged residue facing the subsequent subunit. Each of these events triggers the following process and propagates across the complete ca. 200 Å membrane domain. We also suggested based on thermodynamic considerations [46] that the proton pumping process involves a “backwave” that couples to proton release across the membrane.

To study the energetics and dynamics of this coupling principle, we perform here classical molecular dynamics (MD) simulations in combination with quantum chemical density functional theory (DFT)

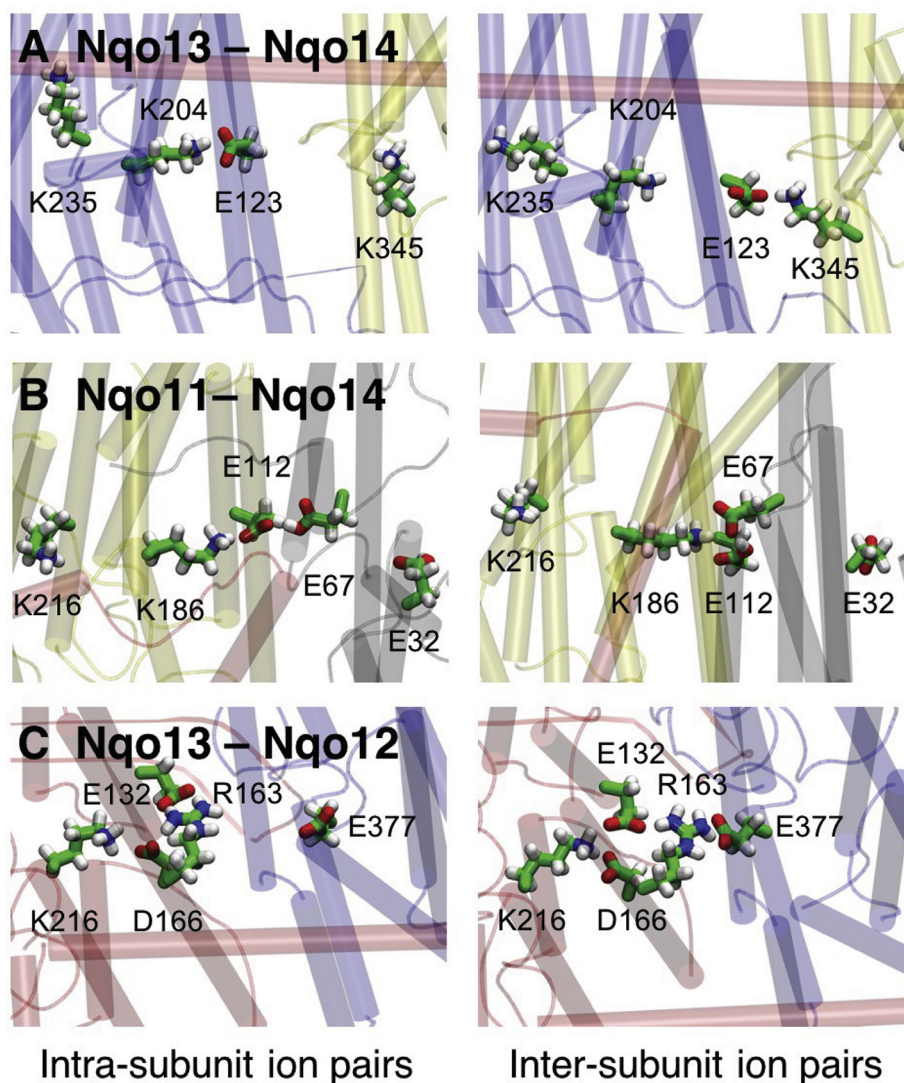


Fig. 2. Structure of the antiporter-like subunit interfaces in the membrane domain of complex I. The figure shows A) the Nqo13-Nqo14 interface, B) the Nqo11-Nqo14 interface, and C) the Nqo13-Nqo12 interface. Ion pairs between conserved charged residues form alternating contacts within the same and neighboring subunits, with snapshots of *intra*-subunit contacts (to the left) and *inter*-subunit contacts (to the right).

calculations on the experimentally resolved X-ray structure of complex I from *Thermus thermophilus*. Our data suggest how the *inter*-subunit contacts are established and how these interactions could modulate the proton transfer energetics.

2. Models and methods

2.1. Classical molecular dynamics

The X-ray structure of *T. thermophilus* complex I [9] was embedded in a POPC membrane and solvated with TIP3P water, and the system was neutralized with a ca. 100 mM NaCl concentration. Ubiquinone (Q_{10}) was modeled in the Q-cavity, which was identified using the HOLE [47] software, and the Q headgroup was placed between His₄₋₃₈ and Tyr₄₋₈₇ of the Nqo4 subunit. The system comprised ca. 830,000 atoms. A constant temperature of 310 K and pressure of 1 bar were modeled in an *NPT* ensemble, and long-range electrostatics were treated by the Particle Mesh Ewald (PME) method [48]. The simulations were performed using NAMD2 [49] and the CHARMM27 force field [50,51] using a 2 fs integration timestep. Force field parameters for the cofactors were derived from density functional theory (DFT/B3LYP/def2-TZVP) calculations. pK_a values were estimated using

Poisson-Boltzmann (PB) continuum electrostatic calculations and performed using the Adaptive PB solver (APBS) [52], by performing the Monte Carlo sampling of the 2^N possible protonation states with Karlsberg+ [53]. The system was described by explicit partial atomic charges embedded in an inhomogeneous medium with an $\epsilon = 4$, and bulk water by a homogeneous medium with $\epsilon = 80$. Part of the simulation data were also employed in Ref. [33] and are reported in Table S1. Principal component analysis (PCA) [54,55] of the MD data was performed using the position of the C α atoms of subunits during 200–600 ns of dynamics (simulations 2 and 4) after 400 ns of simulation. The PCA and related analyses were performed with ProDy [56].

2.2. Quantum chemical density functional theory models

Quantum chemical DFT models consisting of the Lys₁₃₋₂₃₅, His₁₃₋₂₉₂ and Glu₁₃₋₃₇₇ residues of Nqo13 and Lys₁₄₋₂₁₆, His₁₄₋₂₆₅ and Lys₁₄₋₃₄₅ residues of Nqo14, with three intervening water molecules in each case, were constructed based on 100 ns relaxed MD simulations. The amino acid residues were cut at the C β (for His and Glu) or C δ (for Lys) positions, which were fixed during structure optimization at the B3LYP-D3/def2-SVP level [57–60]. The protein environment was treated as a polarizable medium with $\epsilon = 4$ using the conductor-like

screening model (COSMO) [61]. Transition states were also optimized at the same level of theory. Electronic energies were computed at B3LYP-D3/def2-TZVP/ $\epsilon = 4$ level with zero-point vibrational (ZPE) energy corrections obtained at B3LYP-D3/def2-SVP/ $\epsilon = 4$ level. To study the effect of the Lys₁₃-204/Lys₁₄-186, we added a Lys α -amino group at 10.5 Å from the Lys₁₃-235/Lys₁₄-216. All calculations were performed with TURBOMOLE v 6.6 [62].

3. Results and discussion

3.1. Inter-subunit contacts affect intra-subunit residue conformations

Starting from the crystal structure of complex I from *Thermus thermophilus*, we performed ca. 3 μ s classical molecular dynamics (MD) simulations with the central polar residues modeled in both their protonated and deprotonated forms (Table S1). In the MD simulations, we find that the Lys/Arg-Glu ion pairs in each antiporter-like subunit form transient contacts with neighboring subunits that are stabilized by interactions with oppositely charged residues at their interface (Fig. 2). For each of the three interfaces Nqo11/Nqo14, Nqo13/Nqo14, and Nqo12/Nqo13, we observe a qualitatively similar behavior, with the key residues showing a two-state conformational switching behavior, which could be important for the signal propagation in the membrane domain of complex I.

We find that the Nqo13/Nqo14 interface has the clearest switching behavior (Fig. 3A). The MD simulations suggest that Glu₁₃-123 can form both an *intra*-subunit salt-bridge with Lys₁₃-204 and an *inter*-subunit contact with Lys₁₄-345 upon conformational switching (Fig. 2A, Fig. 3A). This switching is coupled with a decrease in the distance between Lys₁₃-204 and the central Lys₁₃-235 (Fig. 3A, upper panel), which could function as a primary proton donor in the pumping process [33], as also supported by site-directed mutagenesis experiments [41]. The distance distribution for the Lys₁₃-235/Lys₁₃-204 pair shows two major sidechain conformations (Fig. 2A). When either of the residues is modeled in the deprotonated state, we obtain a mean distance of ca. 10.5 Å, whereas when both residues are in their protonated (charged) states, the electrostatic repulsion increases their mean distance to ca. 14.0 Å (Fig. 3A, lower panel), showing that the ion-pair dynamics is tightly coupled to the protonation state of the residues (Fig. 3A). Interestingly, in the crystal structure of complex I from *Thermus thermophilus* (PDB ID: 4HEA), the Lys₁₃-204/Glu₁₃-123 ion pair has been refined in the dissociated state, with Glu₁₃-123 flipped toward the Nqo14 subunit.

Similar as for Nqo13/Nqo14, we also observe a conformational switching at the interface between Nqo14 and Nqo11 (Fig. 3B). However, in contrast to the Nqo13/Nqo14 interface where Glu₁₃-123 interacts with two oppositely charged residues, Glu₁₄-112 is surrounded by Lys₁₄-186 and two acidic residues, Glu₁₁-67 and Glu₁₁-32 (Fig. 2B). Here we observe two distinct conformations of Glu₁₁-67, which result in a ca. 7 Å distance to Lys₁₄-186 when it is modeled in a protonated state, and a ca. 3.2 Å distance when it is modeled in a deprotonated state (Fig. 3B), forming a hydrogen-bonded contact. When deprotonated, Glu₁₁-67 faces away from Glu₁₁-32, possibly due to electrostatic repulsion. Similar as in Nqo13, this conformational change correlates with an increase in the Lys₁₄-186/Lys₁₄-216 distance (Fig. 3B). Lys₁₄-216 is likely to function as the proton donor in subunit Nqo14 [33,44].

We next analyzed the ion-pair dynamics at the Nqo12/Nqo13 interface, which is structurally different from the other antiporter-like subunits. The Nqo12/Nqo13 interface comprises two positively charged residues, Lys₁₂-216 and Arg₁₂-163, which are compensated by three negatively charged residues, Glu₁₃-377, Glu₁₂-132, and Asp₁₂-166 (Fig. 2C). Arg₁₂-163 has been suggested to replace a putative Na⁺-binding site in their evolutionary ancestral Na⁺-pumping Mrp transporters [4]. Our MD simulations suggest that Arg₁₂-163 can form a salt-bridge with the surrounding acidic residues. In addition to the two distinct conformational states observed for the other interfaces, we also

observe a third intermediate state (Fig. 3C, lower panel), which could arise from simultaneous interaction with both its acidic neighbors at the same time. We find that the strong interaction between Arg₁₂-163 and Asp₁₂-166 is anti-correlated with the opening of the Glu₁₃-377/Arg₁₂-163 ion pair (Fig. 3C, upper panel).

In Nqo12, the distance between the putative proton donor, Lys₁₂-329, and residues Lys₁₂-216/Asp₁₂-166 at the *inter*-subunit interface is larger than 18 Å, making a direct electrostatic coupling between the two sites somewhat weaker, as compared to the interaction in the other subunits. However, the conserved His₁₂-241 of TM8, located in the same position as the central Lys in Nqo13 and Nqo14, could provide a link necessary to couple the ion-pair dynamics to the proton transfer process. The higher complexity in Nqo12 might be related to the fact that it is the terminal antiporter-like subunit, and the coupling might be weaker than for the other subunits [4] (however, cf. also [46]).

3.2. Channel hydration and coupling between subunits

We recently observed a connection between the protonation state of the central Lys residues and the opening/closing dynamics of the proton channels in the antiporter-like subunits [33]. Our MD simulations suggest that in Nqo13, the water connectivity next to the broken helix TM7a is established when Lys₁₃-235 is protonated, and is lost upon its deprotonation (Fig. 4A). The hydration state of the antiporter-like subunit is, interestingly, also coupled with subtle structural changes (Fig. 4A). We observe that upon deprotonation of the middle Lys, the hydration level of the channel next to the broken helix TM7a drops significantly (Fig. 4B). Only one water molecule, which interacts with the deprotonated Lys, remains close to the TM4-8 helix bundle, whereas the remaining channel water is pushed toward the N-side by His₁₃-211 and Leu₁₃-214, which move closer together to form a gating element. Both residues are located on the upper part of the broken helix (TM7) of the antiporter-like subunit. Communication between the two parts of the helix (TM7a and TM7b) could be mediated by Trp₁₃-213 in TM7a, as indicated by a hydrogen-bond between the tryptophan and the backbone of Leu₁₃-203 of the lower helix (TM7b). The tilt of TM7a relative to TM7b also changes significantly after deprotonation of Lys₁₃-235 (Fig. 4B), which also couples to a subtle shift in the π -kink of TM8 with Lys₁₃-235. These conformational changes lead to a decrease in the channel radius by ca. 2 Å measured from the distance between Leu₁₃-214 on TM7a and Lys₁₃-287 on TM8, and a decrease in mean channel hydration by ca. 50% (within 4 Å of the gate, Fig. 4B).

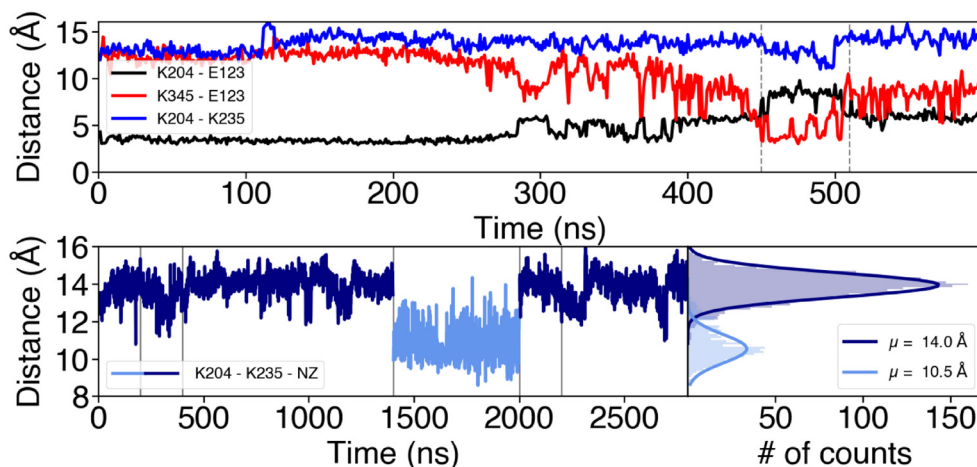
The MD simulations performed with different protonation states reveal possible effects driving the opening/closure of channels. To further probe global coupling effects, we performed a principal component analysis (PCA) that projects out global slow relaxing degrees of freedom. The dynamical correlation between the antiporter-like subunits, calculated based on the PCA correlation matrix on different trajectories, shows how the coupling of motions could depend on both the hydration level of the subunits and the protonation states of the buried residues. When the putative proton channels are hydrated, we observe a strong coupling between the subunits (Fig. 5A). However, upon dehydration of the water channels by deprotonation of the central Lys, the coupling between subunits weakens as indicated by a reduced correlation (Fig. 5B).

Although analysis of more intermediate states is needed to clarify details of the coupling between subunits, our data nevertheless indicate that the protonation state of conserved residues and channel hydration affect complex I dynamics. This suggests that not only the protein structure, but also the water molecules play an active role in the pumping process, by providing essential coupling elements that transmit the signal in addition to their role as “proton wires”.

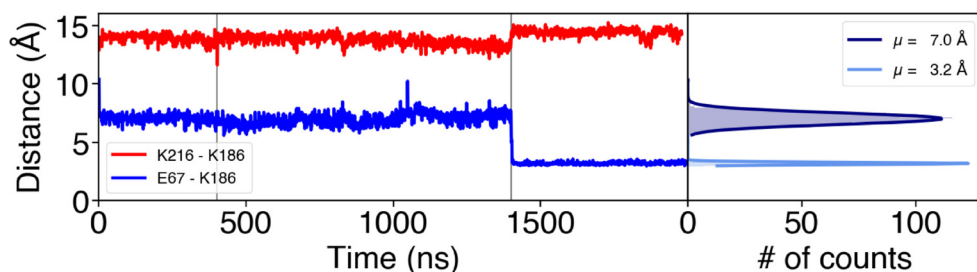
3.3. Inter-subunit contacts modulate proton transfer energetics

To “push” the proton horizontally within the antiporter-like

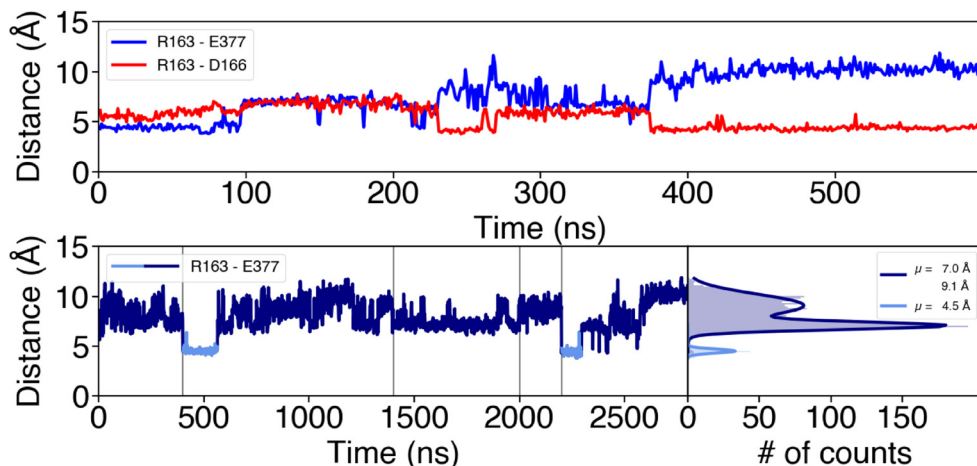
A Nqo13 - Nqo14



B Nqo11 - Nqo14



C Nqo12 - Nqo13



domains, complex I must invest energy by destabilizing the protonated middle Lys. As described above, this could be achieved by opening of the Lys/Glu ion pair, which is expected to result in a charge repulsion between Lys₁₃-204 and Lys₁₃-235 (in Nqo13), as also suggested by recent free energy simulations [33]. To qualitatively probe such coupling principles, we built quantum chemical model systems comprising the sidechains of residues Lys-His-Lys and Lys-His-Glu (proton donor – bridging residue – proton acceptor), bridging water molecules, and using a protonated Lys sidechain as “triggering signal” for the proton transfer process (Fig. 6). The structures were extracted from the classical MD simulations of complex I after ca. 100 ns. After DFT optimization of intermediate and transition states, we studied the energetics of the proton transfer reaction for this model system by probing the

effect of a positive charge next to the proton donor, mimicking the charge of the unpaired Lys resulting from the flip of the interface salt-bridges to the neighboring subunit.

In subunit Nqo13 (Fig. 6A), our quantum chemical models suggest that proton transfer from Lys₁₃-235 to His₁₃-292 via one bridging water molecule (A1 to A3 in Fig. 6A) has an energy barrier of ca. 4 kcal mol⁻¹ and takes place via a hydronium-like transition state structure. The process is exergonic by ca. 5 kcal mol⁻¹. The effect of the uncompensated Lys₁₃-204 lowers this barrier by ca. 2 kcal mol⁻¹ and renders the reaction more exergonic by ca. 3 kcal mol⁻¹. The subsequent proton transfer from His₁₃-292 to Glu₁₃-377 via two water molecules (A3 to A5 in Fig. 6A) has a barrier of ca. 6 kcal mol⁻¹, with a transition state resembling a Zundel ion, and a reaction energy of ca.

Fig. 3. Ion-pair dynamics at the subunit interfaces in the membrane domain of complex I. A) The Nqo13 – Nqo14 interface. *Upper panel*: Distances of the Lys₁₃-204/Glu₁₃-123 ion pair (in black), the Lys₁₄-345/Glu₁₃-123 ion pair (in red) during 0.6 μs of MD simulation (simulation 5). The flip of the ion pair correlates with a decrease in the Lys₁₃-204/Lys₁₃-235 distance (in blue). *Lower panel*: Lys₁₃-235/Lys₁₃-204 distances and their distribution during 2.8 μs of dynamics (simulations 1 to 5). The dark blue curve represents a ‘repulsive’ regime where both residues are positively charged, whereas deprotonation of Lys₁₃-204 leads to sampling of shorter distances (light blue curve). Vertical grey lines indicate boundaries between separate simulations. B) The Nqo11 – Nqo14 interface. Distances of the Glu₁₁-67/Lys₁₄-186 (in blue) and Lys₁₄-186/Lys₁₄-216 (in red) ion pairs during 2 μs MD simulations (simulations 1, 2, 5) and the distribution of the distances between Glu₁₁-67/Lys₁₄-186 during the MD simulations. Upon deprotonation of Glu₁₁-67 at 1.4 μs, the inter-subunit contact forms, which correlates with an increase in the Lys₁₄-186/Lys₁₄-216 distance. Deprotonation of Glu₁₁-67 leads to formation of contact with Lys₁₄-186. C) The Nqo12 – Nqo13 interface. *Upper panel*: Distances of the Arg₁₂-163/Glu₁₃-377 (in blue) and the Arg₁₂-163/Asp₁₂-166 (in red) ion pairs during 0.6 μs of MD simulation (simulation 5). Arg₁₂-163 forms *inter*- and *intra*-subunit salt-bridges. *Lower panel*: Distances between Arg₁₂-163/Glu₁₃-377 during 2.8 μs of simulation (simulations 1 to 5) and their respective distribution. *Intra*-subunit and intermediate states are drawn in dark-blue and the *inter*-subunit state in light blue. Vertical grey lines indicate boundaries between separate simulations.

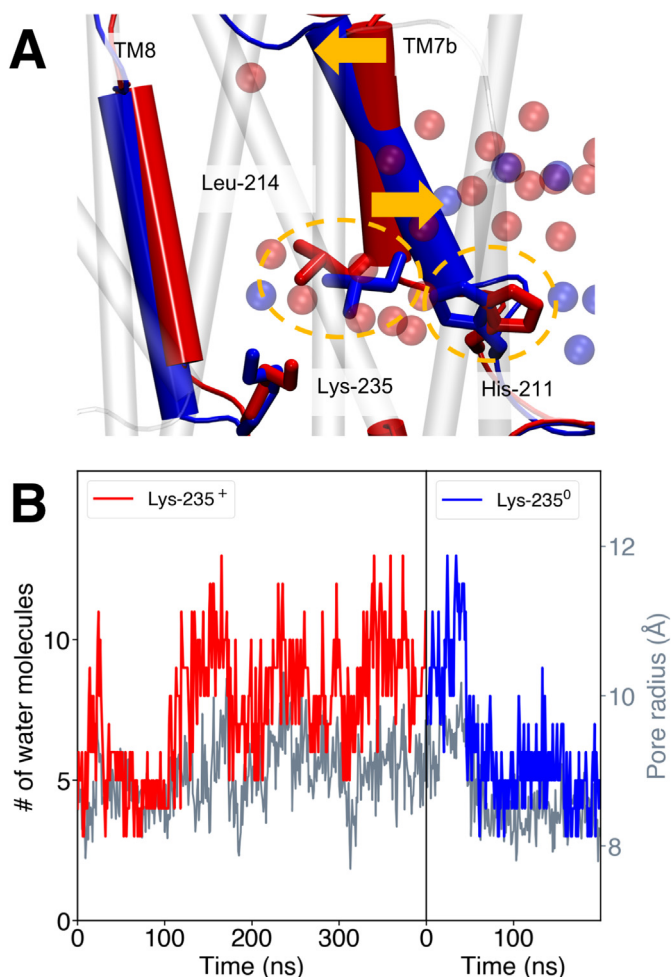


Fig. 4. A) Conformational changes associated with deprotonation of the central Lys₁₃₋₂₃₅ in subunit Nqo13 in the membrane domain of complex I. When Lys₁₃₋₂₃₅ is simulated in a protonated state (red, simulation 2) the gate formed by the sidechains of Leu₁₃₋₂₁₄ and His₁₃₋₂₁₁ (yellow circles) is open, whereas the gate closes when Lys₁₃₋₂₃₅ is modeled in a deprotonated state (blue, simulation 4), inducing dehydration of the channel. The opening/closing is coupled to a structural tilting of the TM7b broken helix (yellow arrows). B) The channel hydration state (in red and blue) and channel radius (in grey) for subunit Nqo13 with Lys₁₃₋₂₃₅ modeled the protonated (simulation 2, red) and deprotonated states (simulation 4, blue), respectively. The hydration state is measured as the number of water molecules within 4 Å around the Leu₁₃₋₂₁₄/His₁₃₋₂₁₁ gate. The pore radius is measured as distance between Leu₁₃₋₂₁₄ (TM7a) and Lys₁₃₋₂₈₇ (TM8). The mean hydration changes from four water molecules to eight water molecules during channel opening.

–14 kcal mol⁻¹. The exergonicity of this process is likely to be overestimated due to the neutralization of an uncompensated negative Glu in a low dielectric environment. The presence of the unpaired Lys sidechain at 10.5 Å from Lys₁₃₋₂₃₅ stabilizes both the intermediate and final states (A3 to A5 in Fig. 6A) by ca. 3 kcal mol⁻¹, conserving the barrier of ca. 5 kcal mol⁻¹. In subunit Nqo14 (Fig. 6B), the energetics of the proton transfer from Lys₁₄₋₂₁₆ to His₁₄₋₂₉₂ (B1 to B3 in Fig. 6B) is similar as in the Nqo13 subunit model, with a barrier of ca. 5 kcal mol⁻¹ and an exergonicity of ca. 5 kcal mol⁻¹. We find that the charge of Lys₁₄₋₁₈₆ could stabilize this hydronium-ion transition state by ca. 2 kcal mol⁻¹, but it slightly destabilizes the protonated His₁₄₋₂₆₂ by ca. 1 kcal mol⁻¹. Proton transfer from His₁₄₋₂₆₂ to Lys₁₄₋₃₄₅ via two water molecules (B3 to B5 in Fig. 6B) has a relatively high energy barrier of ca. 15 kcal mol⁻¹ in the model system, and it is endergonic by ca. 7 kcal mol⁻¹, yielding a final state which is ca. 2 kcal mol⁻¹ higher in energy than the initial state. However, the charge of the

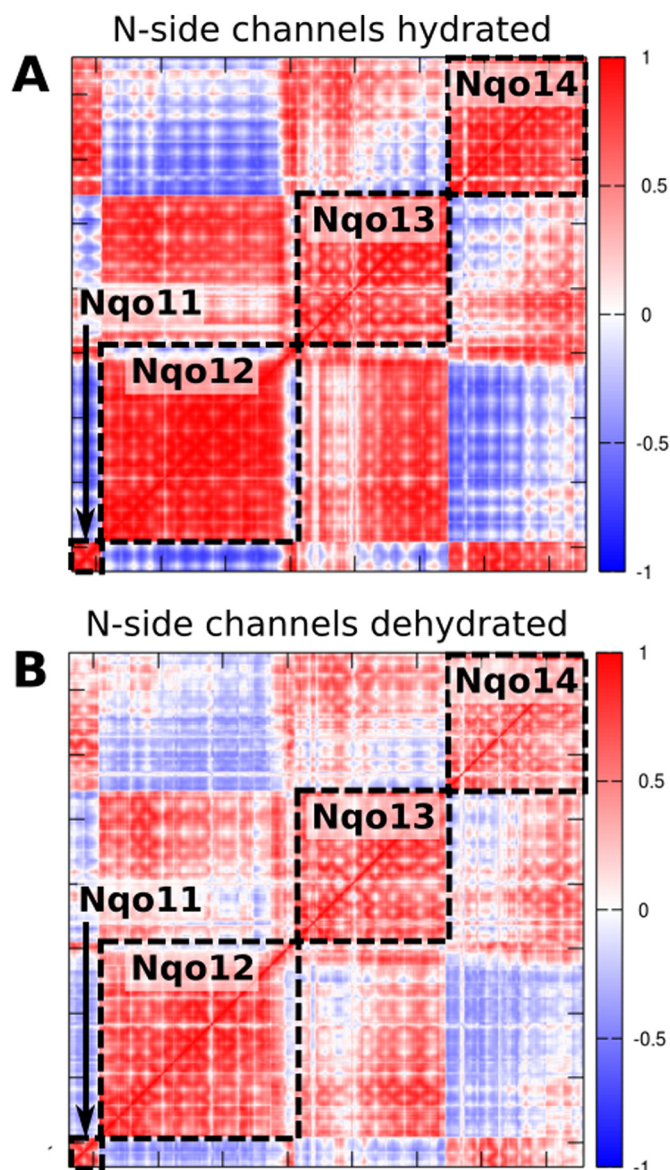


Fig. 5. Correlation matrices of the membrane domain of complex I obtained from principal component (PC) analysis of the MD simulation data (simulations 2 and 4, Table S1). Correlation of PCs A) with open channels between the N-side bulk water and central Lys residues, B) upon deprotonation of the central Lys residues and closure of the N-side channels. Neutralization and/or dehydration of the subunits cause a decreased coupling between subunits. The colors refer to the inter-residue correlation from 1 (directly correlated), to –1 (anti-correlated).

uncompensated Lys lowers the barrier by ca. 4 kcal mol⁻¹ and makes the reaction more exergonic by ca. 4 kcal mol⁻¹, now being stabilized relative to the initial state by ca. 2 kcal mol⁻¹. The intermediates and transition states observed in our QM cluster models are consistent with the ones from QM/MM-models probed in our previous study [33].

Although the proton transfer model employed here is simple, it nevertheless captures qualitative features on how modulation of the proton transfer energetics could be achieved in complex I. The model also describes some of the essential features needed to create a tightly coupled proton pumping machinery. While here the cost to change the energetics of the system is given by the energy required to “create” a positive charge next to our proton transfer chain, the same process in the complete system would also include the free energy caused by the separation of the ion pair and creation of the “excess charge”. Free energy computations [33] estimate these effects as 3–4 kcal mol⁻¹ with

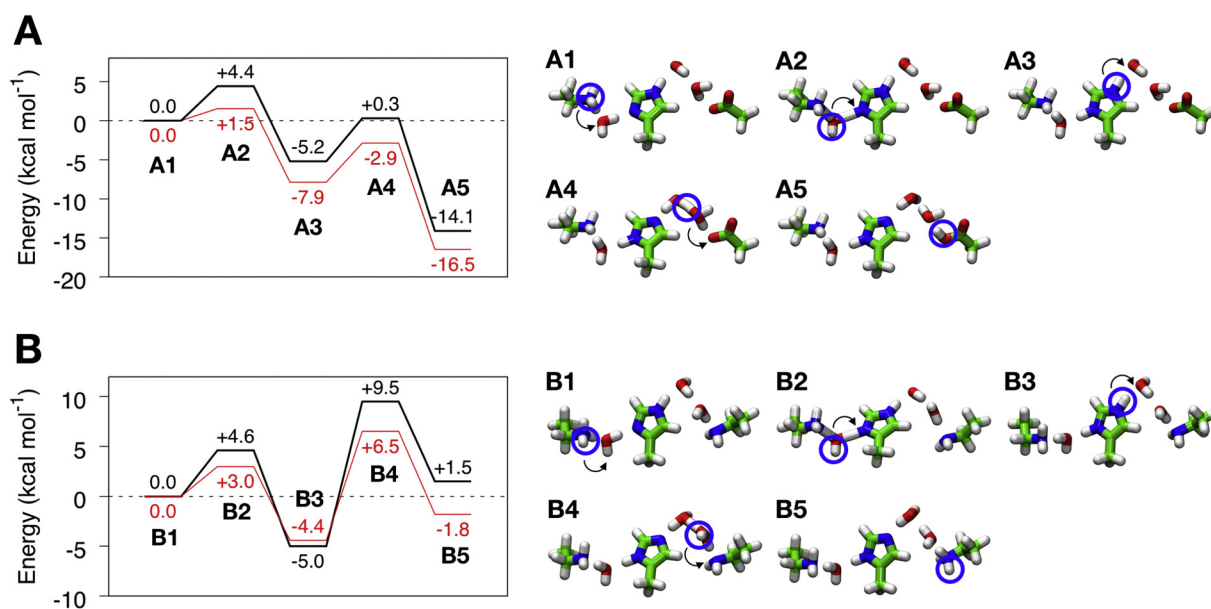


Fig. 6. Effect of dissociation of the “Glu-Lys” ion pair on proton transfer energetics in minimal quantum chemical models of the proton pathway in A) Nqo13 and B) Nqo14. Energy profiles (in kcal mol⁻¹) are shown to the left and the corresponding structures of optimized intermediates and transition states are drawn on the right. Profiles in red show the effect of adding a polarizing Lys⁺, that forms upon dissociation of the Lys/Glu ion pair. The energetics were estimated at B3LYP/def2-TZVP/ $\epsilon = 4$ level with zero-point energy (ZPE) corrections calculated at the B3LYP/def2-SVP level.

a deprotonated central Lys and > 8 kcal mol⁻¹ when Lys_{13–235} is modeled in a protonated state. The models thus support that the inter-subunit ion pairs could influence the lateral proton transfer reaction and coupling between subunits propagated.

4. Conclusions

Our molecular simulations presented here provide insight on how *inter*-subunit contacts can modulate the proton transfer processes in the antiporter-like subunits of complex I. The proton pumping in the membrane domain of complex I, involves proton transfer reactions between the N-side bulk and buried titratable lysine residues, as well as horizontally across the membrane domain. Our simulations indicate that the energetics and kinetics of these proton transfer reactions are coupled to the charged state of a conserved ion pair within each antiporter-like subunit, and in turn, regulated by the charged state of the neighboring subunit. Our data indicate that Glu_{13–123} in Nqo13, could act as a two-state conformational switch, by interaction with Lys_{14–345} of Nqo14. The resulting uncompensated charge of Lys_{13–204} could therefore lead to proton transfer to Glu_{13–377}. Quantum chemical model calculations of these processes further support that the conformational state of the Lys/Glu ion pair indeed strongly modulates the proton transfer energetics. Moreover, our principal component analysis of global dynamics in the membrane domain of complex I suggests that the ion-pair dynamics is also linked with channel hydration and *inter*-subunit couplings. We observe similar effects as described for Nqo13, also in the other antiporter-like subunits. Our combined results suggest that a combination of conformational and electrostatic switching provide an important functional principle to achieve an *action-at-a-distance* effect in complex I.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbabo.2018.06.001>.

Transparency document

The [Transparency document](#) associated with this article can be found in online version.

Acknowledgements

This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program/grant agreement 715311. The Leibniz-Rechenzentrum (LRZ), SuperMuc (projects: pr48de and pr27xu) provided computational resources.

References

- [1] M. Wikström, Two protons are pumped from the mitochondrial matrix per electron transferred between NADH and ubiquinone, *FEBS Lett.* 169 (2) (1984) 300–304.
- [2] J. Hirst, Mitochondrial complex I, *Annu. Rev. Biochem.* 82 (1) (2013) 551–575.
- [3] U. Brandt, Energy converting NADH: ubiquinone oxidoreductase (complex I), *Annu. Rev. Biochem.* 75 (1) (2006) 69–92.
- [4] L.A. Sazanov, A giant molecular proton pump: structure and mechanism of respiratory complex I, *Nat. Rev. Mol. Cell Biol.* 16 (6) (2015) 375–388.
- [5] A.J.Y. Jones, J.N. Blaza, F. Varghese, J. Hirst, Respiratory complex I in *Bos taurus* and *Paracoccus denitrificans* pumps four protons across the membrane for every NADH oxidised, *J. Biol. Chem.* 292 (12) (2017) 4987–4995.
- [6] A.S. Galkin, V.G. Grivennikova, A.D. Vinogradov, H⁺/2e⁻ stoichiometry in NADH-quinone reductase reactions catalyzed by bovine heart submitochondrial particles, *FEBS Lett.* 451 (2) (1999) 157–161.
- [7] P. Mitchell, Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism, *Nature* 191 (4784) (1961) 144–148.
- [8] M. Yoshida, E. Muneyoki, T. Hisabori, ATP synthase — a marvelous rotary engine of the cell, *Nat. Rev. Mol. Cell Biol.* 2 (9) (2001) 669–677.
- [9] R. Baradaran, J.M. Berrisford, G.S. Minhas, L.A. Sazanov, Crystal structure of the entire respiratory complex I, *Nature* 494 (7438) (2013) 443–448.
- [10] V. Zickermann, C. Wirth, H. Nasiri, K. Siegmund, H. Schwalbe, C. Hunte, U. Brandt, Mechanistic insight from the crystal structure of mitochondrial complex I, *Science* 347 (6217) (2015) 44–49.
- [11] J. Zhu, K.R. Vinothkumar, J. Hirst, Structure of mammalian respiratory complex I, *Nature* 536 (7616) (2016) 354–358.
- [12] K. Fiedorczuk, J.A. Letts, G. Degliesposti, K. Kaszuba, M. Skehel, L.A. Sazanov, Atomic structure of the entire mammalian mitochondrial complex I, *Nature* 538 (7625) (2016) 406–410.
- [13] J. Carroll, I.M. Fearnley, J.M. Skehel, R.J. Shannon, J. Hirst, J.E. Walker, Bovine complex I is a complex of 45 different subunits, *J. Biol. Chem.* 281 (43) (2006) 32724–32727.
- [14] K.R. Vinothkumar, J. Zhu, J. Hirst, Architecture of mammalian respiratory complex I, *Nature* 515 (7525) (2014) 80–84.
- [15] C. Wirth, U. Brandt, C. Hunte, V. Zickermann, Structure and function of mitochondrial complex I, *Biochim. Biophys. Acta* 1857 (7) (2016) 902–914.
- [16] K. Kmita, V. Zickermann, Accessory subunits of mitochondrial complex I, *Biochem. Soc. Trans.* 41 (5) (2013) 1272–1279.
- [17] E. Maklashina, A.B. Kotlyar, G. Cecchini, Active/de-active transition of respiratory

- complex I in bacteria, fungi and animals, *Biochim. Biophys. Acta* 1606 (1) (2003) 95–103.
- [18] A.D. Vinogradov, Catalytic properties of the mitochondrial NADH-ubiquinone oxidoreductase (complex I) and the pseudo-reversible active/inactive enzyme transition, *Biochem Biophys Acta* 1364 (2) (1998) 169–185.
- [19] S. Dröse, A. Stepanova, A. Galkin, Ischemic A/D transition of mitochondrial complex I and its role in ROS generation, *Biochim. Biophys. Acta* 1857 (7) (2016) 946–957.
- [20] A. Di Luca, V.R.I. Kaila, Global collective motions in the mammalian and bacterial respiratory complex I, *Biochim. Biophys. Acta* 1859 (5) (2018) 326–332.
- [21] C. Hunte, V. Zickermann, U. Brandt, Functional modules and structural basis of conformational coupling in mitochondrial complex I, *Science* 329 (5990) (2010) 448–451.
- [22] L.A. Sazanov, P. Hinchliffe, Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*, *Science* 311 (5766) (2006) 1430–1536.
- [23] M.L. Verkhovskaya, N. Belevich, L. Euro, M. Wikström, Real-time electron transfer in respiratory complex I, *Proc. Natl. Acad. Sci. U. S. A.* 105 (10) (2008) 3763–3767.
- [24] S. De Vries, K. Dörner, M.J.F. Strampraad, T. Friedrich, Electron tunneling rates in respiratory complex I are tuned for efficient energy conversion, *Angew. Chem. Int. Ed.* 54 (2015) 2844–2848.
- [25] M. Verkhovskaya, 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 (2) (2013) 491–511.
- [26] V. Sharma, G. Belevich, A.P. Gamiz-Hernandez, T. Róg, I. Vattulainen, M.L. Verkhovskaya, M. Wikström, G. Hummer, V.R.I. Kaila, Redox-induced activation of the proton pump in the respiratory complex I, *Proc. Natl. Acad. Sci. U. S. A.* 112 (37) (2015) 11571–11576.
- [27] A.P. Gamiz-Hernandez, A. Jussupow, M.P. Johansson, V.R.I. Kaila, Terminal electron-proton transfer dynamics in the quinone reduction of respiratory complex I, *J. Am. Chem. Soc.* 139 (45) (2017) 16282–16288.
- [28] M.A. Tocilescu, U. Fendel, K. Zwicker, S. Dröse, S. Kerscher, U. Brandt, The role of a conserved tyrosine in the 49-kDa subunit of complex I for ubiquinone binding and reduction, *Biochim. Biophys. Acta* 1797 (22) (2010) 625–632.
- [29] P.K. Sinha, N. Castro-Guerrero, G. Patki, M. Sato, J. Torres-Bacete, S. Sinha, H. Miyoshi, A. Mitsuno-Yagi, T. Yagi, Conserved amino acid residues of the NuoD segment important for structure and function of *Escherichia coli* NDH-1 (complex I), *Biochemistry* 54 (3) (2015) 753–764.
- [30] J.G. Fedor, A.J.Y. Jones, A. Di Luca, V.R.I. Kaila, J. Hirst, Correlating kinetic and structural data on ubiquinone binding and reduction by respiratory complex I, *Proc. Natl. Acad. Sci. U. S. A.* 114 (48) (2017) 12737–12742.
- [31] R.G. Efremov, L.A. Sazanov, Structure of the membrane domain of the respiratory complex I, *Nature* 465 (7297) (2011) 414–420.
- [32] V.R.I. Kaila, M. Wikström, G. Hummer, Electrostatics, hydration and proton transfer dynamics in the membrane domain of respiratory complex I, *Proc. Natl. Acad. Sci. U. S. A.* 111 (19) (2014) 6988–6993.
- [33] A. Di Luca, A.P. Gamiz-Hernandez, V.R.I. Kaila, Symmetry-related proton transfer pathways in respiratory complex I, *Proc. Natl. Acad. Sci. U. S. A.* 114 (31) (2017) E6314–E6321.
- [34] O. Haapanen, V. Sharma, Role of water and protein dynamics in proton pumping by respiratory complex I, *Sci. Rep.* 7 (7747) (2017) 1–12.
- [35] E. Screpanti, C. Hunte, Discontinuous membrane helices in transport proteins and their correlation with function, *J. Struct. Biol.* 159 (2007) 261–267.
- [36] L.R. Forrest, et al., Mechanism for alternating access in neurotransmitter transporters, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 10338–10343.
- [37] J. Torres-Bacete, E. Nakamaru-Ogiso, A. Matsuno-Yagi, T. Yagi, Characterization of the NuoM (ND4) subunit in *Escherichia coli* NDH-1: conserved charged residues essential for energy-coupled activities, *J. Biol. Chem.* 282 (51) (2007) 36914–36922.
- [38] M.C. Kao, S. Di Bernardo, M. Perego, E. Nakamaru-Ogiso, A. Matsuno-Yagi, T. Yagi, Functional role of four conserved charged residues on the membrane domain subunit NuoA of the proton-translocating NADH-quinone oxidoreductase from *Escherichia coli*, *J. Biol. Chem.* 279 (31) (2004) 32360–32366.
- [39] J. Torres-Bacete, P.K. Sinha, N. Castro-Guerrero, A. Matsuno-Yagi, T. Yagi, Features of subunit NuoM (ND4) subunit in *Escherichia coli* NDH-1: topology and implication of conserved Glu144 for coupling site 1, *J. Biol. Chem.* 284 (48) (2009) 33062–33069.
- [40] E. Nakamaru-Ogiso, M.C. Kao, H. Chen, S.C. Sinha, T. Yagi, T. Ohnishi, The membrane subunit NuoL (ND5) is involved in the indirect proton pumping mechanism of *Escherichia coli* Complex I, *J. Biol. Chem.* 285 (50) (2010) 39070–39078.
- [41] J. Michel, J. Deleon-Rangel, S. Zhu, K. Van Ree, S.B. Vik, Mutagenesis of the L, M, and N subunits of complex I from *Escherichia coli* indicates a common role in function, *PLoS One* 6 (2) (2011) e17420.
- [42] L. Euro, G. Belevich, M.I. Verkhovskaya, M. Wikström, M. Verkhovskaya, Conserved lysine residues of the membrane subunit NuoM are involved in energy conversion by the proton-pumping NADH:ubiquinone oxidoreductase (complex I), *Biochim. Biophys. Acta* 1777 (9) (2008) 1166–1172.
- [43] M. Wikström, G. Hummer, Stoichiometry of the proton translocation by respiratory complex I and its mechanistic implications, *Proc. Natl. Acad. Sci. U. S. A.* 109 (12) (2012) 4431–4436.
- [44] M. Wikström, V. Sharma, V.R.I. Kaila, J.P. Hosler, G. Hummer, New perspectives on the proton pumping in cellular respiration, *Chem. Rev.* 115 (5) (2015) 2196–2221.
- [45] U. Brandt, A two-state stabilization-change mechanism for proton-pumping complex I, *Biochim. Biophys. Acta* 1807 (10) (2011) 1364–1369.
- [46] V.R.I. Kaila, Long-range proton-coupled electron transfer in biological energy conversion: towards mechanistic understanding of respiratory complex I, *J. Roy. Soc. Interfaces* 14 (15) (2018) 20170916.
- [47] O.S. Smart, J.G. Neduveilil, X. Wang, B.A. Wallace, M.S.P. Sansom, HOLE: a program for the analysis of the pore dimensions of ion channel structural models, *J. Mol. Graph.* 14 (6) (1996) 354–360.
- [48] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N -log(N) method for Ewald sums in large systems, *J. Chem. Phys.* 98 (12) (1993) 10089–10092.
- [49] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, K. Laxmikant, K. Schulten, Scalable molecular dynamics with NAMD, *J. Comput. Chem.* 26 (16) (2005) 1781–1802.
- [50] 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. 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, M. Karplus, All atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* 102 (18) (1998) 3586–3616.
- [51] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell Jr., R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation of six lipid types, *J. Phys. Chem. B* 114 (23) (2010) 7830–7843.
- [52] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nano-systems: application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 10037–10041.
- [53] G. Kieseritzky, E.W. Knapp, Optimizing pKa computation in proteins with pH adapted conformations, *Proteins* 71 (3) (2008) 1335–1348.
- [54] I. Bahar, T.R. Lezon, A. Bakan, I.H. Shrivastava, Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins, *Chem. Rev.* 110 (3) (2010) 1463–1497.
- [55] A.W. Van Wynsberghe, Q. Cui, Interpreting correlated motions using normal mode analysis, *Structure* 14 (1) (2006) 1647–1653.
- [56] A. Bakan, L.M. Meireles, I. Bahar, Prody: protein dynamics inferred from theory and experiments, *Bioinformatics* 27 (11) (2011) 1575–1577.
- [57] A.D. Becke, Density-functional thermochemistry. III. The role of exact-exchange, *J. Chem. Phys.* 98 (7) (1993) 5648–5652.
- [58] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B Condens. Matter* 37 (2) (1988) 785–789.
- [59] F. Weigend, 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 (18) (2005) 3297–3305.
- [60] S. Grimme, J. Antony, S. Ehrlich, H.J. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, *J. Chem. Phys.* 132 (15) (2010) 154104.
- [61] A. Klamt, Schüürmann, COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, *J. Chem. Soc. Perkin Trans. 2* (1993) 799–805.
- [62] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, Electronic structure calculations on workstation computers: the program system Turbomole, *Chem. Phys. Lett.* 162 (3) (1989) 165–169.

Article II

Water-Gated Proton Transfer Dynamics in Respiratory Complex I

Max E. Mühlbauer,^{||} Patricia Saura,^{||} Franziska Nuber, Andrea Di Luca, Thorsten Friedrich, and Ville R. I. Kaila*



Cite This: *J. Am. Chem. Soc.* 2020, 142, 13718–13728



Read Online

ACCESS |



Metrics & More

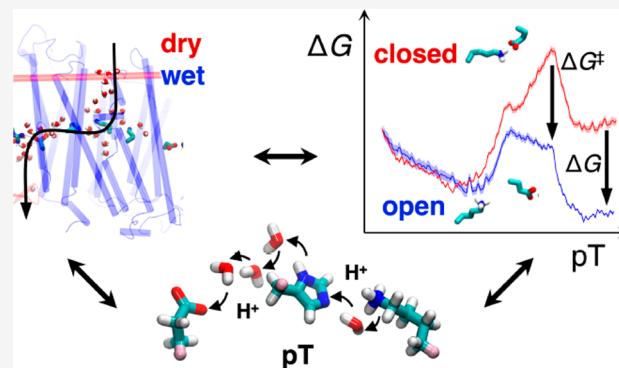


Article Recommendations



Supporting Information

ABSTRACT: The respiratory complex I transduces redox energy into an electrochemical proton gradient in aerobic respiratory chains, powering energy-requiring processes in the cell. However, despite recently resolved molecular structures, the mechanism of this gigantic enzyme remains poorly understood. By combining large-scale quantum and classical simulations with site-directed mutagenesis and biophysical experiments, we show here how the conformational state of buried ion-pairs and water molecules control the protonation dynamics in the membrane domain of complex I and establish evolutionary conserved long-range coupling elements. We suggest that an electrostatic wave propagates in forward and reverse directions across the 200 Å long membrane domain during enzyme turnover, without significant dissipation of energy. Our findings demonstrate molecular principles that enable efficient long-range proton–electron coupling (PCET) and how perturbation of this PCET machinery may lead to development of mitochondrial disease.



INTRODUCTION

With up to 45 subunits and a molecular mass of 1.0 MDa, the respiratory complex I (NADH:ubiquinone oxidoreductase) is one of the largest and most complicated membrane-bound proteins.^{1–4} This gigantic enzyme functions as the initial electron acceptor in respiratory chains of mitochondria and aerobic bacteria by transferring electrons between nicotinamide adenine dinucleotide (NADH, $E_m = -320$ mV) and ubiquinone (Q, $E_m = +90$ mV), and the 0.8 V redox span is employed to pump four protons across the membrane.⁵ The established proton motive force (pmf) is further used for active transport and synthesis of adenosine triphosphate (ATP) by the rotary motion of F_0F_1 -ATP synthase.^{6,7}

In recent years, several structures of complex I from different organisms have been resolved,^{8–11} showing that the electron transfer (eT) and proton pumping machineries are separated into unique domains of the enzyme. The eT module in the hydrophilic domain extends up to 100 Å above the membrane plane, whereas proton pumping is catalyzed by the 200 Å long membrane domain (Figure 1).^{8,9} The complete charge transfer (CT) process thus spans nearly 300 Å, but despite its large spatial extent, the process is fully reversible and strongly coupled. Consistent with the principle of microscopic reversibility, perturbations at the far end of the pumping module result in decreased Q oxidoreduction activity.^{12–17} Although molecular principles of this remarkable long-range proton-coupled electron transfer (PCET) process have recently started to emerge,^{1–4,8–11,18–21} the overall mechanism

still remains one of the most complex unsolved questions in biochemistry (cf. ref 22 and references therein for definition and discussion of both short- and long-range PCET).

Structural data in combination with computational and biophysical experiments provide a starting point to derive a molecular understanding of the functional principles in complex I.^{1–4,8–10,23–30} Recent studies suggest that proton pumping is triggered by electrostatic and conformational transitions resulting from the reduction of quinone,^{23,28} bound around 20 Å above the membrane plane in a protein cavity near the N2 iron–sulfur (FeS) center (Figure 1A).³⁰ The thermodynamic driving force ($\Delta G = 0.8$ V) for the pumping process is likely to originate from motion and coupled conformational transitions of the Q-species from a low potential site near the FeS chain (–300 mV) to a second, transient Q-binding site in the membrane domain with a redox potential most likely close to that of Q in membranes (+90 mV),^{1,30–32} which still remains to be experimentally validated.

The membrane domain comprises the antiporter-like subunits Nqo12/13/14 (*T. thermophilus* nomenclature) that contain a 200 Å long array of buried charged residues located

Received: March 13, 2020

Published: July 9, 2020



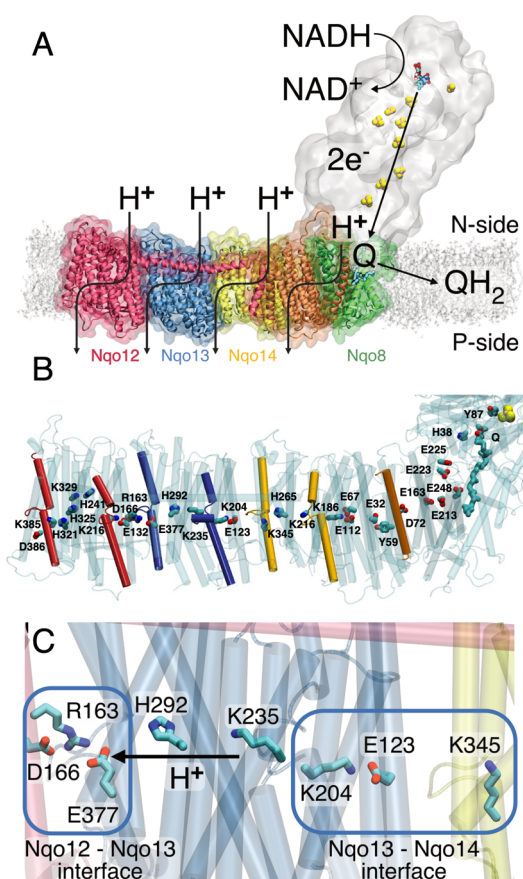


Figure 1. (A) Complex I catalyzes electron transfer from NADH to Q in its hydrophilic domain (in white) and pumps protons across the Nqo12 (red), Nqo13 (blue), and Nqo14 (yellow), as well as in the Nqo8 (green) and Nqo7/10/11 (in orange) subunits of the membrane domain. (B) An axis of conserved charged and polar residues spans the membrane domain from the Q channel to the terminal Nqo12 subunit. (C) Nqo13 is connected to its neighboring subunits by charged interfaces (blue squares) that enable formation of ion-pairs between the subunits. The lateral pT from Lys235₁₃ via His292₁₃ to Glu377₁₃ is marked with an arrow.

at the center of the membrane and provide key elements in the pumping machinery (Figure 1B). Molecular simulations^{1,25,27,29,33} suggest that proton channels are established around broken helix motifs TM7a/b and TM12a/b within five-helical transmembrane (TM) bundle segments of the antiporter-like subunits, with several residues along these pathways indirectly supported by site-directed mutagenesis experiments.^{12–17} The proton transfer (pT) involves hydration changes within TM4–8 and TM9–13 in Nqo12/13/14, and indications of a fourth proton channel have been found near Nqo8/Nqo10 that also involves the TM2–6 bundle segment of Nqo8.^{8,10,27} The proton channels start from the negatively charged side (N-side) of the membrane and connect via a horizontal axis along the membrane plane to output sites at the positively charged side (P-side) of the membrane. Recent data^{25,27} suggest that the pumping process is mediated by a central lysine residue (Lys216₁₄/Lys235₁₃/Lys329₁₂) that pushes the protons laterally toward a terminal lysine (Lys345₁₄/Lys385₁₂) in Nqo12/14 or a glutamate (Glu377₁₃) residue in Nqo13, with pT reactions that are controlled by the conformational state of the conserved Lys/Glu (Asp) ion-pair at the interface of each subunit (Figure

1).^{1,25,27} Long-range proton pumping was suggested to propagate by an electrostatic wave in forward and reverse pulses across the membrane domain.¹ It was also suggested that complex I could drive pumping by direct conformational changes, e.g., along the 100 Å horizontal helix connecting Nqo12 to Nqo14^{8,26} or β -sheet motifs between subunits.^{4,8,26} Site-directed mutagenesis experiments^{34–36} do not currently support such “steam engine/piston”-models, whereas local conformational/electrostatic coupling between the Q site and Nqo8 has indirectly been supported by structural and biochemical studies.^{25,28}

Our previous work has shown that complex I employs proton pathways in symmetry-related locations^{25,27,29} to transfer protons across the membrane, with both hydration changes^{27,29} and ion-pair dynamics^{25,27} forming important elements for the long-range charge transfer process. Here we elucidate how the ion-pair dynamics determines both the wetting/dewetting transitions and proton transfer energetics, and *vice versa*, using both classical and hybrid quantum/classical (QM/MM) free energy simulations. We show how the hydration state of the antiporter-like subunit Nqo13 strongly modulates the protonation and ion-pair dynamics and further validate the findings by site-directed mutagenesis experiments, pumping measurements, and kinetic models.

RESULTS

Ion-Pair Conformation Is Determined by the Hydration State of the Proton Channels. To probe the coupling between the hydration state and the proton transfer reactions, we performed classical molecular dynamics (MD) simulations on complex I from *Thermus thermophilus* (PDB ID: 4HEA⁸) embedded in a lipid membrane environment. Consistent with previous simulations,^{25,27} the membrane domain spontaneously hydrates on 0.2 μ s time scales, and water molecules establish quasi-one-dimensional wires that connect the N-side bulk with buried charged and polar residues at the center of the antiporter-like subunits (Figure S1A). In Nqo13, the water molecules enter TM7a from the N-side solvent connecting to Lys235₁₃, from which the proton wire continues further via His292₁₃ to Glu377₁₃ of TM12a and to the P-side of the membrane (Figure S1A), with similar hydration patterns observed also in Nqo12 and Nqo14.^{25,27} The simulations support that the channel hydration strongly depends on the protonation state of the middle Lys235₁₃ (Figure S1B), consistent with previous data.^{25,27}

On the basis of selected snapshots of the hydration trajectory, we next computed free energy profiles for opening the central Lys/Glu ion-pair in Nqo13 (Lys204₁₃/Glu123₁₃) in a dry state ($N < 5$ water molecules), at intermediate channel hydration levels ($N \approx 10$ water molecules), and in a fully hydrated “wet state” of the channel ($N > 15$ water molecules) (Figure 2). The wet/dry transition is likely to represent key intermediate states in the pumping process of complex I, as supported by the steady state equilibrium reached during the MD and free energy simulations (Figures S1, S10). Rigorous grand-canonical hydration free energy calculations³⁷ are outside the scope of our large complex I simulations, with close to 1 million atoms. In the open ion-pair conformation, formed by shifting Glu123₁₃ to establish a new ion-pair with Lys345₁₄ (Figure 2B), Lys204₁₃ is stabilized by π -cation interactions with Tyr234₁₃/Trp213₁₃ (Figure S2B), which pushes Lys235₁₃ closer to His292₁₃ by ~ 1 Å (Figure S2D) and enabled by the flexibility of the broken helix TM7a/b.

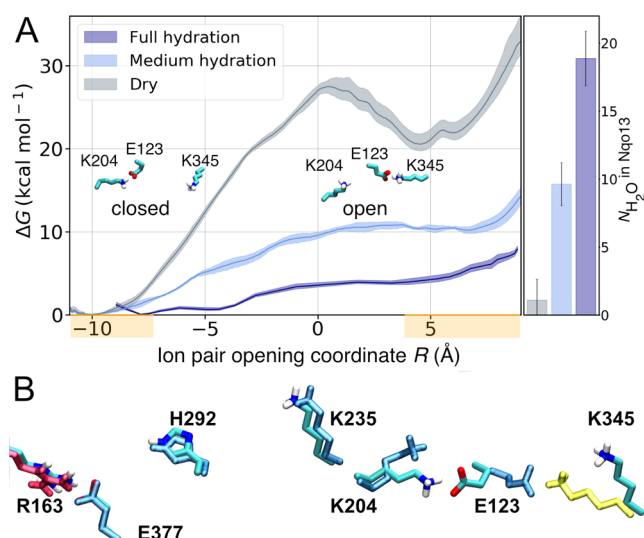


Figure 2. (A) Free energy profiles for opening the Lys204₁₃/Glu123₁₃ ion-pair at the Nqo13/Nqo14 interface in the fully hydrated state (dark blue), at medium hydration levels (light blue), and in the dry state (gray) of Nqo13. The Glu/Lys residues were modeled in their charged state (see Table S10). The ion-pair opening (shifting) reaction coordinate, R , is defined as $d(\text{E123}_{13}\text{-K204}_{13}) - d(\text{E123}_{13}\text{-K345}_{14})$. Closed ion-pair $R \sim (-9 \pm 2)$ Å and open ion pair $R \sim (+5 \pm 2)$ Å are marked in orange on the R -axis. Ion-pair opening is favored by the increased hydration state of the proton channel. (B) In the closed conformation, the ion-pair interaction is established between Lys204₁₃ and Glu123₁₃, whereas in the open conformation, Glu123₁₃ forms a contact with Lys345₁₄ (in yellow).

However, despite key conformational changes at the local side-chain level, the average (*maximum*) root-mean square deviation on the protein backbone is <1.5 Å (<2 Å) (Figure S2C), suggesting that subtle conformational changes are sufficient to support the switching dynamics.

In the dry state, our replica-exchange umbrella sampling (REUS) calculations result in a high dissociation free energy barrier of Lys204₁₃/Glu123₁₃ ion-pair of 27 kcal mol⁻¹ with an open state free energy of >20 kcal mol⁻¹ above the closed ion-pair configuration (Figure 2) that is both kinetically and thermodynamically expected to be blocked during complex I turnover on the millisecond time scale.^{1–4} Interestingly, the free energy barrier drops to ~ 10 kcal mol⁻¹ at intermediate hydration levels, whereas in the fully hydrated state, the ion-pair opening barrier is only ~ 4 kcal mol⁻¹ (Figure 2). This suggests that the opening dynamics becomes kinetically accessible in the nano- to-microsecond time scales upon channel wetting but is kinetically prevented when the proton channels are dry.

To further validate the computed free energies, we performed unbiased MD simulations of the fully hydrated state, mimicking states prior to and after the pT reaction. When Lys235₁₃ is protonated, the electrostatic repulsion between the middle lysine and Lys204₁₃ rapidly closes the Lys204₁₃/Glu123₁₃ ion-pair. However, in stark contrast, when the proton is moved from the middle lysine to Glu377₁₃, the ion-pair remains open for at least 100 ns (Figure S3A,B), further supporting the tight coupling between the protonation and ion-pair dynamics.

Ion-Pair Dynamics Triggers Proton Transfer. Having established that the channel hydration controls the ion-pair dynamics, we next probed how the pT reaction itself is affected

by the conformational state of the ion-pair. As classical simulations cannot capture bond-formation/breaking processes, we employed density functional theory (DFT)-based QM/MM molecular dynamics simulations and free energy calculations to elucidate reaction barriers and thermodynamic driving forces for the pT reaction (see Materials and Methods). In the QM/MM calculations, we opened up a reactive region of nearby polar protein residues and water molecules connecting Lys235₁₃ via His292₁₃ to Glu377₁₃ while electrostatically coupling the system to the surrounding protein (Figure 3A). When the ion-pair is in the closed conformation, we obtain an overall endergonic free energy profile for the pT to Glu377₁₃ with a barrier of ~ 10 kcal mol⁻¹, whereas dissociation of the ion-pair leads to an exergonic pT reaction of approximately -2 kcal mol⁻¹, and the reaction barrier drops to ~ 4 kcal mol⁻¹, making the overall reaction accessible to nanosecond time scales (Figure 3C). This suggests that the formation of the open ion-pair (Glu123₁₃/Lys204₁₃) conformation rather than pT between Lys235₁₃ and Glu377₁₃ itself becomes rate-limiting for the overall process. The dynamics observed in our unbiased QM/MM MD simulations is consistent with the calculated free energy profiles, with the proton remaining bound to the initial Lys235₁₃/His292₁₃ bridge in the closed ion-pair conformation, whereas spontaneous pT to Glu377₁₃ is observed in selected snapshots of the open ion-pair state (Figure 3D). The free energy barrier for the dry-state pT reaction is >20 kcal mol⁻¹ in both closed and open ion-pair conformations (Figure S3D), suggesting that the pT reaction is water-gated.

Perturbing the Proton Pumping Machinery. To perturb the lateral pT process in Nqo13, we mutated the bridging His292₁₃ that provides an important intermediate site in the pT process into an alanine residue, as well as the homologous His322_M and another bridging histidine, His348_M of the NuoM subunit in *E. coli* complex I²⁶ (Ser318₁₃, in the *T. thermophilus*, Figure S9; see Figure S11 and Table S11 for comparison of structure and sequence of the two organisms). The mutagenesis studies were feasible only for *E. coli* complex I. The proton pumping subunits share a $\sim 30\%$ sequence similarity but have a high structural similarity and are thus expected to employ similar pumping mechanisms (see Table S12 and Figure S9). His348_M stabilizes the proton wire in MD simulations of the *E. coli* complex I (Figure S4C). In classical MD simulations of the H292A₁₃ and H322A_M/H348A_M variants, the introduced substitution results in a long proton wire with $N > 5$ water molecules relative to 2–3 bridging water molecules between the lysine and glutamate in the wild type (WT) system (Figure 4A, Figure S4A–F). The water wire is partially disrupted by the histidine substitutions, leading to an overall lowered hydrogen-bonded connectivity relative to wild type model (Figure 4B, Figure S4). The substitution also results in conformational changes that could electrostatically modulate the proton affinity of the proton acceptor Glu377₁₃ (Figure 4A, Figure S4), in particular by shorter distances to Arg163₁₂ of Nqo12 (Arg175_L in *E. coli*, Figure 4A,B), which could electrostatically tune the pK_a of the Glu377₁₃.

In QM/MM simulations of the H292A₁₃ variant, we do not observe spontaneous deprotonation from Lys235₁₃, resembling the pT behavior in the dry state (Figure 4C), with drastically increased pT barrier in DFT models (Figure S5). The perturbed barriers are also supported in DFT models constructed based on the *E. coli* complex I simulations (Figure S12). The simulations on the *E. coli* complex I support that

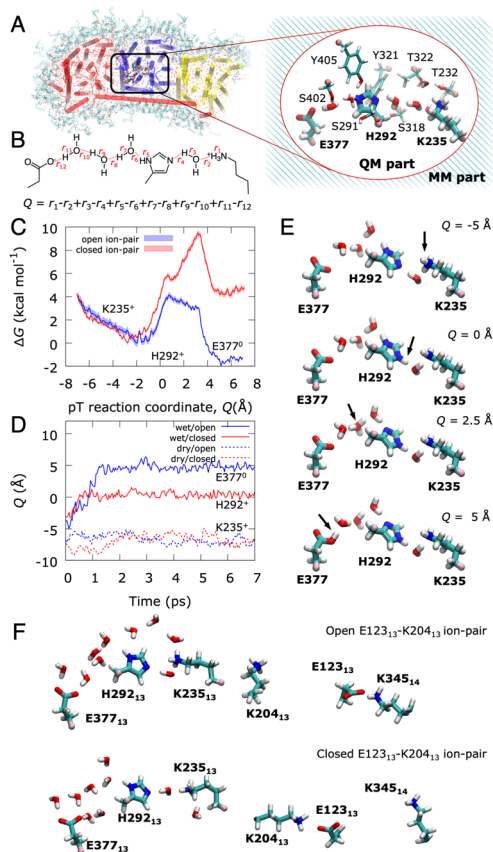


Figure 3. Free energy and dynamics of lateral proton transfer in Nqo13. Prior to the pT reaction, all Glu/Lys residues were modeled in their charged states (see Table S10). (A) QM/MM models for the pT reaction in the Nqo13 subunit, showing the atoms in the QM region (inset). (B) The proton transfer (pT) reaction coordinate (Q) is represented by a linear combination of bond breaking and bond forming distances. (C) QM/MM free energy profiles for the lateral pT from Lys235₁₃ via His292₁₃ to Glu377₁₃ in the fully hydrated state with closed (in red) and open (in blue) ion-pair conformations, respectively. (D) Unbiased QM/MM MD simulations for the pT process. In the medium-hydrated state (red and blue dashed lines), the pT stalls at Lys235₁₃ in both open and closed ion-pair conformations. In the fully hydrated state with a closed ion-pair, the proton equilibrates between Lys235₁₃ and His292₁₃ within ~ 1 ps (solid red line) but does not progress onward during the remaining simulation. In stark contrast, the ion-pair opening induces full pT to Glu377₁₃ within ~ 2 ps (solid blue line). (E) Snapshots of the QM/MM free energy profiles showing the transferred proton on Lys235₁₃ (top), His292₁₃ (second panel), at an intermediate Zundel ion (third panel), and Glu377₁₃ (bottom). See Movie S1 for animation of the pT dynamics. (F) Snapshots of the open ion-pair conformation (top, with intact Glu123₁₃/Lys345₁₄ ion-pair) and the closed ion-pair conformation (bottom, intact Lys204₁₃/Glu123₁₃ ion-pair).

both histidine residues provide important gating residues for the pT process (Figure 4C–F, Figure S12).

To experimentally assess the effect of a perturbed pT reaction, we introduced both single (H322A_M, H348A_M) and the double substitutions (H322A_M/H348A_M) into NuoM of *E. coli* complex I. Figure 4D and Figure 4E show pumping experiments, where we monitored the proton translocation of complex I reconstituted into proteoliposomes. To this end, we used both an ACMA fluorescence assay that is quenched as a response to pT across the membrane (ΔpH), as well as by monitoring absorbance changes using the membrane-potential

($\Delta\Psi$) sensitive dye, oxonol-VI, providing an independent measure of the pumping process. The absorbance signal is linear to the $\Delta\Psi$, whereas the fluorescence signal has a more complex dependence on the established ΔpH .³⁸ The $\Delta\text{pH}/\Delta\Psi$ generation for both the single and double mutant variants was measured and compared to that of wild type complex I. All variants were fully assembled, stable and active, and reconstituted with similar orientations into liposomes (Figure S6, Tables S1 and S2). On the basis of the ACMA quenching, the histidine substitutions lead to a significant decrease in the proton pumping activity by approximately 40% for the double mutant (Figure 4D,E, Table 1, Table S2), whereas the individual substitutions have a smaller effect of 10–30%. The MD simulations suggest that the protons might take alternative pathways in the single variants to partially rescue the perturbed wiring (Figure 4C–F). A similar but slightly stronger effect on the $\Delta\Psi$ generation was inferred by monitoring absorbance changes with the oxonol assay, where the double mutant is reduced to 40% of the WT, and the single mutants show a residual absorbance of 61% and 74%, respectively (Figure 4D,E, Table S2). Part of the difference might arise from proton leaks in the vesicles, indicative of the initial burst phase followed by constant decrease in the signal (Figure 4D), in addition to the nonlinearity of the fluorescence signal with the proton concentration. However, the proteoliposome background leak levels, assembly, and orientation of proteoliposomes containing the WT and mutant variants are similar. Our data thus show that the characterized variants, with mutations around 100 Å from the electron transfer module, result in lower electron transfer rates and reduced proton pumping (Figure S6, Supporting Information Methods). The histidine substitutions result in a significant decrease of the oxidoreduction activity up to 52% for the double variant, whereas a 20–40% reduction in activity is observed with the single substitutions (Figure 4E, Table 1). The experiments thus support that the simulated proton transfer reactions in Nqo13 (NuoM) could affect complex I activity.

Proton Pumping Kinetics. To better understand the role of the histidine residues in proton pumping, we constructed a kinetic model of the pumping process using a kinetic master equation approach, with relative barriers derived from our atomistic simulations (Figure 4F,G, Figure S7, Tables S3 and S4; see Materials and Methods). To this end, we modeled the Glu/Lys ion-pair and the pumping elements in Nqo13 with proton acceptor/donor sites, mimicking the middle lysine residue (Lys235₁₃), connected to a proton acceptor site (Glu377₁₃). We coupled the ion-pair opening and the pT transfer reaction with relative rates similar to the wet/dry states transition and induced opening of the ion-pair by applying a nonequilibrium external force (see Materials and Methods). The kinetic model predicts that one proton is taken up from the N-side and pumped across to the P-side (Figure 4F,G) when the coupling between the ion-pair and the middle lysine residue reaches a threshold of >5 kcal mol⁻¹. To mathematically model the effect of the histidine substitutions, we next slowed down the pT rate along the channel (k_{pT}). Interestingly, when the pT rate in the channel becomes competing with pT uptake/release rates to the bulk, we observed a strong lowering of the proton pumping stoichiometry, consistent with our experimental observations. The pumping is also sensitive to the $\text{p}K_{\text{a}}$ of the middle lysine, with efficient proton translocation when the $\text{p}K_{\text{a}}$ resides in the 10 ± 2 range, within the optimized parameter set. The kinetic

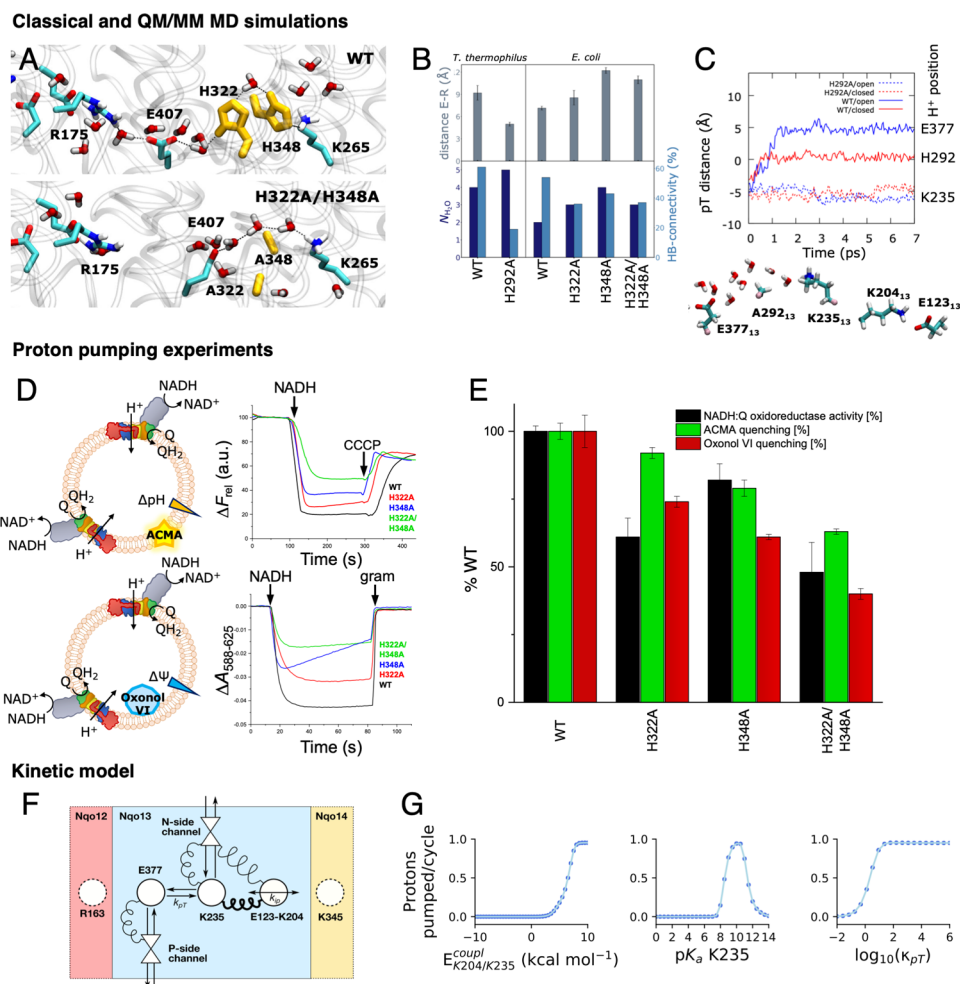


Figure 4. Perturbing the proton transfer dynamics in complex I. (A) Snapshots of pT wires in the WT (top) and H322A/H348A variant (bottom) of *E. coli* complex I after 100 ns MD simulations (see Figure S4). (B) Substitution of the bridging histidine residues leads to perturbation in the Glu377₁₃/Arg163₁₂ (Glu407_M/Arg175_L) and Lys235₁₃/Glu377₁₃ (Lys265_M/Glu407_M) distances in classical MD simulations of *T. thermophilus* (T) and *E. coli* (E) complex I. Additional water molecules close the gap between the lysine and glutamate residues but with a lowered occupancy relative to the WT. (C) pT dynamics from QM/MM MD simulations of the WT and H292A variant (top) and snapshot of the H292A variant, showing the position of the pT wire and the Lys204₁₃-Glu123₁₃ ion-pair in closed conformation (bottom). (D) ACMA (top) and oxonol-VI (bottom) assays from proton pumping experiments in *E. coli* complex I. The ACMA and oxonol-VI assays monitor the ΔpH and $\Delta\Psi$ gradient across the proteoliposome membrane, respectively (see also Figure S6D). (E) Summary of NADH:Q oxidoreduction activity, ACMA, and oxonol quenching experiments of WT and alanine variants. All measurements were performed in triplicates, with errors given as standard deviations. The NADH:Q oxidoreductase activity was determined from monitoring NADH decrease at 340 nm. See Supporting Information Methods and Figure S6 for further details. (F) Schematic representation of the proton pumping kinetic model of complex I with N- and P-side channels, Glu377₁₃, Lys235₁₃, and the ion-pair Glu123₁₃/Lys204₁₃ (see Materials and Methods). (G) Effect of the parameters of the kinetic model on the proton pumping across the membrane (see Materials and Methods). Left: coupling energy between ion-pair and Lys235₁₃. Center: the $\text{p}K_a$ of Lys235₁₃. Right: the K235/E377 intrinsic pT rate. See Supporting Information Methods and Figures S7, S13 for a detailed description of the kinetic model.

Table 1. Proton Pumping and NADH:Q Oxidoreductase Activity of Wild Type and NuOM Variants of Complex I^a

construct	pumping		activity
	ACMA (rel WT [%])	oxonol-VI (rel WT [%])	NADH:Q oxidoreductase (rel WT [%])
WT	100 ± 3	100 ± 6	100 ± 2
H322A	92 ± 2	74 ± 2	61 ± 7
H348A	79 ± 3	61 ± 1	82 ± 6
H322A/H348A	63 ± 1	40 ± 7	48 ± 11

^aThe values are shown relative to the wild type complex I. See Tables S1 and S2 for absolute values.

model thus further supports that the ion-pair opening modulates both kinetically and thermodynamically the protonation signal to effectively propagate across the membrane domain to achieve pumping.

DISCUSSION

Our combined findings suggest that proton pumping along the 200 Å long membrane domain in complex I is established by conformational switching of the conserved ion-pairs that enables lateral pT within each proton pumping module. The pT reactions are gated by water molecules, whereas the channel wetting is in turn determined by the protonation state of the middle lysine residues.^{12,25,27} We could show that the

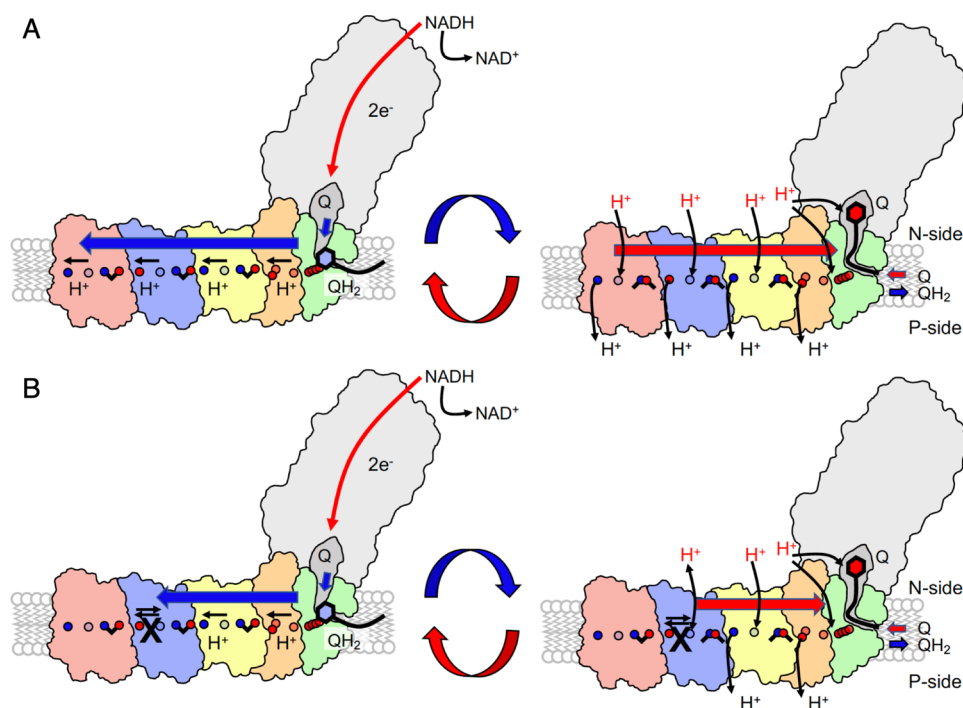


Figure 5. Putative proton pumping mechanism in complex I. (A) Left: reduction of quinone (Q) to quinol (QH₂) triggers stepwise proton transfer and ion-pair opening steps that propagate as a forward electrostatic pulse (blue arrow) to the terminal Nqo12 edge (in red) of the membrane domain. Right: The pumped protons are released by proton uptake from the N-side (H⁺ in red) and ion-pair closure during the backward signal (red arrow) to the membrane-bound Q binding site. Quinol/quinone exchange and reprotonation of the Q-site reset the pumping machinery for the next catalytic cycle. (B) Introduced mutations in Nqo13 (black cross) perturb the lateral pT from Nqo13 onward. The electrostatic backward pulse is reflected prematurely back from Nqo13 to the Q-channel, establishing pumping by Nqo14 (in yellow) and Nqo8/7/10/11 (green/orange) but at perturbed timing that affects the Q oxidoreduction activity.

ion-pair dissociation is kinetically accessible only when the proton channel is in a hydrated state but kinetically blocked in the dry state.

Our findings provide support for a recently proposed long-range pumping mechanism¹ in which pumping is established by forward and backward propagation of a protonation signal across the membrane domain (Figure 5A). In the proposed mechanism, the thermodynamic driving force for the pumping originates from quinone reduction that could trigger proton uptake and conformational changes at or near the Nqo8 subunit.^{23,28} After quinone motion to a putative second binding site within the membrane domain,³⁰ the free energy is transduced to the pumping module by pushing the proton toward the terminal end of the Nqo8/10/11 subunits, near the first ion-pair in Nqo14. This positive partial charge accumulates near the Nqo14 edge and triggers opening of the Glu-Lys ion-pair in Nqo14, pushing the proton laterally toward the Nqo13 interface. The charge imbalance further induces sequential ion-pair opening, hydration changes, and pT in Nqo13 and similarly in Nqo12. Upon ejection of a proton at the terminal end of Nqo12 to the P-side, a new proton is taken up from the N-side followed by ion-pair closure within the same subunit. This destabilizes the proton at the terminal edge of Nqo13, which is then ejected to the P-side. Reprotonation from the N-side and ion-pair closure at Nqo13 ejects the pumped proton at Nqo14 and a similar event could lead to ejection of the proton at the Nqo8/10/11 interface. Exchange of the quinol with a new quinone and reprotonation of the initial quinone reduction site lead to a new reaction cycle.

By partially blocking the pT elements in Nqo13, we created a complex I variant that is unable to effectively propagate the electrostatic pulse from the Nqo13 subunit onward, with terminal protons most likely ejected from the Nqo13/14 interface (Figure 5B). We expect that the Nqo14 and the Nqo8/10/11 modules can, nevertheless, still pump protons at near wild type activities in the mutant variants, thus resulting in roughly halved pT activity observed in our pumping experiments. The perturbed quinone oxidoreductase activity provides possible experimental support for the proposed electrostatic backwave¹ upon which the protons are ejected to the P-side of the membrane. The reduced activity could arise from slips in forward/reverse signals that partially propagate in two directions or in a too rapid backwave-signal to the putative second quinone binding site³⁰ that results in the uncoupling effects, although detailed experiments are still required to fully quantify the effects. Nevertheless, we argue that both the experimental validation and kinetic models provide important boundaries for the computational prediction and allow testing implication beyond the microscopic description of the proton pumping process. Interestingly, perturbed timing in protonation reactions has been shown to uncouple eT and pT steps also in cytochrome oxidase (CcO), another redox-driven proton pump in respiratory chains.³⁹ We have demonstrated here that water molecules, although not yet experimentally resolved in complex I, provide key elements in gating the proton translocation process by conformational switching and modulation of the kinetic proton transfer barriers. Similar electrostatically driven functional elements based on hydration dynamics⁴⁰ could be mechanistically important in charge transfer reactions of cytochrome oxidase,⁴¹

light-driven ion-pumps,⁴² photosystem II,⁴³ or other PCET-catalyzing enzymes.⁴⁴ Our findings illustrate how complex I enables an action-at-a-distance coupling of elementary electron and proton transfer processes by combining electrostatic and conformational switching with specific wetting-transitions of its proton channels. The findings provide a blueprint for understanding conserved mechanisms in the complex I superfamily and how mutations within this charge-transfer machinery may result in human disease.

MATERIALS AND METHODS

Molecular Dynamics Simulations. The X-ray structure of complex I from *T. thermophilus* (PDB ID: 4HEA⁸) was embedded in a POPC/POPE-membrane, solvated with TIP3P water, 150 mM NaCl (see Table S7), with protonation states based on Poisson–Boltzmann continuum electrostatic calculations and previous studies.²⁸ The system was modeled using the CHARMM36m force field,⁴⁵ with cofactor parameters derived from DFT calculations.^{27,28} The FeS centers were modeled in the oxidized states except for N2, which was kept reduced in all simulations. The simulation setups comprised about 900 000 atoms. MD simulations were performed in the NPT ensemble at 310 K and 1 bar pressure, using a 2 fs integration time step and modeling long-range electrostatics using the particle mesh Ewald (PME) method.⁴⁶ After initial minimization and equilibration, the systems were simulated for 200 ns to establish hydration of the water channels in the membrane domain. Simulations were also carried out with His292₁₃ substituted by an alanine and for the WT and NuoM-variants of the membrane domain of *E. coli* complex I (PDB ID: 3RKO²⁶), constructed using the same conditions as for *T. thermophilus* complex I. All classical MD simulations were performed using NAMD2.⁴⁷ See Table S7 for a summary of all performed classical simulations.

Classical Free Energy Calculations. Free energy profiles for the Lys/Glu ion-pair dissociation in Nqo13 were computed using replica exchange umbrella sampling (REUS). To this end, the center-of-mass difference between Lys204₁₃/Glu 123₁₃ (r_1) and Glu123₁₃/Lys345₁₄ (r_2) headgroups was used as reaction coordinate, $R_c = r_2 - r_1$. To generate initial structures for the 20 REUS windows, 10 ns steered molecular dynamics (SMD) were performed. The SMDs were started from a 100 ns relaxed MD simulation of complex I of the dry state or after 200 ns MD simulations of the hydrated state. A harmonic bias with a force constant of 10 kcal mol⁻¹ Å⁻¹ was applied on the R_c . The center of the harmonic potential was scanned from -9.5 Å to +9.5 Å during the SMD simulations, from which 20 windows with equidistant values of the R_c from -10 Å (closed) to +8 Å (open) were equilibrated for 1 ns each. A harmonic force constant of 2.5 kcal mol⁻¹ Å⁻¹ was used for the bias, and exchange between neighboring windows was attempted every 10 ps. Each REUS window was simulated for 36 ns (720 ns in total for each free energy profile). Free energy profiles were computed using the dynamic histogram analysis method (DHAM)⁴⁸ with a bin width of 0.01 Å. Previously calculated free energies²⁷ were also remapped on the R_c based on REUS simulations performed using the complementary r_1 reaction coordinate. Free energy barriers were converted into rates using transition state theory with a standard pre-exponential factor of 0.16 ps⁻¹, a reflection coefficient $\kappa = 1$ of and at $T = 310$ K. See Figure S8 for convergence of free energy simulations, Table S7 for summary of all performed classical simulations, and Table S10 for summary of key protonation states.

QM/MM Dynamics and Free Energies. Hybrid DFT-based QM/MM MD simulations on the pT in Nqo13 were performed based on structures selected from the classical MD simulations in different hydration states, with open and closed ion-pair conformations. The system was trimmed to ~50 000 atoms, including the Nqo12, Nqo13, and Nqo14 subunits, and lipid/water/ions in their vicinity (Figure 3A). The QM region comprised 125–137 atoms, including Lys235, His292, Glu377, Ser291, Ser318, Thr232, Thr322, Tyr321, Tyr405, Ser402 of Nqo13 and up to 10 water molecules obtained from the

MD simulations. Link-atoms were introduced between $C\alpha$ and $C\beta$ atoms of the protein residues. The QM region was described at the B3LYP-D3/def2-SVP level,^{49–52} and the MM region with the CHARMM36m force field.⁴⁵ The classically relaxed structures were optimized at the QM/MM level using the adopted-basis Newton–Raphson algorithm and allowing all QM and MM atoms around 10 Å of the QM region to relax. QM/MM potential energy surfaces of the pT reaction were computed by scanning the reaction coordinate Q , defined as a linear combination of all breaking and forming bond distances, between -7.0 Å and +7.0 Å, with a harmonic restraint of 3000 kcal mol⁻¹ Å⁻² on Q . QM/MM free energy profiles were calculated at the B3LYP-D3/def2-SVP level^{49–52} using umbrella sampling (US)/weighted histogram analysis method (WHAM), with 58 windows spanning Q and a harmonic restraint of 100 kcal mol⁻¹ Å⁻². Each window was simulated for 1.6 ps at $T = 310$ K using a 1 fs integration time step, with total sampling of ~93 ps for each free energy profile. The pT reaction was also studied using unbiased QM/MM MD simulations starting from selected snapshots of wet and dry states and for the H292A mutant. All QM/MM calculations were performed with the CHARMM/TURBOMOLE interface.⁵³ See Figure S8 for convergence of free energy simulations and Table S8 for summary of all performed QM/MM simulations.

DFT Energy Profiles. DFT models were built based on MD snapshots from the fully hydrated trajectories of the WT and the H292A variant. The models included residues Lys235, His292/Ala292, Glu377, Ser291, Ser318, Thr232, Thr322, Tyr321, Tyr405, Ser402, Ala288, Met293, Leu314, Val399 of Nqo13 subunit and $N = 10$ (WT)/13 (H292A) water molecules, comprising 179–181 atoms. $C\alpha$ – $C\beta$ positions were cut and saturated with hydrogen atoms, with $C\beta$ positions fixed during geometry optimizations to simulate the protein framework. Geometry optimizations were performed at the B3LYP-D3/def2-SVP level, using an implicit polarizable medium with $\epsilon = 4$ to model the protein environment,^{49–52,54,55} with single point energies computed at the B3LYP-D3/def2-TZVP/ $\epsilon = 4$ level. Entropic effects ($T\Delta S$ at $T = 310$ K) and zero-point energies (ZPE) were evaluated at the B3LYP-D3/def2-SVP/ $\epsilon = 4$ level. Reaction pathways for the pT process were optimized using a chain-of-states method.^{55,56} All QM calculations were performed with TURBOMOLE version 7.3.⁵⁷

Kinetic Model and Master Equation. A five-site kinetic master equation model^{58–60} was created to probe the proton pumping kinetics in Nqo13 by modeling proton channels from the N- and P-sides of the membrane, connected to middle (Lys235₁₃) and terminal (Glu377₁₃) proton acceptor sites, respectively, and a switchable ion-pair that was opened upon application of external force (Figure 4F). The energy of each microstate was expressed as a sum of site energies (E_n^{intr}) and coupling energies (E_{nm}),

$$E = \sum_{n=1}^{N_{\text{sites}}} \chi_n E_n^{\text{intr}} + \sum_{n>m}^{N_{\text{sites}}, N_{\text{sites}}-1} \chi_n \chi_m E_{nm}^{\text{coupl}} \quad (1)$$

where occupied/empty (protonated/deprotonated, open/closed) states correspond to $\chi = 1/0$, resulting in a total of $2^5 = 32$ microstates. After parameter optimization (see below), the intrinsic energy of all sites was scanned across a pK_a range of 0–14, with $E_n^{\text{intr}} = -RT(pH - pK_a)$ model for protonation sites, and in the range of -10 kcal mol⁻¹ to +10 kcal mol⁻¹ for channel and ion-pair sites. The coupling strengths were scanned in the range of -10 kcal mol⁻¹ to +10 kcal mol⁻¹. Transition rates, k_{ij} , between microstates i and j , were modeled as

$$k_{ij} = \kappa_{ij} \exp\left[-\frac{(E_j - E_i)}{2RT}\right] \quad (2)$$

with optimized parameters and intrinsic rate coefficients κ_{ij} given in Tables S3 and S4. The population dynamics was described with the master equation,

$$\frac{dp_i}{dt} = \sum_{j=1}^{N_{\text{states}}} k_{ji} p_j - \sum_{j=1}^{N_{\text{states}}} k_{ij} p_i \quad (3)$$

that was numerically integrated using an ordinary differential equation (ODE) solver as implemented in the SciPy library.⁶¹ After an initial equilibrium distribution obtained at steady state ($dp_i/dt = 0$), a biasing force was applied to open the ion-pair for t time steps,

$$\Delta E_{\text{bias}} = \Delta E_{\text{bias}}^{\text{max}} \sin^2(2\pi t/t_{\text{period}}) \quad (4)$$

where $\Delta E_{\text{bias}}^{\text{max}}$ is 400 mV and t_{period} is set to 1 ms based on the turnover of complex I. The pumping flux from the N-side to the P-side was measured as the net number of protons exiting (entering) the system in one cycle during the application of the external force. See [Supporting Information Methods](#) for more detailed description of the kinetic model and dependence on the model parameters.

Construction of Complex I Mutants. A derivative of the *E. coli* strain BW25113⁶² with deleted *ndh* gene and the *nuo* operon replaced by a resistance cartridge (*nptII*) was created by genomic replacement⁶³ and used to overproduce complex I and the variants. *E. coli* DH5 α Δ *nuo*⁶⁴ and plasmid pBAD*nuo*_{His}⁶⁵ were used for generation of the mutants (see [Table S5](#)). All mutations were confirmed by DNA sequencing (GATC Eurofins, Konstanz, Germany). Cells were grown at 37 °C in baffled flasks using autoinduction medium⁶⁶ ([Table S6](#)). Cells were disrupted by three passages through an EmulsiFlex (Avestin), and cytoplasmic membranes were isolated by differential centrifugation as described before.⁶² Cytoplasmic membranes were suspended in buffer A*_{pH6.8} ([Table S6](#)).

Purification and Reconstitution of Complex I. All steps were carried out at 4 °C. 2% (w/v) LMNG (Anatrace) was added to the membrane suspension, followed by incubation for 1 h with gentle stirring at RT and 20 min centrifugation at 160 000g. The supernatant was adjusted to 20 mM imidazole and applied on a ProBond Ni²⁺-IDA column (35 mL, Invitrogen) equilibrated in binding buffer ([Table S6](#)). Bound proteins were eluted with 308 mM imidazole (60% elution buffer; [Table S6](#)). Fractions with NADH/ferricyanide oxidoreductase activity were concentrated by ultrafiltration (100 kDa MWCO Amicon Ultra-15 centrifugal filter; Millipore) and further applied onto a Superose 6 size-exclusion chromatography column (300 mL, GE Healthcare) equilibrated in buffer A*_{MNG} ([Table S6](#)). The fractions with the highest NADH/ferricyanide oxidoreductase activity were concentrated, and the protein was either directly used for biophysical measurements or stored at -80 °C.

Reconstitution of Complex I into Liposomes. *E. coli* polar lipids (25 mg mL⁻¹ in CHCl₃; Avanti) were evaporated and dissolved in a 5-fold volume lipid buffer ([Table S6](#)). The suspension was frozen in liquid nitrogen and thawed at 29 °C seven times to obtain unilamellar vesicles, followed by extrusion (0.1 μ M polycarbonate membrane, Mini Extruder; Avanti).⁷¹ For reconstitution, 0.5–1 mg of complex I was mixed with reconstitution buffer ([Table S6](#)) and incubated for 5 min on ice. An amount of 250 μ L of liposomes was mixed with 8 μ L of sodium cholate (20%, w/v) and mixed with complex I in the reconstitution buffer and incubated for 15 min at RT. The solution was applied onto a size-exclusion column (PD-10 desalting column, 8.3 mL, Sephadex G-25, GE Healthcare) equilibrated in lipid buffer in order to remove the excess detergent. The eluate was centrifuged (4 °C, 2 bar air pressure, 30 min, 150 000g, Rotor A-100, Beckman Airfuge), and sedimented proteoliposomes were resuspended in 500 μ L of lipid buffer. The NADH/ferricyanide oxidoreductase activity of the proteoliposomes was used to adjust an equal complex I content in the pumping assays.

Measurement of Δ pH and $\Delta\Psi$. The proton gradient (Δ pH) generated by WT and variants was probed by monitoring the fluorescence quenching of 9-amino-6-chloro-2-methoxyacridine (ACMA, Sigma), with an assay containing 100 μ M decylubiquinone (Sigma), 0.2 μ M ACMA, and 50 μ L of proteoliposomes (with 50–100 μ g of complex I) in the ACMA buffer ([Table S6](#)). Measurements were performed at 30 °C by starting the reaction with 130 μ M NADH. The fluorescence was detected with an LS 55 fluorescence

spectrometer (PerkinElmer) at excitation/emission wavelengths of 430 nm/480 nm. Addition of 10 μ M carbonyl cyanide 3-chlorophenylhydrazone (CCCP) was used to dissipate Δ pH. The generation of $\Delta\Psi$ was determined by monitoring absorption change at 588–625 nm of oxonol-VI (Sigma), with an assay containing 0.5 μ M oxonol-VI, 50 μ M Q₀, and proteoliposomes in the oxonol buffer ([Table S6](#)).⁶⁷ The process was studied at 30 °C by starting the reaction with 100 μ M NADH. The absorbance changes were measured with a diode-array spectrometer (Hellma; TIDAS II, J&M Aalen). Addition of 2 μ g mL⁻¹ gramicidin was used to dissipate the $\Delta\Psi$. Pumping activities were characterized from the maximum level of the optical changes without extrapolation to zero time point using the plateau levels as a base. Exact proton pumping stoichiometry was not characterized (cf. e.g., refs 68–70). All measurements were performed in triplicates for all three constructs and errors are reported as standard deviations. See [Supporting Information Methods](#) for further experimental details.

Activity Measurements. The NADH:decylubiquinone oxidoreductase activity was measured by the decrease of the NADH concentration at 340 nm.⁶⁵ The assay contained 60 μ M decylubiquinone, 2 μ g of complex I, and a 10-fold molar excess (5 μ g) of *E. coli* cytochrome *b*₀3 in buffer A*_{MNG}. The reaction was started by an addition of 150 μ M NADH. All activity assays were performed at 30 °C. See [Supporting Information Methods](#) for further details.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.0c02789>.

Supplementary methods; figures showing hydration, proton transfer and conformational dynamics, DFT reaction barriers, kinetic model, convergence of free energy calculations, and multiple sequence alignments; experimental details of biophysical experiments, oligonucleotide sequences of primers used for site-directed mutagenesis, buffers used for expression, purification, reconstitution, and biophysical experiments; and summary of QM/MM and MD simulations (PDF)

Movie showing pT reaction in QM/MM models (MP4)

■ AUTHOR INFORMATION

Corresponding Author

Ville R. I. Kaila – Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden; Center for Integrated Protein Science Munich at the Department of Chemistry, Technical University of Munich, D85748 Garching, Germany; orcid.org/0000-0003-4464-6324; Email: ville.kaila@dbb.su.se

Authors

Max E. Mühlbauer – Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden; Center for Integrated Protein Science Munich at the Department of Chemistry, Technical University of Munich, D85748 Garching, Germany

Patricia Saura – Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden; Center for Integrated Protein Science Munich at the Department of Chemistry, Technical University of Munich, D85748 Garching, Germany

Franziska Nuber – Institut für Biochemie, Albert-Ludwigs-Universität Freiburg, 79104 Freiburg, Germany

Andrea Di Luca – Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden; Center

for Integrated Protein Science Munich at the Department of Chemistry, Technical University of Munich, D85748 Garching, Germany

Thorsten Friedrich – Institut für Biochemie, Albert-Ludwigs-Universität Freiburg, 79104 Freiburg, Germany; orcid.org/0000-0002-0813-7883

Complete contact information is available at:
<https://pubs.acs.org/10.1021/jacs.0c02789>

Author Contributions

^{||}M.E.M. and P.S. contributed equally.

Notes

The authors declare no competing financial interest.

Data supporting the findings of this manuscript are available from the corresponding authors upon reasonable request.

ACKNOWLEDGMENTS

This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement 715311, to V.R.I.K.), and the Deutsche Forschungsgemeinschaft within the SPP 1927 (Grant FR 1140/11-1, to T.F.). V.R.I.K. is also supported by the Knut and Alice Wallenberg (KAW) Foundation. Computing time was provided by the Leibniz Rechenzentrum (LRZ) (SuperMUC Project pr27xu) and PRACE (Project pr1ejk).

REFERENCES

- (1) Kaila, V. R. I. Long-Range Proton-Coupled Electron Transfer in Biological Energy Conversion: Towards Mechanistic Understanding of Respiratory Complex I. *J. R. Soc., Interface* **2018**, *15*, 20170916.
- (2) Hirst, J. Mitochondrial Complex I. *Annu. Rev. Biochem.* **2013**, *82*, 551–575.
- (3) Brandt, U. Energy Converting NADH: Quinone Oxidoreductase (Complex I). *Annu. Rev. Biochem.* **2006**, *75*, 69–92.
- (4) Sazanov, L. A. A giant molecular proton pump: Structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 375–388.
- (5) Wikström, M. Two protons are pumped from the mitochondrial matrix per electron transferred between NADH and ubiquinone. *FEBS Lett.* **1984**, *169*, 300–304.
- (6) Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **1961**, *191*, 144–148.
- (7) Yoshida, M.; Muneyuki, E.; Hisabori, T. ATP synthase - A marvellous rotary engine of the cell. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 669–677.
- (8) Baradaran, R.; Berrisford, J. M.; Minhas, G. S.; Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **2013**, *494*, 443–448.
- (9) Agip, A. N. A.; Blaza, J. N.; Bridges, H. R.; Viscomi, C.; Rawson, S.; Muench, S. P.; Hirst, J. Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states. *Nat. Struct. Mol. Biol.* **2018**, *25*, 548–556.
- (10) Zickermann, V.; Wirth, C.; Nasiri, H.; Siegmund, K.; Schwalbe, H.; Hunte, C.; Brandt, U. Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* **2015**, *347*, 44–49.
- (11) Schuller, J. M.; Saura, P.; Thiemann, J.; Schuller, S. K.; Gamiz-Hernandez, A. P.; Kurisu, G.; Nowaczyk, M. M.; Kaila, V. R. I. Redox-coupled proton pumping drives carbon concentration in the photosynthetic complex I. *Nat. Commun.* **2020**, *11*, 494.
- (12) Euro, L.; Belevich, G.; Verkhovskaya, M. I.; Wikström, M.; Verkhovskaya, M. Conserved lysine residues of the membrane subunit NuoM are involved in energy conversion by the proton-pumping NADH:ubiquinone oxidoreductase (Complex I). *Biochim. Biophys. Acta, Bioenerg.* **2008**, *1777*, 1166–1172.
- (13) Nakamaru-Ogiso, E.; Kao, M. C.; Chen, H.; Sinha, S. C.; Yagi, T.; Ohnishi, T. The membrane subunit NuoL(ND5) is involved in the indirect proton pumping mechanism of Escherichia coli complex I. *J. Biol. Chem.* **2010**, *285*, 39070–39078.
- (14) Amarneh, B.; Vik, S. B. Mutagenesis of subunit N of the Escherichia coli Complex I. Identification of the initiation codon and the sensitivity of mutants to decylubiquinone. *Biochemistry* **2003**, *42*, 4800–4808.
- (15) Michel, J.; DeLeon-Rangel, J.; Zhu, S.; van Ree, K.; Vik, S. B. Mutagenesis of the L, M, and N subunits of complex I from Escherichia coli indicates a common role in function. *PLoS One* **2011**, *6*, e17420.
- (16) Sinha, P. K.; Torres-Bacete, J.; Nakamaru-Ogiso, E.; Castro-Guerrero, N.; Matsuno-Yagi, A.; Yagi, T. Critical roles of subunit NuoH (ND1) in the assembly of peripheral subunits with the membrane domain of Escherichia coli NDH-1. *J. Biol. Chem.* **2009**, *284*, 9814–9823.
- (17) Torres-Bacete, J.; Nakamaru-Ogiso, E.; Matsuno-Yagi, A.; Yagi, T. Characterization of the NuoM (ND4) subunit in Escherichia coli NDH-1: Conserved charged residues essential for energy-coupled activities. *J. Biol. Chem.* **2007**, *282*, 36914–36922.
- (18) Sazanov, L. A. The mechanism of coupling between electron transfer and proton translocation in respiratory complex I. *Bioenerg. Biomembr.* **2014**, *46*, 247–253.
- (19) Wikström, M.; Sharma, V.; Kaila, V. R. I.; Hosler, J. P.; Hummer, G. New Perspectives on Proton Pumping in Cellular Respiration. *Chem. Rev.* **2015**, *115*, 2196–2221.
- (20) Le Breton, N.; Wright, J. J.; Jones, A. J. Y.; Salvadori, E.; Bridges, H. R.; Hirst, J.; Roessler, M. M. Spectroscopy to define the proton-coupled electron transfer reaction at Fe-S cluster N2 in respiratory complex I. *J. Am. Chem. Soc.* **2017**, *139*, 16319–16326.
- (21) Saura, P.; Kaila, V. R. I. Energetics and dynamics of proton-coupled electron transfer in the NADH/FMN site of respiratory complex I. *J. Am. Chem. Soc.* **2019**, *141*, 5710–5719.
- (22) Hammes-Schiffer, S. Proton-coupled electron transfer: classification scheme and guide to theoretical methods. *Energy Environ. Sci.* **2012**, *5*, 7696–7703.
- (23) Cabrera-Orefice, A.; Yoga, E. G.; Wirth, C.; Siegmund, K.; Zwicker, K.; Guerrero-Castillo, S.; Zickermann, V.; Hunte, C.; Brandt, U. Locking loop movement in the ubiquinone pocket of complex I disengages the proton pumps. *Nat. Commun.* **2018**, *9*, 4500.
- (24) Fedor, J. G.; Jones, A. J. Y.; Di Luca, A.; Kaila, V. R. I.; Hirst, J. Correlating kinetic and structural data on ubiquinone binding and reduction by respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 12737–12742.
- (25) Di Luca, A.; Mühlbauer, M. E.; Saura, P.; Kaila, V. R. I. How inter-subunit contacts in the membrane domain of complex I affect proton transfer energetics. *Biochim. Biophys. Acta, Bioenerg.* **2018**, *1859*, 734–741.
- (26) Efremov, R. G.; Sazanov, L. A. Structure of the membrane domain of respiratory complex I. *Nature* **2011**, *476*, 414–420.
- (27) Di Luca, A.; Gamiz-Hernandez, A. P.; Kaila, V. R. I. Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E6314–E6321.
- (28) Sharma, V.; Belevich, G.; Gamiz-Hernandez, A. P.; Róg, T.; Vattulainen, I.; Verkhovskaya, M. L.; Wikström, M.; Hummer, G.; Kaila, V. R. I. Redox-induced activation of the proton pump in the respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 11571–11576.
- (29) Kaila, V. R. I.; Wikström, M.; Hummer, G. Electrostatics, hydration, and proton transfer dynamics in the membrane domain of respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 6988–6993.
- (30) Warnau, J.; Sharma, V.; Gamiz-Hernandez, A. P.; Di Luca, A.; Haapanen, O.; Vattulainen, I.; Wikström, M.; Hummer, G.; Kaila, V. R. I. Redox-Coupled Quinone Dynamics in the Respiratory Complex I. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E8413–E8420.

- (31) Hunte, C.; Zickermann, V.; Brandt, U. Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science* **2010**, *329*, 448–451.
- (32) Gamiz-Hernandez, A. P.; Jussupow, A.; Johansson, M. P.; Kaila, V. R. I. Terminal Electron-Proton Transfer Dynamics in the Quinone Reduction of Respiratory Complex I. *J. Am. Chem. Soc.* **2017**, *139*, 16282–16288.
- (33) Haapanen, O.; Sharma, V. Role of water and protein dynamics in proton pumping by respiratory complex I. *Sci. Rep.* **2017**, *7*, 7747.
- (34) Belevich, G.; Knuuti, J.; Verkhovskiy, M. I.; Wikström, M.; Verkhovskaya, M. Probing the mechanistic role of the long α -helix in subunit L of respiratory Complex I from *Escherichia coli* by site-directed mutagenesis. *Mol. Microbiol.* **2011**, *82*, 1086–1095.
- (35) Zhu, S.; Vik, S. B. Constraining the lateral helix of respiratory Complex I by cross-linking does not impair enzyme activity or proton translocation. *J. Biol. Chem.* **2015**, *290*, 20761–20773.
- (36) Steimle, S.; Schnick, C.; Burger, E. M.; Nuber, F.; Krämer, D.; Dawitz, H.; Brander, S.; Matlosz, B.; Schäfer, J.; Maurer, K.; Glessner, U.; Friedrich, T. Cysteine scanning reveals minor local rearrangements of the horizontal helix of respiratory complex I. *Mol. Microbiol.* **2015**, *98*, 151–161.
- (37) Rasaiah, J. C.; Garde, S.; Hummer, G. Water in Nonpolar Confinement: From Nanotubes to Proteins and Beyond. *Annu. Rev. Phys. Chem.* **2008**, *59*, 713–740.
- (38) D'Alessandro, M.; Turina, P.; Melandri, B. A. Quantitative evaluation of the intrinsic uncoupling modulated by ADP and Pi in the reconstituted ATP synthase of *Escherichia coli*. *Biochim. Biophys. Acta, Bioenerg.* **2011**, *1807*, 130–143.
- (39) Pawate, A. S.; Morgan, J.; Namslauer, A.; Mills, D.; Brzezinski, P.; Ferguson-Miller, S.; Gennis, R. B. A mutation in subunit I of cytochrome oxidase from *Rhodobacter sphaeroides* results in an increase in steady-state activity but completely eliminates proton pumping. *Biochemistry* **2002**, *41*, 13417–13423.
- (40) Saura, P.; Frey, D. M.; Gamiz-Hernandez, A. P.; Kaila, V. R. I. Electric field modulated redox-driven protonation and hydration energetics in energy converting enzymes. *Chem. Commun.* **2019**, *55*, 6078–6081.
- (41) Supekar, S.; Kaila, V. R. I. Dewetting transitions coupled to K-channel activation in cytochrome *c* oxidase. *Chem. Sci.* **2018**, *9*, 6703–6710.
- (42) Suomivuori, C.-M.; Gamiz-Hernandez, A. P.; Sundholm, D.; Kaila, V. R. I. Energetics and dynamics of a light-driven sodium-pumping rhodopsin. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 7043–7048.
- (43) Ugur, I.; Rutherford, A. W.; Kaila, V. R. I. Redox-coupled substrate water reorganization in the active site of Photosystem II - The role of calcium in substrate water delivery. *Biochim. Biophys. Acta, Bioenerg.* **2016**, *1857*, 740–748.
- (44) Kaila, V. R. I.; Hummer, G. Energetics of Direct and Water-Mediated Proton-Coupled Electron Transfer. *J. Am. Chem. Soc.* **2011**, *133*, 19040–19043.
- (45) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (46) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (47) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (48) Stelzl, L. S.; Kells, A.; Rosta, E.; Hummer, G. Dynamic Histogram Analysis To Determine Free Energies and Rates from Biased Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 6328–6342.
- (49) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.
- (50) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (51) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577.
- (52) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104–1–19.
- (53) Riahi, S.; Rowley, C. N. The CHARMM-TURBOMOLE interface for efficient and accurate QM/MM molecular dynamics, free energies, and excited state properties. *J. Comput. Chem.* **2014**, *35*, 2076–2086.
- (54) Klamt, A.; Schüürmann, G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799–805.
- (55) Plessow, P. Reaction path optimization without NEB springs or interpolation algorithms. *J. Chem. Theory Comput.* **2013**, *9*, 1305–1310.
- (56) E, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *66*, 052301.
- (57) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
- (58) Hill, T. L. *Free Energy Transduction in Biology*; Academic: New York, 1977.
- (59) Kim, Y. C.; Wikström, M.; Hummer, G. Kinetic models of redox-coupled proton pumping. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2169–2174.
- (60) Kim, Y. C.; Hummer, G. Proton-pumping mechanism of cytochrome *c* oxidase: A kinetic master-equation approach. *Biochim. Biophys. Acta, Bioenerg.* **2012**, *1817*, 526–536.
- (61) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, M.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P. and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17* (3), 261–272.
- (62) Datsenko, K. A.; Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 6640–6645.
- (63) Burschel, S.; Kreuzer Decovic, D.; Nuber, F.; Stiller, M.; Hofmann, M.; Zupok, A.; Siemiatkowska, B.; Gorka, M.; Leimkühler, S.; Friedrich, T. Iron-sulfur cluster carrier proteins involved in the assembly of *Escherichia coli* NADH:ubiquinone oxidoreductase (complex I). *Mol. Microbiol.* **2019**, *111*, 31–45.
- (64) Pohl, T.; Uhlmann, M.; Kaufenstein, M.; Friedrich, T. Lambda red-mediated mutagenesis and efficient large scale affinity purification of the *Escherichia coli* NADH:Ubiquinone oxidoreductase (complex I). *Biochemistry* **2007**, *46*, 10694–10702.
- (65) Pohl, T.; Bauer, T.; Dörner, K.; Stolpe, S.; Sell, P.; Zocher, G.; Friedrich, T. Iron-sulfur cluster N7 of the NADH:Ubiquinone oxidoreductase (complex I) is essential for stability but not involved in electron transfer. *Biochemistry* **2007**, *46*, 6588–6596.
- (66) Studier, W. F. Stable expression clones and auto-induction for protein production in *E. coli*. *Methods Mol. Biol.* **2014**, *1091*, 17–32.
- (67) Verkhovskaya, M.; Knuuti, J.; Wikström, M. Role of Ca²⁺ in structure and function of Complex I from *Escherichia coli*. *Biochim. Biophys. Acta, Bioenerg.* **2011**, *1807*, 36–41.
- (68) Dröse, S.; Krack, S.; Sokolova, L.; Zwicker, K.; Barth, H. D.; Morgner, N.; Heide, H.; Steger, M.; Nübel, E.; Zickermann, V.; Kerscher, S.; Brutschy, B.; Radermacher, M.; Brandt, U. Functional

Dissection of the Proton Pumping Modules of Mitochondrial Complex I. *PLoS Biol.* **2011**, *9*, e1001128–e1001128.

(69) Jones, A. J. Y.; Blaza, J. N.; Varghese, F.; Hirst, J. Respiratory Complex I in *Bos Taurus* and *Paracoccus Denitrificans* Pumps Four Protons across the Membrane for Every NADH Oxidized. *J. Biol. Chem.* **2017**, *292*, 4987–4995.

(70) Galkin, A.; Dröse, S.; Brandt, U. The Proton Pumping Stoichiometry of Purified Mitochondrial Complex I Reconstituted into Proteoliposomes. *Biochim. Biophys. Acta, Bioenerg.* **2006**, *1757* (12), 1575–1581.

(71) von Ballmoos, C.; Biner, O.; Nilsson, T.; Brzezinski, P. Mimicking respiratory phosphorylation using purified enzymes. *Biochim. Biophys. Acta, Bioenerg.* **2016**, *1857*, 321–331.

Article III



Deactivation blocks proton pathways in the mitochondrial complex I

Michael Röpke^a, Daniel Riepl^{b,1}, Patricia Saura^{b,1}, Andrea Di Luca^{b,1}, Max E. Mühlbauer^{a,b}, Alexander Jussupow^{a,b}, Ana P. Gamiz-Hernandez^b, and Ville R. I. Kaila^{a,b,2}

^aDepartment Chemie, Technische Universität München, D-85747 Garching, Germany; and ^bDepartment of Biochemistry and Biophysics, Stockholm University, 10691 Stockholm, Sweden

Edited by Harry B. Gray, California Institute of Technology, Pasadena, CA, and approved May 11, 2021 (received for review September 16, 2020)

Cellular respiration is powered by membrane-bound redox enzymes that convert chemical energy into an electrochemical proton gradient and drive the energy metabolism. By combining large-scale classical and quantum mechanical simulations with cryo-electron microscopy data, we resolve here molecular details of conformational changes linked to proton pumping in the mammalian complex I. Our data suggest that complex I deactivation blocks water-mediated proton transfer between a membrane-bound quinone site and proton-pumping modules, decoupling the energy-transduction machinery. We identify a putative gating region at the interface between membrane domain subunits ND1 and ND3/ND4L/ND6 that modulates the proton transfer by conformational changes in transmembrane helices and bulky residues. The region is perturbed by mutations linked to human mitochondrial disorders and is suggested to also undergo conformational changes during catalysis of simpler complex I variants that lack the “active”-to-“deactive” transition. Our findings suggest that conformational changes in transmembrane helices modulate the proton transfer dynamics by wetting/dewetting transitions and provide important functional insight into the mammalian respiratory complex I.

cell respiration | bioenergetics | molecular simulations | QM/MM | cryoEM

In mitochondrial cellular respiration, the membrane-bound enzyme complexes I, III, and IV convert chemical energy into a flux of electrons toward dioxygen (1–6). The free energy of the process is transduced by pumping protons across the inner mitochondrial membrane (IMM), powering oxidative phosphorylation and active transport (7, 8). The electron transport process is initiated by the respiratory complex I (NADH:ubiquinone oxidoreductase), a 45-subunit modular enzyme machinery that shuttles electrons from nicotinamide adenine dinucleotide (NADH) to ubiquinone (Q₁₀) and transduces the free energy by pumping protons across the IMM, generating a proton motive force (*pmf*) (1, 4–6) (Fig. 1). This proton-coupled electron transfer reaction is fully reversible, and complex I can also operate in reverse electron transfer (RET) mode, powering ubiquinol oxidization by consumption of the *pmf*. Such RET modes become prevalent under hypoxic or anoxic conditions that may result, e.g., from stroke or tissue damage (9), during which the electrons leak from complex I to molecular oxygen and result in the formation of reactive oxygen species (ROS) with physiologically harmful consequences (9–11). To regulate this potentially dangerous operation mode, the mammalian complex I can transition into a “deactive” (D) state with low Q₁₀-turnover activity (12, 13). Although some structural changes involved in the “active”-to-“deactive” (A/D) transition were recently resolved (14–16), the molecular details of how this transition regulates enzyme turnover and its relevance during *in vivo* conditions still remain puzzling. Moreover, it is also debated whether conformational changes linked to this transition are involved in the native catalytic cycle of all members of the complex I superfamily or whether this transition is specific for the mitochondrial enzyme (12, 13, 17).

The recently resolved cryo-electron microscopy (cryoEM) structures of the “deactive” mammalian complex I at around 4-Å resolution highlighted conformational changes around several subunits close to the interface between the hydrophilic and membrane domains of complex I. Particularly interesting are the conformational changes around the membrane domain subunits ND4L, ND3, and ND6 (ND for NADH Dehydrogenase) that form a bundle of 11 transmembrane (TM) helices, connected by long loop regions (14–16). Notably, it was observed that TM3 of ND6 transitions from a fully α -helical form in the “active” state to a π -bulge around residues 60 to 65 during the deactivation process (14–16). Although the exact relevance of these conformational transitions remains debated, it is notable that several point mutations in the vicinity of these regions have been linked to mitochondrial disease (11), supporting their possible functional relevance. Structural changes during complex I deactivation, inferred from the lack of density in the cryoEM maps (14–16), were also suggested to take place in several loop regions of the membrane domain, near the Q₁₀ binding site tunnel, and around the supernumerary subunits NDUFA5/NDUFA10 (16, 18). However, the functional consequences of these structural changes and their coupling to the biological activity still remain unclear.

Significance

The electron transport chain of mitochondria is initiated by the respiratory complex I that converts chemical energy into a proton motive force to power synthesis of adenosine triphosphate. On a chemical level, complex I catalyzes elementary electron and proton transfer processes that couple across large molecular distances of >300 Å. However, under low oxygen concentrations, the respiratory chain operates in reverse mode and produces harmful reactive oxygen species. To avoid cell damage, the mitochondrial complex I transitions into a deactive state that inhibits turnover by molecular principles that remain elusive. By combining large-scale molecular simulations with cryo-electron microscopy data, we show here that complex I deactivation blocks the communication between proton pumping and redox modules by conformational and hydration changes.

Author contributions: V.R.I.K. designed research; M.R., D.R., P.S., A.D.L., M.E.M., A.J., and A.P.G.-H. performed research; M.R., D.R., P.S., A.D.L., M.E.M., A.J., and A.P.G.-H. contributed new reagents/analytic tools; M.R., D.R., P.S., A.D.L., M.E.M., A.J., A.P.G.-H., and V.R.I.K. analyzed data; and V.R.I.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹D.R., P.S., and A.D.L. contributed equally to this work.

²To whom correspondence may be addressed. Email: ville.kaila@dbb.su.se.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2019498118/-DCSupplemental>.

Published July 16, 2021.

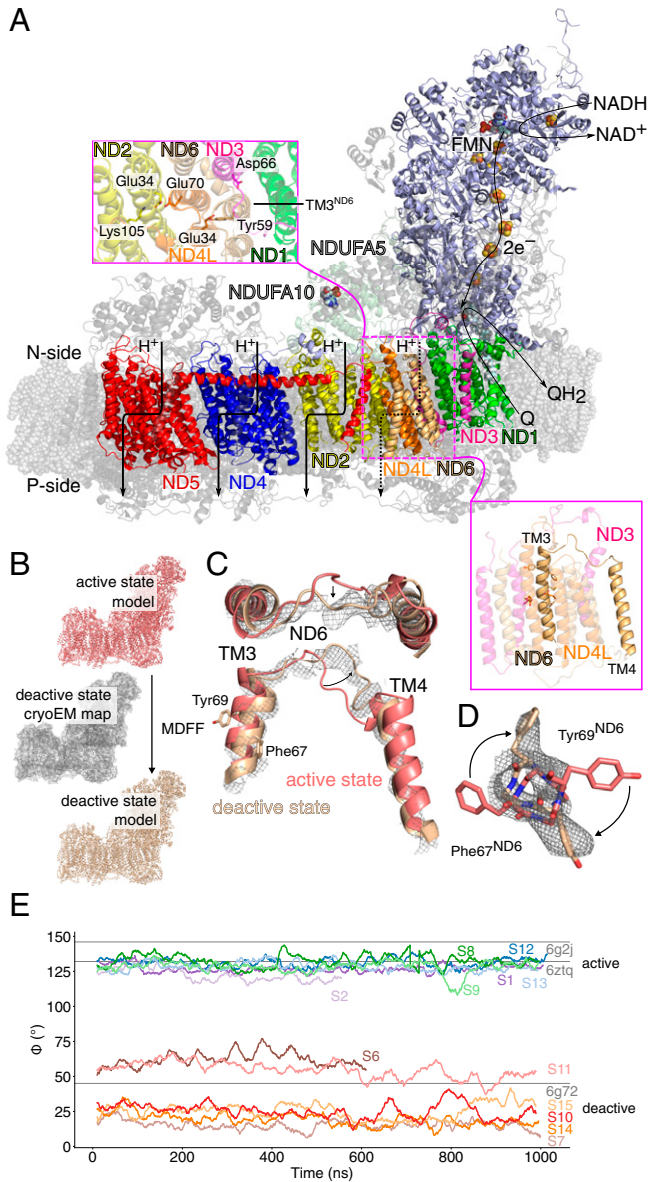


Fig. 1. Structure and function of the mammalian complex I. (A) Electron transfer from NADH reduces quinone (Q) to quinol (QH₂) and triggers proton pumping across the membrane domain. (Inset) Closeup of the ND1/ND3/ND4L/ND6 interface involved in the “active” to “deactive” transition. TM3^{ND6}, which has been linked to conformational changes in the A/D transitions, is marked. (B) An intact atomic model of the deactive state was constructed using MDFF based on the cryoEM structure of the “active” state and the density map of the “deactive” state. (C and D) Conformational changes during the A (in pink)/D (in brown) transition during the MDFF simulations at the TM3^{ND6} region, with the D state density map shown. Refer to *SI Appendix, Figs. S2, S3, and S12* for other conformational changes. (E) The dihedral angle, ϕ , for ND6 residues Leu51(C β)-Leu51(C α)-Phe67(C α)-Phe67(C β) during MD simulations in the “active” and “deactive” states in comparison to refined cryoEM models.

To probe how structural changes linked to deactivation could affect the protonation and quinone dynamics of the mammalian complex I, we combine here atomistic molecular dynamics (MD) simulations and hybrid quantum/classical (QM/MM) free energy calculations with cryoEM data (16, 19). Our combined findings suggest that conformational changes around the ND1/ND3/ND4L/ND6 interface and conserved loop regions could block

the coupling between proton pumping- and redox-modules upon complex I deactivation. The explored molecular principles are of general importance for elucidating energy transduction mechanisms in the mammalian respiratory complex I and possibly other bioenergetic enzyme complexes, but also for understanding the development of mitochondrial diseases.

Results

Conformational Dynamics Modulate Water-Mediated Proton Transfer Reactions. To probe the dynamics of the mammalian complex I, we embedded the 45-subunit “active” state of mouse complex I (19) in a mitochondrial inner lipid membrane-model with POPC/POPE/cardiolipin (2:2:1), solvated the model with water molecules and 150 mM NaCl, and simulated the ca. 1 million atom system in total for around 8.5 μ s using atomistic MD simulations. A quinone or quinol molecule was modeled in the primary binding site near the N2 iron–sulfur center (20) or in a membrane-bound binding site located near the Q-tunnel kink region at ND1 that was recently predicted based on simulations and validated experimentally (19, 21). Simulations were also performed with the Q-cavity left in an empty (apo) state (refer to *SI Appendix, Fig. S1* and Table S1 for simulation setups and Table S2 for modeled protonation states). To obtain molecular insight into the “deactive” state, we constructed an intact atomistic model of this form by targeting the “active” state model toward the experimental cryoEM “deactive” density (EMD: 4356) (16) using an MD flexible fitting (MDFF) approach, followed by unrestrained MD simulation for a total simulation time of ca. 6 μ s (Fig. 1B). In the targeted MDFF simulations, the cryoEM density acts as an external biasing potential that guides the dynamics of the residues, whereas for unresolved regions, the dynamics is directed only by the biomolecular force field (22). This modeling approach is likely to provide a more balanced description of the “deactive” state and a better comparison to our “active” state simulations, as the former state could not be experimentally resolved at the same atomic level of detail as the “active” state. The obtained “deactive” model, created from our “active” state simulation setup, resembles the previously refined 3.9-Å structure of the former [Protein Data Bank (PDB) ID: 6G72] (16) for the experimentally resolved regions but contains more atomic detail for unresolved parts (Fig. 1C and D and *SI Appendix, Figs. S2, S3, and S12, Materials and Methods*). The global dynamics observed in the MD simulations resemble the dynamics inferred from the local resolution of the cryoEM map (*SI Appendix, Fig. S24*), including a relative twist of the hydrophilic domain relative to the membrane domain experimentally described before (14–16, 23) (*SI Appendix, Fig. S3F*). We also observe pronounced conformational differences around the ND3/ND4L/ND6 subunits between the “active” and “deactive” state simulations, particularly at Phe67^{ND6} and Tyr69^{ND6} of ND6 (Fig. 1C–E). The displacement rotates TM3^{ND6} helix and results in the formation of a π -bulge, closely resembling the structural changes observed in previous cryoEM structures (Fig. 1E) (16) and further supporting that the model captures key features of the “deactive” state around this region. The “deactive” state model remains stable during unrestrained simulations, suggesting that the structure is a local free-energy minimum.

Water molecules establish S-shaped proton pathways around broken-helix elements in the three antiporter-like subunits, ND4, ND5, and ND2, in both our “active” and “deactive” state simulations (*SI Appendix, Fig. S4*), closely resembling the proton pathways described before (24–28). However, close to the ND1/ND3/ND4L/ND6 subunits, the conformational changes between the “active” and “deactive” states strongly alter the protonation dynamics: in the “active” state simulations, we observe a proton wire that forms on ca. 0.5- μ s timescales and establishes a hydrogen-bonded connection from the negatively charged (N) side of the membrane to a network of carboxylates in ND1 (Asp199^{ND1}, Glu202^{ND1}, Glu204^{ND1}, Glu227^{ND1}, Glu192^{ND1}, and Glu143^{ND1})

(Fig. 2 *A* and *B*). This proton wire continues toward Glu34^{ND4L}/Tyr59^{ND6} and extends further to Glu70^{ND4L}, located next to the Glu34^{ND2}/Lys105^{ND2} ion-pair of ND2 (Figs. 2 *A* and *B* and 3 *A* and *E*). However, despite rather extensive hydration in ND1, we observe large (around 8- to 10-Å) gaps in the proton wire in the ND3/ND6/ND4L region between Asp66^{ND3} and Glu34^{ND4L} and Glu34^{ND4L} and Glu70^{ND4L}/Glu34^{ND2} that could prevent efficient proton transfer between the residues (Fig. 3*A* and *SI Appendix*, Fig. S5). When we protonate Asp66^{ND3}, mimicking a putative proton transfer from the E-channel, a water wire is established during the 100-ns simulation between Asp66^{ND3} and Glu34^{ND4L}/

Tyr59^{ND6} (Fig. 3*B* and *SI Appendix*, Fig. S5*B*), and by further transferring the proton to Glu34^{ND4L}, we obtain a proton wire between Glu34^{ND4L} and Glu70^{ND4L} (Fig. 3*D*) that stabilizes during the 100-ns MD simulations initiated after the microsecond trajectory (*SI Appendix*, Fig. S6). Shifting a proton through the acidic residues in this area increases the overall hydration level during the simulations (Fig. 3*B*). These findings suggest that the proton transfer reaction could create its own conduction pathway through the channel from the putative second membrane-bound Q-binding region to the interface of the first ion pair Glu34^{ND2}/Lys105^{ND2} in ND2. Proton transfer reactions and protonated

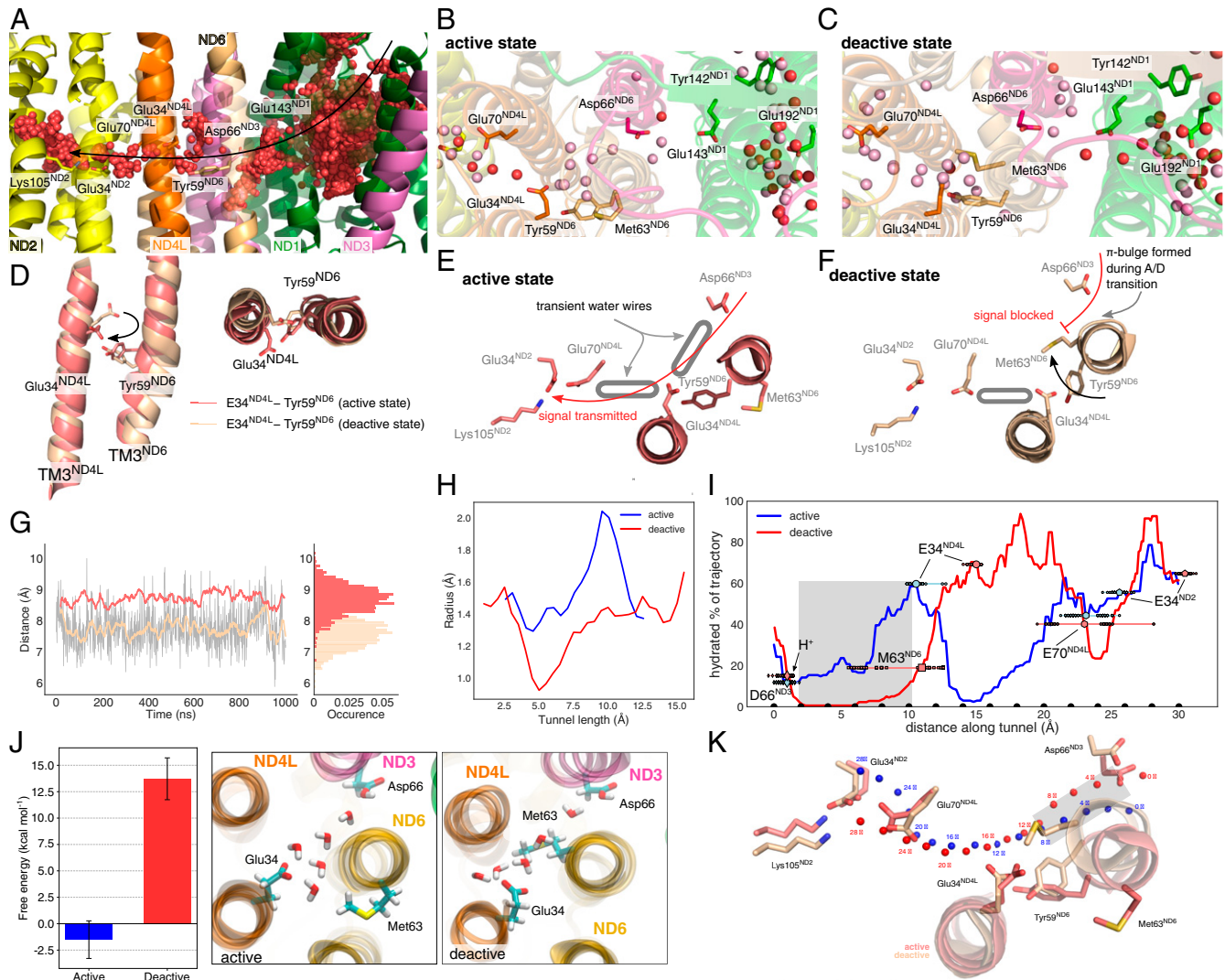


Fig. 2. Hydration dynamics in the “active” and “deactive” states. (*A*) Ensemble average of water molecules during 1,000-ns MD simulations of the “active” state. The water molecules form an effective proton wire via the E-channel in ND1 and connect via the ND3/ND4L/ND6 region to the ND2 interface. (*B* and *C*) Snapshot of individual water molecules (in red) after the 1,000-ns MD simulations in the ND1/ND3/ND4L/ND6 region in the “active” (*B*) and “deactive” (*C*) states. Water molecules in pink are obtained from simulations with shifted protonation states (simulations S19 and S23; *SI Appendix*, Table S1). Refer also to *SI Appendix*, Fig. S6 for cluster analysis. (*D*) Rotation of TM3^{ND4L/ND6} helices induce conformational changes around Glu34^{ND4L}/Tyr59^{ND6} in the “active” (red) and “deactive” (brown) states. Proton wires are established from Asp66^{ND3} via Glu34^{ND4L} to Glu70^{ND4L} and further to the Glu34/Lys105 ion pair of ND2 in the “active” state (*E*), but are blocked in the “deactive” state (*F*). (*G*) Dynamics of the Glu34^{ND4L}/Tyr59^{ND6} during the “active” (red) and “deactive” (brown) states. (*H*) Tunnel radius connecting Asp66^{ND3} and Glu34^{ND4L} in the “active” and “deactive” state simulations, predicted using CAVER tunnel analysis (*Materials and Methods*). (*I*) Hydration fraction along the tunnels connecting the acidic residues along the ND3/ND6/ND4L/ND2 gating region. The tunnel distances correspond to the beads shown in *K*. (*J*) (Left) Hydration free energy for introducing a water wire in the gating region between Asp66^{ND3} and Glu34^{ND4L}, obtained using alchemical FEP simulations, suggesting that the formation of a water wire in the “deactive” state is disfavored by ca. 15 kcal · mol⁻¹ as compared to the “active” state. Error bars indicate a 95% CI based on three independent simulations (*SI Appendix*, *SI Methods*). (Middle/Right) Snapshots of the fully formed water wires in the “active” and “deactive” states during the FEP calculations. (*K*) Tunnel positions shown in panel *I* relative to the structure in the “active” and “deactive” states. Beads are placed 2 Å apart to highlight the channel pathway.

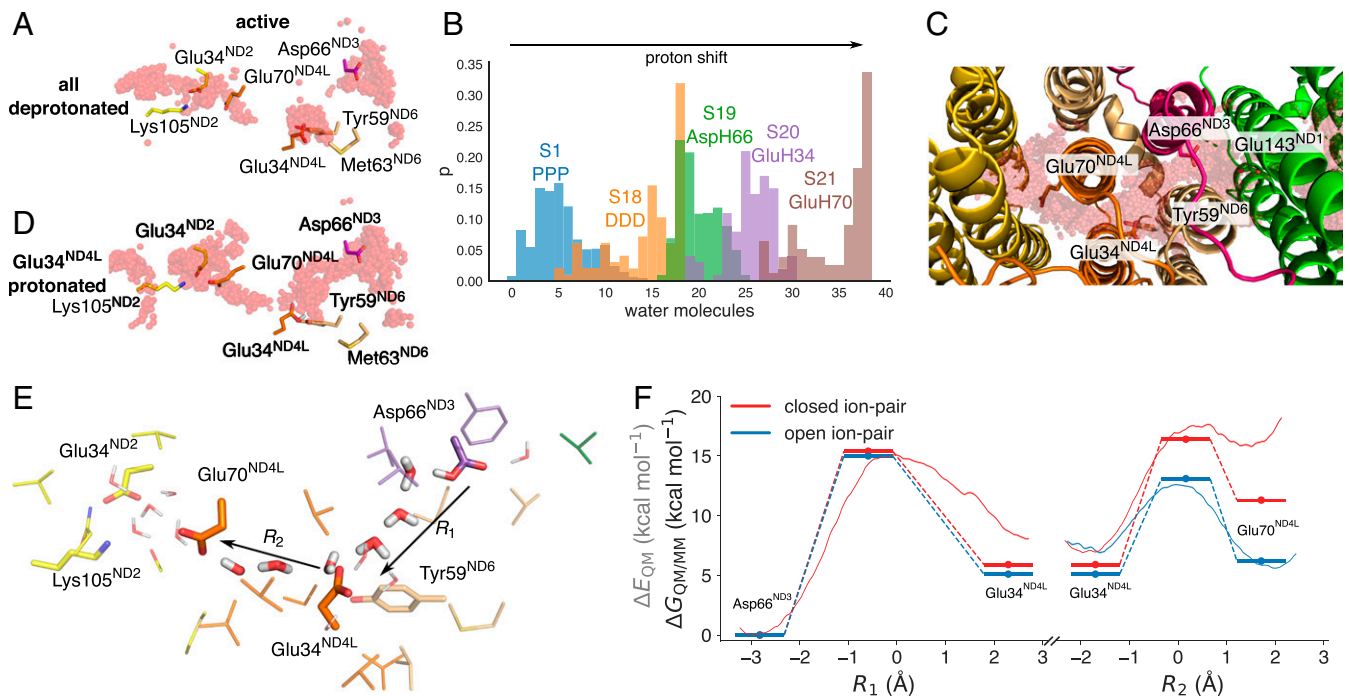


Fig. 3. Proton transfer in the ND3/ND4L/ND6 deactivation gate. (A) Proton pathways from Asp66^{ND3} via Glu34^{ND4L} to Glu70^{ND4L} and the Glu34^{ND2}/Lys105^{ND2} ion pair of ND2. (B) Shifting the protonation state in the ND4L/ND6/ND3 region favors enhanced hydration and hydrogen-bonded wiring between the putative membrane-bound Q-site and the ND2 interface (see *SI Appendix, Table S1*). The figure shows the count of water molecules between the tunnel region extending from ND3 to ND2 (C) during the MD simulations of each state. The distributions are normalized to 1. (C) Top view of the gating region and ensemble average of water molecules in simulations S18 to S21 (*SI Appendix, Table S1*). (D) Shifting the proton to Glu34^{ND4L} enhances water-mediated contacts to Glu70^{ND4L}. Refer also to *SI Appendix, Fig. S5* for exploration of different protonation states in the MD simulations. (E) Residues surrounding the ND3/ND4L/ND6 water chain that were included in the quantum chemical DFT and QM/MM free energy calculations. (F) Proton transfer energetics based on QM/MM free energy simulations (red/blue lines) and quantum chemical DFT models (energy level diagrams) and effect of the ND2 ion pair conformation on the proton transfer profile. Reaction coordinates obtained from the DFT models are marked with a filled circle. Statistical errors in the free energy profiles are ca. 0.12 kcal · mol⁻¹ (transparent red), and the effects of DFT-sampling on the convergence of the free energy profiles are shown in *SI Appendix, Fig. S8*. Benchmarking calculations suggest that the transition states and reaction energies have a few kcal · mol⁻¹ error relative to correlated ab initio calculations (*SI Appendix, Table S7*).

water molecules themselves have also previously been suggested to create their own proton wires (29).

In stark contrast to the well-wired “active” state, we observe that Met63^{ND6} located in the middle of the π -bulge on TM3^{ND6} blocks the proton pathway between Asp66^{ND3} and Glu34^{ND4L} in all “deactive” state simulations (Fig. 2 C and F and *SI Appendix, Fig. S5 E–H*). Rotation of TM3^{ND4L} positions Glu34^{ND4L} toward the Glu70^{ND4L} in the “active” state, whereas the residue points away from the ND2 site in the “deactive” state (Fig. 2 D and G). Analysis of the hydration frequency across the simulation trajectories suggests that the proton pathway between Asp66^{ND3} and Glu34^{ND4L} is well hydrated in our “active” state simulations, whereas in the “deactive” state, Met63^{ND6} completely blocks the formation of a water wire by closing the tunnel (Fig. 2 H and I and *SI Appendix, Fig. S5 I, J, and K*). This ca. 10-Å-long tunnel region comprises on average around five water molecules in the “active” state MD simulations (Fig. 2 B and J). In the “deactive” state, the gating region remains dry in all simulated protonation states, an observation that is well reproduced in our 6- μ s simulation trajectories (*SI Appendix, Fig. S5 E–H and J*).

To further probe whether a proton wire could form on much longer timescales, inaccessible to the MD simulations, we computed hydration free energies using alchemical free energy perturbation (FEP) calculations by creating a water wire into the gating region in both the “active” and “deactive” states (*Materials and Methods*). No restraints were applied on the protein to allow the system to structurally relax upon introduction of the water chain in the cavity (*Materials and Methods*). The FEP

calculations suggest that the hydration free energy is weakly exergonic in the “active” state (–2 kcal · mol⁻¹; Fig. 2J), supporting the MD results for the well-hydrated gating region. In stark contrast, the hydration free energy is strongly endergonic (+14 kcal · mol⁻¹) in the “deactive” state (Fig. 2J) and arises from repulsive van der Waals contributions, indicating steric clashes (*SI Appendix, Fig. S7*). Notably, Met63^{ND6} remains in its inward position during the FEP calculations, leading to a partially broken water wire (Fig. 2J and *SI Appendix, Fig. S7*). During the FEP calculations, we observe small conformational changes, particularly in TM3^{ND6} adjacent to the hydrated cavity (*SI Appendix, Fig. S7F*). We expect that introduction of a protonated water species in this gating region that could favor the hydration process (29) would also experience similar steric repulsion as the neutral water chain introduced here. To probe how introduction of a protonated water species could favor the hydration process, as shown in simulations of other systems (29), we performed MD simulations with an explicit H₃O⁺ species introduced in the cavity next to Asp66^{ND3} or Glu34^{ND4L} and modeling the respective carboxylate in the deprotonated state (*SI Appendix, SI Methods*). During the 110-ns simulations of these charge-separated states, the cavity remains dry (*SI Appendix, Fig. S5 L–O*), supporting that the thermodynamic cost of hydrating this region could be high in the “deactive” state. The combined simulation results thus support that the proton pathways are likely to be blocked in the “deactive” state also on timescales comparable to the turnover of complex I.

To probe the bond formation/breaking energetics during the proton transfer reaction, we next performed hybrid density functional theory (DFT)-based QM/MM free energy simulations as well as quantum chemical DFT calculations based on the MD simulation trajectories (*Materials and Methods*). These QM/MM free energy simulations are expected to describe local free energy changes due to their limited sampling in comparison to computationally faster reactive force field (29–33) or semiempirical calculations (34, 35), derived by pre-parametrizations of the protein residues and reaction intermediates involved (33, 36, 37). Hybrid functionals, such as the dispersion-corrected B3LYP-D3 used here, often provide a good balance between computational accuracy and cost, with an error of a few kilocalories \cdot mole⁻¹ based on benchmarking calculations relative to correlated ab initio calculations (*SI Appendix, Table S7*; refer to ref. 38 for further discussion on challenges in DFT-modeling). Here, we computed the free energies based on 320-ps DFT-based QM/MM umbrella sampling (US) simulations that yielded converged local free energy profiles, with structures selected from the microsecond MD simulations to account for dynamical effects (*Materials and Methods*). We expect that the explored free energies provide an accurate description of the local proton transfer energetics in the ND1/ND4L/ND3/ND6 region, although slow-relaxing degrees of freedom arising from possible large-scale dynamical effects are difficult to accurately account for with our current models. The MD simulations are here limited to a total sampling of around 15 μ s, which is unlikely to capture possible slower conformational dynamics on the millisecond turnover timescale. However, part of this problem is possibly circumvented, as we have modeled several transient states along a putative pumping cycle and created simulations of the “active” and “deactive” states based on experimental data.

In the “active” state, we obtain an endergonic proton transfer reaction ($\Delta G \sim +6$ kcal \cdot mol⁻¹) between Asp66^{ND3} and Glu34^{ND4L} followed by a nearly isoenergetic proton transfer to Glu70^{ND4L} ($\Delta G \sim -1$ kcal \cdot mol⁻¹) (Fig. 3 *E* and *F*). The overall reaction barrier is around 15 kcal \cdot mol⁻¹, suggesting that the reaction is kinetically feasible in millisecond timescales based on transition state theory but close to the turnover timescale of complex I. The proton transfer reaction takes place in both our DFT reaction pathway optimizations and QM/MM models via a Grotthuss-type transfer process, with Zundel/hydronium intermediates (*Materials and Methods* and *SI Appendix, Fig. S8*). However, further exploration of other reactions coordinates [e.g., the center of excess charge (29–31, 34, 35) or energy difference between diabatic states (32)], often used in semiempirical or reactive force field calculations, are beyond the scope of the present work but could provide additional insight into the proton transfer dynamics on longer simulation timescales, and beyond pre- and post-proton transfer water configurations.

This endergonic reaction step for the Asp66^{ND3} \rightarrow Glu34^{ND4L} transfer could be powered by the oxidoreduction chemistry and subsequent Q dynamics that has been suggested to release part of the 800-mV ($\Delta G = -18.4$ kcal \cdot mol⁻¹) driving force when the quinol docks into the second membrane-bound binding site (1, 21). The reaction energetics could be further modulated by both hydration and conformational states of the channel.

In contrast to our “active” state simulations, the proton transfer reaction is blocked in the “deactive” state by the TM3^{ND6} helix rotation that prevents protonic connectivity between Asp66^{ND3} and Glu34^{ND4L} (Fig. 2 *C* and *F* and *SI Appendix, Fig. S5 E–H*). Due to the absence of this pathway, we are unable to probe the proton transfer free energy profiles in the “deactive” state model, as the water wire forms a prerequisite for the reaction.

We find that the proton transfer energetics in the “active” state is further modulated by the conformational state of the Glu34^{ND2}/Lys105^{ND2} ion pair in ND2, with the ion pair opening favoring the proton transfer reactions toward the ND2 interface

(Fig. 3*F* and *SI Appendix, Fig. S8H*; *Materials and Methods*). Protonation of Glu70^{ND4L} could further trigger dissociation of the Glu/Lys ion pair, push the proton laterally toward the ND2 subunit (Fig. 3*E*), and initiate a long-range charge transfer cascade in the membrane domain in an electrostatic-cradle motion, as suggested before (1).

Perturbation of the Deactivation Gate by Mitochondrial Disease-Related Mutations. Bioinformatic analysis suggests that key residues along the proposed proton pathways at the ND3/ND4L/ND6-ND2 interface are conserved and involved in multiple human mitochondrial disorders (10, 11, 39, 40) (*SI Appendix, Tables S5 and S6* and Fig. S9 *A–C*). To probe how these residues could affect the proposed proton pathways, we modeled disease-causing point mutations in silico by locally relaxing the structure by short MD simulations based on the hydrated 1- μ s simulation of the “active” state (*SI Appendix, Fig. S9* and Table S1). The models are likely to account for local relaxation effects around the channel, but do not capture, for example, possible large-scale conformational changes linked to the mutations.

We find that substitution of Glu34^{ND4L} for alanine, which has been linked to Leber’s Hereditary Optic Neuropathy (LHON) (39, 40), blocks the proton pathway between Asp66^{ND3} and Tyr59^{ND6}/Glu34^{ND4L} (*SI Appendix, Fig. S9E*), a substitution that also experimentally strongly inhibits the activity of *Escherichia coli* complex I (40) (*SI Appendix, Fig. S9I*). Substitution of Glu34^{ND4L} for glutamine results in a well-hydrated cavity, but the amide headgroup prevents the residue from participating in the proton transfer reaction (*SI Appendix, Fig. S9F*) that could explain the drastic loss of activity of E34Q^{ND4L} in *E. coli* complex I (*SI Appendix, Fig. S9 H and I*) (39, 40). Similarly, the E70A^{ND4L} substitution lowers the complex I activity by around 20% (*SI Appendix, Fig. S9I*), whereas in our MD simulations of this mutant, we observe a longer water-mediated wire to Glu34^{ND2} that could increase the reaction barriers (*SI Appendix, Fig. S9G*). Upon substitution of Glu34^{ND2} for aspartate, the channel remains well hydrated, consistent with experimental activities that are comparable to that of wild-type complex I (*SI Appendix, Fig. S9 D and H*).

Interestingly, several point-mutations linked to LHON and other mitochondrial disorders introduce substitutions that result in hydrogen-bonded contacts with either water molecules within the proton wires and/or the putative proton donor/acceptor groups identified in the MD simulations that could perturb the proton transfer kinetics (*SI Appendix, Fig. S9 B and C*). Similar mutations around the D-channel in cytochrome *c* oxidase lead to uncoupling of the redox-driven proton pump (41).

Quinone and Loop Dynamics. To probe the crosstalk between the proton pathways in the proposed ND3/ND4L/ND6 gating region and the quinone site, we next analyzed the dynamics of Q₁₀ modeled in its binding cavity (Fig. 4). We find that Q₁₀ can form contacts with His59^{NDUFS2}/Tyr108^{NDUFS2} and stays bound within the 30-Å-long Q-tunnel during the MD simulations (Fig. 4 *A* and *C* and *SI Appendix, Fig. S10*). Q₁₀ dissociates during some of the simulations from its original position to a location ca. 5 to 10 Å from His59^{NDUFS2}/Tyr108^{NDUFS2}. Both binding modes are supported by several independent shorter MD simulations (*SI Appendix, Fig. S10 C and D*) that are overall consistent with results from previous studies on the simpler bacterial complex I (20, 21, 42–45). Moreover, upon loss of atomistic detail, our coarse-grained (CGMD) simulations suggest that the Q₁₀ could also stay bound on longer (10- μ s) timescales (*SI Appendix, Fig. S10 C and D*). We note that Q₁₀ can find a stable binding pose at around 10 Å from Tyr108^{NDUFS2}, resembling a previously described binding site 1’ observed in computational studies (21, 42–45) (Fig. 4*C*), as well as in an experimentally resolved inhibitor-bound form of complex I (46). Our MD simulations suggest that

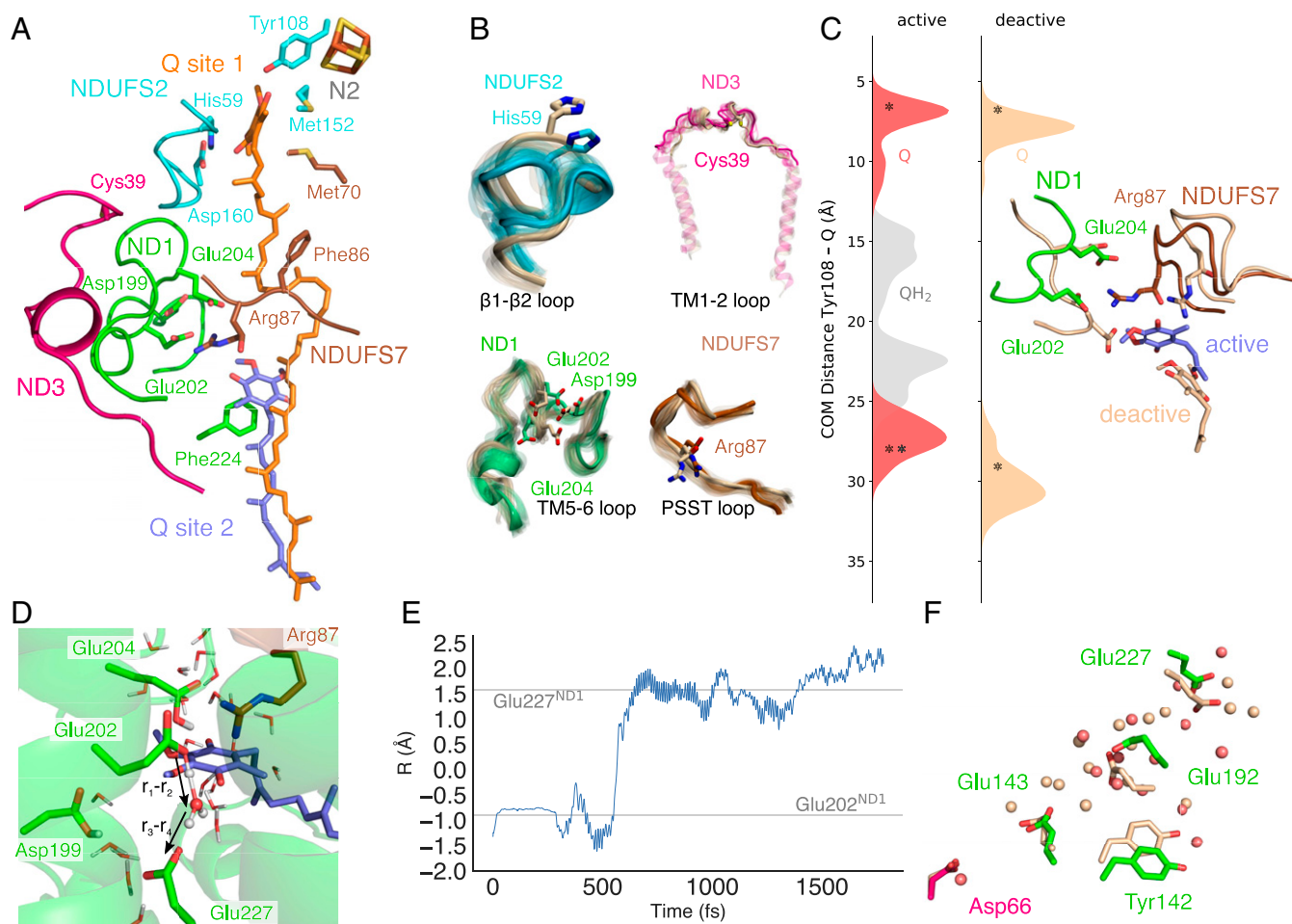


Fig. 4. Conformational changes, Q dynamics, and proton transfer around the substrate channel region. (A) Key regions interacting with Q₁₀ in the primary (site 1) and membrane-bound (site 2) Q-binding sites. (B) Conformational changes in loop regions during 1,000-ns MD simulations of the “active” (in subunit color) and “deactive” (in light brown) states. (C) Q₁₀ dynamics measured from the center-of-mass between the Q and Tyr108^{NDUFS2} headgroups in the “active” and “deactive” states with the Q/QH₂ modeled in the first/second binding regions. Starting positions are marked with an asterisk. (D and E) Proton transfer from Glu202^{ND1}/QH⁻ to Glu227^{ND1} during QM/MM MD simulations. The proton transfer coordinate, $R = r_1 - r_2 + r_3 - r_4$, was computed as a difference between bond-forming (H•••O) and bond-breaking (O-H) distances. (F) Water molecules and conformational changes in the “active” (red/colored) and “deactive” (light brown) states leading from Glu227^{ND1} to Asp66^{ND3}.

both quinone and quinol can also interact with residues around a putative membrane-bound Q-site (Fig. 4A and C), which could be important for transmitting the redox signal to the membrane domain (1, 21). In this second binding region, the Q₁₀ headgroup forms contact with Arg87^{NDUFS7} and Glu204^{ND1}/Glu202^{ND1} together with a network of other charged residues (Fig. 4C and *SI Appendix*, Fig. S11), but Q remains overall dynamically flexible, possibly due to its rather shallow free energy surface (21, 43–45). Glu202^{ND1} and Glu204^{ND1} were previously suggested to undergo conformational changes upon Q reduction (47), and the residues are observed in different conformations in various experimentally resolved structures of complex I (46, 48). These carboxylates, together with Asp199^{ND1}, are located on the flexible TM5-6 loop region of ND1 that contains several charged residues. This loop samples different conformational dynamics in our “active” and “deactive” simulations (Fig. 4A–C and *SI Appendix*, Fig. S12) (14–16, 23): in the “active” state, part of this loop transitions into an α -helical segment around residues 209 to 216 (Fig. 4B and *SI Appendix*, Fig. S12B) that is stabilized by a network of serine residues (Ser209/123/125/128^{ND1}), whereas it remains unstructured and dynamically more flexible in the “deactive” state simulations (Fig. 4B). We note that Glu202^{ND1} and Glu204^{ND1} form water-

wired contacts with QH₂ during the MD simulations (*SI Appendix*, Figs. S6, S10, and S11). Our QM/MM MD simulations (*Materials and Methods*) suggest that the water molecules can catalyze proton transfer from the quinol to this carboxylate cluster and result in the formation of a QH⁻ species (Fig. 4D), but the pK_a of this cluster is likely to be sensitive to the conformation of the surrounding loop regions and Arg87^{NDUFS7}, which also interacts with Glu202^{ND1}, Glu204^{ND1}, and Q₁₀ in the simulations (*SI Appendix*, Fig. S11). The carboxylate cluster is connected via water molecules to a rich network of protonatable residues within the TM5-6 loop region that could transfer the proton to Glu227^{ND1} and further to Glu192^{ND1} (Fig. 4D–F). In the “active” state simulations, Glu192^{ND1} forms contacts with Tyr142^{ND1}, but the tyrosine swings outwards in the “deactive” simulations (Fig. 4F). The Glu192^{ND1}/Tyr142^{ND1} cluster is bridged via one to two water molecules to Glu143^{ND1} that in turn forms a water wire to Asp66^{ND3} next to the deactivation gate, described above (Fig. 4F).

In addition to the TM5-6 loop of ND1, the “active” and “deactive” state simulations sample different conformational dynamics around the β 1- β 2 loop of NDUFS2 that comprises the Q-binding His59^{NDUFS2} and a loop region in NDUFS7 that holds Arg87^{NDUFS7} (the “PSST-loop”, Fig. 4B and *SI Appendix*, Fig.

S12). Motion in the β 1- β 2 loop swings His59^{NDUFS2} out from the binding pocket in the “deactive” state simulations that could reduce the affinity of Q for the primary binding site. These regions were also previously suggested to become more dynamic in the “deactive” state based on loss of electron density in the D-state maps in the ND1, ND3, and NDUFS2 regions (*SI Appendix, Fig. S12A*) (16, 23), and the PSST-loop has also been linked to Q-binding (49). The conformational changes around the TM5-6 loop of ND1, the PSST-loop of NDUFS7, and the helices embracing the Q cavity could account for the partial dissociation of the Q₁₀ from its membrane-bound site that we observe in our “deactive” state simulations but not in the “active” state (Fig. 4C). We also note some enhanced conformational dynamics in the TM1-2 loop of ND3 that contains a solvent exposed Cys-residue (Fig. 4B), a region previously shown to be relevant for coupling between the Q-oxidoreduction function and proton pumping activity (50). However, it is unlikely that our microsecond simulations exhaustively sample the rich dynamics of these loop regions.

These local conformational changes around the Q tunnel are accompanied by motion in some supernumerary subunits, in particular NDUFA10 and NDUFA5 that sample different conformational dynamics in our “active” and “deactive” state simulations, further supporting their possible functional role in the A/D transition (16, 18). In our MD simulations, this shift leads to a rearrangement of a network of ion pairs at the subunit interface that could stabilize the respective states in specific conformations (*SI Appendix, Fig. S2 B–D*).

Discussion

We have shown here by combination of multiscale molecular simulations with cryoEM data how structural changes between “active” and “deactive” states of the mammalian complex I could lead to a functional decoupling of the redox domain from the proton-pumping modules. We observe a molecular gate that is established by conformational changes at the ND3/ND4L/ND6 interface that regulate proton transfer toward the ND4L/ND2 interface and vice versa under RET conditions (Fig. 5). The

protonation-driven hydration dynamics could be driven by the electric field created by the proton itself that, in turn, interacts with the dipole field of the water wire to minimize the overall energetic cost (51). Similar field-induced gating effects could also be employed in other energy-transducing proteins (1, 52, 53). We further suggest that protonation of the ND2 interface triggers long-range proton pumping by sequential ion pair opening, proton transfer reactions, and changes in the hydration state of the proton channels (Fig. 5) (1, 24–27).

We found here that the membrane-bound Q-binding site within ND1 (19, 21, 42–45, 54) is coupled with protonatable residues in the ND3/ND4L/ND6 region. The protonation cascade could be initiated by the formation of a QH⁺ species that remains locked at the membrane-bound Q-site (e.g., until a protonation backwave restores and releases the QH₂) (1). Interestingly, several residues involved in the proposed deactivation gate are linked to human mitochondrial disorders, further supporting the functional relevance of the region (11, 39, 40).

We expect that the membrane-bound Q-site and its linked protonation reactions are unable to equilibrate when the deactivation gate is in a closed state, which could in turn inhibit redox turnover during the complex I deactivation and RET conditions. Conformational changes around the Q channel and its surrounding TM helices and loop regions could further block the substrate entry to complex I and modulate the substrate affinity. In addition to the changes in the proton pumping and local Q-binding region explored here, it is important to note that complex I deactivation is likely to have also more global effects (e.g., by affecting the equilibrium redox states of the cofactors in the peripheral arm and the association of NADH with the protein) (12, 13, 23).

The bacterial complex I lacks a prevalent A/D transition (12, 13, 55); although the functional residues providing wiring between ND1 (Nqo8/NuoH in *Thermus thermophilus*/*E. coli*; *SI Appendix, Table S8*) and ND2 (Nqo14/NuoN in *Thermus thermophilus*/*E. coli*) are highly conserved (*SI Appendix, Table S5*), Met63^{ND6}, which is involved in blocking the proton wire at the ND4L/ND6 interface, is replaced by bulky nonpolar residues in the bacterial

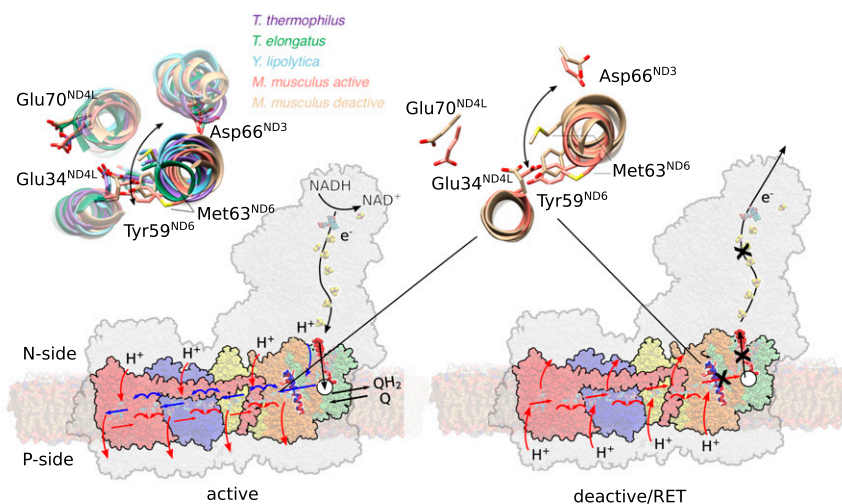


Fig. 5. Deactivation and proton pumping in the mammalian complex I. (*Left*) Quinone reduction triggers conformational and electrostatic changes around ND1 (in green), charging up the proton pump. Upon quinol binding to the membrane bound site (white circle), a proton is pushed from the second Q binding region along the ND3/ND4L/ND6 channel to the ND4L/ND2 interface (orange). The proton transfer can propagate only when the deactivation gate at TM3^{ND6} is in an open state. Protonation changes at the ND4L/ND2 interface trigger (blue arrows) consecutive opening of ion pairs, protonation, and hydration changes in ND2 (yellow), ND4 (blue), and ND5 (red). Upon propagation of the signal in a backward wave to the ND2 interface (red arrows), the loaded protons are released to the P-side, and QH₂ is released to the membrane. (*Right*) Upon RET conditions, rotation of the TM3^{ND6} (helix in red/blue) blocks the proton transfer signal to propagate between ND2 and the second quinone-binding site (cross) that decouples the *pmf*-driven quinone oxidoreduction. (*Inset*) Conformational changes around TM3^{ND6} in the “active” state (PDB ID: 6TZQ)/“deactive” state model (*Right*) and comparison of the region in various experimental structures (*Left*), some of which lack the A/D transition.

isoforms that could also serve a similar gating function. Based on these observations, we suggest that the homologous ND4L/ND6 interface undergoes similar but perhaps a more subtle conformational change also in the bacterial complex I, to enable proton transfer toward the ND2 (Nqo14/NuoN) interface (Fig. 5, *Inset*, *SI Appendix*, Fig. S13). Moreover, NDUFA5 (NUFM/B13), which was found to undergo conformational changes between our “active” and “deactive” simulations, is not present in *Yarrowia lipolytica* complex I (18, 46) and could result in smaller differences between the two states, as also suggested by recent cryoEM structures (17, 28, 54).

Our combined findings indicate that at least some conformational changes linked to the “deactive” state in the mammalian complex I could be involved in the native catalytic cycle of all canonical members of the complex I superfamily, some of which lack the A/D transition (Fig. 5 and *SI Appendix*, Fig. S13) (17). We suggest that the proposed deactivation gate at the ND3/ND4L/ND6 interface modulates proton pumping but has further evolved by fine-tuned modular adaptation in the mammalian complex I to enable enzyme regulation and protect against ROS formation under rapidly changing respiratory chain conditions.

Materials and Methods

Classical MD Simulations. The 3.0-Å piericidin-bound active state cryoEM structure of mouse complex I (PDB ID:6ZTQ) was embedded in a lipid membrane comprising a 2:2:1 mixture of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-ethanolamine (POPE), and 1,3-bis(sn-3'-phosphatidyl)-sn-glycerol (cardiolipin). Ubiquinone or ubiquinol was modeled in the first or second binding site based on starting positions extracted from earlier studies (21). For simulations S1 through S15 (*SI Appendix*, Table S1), the N2 iron-sulfur center was modeled in a reduced state, whereas all other iron-sulfur clusters were modeled in their oxidized states. In simulations S16/S17, the N2 center was also modeled in the oxidized state. The quinone was modeled in both the oxidized (Q) and reduced (QH₂) states and flavin mononucleotide (FMN) in its oxidized form. Initial protonation states of titratable groups were determined based on Poisson-Boltzmann continuum electrostatic calculations coupled to Monte Carlo sampling (*SI Appendix*, Table S2) using the adaptive Poisson-Boltzmann solver (56) in combination with Karlsberg+ (57, 58). Atoms were described by point charges, embedded in a dielectric continuum with $\epsilon = 4$ employed to model the protein surroundings. Bulk water was modeled with $\epsilon = 80$, with a solvent probe radius of 1.4 Å and ionic strength of 100 mM potassium chloride. Complex I was divided into three overlapping parts (*SI Appendix*, Fig. S1B), which were titrated independently. The MD simulations were performed by considering multiple different protonation states along the observed proton pathways in the ND1/ND4L/ND6/ND3 region (*SI Appendix*, Table S2). To this end, we sequentially moved the proton along the observed proton wire in this region in addition to simulations in which key E-channel residues (Glu34^{ND4L}, Glu70^{ND4L}, and Asp66^{ND3}) were modeled in both their fully deprotonated and protonated states (*SI Appendix*, Table S1). Arg87^{NDUF57} was modeled in its γ -hydroxylated form and Arg85^{NDUF52} as symmetric dimethylarginine. The model was solvated with TIP3P water molecules, and 632 Na⁺ and 497 Cl⁻ ions were added to neutralize the system to achieve a 150 mM NaCl concentration. The full system comprises ca. 991,000 atoms. Atomistic MD simulations were carried out using the CHARMM36 force field (59) in combination with DFT-derived parameters for the cofactors (20, 26, 27). The deactive state model was prepared using the same procedure based on the cryoEM density of the deactive state map by MDFF (22) with the cryoEM density map of the “deactive” state (Electron Microscopy Data Bank, EMDB: 4356) (see section *Atomic Model of the “Deactive” State*). The MD simulations were propagated in an *NPT* ensemble at $T = 310$ K using the Langevin thermostat at $P = 1$ atm pressure with Nosé-Hoover-Langevin piston pressure control for 1 μ s with a 2-fs integration timestep. Long-range electrostatics were treated using the Particle Mesh Ewald method with a grid size of 1 Å and by computing the long-range contributions in reciprocal space for pair-distances above 12 Å. The Lennard-Jones cutoff was set to 12 Å using Nanoscale Molecular Dynamics' (NAMD) built-in switching function starting at 10 Å, and rigid bonds were modeled using the ShakeH algorithm. Classical MD simulations were performed in different states and in duplicates, leading to a total MD simulation time ca. 15 μ s. All MD simulations were carried out with NAMD version 2.12 (60). Visual Molecular Dynamics (61) and PyMol v. 2.4.1 (62) were used for analysis and visualization, and further clustering analysis of the hydration dynamics was performed using WATCLUST (63). Hydration content (Fig. 2I and *SI Appendix*,

Fig. S5 I, J, and K) was also determined based on MD hydration levels along tunnel coordinates identified using CAVER (64). Tunnels corresponding to productive pathways were extracted based on snapshots extracted every 1 ns of “active” and “deactive” MD simulations using Caver version 3.0 (64) and a probe radius set to 0.9 Å. The gathered tunnel coordinates were evaluated for their individual hydration in a 2 Å sphere over the MD trajectory and visualized with a surface proportional to the average hydration over the MD trajectory. Hydration frequencies (Fig. 2I and *SI Appendix*, Fig. S6) were computed by identifying connections between the head groups of functional residues using CAVER3.0 and a probe radius of 0.9 Å. The “hydrated-% of trajectory” was obtained by counting the fraction of water molecules within 2 Å from each point along the tunnel coordinate during the simulation trajectory. Refer to *SI Appendix*, Table S1 for summary of classical MD simulations.

Atomic Model of the “Deactive” State. A “deactive” state model of the mouse complex I was built based on “active” state mouse complex I (PDB ID:6ZTQ) model (19) that was targeted to the complete cryoEM density map of the “deactive” state (EMDB: 4356) (16) using the MDFF (22) approach to construct an intact simulation model. The MDFF approach converts the cryoEM density map (ϕ) into an external restraining potential according to the following:

$$V(r) = \sum_i w_i \phi(r_i)$$

and

$$\phi(r) = \begin{cases} \zeta \left[1 - \frac{\phi(r) - \phi_{\text{thr}}}{\phi_{\text{max}} - \phi_{\text{thr}}} \right] & \text{for } \phi(r) \geq \phi_{\text{thr}}, \\ \zeta & \text{for } \phi(r) < \phi_{\text{thr}} \end{cases}$$

where $\phi(r)$ is the cryoEM density map at position r , ϕ_{thr} is the threshold density used for flattening the solvent beyond the solvent peak (here $\phi_{\text{thr}} = 0$), and w_i and ζ are scaling factors that tune the strength of the cryoEM map (here, w_i is an atom-specific scaling factor set to the atomic mass and $\zeta = 0.3$). Harmonic restraints were placed on the protein secondary structure elements during the MDFF simulations. Restraints for ND6 included the first helix turn-up to residue 65 to account for unresolved density (*SI Appendix*, Table S3). The active state model was targeted into the “deactive” state by MDFF minimization followed by 5 ns of MDFF dynamics in implicit solvent (22). The model was characterized by computing cross-correlation of computed and experimental cryoEM densities and RMSD differences to experimentally refined models (*SI Appendix*, Fig. S2 A and B). The obtained deactive state model was studied in subsequent unrestrained MD simulation using the same setup as for the active state (*SI Appendix*, Table S1).

FEP Calculations. The hydration free energy of the gating region was estimated using alchemical FEP calculations using a complex I model comprising 120,000 atoms constructed from the MD simulations of the “active” and “deactive” states. Briefly, a chain comprising five water molecules was created between Asp66^{ND3} and Glu34^{ND4L} in 20 λ -steps during 220-ns MD simulations. The free energy was computed according to the following:

$$\Delta G_{A \rightarrow B} = -k_B T \ln \left\langle \exp \left[-\frac{E_B - E_A}{k_B T} \right] \right\rangle,$$

where $\Delta G_{A \rightarrow B}$ is the free energy difference between states A and B, k_B is the Boltzmann constant, T is the temperature, $E_B - E_A$ is the energy differences state A and B, and $\langle \rangle$ is the ensemble average for the simulation propagated in state A. Free energies based on forward and backward sampling were obtained via Bennet's acceptance ratio. The hydration free energy in Fig. 2J is reported relative to removal of five water molecules from the bulk solvent. Refer to *SI Appendix* for detailed methods description of the FEP calculations and *SI Appendix*, Fig. S7 for decomposition and convergence of the FEP calculations.

Exploration of the Quinone Dynamics. To further probe the quinone/quinol dynamics we performed additional classical MD simulations as well as CGMD based on a model constructed from the complete mouse complex I, comprising chains A, B, C, D (residues 34 to 430), H, I, J, K, P, a, b, e, r, W, X (residues 1 to 150), and Z. This ~396,000-atom model was embedded in a POPC membrane, solvated with TIP3P water molecules, and 100 mM NaCl. The Q/QH₂ near Tyr108^{NDUF52} was modeled based on the cryoEM model (simulations S62 and S63), or docked in (simulations S64 to S77) based on the local binding minima identified in simulations of *T. thermophilus* complex I (21).

After minimization, the system was heated up to $T = 310$ K with harmonic restraints ($2 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$) on the entire protein structure and cofactors (0.5 ns). In a second step, restraints were applied only to the backbone atoms (1 ns), followed by release of restraints. Simulations were performed with the same simulation parameters as applied for the full-sized model, using NAMD 2.9/2.13 (60). $10 \times 10 \mu\text{s}$ CGMD simulations (simulations S78 to S87) were further performed with Q docked to different locations in the substrate tunnel using the MARTINI 2.2 force field (65), by employing a 20-fs timestep, a Parrinello–Rahman barostat with $\tau_p = 12.0$ ps, and a compressibility of $\chi = 3.0 \times 10^{-4} \text{ bar}^{-1}$. CGMD simulation were performed at $T = 310$ K with a thermostat coupling constant of $\tau_t = 1.0$ ps. Nonbonded interactions were treated using a cutoff distance of 11 \AA and $\epsilon = 15$. All CGMD simulations were performed using Gromacs 2016.3 (66). Refer to *SI Appendix, Table S1* for a list of performed simulations.

QM/MM and DFT Calculations. QM/MM MD simulations were performed to explore the Q binding near the membrane-bound binding site (site 2) as well as for the proton transfer in the ND3/ND4L/ND6 region. For probing the dynamics in Q site 2, a QM/MM model including the quinol and nearest residues and water molecules with $n = 208$ QM atoms was used (*SI Appendix, Table S4*). The QM region was treated at the B3LYP-D3/def2-SVP level (67–69) using a 1-fs integration timestep at $T = 310$ K. The remaining MM system was modeled at the CHARMM36 level in combination with in-house DFT-derived parameters for the cofactors (20, 26, 27). Link atoms were introduced at the C α -C β position, with exception of arginine residues, in which the link atom was placed at C β -C γ position, whereas Q₁₀ was cut at the C₉-C₁₁ bond and saturated with a link atom. The total QM/MM system was trimmed to include around 19,527 atoms. Refer to *SI Appendix, Fig. S1* and Table S4 for further details.

QM/MM free energy calculations and DFT cluster models of the proton transfer at the ND4L/ND3/ND6/ND2 interface were built based on MD snapshots of the “active” state (simulation S21 after 1,000 ns), with protonic connectivity from Asp66^{ND3} to Glu34^{ND4L}/Tyr59^{ND6} and from Glu34^{ND4L} to Glu70^{ND4L}. QM/MM US simulations were performed by splitting the proton transfer reaction into two steps with 90 and 139 QM atoms, respectively (*SI Appendix, Table S4*). The QM region was coupled to the MM region (modeled at the CHARMM36 level), by link atoms that were introduced at the C α -C β positions. The QM/MM/US simulations comprised 25 and 19 windows for the first and second pT step, respectively. Each US window was sampled for 5 ps at $T = 310$ K using a restraint of $100 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ on the proton transfer reaction coordinate, $R_1 = [-3 \text{ \AA}, +3 \text{ \AA}]$ and $R_2 = [-2.5 \text{ \AA}, +2.5 \text{ \AA}]$, for the first and second part with the Glu34/Lys105^{ND2} model in the closed and

open states. The free energy profiles were calculated using the weighted histogram analysis method. Refer to *SI Appendix, Fig. S8* for the definition of R_1 and R_2 and convergence of the free energies. The same proton transfer reactions were also studied using DFT cluster models but with larger QM regions comprising 190 to 250 QM atoms (*SI Appendix, Table S4*). Protein residues were cut at the C α -C β position and saturated with hydrogen atoms. C β positions and some carbon atoms of aliphatic residues surrounding the proton transfer pathway were kept fixed during geometry optimizations. Geometry optimizations were performed at the B3LYP-D3/def2-SVP level of theory (67–69), using an implicit polarizable medium with $\epsilon = 4$. Reaction pathways of the proton transfer reactions in the DFT models were optimized using minimum energy pathway optimization approach (70) that resembles the zero-temperature string method (*SI Appendix, SI Methods*). Similar Grothuss-type transfer reactions are predicted in both the QM/MM models using a geometric reaction coordinate and the DFT models employing reaction pathway optimization, suggesting that the choice of the reaction coordinate is not sensitive to the predicted transfer process (*SI Appendix, Fig. S8*). The energetics of the proton transfer processes were calculated at B3LYP-D3/def2-TZVP/ $\epsilon = 4$ level by combining the two optimized DFT models of the first (Asp66^{ND3} to Glu34^{ND4L}) and second (Glu34^{ND4L} to Glu70^{ND4L}) proton transfer reactions, with a combined model size of ca. 380 atoms (Fig. 3E). The DFT energetics was benchmarked against correlated ab initio theory model calculations, suggesting that the reaction barriers at the B3LYP-D3 level are underestimated by a few kilocalories $\cdot \text{mole}^{-1}$ (*SI Appendix, Table S7*), whereas the limited free energy sampling is likely to have an opposite effect. All QM/MM or DFT calculations were performed using the CHARMM/TURBOMOLE interface (71) CHARMM version 38 (72), and/or TURBOMOLE version 7.2 to 7.5 (73), and ORCA version 4.2 (74).

Data Availability. All study data are included in the article and/or *SI Appendix*.








ACKNOWLEDGMENTS. This work received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program/Grant Agreement 715311. The project was also supported by the Knut and Alice Wallenberg Foundation. We are thankful for the computing time provided by Partnership for Advance Computing in Europe (PRACE) for awarding us access to MareNostrum hosted by Barcelona Supercomputing Center. This work was also supported by the Swedish National Infrastructure for Computing (SNIC 2020/1-38) at Center for High Performance Computing (PDC) Centre, partially funded by the Swedish Research Council through Grant Agreement 016-07213.

- V. R. I. Kaila, Long-range proton-coupled electron transfer in biological energy conversion: Towards mechanistic understanding of respiratory complex I. *J. R. Soc. Interface* **15**, 20170916 (2018).
- V. R. I. Kaila, M. I. Verkhovskiy, M. Wikström, Proton-coupled electron transfer in cytochrome oxidase. *Chem. Rev.* **110**, 7062–7081 (2010).
- U. Brandt, Energy converting NADH:quinone oxidoreductase (complex I). *Annu. Rev. Biochem.* **75**, 69–92 (2006).
- M. Wikström, V. Sharma, V. R. I. Kaila, J. P. J. Hosler, G. Hummer, New perspectives on proton pumping in cellular respiration. *Chem. Rev.* **115**, 2196–2221 (2015).
- J. Hirst, Mitochondrial complex I. *Annu. Rev. Biochem.* **82**, 551–575 (2013).
- L. A. Sazanov, A giant molecular proton pump: Structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **16**, 375–388 (2015).
- P. Mitchell, Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144–148 (1961).
- M. Yoshida, E. Muneyuki, T. Hisabori, ATP synthase –A marvellous rotary engine of the cell. *Nat. Rev. Mol. Cell Biol.* **2**, 669–677 (2001).
- K. R. Pryde, J. Hirst, Superoxide is produced by the reduced flavin in mitochondrial complex I: A single, unified mechanism that applies during both forward and reverse electron transfer. *J. Biol. Chem.* **286**, 18056–18065 (2011).
- M. P. Murphy, How mitochondria produce reactive oxygen species. *Biochem. J.* **417**, 1–13 (2009).
- K. Fiedorczuk, L. A. Sazanov, Mammalian mitochondrial complex I structure and disease-causing mutations. *Trends Cell Biol.* **28**, 835–867 (2018).
- M. Babot, A. Galkin, Molecular mechanism and physiological role of active-deactive transition of mitochondrial complex I. *Biochem. Soc. Trans.* **41**, 1325–1330 (2013).
- M. Babot, A. Birch, P. Labarbuta, A. Galkin, Characterisation of the active/de-active transition of mitochondrial complex I. *Biochim. Biophys. Acta* **1837**, 1083–1092 (2014).
- J. Zhu, K. R. Vinothkumar, J. Hirst, Structure of mammalian respiratory complex I. *Nature* **536**, 354–358 (2016).
- K. Fiedorczuk *et al.*, Atomic structure of the entire mammalian mitochondrial complex I. *Nature* **538**, 406–410 (2016).
- A. A. Agip *et al.*, Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states. *Nat. Struct. Mol. Biol.* **25**, 548–556 (2018).
- K. Parey *et al.*, Cryo-EM structure of respiratory complex I at work. *eLife* **7**, e39213 (2018).
- A. Di Luca, V. R. I. Kaila, Global collective motions in the mammalian and bacterial respiratory complex I. *Biochim. Biophys. Acta Bioenerg.* **1859**, 326–332 (2018).
- H. R. Bridges *et al.*, Structure of inhibitor-bound mammalian complex I. *Nat. Commun.* **11**, 5261 (2020).
- A. P. Gamiz-Hernandez, A. Jussupow, M. P. Johansson, V. R. I. Kaila, Terminal electron-proton transfer dynamics in the quinone reduction of respiratory complex I. *J. Am. Chem. Soc.* **139**, 16282–16288 (2017).
- J. Warnau *et al.*, Redox-coupled quinone dynamics in the respiratory complex I. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8413–E8420 (2018).
- L. G. Trabuco, E. Villa, K. Mitra, J. Frank, K. Schulten, Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
- J. N. Blaza, K. R. Vinothkumar, J. Hirst, Structure of the deactive state of mammalian respiratory complex I. *Structure* **26**, 312–319.e3 (2018).
- A. Di Luca, A. P. Gamiz-Hernandez, V. R. I. Kaila, Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6314–E6321 (2017).
- O. Haapanen, V. Sharma, Role of water and protein dynamics in proton pumping by respiratory complex I. *Sci. Rep.* **7**, 7747 (2017).
- M. E. Mühlbauer *et al.*, Water-gated proton transfer dynamics in respiratory complex I. *J. Am. Chem. Soc.* **142**, 13718–13728 (2020).
- J. M. Schuller *et al.*, Redox-coupled proton pumping drives carbon concentration in the photosynthetic complex I. *Nat. Commun.* **11**, 494 (2020).
- D. N. Grba, J. Hirst, Mitochondrial complex I structure reveals ordered water molecules for catalysis and proton translocation. *Nat. Struct. Mol. Biol.* **27**, 892–900 (2020).
- Y. Peng, J. M. Swanson, S. G. Kang, R. Zhou, G. A. Voth, Hydrated excess protons can create their own water wires. *J. Phys. Chem. B* **119**, 9212–9218 (2015).
- C. Li, Z. Yue, L. M. Espinoza-Fonseca, G. A. Voth, Multiscale simulation reveals passive proton transport through SERCA on the microsecond timescale. *Biophys. J.* **119**, 1033–1040 (2020).
- R. Liang, J. M. Swanson, Y. Peng, M. Wikström, G. A. Voth, Multiscale simulations reveal key features of the proton-pumping mechanism in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7420–7425 (2016).

32. A. V. Pislakov, P. K. Sharma, Z. T. Chu, M. Haranczyk, A. Warshel, Electrostatic basis for the unidirectionality of the primary proton transfer in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7726–7731 (2008).
33. A. Warshel, R. M. Weiss, An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.* **102**, 6218–6226 (1980).
34. D. Riccardi *et al.*, Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B* **110**, 6458–6469 (2006).
35. P. Goyal, S. Yang, Q. Cui, Microscopic basis for kinetic gating in cytochrome c oxidase: Insights from QM/MM analysis. *Chem. Sci. (Camb.)* **6**, 826–841 (2015).
36. S. Lee, R. Liang, G. A. Voth, J. M. J. Swanson, Computationally efficient multiscale reactive molecular dynamics to describe amino acid deprotonation in proteins. *J. Chem. Theory Comput.* **12**, 879–891 (2016).
37. M. Gaus, A. Goez, M. Elstner, Parametrization and benchmark of DFTB3 for organic molecules. *J. Chem. Theory Comput.* **9**, 338–354 (2013).
38. A. J. Cohen, P. Mori-Sánchez, W. Yang, Challenges for density functional theory. *Chem. Rev.* **112**, 289–320 (2012).
39. M. Kervinen, J. Pätsi, M. Finel, 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).
40. M. C. Kao, E. Nakamaru-Ogiso, A. Matsuno-Yagi, T. Yagi, Characterization of the membrane domain subunit NuoK (ND4L) of the NADH-quinone oxidoreductase from *Escherichia coli*. *Biochemistry* **44**, 9545–9554 (2005).
41. A. S. Pawate *et al.*, A mutation in subunit I of cytochrome oxidase from *Rhodobacter sphaeroides* results in an increase in steady-state activity but completely eliminates proton pumping. *Biochemistry* **41**, 13417–13423 (2002).
42. M. Hoias Teixeira, G. Menegon Arantes, Balanced internal hydration discriminates substrate binding to respiratory complex I. *Biochim. Biophys. Acta Bioenerg.* **1860**, 541–548 (2019).
43. O. Haapanen, A. Djurabekova, V. Sharma, Role of second quinone binding site in proton pumping by respiratory complex I. *Front Chem.* **7**, 221 (2019).
44. C. Gupta *et al.*, Charge transfer and chemo-mechanical coupling in respiratory complex I. *J. Am. Chem. Soc.* **142**, 9220–9230 (2020).
45. A. Jussupow, A. Di Luca, V. R. I. Kaila, How cardiolipin modulates the activity of complex I. *Sci. Adv.* **5**, eaav1850 (2019).
46. V. Zickermann *et al.*, Structural biology. Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* **347**, 44–49 (2015).
47. V. Sharma *et al.*, Redox-induced activation of the proton pump in the respiratory complex I. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11571–11576 (2015).
48. R. Baradaran, J. M. Berrisford, G. S. Minhas, L. A. Sazanov, Crystal structure of the entire respiratory complex I. *Nature* **494**, 443–448 (2013).
49. E. Galemou Yoga *et al.*, Mutations in a conserved loop in the PSST subunit of respiratory complex I affect ubiquinone binding and dynamics. *Biochim. Biophys. Acta Bioenerg.* **1860**, 573–581 (2019).
50. A. Cabrera-Orefice *et al.*, Locking loop movement in the ubiquinone pocket of complex I disengages the proton pumps. *Nat. Commun.* **9**, 4500 (2018).
51. P. Saura, D. M. Frey, A. P. Gamiz-Hernandez, V. R. I. Kaila, Electric field modulated redox-driven protonation and hydration energetics in energy converting enzymes. *Chem. Commun. (Camb.)* **55**, 6078–6081 (2019).
52. S. Supekar, V. R. I. Kaila, Dewetting transitions coupled to K-channel activation in cytochrome c oxidase. *Chem. Sci. (Camb.)* **9**, 6703–6710 (2018).
53. C. M. Suomivuori, A. P. Gamiz-Hernandez, D. Sundholm, V. R. I. Kaila, Energetics and dynamics of a light-driven sodium-pumping rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7043–7048 (2017).
54. K. Parey *et al.*, High-resolution cryo-EM structures of respiratory complex I: Mechanism, assembly, and disease. *Sci. Adv.* **5**, eaax9484 (2019).
55. N. Belevich, G. Belevich, Z. Chen, S. C. Sinha, M. Verkhovskaya, Activation of respiratory complex I from *Escherichia coli* studied by fluorescent probes. *Heliyon* **3**, e00224 (2017).
56. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nano-systems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041 (2001).
57. G. Kieseritzky, E. W. Knapp, Optimizing pKa computation in proteins with pH adapted conformations. *Proteins* **71**, 1335–1348 (2008).
58. B. Rabenstein, G. M. Ullmann, E.-W. Knapp, Calculation of protonation patterns in proteins with structural relaxation and molecular ensembles – Application to the photosynthetic reaction center. *Eur. Biophys. J.* **27**, 626–637 (1998).
59. R. B. Best *et al.*, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
60. J. C. Phillips *et al.*, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
61. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 (1996).
62. W. L. Delano, The PyMOL Molecular Graphics System, Schrodinger. (2002). <https://pymol.org/2/>.
63. E. D. López *et al.*, WATCLUST: A tool for improving the design of drugs based on protein-water interactions. *Bioinformatics* **31**, 3697–3699 (2015).
64. E. Chovanová *et al.*, CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **8**, e1002708 (2012).
65. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
66. H. J. C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).
67. A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
68. A. Schäfer, H. Horn, R. Ahlrichs, Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **97**, 2571–2577 (1992).
69. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
70. P. Plessow, Reaction path optimization without NEB springs or interpolation algorithms. *J. Chem. Theory Comput.* **9**, 1305–1310 (2013).
71. S. Riahi, 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).
72. B. R. Brooks *et al.*, CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
73. R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **162**, 165–169 (1989).
74. F. Neese, F. Wennmohs, U. Becker, C. Riplinger, The ORCA quantum chemistry program package. *J. Chem. Phys.* **152**, 224108 (2020).

Article IV

Design of buried charged networks in artificial proteins

Mona Baumgart¹, Michael Röpke ^{1,3}, Max E. Mühlbauer ^{1,3}, Sam Asami ¹, Sophie L. Mader ¹, Kai Fredriksson¹, Michael Groll ¹, Ana P. Gamiz-Hernandez ^{1,2} & Ville R. I. Kaila ^{1,2}✉

Soluble proteins are universally packed with a hydrophobic core and a polar surface that drive the protein folding process. Yet charged networks within the central protein core are often indispensable for the biological function. Here, we show that natural buried ion-pairs are stabilised by amphiphilic residues that electrostatically shield the charged motif from its surroundings to gain structural stability. To explore this effect, we build artificial proteins with buried ion-pairs by combining directed computational design and biophysical experiments. Our findings illustrate how perturbation in charged networks can introduce structural rearrangements to compensate for desolvation effects. We validate the physical principles by resolving high-resolution atomic structures of the artificial proteins that are resistant towards unfolding at extreme temperatures and harsh chemical conditions. Our findings provide a molecular understanding of functional charged networks and how point mutations may alter the protein's conformational landscape.

¹Center for Integrated Protein Science Munich (CIPSM) at the Department Chemie, Technische Universität München, Lichtenbergstraße 4, 85748 Garching, Germany. ²Department of Biochemistry and Biophysics, Stockholm University, 10691 Stockholm, Sweden. ³These authors contributed equally: Michael Röpke, Max E. Mühlbauer. ✉email: ville.kaila@dbb.su.se

Protein structures are universally folded by shielding hydrophobic amino acids from the aqueous environment to reach a global free energy minimum¹. However, buried charged networks provide central functional elements responsible for the biological activity in many proteins^{2–4}. In complex I, a redox-driven proton pump essential for cell respiration, a 200 Å network of buried ion-pairs and charged residues provide functional elements responsible for the enzyme's proton pumping machinery^{2,5}. Similarly, the heat shock protein 90, an essential molecular chaperone in the eukaryotic cell, employs an extended network of charged amino acids to regulate the enzymatic activity and the chaperone function^{3,6}.

Charged residues buried within the low-dielectric protein core ($\epsilon_{\text{protein}} = 4–10$) are thermodynamically destabilised relative to the high-dielectric aqueous environment ($\epsilon_{\text{aq}} = 80$) by the (Born) desolvation free energy associated with moving charged groups towards lower dielectric surroundings. This effect can in part be compensated by the electrostatic interactions between oppositely charged residues, but recent studies^{4,7} suggest that each ion-pair destabilises the overall native protein fold by up to 5 kcal mol⁻¹. This raises the important question on how proteins with large extended networks with hundreds of charged residues^{2,5} are structurally and thermodynamically stabilised. To address this puzzling question without the potential bias of natural protein scaffolds, we introduce here buried ion-pairs into the hydrophobic core of artificial helical bundle scaffolds^{8,9} by combining directed computational design and molecular simulations with biophysical, spectroscopic, and structural experiments. These experimentally well-behaved, minimal de novo models do not structurally resemble highly intricate natural proteins, but despite their simple structure, four-helical bundles have been used to probe, e.g., oxidoreductase^{9,10}, oxygen-binding⁸, light-capturing¹¹ and ion-transport properties¹². Here they have been used to deduce interaction principles of buried ion-pairs without the complicated effect of complex natural surroundings based on rather simple biophysical and computational experiments.

Results and discussion

Structural stability of buried ion-pairs in artificial bundle proteins. To study how the protein stability is affected by buried charged residues, we introduced a glutamate (Glu)/lysine (Lys) ion-pair into the hydrophobic protein core of an artificial 4 α -helical protein scaffold¹⁰ at positions 49 and 84 (F49E and F84K), where helix 1 forms contacts with helices 2 and 3 (see Fig. 1a, Methods and Supplementary Table 1). Atomistic molecular dynamics (MD) simulations of this *Maquette 1* model suggest that the overall structure remains stable on the simulation timescale (Fig. 1b), and energetically favours the charged (Glu⁻/Lys⁺) state over the neutral (Glu⁰/Lys⁰) state (Supplementary Fig. 1d). During the MD simulations, a few water molecules transiently interact with the buried charges and partially destabilise the hydrophobic core relative to the *Maquette 1* model with a non-polar core (Fig. 1c). To further characterise the properties of the constructs, we expressed the proteins in *E. coli*, and studied their stability using circular dichroism (CD) spectroscopy and chemical unfolding experiments. The proteins remain α -helical, as indicated by the minima at 208 and 222 nm (Fig. 1e). However, consistent with findings on ion-pairs in natural proteins^{4,7}, the introduced charges decrease both the melting temperature by $\Delta T_m > 40$ °C and the protein overall stability by $\Delta\Delta G \approx 5$ kcal mol⁻¹ relative to our model with an intact hydrophobic core (Fig. 1d–g).

Removal of the Glu or Lys counter charges decreases the protein stability by around 3 kcal mol⁻¹ based on unfolding experiments (Supplementary Figs. 1, 2a). However, despite the

well-packed protein models in our MD simulations and a two-state folding transition in the chemical unfolding experiments, our nuclear magnetic resonance (NMR) experiments show heteronuclear single quantum coherence (HSQC) spectra that are not too well-dispersed and contain some overlapping signals (Supplementary Fig. 2f), an effect that may arise from the highly symmetric structure of these models, and/or involvement of possible molten-globule like states¹⁰. Although helix dipole-sidechain charge interactions may affect the protein stability^{13–15}, we find here that inverting the polarity of the ion-pair in the *Maquette 1* model only leads to a minor shift in the melting temperature ($\Delta\Delta T_m = 0.1$ °C) (Supplementary Fig. 1f, 2a). Although we currently lack experimental structural data to draw definite conclusions about the buried ion-pairs in the *Maquette 1* models, our data support that buried charges decrease the overall protein stability similarly to what has been described for natural proteins^{4,7} and consistent with the data for other de novo protein models (see below).

To improve the accuracy of the design, we next employed another artificial protein scaffold (*Maquette 2*) with a resolved core scaffold¹¹, but a different helical topology (helix 1 interacting with helices 2 and 4) as compared to the *Maquette 1* model (Fig. 1h, *inset*). Insertion of the ion-pair at positions 17 and 72 (F17E/F72K relative to *Maquette 2*/hc) yields a stable structure in the MD simulations (Fig. 1i, Supplementary Fig. 3), but similar to the *Maquette 1* model, the charged element forms transient interactions with a few water molecules during the simulations (Fig. 1j). The ion-pair destabilises the protein by $\Delta\Delta G \approx 2$ kcal mol⁻¹ in chemical unfolding experiments (Fig. 1k, n), suggesting an improved stability relative to *Maquette 1*, but an overall destabilisation relative to the construct with the non-polar core. Moreover, the overall protein stability remains unaffected when the charged residues are introduced into the more water accessible region of the bundle (I101E/F50K) (Supplementary Fig. 3c).

Characterisation of buried ion-pairs. The *Maquette 2* proteins have well-dispersed HSQC spectra, but the introduced ion-pair leads to significant chemical shift perturbations (Fig. 2c, Supplementary Fig. 2b), indicating possible structural rearrangements to compensate for the electrostatic penalty. The ion-pair is supported by NH₃-selective heteronuclear in-phase single quantum coherence spectroscopy (HISQC) measurements at 275 K that reveal slow proton exchange of the shielded sidechains. This NMR signal further disappears upon substitution of either Lys72 or Glu17 with phenylalanine (Fig. 2d), whereas the intensity of the peak indicates dynamic flexibility consistent with our MD simulations (Fig. 2b, Supplementary Fig. 3). Based on 3D-NMR experiments at 293 K, we resolved a structural model of *Maquette 2* with the introduced charged element (Fig. 2a, Supplementary Fig. 2c, Supplementary Table 2). The model reveals a subtle bending of helices 1 and 3 that could arise from the thermodynamic cost associated with the introduced charged element (Figs. 1n and 2a). However, the mixed open/closed ion-pair conformations as indicated by NH₃-selective HISQC spectrum (Fig. 2d) and MD simulation data (Fig. 1h, Supplementary Fig. 7b), are not reflected in the model building, which favours dissociation of the ion-pair in the structure calculation (see “Methods” section). Nevertheless, the MD simulations suggest that the closed ion-pair conformation can easily form in the NMR-models by a <1 Å inward motion of helices 1 and 3. These conformational changes are also qualitatively supported by structural modelling (Fig. 2a, Supplementary Fig. 2c). Taken together, our data suggest that the ion-pair introduces some dynamical flexibility and subtle structural rearrangements in the

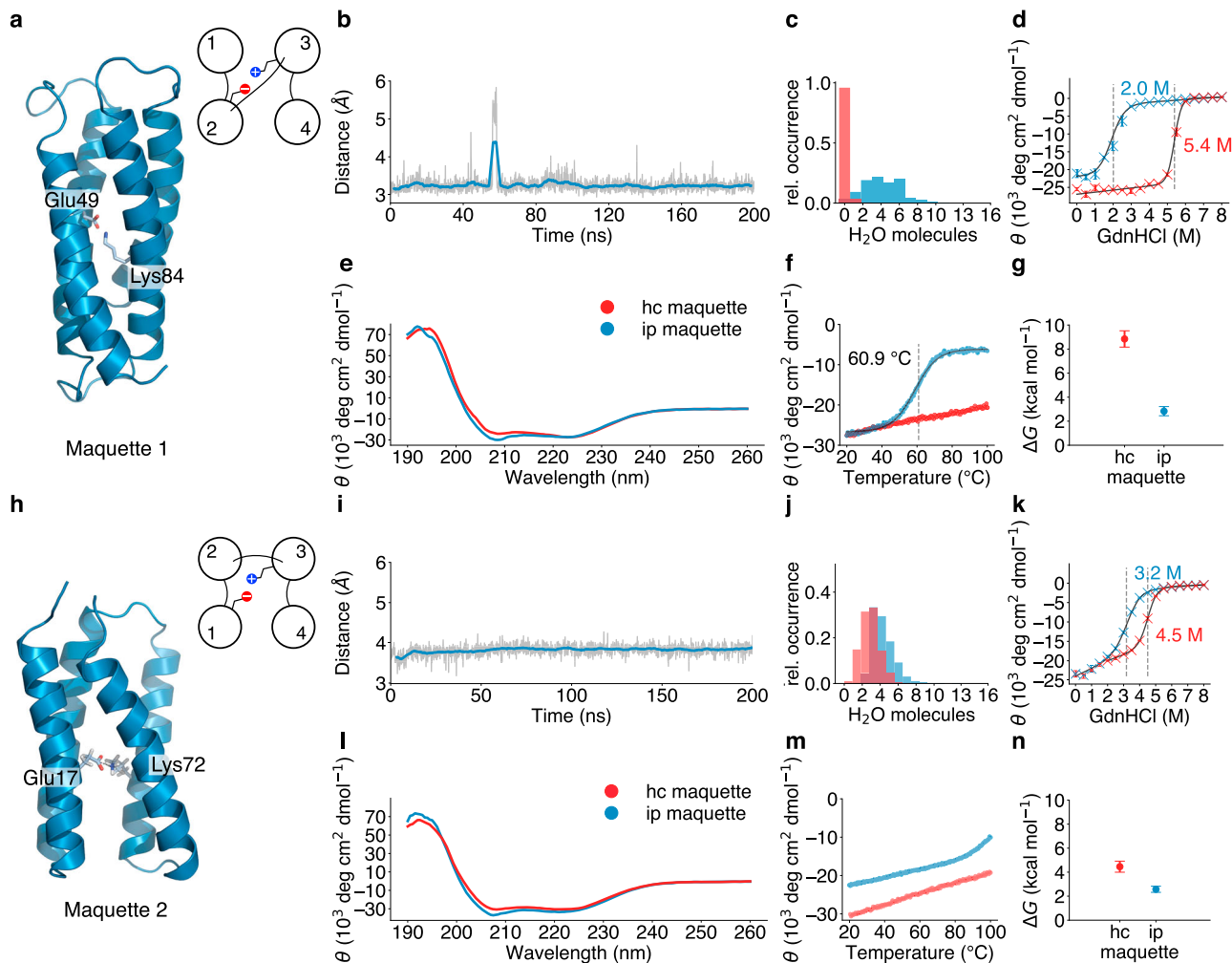


Fig. 1 Structural stability of buried ion-pairs in artificial bundle proteins. **a, h** Structural snapshot of protein models; **b, i** distances between introduced ion-pairs (the shaded line represents the actual data, the coloured line is a moving average over 50 frames), and **c, j** number of buried water molecules during MD simulations within 9 Å of helices 1–4; **e, l** CD spectra at 20 °C of protein *Maquettes* 1 and 2 with a hydrophobic core (hc) and introduced ion-pair (ip); **f, m** melting curves measured by following the temperature dependence of the 222-nm trough in the CD spectra; **d, k** chemical unfolding with GdnHCl up to 8 M monitored by CD spectroscopy. Data points are presented as mean values of triplicates ($n = 3$ independent experiments) with \pm standard deviation shown as error bars; and **g, n** thermodynamic stability (ΔG) values calculated from the chemical unfolding experiments, with error bars computed from the standard deviation of the chemical unfolding profiles.

bundle scaffold, which might enable exploration of both open and closed ion-pair conformations (Fig. 2a, Supplementary Fig. 3). Such dynamic flexibility coupled to subtle conformational changes are critical for the function of ion-pairs in natural proteins^{5,6}.

Analysis and design of buried charged elements in proteins. To better understand the charge-compensation principles, we next developed a computational design approach, where we directed the residue search based on atomistic molecular simulations, by explicitly modelling both solvation and electrostatic effects underlying the thermodynamic stability of the ion-pairs. To this end, we optimised the Glu17-Lys72 ion-pair (F17E/F72K) by introducing random *in silico*-point mutations in the protein surroundings, and directed the residue search using a Metropolis Monte Carlo (MC) criterion towards an increased interaction between the ion-pair and the protein structure as an optimisation target. At the same time, we selected for structures with an increased overall protein stability (Fig. 3a, b, see Methods section).

In our directed design, we find that polar and charged residues that form hydrogen-bonded contacts with the buried ion-pair are

introduced in 2230 of 2838 ($\approx 80\%$) designed constructs probed within the first solvation sphere of the charged element (Fig. 3c). In addition to other charged residues, Gln/Asn, Tyr, and Ser/Thr constitute 20% of the sampled substitutions. These mutations insert amphiphilic elements around the charged centre (Fig. 3c) that could enhance the protein stability. In statistical analysis of ca. 180,000 ion-pairs from around 6000 resolved membrane protein structures (Fig. 3d, Supplementary Figs. 4 and 5), we find similar charge-stabilising motifs in natural proteins as predicted in the computational design. Interestingly, we find a 10–30% increased probability of observing charge-stabilising residues (Gln, Asn, Tyr, Ser, Thr) within the first solvation sphere of the ion-pair in the natural proteins (Fig. 3d, Supplementary Figs. 4 and 5), closely resembling the distribution predicted by our directed design (Fig. 3c). Moreover, tyrosine residues that have an overall slightly higher natural abundance, are somewhat over-represented within the membrane proteins dataset (Supplementary Fig. 5b), possibly as they comprise both bulky non-polar/aromatic properties that can be used for packing the protein core or form cation- π interactions with positively charged residues, but also a polar hydroxy headgroup that can form hydrogen

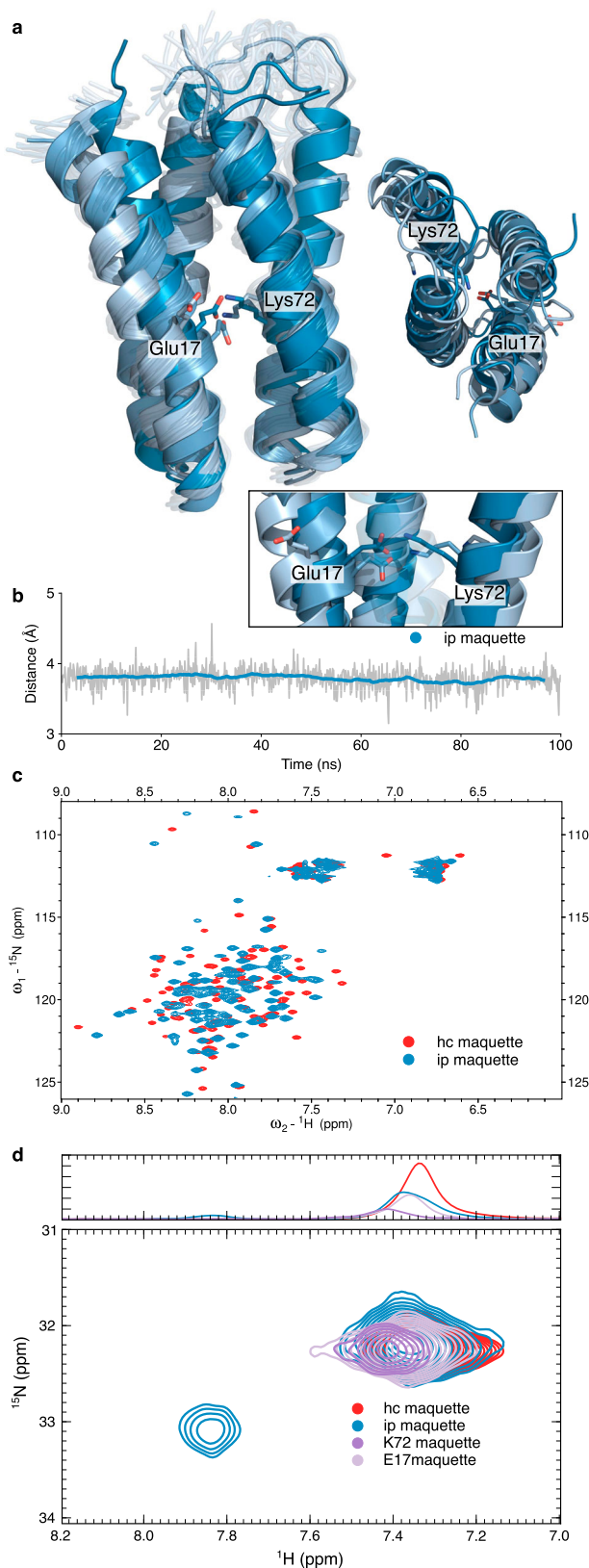


Fig. 2 Biophysical and structural characterisation of ion-pair in Maquette 2.

a NMR-structural model of *Maquette 2* with an ion-pair at positions 17 and 72 (F17E/F72K) modelled in both closed/open conformations (in dark/pale blue). **b** Dynamics of the Glu17 and Lys72 ion-pair modelled based on the NMR structure. The shaded line represents the actual data, the coloured line is a moving average over 50 frames. **c** Heteronuclear single quantum coherence (HSQC) NMR spectrum of *Maquette 2* with Glu17-Lys72 ion-pair (ip, F17E/F72K, in blue), and hydrophobic core (hc, in red). **d** NH_3 -selective HSQC experiment at 275 K, indicative of a protected sidechain with slow proton exchange. Glu17-Lys72 ion-pair (ip, F17E/F72K, in blue), hydrophobic core (hc, in red), and single residue substitutions (K72 *maquette* (E17F) in purple); E17 *maquette* (K72F), in pale purple.

residues¹⁸, are also clearly pronounced in the distribution (Fig. 3d, Supplementary Fig. 4). We note that the buried Trp68 might stabilise the Glu17-Lys72 ion-pair in the *Maquette 2* model (F17E/F72K, Supplementary Fig. 2c). The ion-pair distributions further suggest that glutamine residues seem to favour Arg-Asp ion-pairs, whilst Lys-Glu ion-pairs favour asparagine residues (Supplementary Fig. 5a), possibly due to steric constraints in natural proteins.

To understand the physical basis of the shielding motifs, we further studied the designed charged elements by electrostatic calculations. For the unshielded ion-pair, the Born desolvation penalty ($\Delta G = +12 \text{ kcal mol}^{-1}$) becomes partially compensated by the electrostatic interaction with the protein surroundings ($\Delta G = -6 \text{ kcal mol}^{-1}$), leading to an overall destabilising effect in the model calculations (Fig. 3e, see Supplementary Information). The partially uncompensated charges expose a non-uniform electric field, which could be responsible for the partial wetting of the protein core during the MD simulations and the overall reduced protein stability (Fig. 1). However, the shielding element provides a clear electrostatic compensation in the model systems that results in enhanced stability (Fig. 3e). The thermodynamic cost of embedding a charge-shielding Gln or Asn residue next to an ion-pair within the protein core is nearly isoenergetic in the model calculations (Supplementary Fig. 2e).

Inspired by these findings, potential charge-stabilising motifs were introduced in the ion-pair surroundings of the *Maquette 2* (Supplementary Fig. 2d). We find that introduction of Asn at position 69 (*Maquette 2/V69N*) leads to a small, but statistically significant enhancement of the protein stability in chemical unfolding experiments (Fig. 3f). Moreover, in NH_3 -selective HSQC measurements, a significant increase in the population of the closed ion-pair is observed, suggesting that the asparagine could stabilise the closed ion-pair conformation (Fig. 3f), as also qualitatively supported by MD simulations (Fig. 3g). The HSQC spectrum of *Maquette 2/V69N* indicates that the protein is well-folded, with a downfield shift of most peaks relative to the ion-pair model (Supplementary Fig. 2g). To further investigate the conformational flexibility of the *Maquette 2* models, we performed heteronuclear NOE (hetNOE), as well as longitudinal (T_1) and transverse (T_2) relaxation experiments¹⁹. The hetNOE and relaxation data of the hydrophobic core, ion-pair, and charge-stabilised ion-pair constructs are very similar (Supplementary Fig. 8), indicating that all models have similar dynamics on the ps-ns timescale, maintain a similar fold on the studied conditions, and are likely to have overall similar structures.

bonds. We note, however, that certain amino acids may have special preferred locations in the protein structure¹⁶. Moreover, charged elements are pronounced with an 10–40% increased probability¹⁷, whereas hydrophobic residues occur with a 10–60% reduced probability around the ion-pair. Interestingly, tryptophan residues that can form cation- π interactions with, e.g., Lys

Rational design of an ultra-stable buried charged network.

To test the physical (Fig. 3d, e) and evolutionary (Fig. 3c) insights, we created a charged network, with a double ion-pair surrounded by four charge-stabilising Gln-motifs embedded

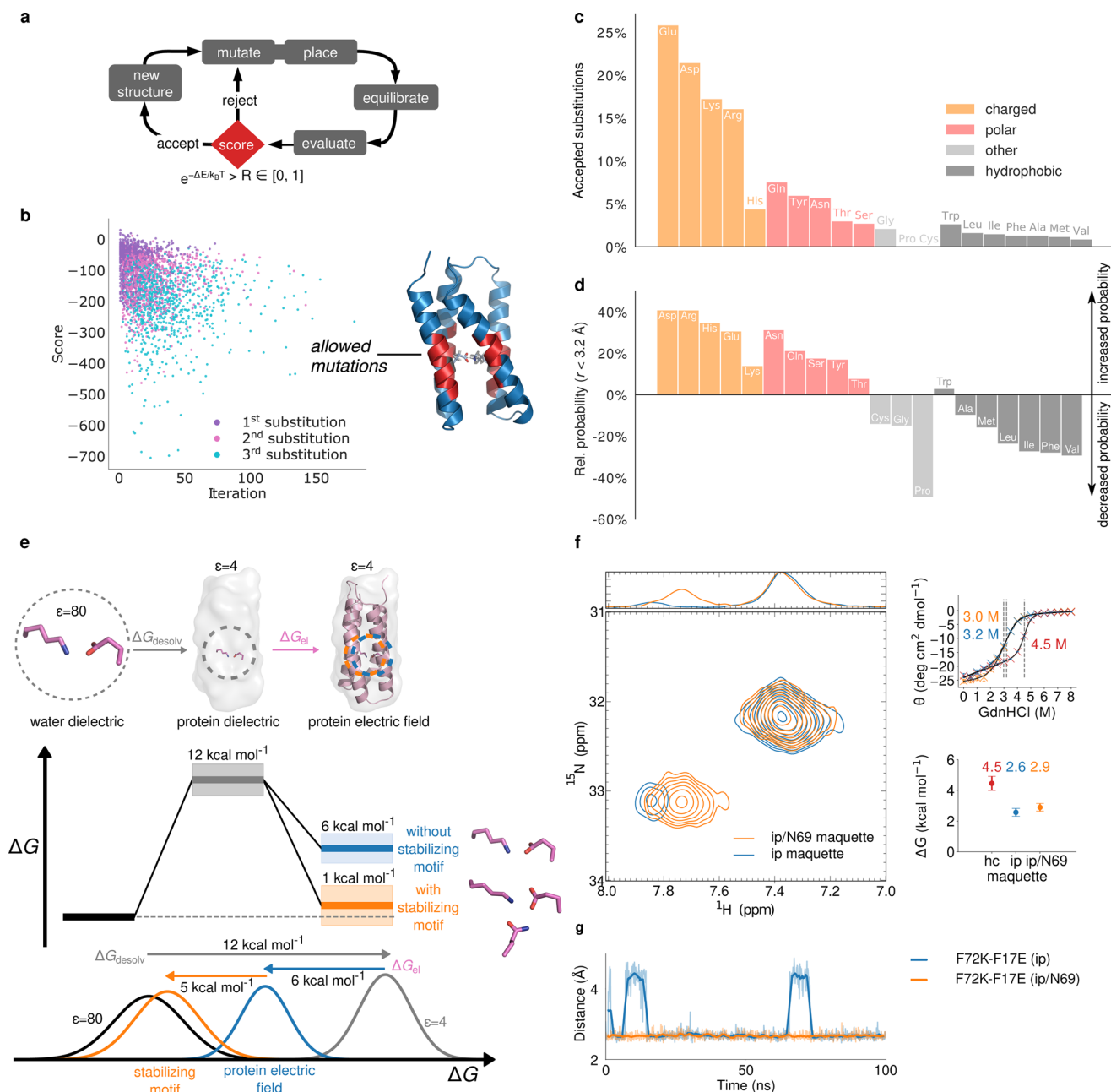


Fig. 3 Directed design, evolutionary conservation, and biophysical analysis of buried charged elements in proteins. **a** Directed computational design algorithm employed to optimise buried charge elements. **b** Sampled energy-score during directed design (see “Methods” section). Inset: *Maquette 2* model with area of introduced substitutions (in red). **c** Distribution of substituted amino acids in 1000 directed simulations. **d** Relative probability of amino acids surrounding ca. 182,000 natural ion-pairs from ~6000 structures. **e** Electrostatic calculations on the energetics of the ion-pair with and without charge-stabilising motif. **f** NH_3 -selective HISQC experiment at 275 K of *Maquette 2* with charge-shielded ion-pair. Right: chemical unfolding with GdnHCl monitored by CD spectroscopy of the *Maquette 2*-protein with stabilising shielding motif (top), and thermodynamic stability (ΔG) values calculated from the chemical unfolding experiments (bottom). Data points of the chemical unfolding profiles are presented as mean values of triplicates ($n = 3$ independent experiments) with \pm standard deviation shown as error bars. **g** MD simulations of *Maquette 2* models at 275 K. The figure shows distances (Glu/OE2-Lys/NZ) between the ion-pair without (in blue) and with charged-stabilising Asn69 (in orange). The shaded line represents the actual data, and the coloured line is a moving average over 50 frames.

within a ca. 70 Å-long α -helical *Maquette*-protein framework (*Maquette 3*, Fig. 4c, d, see Extended Methods section), originally a Zn^{2+} -site binding scaffold (PDB ID: 5VJT), that provides a larger non-polar interaction surface relative to the minimal 4 α -helical *Maquette* models 1 and 2. The designed construct is stable in MD simulations with intact ion-pairs and minimal water penetration (Fig. 4a, Supplementary Fig. 6), and the pronounced

minima at 208 and 222 nm in the CD spectrum support its α -helical secondary structure (Fig. 4c). Remarkably, the *Maquette 3* model is resistant towards unfolding at >100 °C and 6 M guanidine chloride concentrations, indicating a high overall protein stability, despite the large buried charge density. To validate the underlying molecular structure behind the high stability, we resolved the x-ray structure of the *Maquette 3* model at 1.85 Å

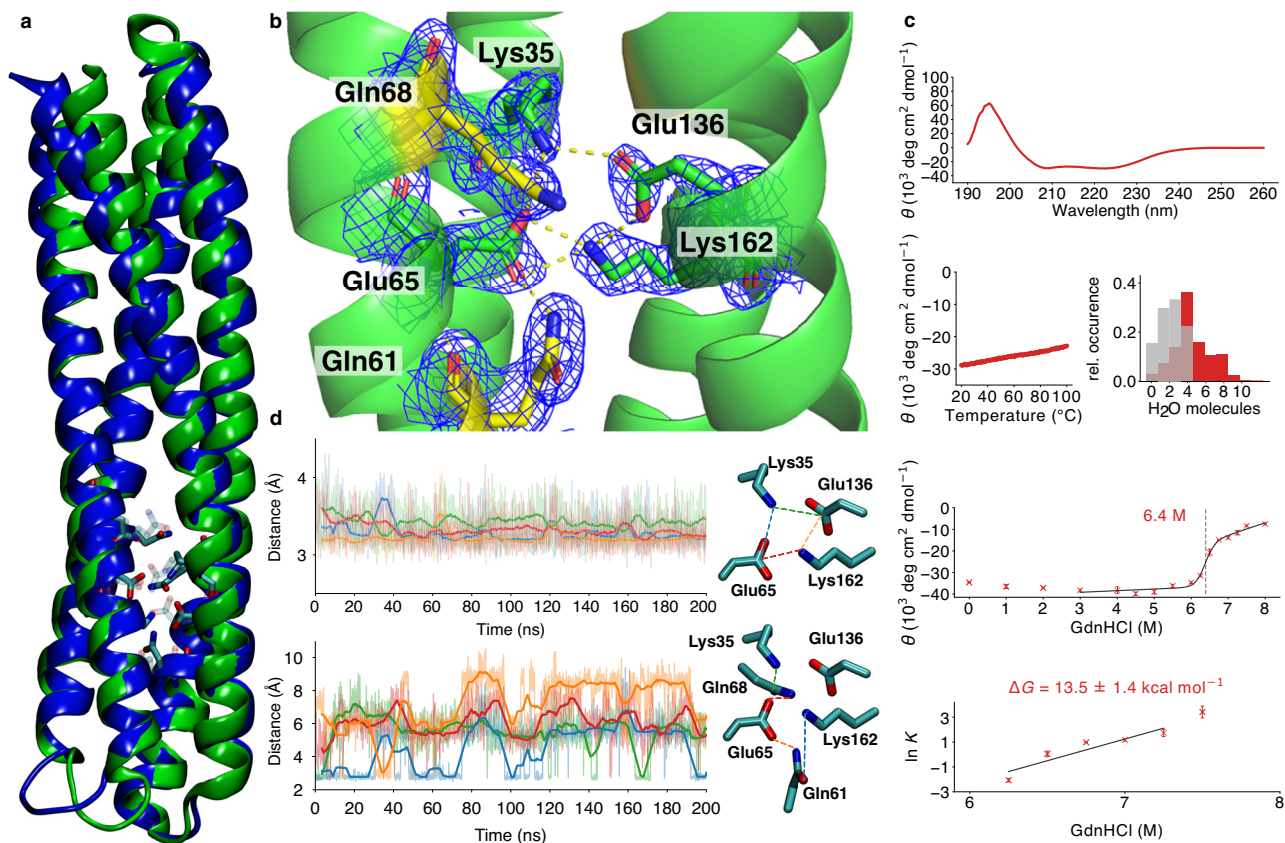


Fig. 4 Rational design of a buried ultra-stable charged motif. **a** Structural snapshots from the in silico model of the *Maquette 3* protein with a charge-shielded buried ion-pair network (in blue) and the resolved x-ray structure (in green). **b** Closeup of the designed charged network in the 1.85 Å resolved x-ray structure. **c** CD spectrum at 20 °C of *Maquette 3* (top); melting curves measured by following the temperature dependence of the 222-nm trough from CD spectroscopy (left); count of buried water molecules 5 Å around the ion-pair/Asn surroundings (grey) and water molecules within 5 Å of the complete protein (in red) (right); chemical unfolding with GdnHCl up to 8 M monitored by CD spectroscopy; thermodynamic stability inferred from the chemical unfolding experiments (bottom). Data points of the chemical unfolding profile are presented as mean values of triplicates with \pm standard deviation as error bars. **d** MD simulation of the ion-pair dynamics. Top: Distances between Lys35-Glu65 (blue), Lys35-Glu136 (green), Glu65-Lys162 (red), and Glu136-Lys162 (orange). Bottom: Distances between Lys35-Gln68 (green), Gln61-Glu65 (orange), Gln61-Lys162 (blue), and Glu65-Gln68 (red). The measured atomic distances are depicted to the right of the plots. See Supplementary Fig. 6 for further details. The shaded line represents the actual data, the coloured line is a moving average over 50 frames.

resolution (Fig. 4a, b, Supplementary Fig. 6, Supplementary Table 3). The $2F_o - F_c$ electron density map closely resembles the designed in silico model (backbone RMSD (in silico/x-ray) = 1.7 Å), with a charged core forming tight ion-pairs between Lys162/Glu65 and Lys35/Glu136 at distances of 2.7–2.9 Å. The charged residues are stabilised by hydrogen bonds to Gln61/68/169/132 (Fig. 4b, d, Supplementary Fig. 6c, e) that dynamically exchange in the MD simulations (Fig. 4d, Supplementary Fig. 7c). Although other charge-stabilising residues, such as tyrosine and histidine, are often present around natural ion-pairs (Supplementary Figs. 4 and 5), these residues were too bulky to be inserted around the charged network in the current minimal *Maquette 3* model. Intriguingly, the ion-pairs remain tightly sealed from water molecules in the crystal structure, consistent with the high stability indicated by our unfolding experiments and MD simulation (Fig. 4c, Supplementary Fig. 6b). Similar to the *Maquette 1* models, the HSQC spectrum of *Maquette 3* (Supplementary Fig. 6f) shows several overlapping and less sharp signals possibly due to the higher internal symmetry of the model. We could therefore not use NMR to study these models further. Nevertheless, to quantify the stabilising effect of the charge-shielding residues surrounding the ion-pairs, we further substituted two of

the glutamines with alanine residues (Q61A, Q68A). Our chemical unfolding measurements suggest that these substitutions indeed destabilise the protein by around 5 kcal mol⁻¹, supporting the proposed charge-shielding effect (Supplementary Fig. 6g). Taken together, the *Maquette 3* design supports the proposed electrostatic principles in which amphiphilic residues shield charged networks from the non-polar surroundings to gain thermodynamic stability.

In summary, we have characterised here the energetics of buried ion-pairs using artificial protein design in combination with computational, biophysical, and structural experiments. We identified that amphiphilic shielding motifs stabilise the buried charged networks by a fine-tuned balance between desolvation and electrostatic effects. The motifs could represent an evolutionary prerequisite for buried ion-pairs, as they are found in a majority of the analysed natural protein structures, and independently supported by our directed computational design. Our de novo proteins built based on these physical principles, resulted in ultra-stable charged networks that sustain harsh chemical and physical conditions. Our findings may provide a basis for understanding the structural stability of extended charged networks in molecular gates, switches, and allosteric

regulation sites in different protein machineries^{5,6,20,21}, and how disease-causing mutations may destabilise the protein structure and result in an altered conformational landscape^{22–24}.

Methods

In silico modelling of Maquette models. An atomic model of *Maquette 1* was built in silico with α -helical secondary structure restraints placed on the residues 1–27, 36–62, 71–97, and 106–130 based on PsiPred²⁵. The helices were packed to maximise the non-polar interaction within the core, in which Phe-15 and Phe-21 were restrained to form a stacking interaction with the nearest anti-symmetric helix, following computational modelling principles applied before²⁶. After the initial restrained relaxation, loops were built to connect the helices, followed by restrained minimisation and MD simulations using the CHARMM36 force field²⁷ at $T = 300$ K and using a 1 fs integration timestep. The *Maquette 1* models were built and relaxed using CHARMM²⁸, followed by MD simulations (see below). The *Maquette 2* models were constructed based on scaffold structures resolved in ref. ¹¹. The *Maquette 3* model was built in silico based on a helical coiled-coil template (PDB ID: 5VJT). First, all prosthetic groups were removed from the template, and glycine and alanine residues were replaced by valine and isoleucine residues, respectively, to stepwise maximise the non-polar packing, followed by a short (20 ns) MD simulation to test its stability. To further maximise the non-polar packing, hydrophobic residues at positions H10F, H111F, H125F, H139F, and the charged network was modelled manually by substitutions E162K, E35K, I61Q, H68Q, I132Q, Y169Q according to the charge-shielding principles described in Fig. 3e. The final sequence of the *Maquette 3* model is shown in Supplementary Table 1.

Molecular simulations. Classical MD simulations of the designed proteins were performed by solvating the protein in a TIP3P water box and 150 mM NaCl concentration. MD simulations were performed in duplicates for 200 ns (Supplementary Fig. 7) at $T = 310$ K using a 2 fs integration timestep and the CHARMM36 force field²⁷, and modelling the long-range electrostatics using the particle-mesh Ewald approach. Solvation free energies were estimated using Poisson-Boltzmann (PB) continuum electrostatic calculations with MC sampling. Protein residues were modelled as atomic point charges, and the protein surroundings were treated as polarisable medium with $\epsilon = 4$ that was embedded in a water dielectric of $\epsilon = 80$. The linearized-PB equations were solved numerically using the adaptive PB-solver (APBS) and Karlsberg^{29,30}. Visual Molecular Dynamics (VMD)³¹ and PyMOL³² were used for visualisation and analysis, MD simulations were performed using NAMD³³. See Supplementary Information for summary of the performed simulations (Supplementary Fig. 6, Supplementary Table 4).

Directed computational design. Charged elements in the artificial protein models were optimised by introducing random mutations into the *Maquette* structures, followed by 10 ps classical MD simulations in explicit solvent. The protein models were scored by calculating the self-interaction of the ion-pair (weight = 0.48) and its interaction with the remaining protein surroundings (weight = 0.48) and all nonbonded interactions within the protein (weight = 0.02). The sum of the weights were normalized to 1.0. The substitutions were accepted or rejected based on weighted interaction scores using a Metropolis Monte Carlo criterion. All substitutions were subjected to 5000 steepest-gradient minimisation steps, followed by 5 ps restrained relaxation with harmonic protein backbone restraints with force constants of 1 kcal mol⁻¹ Å⁻². The protein interactions were modelled using the CHARMM36 force field²⁷, and the MD relaxation was performed using a 1 fs integration timestep at $T = 310$ K using NAMD³³. The directed computational design approach was implemented in Python in our in-house code PACMAN (Packing with MC-assisted networks).

Statistical analysis of ion-pairs. Statistical analysis of naturally occurring ion-pairs was carried out based on structures available in the orientations of proteins in membranes (OMP) database³⁴. From the 7230 available structures, 6045 with a resolution <3.5 Å were included. The protein surface was calculated using the MSMS program³⁵, and ion-pairs were identified using the Salt Bridges plugin of VMD³¹. All residues within a 15 Å sphere of the identified ion-pairs were counted and probabilities for their occurrence within a certain distance to the centre-of-mass of the ion-pair was calculated as,

$$p^{aa}(r_{\text{thr}}) = \frac{N_{\text{thr}}^{aa}}{\sum_i^{AA} N_{\text{thr}}^i} \quad (1)$$

where aa is the identified amino acid, r_{thr} is the threshold radius, AA the full set of amino acids, and N_{thr}^{aa} is the number of amino acids within a given threshold radius.

Protein expression and purification. Genes for the computationally designed proteins were synthesized and cloned into pET21a(+) vector (Genscript). F17E, F72K, 1ip/I43Y, 1ip/I39Q, 1ip/R76S, and 1ip/L91H of *Maquette 2* were produced using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's manual and the primers listed in Supplementary Table 5.

The C-terminal His₆-tagged proteins were overexpressed in *E. coli* BL21 (DE3) for 4 h at 37 °C after induction with 0.5 mM IPTG at OD₆₀₀ 0.7. After sonication, the lysate was applied on a Ni-NTA affinity column. The His-tag was cleaved by a Tobacco Etch Virus protease, and the proteins were purified by size exclusion chromatography using a preparative Superdex 75 pg HiLoad 16/600 column (GE Healthcare) with a flow rate of 1.0 mL min⁻¹. See Supplementary Table 1 for amino acid sequences of all artificial proteins and Supplementary Table 6 for DNA sequences. The buffers used during protein purification are summarised in Supplementary Table 7.

Circular dichroism spectroscopy. CD spectra and melting curves were measured on a Jasco J-715 system combined with a cooling-heating unit using a 1-mm-path quartz cuvette. Protein concentrations of around 5–10 μ M were used for measurements at 20 °C. For the determination of melting curve, samples were heated up from 20 to 100 °C, while the alpha-helical CD-signal at 222 nm was monitored. The protein stability was also studied by guanidine-hydrochloride (GdnHCl) or urea induced unfolding with GdnHCl/urea concentrations up to 8 M. Folding free energies were obtained from linear least-square fits of $-RT \ln K$ that were extrapolated to 0 M GdnHCl. The apparent equilibrium constant K was determined from the fraction of the denatured protein, $f_D = (y - y_N) / (y_D - y_N)$, where y is the CD raw signal, and y_N and y_D are the signals for the native and denatured states, respectively. Error bars were computed from the standard error of the least-square fits from the average unfolding profiles.

Nucleic magnetic resonance spectroscopy. Heteronuclear NMR spectra were recorded using ¹⁵N- and ¹⁵N,¹³C-labelled protein at 20 °C in a buffer with 50 mM sodium phosphate, 100 mM NaCl, and 1 mM EDTA at pH 7.5. NMR experiments were performed on Bruker Avance III spectrometers operating at a ¹H Larmor frequency of 600 MHz (14.09 T) and 950 MHz (22.31 T) using a CPTCI triple-resonance cryoprobe and the Bruker TopSpin v. 3.5 software. Sequence-specific backbone and side-chain assignment were obtained using 3D HNC(A), HN(CO)CA, HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH, HBHA(CO)NH, H(CCO)NH, CC(CO)NH, HC(C)H/(H)CCH TOCSY, ¹⁵N/¹³Cali/¹³Caro.-edited ¹H,¹H NOESY (120 ms mixing time) as well as 2D (HB)CB(CGCD)HD. 2D ¹H,¹⁵N HSQC spectra were recorded for the detection of Lys ¹⁵NH₃ groups³⁶. ¹H, ¹³C, ¹⁵N chemical shifts were referenced via DSS and spectra were assigned using Sparky³⁷. ¹⁵N T_1 , $T_{1\rho}$, and steady-state hetNOE experiments were collected at an external magnetic field of 11.74 T (¹H Larmor frequency of 500 MHz) and 293 K. Temperature compensation was employed, according to Lakomek et al.³⁸ The ¹⁵N rf amplitude for the $T_{1\rho}$ spin-lock was set to 1.5 kHz. For the steady-state hetNOE experiment a saturation time of 4 s was employed. T_2 was calculated according to

$$\frac{1}{T_2} = \frac{R_{1\rho}}{\sin^2\theta} - \frac{R_1}{\tan^2\theta} \quad (2)$$

with

$$R_{1\rho} = 1/T_{1\rho}, \quad (3)$$

$$R_1 = 1/T_1, \quad (4)$$

$$\tan\theta = \frac{\omega_1}{\Omega}. \quad (5)$$

ω_1 is the amplitude of the ¹⁵N spin-lock field and Ω is the ¹⁵N resonance offset from the spin-lock carrier frequency. The experimental error was set to two times the standard deviation of the spectral noise. Uncertainties were estimated by 1000 MC runs.

NMR structure calculation. Automated NOE assignment was conducted in CYANA^{39,40} to generate restraints and calculate de novo structures of the *Maquette 2* construct with ion-pair at Glu17/Lys72. The closed ion-pair conformation was modelled by restraining lower and upper distance limits between Glu17-OE1 and Lys72-NZ of 2.8 and 3.0 Å, respectively, as guided by the MD simulations. The calculations with torsion angle dynamics were started from 200 random structures, employing 20,000 torsion angle dynamics steps. In addition to distance restraints, dihedral angle restraints derived from chemical shifts were obtained using TALOS-N⁴¹. The restrained energy minimisation of the consensus bundle of 20 conformers in explicit solvent against the AMBER force field was carried out using OPALp^{42,43}. The statistics of the structure calculations is given in the Supplementary Table 2.

X-ray crystallography. Crystals of the *Maquette 3* construct were grown at 20 °C using the sitting drop vapour diffusion method. Drops contained a 1:2 mixture of protein solution (40 mg ml⁻¹ protein) and reservoir solution (100 mM HEPES buffer pH 7.5, 1.5 M potassium phosphate). The crystals were cryo-protected with 25% glycerine. Diffraction data were collected at the beamline X06SA at the Paul Scherrer Institute, SLS, Villigen, Switzerland ($\lambda = 1.0$ Å). Evaluation of reflection intensities and data reduction were performed with the program package XDS (v. January 31, 2020)⁴⁴. Refinement of the initial model was carried out with the coordinates of the in silico model of *Maquette 3* using REFMAC⁴⁵, followed by

model building using COOT⁴⁶ and CCP4 v. 7.0⁴⁷. See Supplementary Table 3 for further details.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The structures have been deposited to the Protein Data Bank, PDB IDs: 6ZOC [rcsb.org/structure/6zoc], 6Z35 [rcsb.org/structure/6z35]. The NMR assignments are deposited in the BMRB under the entry number 34518 [bmrb.io/data_library/summary/?bmrblid=534518]. Other data are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

The pacman code is available under MIT license at <https://github.com/KailaLab/pacman-protein-design>.

Received: 30 May 2020; Accepted: 19 February 2021;

Published online: 25 March 2021

References

- Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
- Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494**, 443–448 (2013).
- Ali, M. M. U. et al. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 (2006).
- Robinson, A. C., Castaneda, C. A., Schlessman, J. L. & Garcia-Moreno, E. B. Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. *Proc. Natl Acad. Sci. USA* **111**, 11685–11690 (2014).
- Kaila, V. R. I. Long-range proton-coupled electron transfer in biological energy conversion: Towards mechanistic understanding of respiratory complex I. *J. R. Soc. Interface* **15**, 20170916 (2018).
- Mader, S. L. et al. Conformational dynamics modulate the catalytic of the molecular chaperone Hsp90. *Nat. Commun.* **11**, 1410 (2020).
- Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D. & García-Moreno, E. B. Charges in the hydrophobic interior of proteins. *Proc. Natl Acad. Sci. USA* **107**, 16096–16100 (2010).
- Koder, R. L. et al. Design and engineering of an O₂ transport protein. *Nature* **458**, 305–309 (2009).
- Robertson, D. E. et al. Design and synthesis of multi-haem proteins. *Nature* **368**, 425–432 (1994).
- Farid, T. A. et al. Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat. Chem. Biol.* **9**, 826–833 (2013).
- Polizzi, N. F. et al. De novo design of a hyperstable non-natural protein-ligand complex with sub-Å accuracy. *Nat. Chem.* **9**, 1157–1164 (2017).
- Joh, N. H. et al. De novo design of a transmembrane Zn²⁺-transporting four-helix bundle. *Science* **346**, 1520–1524 (2014).
- Hwang, J.-L. & Warshel, A. Why ion pair reversal by protein engineering is unlikely to succeed. *Nature* **334**, 270–272 (1988).
- Robinson, A. C., Schlessman, J. L. & García-Moreno, E. B. Dielectric properties of a protein probed by reversal of a buried ion pair. *J. Phys. Chem. B* **122**, 2516–2524 (2018).
- Sali, D., Bycroft, M. & Fersht, A. R. Stabilization of protein structure by interaction of α -helix dipole with a charged side chain. *Nature* **335**, 740–743 (1988).
- Hessa, T. et al. Molecular code for transmembrane-helix recognition by the SecE1 translocon. *Nature* **450**, 1026–1030 (2007).
- Bush, J. & Makhatazde, G. I. Statistical analysis of protein structures suggests that buried ionizable residues in proteins are hydrogen bonded or form salt bridges. *Proteins* **79**, 2027–2032 (2011).
- Dougherty, D. A. Cation- π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **271**, 163–168 (1996).
- Wagner, G. The importance of being floppy. *Nat. Struct. Biol.* **2**, 255–257 (1995).
- Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331–339 (2014).
- Di Luca, A., Gamiz-Hernandez, A. P. & Kaila, V. R. I. Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl Acad. Sci. USA* **114**, E6314–E6321 (2017).
- Fiedorczuk, K. & Sazanov, L. A. Mammalian mitochondrial complex I structure and disease-causing mutations. *Trends Cell Biol.* **28**, 835–867 (2018).
- Boczek, E. E. et al. Conformational processing of oncogenic v-Src kinase by the molecular chaperone Hsp90. *Proc. Natl Acad. Sci. USA* **112**, E3189–E3198 (2015).
- Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
- Buchan, D. W. A. & Jones, D. T. The PSIREN protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
- Gámiz-Hernández, A. P., Kieseritzky, G., Galstyan, A. S., Demir-Kavuk, O. & Knapp, E. W. Understanding properties of cofactors in proteins: redox potentials of synthetic cytochromes b. *Chem. Phys. Chem.* **11**, 1196–1206 (2010).
- Huang, J. & Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
- Brooks, B. R. et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1174–1178 (2009).
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
- Kieseritzky, G. & Knapp, E. W. Optimizing pKa computation in proteins with pH adapted conformations. *Proteins* **71**, 1335–1348 (2008).
- Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
- The PyMol Molecular Graphics System. Version 1.8 Schrödinger LLC (2015).
- Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
- Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
- Sanner, M. F., Olson, A. J. & Spehner, J.-C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996).
- Iwahara, J., Jung, Y. S. & Clore, G. M. Heteronuclear NMR spectroscopy for lysine NH₃ groups in proteins: Unique effect of water exchange on 15N transverse relaxation. *J. Am. Chem. Soc.* **129**, 2971–2980 (2007).
- Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
- Lakomek, N. A., Ying, J. & Bax, A. Measurement of 15N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR* **53**, 209–221 (2012).
- Güntert, P. & Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62**, 453–471 (2015).
- Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).
- Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).
- Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **66**, 27–85 (2003).
- Koradi, R., Billeter, M. & Güntert, P. Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun.* **124**, 139–147 (2000).
- Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
- Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **60**, 2184–2195 (2004).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
- Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Cryst.* **D67**, 235–242 (2011).

Acknowledgements

We acknowledge Prof. Michael Sattler and Prof. Franz Hagn for providing us access to the Bavarian NMR Facility, Astrid König for technical assistance in crystallisation trials, and Dr. Sina Kazemi for helpful discussions regarding NMR structure calculations. This work was supported by the collaborative research centre, SFB1035 (B12) and the Centre for Integrated Protein Science Munich (CIPSM). This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program/grant agreement no. 715311. VRIK is supported by the Knut and Alice Wallenberg Foundation. Computational resources were provided by the Leibniz-Rechenzentrum (LRZ), SuperMuc (project: pr27xu), and the Swedish National Infrastructure for Computing (SNIC) at PDC partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

Author contributions

M.B. expressed, purified, and characterised the designed proteins, and performed computational and biophysical characterisation. M.B., M.R., M.E.M., S.L.M., A.P.G.H., and V.R.I.K. performed molecular simulations and design. M.R., M.E.M., S.L.M., and V.R.I.K. developed analytic computational tools. M.B. and S.A. performed NMR experiments and calculated NMR structure ensembles. M.B., M.G., and V.R.I.K. resolved the x-ray structures. A.P.G.H. performed initial computational model building. M.B., M.R., M.E.M., S.A., S.L.M., K.F., M.G., A.P.G.H., and V.R.I.K.

analysed results. V.R.I.K. designed and directed the project, performed initial computational design and experimental characterisation, and wrote the manuscript with input from all authors.

Funding

Open access funding provided by Stockholm University.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21909-7>.

Correspondence and requests for materials should be addressed to V.R.I.K.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Article V

Functional Dynamics of an Ancient Membrane-Bound Hydrogenase

Max E. Mühlbauer, Ana P. Gamiz-Hernandez, and Ville R. I. Kaila*

Cite This: *J. Am. Chem. Soc.* 2021, 143, 20873–20883

Read Online

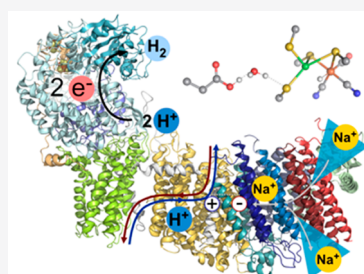
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The membrane-bound hydrogenase (Mbh) is a redox-driven Na^+/H^+ transporter that employs the energy from hydrogen gas (H_2) production to catalyze proton pumping and Na^+/H^+ exchange across cytoplasmic membranes of archaea. Despite a recently resolved structure of this ancient energy-transducing enzyme [Yu et al. *Cell* 2018, 173, 1636–1649], the molecular principles of its redox-driven ion-transport mechanism remain puzzling and of major interest for understanding bioenergetic principles of early cells. Here we use atomistic molecular dynamics (MD) simulations in combination with data clustering methods and quantum chemical calculations to probe principles underlying proton reduction as well as proton and sodium transport in Mbh from the hyperthermophilic archaeon *Pyrococcus furiosus*. We identify putative Na^+ binding sites and proton pathways leading across the membrane and to the NiFe-active center as well as conformational changes that regulate ion uptake. We suggest that Na^+ binding and protonation changes at a putative ion-binding site couple to proton transfer across the antiporter-like MbhH subunit by modulating the conformational state of a conserved ion pair at the subunit interface. Our findings illustrate conserved coupling principles within the complex I superfamily and provide functional insight into archaeal energy transduction mechanisms.



INTRODUCTION

The membrane-bound hydrogenase (Mbh) is a primordial enzyme that powers energy transduction in the thermophilic archaeon *Pyrococcus furiosus*.^{1,2} Mbh catalyzes ferredoxin (Fd)-driven ($E_{m,7} = -450$ mV, redox midpoint potential at pH = 7) hydrogen gas production ($E_{m,7} = -420$ mV)^{1–4} and employs the small thermodynamic driving force ($\Delta G \sim -60$ mV) for proton pumping and Na^+/H^+ exchange.^{3,4} The sodium motive force (smf) generated across the archaeal membrane powers subsequent Na^+ -driven ATP synthesis.⁵

Mbh is a 14-subunit enzyme complex^{1,2,6} comprising a hydrophilic domain, responsible for electron transfer and H_2 production, and a membrane domain driving Na^+/H^+ transport, with the recently resolved cryoEM structure revealing key features of its molecular architecture¹ (cf. also ref 7). Mbh is a predecessor of the modern complex I superfamily,^{1,7,11} which catalyzes NADH- or Fd-driven quinone (Q) reduction and couples the much larger ($\Delta G \sim -800$ to -1200 mV) driving force to proton pumping across the membrane domain. Despite intensive work in recent years,^{8–12} the molecular principles of the fascinating energy transduction mechanism employed by this superfamily remain unsolved.

The catalytic cycle of Mbh is initiated by association of the reduced Fd to the positively charged MbhN at the top of the hydrophilic domain, comprising subunits MbhI–MbhL and MbhN (Figure 1). The electrons are transferred from Fd via three tetranuclear iron–sulfur (FeS) centers to the NiFe core of MbhL, responsible for the proton reduction.¹ The binuclear NiFe catalytic center is coordinated by four cysteine residues (Cys68_L, Cys71_L, Cys374_L, Cys377_L), in addition to two CN^-

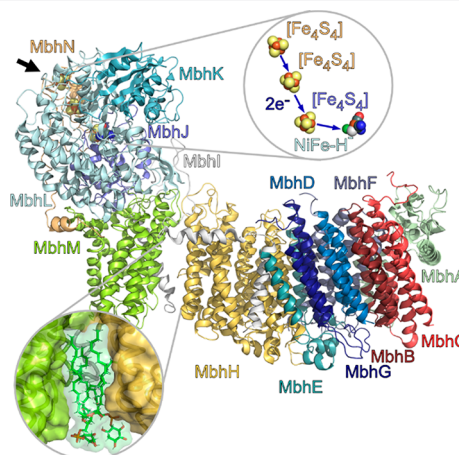


Figure 1. Structure and function of Mbh. Ferredoxin docks to MbhN (position indicated by black arrow) and transfers electrons via the FeS centers to the NiFe active site (inset), where protons are reduced to hydrogen gas ($2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2$). Proton transport and Na^+/H^+ exchange could take place in MbhM/MbhH and MbhA–G, respectively (see the main text). A lipid gap with modeled phosphatidylinositol is shown in the bottom inset.

Received: September 3, 2021

Published: November 30, 2021



and a CO ligand, and closely related to soluble NiFe-hydrogenases.^{1,13} The membrane domain of Mbh comprises eight subunits responsible for ion transport (Figure 1). Based on mutagenesis and structural studies of the related Mrp (multiple-resistance and pH adaption) transporters,^{1,14,15} the Na⁺/H⁺ exchange was suggested to take place in MbhB/C/D/G (MrpF/G/A_{nt}/C in Mrps).^{14,15} Together with MbhA/F/G (Figure 1), these subunits located on the terminal edge of Mbh establish a bundle of 3 × 4 transmembrane (TM) helices—a symmetric motif common in many antiporters.^{8,9,11,14,15}

MbhH is related to the proton pumping NuoN module of complex I (*E. coli* nomenclature; Nqo14/ND2 in other species), but it is rotated by 180° in the membrane relative to the former (Figure 1). This antiporter-like subunit comprises two symmetry-related TM-helix bundles with one broken helix each, buried charged residues, and a conserved ion pair. These features form functional elements for water-mediated proton pumping in complex I^{16–24} and could have a similar function in Mbh. MbhH is further clamped by an amphipathic transverse helix of MbhI that could secure tight electrostatic interaction between the subunits.⁸

MbhG and MbhD+E are related to NuoK/J (Nqo11/10, ND4L/ND6)—subunits that are also most likely involved in proton transport in complex I.^{8,9,16,18,20–24} However, in contrast to the latter where they reside close to the interface between the hydrophilic and membrane domains, these subunits are located in the middle of the membrane domain in Mbh (Figure 1). MbhG/D/E could catalyze Na⁺/H⁺ transport, although the molecular principles remain debated. Yu et al.¹ suggested that MbhC is responsible for Na⁺ transport, while proton transfer was suggested to occur in MbhH and MbhD/G. In contrast, Steiner and Sazanov¹⁵ resolved sodium ions in Mrp subunits homologous to MbhD (N-terminal part of MrpA) and MbhG (MrpC). A sodium pathway was also suggested to reside in MbhA (MrpE), with a putative input channel from the negatively charged side (cytoplasmic side) in MbhA and an output site to the positively charged side (periplasmic side) of the membrane at the MbhG/F/H interface (Figure 1). Sodium ions have also been suggested to enter MbhH (MrpD) and exit to the periplasmic side at MbhF/G/H.¹⁴

The membrane-bound MbhM (related to NuoH/Nqo8/ND1 of complex I) forms an “ankle” region at the interface of the hydrophilic and membrane domain. This region has a special functional role in complex I,^{18,23–25} whereas a large, ca. 10 Å wide cleft is observed in the cryoEM structure at the MbhM/MbhH interface (Figure 1, bottom inset).¹ The structure also revealed conserved loop regions within MbhM (NuoH/Nqo8), MbhI (N-terminal part of NuoA/Nqo7), and MbhL (NuoD/Nqo4), which undergo conformational changes in complex I.^{20,23,26,27}

To shed light on the elusive ion transport mechanism of Mbh, we probe here the coupling and dynamics between the redox and proton pumping domains and the sodium/proton exchange domain in the membrane-bound hydrogenase from *P. furiosus*¹ by combining classical molecular dynamics simulations and data clustering methods, with quantum chemical calculations. Our findings illustrate detailed hydration dynamics, putative sodium and proton binding sites, and possible locations of the proton channels. We also present a mechanistic model that could explain how the coupling between redox-driven proton reduction, proton pumping, and ion transport is achieved.

RESULTS

Global Dynamics and Proton Wires Leading to the Active Site. To gain insight into the functional dynamics of Mbh, we performed 8 μs of MD simulations of the solvated 14-subunit enzyme modeled in a phosphatidylinositol (PI) membrane, which is abundant in *P. furiosus*,²⁸ and explored the effects of modeling key residues in different protonation states (Tables S1 and S2). Mbh remains structurally stable during the MD simulations, with an overall root-mean-square deviation (RMSD) of <4 Å relative to the refined cryoEM structure (Figure S1A). The protein shows a high flexibility toward the cytoplasmic side surface of MbhA/B/C and the upper part of the hydrophilic domain (Figure S1B).

The conserved loops connecting the active site and a charged funnel in MbhM are also highly flexible, particularly the long MbhI loop, which we modeled *in silico* based on the experimentally resolved backbone coordinates. We note that the dynamics projected from the MD simulations are in good overall agreement with B-factors extracted from the cryoEM maps (Figure S1B), supporting that the simulations capture the global dynamics of the protein.

To probe possible H⁺ and Na⁺ transfer pathways, we next analyzed buried water networks formed during the MD simulations. The dry cryoEM structure reaches a highly hydrated state within 100 ns of the MD simulations, where a few hundred water molecules form transient interactions within the hydrophilic domain as well as buried parts of the membrane domain (Figure 2A,B and Figure S2). The water

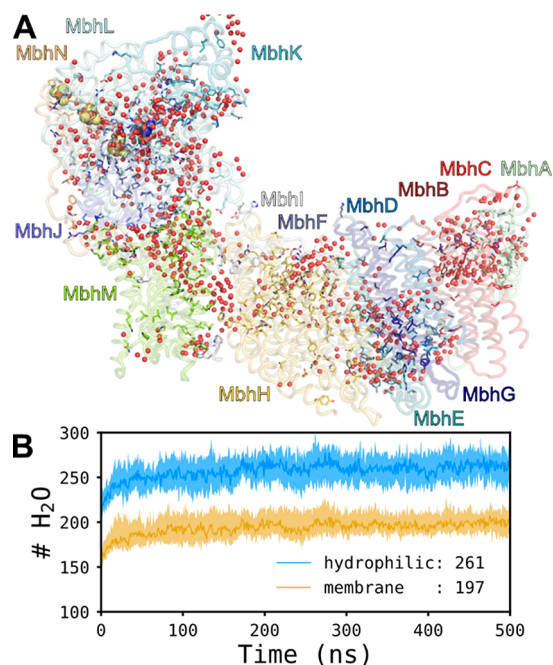


Figure 2. Global hydration dynamics of Mbh. (A) Conserved residues (see the Methods section) within the complex I superfamily are shown in stick representation and buried water molecules from MD simulations within 6 Å of these residues as red spheres. (B) Around 200–250 water molecules associate with the hydrophilic (MbhI–N, blue) and the membrane domains (MbhA–H, yellow) of Mbh during MD simulations. The plot shows the median over 16 independent simulations (8 μs in total) for each time point, with 25 and 75 percentiles, represented by the shaded areas. Hydration dynamics for individual subunits are reported in Figure S2.

wires connecting the cytoplasmic side with the NiFe site could conduct protons for H₂ production. On the basis of clustering analysis (see the [Methods](#) section), we observe four possible proton channels (see definition in the [Supporting Information](#)), comprising several charged residues and water molecules, leading to Glu21_L (Figure 3A,B, Figure 4, and

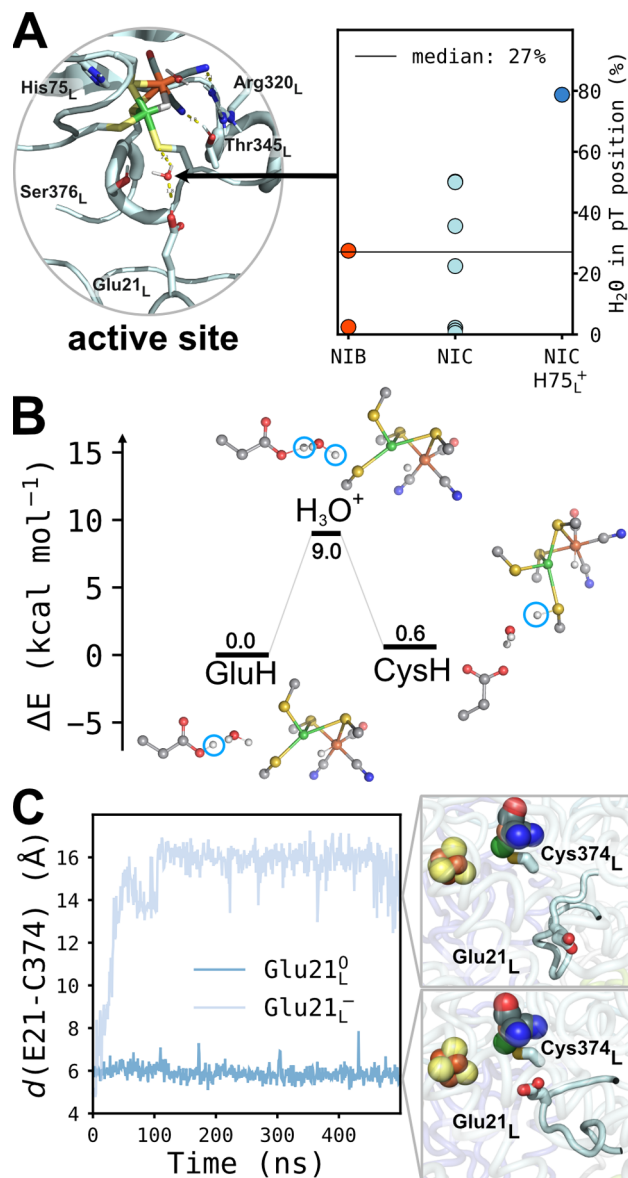


Figure 3. Hydration and protonation dynamics of the Mbh active site. (A) The NiFe cluster is stabilized by conserved contacts to surrounding residues. Glu21_L on the β 1– β 2 loop is bridged to Cys374_L by a water molecule in up to 85% of the MD simulation time depending on the state (NIB: OH[−] ligand; NIC: H[−] ligand; NIC H75⁺: NIC with cationic His75_L). (B) DFT calculations suggest that water-mediated proton transfer from Glu21_L to Cys374_L is energetically and kinetically feasible. Only the chemically active atoms are shown. See [Figure S4](#), [Table S5](#), and [Movie S1](#) for further details. (C) Glu21_L flips away from the active site upon deprotonation.

[Figure S3A](#)), a highly conserved residue within hydrogenases¹³ located on the β 1– β 2 loop next to the NiFe site. The active site Cys374_L is further bridged by a water molecule to Glu21_L, particularly when His75_L is modeled in a protonated state ([Figure 3A](#)). Interestingly, the neighboring Glu20_L corre-

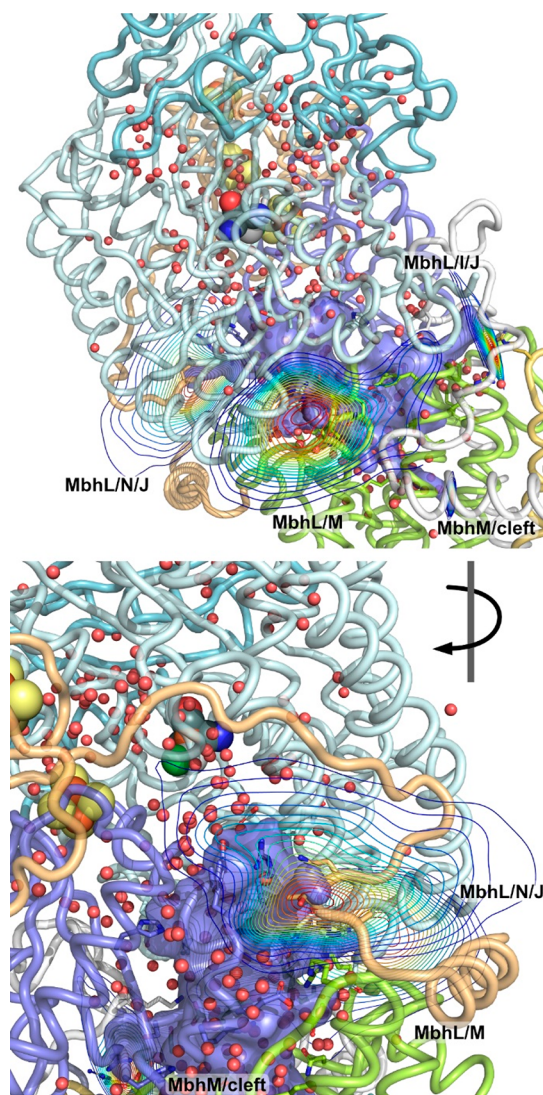


Figure 4. Water clusters leading to the cytoplasmic bulk and the cleft between MbhH and MbhM could allow for reprotonation of Glu21_L from the bulk. The clusters are present in all simulations, but the number of water molecules varies depending on the protonation states. The figure shows centroid pathways based on cluster analysis (see the [Methods](#) section).

sponds to the quinone-coordinating active site histidine (His38_{Nq04} in *T. thermophilus*) in complex I and is conformationally flexible.^{1,7–12,18} Water influx occurs at the interface between MbhL and MbhM formed by the lower edge of the β -sheet in MbhL (around 25%; median over all simulations, [Figure 4](#) and [Figure S3B](#)). Channels also form along the interface of the β 1– β 2 and MbhI loops (25%, [Figure 4](#) and [Figure S3B](#)) and at the MbhL/MbhJ interface (34%, [Figure 4](#) and [Figure S3B](#)). A cleft between MbhM and MbhH (12%) also weakly contributes to the overall water influx into the NiFe-site ([Figure 4](#) and [Figure S3B](#)), similar to the E-channel in complex I.^{8,18,20–25} These channels, particularly around the β 1– β 2 loop, comprise functionally important motifs in the canonical complex I,^{8,11} thus also supporting their relevance in Mbh.

Interestingly, when Glu21_L is modeled in its protonated state, as predicted by our electrostatic calculations (see the [Methods](#) section), the residue forms a water-mediated contact with Cys374_L of the NiFe site ([Figure 3A,C](#)). In contrast, upon

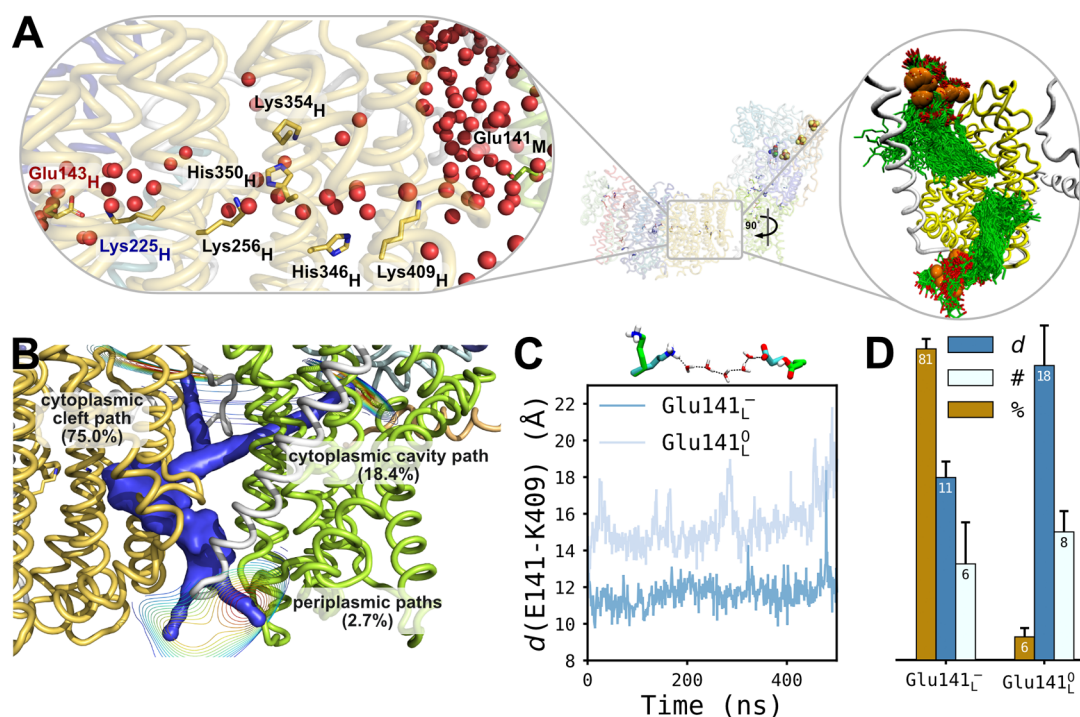


Figure 5. Structure and dynamics of MbhH. (A) Left inset: water molecules establish a hydrogen-bonded proton array between the cytoplasmic bulk and the hydrophilic axis, comprising histidine and lysine residues and leading to Lys256_H. The figure shows the last snapshot (500 ns) from simulation S3 (Table S1). Right inset: lipid binding sites on the cytoplasmic and periplasmic sides of the membrane showing the 20 most probable binding poses from eight independent simulations, as determined by pyLipid (see Supporting Information, Methods). (B) Water pathways based on clustering analysis of 8 μ s MD data (see the Methods section). Three water inlet clusters serve as water conduits to the hydrophilic axis in MbhH: (I) from the cytoplasm via the lipid cleft (contribution 75%), (II) from the cytoplasmic side via the E-channel homologue in MbhM (18.4%), and (III) from the periplasmic side by a water wire between TM1/2 of MbhI and TM2/3 of MbhM via the broken helix TM12 of MbhH (2.7%). (C) Distances between Glu141_M and Lys409_H across the lipid filled cleft during MD simulations. Protonation of Glu141_M increases the distances to Lys409_H. See Figure S9 for distance distribution in all simulations. The inset shows the shortest water-mediated pathway between the two residues in the deprotonated state (cyan sticks) and the lack of connectivity in the protonated state (green sticks). (D) Protonation of Glu141_M increases the distance to Lys409_H (dark blue, in Å) and disrupts the hydrogen-bonding (HB) connectivity, as shown by the probability of the stable HB wire between the two residues (in tan) and the number of HBs along the shortest path (in light blue).

deprotonation, Glu21_L flips away from the NiFe site toward the hydrated channels leading to the cytoplasmic bulk phase at the MbhM/MbhL interface (Figure 3C and Figure S3).

To probe whether Glu21_L could act as a proton donor for Cys374_L, we performed quantum chemical density functional theory (DFT) calculations (Figure 3B), which allowed us to address the reaction energetics along key steps of the NiFe catalytic cycle (see the Methods section, Figure S4). The water-mediated proton transfer from Glu21_L to Cys374_L has a reaction barrier of around 9 kcal mol⁻¹, and the two states are nearly isoenergetic by $\Delta E = +0.6$ kcal mol⁻¹ in the Ni^{II}/Fe^{II} state (Figure 3B, Table S5, and Movie S1).

These findings suggest that the proton uptake is kinetically accessible on physiologically relevant time scales along the conserved Glu21_L/Cys374_L pathway (Figure 3B), residues that may also be functionally important in canonical hydrogenases.^{29,30,35} Formation of the protonated Cys374_L leads to an elongation of the Ni–Cys374_L bond (Figure 3B and Figure S3), which in turn opens up the contact for the proton to the H⁻ bond between the Ni and the Fe. This could lead to the formation of the H₂ species in a process, enabled by a subtle (130° rotation of the Cys–H bond (Figure 3B, Figure S4, Table S5, and Movie S1; cf. also refs 29 and 30). We note that the conformation of Arg320_L, which forms contacts with both the NiFe center and Asp372_L, strongly influences the energetics of these reaction steps, and samples both inward

and outward conformations in the MD simulations (Figure S9D), while the conserved salt-bridge to Asp372_L remains intact. On the basis of the water network analysis and DFT calculations, we suggest that at least one proton could be transferred from the cytoplasmic bulk phase to the NiFe center for the H₂ production via Glu21_L, Cys374_L, and the water channels formed around the β 1– β 2 loop.

Functional Hydration of the Mbh Membrane Domain. The membrane domain of Mbh undergoes a significant hydration change during the MD simulations, in which around 200 water molecules establish pathways that could enable both proton and Na⁺ transport across the membrane (Figure 2A,B and Figure S2). We observe a hydration site from the cytoplasmic side at the large cleft between MbhH and MbhM (Figure S5A,B). These water molecules form hydrogen-bonded arrays toward the central axis of MbhH connecting Lys256_H with His346_H, His350_H, and Lys354_H. This pathway leads to the cytoplasmic bulk between MbhH and MbhM via Lys409_H (Figure 5A) and accounts for a significant portion of the water influx toward MbhH. The same site was recently suggested to support Na⁺ transport in the related Mrps,¹⁴ although the chain of three conserved lysine residues along the pathway is expected to electrostatically disfavor Na⁺ transport.

A transient water chain also connects Glu141_M with Lys409_H across the nonpolar lipid-filled cleft (Figure 5C,D). However, these residues are ca. 10–12 Å apart when Glu141_M

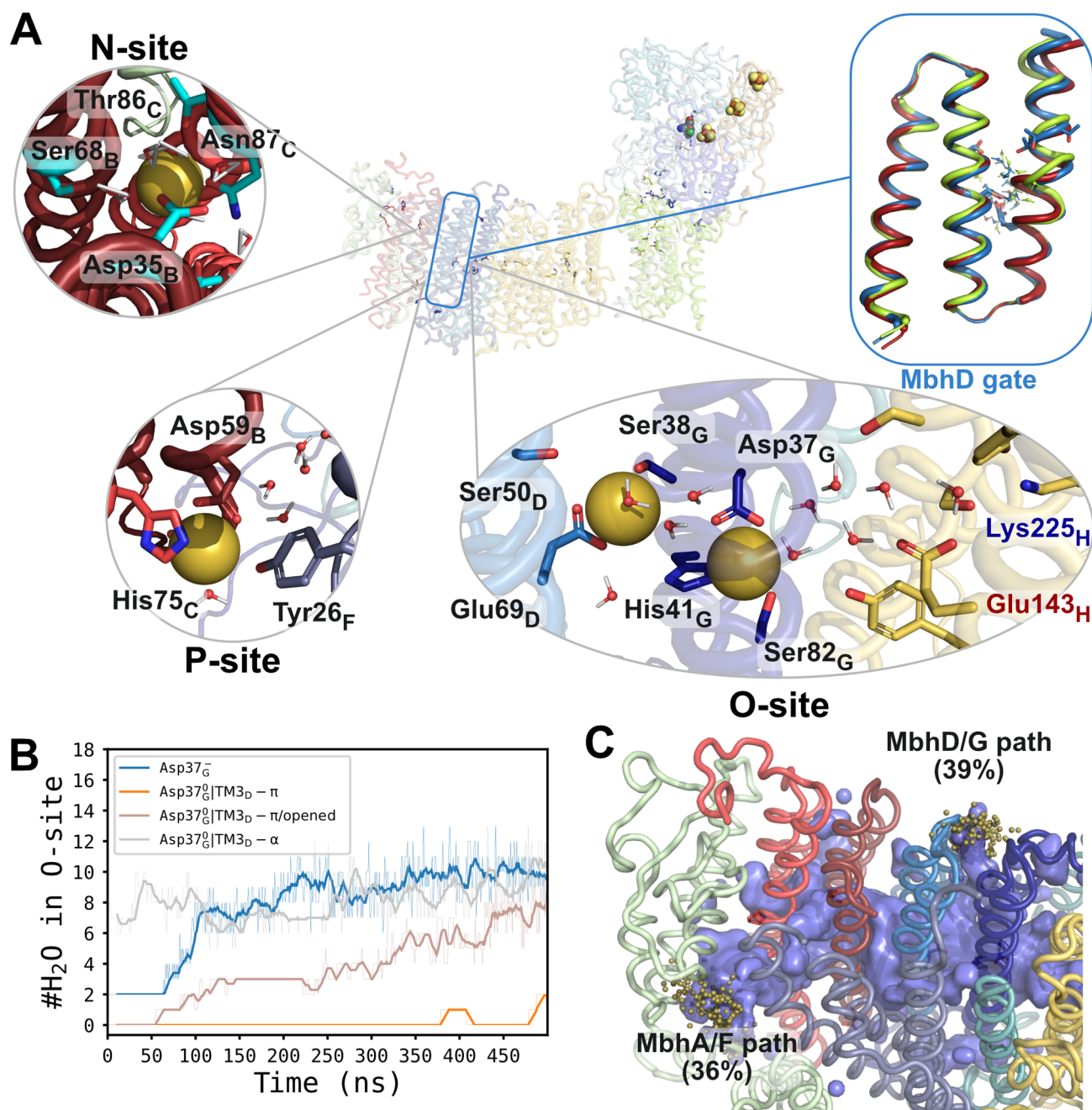


Figure 6. Na⁺/H⁺ transport in MbhA-G. (A) Putative sodium binding sites and functional motifs in Mbh based on MD simulations. Insets: the N-site at MbhA/C, here with a bound Na⁺, is accessible from the cytoplasmic bulk; O-site in MbhG with two modeled Na⁺ ions; P-site, which spontaneously binds bulk Na⁺ ions during most MD simulations; and a putative conformational gate at TM2/TM3 of MbhD modulating ion transport between the N- and O-sites. (B) Hydration of the O-site in different protonation and conformational states. Asp37_G deprotonation (blue) and conformational changes in the MbhD gate (brown) favor hydration of the O-site. Hydration in the closed-gate conformation (π -bulge form) is low (orange), whereas biasing toward α -helical conformation (gray) increases the hydration levels (see Figure S6). (C) O-site hydration in MD simulations. Influx via MbhD/G and MbhA/F accounts for ca. 75% of the observed water molecules in the region. The MbhA/F pathway is connected from the N-site via the MbhD gate to the O-site. Inlets from pathway cluster analysis are shown as spheres, and areas with volumes of greater than median hydration are depicted as surface.

is modeled in a deprotonated state and ca. 18 Å upon protonation of the latter, rendering proton exchange between the residues unlikely, especially since no titratable groups stabilize the long hydrogen-bonding wire.

Three to four PI lipids bind to the cleft region (Figure 5A), sealing the MbhH/MbhM gap from the periplasmic side and the horizontal gap at the MbhH/MbhM interface near the broken helix TM12 (Figure 5A). This observation is consistent

with the blurred density around the region in the cryoEM map.¹

Our simulations also indicate that binding of at least one additional lipid from the cytoplasmic side is sterically possible, which could block the proton transfer across the MbhM/MbhH interface (Figure S3D).

We observe another water influx site from the periplasmic side that reaches Lys409_H between MbhI and MbhH (TM12),

with many conserved polar residues lining up along the pathway (Figure 5A,B). This cluster, which has an overall occupancy of around 5%, is not present in all simulations (Figure S3B), but it could nevertheless be functionally relevant. Although no water molecules enter the hydrophilic axis of MbhH via TM7, we observe a partial pathway leading to the cytoplasmic side around Lys256_H, His350_H, Lys354_H, and water molecules (Figure 5B). These findings thus indicate that a canonical S-shaped pathway, analogous to those observed in complex I,^{20,21,23,24} could also establish a proton pathway in MbhH.

Residues along the lateral proton transfer wire in MbhH strongly interact with the conserved Lys225_H/Glu143_H ion pair at the interface of MbhH and a putative Na⁺-binding MbhG/D (see below), but the two regions do not exchange water molecules (Figure 5A). The conformation of this ion pair could modulate the proton transfer barrier along the lateral pathway in MbhH, and vice versa, similar to what has been observed in complex I (see below).^{8,16,19,20,36}

This ion pair also forms strong electrostatic interactions with a charged cluster in MbhG/D, comprising Asp37_G, His41_G, and Glu69_D residues that bind Na⁺ in Mrp (Figure 6A, see below).¹⁵ This region, called here the “O-site” (for the occluded Na⁺ binding site) together with the ion pair, is hydrated from the cytoplasmic side of the membrane during our simulations and could form key elements enabling proton-coupled Na⁺ transport.

Hydration Dynamics of the Putative Na⁺-Binding Site. We find that hydration of the putative O-site occurs via two main pathways, which account for >75% of the water molecules observed in this region. The major channel leads from the cytoplasmic side at the MbhA/MbhF interface and is established around a kink region at Pro88_C of the broken helix TM3 of MbhC (Figure 6A). The water molecules flow in via a cluster of conserved polar residues in MbhB and MbhC (Thr39_B/Thr42_C and Asp35_B/Asn38_B/Thr86_C) that could form a primary Na⁺ binding site (Figure 6A), here called the “N-site” (for negatively charged side Na⁺-binding site). Interestingly, a similar motif establishes a Na⁺ binding site in the unrelated light-triggered Na⁺-pump KR2.³² In the related Mrps, a Na⁺ pathway was recently suggested to also involve a Thr/Asn cluster,¹⁵ but leading via His137_A toward the proposed N-site.¹⁵ In our simulations, the latter pathway remains sealed from water molecules, whereas in Mrps, two detached TM-helices (in the subunit homologous to MbhA) could open up this channel.

To further probe the principles underlying sodium binding, we placed a Na⁺ ion around the N-site in MbhB/MbhC. In these simulations, the Na⁺ rapidly finds a binding pose that is stabilized by Asp35_B, Asn87_B, Ser68_B, Thr86_C, and water molecules (Figure 6A and Figure S5). Mutations of the homologous residues in Mrps block Na⁺ transport activity, thus further supporting the functional relevance of this site.¹⁴

Water molecules also enter from the cytoplasmic side between MbhD and MbhG close to Lys23_G and Lys28_G and lead further to the O-site that could support proton transfer across this region. The MbhD/G pathway is favored in simulations, where Asp37_G is modeled in a deprotonated (anionic) state. We note, however, that the high positive charge around this region is unlikely to support Na⁺ transport (cf. ref 14), unless protonation changes are involved (cf. also ref 1).

MbhD Regulates Ion Transport to the Na⁺/H⁺ Coupling Site. The conformation of the π -bulge in TM3 of MbhD (residues 71–79) correlates with the overall water influx toward the putative Na⁺/H⁺ coupling site in MbhD/G (Figure 6B). In the hydrated state, the π -bulge moves toward MbhF and opens up a gap between TM2 and TM3 of MbhD (Figure 6B). This leads to an increase in the hydration level by a factor of 2 (Figure 6B and Figure S6B), whereas in the dry state, the TM2 and TM3 helices remain in close contact and prevent water diffusion across the site (Figure 6A).

The conformational state of the TM3 helix of ND3 in the canonical complex I was recently suggested to regulate proton transfer during the active-to-deactive transition and possibly during turnover.^{23–27} More specifically, the α -helical form, present in the active state of complex I, favors well-wired proton pathways, whereas the π -bulge, observed in the deactive state, could block proton transfer.²⁴ Interestingly, the homologous TM3 of MrpA is captured in an α -helical form,¹⁵ suggesting that conformational transitions between a π -bulge and α -helix could also be involved in Mbh.

To further probe how such conformational transitions affect the hydration dynamics in Mbh, we perturbed TM3 of MbhD to form an α -helix during the MD simulations (see Supporting Information Methods). The modeled α -helix remains dynamically stable for 0.5 μ s during unrestrained MD simulations (Figure S7) and results in rapid hydration of the sodium cavity from the cytoplasmic side via MbhA (Figure 6B,C). In stark contrast, in simulations where TM3 forms a π -bulge, the region becomes 90% less hydrated (Figure 6B). These findings suggest that TM3 of MbhD could act as a gate that controls water and Na⁺ exchange between the proposed N- and O-binding sites.

Sodium Binding at the P-Site. We observe spontaneous sodium binding in nearly all MD simulations at the periplasmic side surface in a cavity formed by MbhB, MbhC, and MbhF (cf. also ref 1). The Na⁺ binds to the conserved Asp59_B, Tyr26_F, and His75_C or, in a few simulations, to the nearby loop of MbhA. Both binding modes remain highly stable throughout the 0.5 μ s simulations (Figure S5G), with rapid water exchange with the bulk. However, in contrast to the N-site at the MbhB/C interface, we do not observe pathways leading toward the interior of Mbh or to the O-site, located ca. 18 Å away. However, Na⁺ transport to/from the P-site could be achieved by conformational switching into an alternate access state,³⁷ possibly supported by the structurally similar TM motifs also found in other transporter proteins.³⁸

Sodium Transport between Binding Sites. To probe the Na⁺ transport mechanism between the putative N- and O-sites, we placed sodium ions at the respective sites or at the MbhC/MbhD interface. The Na⁺ ions remain tightly bound at the N- or O-sites on 0.5 μ s time scales (Figure 7A), but at the interface region, the Na⁺ interacts with TM2 and TM3 of MbhD for ca. 0.5 μ s, after which it moves to the O-site to a position that closely resembles the binding mode in Mrp (Figure 7B and Figure S10).¹⁵ These findings support that the MbhD gate could be involved in Na⁺ transport in Mbh.

We also probed the reverse Na⁺ transport direction by placing Na⁺ ions at both sides of His41_G, as experimentally resolved for Mrp.¹⁵ When the carboxylates are modeled in their deprotonated states, the Na⁺ ions remain strongly bound at the O-site (Figures 6A and 7A), whereas upon protonation of Asp37_G and Glu69_D, the Na⁺ moves to the N-site within 50 ns (Figure 7A). The Na⁺ diffusion leads to a subtle

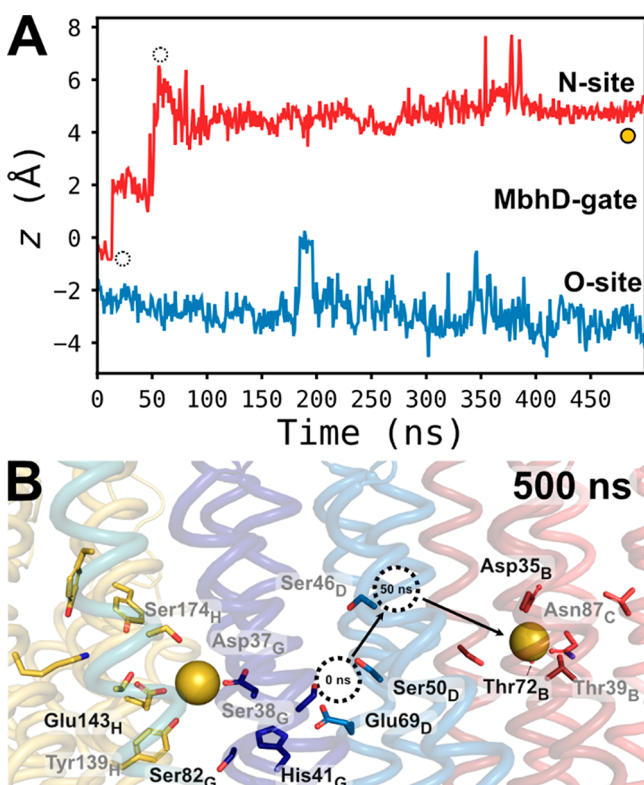


Figure 7. Sodium transport is influenced by the protonation states in the O-site. (A) Dynamics of Na^+ ions between the O- and N-sites in unbiased MD simulations from the center of the membrane ($z = 0 \text{ \AA}$) with Asp37_G modeled in deprotonated (blue) and protonated (red) states. The structure depicted in (B) is marked with a yellow circle. (B) Structural snapshot of the sodium motion after 500 ns of classical MD simulation. Intermediate sodium positions during the simulation are marked with dashed circles in the structure and the time trace depicted in panel A. Residues forming contacts with sodium ions during the simulation are listed in black if the ConSurf score is at least eight and otherwise in gray (see Table S4).

conformational change in the π -bulge region of MbhD-TM3 (Figure S10). Na^+ binding drastically reduces the proton affinity of both Glu69_D and Asp37_D, suggesting that the Na^+/H^+ binding events are tightly coupled (Figure S8).

The water network analysis suggests that Asp37_G and Glu143_H are in direct hydrogen-bonded contact via 2–3 water molecules, which could enable proton transfer between the residues upon Na^+ binding. To test this process, we transferred the proton from Asp37_G to Glu143_H, which leads to rapid dissociation of Lys225_H toward Lys256_H. These conformational changes lower the pK_a of Lys256_H and could thus trigger proton transfer in the MbhH subunit (Figure 8A and Figure S8).

DISCUSSION

Mbh was originally suggested to function as a proton pump based on experiments performed in membrane vesicles.³ However, it is also possible that the enzyme employs the small energy transduced from Fd-driven hydrogen production to drive secondary active sodium/proton exchange,^{4,6} an operational mode that is consistent with the sodium-dependent F_0F_1 -ATP synthase in *P. furiosus*.³³ Sodium functions as the coupling ion in the homologous hydrogenase in *Thermococcus*

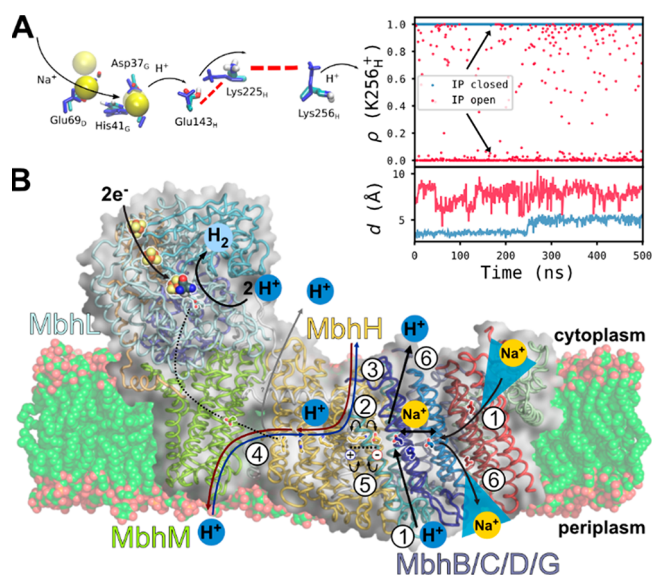


Figure 8. Putative redox-driven Na^+/H^+ transport and proton pumping in Mbh. (A) Sodium binding to the O-site triggers protonation changes, conformational switching of the Glu143_H/Lys225_H ion pair and deprotonation of Lys256_H. Inset: the predicted protonation fraction of Lys256_H for the open ion pair (blue) and closed ion pair conformations (red). (B) Summary of mechanistic model: (1) Na^+ uptake to the O-site via MbhB/C/D (N-site) and H^+ uptake via MbhB, (2) modulate the conformational state of the buried ion pair in MbhH, and (3) trigger proton transfer across the membrane in MbhH. (4) Reprotonation of Lys256_H induces (5) association of the ion pair and (6) ejection of the Na^+ and H^+ across opposites sides of the membrane. Protonation of the O-site lowers the affinity of the Na^+ , which could leave via the P-site by conformational changes in the MbhD gate (see text). Protons could exit to the cytoplasmic side via a water cluster observed in MbhG.

onnurineus NA1, supporting a possible similar role also in Mbh.³⁴

Although the exact Na^+/H^+ stoichiometry of Mbh is unknown, the bioenergetic boundary conditions thermodynamically allow Mbh to transfer one ion (Na^+ or H^+) per $2e^-$ at 120 mV sodium motive force (smf) and thus enable the possible electrogenic function of Mbh (cf. also refs 3 and 35). In contrast, for the related Mrps, which lack the redox module, mechanisms based on one Na^+ exchanged per proton¹⁵ or even sodium transport without H^+ exchange were recently suggested.¹⁴

We propose that the terminal MbhA-MbhG module functions as the sodium translocation domain in Mbh (cf. also refs 1 and 15), whereas the hydration dynamics in MbhH supports its involvement in proton transfer. Our findings indicate that the conserved Lys225_H/Glu143_H ion pair of MbhH and the putative Na^+/H^+ binding O-site establish a coupling element between the proton and sodium transport and proton pumping (Figure 8A). To this end, the sodium affinity at the O-site is modulated by the protonation state of the Glu69_D/His41_G/Asp37_G cluster, and vice versa. Na^+ binding to this site could trigger dissociation of the Lys225_H/Glu143_H ion pair (Figures 7 and 8A) and, in turn, induce proton transfer from Lys256_H (Figure 8A), either directly across the membrane via His350_H/Lys354_H and water molecules around TM7 (Figure 8B) or laterally toward the MbhH/M interface (Figure 8B). However, because the unresolved lipid cleft poses challenges in the current modeling

and the region lacks titratable groups, it remains possible that the MbhH/M interface is not used for proton transfer to the cytoplasmic side. As an alternative, we propose that the protons are taken up via the water-mediated pathway formed along TM12 and transferred via the Lys409_H-His346_H-His350_H chain to reprotonate Lys256_H. This is expected to result in reassociation of the Lys225_H/Glu143_H ion pair and ejection of the proton from the O-site to the cytoplasmic side as well as Na⁺ ejection to the periplasmic side. The ion release could involve conformational switching into an alternate access state, e.g., at the MbhB/D interface. Conformational changes in MbhD were found to favor Na⁺ transfer between the N- and O-sites (Figure S10), whereas the proton release could occur via the water cluster observed at MbhD/G.

In this putative model, the transport of Na⁺ and H⁺ across the membrane could be triggered by reprotonation of Lys256_H in MbhH, thus following overall similar, although simpler, physical principles as proposed for complex I.⁸ Although the exact molecular principles of the redox-driven conformational changes in the conserved loop regions and charge arrays in MbhM and MbhI/L remain unclear, we note that the so-called E-tunnel could regulate the proton affinity and accessibility of the terminal lysine in MbhH and be involved in coupling the redox reactions with the charge transport process.

On the basis of the structural similarity to complex I, we note that the functional elements in MbhH are also expected to support proton conduction in the reverse direction, depending on the external conditions, and therefore do not exclude the possibility of its involvement in establishing a secondary proton gradient (Figure 8B, cf. refs 4 and 6).

Our study also found evidence for proton pathways leading to the NiFe center, with input sites, around the highly conserved $\beta 1$ - $\beta 2$ loop interface, leading to Cys374_L and the metal bound hydride in the active site via Glu21_L. Glu21_L undergoes a protonation-state-dependent conformational switching, which could help shuttle protons to the active site. A similar conformational switching of the functionally central Glu242 in cytochrome *c* oxidase has been suggested to favor kinetic gating and prevent possible back-leaks.³⁹

Previous MD simulations on the NiFe-hydrogenase from *D. vulgaris* revealed three proton pathways leading to Glu34_L (equivalent of Glu21_{MbhL}),⁴⁰ a residue that is also supported by other experimental⁴¹ and computational³⁰ studies. More specifically, two pathways were observed in NiFe-hydrogenases that correspond to our channels at the MbhL/I/J interface, whereas another pathway showed overall resemblance to our MbhL/N/J channel.⁴⁰ To this end, we note that the membrane subunit MbhM could lead to some structural differences at the MbhL/M interface. A pathway resembling the MbhL/M site was also observed in respiratory complex I (Figure S3),⁴² whereas hydration through the E-channel from the cleft region in Mbh has not been described before. Interestingly, His75_{MbhL} has also been suggested to form an alternative pathway to the active site.^{21,43} Although we find no significant contribution of the latter channel in the modeled states (Figure S3A), our data provide further support for several of the other pathways previously observed in NiFe-hydrogenases.

We note that oxygen sensitivity in hydrogenases has been linked to a proximal 3Fe4S iron sulfur center as well as to the topology of the hydrophobic gas channels leading to the active site.^{41,43,44} Mbh shows an oxygen tolerance with a half-life of around 14 h in oxygen sensitivity assays,⁶ despite comprising only 4Fe4S centers. Mutational studies suggest that con-

striction of the hydrophilic tunnel around the corresponding MbhI/J/L pathway observed here can significantly enhance the oxygen tolerance in NiFeSe-hydrogenases from *D. vulgaris*.⁴⁵ These findings indicate that the MbhI/J/L pathway could also be relevant for the oxygen tolerance of Mbh, whereas explicit diffusion of O₂ or H₂ along the channels would be necessary to study the effect.

CONCLUSIONS

We have presented here functional dynamics of the membrane-bound hydrogenase (Mbh) from *Pyrococcus furiosus* by using large-scale molecular simulations. In summary, we observed putative water-mediated proton pathways leading along the MbhL/MbhM interface to Glu21_L, which could shuttle protons to the NiFe center, responsible for the H₂ production. We also observed significant hydration changes in the MbhH subunit during the MD simulations as well as pK_a shifts upon conformational changes in a buried ion-pair at the MbhH/MbhG interface—functional elements that could support proton transfer across the archaeal membrane. The simulations also revealed three putative Na⁺ binding sites that could be responsible for the Na⁺/H⁺ transport activity of Mbh. The N-site at the MbhC/B interface is a Thr/Ser-rich region, which, via conformational changes in a transmembrane helix of MbhD, can transfer Na⁺ ions to the buried carboxylate rich O-site, located at the MbhG/MbhD interface. Conformational switching of an analogous TM-helix was recently suggested to regulate proton transfer also in the canonical complex I²⁴ (cf. also ref 23). We further found that the Na⁺ affinity of the O-site is sensitive to the protonation state of Asp37_G, and vice versa. This putative Na⁺/H⁺ binding site is electrostatically strongly coupled to the buried ion pair at the MbhH/MbhG interface, which in turn could trigger proton transfer along MbhH. We also found evidence for a putative Na⁺ binding site at the periplasmic side of the membrane in MbhC/B/F, which upon conformational changes could become accessible to the O-site binding region. Our combined findings provide insight into key conserved coupling principles within the complex I superfamily and detailed functional insight into archaeal energy transduction mechanisms.

MATERIALS AND METHODS

The cryo-EM structure of Mbh from *Pyrococcus furiosus* (PDB ID: 6CFW)¹ was embedded in a 1-palmitoyl-2-palmitoleoyl-*sn*-glycero-3-phosphoinositol (PYPI) membrane by using CHARMM-GUI.⁴⁶ We additionally modeled two PYPI lipid molecules in the cleft between MbhH and MbhM, unresolved side chains in MbhI, and missing N/C-terminal residues. The model was embedded in a 200 × 100 × 168 Å³ box comprising TIP3P water molecules and ions to mimic a 250 mM NaCl concentration. MD simulations were performed by using the CHARMM36m force field⁴⁷ in combination with force field parameters for the NiFe site for the NIB (Fe^{II}/Ni^{III}-OH⁻) and NIC (Fe^{II}/Ni^{III}-H⁻ ligand) states.⁴⁸ The MD simulations were performed by using NAMD ver. 2.13⁴⁹ with periodic boundary conditions (PBC) and long-range electrostatics modeled with the particle mesh Ewald (PME) approach with a grid size of 1 Å, at 1 bar and 310 K, and with a 2 fs integration time step. The membrane was first equilibrated around the protein, with heavy atoms restrained by a harmonic force constant of 10 kcal mol⁻¹ Å⁻². Lipid tails were initially melted with a harmonic restraint on the headgroups. After gradual heating to 310 K, all restraints were removed, followed by creation of the NIB and NIC states and models with different protonation states of the titratable residues. Initial protonation states were assigned based on electrostatic calculations (see the Supporting Information and Table S2). MDAnalysis,⁵⁰ Visual Molecular Dynamics,⁵¹ and PyMol⁵² were used

for analysis and visualization (see the Supporting Information Methods and Table S1 for further simulation details).

Clustering Analysis. Hydration dynamics in the MD simulations were analyzed in all states by aligning the trajectories around subunits MbhA-MbhH, MbhI, and MbhM, or MbhJ-MbhN. C_{α} atoms of residues in helices and sheets were used to define the scope of the analysis by a convex hull. Water molecules around putative sodium sites were identified within 6 Å of the center of mass of the O-site (E143_H, K225_H, D37_G, H41_G, E69_D), N-site (D35_B), or P-site (D59_B). For MbhH, water molecules within 4 Å of K409_H, H350_H, K354_H, or K256_H, and for the NiFe site, water molecules within 4 Å of E21_L were considered in the analysis. The clustering analysis was performed with Aquaduct ver. 1.0.11⁵⁵ using the Barber algorithm (at a cutoff of 1.4 Å), with path trimming for the sodium site analysis. Water analysis was performed on the full data set for all sites in the membrane domain. Water clusters in the active site were analyzed based on simulations S3, S5, S10, and S11 (Table S1). The largest cluster was recursively divided into two subclusters by using the balanced-iterative reducing and clustering using the hierarchies (BIRCH) method for simulations S3 and S10 (for NiFe site) and simulations S1, S3, S4, and S9 (for MbhH). Clusters with <10 members were assigned to the outlier cluster, and conserved residues were identified by using ConSurf.⁵⁴ For the visualization of cluster medoid paths, paths with the same input and output cluster were chosen.

DFT Models. Quantum chemical DFT models of the NiFe site were created based on the MD-minimized Mbh model. The DFT model comprised E21_L, C68_L, I70_L, C71_L, H75_L, T376_L, R320_L, D372_L, C374_L, C377_L, and N36_J, the NiFe core, and three H₂O molecules. Amino acids were cut and saturated with hydrogen atoms at the α - $C\beta$ bond. The DFT model comprised 126 atoms, which were structure optimized at the B3LYP-D3/ $\epsilon=4$ level^{55,56} by using the def2-SVP basis sets for all atoms except Fe and Ni, which were modeled with the def2-TZVP basis sets.⁵⁷ The reaction pathway for proton transfer from E21_L to C374_L and C374_L to H⁻ were optimized along minimum-energy pathways, followed by optimization of H₃O⁺ transition state, which showed one imaginary frequency at -924.6 cm⁻¹ (see the Supporting Information Methods). All systems were optimized in the triplet spin state. TURBOMOLE ver. 7.5⁵⁸ was used for the DFT calculations (see Supporting Information Methods, Figure S4, and Table S5 for details of all QM calculations).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.1c09356>.

Figures showing global dynamics, lipid binding, hydration analysis, hydration pathways, ion-pair distance distributions, radial distribution functions, surface contact area and its anticorrelation with O-site hydration, helicity measures, sodium dynamics, QM models, conserved residues, pK_a distributions and simulation details (PDF)

Movie S1 (MPG)

■ AUTHOR INFORMATION

Corresponding Author

Ville R. I. Kaila – Department of Biochemistry and Biophysics, Stockholm University, 10691 Stockholm, Sweden; Department of Chemistry, Technical University of Munich, 85748 Garching, Germany; orcid.org/0000-0003-4464-6324; Email: ville.kaila@dbb.su.se

Authors

Max E. Mühlbauer – Department of Biochemistry and Biophysics, Stockholm University, 10691 Stockholm, Sweden;

Department of Chemistry, Technical University of Munich, 85748 Garching, Germany

Ana P. Gamiz-Hernandez – Department of Biochemistry and Biophysics, Stockholm University, 10691 Stockholm, Sweden

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/jacs.1c09356>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement number 715311), the Knut and Alice Wallenberg Foundation, the Swedish Research Council, and the German Research Foundation, TRR235 (Emergence of life). Computing resources were provided by LRZ/SuperMuc (grant: pn34he) and SNIC/PDC (SNIC 2020/1-38) at PDC Centre, partially funded by the Swedish Research Council through grant agreement no. 2016-07213.

■ REFERENCES

- (1) Yu, H.; Wu, C.-H.; Schut, G. J.; Haja, D. K.; Zhao, G.; Peters, J. W.; Adams, M. W. W.; Li, H. Structure of an Ancient Respiratory System. *Cell* **2018**, *173*, 1636–1649.
- (2) Sapra, R.; Verhagen, M. F.; Adams, M. W. W. Purification and characterization of a membrane-bound hydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* **2000**, *182*, 3423–3428.
- (3) Sapra, R.; Bagramyan, K.; Adams, M. W. W. A simple energy-conserving system: proton reduction coupled to proton translocation. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 7545–7550.
- (4) Schut, G. J.; Boyd, E. S.; Peters, J. W.; Adams, M. W. W. The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiol. Rev.* **2013**, *37*, 182–203.
- (5) Mayer, F.; Müller, V. Adaptations of anaerobic archaea to life under extreme energy limitation. *FEMS Microbiol. Rev.* **2014**, *38*, 449–472.
- (6) McTernan, P. M.; Chandrayan, S. K.; Wu, C. H.; Vaccaro, B. J.; Lancaster, W. A.; Yang, Q.; Fu, D.; Hura, G. L.; Tainer, J. A.; Adams, M. W. W. Intact functional fourteen-subunit respiratory membrane-bound [NiFe]-hydrogenase complex of the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **2014**, *289*, 19364–19372.
- (7) Yu, H.; Haja, D. K.; Schut, G. J.; Wu, C. H.; Meng, X.; Zhao, G.; Li, H.; Adams, M. W. W. Structure of the respiratory MBS complex reveals iron-sulfur cluster catalyzed sulfane sulfur reduction in ancient life. *Nat. Commun.* **2020**, *11*, 5953.
- (8) Kaila, V. R. I. Long-range proton-coupled electron transfer in biological energy conversion: Towards mechanistic understanding of respiratory complex I. *J. R. Soc., Interface* **2018**, *15*, 20170916.
- (9) Sazanov, L. A. A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 375–388.
- (10) Hirst, J. Mitochondrial Complex I. *Annu. Rev. Biochem.* **2013**, *82*, 551–575.
- (11) Yu, H.; Schut, G. J.; Haja, D. K.; Adams, M. W. W.; Li, H. Evolution of complex I-like respiratory complexes. *J. Biol. Chem.* **2021**, *296*, 100740.
- (12) Kaila, V. R. I.; Wikström, M. Architecture of bacterial respiratory chains. *Nat. Rev. Microbiol.* **2021**, *19*, 319–330.
- (13) Lubitz, W.; Ogata, H.; Rüdiger, O.; Reijerse, E. Hydrogenases. *Chem. Rev.* **2014**, *114*, 4081–4148.
- (14) Li, B.; Zhang, K.; Nie, Y.; Wang, X.; Zhao, Y.; Zhang, X. C.; Wu, X. L. Structure of the *Dietzia* Mrp complex reveals molecular

mechanism of this giant bacterial sodium proton pump. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 31166–31176.

(15) Steiner, J.; Sazanov, L. Structure and mechanism of the Mrp complex, an ancient cation/proton antiporter. *eLife* **2020**, *9*, e59407.

(16) Di Luca, A.; Gamiz-Hernandez, A. P.; Kaila, V. R. I. Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E6314–E6321.

(17) Kaila, V. R. I.; Wikström, M.; Hummer, G. Electrostatics, hydration, and proton transfer dynamics in the membrane domain of respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 6988–6993.

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

(19) Mühlbauer, M. E.; Saura, P.; Nuber, F.; Di Luca, A.; Friedrich, T.; Kaila, V. R. I. Water-gated proton transfer dynamics in respiratory complex I. *J. Am. Chem. Soc.* **2020**, *142*, 13718–13728.

(20) Röpke, M.; Saura, P.; Riepl, D.; Pöverlein, M. C.; Kaila, V. R. I. Functional Water Wires Catalyze Long-Range Proton Pumping in the Mammalian Respiratory Complex I. *J. Am. Chem. Soc.* **2020**, *142*, 21758–21766.

(21) Grba, D. N.; Hirst, J. Mitochondrial complex I structure reveals ordered water molecules for catalysis and proton translocation. *Nat. Struct. Mol. Biol.* **2020**, *27*, 892–900.

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

(23) Kampjut, D.; Sazanov, L. A. The coupling mechanism of mammalian respiratory complex I. *Science* **2020**, *370*, eabc4209.

(24) Röpke, M.; Riepl, D.; Saura, P.; Di Luca, A.; Mühlbauer, M. E.; Jussupow, A.; Gamiz-Hernandez, A. P.; Kaila, V. R. I. Deactivation blocks proton pathways in the mitochondrial complex I. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2019498118.

(25) Di Luca, A.; Kaila, V. R. I. Molecular strain in the active/deactive-transition modulates domain coupling in respiratory complex I. *Biochim. Biophys. Acta, Bioenerg.* **2021**, *1862*, 148382.

(26) Agip, A. N. A.; Blaza, J. N.; Bridges, H. R.; Viscomi, C.; Rawson, S.; Muench, S. P.; Hirst, J. Cryo-EM structures of complex I from mouse heart Mitochondria in two Biochemically Defined States. *Nat. Struct. Mol. Biol.* **2018**, *25*, 548–556.

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

(28) Lobasso, S.; Lopalco, P.; Angelini, R.; Vitale, R.; Huber, H.; Müller, V.; Corcelli, A. Coupled TLC and MALDI-TOF/MS analyses of the lipid extract of the hyperthermophilic archaeon *Pyrococcus furiosus*. *Archaea* **2012**, *2012*, 957852.

(29) Siegbahn, P. E. M.; Liao, R. Z. The Energetics of Hydrogen Molecule Oxidation in NiFe-hydrogenase. *ACS Catal.* **2020**, *10*, 5603–5613.

(30) Siegbahn, P. E. M.; Tye, J. W.; Hall, M. B. Computational studies of [NiFe] and [FeFe] hydrogenases. *Chem. Rev.* **2007**, *107*, 4414–4135.

(31) Rich, P. R.; Maréchal, A. Functions of the hydrophilic channels in protonmotive cytochrome c oxidase. *J. R. Soc., Interface* **2013**, *10*, 20130183.

(32) Suomivuori, C. M.; Gamiz-Hernandez, A. P.; Sundholm, D.; Kaila, V. R. I. Energetics and dynamics of a light-driven sodium-pumping rhodopsin. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 7043–7048.

(33) Pisa, K. Y.; Huber, H.; Thomm, M.; Müller, V. A sodium ion-dependent A1AO ATP synthase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *FEBS J.* **2007**, *274*, 3928–3938.

(34) Lim, J. K.; Mayer, F.; Kang, S. G.; Müller, V. Energy conservation by oxidation of formate to carbon dioxide and hydrogen via a sodium current in a hyperthermophilic archaeon. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 11497–502.

(35) Calisto, F.; Sousa, F. M.; Sena, F. V.; Refojo, P. N.; Pereira, M. M. Mechanisms of Energy Transduction by Charge Translocating Membrane Proteins. *Chem. Rev.* **2021**, *121*, 1804–1844.

(36) Haapanen, O.; Sharma, V. Role of water and protein dynamics in proton pumping by respiratory complex I. *Sci. Rep.* **2017**, *7*, 7747.

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

(38) Forrest, L. R.; Zhang, Y.; Jacobs, M. T.; Gesmonde, J.; Xie, L.; Honig, B. H.; Rudnick, G. Mechanism for alternating access in neurotransmitter transporters. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10338–10343.

(39) Kaila, V. R. I.; Verkhovsky, M. I.; Hummer, G.; Wikström, M. Glutamic acid 242 is a valve in the proton pump of cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 6255–6259.

(40) Sumner, I.; Voth, G. A. Proton Transport Pathways in [NiFe]-Hydrogenase. *J. Phys. Chem. B* **2012**, *116*, 2917–2926.

(41) Tai, H.; Hirota, S.; Stripp, S. T. Proton Transfer Mechanisms in Bimetallic Hydrogenases. *Acc. Chem. Res.* **2021**, *54*, 232–241.

(42) Jussupow, A.; Di Luca, A.; Kaila, V. R. I. How cardiolipin modulates the dynamics of respiratory complex I. *Sci. Adv.* **2019**, *5* (3), 1850.

(43) Fontecilla-Camps, J. C.; Volbeda, A.; Cavazza, C.; Nicolet, Y. Structure/Function Relationships of [NiFe]- and [FeFe]-Hydrogenases. *Chem. Rev.* **2007**, *107*, 4273–4303.

(44) Kalms, J.; Schmidt, A.; Frielingsdorf, S.; Utesch, T.; Gotthard, G.; von Stetten, D.; van der Linden, P.; Royant, A.; Mroginski, M. A.; Carpentier, P.; Lenz, O.; Scheerer, P. Tracking the Route of Molecular Oxygen in O₂-Tolerant Membrane-Bound [NiFe] Hydrogenase. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E2229–E2237.

(45) Zacarias, S.; Temporão, A.; del Barrio, M.; Fourmond, V.; Léger, C.; Matias, P. M.; Pereira, I. A. C. A Hydrophilic Channel Is Involved in Oxidative Inactivation of a [NiFeSe] Hydrogenase. *ACS Catal.* **2019**, *9*, 8509–8519.

(46) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.

(47) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

(48) Smith, D. M.; Xiong, Y.; Straatsma, T. P.; Rosso, K. M.; Squier, T. C. Force-Field Development and Molecular Dynamics of [NiFe] Hydrogenase. *J. Chem. Theory Comput.* **2012**, *8*, 2103–2114.

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

(50) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.

(51) Humphrey, W. A.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(52) Delano, W. L. The PyMOL Molecular Graphics System, ver. 2.5; <https://pymol.org/2/> (accessed 2021-10-29).

(53) Magdziarz, T.; Mitusińska, K.; Bzówka, M.; Raczyńska, A.; Stańczak, A.; Banas, M.; Bagrowska, W.; Góra, A. AQUA-DUCT 1.0: structural and functional analysis of macromolecules from an intramolecular voids perspective. *Bioinformatics* **2020**, *36*, 2599–2601.

(54) Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **2016**, *44*, W344–W350.

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

(56) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A. Consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **2010**, *132*, 154104–19.

(57) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(58) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.