Supplementary Information

Water-Gated Proton Transfer Dynamics in Respiratory Complex I.

Max E. Mühlbauer[1,2,a], Patricia Saura[1,2,a], Franziska Nuber[3], Andrea Di Luca[1,2], Thorsten Friedrich[3], Ville R. I. Kaila[1,2,*]

[1] Department of Biochemistry and Biophysics, SE-106 91 Stockholm University, Stockholm, Sweden.
[2] Center for Integrated Protein Science Munich at the Department of Chemistry, Technical University of Munich, Lichtenbergstr. 4, D85748 Garching, Germany.
[3] Institut für Biochemie, Albert-Ludwigs-Universität Freiburg, Albertstrasse 21, 79104, Freiburg Germany.
[a] Contributed equally to this work.
[*]Corresponding author: ville.kaila@dbb.su.se

Supplementary Methods

Cell growth and isolation of cytoplasmic membranes. Strains were grown at 37°C in baffled flasks using minimal medium[1] with 25 mM acetate as sole carbon source. After 3 h growth expression of the *nuo* operon was induced by an addition of 0.02% (w/v) L-arabinose. Cells that were used to prepare cytoplasmic membranes were grown on auto-induction medium (1% (w/v) peptone, 0.5% (w/v) yeast extract, 0.4% glycerol, 25 mM $Na_2HPO_4 \cdot 2\ H_2O$, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4 \cdot 7\ H_2O$, 0.2% (w/v) L-arabinose, 0.05% (w/v) glucose, 30 $mg \cdot L^{-1}$ Fe-NH_4-citrate, 0.5 mM l-cysteine, 50 $mg \cdot L^{-1}$ riboflavine) containing chloramphenicol (34 µg/mL). Cells were harvested and cytoplasmic membranes were isolated using an EmulsiFlex (Avestin) as described in Ref 2. When cytoplasmic membranes were used for protein preparation, the membrane fraction was suspended after ultracentrifugation in buffer $A^{*}_{pH6.8}$ (50 mM MES/NaOH, 50 mM NaCl, 5 mM $MgCl_2$, 10% (v/v) glycerol, pH 6.8).

Other Analytical Procedures. Protein concentration was determined by the biuret method using BSA as a standard.[3] The concentration of purified complex I was determined by the difference of absorbance at 280-310 nm (TIDAS II, J&M Aalen) using an ε of 781 $mM^{-1}cm^{-1}$ as derived from the amino acid sequence.[4] SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) was performed according to Schägger[5] with a 10% separating gel and a 3.9% stacking gel.

Graph analysis. Graphs were built by modelling the sidechains of charged/polar residues and water molecules in subunit Nqo13/Nqo14 as nodes, and hydrogen-bonds between the nodes of as edges. Edges were defined if the heavy atom (X, Y) distance was < 3 Å and the hydrogen-bond angle (X – H – Y) > 120º. To account for the two histidine tautomers (Nδ/Nε) and allow for rotameric flips of the sidechain, we also introduce an additional edge between the nitrogen atoms of the bridging histidine residues. The edge weight connecting two vertices was set to the distance between the nodes, and shortest paths were calculated using the Dijkstra's algorithm[6] based on the edge weights. Connectivity within the graph was calculated based as the fraction of unbroken hydrogen-bonding paths along the MD trajectory. Average connectivities (Figure 4B) were estimated from the Lys/His and His/Glu pathways, and additionally between the bridging histidine residues for *E. coli* complex I. The algorithm was implemented based on the network[7] and mdtraj[8] python libraries.

DFT calculations of subunit NuoM of *E. coli* complex I. DFT models were built based on snapshots from the MD trajectories of WT, H322A, H348A and H322A/H348A variants of *E. coli* complex I. The models included the following residues of subunit NuoM: Lys265, His(Ala)322, His(Ala)348, Glu407, Ser319, Ala352, Ala260, Leu264, Met323, Ser321, Ser351, Gln344, Leu396, Asn403, Thr318, Thr395, Tyr435, Ala432, Leu429 (Leu396/Leu429 were not included in H322A variant, and Leu396/Ala260 were not included in H322A/H348A variant, due to less proximity to pT wire residues), in addition to around 10 nearby water molecules, comprising 185-223 atoms in total. Cα-Cβ positions were cut and saturated with hydrogen atoms, and Cβ positions were kept fixed during geometry optimizations to simulate the protein framework. Geometry optimizations were performed at the B3LYP-D3/def2-SVP/ε=4 level,[9-13] and single point energies were computed at the B3LYP-D3/def2-TZVP/ε=4 level. Reaction pathways were optimized using a *chain-of-states* method.[14,15] The QM calculations were performed with TURBOMOLE v7.3.[16]

Kinetic Model and Master Equation. A kinetic master equation model was created, following the method by Kim *et al*.[17,18] The model contained protonation sites, channel sites, and ion-pair sites. The protonation sites were allowed to exchange protons with other protonation site and with the bulk, which was modelled as a proton reservoir with pH=7. The channel sites were linked to the protonation sites, and allowed to transfer protons when the channel is in the open state. The ion-pair site was allowed to switch between open and closed conformations, and the coupling interactions were treated as an input parameter of the kinetic model.

A five sites model system was built, as shown in Figure 4F, with C_{out} (channel site), E377 (protonation site), K235 (protonation site), C_{in} (channel site), K204/E123 (ion-pair site), according to residue numbering in Nqo13 subunit. Each site was modelled in occupied (open, χ =1) or empty (close, χ =0) states, leading to a total of 2^5= 32 microstates. A transition-rate matrix that satisfies detailed balance was built by allowing for 1) proton uptake from the N- and P-sides of the membrane with the "K235"- and "E377" sites, respectively; 2) proton transfer between K235 and E377; 3) ion-pair opening in all states but unique state-dependent transition rates; and 3) block all pT reactions to the bulk when the channel is in the closed states. The transition rate ($\boldsymbol{k_{ij}}$) between states $\boldsymbol{i}$ and $\boldsymbol{j}$ was defined as,

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

where $\boldsymbol{\kappa_{ij}}$ is the intrinsic rate between the two processes in absence of driving force, and E_i / E_j are the energies for unique microstates. The intrinsic rates were treated as model parameters. The factor ½ in the exponent accounts for equal contribution of product and reactant states on the forward (k_{ij}) / backward (k_{ji}) rates. The energy of single kinetic states was expressed as the sum of the single site energies and their couplings as,

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

where $\boldsymbol{E_n^{intr}}$ is the intrinsic energetic cost of protonating (or opening) the site, introduced as a parameter in the model. For the protonation sites the intrinsic energy was defined as $\boldsymbol{\Delta E_{prot} = -RT(p\mathrm{H} - pK_a)}$. Active directional proton pumping from the N-side to P-side was achieved by biasing the K204/E123 opening rate with a time-dependent biasing potential of the form,

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

A complete set of the parameters used for the kinetic simulations are listed in Table S3 and S4. The master equation was defined as,

$$\boldsymbol{\frac{dp_i}{dt} = \sum_{j=1}^{N_{states}} k_{ji} p_j - \sum_{j=1}^{N_{states}} k_{ij} p_i}$$

Given an initial equilibrium distribution of the site occupation ($\boldsymbol{dp_i/dt}$ = 0), the bias was turned on and the model was propagated for $\boldsymbol{t_{tot}}$ timesteps that corresponds to an integer number of bias cycles ($\boldsymbol{t_{tot} = nt_{period}}$). The pumping flux was measured as the number of protons transferred from E377 to the P-side (or alternatively from N-side to K235, equivalent within numerical accuracy of the model and in steady state conditions, measured as integral of the flux over the last cycle of pumping during the simulation.

$$\boldsymbol{E_{pump} = \int_{(n-1)t_{period}}^{nt_{period}} \sum_{i=1}^{N_{states}} \frac{dp_i(E377 \rightarrow P_{side})}{dt} dt}$$

Optimization of the model was obtained by maximizing $\boldsymbol{E_{pump}}$. Due to the high-dimensionality of the function, we optimized the solution starting from an estimate of the parameters based on the rates obtained from simulations. Starting conditions and results of the optimization are reported in Supplementary Table 4. We find an optimized $\boldsymbol{E_{pump}}$=0.95 protons/cycle, indicative of high efficiency of the kinetic model. The pumping fluxes and occupation sites profiles of the optimized solution are shown in Figure S13, and the parameter dependence are shown in Figure S7.

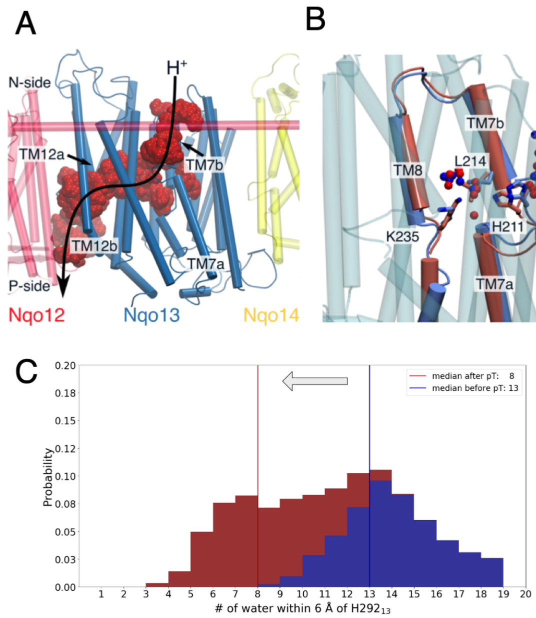

Figure S1. Hydration dynamics in Nqo13. (A) Proton channels are established at symmetry-related location around TM7a/b and TM12a/b in Nqo13, consistent with previous MD studies[19,20]. (B) MD snapshots of the $Lys235_{13}^{(+)}/Glu377_{13}^{(-)}$ state (in blue) before the pT reaction, and the $Lys235_{13}^{(0)}/Glu377_{13}^{(0)}$ state (in red) after the pT reaction. Water molecules, shown as spheres, are obtained from the overlap of snapshots of simulations S9 and S10 after 50 ns (Table S7). (C) The channel hydration depends on the protonation state of the middle $Lys235_{13}$. The cumulative histograms show the probability distribution of water occupancy before pT ($Lys235_{13}^{(+)}/Glu377_{13}^{(-)}$, in blue) and after pT ($Lys235_{13}^{(0)}/Glu377_{13}^{(0)}$, in red). The water occupancies were computed within 6 Å from $His292_{13}$.

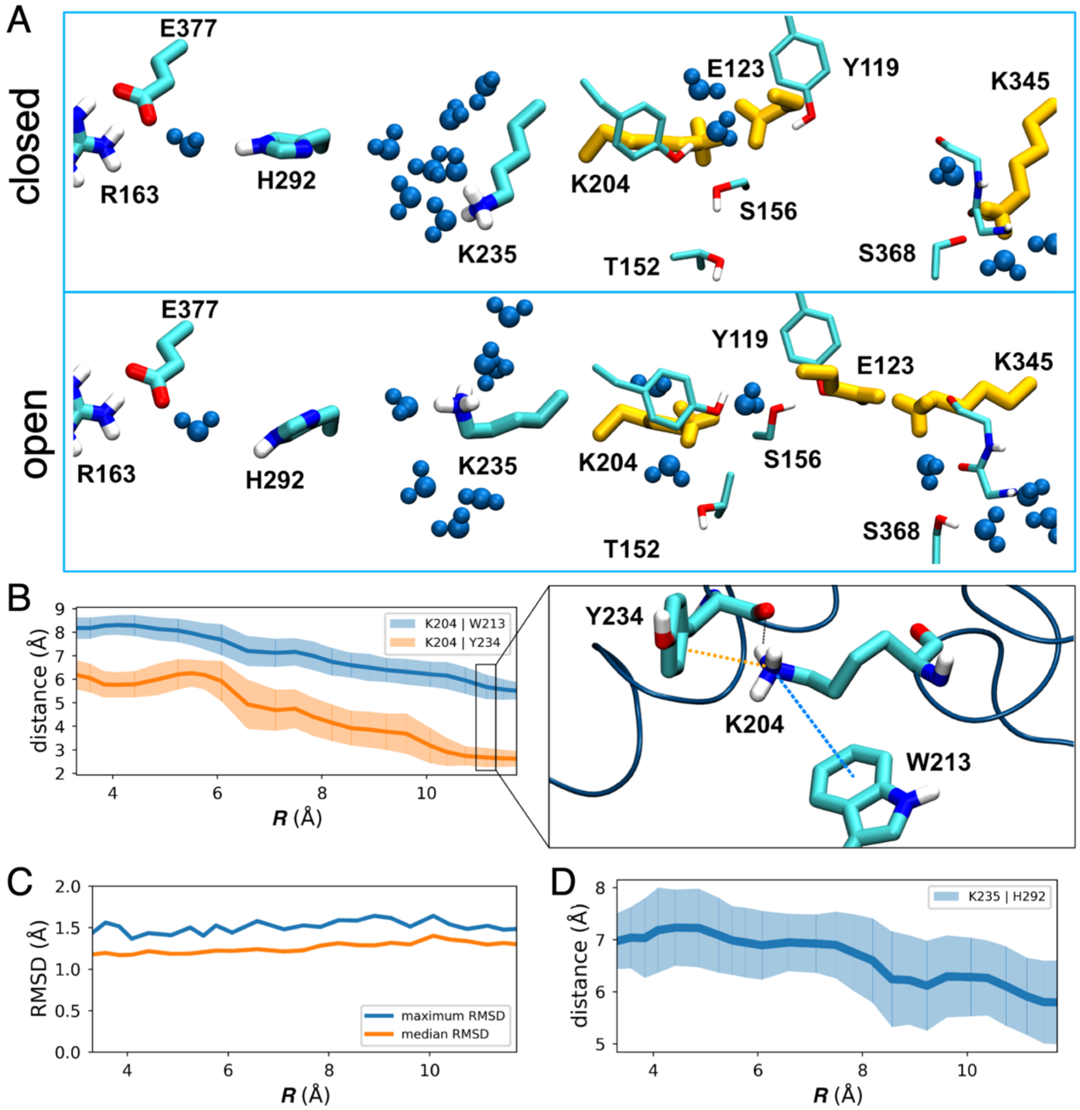

Figure S2. Conformational dynamics linked to ion-pair opening. (A) Hydrogen-bonding network in the hydrated state of Nqo13 with closed (*top*) and open (*bottom*) $Glu123_{13}$-$Lys204_{13}$ ion-pair conformations. Key residues are drawn in thick stick representation, stabilising residues in thinner stick representation, and water molecules forming hydrogen-bonding interactions are shown as blue spheres. The ion-pair is highlighted in yellow. (B) In the open ion-pair conformation (high *R* values), $Lys204_{13}$ is stabilised by hydrogen-bonds and π-cation interactions to $Tyr234_{13}$/$Trp213_{13}$. (C) The ion-pair opening does not induce large conformational changes, as indicated by a low RMSD as a function of the opening distance (*R*). (D) The ion-pair opening brings $Lys235_{13}$ closer to $His292_{13}$ by *ca*. 1 Å.

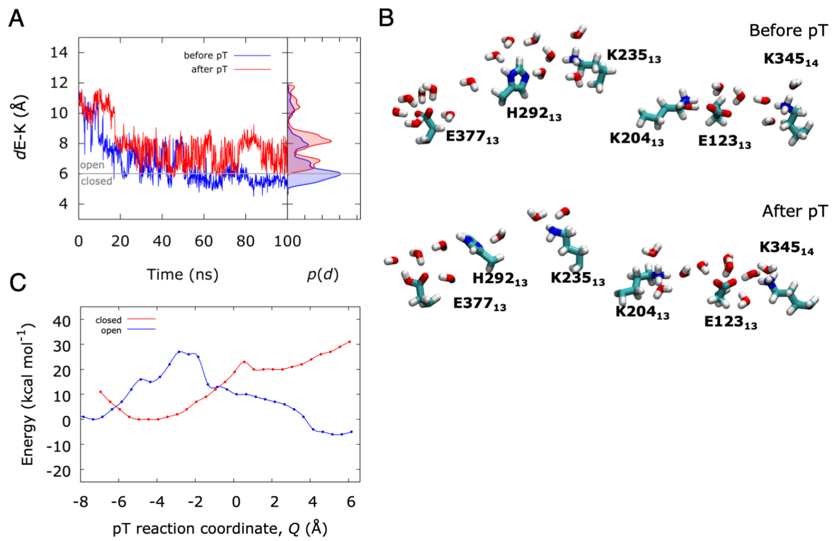

Figure S3. Connection between proton transfer dynamics, hydration states, and ion-pair conformation in Nqo13. (A) The Glu123$_{13}$/Lys204$_{13}$ ion-pair distance before (Lys235$_{13}^{(+)}$/Glu377$_{13}^{(-)}$, in red) and after (Lys235$_{13}^{(0)}$/Glu377$_{13}^{(0)}$, in blue) the pT reaction. Both simulations were initiated from the open Lys204$_{13}$-Glu123$_{13}$ ion-pair conformation in the fully hydrated state. Closed and open ion-pairs are indicative of distances < 5 Å and 6-8 Å, respectively. Deprotonation of Lys235$_{13}$ stabilises the open ion-pair conformation. (B) MD snapshots of the Lys235$_{13}^{(+)}$/Glu377$_{13}^{(-)}$ state, before pT reaction has taken place (*top*) and the Lys235$_{13}^{(0)}$-Glu377$_{13}^{(0)}$ state after the pT reaction *(bottom*), showing the Lys204$_{13}$-Glu123$_{13}$-Lys345$_{14}$ interface. (C) QM/MM potential energy profiles of the pT reaction coordinate *Q* (see main text Figure 3) at medium hydration levels with open- (in blue) and closed (in red) Lys204$_{13}$-Glu123$_{13}$ ion-pairs conformations. Partial dehydration of the proton channel drastically increases the pT barrier in both ion-pair conformations.

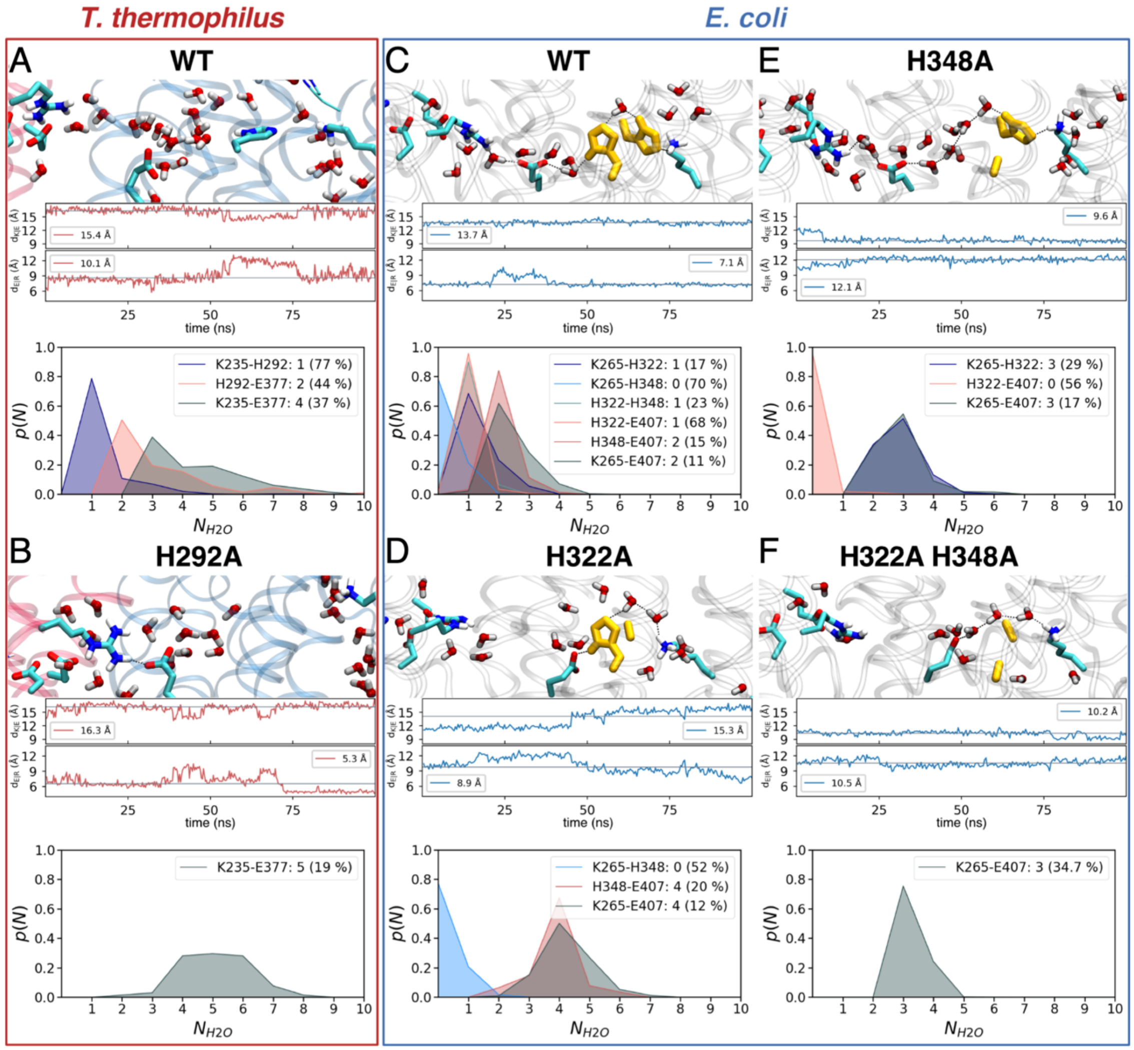

Figure S4. Proton wires in wild type and alanine substituted complex I from *T. thermophilus* and *E. coli*. Snapshot of MD simulation (*top*), Lys235$_{13}$/Glu377$_{13}$ ($d_{K\text{-}E}$) and Glu377$_{13}$/Arg163$_{12}$ ($d_{E\text{-}R}$) distances (*middle*), and water occupancy $p(N)$ between residues (*bottom*), with median water occupancy and connectivity of the wire given in the legend for (A) WT complex I from *T. thermophilus*; (B) the H292A variant of complex I from *T. thermophilus*; (C) WT complex I from *E. coli;* (D) the H322A variant of complex I from *E. coli;* (E) the H348A variant of complex I from *E. coli;* and (*F*) the H322A/ H348A variant of complex I from *E. coli.* Data was analysed from simulations 12 to 14 (see Table S7).

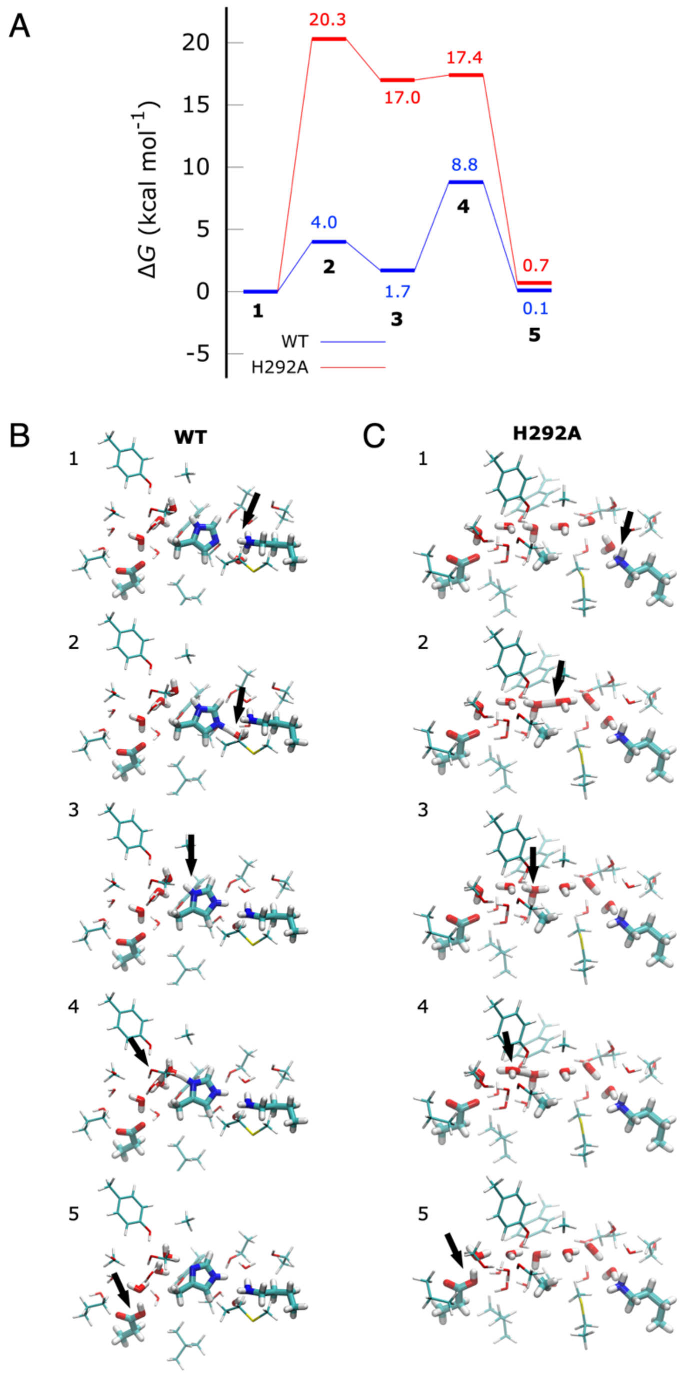

Figure S5. Proton transfer barriers in the WT and the H292A variant of complex I studied by DFT models. (A) Free energy profiles for pT reaction obtained at the B3LYP-D3/def2-TZVP/ε=4 level (in kcal mol^{-1}), with entropic and zero-point corrections estimated at B3LYP-D3/def2-SVP/ε=4 level. (B) Structure of intermediates along the pT reaction.

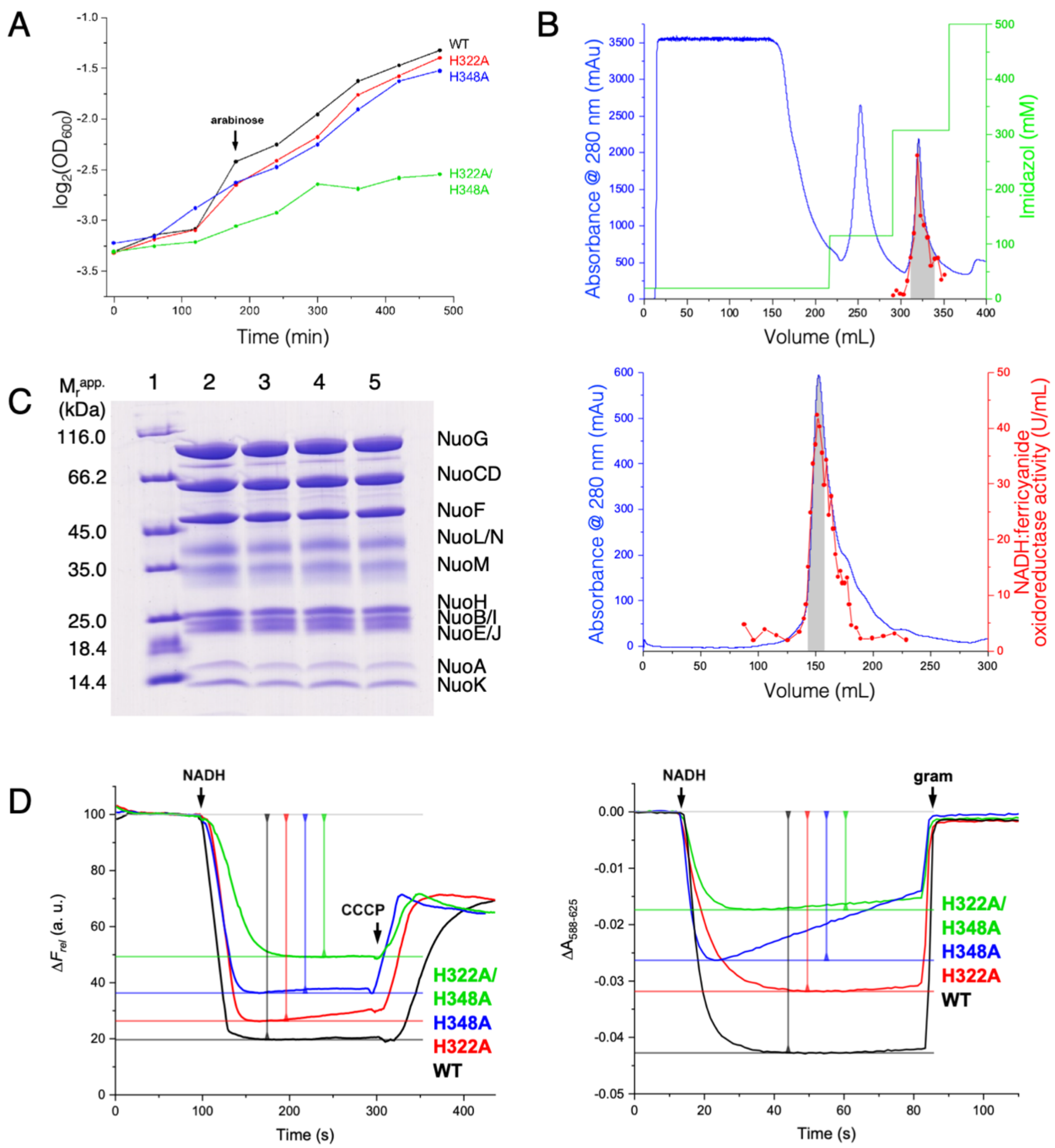

Figure S6. Isolation of WT and variants of complex I. (A) Growth curves of *E. coli* wild type and mutants in minimal medium with acetate as sole carbon source. The arrow indicates induction of gene expression after 3 h by an addition of arabinose. (B) Isolation of the H348A variant by affinity-chromatography on ProBond Ni^{2+}-IDA (*top*) and by size-exclusion chromatography on Superose 6 (*bottom*). Fractions indicated with a grey background were used in the next step. Preparations of the wild type and the variants showed virtually identical elution profiles. (C) SDS-PAGE of isolated complex I and the variants revealed the presence of all complex I subunits (lane 1: marker, 2: WT; 3: H322A; 4: H348A; 5: H322A/H348A). The faint band at around 80 kDa is a proteolytic digestion product of NuoG. See Table S9 for subunit naming in *E. coli* and *T. thermophilus* complex I. (D) Quantification of pumping activities (*left*: ACMA assay, *right*: oxonol-VI assay) was characterized from the maximum level of the optical changes without extrapolation to zero time point using the plateau levels as a base.

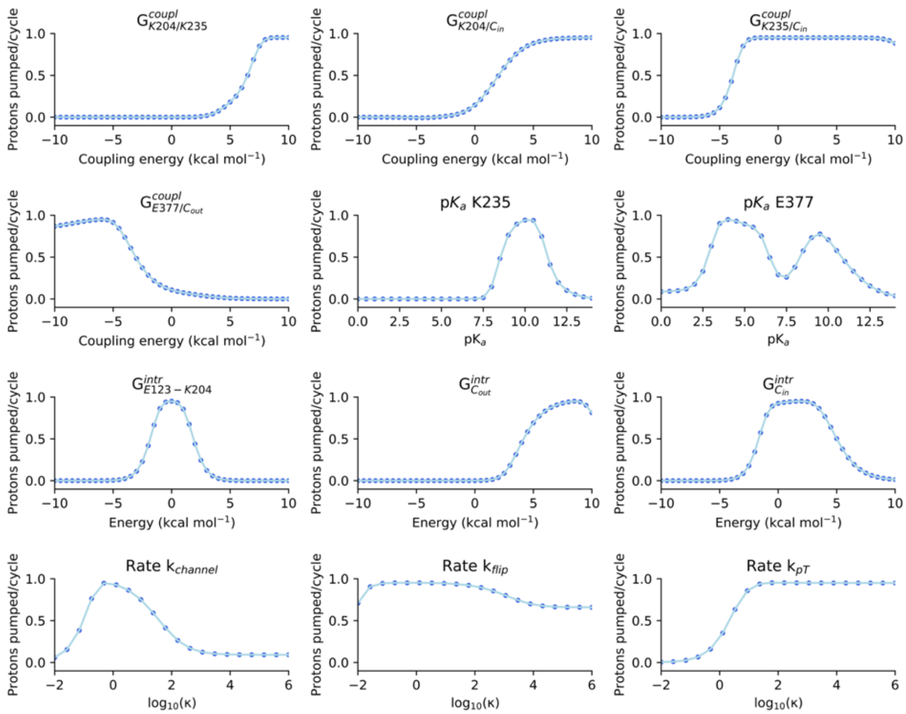

Figure S7. Dependence of proton pumping on different parameters in the kinetic model. The efficiency of the proton pumping is calculated varying single parameters values (x-axis) while maintaining the other parameters fixed. Each subplot refers to a single parameter scan (see Table S4 for nomenclature and complete list of optimized values).

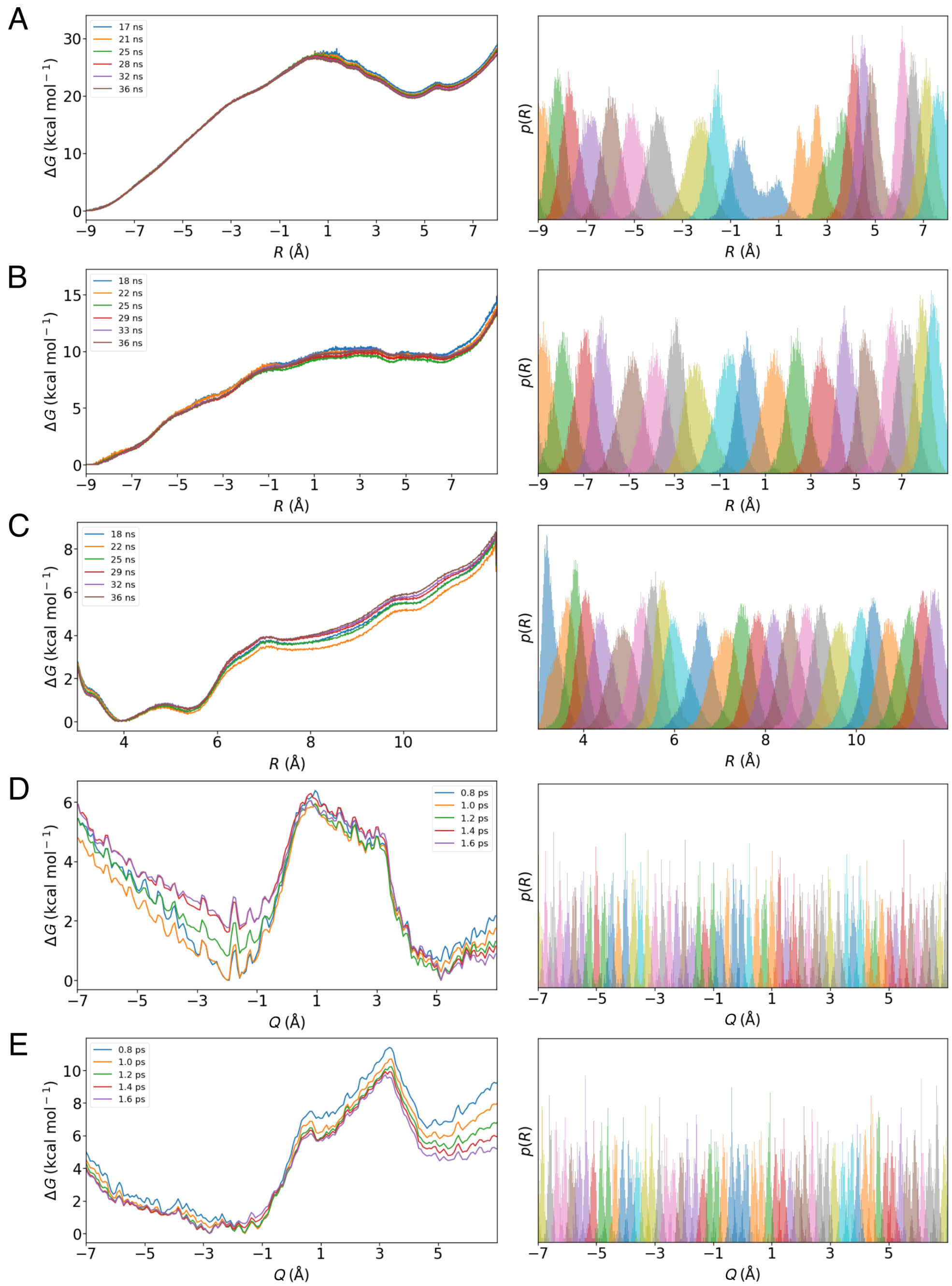

Figure S8. Convergence of classical and QM/MM-based free energy calculations. Time evolution (left) and window overlap (right) as a function of the reaction coordinates (ion-pair opening coordinate, $R = d$(E123$_{13}$-K204$_{13}$) – d(E123$_{13}$-K345$_{14}$), and proton transfer reaction coordinate, Q see Figure 3), showing the convergence of classical replica exchange umbrella sampling (REUS) and DFT-based QM/MM umbrella sampling (QM/MM/ US) simulations. Classical REUS simulation of (A) the dry state (simulation S4, Table S7), (B) at medium hydration levels (simulation S5, Table S7), and (C) in the fully hydrated state (simulations S8, Table S7). The ±1 ns variation in shown simulation timepoints minimises boundary errors in the DHAM analysis. Convergence of the QM/MM/US simulations of the pT reaction in Nqo13 with (D) open ion-pair (simulation 5, Table S8), and (*E*) closed ion-pair (simulation 6, Table S8).

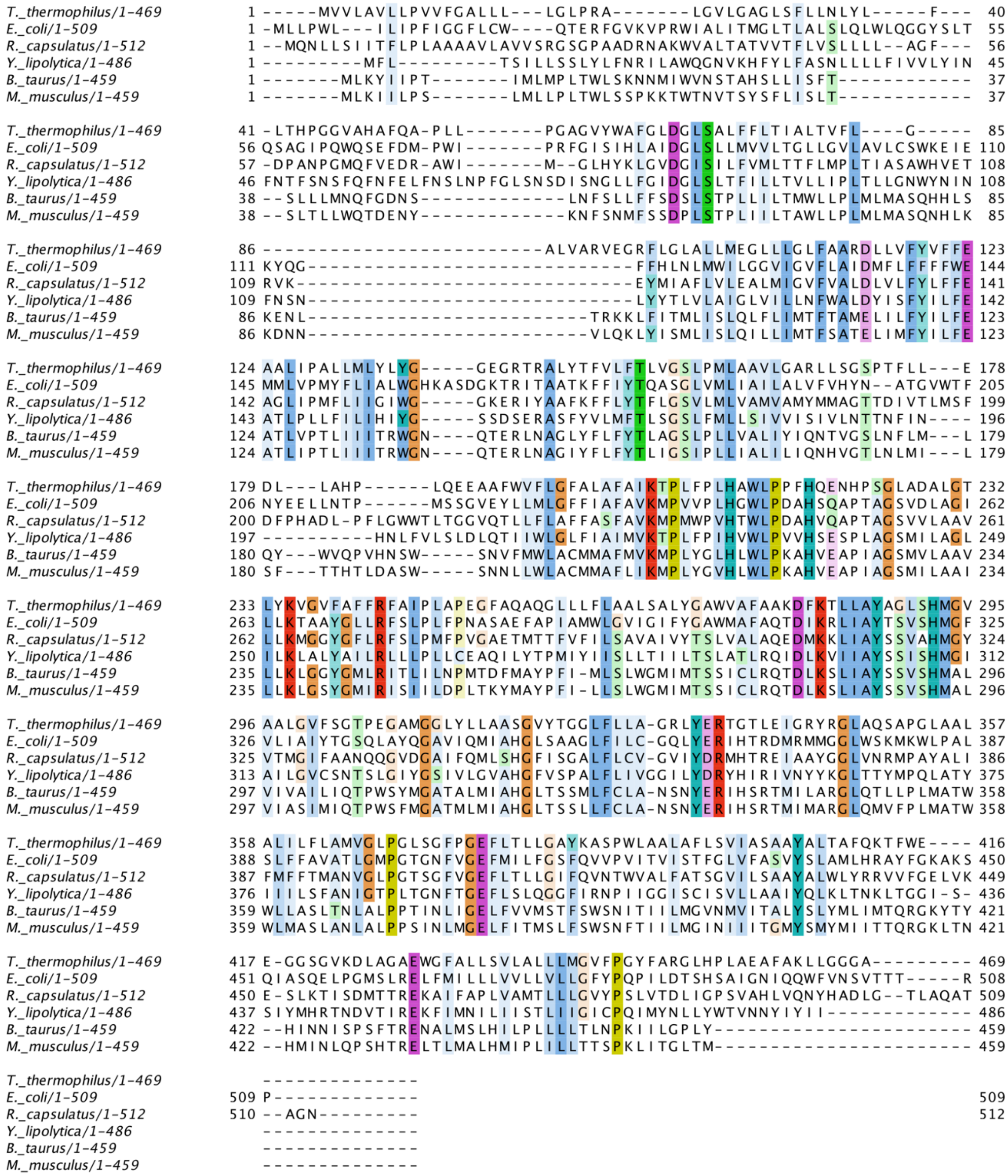

Figure S9. Multiple sequence alignments (MSA) of subunit Nqo13 from *Thermus thermophilus, Escherichia coli*, *Rhodobacter capsulatus*, *Yarrowia lipolytica*, *Bos taurus,* and *Mus musculus*. Key conserved charged and polar residues are marked. The MSA was performed using ClustalW.[21]

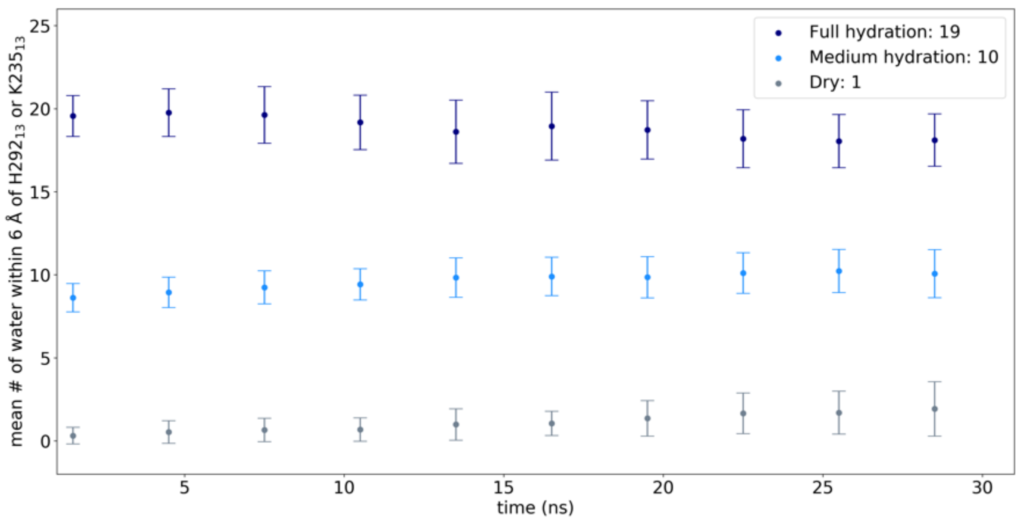

Figure S10. Hydration states of Nqo13 during the REUS free energy simulations. The average of the water occupancy averaged over the 20 independent windows used for each free energy calculations is shown. To correct for fast fluctuations, the hydration is time-averaged in bins of 3 ns each. The reported values are the number of water molecules within 6 Å of residues His292$_{13}$ and Lys235$_{13}$.

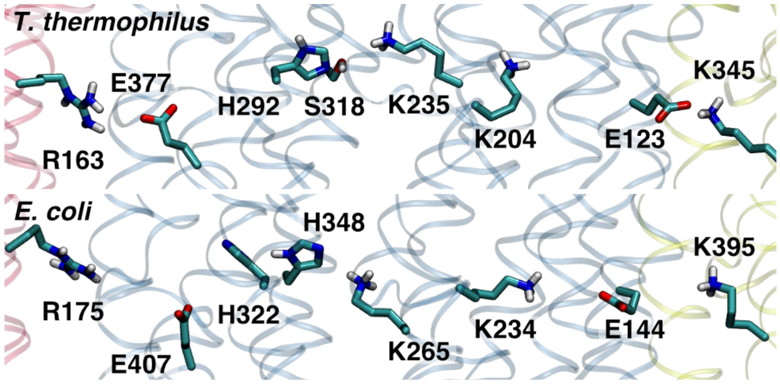

Figure S11. Key residues involved in the pT reaction in *T. thermophilus* (*top*) and *E. coli* (*bottom*) complex I isoforms.

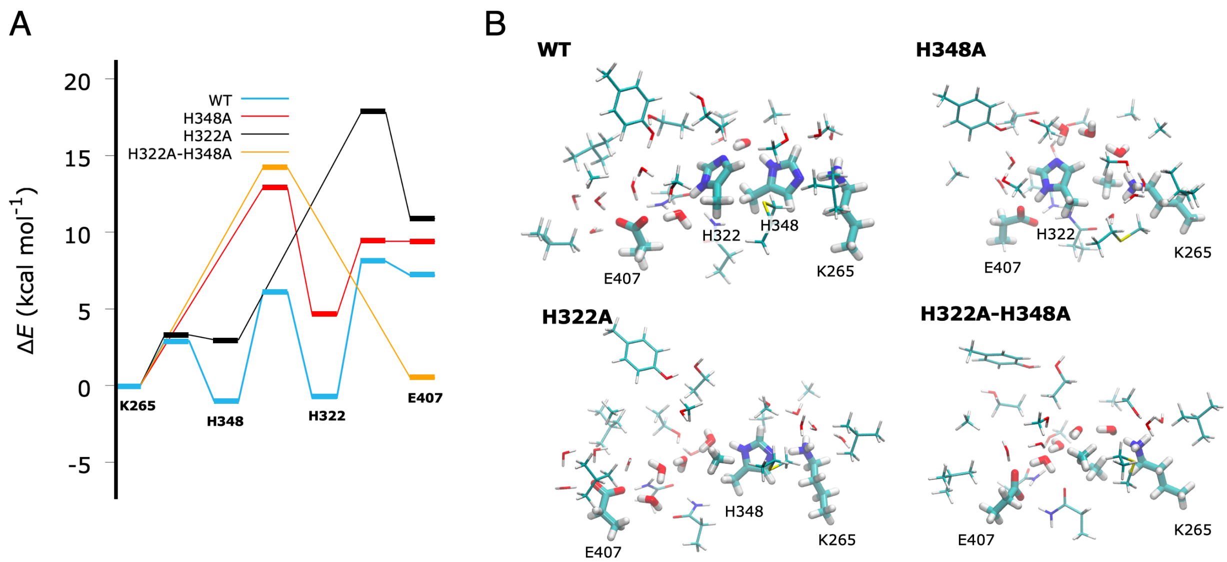

Figure S12. Proton transfer barriers in the *E. coli* complex I variants. (A) Energetics of the proton transfer reaction in the WT and H322A, H348A, and H322A/H348A variants, calculated at the B3LYP-D3/def2-TZVP/ε =4 level. (B) Snapshots of the DFT models of WT and H322A, H348A, and H322A/H348A variants.

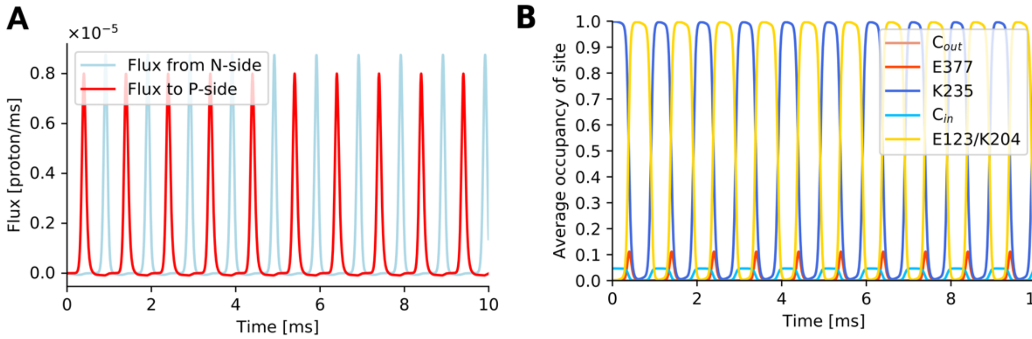

Figure S13. (A) Proton pumping profiles and (B) occupation of single sites of the optimized kinetic model during the simulation time (see also Table S4). E123/K204 is the biased state.

Table S1. Accessibility of the NADH binding site in reconstitutions determined from the ratio of outward-facing to inward-facing complexes by measuring the NADH/FeCN activity of liposomes before and after addition of DDM.

	accessible NADH binding site
	[%]
WT	60
H322A_M	55
H348A_M	55
H322A_M/H348A_M	61

Table S2. NADH/Q, NADH/FeCN, and proton pumping activity of WT and alanine variants of *E. coli* complex I.

	NADH:Q oxidoreductase activity		**NADH/ferricyanide oxidoreductase activity**	
	[U/mg]	[%]	[U/mg]	[%]
WT	36.0 ± 0.8	100 ± 2	89.6 ± 15.5	100 ± 17
$H322A_M$	21.9 ± 1.6	61 ± 7	83.9 ± 9.3	94 ± 11
$H348A_M$	29.4 ± 1.8	82 ± 6	94.9 ± 20.3	106 ± 21
$H322A_M/H348A_M$	17.2 ± 1.9	48 ± 11	111.1 ± 16.2	124 ± 15

	ACMA quenching	
	[%]	rel. to WT [%]
WT	80.5 ± 2.1	100 ± 3
$H322A_M$	74.0 ± 1.4	92 ± 2
$H348A_M$	63.5 ± 2.1	79 ± 3
$H322A_M/H348A_M$	50.5 ± 0.7	63 ± 1

	Oxonol quenching	
	$\Delta A_{588\text{-}625}$	rel. to WT [%]
WT	0.0429 ± 0.0027	100 ± 6
$H322A_M$	0.0319 ± 0.0005	74 ± 2
$H348A_M$	0.0263 ± 0.0004	61 ± 1
$H322A_M/H348A_M$	0.0172 ± 0.0012	40 ± 7

Table S3. Parameters employed in the kinetic model.

Parameter	Value
ΔG_{bias}^{max}	-400 mV
T	310 K
t_{period}	1 ms
n	10
pH	7

Table S4. Rate coefficients employed in the kinetic model.

	Starting value	Optimized value
k_{chan} (μs^{-1})	1	0.611
k_{flip} (μs^{-1})	100	0.130
k_{pT} (μs^{-1})	10000	6426.4
pK_a (K235)	9.00	10.2
pK_a (E377)	5.00	4.00
E (E123-K204) (kcal mol^{-1})	2	-0.19
E (C_{in}) (kcal mol^{-1})	2	2.01
E (C_{out}) (kcal mol^{-1})	4	8.47
E_{coupl} (C_{in}/K235) (kcal mol^{-1})	-3	-0.16
E_{coupl} (C_{out}/E377) (kcal mol^{-1})	-3	-5.99
E_{coupl} (K204-K235) (kcal mol^{-1})	8	8.69
E_{coupl} (K204-C_{in}) (kcal mol^{-1})	8	10.34

Table S5. Oligonucleotide sequences of primers used for site-directed mutagenesis. Changed codons are shown in bold, exchanged bases are in italics. Silent mutations were introduced to generate novel restriction sites to check the success of site-directed mutagenesis. Restriction sites are underlined. The 'Check_nuoM_*' oligonucleotides were used for DNA sequencing.

Oligonucleotide	**sequence (5'-3')**	**Novel restriction site**
nuoM_H322A_fwd	CGCCTACACCTCGGTTTCC**GCC**ATGGGCTTCGTGTCTGATTG	*NcoI*
nuoM_H322A_rev	CAATCAGCACGAAGCCCATGGCGGAAACCGAGGTGTAGGCG	
nuoM_H348A_fwd	CGGTAATCCAGATGATTGCG**GCC**GG*C*TTGTCGGCGGCGGGTCTG	*PdiI*
nuoM_H348A_rev	CAGACCCGCCGCCGACAAGCCGGCCGCAATCATCTGGATTACCG	
Check_nuoM_H322A	GCGCCAATTGCTATGTGGCTG	
Check_nuoM_H348A	CCGATATCAAACGTCTGATC	

Table S6. Buffers used for expression, purification, reconstitution, and biophysical experiments of *E. coli* complex I.

Buffer	Content
autoinduction medium	1% (w/v) peptone, 0.5% (w/v) yeast extract, 0.4% glycerol, 25 mM $Na_2HPO_4 \cdot 2\ H_2O$, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4 \cdot 7\ H_2O$, 0.2% (w/v) L-arabinose, 0.05% (w/v) glucose, 30 mg·L^{-1} Fe-NH_4-citrate, 0.5 mM l-cysteine, 50 mg·L^{-1} riboflavine, containing chloramphenicol (34 µg/mL)
buffer $A^*_{pH6.8}$	50 mM MES/NaOH, 50 mM NaCl, 5 mM $MgCl_2$, 10% (v/v) glycerol, pH 6.8
binding buffer	20 mM imidazole, 50 mM MES/NaOH, 50 mM NaCl, 5 mM $MgCl_2$, 10% (v/v) glycerol, 0.005% (w/v) LMNG, pH 6.8
elution buffer	500 mM imidazole, 50 mM MES/NaOH, 50 mM NaCl, 5 mM $MgCl_2$, 10% (v/v) glycerol, 0.005% (w/v) LMNG, pH 6.8
buffer A^*_{MNG}	50 mM MES/NaOH, 50 mM NaCl, 5 mM $MgCl_2$, 10% (v/v) glycerol, 0.005% (w/v) LMNG, pH 6.0
lipid buffer	5 mM MES/NaOH, 50 mM NaCl, pH 6.7
reconstitution buffer	20 mM HEPES, 200 mM KCl, 73 mM sucrose, 5 mM $MgSO_4$, 0.05% (w/v) l-α-phosphatidylcholine, 1.1% (w/v) *n*-octylglucoside, 0.6% (w/v) sodium deoxycholate, 0.6% (w/v) sodium cholate, pH 7.5
ACMA buffer	5 mM KH_2PO_4/KOH, 50 mM KCl, 1 mM $MgCl_2$, pH 7.5
oxonol buffer	200 mM MES/KOH, 1 mM $MgSO_4$, 0.1 µM monensin, 300 mM mannitol, pH 6.75

Table S7. Summary of classical molecular dynamics simulations. Biases in SMD and REUS simulations were introduced on the $Lys204_{13}/Glu123_{13}/Lys345_{14}$ distances in simulations 2 to 4 and on $Lys204_{13}/Glu123_{13}$ in simulations 6 and 7 (see Methods). MD data[9] from simulations 6 to 8 was remapped onto the same reaction coordinate as simulations 2 to 4 for direct comparison.

Simulation	Length (ns)	Initiated from	Description
S1	200	PDB: 4HEA, after equilibration	MD
S2	10	S1 after 100ns	SMD, dry state
S3	10	S1 after 200 ns	SMD, N-side connected
S4	20 x 36	20 frames of S2	REUS, dry state
S5	20 x 36	20 frames of S3	REUS, N-side connected
S6	600	PDB: 4HEA, after equilibration	MD
S7	16	S6 after 600 ns	SMD, full hydration
S8	25 x 36	25 frames of S7	REUS, full hydration
S9	100	S6 after 600 ns	MD, full hydration, unbiased ion-pair
S10	100	S6 after 600 ns	MD, full hydration, after pT from $Lys235_{13}$ to $Glu377_{13}$ ($Lys235_{13}^{0}/Glu377_{13}^{0}$ state)
S11	100	S6 after 600 ns	MD, Nqo13-H292A
S12	200	PDB: 3RKO, after equilibration	MD, NuoM-H322A-H348A
S13	110	S12 after 100 ns	MD, NuoM-H322A
S14	110	S12 after 100 ns	MD, NuoM-H348A
S15	110	S12 after 100 ns	MD
Total	4006		

Table S8. Summary of QM/MM MD and QM/MM free energy simulations sampled at the B3LYP-D3/def2-SVP level (see Methods).

Simulation	**Length (ps)**	**State**	**Description**
1	7	Open ion-pair, medium hydration	Unbiased QM/MM MD
2	7	Closed ion-pair, medium hydration	Unbiased QM/MM MD
3	7	Open ion-pair, full hydration	Unbiased QM/MM MD
4	7	Closed ion-pair, full hydration	Unbiased QM/MM MD
5	58 x 1.6	Open ion-pair, full hydration	QM/MM US
6	58 x 1.6	Closed ion-pair, full hydration	QM/MM US
7	7	Open ion-pair, H292A mutant	Unbiased QM/MM MD
8	7	Closed ion-pair, H292A mutant	Unbiased QM/MM MD
Total	227.6		

Table S9. Subunit naming in *T. thermophilus* and *E. coli* complex I.

Subunit in *E. coli*	**Subunit in *T. thermophilus***
NuoF	Nqo1
NuoE	Nqo2
NuoG	Nqo3
NuoCD	Nqo5/4
NuoB	Nqo6
NuoA	Nqo7
NuoH	Nqo8
NuoI	Nqo9
NuoJ	Nqo10
NuoK	Nqo11
NuoL	Nqo12
NuoM	Nqo13
NuoN	Nqo14

Table S10. Protonation states of key residues used in the classical MD simulations. Residue numbers and subunit names are given in both *T. thermophilus* (TT) and *E. coli* (EC) nomenclature to enable comparison of simulations on both organisms. Residues substituted with alanine are indicated with A (simulations S12-S14).

TT	EC	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
K345$_{14}$	K395$_{N}$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
E123$_{13}$	K144$_{M}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
K204$_{13}$	K234$_{M}$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
K235$_{13}$	K265$_{M}$	+	+	+	+	+	+	+	+	+	0	+	+	+	+	+
H292$_{13}$	H322$_{M}$	0	0	0	0	0	0	0	0	0	0	0	A	A	0	0
S318$_{13}$	H348$_{M}$	0	0	0	0	0	0	0	0	0	0	0	A	0	A	0
E377$_{13}$	E407$_{M}$	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-
R163$_{12}$	R175$_{L}$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table S11. Nomenclature of key residues in *T. thermophilus* and *E. coli* complex I.

T. thermophilus	***E. coli***
$K345_{14}$	$K395_{N}$
$E123_{13}$	$E144_{M}$
$K204_{13}$	$K234_{M}$
$K235_{13}$	$K265_{M}$
$H292_{13}$	$H322_{M}$
$S318_{13}$	$H348_{M}$
$E377_{13}$	$E407_{M}$
$R163_{12}$	$R175_{L}$

Table S12. Sequence and structural similarity[22] of antiporter-like subunits in *T. thermophilus* and *E. coli* complex I. [a] First / second similarity scores represent structural alignment based on the respective *E. coli* / *T. thermophilus* subunits.

T. thermophilus	***E. coli***	**% of sequence similarity**	**TM structural similarity score[a]**
Nqo12	NuoL	41.3	0.91/0.92
Nqo13	NuoM	27.5	0.86/0.94
Nqo14	NuoN	34.4	0.86/0.95

Movie S1 (MovieS1.mp4). Water-mediated proton transfer dynamics from $Lys235_{13}$ via $His292_{13}$ to $Glu377_{13}$, from QM/MM MD simulations.

SI References

1. Miller, J. H. A Short Course in Bacterial Genetics – A Laboratory Manual and Handbook for *Escherichia coli* and Related Bacteria. *J. Basic Microbiol.* **1992***, 33*, 278–278.
2. Burschel, S. *et al.*, Iron-sulfur cluster carrier proteins involved in the assembly of *Escherichia coli* NADH:ubiquinone oxidoreductase (complex I). *Mol. Microbiol.* **2019**, *111*, 31–45.
3. Gornall, A. G.; Bardawill, C. J.; David, N. M. Determination of serum proteins by means of the biuret reaction. *J. Biol. Chem.***1949**, *177*, 751–766.
4. Gill, S. C.; von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **1989**, *182*, 319–326.
5. Schägger, H.; von Jagow, G.Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **1987***, 166,* 368–379.
6. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numerische Mathematik*. **1959,** *1,* 269–271.
7. Hagberg, A. A.; Schult, D. A.; Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy2008)*, **2018**, 11–15.
8. McGibbon, R. T. et al., MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories, *Biophys. J.* **2015**, *109*, 1528 – 1532.
9. 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* **1988**, **37**, 785–789.
10. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, 98, 5648–5652.
11. Schäfer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577.
12. 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.
13. 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.* **1993**, *2*, 799–805.
14. Plessow, P. Reaction path optimization without NEB springs or interpolation algorithms. *J. Chem. Theory Comput.* **2013**, *9*, 1305–1310.
15. E, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B - Condens. Matter Mater. Phys.* **2002**, *66*, 523011–523014.
16. 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.
17. Kim, Y. C.; Wikström, M.; Hummer, G. Kinetic models of redox-coupled proton pumping. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2169–2174.
18. 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.
19. 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.
20. Di Luca, A.; Gamiz-Hernandez, A. P.; Kaila, V. R. I. Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl. Acad. Sci.* **2017**, *114*, E6314–E6321.
21. F. Sievers, *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal. *Omega. Mol. Syst. Biol.* **2011**, *7*, 1-6.
22. Zhang, Y; Skolnick, J. TM-align: A protein structure alignment algorithm based on TM-score. *Nucleic Acids Research,* **2005**, *33,* 2302-2309.