Role of protonation states in stability of molecular dynamics simulations of high-resolution membrane protein structures

Jonathan Lasham[1], Amina Djurabekova[1], Volker Zickermann[2,3], Janet Vonck[4], Vivek Sharma[1,5]*

[1]Department of Physics, University of Helsinki, 00014 Helsinki, Finland.

[2]Institute of Biochemistry II, University Hospital, Goethe University, 60438 Frankfurt am Main, Germany.

[3]Centre for Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe University, 60438 Frankfurt am Main, Germany.

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

[5]HiLIFE Institute of Biotechnology, University of Helsinki, 00014 Helsinki, Finland.

* correspondence: vivek.sharma@helsinki.fi

Abstract

Classical molecular dynamics (MD) simulations provide unmatched spatial and time resolution of protein structure and function. However, accuracy of MD simulations often depends on the quality of force field parameters and the time scale of sampling. Another limitation of conventional MD simulations is that the protonation states of titratable amino acid residues remain fixed during simulations, even though protonation state changes coupled to conformational dynamics are central to protein function. Due to the uncertainty in selecting protonation states, classical MD simulations are sometimes performed with all amino acids modeled in their standard charged states at pH 7. Here we performed and analyzed classical MD simulations on high-resolution cryo-EM structures of two membrane proteins that transfer protons by catalyzing protonation/deprotonation reactions. In simulations performed with amino acids modeled in their standard protonation state the structure diverges far from its starting conformation. In comparison, MD simulations performed with pre-determined protonation states of amino acid residues reproduce the structural conformation, protein hydration, and protein-water and protein-protein interactions of the structure much better. The results suggest it is crucial to perform basic protonation state calculations, especially on structures where protonation changes play an important functional role, prior to launching any MD simulations. Furthermore, the combined approach of protonation state prediction and MD simulations can provide valuable information on the charge states of amino acids in the cryo-EM sample. Even though accurate prediction of protonation states currently remains a challenge, we introduce an approach of combining pKa prediction with cryo-EM density map analysis that helps in improving not only the protonation state predictions, but also the atomic modeling of density data.

Keywords; molecular dynamics, protonation states, pKa calculations, cryo-electron microscopy, electrostatics.

Introduction

Within the past decade, the resolution of single-particle cryo-EM structures has improved dramatically, largely due to the improvements in direct electron detectors and processing software [1]. The resolution of single-particle cryo-EM structures is now comparable to x-ray crystallography and NMR structures, and the so-called "resolution revolution" has made it possible to determine structures of many previously inaccessible complexes, particularly for membrane proteins [2]. Many of these structures have a resolution below 2.5 Å, allowing accurate modeling of protein conformations, including ordered water molecules, which has significant implications in drug design [3] and is also central to our understanding of the molecular mechanism of enzymes that catalyze proton transfer reactions [4].

Proton transfer can take place across a chain of water molecules via a Grotthuss-type mechanism [5]. However, proton transfer routes through proteins do not consist only of water molecules but are often made up of hydrogen-bonded networks of polar and charged amino acid residues. Besides titratable residues such as glutamic acid and lysine, which can donate or accept protons, other residues that are part of these networks include asparagine and glutamine, and serine, threonine, and tyrosine with hydroxyl groups in their sidechains. These polar residues can not only assist in proton transfer reactions by stabilizing charged intermediates but may also undergo protonation/deprotonation reactions in proteinaceous environments [4, 6-8]. Moreover, titratable acidic and basic residues are well-known to act as proton transfer elements and proton loading sites in several enzymes [4, 9, 10]. However, the protonation states in protein structures are usually not explicitly modelled. Such information can be obtained with neutron diffraction techniques, but only for relatively small proteins [11].

The two main structure determination methods for protein complexes, x-ray crystallography and cryo-EM, produce superficially similar results. However, they differ fundamentally in the way atomic structures are imaged. X-rays are deflected by electrons and thus produce electron density maps, but electrons are scattered by Coulomb potential, and they are thus sensitive to charges. While positively charged ions just add extra density to an already positive signal, negative charges, in particular O^-, can give rise to negative electron scattering amplitudes [12-14]. As a result, negatively charged sidechains of glutamate and aspartate residues are not visible in cryo-EM maps. Although the weak density of these sidechains has sometimes been interpreted as a result of the radiation sensitivity of the carboxyl group [15, 16], the former mechanism appears to have a larger contribution [14, 17-19]. The absence of sidechain density hampers accurate model building. On the other hand, it makes it possible to determine the charge state of acidic residues.

To complement cryo-EM structures, molecular dynamics (MD) simulations are often performed [20-30]. These help in studying the dynamics of the protein and the solvent, as well as the binding and unbinding of lipids, ligands, and ions. The approach of combining structural data with MD simulations is a powerful technique to understand protein structure and function. The abundance of new structures gives plenty of opportunity to perform these simulations routinely at various levels of computational approximations. However, one limitation of conventional molecular dynamics is that covalent bonds are fixed throughout the simulation, therefore the charge states of titratable residues (often selected as the standard charge state that is Asp/Glu deprotonated, Lys/Arg protonated and His neutral) remain unchanged. This results in a biased scenario, and biologically relevant conformational

96   states with alternative protonation states are not populated. Moreover, situations where
97   changes in protonation states can occur as a function of conformational dynamics (e.g.
98   membrane proteins catalyzing proton transfer) are also not captured. However, there are
99   various methods that can allow charge states of amino acids to change during MD
100   simulations, such as hybrid QM/MM MD, and constant pH MD, but these are generally
101   computationally costly [31], although there have been recent improvements in the enhanced
102   scalability of constant pH MD [32].

104   For many proteins, performing MD simulations by systematic altering of charge states of
105   individual titratable residues is unrealistic, due to the large number of such residues. Of
106   course, one can alter protonation states of selected conserved residues to study specific
107   questions, and this has indeed been performed to obtain valuable functional insights [33-37].
108   However, due to the long-range nature of electrostatic interactions the protonation state of
109   one titratable residue affects another's by ca. 8 kcal/mol (with a separation of 10 Å at $\varepsilon=4$)
110   and such aspects are often ignored when performing simulations.

112   Alternatively, one can calculate the pKa of all titratable residues in a protein for a given
113   conformational state (obtained from a cryo-EM experiment, for example) and perform MD
114   simulations in that fixed protonation state. There are several methods to perform pKa
115   calculations, many of which are based on continuum electrostatics approaches [38]. However,
116   these require a significant amount of preprocessing and can display large variation in pKa
117   values due to subtle conformational changes. Alternatively, empirical methods, such as
118   Propka can be employed relatively easily, to give fast and sufficiently accurate predictions of
119   the pKa of all ionizable groups present in a protein, even if hundreds in number [39]. Methods
120   like Propka [40], often give reasonable estimates for the pKa of buried titratable residues, in
121   particular if the sites are far from any redox active cofactors. Due to their rapid pKa
122   estimations, they can be used on thousands of simulation snapshots to obtain profiles of pKa
123   change as a function of conformational dynamics [35, 41]. Similarly, Monte Carlo-based pKa
124   prediction methods have also been used to predict charged states of systems, either using
125   standalone PDB files or simulation snapshots [42].

127   In this study, we present MD simulations on two high-resolution cryo-EM structures of
128   membrane proteins. The two structures are the respiratory complex I from *Yarrowia lipolytica*
129   (PDB 7O71, EMD-12742) [43] and the multiple resistance and pH adaptation (Mrp)
130   cation/proton antiporter from *Bacillus pseudofirmus* (PDB 7QRU, EMD-14124) [44]. Both
131   proteins facilitate proton transfer reactions that involve amino acid residues and water
132   molecules, many of which have been resolved in the two structures. The role of protein
133   hydration, water dynamics as well as change in protonation state is central to the structure
134   and function of these proteins. By performing long time scale MD simulations in multiple
135   charge states of these proteins and extensively analyzing protein and solvent dynamics, we
136   show that both proteins deviate from the original cryo-EM conformation when simulated in
137   the standard protonation state of titratable amino acid residues, while MD simulations in pre-
138   defined protonation states stabilize the protein conformation much better. We propose that
139   the approach of protonation state prediction combined with MD simulations can give insights
140   into the charge state of residues in a cryo-EM structure. Furthermore, we find that the
141   prediction of protonation states of acidic residues agrees well with the charge state

142 assignment based on cryo-EM density maps, and that outliers can be identified with the
143 approach discussed here, leading to an improved modeling of cryo-EM density data.

144 **Results**

145 Two high-resolution membrane protein structures were chosen for this investigation. First,
146 the 2.1 Å resolution structure of complex I from *Y. lipolytica*, which is over 1 MDa in size, and
147 has more than 1600 structural water molecules resolved [43]. Around 100 of these are in the
148 potential proton transfer pathways present in its membrane-bound subunits (Fig. 1). Second,
149 a structure from the evolutionarily related Na^+/H^+ Mrp antiporter [44] at a similar resolution of
150 2.2 Å. This protein is much smaller than complex I (ca. 213 kDa) and consists of membrane-
151 bound subunits only, without redox active cofactors. There are 360 water molecules resolved
152 in the structure of the Mrp antiporter, around 70 of them in the potential proton and sodium
153 transfer pathways (Fig. 1). The structure of complex I contains more than 1300 titratable
154 residues, while the Mrp structure has around 200. We performed MD simulations on both
155 complexes with all titratable sites either modeled in their standard states (S state) or based
156 on pKa calculations at pH 7 (P7 state). We find that in the P7 state, complex I has 68
157 neutralized titratable amino acids, out of which 25 are in the membrane core of the enzyme
158 (Fig. 1), while in the Mrp antiporter 13 residues are neutralized in the P7 state. Accordingly,
159 the charge reduces from +90/-94 to +83/-76 in the complex I membrane arm, and from +116/-
160 118 to +110/-112 in the Mrp upon neutralization of titratable sites as part of pKa calculations
161 (see Tables S1 and S2 for lists of residues neutralized in complex I and Mrp antiporter,
162 respectively).

163
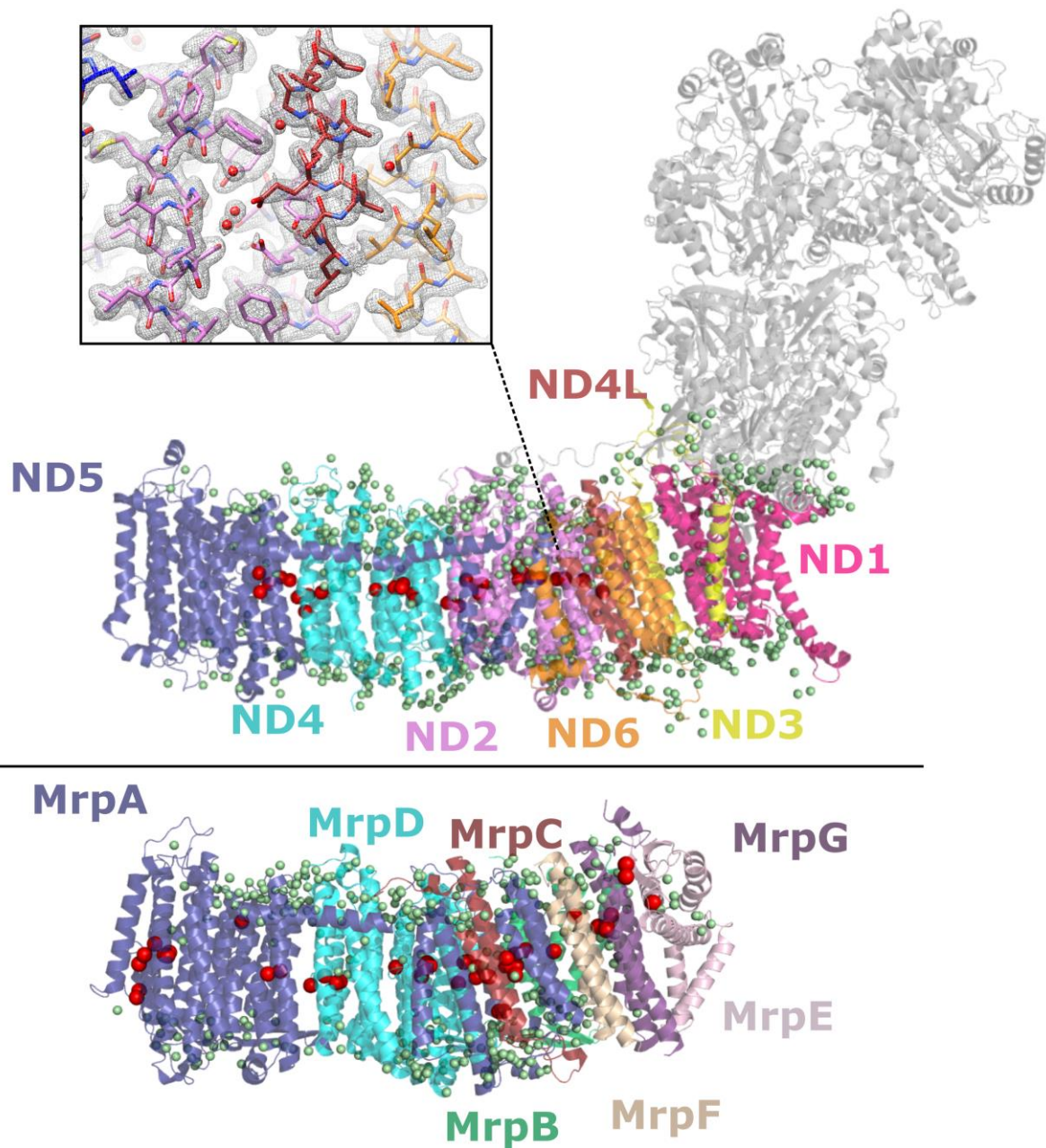
Fig. 1 – Structure of the membrane domain of respiratory complex I (upper) and Mrp antiporter (lower). The protein is shown in ribbon representation and colored by subunit. Only core membrane-bound subunits are shown in color for complex I, with the hydrophilic core subunits shown as gray. Structurally resolved water molecules are shown as small green spheres, with those in the functionally relevant hydrophilic axes of complex I and Mrp antiporter shown as larger red spheres. Proton translocation is suggested to take place across this axis, which is a vital part of the mechanism for both proteins. The inset in the upper panel shows a region of the complex I membrane arm within the high-resolution cryo-EM density (grey mesh), including density for water molecules.

Global mobility of proteins in P7 and S states

First, the overall global mobility of both systems was analyzed in the S and P7 states. Fig. 2 shows the time series as well as the distribution of the RMSD (root mean square deviation) of the backbone atoms for the core membrane-bound subunits of complex I and Mrp antiporter in the two simulation states. The P7 state systems with lower charge have overall lower RMSD for both protein complexes, which means that the backbone atoms stay closer to the starting conformation throughout the simulation. This notion is also true if a similar RMSD analysis is performed on the Cα backbone atoms and all protein atoms excluding hydrogens, even with the inclusion of the hydrophilic domain of complex I (Fig. S1).
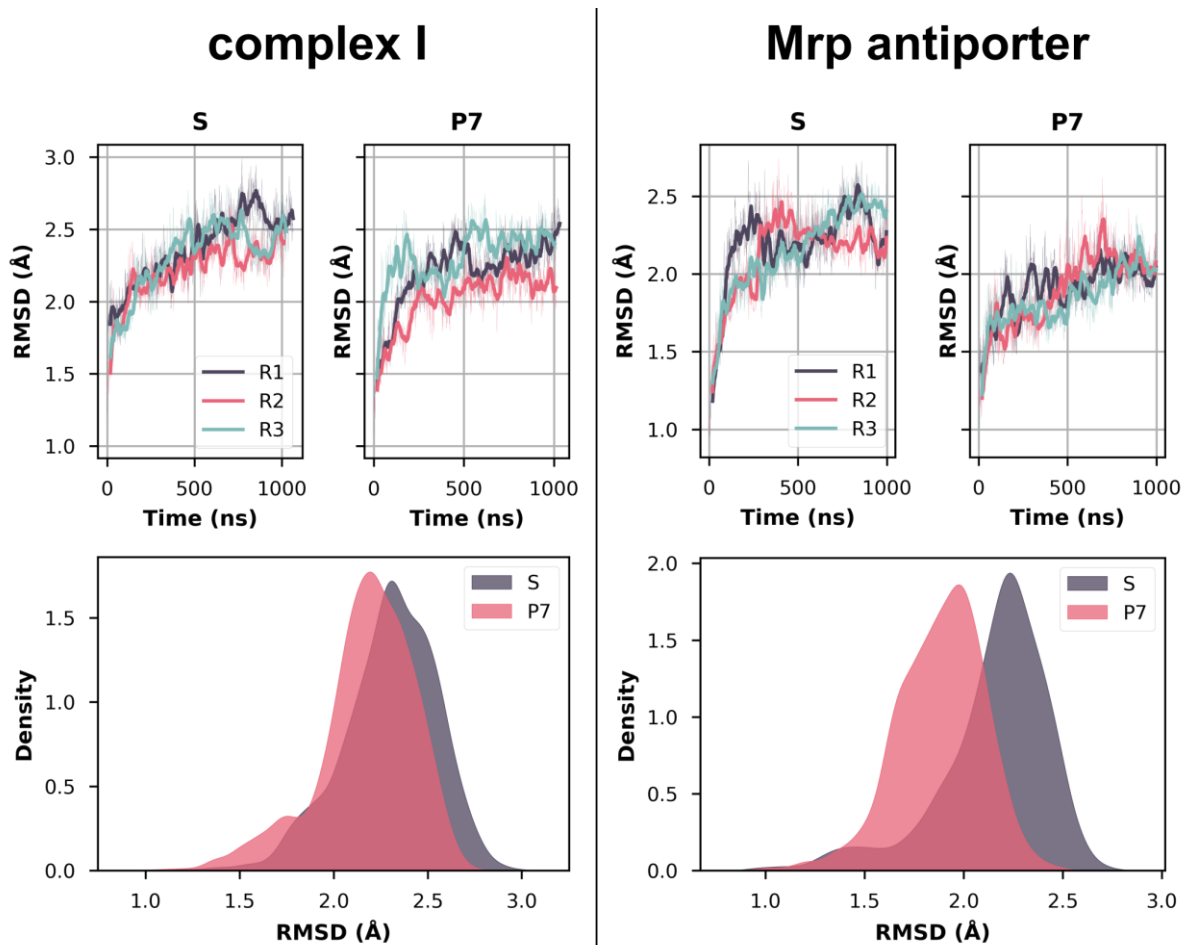
**Fig. 2 –** RMSD (root mean square deviation) of the backbone atoms over time for both complex I (core membrane subunits, left panel) and Mrp antiporter (right panel) in S and P7 states. The upper plots show the RMSD as a time series, with each colored trace representing a different simulation replica. The thick lines show the moving average of 20 ns, while the thinner lines show the RMSD for every 1 ns. The bottom panels show the distribution of RMSD values in the S and P7 states using a kernel density estimate (KDE) function with combined data of all three replicas.

Similarly, the RMSF (root mean square fluctuation), which measures the average amount an atom moves during the entire simulation, is also consistently found to be lower in the P7 states than in the S states. The boxplots in Fig. 3 show the RMSF of Cα atoms in each individual subunit of the Mrp antiporter. For all antiporter chains, lower RMSF values are observed in

202 the P7 state, indicating that there is an overall stability in the system as charges of titratable
203 sites are neutralized (based on pKa estimates). A similar trend was also observed in complex
204 I (Fig. S2). The spatial distribution of the RMSF data is shown in the lower panel of Fig. 3. MrpA
205 subunit (Figs. 1 and 3) has the largest number of titratable residues that undergo protonation
206 state changes to become neutralized and this has a clear impact in reducing the RMSF of the
207 subunit. Interestingly, the stability of the protein also increases in areas somewhat far from
208 the residues that change protonation states (magenta spheres), highlighting that the charge
209 change of just a few titratable residues can have conformational changes imparted both at
210 the local and the global level.
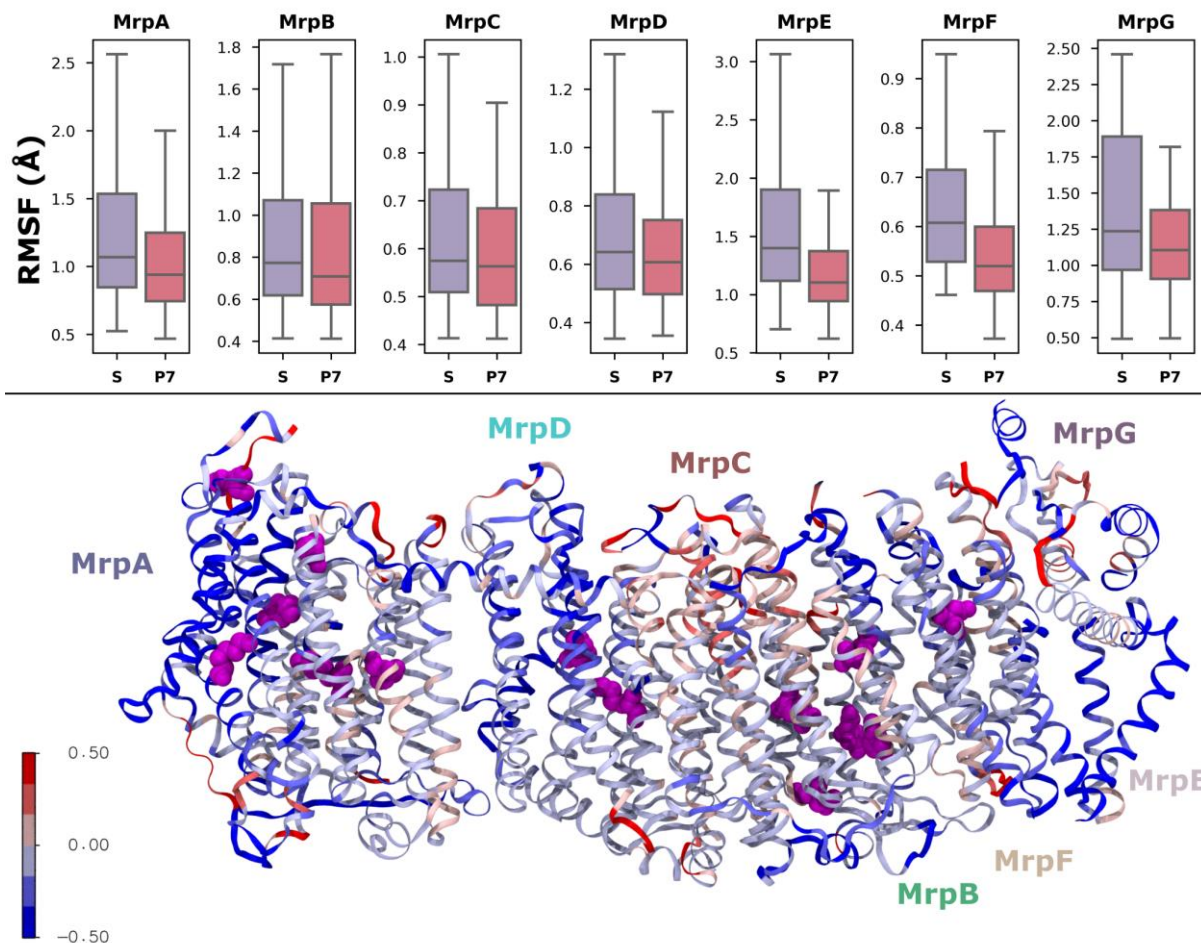
211
212 **Fig. 3 –** (Top) RMSF (root mean square fluctuation) of Cα atoms in Mrp antiporter subunits,
213 calculated for all simulation data in S and P7 states. The box plots represent the distribution
214 of RMSF values for each subunit, in both states. The shaded box represents the interquartile
215 range, with the middle line showing the median. The upper and lower lines are the maximum
216 and minimum values respectively. (Bottom) The panel shows the protein structure colored
217 according to the extent of conformational change in the P7 state compared to the S state.
218 Red represents Cα RMSF increased in the P7 state and blue represents a decrease. The
219 magenta spheres show the positions of residues that undergo a protonation state change in
220 P7 simulations.
221
222 Overall, the key observation is that MD simulations performed in predefined protonation
223 states retain a structure closer to the cryo-EM conformation for both protein complexes
224 studied here. This in turn means that the simulated charge state is closer to the charge state

225 of the protein during cryo-EM sample preparation. To understand why P7 state simulations
226 show overall stability relative to the S state, we performed additional analyses.
227

Water-protein interactions and protein hydration in P7 and S states
229 In the high-resolution structures of complex I and Mrp antiporter, the positions of several
230 water molecules are resolved. We therefore next analyzed how water-protein interactions
231 and protein hydration are affected in the S and P7 simulation states. The contacts between
232 all water molecules and protein were clearly reduced in the P7 states compared with the S
233 states (Fig. 3, upper panels). The lower charge in P7 states (+83/-76 compared to +90/-94 in
234 S state of complex I and +110/-112 compared to +116/-118 in S state of Mrp antiporter)
235 prevents extensive hydration of the protein resulting in lower number of contacts between
236 water oxygens and protein. This is also reflected in the clear reduction of water contacts that
237 take place with the residues that change protonation state (Fig 3, middle panel).
238
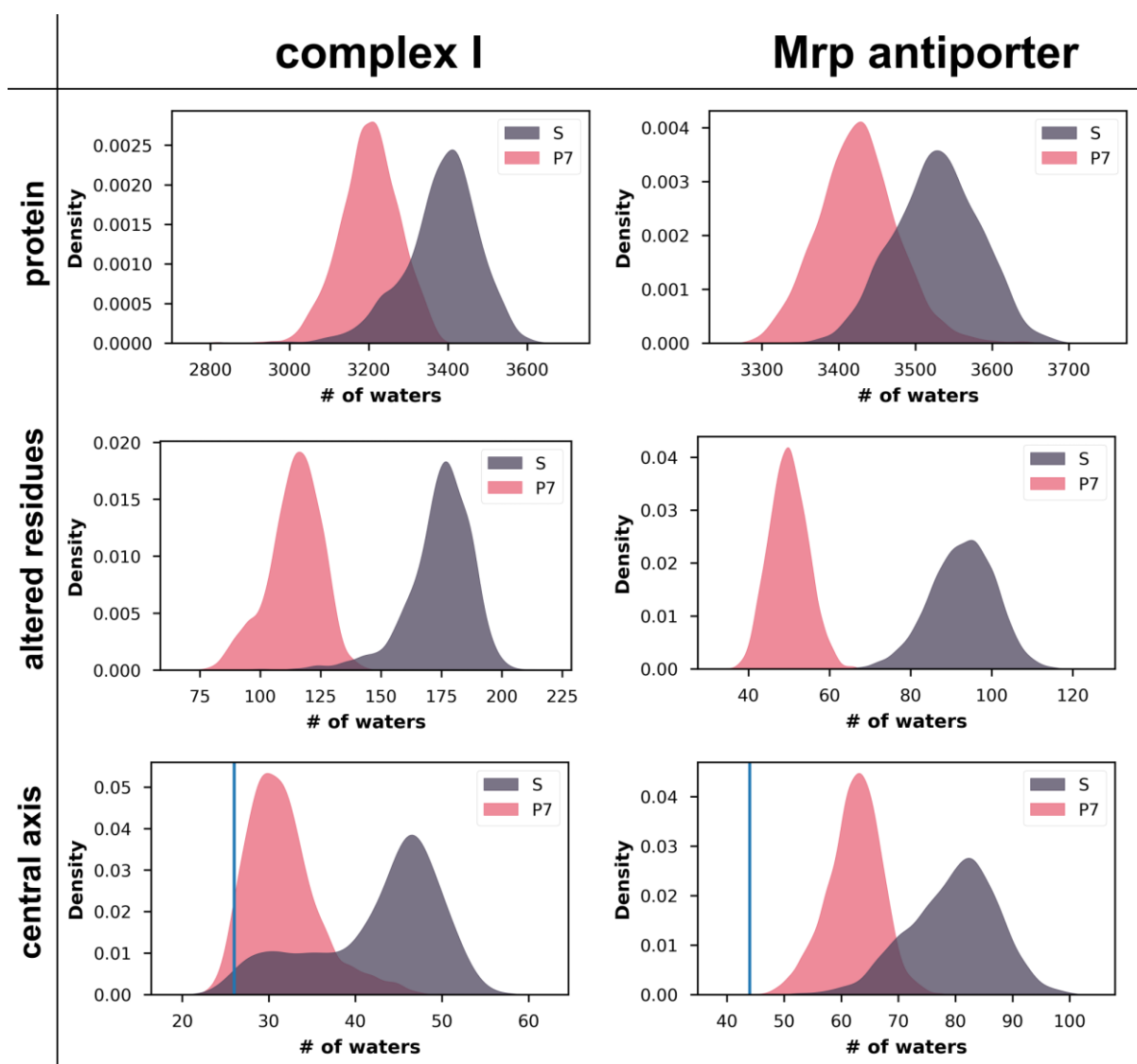
240 **Fig. 3 –** KDE plots for the number of water oxygen atoms in contact within 4 Å of protein
241 residues, with the left column showing data for complex I (core membrane subunits) and the
242 right for Mrp antiporter. Top panels show the number of waters in contact with all protein
243 atoms, middle panels show for protein residues that change protonation state in P7 state
244 simulations, and lower panels show the number of waters in contact with residues in the

245 functionally relevant horizontal axis (Fig. 1). Blue vertical lines in the lower panels indicate the
246 number of waters resolved in the cryo-EM structures that are in contact with the residues in
247 the central hydrophilic axis.
248
249 In both proteins, a hydrophilic axis runs through the central part of the membrane domain,
250 which is thought to be the structural basis for proton transfer reactions [43-48]. We next
251 evaluated the time evolution of internal water molecules in this hydrophilic axis for both
252 states in both proteins. We observed that the number of internal water molecules in the P7
253 state remains lower and closer to the water content observed in the structures, whereas in S
254 state runs, extensive hydration of the proteins is observed (Fig. 3, lower panel and Fig. S3).
255
256 Next, we analyzed the hydrogen bonding between the water molecules and the protein. The
257 time evolution of the hydrogen bonds between protein and structurally resolved water
258 molecules showed a more rapid decline in S states compared to the P7 states, with the latter
259 also displaying a slightly higher number of such interactions in long time scales (Fig. 4). This
260 indicates that although water exchange does take place in the P7 state, the exchange is less
261 rapid and a greater number of hydrogen bonds that existed in the initial structural state are
262 maintained for long time scales. The stable water-protein hydrogen bonds combined with
263 lower hydration in the P7 state also suggests a lower water exchange rate in these
264 simulations. The details of which protein-structural water hydrogen bonds are stabilized in
265 the Mrp antiporter is shown in Table S3, and notably the stabilization of these hydrogen bonds
266 is associated with a decrease in RMSF.
267
268 In Mrp antiporter, in the S state most structural waters loose hydrogen bonds with the protein
269 residues in the μs time scale, whereas in P7 state the loss is relatively slower and less
270 extensive in the given time scales (Fig. 4). Interestingly, the loss of hydrogen bonds is much
271 more drastic in complex I simulations, especially in the shorter time scales. To further probe
272 into the kinetics of loss of hydrogen bonding between the structural waters and protein, we
273 fitted the profiles (Fig. 4, upper panels) to exponential function (see methods) and obtained
274 the half-life of protein-water hydrogen bonds ($t_{1/2}$). The results show that on average in
275 complex I P7 state hydrogen bonds survive longer (average $t_{1/2}$ ~162 ns) compared to S state
276 (average $t_{1/2}$ ~113 ns). For relatively shorter time scales (ca. 500 ns), a more subtle but similar
277 effect was also observed for Mrp antiporter simulations (average $t_{1/2}$ ~135 ns in S state
278 compared to ~145 ns in P7 state).
279
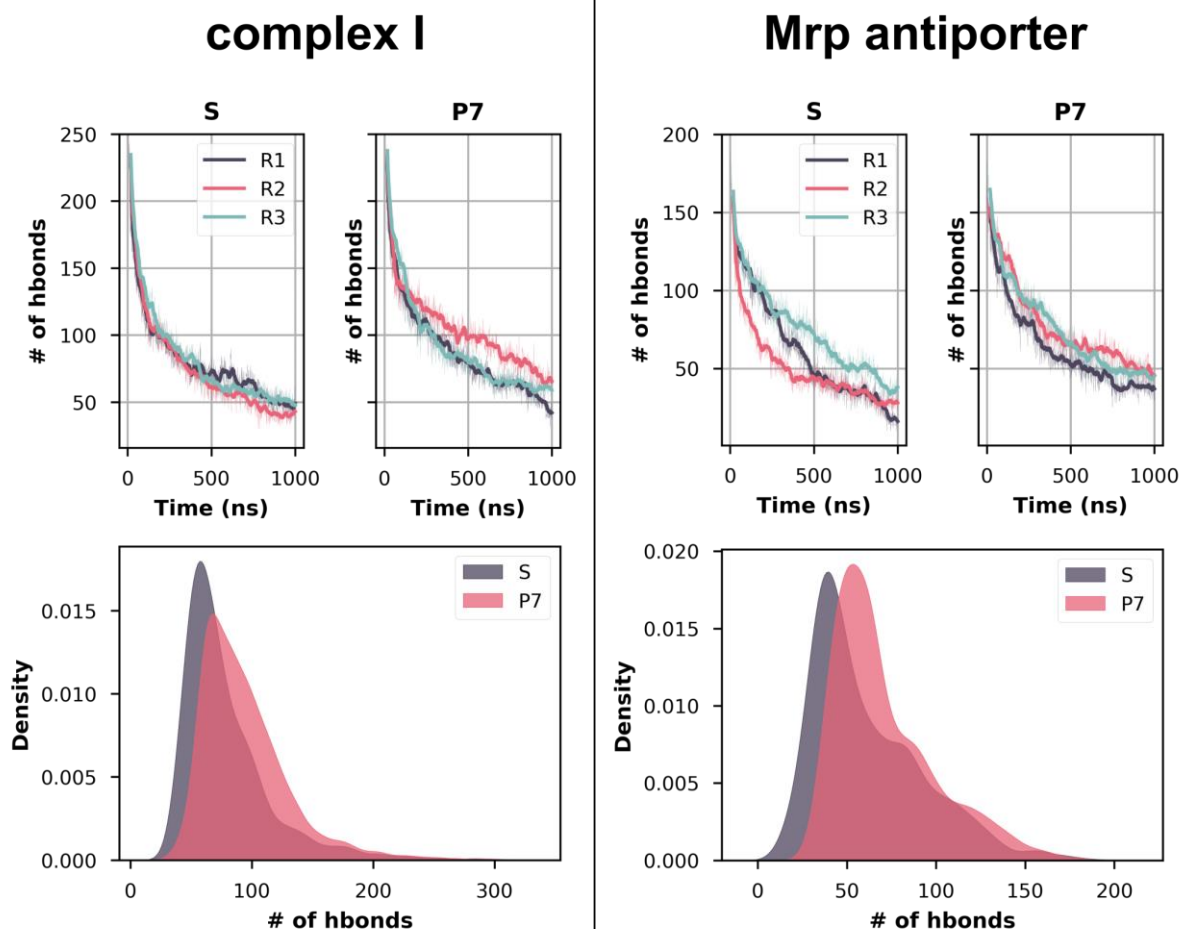
Fig 4 – Hydrogen bonds between protein and structural water molecules. (Top) The number of hydrogen bonds throughout the S and P7 simulation trajectories are shown as a time series. The different colors of traces represent different simulation replicas. (Bottom) The data from all trajectories is shown in the lower panels as a density. Only the core membrane subunits of complex I were considered for analysis. The hydrogen bonding distance was cut off at 3.5 Å and the angle at 150°.

**Protein-protein interactions**

In addition to the protein-water hydrogen bonding interactions, protein-protein hydrogen bonds were also analyzed. The number of protein-protein sidechain hydrogen bonds are consistently found to be higher throughout the P7 state trajectories (Fig. 5). This implies that structural stability is higher in P7 states compared to the S states for both Mrp antiporter and complex I, as also observed in RMSD and RMSF analysis (see above). The lower number of protein-protein interactions is also in agreement with the higher level of hydration observed in the S state. In other words, enhanced water entry/exit in the protein causes protein-protein interactions to be perturbed, resulting in structural destabilization. Specific pairs of residues from the Mrp antiporter that showed a significant increase (>50 %) in hydrogen bonding are shown in Table S4. Interestingly, although only two of the residue pairs undergo changes in protonation state between S and P7 states, still several other pairs show enhancement in hydrogen bonding occupancy, pointing out that long-range effects can occur upon charge change.

304
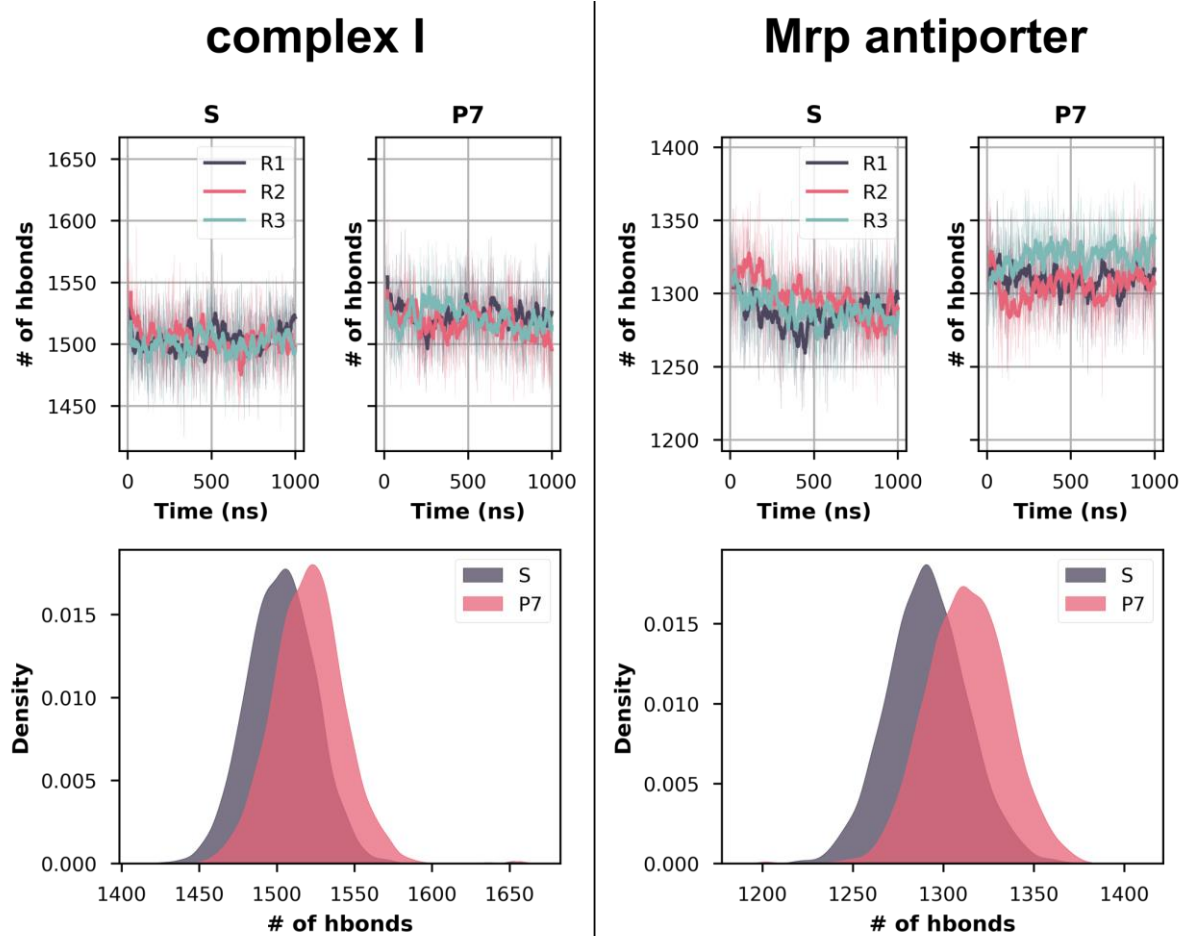
Fig. 5 – Number of hydrogen bonds between protein sidechains throughout the simulation for complex I and Mrp antiporter. Top panels show the time evolution of all hydrogen bonds for three different simulation replicas, with the hydrogen bonding distance cut off at 3.5 Å and the angle at 150°. Bottom panels show the same data as KDE plots. In case of complex I, only core subunits of the membrane arm are considered (see Fig. S4 for full protein).

The above data on protein dynamics and hydration, as well as hydrogen bonding analysis in two different charge states lead to the conclusion that higher protein fluctuations are caused by extensive hydration, which in turn is caused by the highly charged state of the protein, as in the S state. The correct (or near-to-correct) charge state of the protein is thus critical for its structural stability. Hence, prior to launching long time scale MD simulations of high-resolution membrane proteins, a proper charge analysis by means of pKa calculations is recommended.

Simulations with different pH values

It is clear from the above analysis that the protein simulations that have charge states defined by predetermined protonation states of titratable residues set at pH 7 are well-behaved and are closer to the conformation obtained from cryo-EM. This is true for both global and local mobility, as well as protein hydration. To further explore how altering protonation state affects these characteristics, we performed a series of simulations of the Mrp antiporter complex at 4 more pH values: 5, 6, 8, and 9. The RMSD of the protein (Fig. 6A) was lowest at pH 6 and pH 7, with pH 5 showing the most instability. Similarly, both the number of water

328 molecules in contact with protein (Fig. 6B) and the number of water molecules in the central
329 hydrophilic axis (Fig. 6C) was lowest for pH 6 and pH 7, with pH 5 deviating the most. At pH 5,
330 many of the negatively charged residues become neutral, and positive residues stay positive.
331 Importantly, many histidine residues become doubly protonated, which introduces excess
332 positive charge in several regions of the protein. As a result, bulk water molecules enter the
333 protein, causing instability in protein sidechains and loops, even more than in the S state.
334
335
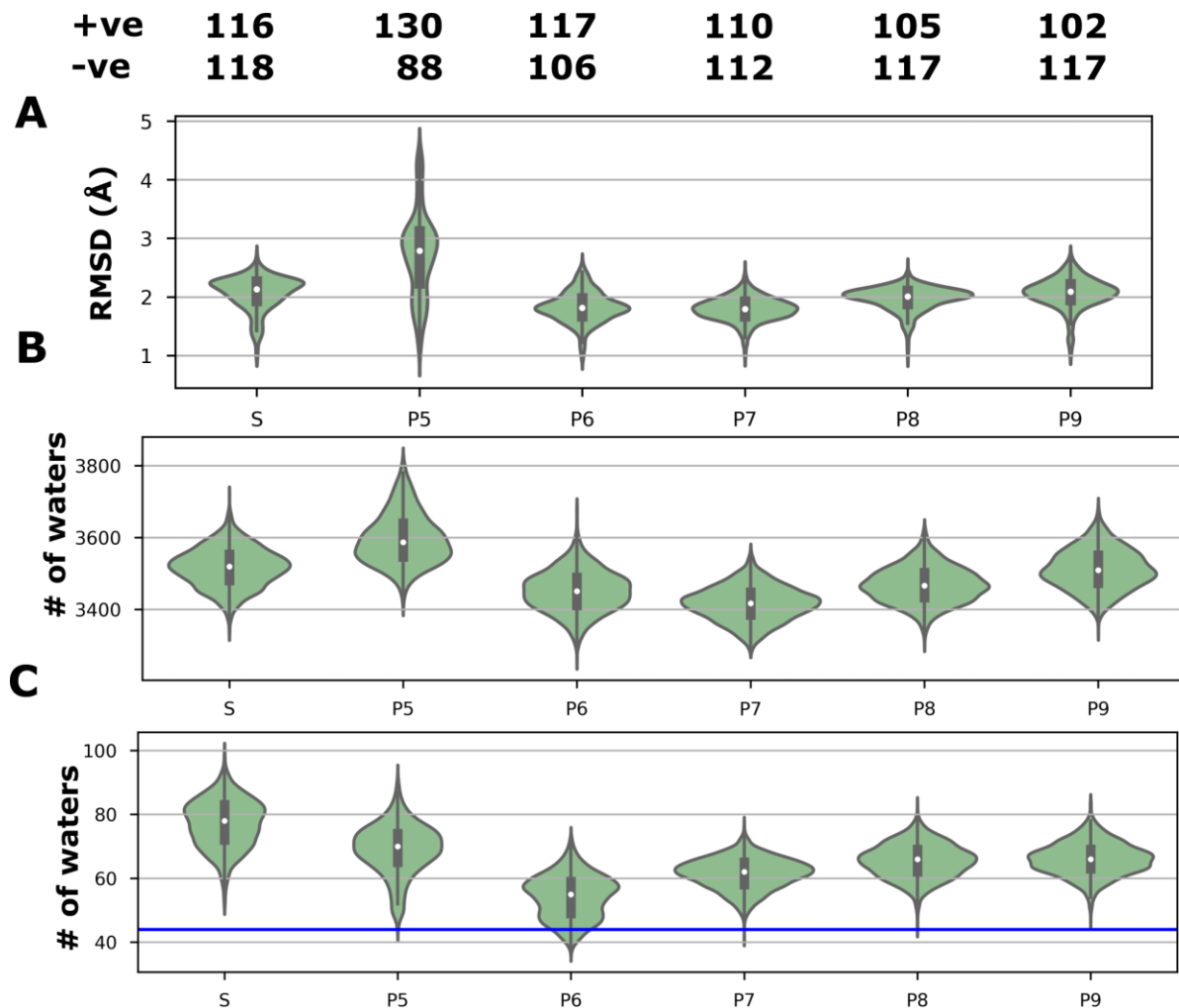
336
337 **Fig. 6** – Comparison of Mrp antiporter MD simulations with protonation states set at different
338 pH values. Panel A shows the RMSD of backbone atoms, panel B shows the number of water
339 molecules within 4 Å of all protein residues, and panel C shows the number of waters in the
340 functionally relevant central axis (Fig. 1), with the blue line showing the number of waters in
341 contact with protein observed in the cryo-EM structure. Data is shown as violin plots, where
342 the shaded area is a KDE plot of all trajectory data and in the center is a box plot displaying
343 the median (white dot with the black box indicating the interquartile range). The total number
344 of positively and negatively charged residues at different pH are listed above panel A.
345
346 **Charge state assignment based on cryo-EM density**
347 The above analysis highlights the importance of correct charge state description of protein,
348 which is necessary for its stability during MD simulations. However, prediction of charge state
349 by pKa calculations has its limitations. We thus attempted to deduce the protonation states

350   of amino acids from the high resolution cryo-EM data. Cryo-EM density maps can be used to
351   identify the charge states of acidic amino acid residues (glutamates and aspartates), because
352   neutral and negatively charged atoms have vastly different scattering amplitudes [14, 17] and
353   thus different cryo-EM density profiles. By carefully analyzing the cryo-EM density in our high-
354   resolution maps, we obtained the charge states of selected glutamate and aspartate residues
355   and compared those to the pKa values obtained based on the modeled structure. We found
356   that there was an agreement between the charge state assignment based on the cryo-EM
357   maps (EMD-12742 for complex I and EMD-14124 for the Mrp antiporter) and the predicted
358   pKa based on the derived atomic models (PDBs 7O71 and 7QRU, respectively). However,
359   several acidic residues showed discrepancies; the pKa calculations predicted some
360   carboxylates to be charge neutral that were assigned to be anionic by the cryo-EM density
361   analysis (Table 1). A possible explanation for this discrepancy is that although the protein
362   model-based pKa predictions are largely correct (especially for buried residues), the positions
363   of the (invisible) sidechains of charged glutamates and aspartates were incorrectly modeled,
364   resulting in overestimated structure-based pKas. To fix this discrepancy between the two
365   estimates, the cryo-EM density map of respiratory complex I (EMD-12742) was revisited. We
366   remodeled the sidechains of selected residues that are predicted to be anionic based on
367   missing cryo-EM densities using Coot [49] and re-performed the pKa prediction on the new
368   model. A total of 7 glutamates and aspartates were remodeled one at a time (Table 1).
369

370   **Table 1 -** Charge state of selected acidic amino acid residues in complex I from cryo-EM
371   density data (EMD 12742) and pKa calculations, based on PDB 7O71 coordinates or on a
372   remodeled carboxylate sidechain. All carboxylates were remodeled independently.
373   [a] D67^{ND3} was not remodeled, but instead the new pKa value is based on the structure with
374   E147^{ND1} remodeled.

	Density charge estimate	Original pKa	Remodeled pKa
E147^{ND1}	Charged	9.11	7.03
E206^{ND1}	Charged	8.85	6.37
E210^{ND1}	Charged	7.96	5.69
E231^{ND1}	Charged	7.44	6.56
D67^{ND3}	Neutral	7.18	9.34[a]
E69^{ND3}	Charged	9.27	9.16
E30^{ND4L}	Charged	7.59	8.69
E66^{ND4L}	Charged	9.81	10.16

375

376   The residues E231^{ND1} and E147^{ND1} of the functionally relevant E-channel in complex I [43] were
377   both predicted to be neutral based on pKa calculations but charged based on density analysis.
378   When pKa calculations were performed on re-modeled sidechains, it led to a drop in their
379   sidechain pKa and stabilization of their deprotonated states in good agreement with cryo-EM
380   density-based assignment (Table 1). Interestingly, the remodeled sidechain conformation of
381   E147^{ND1} not only improved the pKa of the residue itself, but also that of residues in its
382   surroundings, e.g. D67^{ND3} (Fig. 7, Table 1). A similar improved agreement between pKa
383   prediction and cryo-EM-based charge state assignment was also observed for E206 and E210
384   from the ND1 subunit (Table 1), both of which are known to be central for protein function
385   [43]. However, pKa predictions of E30 and E66 from ND4L, as well as E69 from ND3 could not be

386 improved despite their remodeled sidechains. One potential reason for this could be that the
387 surroundings in two different conformations do not differ as drastically as in the other cases.
388 Furthermore, the approximate nature of pKa predictions is well known, and with no explicit
389 treatment of neighboring solvation can result in the overestimation of pKa values. All in all,
390 remodeling of several sidechains not only improved the charged state assignment, but also
391 improved the atomic modeling of cryo-EM density data (see discussion).
392
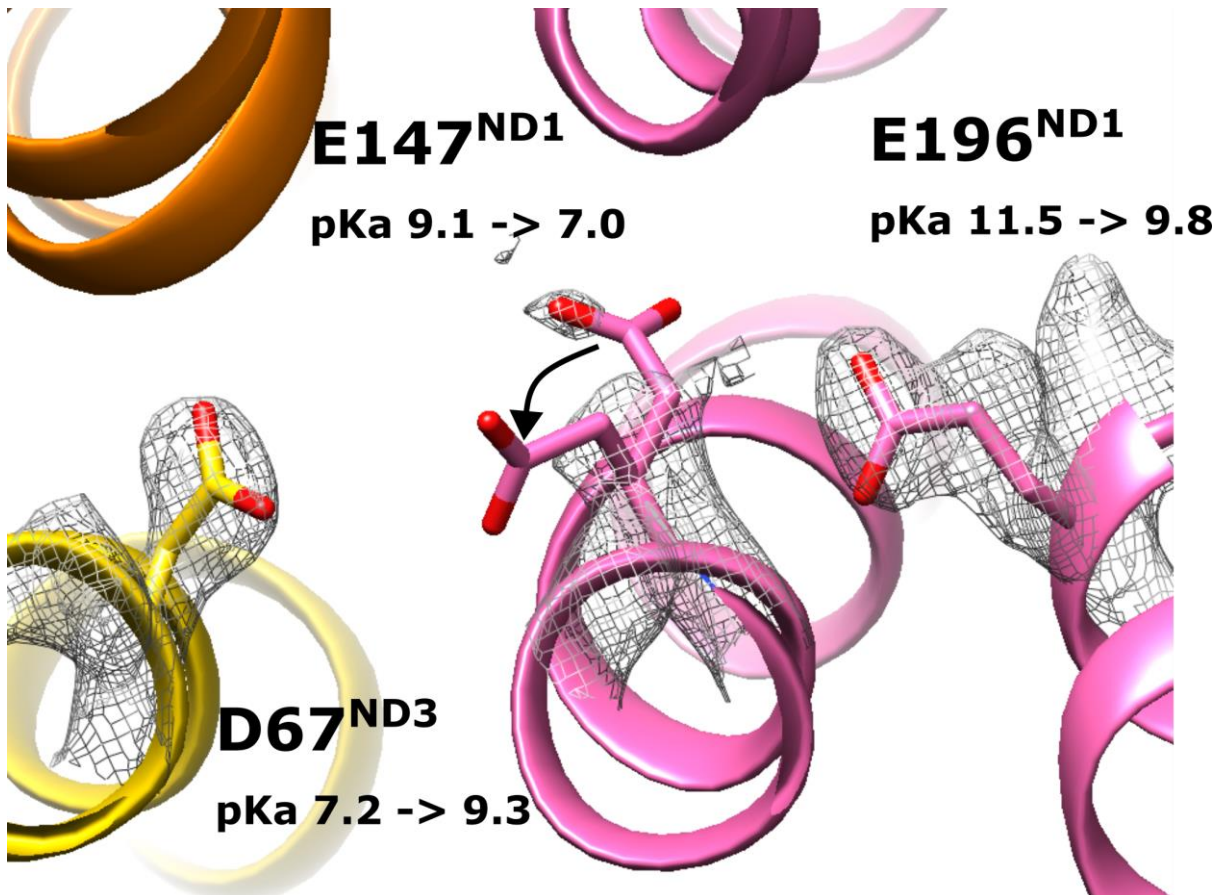
393
394 **Fig 7 -** The remodeling of the sidechain of E147^{ND1} (indicated by black arrow) improved its pKa
395 as well as that of neighboring D67^{ND3}. The first value of pKa is based on the original
396 coordinates (PDB 7O71), while the second value is after remodeling the sidechain of E147^{ND1}.
397 The cryo-EM density of the three residues is shown as grey mesh.
398

Discussion

400 Protonation states of amino acids play a key role in protein structure stability and function.
401 However, it is extremely challenging to obtain an accurate estimate of the pKa of an amino
402 acid both by experiments and computations, and as a result in most cases, pKa values and
403 how they change during enzymatic action remains unknown. Because of this, MD simulations
404 are often performed by modeling amino acids in their fixed protonation states, including
405 standard charged states. Here we show that MD simulations of a membrane protein with all
406 amino acids modeled in their standard charged states results in lower stability of the protein
407 overall and drives the conformation of the protein far from its starting point that is the cryo-
408 EM conformation. Instead, if MD simulation is performed by fixing the charged states of
409 titratable amino acids based on pKa calculations, the conformational state remains closer to

410   the original structure. Our analysis of protein mobility and hydration, as well as residue-water
411   and residue-residue interactions supports this notion. Also, since the structural conformation
412   is retained in MD simulations performed after pKa calculations, we propose these more
413   closely match the charge states of amino acids that existed during the cryo-EM experiment.
414

415   Our results also indicate that the relationship between protein stability and hydration is
416   strong, and that reduced hydration leads to a more stable simulation system which can
417   maintain important interactions. When leaving all protonatable residues in their standard
418   charged states, hydration is considerably enhanced, resulting in important protein-protein
419   interactions being broken, thereby reducing stability.
420

421   The higher level of hydration observed in standard state simulations is primarily due to the
422   charged states of buried titratable amino acid residues. These charged amino acid residues in
423   the low protein dielectric interior create an unfavorable high-energy scenario due to poor
424   solvation. This establishes an electric field, pulling bulk water molecules into the protein and
425   enhancing solvation similar to the nanoscopic electroporation described earlier [50], see also [33].
426   We also note that protein hydration obtained from MD simulations performed in pre-defined
427   charge state is much closer to the structural hydration than that from the MD simulations
428   performed in standard state of the residues. Despite this, the hydration obtained from MD
429   simulations is in general higher than the hydration observed in the structure. The reason for
430   this discrepancy is in part due to the underestimation of protein hydration in the cryo-EM
431   maps, where only highly occupied and tightly bound water molecules can be observed.
432

433   By performing MD simulations in pre-determined charge state from pKa calculations, one not
434   only investigates the state that was captured during structure resolution, but also its time
435   evolution more accurately. This is particularly useful in situations where there are a
436   particularly large number of titratable residues present in the protein with possible functional
437   relevance, such as in photosynthetic and respiratory enzymes, as well as proteins that couple
438   translocation of ions or metabolites to proton transfer reactions. However, empirical pKa
439   calculations or more thorough continuum electrostatics based pKa estimations have limited
440   accuracy and site-site interaction energy terms can often dominate, leading to large scale pKa
441   shifts with subtle changes in structure. Thus, it is important to use an accurate input model,
442   based on high-resolution data where there is less uncertainty in sidechain placements. Below
443   ~2.5 Å resolution the correct rotamers of most sidechains can be confidently modeled, with
444   the key exception being negatively charged sidechains in cryo-EM maps, which have negative
445   atomic scattering factors [13, 17, 18] and thus are usually invisible beyond the Cβ atoms. As
446   modeling and refinement programs do not take this into account, the sidechain orientation
447   of these residues is unreliable [18]. On the other hand, this fact makes it possible to deduce the
448   charge state of acidic residues from the cryo-EM density.
449

450   Here we propose a method that may assist to improve modeling of sidechains of acidic
451   residues. By performing fast empirical pKa calculations on the 3D model, pKas of all acidic
452   residues can be estimated. If the predicted pKa value suggests a neutral state of an acidic
453   residue that is expected to be anionic based on density analysis, this warrants remodeling of
454   the sidechain of the amino acid. If the remodeled sidechain conformation yields a lower pKa,
455   this is likely to be the more accurate conformation. By combining pKa calculations with
456   sidechain remodeling on selected acidic residues of complex I, we show here that this is

457 feasible. In our future work, we aim to automate this process and obtain more accurate
458 models of sidechains of acidic amino acid residues from the cryo-EM data.
459

Methods

461 We performed all-atom molecular dynamics simulations of complex I from *Yarrowia lipolytica*
462 (PDB: 7O71)[43] and Mrp antiporter from *Bacillus pseudofirmus* (PDB: 7QRU, altloc B)[44]. Both
463 proteins were placed in a mixed lipid bilayer and solvated with TIP3P and NaCl. Gromacs
464 versions 2020 and 2021 [51] were used for minimization, equilibration, and production runs.
465 Full details of the molecular dynamics set ups can be found in previous work on complex I[43]
466 and Mrp antiporter[44]. The production runs were performed without any constraints in an NPT
467 ensemble, using Nosé–Hoover thermostat[52, 53] at 310 K and Parrinello-Rahman barostat[54] at
468 1 atm. We employed the LINCS algorithm to achieve a 2 fs timestep[55], and the particle-mesh
469 Ewald method[56] with a cut-off of 12 Å to handle electrostatic interactions. The cut-off for van
470 der Waals interactions was 12 Å, with a switching distance of 10 Å. Table S5 lists the
471 simulations setups, and the S and P7 states for complex I and Mrp antiporter are extended
472 simulations from previous work. The P5, P6, P8, and P9 simulations of Mrp are new systems
473 constructed using the same protocol.
474

475 Estimates of pKa were performed using the Propka software package[39]. The calculations were
476 performed on the PDB structures of both complex I and Mrp. Asp, Glu and His were
477 considered protonated (neutral) if their pKa was more than 7 in the P7 simulations, while Lys
478 was considered deprotonated if its pKa was less than 7. The same process was performed in
479 the P5, P6, P8, and P9 setups, with their respective pKa cutoffs changed. His was considered
480 with the δ-nitrogen protonated when neutral.
481

482 Exponential fit of hydrogen bond decay was calculated using the SciPy *curve_fit* function[57].
483 We used a mono-exponential of the form $y = me^{(-tx)} + b$, where m, t, and b were
484 parameters that could vary. The quality of the fit was determined using R², obtaining values
485 > 0.9. The half-life was calculated as $t_{1/2} = t \times \ln 2$.
486

487 The electron density analysis and remodeling of the resolved structures was done using
488 Coot[49]. The change in the glutamic and aspartic acid conformations was achieved with the
489 rotamers function in Coot. The key requirement for the new position of the sidechain was a
490 substantial deviation from its previous position, as well as avoiding clashes with its
491 surroundings. Care was taken not to position the side chain in positive density.
492

493 Visualization, analysis, and figure preparation was performed using VMD, Pymol and UCSF
494 Chimera [58].
495

496

Acknowledgements

References

1. Henderson, R., Overview and future of single particle electron cryomicroscopy. *Archives of Biochemistry and Biophysics* **2015,** *581*, 19-24.

2. Kühlbrandt, W., Forty years in cryoEM of membrane proteins. *Microscopy* **2022,** *71* (Supplement_1), i30-i50.

3. Bartesaghi, A.; Merk, A.; Banerjee, S.; Matthies, D.; Wu, X.; Milne, J. L. S.; Subramaniam, S., 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science* **2015,** *348* (6239), 1147-1151.

4. Ishikita, H.; Saito, K., Proton transfer reactions and hydrogen-bond networks in protein environments. *Journal of The Royal Society Interface* **2014,** *11* (91), 20130518.

5. Cukierman, S., Et tu, Grotthuss! and other unfinished stories. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2006,** *1757* (8), 876-885.

6. Hong, G.; Cornish, A. J.; Hegg, E. L.; Pachter, R., On understanding proton transfer to the biocatalytic [Fe―Fe]H sub-cluster in [Fe―Fe]H2ases: QM/MM MD simulations. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2011,** *1807* (5), 510-517.

7. Zdorevskyi, O.; Djurabekova, A.; Lasham, J.; Sharma, V., Horizontal proton transfer across the antiporter-like subunits in mitochondrial respiratory complex I. *Chemical Science* **2023,** *14* (23), 6309-6318.

8. Sugo, Y.; Ishikita, H., Mechanism of Asparagine-Mediated Proton Transfer in Photosynthetic Reaction Centers. *Biochemistry* **2023,** *62* (10), 1544-1552.

9. Kaila, V. R.; Sharma, V.; Wikström, M., The identity of the transient proton loading site of the proton-pumping mechanism of cytochrome c oxidase. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **2011,** *1807* (1), 80-84.

10. Kaur, D.; Khaniya, U.; Zhang, Y.; Gunner, M. R., Protein Motifs for Proton Transfers That Build the Transmembrane Proton Gradient. *Frontiers in Chemistry* **2021,** *9*.

11. Fukuda, Y.; Hirano, Y.; Kusaka, K.; Inoue, T.; Tamada, T., High-resolution neutron crystallography visualizes an OH-bound resting state of a copper-containing nitrite reductase. *Proceedings of the National Academy of Sciences* **2020,** *117* (8), 4071-4077.

12. Kühlbrandt, W.; Wang, D. N.; Fujiyoshi, Y., Atomic model of plant light-harvesting complex by electron crystallography. *Nature* **1994,** *367* (6464), 614-621.

13. Kimura, Y.; Vassylyev, D. G.; Miyazawa, A.; Kidera, A.; Matsushima, M.; Mitsuoka, K.; Murata, K.; Hirai, T.; Fujiyoshi, Y., Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature* **1997,** *389* (6647), 206-211.

14. Marques, M. A.; Purdy, M. D.; Yeager, M., CryoEM maps are full of potential. *Current Opinion in Structural Biology* **2019,** *58*, 214-223.

15. Allegretti, M.; Mills, D. J.; McMullan, G.; Kühlbrandt, W.; Vonck, J., Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *Elife* **2014,** *3*, e01963.

16. Bartesaghi, A.; Matthies, D.; Banerjee, S.; Merk, A.; Subramaniam, S., Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences* **2014,** *111* (32), 11709-11714.

17. Yonekura, K.; Kato, K.; Ogasawara, M.; Tomita, M.; Toyoshima, C., Electron crystallography of ultrathin 3D protein crystals: atomic model with charges. *Proceedings of the National Academy of Sciences* **2015,** *112* (11), 3368-3373.

18. Wang, J., On the appearance of carboxylates in electrostatic potential maps. *Protein Science* **2017,** *26* (3), 396-402.

19. Maki-Yonekura, S.; Kawakami, K.; Takaba, K.; Hamaguchi, T.; Yonekura, K., Measurement of charges and chemical bonding in a cryo-EM structure. *Communications Chemistry* **2023,** *6* (1), 98.

20. Lundborg, M.; Narangifard, A.; Wennberg, C. L.; Lindahl, E.; Daneholt, B.; Norlén, L., Human skin barrier structure and function analyzed by cryo-EM and molecular dynamics simulation. *Journal of Structural Biology* **2018,** *203* (2), 149-161.

21. Arenz, S.; Bock, L. V.; Graf, M.; Innis, C. A.; Beckmann, R.; Grubmüller, H.; Vaiana, A. C.; Wilson, D. N., A combined cryo-EM and molecular dynamics approach reveals the mechanism of ErmBL-mediated translation arrest. *Nature Communications* **2016,** *7* (1), 12026.

22. Hadden, J. A.; Perilla, J. R.; Schlicksup, C. J.; Venkatakrishnan, B.; Zlotnick, A.; Schulten, K., All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits. *eLife* **2018,** *7*, e32478.

23. Ahmad, K.; Javed, A.; Lanphere, C.; Coveney, P. V.; Orlova, E. V.; Howorka, S., Structure and dynamics of an archetypal DNA nanoarchitecture revealed via cryo-EM and molecular dynamics simulations. *Nature Communications* **2023,** *14* (1), 3630.

24. Li, W.; Frank, J., Transfer RNA in the hybrid P/E state: Correlating molecular dynamics simulations with cryo-EM data. *Proceedings of the National Academy of Sciences* **2007,** *104* (42), 16540-16545.

25. Réblová, K.; Rázga, F.; Li, W.; Gao, H.; Frank, J.; Šponer, J., Dynamics of the base of ribosomal A-site finger revealed by molecular dynamics simulations and Cryo-EM. *Nucleic Acids Research* **2009,** *38* (4), 1325-1340.

26. Jojoa-Cruz, S.; Saotome, K.; Murthy, S. E.; Tsui, C. C. A.; Sansom, M. S. P.; Patapoutian, A.; Ward, A. B., Cryo-EM structure of the mechanically activated ion channel OSCA1.2. *eLife* **2018,** *7*, e41845.

27. Johnson, R. M.; Fais, C.; Parmar, M.; Cheruvara, H.; Marshall, R. L.; Hesketh, S. J.; Feasey, M. C.; Ruggerone, P.; Vargiu, A. V.; Postis, V. L. G.; Muench, S. P.; Bavro, V. N., Cryo-EM Structure and Molecular Dynamics Analysis of the Fluoroquinolone Resistant Mutant of the AcrB Transporter from Salmonella. *Microorganisms* **2020,** *8* (6), 943.

28. Zhang, M.; Gui, M.; Wang, Z.-F.; Gorgulla, C.; Yu, J. J.; Wu, H.; Sun, Z.-y. J.; Klenk, C.; Merklinger, L.; Morstein, L.; Hagn, F.; Plückthun, A.; Brown, A.; Nasr, M. L.; Wagner, G., Cryo-EM structure of an activated GPCR–G protein complex in lipid nanodiscs. *Nature Structural & Molecular Biology* **2021,** *28* (3), 258-267.

29. Fuss, M. F.; Wieferig, J.-P.; Corey, R. A.; Hellmich, Y.; Tascón, I.; Sousa, J. S.; Stansfeld, P. J.; Vonck, J.; Hänelt, I., Cyclic di-AMP traps proton-coupled K+ transporters of the KUP family in an inward-occluded conformation. *Nature Communications* **2023,** *14* (1), 3683.

30. Tascón, I.; Sousa, J. S.; Corey, R. A.; Mills, D. J.; Griwatz, D.; Aumüller, N.; Mikusevic, V.; Stansfeld, P. J.; Vonck, J.; Hänelt, I., Structural basis of proton-coupled potassium transport in the KUP family. *Nature communications* **2020,** *11* (1), 626.

31. Wei, R. J.; Khaniya, U.; Mao, J.; Liu, J.; Batista, V. S.; Gunner, M. R., Tools for analyzing protonation states and for tracing proton transfer pathways with examples from the Rb. sphaeroides photosynthetic reaction centers. *Photosynthesis Research* **2022**.

32. Aho, N.; Buslaev, P.; Jansen, A.; Bauer, P.; Groenhof, G.; Hess, B., Scalable Constant pH Molecular Dynamics in GROMACS. *Journal of Chemical Theory and Computation* **2022,** *18* (10), 6148-6160.

33. Kaila, V. R.; Wikström, M.; Hummer, G., Electrostatics, hydration, and proton transfer dynamics in the membrane domain of respiratory complex I. *Proceedings of the National Academy of Sciences* **2014,** *111* (19), 6988-6993.

34. Roh, S.-H.; Shekhar, M.; Pintilie, G.; Chipot, C.; Wilkens, S.; Singharoy, A.; Chiu, W., Cryo-EM and MD infer water-mediated proton transport and autoinhibition mechanisms of V_o complex. *Science Advances* **2020,** *6* (41), eabb9605.

604   35.   Haapanen, O.; Sharma, V., Role of water and protein dynamics in proton pumping by
605   respiratory complex I. *Scientific Reports* **2017,** *7* (1), 7747.
606   36.   Kurnikov, I. V.; Kyrychenko, A.; Flores-Canales, J. C.; Rodnin, M. V.; Simakov,
607   N.; Vargas-Uribe, M.; Posokhov, Y. O.; Kurnikova, M.; Ladokhin, A. S., pH-Triggered
608   Conformational Switching of the Diphtheria Toxin T-Domain: The Roles of N-Terminal
609   Histidines. *Journal of Molecular Biology* **2013,** *425* (15), 2752-2764.
610   37.   Lazaratos, M.; Siemers, M.; Brown, L. S.; Bondar, A.-N., Conserved hydrogen-bond
611   motifs of membrane transporters and receptors. *Biochimica et Biophysica Acta (BBA) -*
612   *Biomembranes* **2022,** *1864* (6), 183896.
613   38.   Gunner, M. R.; Mao, J.; Song, Y.; Kim, J., Factors influencing the energetics of
614   electron and proton transfers in proteins. What can be learned from calculations. *Biochimica*
615   *et Biophysica Acta (BBA) - Bioenergetics* **2006,** *1757* (8), 942-968.
616   39.   Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3:
617   consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of*
618   *chemical theory and computation* **2011,** *7* (2), 525-537.
619   40.   Davies, M. N.; Toseland, C. P.; Moss, D. S.; Flower, D. R., Benchmarking pK(a)
620   prediction. *BMC Biochem* **2006,** *7*, 18.
621   41.   Haapanen, O.; Djurabekova, A.; Sharma, V., Role of Second Quinone Binding Site in
622   Proton Pumping by Respiratory Complex I. *Frontiers in chemistry* **2019,** *7*, 221.
623   42.   Khaniya, U.; Gupta, C.; Cai, X.; Mao, J.; Kaur, D.; Zhang, Y.; Singharoy, A.;
624   Gunner, M., Hydrogen bond network analysis reveals the pathway for the proton transfer in
625   the E-channel of T. thermophilus Complex I. *Biochimica et Biophysica Acta (BBA)-*
626   *Bioenergetics* **2020**, 148240.
627   43.   Parey, K.; Lasham, J.; Mills, D. J.; Djurabekova, A.; Haapanen, O.; Yoga, E. G.;
628   Xie, H.; Kühlbrandt, W.; Sharma, V.; Vonck, J.; Zickermann, V., High-resolution structure
629   and dynamics of mitochondrial complex I—Insights into the proton pumping
630   mechanism. *Sci. Adv.* **2021,** *7* (46), eabj3221.
631   44.   Lee, Y.; Haapanen, O.; Altmeyer, A.; Kühlbrandt, W.; Sharma, V.; Zickermann,
632   V., Ion transfer mechanisms in Mrp-type antiporters from high resolution cryoEM and
633   molecular dynamics simulations. *Nature Communications* **2022,** *13* (1), 6091.
634   45.   Grba, D. N.; Hirst, J., Mitochondrial complex I structure reveals ordered water
635   molecules for catalysis and proton translocation. *Nature Structural & Molecular Biology*
636   **2020,** *27* (10), 892-900.
637   46.   Kampjut, D.; Sazanov, L. A., The coupling mechanism of mammalian respiratory
638   complex I. *Science* **2020,** *370* (6516).
639   47.   Chung, I.; Wright, J. J.; Bridges, H. R.; Ivanov, B. S.; Biner, O.; Pereira, C. S.;
640   Arantes, G. M.; Hirst, J., Cryo-EM structures define ubiquinone-10 binding to mitochondrial
641   complex I and conformational transitions accompanying Q-site occupancy. *Nature*
642   *Communications* **2022,** *13* (1), 2758.
643   48.   Gu, J.; Liu, T.; Guo, R.; Zhang, L.; Yang, M., The coupling mechanism of
644   mammalian mitochondrial complex I. *Nature Structural & Molecular Biology* **2022,** *29* (2),
645   172-182.
646   49.   Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta*
647   *crystallographica section D: biological crystallography* **2004,** *60* (12), 2126-2132.
648   50.   Hummer, G.; Wikström, M., Molecular simulation and modeling of complex I.
649   *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **2016,** *1857* (7), 915-921.
650   51.   Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl,
651   E., GROMACS: High performance molecular simulations through multi-level parallelism
652   from laptops to supercomputers. *SoftwareX* **2015,** *1*, 19-25.

653 52.   Nosé, S., A unified formulation of the constant temperature molecular dynamics
654 methods. *The Journal of chemical physics* **1984,** *81* (1), 511-519.
655 53.   Hoover, W. G., Canonical dynamics: equilibrium phase-space distributions. *Physical*
656 *review A* **1985,** *31* (3), 1695.
657 54.   Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new
658 molecular dynamics method. *Journal of Applied physics* **1981,** *52* (12), 7182-7190.
659 55.   Hess, B., P-LINCS: A parallel linear constraint solver for molecular simulation.
660 *Journal of Chemical Theory and Computation* **2008,** *4* (1), 116-122.
661 56.   Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N· log (N) method for
662 Ewald sums in large systems. *The Journal of chemical physics* **1993,** *98* (12), 10089-10092.
663 57.   Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.;
664 Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J., SciPy 1.0:
665 fundamental algorithms for scientific computing in Python. *Nature methods* **2020,** *17* (3),
666 261-272.
667 58.   Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.;
668 Meng, E. C.; Ferrin, T. E., UCSF Chimera—a visualization system for exploratory research
669 and analysis. *Journal of computational chemistry* **2004,** *25* (13), 1605-1612.
670
671

672