doi:10.1016/j.jmb.2009.02.042 J. Mol. Biol. (2009) 387, 1165–1185

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Functional Hydration and Conformational Gating of Proton Uptake in Cytochrome *c* Oxidase

Rowan M. Henry, Ching-Hsing Yu, Tomas Rodinger and Régis Pomès*

Molecular Structure and Function, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8

Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8

Received 12 December 2008; received in revised form 10 February 2009; accepted 14 February 2009
Available online
24 February 2009

Cytochrome *c* oxidase couples the reduction of dioxygen to proton pumping against an electrochemical gradient. The D-channel, a 25-Å-long cavity, provides the principal pathway for the uptake of chemical and pumped protons. A water chain is thought to mediate the relay of protons via a Grotthuss mechanism through the D-channel, but it is interrupted at N139 in all available crystallographic structures. We use free-energy simulations to examine the proton uptake pathway in the wild type and in single-point mutants N139V and N139A, in which redox and pumping activities are compromised. We present a general approach for the calculation of water occupancy in protein cavities and demonstrate that combining efficient sampling algorithms with long simulation times (hundreds of nanoseconds) is required to achieve statistical convergence of equilibrium properties in the protein interior. The relative population of different conformational and hydration states of the D-channel is characterized. Results shed light on the role of N139 in the mechanism of proton uptake and clarify the physical basis for inactive phenotypes. The conformational isomerization of the N139 side chain is shown to act as a gate controlling the formation of a functional water chain or "proton wire." In the closed state of N139, the spatial distribution of water in the D-channel is consistent with available crystallographic models. However, a metastable state of N139 opens up a narrow bottleneck in which 50% occupancy by a water molecule establishes a proton pathway throughout the D-channel. Results for N139V suggest that blockage of proton uptake resulting from persistent interruption of the water pathway is the cause of this mutant's marginal oxidase activity. In contrast, results for N139A indicate that the D-channel is a continuously hydrated cavity, implying that the decoupling of oxidase activity from proton pumping measured in this mutant is not due to interruption of the proton relay chain.

Keywords: water-mediated proton transport; free energy simulations; distributed replica sampling; protein hydration; redox-coupled proton pump

Edited by D. Case

Introduction

Cytochrome *c* oxidase (C*c*O), an intrinsic membrane protein found in eukaryotes and in many bacteria, is the terminal enzyme in the electron transfer chain. Its role is to convert molecular oxygen to water and use the energy liberated from this redox reaction to pump protons across the membrane against an electrochemical gradient. The resulting proton motive force is then utilized by ATP synthase to drive the synthesis of ATP. Extensive studies have been carried out in an attempt to understand the mechanism of the redox-coupled pumping of protons by C*c*O,[1–4] and a recurrent electrostatic mechanism linking electron and proton movement

*Corresponding author. E-mail address: pomes@sickkids.ca.

Present addresses: C.-H. Yu, SciNet High Performance Computing Consortium, Toronto, Ontario, Canada M5T 1W5; T. Rodinger, ZymeWorks, Inc., Vancouver, British Columbia, Canada V6Z 2P5.

Abbreviations used: C*c*O, cytochrome *c* oxidase; PMF, potential of mean force; GCMC, grand-canonical Monte Carlo; DRPE, distributed replica potential energy; MD, molecular dynamics.

throughout the entire redox cycle has recently been proposed.[5] The overall reaction catalyzed by C*c*O is

$$O_2 + 4e^-_{out} + 8H^+_{in} \rightarrow 2H_2O + 4H^+_{out},$$

where "in" and "out" denote the inner mitochondrial matrix (cytosol in bacteria) and the intermembrane space (periplasmic space), respectively.[6] It is believed that all protons enter C*c*O through one of two channels, the K-channel or the D-channel.[7–9] The K-channel, contained within subunit II, spans from E101 at the entrance to Y288 at the active site and contains a critical lysine residue, K362, for which the channel is named (unless otherwise noted, all numbering corresponds to the *Rhodobacter sphaeroides* enzyme). The K-channel is responsible for the delivery of one or two protons to the reaction site (binuclear centre), with the remaining six to seven protons transferred to the active site through the D-channel.[10,11]

The D-channel is located in subunit I and extends approximately 25 Å from residue D132 at the entrance, to residue E286, which is approximately 12 Å from the binuclear centre. Residue D132, for which the D-channel is named, is a highly conserved residue that has been proposed to act as a "proton antenna", recruiting protons from the cytosol to be transferred through the D-channel.[12] Residue E286 plays an essential role in the catalytic activity by shuttling protons from the D-channel to both the binuclear centre and the proton loading site.[13–16] Point mutations of some of the residues lining the D-channel have been observed to significantly affect the activity of the enzyme and highlight the importance of the D-channel to the mechanism of redox-coupled proton pumping.[9,10,12,17–25] In particular, single-point mutations of residues N139 and N207 to an aspartate residue produce a decoupled phenotype, whereby the protein maintains wild-type turnover while completely abolishing proton pumping.[20,23] Moreover, the single-point mutation G204D eliminates all protein activity.[22] These mutant phenotypes appear to be the result of the introduction of a charged group in the D-channel, which induces an increase in the pK_a of E286, perturbing its ability to relay protons to the binuclear centre and the proton loading site.[5] However, decoupled and inactive phenotypes have also been observed for several uncharged single point mutations of N139 or its equivalent in C*c*O from *Paracoccus denitrificans* and *Escherichia coli*. A summary of N139 mutants and their phenotypes is presented in Table 1. For example, N139A is decoupled but retains significant oxidase activity, while both pumping and oxidase activity are compromised in N139V.[25,27] These altered phenotypes suggest that N139 plays a special role in the uptake of protons. Understanding the physical and molecular basis of this residue's function is likely to provide insight into the pumping mechanism itself.

The three-dimensional (3D) structure of C*c*O is known from several high-resolution crystallographic structures (Fig. 1).[7,28–30] Residue N139 lies approximately one-quarter of the way in what appears to be the narrowest portion of the D-channel. It is separated from E286 by a water-filled cavity making up the rest of the D-channel.[29,30] The conformation of N139 in the crystallographic structures appears to interrupt the hydrogen-bonded network of water in the D-channel [Protein Data Bank (PDB) ID codes 1M56 and 2GSM].[29,30]

The hydration of the D-channel is of high functional relevance. The relatively non-polar character of the D-channel, together with the absence of titratable residues between D132 and E286, indicate that water is required for proton uptake up to E286. The long-range translocation of protons can occur by a Grotthuss mechanism[31,32] involving the relay

Table 1. Selected D-channel point mutations and resultant oxidase activity and proton pumping

Organism	Mutation	Activity (%)	Pumping	Reference
R. sphaeroides	D132N, A	<5	No	12
R. sphaeroides	D132N/N139D	~20	Yes	17
R. sphaeroides	N139D	150–300	No	20
R. sphaeroides	N139A	~36	No	[a]
R. sphaeroides	G204D	~2	No	22
R. sphaeroides	N207D	~100	No	23
R. sphaeroides	N139T	~40	No	26
P. denitrificans	N131D (N139D)[b]	100	No	25
P. denitrificans	N199D (N207D)	50	No	25
P. denitrificans	N131V (N139V)	6	No	25
P. denitrificans	N131V (N139V)	3	—	27
P. denitrificans	N131A (N139A)	11	No	27
P. denitrificans	N131C (N139C)	86	No	27
P. denitrificans	N131Q (N139Q)	52	Reduced	27
E. coli	N142D (N139D)	~50	Yes	9
E. coli	N142V (N139V)	22	Reduced	9
E. coli	D135N (D132N)	45	Reduced	9
E. coli	D135N/N142D (D132N/N139D)	33	Yes	9

[a] R. B. Gennis, personal communication.
[b] *R. sphaeroides* numbering provided in parentheses.

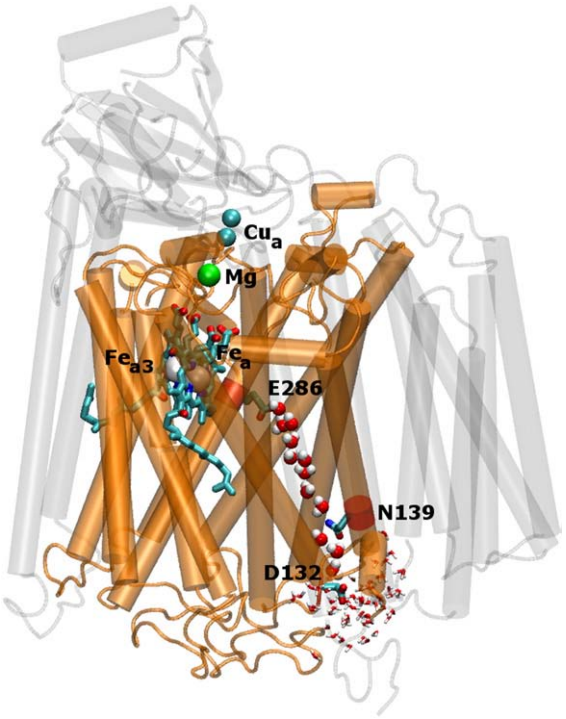

Fig. 1. C*c*O of *R. sphaeroides*. Subunit I, which contains the D-channel and the binuclear centre, is highlighted in orange. The simulated system consists of subunit I together with metal cofactors and a cap of 57 water molecules added to the entrance of the D-channel. Also highlighted are key residues of the D-channel, which extends from D132 to E286. This conformation of the D-channel is taken from one of our simulations with N139 in the open state and a continuous chain of 12 water molecules providing a proton pathway from the entrance of the D-channel to E286.

or transfer of H nuclei between hydrogen-bonded water molecules. In this process, small-amplitude fluctuations of H-bearing atoms occurring in the picosecond-to-nanosecond time range give rise to long-range displacement of protons. For this reason, elucidation of the molecular basis of proton transport in biomolecules requires a high level of detail that is not available from static crystallographic structures but is amenable to computer simulations. Accordingly, computational studies have played an essential role in deciphering the molecular mechanism of proton transport, both in bulk water[33] and in hydrogen-bonded water chains or water wires found in the interior of biological pores such as the gramicidin channel.[34] These studies have shown that the presence of hydrogen bonds, although not always sufficient,[35,36] is necessary for water-mediated proton relay.[34] In turn, this suggests that the interruption of the water chain at N139 in the crystallographic structures of the D-channel is functionally significant.

Several computational studies have examined the hydration of the D-channel and its relation to function. Two studies predating the detection of water in the cavity in crystallographic structures predicted the presence of a chain of water molecules in the D-channel.[13,15] Conformational isomerization of E286 was examined systematically with free-energy calculations and was shown to be compatible with the shuttling of protons between water molecules in the D-channel and in the active site of the enzyme.[14] In addition, it has recently been proposed that E286 could function as a proton valve.[37] Other simulation studies aimed at identifying possible proton pathways examined water movement in the enzyme's cavities.[38–40] Most recently, several computational studies have begun to probe the mechanism of water-mediated proton relay up to E286, either from N139[41,42] or from the entrance.[43] The latter study showed the presence of a free-energy barrier opposing proton uptake at the location of N139. However, while the existence of a conformational gate involving N139 has been postulated,[27,43] the putative gating mechanism and its functional role have not been investigated.

In addition, the equilibrium hydration state of the D-channel remains unclear. In two recently determined crystal structures of *R. sphaeroides* C*c*O, the D-channel contains 10 to 11 water molecules.[29,30] Although crystallographic models are an invaluable tool for providing insight into the structure and function of proteins, detecting the location of buried water is not without difficulties.[44–46] Crystal structures of the same protein can vary in the number and location of water molecules, even within the same crystal subunit, and the mobility of water in some protein cavities may lead to underestimates of their hydration state. Simulation studies have suggested that the D-channel may contain as many as 16 water molecules.[39]

Given the role of water in mediating proton relay, a detailed assessment of D-channel hydration is needed to understand the structural factors modulating proton uptake. In particular, understanding the effect of the apparent bottleneck in the D-channel at position 139 on the structure and fluctuations of the water wire is a prerequisite to elucidating the functional role of N139 and the molecular origin of decoupled and inactive phenotypes in single point mutants of this residue.

Here we examine the structural and thermodynamic basis for the functional hydration of the D-channel in a systematic way. Specifically, we use molecular simulations to gain insight into the functional role of N139 and two single-point mutants, N139A and N139V. We first analyze the spatial distribution of water molecules in various hydration states of the D-channel and compare these results to the location of water molecules in the available crystallographic structures. The free-energy profile for the conformational isomerization of N139 reveals the presence of a metastable "open" state compatible with the presence of a chain of water molecules. The water occupancy of this hydration bottleneck is estimated using free-energy calculations for transferring water molecules from the D-channel to bulk solution. These calculations

are repeated for N139A and N139V mutants. We then discuss methodological aspects of free-energy simulations, the thermodynamic basis for the hydration of the D-channel, and the functional implications of our findings. Finally, the theoretical and methodological basis for large-scale free-energy simulations using distributed replica sampling is presented.[47]

Results

The overall view of the system is shown in Fig. 1. The simulations contained subunit I together with a hemispheric cap of water molecules at the entrance of the D-channel. In this systematic comparative study, we examine the structure of the proton pathway in different conformational and hydration states of the D-channel, successively in the wild type and in two single-point mutants of N139.

Wild-type protein

Conformational gating of N139

In the two recent crystallographic structures of C*c*O,[29,30] the side chain of residue N139 is found in a conformation that appears to occlude the D-channel (Fig. 2a). We used free-energy simulations to identify other energetically favourable rotameric states of this side chain. Conventional umbrella sampling and large-scale distributed replica sampling simulations were used respectively in the absence and the presence of water molecules in the D-channel. In the absence of D-channel water molecules, we computed the 2D free-energy map or potential of mean force (PMF) for the conformational isomerization of N139 using 2D umbrella sampling where both χ_1 and χ_2 were biased (Fig. 3a). From these results, it appears that the crystallographic rotamer is a metastable state related to the global minimum by a 180° flip in the χ_2 torsion (Fig. 3a), suggesting that the conformation of the side chain may have been misassigned relative to χ_2 in the crystallographic

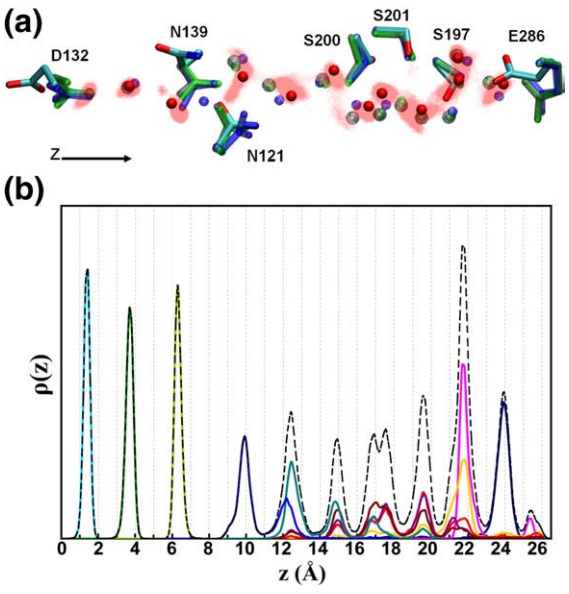

Fig. 2. (a) Spatial distribution of water oxygen atoms from a 5-ns simulation overlaid onto a representative snapshot of 12 water molecules (red) and the crystallographic water positions from Svensson-Ek *et al.* (blue),[29] and Qin *et al.* (black, green).[30] Key residues are depicted, including D132 at the opening of the D-channel, N139 at the bottleneck, and E286 at the end of the D-channel. (b) Individual axial distributions of 11 water molecules (continuous lines) and total O atom distribution (dashed line) from a 600-ns free-energy simulation. Movement of water in the wider part of the channel ($z > 8$ Å) results in multimodal distributions.

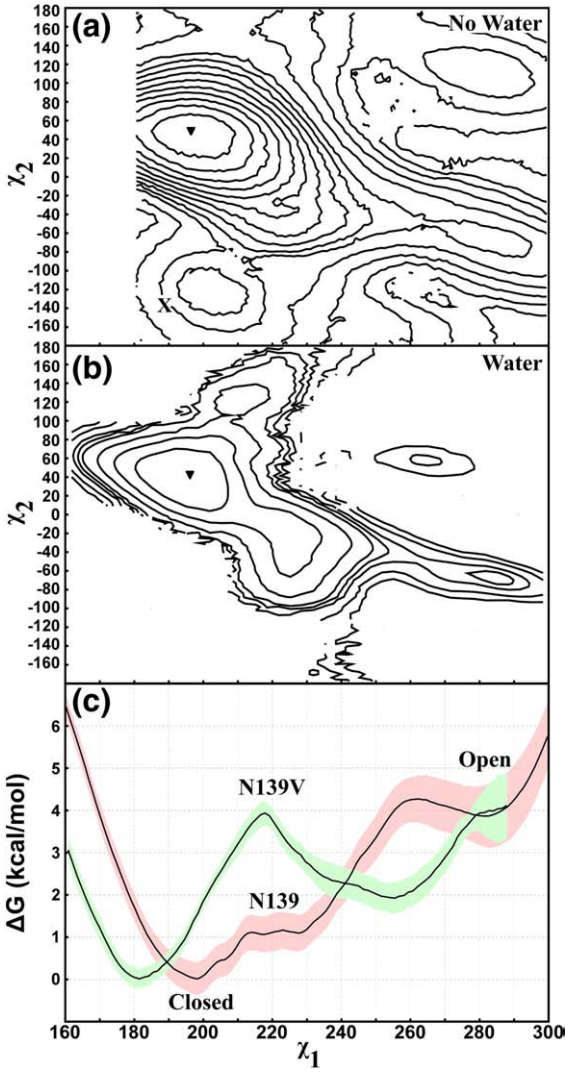

Fig. 3. PMF free-energy profiles for the rotation of N139 and N139V side chains. (a) PMF of side chain torsions χ_1 and χ_2 of N139 from 2D umbrella sampling in the absence of water in the D-channel. The crystallographic conformation is marked by X. (b) 2D PMF of N139 from 1D umbrella sampling along χ_1 in the presence of water. Contour spacing is 1 kcal/mol, with an additional contour at 4.5 kcal/mol in (b). (c) PMF of N139 (red error bars) and N139 V (green error bars) from umbrella sampling along χ_1.

model. A misassignment is possible given the difficulty in distinguishing between the electron densities of the oxygen and nitrogen moieties of the amide group at crystallographic resolutions. This discrepancy in the χ_2 angle of N139 is also present in the two 1.8 Å structures of bovine C*c*O (1V54 and 2DYR)[44,48], one of which (1V54) is consistent with our findings and the other (2DYR) with the 2.3 Å structure of Svensson-Ek *et al*.[29] Consistent with our findings, a 180° χ_2 flip in the latter structure is also recommended by the online program MolProbity.[49,50] The global minimum at $(\chi_1,\chi_2)=(195°, 41°)$ is henceforth referred to in this study as the "closed" conformation of N139. In addition, the PMF reveals a metastable rotamer at $(\chi_1,\chi_2)=(285°, -70°)$, in which the amide group of N139 has moved away from the channel axis, thus opening up the channel (Fig. 2a). To further characterize these newly identified rotameric states, we performed a more rigorous 1D umbrella sampling simulation with water molecules in the D-channel, using distributed replica sampling to induce a random walk along χ_1. Two- and one-dimensional PMF profiles obtained from this simulation are shown in Fig. 3b and c, respectively. The addition of water (Fig. 3b) preserves the qualitative features of the PMF obtained in the dry channel (Fig. 3a), but results in an approximately twofold reduction in free-energy differences. The fact that the crystallographic conformer was not sampled despite the length of this multiple-replica simulation (374 ns) suggests that the X-ray conformation of N139 remains disfavoured in the presence of water. The free energy of the putative open state, relative to the global minimum, drops from approximately 9.0 to 4.0 kcal/mol upon hydration of the D-channel due to hydrogen bonds with water molecules. We now consider the effect of the conformational isomerization of N139 on D-channel hydration.

D-channel hydration

Figure 4a shows the average O atom distribution of 11 water molecules along the axis of the D-channel computed from a 5-ns simulation in which the side chain of residue N139 retained a closed conformation. This distribution is largely consistent with the location of water molecules in the crystal structures (shown as symbols in Fig. 4). In particular, both the computed distribution and the crystallographic models contain a large gap or interruption of the water chain at the location of residue 139 ($4<z<9$ Å). Beyond this defect ($z>9$ Å), the computed and experimentally-determined spatial locations of water molecules are in good overall agreement, with some discrepancies between 14 and 20 Å that are also found in the open state of the channel, as discussed in more detail below.

Below the water gap ($z<4$ Å), the computed distribution is in good agreement with the two high-resolution crystal structures at $z\approx3.5$ Å. However, the calculated distribution differs from the crystal structures at $z=1.5$ Å due to a discrepancy in the

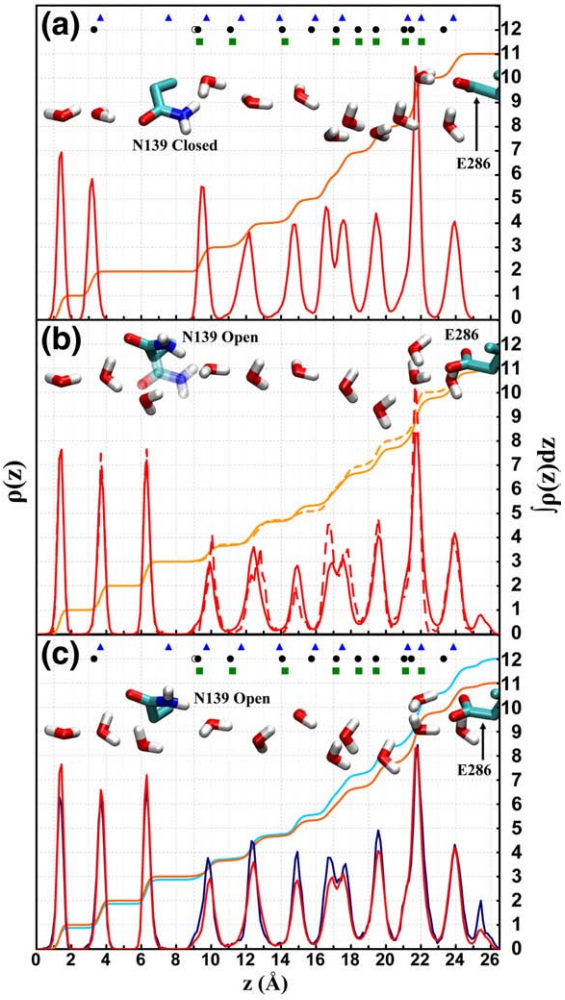

Fig. 4. Axial distribution and cumulative occupancy of water in the D-channel from MD simulations with (blue) 12 and (red) 11 water molecules. (a) Closed state of residue N139 in a 5-ns simulation of 11 water molecules. (b) Open state of N139 in two simulations of 5 ns (dotted line) and 600 ns (continuous line) with N139 restrained in the open position. Both simulations contained 11 water molecules. (c) Open state of N139 in two simulations with 11 and 12 water molecules. Representative simulation snapshots and axial positions of crystallographic water molecules from Svensson-Ek *et al.* (triangles)[29] and Qin *et al.* (circles, squares)[30] are also shown. The open circle denotes the location of a possible 12th water molecule from Qin *et al.*

conformational state of D132 (Fig. 2a). In the crystallographic structures, the side chain of D132 resides in the D-channel. A change in side-chain conformation early in our simulations led the carboxylate group to point into the bulk solution. This group was replaced by a spatially-localized water molecule at the channel entrance, leading to full hydrogen-bond connectivity between water in the bulk and in the D-channel (Fig. 2a, bulk not shown). In subsequent comparisons of computational and crystallographic models, only the portions of the channel located at $z>2$ Å will be considered.

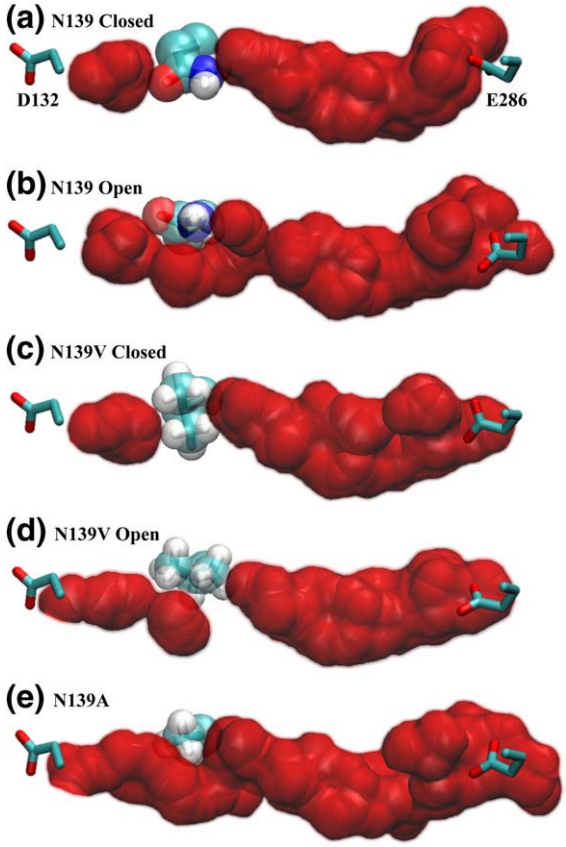

Fig. 5. Inner surface representation of the D-channel including residues D132, E286, and (a) wild-type, closed state; (b) wild-type, open state; (c) N139V mutant, closed state; (d) N139V mutant, open state; (e) N139A mutant. Surfaces were generated from simulation snapshots using VMD.[51]

The closed conformation of N139 precludes the formation of a hydrogen-bonded chain of water molecules (Fig. 5a). However, the metastable conformation of N139 at $(\chi_1,\chi_2)=(285°, -70°)$ opens up the D-channel (Fig. 5b). Since hydrogen bonding is required for proton relay,[34] the remainder of our simulations were performed with N139 in its open conformation. Figure 4b depicts 11-water-molecule distributions computed from two independent simulations in which the side chain of residue N139 was restrained to its open conformation. The opening of N139 creates a cavity at $z \approx 6$ Å that is filled up with a water molecule from the top of the D-channel ($z > 9$ Å). Despite this rearrangement, the water distribution in the rest of the channel is consistent with that obtained in the closed state of N139 and is conserved over a significant range of time scales, from 100 ps (data not shown), to 5 ns, and on to the longer cumulative time of the free-energy simulations (600 ns) (Fig. 4b). Moreover, adding a 12th water molecule to the D-channel in its open state also preserves the average water distribution, with changes limited to the magnitude of the peaks above $z=9$ Å (Fig. 4c). The consistency of water distributions across time scales and varying hydration states suggests that these peaks represent the equilibrium positions of the water molecules, with the only significant difference being a local change in water density at residue N139 upon opening and closing of the channel.

A comparison of axial water distributions obtained from simulations of 11 and 12 water molecules to the crystallographic structures by Svensson-Ek *et al.*[29] and by Qin *et al.*[30] is shown in Fig. 4c. The two structures present in the crystal subunit of Qin *et al.* differ in their resolution of bound molecules, including water. We refer to the structure described in the article[30] as structure QA, and to the alternate structure as structure QB. The three crystallographic structures differ in the number and location of water molecules in the D-channel. In particular, QA contains up to two more water molecules than both QB and the Svensson-Ek structure, henceforth denoted SE. For a detailed comparison, we have divided the D-channel into six sections as listed in Table 2.

Beyond the entrance of the D-channel ($z > 2.5$ Å), the computed water distributions are consistent with experimental data both locally and globally. In particular, the running total remains within one water molecule of the crystallographic models through consecutive sections of the D-channel (Table 2). Differences in hydration of the lowest three sections ($z < 8$ Å) have already been discussed above. In the next section ($8.0 < z < 14.5$ Å), water molecules are located within 1 Å from each other in the various models with one exception, which is due to the presence of an extra water molecule at $z \approx 9$ Å in structure QA. However, this water molecule is not identified as being in the D-channel in the article by Qin *et al.*,[30] and occupies a cavity that is off-axis from the other water molecules in the channel. In addition, due to the absence of subunits III and IV in the crystal structures of Qin *et al.*, this water

Table 2. Comparison of computed and crystallographic water distributions in the D-channel

	No. of water molecules per region				Total water molecules			
Region (Å)	Computed	QA	QB	SE	Computed	QA	QB	SE
20.5–25.5	3	3	4	3	3	3	3	3
14.5–20.5	3–4	5	4	3	6–7	8	7	6
8.0–14.5	2	2	2	2	8–9	10	9	8
5.0–8.0	1	0	0	1	9–10	10	9	9
2.5–5.0	1	1	0	1	10–11	11	9	10
0–2.5	1	0	0	0	11–12	11	9	10

molecule is but one residue away from the outside of the protein. This extra water molecule is likely due to the channel's proximity to the protein exterior and is omitted from the remaining analyses. The next section of the D-channel ($14.5<z<20.5$ Å) holds different amounts of water in simulations with 11 and 12 water molecules: specifically, this "sponge-like" region accommodates 3 or 4 water molecules with hardly any change in the average peak distribution (Fig. 4c). This property is consistent with the variance in the number of water molecules (3 to 5) in different crystal structures in this region (Table 2). Despite this variability, the number of water molecules computed for this section matches the number found in structures QB and SE. Finally, computed water distributions near E286 at the top of the D-channel ($20.5<z<25.5$ Å) are in excellent agreement with all three structures, with two water molecules at $z \approx 21$–22 Å matching a sharp peak in the computed distribution (Fig. 4c).

Although the peaks in the computed axial distribution of water do not match exactly the placement of water in the crystallographic structures, the latter are found in regions of high density, as best shown in the 2D projection depicted in Fig. 2a. The overlap of most crystallographic water molecules with the computed water density provides a strong indication that the differences between the various models are insignificant. Additionally, the region that demonstrates the most variance when the models are compared (Table 2) is seen here to have the most mobile water molecules, suggesting a possible reason for this variation.

A detailed analysis of the axial spatial distribution of individual water molecules provides further insight into the mobility of water in the D-channel (Fig. 2b). From the channel entrance to approximately 8 Å, three discrete peaks define a "solid-like" region, in which water molecules do not interconvert. This solid-like region, which consistently contains 3 water molecules, corresponds to the narrower portion of the D-channel (Fig. 5b). By contrast, water molecules located in the wider segment of the channel ($8<z<26$ Å) readily diffuse and exchange positions with one another (Fig. 2b). This "fluid-like" region can accommodate at least 8 or 9 water molecules. Because of differences in the mobility of water in the fluid- and solid-like regions, removal of 1 water molecule from the 12-water hydration state produced significantly different distributions depending on which water molecule was removed (Fig. 6). Removal of a water molecule from the fluid region conserved the peak distribution and did not result in a defect (Fig. 6b). In the fluid region, D-channel water molecules replaced the extracted water molecule within 100 ps (data not shown). By contrast, removal of a water molecule from the solid-like region (at $z \approx 6$ Å) led to a defect that persisted over the rest of the 5-ns simulation (Fig. 6c).

The above results reveal the presence of a possible hydration bottleneck at residue N139 (near $z \approx 6$ Å)

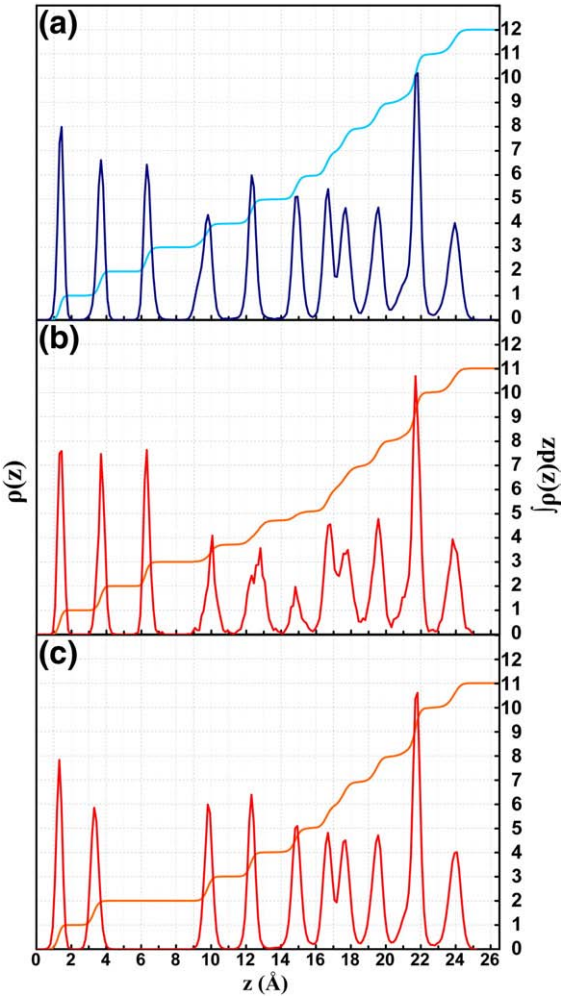

Fig. 6. Axial distribution of water in the D-channel from 5-ns simulations with (a) 12 water molecules, (b) 11 water molecules following removal of a 12th water molecule from the wider part of the D-channel ($8.0<z<25.0$ Å); and (c) 11 water molecules following removal of a 12th water molecule from the bottleneck ($z \approx 6.0$ Å).

even in the open state of the channel. However, the long time scale of water movement prevents the direct estimate by brute-force simulations of the likelihood that the narrow part of the D-channel is hydrated. Thus, the outcome of the above simulations depends on the initial hydration state of the channel's bottleneck. Although the initial placement of a water molecule at that location leads to a persistent hydrogen-bonded water chain in both 11- and 12-water hydration states (Fig. 6a and b), the gap created by the removal of that water molecule persists over long time scales (Fig. 6c). Hydration of the channel at N139 therefore emerges as the single most critical "link" in the ability of water to relay protons in the D-channel. In the next section, we present free-energy calculations to determine the probability of hydrating this apparent bottleneck, and therefore, of forming a water wire for the rapid uptake of protons.

Hydration free-energy simulations

The above results suggest that a proton wire may form in the open state of the D-channel with either 11 or 12 water molecules. Thermodynamic integration in four spatial dimensions[52] was utilized to calculate the free energy for extracting the water molecule located at the bottleneck of the D-channel successively in the 12- and 11-water hydration states of the channel. Distributed replica sampling[47] was used to perform efficient Boltzmann sampling of conformational space. Both simulations allowed the water molecule at $z \approx 6$ Å to undergo a random walk in an extra, non-physical spatial dimension, the w coordinate, between the fully inserted-state at $w=0$ and the fully-extracted state at $w=20$ Å. Figure 7 shows the distribution of water molecules in the fully inserted and fully extracted states in each of the two simulations (note that the test water molecule is not included in the $w=20$ Å distributions).

As expected, the distribution of water molecules in the fluid region of the D-channel was largely unaffected by the removal of the 12th and 11th water molecules (Fig. 7). In the free-energy simulation for the extraction of the 11th water molecule, the 10 remaining D-channel water molecules did not bridge the resulting defect in the proton uptake pathway (Fig. 7b). However, the defect resulting from the removal of the 12th water molecule was transiently filled by the two water molecules usually located at $z=1.5$ Å and $z=3.5$ Å (Fig. 7a). Qualitatively, the resulting distribution resembles that of an 11-water-molecule wire capable of proton relay.

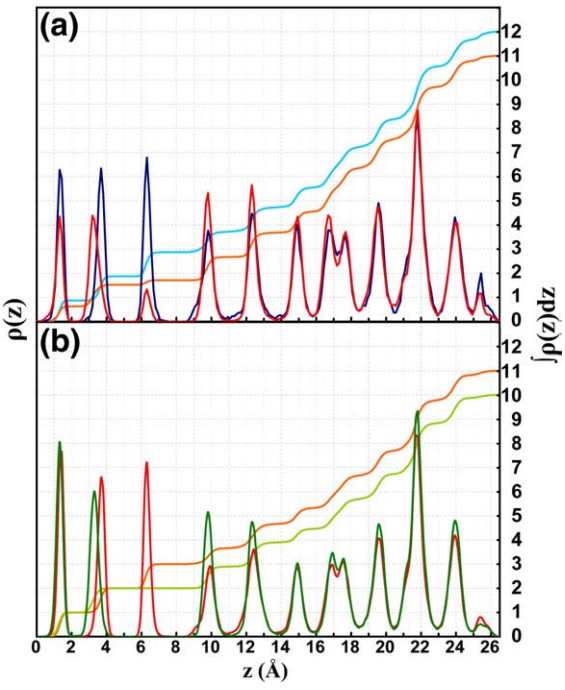

Fig. 7. Axial distribution of water from free-energy simulations in which a test water molecule is reversibly extracted from the bottleneck of the D-channel. (a) Reversible transfer between (blue) 12 and (red) 11 water molecules. (b) Reversible transfer between (red) 11 and (green) 10 water molecules.

The calculated free-energy change for the transfer of the 12th and 11th water molecules at $z \approx 6$ Å into vacuo was found to be 7.4±0.05 and 7.7±0.05 kcal/mol, respectively. A calculated standard-state correction value of 0.67 kcal/mol and the hydration free energy of a TIP3P water molecule of -6.7±0.04 kcal/mol[52] were added to obtain the total free energy of transfer from the D-channel to bulk solution, yielding estimates of −0.24±0.06 and 0.07±0.06 kcal/mol, respectively. These values correspond to the probability of finding the 12th and 11th water molecules in the bottleneck of the D-channel, rather than in bulk solution, of $p_{in}/p_{bulk}=\exp(-\beta\Delta G)=0.67$ and 1.1, respectively. The corresponding fractional occupancies of the N139 bottleneck by water are $x_{in}=p_{in}/(p_{in}+p_{bulk})=40$ and 53%.

Although the 11-water-molecule distributions in Fig. 7a and b are qualitatively similar, it would be incorrect to assume that they correspond to the same thermodynamic state and that the difference in occupancy of peak 3 is necessarily due to insufficient convergence. The fully extracted 12-water state and the fully inserted 11-water state are thermodynamically distinct. This is because the fully inserted states do not allow for partial occupancy of the defect, whereas fully extracted states do. Thus, in the inserted state of the 11-water simulation, the defect is occupied 100% of the time by construction. By contrast, the extracted state of the 12-water simulation corresponds to an 11-water state with ~20% occupancy of the bottleneck. The latter result is not a rigorous estimate due to infrequent exchanges of water molecules in the solid-like bottleneck region (see Discussion). Nevertheless, this value is commensurate with the more rigorous estimates obtained above from transfer free-energy calculations.

Valine mutant

The above study provided insight into the thermodynamic basis of hydration of the D-channel and the number and spatial distribution of water molecules in the wild-type protein. In this and the following sections, we describe analogous simulations performed on valine and alanine mutants of N139 in order to characterize the ability of these mutants to host a functional proton wire.

Spatial distribution of water

Representative snapshots and axial water distributions obtained from two 5-ns simulations with N139V in open and closed states are shown in Fig. 8. Initially, the side chain of N139V was constructed in a putative open conformation by analogy with the open state of the wild-type protein (Fig. 8b). However, on longer simulation time scales (>5 ns), residue N139V occasionally reached a closed state (Figs. 5c and 8a). To characterize the conformational

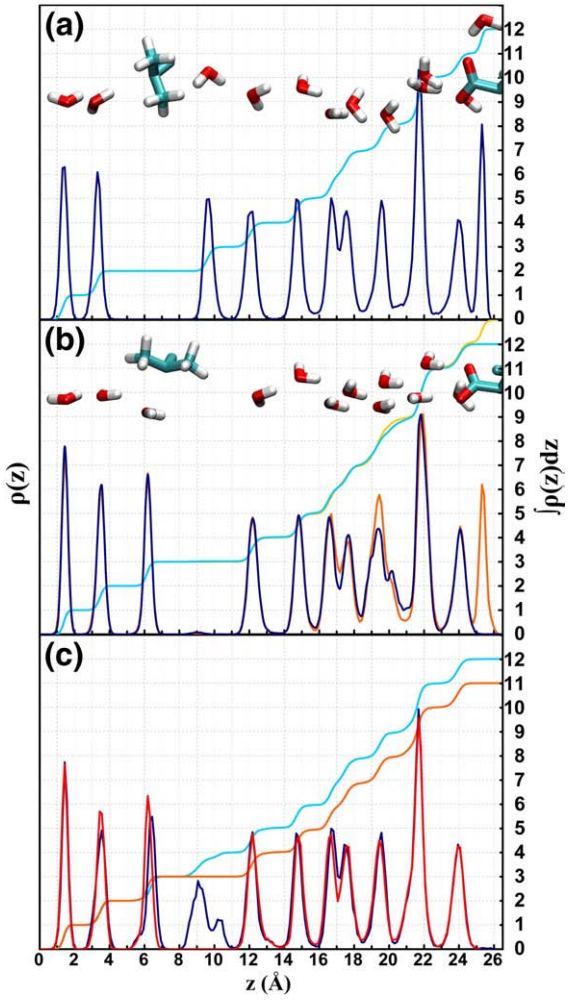

Fig. 8. Axial distribution of water in the N139V mutant. (a) Five-nanosecond simulation in the closed state of N139V. (b) Five-nanosecond simulation in the open state of N139V with (blue) 12 or (orange) 13 water molecules. (c) Free-energy simulation from (blue) 12 to (red) 11 water molecules.

properties of N139V, we computed the PMF for the χ_1 torsion of N139V using umbrella sampling (Fig. 3c). The closed state of N139V is favoured by approximately 2 kcal/mol over the open state. Despite this preference, the closed state of N139V was not investigated further since, like the wild type, it corresponds to steric occlusion of the D-channel and is therefore incapable of mediating proton uptake (Figs. 5c and 8a). To examine whether or not a water wire can form in the open state of N139V, all further simulations were performed with N139V restrained to its open conformation ($\chi_1 = 255°$).

Simulations to determine the hydration state of the N139V mutant were performed successively with 10, 11, and 12 water molecules in the D-channel. In all cases the simulation revealed a defect at $z \approx 9$ Å (Fig. 8b). An additional 5-ns simulation was performed with a 13th water molecule initially placed at $z \approx 9$ Å. The resulting distribution (Fig. 8b) also displays a defect at $z \approx 9$ Å, with the 13th water molecule residing within the fluid-like region of the D-channel, producing a new peak at $z \approx 25.5$ Å while leaving the remaining peak locations essentially unchanged. Like the wild-type protein, the distribution in the solid-like region is largely unaffected by the number of water molecules in the D-channel, and the fluid-like region can accommodate up to 9 water molecules without significant changes in the spatial distribution of water. However, in the valine mutant, the defect at $z \approx 9$ Å does not readily fill up within 5-ns even if the D-channel is saturated with excess water molecules. Upon inspection of the D-channel inner surface, it appears that this defect is caused by a steric occlusion of the D-channel analogous to that of the closed state of residue 139 (Fig. 5d).

Due to the persistence of the defect, an additional 5-ns simulation was performed with 12 water molecules, one of which was restrained at $z = 9$ Å. This procedure produced a water distribution consistent with the capacity to relay an excess proton (data not shown). In the next section we estimate the probability of hydrating this apparent hydration bottleneck.

Hydration free-energy simulation

Unlike the wild type, in which a water molecule placed at the bottleneck remains there in both 11- and 12-water molecule states, in the N139V mutant a defect persists even with 13 water molecules. However, a hydrogen-bonded water wire can be formed with 12 water molecules provided the water molecule at $z \approx 9$ Å is restrained to that region. Thus, a flat-bottomed potential was used to restrain the test water molecule to that region at all values of w. This contrasts with our simulations of the other protein systems, in which such a bias was only used to restrain the test water molecule to the binding region at large values of w (the "funnel", see Theory). We performed free-energy simulations to determine the probability that a water molecule would be in this region of the D-channel rather than bulk solution in the presence of such a restraint.

The D-channel contained 12 water molecules in the fully-inserted state ($w = 0$) and the water molecule at $z \approx 9$ Å was subjected to a random walk in w coordinate space. Figure 8c depicts the axial distribution of water molecules in the coupled and decoupled states. Water molecules in both fluid-like and solid-like regions of the D-channel are largely unaffected by the removal of the 12th water molecule. The water distributions of inserted and extracted states are nearly identical, with only a slight shift of peak 3 toward peak 2.

The free-energy change for the transfer of the water molecule at $z \approx 9$ Å to vacuo was found to be 5.7 ± 0.1 kcal/mol. Applying the same standard state correction and water hydration free-energy values as before, the total free-energy change was calculated to be -1.8 ± 0.1 kcal/mol. Thus, the probability of transferring a 12th water molecule from bulk solution to the bottleneck of the D-channel is appro-

ximately 4%. It should be noted that this result does not take into account the work applied to restrain the water molecule near $z=9$ Å. An underestimate of this work was obtained by analyzing the axial distribution of this water molecule in the presence of the flat-bottom restraint. The work exerted by the biasing potential on the system was calculated to be −0.14 kcal/mol. This correction results in a transfer free energy of -2.0 ± 0.1 kcal/mol, leading to an upper-bound estimate of 3% for the water occupancy of the bottleneck.

Alanine mutant

Spatial distribution of water

Finally, we examined the probability of forming a hydrogen-bonded water wire in the D-channel in the N139A mutant. Five-nanosecond simulations suggested that 12 water molecules were sufficient for forming a water wire (Fig. 9a). The axial distribution of water molecules differs at the top of the D-channel, with the commonly observed double peak at $z\approx21.5$ Å split into two peaks. An extra peak at $z\approx25.5$ Å is also visible in the distribution. This peak was observed previously in the N139 V mutant with 13 water molecules and in the closed state (Fig. 8a and b), and likely corresponds to a metastable state that can be promoted by the crowding of water molecules in the D-channel.

As in the wild-type protein, removal of a water molecule from the solid-like region of the D-channel at $z\approx6$ Å produced a defect that persisted for the duration of the simulation (Fig. 9c). In addition, removal of a water molecule from the fluid-like region of the channel led to the transfer of water molecules from the solid region to the fluid region, producing defects at $z\approx6$ Å and $z\approx1.5$ Å (Fig. 9b). Again, this result suggested that the region of the D-channel near residue 139 is the most likely region to be dehydrated and is the weakest link in forming a water wire capable of relaying protons. Therefore, we computed the free energy for hydrating the bottleneck at N139A to determine the probability of forming a putative water wire in the D-channel.

Hydration free-energy simulation

In the preliminary simulations of the N139A mutant, a defect was formed at $z\approx6$ Å independently of which water molecule was removed, suggesting that this is a likely region for local dehydration. Free energy simulations were performed with 12 water molecules and with the test water molecule located at $z\approx6$ Å undergoing a random walk in the fourth dimension. The free-energy change for the removal of this water molecule into vacuo was found to be 8.8 ± 0.07 kcal/mol. After addition of the standard-state correction and TIP3P hydration free energy, the free-energy change for the transfer of a water molecule from the D-channel to bulk is 1.2 ± 0.08 kcal/mol. Thus, the probability that a 12th water molecule occupies the bottleneck of the D-channel rather than bulk is $p_{in}/p_{bulk}=7.7$, indicating that the D-channel is fully hydrated 88% of the time.

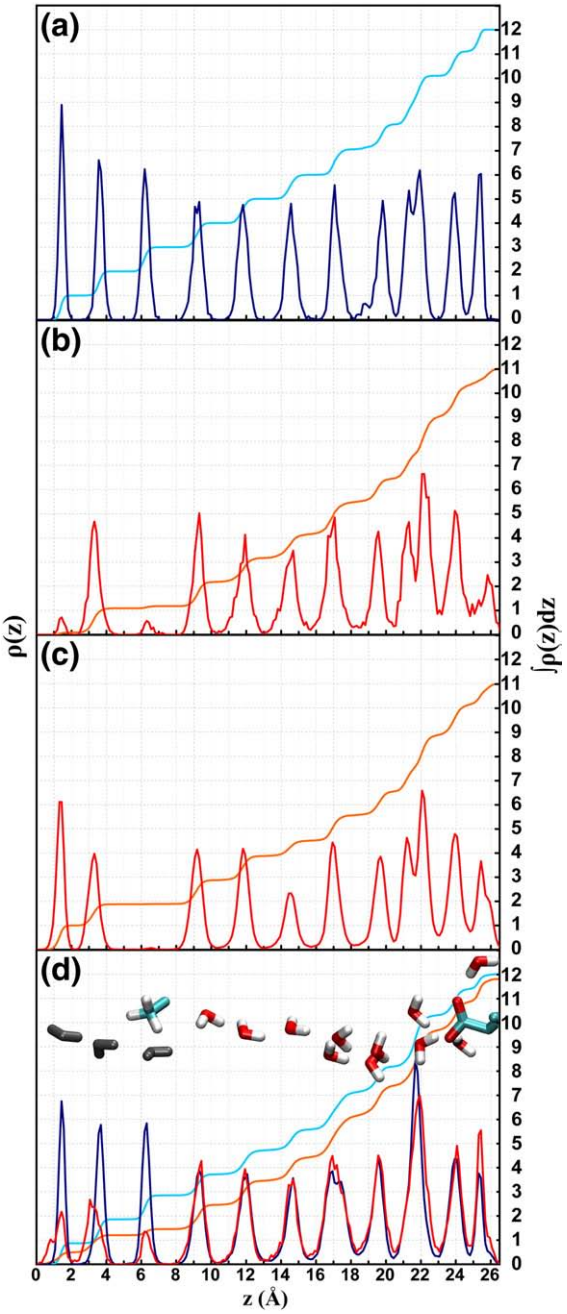

Fig. 9. Axial distribution of water in the N139A mutant from 5-ns simulations with (a) 12 water molecules; (b) 11 water molecules after removal of 1 water molecule from the wider part of the channel; (c) 11 water molecules following removal of a 12th water molecule from the bottleneck. (d) Water distributions before and after reversible removal of the 12th water molecule from the bottleneck in free-energy simulations. In the latter simulations, water molecules spontaneously diffused from the solid-like region into the large channel cavity and were replaced by bulk water (grey). The fully-extracted distribution (red) is the sum of the D-channel and bulk water distributions.

Figure 9d depicts the distribution of water molecules in both inserted and extracted states. As predicted by the preliminary 5-ns simulations, in the fully-extracted system the remaining water molecules are insufficient for forming a water wire capable of relaying protons. Upon removal of the 12th water molecule, one of the water molecules from the solid-like region of the channel transferred to the fluid-like region. One water molecule remained mostly in the solid-like region and transiently occupied all three peak locations. This confirms that 11 water molecules are insufficient to form a proton wire and suggests that in the case of the alanine mutant, there is no solid-like region (Fig. 5e).

Even more interesting is the behaviour of the bulk water molecules in our simulation. At long time scales, bulk water molecules entered the D-channel in the fully-extracted state (Fig. 9d, small population at 160 and 375 ns, significant presence after 500 ns). The fully-extracted water distribution in Fig. 9d is the sum of the D-channel and bulk water distributions. Bulk water molecules intrude as far as 9 Å, further indicating a lack of a solid region, and represent an addition of nearly one water molecule to the fully-extracted distribution. This phenomenon did not occur with the other protein systems on the time scales we simulated. Along with reestablishing a water wire capable of proton relay, the diffusion of bulk water molecules into the D-channel coincided with a reduction in the calculated free energy of transfer. If the simulation was allowed to proceed over much longer time scales, we expect this trend to continue, with the calculated free energy of transfer approaching zero due to the replacement of the test water molecule by bulk water. The free-energy value obtained before the intrusion of bulk water predicts a fully hydrated D-channel in the N139A mutant. This result is corroborated by the spontaneous movement of bulk water molecules into the D-channel at longer simulation time scales.

Discussion

Methodological considerations

Using free-energy simulations to estimate the hydration of protein cavities

We have used a general method to compute the free energy of water transfer into a protein cavity. A rigorous and practical way to couple multiple simulations on an arbitrary computing platform, distributed replica sampling,[47] was combined with thermodynamic integration along a non-physical, "fourth" spatial dimension, which was previously shown to be an effective reaction pathway for the calculation of chemical potentials.[52] Using distributed replica sampling, we achieved a random walk along the reaction coordinate to enhance the efficiency of the calculation and reduce the risk of systematic sampling errors due to ruggedness in other degrees of freedom.[53] This general method is also suitable to calculate the affinity of a molecular ligand for a protein binding site.[54,55] Despite this approach, spatial restraints had to be applied in order to reduce the size of conformational space. These restraints were necessary to achieve convergence in the free-energy calculation because of sampling bottlenecks arising from the conformational isomerization of the protein, from the exchange of water molecules within the cavity and with the bulk, and from the vanishing size of the water molecule as it is decoupled (extracted) from the protein.

The principal challenge of biomolecular simulations is that biomolecules evolve over a wide range of time scales. The successful calculation of free-energy changes rests on the separation between the time scale of the simulation and that of the process under consideration. Three cases may arise. If the simulation is much longer than the slowest relevant exchanges or relaxation times in the system, brute-force sampling is adequate. Inversely, if the slowest events of relevance occur on time scales much longer than that of the simulation, biased sampling or separate free-energy calculations can be performed for distinct substates of the system. However, processes involving rearrangements occurring on the time scale of the simulations present difficulties. Statistical significance is hard to achieve due to an insufficient number of infrequent transitions between different substates of the system. In such cases, artificial restraining forces may be applied to confine conformational sampling to predefined regions of phase space, thus facilitating the application of free-energy calculations to distinct substates.[56]

In the present study, such challenges resulted from the large size of the protein cavity and from the fact that the time scale of the simulation is commensurate with the relaxation rate of amino acid side chains in the bottleneck and with the rate of exchange of water between the bottleneck and the wider part of the channel or bulk water. Water dynamics in the D-channel is heterogeneous: in the larger part of the cavity, water molecules readily exchange with one another on nanosecond time scales (Fig. 2b), whereas spontaneous exchanges with the bottleneck, like conformational gating, may take hundreds of nanoseconds or more. The infrequent opening/closing of the bottleneck in the wild-type and the N139 V mutant was resolved by holding the bottleneck in the open state and by performing separate calculations for the gating itself. Due to slow water exchange with the bulk, multiple hydration states of the D-channel also had to be considered separately. In addition, the need to circumvent infrequent exchanges of water molecules between the bottleneck and other regions of the channel was resolved by imposing artificial restraints upon the test water molecule at residue 139. This strategy was justified because the narrower part of the channel is the most likely location of a defect in all three cases studied. The spatial restraint

avoided the need to sample all possible locations for insertion of the test water molecule in the large cavity of the D-channel.

Finally, in free-energy calculations where the effective size of the ligand decreases as it is artificially pulled out of the protein, the increase in the number of ways to insert a vanishingly small molecule in a protein cavity with a rugged shape also results in an entropic sampling bottleneck. This is not an issue in calculations of hydration free energy, where the solvent rapidly closes in to occupy the space vacated by the solute molecule. However, in the glassy protein interior, a molecule of vanishing size can get trapped in interatomic crevices. Here this problem was addressed by using a "funnel-like" restraint designed to shrink the size of the 3D volume sampled by the test water molecule progressively as it is gradually pulled into the fourth dimension. A restraint of constant size is often utilized in the more conventional shifted-scaling or double-decoupling methods, where the ligand molecule is decoupled or annihilated from the binding pocket.[56–58] The harmonic biasing force constant is usually calculated based on the mean-square fluctuations of the ligand in the binding pocket computed in short unrestrained simulations which themselves present the burden of convergence.[59] The use of a funnel restraint at once removes the need for a force constant estimate and reduces the sampling bottleneck due to the vanishingly small ligand, as the restraint is designed to be non-interacting when the ligand is fully coupled to the protein and highly restrictive once the ligand is decoupled. The dual benefit of this funnel scheme was obvious in the present study, as attempts to obtain statistically converged results using constant-radius restraints failed even in simulations extending into hundreds of nanoseconds (results not shown).

Other computational methods for predicting water occupancy

Other free-energy simulation methods have been used in recent years to estimate the hydration state of buried protein cavities.[56–58,60–63] Commonly, thermodynamic integration is used with shifting-scaling[57] or double-decoupling[58] pathways, and up to three separate consecutive free-energy simulations are used to decouple successively (i) Coulombic, then (ii) attractive and (iii) repulsive Lennard–Jones interactions between ligand and protein. Compared to the fourth-dimension pathway used here, in which decoupling of all non-bonded interactions is performed at once, such staging may lead to a greater accumulation of error. Independently of the decoupling pathway, sampling methods based on generalized ensembles, such as distributed replicas, facilitate transitions in orthogonal degrees of freedom, thereby decreasing systematic sampling error.[53]

To avoid the need to specify the hydration state of a protein cavity *a priori*, grand-canonical Monte Carlo (GCMC) simulations can be used to compute the relative probabilities of multiple hydration states.[60,61,63] In GCMC simulations, a protein cavity of finite volume is in open equilibrium with a bulk reservoir, and water molecules can exchange between the two. This approach is useful when little information of the hydration state of a cavity is known, and it has been used in conjunction with binding simulations of molecular ligands, which may displace internal water molecules.[61] However, it is unclear that GCMC simulations would present an advantage in the rugged cavity of the D-channel, since they would also be limited by the long equilibration times in the bottleneck region.

Effect of water equilibration and relaxation on convergence

As discussed above, the convergence of free-energy simulations is intimately related to the inherent relaxation time scales of the system. In this study, convergence was gauged based on the decreasing statistical error about the mean of the free energy. However, it is not guaranteed that the statistically converged state is the true free-energy minimum and we cannot predict whether the computed free energy would not change upon further sampling. We used long simulation times extending to hundreds of nanoseconds together with multiple replicas and a random walk between inserted and extracted states in an attempt to minimize systematic sampling error.[54] A caveat to reaching such long time scales is that the simulation may begin to sample new phenomena with long relaxation times. A case in point is provided by the simulation of the alanine mutant. Although the calculation had apparently converged, the calculated free energy began to decrease after several hundred nanoseconds as bulk water spontaneously diffused into the D-channel, reducing the probability of reinserting the test water molecule (see Results and Fig. 9d). With even longer time scales extending into the microsecond range, the channel-to-bulk transfer free-energy should reflect the partial or total replacement of the test water molecule by bulk, eliminating the need for a non-physical pathway. These observations further underscore the importance of tailoring the methodology to the relaxation properties of the system.

Effect of protein conformational restraints on computed properties

This study rests on the assumption that structural differences of the single-point mutants of N139 considered, as well as the conformational fluctuations modulating the access of protons in the wild-type protein, are local. The spatial restrictions imposed on the protein in our finite-size simulation system allowed us to reach long simulation times, but may result in longer relaxation times and could affect computed equilibrium properties. Estimating the effect of spatial restraints made necessary by our finite-size system would require simulation where

the protein is allowed to move in an explicit hydrated lipid bilayer, as done by Olkhova *et al.* in their much shorter (nanosecond) simulation of C*c*O.[39] Unfortunately, such a large system would preclude the long time scales required to converge the free-energy calculations presented here. The work presented here is the result of a necessary compromise between spatial and temporal scales.

Nevertheless, the validity of the present set up is supported by the agreement between computed and crystallographic water positions and by the metastable nature of the open state of N139. Moreover, it is unlikely that the tertiary architecture of the D-channel, which rests on extensive contacts between several alpha-helices, would be affected significantly by the two single-point mutations considered here, since the packing of these helices does not involve residue 139. Examples of protein cavities created or enlarged by point mutations to smaller residues without perturbing the protein fold include non-polar mutants of T4 lysozyme.[46]

Thermodynamic basis of D-channel gating and hydration

Anatomy of a hydration bottleneck: cavity size, shape, and polarity

Both enthalpy and entropy play a role in the thermodynamic basis for cavity hydration.[63] The hydration state of a protein cavity depends not only on the polarity of the chemical groups lining its walls, but also on the size and the shape of the cavity. A polar cavity is more likely to be hydrated than a hydrophobic one of the same size and shape because of favourable interactions with water. In addition, the presence of other water molecules also contributes to a favourable enthalpic change, so that water occupancy is conditional upon the total number of water molecules already in the cavity.[46,56,63] The size of the cavity modulates the entropic component of the free-energy change. Due to the entropic cost of excluding the 55-molar solvent, even completely hydrophobic cavities of sufficient volume do contain water.[62,63] However, the shape of the cavity is also a factor, as seen in the present study, where the steric bottleneck of the D-channel also corresponds to a hydration bottleneck, since it is the most likely segment to be dehydrated. The trade-off between cavity size and polarity is also illustrated within the bottleneck itself. Although it is less polar than in the wild-type protein, the larger cavity created by the N139A mutation (Fig. 5e) results in higher water occupancy (90% *versus* 50%), whereas the narrow non-polar bottleneck from the N139V mutation (Fig. 5d) is predicted to be mostly dehydrated (<3%).

Overall hydration state of the D-channel

In principle, the hydration state of the bottleneck is coupled to that of the rest of the channel, since water molecules in the bottleneck can form hydrogen bonds with other water molecules in both the solid and fluid-like regions. In this study, we focused on the weakest link of the water wire due to its presumed functional significance, and we did not explicitly compute the relative probabilities of hydration states differing in the total number of water molecules. As a consequence, the above results are conditional to total hydration numbers of 10, 11, and 12 (8 and 9 in the fluid-like region of the D-channel).

The quantitative agreement between the two free-energy simulations of the wild-type protein (Fig. 7) suggests that the hydration state of the bottleneck is relatively insensitive to the hydration number in the rest of the channel. Specifically, we found that the water occupancy of the bottleneck is 53% and 40% whether 8 or 9 water molecules occupy the large fluid-like cavity. However, this result may not hold in different hydration states. This fractional occupancy could drop if there were fewer than 8 or increase if there were more than 9 water molecules. In the limiting case where the D-channel was devoid of water, a water molecule placed at the bottleneck would be entropically driven to the wider region. The question is then whether these two hydration numbers are representative of the equilibrium hydration state of the cavity. Given that the fluid-like region consistently contains 8 to 10 water molecules in the crystallographic structures, it is unlikely that fewer than 8 water molecules would ever reside in that region (see Table 2 for $z > 8$ Å). Saturation of the fluid region with more than 10 water molecules cannot be discounted on the basis of the present study. As a result, it is possible that bottleneck hydration rises beyond predicted values, increasing the likelihood of forming a water wire in the open state of the channel.

The fact that bottleneck occupancy should decrease with fewer water molecules in the wider part of the channel suggests that the relative free-energy difference of ~0.3 kcal/mol favouring the 11-water state over the 12-water state is due to remaining systematic sampling errors. Nevertheless, if we assume that the 10-, 11-, and 12-water states are the only three hydration states with significant populations (which is reasonable given the excellent overall agreement with the crystallographic structures), the probability that the open state of the D-channel contains a sufficient number of molecules to form a water wire is approximately 50%.

The above analysis suggests that although fluctuations of the hydration number in the wider part of the channel are likely, their effect on the hydration state of the bottleneck is moderate. Inversely, the fact that the distribution of water molecules in the fluid-like region in the two mutants considered in this study is nearly identical to that of the wild-type enzyme suggests that the hydration state of the fluid region of the D-channel is independent of the steric, chemical, and hydration state of the bottleneck. As a result, the hydration of the

bottleneck is a local property, which justifies *a posteriori* the fact that we considered different hydration states of different regions of the D-channel separately in the present study. The functional implications of these findings are discussed below.

Thermodynamics of conformational gating

The present study has revealed the existence of an open rotameric state of N139 that enables the formation of a pathway for the uptake of protons. The metastable nature of the open state is consistent with the fact that all the crystallographic structures model N139 in its closed state.

The gating free-energy calculations underscored the interplay of hydration and gating. Specifically, D-channel hydration results in the dramatic stabilization of the open state relative to the closed state (compare Fig. 3a and b). A comparison of the gating PMF profiles for the wild-type and the valine side chains suggests that the latter is more likely to be open (Fig. 3c). However, this conclusion may be premature in light of (a) the lesser computational effort devoted to the N139 V gating calculation and (b) uncertainties due to the slow relaxation of water in the bottleneck, which required a systematic study of the open state in the first place. Although water is sterically excluded from the bottleneck in the closed state of residue 139, the population of the open state depends in principle on the hydration state of the bottleneck. As a result, convergence of the free-energy profiles for conformational isomerization of N139 and N139V depends on equilibration of water in the bottleneck. Although the N139 bottleneck was partially hydrated in the open state during the gating simulations, this is not a guarantee that equilibrium has been attained. A systematic examination of gating and proton uptake kinetics is in order and will be the object of a subsequent study.

Mechanistic implications

The present study has multiple implications to the mechanism of proton uptake in the D-channel and sheds light on the molecular origin for the effect of mutations, respectively in the bottleneck and in the wider regions of the channel.

First, the agreement between the computed distributions of water molecules in the D-channel and the location of buried water molecules in high-resolution crystallographic structures of the protein is as good as that between the crystallographic models themselves (Figs. 2a and 4 and Table 2). By far the most significant difference is the presence of a water molecule at the channel bottleneck in the open conformation of residue N139. This conformational change enables the link in the proton relay pathway that is missing in the X-ray structures and suggests, as discussed below, that conformational gating by N139 plays an important role in the proper function of the enzyme.

By contrast, the high mobility of water in the wider region of the D-channel observed in the present study (Fig. 2b), together with the multimodal character of the spatial water density distribution and its tolerance to changes in hydration number, are consistent with the relative tolerance of the enzyme to mutations in this region of the D-channel.[25,44,64] In addition, these properties have several implications to the mechanism of proton uptake, both in the wild type and in N139 mutants.

In the wild-type protein, a consequence of the above findings is to question the validity of evidence for the existence of a "proton reservoir" in the D-channel. It has been proposed, on the basis of computed free-energy profiles for proton movement in the large cavity of the D-channel, that water holds an excess proton when E286 is protonated.[42,43] Agreement between the simulated time-averaged position of oxygen atoms in a protonated water chain and the location of crystallographic water was invoked as supporting evidence for this hypothesis.[42] However, the excellent agreement between simulated and crystallographic water distributions obtained in the current study in the absence of an excess proton (Fig. 2a) suggests that the purported evidence for a "proton reservoir" is not meaningful.

To set the stage for further mechanistic considerations, we first review the three hypotheses proposed previously to explain the non-native phenotypes of N139 mutants. Decoupled and inactive phenotypes have been proposed to arise from three different factors that are not mutually exclusive and that have yet to be resolved.[26,27] Residue E286, which shuttles chemical and pumped protons between the D-channel and the active site of the enzyme,[13,14] plays a central role in all three cases. Both its conformational isomerization[14] and its proton affinity[5] are key to its proton shuttling ability and are therefore thought to be important for the proper function of the enzyme. The three factors are (i) changes in the pK_a and/or conformational equilibrium of E286 due to long-range electrostatic interactions;[20,23,25,27,65] (ii) changes in the pK_a and/or conformational equilibrium of E286 induced by local perturbations of its environment;[23,26,27] and (iii) perturbations of the kinetics of proton delivery to E286.[26]

The effect of long-range electrostatic interactions between E286 and three charged mutants, N139D, N207D, and G204D, on the decoupled and (in the latter case) inactive phenotypes of these mutants, is supported by pK_a calculations in a detailed model of the entire redox cycle of the enzyme.[5] Specifically, the introduction of a charged residue in the D-channel raises the pK_a of E286, thereby impairing its capacity to reload the proton loading site (in all three mutants) and even to deliver chemical protons to the binuclear centre (in the G204D mutant). However, this explanation does not apply to neutral mutants of N139 such as A, V, T, C, or Q, in which both pumping and oxidase activity are reduced or eliminated

(Table 1), given the absence of long-range coulombic interactions with E286. *A fortiori*, any long-range electrostatic coupling between residue N139 and the cofactors in various redox/charge states should be discounted. Likewise, the fact that the conformational isomerization of N139 and its replacement by A and V only result in local perturbations of water structure and leaves the rest of the channel unchanged argues against any significant perturbation of the conformational equilibrium of E286 by uncharged mutants of N139.

Since neither the local environment of E286 nor the long-range electrostatic field are affected, it is likely instead that the primary effect of replacing N139 by neutral residues, polar or not, is to compromise the kinetics of proton uptake in the D-channel. In particular, slowing down the reprotonation of E286 may lead to slippage or bypass of the proton pumping step, resulting in decoupling of pumping and oxidase activity. Two mechanisms to that effect have been discussed in the framework of a detailed sequence of electron and proton transfer steps.[5] In this scheme, the ejection (i.e., pumping) of a vectorial proton from a putative proton-loading site (PLS) is electrostatically triggered by the reprotonation of E286 from the D-channel following delivery of a chemical proton to the binuclear centre. If reprotonation of E286 is slower than electron transfer to heme *a*, the PLS may not become sufficiently activated to pump a proton, effectively bypassing the pumping step and decoupling pumping from oxidase activity.[5] In addition, fast proton reload of E286 from the D-channel may be required to prevent the backflow or slippage of protons from the PLS to E286, as also pointed out by others.[3,65]

Here we saw that N139V is unlikely to host a water wire, suggesting that the absence of proton pumping and the drastic reduction in oxidase activity[25,27] are both due to a persistent defect in the proton pathway and subsequent impairment of proton uptake. By contrast, the present study predicts that water occupies the bottleneck most of the time in the N139A mutant. The presence of a continuous water chain in a pore-like D-channel suggests that altered proton uptake kinetics could arise instead from the energetic profile of proton transport at the bottleneck. However, the fact that N139A is apparently an "open" pore, together with previous mutagenesis studies, also raises the question of the role of this highly conserved asparagine residue.

Indeed, the kinetic hypothesis resonates with the finding that the two rotameric states of the N139 side chain appear to act as a gate controlling water and proton access into the D-channel, which in itself strongly suggests that the conformational isomerization of N139 plays a role in the kinetics of proton uptake. In support of this hypothesis, a recent study of the free-energy profile for proton uptake in the D-channel revealed the presence of a 5-kcal/mol free-energy barrier at N139. Although gating *per se* was not described, the barrier was tentatively attributed to conformational fluctuations involving N139.[43] The effect of conformational isomerization on proton uptake kinetics awaits further investigation.

In any event, the fact that the gate apparently disappears in N139A, a decoupled mutant, suggests that a gate may indeed be required for the proper function of the enzyme. Although, as stated above, it is possible that this mutation delays proton uptake by raising the energetic barrier of crossing the bottleneck compared to the wild type, an alternative hypothesis is that eliminating the gate could compromise proton selectivity. Thus, it is conceivable that the gated bottleneck of the D-channel acts as a "proton filter", preventing the intrusion of physiological cations other than H^+ into the large cavity forming the top of the D-channel, where they could interfere with proton uptake.

Conclusions and Perspectives

We have examined the structural and thermodynamic basis for the gating and hydration of the D-channel in the wild-type C*c*O and in two single-point mutants, N139A and N139V. Our results clarify the molecular basis for the relay of protons in the hydrated, channel-like cavity. The detailed comparison of water placement derived from crystallographic studies with the spatial distribution of water in the wider part of the D-channel obtained from the simulations offers an explanation for the relative tolerance of the top of the channel to mutations.

Inversely, our results show that residue N139 is both a steric and a hydration bottleneck and that the spontaneous conformational isomerization of its side chain acts as a gate that controls the formation of a putative water wire. These results shed light onto the role of N139 in the mechanism of proton uptake and clarify the physical basis for the phenotypes of single-point mutants of residue 139. Results for N139V suggest that blockage of proton uptake resulting from interruption of the water pathway is the cause of this mutant's marginal activity. Results for N139A indicate that the D-channel is a continuously hydrated cavity, implying that the decoupled phenotype is not the result of a break in the proton relay chain.

While important mechanistic questions remain, this study underlines the need for the systematic examination of proton conduction pathways to understand the molecular mechanism of proton movement in C*c*O. More generally, the present study provides insight into the physical basis of biomolecular hydration and presents a general approach for the calculation of water occupancy in protein cavities. We show that combining efficient sampling algorithms with long simulation times is required to achieve statistical convergence of equilibrium properties in protein interiors. Accordingly, the method described in this study can be used to predict the binding affinity of molecular ligands or inhibitors to the active site of enzymes.[54]

Elucidating the interplay between conformational fluctuations, hydration, and proton relay is required to understand molecular mechanisms of proton movement in water-filled channels and channel-like protein cavities—especially in a pump, which requires kinetic as well as thermodynamic control of proton movement. By identifying a conformational gate that controls the pathway of protons in the D-channel, this study paves the way for a detailed examination of proton uptake kinetics, in which the time dependence of conformational gating and proton relay will be characterized successively in the wild type and in mutant forms of N139 in which the enzyme phenotype is compromised.

Theory

Free-energy calculations

To calculate the free-energy change between a bound and unbound ligand molecule from a receptor molecule, we employed thermodynamic integration in four spatial dimensions, a method described in detail by Rodinger *et al.*[47] In this approach, the coupling between the ligand (a water molecule in this study) being extracted and the remainder of the molecular system (C*c*O in this study) is modulated by a non-physical degree of freedom, their spatial separation in the fourth dimension, w. The free-energy change for the extraction of the ligand molecule from the receptor corresponds to the energy difference between the fully-coupled (fully-inserted) and the fully-decoupled (fully-extracted) states at $w=0$ and $w=\infty$, respectively. The PMF along the w coordinate can be computed by either umbrella sampling[66] or thermodynamic integration[67] simulations, but for the purpose of this work, thermodynamic integration was utilized.

The mean force is acquired by averaging the fourth-dimensional component of the force ($-\partial H/\partial w$) between the ligand and receptor over the sampling run of a molecular dynamics (MD) simulation. Simulations are carried out at discrete steps of w, keeping in mind that adjacent steps must be close enough together such that the numerically integrated data accurately reconstructs the PMF. In principle, sampling should extend from $w=0$ to $w=\infty$, but this is not feasible. However, it has been shown that a PMF along w converges toward the $w=\infty$ value at relatively low values of w, allowing extrapolation to be used past this point (typically ~20 Å).[52] Further details of free-energy calculations using thermodynamic integration in four spatial dimensions along with the advantages afforded by this technique can be found elsewhere.[52]

Distributed replica sampling

Distributed replica sampling[47] is utilized in this study to perform efficient Boltzmann sampling of conformational space. This is a general technique to achieve sampling uniformity along the temperature (T) or reaction coordinate (λ) space, but is described here in terms of the fourth-dimensional reaction coordinate (w) used in this study. Multiple replicas of a protein system differing in reaction coordinate (in this case, the w coordinate) are simulated independently. Periodically, individual replicas are halted and a stochastic move in w space is attempted. The probability of accepting this move is given by the Metropolis Monte Carlo criterion:

$$\rho = \min[1, \exp(-\beta\Delta)], \tag{1}$$

where β is the inverse of the product of the Boltzmann constant and the absolute temperature (298 K), and

$$\begin{aligned}\Delta = {} & E(q_m, (w_m + \partial w_m)) \\ & - E(q_m, w_m) + D(w_1, w_2, \ldots, (w_m + \partial w), \ldots, w_n) \\ & - D(w_1, w_2, \ldots, w_m, \ldots, w_n),\end{aligned} \tag{2}$$

where q_m represents the 3D atomic coordinates of the atoms of replica m, w_m is the coordinate of replica m in the fourth dimension, $w_m + \partial w_m$ is the proposed new fourth dimensional coordinate, E is the potential energy of the system, and D is the distributed replica potential energy (DRPE), which enforces the distribution of replicas across the range of the reaction coordinate. The DRPE represents an energy penalty associated with a non-ideal distribution of the replicas.

The DRPE was calculated as outlined by Rodinger *et al.*[47] The number of time steps between move attempts and replica positions used in this study are listed in Supplementary Table S1.

Reduction of ligand sampling volume

By combining thermodynamic integration in four spatial dimensions and distributed replica sampling, our simulations produce a random walk of the extracted ligand molecule in the fourth dimension. For this to occur, the path needs to be reversible, and therefore the ligand molecule needs to return to the same region of the receptor from which it was extracted. However, once the ligand is extracted from the receptor (e.g., $w=20$ Å), it is free to sample all 3D space in principle. As a result, the probability that upon reinsertion into the receptor the ligand molecule would return to the location from which it was extracted is very low. In addition, as the ligand is abstracted into the fourth dimension, its effective size in the receptor diminishes. This increases the number of available states in the first three spatial dimensions, and therefore increases the amount of sampling required.

To address both issues, a flat-bottomed radial restraint is applied to the centre of mass of the ligand molecule to restrict it to a specific region of 3D space. The size of the restraint is modulated along with the w coordinate of the ligand. The sizes are chosen in such a way as to have no influence on the simulation at low w coordinates, when the ligand molecule is fully inserted into the receptor,

and yet be maximally restrictive at high w coordinates to limit the free volume sampled by the ligand in 3D space. Because the size of the restraint decreases as the unbinding process progresses, the restraint is referred to as the "funnel". The force exerted by the funnel on the ligand is recorded throughout the simulation and integrated with respect to the funnel radius to yield the work contributed by the funnel restraint. This funnel work is added to the free energy obtained by integrating the average force exerted on the ligand by the receptor with respect to w. This total is the free energy of transfer of a ligand from the receptor into vacuum. The funnel method described here is of general application and can be used for any ligand binding simulation utilizing extraction along a non-physical pathway, such as "scaling–shifting" methods.[68,69]

Standard-state correction

Since protein–ligand binding free energies depend on the concentration of the ligand, a standard-state correction is needed to account for the restricted volume of the funnel when the ligand molecule is in vacuo ($w=\infty$). Thus, the free-energy change (ΔA) for expanding the volume available to the fully extracted ligand in the funnel to the volume that the ligand occupies in its standard state was calculated. For a water molecule, the standard state volume is 30.0 Å^3, assuming a bulk concentration of 55.3 mol/L.

The difference in absolute energies A_f and A_s, corresponding to the volume allowed by the funnel restraint and of the standard state, was calculated, where

$$A = -RT\ln(Z), \tag{3}$$

and Z is the partition function of volume s or f. Then

$$Z = \sum_j e^{-E_j/RT}, \tag{4}$$

where E_j is the energy contributed by the volumetric restraints on the ligand molecule for all possible microstates, j. To calculate Z for the funnel state and the standard state, we need only to consider the energy function used to restrain the ligand molecule to the vacuum. In the case of the standard state, which is composed of infinitely "hard" walls, the energy applied to the ligand molecule is $E=0$ when the water molecule is contained within the volume and $E=\infty$ outside the boundaries of the volume. Thus, using Eq. (4), Z is equal to the free volume of the standard state (30.0 in this study), and A_s is equal to -2.01 kcal/mol at $T=298$ K.

Similarly, in the case of the funnel state, $E=0$ when the water molecule is contained within the funnel restraints. However, $E\neq\infty$ at coordinates outside of the funnel restraint, but rather E depends on the flat-bottomed harmonic restraint. Thus, a program was created to numerically calculate Z for an arbitrary geometric volume. The program divides an area of space significantly larger than the space defined by the funnel restraints into small cubes. It then proceeds to traverse systematically through all the cubes, calculating the energy of each cube based on the equation of the funnel restraint and the coordinates of the centre of the cube ($E=0$ if the cube resides inside the funnel boundaries). Subsequently, the Boltzmann probability of each cube is calculated based on its energy, and these probabilities are each multiplied by the volume of the cube to obtain the unnormalized probability. Finally, these values are summed to obtain the partition function, Z. The program is iterated with decreasing cube volumes and increasing search space until the calculated partition function converges upon a single value. Using this method, the partition function of our funnel volume was calculated as $Z=6.6$ and A_f was calculated to be -1.12 kcal/mol. Thus, the standard-state correction amounted to a free-energy change of $\Delta A=$-0.9 kcal/mol. This value was added to the calculated free energy for transfer from the D-channel to vacuo, along with the hydration free energy of a TIP3P water molecule, to obtain the total transfer free energy of a water molecule from the D-channel to the bulk.

The funnel volume at small values of w is large enough that it does not interact with the test water molecule in any of the wild-type and mutant systems investigated here. As a consequence, the same funnel potential was used in all three cases. Additionally, although the funnel restraint used here was a soft-walled cylindrical volume, the program written to calculate the partition function is general and can be used to calculate the partition function of many 1D, 2D, and 3D geometric restraints with user-specified force constants.

Methods

Molecular system

The initial conformation of the protein was obtained from the structure of *R. sphaeroides* C*c*O solved at 2.3 Å resolution by X-ray crystallography (PDB ID code 1M56).[29] Subunit I, which contains the D-channel and the active site of the enzyme, was used as the starting wild-type structure for the simulations performed in this study. The initial system consisted of 8796 protein atoms (including heme *a*, heme *a3*, Mg and Cu cofactors), 141 crystallographic water molecules (not including those present in the D-channel), and 10 to 12 water molecules in the D-channel. To model bulk solution on the matrix side of the protein, a hemispherical "cap" of water molecules was placed at the entrance of the D-channel. The cap, with a radius of 9 Å, contained 57 water molecules, for a total of 9420 to 9426 atoms in the wild-type system. An axis connecting the C_α atoms of D132 and E286 was used to define the D-channel and was aligned with the z-axis of the system. Residues with at least one heavy atom within 5 Å from the D-channel axis as well as the D-channel water molecules and cap water molecules were allowed to move during the MD simulations. All remaining atoms in the system were held fixed. This produced a total of 876 to 879 moving atoms during the MD simulations.

The CHARMM force field, version 22,[69] was used to model the protein, and the TIP3P force field[70] was used to model all water molecules. Previous studies have shown that the TIP3P water model consistently describes the structure and fluctuations of a single file of water molecules in the interior of transmembrane proteins.[36] The enzyme was simulated in the fully-reduced R state. The charge distribution of the binuclear centre was calculated as described by Fadda *et al.*[5] Titratable residues were simulated at standard protonation states. Geometric constraints were placed on the cap water molecules as well as the D-channel water molecules. The cap water molecules were restrained by a spherical boundary potential with a radius of 9 Å centred at the channel opening and a force constant of 5 kcal mol^{-1} Å^{-2}. To prevent water molecules from travelling past E286 and exiting the D-channel into the active-site region, a half-harmonic restraint was placed on the D-channel water molecules at $z=25.5$ Å with a force constant of 5 kcal mol^{-1} Å^{-2}. Finally, as discussed above, a funnel restraint was placed on the test water molecule. This restraint was chosen to be cylindrical in shape, with the sides of the cylinder composed of a flat-bottomed potential running parallel to the axis of the D-channel, and the caps composed of a second flat-bottomed potential 2.5 Å wide, centred at the initial z-coordinate of the test water molecule. The radius of the cylinder decreased with increasing w values (Table S1), and all sides of the cylinder had a harmonic force constant of 5.0 kcal mol^{-1} Å^{-2}.

Geometric restraints were imposed on the χ_1 angle of N139 and N139V in order to prevent the side chain of these residues from adopting a rotameric state that would block the D-channel and prevent the formation of a hydrogen-bonded water wire. Residue N139 was held open with a harmonic restraint applied to χ_1 centred at 285° with a force constant of 0.002 kcal mol^{-1} deg^{-2}. The χ_1 angle of N139V was restrained with a harmonic potential centred at 255° with a force constant of 0.002 kcal mol^{-1} deg^{-2}.

Molecular dynamics simulations

The MD trajectories were generated using the program CHARMM, version 28.[71] The Langevin equations of motion were propagated at 298 K with an integration step of 2 fs and a friction coefficient of 5 ps^{-1} applied to all heavy atoms. The SHAKE algorithm[72] was employed to fix all bonds involving hydrogen atoms with a bond deviation tolerance of 1.0×10^{-6}. Non-bonded interactions were calculated with a force-based switching function acting between 14 and 16 Å. Trajectories and structures were viewed using Visual Molecular Dynamics (VMD).[51]

Initial determination of hydration states

A series of simulations were executed for all three protein systems containing 10, 11, or 12 water molecules in the D-channel. Each protein system was initially set up with 12 water molecules in the D-channel and equilibrated for 2 ns. Multiple simulations were executed for each protein system with 11 water molecules, each differing in the initial spatial distribution of water molecules. Analogously, this was done for each protein system with 10 water molecules in the D-channel. All simulations were run for 5 ns using the setup outlined above. During the course of the simulations, the z-coordinates of the water oxygen atoms were recorded to compute the axial distribution of the water molecules in the D-channel.

Free-energy simulations of conformational isomerization of residue 139

The reversible work or PMF for the rotation about χ_1 of residue N139 was calculated using two methods. First, a 2D PMF was generated using umbrella sampling[66] with biasing potentials applied to both side-chain torsions χ_1 and χ_2 in the absence of water in the D-channel. Harmonic restraints with a force constant of 0.01 kcal mol^{-1} deg^{-2} were imposed at 15° intervals from 180° to 300° and −165° to 165° for χ_1 and χ_2, respectively. Starting from the crystallographic closed state of residue N139, each (χ_1, χ_2) window was simulated for 100 ps of equilibration, followed by 400 ps of data collection. The PMF was generated using Alan Grossfield's implementation of WHAM.[73]

The PMF along χ_1 was also calculated using 1D umbrella sampling together with distributed replica sampling[47] to reduce systematic sampling errors.[53] Replicas of the system were created with harmonic restraints imposed on χ_1, centred between 170° and 310° in 10° increments, with a force constant of 0.03 kcal mol^{-1} deg^{-2}. Construction of all starting structures began from an equilibrated structure with N139 in the closed position ($\chi_1\approx195°$), and 12 water molecules in the D-channel. The crystal coordinates of subunits II, III, and IV were added, and these atoms remained fixed for the extent of all simulations. From this starting structure, the remaining structures were created by running successive 400-ps simulations. Here, the final structure of a replica with $\chi_1=X$ was used as the starting structure for the replica with $\chi_1=X+10°$ or $\chi_1=X-10°$, as required. All replicas were then equilibrated for another 400 ps with a force constant of 0.02 kcal mol^{-1} deg^{-2}. Next, a series of 2-ns simulations of all replicas were performed while adjusting both the umbrella centres and force constants in an attempt to obtain near 20% overlap between all adjacent umbrellas in the sampled χ_1 space. This resulted in the final 23 umbrella positions and force constants found in Supplementary Table S2. Adaptive A values were calculated as described before[53] using 1 ns of simulation time per replica with no exchanges. All replicas were then run while allowing umbrella changes every 2 ps. The total sampling time at each replica position varied from 11 to 22 ns, with an average of 16 ns and a total sampling time of 374 ns. Throughout the simulation, A values were adjusted based on the accumulated free-energy surface. Error was calculated using block averaging, excluding the first 20% as equilibration. A separate PMF was generated for each block of data, and a global fit was used to align all generated PMFs. The standard deviation between the PMFs was then calculated.

Finally, umbrella sampling was utilized to construct the PMF for the rotation about χ_1 of residue N139V with harmonic biasing potentials centred 10° apart from 160° to 290°, with two additional umbrellas at 205° and 215°. A comparatively smaller sampling effort was devoted to N139V than to the wild type because our primary purpose was to determine if the closed conformation of N139V was a stable conformer or an artefact of our free-energy simulation protocol. A force constant of 0.02 kcal mol^{-1} deg^{-2} was used for all umbrellas, except those centred at 205° and 215°, which had a force constant of 0.04 kcal mol^{-1} deg^{-2}. Starting structures were obtained from previous free-energy simulations in which N139V had sampled conformations between the open and closed states. Each structure was equilibrated for 1 ns before data collection. The values of χ_1 were recorded every 200 fs and each window was simulated for 4 ns, for a total of 56 ns.

Free-energy simulations of water transfer

Free-energy simulations were performed for all three protein systems using thermodynamic integration in four spatial dimensions along with distributed replica sampling. Based on the results of the initial 5-ns simulations, two free-energy simulations were run for the wild-type system. In one simulation, the fully-hydrated state ($w=0$) contained 12 water molecules. In the second simulation, 11 water molecules were present in the D-channel in the fully-hydrated state. In both cases, the water molecule initially present at $z \approx 6.0$ Å was extracted from the D-channel along the fourth-dimensional reaction coordinate into vacuo.

Based on the results of the 5-ns simulations, one free-energy simulation was run for both the N139A and N139V mutant systems with 12 water molecules present in the fully hydrated state ($w=0$). For the N139A mutant system, the water molecule initially present at $z \approx 6.0$ Å was assigned as the test water molecule. In the case of the N139V mutant, the water molecule initially present at $z \approx 9.0$ Å was extracted from the D-channel. The decision of where to extract a water molecule from the D-channel was based on where defects tended to form during the 5-ns simulations.

Fifty-six replicas of each protein system were created, each differing in the initial fourth-dimensional coordinate (w) of the test water molecule oxygen atom (Table S1). The radius of the restraining funnel was made progressively smaller as the water molecule was extracted further into the fourth-dimensional coordinate space. Each fourth-dimensional w coordinate had a corresponding funnel radius (Table S1).

Each replica position was simulated for 400 fs (200 steps, 2 fs per step) before a move attempt was made. Table S1 displays the total move attempts initially allotted at each replica position. The simulation was allowed to run until convergence had been attained. For the wild-type 11-water-molecule system, wild-type 12-water-molecule system, and the N139V system, convergence was attained after each replica had performed approximately 11,200–22,500, 7100–23,500, and 12,200–20,000 move attempts, respectively (430, 600, and 400 ns total simulation time). In the case of the N139A system, more move attempts were needed than were initially allotted due to the difficulty in reaching convergence (21,200–50,000 move attempts, 908 ns total simulation time). Convergence was gauged based on the statistical error obtained by block-averaging the mean-force values. Although an acceptably low value for the statistical error was obtained early in time, the simulations were allowed to continue for much longer time periods to minimize systematic error.

Error calculations and convergence

Convergence was gauged based on the decreasing statistical error about the mean of the free energy. Statistical error was calculated using block averaging of the recorded force values along the w coordinate, along with standard error propagation techniques to obtain the statistical error about the mean of the free energy. The data were divided into 50 evenly-sized time windows, with approximately 30% of the initial simulation data discarded as equilibration, leaving 35 windows to be used for block averaging error calculations. The statistical error for the wild-type (11 and 12 water molecules), N139A, and N139V protein systems was calculated to be 0.05, 0.05, 0.07, and 0.1 kcal/mol, respectively. Systematic error was minimized by extensive simulation times while allowing the system to undergo a random walk in w-coordinate space. During the course of the simulations, each replica performed many full transitions from $w=0$ Å to $w=20$ Å and back.

Acknowledgements

We thank R. B. Gennis for communicating unpublished results and gratefully acknowledge the high-performance facility of the Centre for Computational Biology at the Hospital for Sick Children for generous access to computer resources. This work was supported by Operating Grant MOP43949 from the Canadian Institutes of Health Research. R.P. is a CRCP Chairholder.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.02.042

References

1. Wikström, M. (2004). Cytochrome *c* oxidase: 25 years of the elusive proton pump. *Biochim. Biophys. Acta*, **1655**, 241–247.
2. Michel, H. (1999). Cytochrome *c* oxidase: catalytic cycle and mechanisms of proton pumping—a discussion. *Biochemistry*, **38**, 15129–15140.
3. Wikström, M. & Verkhovsky, M. I. (2007). Mechanism and energetics of proton translocation by the respiratory heme-copper oxidases. *Biochim. Biophys. Acta*, **1767**, 1200–1214.
4. Brzezinski, P. & Gennis, R. B. (2008). Cytochrome *c* oxidase: exciting progress and remaining mysteries. *J. Bioenerg. Biomembr.* **40**, 521–531.
5. Fadda, E., Yu, C. H. & Pomès, R. (2008). Electrostatic control of proton pumping in cytochrome *c* oxidase. *Biochim. Biophys. Acta*, **1777**, 277–284.
6. Wikström, M. (1977). Proton pump coupled to cytochrome *c* oxidase in mitochondria. *Nature*, **266**, 271–273.
7. Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. (1995). Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature*, **376**, 660–669.
8. Gennis, R. B. (1998). Multiple proton-conducting pathways in cytochrome oxidase and a proposed role for the active-site tyrosine. *Biochim. Biophys. Acta*, **1365**, 241.
9. Garcia-Horsman, J. A., Puustinen, A., Gennis, R. B. & Wikström, M. (1995). Proton transfer in cytochrome *bo3* ubiquinol oxidase of *Escherichia coli*: second-site mutations in subunit I that restore proton pumping in the mutant Asp135. *Biochemistry*, **34**, 4428–4433.
10. Konstantinov, A. A., Siletsky, S., Mitchell, D., Kaulem, A. & Gennis, R. B. (1997). The roles of the two proton input channels in cytochrome *c* oxidase from *Rhodobacter sphaeroides* probed by the effects of site-directed mutations on time-resolved electrogenic intraprotein proton transfer. *Proc. Natl Acad. Sci. USA*, **94**, 9085–9090.

11. Brzezinski, P. & Adelroth, P. (1998). Pathways of proton transfer in cytochrome *c* oxidase. *J. Bioenerg. Biomembr.* **30**, 99–107.
12. Mills, D. A., Florens, L., Hiser, C., Qian, J. & Ferguson-Miller, S. (2000). Where is 'outside' in cytochrome *c* oxidase and how and when do protons get there? *Biochim. Biophys. Acta*, **1458**, 180–187.
13. Riistama, S., Hummer, G., Puustinen, A., Dyer, R. B., Woodruff, W. H. & Wikström, M. (1997). Bound water in the proton translocation mechanism of haem-copper oxidases. *FEBS Lett.* **414**, 275–280.
14. Pomès, R., Hummer, G. & Wikström, M. (1998). Structure and dynamics of a proton shuttle in cytochrome *c* oxidase. *Biochim. Biophys. Acta*, **1365**, 255–260.
15. Hofacker, I. & Schulten, K. (1998). Oxygen and proton pathways in cytochrome *c* oxidase. *Proteins: Struct. Funct. Genet.* **30**, 100–107.
16. Adelroth, P., Karpefors, M., Gilderson, G., Tomson, F. L., Gennis, R. B. & Brzezinski, P. (2000). Proton transfer from glutamate 286 determines the transition rates between oxygen intermediates in cytochrome *c* oxidase. *Biochim. Biophys. Acta*, **1459**, 533–539.
17. Brändén, G., Pawate, A. S., Gennis, R. B. & Brzezinski, P. (2006). Controlled uncoupling and recoupling of proton pumping in cytochrome *c* oxidase. *Proc. Natl Acad. Sci. USA*, **103**, 317–322.
18. Adelroth, P., Ek, M. S., Mitchell, D. M., Gennis, R. B. & Brzezinski, P. (1997). Glutamate 286 in cytochrome *aa3* from *Rhodobacter sphaeroides* is involved in proton uptake during the reaction of the fully-reduced enzyme with dioxygen. *Biochemistry*, **36**, 13824–13829.
19. Aagaard, A., Gilderson, G., Mills, D. A., Ferguson-Miller, S. & Brzezinski, P. (2000). Redesign of the proton-pumping machinery of cytochrome *c* oxidase: proton pumping does not require Glu(I-286). *Biochemistry*, **39**, 15847–15850.
20. Pawate, A. S., Morgan, J., Namslauer, A., Mills, D., Brzezinski, P., Ferguson-Miller, S. & Gennis, R. B. (2002). A mutation in subunit I of cytochrome *c* oxidase from *Rhodobacter sphaeroides* results in an increase in steady state activity but completely eliminates proton pumping. *Biochemistry*, **41**, 13417–13423.
21. Mitchell, D. M., Fetter, J. R., Mills, D. A., Adelroth, P., Pressler, M. A., Kim, Y. *et al.* (1996). Site-directed mutagenesis of residues lining a putative proton transfer pathway in cytochrome *c* oxidase from *Rhodobacter sphaeroides*. *Biochemistry*, **35**, 13089–13093.
22. Han, D., Morgan, J. E. & Gennis, R. B. (2005). G204D, a mutation that blocks the proton-conducting D-channel of the aa3-type cytochrome *c* oxidase from *Rhodobacter sphaeroides*. *Biochemistry*, **44**, 12767–12774.
23. Han, D., Namslauer, A., Pawate, A. S., Morgan, J. E., Nagy, S., Vakkasoglu, A. S. *et al.* (2006). Replacing Asn207 by aspartate at the neck of the D channel in the aa_3-type cytochrome *c* oxidase from *Rhodobacter sphaeroides* results in decoupling the proton pump. *Biochemistry*, **45**, 14064–14074.
24. Backgren, C., Hummer, G., Wikström, M. & Puustinen, A. (2000). Proton translocation by cytochrome *c* oxidase can take place without the conserved glutamic acid in subunit I. *Biochemistry*, **39**, 7863–7867.
25. Pfitzner, U., Hoffmeier, K., Harrenga, A., Kannt, A., Michel, H., Bamberg, E. *et al.* (2000). Tracing the D-pathway in reconstituted site-directed mutants of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Biochemistry*, **39**, 6756–6762.
26. Lepp, H., Salomonsson, L., Zhu, J. P., Gennis, R. B. & Brzezinski, P. (2008). Impaired proton pumping in cytochrome *c* oxidase upon structural alteration of the D pathway. *Biochim. Biophys. Acta*, **1777**, 897–903.
27. Dürr, K. L., Koepke, J., Hellwig, P., Müller, H., Angerer, H., Peng, G. *et al.* (2008). A D-pathway mutation decouples the *Paracoccus denitrificans* cytochrome *c* oxidase by altering the side-chain orientation of a distant conserved glutamate. *J. Mol. Biol.* **384**, 865–877.
28. Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K. *et al.* (1996). The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å. *Science*, **272**, 1136–1144.
29. Svensson-Ek, M., Abramson, J., Larsson, G., Tornroth, S., Brezezinski, P. & Iwata, S. (2002). The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome *c* oxidases from *Rhodobacter sphaeroides*. *J. Mol. Biol.* **321**, 329–339.
30. Qin, L., Hiser, C., Mulichak, A., Garavito, R. M. & Ferguson-Miller, S. (2006). Identification of conserved lipid/detergent-binding sites in a high-resolution structure of the membrane protein cytochrome *c* oxidase. *Proc. Natl Acad. Sci. USA*, **103**, 16117–16122.
31. de Grotthuss, C. J. T. (1806). Mémoire sur la décompostion de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann. Chim. (Paris)*, **58**, 54–74.
32. de Grotthuss, C. J. T. (2006). Memoir on the decomposition of water and of the bodies that it holds in solution by means of galvanic electricity. *Biochim. Biophys. Acta*, **1757**, 871–875.
33. Tuckerman, M., Laasonen, K., Sprik, M. & Parrinello, M. (1995). *Ab initio* molecular dynamics simulation of the solvation and transport of H_3O^+ and OH^- ions in water. *J. Phys. Chem.* **99**, 5749–5752.
34. Pomès, R. & Roux, B. (1996). Structure and dynamics of a proton wire: a theoretical study of H^+ translocation along the single-file water chain in the gramicidin channel. *Biophys. J.* **71**, 19–39.
35. de Groot, B. L., Frigato, T., Helms, V. & Grubmüller, H. (2003). The mechanism of proton exclusion in the aquaporin-1 water channel. *J. Mol. Biol.* **333**, 279–293.
36. Chakrabarti, N., Tajkhorshid, E., Roux, B. & Pomès, R. (2004). Molecular basis of proton blockage in aquaporins. *Structure*, **12**, 65–74.
37. Kaila, V. R., Verkhovsky, M. I., Hummer, G. & Wikström, M. (2008). Glutamic acid 242 is a valve in the proton pump of cytochrome *c* oxidase. *Proc. Natl Acad. Sci. USA*, **105**, 6255–6259.
38. Zheng, X., Medvedev, D. M., Swanson, J. & Stuchebrukhov, A. A. (2003). Computer simulation of water in cytochrome *c* oxidase. *Biochim. Biophys. Acta*, **1557**, 99–107.
39. Olkhova, E., Hutter, M. C., Lill, M. A., Helms, V. & Michel, H. (2004). Dynamic water networks in cytochrome *c* oxidase from *Paracoccus denitrificans* investigated by molecular dynamics simulations. *Biophys. J.* **86**, 1873–1889.
40. Seibold, S. A., Mills, D. A., Ferguson-Miller, S. & Cukier, R. I. (2005). Water chain formation and possible proton pumping routes in *Rhodobacter sphaeroides* cytochrome *c* oxidase: a molecular dynamics comparison of the wild type and R481K mutant. *Biochemistry*, **44**, 10475–10485.
41. Xu, J. & Voth, G. A. (2005). Computer simulation of explicit proton translocation in cytochrome *c* oxidase: the D-pathway. *Proc. Natl Acad. Sci. USA*, **102**, 6795–6800.
42. Xu, J., Sharpe, M. A., Qin, L., Ferguson-Miller, S. & Voth, G. A. (2007). Storage of an excess proton in the

hydrogen-bonded network of the D-pathway of cytochrome *c* oxidase: identification of a protonated water cluster. *J. Am. Chem. Soc.* **129**, 2910–2913.
43. Xu, J. & Voth, G. A. (2006). Free energy profiles of H^+ conduction in the D-pathway of Cytochrome *c* oxidase: a study of the wild type and N98D mutant enzymes. *Biochim. Biophys. Acta*, **1757**, 852–859.
44. Tsukihara, T., Shimokata, K., Katayama, Y., Shimada, H., Muramoto, K., Aoyama, H. *et al.* (2003). The low-spin heme of cytochrome *c* oxidase as the driving element of the proton-pumping process. *Proc. Natl Acad. Sci. USA*, **100**, 15304–15309.
45. Ernst, J. A., Clubb, R. T., Zhou, H. X., Gronenborn, A. M. & Clore, G. M. (1995). Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science*, **267**, 1813–1817.
46. Liu, L., Quillin, M. L. & Matthews, B. W. (2008). Use of experimental crystallographic phases to examine the hydration of polar and nonpolar cavities in T4 lysozyme. *Proc. Natl Acad. Sci. USA*, **105**, 14406–14411.
47. Rodinger, T., Howell, P. L. & Pomès, R. (2006). Distributed replica sampling. *J. Chem. Theory and Comput.* **2**, 725–731.
48. Shinzawa-Itoh, K., Aoyama, H., Muramoto, K., Terada, H., Kurauchi, T., Tadehara, Y. *et al.* (2007). Structures and physiological roles of 13 integral lipids of bovine heart cytochrome *c* oxidase. *EMBO J.* **26**, 1713–1725.
49. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (1999). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **285**, 1735–1747.
50. Lovell, S. C., Davis, I. W., Adrendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G. *et al.* (2003). Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins: Struct. Funct. Genet.* **50**, 437–450.
51. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28.
52. Rodinger, T., Howell, P. L. & Pomès, R. (2005). Absolute free energy calculations by thermodynamic integration in four spatial dimensions. *J. Chem. Phys.* **123**, 34104.
53. Neale, C., Rodinger, T. & Pomès, R. (2008). Equilibrium exchange enhances the convergence rate of umbrella sampling. *Chem. Phys. Lett.* **460**, 375–381.
54. Rodinger, T., Howell, P. L. & Pomès, R. (2008). Calculation of absolute protein-ligand binding free energy using distributed replica sampling. *J. Phys. Chem.* **129**, 155102.
55. Pomès, R., Eisenmesser, E., Post, C. B. & Roux, B. (1999). Calculating excess chemical potentials using dynamic simulations in the fourth dimension. *J. Chem. Phys.* **111**, 3387–3395.
56. Roux, B., Nina, M., Pomès, R. & Smith, J. C. (1996). Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. *Biophys. J.* **71**, 670–681.
57. Olano, L. R. & Rick, S. W. (2004). Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.* **126**, 7991–8000.
58. Hamelberg, D. & McCammon, J. A. (2004). Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method. *J. Am. Chem. Soc.* **126**, 7683–7689.
59. Mobley, D. L., Chodera, J. D. & Dill, K. A. (2006). On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **125**, 084902.
60. Woo, H. J., Dinner, A. R. & Roux, B. (2004). Grand canonical Monte Carlo simulations of water in protein environments. *J. Chem. Phys.* **121**, 6392–6400.
61. Deng, Y. & Roux, B. (2008). Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. *J. Chem. Phys.* **128**, 115103.
62. Vaitheeswaran, S., Yin, H., Rasaiah, J. C. & Hummer, G. (2004). Water clusters in nonpolar cavities. *Proc. Natl Acad. Sci. USA*, **101**, 17002–17005.
63. Rasaiah, J. C., Garde, S. & Hummer, G. (2008). Water in nonpolar confinement: from nanotubes to proteins and beyond. *Annu. Rev. Phys. Chem.* **59**, 713–740.
64. Namslauer, A., Lepp, H., Brändén, G., Jasaitis, A., Verkhovsky, M. I. & Brzezinski, P. (2007). Plasticity of proton pathway structure and water coordination in cytochrome *c* oxidase. *J. Biol. Chem.* **282**, 15148–15158.
65. Namslauer, A., Pawate, A. S., Gennis, R. B. & Brzezinski, P. (2003). Redox-coupled proton translocation in biological systems: proton shuttling in cytochrome *c* oxidase. *Proc. Natl Acad. Sci. USA*, **100**, 15543–15547.
66. Torrie, G. M. & Valleau, J. P. (1974). Monte-Carlo free energy estimates using non-Boltzmann sampling—application to subcritical Lennard–Jones fluid. *Chem. Phys. Lett.* **28**, 578–581.
67. Straatsma, T. P., Berendsen, H. J. C. & Postma, J. P. M. (1986). Free-energy of hydrophobic hydration—a molecular-dynamics study of noble-gases in water. *J. Chem. Phys.* **85**, 6720–6727.
68. Zacharias, M., Straatsma, T. P. & McCammon, J. A. (1994). Separation-shifted scaling, a new scaling method for Lennard–Jones interactions in thermodynamic integration. *J. Chem. Phys.* **100**, 9025–9031.
69. MacKerell, A. D. J., Bashford, D., Bellott, M., Dunbrack, R. L. J., Evanseck, J. D., Field, M. J. *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
70. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. E. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
71. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
72. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical-integration of Cartesian equations of motion of a system with constraints—molecular-dynamics of N-alkanes. *J. Comp. Phys.* **23**, 327–341.
73. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. (1995). Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **16**, 1339–1350.