Supporting Information: Changing hydration level in an internal cavity modulates the proton affinity of a key glutamate in Cytochrome c Oxidase

Puja Goyal [*], **Jianxun Lu** [†], **Shuo Yang** [*] **M. R. Gunner** [†] **Qiang Cui** [*]

[*]Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin-Madison, 1101 University Avenue, Madison, WI 53706, and [†]Department of Physics, City College of New York, 160 Convent Ave., New York NY 10031

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Detailed Descriptions of Computational Methods

Molecular models for CcO and force field parameters. The Cytochrome c Oxidase (CcO) models are obtained starting with the four subunit, X-ray structure for the *R. sphaeroides* CcO in the fully oxidized state (PDB entry 1M56) at 2.3 Å resolution [1]. The comparison of different structures show little change near Trp172, the region of interest here (see Fig.S1a, also Table S2). Rather the main difference is that a hydrogen bond between Tyr 288 in the active site and the OH of the heme a3 farnesyl group [2] is not found in all structures. The current hypothesis is that this hydrogen bond controls access to the K-pathway [3], which is only open in the reductive half cycle. We focus on the reactions in the oxidative half cycle where the K-pathway, which donates protons to the BNC, is closed. However, the modulation of the Glu286 proton affinity for proton transfer to the BNC is proposed to be the same for all four steps of the redox cycle. Pumped protons always move into CcO through the D channel via Glu286 [4].

The co-factors in CcO include Cu_A and Heme a. The active site is made up of the Binuclear Center (BNC: Heme a_3, Cu_B) and the nearby Tyr 288 of chain A. Heme a_3 has 1 His ligand (His 419) and an open site to bind feryl oxygen, a hydroxyl or water. Cu_B has 3 His ligands (His 284, His 333, His 334) and an open site to bind a hydroxyl or water (Fig.S1a).

Molecular dynamics (MD) simulations are carried out with two different set-ups. PBC, periodic boundary calculations, are used for longer (15-50 ns) unrestrained MD with explicit solvent and a full atomistic lipid membrane. GSBP, the generalized solvent boundary potential [5], is used for shorter (1-3 ns), local MD simulations where a region that contains about 8,000 atoms around Glu286 is free to move (see below for details).

Simulations prepare structures in different CcO assigned protonation and redox states, which are labeled with a 5 character notation such as PDD-RO; the first three letters indicate the protonation state (Protonated or Deprotonated) of Glu286, the propionate D of heme a_3 ($PRDa_3$) and the ligand of Cu_B (hydroxide (D) or water (P)). The last two letters indicate the redox state (Reduced or Oxidized) of heme a and Cu_B, respectively. In all calculations the ferryl iron of heme a_3 is bound to an oxygen ($Fe^{4+}=O^{2-}$) and Cu_A is oxidized. The protonation states of the titratable groups from previous multiconformation continuum electrostatics (MCCE) results [6] on CcO are used as listed in the footnote of Table S4 and are consistent with the MCCE calculations presented here. His 334 is assumed to be neutral, as supported by our previous pK_a analysis [7].

Two sets of force fields are used for the metal co-factors and active site residues (heme a, Cu_A, BNC and their ligand residues) in the MD simulations. The "Johansson-set", developed by Johansson et al. [8] (simulations are labeled by a "j" following the state specification (e.g., PDD-OOj)), and the Ghosh set developed locally in our previous work [7](simulations are labeled by a "g" (e.g., PDD-ROg)). Most unrestrained, PBC simulations use the Johansson-set of parameters, while all GSBP simulations use the Ghosh set; several PBC simulations have been carried out with the Ghosh-set parameters to demonstrate the general robustness of the results (Fig.S4). It should be noted that Tyr 288 of subunit A is in the deprotonated, negatively charged state in the Johansson-set parameters, while it is in the protonated, neutral state in the Ghosh-set parameters. Thus, the net charge of the CcO active site (consisting of Heme a_3, Cu_B and Tyr 288) plus heme a in the $\mathbf{P}_R$ (PDD-OO) state with the Johansson parameters is identical to that of the PDD-RO state with the Ghosh parameters (i.e., PDD-OOj=PDD-ROg). We note that in the 1M56 crystal structure, the Fe atom of heme a and the Tyr288 sidechain O atom are located at distances of 13.5 Å and 12.5 Å, respectively, from the carboxylate C atom of Glu286. Hence the precise location of the electron should make little difference to the results. For a summary of all simulated states, see Table S1.

Unrestrained, PBC simulations with explicit membrane and bulk solvent. Four protonation states of the key groups are studied with unrestrained MD:

- The $\mathbf{P}_R$ (PDD-OO) state represents the protein before the proton is transferred from the protonated Glu286 to the the D-propionate of heme a_3 ($PRDa_3$), which we assume to be the proton loading site (PLS) here. Cu_B in its cupric state is bound to a hydroxide (Cu^{2+}-OH^-).
- The $\mathbf{P}'_R$ (DPD-OO) state is the intermediate following $\mathbf{P}_R$. The proton has moved to the $PRDa_3$ from Glu286. Thus, we assume that loading the PLS with the pumped proton precedes proton transfer to the substrate in the BNC. Although this model is also considered by many researchers in the field [4, 9, 10], the $\mathbf{P}'_R$ state has not been directly observed.
- The $\mathbf{P}''_R$ (PPD-OO) state, in which both Glu286 and $PRDa_3$ are in the protonated (charge-neutral) state can be considered as the state following $\mathbf{P}'_R$ in the pumping cycle where Glu286 has been reprotonated after giving a proton to the $PRDa_3$. This state also allows us to see if

---

Reserved for Publication Footnotes

the transition from small to big cavity form is dependent on the protonation state of PRDa$_3$ or Glu286.

- The $'\mathbf{F}$ (DPP-OO) state retains a deprotonated Glu286 and has a protonated PRDa$_3$ and a water bound to the cupric Cu$_B$ (Cu^{2+}-HOH). Thus, the pumped and chemical protons have been transferred to the PLS and the BNC. This state is most vulnerable to the proton back-flow, if Glu286 is not rapidly reprotonated through the D channel [11].

The setup of the PBC simulation is similar to that reported in our recent study [12]. The initial structure is prepared by first patching the protein into the desired redox state, and then embedding it into a pre-equilibrated lipid bilayer. Lipid molecules with head group atoms within 3 Å from the protein are removed, leading to 326 DPPC and 6 POPE molecules; the POPE molecules are included here because they were resolved in the original crystal structure [1]. Water molecules are then added to solvate the system to obtain a rectangular unit cell with dimensions of 120 Å×120 Å×130 Å. The system contains 17,701 protein atoms, 175 water molecules inside the protein and 40,288 water in the bulk. Next, 116 potassium and 109 chloride ions are added to make the system charge neutral with a near physiological salt concentration. CHARMM22 force field [13] (with the CMAP corrections [14] included) is used for the standard protein residues, while lipid molecules are treated using the CHARMM36 force field for lipids [15]. Electrostatic interactions are calculated using Particle Mesh Ewald [16], and van der Waals interaction using a cutoff scheme with the switch function turned on between 10 and 12 Å. All bonds involving hydrogen are constrained using LINCS [17] to allow a 1 fs time step. The system is coupled to the Nose-Hoover thermostat to maintain a constant temperature of 323 K and pressure control is achieved using the Parrinello-Rahman extended ensemble pressure coupling [18]. All calculations are done using Gromacs version 4.5.5 [19].

As summarized in Table S1, most PBC simulations are on the order of 15-50 ns, and independent trajectories are run for several states (e.g., $\mathbf{P}_R$ and $'\mathbf{F}$). Most properties of interest converge after ∼10ns, although subtle differences are also observed among independent simulations for the same state. The multi-subunit protein remains structurally stable throughout the simulations; e.g., the Cα RMSD for subunit I, which holds all important metal co-factors and the active site, is typically below 1.5 Å relative to the starting crystal structure.

Local, GSBP MD simulations. The local, GSBP MD simulations start with either the 1M56 crystal structure, which has a small cavity, or with a $'\mathbf{F}$ state structure generated with unrestrained PBC MD simulations, which has a large cavity (see Table S1). As these simulations constrain some part of the protein, the cavity size and number of waters bound near Glu286 do not undergo significant changes during the simulations. Thus the GSBP simulations allow analysis of the proton affinity of Glu286 in microscopic simulations within structures with a defined cavity size and hydration and fixed ionization states of PRDa$_3$ or other groups.

In the GSBP setup, the system is partitioned into inner and outer regions (see Fig.S1b) and only the microscopic dynamics of the inner region are followed explicitly; the contributions from the outer region, including bulk solvation effects, are approximated at the Poisson-Boltzmann (PB) continuum electrostatics level. We use a rectangular boundary condition for the GSBP setup with dimensions of 40 Å×38 Å×56 Å for the inner region centered at Glu 286. The rest of the protein is treated as the outer region. With the 1M56 structure as the starting configuration, the outer region has 10,562 atoms that are held fixed, while 8,447 atoms (7,922 protein and 525 waters) that belong to the inner region are explicitly simulated with molecular dynamics. Protein atoms at the boundary of the inner and the outer regions are constrained according to the previously described protocol [5, 7]. Thus, in the GSBP-IM56 and GSBP-1M56(+9w) simulations, the outer region keeps the crystal coordinates. To be consistent with the GSBP protocol, the extended electrostatics model [20] is used to treat the electrostatic interactions among inner region atoms, where interactions beyond 12 Å are treated with multipolar expansions that include the dipolar and quadrupolar terms.

During the molecular dynamics simulations, all bonds involving hydrogen are constrained with SHAKE to allow a 1 fs time step. Langevin dynamics with a temperature bath of 300 K is carried out for mobile inner region atoms within 4 Å of the boundary between inner and outer regions (i.e. in the so-called buffer region), while Newton's equations of motion are solved for the rest of the mobile inner region atoms. Protein atoms in the buffer region are harmonically restrained with force constants determined directly from the B-factors in the 1M56 PDB file [21]. The entire system is heated to 300 K and equilibrated for at least 100 ps prior to any production simulations. All GSBP calculations [5] are carried out using the program CHARMM [22].

The number of water molecules in the GSBP simulations that start with the crystal structure is determined by Grand Canonical Monte Carlo (GCMC) simulations [23] as described in Ref. [7]. In GSBP-1M56, five water molecules are added to the cavity [7], which is empty in the crystal structure. In the GSBP-1M56+9w simulations, in addition to the water molecules in the GSBP-1M56 setup, 6 water molecules are added near Glu286 and in the D-channel, while 3 are added near PRDa$_3$: 8 water molecules are first introduced by inspection near Glu286 (near Trp172 and in the D-channel). After 13 cycles of 10,000 steps of GCMC and 10,000 steps of MD (2 fs time-step) each, 6 of the added water molecules are retained. In addition, 3 water molecules are added between PRDa$_3$ and Mg^{2+}.

In the GSBP simulations that start with a snapshot from the $'\mathbf{F}$ state PBC simulation (GSBP-PBC$'\mathbf{F}$), the cavity is larger and occupied by a significantly larger number of water molecules. These features are retained during the subsequent MD simulations (see below).

Additional analyses of structural/hydration features from PBC and GSBP simulations

PBC results are not sensitive to the force field parameters of metal co-factors. To test whether the observations from the MD simulations are sensitive to the force field used for the metal co-factors and active-site residues (heme a, Cu$_A$, BNC and their ligand residues), we have carried out additional PBC simulations using the Johansson [8] and Ghosh [7] sets of parameters for two enzyme states ($\mathbf{P}_R$=PDD-OOj vs. PDD-ROg and $'\mathbf{F}$=DPP-OOj vs. DPP-ORg); the comparison for the PDD-OOj vs. PDD-ROg is shown in Fig.S4 as an example. The qualitative results concerning the level of hydration and distance distributions between key residues are very consistent among the different sets of simulations. Thus, the protonation of PRDa$_3$ leads to the displacement of Trp172 away from PRDa$_3$ and a higher level of hydration in the hydrophobic cavity, independent of the parameters used to describe a given BNC redox state.

We note that the current simulations use a standard nonpolarizable force field. There have been discussions regarding the potential importance of including electronic polarization effects for charged residues in the protein interior; e.g., it was proposed that the partial charges for charged residues should be reduced as a simple scheme to approximate the effect of electronic polarization [24,25]. We have explored this scheme in the PBC simulations for the $\mathbf{P}_R$ state by scaling down (by a factor of $\sqrt{2}$ as recommended in Refs. [24, 25]) the partial charges for Arg481 and PRDa$_3$, which form the key salt-bridge in the active site. As also shown in Fig.S4, the qualitative results for the level of hydration and key distance distributions remain unchanged, again demonstrating the robustness of the observations.

Results for the $\mathbf{P}''_R$ state show PRDa$_3$ not Glu286 protonation is important. The results in the main text show the expansion of the cavity and increase of hydration occur in both the $\mathbf{P}'_R$ and $'\mathbf{F}$ states. These states both have a protonated PRDa$_3$ and deprotonated Glu 286, while the small cavity $\mathbf{P}_R$ state has a deprotonated PRDa$_3$ and protonated Glu 286. To establish which protonation state change is responsible for the opening of the hydrophobic cavity, we have carried out PBC simulations for the $\mathbf{P}''_R$ state, in which the PRDa$_3$ is protonated while Glu286 is charge neutral. As shown in Fig.S5, the general features resemble those observed for the $\mathbf{P}'_R$ and $'\mathbf{F}$ states. Therefore, these simulations provide support for the proposal that protonation of PRDa$_3$, rather than deprotonation of Glu 286, leads to the displacement of Trp172 and expansion of the hydrophobic cavity.

GSBP simulations maintain cavity size of the input structure. As shown in Figs.S6-S8, due to the fact that part of the loop that bears Trp172 is held fixed in the GSBP set up (Fig.S1b), the properties of the hydrophobic cavity depend on the initial structure in the local MD simulations. Thus, in GSBP simulations where PRDa$_3$ is protonated, the size and hydration level of the cavity in the GSBP-1M56, DPD-ROg simulations remain more similar to those observed in the PBC simulations for the $\mathbf{P}_R$ state. Increasing the number of water molecules near Glu286 but still using the crystal structure for the starting coordinates (1M56+9w) does not lead to major differences in the active site structures (Fig.S7), with little change in the salt-bridge between Arg481/PRDa$_3$ or the position of Trp172. The minor exception is that the Glu286-PRDa$_3$ distance shifts to longer distances (by ∼1 Å) as compared to GSBP-1M56 results (Fig.S7a). This is likely because several water molecules are also added to the top of the D-channel, and the deprotonated Glu286 has a significant population of the "downward" orientation, which is further from the PRDa$_3$(see Fig.S6b).

When a snapshot from PBC$'\mathbf{F}$ simulation is used for the initial coordinates of the DPD-ROg GSBP simulations, the cavity properties remain close to the initial PBC$'\mathbf{F}$ state structure. This is clearly illustrated by the distance between Glu286 and PRDa$_3$ (Fig.S7a), the position of Trp172 relative to PRDa/PRDa$_3$ (Figs.S7d-f) and the level of hydration near Glu286 and PRDa$_3$ (compare Figs.S8a-b and Figs.3-4 in the main text). An interesting observation is that the GSBP-PBC$'\mathbf{F}$ simulation has a high probability of forming a direct hydrogen bonding interaction between the protonated PRDa$_3$ and deprotonated PRAa$_3$ (Fig.S6d), despite the high level of solvation near PRDa$_3$ (Fig.S8b).

In summary, comparing the PBC and various GSBP simulations illustrates that, the structural and solvation properties of the hydrophobic cavity, are generally consistent as long as the loop bearing Trp172 is allowed to move. These results illustrate the importance of allowing full protein flexibility in describing the response to redox/titration state changes; with only local MD simulations the loop does not open. On the other hand, the results support using the local, GSBP setup to dissect the impact of cavity size/hydration level on the proton affinity of Glu286 with microscopic pK$'_7$ simulations.

Local dielectric constant. To better understand local dielectric properties of CcO, we focus on four regions: Glu286, PRDa$_3$, Ser200 in the D-channel and Asp132 at the entrance of the D-channel. In each case, we select a spherical region of 10 Å radius (by residue) around a specific atom in the residue of interest (see Table S3) and then calculate the fluctuation of the total dipole moment associated with the spherical region (ΔM_p^2) and thus the corresponding G factor. Following a Kirkwood-Fröhlich model [26, 27], the G factor is used to estimate the effective dielectric constant, ϵ_1, for the region of interest (also see discussions in Ref. [28]):

$$\begin{aligned} G &= \frac{\Delta M_p^2}{k_B T\ r_1^3} \\ &= \frac{(\epsilon_1-1)[(1+2\epsilon_2)(2\epsilon_W+2\epsilon_2)-2(r_1/r_2)^3(\epsilon_W+\epsilon_2)(1-\epsilon_2)]}{(\epsilon_1+2\epsilon_2)(2\epsilon_W+2\epsilon_2)-2(r_1/r_2)^3(\epsilon_W-\epsilon_2)(\epsilon_1-\epsilon_2)}, \end{aligned} \qquad \mathbf{[1]}$$

where r_1 is the radius of the region of interest (10 Å), r_2 is the effective radius of the surrounding protein; ϵ_W and ϵ_2 are the dielectric constant for bulk water and the protein in the surrounding region. Since CcO is large, $r_1 << r_2$, thus the simplified expression becomes,

$$G = \frac{\Delta M_p^2}{k_B T\ r_1^3} = \frac{(\epsilon_1-1)[(1+2\epsilon_2)]}{(\epsilon_1+2\epsilon_2)}. \qquad \mathbf{[2]}$$

Thus the only parameter we need to specify for determining ϵ_1 is the value of ϵ_2, for which we explore the use of two "limiting" values for ϵ_2: 4 and 20. Test calculations indicate that for regions near Glu286, PRDa$_3$ and Ser200, using $\epsilon_2 = 4$ gives more stable results for ϵ_1, while for Asp132 only $\epsilon_2 = 20$ gives sensible results for the local ϵ_1. These observations are qualitatively consistent with the fact that Asp132 is close to the mouth of the D-channel and therefore surrounded (towards the N-side of the membrane) by more water molecules and flexible residues. The convergence of the computed G factor is reached typically after 10 ns.

In general, as shown in Table S3, the local dielectric constant (ϵ_1) is fairly low except for the region near Asp132, which is fairly close to the bulk solvent. For the small cavity, $\mathbf{P}_R$ state, the estimated ϵ_1 is about 4 for regions near Glu286, PRDa$_3$ and Ser200; for Asp132, the value is about 15. As the titration state of the enzyme changes, the level of solvation and local flexibility also change and therefore the estimated ϵ_1 varies. In the large cavity, $'\mathbf{F}$ state, for example, the estimated ϵ_1 near Glu286 and PRDa$_3$ increases to 7-9. The trend is consistent with the higher level of solvation near these residues in the PBC simulations of the $'\mathbf{F}$ state relative to the other states.

Calculation of pK$'_7$ and pK$_a$ for Glu286

The true pK$_a$ is computed with the MCCE approach with titration calculations that keep all protonation states in equilibrium with the structure and fixed redox and protonation states as a function of pH [29]. However, much of the results describing the proton affinity of Glu 286 is reported as pK$'_7$, a value that can be computed by multiple methods as it fixes the protonation states of all other residues at equilibrium at pH 7. Thus, it does not rely on methods that can keep the system in equilibrium with the pH (see Table S4). The pK$'_7$ estimates the free energy of deprotonation of the acid at the physiological pH. The pK$'_7$ analysis with a simple single conformer continuum electrostatics (SCCE) approach [30,31] focuses on the change of electrostatic free energy when Glu286 adopts different protonation states. The simplicity of the model allows an investigation of effects associated with key parameters of such calculations, i.e., the dielectric constant for the protein (ϵ_{prot}) and the hydrophobic cavity (ϵ_{cav}). The more sophisticated MCCE approach [32] goes beyond the SCCE approach as it samples the conformations and protonation states of titratable groups in the protein in a single Monte Carlo simulation. MCCE has been tested for reduction potential and pK$_\mathrm{a}$ calculations in a large number of proteins [29,33–36]. The microscopic QM/MM-TI pK$'_7$ calculation [7,37] treats the nearby environment of the titratable group in a microscopic fashion and therefore does not need phenomenological parameters such as ϵ_{cav}. On the other hand, the results of such simulations are sensitive to the degree of sampling and description of electrostatics/polarization effects [37,38]. In the following, we summarize in more detail the calculation of pK$'_7$ by the various methods.

Definition of pK$'_7$. The free energy of ionization of an acid (A) equilibrated in the protein at pH 7 is:

$$\Delta G(AH \rightarrow A^-) = 1.36 \times (pK'_7 - 7)\mathrm{kcal/mol} \qquad [3]$$

$$pK'_7 = pK_{a,sol} + \Delta\Delta G_{prot}/(1.36 \times k_B T) \qquad [4]$$

The $\Delta\Delta\mathrm{G}_{prot}$ is the difference in interaction of the protein with the ionized and neutral acid leading to the shift in the proton affinity. It is determined with all other residues at their equilibrium protonation state at pH 7. In MCCE, Monte Carlo simulations sample distributions of protonation micro states so residues can have fractional ionization. In SCCE and QM/MM-TI calculations, the protonation states are fixed at integer values representing one protonation microstate of the protein. See Table S4 for the list of the residues that are not in their standard ionization states at pH 7 as determined by MCCE calculations [6].

QM/MM-TI calculations of pK$'_7$: methods. Each method of calculating the pK$'_7$ of Glu286 shows a significant loss of proton affinity when CcO has a large cavity and when the PRDa$_3$ is protonated. The QM/MM-TI calculations generally give a very high pK$'_7$. The SI explores several ways to investigate the sensitivity of the results to various methodological details. Details of the QM/MM Thermodynamic Integration (QM/MM-TI) pK$'_7$ calculation scheme can be found in our previous works [37, 39]. Briefly, the dual-topology single-coordinate based TI approach (DTSC-TI) is used in a QM/MM framework where the titratable group is treated with the SCC-DFTB approach [40, 41]. The total free energy of deprotonation is dominated by $\Delta\mathrm{G}^{(1)}_{E\cdot A(H>D)}$, the electrostatic free energy of converting the acidic proton to a dummy atom(D); here E represents the enzyme environment and AH is the titratable acidic residue (Glu286). The free energy derivative for a given λ window is expressed as:

$$\left(\frac{\partial \mathrm{G}^{(1)}_{E\cdot A(H>D)}}{\partial\lambda}\right)_\lambda = \langle \mathrm{U}^{QM/MMelec}_{E\cdot AD^-} - \mathrm{U}^{QM/MMelec}_{E\cdot AH} + \mathrm{U}^{bonded}_D \rangle_\lambda, \qquad [5]$$

where

$$\mathrm{U}^{QM/MMelec.}_{E\cdot AX} = \langle \mathit{\Psi}_{E\cdot AX} | \widehat{H}^{QM}_{AX} + \widehat{H}^{QM/MMelec}_{E\cdot AX} | \mathit{\Psi}_{E\cdot AX} \rangle. \qquad [6]$$

Eq.5 represents the energy gap between the final (E·AD$^-$) and initial (E·AH) states averaged over the configurations sampled in a particular λ window. The principal contribution to the energy gap comes from the QM/MM electrostatic terms while the bonded terms between the dummy atom and E·A$^-$ (represented by U^{bonded}_D) are in practice very small. The total electrostatic free energy of deprotonation is given by

$$\Delta\mathrm{G}^{(1)}_{E\cdot A(H>D)} = \int_0^1 \left(\frac{\partial \mathrm{G}^{(1)}_{E\cdot A(H>D)}}{\partial\lambda}\right)_\lambda d\lambda. \qquad [7]$$

Instead of calculating the absolute pK$_\mathrm{a}$ which requires estimation of the solvation free energy of a proton, which is difficult to measure or compute accurately, we calculate the pK$_\mathrm{a}$ shift relative to acetic acid in solution, with an experimental pK$_\mathrm{a}$ of 4.74 [42]. This also simplifies the calculation by helping to cancel out other contributions, like the zero-point energy difference between the protonated and deprotonated states as well as van der Waals interactions involving the acidic proton [37,39].

The QM/MM-TI simulations are carried out in the GSBP framework [7]; this assumes that the key factors that dictate the pK$'_7$ value of Glu 286 are local interactions. Importantly, by using different initial structures for the GSBP setup (see Table S6), one can explore the impact of cavity properties on the pK$'_7$ value of this key residue.

TI-US: TI coupled to umbrella sampling in the energy gap coordinate

Numerous studies have indicated that conducting extensive sampling is crucial to the reliability of microscopic pK_a calculations [38,43,44]. To explore effects of enhanced sampling of the degrees of freedom tightly coupled to the titration of Glu286, we couple the TI protocol with umbrella sampling in the energy gap coordinate. This has been used in different forms in several previous studies [43,45–49], perhaps most notably by Warshel and co-workers [49] who used the EVB potential function and by Yang and co-workers [43] to overcome "hidden barriers" in alchemical free energy simulations. In the specific context of pK$'_7$ calculations for Glu286, many motions are likely to respond to the titration process, including the rotation/translation of water molecules in the vicinity of Glu286 and the re-orientations of Glu286 and PRDa$_3$. Therefore, using the energy gap between the protonated and deprotonated states as a collective coordinate is more effective than biasing a specific set of conformational degrees of freedom. The specific form of the energy gap in the current DTSC-TI simulation is given by,

$$\begin{aligned} \Delta U &= \mathrm{U}^{QM/MMelec}_{E\cdot AD^-} - \mathrm{U}^{QM/MMelec}_{E\cdot AH} + \mathrm{U}^{bonded}_D \\ &= \Delta U^{QM/MMelec} + \mathrm{U}^{bonded}_D, \end{aligned} \qquad [8]$$

in which the bonded terms associated with the dummy atom (U^{bonded}_D) are expected to be small in magnitude and therefore only $\Delta U^{QM/MMelec}$ is used in the umbrella sampling calculations.

For each λ window in the TI, umbrella sampling along $\Delta U^{QM/MMelec}$ is followed by a WHAM analysis [50] to obtain the potential of mean force (PMF) and the unbiased probability distribution of $\Delta U^{QM/MMelec}$, given by $\rho_\lambda(\Delta U^{QM/MMelec})$. This probability distribution is then converted to the probability distribution of the total energy gap, $\rho_\lambda(\Delta U)$, using Eq.9. $\rho_\lambda(\Delta U|\Delta U^{QM/MMelec})$, which is the conditional probability of ΔU given a particular value of

$\Delta U^{QM/MMelec}$. This is estimated by combining data from all the umbrella windows for a particular λ window.

$$\rho_\lambda(\Delta U) = \int d\Delta U^{QM/MMelec} \quad \rho_\lambda(\Delta U|\Delta U^{QM/MMelec}) \, \rho_\lambda(\Delta U^{QM/MMelec}) \qquad \textbf{[9]}$$

Knowledge of $\rho_\lambda(\Delta U)$ allows the average energy gap $\langle \Delta U \rangle_\lambda$, which is equal to the free energy derivative $(\partial G^{(1)}/\partial\lambda)_\lambda$, to be calculated. Giving the scheme which couples umbrella sampling to thermodynamic integration as TI-US, $\rho_\lambda(\Delta U)$ and $(\partial G^{(1)}/\partial\lambda)_\lambda$ obtained from TI-US and conventional TI can be compared. Any significant differences observed in results from the two schemes highlight limitations in configurational sampling with conventional TI. Following Eq.7, the electrostatic free energy of deprotonation and hence the pK'_7 can also be compared between the two schemes.

Bennett Overlapping Histograms (BOH) analysis

BOH analysis for the TI and TI-US data provides another way to evaluate the statistical uncertainty of the estimated free energy changes for Glu ionization. Extending the BOH equations [45,51] to two arbitrary windows λ and λ' in the TI calculation, we get the following expressions:

$$\ln \frac{\rho_\lambda(\Delta U)}{\rho_{\lambda'}(\Delta U)} = (\lambda' - \lambda)\beta\Delta U - \beta\Delta G_{\lambda',\lambda} \qquad \textbf{[10]}$$

$$P_\lambda(\Delta U) = \ln \rho_\lambda(\Delta U) - \frac{1}{2}(\lambda' - \lambda)\beta\Delta U \qquad \textbf{[11]}$$

$$P_{\lambda'}(\Delta U) = \ln \rho_{\lambda'}(\Delta U) + \frac{1}{2}(\lambda' - \lambda)\beta\Delta U \qquad \textbf{[12]}$$

$$P_{\lambda'}(\Delta U) - P_\lambda(\Delta U) = \beta\Delta G_{\lambda',\lambda} \qquad \textbf{[13]}$$

The free energy difference between the λ and λ' windows ($\Delta G_{\lambda',\lambda} = G_{\lambda'} - G_\lambda$) can be estimated from the plateau region in the function $P_{\lambda'}(\Delta U) - P_\lambda(\Delta U)$ over the range of ΔU in which $\rho_{\lambda'}(\Delta U)$ and $\rho_\lambda(\Delta U)$ overlap. The absence of such a plateau region indicates sampling related problems, making BOH an independent graphical estimator for the convergence of free energy simulations.

The sum of $\Delta G_{\lambda',\lambda}$ values for pairs of adjacent windows in the TI/TI-US calculation yields the total free energy of deprotonation which can be compared to the value obtained using Eq.7. For properly converged simulations, the values obtained from the two methods should agree with each other (within statistical uncertainties).

Additional analysis of QM/MM-TI calculations of pK'_7. The main results of QM/MM-TI pK'_7 calculations are summarized in Table 1 of the main text and discussed in light of the CcO mechanism. Here we present additional technical analyses of these calculations such as their statistical convergence and the effects of the size of QM region, as well as provide additional information about interactions within CcO that contribute to the computed pK'_7 values.

Statistical analysis and sampling

Figure S9a shows the PMFs along the QM/MM electrostatic component of the total energy gap, $\Delta U_{QM/MMelec.}$, for different λ windows. The PMFs are largely parabolic in nature, except for $\lambda = 0.25$, thus not showing any significant signature of hidden barrier in the "orthogonal space". Figure S9b shows how the probability distributions of the total energy gap, ΔU, for the various λ windows differ between TI-US and conventional TI. The differences in the distributions are very minor for most windows except for λ=0.25, for which even the peak positions in the TI and TI-US distributions differ significantly. This is consistent with the flatter PMF along $\Delta U_{QM/MMelec.}$ for $\lambda = 0.25$, and indicates that a diverse set of conformations are important but only accessible with the TI-US sampling (see below).

Consistent with the trends in the energy gap distributions, Table S7 (top table) shows that for the λ=0.25 window, the free energy derivatives from TI-US and TI differ by ~6.5 kcal/mol while the difference for the other windows is smaller (2-3 kcal/mol). Using the Linear Response Approximation (LRA) [52], the estimated $\Delta G^{(1)}$ from TI-US is found to be ~2 kcal/mol lower than that from conventional TI. Thus, the difference in the pK'_7 estimated by the two methods should be less than 2 pH units. There is a slight improvement in the R^2 value of the linear fit to the free energy derivatives with TI-US. This suggests that with LRA, despite some significant differences in the free energy derivatives in the two methods of sampling, the final $\Delta G^{(1)}$ does not vary much. LRA is thus shown to be useful since it reduces the importance of a λ window with insufficient sampling in conventional TI in the final $\mathrm{pK_a}$ value.

Figure S10a-b shows plots of the function $P_{\lambda'}(\Delta U) - P_\lambda(\Delta U)$ (Eq.13) for several pairs of windows. The functions have a clear plateau (more obvious for TI-US simulations), with standard deviations from the average ranging of ~0.4-1.8 kcal/mol (Table S7). The BOH approach is used to obtain an independent estimate of $\Delta G^{(1)}$, employing the overlap region between the energy gap probability distributions for adjacent λ windows. Table S7 shows that BOH estimates of $\Delta G^{(1)}$ (bottom table) from TI and TI-US agree well with previous estimates (using the $(\partial G^{(1)}/\partial\lambda)_\lambda$ values and LRA, top table) from TI and TI-US, respectively.

Justifying the choice of using $\Delta U_{QM/MMelec.}$ as an effective coordinate for enhanced sampling of motions coupled to titration, we find that biasing the energy gap brings about several structural changes. Fig. S10c-d shows that, in TI-US windows where $\Delta U_{QM/MMelec.}$ is biased to values not sampled (or sampled with a very low probability) in standard TI simulations, the probability distribution of Glu286-PRD$\mathrm{a_3}$ separation is different from that in unbiased simulations. In general, the smaller the energy gap the longer the Glu286-PRD$\mathrm{a_3}$ distances. Fig. S10d reflects the rise in the level of solvation of Glu286 on lowering the energy gap and vice versa.

In short, the analysis of the TI and TI-US simulations suggests that although umbrella sampling along the energy gap coordinate does lead to interesting effects in some λ windows, the overall impact on the estimated pK'_7 value for Glu286 is in the range of 1-2 pH units. This result along with the reasonably behaved $P_{\lambda'}(\Delta U) - P_\lambda(\Delta U)$ plots suggests that the level of sampling we have performed appear to be quite adequate for the purpose of understanding factors that determine the proton affinity of Glu286. On the other hand, it remains possible that the apparent 'convergence' of our simulations is partially due to the use of the GSBP protocol, which does not allow collective motions of the protein; fairly subtle structural rearrangements at the backbone level have been proposed to contribute to $\mathrm{pK_a}$ of buried residues in proteins [53,54]. This possibility is explored to some degree by using a protein structure taken from a PBC simulation (GSBP-PBC$'$**F**); as shown in Table S6, the GSBP-PBC$'$**F** and GSBP-1M56+9w give similar pK'_7 values, suggesting that the level of hydration and local electrostatics play the dominant role in determining the proton affinity of Glu 286.

Effect of the size of the QM region

In general, the treatment of interactions between QM and MM atoms is important for the accuracy of QM/MM simulations [55–60]. Several authors have shown the limitation of using point-charge type of models for QM-MM electrostatics, especially when the QM region is highly charged; for the specific case of using SCC-DFTB for the QM, we have shown that a Klopman-Ohno (KO) approximation, which includes an approximate treatment of charge penetration effects, gives substantial improvement for the interaction between charged QM solutes and MM water [60]. Since parameters in the KO model were developed in Ref. [60] for a specific parameterization of SCC-DFTB, we have not used that approach here. Instead, we explore the dependence of the free energy derivatives in pK'_7 calculations on the size of the QM region by adding nearby water molecules into the QM region; we note that the SCC-DFTB variant used here has been shown [61] to give generally reliable hydrogen-bonding interactions when compared to high-level *ab initio* calculations.

As shown in Table S8, it is found that the average correction to the energy gap is different for λ=0.0 and λ=1.0 windows of a particular pK'_7 calculation, while also being different for different pK'_7 calculation sets. Fig. S11 shows that the level of solvation of the carboxylic acid is higher in the λ=1.0 window than in the λ=0.0 window for each pK'_7 calculation set, and for a particular λ window, the level of solvation increases in the order GSBP-1M56, GSPB-PBC′**F** and solution. It appears that the magnitude of the dependence on the size of the QM region correlates with the level of solvation.

The negative sign of the correction for the λ=0.0 window for all three pK'_7 sets indicates that enlarging the QM region helps lower the energy gap by increasing the electronic polarization of the negatively charged state of the carboxylic acid. The magnitude of the correction is smaller for the GSBP-1M56, XDD-ROg setup, consistent with the fact that Glu286 in this model is relatively less solvated. On the other hand, in the λ=1.0 window, the population of water molecules close to the carboxylate is higher and hence MM waters around the QM carboxylate cause over-polarization, thus underestimating the energy gap. When the water molecules around the carboxylic acid are treated as QM, the energy gap rises, and the rise decreases in the order of bulk < GSBP-PBC′**F** < GSBP-1M56.

The net correction of adding waters to the QM region to the free energy of deprotonation is relatively minor for the CcO simulations. It is ∼-1 and -0.4 pH units for the two cases analyzed here (Table S8). The effect on the solution reference, however, is quite notable, as it is about ∼2 pH units. Therefore, it seems that for protein sites that have a significantly different degree of solvation compared to solution, errors in the QM/MM interactions do not cancel well even for $\mathrm{pK_a}$ shift calculations. This further highlights the importance of carefully evaluating the accuracy of QM/MM interactions under different environments [62], an issue we are actively pursuing [60].

The analysis presented here suggests the computed, absolute value of the pK'_7 of Glu286 should be interpreted with care, even when derived from the $\mathrm{pK_a}$ shift relative to a solution reference. Nevertheless, the analysis also suggests that the benefit of extending the QM size in the current protein simulations is very limited. In this work, therefore, we limit to calculations with the small QM region including only Glu286 itself.

Perturbation analysis of residue contribution

To gain insight into the importance of a particular set of residues to the computed pK'_7 of Glu286, we carry out a set of "perturbative analyses" in which the energy gap is re-evaluated for snapshots from selected TI trajectories after some parameters of the model are modified. For example, the effect of a set of residues on the calculated pK'_7 of Glu286 can be estimated by setting their partial charges to zero and re-evaluating the free energy derivatives (by calculating the new energy gap) in the original trajectories for the different λ windows. Although the effects of individual residues are not strictly additive, because of the QM/MM framework, and the relaxation of the environment after the charge perturbation is not included. Thus, the results should be treated qualitatively [37,38]. However, this perturbative analysis can provide valuable information about how the environment influences the pK'_7 of Glu286.

Another approximation we use in these perturbative analyses is the linear response approximation (LRA), which has been observed to hold well in the DTSC-TI framework in previous [7, 37, 63] and current simulations. This is due largely to the fact that our DTSC-TI framework treats electrostatic and van der Waals components in two separate steps and here we focus on the electrostatic component, which is found to dominate $\mathrm{pK_a}$ shifts between protein and the solution reference [39]. In this way, changes in the free energy derivatives following a specific perturbation only need to be estimated for the λ = 0.0 and 1.0 windows.

To gain insights into residues that dictate the pK'_7 of Glu286, we focus on the GSBP-1M56, XDD-ROg setup. The charge of a particular group of residues is switched off and the free energy derivative is re-evaluated for the λ=0.0 and λ=1.0 windows, using 5,500 snapshots separated by 0.2 ps from the trajectories obtained with the original charges. Results from the perturbative analysis are summarized in Table S10. The analysis shows that the only residues with non-zero net charge in the vicinity of Glu286 are heme a, heme a_3, Arg481, Arg482 and Cu_B with ligands. Of these, the ones that raise the pK'_7 of Glu286 are heme a and heme a_3. This is expected given that the net charge on Arg481, Arg482 and Cu_B(with ligands) is +1 while the net charge on each reduced heme is -2, derived from the deprotonated propionic acids . More detailed decomposition of residual contributions (Table S11) shows the heme propionates are largely responsible for raising the pK'_7 of Glu286.

The use of scaled partial charges to estimate for the absence of electronic polarization

The current simulations use a standard non-polarizable force field. There has been discussions regarding the potential importance of including electronic polarization effects for charged residues in the protein interior [24, 25]. It has been proposed that reducing the partial charges for charged residues represents a simple scheme to approximate the effect of electronic polarization. Specifically, they suggested that to account for proper screening between such groups due to the high-frequency electronic dielectric constant of 2, charges of these groups should be scaled by $1/\sqrt{2}$.

Table S9 shows the effect that this proposed charge scaling has on the computed pK'_7 of Glu286 in the GSBP-1M56, XDD-ROg model. Three different charge-scaling schemes are tested, all of which have quite a modest effect of around 2 pH units on the pK'_7 of Glu286. It should be noted that the charge-scaling done here is not rigorously consistent with Stuchebrukhov et al.'s proposal since the charges of the titratable residue, Glu286, are not scaled in its deprotonated state (to maintain a meaningful definition of pK'_7). Therefore, we conclude that although it is interesting to test the effect of electronic polarization with more elaborate force field models, it is unlikely that including the effect will greatly change the computed pK'_7 of Glu286; this is due mainly to the fact that Glu286 is not in the immediate neighborhood of charged residues. Along this line, it is possible that the pK'_7 of $PRDa_3$ and therefore proton transfer energetics from Glu286 to $PRDa_3$ will be more sensitive to the description of electronic polarization, because $PRDa_3$ is very close to several charged groups, especially the pair of Arg residues (Arg481, 482) that forms salt bridges with propionates in the hemes.

Summary of the analyses of QM/MM-TI calculations for Glu286 pK'_7

In summary, we have carried out systematic analyses to probe how sampling, approximate treatment of electronic polarization and use of different QM regions impact the results of QM/MM-TI calculations for the Glu286 pK'_7. We find that these methodological details generally have rather limited effects on the computed pK'_7. Therefore, the general trends discussed in the main text, especially regarding the importance of the hydration level of the cavity and protonation state of $PRDa_3$ are found to be robust.

Continuum Electrostatics (CE) Calculations of pK'_7. In the SCCE-LRA and MCCE calculations, $\Delta\Delta G_{prot}$ (Eq.4) includes the shift in solvation energy when the ionized or neutral acid is moved from solution into the protein as well as the electrostatic interactions with the backbone, polar and ionized side chains and the CcO co-factors.

Both SCCE and MCCE calculations are carried out using 6 (MCCE) or 10-20 (SCCE) snapshots from the GSBP calculations subjected to local dynamics in different redox and protonation states. The input structures to generate the snapshots start from the 1M56 crystal structure (GSBP-1M56) and a snapshot of the unrestrained PBC calculations in the $'\mathbf{F}$ state with a large cavity (GSBP-PBC$'\mathbf{F}$). These restrained MD simulations maintain the initial cavity size, but allow local relaxation in the core of the protein around the imposed protonation states (Table S1, Fig S1). By using these snapshots, the impact of cavity properties on the pK'_7 of Glu 286 can be evaluated.

Both CE calculations use many of the same or similar parameters. All the explicit water molecules are deleted. A value of 1.4 Å is used for the water probe radius, with the dielectric boundary taken to be the contact+reentrant surface to determine the boundary between regions of low and high dielectric constant. The SCCE calculations, as in the GSBP calculations, use the atomic radii of Nina and Roux [64] to define the protein-solvent boundary, while the MCCE calculations use the Parse radii [65]. The solvent dielectric constant is 80 and the salt concentration is 150 mM. The SCCE calculations use a coarse cubic grid of 1.2 Å spacing and a fine grid of 0.4 Å spacing in focusing calculations. MCCE increases the scale by a factor of 2 in each focusing run to achieve a final scale of 0.5Å per grid.

Both calculations use a rectangular region at a low dielectric constant for the membrane. In the SCCE calculations the membrane thickness is 32 Å thick with a dielectric constant of 2, while in the MCCE calculations a 33 Å membrane is added with IPECE [35] with the dielectric constant being the same as that of the protein. For the SCCE calculations the protein dielectric constant is 2 or 4 (Table S12) while in MCCE values of 2, 4, 8 and 20 are explored (Table S13). In CE calculations cavities within the protein are generally given the dielectric constant of the solvent. In MCCE calculations, a value of 80 is always used for this. However, the PBEQ module in CHARMM allows SCCE calculations to change the dielectric constant of all the internal cavities which lie within the membrane width; values of 4, 80 or 9 have been explored (Table S12), the latter being an estimate based on PBC calculations (Table S3).

The SCCE calculations use the Ghosh set of charges [7] for the co-factors and standard CHARMM charges for the amino acids. The MCCE calculations use Parse charges for amino acids [65] and co-factor charges from Refs. [6] and [66]. The heme a_3 ferryl state uses the charges from Ghosh et al. [7]. As MCCE samples conformer positions as well as protonation states, additional parameters for non-bonded interactions are needed. Full Amber van der Waals parameters as well as an implicit van der Waals interaction with the implicit water are used as described in Ref. [32].

SCCE-LRA, Single Conformation Continuum Electrostatics with Linear Response

In the SCCE-LRA protocol, 10-20 snapshots are taken from the trajectories for the λ=0.0 (Glu neutral) and λ=1.0 (Glu ionized) windows of the corresponding microscopic, QM/MM-TI simulation using local, GSBP MD simulations of pK'_7. Electrostatic interactions with all 18,485 protein atoms belonging to all the four sub-units are included for Poisson-Boltzmann calculations, which are carried out using the PBEQ module [67] in CHARMM.

Four Poisson-Boltzmann calculations are carried out for each snapshot so as to compute the following: 1) $G_{GluH,prot}$, the total electrostatic energy of the system with Glu286 neutral, 2) $G_{GluH,aq}$, the total electrostatic energy for just the neutral Glu286 residue in a dielectric continuum with ϵ=80, 3) $G_{Glu^-,prot}$, the total electrostatic energy of the system with Glu286 negatively charged and 4) $G_{Glu^-,aq}$, the total electrostatic energy for just the negatively charged Glu286 residue in a dielectric continuum with ϵ=80. The pK'_7 shift relative to aqueous solution is then calculated as:

$$\Delta pK_a = \frac{(G_{Glu^-,prot} - G_{GluH,prot}) - (G_{Glu^-,aq} - G_{GluH,aq})}{2.303 k_B T} \qquad \mathbf{[14]}$$

To take into account the structural relaxations of the environment for different protonation states of Glu286, results for snapshots from the λ=0.0 and λ=1.0 windows are averaged in a LRA framework, which has been shown to be effective in previous continuum electrostatics studies of protein pK_a problems [31, 68].

MCCE, Multi Conformation Continuum Electrostatics

In MCCE, the backbone is rigid but side chains and polar hydrogens can be found in different conformations [32]. Atomic conformational degrees of freedom and residue protonation states are sampled in a single Monte Carlo simulation. Thus, the active site groups are fixed as described in Table S1 and sampled with the degrees of freedom described in S4. The positions and the protonation states of all groups that are not explicitly fixed come to equilibrium at the imposed pH.

As there can be several conformations of the ionized and neutral Glu286 as well as different conformation and protonation states of the surrounding residues the interaction energies are calculated taking into account the probability of each conformer. The pK'_7 is a mean field calculation of the energy using the averaged conformer occupancy in the Monte Carlo sampling. It thus misses the correlation of individual conformers of the Glu with conformers of other residues. For example, an unfavorable interaction may be seen between two conformers in the mean field analysis that would never be found within the Monte Carlo sampling as these conformers would never be found together in the same accepted microstate [32]. These mean-field errors are not likely to be very serious here. The MCCE calculations as well as the microscopic (QM/MM-TI) pK'_7 calculations show Glu286 is not strongly interacting with other titrating residues, other than $\mathrm{PRDa_3}$, whose effect is explicitly analyzed here (see Table S11). The calculation of the true $\mathrm{pK_a}$ rely on evaluating the true microstate energy so do not suffer from errors due to the mean field approach.

In the framework of Eq.4, the $\Delta\Delta\mathrm{G}_{prot}$ has contributions from both desolvation and pair-wise interactions,

$$\Delta\Delta G_{prot} = \Delta\Delta G_{desolv} + \Delta G_{pairwise}, \qquad \textbf{[15]}$$

where the desolvation energy compares the loss of solvation energy of the ionized and neutral Glu in the protein compared with that found in a medium with a dielectric constant of 80.

$$\begin{aligned}\Delta\Delta G_{desolv} = {} & (\Delta G_{solv,Glu^-,prot} - \Delta G_{solv,GluH,prot}) \\ & -(\Delta G_{solv,Glu^-,aq} - \Delta G_{solv,GluH,aq}),\end{aligned} \qquad \textbf{[16]}$$

and the difference in pairwise interactions of the ionized and neutral Glu with the rigid amide backbone dipoles (bkb) and the conformational sampled protein side chains and CcO cofactors (res).

$$\Delta G_{pairwise} = \Delta G_{res+bkb,Glu^-,prot} - \Delta G_{res+bkb,GluH,prot}. \qquad \textbf{[17]}$$

The pK'_7 values calculated with in Tables 1, S13 and 14 represent the average of multiple GSBP trajectories. For the local MD simulations initiated with large or small cavities, the results from with $\mathrm{PRDa_3}$ protonated or deprotonated in the GSBP simulations are averaged. For the $\epsilon_{prot} = 4$ calculations the results represent the average of 6 snapshots in each state. MCCE Monte Carlo sampling then defines the $\mathrm{PRDa_3}$ protonation state for each snapshot, which may be different than that in the GSPB simulation. For calculations with the other dielectric constants only one snapshot was averaged for each GSPB state. In Table 1 the results from trajectories with Glu ionized and neutral are averaged, as suggested by the LRA approximation. In Tables S13 and S14 the results with the different Glu ionization states are reported explicitly. The Glu proton affinity is ~1 pH unit higher when evaluated in MD trajectories where the Glu is neutral.

Table S14 breaks down the changes in proton affinity into those caused by changes in solvation energy or by changing interaction with backbone or side chains and cofactors (Eqn.S15-17). Protonating $\mathrm{PRDa_3}$ lowers the proton affinity of the Glu by 1.5-1.7 pH units, a value that is independent of the cavity size. There is no change in Glu solvation energy or interaction with the backbone dipoles. However, the change in cavity size influences the Glu pK'_7 predominantly by reducing the solvation penalty for ionization. Changes in both Glu conformation in the cavity and dielectric screening cause the ionized Glu to have less favorable interactions with the backbone and more favorable ones with the propionic acid. However, as the changes in pairwise interactions largely cancel, the shift in pK'_7 with cavity size is almost entirely due to changes in solvation of the Glu by the implicit solvent in the cavity.

1. **M. Svensson-Ek, J. Abramson, G. Larsson, S. Tornroth, P. Brzezinski, S. Iwata, 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).**
2. **S. Ferguson-Miller, C. Hiser, J. Liu, Gating and regulation of the cytochrome c oxidase proton pump, Biochim. Biophys. Acta - Bioeng. 1817 (2012) 489–494.**
3. **L. Qin, J. Liu, D. A. Mills, D. A. Proshlyakov, C. Hiser, S. Ferguson-Miller, Redox-dependent conformational changes in cytochrome c oxidase suggest a gating mechanism for proton uptake, Biochemistry 48 (2009) 5121–5130.**
4. **V. R. Kaila, M. V. Verkhovsky, M. Wikström, Proton-coupled electron transfer in cytochrome c oxidase, Chem. Rev. 110 (2010) 7062–7081.**
5. **W. Im, S. Bernéche, B. Roux, Generalized solvent boundary potential for computer simulations, J. Chem. Phys. 114 (7) (2001) 2924–2937.**
6. **Y. Song, E. Michonova-Alexova, G. Gunner, M., Calculated proton uptake on anaerobic reduction of cytochrome c oxidase: Is the reaction electroneutral?, Biochemistry 45 (2006) 7959.**
7. **N. Ghosh, X. Prat-Resina, M. Gunner, Q. Cui, Microscopic pk_a analysis of glu286 in cytochrome c oxidase (rhodobacter sphaeroides): Toward a calibrated molecular model, Biochem. 48 (2009) 2468–2485.**
8. **M. P. Johansson, V. R. I. Kaila, L. Laakkonen, Charge parameterization of the metal centers in cytochrome c oxidase, J. Comp. Chem. 29 (2007) 753–767.**
9. **J. P. Hosler, S. Ferguson-Miller, D. A. Mills, Energy transduction: proton transfer through the respiratory complexes, Annu. Rev. Biochem. 75 (2006) 165–187.**
10. **P. Brzezinski, R. B. Gennis, Cytochrome c oxidase: exciting progress and remaining mysteries, J. Bioenerg. Biomember 40 (2008) 521–531.**
11. **V. R. I. Kaila, M. I. Verkhovsky, G. Hummer, M. Wikstrom, Glutamic acid 242 is a valve in the proton pump of cytochrome c oxidase, Proc. Natl. Acad. Sci. U.S.A. 105 (17) (2008) 6255.**
12. **S. Yang, Q. Cui, Glu-286 rotation and water wire reorientation are unlikely the gating elements for proton pumping in cytochrome c oxidase, Biophysical Journal 101 (2011) 61–69.**
13. **A. D. MacKerell Jr. et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.**
14. **A. D. Mackerell, M. Feig, C. L. Brooks, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, J. Comput. Chem. 25 (11) (2004) 1400–15.**
15. **J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell Jr., R. W. Pastor, Update of the charmm all-atom additive force field for lipids: Validation on six lipid types, J. Phys. Chem. B 114 (2010) 7830–7843.**
16. **U. Essmann, L. Perera, M. L. Berkowitz, A smooth particle mesh ewald method, J. Chem. Phys. 103 (1995) 8577–8593.**
17. **B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, Lincs: A linear constraint solver for molecular simulations, J. Comput. Chem. 18 (12) (1997) 1463–72.**
18. **M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, J. Appl. Phys. 52 (12) (1981) 7182–90.**
19. **B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (3) (2008) 435–47.**
20. **R. Stote, S. D.J., M. Karplus, On the treatment of electrostatics interactions in biomolecular simulations, J. Chim. Phys. 88 (2419).**
21. **Brooks, C. L. III, M. Karplus, Solvent effects on protein motion and protein effects on solvent motion. dynamics of the active site region of lysozyme., J. Mol. Biol. 208 (1989) 159.**
22. **B. R. Brooks, C. L. B. III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, Charmm: The biomolecular simulation program, J. Comp. Chem. 30 (2009) 1545–1614.**
23. **H. J. Woo, A. R. Dinner, B. Roux, Grand canonical monte carlo simulations of water in protein environments, J. Chem. Phys. 121 (2004) 6392–6400.**
24. **I. V. Leontyev, A. A. Stuchebrukhov, Electronic continuum model for molecular dynamics simulations of biological molecules, J. Chem. Theory Comput. 6 (2010) 1498–1508.**
25. **I. V. Leontyev, A. A. Stuchebrukhov, Accounting for electronic polarization in non-polarizable force fields, Phys. Chem. Chem. Phys. 13 (2011) 2613–2626.**
26. **T. Simonson, D. Perahia, Internal and interfacial dielectric properties of cytochrome c from molecular dynamics in aqueous solution, Proc. Natl. Acad. Sci. U.S.A. 92 (1995) 1082–1086.**
27. **G. B. Goh, B. Garcia-Moreno E., C. L. Brooks III, The high dielectric constant of staphylococcal nuclease is encoded in its structural architecture, J. Am. Chem. Soc. 133 (2011) 20072–20075.**
28. **G. King, F. S. Lee, A. Warshel, Microscopic simulations of macroscopic dielectric constants of solvated proteins, J. Chem. Phys. 95 (1991) 4366–4377.**
29. **M. R. Gunner, J. Mao, Y. Song, J. Kim, Factors influencing the energetics of electron and proton transfers in proteins. what can be learned from calculations?, Biochim. Biophys. Acta. 1757 (2006) 942–968.**
30. **J. Antosiewicz, J. A. McCammon, M. K. Gilson, Prediction of ph-dependence properties of proteins, J. Mol. Biol. 238 (1994) 415–436.**

31. G. Archontis, T. Simonson, Proton binding to proteins: A free-energy component analysis using a dielectric continuum model, Biophysical Journal 88 (2005) 3888–3904.
32. Y. F. Song, J. J. Mao, M. R. Gunner, MCCE2: improving protein pK$_a$ calculations with extensive side chain rotamer sampling, J. Comp. Chem. 30 (2009) 2231–2247.
33. Z. Zheng, M. R. Gunner, Analysis of the electrochemistry of hemes with ems spanning 800 mv, Proteins 75 (2009) 719–734.
34. R. E. Georgescu, E. G. Alexov, M. R. Gunner, Combining conformational flexibility and continuum electrostatics for calculating pk(a)s in proteins, Biophys. J. 83 (2002) 1731–1748.
35. Y. Song, J. Mao, M. R. Gunner, Calculation of proton transfers in bacteriorhodopsin br and m intermediates, Biochemistry 42 (2003) 9875–9888.
36. M. R. Gunner, X. Zhu, M. C. Klein, Mcce analysis of the pkas of introduced buried acids and bases in staphylococcal nuclease., Proteins: Struct., Funct., & Bioinf. 79 (2011) 3306–3319.
37. D. Riccardi, P. Schaefer, Q. Cui, pk_a calculations in solution and proteins with QM/MM free energy perturbation simulations, J. Phys. Chem. B 109 (2005) 17715–17733.
38. N. Ghosh, Q. Cui, pk_a of residue 66 in staphylococal nuclease: insights from QM/MM simulations with conventional sampling, J. Phys. Chem. B 112 (2008) 8387–8397.
39. G. Li, Q. Cui, pk_a calculations with QM/MM free energy perturbations, J. Phys. Chem. B 107 (2003) 14521–14528.
40. Elstner, M.; Porezag, D.; Jungnickel, G.; Elstner, J.; Haugk, M.; Frauenheim, Th.; Suhai, S. and Seifert, G., Self-consistent-charge density-functional tight-binding method for simulations of complex materials pro perties, Phys. Rev. B 58 (11) (1998) 7260–7268.
41. Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M. and Qiang Cui, Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes (Feature Article), J. Phys. Chem. B 110 (2006) 6458–6469.
42. D. R. Lide (Ed.), CRC Handbook Chemistry and Physics, 85th Edition, CRC Press, 2005.
43. L. Q. Zheng, M. G. Chen, W. Yang, Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems, Proc. Natl. Acad. Sci. USA 105 (2008) 20227–20232.
44. J. E. Nielsen, M. R. Gunner, E. B. Garcia-Moreno, The pk(a) cooperative: A collaborative effort to advance structure-based calculations of pk(a) values and electrostatic effects in proteins, Proteins: Struct., Funct., & Bioinf. 79 (2011) 3249–3259.
45. I. Kim, T. W. Allen, Bennett's acceptance ratio and histogram analysis methods enhanced by umbrella sampling along a reaction coordinate in configurational space, J. Chem. Phys. 136 (2012) 164103.
46. N. Ota, A. T. Brunger, Overcoming barriers in macromolecular simulations: non-boltzmann thermodynamic integration, Theor. Chem. Acc. 98 (1997) 171–181.
47. M. Leitgeb, C. Schroder, S. Boresch, Alchemical free energy calculations and multiple conformational substates, J. Chem. Phys. 122 (2005) 084109.
48. H.-X. Zhou, Calculation of free-energy differences and potentials of mean force by a multi-energy gap method, J. Chem. Phys. 128 (2008) 114104.
49. A. Warshel, Computer Modeling of Chemical Reactions in Enzymes and Solution, Wiley, New York, 1991.
50. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, Multidimensional free-energy calculations using the weighted histogram analysis method, J. Comput. Chem. 16 (1995) 1339–1350.
51. A. Pohorille, C. Jarzynski, C. Chipot, Good practices in free-energy calculations, J. Phys. Chem. B 114 (2010) 10235–10253.
52. A. Warshel, S. T. Russell, Calculations of electrostatic interactions in biological systems and in solutions, Quart. Rev. Biophys. 17 (1984) 283.
53. D. G. Isom, C. A. Castaneda, P. D. Velu, B. Garcia-Moreno, Charges in the hydrophobic interior of proteins, Proc. Natl. Acad. Sci. USA 107 (2010) 16096–16100.
54. M. S. Chimenti, V. S. Khangulov, A. C. Robinson, A. Heroux, A. Majumdar, J. L. Schlessman, B. Garcia-Moreno, Structural reorganization triggered by charging of lys residues in the hydrophobic interior of a protein, Struct. 20 (2012) 1071–1085.
55. D. Riccardi, G. Li, Q. Cui, The importance of van der waals interactions in QM/MM simulations, J. Phys. Chem. B 108 (2004) 6467–6478.
56. M. Freindorg, J. Gao, Optimization of the lennard-jones parameters for a combined ab initio quantum mechanical and molecular mechanical potential using the 3-21g basis set, J. Comput. Chem. 17 (3) (1996) 386–395.
57. D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chattfield, M. Hodošček, B. R. Brooks, Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method, J. Chem. Phys. 117 (2002) 10534–10547.
58. K. E. Shaw, C. J. Woods, A. J. Mulholland, Compatibility of quantum chemical methods and empirical (mm) water models in quantum mechanics/molecular mechanics liquid water simulations, J. Phys. Chem. Lett. 1 (2010) 219–223.
59. B. Wang, D. G. Truhlar, Including charge penetration effects in molecular modeling, J. Chem. Theo. Comp. 6 (2010) 3330–3342.
60. G. Hou, X. Zhu, M. Elstner, Q. Cui, A modified QM/MM hamiltonian with the self-consistent-charge density-functional-tight-binding theory for highly charged QM regions, J. Chem. Theo. Comp. 8 (2012) 4293–4304.
61. Yang, Y.; Yu, H.; York, D.; Cui, Q. and Elstner, M., Extension of the self-consistent-charge density-functional tight-binding third-order expansion of the density functional theory total energy an introduction of a modified effective coulomb interaction, J. Phys. Chem. A 111 (2007) 10861–10873.
62. T. J. Giese, D. M. York, Charge-dependent model for many-body polarization, exchange, and dispersion interactions in hybrid quantum mechanical/molecular mechanical calculations, J. Chem. Phys. 127 (2007) 194101.
63. D. Riccardi, Q. Cui, pk_a analysis for the zinc-bound water in human carbonic anhydrase ii: benchmark for "multi-scale" QM/MM simulations and mechanistic implications, J. Phys. Chem. A 111 (2007) 5703–5711.
64. M. Nina, D. Beglov, B. Roux, Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations, J. Phys. Chem. B 101 (1997) 5239–5248.
65. D. Sitkoff, K. A. Sharp, B. Honig, Accurate calculation of hydration free-energies using macroscopic solvent models, J. Phys. Chem. 98 (1994) 1978–1988.
66. J. Zhang, M. R. Gunner, Multiconformation continuum electrostatics analysis of the effects of a buried asp introduced near heme a in rhodobacter sphaeroides cytochrome c oxidase, Biochem. 49 (2010) 8043–8052.
67. W. Im, D. Beglov, B. Roux, Continuum solvation model: computation of electrostatic forces from numerical solutions to the poisson-boltzmann equation, Comput. Phys. Commun. 111 (1998) 59–75.
68. G. Archontis, T. Simonson, Dielectric relaxation in an enzyme active site: Molecular dynamics simulations interpreted with a macroscopic continuum model, J. Am. Chem. Soc. 123 (2001) 11047–11056.
69. W. Yang, R. Bitetti-Putzer, M. Karplus, Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence, J. Chem. Phys. 120 (2004) 2618–2628.
70. L. Qin, C. Hiser, A. Mulichak, R. M. Garavito, S. Ferguson-Miller, Identification of conserved lipid/detergent-binding sites in a high-resolution structure of the membrane protein cytochrome c oxidase, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 16117–16122.
71. S. T. Russell, A. Warshel, Calculations of electrostatic energies in proteins: the energetics of ionized groups in bovine pancreatic trypsin inhibitor, J. Mol. Biol. 185 (1985) 389–404.
72. A. Warshel, P. K. Sharma, M. Kato, W. W. Parson, Modeling electrostatic effects in proteins, Biochim. Biophys. Acta. 1764 (2006) 1647–1676.
73. S. Chakrabarty, A. Warshel, Capturing the energetics of water insertion in biological systems: The water flooding approach, Proteins: Struct., Funct., & Bioinf. 81 (2013) 93–106.

Table S1. Summary of different simulation setups

BC^a	Input^b	State^c	Redox/titration patterns^d	Length (ns)	Parameters^e	cavity
PBC	1M56	$\mathbf{P}_R$=PDD-OO	E286H; PRDa$_3^-$; Cu$_B^{2+}$-OH$^-$; Fe$_a$(III); Tyr288$^-$	15×2 + 3×50	j	small
PBC	1M56	$\mathbf{P}'_R$=DPD-OO	E286$^-$; PRDa$_3$H; Cu$_B^{2+}$-OH$^-$; Fe$_a$(III); Tyr288$^-$	15	j	large
PBC	1M56	$\mathbf{P}''_R$=PPD-OO	E286H; PRDa$_3$H; Cu$_B^{2+}$-OH$^-$; Fe$_a$(III); Tyr288$^-$	50	j	large
PBC	1M56	$'\mathbf{F}$=DPP-OO	E286$^-$; PRDa$_3$H; Cu$_B^{2+}$-H$_2$O; Fe$_a$(III); Tyr288$^-$	15 + 2×50	j	large
PBC	1M56	PDD-RO	E286H; PRDa$_3^-$; Cu$_B^{2+}$-OH$^-$; Fe$_a$(II); Tyr288H	50	g	small
PBC	1M56	DPP-OR	E286$^-$; PRDa$_3$H; Cu$_B^{+}$-H$_2$O; Fe$_a$(III); Tyr288H	15	g	large
GSBP	1M56	XDD-RO	E286X; PRDa$_3^-$; Cu$_B^{2+}$-OH$^-$; Fe$_a$(II); Tyr288H	6×∼3	g	small
GSBP	1M56	DPD-RO	E286$^-$; PRDa$_3$H; Cu$_B^{2+}$-OH$^-$; Fe$_a$(II); Tyr288H	∼1	g	small
GSBP	1M56	$'\mathbf{F}$=DPP-OO	E286$^-$; PRDa$_3$H; Cu$_B^{2+}$-H$_2$O; Fe$_a$(III); Tyr288$^-$	∼2	j	small
GSBP	1M56+9w	XDD-RO	E286X; PRDa$_3^-$; Cu$_B^{2+}$-OH$^-$; Fe$_a$(II); Tyr288H	6×∼3	g	small
GSBP	1M56+9w	XPD-RO	E286X; PRDa$_3$H; Cu$_B^{2+}$-OH$^-$; Fe$_a$(II); Tyr288H	6×∼3	g	small
GSBP	PBC$'\mathbf{F}$ state	XDD-RO	E286X; PRDa$_3^-$; Cu$_B^{2+}$-OH$^-$; Fe$_a$(II); Tyr288H	6×∼3	g	large
GSBP	PBC$'\mathbf{F}$ state	XPD-RO	E286X; PRDa$_3$H; Cu$_B^{2+}$-OH$^-$; Fe$_a$(II); Tyr288H	6×∼3	g	large

a. PBC: Periodic Boundary Condition used for unrestrained MD; GSBP: Generalized Solvent Boundary Potential used for local MD.

b. Input structure: 1M56: starting coordinates taken from the crystal structure; 1M56+9w: 6 additional water molecules added to 1M56 structure in the region near Glu286 and 3 near PRDa$_3$; PBC$'\mathbf{F}$: local GSBP simulation starting coordinates taken from a snapshot of the $'\mathbf{F}$-state PBC simulation. See text for more details. In local, GSBP MD the cavity size of the input structure is maintained.

c. The states are labeled with a 5 character notation. The first three letters indicate the protonation state (Protonated or Deprotonated) of Glu286, propionate D of heme a$_3$ (PRDa$_3$), the ligand of Cu$_B$ (hydroxide (D) or water (P)). The last two letters indicate the reduction state (Reduced or Oxidized) of heme a and Cu$_B$, respectively. The first letter of "X" indicates pK$_7'$ simulations in which the protonation state of Glu 286 is varied.

d. Other co-factors are fixed as: Cu$_A$ oxidized, Fe$_{a3}$(IV)=O^{2-}, His334H.

e. Parameters for the metal co-factors: "j" uses the Johansson set [8] and "g" uses the Ghosh set [7]. The Ghosh parameters have a neutral Tyr 288 and the Johansson parameters have a deprotonated, anionic Tyr 288. Therefore, the net charge of hemes a and a3, Cu$_B$ and Tyr 288 in the $\mathbf{P}_R$ (PDD-OO) state with the Johansson parameters is identical to that of the PDD-RO state with the Ghosh parameters. In the latter, the extra electron resides on heme a.

Table S2. Active site features for different crystal structures of CcO[a]

PDB code	Resolution (Å)	Glu286-PRDa$_3$ (Å)	PRDa$_3$-Arg481 (Å)	Trp172-PRDa$_3$ (Å)	Trp172-PRDa (Å)	# of water
	Paracoccus denitrificans					
3HB3	2.25	10.3	2.9	2.7	4.3	0
1QLE	3.00	10.3	2.8	3.4	5.1	0
	Rhodobacter Sphaeroides					
3FYE	2.15	10.4	2.8	2.9	4.4	0
2GSM	2.00	10.2	2.8	2.8	4.5	1
1M56	2.30	10.3	3.2	3.1	4.8	0
	Bovine					
3ASO	2.30	11.3	2.9	2.8	4.5	0
2Y69	1.95	11.5	2.8	2.8	4.5	0
3ABM	1.95	11.4	2.9	2.8	4.4	0
2EIJ	1.90	11.4	2.9	2.8	4.4	0
2DYR	1.80	11.4	2.9	2.8	4.4	0
1V54	1.80	11.4	2.9	2.7	4.4	0
1V55	1.90	11.4	2.8	2.8	4.4	0

a. The four distances reported in Fig.4 of the main text (Cδ-Glu286 to C$\alpha\delta$ of PRDa$_3$; minimal distance between propionate acidic oxygens of PRDa$_3$ and side chain NH of Arg481; minimal distance between propionate acidic oxygens of PRDa$_3$ and Nϵ1 of Trp172; minimal distance between propionate acidic oxygens of PRDa and Nϵ1 of Trp172) and the number of water molecules in the cavity in various CcO crystal structures. Note that these distances, especially those that involve Trp172, change significantly in PBC based MD simulations when PRDa$_3$ is protonated (see main text).

Table S3. Computed local dielectric constants (ϵ_1) in PBC simulations around several key residues[a]

Region[b]	$\mathbf{P}_R$	$\mathbf{P}'_R$	$'\mathbf{F}$
Glu286 Cδ	4.2	5.0	8.8
PRDa$_3$ C$\gamma\delta$	4.1	4.5	7.1
Ser200 Oγ	3.4	3.9	3.1
Asp132 Cγ	14.5	14.2	15.7

a. For the computation of ϵ_1 with Eq.2, ϵ_2 is assigned a value of 4 except for the region near Asp132, where an $\epsilon_2 = 20$ is used due to the proximity to bulk solvent.

b. For the calculation of the dipole fluctuation, contributions from all residues with an atom within 10 Å from the specified reference point in the input structure (average structure from the MD) are included.

Table S4. Summary of degrees of freedom for pK$_\mathrm{a}$ and pK$'_7$ calculations

Method	calculation[a]	protonation[b]	backbone[c]	side chains[c]	solvent model[d]	internal dielectric[d]	num snapshots[e]
QM/MM-TI	pK$'_7$	fix	free	free	explicit	ϵ=1	
SCCE-LRA	pK$'_7$	fix	fix	fix	ϵ=80	ϵ=2,4	10-20
MCCE	pK$'_7$	free at pH 7	fix	free	ϵ=80	ϵ=4	6
MCCE	pK$_\mathrm{a}$	free pH titration	fix	free	ϵ=80	ϵ=4	6

a. pK$'_7$ is obtained from the energy for ionization with all protonation states fixed; pK$_\mathrm{a}$ is calculated by a pH titration with the protein remaining in equilibrium with the solution pH.

b. protonation states: Fix have all Asp, Glu, Arg, Lys and propionic acids in their ionized states with the exception of Glu286, Lys362, Asp407, Lys442 of subunit I, Glu90, Glu185 and Asp251 of subunit III. All His, Tyr and Cys are neutral with the exception of His67, His534 of subunit I, Cys252 and Cys256 of subunit II and His37, His132 and His188 of subunit III, which are charged; free at pH 7: all residues are equilibrated in the defined redox state and Glu286, PRDa$_3$, Tyr 288 and Cu$_B$ water protonation states (see Table S1).

c. The backbone and side chains can be fixed in a single position or free to move. For QM/MM-TI calculations, local, GSBP MD is used. For MCCE, side chain rotamers are subjected to Monte Carlo sampling.

d. Solvent model and internal dielectric: The QM/MM-TI calculations have explicit water in the inner GSBP region subjected to MD simulations; the outer GSBP region is treated with Poisson-Boltzmann electrostatics. The GSBP dynamics maintains the cavity characteristics of the input structure. The SCCE-LRA and MCCE calculations use the Poisson-Boltzmann equation with an external dielectric constant of 80 and an internal dielectric constant of 2 or 4. Additional calculations are also reported in Tables S12,S13 for SCCE-LRA and MCCE with a range of internal dielectric constants.

e. Num snapshots: The results of these numbers of individual snapshots are averaged for the reported pK$'_7$ or pK$_\mathrm{a}$.

Table S5. Statistical analysis of pK$'_7$ simulations for Glu286 using QM/MM-TI^a

	GSBP-1M56, XDD-ROg		GSBP-1M56+9w, XDD-ROg		GSBP-PBC$'\mathbf{F}$, XDD-ROg		GSBP-1M56+9w, XPD-ROg		GSBP-PBC$'\mathbf{F}$, XPD-ROg	
Cavity	small		small		large		small		large	
Hydration	low		high		high		high		high	
Propionate	PRDa$_3^{(-)}$		PRDa$_3^{(-)}$		PRDa$_3^{(-)}$		PRDa$_3$H		PRDa$_3$H	
λ	prod(equ)^e	τ(n)^e	prod(equ)	τ(n)	prod(equ)	τ(n)	prod(equ)	τ(n)	prod(equ)	τ(n)
0.00	1.1(0.170)	2(456)	3.0(1.471)	21(73)	1.0(0.731)	4(65)	1.0(0.153)	20(43)	3.0(1.938)	10(104)
0.25	1.1(0.160)	58(17)	3.0(1.732)	13(99)	1.0(0.435)	8(76)	1.0(0.396)	11(54)	3.0(2.679)	13(25)
0.50	1.1(0.668)	12(38)	3.0(1.461)	44(35)	1.0(0.800)	10(21)	1.0(0.438)	8(69)	3.0(2.263)	13(56)
0.75	1.1(0.806)	5(62)	3.0(1.752)	26(49)			1.0(0.710)	14(21)	3.0(2.113)	17(52)
1.00	1.1(0.253)	90(10)	3.0(1.860)	11(102)	1.0(0.375)	10(65)	1.0(0.855)	5(29)	3.0(1.032)	46(44)

a. See Table S1 for definition of CcO redox and protonation states and GSBP setup; prod(equ) gives the total simulation time (in nanoseconds) and the segment identified as equilibration (in parentheses). τ gives the size of the block (in picoseconds), and n gives the total number of blocks in the final free energy derivative calculations. The "reverse cumulative averaging" protocol of Yang et al. [69] was employed.

Table S6. Free energy derivatives, $\Delta G^{(1)}_{E\cdot A(D>H)}$, and computed pK$'_7$ from QM/MM-TI simulations for Glu286^a

λ	GSBP-1M56, XDD-ROg	GSBP-1M56+9w, XDD-ROg	GSBP-PBC$'\mathbf{F}$, XDD-ROg	GSBP-1M56+9w, XPD-ROg	GSBP-PBC$'\mathbf{F}$, XPD-ROg
Cavity	small	small	large	small	large
Hydration	low	high	high	high	high
0.00	233.8±0.2	231.8±0.8	233.1±0.4	224.2±1.8	218.5±1.1
0.25	212.2±0.9	195.7±1.0	195.5±0.8	188.7±1.2	187.8±1.5
0.50	160.7±1.6	163.0±1.0	158.2 ±1.1	152.3±0.9	158.1±0.9
0.75	128.2±0.8	112.5±1.4		119.1±2.4	119.1±1.0
1.00	93.5±1.2	85.7±0.4	83.1±1.3	85.2±1.6	84.6±1.67
$\Delta G^{(1)}_{E\cdot A(D>H)}$ ^{b}	165.7±0.99	157.8±0.99	158.1±1.0	153.9±1.0	153.6±1.0
ΔpK$'_7$(QM size)^c	-3.6				-2.8
ΔpK$'_7$([His277H]$^+$)^c	-3.6				-3.5
Final pK$'_7$ estimate^c	18.5	14.0^d	14.3^d	11.2^d	10.6

a. The free energy derivatives are given in kcal/mol, and the statistical errors are based on block average (see Table S6).

b. Computed on the basis of the linear fit of the free energy derivatives vs λ and subsequent integration over λ; the values in parentheses are the R^2 values for the linear fit.

c. The pK$'_7$ is computed using the calculated pK$_a$ shift relative to acetic acid in solution (experimental value of 4.74), including correction due to the QM size (see Table S8). His 277 is the only residue within 20 Å of Glu 286 that has an ambiguous titration state. Since His 277 is close to a lipid head-group in the crystal structure as well as close to the protein surface, it is fixed in the protonated state.

d. Including ΔpK$'_7$(QM size)= -2.4(solution correction) and ΔpK$'_7$([His277H]$^+$)= -3.5.

Table S7. Comparison of results from QM/MM-TI and TI-US calculations for Glu286 pK'_7 with the GSBP-1M56, XDD-ROg (small cavity, low-hydration, $PRDa_3{}^{(-)}$) model[a]

λ	TI[b]	TI-US
0.00	233.9	233.9
0.25	212.2	205.7
0.50	160.9	162.2
0.75	128.2	126.4
1.00	93.5	90.8
$\Delta G^{(1)}_{E\cdot A(D>H)}$ [c]	165.7(0.99)	163.8(1.00)

	TI[d]	TI-US[d]
$\beta^*\Delta G_{0.25,0.00}$	93.8 (0.4)	93.4 (0.7)
$\beta^*\Delta G_{0.50,0.25}$	77.4 (0.7)	75.0 (1.6)
$\beta^*\Delta G_{0.75,0.50}$	60.6 (0.9)	59.7 (1.8)
$\beta^*\Delta G_{1.00,0.75}$	47.1 (0.5)	46.8 (0.8)
$\beta^*\Delta G_{1.00,0.00}$	278.7	274.9
$\Delta G_{1.00,0.00}$	166.2	163.9

a. All free energy values are in the unit of kcal/mol. The upper table shows the free energy derivatives for the different λ windows and the resulting $\Delta G^{(1)}_{E\cdot A(D>H)}$ computed on the basis of the linear fit of the free energy derivatives vs λ and subsequent integration over λ. The lower table uses Eq.13 to treat the average value of the function $P_{\lambda'} - P_{\lambda}$ (illustrated in Fig. S10) as $\beta^*\Delta G_{\lambda',\lambda}$. $\Delta G_{1.00,0.00}$ is equivalent to $\Delta G^{(1)}_{E\cdot A(D>H)}$.

b. Computed using the part of the unbiased trajectory not rejected as equilibration in the block averaging scheme.

c. Values in parentheses are the R^2 values of the linear fit to the free energy derivatives.

d. Values in parentheses are standard deviations.

Table S8. Effect of QM region size on computed Glu286 pK'_7 from QM/MM-TI simulations

	$\langle\Delta\Delta U\rangle_{\lambda=0.0}$ [a] (kcal/mol)	$\langle\Delta\Delta U\rangle_{\lambda=1.0}$ [a] (kcal/mol)	$\Delta\Delta G^{(1)}$ (kcal/mol)	$\Delta\Delta G^{(1)}$ (pH units)
Acetic acid pK_a, solution	-5.3±3.7	11.7±3.3	3.2	2.4
Glu286 pK'_7, GSBP-1M56, XDD-ROg	-3.8±2.4	0.7±2.3)	-1.6	-1.2
Glu286 pK'_7, GSBP-PBC$'\mathbf{F}$, XPD-ROg	-5.3±2.5	4.2±2.9	-0.5	-0.4

a. $\Delta\Delta U$ is the difference in energy gap calculated with a large QM region (including water molecules within 5 Å from the carboxylate oxygens) and that calculated with a small QM region including the Glu286 side chain only.

Table S9. Effect of scaling charges of charged residues by $1/\sqrt{2}$ on computed Glu286 pK'_7 with the GSBP-1M56, XDD-ROg (small cavity, low-hydration, $PRDa_3{}^{(-)}$) model in QMMM-TI simulation[a]

	$\Delta G^{(1)}$(LRA)(kcal/mol)	$\Delta\Delta G^{(1)}$(kcal/mol)[e]	$\Delta\Delta G^{(1)}$ (pH units)
Original charges	-26.0		
Scaled charges I[b]	-28.3	-2.3	-1.7
Scaled charges II[c]	-28.4	-2.4	-1.8
Scaled charges III[d]	-23.3	2.7	1.9

a. Glu286 charges are not scaled.

b. Charges scaled: ionized Arg, Lys, Glu, Asp, His, Cys; Cu_B with ligands; Mg; Cu_A; Ca; heme a; heme a_3.

c. Charges scaled: ionized Arg, Lys, Glu, Asp, His, Cys; Cu_B with ligands; Mg; Cu_A; Ca; PRD_a; PRA_a; $PRDa_3$; $PRAa_3$.

d. Charges scaled: ionized Arg, Lys, Glu, Asp, His, Cys; Cu_B with ligands; Mg; Cu_A; Ca; PRD_a; PRA_a; $PRAa_3$.

e. $\Delta\Delta G^{(1)}$ is calculated as the difference between $\Delta G^{(1)}$ with scaled charges and $\Delta G^{(1)}$ with original charges.

Table S10. Perturbative analysis of Glu286 pK$'_7$ with the GSBP-1M56, XDD-ROg (small cavity, low-hydration, PRDa$_3{}^{(-)}$) model[a]

Charges zeroed out[b]	Charged residues[c]	$\Delta\Delta G^{(1)}$(kcal/mol)[d]	$\Delta\Delta G^{(1)}$ (pH units)
Within 5 Å	None	-9.8	-7.2
Between 5 Å and 8 Å	heme a, a$_3$	34.1	24.9
Between 8 Å and 11 Å	Cu$_B$ with ligands	-29.6	-21.6
Between 11 Å and 14 Å	Arg481, Arg482	-15.2	-11.1
Between 14 Å and 17 Å	None	-8.1	-5.9
Between 17 Å and 20 Å		-2.2	-1.6

a. The calculations are done in the QM/MM framework with the same QM region as in the microscopic pK$'_7$ calculation. All partial charges on atoms within a specific region are set to zero in the perturbative analysis.

b. Distances are measured from the carboxylate C atom of Glu286. The entire residue with at least one atom within the particular distance range is selected.

c. Residues with a non-zero net charge with at least one atom in the particular distance range

d. $\Delta\Delta G^{(1)}$ is calculated as the difference between $\Delta G^{(1)}$ with no charges zeroed out and $\Delta G^{(1)}$ with the charges on selected residues zeroed out. A positive value of $\Delta\Delta G^{(1)}$ for a set of residues indicates that those residues increase the proton affinity and thus increase pK$'_7$.

Table S11. Perturbative analysis of Glu286 pK$'_7$ with GSBP-1M56, XDD-ROg model (small cavity, low-hydration, PRDa$_3{}^{(-)}$) for residues with at least one atom between 5 Å and 8 Å from Glu286

Residue(s)	$\Delta G^{(1)}$(LRA)(kcal/mol)	$\Delta\Delta G^{(1)}$(kcal/mol)	$\Delta\Delta G^{(1)}$ (pH units)
Heme a$_3$	136.0	27.7	20.2
PRDa$_3$	146.1	17.6	12.8
PRAa$_3$	155.5	8.2	6.0
Heme a	149.8	13.9	10.1
PRDa	151.9	11.8	8.6
PRAa	159.4	4.3	3.1
Water	168.1	-4.4	-3.2
Met106	163.1	0.6	0.4
Phe109	163.9	-0.2	-0.2
Val111	165.4	-1.7	-1.3
Pro113	164.6	-0.9	-0.7
Trp172	163.4	0.3	0.2
Val194	163.5	0.3	0.2
Ser197	164.6	-0.9	-0.7
Tyr288	164.4	-0.7	-0.5
Ser201	165.6	-1.9	-1.4
Ile239	163.5	0.2	0.1
Ala242	164.3	-0.6	-0.4
Leu243	163.6	0.1	0.1
Leu246	165.6	-1.9	-1.4
Gly283	161.1	2.6	1.9
His284	161.6	2.1	1.5
Pro285	163.9	-0.2	-0.1
Met424	163.7	0.0	0.0

Table S12. Glu286 pK$'_7$ from SCCE-LRA with different dielectric parameters[a]

Snapshots[b]	$\epsilon_{prot} = 4, \epsilon_{cav} = 4$	$\epsilon_{prot} = 4, \epsilon_{cav} = 9$	$\epsilon_{prot} = 4, \epsilon_{cav} = 80$	$\epsilon_{prot} = 2, \epsilon_{cav} = 80$
GSBP-1M56, XDD-RO	14.6±0.7	13.7±0.6	10.2±0.7	15.1±1.6
GSBP-PBC$'\mathbf{F}$, XDD-RO	15.4±0.9	13.8±0.8	8.7±0.8	11.8±1.6
GSBP-PBC$'\mathbf{F}$, XPD-RO	11.6±1.3	9.7±1.3	6.5±1.2	8.3±2.7

a. The results are averaged over 10-20 snapshots (separated by 70-100 ps) from MD simulations with equal numbers of snapshots with Glu286 protonated and deprotonated; the standard deviations do not change significantly when more (100) snapshots are used. $\epsilon_{prot}, \epsilon_{cav}$ are the dielectric constants used for the protein and the hydrophobic cavity, respectively; ϵ_w for bulk is always set to 80. The dielectric constant used for the membrane slab (taken to be the same as ϵ_{cav}) has a minimal impact on the result since Glu286 is far from the protein/membrane interface.

b. The snapshots are from GSBP based MD simulations with different initial coordinates (1M56: crystal structure; PBC$'\mathbf{F}$: an equilibrated snapshot from PBC simulation for the $'\mathbf{F}$ state) and PRDa$_3$ either deprotonated or protonated. See Table S1 for details.

Table S13. Glu286 pK_7' from MCCE with different protein dielectric constants[a]

Cavity	Input struc.[b]	State[c]	$\epsilon_{prot}=2$ E$^-$/EH[d]	$\epsilon_{prot}=4$ E$^-$/EH[d]	$\epsilon_{prot}=8$ E$^-$/EH[d]	$\epsilon_{prot}=20$ E$^-$/EH[d]
small	1M56	XDD-RO	15.4/20.5	10.9/11.9	8.6/9.5	5.5/5.5
small	1M56	XPD-RO	13.4/18.4	9.3/10.2	8.0/8.8	5.1/5.2
big	PBC′**F**	XDD-RO	12.5/13.1	8.9/9.6	7.4/7.8	5.0/5.0
big	PBC′**F**	XPD-RO	10.6/10.8	7.3/8.1	6.5/6.7	4.6/4.5
		pK changes[e]				
Change due to $PRDa_3$ protonation in different structures						
		1M56	2.1/2.1	1.7/1.7	0.7/0.7	0.3/0.3
		PBC′**F**	2.0/2.3	1.6/1.6	0.8/1.0	0.5/0.5
Dependence of effect of cavity hydration on the ionization of $PRDa_3$						
		$PRDa_3^{(-)}$	2.9/7.4	2.1/2.3	1.3/1.7	0.5/0.5
		$PRDa_3H$	2.8/7.6	2.0/2.2	1.4/2.0	0.6/0.7
		Combined Effect	4.9/9.7	3.6/3.9	2.1/2.7	0.9/1.0

a. All MCCE calculations use a dielectric constant for water of 80, including within internal protein cavities.
b. The GSBP based MD simulations started with different initial coordinates. 1M56: the crystal structure; PBC′**F**: an equilibrated snapshot from PBC simulation for the ′**F** state.
c. The "X" highlights that the protonation state of Glu286 is varied in the QM/MM-TI calculations calculations that yield the snapshots.
d. The pK_7' values before and after the slashes are computed with snapshots from QM/MM-TI calculations with an ionized and neutral Glu 286, respectively. The averaged results are shown as MCCE pK_7' in Table 1 of the main text.
e. The effects of cavity size (i.e. change of hydration level of the cavity), protonation of $PRDa_3$ are calculated based on the computed pK_7' values from different setups. The combined effect is obtained by taking the difference between pK_7' values computed with a small cavity (low hydration), $PRDa_3{}^{(-)}$ and a large cavity (high hydration), $PRDa_3H$.

Table S14. Breakdown of $\Delta\Delta G_{prot}$ (in pH units) for MCCE pK_7' calculated with $\epsilon_{prot}=4$[a]

Cavity	Input struc. [b]	State[c]	$\Delta\Delta G_{dsolv}$ E$^-$/EH[d]	ΔG_{bkb} E$^-$/EH[d]	ΔG_{res} E$^-$/EH[d]
small	1M56	XDD-RO	5.9/6.0	-1.3/-1.0	1.3/1.7
small	1M56	XPD-RO	5.9/6.0	-1.3/-1.0	-0.4/0.0
big	PBC′**F**	XDD-RO	3.8/4.1	-0.8/-0.2	0.9/0.7
big	PBC′**F**	XPD-RO	3.8/4.1	-0.8/-0.2	-0.7/-0.8
		Contribution changes[e]			
Change due to $PRDa_3$ protonation in different structures					
		1M56	0.0/0.0	0.0/0.0	1.7/1.7
		PBC′**F**	0.0/0.0	0.0/0.0	1.5/1.5
Dependence of effect of cavity hydration on the ionization of $PRDa_3$					
		$PRDa_3^{(-)}$	2.1/2.0	-0.5/-0.8	0.5/1.0
		$PRDa_3H$	2.1/2.0	-0.5/-0.8	0.3/0.8
		Combined Effect	2.1/2.0	-0.5/-0.8	2.0/2.5

a. For definition of components, see Eqs.15-17; ΔG_{bkb} and ΔG_{res} are contributions from protein backbone and side chains to $\Delta G_{pairwise}$.
b. The snapshots are from GSBP based MD simulations with different initial coordinates. 1M56: the crystal structure; PBC′**F**: an equilibrated snapshot from PBC simulation for the ′**F** state.
c. The "X" highlights that the protonation state of Glu286 is varied in the QM/MM-TI calculations calculations that yield the snapshots.
d. The values before and after the slashes are computed from snapshots from QM/MM-TI calculations with an ionized and neutral Glu 286, respectively.
e. The effects of cavity size yielding a change of hydration level of the cavity and protonation of $PRDa_3$ are calculated based on the computed free energy contributions from different setups. The combined effect is obtained by taking the difference between values computed with a small cavity (low hydration), $PRDa_3{}^{(-)}$ and a large cavity (high hydration), $PRDa_3H$. Positive values indicate increase in Glu proton affinity.

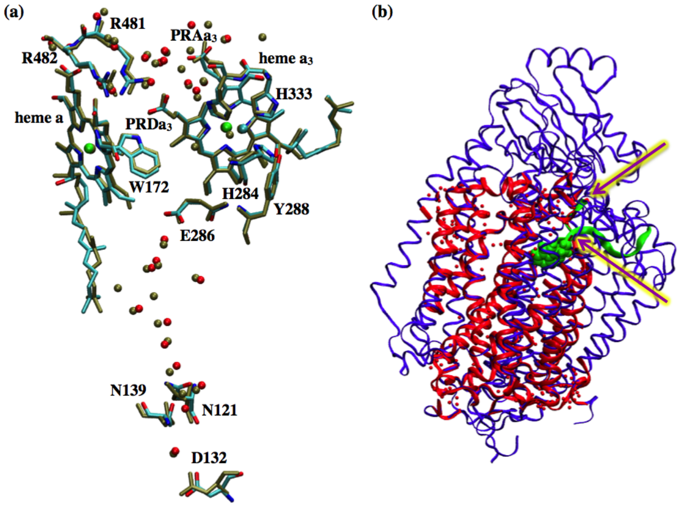
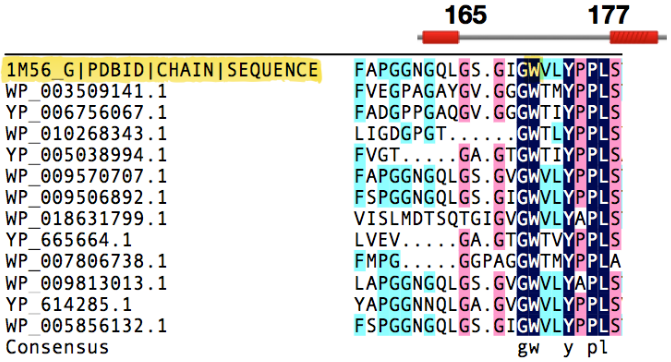


Fig. S1. Structure of the computational model for CcO. (a) Comparison of water molecules resolved in two crystal structures for *Rhodobacter sphaeroides* CcO; the PDB codes are 1M56 [1] (4 sub-unit, 2.3 Å resolution, colored by atom type) and 2GSM [70] (2 sub-unit, 2.0 Å resolution, colored in tan). The positions of the redox co-factors, amino-acid side-chains and water molecules (in and around the active site) generally agree well. However, the latter has 2 extra water molecules in the D-channel (which is possibly related to the absence of subunit III in the construct [70]) and one resolved water molecule hydrogen-bonded to the Cu$_B$ ligand OH$^-$. (b) Demonstration of the location of the inner region in the GSBP-1M56(+9w) setup. Protein in the inner region is shown in red ribbons while water O atoms are displayed as red dots. The rest of the protein is shown in blue ribbons. Trp172 and the loop bearing it (Residues 165-177 in subunit I) are shown in green. The purple arrows indicate the points on this loop where the inner and outer regions intersect.



```
1M56_G|PDBID|CHAIN|SEQUENCE     FAPGGNGQLGS.GIGWVLYPPLS
WP_003509141.1                  FVEGPAGAYGV.GGGWTMYPPLS
YP_006756067.1                  FADGPPGAQGV.GGGWTIYPPLS
WP_010268343.1                  LIGDGPGT......GWTLYPPLS
YP_005038994.1                  FVGT.....GA.GTGWTIYPPLS.
WP_009570707.1                  FAPGGNGQLGS.GVGWVLYPPLS
WP_009506892.1                  FSPGGNGQLGS.GIGWVLYPPLS
WP_018631799.1                  VISLMDTSQTGIGVGWVLYAPLS
YP_665664.1                     LVEV.....GA.GTGWTVYPPLS
WP_007806738.1                  FMPG.....GGPAGGWTMYPPLA
WP_009813013.1                  LAPGGNGQLGS.GVGWVLYAPLS
YP_614285.1                     YAPGGNNQLGA.GVGWVLYPPLS
WP_005856132.1                  FSPGGNGQLGS.GIGWVLYPPLS
Consensus                                     gw  y pl
```

Fig. S2. The loop region (165-177 based on *Rhodobacter sphaeroides* residue numbers) is highly conserved. The shown sequences are randomly chosen from the result of a BLAST search that retrieved 1000 sequences. The sequences have identity to the query sequence (*Rhodobacter sphaeroides*) ranging from about 50% to 80% and are aligned using the DNAMAN software package (Version 7.358, Lynnon Corporation, Canada).

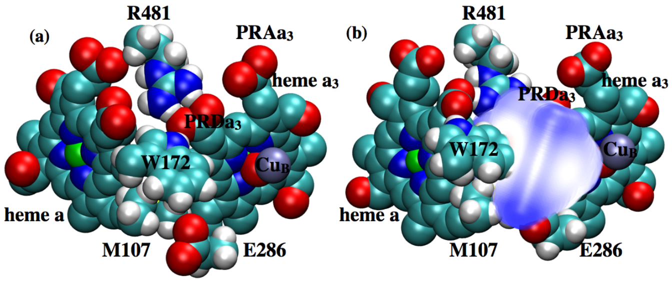


Fig. S3. Illustration for the change of cavity size in MD simulations where PRDa3 is protonated. (a) Space-filling model for the active site region in the crystal (1M56) structure, which has no free volume when the active site region is probed with a sphere of 1.4 Å radius. (b) Similar plot for a snapshot of the $'\mathbf{F}$-state PBC simulation. The cavity accessible to water is revealed by a probing sphere of 1.4 Å radius and illustrated in light blue; the volume of the cavity is 155±21Å^3.

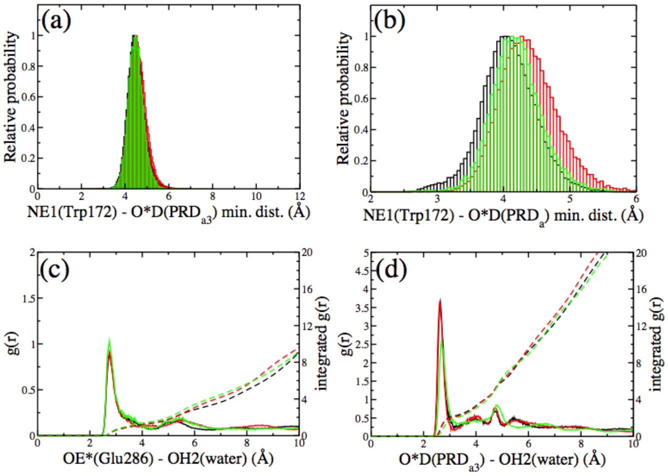


Fig. S4. Comparison of key cavity properties: distance distributions calculated using unrestrained PBC simulations for several chemical states (black: PDD-ROg; red: $\mathbf{P}_R$=PDD-OOj; green: $\mathbf{P}_R$=PDD-OOj scaled) with different force field parameters for the metal co-factors and active-site residues. (a) W172 side chain-PRDa3; (b) W172 side chain-PRDa; (c-d) radial distribution functions (in solid) and integrated radial distribution functions (in dash) of water oxygen around Glu 286 and PRDa3. In "PDD-OOj scaled", the partial charges for PRDa3 and Arg481 are scaled by $1/\sqrt{2}$ as an approximate way to evaluate the effect of including electronic polarization [25]. Similar comparisons are made also for $'\mathbf{F}$=DPP-OOj and DPP-ORg, and similar agreement between the results is observed.

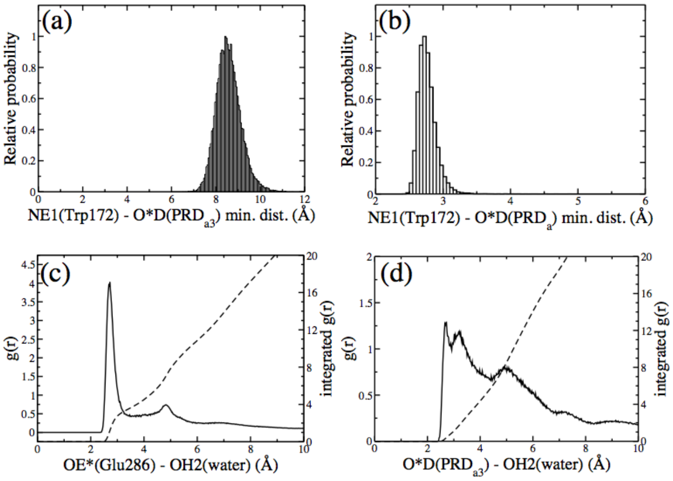


Fig. S5. Cavity properties for the $\mathbf{P}''_R$ state (Table S1) PBC simulations. (a) W172 side chain-PRDa3; (b) W172 side chain-PRDa; (c) solvation around the average position of OE1 and OE2 of E286; (d) solvation around the average position of O1D and O2D of PRDa3.

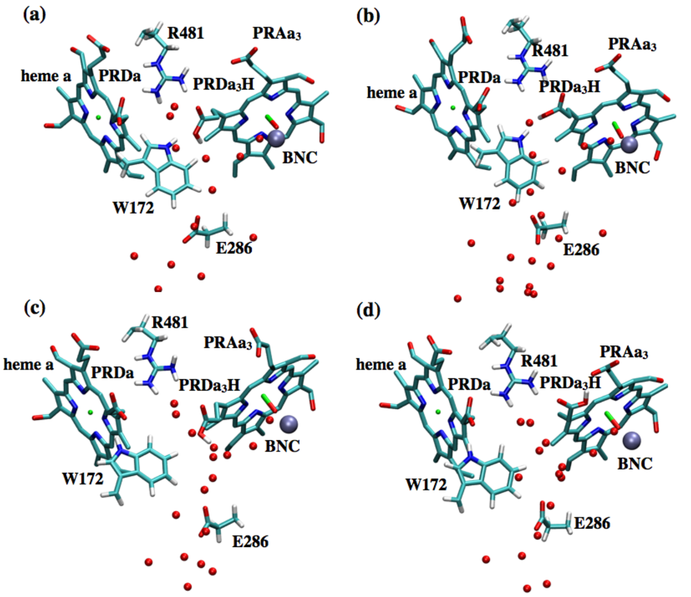

(a)
R481
PRAa₃
heme a
PRDa
PRDa₃H
BNC
W172
E286
(b)
R481
PRAa₃
heme a
PRDa
PRDa₃H
BNC
W172
E286
(c)
R481
PRAa₃
heme a
PRDa
PRDa₃H
BNC
W172
E286
(d)
R481
PRAa₃
heme a
PRDa
PRDa₃H
BNC
W172
E286

Fig. S6. Snapshots from GSBP simulations to illustrate that the hydration level and local conformational properties of the hydrophobic cavity in the DPD-ROg state depend on the initial structure used in the GSBP setup. (a) 1M56 with a small cavity; (b) 1M56+9w; (c-d) PBC$'\mathbf{F}$ with a large cavity.

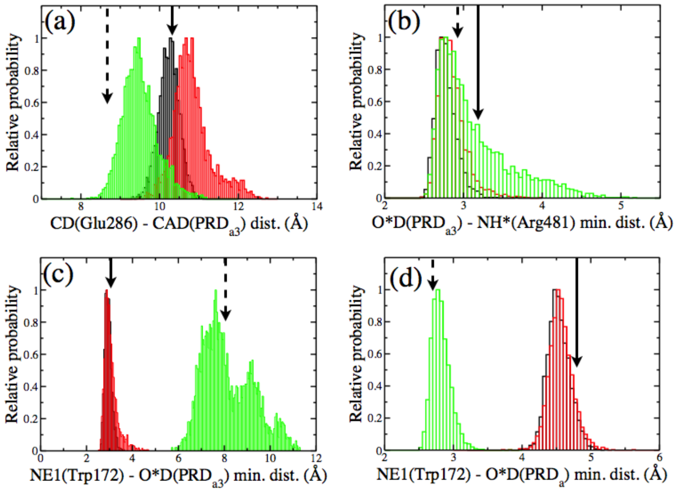


Fig. S7. Dependence of the key distance distributions for residues near the hydrophobic cavity on the initial structures used in local GSPB simulations in the DPD-ROg state (black: GSBP-1M56; red: GSBP-1M56+9w; green: GSBP-PBC$'\mathbf{F}$). Compare with Fig.4 in main text for distances in the unrestrained PBC simulations . (a) E286-PRDa3; (b) R481-PRDa3; (c) W172 side chain-PRDa3; (d) W172 side chain-PRDa. The solid arrows indicate the corresponding values in the crystal structure, and the dashed arrows indicate the corresponding values in the starting snapshot from a PBC simulation for the $'\mathbf{F}$ state.

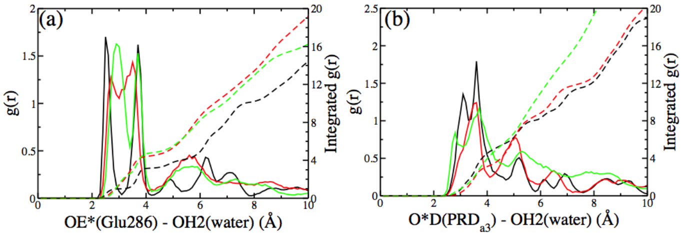


Fig. S8. Comparison of radial distribution functions (in solid) and integrated radial distribution functions (in dash) of water oxygen around Glu 286 and PRDa3 sites calculated using GSBP setups for the DPD-ROg state (black: GSBP-1M56; red: GSBP-1M56+9w; green: GSBP-PBC$'\mathbf{F}$). (a) the average position of OE1 and OE2 of E286; (b) the average position of O1D and O2D of PRDa3.

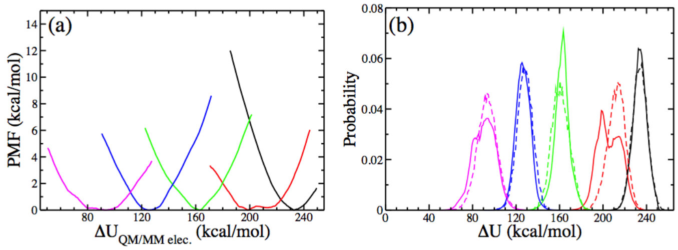


Fig. S9. Results of QM/MM-TI pK$_7'$ calculations with energy-gap umbrella sampling using the GSBP-1M56, XDD-ROg model. (a) PMFs along $\Delta U_{QM/MMelec.}$ for different λ windows obtained using TI-US. (b). Comparison of the probability distribution of the total energy gap, ΔU, from TI (dashed lines) and TI-US (full lines). The rightmost PMF and probability distribution correspond to the λ=0.0 window while the leftmost ones correspond to λ=1.0.

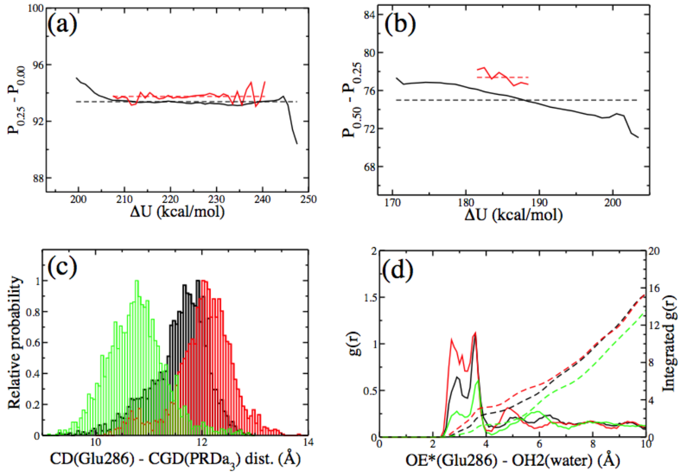


Fig. S10. Comparison of properties from regular thermodynamic integration (TI) and energy-gap umbrella sampling (TI-US) pK$_7'$ simulations. (a-b) P$_{\lambda'}$-P$_{\lambda}$ plots (Eq.13); black: TI-US; red: TI. Each dashed line represents the average value of the full curve of the same color. (c-d) Distribution of Glu286-PRDa$_3$ distance and the level of solvation of Glu286 (measured in terms of the radial distribution function of water O atoms around the center of mass of the Glu286 side-chain oxygen atoms) for the λ=0.5 window. Black: TI, with an average energy gap of 160.7 kcal/mol; red: TI-US with the energy gap restraint centered at 138.0 kcal/mol; green: TI-US with the energy gap restraint centered at 191.0 kcal/mol.

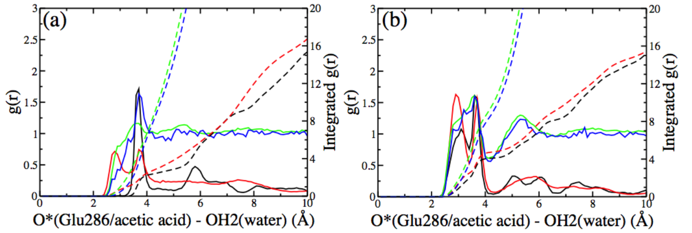


Fig. S11. Radial distribution function of water O atoms around the center of mass of the Glu286/acetic acid (side-chain) oxygen atoms for λ=0.0 and λ=1.0 windows of Glu286 pK$_a$ calculations in the GSBP-1M56, XDD-ROg (black) and GSBP-PBC$'\mathbf{F}$, XPD-ROg (red) models, and of acetic acid pK$_7'$ calculation in solution using a QM/MM (green) or MM (blue) potential. The solid and dotted curves represent g(r) and integrated g(r), respectively.

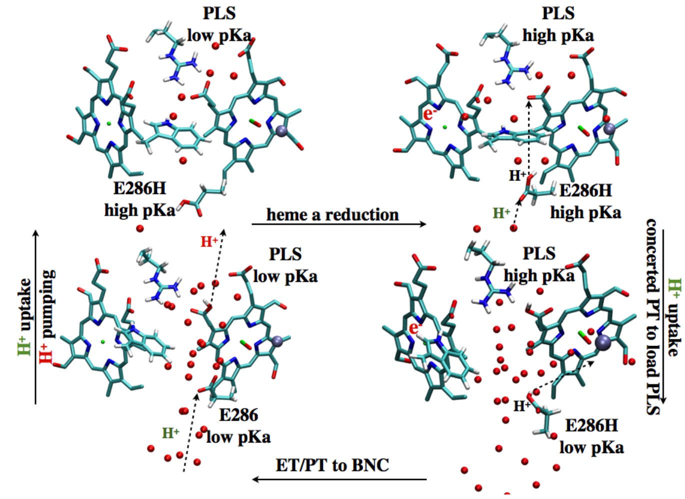

PLS
low pKa
PLS
high pKa
e
H+
E286H
high pKa
heme a reduction
E286H
high pKa
H+
H+
PLS
low pKa
H+ uptake
H+ pumping
H+ uptake
concerted PT to load PLS
PLS
high pKa
e
H+
E286
low pKa
H+
E286H
low pKa
ET/PT to BNC

Fig. S12. A scheme that illustrates how change of hydration level in the hydrophobic cavity coupled to PRDa3 protonation modulates the proton affinity of Glu286 and therefore drives the proton pumping cycle in CcO. As emphasized in the main text, the role for changing hydration in determining proton/electron transfer activities has been considered as one general mechanism to modulate the proton affinity of buried charges [71, 72], including specifically for the stabilization of a deprotonated Glu286 in CcO [73]. Our proposal illustrated here is distinct in that it captures a specific local loop motion coupled to the protonation of a remote group, 10 Å from Glu286, that triggers the change of cavity hydration level, which in turn modulates the proton affinity of Glu286. This proposes a specific molecular mechanism to control the hydration level and proton affinity of this key residue.