Proton storage site in bacteriorhodopsin: new insights from QM/MM simulations of microscopic pK_a and infrared spectra

[1]Puja Goyal[‡], [1]Nilanjan Ghosh[‡], [2]Prasad Phatak[‡], [2]Maike Clemens, [3]Michael Gaus, [2,3]Marcus Elstner[*] and [1]Qiang Cui[*]

[1]*Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, 1101 University Ave, Madison, WI 53706*

[2]*Department of Physical and Theoretical Chemistry TU Braunschweig, Hans-Sommer-Straβe 10 D-38106 Braunschweig, Germany*

[3]*Institute of Physical Chemistry, Karlsruhe Institute of Technology, Kaiserstr. 12, 76131 Karlsruhe, Germany and*

[‡]*Contributed equally*

The following results are included: (1) Geometries and normal modes of gas-phase models; (2) Ring-polymer molecular dynamics simulations for the nuclear quantum effects on continuum band in a gas-phase model compound; (3) Effects of choosing different origins for dipole autocorrelation function calculations; (4) Additional results for active site behaviors; (5) the possibility of having two protons in the PRG region; (6) gas-phase NMR calculations that suggest that ^{17}O NMR experiments can potentially be used for distinguishing the two PRG models; (7) A table (Table S1) that summarizes the statistical analysis of sampling in pK_a simulations; (8) A table (Table S2) that compares hydrogen-bonding interaction strength between a carboxylate and a Ser vs. a Cys residue.

I. OPTIMIZED GEOMETRIES OF ACTIVE SITE MODEL SYSTEMS

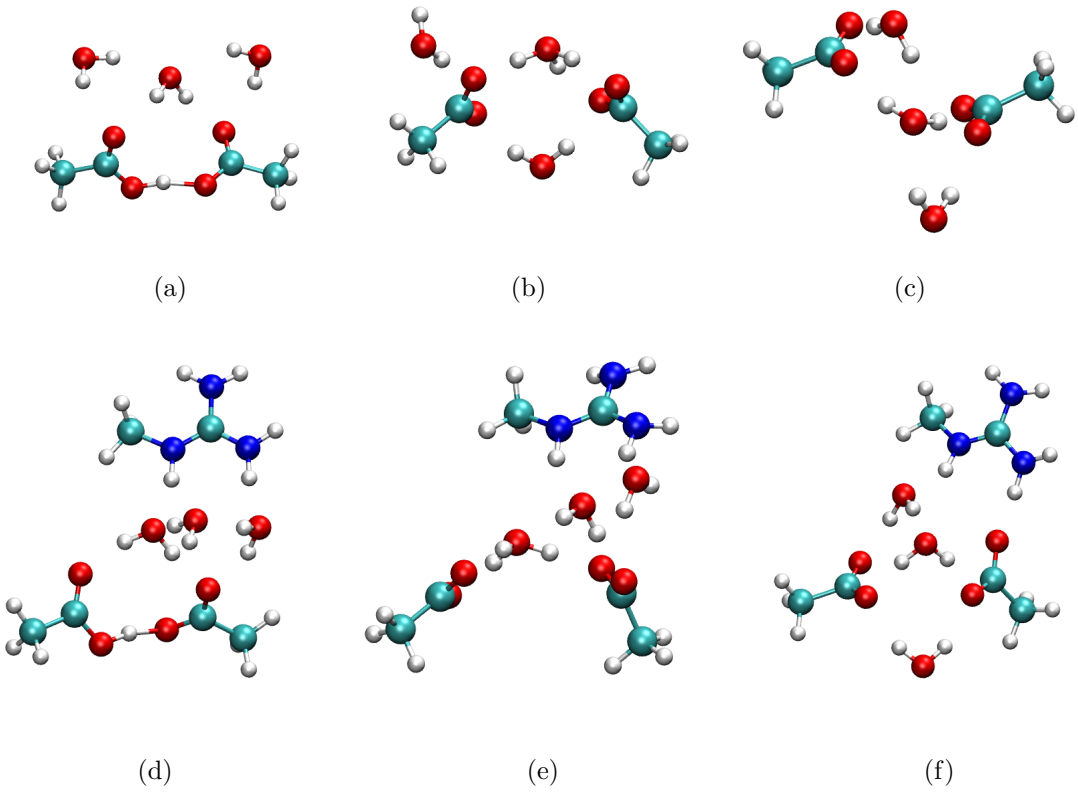


FIG. S1. Optimized structures for the active site models with different protonation patterns and states; (a-c) are for **M1** and (d-f) are for **M2R**.

Optimized geometries at the B3LYP level (for basis set, see main text) for the protonated and deprotonated forms of active site models are shown in Fig.S1; the SCC-DFTB optimized

geometries are generally very similar and therefore not shown. In general, these structures also reflect features in the crystal structures of bR, which justifies the relevance of the benchmark calculations. With the large model (**MLR**), **the hydrogen-bonding patterns in B3LYP and SCC-DFTB optimized structures are not identical, which may also contribute to the apparent PA errors shown in the main text. For example, in Fig.S2a, the hydrogen-bonding between the "Glu204" and the "backbone" of "Ser193" is broken in the B3LYP structure (2.90 Å vs. 2.14 Å in DFTB3-diag), which leads to the excess proton to better localize on Glu204 at the B3LYP level (1.05 Å vs. 1.27 Å with DFTB3-diag).**

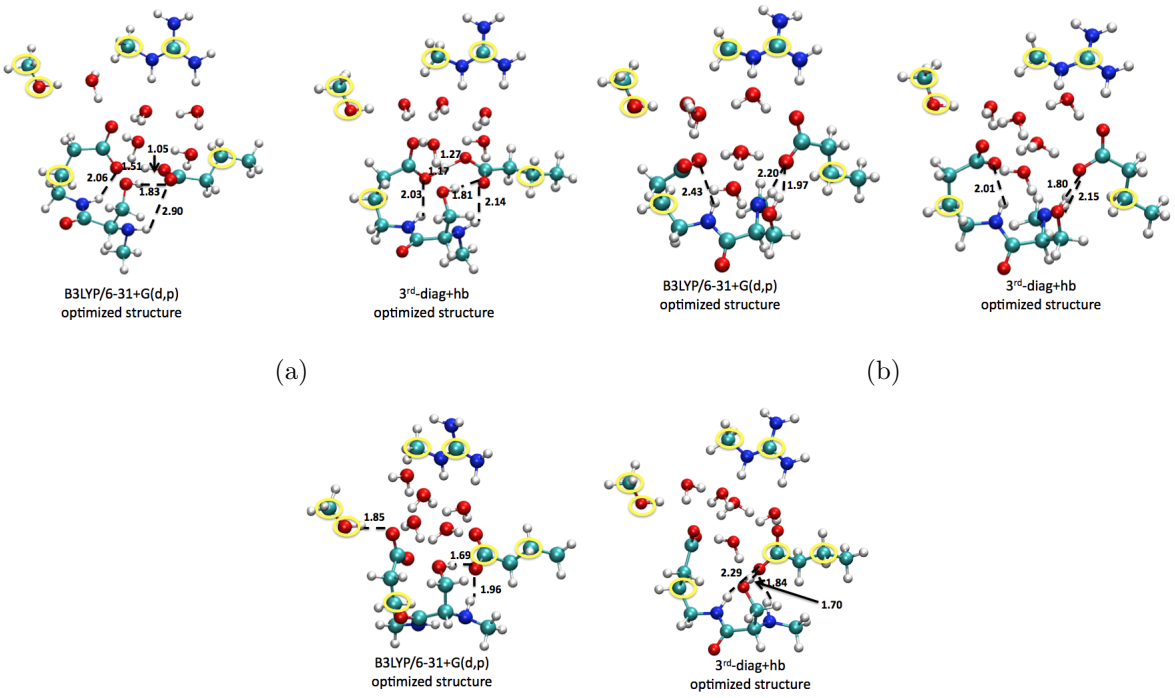


FIG. S2. Optimized structures for the large active site model (**MLR**) with different protonation patterns and states; (a) for the protonated Glu pair, (b) for the protonated water cluster and (c) for the deprotonated state. Atoms circled in yellow are fixed during geometry optimization (based on the 1UCQ model for a,b and the 2NTW model for c). 3^{rd}-diag+hb is the same as the DFTB3-diag[1] approach.

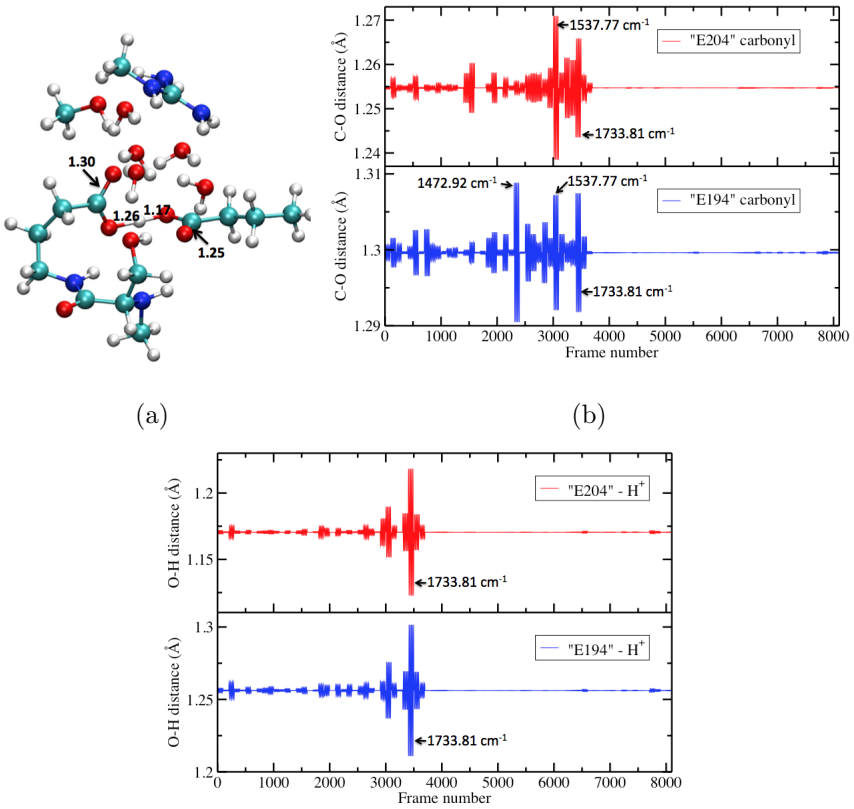


FIG. S3. Structural and normal mode results for the large active site model at the SCC-DFTB (DFTB3-diag) level. (a) To make sure the normal mode results are well defined, no constraints are included in the geometry optimization, which start with the structure optimized with constraints (Fig.S2). In (b-c), the variations of carbonyl C-O distances and O-H distances (H being the excess proton) of Glu204/194 in different normal modes are plotted; the modes have significant CO or HO stretch components are labeled by their frequencies. Upon substituting the oxygens in the sidechains of Glu204/194 to ^{18}O, these modes undergo the following changes: 1472.9 to 1469.1, 1537.8 to 1528.0 and 1733.8 to 1722.6 cm^{-1}.

II. NORMAL MODES OF THE LARGE ACTIVE SITE MODEL

To investigate the frequency range of carbonyl stretches in Glu204 and 194 when they are bonded by a shared proton, we fully optimize the protonated **MLR** model and carry out normal mode calculations at the DFTB3-diag level. As shown in Fig.S3a, with this specific hydrogen-bonding environment, the optimized carbonyl distances are shorter for "Glu204" than for "Glu194". As to the character of the normal modes, Fig.S3c clearly

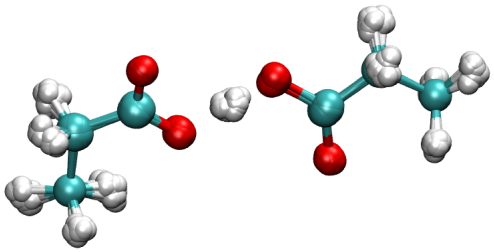
indicates that the mode with 1733.8 cm^{-1} corresponds to the anti-symmetric stretching mode of the excess proton between the two carboxylates; it's interesting that this mode also has significant carbonyl stretch components (see panel b of Fig.S3). Two other modes also have significant carbonyl stretch components: the mode of 1537.8 cm^{-1} involves carbonyl stretches of both "Glu194" and "Glu204", while the mode of 1472.9cm^{-1} is more dominated by the carbonyl stretch of "Glu194". This trend is consistent with the longer carbonyl distance in "Glu194". Although this analysis uses only one structure and the SCC-DFTB approach tends to red-shift carbonyl stretches as found in previous studies[2,3], the model analysis clearly indicates that the "intermolecular bond" significantly red-shift the carbonyl stretches to far below 1700 cm^{-1}. More systematic integrated experiments and calculations for model compounds[4] are needed to firmly establish the frequency range of the carbonyl stretches for Glu pairs with a shared proton.

III. RING POLYMER MOLECULAR DYNAMICS (RPMD) SIMULATIONS FOR A MODEL FOR THE PRG IN BR

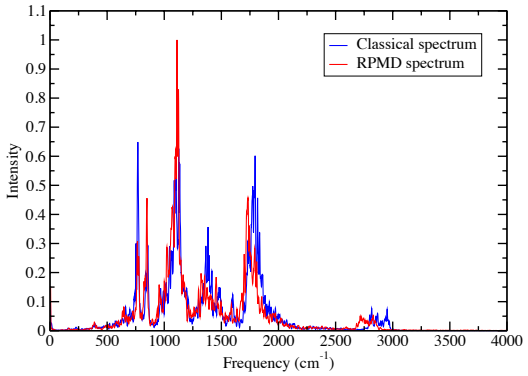
To investigate the impact of nuclear quantum effects on the calculated IR spectra, we carry out RPMD[5,6] for a model system in the gas phase. The model contains a pair of propanoic acids with a shared proton. To avoid too much structural fluctuations in the gas phase, Cβ atoms are fixed in space; water molecules are not included to prevent any complication due to potential "water evaporation".

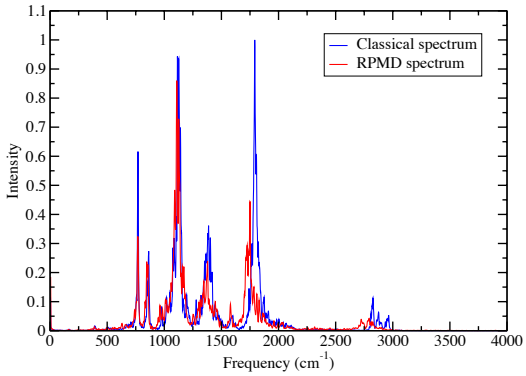
In the RPMD simulations, the system is described with standard second-order SCC-DFTB[7] and the calculations are done with a locally modified version of CHARMM. Each atom is quantized by 32 beads at 300 K and 40 beads at 170 K (see Fig.S4a for an illustration). The system is first equilibrated using the Berendsen weak-coupling algorithm with a coupling constant of 0.2 ps, for 50 ps at 300 K and 80 ps at 170 K. A time step of 0.1 fs is used. A long production run in the NVE ensemble is carried out and several snapshots are taken from this trajectory and used as starting structures for independent runs, each being 150 ps at 300 K and 200 ps at 170 K. The dipole moment for a given step is an average over the dipole moments for all replicas. A total of 70 and 60 independent trajectories are found to give converged dipole-dipole correlation functions at 300 K and 170 K, respectively.

As shown in Figs.S4b-c, the effect of nuclear quantization (at the level of RPMD) on



(a)





(b) (c)

FIG. S4. Set up (a) and calculated IR spectra (b for 300 K and c for 170 K) with RPMD.

the calculated IR spectra is rather minor, even for the continuum band in the 1800-2000 cm^{-1} region; this is probably because the "classical" spectra already included the quantum correction factor that accounts for a significant part of the nuclear quantum effects on absorption intensity. There are minor red-shifts for high-frequency bands, especially those above 1750 cm^{-1}, although the magnitude is very modest for the purpose of this study. Interestingly, for 170 K, the intensity for the continuum band from RPMD is quite a bit lower than the prediction from classical MD simulations; this is also seen at 300 K, although the effect is less striking.

Overall, the calculations suggest that the impact of nuclear quantum effects on the IR spectra are not significant for the purpose of this work. Therefore, we did not pursue this further for the real bR system, although such calculations are readily done with our revised CHARMM program.

IV. CHOICES OF ORIGINS FOR DIPOLE AUTOCORRELATION FUNCTION CALCULATIONS

IR spectra for the L-state of WT bR (using 1UCQ as the starting crystal structure) have been calculated based on SCC-DFTB/MM trajectories using three different coordinate origins for the QM dipole moment: the center of charge (COC), the center of mass (COM) for the QM atoms and the geometric center (GO) of the entire system. As seen in Fig. S5, the calculated IR spectra show very little origin dependence for the region of interest (1800-2000 cm^{-1}).

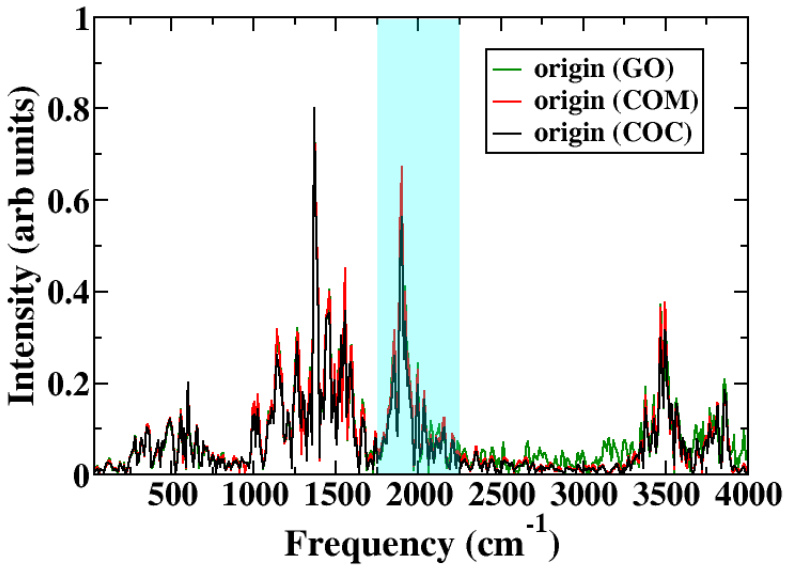


FIG. S5. Comparision of IR spectra from Fourier transform of dipole-dipole correlation function obtained using three different choice of origin for the QM region atoms. The IR spectra is from SCC-DFTB/MM simulations for the L-state of WT bR using the 1UCQ structure.

V. ADDITIONAL RESULTS ON ACTIVE SITE PROPERTIES

A. The distribution of Arg82 sidechain with 5 water molecules in the active site

As shown in Fig.S6, the distribution of the Arg82 sidechain is fairly broad during the MD simulations when there are five water molecules in the PRG region; as explained in the main text and previous study[3], three water molecules are in the starting structure which is

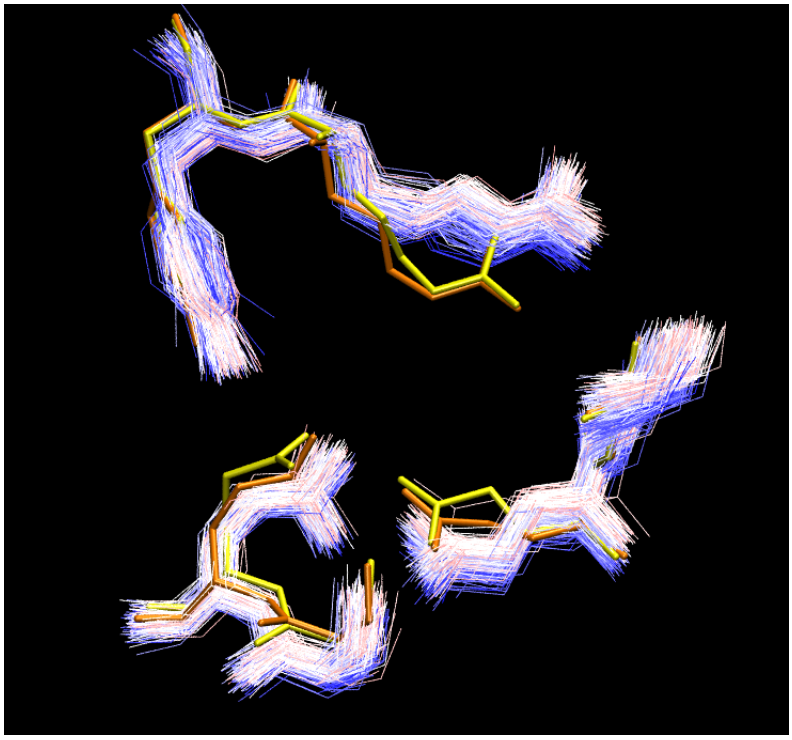


FIG. S6. Distribution of sidechain configurations from MD overlapped with those from crystal structures (yellow: 1C3W for the ground state, orange: 1UCQ for the L-state). For the MD results, the red lines show the beginning of the simulation, white lines show the middle and blue line show the end of the simulation.

based on the crystal structure for the L-state (1UCQ)[8], and two water molecules penetrate into the PRG during the nanosecond simulations. Although the sidechain remains close to the crystal structure orientations (yellow and orange for the ground and L-state structures, respectively), it drifts upwards. All residues in this region have notable fluctuations (e.g., shown are Tyr83, Glu194, Glu204 and Ser193), and the separation between Arg82 and the Glu pairs is observed to increase during the simulation (as also shown by Fig. 5b in the main text), though not in a monotonic fashion. As the model calculations in the main text show, Arg82 tends to further stabilize the excess proton on the Glu pair rather than on the water cluster. Thus the fact that the excess proton clearly favors the Glu pair in QM/MM simulations despite the upward motion of Arg82 sidechain further supports the intermolecular proton bond model for the PRG.

B. The distribution of Arg82 sidechain with 3 water molecules in the active site

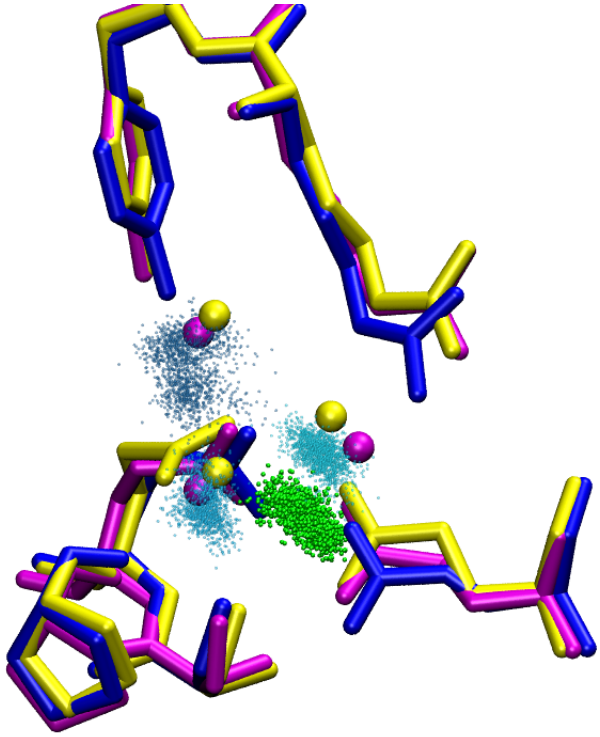
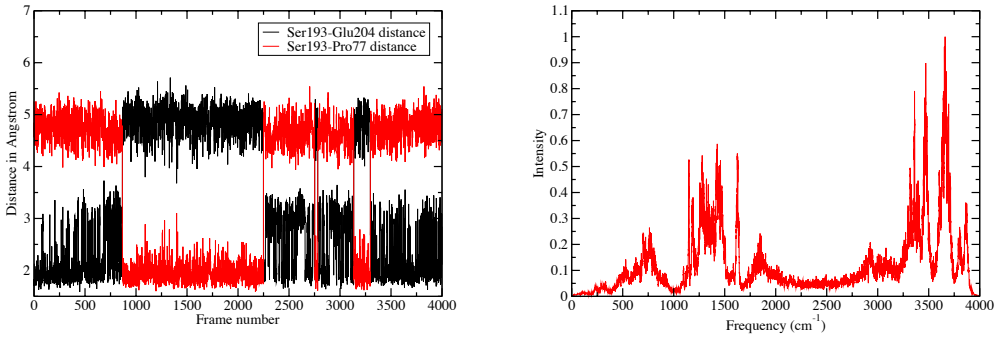


FIG. S7. Average structure of the PRG region (in blue) and water/proton distribution from a 1.75 ns MD simulation with three water molecules in the L-state, as compared to crystal structures (yellow: 1C3W for the ground state; orange: 1UCQ for the L-state). The green dots indicate distribution of the excess proton during the simulation.

To further explore factors that dictate the orientation of Arg82, we perform simulations for the L state after removing the two additional water molecules from the PRG region. For 1.75 ns, no additional water molecules are observed to penetrate into the region. The average orientations of key sidechains, including Arg82, from this trajectory agrees much better with the crystal structures (Fig.S7); improvement in the average water positions is also observed. The most important observation, however, is that the excess proton (green dots in Fig.S6) still prefers to be delocalized between the pair of Glu residues rather than on the water molecules. Therefore, our model holds regardless of the number of water molecules included in the PRG region.

C. An "in" and "out" Ser193 simulation at 170K



(a) (b)

FIG. S8. Orientation of Ser193 sidechain and calculated IR spectra for the PRG region in a MD simulation at 170K, in which Ser193 is observed to switch spontaneously between the "in" and "out" configurations. The trajectory is 2 ns long, frames in (a) are separated by 0.5 ps.

The PMF results presented in the main text suggests that the sidechain of Ser193 tends to have more similar populations for the "in" and "out" configurations at 170 K compared to higher temperatures. Indeed, in one of the trajectories, we observe spontaneous isomerization of the Ser193 sidechain between these two orientations (Fig.S8a); more than five isomerization events have been observed in 2 ns of simulations. It is interesting that even with the mixed populations for the "in" and "out" configurations the intensity of the continuum band gets quenched substantially (Fig.S8b). In other words, it is not necessary to have the Ser193 sidechain to be exclusively in the "out" configuration to observe a significant quench in the intensity of the continuum band. This observation is important since the PMF results suggest that the "out" configuration is unlikely dominant even at 170 K.

D. IR spectra from 230 K simulation

At 230 K, which is the temperature where bR was observed to have very different dynamics[9,10], the calculations spectra (Fig.S9) are qualitatively very similar to those at 300 K, especially the continuum band region ~ 2000 cm^{-1}. This is also consistent with the Ser193 PMF results shown in the main text that at 230K, the behavior of Ser193 sidechain

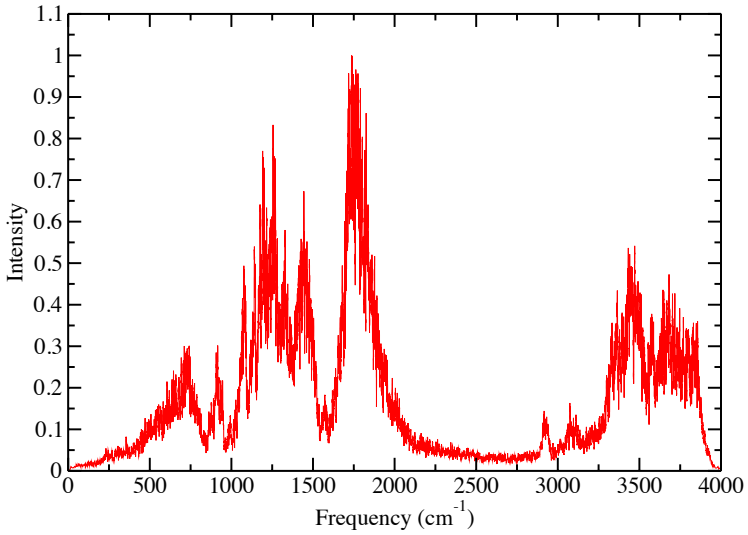


FIG. S9. Calculated IR spectra for the PRG based on SCC-DFTB/MM simulations at 230K.

(thus the degree of delocalization for the excess proton) is very similar to that at 300 K. As discussed in the main text, the discrepancy between calculations and experimental observation of Kandori et al.[11] is possibly due to the fact that the force field used here does not quantitatively capture the temperature dependence of water/protein dynamics.

VI. TWO PROTON MODEL

An interesting alternative for the PRG is a doubly protonated model that reconciles both the delocalized water cluster and our proposed "intermolecular proton bond" model. In such a model two excess protons would reside in the PRG region; a "functional" proton that binds to the water cluster as in the proposal of Gerwert and co-workers[12], and a "structural" proton that stabilizes the pair of Glu residues as in our proposal[3]. The released proton during the L-to-M transition could in principle originate from either of these two sites. If the "structural" proton from the intermolecular proton bond is released, the "functional" proton stored in the nearby water cluster would transfer and get delocalized between the two glutamate residues thereby stabilizing the negative charge.

To test this hypothesis we carry out simulations in the L-state (1UCQ) using a structure taken from the WT L-state simulations with the first proton shared between Glu194/204 and

adding a second excess on the PRG water cluster (Fig S10a). During simulation the second proton quickly hops from the water cluster to Glu194; as a result, the proton originally shared between the Glu sidechains becomes localized onto Glu204 as shown in Fig S10b. The side chain of the two residues remain close to each other during the ∼2 ns simulation. Due to the localized nature of the protons, the corresponding IR spectra lack the continuum band but shows a peak corresponding to the ν_{COOH} stretch (Fig S10c). Therefore, it is highly unlikely that the doubly protonated state model is valid.

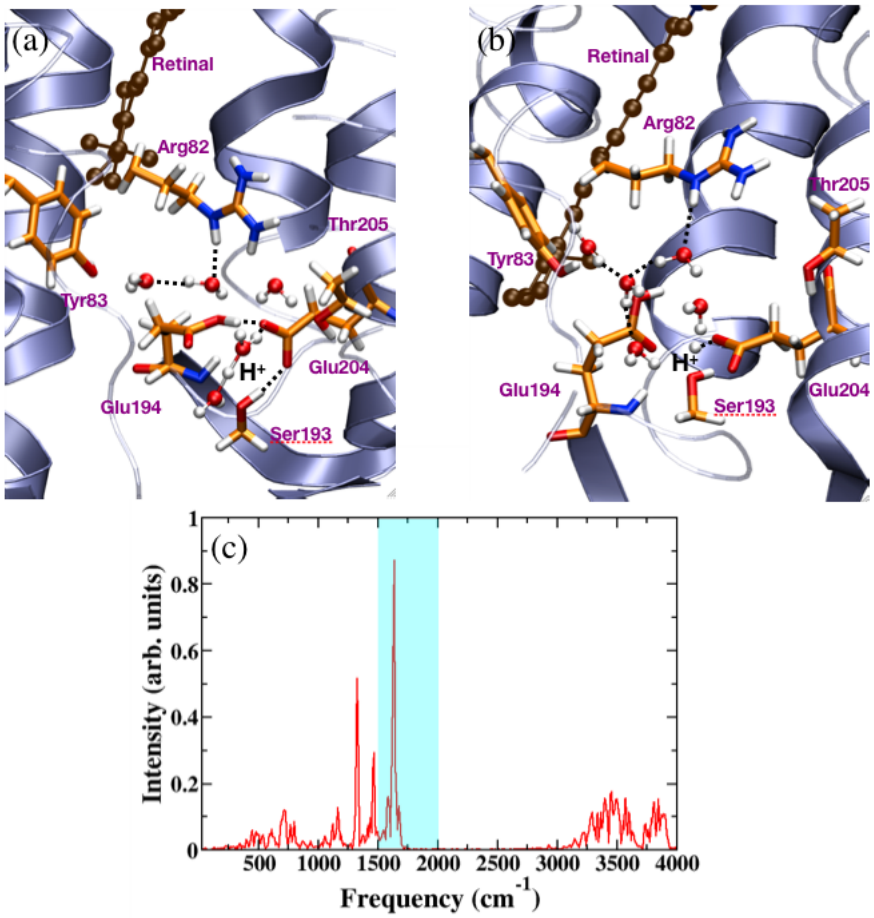


FIG. S10. Results from SCC-DFTB/MM simulations with two protons in the PRG. (a) shows a snapshot of the starting structure used to initiate the simulations; (b) shows a snapshot after 1 ns of simulation, showing both protons localized onto the Glu residues; (c) shows the corresponding IR spectra, which do not exhibit the continuum feature in the 1800-2000 cm^{-1} region.

VII. NMR CHEMICAL SHIFT PATTERNS FOR THE TWO PRG MODELS

As mentioned in the main text, one potential experimental approach to distinguish the two PRG models is to use ^{17}O-NMR with ^{17}O-labeled Glu residues in bR. To illustrate this point, we carry out model calculations in the gas phase; in principle, chemical shifts can also be calculated for the PRG in bR with QM/MM approach[13].

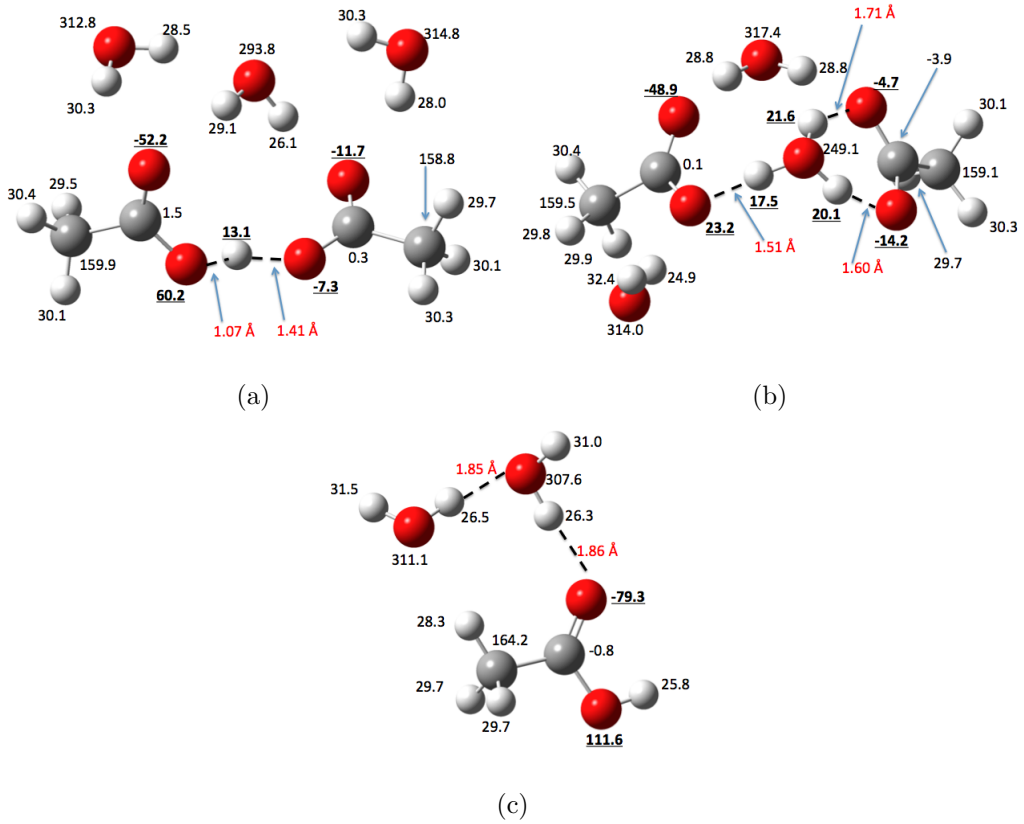


FIG. S11. Absolute chemical shifts calculated at the B3LYP/6-311++G(2d,2p) level with geometries optimized with the same method in the gas phase. Several key shifts are labeled in bold, and key geometrical parameters are in red.

Although these are only gas-phase models, the results help highlight that the two PRG models are expected to have different ^{17}O and proton chemical shifts patterns. For the Glu-pair model, only the shared proton has significantly different shift compared to other protons in the region; the oxygen atom engaged in strong interaction with the excess proton also has very different chemical shift compared to either fully deprotonated or protonated Glu (e.g., compare Fig.S11a and c). For the water cluster model that contains the hydronium between the two Glu sidechains, three protons exhibit significantly different shifts and the

Glu oxygens also have different chemical shifts compared to the Glu-pair model. Therefore, if it is indeed feasible to carry out ^{17}O NMR study for ^{17}O-labeled protein, comparing measured chemical shifts with QM/MM calculations can be very effective for distinguishing the two PRG models.

REFERENCES

* marcus.elstner@kit.edu, cui@chem.wisc.edu

[1] Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theo. Comp.* **2011,** *7,* 931-948.

[2] Witek, H. A.; Morokuma, K. *J. Comp. Chem.* **2004,** *25,* 1858-1864.

[3] Phatak, P.; Ghosh, N.; Yu, H.; Cui, Q.; Elstner, M. *Proc. Acad. Natl. Sci. U.S.A.* **2008,** *105,* 19672-19677.

[4] Roscioli, J. R.; McCunn, L. R.; Johnson, M. A. *Science* **2007,** *316,* 249-254.

[5] Craig, I. R.; Manolopoulos, D. E. *J. Chem. Phys.* **2004,** *121,* 3368-3373.

[6] Witt, A.; Ivanov, S. D.; Shiga, M.; Forbert, H.; Marx, D. *J. Chem. Phys.* **2009,** *130,* 194510.

[7] Elstner, M.; Porezag, D.; Jungnickel, G.; Elstner, J.; Haugk, M.; Frauenheim, Th.; Suhai, S. and Seifert, G., *Phys. Rev. B* **1998,** *58,* 7260-7268.

[8] Kouyama, T.; Nishikawa, T.; Tokuhisa, T. and Okumura, H., *J. Mol. Biol.* **2004,** *335,* 531-546.

[9] Ferrand, M.; Dianoux, A. J.; Petry, W.; Zaccai, G. *Proc. Natl. Acad. Sci. USA* **1993,** *90,* 9668-9672.

[10] Lechnert, U.; Reat, V.; Weik, M.; Zaccai, G.; Pfister, C. *Biophys. J.* **1998,** *75,* 1945-1952.

[11] Lorenz-Fonfria, V. A.; Furutani, Y.; Kandori, H. *Biochem.* **2008,** *47,* 4071-4081.

[12] Garczarek, F.; Gerwert, K. *Nature* **2006,** *439,* 109-112.

[13] Cui, Q.; Karplus, M. *J. Phys. Chem B* **2000,** *104,* 3721-3743.

TABLE S1. Statistical Analysis of pK_a simulations for bR[a]

λ	L-state-Set1		L-state-Set2		Ground state		"early M"-state[b]	
	prod(equi)	τ (n)	prod(equi)	τ (n)	prod(equi)	τ (n)	prod(equi)	τ (n)
0.00	1.5(0.6)	21(42)	1.5(0.5)	24(42)	1.5(0.3)	21(57)	1.5(0.5)	24(41)
0.25	1.5(0.5)	20(50)	1.5(0.6)	21(43)	1.5(0.6)	20(45)	1.5(0.3)	26(46)
0.50	1.5(0.4)	24(46)	1.5(0.6)	22(45)	1.5(0.4)	18(61)	1.5(0.4)	20(55)
0.75	1.5(0.3)	20(60)	1.5(0.5)	19(52)	1.5(0.6)	19(47)	1.5(0.3)	28(43)
1.00	1.5(0.6)	21(42)	1.5(0.6)	24(37)	1.5(0.5)	20(50)	1.5(0.5)	21(48)

a. "prod(equi)" gives the total amount of simulation time (in ns) and the segment identified as equilibration (in parentheses). τ gives the size of the block (in ps) and n gives the total number of blocks in the final free energy derivative calculations. b. a model for the "early M" state by equilibrating after modifying the protonation state of the retinal and Asp85 (see **Methods**).

TABLE S2. Calculated hydrogen bonding interactions (in kcal/mol) between Ser/Cys sidechains and a deprotonated carboxylate group[a]

Interaction	B3LYP	MP2
Ser $(CH_3OH)\cdots CH_3COO^-$	-16.2	-16.1
Cys $(CH_3SH)\cdots CH_3COO^-$	-11.6	-11.8

a. The geometries are fully optimized at the B3LYP/6-31+G(d,p) level; for single point energy calculations, a large basis set, 6-311++G(2d,2p) is used for both B3LYP and MP2.

VIII. COMPLETE REFERENCE FOR REF. 40 IN THE MAIN TEXT

Ref. 40: A. D. J. MacKerell, D. Bashford, M. Bellott, R. L. Jr.Dunbrack, J. D. Evenseck, M. J. Field, S. Fis- cher, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. I. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiókiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B, 102:3586-3616 (1998).