Supplemental Information

Deprotonation of D96 in bacteriorhodopsin opens the proton uptake pathway

Ting Wang[1#], Ayla O. Sessions[1, 3#], Christopher S. Lunde[1], Shahab Rouhani[2], Robert M. Glaeser[2], Yong Duan[1], and Marc T. Facciotti[1*]*

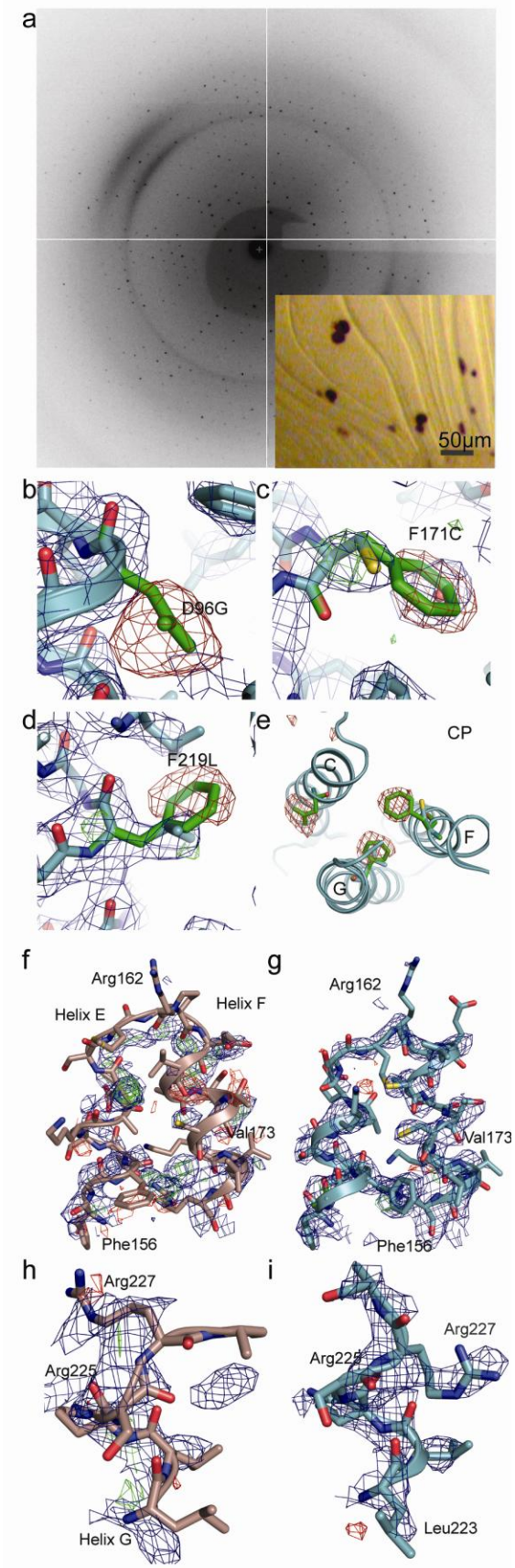
a
50μm
b
D96G
c
F171C
d
F219L
e
CP
C
F
G
f
Arg162
Helix E
Helix F
Val173
Phe156
g
Arg162
Val173
Phe156
h
Arg227
Arg225
Helix G
i
Arg227
Arg225
Leu223

Figure S1. **Crystals of the triple mutant and an associated diffraction pattern.** (a) A diffraction pattern from one of the crystals used in this study. The inset shows triple mutant crystals (purple) embedded in retinal saturated phosphatidylethanolamine-based continuous lipid bilayer gel (yellow). A 50 μm scale bar is included. (b) Blue electron density maps were calculated for *in silico* mutants of the triple mutant structure reverted to the wild-type amino acid composition. 2Fo-Fc electron density maps contoured at 1.1σ are drawn in blue, positive Fo-Fc maps contoured at 3σ are drawn in green while negative Fo-Fc maps contoured at 3σ are drawn in red. Triple mutant structure is colored cyan while mutant side-chains are overlaid in green. The electron density around G96. The introduction of wild-type D96 induces a large negative feature in the Fo-Fc map and no positive feature. (c) The density around C171. Substitution of wild-type F171 induces a large negative peak and corresponding positive peak that is well modeled with the mutated residue. (d) The density around L219. Substitution of wild-type F219 induces a large negative peak and corresponding positive peak that is well modeled with the mutated residue. (e) A view of the negative Fo-Fc peaks and corresponding wild-type side-chains shown from the cytoplasmic side of the protein. (f) For panels f, g, h, and i the EM-derived structure of the triple mutant 1FBK was aligned to 4PFD using the alpha carbon atoms of residues located in the structurally invariant transmembrane core of the proteins. 2Fo-Fc maps **(**contoured at 1.1σ**)** generated with the 4PFD structure factors and atomic coordinates for the model highlighted in each panel are drawn as blue isosurfaces. Positive Fo-Fc maps (contoured at 3σ) are drawn in green while negative Fo-Fc maps are drawn in red. **(**f**)** Residues 154-175 in the EF loop of EM-derived model 1FBK are shown. Note the EM-derived model does not fit well in the electron density derived from x-ray data. **(**g**)** Residues 154-175 in the EF loop of model 4PFD are shown. **(**h**)** Residues 223-228, the terminus of helix-G, of model 1FBK are shown. **(**i**)** Residues 223-228, the terminal of helix-G, of model 4PFD are shown.

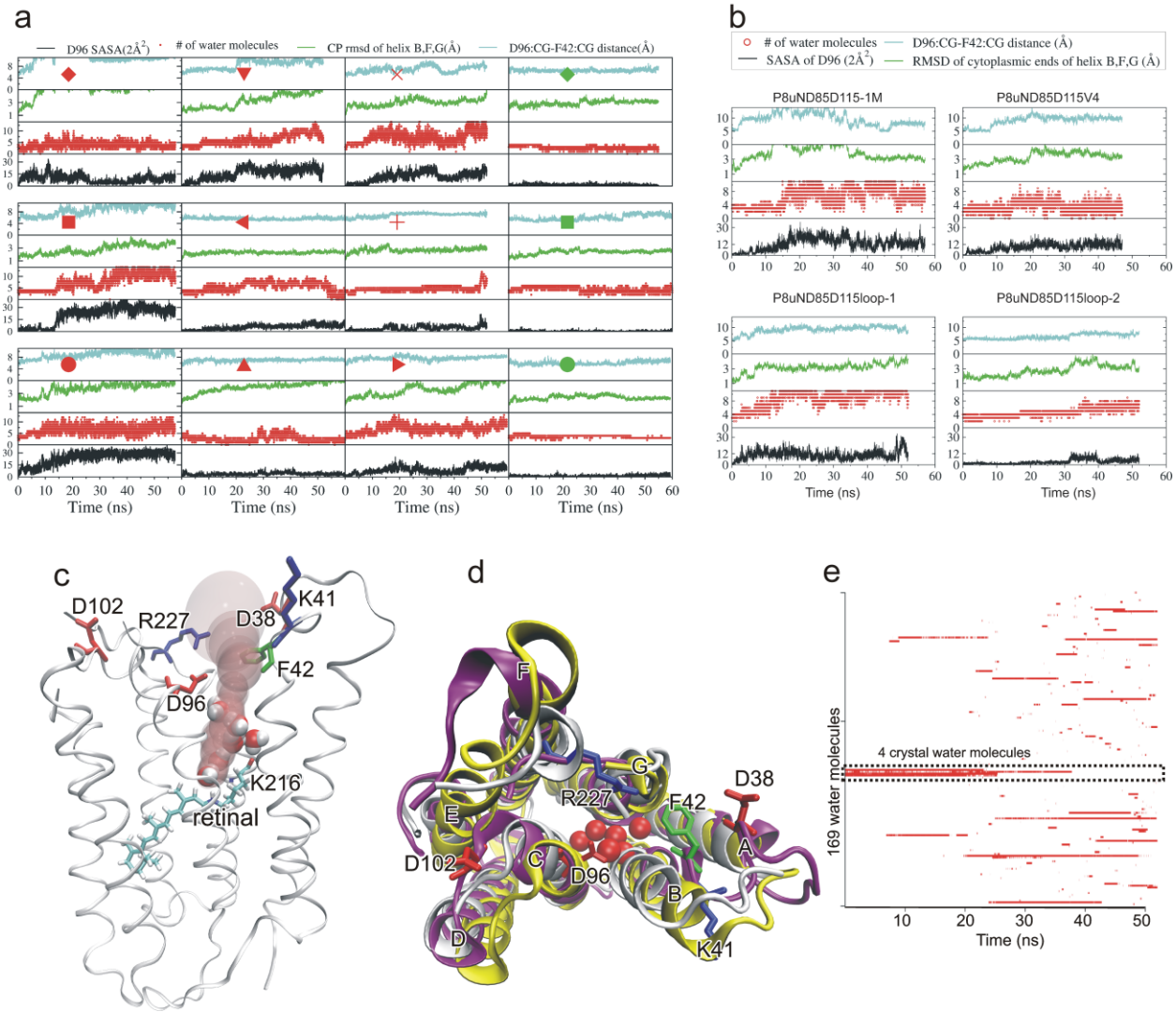

Figure S2. Simulation results based on structure 1P8U. (a) Dynamics in the vicinity of D96. Left three columns show 9 simulations with deprotonated D96 and each is labeled by a *red* symbol. The most right column shows 3 simulations with protonated D96, each labeled with a *green* symbol. The detail of the simulation that each symbol represents is found in Table S4. For each simulation, the D96-F42 side chain Cγ-Cγ distance (*cyan*), the RMSD of the cytoplasmic ends of helices B, F and G (*green*), the number of water molecules in the D96-K216 cavity (*red*), and the SASA of D96 (*black*), are each plotted against the simulation time.

(b) Trajectory analysis of four additional control simulations based on structure 1P8U. In these four simulations, D96 was deprotonated and D85 and D115 were protonated to mimic the N state. The differences between these four simulations and the simulations (p8uND85D115-1, -2 and -3) listed in Table S4 are as follows: In simulation p8uND85D115-1M, the salt concentration was set to 1M. In simulation p8uND85D115V49, the alanine residue at position 49 was mutated back to valine as in the wild type and only three crystal water molecules (WAT502, WAT505

and WAT506) were kept in the starting structure. In simulations p8uND85D115loop-1 and p8uND85D115loop-2, the EF loop (residues 157 to 161) was modeling in.

(c-d) Simulated structure of 1P8U with deprotonated D96 and protonated D85 and D115. The atomic coordinates of 1P8U-derived simulations shown in panels c and d are from the last snapshot of the 52 ns simulation. (c) A contiguous water channel depicted as a pink surface reached the Schiff base and six water molecules (*red ball*) filled the D96-K216 cavity, forming hydrogen bonds with each other and with the Schiff base. (d) The cytoplasmic end of helix F tilted outwards and that of helix B unfolded to a coil. The starting structure is shown in white, the simulated 1P8U structure is drawn in purple, and the crystal structure 4PFD is shown in yellow. (e) The binary occurrence of each of the 169 water molecules that accessed the D96-K216 cavity. Red dots indicate presence of water molecule. The four crystal water molecules had residence time longer than 20 ns.

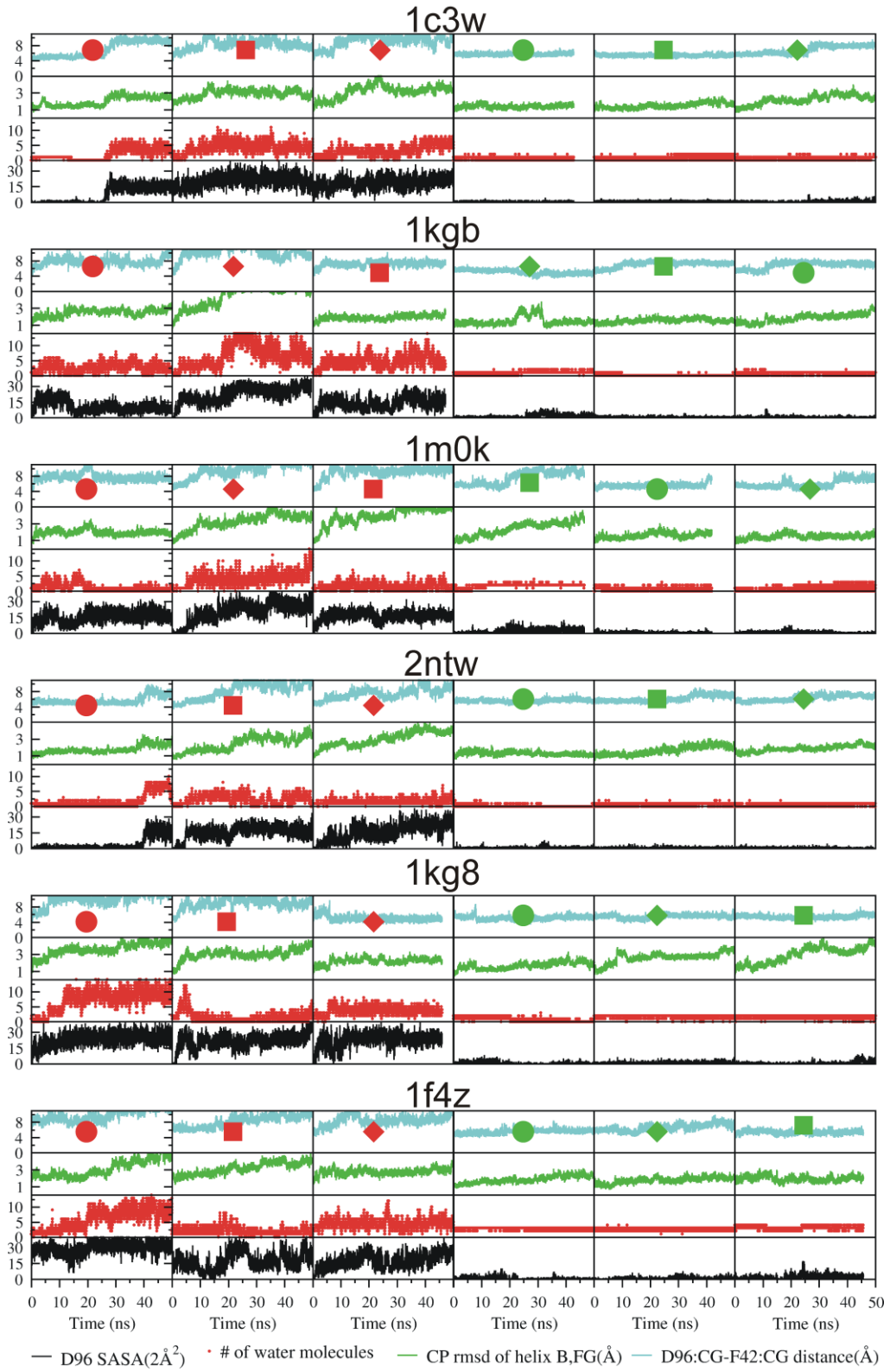
1c3w
1kgb
1m0k
2ntw
1kg8
1f4z
Time (ns)
D96 SASA(2Å^2)
# of water molecules
CP rmsd of helix B,FG(Å)
D96:CG-F42:CG distance(Å)

Figure S3. **Trajectory analyses of the simulations based on the x-ray structures of the ground (1c3w and 1kgb), K (1m0k), L (2ntw), M1 (1kg8), and M2 (1f4z) states.** Left three columns show the simulations with deprotonated D96 and each is labeled by a *red* symbol. The right three columns show the simulations with protonated D96, each labeled with a *green* symbol. The detail of the simulation that each symbol represents is in Table S4. For each simulation, the D96-F42 side chain Cγ-Cγ distance (*cyan*), the RMSD of the cytoplasmic ends of helices B, F and G (*green*), the number of water molecules in the D96-K216 cavity (*red*), and the SASA of D96 (*black*), are each plotted against the simulation time.

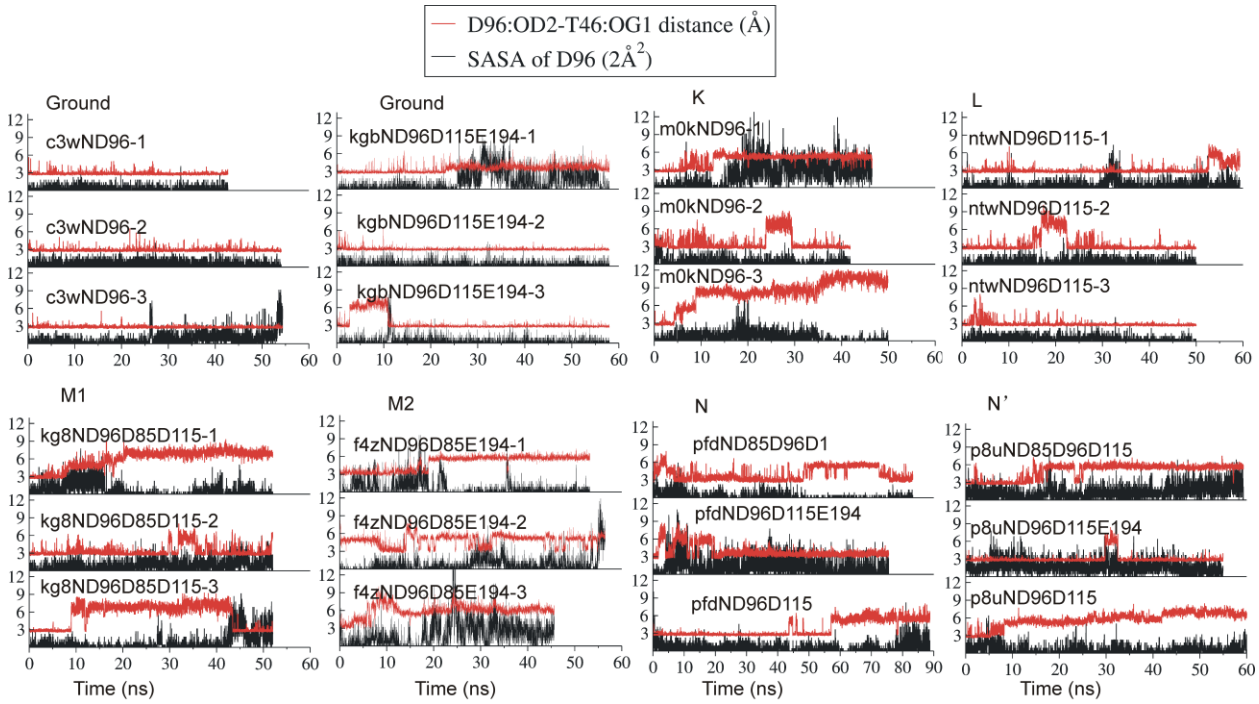

Figure S4. **Time series of the D96-T46 distance and solvent accessible surface around D96 in the simulations with protonated D96.** Red lines indicate the D96-T46 distance while the SASA around D96 is rendered as a black line. Note lack of correlation between D96-T46 bond distance and SASA. This indicates that even when the D96-T46 bond is broken the channel remains functionally closed when D96 is protonated.

SupplementalTables

Table S1-A: Results of Perfect Merohedral Twinning Test

Resolution	$\langle I^2 \rangle / \langle I \rangle^2$	No. reflections
13.195	6.38	424
5.798	2.33	423
4.830	2.64	423
4.308	2.07	424
3.959	2.18	423
3.701	2.13	423
3.500	2.06	424
3.340	2.15	423
3.204	2.12	423
3.088	1.88	424
2.987	2.02	423
2.898	2.05	423
2.818	1.79	424
2.747	1.71	423
2.682	1.73	423

Resolution	$<I^2>/<I>^2$	No. reflections
Data reflect results of the "Perfect Merohedral Twinning Test" available at (http://nihserver.mbi.ucla.edu/Twinning/). The applied test looks for deviations from Wilson statistics. The webserver computes the following expression $<I^2>/<I>^2$, where the variable I corresponds to reflection intensity, in thin resolution shells. The expected value is 1.5 for twinned data and 2.0 for untwined data.		

Table S1-B: Results of Partial Merohedral Twinning Test

Reported Parameter	**Value**
No. Total Data Read	6579
No. Acentric Data	6362
No. Data Within Resolution Range	6350
No. Twin Law Related Reflections	2691 (pairs)
No. Twin Law Pairs Considered	1346
$\langle H \rangle$	0.507351
$\langle H^2 \rangle$	0.334698
Twin Fraction	0.004187 +/- 0.003164
Data reflect results of the "Partial Merohedral Twinning Test" available at (http://nihserver.mbi.ucla.edu/Twinning/). This test provides a test of partial merohedral twinning. $H=\|I^1_{obs}-I^2_{obs}\|/(I^1_{obs}-I^2_{obs})$ where I^1 and I^2 are spot intensities for twin related reflections. $\langle H \rangle = 0.5 - \alpha$ and $\langle H^2 \rangle = (1-2\alpha)^2/3$, where α is the twin fraction.	

Table S2: Properties of the eight crystal structures used in our simulations.

PDB code	State	Sample	Resolution (Å)	Residues in the model	Number of water in the D96-K216 cavity[a]	D96:OD2-T46:OG1 distance (Å)	D96:CA-K216:CA distance (Å)
1KGB	Br	Wild type	1.65	5-156, 162-231	1	2.63	11.25
1C3W	Br	Wild type	1.55	5-156, 162-231	1	2.58	11.19
1M0K(2)	K	Wild type	2.1	5-156, 162-231	1	2.71	11.39
2NTW(1)	L	Wild type	1.53	5-156, 162-231	1	2.82	11.48
1KG8	M1	Wild type	2.0	5-155, 167-231	2	3.65	11.94
1F4Z	M2	E204Q	1.8	5-156, 162-231	3	3.57	11.72
4PFD	N	D96G, F171C, F219L	2.65	5-229	1	NA	11.52
1P8U	N'	V49A	1.62	5-156, 162-231	4	2.60	11.63

[a]: WAT501 in most of the crystal structures and WAT504 in 1p8u were excluded by the definition of the D96-K216 cavity sphere used in this work.

Table S3. Result summary of 60 MD simulations, complement to **Figs. 2 and 4** in the main text.

Simulation [a]	Symbol in Figs 2 and 4	State	Protonated D/E residues	Number of different water molecules that accessed the D96-K216 cavity [b]	Helix RMSD (Å)	Trajectory length (ns)
kgb-1	●	Br	none	120 (2.5)	2.54[c] (1.32[d])	52.47
kgb-2	◆	Br	none	940 (6.7)	4.54 (1.14)	49.42
kgb-3	■	Br	none	191 (4.6)	1.91 (1.14)	47.17
kgbND96D85D115-1	◆	Br	D96, D85, D115	2 (0.8)	1.52 (0.96)	58.00
kgbND96D85D115-2	■	Br	D96, D85, D115	2 (0.1)	1.55 (0.91)	58.00
kgbND96D85D115-3	●	Br	D96, D85, D115	2 (0.9)	2.03 (0.90)	58.00
C3w-1	●	Br	none	88 (1.6)	2.03 (1.11)	53.64
C3w-2	■	Br	none	356 (4.0)	2.90 (1.08)	55.57
C3w-3	◆	Br	none	188 (3.1)	3.18 (1.21)	50.48
C3wND96-1	●	Br	D96	2 (0.8)	1.40 (0.94)	42.71
C3wND96-2	■	Br	D96	3 (0.9)	1.51 (0.96)	54.06
C3wND96-3	◆	Br	D96	3 (0.8)	2.27 (1.70)	54.38
m0k-1	●	K	none	93 (0.7)	1.97 (1.02)	55.21
m0k-2	◆	K	none	619 (4.1)	3.31 (1.39)	53.96
m0k-3	■	K	none	217 (1.2)	4.02 (1.12)	55.96
m0kND96-1	■	K	D96	5 (0.8)	2.41 (1.16)	46.50
m0kND96-2	●	K	D96	5 (1.8)	1.61 (1.05)	41.83
m0kND96-3	◆	K	D96	3 (0.6)	1.48 (1.16)	49.89
Ntw-1	●	L	none	48 (1.8)	1.83 (1.18)	54.00
Ntw-2	■	L	none	119 (2.1)	2.47 (1.02)	52.00
Ntw-3	◆	L	none	165 (1.6)	3.05 (1.50)	52.00
NtwND96D115-1	●	L	D96, D115	4 (0.1)	1.27 (0.87)	59.33
NtwND96D115-2	■	L	D96, D115	2 (0.8)	1.66 (0.84)	50.00

NtwND96D115-3	◆	L	D96, D115	3 (0.7)	1.88 (1.05)	50.00
Kg8-1	●	M1	none	1002 (6.8)	3.63 (1.56)	50.00
Kg8-2	■	M1	none	230 (1.2)	3.03 (1.53)	50.00
Kg8-3	◆	M1	none	584 (3.7)	2.21 (1.35)	45.93
Kg8ND96D85D115-1	●	M1	D96, D85, D115	11 (0.6)	1.69 (1.29)	52.00
Kg8ND96D85D115-2	◆	M1	D96, D85, D115	2 (1.9)	2.69 (1.09)	52.00
Kg8ND96D85D115-3	■	M1	D96, D85, D115	2 (1.7)	3.01 (1.15)	52.00
f4z-1	●	M2	none	914 (5.4)	3.23 (1.54)	53.35
f4z-2	■	M2	none	219 (1.9)	3.28 (1.64)	57.03
f4z-3	◆	M2	none	273 (4.4)	2.79 (1.24)	52.65
f4zND96D85D115-1	●	M2	D96, D85, D115	3 (2.9)	1.87 (1.23)	53.23
f4zND96D85D115-2	◆	M2	D96, D85, D115	5 (2.6)	1.84 (0.95)	56.46
f4zND96D85D115-3	■	M2	D96, D85, D115	5 (3.2)	1.96 (1.36)	45.63
pfd-1	●	N	none	409 (4.0)	3.66 (1.99)	85.36
pfd-2	■	N	none	280 (4.4)	2.71 (1.04)	59.97
pfd-3	◆	N	none	259 (5.3)	4.30 (1.13)	66.99
pfdND85D115-1	▲	N	D85, D115	261 (5.3)	2.20 (1.13)	87.13
pfdND85D115-2	◀	N	D85, D115	488 (5.5)	2.67 (1.18)	60.00
pfdND85D115-3	▼	N	D85, D115	72 (2.3)	3.43 (1.03)	60.00
pfdND85D115E204-1	▶	N	D85, D115, E204	20 (3.2)	3.17 (1.00)	84.25
pfdND85D115E204-2	+	N	D85, D115, E204	166 (5.7)	2.68 (1.45)	60.00
pfdND85D115E204-3	×	N	D85, D115, E204	380 (3.6)	2.93 (1.2)	60.00
pfdND85D96D115	◆	N	D96, D85, D115	2 (0.1)	2.54 (1.23)	83.29
pfdND96D115	■	N	D96, D115	1 (1.0)	3.01 (1.57)	88.75
pfdND96D115E194	●	N	D96, D115, E194	2 (1.0)	2.43 (0.86)	75.54
P8u-1	●	N'	none	795 (6.6)	3.75 (1.20)	57.57
P8u-2	■	N'	none	622 (7.8)	2.80 (1.24)	57.71
P8u-3	◆	N'	none	223 (3.9)	5.25 (1.09)	57.69
p8uND85D115-1	▲	N'	D85, D115	26 (2.7)	4.00 (1.32)	60.03

p8uND85D115-2	◀	N'	D85, D115	52 (5.5)	2.27 (1.23)	60.00
p8uND85D115-3	▼	N'	D85, D115	169 (6.2)	3.27 (1.06)	52.00
p8uND85D115E194-1	▶	N'	D85, D115, E194	110 (7.1)	3.60 (1.16)	59.34
p8uND85D115E194-2	+	N'	D85, D115, E194	21 (4.7)	2.67 (1.12)	52.00
p8uND85D115E194-3	×	N'	D85, D115, E194	163 (7.2)	2.99 (1.09)	52.00
p8uND85D96D115	●	N'	D96, D85, D115	5 (3.6)	2.35 (1.23)	59.31
p8uND96D115	■	N'	D96, D115	7 (4.3)	2.35 (1.15)	60.19
p8uND96D115E194	◆	N'	D96, D115, E194	6 (2.8)	2.92 (1.13)	54.95

[a]: The name of each simulation began with the first three letters of the crystal structure PDB code and followed by letter N if any D/E residue was protonated. For example, name kgbND96D85D115-1 indicates that this was the first simulation that started from the crystal structure of 1KGB and residues D96, D85 and D115 were protonated while name kgb-1 indicates that this was the first simulation that started from the crystal of 1KGB and none of the D or E residue was protonated. *Green*: simulations with protonated D96; *black*: simulations with deprotonated D96; *red*: simulations to mimic the N state.

[b]: Number of ***different*** water molecules accessing the D96-K216 cavity. Numbers in parenthesis are the averaged number of water molecules per simulation frame.

[c]: averaged RMSD of the cytoplasmic end of helices B, F, and G

[d]: averaged RMSD of the helices except for the cytoplasmic end of helices B, F, and G.

Supplemental Movie Legends

Movie S1. A video clip showing the simulation trajectory pfdND85D115-1(N state, 87ns). D96, F42, and retinal-Schiff base-K216 are shown as sticks in red, green and cyan colors, respectively. Water molecules within 4.0 Å of D96 or within 3.5 Å of K216 are shown as red and white van der Waals balls. A water channel forms immediately after the unrestrained simulation begins.

Movie S2. Simulation trajectory c3w-1 (ground state, 57ns). D96, F42, and retinal-Schiff base-K216 are shown as sticks in red, green and cyan colors, respectively. Water molecules within 4.0 Å of D96 or within 3.5 Å of K216 are shown as red and white van der Waals balls. A water channel forms at ca. 28 ns.

Supplemental Experimental Procedures

Crystallization

Mutant membranes were obtained as a kind gift from Drs. Dieter Oesterhelt and Jörg Tittor. Octyl glucoside was added to the membrane suspension at a final concentration of 1.2% immediately before the membrane suspension was used to hydrate the lipid. In addition, the added octyl glucoside was intentionally saturated in this case with free retinal in order to shift the equilibrium towards the native (retinal-bound) state during incubation.

Crystals were grown in continuous lipid bilayer (CLB) gels composed of phosphatidylethanolamine (PE) lipids. The CLB gel composition consisted of 95% monomethyl-DOPE, 5% DOPE-mPEG350 hydrated to 67% hydration with the detergent/retinal/membrane suspension mixture. Phosphatidylethanolamine (PE) lipids and their conjugates were purchased from Avanti Polar Lipids (Alabaster, Al). Chloroform stocks lipids were mixed according to the desired mass-ratio of lipid and then dried. To form the lipid bilayer gel the dry lipid was mixed with protein solution to a final aqueous concentration of 67% (w/w) using the syringe mixing apparatus described previously(Cheng et al., 1998; Rouhani et al., 2002).

Crystallization screens were carried out in 96 well plates. In short, 0.2-1 µl drops of the gel were dispensed in the bottom of each well using the syringe-dispensing apparatus. The gel drops were then covered with 5 mm glass coverslips (Bellco Glass, Vineland, NJ) and overlaid with 50 ml of crystallization solution (0.1 M Bis-Tris, pH 5.5, 25% PEG 3350). The plate was sealed using Clear Seal tape (Hampton Research, Aliso Viejo, CA). Crystallization was carried out at room temperature (22°C). Figure S1 shows a photograph of the typical crystals that were grown in these conditions. The crystals were mechanically removed from the gel using 20 µm diameter nylon cryo-loops (Hampton Research, Laguna Niguel, CA) and plunge frozen in liquid nitrogen.

In silico mutant verification

At the beginning of the refinement process electron density maps were calculated with wild-type residues at position 96, 171 and 219. Both 2Fo-Fc and Fo-Fc maps showed evidence of mutations at these sites in the structure. Figure S1 illustrates this point by showing electron density maps calculated from our experimental structure factors and our final model in which wild type amino acid residues were substituted at positions 96, 171, and 219.

Data collection, reduction, and refinement

The X-ray diffraction data were collected at Beamline 8.3.1 of the Advanced Light Source at the Lawrence Berkeley National Laboratory. Crystals were screened using a 30 µm pinhole collimator. Exposure times depended on the crystal size and varied from 15 seconds to 3 minutes for 1 degree of rotation. Data reduction was accomplished with the ELVES scripts(Holton and Alber, 2004). Diffraction data from two crystals were integrated with MOSFLM(Leslie, 1992) and subsequently scaled with SCALA(1994). The twin fraction, as determined by the UCLA twinning server (http://nihserver.mbi.ucla.edu/Twinning/), is extremely low (Tables S2 A and 2B). Molecular replacement using 1KGB(Facciotti et al., 2001) as the starting search model, without the retinal, water, and lipid molecules, was performed by the program CNS Version 1.2(Brunger, 2007). Refinement with CNS and model building using the program O(Jones et al., 2001) together with annealed simulated omit, |Fo| - |Fc|, and 3|Fo| - 2|Fc| maps reduced the values of R and Rfree to their final values. Figures were generated using MacPyMol(Nishizawa et al.) version 10.

Evidence of Movement in EF-loop and Helix G regions

Figure S1 illustrates the electron density maps in the EF-loop and Helix G regions. Both 2Fo-Fc and Fo-Fc maps showed evidence of the movement in these regions in the structure. It is seen that the electron density maps calculated from our experimental structure factors are poorly fitted by the 1FBK structure and are fitted reasonably well by our final model 4PFD.

bR structures and modeling used in the simulations.

Structures in the ground-BR, K, L, M1, M2, N, and N' states were from x-ray crystal structures. Except the N and N' states, more than one crystal structures are available for each state. We chose to use those with high resolutions. For the wild-type ground state, we used two structures, the 1.65 Å resolution 1KGB(Facciotti et al., 2001) and 1.55 Å resolution 1C3W(Luecke et al., 1999). For the K state, we used the 1.43 Å resolution 1M0K (model 2) structure(Schobert et al., 2002). For the L state, we used the 1.65 Å resolution 2NTW structure(Lanyi and Schobert, 2007), which is the latest structure of the L state. The M1 and M2 states were the 1.65 Å resolution 1KG8(Facciotti et al., 2001) and 1F4Z(Luecke et al., 2000), respectively. The 1.62 Å resolution 1P8U(Schobert et al., 2003) was the only structure of the N' state. The 2.65 Å resolution 4PFD was the N-like structure. Table S3 lists the crystal structures used in this work.

None of the crystal structures has the complete sequence of 248 residues encoded in the *bop* gene. The first four N-terminal residues are missing in all structures, as well as the last 17 C-terminal residues. In the N state structure the last 19 C-terminal residues were unsolved. In addition, the EF loop (residue 157 to 161) is missing in all the structures except the N state. For each bR model, we included the protein, the retinal molecule and all crystal water molecules. The lipid and organic solvent molecules solved together with the proteins in the crystal structures were not modeled in. The protein residues not solved in the crystal structures were not modeled in.

The E-F loop, connecting helices E and F, may play roles in the opening of the proton uptake channel. However, the E-F loop was not modeled in all but one (4PFD) of the X-ray structures. Unless otherwise noted, the E-F loop was modeled present or absent as it is found in the PDB file. As noted in the main text, we constructed a model of the N' state based on 1P8U in

which the EF loop was added using the loop modeling server(Phatak et al.). Figure S2 illustrates that the addition of the loop did not change the interpretation of the role of 96 in other simulations lacking loops.

Protonation states of the Asp and Glu residues

With D96 as the focus, we made at least two models for each crystal structure: one with deprotonated D96 and the other protonated. Additional models were made with protonated D85, D115, E194 and E204, depending on the intermediate states. To address the question whether the channel opening is simply due to deprotonation of any of the ionizable sites, rather than the effect specifically related to the protonation states of D96, we performed a series of simulations in which other ionizable sites were also varied, including D85, D115, E194, E204, even though biologically these sites may or may not be involved in the proton transfer process. There were 22 models generated in total based on the eight crystal structures (Tables S3 and S4).

D85 is located in the proton release pathway and it becomes protonated during the L to M transition. We therefore modeled D85 as protonated in some of the models in the states after L. D115 is located beside the proton transport pathway and the crystal structures suggest that it is able to form a hydrogen bond with T90 if protonated. E194 and E204 are located in the proton release pathway at the extracellular end of helices F and G, respectively. There is some controversy over the protonation states of E194 and E204 at various points in the photocycle(Brown et al., 1995). Our models included both the deprotonated and protonated states. Although E204, located in the proton release pathway, may not be directly involved in proton release, to investigate its potential influence on the channel opening, we intentionally varied its protonation states. All other Asp and Glu residues were exclusively deprotonated in all models.

Protonation states of the Schiff base

The Schiff base was protonated in the structural models of the ground, K, L, N, and N' states and deprotonated in the M1 and M2 states.

Modeling of the wild type in the N state

The three mutations D96G, F171C and F219L in the crystal structure of the N state were mutated back to the wild type. This was carried out by using the Mutator plugin in the VMD1.8.6 program(Humphrey et al., 1996).

Modeling of bR in solvated membrane environment

First, we used the REDUCE program(Word et al., 1999) to add hydrogen atoms into each bR structure. We then embedded each structure into a lipid bilayer. The palmitoyloleoyl-phosphatidylcholine (POPC) lipid molecule was used and the VMD1.8.6 program(Humphrey et al., 1996) was used to generate a patch of POPC lipid membrane with a dimension of ca. 100 Å × 100Å. Each bR model was then aligned with the axes of the membrane and the center of mass of protein was translated to the center of mass of the membrane by using the combine.tcl script available on the web page of the VMD program. Lipid molecules within 1.5Å of the protein were removed. Depending on the individual bR structure, the number of the lipid molecules in each model was between 120 and 123.

Next, each bR model embedded in the membrane was then solvated in water in the Z-direction, i.e. at the intracellular and extracellular sides parallel to the membrane plane. This was carried out by using the tLeap module in the AMBER8 program(Case et al., 2005) and an in-house script. The salt concentration was set to 200 mM by adding chloride and sodium ions. Depending on the individual bR structure, the number of the water molecules in each model was between 7, 702 and 8, 868 and the total atom number was between 43, 163 and 46, 476.

MD simulations

Each of the solvated bR-membrane systems was first subject to a two-stage energy minimization. In the first stage of 2000 steps, bR was restrained to its crystallographic positions by a harmonic potential with a force constant of 32 kcal/(mol·$Å^2$) while all other atoms were unrestrained. In the second stage of 2000 steps, no restraint was applied. In both stages, the steepest decent method was used for the first 10 steps and the conjugate gradient method was switched on for the rest of the steps.

After energy minimization, the whole system was subjected to a three-stage equilibration MD simulation. In the first stage, the system was gradually heated from 10K to 300K in 50 ps and with constant volume while bR, retinal and the head groups of the lipid molecules were

restrained by a force constant of 32 kcal/(mol·Å^2). The restraints on the head-groups were used to prevent the tails from interdigitating, which would otherwise result in the reduction of the bilayer thickness. In the second stage, the restraints on the protein and the lipid molecules remained but the system was simulated at constant temperature of 300K and constant pressure of 1.0 atm for 500ps. In the third stage, the restraints on the lipid molecules were removed and the system was simulated for 500ps with the restraints remaining on the protein. After the equilibration simulation, the MD simulation continued with all atoms free. The unrestrained MD simulation was the production run used for our analysis. We performed 60 simulations, ranging in length from 41.83 to 88.75 ns, over 3.4 μs in total. The simulation details were summarized in Table S4.

Ewald (PME) method(Essmann et al., 1995) was used for long-range electrostatic interactions with the default parameters. The time step was 1 fs in the equilibration simulation and 2 fs in the production run, and the non-bonded interactions were updated every 10 time steps. The simulation trajectories were saved every 10 ps.

Trajectory Analysis

Water Occupancy Analysis For each of the 60 simulation trajectories, we identified the water molecules that accessed the cytoplasmic proton uptake pathway, i.e. the cavity between D96 and K216. This cavity is approximated by a sphere that is centered at the midpoint of the line connecting the D96:CA atom and the K216:CA atom. The radius of the sphere is the half distance between those two atoms minus 1 Å. As a reference, in the eight crystal structures used in this work, the distances between the D96:CA atom and the K216:CA atom are between 11.2 to 11.9 Å and the numbers of crystal water molecules in the cavities are between 1 and 4 (Table S3). It is worth noting that WAT501 in most of the crystal structures and WAT504 in the N' state (1p8u) were excluded by this definition. For each simulation, we computed the total number of the different water molecules that accessed the D96-K216 cavity as well as the frame-averaged number of the water molecules. The residence time of each water molecule was also recorded. For two trajectories of the N and N' states (pfdND85D115-1 and p8uND85D115-3), we also computed the binary occurrence of each of the cavity water molecules during the simulations.

Other analysis Several structural properties of the protein were analyzed with D96 as the focus. These included the solvent accessible surface area (SASA) of D96, the distance between side chains of D96 and T46, the distance between side chains of D96 and F42, the backbone root-mean-square (RMSD) of the cytoplasmic ends of helices B, F and G, and the backbone RMSD of all helices except for the cytoplasmic end of helices B, F and G. In the calculation of the SASA of D96, only the OD1 and OD2 atoms were used. The cytoplasmic ends of helices B, F, G included residues from D38 to L48, from P165 to F171, and from G220 to L224, respectively.

Channel Analysis We used the Molaxis program(Yaffe et al., 2008) to visualize and analysis the water channel formed in the simulations.

Supplemental References

(1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Brown, L.S., Sasaki, J., Kandori, H., Maeda, A., Needleman, R., and Lanyi, J.K. (1995). Glutamic Acid 204 is the Terminal Proton Release Group at the Extracellular Surface of Bacteriorhodopsin. J Biol Chem *270*, 27122-27126.

Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. Nat. Protocols *2*, 2728-2733.

Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005). The Amber biomolecular simulation programs. J Comput Chem *26*, 1668-1688.

Cheng, A., Hummel, B., Qiu, H., and Caffrey, M. (1998). A simple mechanical mixer for small viscous lipid-containing samples. Chemistry and Physics of Lipids *95*, 11-21.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. J Chem Phys *103*, 8577-8593.

Facciotti, M.T., Rouhani, S., Burkard, F.T., Betancourt, F.M., Downing, K.H., Rose, R.B., McDermott, G., and Glaeser, R.M. (2001). Structure of an early intermediate in the M-state phase of the bacteriorhodopsin photocycle. Biophys J *81*, 3442-3455.

Holton, J., and Alber, T. (2004). Automated protein crystal structure determination using elves. Proc Natl Acad Sci U S A *101*, 1537-1542.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J Mol Graph *14*, 33-38.

Jones, L.J.F., Carballido-Lopez, R., and Errington, J. (2001). Control of Cell Shape in Bacteria: Helical, Actin-like Filaments in Bacillus subtilis. Cell *104*, 913-922.

Lanyi, J.K., and Schobert, B. (2007). Structural Changes in the L Photointermediate of Bacteriorhodopsin. J Mol Biol *365*, 1379-1392.

Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography *26*.

Luecke, H., Schobert, B., Cartailler, J.P., Richter, H.T., Rosengarth, A., Needleman, R., and Lanyi, J.K. (2000). Coupling photoisomerization of retinal to directional transport in bacteriorhodopsin. J Mol Biol *300*, 1237-1255.

Luecke, H., Schobert, B., Richter, H.T., Cartailler, J.P., and Lanyi, J.K. (1999). Structure of bacteriorhodopsin at 1.55 A resolution. J Mol Biol *291*, 899-911.

Nishizawa, R., Nishiyama, T., Hisaichi, K., Matsunaga, N., Minamoto, C., Habashita, H., Takaoka, Y., Toda, M., Shibayama, S., Tada, H., *et al.* (2007). Spirodiketopiperazine-based CCR5 antagonists: Lead optimization from biologically active metabolite. Bioorg Med Chem Lett. *17*, 727-731. Epub 2006 Nov 2001.

Phatak, P., Frahmcke, J.S., Wanko, M., Hoffmann, M., Strodel, P., Smith, J.C., Suhai, S., Bondar, A.N., and Elstner, M. (2009). Long-distance proton transfer with a break in the bacteriorhodopsin active site. J Am Chem Soc *131*, 7064-7078.

Rouhani, S., Facciotti, M.T., Woodcock, G., Cheung, V., Cunningham, C., Nguyen, D., Rad, B., Lin, C.T., Lunde, C.S., and Glaeser, R.M. (2002). Crystallization of membrane proteins from media composed of connected-bilayer gels. Biopolymers *66*, 300-316.

Schobert, B., Brown, L.S., and Lanyi, J.K. (2003). Crystallographic Structures of the M and N Intermediates of Bacteriorhodopsin: Assembly of a Hydrogen-bonded Chain of Water Molecules Between Asp-96 and the Retinal Schiff Base. J Mol Biol *330*, 553-570.

Schobert, B., Cupp-Vickery, J., Hornak, V., Smith, S., and Lanyi, J. (2002). Crystallographic structure of the K intermediate of bacteriorhodopsin: conservation of free energy after photoisomerization of the retinal. J Mol Biol *321*, 715-726.

Word, J.M., Lovell, S.C., Richardson, J.S., and Richardson, D.C. (1999). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol *285*, 1735-1747.

Yaffe, E., Fishelovitch, D., Wolfson, H.J., Halperin, D., and Nussinov, R. (2008). MolAxis: efficient and accurate identification of channels in macromolecules. Proteins *73*, 72-86.