Supporting information for: a-ARM: Automatic Rhodopsin Modeling with Ionization State Selection, Counter-ion Placement and Chromophore Cavity Generation

Laura Pedraza-González,[†] Luca De Vico,[†] María del Carmen Marín,[†] Francesca Fanelli,[‡] and Massimo Olivucci[*,†,¶]

[†]*Department of Biotechnologies, Chemistry and Pharmacy, Università degli Studi di Siena, via A. Moro 2, I-53100 Siena, Italy*

[‡]*Department of Life Sciences, Center for Neuroscience and Neurotechnology, Università degli Studi di Modena e Reggio Emilia, I-41125 Modena, Italy*

[¶]*Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, United States of America*

E-mail: olivucci@unisi.it

Phone: +(39) 0577-234274

Contents

List of Figures **S3**

List of Tables **S4**

S1 Details on the automation limits of the original ARM protocol in terms of the preparation of the `ARM input` **S5**

S2 Workflow of the *a*-ARM protocol **S8**

S3 Detailed workflow of each step for the *a*-ARM automatic input generation **S9**

S3.1 Step 1. Automatic identification of the protein chain, rPSB, chromophore bounded Lys, MC and SC S9

S3.2 Step 2: Automatic generation of the chromophore cavity S10

S3.3 Step 3: Automatic assignment of ionization states S11

S3.4 Step 4: Automatic counter-ion placement S12

S4 Automatic format conversion from LYR to RET plus LYS **S13**

S5 Features of the *a*-ARM QM/MM Models **S14**

S6 Summary of the ARM QM/MM calculations **S17**

S7 Details of the employed comparative modelling protocol for AARh, PoXeR, hMeOp, mMeOp and BCone, GCone, RCone **S26**

S8 Further details on the assignment of ionizable residues protonation state **S28**

S9 Notes **S29**

References **S30**

List of Figures

S1 Automatic Rhodopsin Modeling workflow. The new version (a-ARM) incorporates the original version of the ARM protocol further described in ref. S1. In ARM, N independent simulated annealing (SA) and molecular dynamics (MD) room-temperature relaxations are performed at the MM level using GROMACS on the cavity and the Lys-QM systems. S8

S2 Step 1. Initial preparation of the input file. The red filled boxes represent tasks in which the user may interact with the program. S9

S3 Step 2. Automatic generation of chromophore cavity. The code does not require the user's interaction during its execution. S10

S4 Step 3. Automatic assignment of ionization states. The red filled boxes represent tasks in which the user may interact with the program. S11

S5 Step 4. Automatic selection of counter-ion placement. The code does not require the user's interaction during its execution. S12

List of Tables

S1 Correspondence between atom names in RET and LYR residues. The atom name of LYR and Lys covalently linked to the rPSB are consistent. S13

S2 Main features of the *a*-ARM QM/MM models for the rhodopsins in the benchmark set. S14

S3 Summary of the ten *a*-ARM QM/MM calculations performed for the rhodopsins in the benchmark set. Computed total ground state (S_0) and first excitation (S_1) energies, transition oscillator strength (f_{Osc}), first vertical excitation energy (ΔE_{S1-S0}) and maximum absorption wavelength (λ^a_{max}). Statistical parameters such as average ($\overline{N}$), difference between calculated and experimental data ($|\overline{N}|$) and standard deviation ($\sigma_{\overline{N}}$) are also provided. S17

S4 Trend deviation factor (||Trend Dev.||) for the *a*-ARM$_{\text{default}}$ and *a*-ARM$_{\text{customized}}$ approaches, expressed as mean absolute error (MAE) and mean absolute deviation (MAD) of the x=38 rhodopsins of the benchmark set. S25

S5 Main features of the comparative models S27

S1 Details on the automation limits of the original ARM protocol in terms of the preparation of the `ARM input`

Initial setup of the working PDB file (⌛ ≈ 30 min)

S1.1. The initial template, which is an X-ray structure or a comparative model in PDB format, is cleaned by manual deletion of the lines carrying irrelevant information, so as to extract the information associated with the coordinates of the residues belonging to the QM/MM model. For instance, the selection of the **opsin chain** is handled by manually removing the information on all other chains different from chain A. After this procedure, the `Atom serial number` and the `Residue sequence number` are not consecutive. So, those numerations are fixed manually by checking and changing each row of the template PDB file.

S1.2. There is not a standard procedure to identify and select **rotamers** of a residue. Therefore, for each template structure, the identification of rotamers is made by searching manually residues with non standard four letters (instead of standard three-letter abbreviation) `Residue name`. For instance, in the case of the 1XIO[S2] template, there are two different rotamers for the Lys 210 with `Residue name` ALys and BLys, and two different rotamers for the Retinal with `Residue name` ARET and BRET. This procedure is time demanding and error prone, since for each new model the user must do manually this search considering all 20 standard amino acids and, additionally, the retinal protonated Shiff base (rPSB) chromophore.

S1.3. The selection of the rPSB is done manually, assuming the Retinal as default. Therefore, the information of the RET residue is kept, whereas the information of all other non standard residues (i.e. ions and membrane lipids) is removed by manually deleting the corresponding lines in the file.

S1.4. The selection of the Lys residue covalenty linked to the rPSB, the primary (MC) and

secondary (SC) counter-ions is made by visual inspection.

Generation of the cavity file (⧗ ≈ 15 min)

S1.5. The generation of the `cavity` file is performed by using the CASTp[S3] web-server. Thus, the general procedure for acquiring the ARM cavity file is: i) Upload the PDB file on the web-server page; ii) Submit the job and wait some time for the calculation; iii) Visualize the different pockets and select the one which contains both the rPSB and its linked Lys residue; iv) Write (manually) a list with the residues forming the cavity; v) Generate (manually) the cavity file. This procedure is time demanding and error prone due to the writing of the residues list.

Assignment of ionization states (⧗ ≈ 45 min)

S1.6. The ionization state of the residue side-chains is assigned through the calculation of their pK_a valuess and burying percentage, by manually launching the `PROPKA3.1`[S4] software on the current PDB file. The results contained in the PROPKA output file are then analyzed by doing (manually) a series of operations. These involve manually checking if the difference between the calculated (pK_a^{Calc}) and model (pK_a^{Model}) pK_a values of each amino acid is higher than a shift value (usually 1.5 or 2.0), and if the burying percentage is higher than 55%. However, notice that there is no general consensus for the shift and burying values. Additionally, once the side-chains whose ionization state has to be changed are identified, such change (Asp → ASH, Glu → GLH, His → {HID, HIE, HIP}, Lys→ LYD) is done by manually modifying the `Residue name` in the current PDB file.

Counter-ion placement (⧗ ≈ 180 min)

S1.7. The addition of both Cl^- and Na^+ external counter-ions is necessary to obtain a globally neutral model. The corresponding procedure (Figure 3 in the main text) is handled by a series of manual steps which are time demanding and ambiguous (*i.e.*, subject to the personal user choice):

S1.7.i. Open the working PDB file in a molecular visualization software and identify the rPSB, the covalently linked Lys residue, the MC and the SC.

S1.7.ii. Define -visually- the inner (IS) and outer (OS) surfaces of the protein, based on the positions of both the chromophore and its covalently linked Lys residue. It is difficult (and not rigorous) to establish, visually, a correct criteria to define the border between the IS and the OS.

S1.7.iii. Identify (visually) the positively (Arg, Lys, His) and negatively (Asp, Glu) charged amino acids in the IS and OS of the protein. For instance, if a residue side-chain is above the Lys-QM subsystem it is counted as a IS side-chain, otherwise it is counted as a OS side-chain. Write (manually) a list with the number of positively and negatively charged residues in each protein surface.

S1.7.iv. Sum (manually) the number of IS and OS positive and negative side-chains of the list to calculate the charge in each region, in order to determine the number of Cl^- and Na^+ external counter-ions to be added.

S1.7.v. Identify (visually) the target residue side-chains near to which place the counter-ions. Place manually the Cl^- and Na^+ ions (using a visualization software), save their coordinates in a new PDB file. Open the working PDB and the new PDB file in a text editor and add manually the coordinates of the counter-ion to the working PDB file using the correct format.

S1.7.vi. Open the modified working PDB file to verify the correct placement of the inserted counter-ions.

S2 Workflow of the *a*-ARM protocol

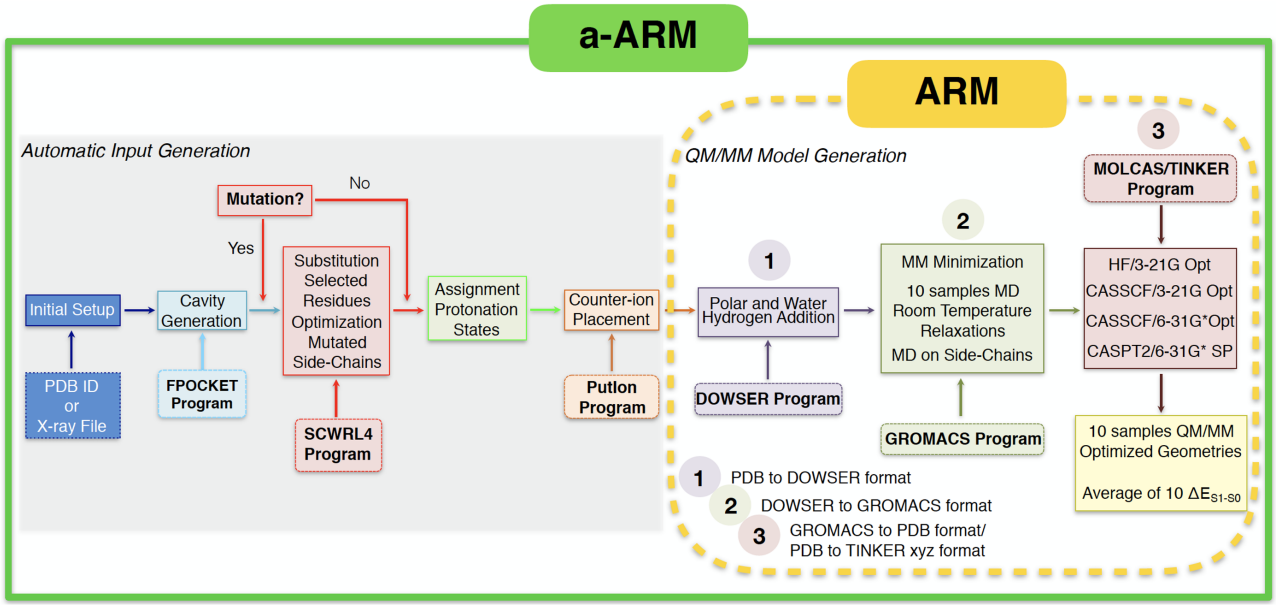

Figure S1: Automatic Rhodopsin Modeling workflow. The new version (*a*-ARM) incorporates the original version of the ARM protocol further described in ref. S1. In ARM, N independent simulated annealing (SA) and molecular dynamics (MD) room-temperature relaxations are performed at the MM level using GROMACS on the cavity and the Lys-QM systems.

S3 Detailed workflow of each step for the a-ARM automatic input generation

S3.1 Step 1. Automatic identification of the protein chain, rPSB, chromophore bounded Lys, MC and SC

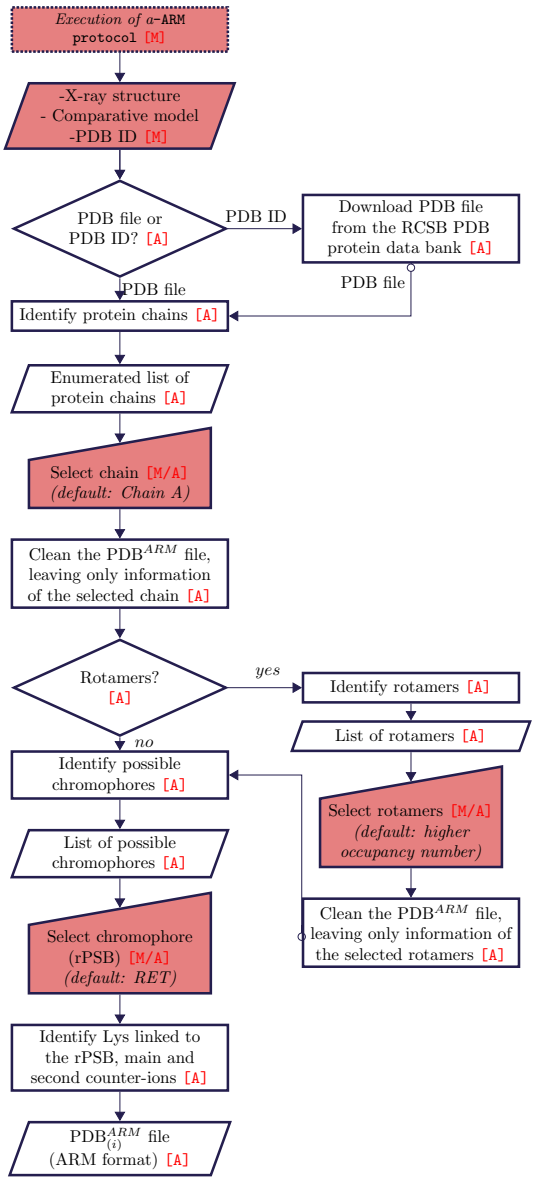

Figure S2: Step 1. Initial preparation of the input file. The red filled boxes represent tasks in which the user may interact with the program.

S3.2 Step 2: Automatic generation of the chromophore cavity

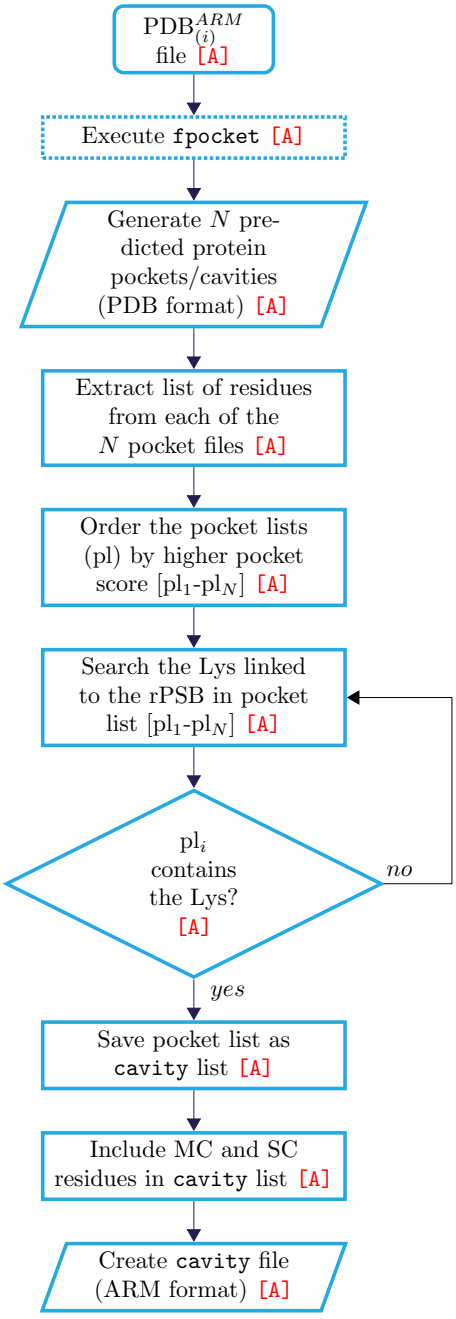

Figure S3: Step 2. Automatic generation of chromophore cavity. The code does not require the user's interaction during its execution.

S3.3 Step 3: Automatic assignment of ionization states

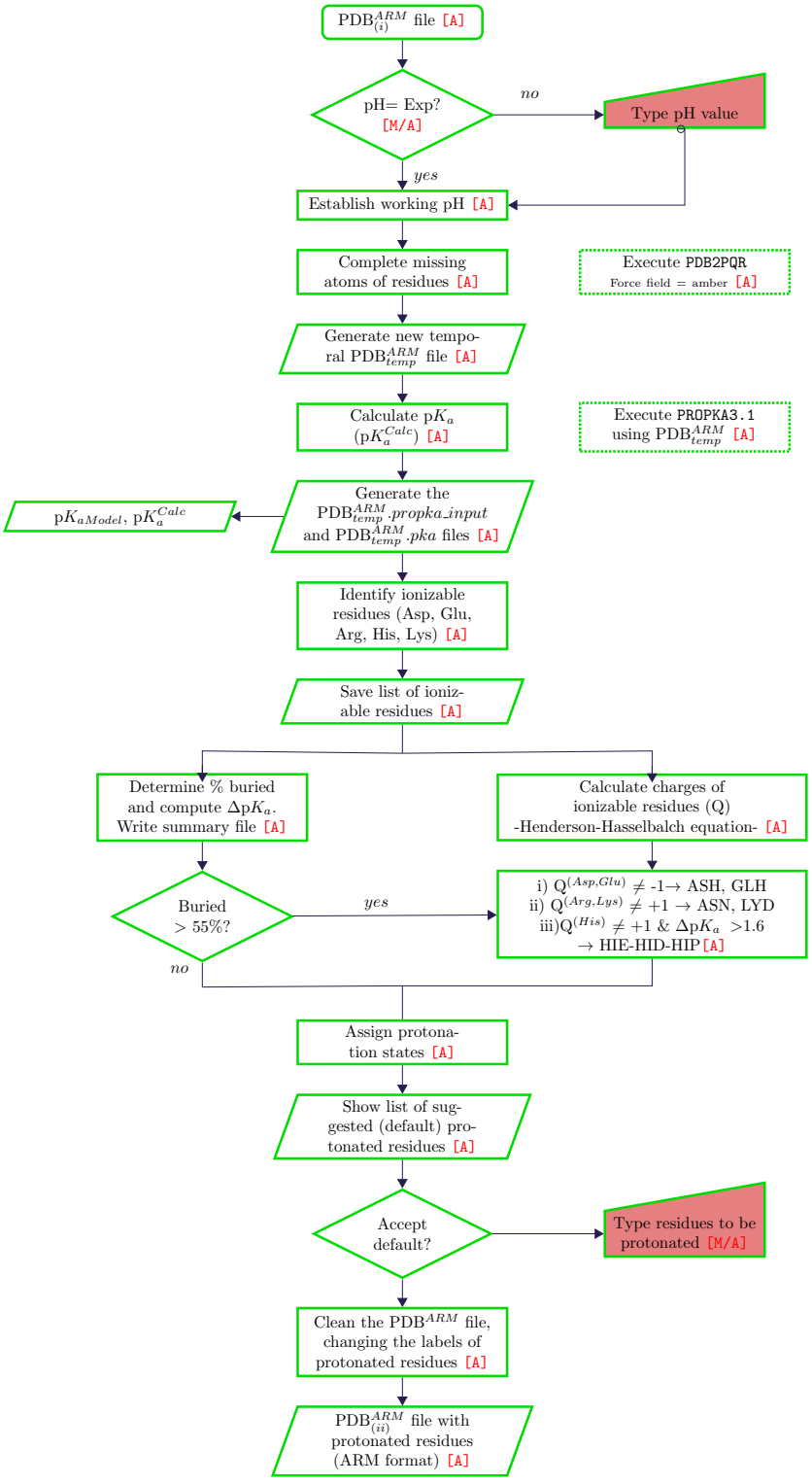

Figure S4: Step 3. Automatic assignment of ionization states. The red filled boxes represent tasks in which the user may interact with the program.

S3.4 Step 4: Automatic counter-ion placement

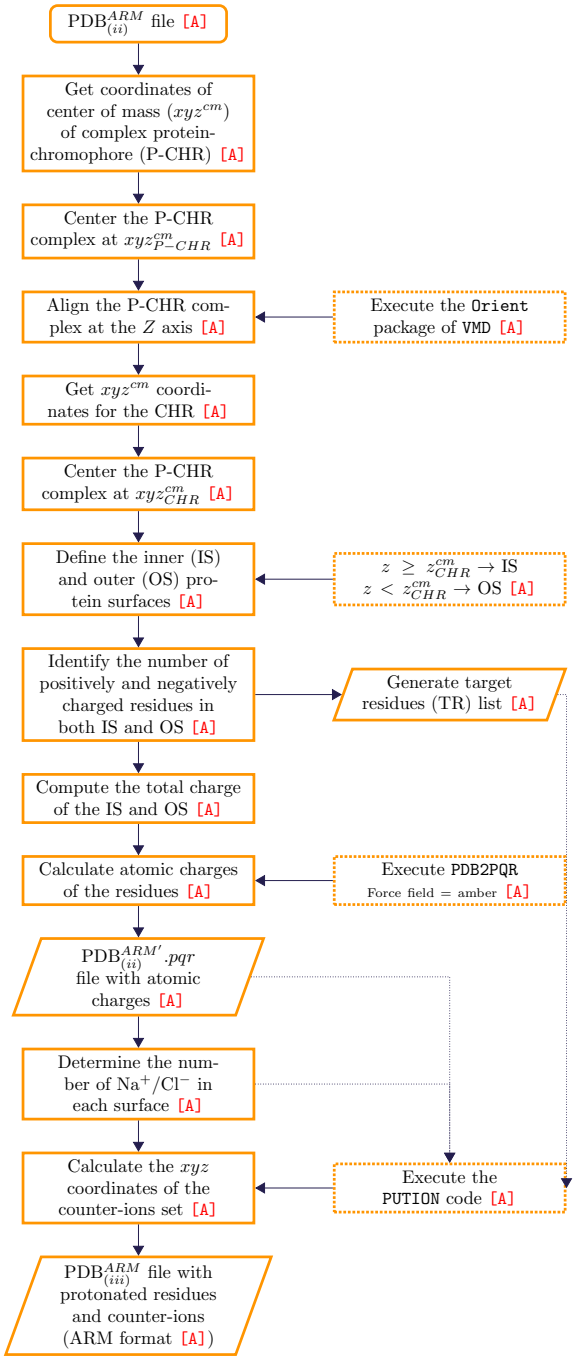

Figure S5: Step 4. Automatic selection of counter-ion placement. The code does not require the user's interaction during its execution.

S4 Automatic format conversion from LYR to RET plus LYS

Table S1: Correspondence between atom names in RET and LYR residues. The atom name of LYR and Lys covalently linked to the rPSB are consistent.

RET	LYR
C1	C17
C2	C16
C3	C15
C4	C14
C5	C12
C6	C11
C7	C10
C8	C9
C9	C80
C10	C7
C11	C6
C12	C5
C13	C3
C14	C2
C15	C1
C16	C18
C17	C19
C18	C13
C19	C8
C20	C4

S5 Features of the *a*-ARM QM/MM Models

Table S2: Main features of the *a*-ARM QM/MM models for the rhodopsins in the benchmark set.

Protein code	PBD ID (Chain)	Res.[a] (Å)	Retinal conf.[b]	Linker Lysine	MC[c]	SC	*a*-ARM QM/MM models					
							Rotamers	pH	Protonated residues	Counter-ions OS[d]	Counter-ions IS[d]	Cavity residues[e]
***a*-ARM$_{default}$**												
ASR$_{AT}$-1	1XIO[S2] (A)	2.0	all-*trans*	210	Asp-75	—	**B**RET **A**Lys-210(0.50)	5.6	ASH(198,217), GLH(36), HID(8,69)	2 Na^+	7 Cl^-	73,**75**,76,79,80,83,109,112,113,116,119,132,135, 136,139,176,179,180,183,**198**,202,209,**210**
ASR$_{AT}$-2	1XIO[S2] (A)	2.0	all-*trans*	210	Asp-75	—	**B**RET **B**Lys-210(0.50)	5.6	ASH(198,217), GLH(36), HID(8,69)	2 Na^+	7 Cl^-	**75**,76,79,80,83,109,112,113,132,135,136,139, 176,179,180,183,**198**,202,209,**210**
ASR$_{13C}$-1	1XIO[S2] (A)	2.0	13-*cis*	210	Asp-75	—	**A**RET **A**Lys-210(0.50)	5.6	ASH(198,217), GLH(36), HID(8,69)	2 Na^+	7 Cl^-	**73**,75,76,79,80,83,109,112,113,116,132,135,136, 139,176,179,180,183,**198**,202,209,**210**
ASR$_{13C}$-2	1XIO[S2] (A)	2.0	13-*cis*	210	Asp-75	—	**A**RET **B**Lys-210(0.50)	5.6	ASH(198,217), GLH(36), HID(8,69)	2 Na^+	7 Cl^-	73,**75**,76,79,80,83,109,112,113,116,132,135,136, 139,176,179,180,183,**198**,202,209,**210**
bR$_{AT}$	6G7H[S5] (A)	1.5	all-*trans*	216	Asp-212	Asp-85	**A**Asp-104(0.80) **A**Leu-15(0.57) **A**Leu-109(0.54)	5.6	ASH(85,96,115) GLH(194)	1 Cl^-	1 Cl^-	83,**85**,86,90,93,118,119,121,122,125,137,138,141, 142,145,182,185,186,189,**212**,215,**216**
bR$_{13C}$	1X0S[S6] (A)	2.5	13-*cis*	216	Asp-212	Asp-85	—	5.2	ASH(85,96,115) GLH(194)	1 Cl^-	3 Cl^-	83,**85**,86,90,93,118,119,121,122,125,134,137,138, 141,142,145,182,185,186,189,208,**212**,215,**216**
bathoRh	2G87[S7] (A)	2.6	all-*trans*	296	Glu-113	Glu-181	—	2.6	ASH(83) HID(211) GLH(122,181,249)	2 Na^+	7 Cl^-	86,91,**113**,114,117,118,121,122,125,167,**181**,186, 187,188,189,191,207,208,211,212,216,261,265,268, 269,272,292,295,**296**,298
BPR	4JQ6[S8] (A)	2.3	all-*trans*	213	Asp-79	Asp-209	—	4.5	GLH(90,124)	4 Na^+	3 Cl^-	77,**79**,80,83,84,87,113,116,117,120,133,134,137, 138,141,180,183,184,187,205,**209**,212,**213**
ChR$_{C1C2}$	3UG9[S9] (A)	2.3	all-*trans*	296	Asp-292	Glu-162		6.0	ASH(195), GLH(121,122, 129,136,274), HID(173,288)	3Cl^-	6 Cl^-	**162**,163,166,167,170,195,198,199,202,216,217, 220, 221,224,225,227,228,229,262,265,266,269, **292**,295,**296**
Rh	1U19[S10] (A)	2.2	11-*cis*	296	Glu-113	Glu-181	—	6.0	ASH(83) HID(211) GLH(122,181)	2 Na^+	6 Cl^-	86,91,**113**,114,117,118,121,122,125,**181**,186,187, 188,189,191,207,208,211,212,216,261,262,265,266, 268,269,272,292,295,**296**,298
SqRh	2Z73[S11] (A)	2.5	11-*cis*	305	Glu-180	—	—	6.6	ASH(80), HID(319)	1 Cl^-	7 Cl^-	80,83,87,111,112,115,116,119,120,123,177,**180**, 186,187,188,204,205,208,209,270,274,277,278, 281,301,304,**305**
hMeOp	T: 2Z73[S11] (A)	2.5	11-*cis*	274	Glu-149	—	—	6.6	ASH(50), HID(288,279)	1 Cl^-	15 Cl^-	53,57,78,81,84,85,88,89,92,95,136,146,**149**,155, 156,157,169,173,174,177,178,181,182,185,236,239, 240,243,244,246,247,270,273,**274**
ASR$_{AT}$-D217E	4TL3[S12] (A)	2.3	all-*trans*	210	75	–	–	5.6	ASH(198,217), GLH(36), HID(8)	–	7 Cl^-	73,**75**,76,79,80,83,109,112,113,116,119,131,132,135, 136,139,176,179,180,183,**198**,202,209,**210**
Arch1	1UAZ[S13] (A)	3.4	all-*trans*	222	Asp-91	Asp-218	—	5.2	ASH(91,102,121), GLH(210)	3 Na^+	4 Cl^-	55,89,**91**,92,95,96,99,102,121,124,125,127,128, 131,143,144,147,148,151,183,184,187,188,191, 192,195,214,**218**,221,**222**,225
Arch2	3WQJ[S14] (A)	1.8	all-*trans*	221	Asp-90	Asp-217	–	7.0	ASH (90,101,120), GLH(209)	3 Na^+	5 Cl^-	88,**90**,91,95,98,123,124,127,146,147,150,187, 190,191,194,213,**217**,220,**221**
ChR2	6EID[S15] (A)	2.4	all-*trans*	257	Glu-123	Asp-253		– def=7.4	GLH(90,101), HID(134,249)	2 Cl^-	9 Cl^-	121,**123**,124,128,131,156,159,160,163,164,177, 178,181,182,185,186,188,189,223,224,226,227, 230,**253**,256,**257**
ChR2-C128T	6EIG[S15] (A)	2.7	all-*trans*	257	Glu-123	Asp-253		– def=7.4	GLH(90,101), HID(134,249)	0 Cl^-	7 Cl^-	121,**123**,124,128,131,156,159,160,163,164,177, 178,181,182,185,186,188,189,223,224,226,227, 230,**253**,256,**257**
KR2-1	3X3C[S16] (A)	2.3	all-*trans*	255	Asp-116	Asp-251	**A**Asp-116(0.65) **A**Gln-157(0.5)	8.0	—	8 Na^+	8 Cl^-	110,113,**116**,117,120,149,150,153,171,174,175, 178,215,218,219,222,**251**,254,**255**
KR2-2	3X3C[S16] (A)	2.3	all-*trans*	255	Asp-116	Asp-251	**A**Asp-116(0.65) **B**Gln-157(0.5)	8.0	—	8 Na^+	8 Cl^-	110,113,**116**,117,120,149,150,153,167,168,171, 172,174,175,178,215,218,219,222,223,247,**251**,

[a] X-Ray diffraction resolution. [b] Retinal conformation. [c] Main (MC) and second (SC) counter-ions. [d] Extracellular (OS) and Intracellular (IS) protein surfaces. [e] In bold Lys, MC and SC.

Continued on next page

Table S2 — continued from previous page

Protein code	PBD ID (Chain)	Res.[a] (Å)	Retinal conf.[b]	Linker Lysine	MC[c]	SC	*a*-ARM QM/MM models					
							Rotamers	pH	Protonated residues	Counter-ions OS[d]	Counter-ions IS[d]	Cavity residues[e]
												254,**255**
NM-R3	5B2N[S17] (A)	1.6	all-*trans*	235	Cl^-	Asp-231	—	8.0	–	7 Na^+	7 Cl^-	2,3,5,6,29,75,76,85,87,88,89,90,91,92,93,94, 96,97,99,103,106,135,136,139,140,143,144, 146,153,154,157,158,160,161,164,201,204, 205,208,209,223,227,**231**,234,**235**
ClR	5G28[S18] (A)	1.6	all-*trans*	235	Cl^-	Asp-231	—	6.0	—	6 Na^+	7 Cl^-	3,5,91,92,95,96,99,103,106,135,136,139,143, 146,153,154,157,158,160,161,164,201,204,205, 207,208,209,223,227,**231**,234,**235**,**403**
SR-II	1JGJ[S19] (A)	2.4	all-*trans*	205	Asp-201	Asp-75	–	5.3	ASH(75,193)	2 Cl^-	1 Cl^-	73,**75**,76,77,80,83,108,109,112,126,127,130, 131,134,171,174,175,178,197,**201**,204,205
SqbathoRh	3AYM[S20]	2.8	all-*trans*	305	180	–	–	6.6	ASH(80), HID(319)	1 Cl^-	7 Cl^-	80,83,87,111,112,115,116,119,120,177,**180**, 186,187,188,204,205,209,270,274,277, 278,301,304,**305**
AARh	T:1U19[S10] (A)	2.2	11-*cis*	296	113	181	–	6.0	ASH(83), HID(211) GLH(122,181)	2 Na^+	7 Cl^-	86,**113**,114,117,118,121,122,125,167,**181**, 186,187,188,189,191,207,208,211,212,261, 265,268,269,272,292,295,**296**,298
BCone	T:1U19[S10] (A)	2.2	11-*cis*	293	Glu-110	Glu-178	–	6.0	GLH(178)	10 Cl^-	5 Cl^-	52,75,76,79,80,82,83,**110**,111,113,114,115,116, 117,118,119,120,121,122,123,125,158,**178**,184, 185,186,188,204,205,208,209,212,213,216,217, 255,258,259,262,263,265,266,267,269,270,289, 292,**293**,295,296,299,300
GCone	T:1U19[S10] (A)	2.2	11-*cis*	312	Glu-129	Glu-102	–	6.0	ASH(99), GLH(102,150), HID(197), LYD(200)	5 Cl^-	11 Cl^-	**129**,130,133,134,137,138,141,184,192,194,197, 202,203,204,205,206,207,216,219,220,223,224, 227,228,277,278,281,282,284,285,288,289,292, 308,311,**312**
RCone	T:1U19[S10] (A)	2.2	11-*cis*	293	Glu-110	Glu-83	–	6.0	ASH(80), GLH(83), HID(178)	4 Cl^-	11 Cl^-	41,44,48,**83**,84,87,88,**110**,111,114,115,117,118, 119,122,164,178,183,184,185,186,188,189,201, 204,205,208,209,210,213,258,259,262,263,265, 266,268,269,270,271,272,273,275,280,284,285, 288,289,292,**293**,295,296
mMeOp	T: 2Z73[S11] (A)	2.5	11-*cis*	283	Glu-160	–	–	6.6	ASH (61), HID(297)	1 Cl^-	17 Cl^-	22,25,26,61,64,65,68,69,91,92,95,96,99,100, 103,157,**160**,166,167,168,180,184,185,188, 189,248,252,253,255,256,279,282,**283**,286
PoXeR$_{AT}$	T:1XIO[S2] (A)	2.0	all-*trans*	209	Asp-74	Asp-108	–	5.6	ASH(108,216), GLH(35)	2 Na^+	6 Cl^-	72,**74**,75,79,82,**108**,111,112,115,118,130, 131,134,135,138,175,178,179,182,201,208,**209**
PoXeR$_{13C}$	T:1XIO[S2] (A)	2.0	all-*trans*	209	Asp-74	Asp-108	–	5.6	ASH(108,216), GLH(35)	2 Na^+	6 Cl^-	72,**74**,75,78,79,82,**108**,111,112,115,118, 130,131,134,135,138,175,178,179,182,201,208,**209**
***a*-ARM$_{customized}$**												
KR2-2$^{(c)}$	3X3C[S16] (A)	2.3	all-*trans*	255	Asp-116	Asp-251	**A**Asp-116(0.65) **B**Gln-157(0.5)	8.0	ASH(251)	7 Na^+	8 Cl^-	110,113,**116**,117,120,149,150,153,167,168,171, 172,174,175,178,215,218,219,222,223,247,**251**, 254,**255**
BPR$^{(c)}$	4JQ6[S8] (A)	2.3	all-*trans*	213	79	209	–	4.5	GLH(90)	4 Na^+	1 Cl^-	77,**79**,80,83,84,87,116,117,120,133,134,137,138, 141,180,183,184,187,205,**209**,212,**213**
RCone$^{(c)}$	T:1U19[S10] (A)	2.2	11-*cis*	293	Glu-110	Glu-83	–	6	ASH(80), GLH(110), HID(178)	4 Cl^-	10 Cl^-	41,44,48,**83**,84,87,88,**110**,111,114,115,117,118, 119,122,164,178,183,184,185,186,188,189, 201,204,205,208,209,210,213,258,259,262, 263,265,266,268,269,270,271,272,273,275, 280,284,285,288,289,292,**293**,295,296
bR$^{(c)}_{AT}$	6G7H[S5] (A)	1.5	all-*trans*	216	85	212	AAsp-104(0.80) ALeu-15(0.57) ALeu-109(0.54)	5.6	ASH(212,96,115) GLH(194)	1 Cl^-	1 Cl^-	83,**85**,86,90,93,118,119,121,122,125,137,138, 141,142,145,182,185,186,189,**212**,215,**216**
bR$^{(c-2)}_{AT}$	6G7H[S5] (A)	1.5	all-*trans*	216	85	212	AAsp-104(0.80) ALeu-15(0.57) ALeu-109(0.54)	5.6	ASH(212,96,115) GLH(194)	1 Cl^-	1 Cl^-	83,**85**,86,90,93,96,115,118,119,121,122,125, 137,138,141,142,145,182,185,186,189,**212**, 215,**216**
ChR$^{c}_{C1C2}$	3UG9[S9] (A)	2.3	all-*trans*	296	292	162	–	5.2	ASH(195), GLH(121,122, 220,221,129,136,140,162,274), HID(173, 288)	5Cl^-	6 Cl^- 269,**292**,295,**296**	**162**,163,166,167,170,195,198,199,202,216,217, 224,225,227,228,229,262,265,266,
ChR2$^{(c)}$	6EID[S15] (A)	2.4	all-*trans*	257	Glu-123	Asp-253	–	5.2 def=7.4	ASH(156,253),GLH(83,90, 101),HID(134,249)	5 Cl^-	10 Cl^-	90,121,**123**,124,128,131,156,159,160,163,164,177, 178,181,182,185,186,188,189,223,224,226,227,

[a] X-Ray diffraction resolution. [b] Retinal conformation. [c] Main (MC) and second (SC) counter-ions. [d] Extracellular (OS) and Intracellular (IS) protein surfaces. [e] In bold Lys, MC and SC.

Continued on next page

Table S2 — continued from previous page

Protein code	PBD ID (Chain)	Res.[a] (Å)	Retinal conf.[b]	Linker Lysine	MC[c]	SC	*a*-ARM QM/MM models					
							Rotamers	pH	Protonated residues	Counter-ions OS[d]	Counter-ions IS[d]	Cavity residues[e]
												230,**253**,256,**257**
ChR2-C128T[(c)]	6EIG[S15] (A)	2.7	all-*trans*	257	Glu-123	Asp-253	–	5.2 def=7.4	ASH(156,253), GLH(82,90, 235), 97,101, HID(134,249)	4 Cl^-	8 Cl^-	121,**123**,124,128,131,156,159,160,163,164,177, 178,181,182,185,186,188,189,223,224,226,227, 230,**253**,256,**257**
***a*-ARM$_{\text{customized}}$ (Effect of chain and pH)**												
ASR$_{AT}$-1 (c-pH)	1XIO[S2] (A)	2.0	all-*trans*	210	75	–	BRET ALys-210	7.4	GLH(36), HID(8,69)	3 Na^+	6 Cl^-	73,**75**,76,79,80,83,109,112,113,116,119,132, 135,136,139,176,179,180,183,198,202,209,**210**
ASR$_{13C}$-2 (c-pH)	1XIO[S2] (A)	2.0	all-*trans*	210	75	–	ARET BLys-210	7.4	GLH(36), HID(8,69)	3 Na^+	6 Cl^-	73,**75**,76,79,80,83,109,112,113,116,119,131,132, 135,136,139,176,179,180,183,198,202,209,**210**
bR$^{\text{c-pH}}_{13C}$	1X0S[S6] (A)	2.5	13-*cis*	216	85	212	–	7.4	ASH(96) GLH(194)	1 Na^+	3 Cl^-	83,**85**,86,90,93,118,119,121,122,125,134,137,138, 141,142,145,182,185,186,189,208,**212**,215,**216**
bR$^{\text{c-pH}}_{AT}$	6G7H[S5] (A)	1.5	all-*trans*	216	85	212	AAsp-104(0.80) ALeu-15(0.57) ALeu-109(0.54)	7.4	ASH(96) GLH(194)	1 Cl^-	1 Na^+	83,**85**,86,90,93,118,119,121,122,125,137,138,141, 142,145,182,185,186,189,**212**,215,**216**
bathoRh (c-pH)	2G87[S7] (A)	7.4	all-*trans*	296	113	181	–	6.0	ASH(83) HID(211) GLH(122,181)	2 Na^+	6 Cl^-	86,91,**113**,114,117,118,121,122,125,167,**181**,186, 187,188,189,191,207,208,211,212,216,261,265,268, 269,272,292,295,**296**,298
bathoRh[(c)]	2G87[S7] (B)	2.6	all-*trans*	296	113	181	–	6.0	ASH(83) HID(211) GLH(122,181)	2 Na^+	6 Cl^-	86,91,**113**,114,117,118,121,122,125,167,**181**,186, 187,188,189,191,207,208,211,212,216,261,265,268, 269,272,292,295,**296**,298
BPR[(c-2)]	4JQ6[S8] (B)	2.3	all-*trans*	213	79	209	–	4.5	GLH(90)	4 Na^+	3 Cl^-	77,**79**,80,83,84,87,113,116,117,120,133,134,137, 138,141,180,183,184,187,205,**209**,212,**213**
BPR (c-pH-2)	4JQ6[S8] (B)	2.3	all-*trans*	213	79	209	–	4.5	GLH(90)	5 Na^+	3 Cl^-	77,**79**,80,83,84,87,113,116,117,120,133,134,137, 138,141,180,183,184,187,205,**209**,212,**213**
Rh[(c)]	1U19[S10] (B)	2.2	11-*cis*	296	113	181	–	6.0	ASH(83) HID(211) GLH(122,181)	2 Na^+	6 Cl^-	83,86,91,**113**,114,117,118,121,122,125,**181**,186, 187,188,189,191,207,208,211,212,216,261,265,268, 269,272,292,295,**296**,298
Rh[(c-pH-2)]	1U19[S10] (B)	2.2	11-*cis*	296	113	181	–	6.0	ASH(83) HID(211) GLH(122)	3 Na^+	6 Cl^-	83,86,91,**113**,114,117,118,121,122,125,**181**,186, 187,188,189,191,207,208,211,212,216,261,265, 268,269,272,292,295,**296**,298
SqRh[c]	2Z73[S11] (B)	2.5	11-*cis*	305	180	–	–	6.6	ASH(80), HID(319)	1 Cl^-	7 Cl^-	80,83,87,111,112,115,116,119,120,123,177,**180**, 186,187,188,204,205,209,270,274,277, 278,281,**305**

[a] X-Ray diffraction resolution. [b] Retinal conformation. [c] Main (MC) and second (SC) counter-ions. [d] Extracellular (OS) and Intracellular (IS) protein surfaces. [e] In bold Lys, MC and SC.

S6 Summary of the ARM QM/MM calculations

Table S3: Summary of the ten *a*-ARM QM/MM calculations performed for the rhodopsins in the benchmark set. Computed total ground state (S_0) and first excitation (S_1) energies, transition oscillator strength (f_{Osc}), first vertical excitation energy (ΔE_{S1-S0}) and maximum absorption wavelength (λ^a_{max}). Statistical parameters such as average ($\overline{N}$), difference between calculated and experimental data ($|\overline{N}|$) and standard deviation ($\sigma_{\overline{N}}$) are also provided.

Seed (N)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)
a-ARM$_{\mathbf{default}}$															
	ASR$_{AT}$-1					**ASR$_{AT}$-2**					**ASR$_{13C}$-1**				
Exp.				*52.0*	*550*				*52.0*	*550*				*53.2*	*537*
1	-871.944185	-871.858715	1.18	53.6	533	-871.856669	-871.769865	1.20	54.5	525	-871.971106	-871.883658	1.04	54.9	521
2	-871.933753	-871.852485	1.35	51.0	561	-871.844720	-871.759667	1.50	53.4	536	-871.967357	-871.878575	1.07	55.7	513
3	-871.933966	-871.852626	1.37	51.0	560	-871.845887	-871.761754	1.33	52.8	542	-871.971365	-871.883857	1.02	54.9	521
4	-871.944074	-871.858426	1.20	53.7	532	-871.844557	-871.751744	1.01	58.2	491	-871.970919	-871.883975	1.03	54.6	524
5	-871.935622	-871.854422	1.34	51.0	561	-871.849651	-871.767147	1.35	51.8	552	-871.967134	-871.878920	1.06	55.4	517
6	-871.933380	-871.852472	1.35	50.8	563	-871.845719	-871.761747	1.33	52.7	543	-871.973374	-871.887370	1.00	54.0	530
7	-871.944236	-871.858925	1.36	53.5	534	-871.863471	-871.776729	1.19	54.4	525	-871.971356	-871.884159	1.02	54.7	523
8	-871.944192	-871.858789	1.19	53.6	534	-871.854097	-871.772424	1.36	51.2	558	-871.967466	-871.879496	1.06	55.2	518
9	-871.933903	-871.852946	1.23	50.8	563	-871.845296	-871.759678	1.34	53.7	532	-871.967513	-871.879555	1.06	55.2	518
10	-871.926741	-871.841506	1.36	53.5	535	-871.847536	-871.756189	0.94	57.3	499	-871.966980	-871.878895	1.06	55.3	517
$\overline{N}$				52.3	548				54.0	530				55.0	520
$\lvert\overline{N}\rvert$				0.3	-2				2.0	-20				1.8	-17
$\sigma_{\overline{N}}$				1.4	15				2.2	21				0.5	5
	ASR$_{13C}$-2					**bR$_{AT}$**					**bR$_{13C}$**				
Exp.				*53.2*	*537*				*50.3*	*568*				*52.2*	*547*
1	-871.957306	-871.870880	1.08	54.2	527	-871.767543	-871.682740	1.25	53.2	537	-871.603497	-871.518338	0.92	53.4	535
2	-871.957423	-871.870771	1.08	54.4	526	-871.767442	-871.682983	1.24	53.0	539	-871.603873	-871.519224	0.94	53.1	538
3	-871.957467	-871.871043	1.08	54.2	527	-871.767725	-871.682885	1.25	53.2	537	-871.604253	-871.519692	0.96	53.1	539
4	-871.956444	-871.871834	1.08	53.1	539	-871.767713	-871.682729	1.25	53.3	536	-871.603550	-871.518503	0.93	53.4	536
5	-871.957499	-871.870947	1.08	54.3	526	-871.767371	-871.682922	1.24	53.0	540	-871.604049	-871.519412	0.95	53.1	538
6	-871.957305	-871.870681	1.09	54.4	526	-871.767501	-871.682631	1.25	53.3	537	-871.603728	-871.518714	0.93	53.3	536
7	-871.956807	-871.869975	1.09	54.5	525	-871.767560	-871.682855	1.24	53.2	538	-871.603783	-871.518817	0.94	53.3	536
8	-871.957236	-871.870434	1.08	54.5	525	-871.767651	-871.683245	1.25	53.0	540	-871.604010	-871.519093	0.94	53.3	537
9	-871.951108	-871.864478	1.03	54.4	526	-871.767701	-871.682689	1.25	53.3	536	-871.603528	-871.518515	0.92	53.3	536
10	-871.956932	-871.870643	1.08	54.1	528	-871.767647	-871.683130	1.25	53.0	539	-871.603293	-871.518417	0.94	53.3	537
$\overline{N}$				54.2	528				53.2	538				53.3	537
$\lvert\overline{N}\rvert$				1.0	-10				2.9	-30				1.1	-10
$\sigma_{\overline{N}}$				0.4	4				0.1	2				0.1	1

Continued on next page

Table S3 — continued from previous page

Seed (N)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)
	bathoRh					**BPR**					**Rh**				
Exp.				*54.0*	*529*				*58.3*	*490*				*57.4*	*498*
1	-871.796873	-871.707343	0.98	56.2	509	-872.092199	-871.990725	0.58	63.7	449	-871.891394	-871.799897	0.84	57.4	498
2	-871.796733	-871.706755	0.96	56.5	506	-872.092759	-871.991196	0.58	63.7	449	-871.891409	-871.799801	0.83	57.5	497
3	-871.797014	-871.707182	0.97	56.4	507	-872.092370	-871.990778	0.57	63.7	448	-871.898920	-871.807462	0.86	57.4	498
4	-871.796624	-871.707482	0.96	55.9	511	-872.091715	-871.990006	0.62	63.8	448	-871.899099	-871.806950	0.88	57.8	494
5	-871.796767	-871.706736	0.96	56.5	506	-872.093278	-871.991460	0.60	63.9	447	-871.892134	-871.800452	0.87	57.5	497
6	-871.797696	-871.709396	0.94	55.4	516	-872.091774	-871.990248	0.57	63.7	449	-871.891939	-871.799669	0.88	57.9	494
7	-871.796909	-871.707454	0.97	56.1	509	-872.092212	-871.990614	0.56	63.8	448	-871.904687	-871.810782	0.97	58.9	485
8	-871.796941	-871.707383	0.97	56.2	509	-872.102907	-872.000990	0.56	64.0	447	-871.900020	-871.809660	0.89	56.7	504
9	-871.833285	-871.743574	0.95	56.3	508	-872.092288	-871.990441	0.57	63.9	447	-871.898517	-871.807389	0.85	57.2	500
10	-871.803384	-871.713400	0.96	56.5	506	-872.089566	-871.988996	0.47	63.1	453	-871.913411	-871.819366	0.84	59.0	484
$\overline{N}$				56.2	509				63.7	449				57.7	495
$\lvert\overline{N}\rvert$				2.2	-20				5.4	-42				0.3	-3
$\sigma_{\overline{N}}$				0.3	3				0.2	2				0.7	6
	SqRh					**hMeOp**					**Arch1**				
Exp.				*58.5*	*489*				*60.4*	*473*				*50.3*	*568*
1	-871.781252	-871.684407	0.80	60.8	470	-871.853694	-871.759998	0.82	58.8	486	-872.150995	-872.071861	1.21	49.7	576
2	-871.780822	-871.684020	0.80	60.7	471	-871.869263	-871.770231	0.82	62.1	460	-872.134891	-872.056149	1.24	49.4	579
3	-871.781270	-871.684471	0.79	60.7	471	-871.868814	-871.769820	0.82	62.1	460	-872.136583	-872.053532	1.17	52.1	549
4	-871.781391	-871.684430	0.81	60.8	470	-871.868208	-871.773950	0.79	59.1	483	-872.138577	-872.059637	1.22	49.5	577
5	-871.781655	-871.684338	0.80	61.1	468	-871.863313	-871.764403	0.76	62.1	461	-872.146883	-872.062362	1.12	53.0	539
6	-871.781443	-871.684495	0.81	60.8	470	-871.874067	-871.773095	0.80	63.4	451	-872.134260	-872.054198	1.22	50.2	569
7	-871.794665	-871.697108	0.79	61.2	467	-871.874090	-871.779508	0.78	59.4	482	-872.134575	-872.053348	1.25	51.0	561
8	-871.780941	-871.684046	0.79	60.8	470	-871.869458	-871.770278	0.81	62.2	459	-872.140807	-872.061822	1.24	49.6	577
9	-871.784194	-871.687478	0.79	60.7	471	-871.862633	-871.764304	0.79	61.7	463	-872.138292	-872.058605	1.22	50.0	572
10	-871.780898	-871.683688	0.80	61.0	469	-871.862633	-871.764304	0.79	61.7	463	-872.137357	-872.057714	1.37	50.0	572
$\overline{N}$				60.9	470				61.3	467				50.5	567
$\lvert\overline{N}\rvert$				2.4	-19				0.9	-6				0.1	-1
$\sigma_{\overline{N}}$				0.2	1				1.6	12				1.2	13
	Arch2					**KR2-1**					**KR2-2**				
Exp.				*52.0*	*550*				*54.5*	*525*				*54.5*	*525*
1	-871.742904	-871.656377	1.22	54.3	527	-871.718190	-871.605973	1.41	70.4	406	-871.881846	-871.771102	1.40	69.5	411
2	-871.761918	-871.673860	1.20	55.3	517	-871.718749	-871.608831	1.48	69.0	415	-871.881184	-871.771286	1.44	69.0	415
3	-871.762079	-871.675040	1.23	54.6	523	-871.718610	-871.608609	1.47	69.0	414	-871.855753	-871.742680	1.39	71.0	403
4	-871.762469	-871.675202	1.23	54.8	522	-871.744476	-871.633994	1.46	69.3	412	-871.855530	-871.742981	1.40	70.6	405
5	-871.761993	-871.674503	1.22	54.9	521	-871.718288	-871.605364	1.45	70.9	403	-871.882383	-871.771524	1.40	69.6	411
6	-871.761602	-871.675097	1.23	54.3	527	-871.722030	-871.612339	1.50	68.8	415	-871.880793	-871.770146	1.42	69.4	412
7	-871.762087	-871.674610	1.23	54.9	521	-871.720750	-871.610400	1.47	69.2	413	-871.882213	-871.771985	1.42	69.2	413
8	-871.760968	-871.675668	1.28	53.5	534	-871.718202	-871.605145	1.43	70.9	403	-871.882437	-871.772269	1.42	69.1	414
9	-871.761597	-871.674473	1.22	54.7	523	-871.718346	-871.605827	1.45	70.6	405	-871.882394	-871.772118	1.41	69.2	413
10	-871.761788	-871.676601	1.28	53.5	535	-871.718550	-871.605579	1.44	70.9	403	-871.884082	-871.773926	1.42	69.1	414
$\overline{N}$				54.5	525				69.9	409				69.6	411
$\lvert\overline{N}\rvert$				2.5	-25				15.4	-116				15.1	-114
$\sigma_{\overline{N}}$				0.6	6				0.9	5				0.7	4

Continued on next page

Table S3 — continued from previous page

Seed (N)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)
	NM-R3					**ClR**					**SR-II**				
Exp.				*55.3*	*517*				*55.3*	*517*				*57.5*	*497*
1	-876.392982	-876.303493	1.01	56.2	509	-873.188913	-873.100453	1.03	55.5	515	-871.779169	-871.690861	1.34	55.4	516
2	-876.372855	-876.283400	1.01	56.1	509	-873.182300	-873.094205	1.04	55.3	517	-871.759098	-871.666148	1.08	58.3	490
3	-876.372546	-876.283065	1.03	56.1	509	-873.182851	-873.095165	1.05	55.0	520	-871.758769	-871.665717	1.11	58.4	490
4	-876.372837	-876.283238	1.01	56.2	509	-873.182582	-873.094987	1.06	55.0	520	-871.763071	-871.670262	1.05	58.2	491
5	-876.408427	-876.318916	1.01	56.2	509	-873.182400	-873.094480	1.06	55.2	518	-871.758942	-871.666012	1.11	58.3	490
6	-876.370355	-876.280882	1.02	56.1	509	-873.182740	-873.095376	1.06	54.8	522	-871.763904	-871.671291	1.07	58.1	492
7	-876.372139	-876.282507	1.01	56.2	508	-873.212629	-873.124515	1.04	55.3	517	-871.763796	-871.671999	1.13	57.6	496
8	-876.372005	-876.282495	1.03	56.2	509	-873.181733	-873.093633	1.05	55.3	517	-871.758395	-871.665243	1.06	58.5	489
9	-876.373257	-876.283571	1.00	56.3	508	-873.182667	-873.095100	1.05	54.9	520	-871.759259	-871.666292	1.12	58.3	490
10	-876.374500	-876.284775	1.01	56.3	508	-873.193237	-873.105228	1.04	55.2	518	-871.759809	-871.666608	1.06	58.5	489
$\overline{N}$				56.2	509				55.2	518				58.0	493
$\lvert\overline{N}\rvert$				0.9	-8				-0.1	1				0.5	-4
$\sigma_{\overline{N}}$				0.1	0				0.2	2				0.9	8
	ChR$_{C1C2}$					**ChR2**					**ChR2-C128T**				
Exp.				62.4	458				60.8	470				59.0	485
1	-871.870140	-871.745569	0.09	78.2	366	-871.815311	-871.691014	1.27	78.0	367	-871.888276	-871.760576	0.06	80.1	357
2	-871.870150	-871.744864	0.08	78.6	364	-871.836024	-871.703610	0.12	83.1	344	-871.880425	-871.753332	0.05	79.8	359
3	-871.851757	-871.727240	0.09	78.1	366	-871.812075	-871.689091	1.28	77.2	370	-871.880270	-871.754419	0.06	79.0	362
4	-871.869494	-871.750353	0.14	74.8	382	-871.817542	-871.685661	0.19	82.8	345	-871.877818	-871.751407	0.06	79.3	360
5	-871.870223	-871.746189	0.10	77.8	367	-871.808973	-871.680495	0.30	80.6	355	-871.880116	-871.753757	0.06	79.3	361
6	-871.869556	-871.750803	0.15	74.5	384	-871.811952	-871.687923	1.30	77.8	367	-871.879732	-871.752082	0.04	80.1	357
7	-871.870109	-871.744002	0.06	79.1	361	-871.816012	-871.684373	0.17	82.6	346	-871.881191	-871.754768	0.05	79.3	360
8	-871.867595	-871.753277	0.25	71.7	399	-871.811320	-871.686689	1.26	78.2	366	-871.878693	-871.754416	0.12	78.0	367
9	-871.878344	-871.753321	0.08	78.5	364	-871.811518	-871.686410	0.76	78.5	364	-871.879972	-871.753295	0.05	79.5	360
10	-871.870245	-871.746581	0.10	77.6	368	-871.808973	-871.680495	0.30	80.6	355	-871.877364	-871.746319	0.04	82.2	348
$\overline{N}$				76.9	372				79.9	358				79.7	359
$\lvert\overline{N}\rvert$				14.5	-86				19.1	-112				20.7	-126
$\sigma_{\overline{N}}$				2.4	12				2.3	10				1.1	5
	SqbathoRh					**AARh**					**BCone**				
Exp.				*54.0*	*530*				*56.3*	*508*				*66.5*	*430*
1	-871.748231	-871.660335	1.11	55.2	518	-871.789121	-871.695758	0.76	58.6	488	-872.457790	-872.349377	0.64	68.0	420
2	-871.751087	-871.662346	1.08	55.7	513	-871.788471	-871.695800	0.80	58.2	492	-872.456892	-872.348952	0.65	67.7	422
3	-871.748516	-871.660100	1.09	55.5	515	-871.787995	-871.692800	0.78	59.7	479	-872.456799	-872.348877	0.65	67.7	422
4	-871.748338	-871.659719	1.09	55.6	514	-871.787008	-871.693466	0.75	58.7	487	-872.456879	-872.348875	0.65	67.8	422
5	-871.748410	-871.659807	1.09	55.6	514	-871.787422	-871.692558	0.78	59.5	480	-872.456762	-872.348799	0.65	67.7	422
6	-871.750155	-871.661742	1.07	55.5	515	-871.788294	-871.696173	0.77	57.8	495	-872.457026	-872.348829	0.64	67.9	421
7	-871.748545	-871.659990	1.08	55.6	515	-871.788960	-871.695320	0.80	58.8	487	-872.456634	-872.348732	0.65	67.7	422
8	-871.748344	-871.660163	1.06	55.3	517	-871.785832	-871.690178	0.95	60.0	476	-872.458406	-872.350480	0.55	67.7	422
9	-871.748545	-871.660065	1.09	55.5	515	-871.786932	-871.693274	0.74	58.8	486	-872.456810	-872.348924	0.65	67.7	422
10	-871.748304	-871.659689	1.09	55.6	514	-871.787764	-871.692417	0.78	59.8	478	-872.458263	-872.349787	0.64	68.1	420
$\overline{N}$				55.5	515				59.0	485				67.8	422
$\lvert\overline{N}\rvert$				1.6	-15				2.7	-23				1.3	-9
$\sigma_{\overline{N}}$				0.2	1				0.7	6				0.1	1

Continued on next page

Table S3 — continued from previous page

Seed (N)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)
	GCone					**RCone**					**mMeOp**				
Exp.				*53.4*	*535*				*49.7*	*575*				*61.2*	*467*
1	-871.823983	-871.738519	0.96	53.6	533	-871.878069	-871.784867	0.82	58.5	489	-871.885083	-871.786096	0.77	62.1	460
2	-871.820043	-871.730651	0.77	56.1	510	-871.879581	-871.782602	0.70	60.9	470	-871.865266	-871.765590	0.77	62.5	457
3	-871.823720	-871.735980	0.82	55.1	519	-871.869267	-871.782100	0.78	54.7	523	-871.873858	-871.774306	0.76	62.5	458
4	-871.849355	-871.761965	0.90	54.8	521	-871.885543	-871.792020	0.82	58.7	487	-871.864992	-871.765271	0.76	62.6	457
5	-871.834350	-871.743110	0.87	57.3	499	-871.874651	-871.781221	0.82	58.6	488	-871.863626	-871.764039	0.77	62.5	458
6	-871.830548	-871.745711	0.97	53.2	537	-871.874561	-871.781004	0.79	58.7	487	-871.879812	-871.780249	0.77	62.5	458
7	-871.823103	-871.733003	0.76	56.5	506	-871.877487	-871.782852	0.70	59.4	481	-871.874444	-871.774610	0.76	62.6	456
8	-871.825817	-871.740184	0.95	53.7	532	-871.874748	-871.781295	0.82	58.6	488	-871.863421	-871.763455	0.76	62.7	456
9	-871.826133	-871.738826	1.07	54.8	522	-871.879733	-871.785604	0.76	59.1	484	-871.863421	-871.763455	0.76	62.7	456
10	-871.856680	-871.768598	0.97	55.3	517	-871.874651	-871.781221	0.82	58.6	488	-871.873599	-871.773768	0.76	62.6	456
$\overline{N}$				55.0	520				58.6	489				62.5	457
$\lvert\overline{N}\rvert$				1.6	-15				8.9	-87				1.3	-10
$\sigma_{\overline{N}}$				1.3	12				1.5	13				0.2	1
	PoXeR$_{AT}$					**PoXeR$_{13C}$**					**Rh-A292S**				
Exp.				*50.3*	*568*				*52.1*	*549*				*58.5*	*489*
1	-871.669668	-871.589645	1.48	50.2	569	-871.753876	-871.667334	1.08	54.3	526	-871.809414	-871.716900	0.87	58.1	492
2	-871.664002	-871.583714	1.48	50.4	567	-871.753676	-871.667115	1.07	54.3	526	-871.824426	-871.730355	0.85	59.0	484
3	-871.663689	-871.581688	1.46	51.5	556	-871.757462	-871.670750	1.07	54.4	525	-871.804696	-871.711665	0.87	58.4	490
4	-871.671756	-871.591043	1.47	50.6	565	-871.761107	-871.674443	1.07	54.4	526	-871.805081	-871.712002	0.87	58.4	490
5	-871.664440	-871.584516	1.58	50.2	570	-871.753713	-871.667134	1.08	54.3	526	-871.805352	-871.711734	0.86	58.7	487
6	-871.662765	-871.581424	1.51	51.0	560	-871.739208	-871.651960	1.08	54.7	522	-871.802251	-871.708188	0.84	59.0	484
7	-871.672916	-871.592235	1.45	50.6	565	-871.759283	-871.672828	1.08	54.3	527	-871.822613	-871.728757	0.85	58.9	485
8	-871.673209	-871.594136	1.46	49.6	576	-871.754385	-871.667478	1.07	54.5	524	-871.805085	-871.711891	0.87	58.5	489
9	-871.665943	-871.585143	1.47	50.7	564	-871.757462	-871.670750	1.07	54.4	525	-871.802904	-871.709334	0.84	58.7	487
10	-871.669499	-871.589398	1.47	50.3	569	-871.755760	-871.668751	1.06	54.6	524	-871.804956	-871.711295	0.86	58.8	486
$\overline{N}$				50.5	566				54.4	525				58.7	487
$\lvert\overline{N}\rvert$				0.2	-2				2.3	-24				0.1	-2
$\sigma_{\overline{N}}$				0.5	6				0.1	1				0.3	3
	Rh-A269T					**Rh-E133D**					**Rh-E122Q**				
Exp.				*55.6*	*514*				*56.1*	*510*				*59.6*	*480*
1	-871.825649	-871.737189	0.92	55.5	515	-871.753815	-871.664801	0.87	55.9	512	-871.800783	-871.706540	0.79	59.1	483
2	-871.838544	-871.748022	0.89	56.8	503	-871.753617	-871.665207	0.92	55.5	515	-871.799270	-871.704077	0.83	59.7	479
3	-871.828123	-871.738697	0.89	56.1	510	-871.753702	-871.665287	0.92	55.5	515	-871.799827	-871.704259	0.77	60.0	477
4	-871.824428	-871.734210	0.91	56.6	505	-871.754417	-871.666875	0.90	54.9	520	-871.814554	-871.717645	0.76	60.8	470
5	-871.841457	-871.750825	0.88	56.9	503	-871.753203	-871.664739	0.91	55.5	515	-871.811682	-871.716775	0.79	59.6	480
6	-871.824293	-871.734484	0.90	56.4	507	-871.753398	-871.664855	0.91	55.6	515	-871.811155	-871.715410	0.81	60.1	476
7	-871.825649	-871.737189	0.92	55.5	515	-871.754538	-871.667667	0.90	54.5	524	-871.800921	-871.706214	0.83	59.4	481
8	-871.825649	-871.737189	0.92	55.5	515	-871.753617	-871.665117	0.92	55.5	515	-871.800004	-871.703944	0.78	60.3	474
9	-871.825649	-871.737189	0.92	55.5	515	-871.753648	-871.665096	0.92	55.6	515	-871.812525	-871.715966	0.97	60.6	472
10	-871.838544	-871.748022	0.89	56.8	503	-871.753651	-871.665211	0.92	55.5	515	-871.815900	-871.720181	0.81	60.1	476
$\overline{N}$				56.2	509				55.4	516				60.0	477
$\lvert\overline{N}\rvert$				0.6	-5				-0.7	6				0.4	-3
$\sigma_{\overline{N}}$				0.6	6				0.4	3				0.5	4

Continued on next page

Table S3 — continued from previous page

Seed (N)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)
	Rh-F261Y					**Rh-G90S**					**Rh-T94S**				
Exp.				*56.1*	*510*				*58.4*	*489*				*57.9*	*494*
1	-871.816196	-871.725707	0.91	56.8	504	-871.921780	-871.831392	0.92	56.7	504	-871.770032	-871.678556	0.87	57.4	498
2	-871.815563	-871.725817	0.93	56.3	508	-871.914877	-871.823459	0.89	57.4	498	-871.770404	-871.677284	0.86	58.4	489
3	-871.816175	-871.725522	0.91	56.9	503	-871.929366	-871.838246	0.86	57.2	500	-871.770444	-871.677505	0.86	58.3	490
4	-871.815776	-871.725423	0.92	56.7	504	-871.938941	-871.849893	0.93	55.9	512	-871.770373	-871.678239	0.88	57.8	495
5	-871.815410	-871.725893	0.94	56.2	509	-871.930656	-871.840075	0.90	56.8	503	-871.782211	-871.689712	0.83	58.0	493
6	-871.815180	-871.726224	0.90	55.8	512	-871.922076	-871.830402	0.91	57.5	497	-871.771338	-871.681823	0.87	56.2	509
7	-871.822480	-871.732189	0.92	56.7	505	-871.917704	-871.826860	0.95	57.0	502	-871.782302	-871.689309	0.87	58.4	490
8	-871.815286	-871.726477	0.92	55.7	513	-871.916073	-871.827523	0.93	55.6	515	-871.774594	-871.681385	0.86	58.5	489
9	-871.816772	-871.730177	0.94	54.3	526	-871.937800	-871.846136	0.88	57.5	497	-871.770751	-871.677628	0.85	58.4	489
10	-871.815358	-871.725838	0.94	56.2	509	-871.940414	-871.850529	0.89	56.4	507	-871.770621	-871.677483	0.86	58.4	489
$\overline{N}$				56.2	509				56.8	504				58.0	493
$\lvert\overline{N}\rvert$				0.1	-1				-1.6	15				0.1	-1
$\sigma_{\overline{N}}$				0.8	7				0.7	6				0.7	6
	Rh-T118A					**Rh-W265Y**					**Rh-W265F**				
Exp.				*59.1*	*484*				*59.2*	*483*				*59.6*	*480*
1	-871.694563	-871.599190	0.87	59.8	478	-871.814494	-871.719463	0.84	59.6	479	-871.774650	-871.679353	0.85	59.8	478
2	-871.694419	-871.599121	0.86	59.8	478	-871.809860	-871.714640	0.85	59.8	479	-871.775609	-871.679323	0.82	60.4	473
3	-871.694311	-871.599251	0.86	59.7	479	-871.795375	-871.702927	0.89	58.0	493	-871.776987	-871.682557	0.84	59.3	483
4	-871.694043	-871.598584	0.88	59.9	477	-871.798040	-871.704184	0.88	58.9	485	-871.775420	-871.679402	0.83	60.3	475
5	-871.694469	-871.599018	0.83	59.9	477	-871.808482	-871.714765	0.88	58.8	486	-871.775999	-871.679993	0.82	60.2	475
6	-871.709362	-871.613468	0.87	60.2	475	-871.799573	-871.706572	0.86	58.4	490	-871.776224	-871.680139	0.81	60.3	474
7	-871.707837	-871.612874	0.87	59.6	480	-871.814696	-871.721797	0.85	58.3	490	-871.775309	-871.680158	0.85	59.7	479
8	-871.694493	-871.599014	0.86	59.9	477	-871.796773	-871.701996	0.99	59.5	481	-871.775819	-871.679532	0.80	60.4	473
9	-871.694177	-871.599129	0.85	59.6	479	-871.799709	-871.706752	0.87	58.3	490	-871.775124	-871.680246	0.86	59.5	480
10	-871.694452	-871.601071	0.86	58.6	488	-871.797972	-871.704619	0.89	58.6	488	-871.775765	-871.679432	0.82	60.4	473
$\overline{N}$				59.7	479				58.8	486				60.0	476
$\lvert\overline{N}\rvert$				0.6	-5				-0.4	3				0.4	-4
$\sigma_{\overline{N}}$				0.4	4				0.6	5				0.4	3
	Rh-D83N-E122Q					**Rh-A292S-A295S-A299C**					**ASR$_{AT}$-D217E**				
Exp.				*60.2*	*475*				*59.1*	*484*				*51.8*	*552*
1	-872.060033	-871.964370	0.78	60.0	476	-871.975981	-871.881896	0.86	59.0	484	-871.930459	-871.846962	1.26	52.4	546
2	-872.062205	-871.964112	0.77	61.6	464	-871.971764	-871.878114	0.87	58.8	487	-871.913533	-871.829757	1.27	52.6	544
3	-872.068860	-871.972036	0.77	60.8	471	-871.973050	-871.878846	0.87	59.1	484	-871.930620	-871.847204	1.26	52.3	546
4	-872.062291	-871.964201	0.74	61.6	465	-871.955098	-871.862576	0.87	58.1	492	-871.930394	-871.846940	1.26	52.4	546
5	-872.070497	-871.974612	0.80	60.2	475	-871.974425	-871.882216	0.87	57.9	494	-871.930497	-871.846922	1.26	52.4	545
6	-872.068335	-871.971752	0.74	60.6	472	-871.954187	-871.862017	0.87	57.8	494	-871.930819	-871.847506	1.28	52.3	547
7	-872.067357	-871.969521	0.74	61.4	466	-871.972978	-871.878222	0.82	59.5	481	-871.930529	-871.847320	1.26	52.2	548
8	-872.068999	-871.971163	0.77	61.4	466	-871.952733	-871.860769	0.87	57.7	495	-871.922630	-871.844577	1.45	49.0	584
9	-872.069687	-871.971518	0.75	61.6	464	-871.952913	-871.859142	0.86	58.8	486	-871.930904	-871.847895	1.28	52.1	549
10	-872.062174	-871.963799	0.74	61.7	463	-871.964745	-871.870067	0.84	59.4	481	-871.931503	-871.848618	1.29	52.0	550
$\overline{N}$				61.1	468				58.6	488				52.0	550
$\lvert\overline{N}\rvert$				0.9	-7				-0.5	4				0.2	-2
$\sigma_{\overline{N}}$				0.6	5				0.7	5				1.1	12

Continued on next page

Table S3 — continued from previous page

Seed (N)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)
a-ARM$_{customized}$															
	KR2-2$^{(c)}$					**BPR$^{(c)}$**					**RCone$^{(c)}$**				
Exp.				*54.5*	*525*				*58.3*	*490*				*49.7*	*575*
1	-871.806421	-871.718965	1.13	54.9	521	-872.093507	-872.002334	0.72	57.2	500	-871.765053	-871.680551	1.10	53.0	539
2	-871.791326	-871.702123	0.87	56.0	511	-872.100809	-872.009175	0.73	57.5	497	-871.754866	-871.675919	1.16	49.5	577
3	-871.817554	-871.728126	0.86	56.1	509	-872.106426	-872.015214	0.77	57.2	500	-871.768755	-871.688682	1.06	50.2	569
4	-871.816829	-871.727536	0.86	56.0	510	-872.094144	-872.002679	0.74	57.4	498	-871.754665	-871.676575	1.15	49.0	583
5	-871.779909	-871.689883	1.00	56.5	506	-872.094386	-872.003224	0.76	57.2	500	-871.752627	-871.674553	1.15	49.0	584
6	-871.816854	-871.727560	0.86	56.0	510	-872.086504	-871.995213	0.76	57.3	499	-871.745093	-871.664668	0.99	50.5	567
7	-871.791072	-871.701981	0.88	55.9	511	-872.086485	-871.995451	0.76	57.1	501	-871.763063	-871.683131	1.20	50.2	570
8	-871.791012	-871.702218	0.89	55.7	513	-872.098137	-872.006542	0.74	57.5	497	-871.754665	-871.676502	1.15	49.0	583
9	-871.816596	-871.727253	0.87	56.1	510	-872.091678	-872.001855	0.76	56.4	507	-871.754604	-871.676427	1.15	49.1	583
10	-871.816596	-871.727253	0.87	56.1	510	-872.096564	-872.005594	0.76	57.1	501	-871.754866	-871.675919	1.16	49.5	577
$\overline{N}$				55.9	511				57.2	500				49.9	573
$\lvert\overline{N}\rvert$				1.4	-14				-1.1	10				0.2	-2
$\sigma_{\overline{N}}$				0.4	4				0.3	3				1.2	14
	ChR$^{(c)}_{C1C2}$					**ChR2$^{(c)}$**					**ChR2-C128T$^{(c)}$**				
Exp.				62.4	458				60.8	470				59.0	485
1	-871.766869	-871.665904	0.64	63.4	451	-871.966578	-871.869065	0.80	61.2	467	-871.968173	-871.873554	0.94	59.4	482
2	-871.765234	-871.662713	0.57	64.3	444	-871.959771	-871.858873	0.78	63.3	452	-871.968639	-871.874623	0.98	59.0	485
3	-871.766757	-871.665228	0.63	63.7	449	-871.966354	-871.868754	0.79	61.2	467	-871.975749	-871.881662	0.94	59.0	484
4	-871.766078	-871.664917	0.63	63.5	450	-871.966776	-871.869326	0.81	61.2	468	-871.968681	-871.873604	0.96	59.7	479
5	-871.765935	-871.666631	0.72	62.3	459	-871.960907	-871.859742	0.75	63.5	450	-871.968570	-871.874403	0.95	59.1	484
6	-871.766677	-871.665276	0.63	63.6	449	-871.959581	-871.858404	0.72	63.5	450	-871.968575	-871.874466	0.95	59.1	484
7	-871.765845	-871.662488	0.57	64.9	441	-871.967420	-871.869935	0.88	61.2	467	-871.974295	-871.880181	0.90	59.1	484
8	-871.766857	-871.665751	0.65	63.4	451	-871.959226	-871.858047	0.64	63.5	450	-871.965017	-871.870650	0.96	59.2	483
9	-871.766838	-871.665506	0.64	63.6	450	-871.936490	-871.839350	0.89	61.0	469	-871.968436	-871.874313	0.96	59.1	484
10	-871.765786	-871.662497	0.57	64.8	441	-871.960510	-871.859972	0.67	63.1	453	-871.969026	-871.873936	0.95	59.7	479
$\overline{N}$				63.8	449				62.3	459				59.2	483
$\lvert\overline{N}\rvert$				1.4	-9				1.5	-11				0.2	-2
$\sigma_{\overline{N}}$				0.8	5				1.2	9				0.3	2
	bR$^{(c)}_{AT}$					**bR$^{(c\text{-}2)}_{AT}$**									
Exp.				*50.3*	*568*				*50.3*	*568*					
1	-871.749782	-871.669143	1.43	50.6	565.0275404	-872.054117	-871.973964	1.38	50.3	568					
2	-871.748107	-871.667915	1.43	50.3	568.1787667	-872.052652	-871.972753	1.37	50.1	570					
3	-871.748042	-871.667100	1.43	50.8	562.9133129	-872.052420	-871.972882	1.40	49.9	573					
4	-871.747525	-871.666578	1.42	50.8	562.8774994	-872.052928	-871.971979	1.38	50.8	563					
5	-871.748246	-871.667982	1.44	50.4	567.6712091	-872.055789	-871.975009	1.36	50.7	564					
6	-871.747018	-871.666745	1.43	50.4	567.59929	-872.055729	-871.974957	1.36	50.7	564					
7	-871.747316	-871.666612	1.43	50.6	564.5677742	-872.056266	-871.975792	1.37	50.5	566					
8	-871.747537	-871.666875	1.42	50.6	564.8664286	-872.054117	-871.973964	1.38	50.3	568					
9	-871.757220	-871.675159	1.39	51.5	555.2339365	-872.055797	-871.975020	1.35	50.7	564					
10	-871.747892	-871.667378	1.43	50.5	565.903774	-872.060348	-871.980071	1.37	50.4	568					
$\overline{N}$				50.7	564				50.4	567					
$\lvert\overline{N}\rvert$				0.4	-4				0.1	-1					

Continued on next page

Table S3 — continued from previous page

Seed (N)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)
$\sigma_{\overline{N}}$				0.3	4				0.3	3					
a-ARM$_{\text{customized}}$ (Effect of chain and pH)															
	ASR$_{AT}$-1$^{(c\text{-}pH)}$					**ASR$_{13C}$-2$^{(c\text{-}pH)}$**					**bR$^{c\text{-}pH}_{13C}$**				
Exp.				52.0	550				53.2	537				52.2	547
1	-872.011057	-871.916625	0.97	59.3	483	-872.030188	-871.934729	1.10	59.9	477	-872.304485	-872.203163	0.62	63.6	450
2	-872.010414	-871.916614	0.95	58.9	486	-872.024819	-871.930528	0.96	59.2	483	-872.300505	-872.199325	0.64	63.5	450
3	-872.009242	-871.915443	1.00	58.9	486	-872.024947	-871.930541	0.92	59.2	483	-872.288999	-872.188334	0.66	63.2	453
4	-872.009660	-871.916203	0.98	58.6	488	-872.024825	-871.930716	0.95	59.1	484	-872.289389	-872.188806	0.67	63.1	453
5	-871.997217	-871.906871	1.19	56.7	504	-872.025047	-871.930507	0.93	59.3	482	-872.288866	-872.188332	0.68	63.1	453
6	-872.009659	-871.915884	0.98	58.8	486	-872.025166	-871.930423	0.93	59.5	481	-872.30111	-872.200178	0.65	63.3	451
7	-872.001978	-871.905460	0.85	60.6	472	-872.057313	-871.964260	0.93	58.4	490	-872.294073	-872.193273	0.64	63.3	452
8	-872.012484	-871.917874	0.94	59.4	482	-872.025158	-871.930510	0.93	59.4	481	-872.288968	-872.188371	0.63	63.1	453
9	-872.009637	-871.915918	0.96	58.8	486	-872.057181	-871.964433	0.92	58.2	491	-872.288981	-872.188391	0.64	63.1	453
10	-872.009541	-871.914735	1.01	59.5	481	-872.024881	-871.930195	0.95	59.4	481	-872.294849	-872.19439	0.67	63	454
$\overline{N}$				59.0	485				59.2	483				63.2	452
$\lvert\overline{N}\rvert$				7.0	-65				6.0	-54				11.0	-95
$\sigma_{\overline{N}}$				1.0	8				0.5	4				0.2	1
	bR$^{(c\text{-}pH)}$					**bathoRh$^{(c\text{-}pH)}$**					**bathoRh$^{(c)}$**				
Exp.				50.3	568				54	529				54	529
1	-871.892865	-871.790117	0.65	64.5	443	-871.806085	-871.716416	0.92	56.3	508	-871.931	-871.844269	1.08	54.4	525
2	-871.892663	-871.790024	0.65	64.4	444	-871.785693	-871.69443	0.93	57.3	499	-871.916126	-871.830093	1.17	54	530
3	-871.892744	-871.790225	0.66	64.3	444	-871.780228	-871.689937	0.97	56.7	505	-871.911142	-871.825087	1.09	54	529
4	-871.892706	-871.789216	0.71	64.9	440	-871.803163	-871.712138	0.95	57.1	501	-871.916011	-871.831531	1.06	53	539
5	-871.892778	-871.790061	0.65	64.5	444	-871.778822	-871.687377	0.91	57.4	498	-871.911302	-871.825178	1.09	54	529
6	-871.892712	-871.790196	0.66	64.3	444	-871.780053	-871.689403	0.96	56.9	503	-871.927654	-871.841412	1.08	54.1	528
7	-871.892806	-871.790431	0.65	64.2	445	-871.779492	-871.688817	0.86	56.9	502	-871.916213	-871.829872	1.15	54.2	528
8	-871.892654	-871.790464	0.65	64.1	446	-871.81367	-871.721809	0.81	57.6	496	-871.911089	-871.825008	1.08	54	529
9	-871.891477	-871.791638	0.69	62.6	456	-871.785693	-871.69443	0.93	57.3	499	-871.927479	-871.841084	1.09	54.2	527
10	-871.892663	-871.790024	0.65	64.4	444	-871.778822	-871.687377	0.91	57.4	498	-871.915855	-871.829806	1.09	54	530
$\overline{N}$				64.2	445				57.1	501				54.0	529
$\lvert\overline{N}\rvert$				13.9	-123				3.1	-28				0.0	0
$\sigma_{\overline{N}}$				0.6	4				0.4	4				0.4	4
	BPR$^{(c\text{-}2)}$					**BPR$^{(c\text{-}pH\text{-}2)}$**					**Rh$^{(c)}$**				
Exp.				58.3	490				58.3	490				57.4	498
1	-872.092199	-871.990725	0.58	63.7	449	-872.141865	-872.051133	0.86	56.9	502	-871.960867	-871.871577	0.94	56	510
2	-872.092759	-871.991196	0.58	63.7	449	-872.133237	-872.042133	0.89	57.2	500	-871.95992	-871.86997	1.11	56.4	507
3	-872.09237	-871.990778	0.57	63.7	448	-872.14175	-872.051126	0.86	56.9	503	-871.972581	-871.882799	0.89	56.3	507
4	-872.091715	-871.990006	0.62	63.8	448	-872.141611	-872.050852	0.86	57	502	-871.960032	-871.870863	1.1	56	511
5	-872.093278	-871.99146	0.60	63.9	447	-872.133057	-872.041955	0.88	57.2	500	-871.961765	-871.873231	0.96	55.6	515
6	-872.091774	-871.990248	0.57	63.7	449	-872.132986	-872.042078	0.87	57	501	-871.971835	-871.882364	0.93	56.1	509
7	-872.092212	-871.990614	0.56	63.8	448	-872.140381	-872.049827	0.86	56.8	503	-871.964434	-871.875355	1.12	55.9	511
8	-872.102907	-872.00099	0.56	64.0	447	-872.133413	-872.042384	0.89	57.1	501	-871.960524	-871.873657	0.94	54.5	525
9	-872.092288	-871.990441	0.57	63.9	447	-872.141516	-872.050709	0.85	57	502	-871.971679	-871.884234	0.94	54.9	521
10	-872.089566	-871.988996	0.47	63.1	453	-872.133237	-872.042133	0.89	57.2	500	-871.969553	-871.879925	0.93	56.2	508
$\overline{N}$				63.7	449				57.0	501				55.8	512

Continued on next page

Table S3 — continued from previous page

Seed (N)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)	S_0 Energy (a.u.)	S_1 Energy (a.u.)	f_{Osc}	ΔE_{S1-S0} (kcal/mol)	λ^a_{max} (nm)
$\lvert\overline{N}\rvert$				5.4	-42				-1.3	11				-1.6	14
$\sigma_{\overline{N}}$				0.2	2				0.1	1				0.6	6
	$Rh^{c\text{-}pH\text{-}2}$					**$SqRh^{c}$**									
Exp.				57.4	498				58.5	489					
1	-872.001068	-871.897257	0.75	65.1	439	-871.756918	-871.662109	0.82	59.5	481					
2	-872.000111	-871.897181	0.68	64.6	443	-871.758008	-871.663224	0.83	59.5	481					
3	-872.000981	-871.897183	0.74	65.1	439	-871.757519	-871.663524	0.84	59	485					
4	-872.000954	-871.897406	0.74	65.0	440	-871.757079	-871.662262	0.83	59.5	481					
5	-872.001058	-871.897215	0.75	65.2	439	-871.757597	-871.66267	0.82	59.6	480					
6	-872.000881	-871.896998	0.74	65.2	439	-871.757229	-871.662417	0.82	59.5	481					
7	-872.001147	-871.897332	0.75	65.1	439	-871.757272	-871.662533	0.82	59.4	481					
8	-872.000922	-871.897069	0.74	65.2	439	-871.756828	-871.66209	0.82	59.4	481					
9	-872.001094	-871.897164	0.74	65.2	438	-871.757211	-871.663164	0.84	59	484					
10	-872.009042	-871.906454	0.75	64.4	444	-871.756865	-871.66208	0.82	59.5	481					
$\overline{N}$				65.0	440				59.4	482					
$\lvert\overline{N}\rvert$				7.6	-58				0.9	-7					
$\sigma_{\overline{N}}$				0.3	2				0.2	2					

Table S4: Trend deviation factor (||Trend Dev.||) for the a-ARM$_{\text{default}}$ and a-ARM$_{\text{customized}}$ approaches, expressed as mean absolute error (MAE) and mean absolute deviation (MAD) of the x=38 rhodopsins of the benchmark set.

			a-**ARM$_{\text{default}}$**			a-**ARM$_{\text{customized}}$**		
Rhodopsin	$\Delta E^{Exp}_{\text{S1-S0}}$	$\delta^{Rh,Exp}_{x,Exp}\Delta E_{\text{S1-S0}}$ [a]	$\Delta E_{\text{S1-S0}}$	$\delta^{Rh,Calc}_{x,Calc}\Delta E_{\text{S1-S0}}$ [b]	||Trend Dev.|| [c]	$\Delta E_{\text{S1-S0}}$	$\delta^{Rh,Calc}_{x,Calc}\Delta E_{\text{S1-S0}}$ [b]	||Trend Dev.|| [c]
Rh	57.4	0.0	57.7	0.0	0.0	57.7	0.0	0.0
bathoRh	54.0	3.4	56.1	1.6	1.8	56.1	1.6	1.8
SqRh	58.5	1.1	60.9	3.2	2.1	60.9	3.2	2.1
BPR	58.3	0.9	63.7	6.0	5.1	57.0	0.7	0.2
ASR$_{AT}$-1	52.1	5.3	52.4	5.3	0.0	52.4	5.3	0.0
ASR$_{13C}$-2	53.2	4.2	54.2	3.5	0.7	54.2	3.5	0.7
bR$_{AT}$	50.3	7.1	53.2	4.5	2.6	50.6	7.1	0.0
bR$_{13C}$	52.2	5.2	53.3	4.4	0.8	53.3	4.4	0.8
ChR$_{C1C2}$	62.4	5.0	76.9	19.2	14.2	63.8	6.1	1.1
hMeOp	60.4	3.0	61.2	3.5	0.5	61.2	3.5	0.5
BCone	66.5	9.1	67.8	10.1	1.0	67.8	10.1	1.0
GCone	53.4	4.0	55.0	2.7	1.3	55.0	2.7	1.3
RCone	49.7	7.7	58.5	0.8	6.9	49.9	7.8	0.1
mMeOp	61.2	3.8	62.3	4.6	0.8	62.3	4.6	0.8
SqbaRh	54.0	3.4	55.5	2.2	1.2	55.5	2.2	1.2
SR-II	57.5	0.1	58.0	0.3	0.2	58.0	0.3	0.2
ASR$_{AT}$-D217E	52.0	5.4	51.8	5.9	0.5	51.8	5.9	0.5
Arch1	50.3	7.1	50.5	7.2	0.1	50.5	7.2	0.1
AARh	56.3	1.1	58.9	1.2	0.1	58.9	1.2	0.1
Arch2	52.0	5.4	54.5	3.2	2.2	54.5	3.2	2.2
ChR2	60.8	3.4	79.9	22.2	18.8	63.3	5.6	2.2
ChR2-C128T	59.0	1.6	79.7	22.0	20.4	59.2	1.5	0.1
KR2-2	54.4	3.0	69.4	11.7	8.7	55.9	1.8	1.2
PoXeR$_{AT}$	50.3	7.1	50.5	7.2	0.1	50.5	7.2	0.1
PoXeR$_{13C}$	52.1	5.3	54.3	3.4	1.9	54.3	3.4	1.9
NM-R3	55.3	2.1	56.1	1.6	0.5	56.1	1.6	0.5
ClR	55.3	2.1	55.2	2.5	0.4	55.2	2.5	0.4
F261Y	56.1	1.3	56.2	1.5	0.2	56.2	1.5	0.2
T94S	57.9	0.5	58.0	0.3	0.2	58.0	0.3	0.2
A292S	58.5	1.1	58.7	1.0	0.1	58.7	1.0	0.1
W265Y	59.0	1.6	58.8	1.1	0.5	58.8	1.1	0.5
W265F	59.6	2.2	60.0	2.3	0.1	60.0	2.3	0.1
T118A	59.1	1.7	59.7	2.0	0.3	59.7	2.0	0.3
G90S	58.4	1.0	57.1	0.6	0.4	57.1	0.6	0.4
E122Q	59.6	2.2	60.0	2.3	0.1	60.0	2.3	0.1
A269T	55.6	1.8	56.1	1.6	0.2	56.1	1.6	0.2
E113D	56.1	1.3	55.4	2.3	1.0	55.4	2.3	1.0
D83N/E122Q	60.2	2.8	60.9	3.2	0.4	60.9	3.2	0.4
A292S/A295S/A299C	59.1	1.7	58.6	0.9	0.8	58.6	0.9	0.8
MAE of ||Trend Dev.||					2.5			0.7
MAD of ||Trend Dev.||					1.2			0.5

[a] Difference between the experimental $\Delta E^{Exp}_{\text{S1-S0}}$ of each of the x rhodopsins with respect to experimental value of Rh ($\delta^{Rh,Exp}_{x,Exp}\Delta E_{\text{S1-S0}}$).

[b] Difference between the calculated $\Delta E_{\text{S1-S0}}$ of each of the x rhodopsins with respect to calculated value of Rh ($\delta^{Rh,Calc}_{x,Calc}\Delta E_{\text{S1-S0}}$).

[c] $||\text{Trend Dev.}|| = |(\delta^{Rh,Exp}_{x,Exp}\Delta E_{\text{S1-S0}})\text{-}(\delta^{Rh,Calc}_{x,Calc}\Delta E_{\text{S1-S0}})|$

S7 Details of the employed comparative modelling protocol for AARh, PoXeR, hMeOp, mMeOp and BCone, GCone, RCone

Comparative modeling was carried out by means of the software MODELLER 1.[?] Sequence identity between target and template protein dictated the modeling strategy. In that respect, values 70% without gaps allowed simple substitution of mutated amino acid side chains while transferring the coordinates of all main chain and of conserved side chain atoms. That was the case of AARh, characterized by 93% of sequence identity to the template, bovine rhodopsin. The same approach was used for modeling PoXeR, in spite of a sequence identity to the Anabaena sensory rhodopsin template lower than the threshold above (i.e. 51%). The reason was that the modeled PoXeR primary sequence aligned without gaps to the sequence of the template and almost all mutated positions point towards the membrane or the extracellular and intracellular water (i.e. only 8 out of the 106 mutated positions point towards the core of the helix bundle).

The same approach was used for modeling PoXeR, whose primary sequence aligned without gaps to the sequence of the Anabaena sensory rhodopsin template, mutated amino acids lying far from the retinal binding site. The other models, i.e. those of human cone opsins (based on bovine rhodopsin) and of mouse melanopsin (based on squid rhodopsin) were achieved by a high degree of model refinement upon randomizing all the Cartesian coordinates of standard residues in the initial model, which produced multiple models from the same alignment (i.e. 250 for each cone opsin and 100 for mouse melanopsin). For each run, the top twenty models, characterized by the lowest values of the MODELLER objective function (which means lowest degree of restraint violation) were subjected to quality checks, which verified the correctness of main chain conformation, leading to selection of one or more models. For each run, the top twenty models were selected, characterized by the lowest values of the MODELLER objective function (which means lowest degree of restraint violation). The set of selected models were subjected to quality checks, which verified the correctness of main chain conformation, leading

to selection of one or more models.

For all modeled opsins, side chains were subjected to automatic adjustment if in non-allowed conformation by using three different backbone-independent and backbone-dependent rotamer libraries. Care was put in keeping as much as possible the conformation of conserved amino acids in the template retinal binding site. Water molecules in conserved amino acid environments were transferred from the template structure to the final model(s) of the target protein.

Table S5: Main features of the comparative models

Code	Protein	Template PDB	Chain	% Identity[a]	N. Water[b]
AARh	Ancestral Archosaur	1U19[S10]	A	93	21
Bcone	Blue Cone opsin	1U19[S10]	A	44	14
Gcone	Green Cone opsin	1U19[S10]	A	41	10
Rcone	Red Cone opsin	1U19[S10]	A	40	10
mMeOp	Mouse melanopsin	2Z73[S11]	A	45	8
PoXeR$_{13C}$	Sensory rhodopsin	4TL3[?]	B[c]	51	5
PoXeR$_{AT}$	Sensory rhodopsin	4TL3[?]	B[c]	51	5

[a]Percentage of sequence identity computed as a ratio between the number of aligned positions.
[b]Number of water molecules translated from the template to the target structures as topologically equivalent.
[c]For PoXe$_{R13}$ and PoXeR$_{AT}$ we selected chain B, rather chain A, because is slightly more complete.

S8 Further details on the assignment of ionizable residues protonation state

As reported in the main text at the end of Section 2.2.3, PROPKA data may not always give a realistic (*i.e.*, conforming to available experimental data or current consensus) results. This is mostly felt with regard to histidines, which show a pK_a value often close to the imposed pH, and thus, it is often unclear which is their actual protonation state. Taking a leaf from the previous ARM work,[S1] we decided to modulate the assignment of the protonation state to ionizable residues as follows. Doing so, we were able to obtain protonation states more in line with currently accepted data.[S1]

Concerning Asp, Glu, Arg and Lys residues, we followed the indication given by the computed charge (as reported in Equation 5), as long as the residue's corresponding buried percentage was higher than 55%. For His residues, we applied the further condition that the computed shift value (ΔpK_a) is higher than 1.6 pH units.

Nevertheless, we would like to stress once more the importance of employing the correct protonation state for ionizable residues, and to follow the available experimental data whenever it is possible, as illustrated in Sections 3.2 and, particularly, 3.2.4.

S9 Notes

Figures of molecular structures were generated with PyMOL (http://www.pymol.org)

References

(S1) Melaccio, F.; del Carmen Marín, M.; Valentini, A.; Montisci, F.; Rinaldi, S.; Cherubini, M.; Yang, X.; Kato, Y.; Stenrup, M.; Orozco-González, Y.; Ferre, N.; Luk, H. L.; Kandori, H.; Olivucci, M. Toward Automatic Rhodopsin Modeling as a Tool for High-Throughput Computational Photobiology. *JCTC* **2016**, *12*, 6020–6034.

(S2) Vogeley, L.; Sineshchekov, O. A.; Trivedi, V. D.; Sasaki, J.; Spudich, J. L.; Luecke, H. Anabaena Sensory Rhodopsin: A Photochromic Color Sensor at 2.0 Å. *Science* **2004**, *306*, 1390–1393.

(S3) Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Res* **2018**, *46*, W363–W367.

(S4) Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

(S5) Nogly, P.; Weinert, T.; James, D.; Carbajo, S.; Ozerov, D.; Furrer, A.; Gashi, D.; Borin, V.; Skopintsev, P.; Jaeger, K.; Nass, K.; Bath, P.; Bosman, R.; Koglin, J.; Seaberg, M.; Lane, T.; Kekilli, D.; Brnle, S.; Tanaka, T.; Wu, W.; Milne, C.; White, T.; Barty, A.; Weierstall, U.; Panneels, V.; Nango, E.; Iwata, S.; Hunter, M.; Schapiro, I.; Schertler, G.; Neutze, R.; Standfuss, J. Retinal Isomerization in Bacteriorhodopsin Captured by a Femtosecond X-Ray Laser. *Science* **2018**, eaat0094.

(S6) Nishikawa, T.; Murakami, M.; Kouyama, T. Crystal Structure of the 13-*cis* Isomer of Bacteriorhodopsin in the Dark-Adapted State. *J. Mol. Biol.* **2005**, *352*, 319–328.

(S7) Nakamichi, H.; Okada, T. Crystallographic Analysis of Primary Visual Photochemistry. *Angew. Chem.* **2006**, *118*, 4376–4379.

(S8) Ran, T.; Ozorowski, G.; Gao, Y.; Sineshchekov, O. A.; Wang, W.; Spudich, J. L.;

Luecke, H. Cross-Protomer Interaction with the Photoactive Site in Oligomeric Proteorhodopsin Complexes. *Acta Crystallogr. D* **2013**, *69*, 1965–1980.

(S9) Kato, H. E.; Zhang, F.; Yizhar, O.; Ramakrishnan, C.; Nishizawa, T.; Hirata, K.; Ito, J.; Aita, Y.; Tsukazaki, T.; Hayashi, S.; Hegemann, P.; Maturana, A. D.; Ishitani, R.; Deisseroth, K.; Nureki, O. Crystal structure of the channelrhodopsin light-gated cation channel. *Nature* **2012**, *482*, 369.

(S10) Okada, T.; Sugihara, M.; Bondar, A. N.; Elstner, M.; Entel, P.; Buss, V. The Retinal Conformation and its Environment in Rhodopsin in Light of a New 2.2 Å crystal structure. *J. Mol. Biol.* **2004**, *342*, 571–583.

(S11) Murakami, M.; Kouyama, T. Crystal Structure of Squid Rhodopsin. *Nature* **2008**, *453*, 363–367.

(S12) Dong, B.; Sánchez-Magraner, L.; Luecke, H. Structure of an Inward Proton-Transporting Anabaena Sensory Rhodopsin Mutant: Mechanistic Insights. *Biophys. J.* **2016**, *111*, 963–972.

(S13) Enami, N.; Yoshimura, K.; Murakami, M.; Okumura, H.; Ihara, K.; Kouyama, T. Crystal Structures of Archaerhodopsin-1 and -2: Common Structural Motif in Archaeal Light-driven Proton Pumps. *J. Mol. Biol.* **2006**, *358*, 675–685.

(S14) Kouyama, T.; Fujii, R.; Kanada, S.; Nakanishi, T.; Chan, S. K.; Murakami, M. Structure of Archaerhodopsin-2 at 1.8 Å Resolution. *Acta Crystallogr. D* **2014**, *70*, 2692–2701.

(S15) Volkov, O.; Kovalev, K.; Polovinkin, V.; Borshchevskiy, V.; Bamann, C.; Astashkin, R.; Marin, E.; Popov, A.; Balandin, T.; Willbold, D.; Büldt, G.; Bamberg, E.; Gordeliy, V. Structural insights into ion conduction by channelrhodopsin 2. *Science* **2017**, *358*, eaan8862.

(S16) Kato, H. E.; Inoue, K.; Abe-Yoshizumi, R.; Kato, Y.; Ono, H.; Konno, M.; Hososhima, S.; Ishizuka, T.; Hoque, M. R.; Kunitomo, H.; Ito, J.; Yoshizawa, S.; Yamashita, K.; Take-

moto, M.; Nishizawa, T.; Taniguchi, R.; Kogure, K.; Maturana, A. D.; Iino, Y.; Yawo, H.; Ishitani, R.; Kandori, H.; Nureki, O. Structural Basis for Na^+ Transport Mechanism by a Light-Driven Na^+ Pump. *Nature* **2015**, *521*, 48–53.

(S17) Hosaka, T.; Yoshizawa, S.; Nakajima, Y.; Ohsawa, N.; Hato, M.; DeLong, E. F.; Kogure, K.; Yokoyama, S.; Kimura-Someya, T.; Iwasaki, W.; Shirouzu, M. Structural Mechanism for Light-Driven Transport by a New Type of Chloride Ion Pump, *Nonlabens marinus* Rhodopsin-3. *J. Biol. Chem.* **2016**, *291*, 17488–17495.

(S18) Kim, K.; Kwon, S.-K.; Jun, S.-H.; Cha, J. S.; Kim, H.; Lee, W.; Kim, J. F.; Cho, H.-S. Crystal Structure and Functional Characterization of a Light-Driven Chloride Pump Having an NTQ Motif. *Nat. Commun.* **2016**, *7*, 12677–12687.

(S19) Luecke, H.; Schobert, B.; Lanyi, J. K.; Spudich, E. N.; Spudich, J. L. Crystal Structure of Sensory Rhodopsin II at 2.4 Angstroms: Insights Into Color Tuning and Transducer Interaction. *Science* **2001**, *293*, 1499–1503.

(S20) Murakami, M.; Kouyama, T. Crystallographic Analysis of the Primary Photochemical Reaction of Squid Rhodopsin. *J. Mol. Biol.* **2011**, *413*, 615–627.