Supporting Information

Identification of the Deprotonated Pyrrole Nitrogen of the Bilin-based Photoreceptor by Raman Spectroscopy with an Advanced Computational Analysis

Shinsuke Osoegawa,[†] Risako Miyoshi,[†] Kouhei Watanabe,[†] Yuu Hirose,[‡] Tomotsumi Fujisawa,[†] Masahiko Ikeuchi,[§] Masashi Unno[†]

[†]*Department of Chemistry and Applied Chemistry, Faculty of Science and Engineering, Saga University, Saga 840-8502, Japan,* [‡]*Department of Environmental and Life Sciences, Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan, and* [§]*Department of Life Sciences (Biology), The University of Tokyo, Meguro, Tokyo 153-8902, Japan*

Experimental Procedures

Sample Preparations and Spectroscopy

Expression and purification of recombinant RcaE holoproteins from PCB-producing *E. coli* were performed as described previously.[1] The buffer solutions used in the spectroscopic measurements for RcaE were 20 mM MOPS-NaOH pH 6.5 and 20 mM Tris-HCl pH 8.0.

Optical absorption spectra were measured at room temperature by using a UV-1800 spectrophotometer (Shimadzu). For excitation, light-emitting diodes (LED) emitting at 535 nm were used as a green light, and LED emitting at 626 nm were used as a red light.

Resonance Raman spectra with 785 nm excitation were obtained as described previously.[2] A diode laser (StarBright 785 XM; Torsana Laser Technologies, Sweden) was used to measure the Raman spectra. The sample was contained in a 3 x 3 x 48 mm quartz cuvette, and the backscattered light from the sample was collected by an aspheric glass condenser lens and refocused by an achromatic lens to the end of a fiber bundle with a core diameter of 100 μm. A long-pass edge filter (Semrock Inc.) was used to reject the laser light before the refocusing lens. The light was directed into a spectrometer (Acton LS 785; Princeton Instruments) equipped with a thermoelectrically cooled CCD detector (PIXIS: 256E; Princeton Instruments). The spectral resolution was ~11 cm^{-1} (FWHM).

Resonance Raman spectra with 441.6 nm excitation were obtained using a spectrometer system composed of a helium-cadmium laser (IK5651R-G; Kimmon Electric Ltd., Tokyo), a 0.5 m single spectrometer (Spex 500M; HORIBA Jobin Yvon, Edison, NJ) equipped with a 1800 groove/mm holographic grating, and a liquid nitrogen-cooled UV-coated CCD detector (Spec-10:400B; Roper Scientific Inc., Trenton, NJ).[3] A 90° scattering geometry was employed, and the scattered photons were collected and focused onto the entrance slit of a spectrometer by using two quartz planoconvex lenses. A Triax190 spectrometer (HORIBA Jobin Yvon, Edison, NJ) was used to remove the excitation light, and the first order of the dispersed light by the 500M spectrometer was imaged on the detector. An entrance slit width of 100 μm corresponded to a spectral resolution of ~4 cm^{-1}, which is better than that for the spectra with 785 nm excitation. All spectra were taken at room temperature (~25 °C), and in-house software was used to eliminate the noise spikes in the spectra caused by cosmic rays. All Raman spectra were calibrated by using neat fenchone as a standard. Sample volumes were 50 μL and were contained in a quartz spinning cell (10 mm in diameter). The cell was spun at 1600 rpm.

MD and QM/MM Calculations

Both the initial setup and the MD runs were performed with the Amber16 program[4] using an explicit representation of solvent molecules and the ff14SB all-atom force field.[5] The initial starting geometry for the ^{15Z}Pg models was taken from the crystal structure of AnPixJg2 (PDB: 3W2Z). Hydrogen atoms were added by the AmberTools xLeap and the protonation state of ionizable groups was set to a neutral pH condition. Note that His322 near the chromophore was protonated at the d-nitrogen. The pyrrole nitrogen of the ring-B or C was deprotonated for models Pg-B and Pg-C, respectively. The protein molecule was solvated in a periodic octahedral box of TIP3P[6] water molecules (10592 waters) and neutralized. The MD calculation was performed as described previously.[7,8] The entire system was subjected to minimization and then heated to 300 K. After heating, the obtained system was simulated for 50 ns at 300 K and 1 atm.

All quantum chemical calculations were performed using the program Gaussian16.[9] A two-layer ONIOM[10] method was used to perform the QM/MM calculations. The QM region consists of the PCB chromophore, while the remainder of the system was treated as the MM region. The initial geometries were obtained from snapshots of the MD simulation, and surrounding water molecules were removed. The QM region was treated by DFT, whereas the MM part was described with the Amber force field[11] and TIP3P parameters for water. The QM part of the system was computed at the B3LYP/6-31G* (geometry optimization and force field calculation) or CAM-B3LYP/6-31G** (Raman intensity) level of theory. In most cases, an electronic embedding scheme that considers the partial charges of the MM region into the QM Hamiltonian was used. For geometry optimizations, the positions of the MM atoms were frozen. Harmonic frequencies as well as Raman intensities with 785 nm excitation were computed, and the simulated spectra were generated assuming a Gaussian band shape with a half-width of 10 cm^{-1}. The calculated vibrational frequencies were uniformly scaled using a factor of 0.98.

Supplemental Results and Discussion

DFT Calculations of the Chromophore Models

The free chromophore models shown in Figure S1 were used to simulate the Raman spectra. As illustrated in Figure S2, the calculated spectra are sensitive to the position of the deprotonated pyrrole NH group. However, the agreement between the observed and computed spectra was poor.

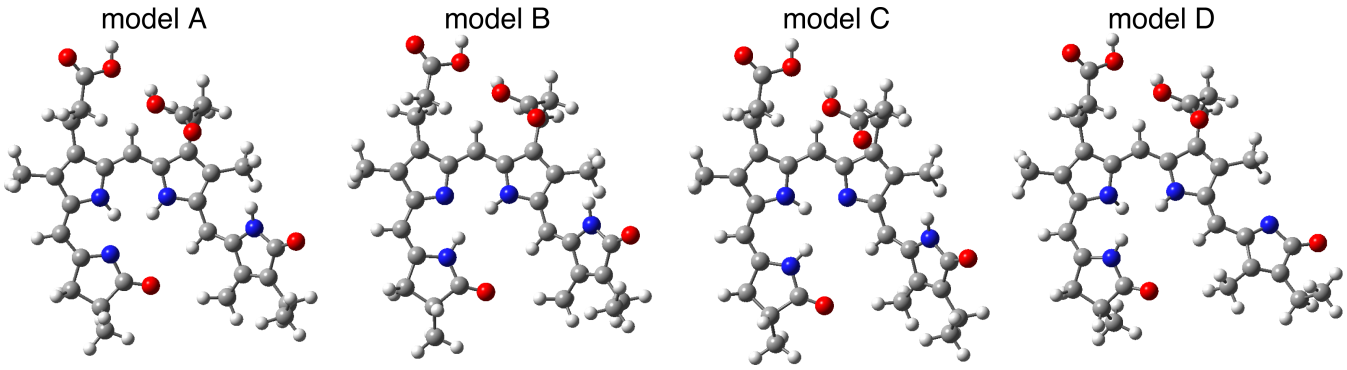

Figure S1. Optimized structures for the active site models of RcaE (^{15Z}Pg). The geometry optimization was performed at the B3LYP/6-31G* level of theory.

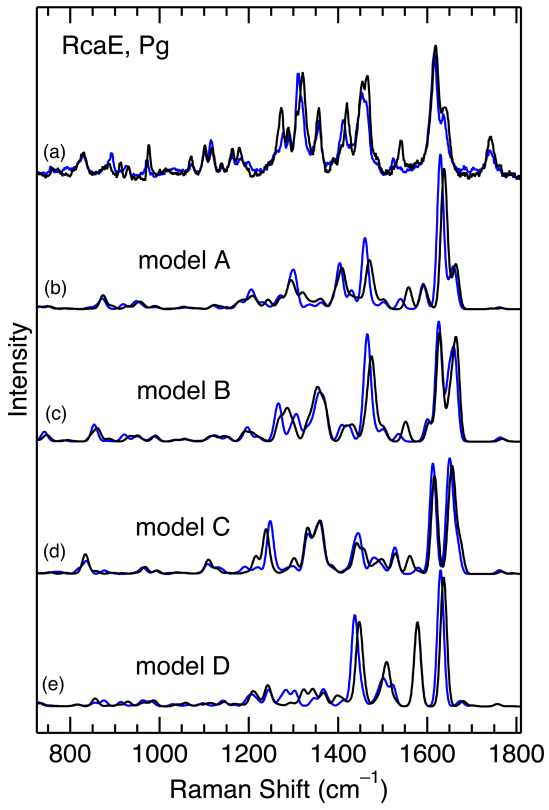

Figure S2. Observed and calculated Raman spectra of RcaE and its models. (a) The observed Raman spectra of ^{15Z}Pg under illumination with a red LED light at pH 8.0 with 441.6 nm excitation. (b–e) The Raman spectra determined for models A, B, C, and D, which are illustrated in Figure S1. The force field (frequency) calculations were performed with B3LYP/6-31G*, while the Raman intensities were obtained at the CAM-B3LYP/6-31G* level of theory.

MD Calculations

The MD trajectories for models Pg-B and Pg-C were analyzed, and the results are summarized in Figures S3–S6. In both models, the overall protein structures are gradually changed during the 50 ns MD runs (Figure S4). The stability of the active site structures is illustrated in Figures S5 and S6.

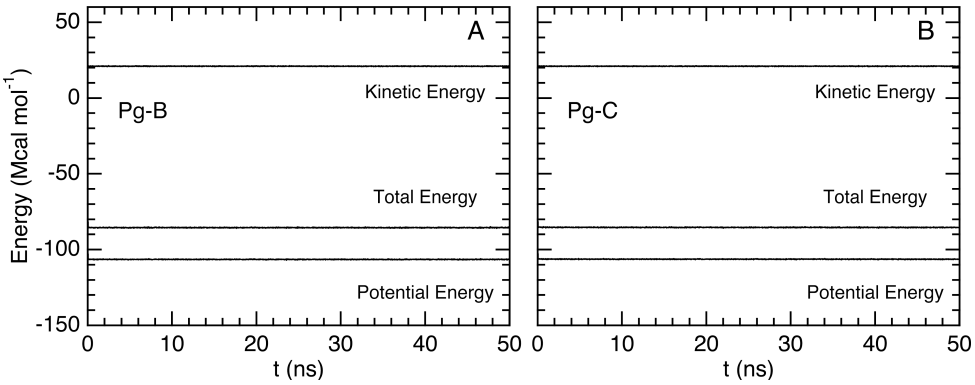

Figure S3. The time dependences of the kinetic energies, potential energies, and total energies of models Pg-B and Pg-C obtained by the 50 ns MD run. (A) model Pg-B. (B) model Pg-C.

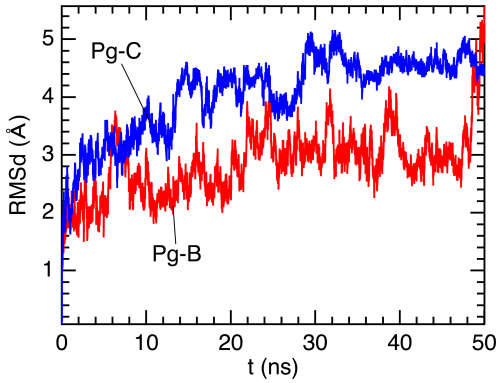

Figure S4. The time dependences of root-mean-square deviation (RMSd) for models Pg-B and Pg-C obtained by the 50 ns MD run. The RMSd of the protein backbone atoms (N, Cα, and C) was obtained from the initial structure.

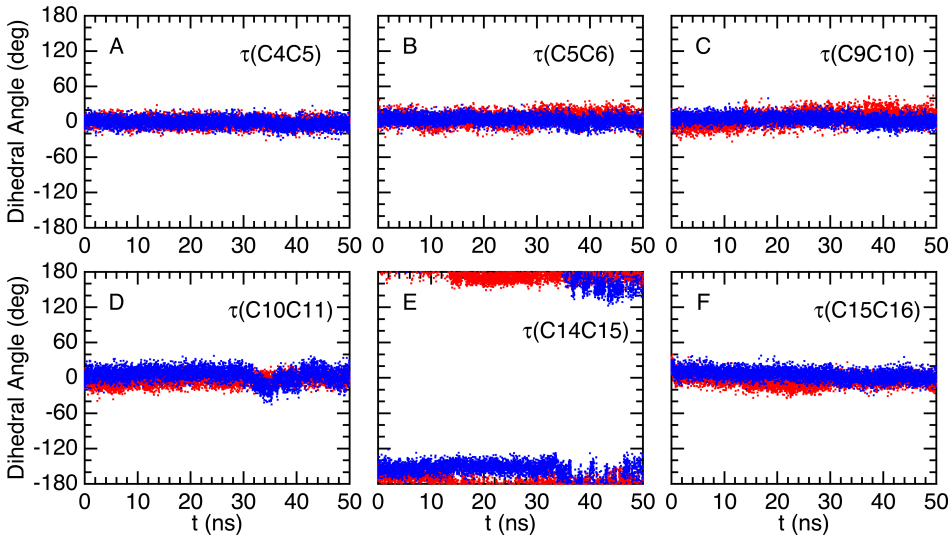

Figure S5. Timelines depicting the dihedral angles around the methine-bridges of the PCB chromophore. The dihedral angles for models Pg-B and Pg-C are illustrated with red and blue dots, respectively. The dihedral angles are defined by the following atoms: (A) NA–C4–C5–C6, (B) C4–C5–C6–NB, (C) NB–C9–C10–C11, (D) C9–C10–C11–NC, (E) NC–C14–C15–C16, and (F) C14–C15–C16–ND.

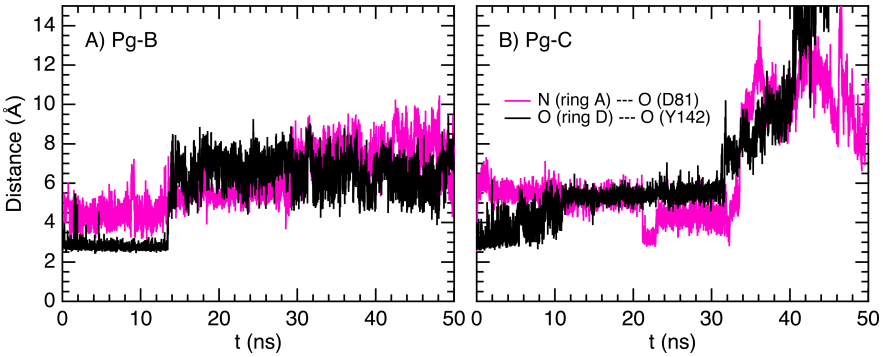

Figure S6. The time dependences of the distances characterizing the chromophore–protein interactions for models (A) Pg-B and (B) Pg-C. The distance between the pyrrole N (ring A) and the side chain O of Asp81 is displayed in magenta, while the black line represents the distance between the carbonyl O (ring D) and the side chain O of Tyr142.

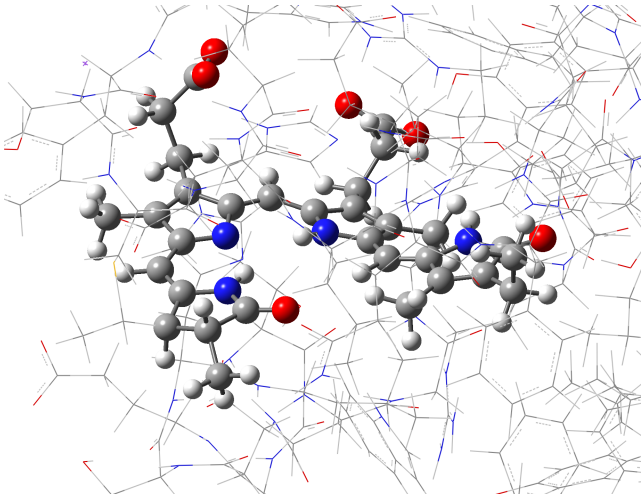

Figure S7. Atoms included in the QM region for model Pg-B. Ball and stick model represents the QM region, while the MM region is displayed by wireframe model. This figure shows the optimized structure based on an MD snapshot at 7.2 ns with electronic embedding scheme.

Structural Parameters for Models

Table S1. Selected structural parameters for the PCB chromophore models.

model	A	B	C	D	Pg-B (EE) [a]	Pg-B/C (EE)[b]	Pg-B (ME) [c]	Pg-C (EE) [d]
				Dihedral Angle (deg)				
NA–C4–C5–C6	–0.9	–1.6	2.6	5.0	0.8	2.9	0.8	–1.2
C4–C5–C6–NB	1.3	–5.4	10.6	22.8	3.1	15.7	3.1	11.7
NB–C9–C10–C11	–3.6	–4.1	–2.6	–10.0	–4.3	–10.4	–4.3	3.5
C9–C10–C11–NC	–2.9	–0.9	2.6	–1.3	–13.0	–8.8	–13.0	3.6
NC–C14–C15–C16	–149.1	–147.3	–149.4	171.9	–159.9	–161.7	–159.9	–148.0
C14–C15–C16–ND	3.8	4.0	2.5	–9.5	0.3	–1.1	0.3	10.4
				Bond Length (Å)				
NA–C4	1.342	1.382	1.394	1.397	1.379	1.390	1.400	1.403
C4–C5	1.405	1.358	1.356	1.352	1.365	1.358	1.360	1.356
C5–C6	1.402	1.445	1.438	1.447	1.445	1.446	1.456	1.450
C6–NB	1.358	1.338	1.361	1.398	1.341	1.373	1.352	1.368
NB–C9	1.381	1.390	1.376	1.397	1.392	1.386	1.382	1.388
C6–C7	1.455	1.466	1.423	1.395	1.470	1.417	1.451	1.416
C7–C8	1.379	1.371	1.405	1.423	1.370	1.411	1.383	1.413
C8–C9	1.452	1.458	1.416	1.399	1.467	1.417	1.460	1.422
C9–C10	1.374	1.377	1.406	1.423	1.376	1.409	1.389	1.420
C10–C11	1.418	1.414	1.384	1.371	1.424	1.387	1.415	1.387
C11–NC	1.383	1.374	1.387	1.391	1.369	1.382	1.374	1.396
NC–C14	1.374	1.369	1.338	1.393	1.379	1.343	1.382	1.332
C11–C12	1.414	1.412	1.450	1.448	1.419	1.461	1.429	1.464
C12–C13	1.411	1.412	1.376	1.376	1.412	1.376	1,410	1.379
C13–C14	1.413	1.417	1.466	1.464	1.420	1.468	1.422	1.466
C14–C15	1.443	1.445	1.449	1.386	1.441	1.450	1.434	1.457
C15–C16	1.360	1.359	1.360	1.423	1.367	1.367	1.374	1.358
C16–ND	1.391	1.391	1.391	1.317	1.402	1.402	1.402	1.400
C16–C17	1.476	1.476	1.477	1.519	1.469	1.473	1.463	1.472
C17–C18	1.359	1.358	1.358	1.344	1.365	1.365	1.368	1.361
C18–C19	1.489	1.490	1.490	1.517	1.474	1.474	1.471	1.484

[a] Model Pg-B based on an MD snapshot at 7.2 ns. An electronic embedding (EE) scheme was used. [b] Model Pg-B/C in which the ring C pyrrole NH moiety is deprotonated, while the protein structure is based on an MD snapshot for model Pg-B at 7.2 ns. An electronic embedding (EE) scheme was used. [c] Model Pg-B based on an MD snapshot at 7.2 ns. A mechanical embedding (ME) scheme was used. [d] Model Pg-C based on an MD snapshot at 7.6 ns. An EE scheme was used.

MD + QM/MM Calculations: Effects of Spectral Averaging

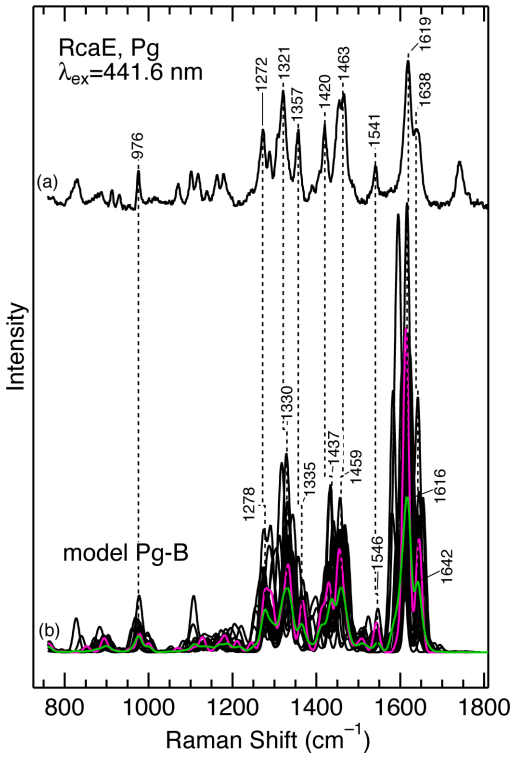

Figure S8. Observed and calculated Raman spectra of RcaE (Pg). (a) Experimental Raman spectrum. (b) Calculated Raman spectra based on the MD + QM/MM method. The averaged spectrum is shown in green, and the calculated spectra for 26 conformations are shown in black. The spectrum for the structure based on an MD snapshot at 7.2 ns is shown in magenta.

Assignment of the Raman Spectra of ^{15Z}Pg State of RcaE

Table S2. Observed and calculated vibrational frequencies (cm^{-1}) of the Pg state of RcaE and model Pg-B.

ν_{obs}[a]	$\nu_{cal,ave}$[b]	$\nu_{cal,7.2ns}$[c]	Assignment[d]
1742	1751	1763	νC=O (ring D)
1639	1642	1645	νC15=C16, νC4=C5
1619	1616	1614	νC9=C10, νC10=C11 (B–C methine bridge)
1541	1546	1543	δNC–H
1465	1459	1463	νC11=NC, νC13=C14 (ring C)
1453		1453	νC11=C12, νC14=C15
1420	1437	1431	νC6=NB, δC–H
1357	1365	1368	δC10–H, δC–H (ring B)
1321	1330	1332	δC–H, δN–H
1272	1278	1280	δC–H, δN–H
976	976	978	CH_3 deform. (ring C)

[a] Observed vibrational frequencies. [b] Calculated vibrational frequencies of the averaged spectrum. [c] Calculated vibrational frequencies of model Pg-B based on an MD snapshot at 7.2 ns. [d] Approximate descriptions of the calculated normal modes. Abbreviations: ν, stretching; δ, bending.

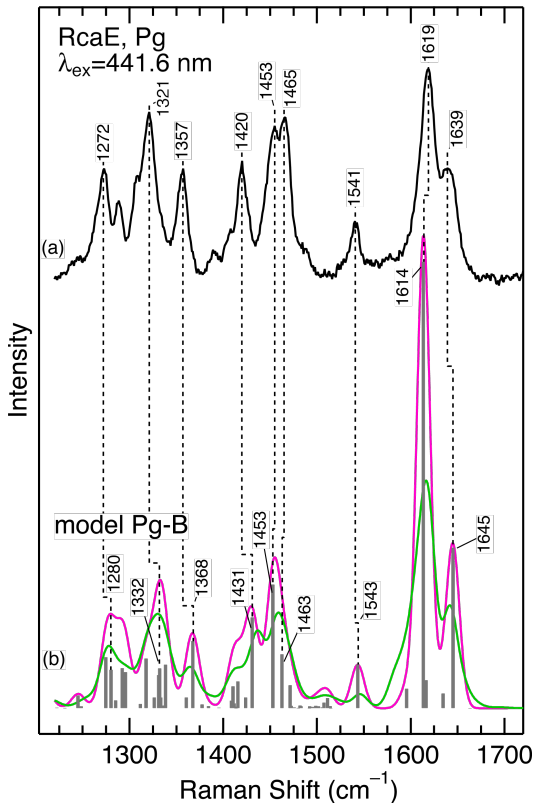

Figure S9. High frequency region of the observed and calculated Raman spectra of the ^{15Z}Pg state of RcaE and model Pg-B. (a) The observed Raman spectrum. (b) The calculated Raman spectra. The averaged spectrum is illustrated with green, whereas the spectrum based on an MD snapshot at 7.2 ns is shown in magenta. Vertical bars (grey) represent computed frequencies and intensities. Gaussian band shapes with a 10 cm^{-1} width are used to simulate the spectra.

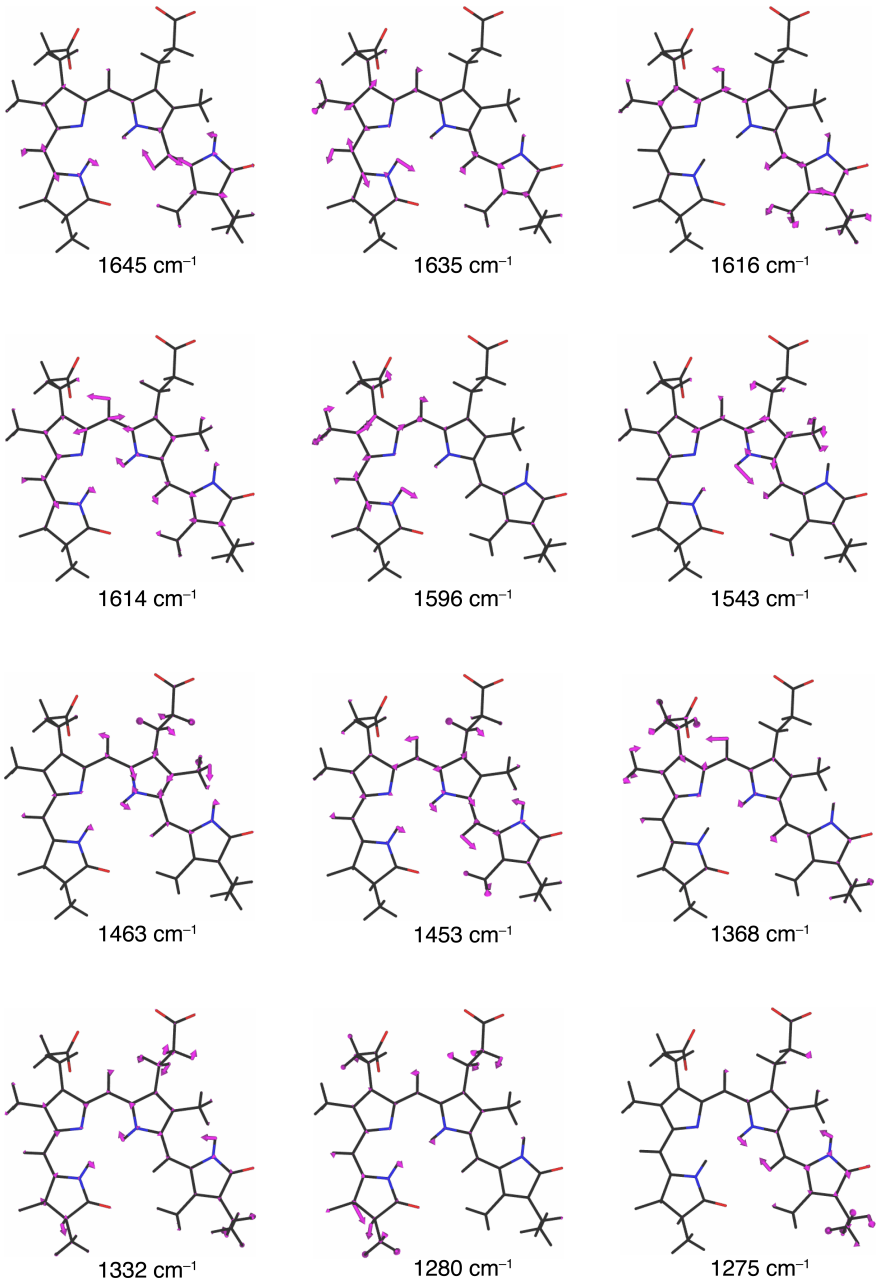

Figure S10. Atomic displacement vectors for selected vibrational modes of the PCB chromophore in model Pg-B.

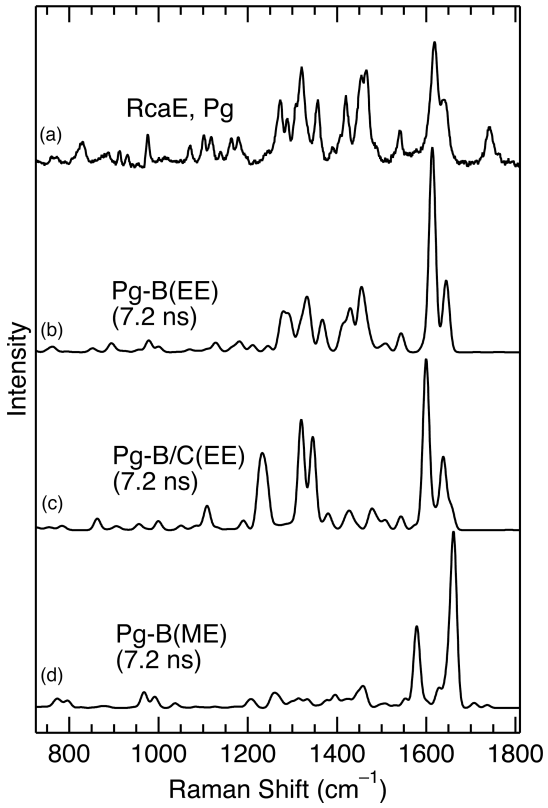

Figure S11. Observed and calculated Raman spectra of RcaE and its models. (a) The observed Raman spectrum of the ^{15Z}Pg state of RcaE. (b, c) The calculated Raman spectra based on an MD snapshot at 7.2 ns with electronic embedding (EE) and mechanical embedding (ME) schemes.

MD + QM/MM Analysis of the 15ZPr State of AnPixJ

The initial starting geometry for model Pr, which contains a protonated *15Z*-PCB chromophore, was taken from the crystal structure of AnPixJg2 (PDB: 3W2Z). The MD calculation was performed for 30 ns at 300 K. Figures S12 shows the time dependence of the selected dihedral angles of the chromophore and the backbone RMSd. 26 snapshots of the initial 10 ns simulations were taken every 400 ps for the subsequent QM/MM calculations. In Figure S13, the simulated Raman spectra of model Pr (trace c) are compared with the observed spectra for the ^{15E}Pr state of RcaE (trace a) and ^{15E}Pr state of AnPixJg2[12] (trace b).

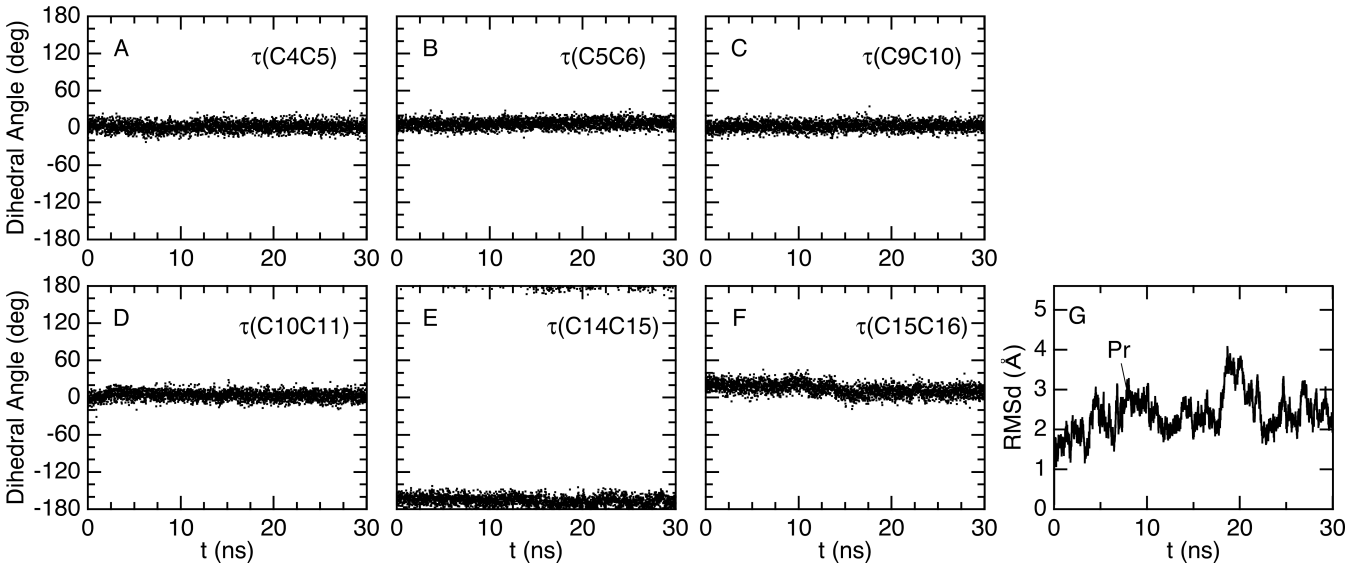

Figure S12. Timelines depicting the dihedral angles around the methine-bridges of the PCB chromophore and RMSd for model Pr obtained by the 30 ns MD run. The dihedral angles are defined by the following atoms: (A) NA–C4–C5–C6, (B) C4–C5–C6–NB, (C) NB–C9–C10–C11, (D) C9–C10–C11–NC, (E) NC–C14–C15–C16, and (F) C14–C15–C16–ND. (G) The RMSd of the protein backbone atoms (N, Cα, and C) was obtained from the initial structure.

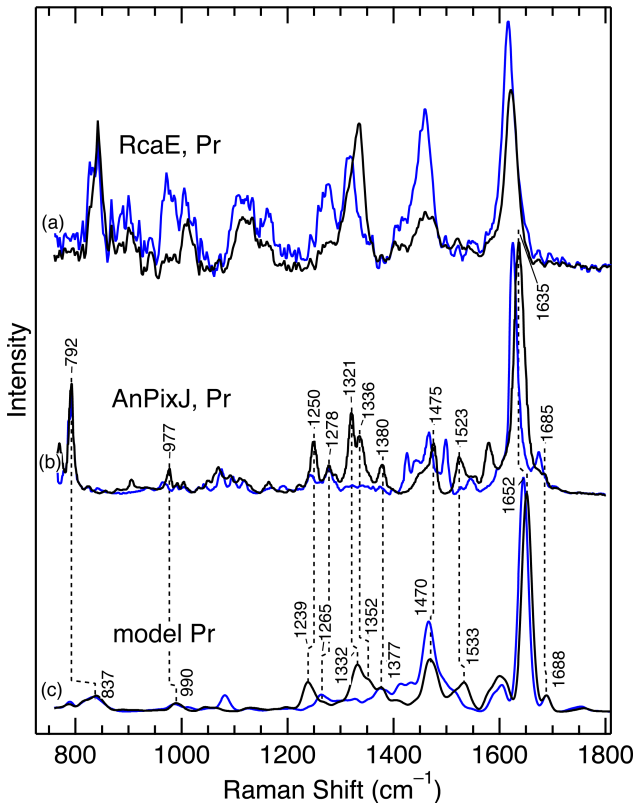

Figure S13. Observed and calculated Raman spectra of Pr. The spectra for samples or models whose exchangeable protons are deuterated are indicated with blue lines. (a) The observed Raman spectra of the ^{15E}Pr state of RcaE. (b) The observed Raman spectra of AnPixJg2, and the data are adapted from Velazquez Escobar et al.[12] (c) The simulated Raman spectra of model Pr (^{15Z}Pr for AnPixJ).

Comparison of the PCB Structures

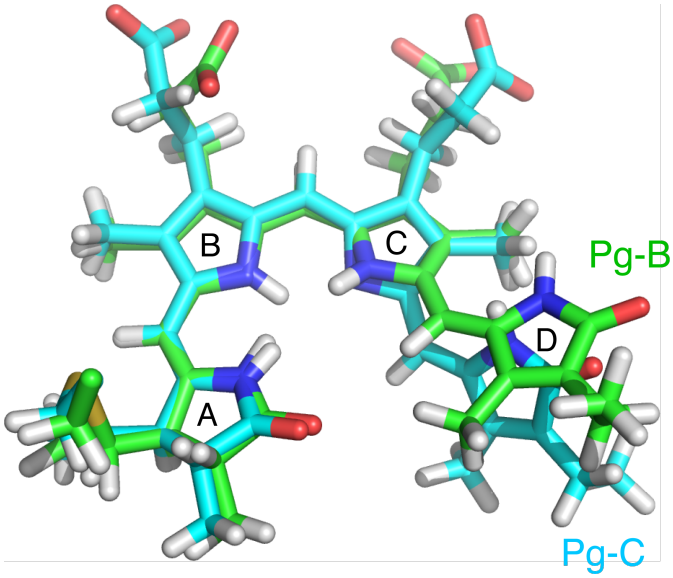

Figure S14. Superimposition of the optimized structures for the PCB chromophore. The structures based on the MD snapshots for models Pg-B (7.2 ns) and Pg-C (7.6 ns) are shown.

References

(1) Hirose, Y.; Rockwell, N. C.; Nishiyama, K.; Narikawa, R.; Ukaji, Y.; Inomata, K.; Lagarias, J. C.; Ikeuchi, M. Green/Red Cyanobacteriochromes Regulate Complementary Chromatic Acclimation via a Protochromic Photocycle. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 4974–4979.
(2) Unno, M.; Kikukawa, T.; Kumauchi, M.; Kamo, N. Exploring the Active Site Structure of a Photoreceptor Protein by Raman Optical Activity. *J. Phys. Chem. B* **2013**, *117*, 1321–1325.
(3) Kajimoto, K.; Kikukawa, T.; Nakashima, H.; Yamaryo, H.; Saito, Y.; Fujisawa, T.; Demura, M.; Unno, M. Transient Resonance Raman Spectroscopy of a Light-Driven Sodium-Ion-Pump Rhodopsin from Indibacter Alkaliphilus. *J. Phys. Chem. B* **2017**, *121*, 4431–4437.
(4) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; et al. *Amber16*; University of California: San Francisco, 2016.
(5) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
(6) Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids. 11. Transferable Intermolecular Potential Functions. Application to Liquid Methanol Including Internal Rotation. *J. Am. Chem. Soc.* **1981**, *103*, 341–345.
(7) Urago, H.; Suga, T.; Hirata, T.; Kodama, H.; Unno, M. Raman Optical Activity of a Cyclic Dipeptide Analyzed by Quantum Chemical Calculations Combined with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2014**, *118*, 6767–6774.
(8) Furuta, M.; Fujisawa, T.; Urago, H.; Eguchi, T.; Shingae, T.; Takahashi, S.; Blanch, E. W.; Unno, M. Raman Optical Activity of Tetra-Alanine in the Poly(L-Proline) II Type Peptide Conformation. *Phys. Chem. Chem. Phys.* **2017**, *19*, 2078–2086.
(9) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16*; Gaussian, Inc.: Wallingford, CT, 2016.
(10) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A New ONIOM Implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives1Dedicated to Professor Keiji Morokuma in Celebration of His 65th Birthday.1. *J. Mol. Struct.: THEOCHEM* **1999**, *461–462*, 1–21.
(11) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(12) Velazquez Escobar, F.; Utesch, T.; Narikawa, R.; Ikeuchi, M.; Mroginski, M. A.; Gärtner, W.; Hildebrandt, P. Photoconversion Mechanism of the Second GAF Domain of Cyanobacteriochrome AnPixJ and the Cofactor Structure of Its Green-Absorbing State. *Biochemistry* **2013**, *52*, 4871–4880.