Supporting Information

The Length and Charge of the N-terminus Regulate the Lifetime of the Signaling State of Photoactive Yellow Protein

Cheolhee Yang,[†,§] Youngmin Kim,[†,§] Seong Ok Kim,[†,‡] Sang Jin Lee,[†,‡] Jungkweon Choi,[†,‡] Hyotcherl Ihee[*,†,‡]

[†]Center for Advanced Reaction Dynamics, Institute for Basic Science (IBS), Daejeon 34141, Republic of Korea

[‡]Department of Chemistry and KI for the BioCentury, KAIST, Daejeon 34141, Republic of Korea

[§] These authors equally contributed.

*Corresponding author. E-mail: hyotcherl.ihee@kaist.ac.kr

Supporting Note 1

Difference between the sequential and parallel models

The kinetics of the PYP photocycle were extensively studied, but a full consensus has not been reached. Especially, regarding the appearance of pR states (red-shifted intermediates of PYP), two different kinetic frameworks were suggested: the formations of pR states occur in (i) a sequential[7-8] (Figure S1) manner or in a parallel manner[9-13] (Figure 1). In the recent TA study of PYP, the sequential formation of pR_1 and pR_2 states was favored, although the parallel scheme cannot be completely ruled out.[14]

To understand the difference between the sequential and parallel models for the formations of pR_1 and pR_2 states, reaction rate equations for each model are explained here. The equilibria in both models are not considered in this stage to simplify the models. In the case of the sequential model (Figure S1), the reaction rate differential equations in the time region from pR_1 existence to pB_2 appearance can be expressed as follows:

$$\begin{aligned}
\frac{d[pR_1]}{dt} &= -k_1[pR_1] \\
\frac{d[pR_2]}{dt} &= k_1[pR_1] - k_2[pR_2] \\
\frac{d[pB_1]}{dt} &= k_2[pR_2] - k_3[pB_1] \\
\frac{d[pB_2]}{dt} &= k_3[pB_1]
\end{aligned} \tag{S1}$$

where $[pR_1]$, $[pR_2]$, $[pB_1]$ and $[pB_2]$ are the concentrations of pR_1, pR_2, pB_1 and pB_2 intermediates, respectively, k_1, k_2 and k_3 are the decay rate constants for pR_1, pR_2 and pB_1, respectively. By solving the differential equations, the concentrations of the intermediates are expressed as follows:

$$\begin{aligned}
[pR_1] &= [pR_1]_0 \exp(-k_1 t) \\
[pR_2] &= -\frac{k_1}{(k_1 - k_2)}[pR_1]_0[\exp(-k_1 t) - \exp(-k_2 t)] \\
[pB_1] &= k_1 k_2 [pR_1]_0 \left[\frac{\exp(-k_1 t)}{(k_1 - k_2)(k_1 - k_3)} - \frac{\exp(-k_2 t)}{(k_1 - k_2)(k_2 - k_3)} + \frac{\exp(-k_3 t)}{(k_1 - k_3)(k_2 - k_3)}\right] \\
[pB_2] &= [pR_1]_0 - k_1 k_2 k_3 [pR_1]_0 \left[\frac{\exp(-k_1 t)}{k_1(k_1 - k_2)(k_1 - k_3)} - \frac{\exp(-k_2 t)}{k_2(k_1 - k_2)(k_2 - k_3)}\right. \\
&\qquad \left. + \frac{\exp(-k_3 t)}{k_3(k_1 - k_3)(k_2 - k_3)}\right]
\end{aligned} \tag{S2}$$

where $[pR_1]_0$ is the initial concentration of pR_1 intermediates. On the other hand, in the case of the parallel model (Figure 1), the reaction rate differential equations in the time region from pR_1 and pR_2 existences to pB_2 appearance can be expressed as follows:

$$\frac{d[pR_1]}{dt} = -k_1[pR_1]$$
$$\frac{d[pR_2]}{dt} = -k_2[pR_2]$$
$$\frac{d[pB_1]}{dt} = k_1[pR_1] + k_2[pR_2] - k_3[pB_1] \tag{S3}$$
$$\frac{d[pB_2]}{dt} = k_3[pB_1]$$

By solving Equation S3, the concentrations of the intermediates are expressed as follows:

$$[pR_1] = [pR_1]_0 \exp(-k_1 t)$$
$$[pR_2] = \left(\frac{1}{f} - 1\right)[pR_1]_0 \exp(-k_2 t)$$
$$[pB_1] = -\frac{k_1}{(k_1 - k_3)}[pR_1]_0[\exp(-k_1 t) - \exp(-k_3 t)]$$
$$-\frac{k_2}{(k_2 - k_3)}\left(\frac{1}{f} - 1\right)[pR_1]_0[\exp(-k_2 t) - \exp(-k_3 t)] \tag{S4}$$
$$[pB_2] = \frac{1}{f}[pR_1]_0 - \frac{k_1 k_3}{(k_1 - k_3)}[pR_1]_0\left[\frac{\exp(-k_1 t)}{k_1} - \frac{\exp(-k_3 t)}{k_3}\right]$$
$$-\frac{k_2 k_3}{(k_2 - k_3)}\left(\frac{1}{f} - 1\right)[pR_1]_0\left[\frac{\exp(-k_2 t)}{k_2} - \frac{\exp(-k_3 t)}{k_3}\right]$$

where f is the concentration ratio of pR_1 to the total concentration of the red-shifted intermediates ($[pR_1]_0$+$[pR_2]_0$, where $[pR_2]_0$ is the initial concentration of pR_2). From Beer-Lambert law, the absorbance of a molecule is linearly proportional to the concentration of the molecule, $A = \varepsilon \cdot l \cdot c$, where A is the absorbance, ε is the extinction coefficient, l is the path length of the container and c is the concentration of the molecule. Consequently, the absorbance difference, ΔA in TA spectra, is also linearly proportional to the concentration of the corresponding intermediate. The total ΔA is expressed as follows:

$$\Delta A_{total} = \left(\Delta A_{pR_1} + \Delta A_{pR_2} + \Delta A_{pB_1} + \Delta A_{pB_2}\right)$$
$$= l(\delta\varepsilon_1[pR_1] + \delta\varepsilon_2[pR_2] + \delta\varepsilon_3[pB_1] + \delta\varepsilon_4[pB_2]) \tag{S5}$$

where ΔAs are the difference of the absorbance corresponding to intermediates, $\delta\varepsilon_1$, $\delta\varepsilon_2$, $\delta\varepsilon_3$ and $\delta\varepsilon_4$ are the difference of the extinction coefficients between intermediates and pG state. By substituting each concentration term, Equation S5 can be used for fitting the experimental data. As shown in Equations S2 and S4, it is obvious that the equations corresponding to both of the sequential and parallel models can be expressed as the simple three-exponential series with the rate constants k_1, k_2 and k_3. Hence, the rate constants for the two models can be determined by fitting the data with a three-exponential series. Each rate constant technically has the same values in the best fitting results for each model.

Consequently, the experimental data can be fitted by both the sequential and the parallel models, but the pre-exponential factors can be different. In this study, we held open the possibility that both the sequential and parallel models for the pR_1 and pR_2 formation can be correct in our experimental conditions and only focused on the change of the rate constants depending on the length of the N-terminal extension.

Supporting Note 2

Theoretical Background of Transient Grating Technique

In a TG experiment, a sample is excited by spatial modulation of light intensity, i.e. transient grating, produced by the interference of two excitation light waves. The refractive index (δn) and the absorbance (δk) changes are induced by the spatially modulated light intensity. The TG signal (I_{TG}) is proportional to the sum of the square of the δn (phase grating) and the square of the δk (amplitude grating) as follows[15-16] :

$$I_{TG} = \alpha[\delta n(t)]^2 + \beta[\delta k(t)]^2 \tag{S6}$$

where α and β are constants determined by the experimental conditions. The refractive index (δn) change mainly comes from the thermal energy releasing (δn_{th}: thermal grating) and generated (or depleted) chemical species (δn_{spe}: species grating). On the other hand, the amplitude grating (δk_{spe}) arises from chemical species that absorb the probe light. The δn_{th} is caused by the thermal relaxation from the excited states and the enthalpy changes by the reaction. Hence, the TG signal can be expressed as follows:

$$I_{TG} = \alpha\left[\delta n_{th}(t) + \sum \delta n_{spe}(t)\right]^2 + \beta\left[\sum \delta k_{spe}(t)\right]^2 \tag{S7}$$

When chemical species possess no absorption at the probe wavelength, the amplitude grating term (δk_{spe}) caused by a chemical species can be neglected. It is known that the chemical species participating in the photoreaction of PYP do not exhibit absorption at the probe wavelength of 835 nm. Accordingly, we exclude the contribution of the amplitude grating term in the TG signals for NE-PYPs. Therefore, the TG signal can be expressed as follows:

$$I_{TG} = \alpha\left[\delta n_{th}(t) + \sum \delta n_{spe}(t)\right]^2 \tag{S8}$$

If the excited state decay and the reaction are fast enough to release the thermal energy, the temporal profile of the thermal grating is given by

$$\delta n_{th}(t) = \delta n_{th}\exp(-D_{th}q^2 t) \tag{S9}$$

where D_{th} is the thermal diffusivity. A rate constant, k_{th}, equal to $D_{th}q^2$, allows us to readily extract the contribution of the thermal grating. On one hand, the δn_{spe} decreases with time by the rates

of not only the reaction but also the diffusion of the molecular species. Hence, the species gratings decay with the reaction rate constant, k and/or the diffusion rate constant, $D_{spe}q^2$, where D_{spe} is the diffusion coefficient of the molecular species. The δn_{spe} for the i species is given by the following equation.

$$\delta n_{spe}(t) = \delta n_i \exp(-D_i q^2 t) \tag{S10}$$

Supporting Note 3

Stokes-Einstein Relation

Stokes-Einstein relation for the translational diffusion coefficient of the molecule in solution is expressed in terms of the viscosity of the solution and the size of the molecule. For a spherical particle, the diffusion coefficient of the molecule can be estimated by using the Stokes-Einstein relation, which is expressed as follows:

$$D = \frac{k_B T}{6\pi\eta r} \tag{S11}$$

where D is the translational diffusion coefficient of the spherical particle, k_B is Boltzmann's constant, T is the absolute temperature, η is the viscosity of the solution, and r is the radius of the spherical particle. From Equation S11, the relationship between the diffusion coefficient and the molecular weight of the particle can be expressed as follows:

$$D \propto \frac{1}{r} \propto \frac{1}{M^{1/3}} \tag{S12}$$

where M is the molecular weight of the protein when the particle has a uniform density. Therefore, the ratio of the diffusion coefficients of two different molecules can be expressed by using the ratio of the molecular weights as follows:

$$\frac{D_1}{D_2} = \frac{M_2^{1/3}}{M_1^{1/3}} \tag{S13}$$

where D_1 and D_2 are the diffusion coefficients of molecules 1 and 2, respectively and M_1 and M_2 are the molecular weights of molecules 1 and 2, respectively. If the conformation of a specific state is similar between molecules 1 and 2, the diffusion coefficient of the state can be estimated by using Equation S13.

Supporting Note 4

Additional interaction between the N-terminus and β-sheets in 5D-PYP

To investigate the additional interaction between the N-terminus and β-sheets in 5D-PYP, we calculated the surface charge distribution for the pG and pB_2 states (details are described in the Methods.). The distribution for the pG state was calculated using the crystal structure of the pG state (PDB ID: 2PHY), while that for the pB_2 state was calculated using the NMR structure of the pB_2 state (PDB ID: 2KX6). The surface charge distributions of the pG and pB_2 states exhibit positive charge within the β-sheets, while showing negative charge in the rest of the structures (Figure S4). Considering that the side chain of the five aspartic acid residues are negatively charged in our experimental condition, the positive charge within the β-sheets of the pG and pB_2 states indicates the possibility that the extended N-terminus could potentially form additional electrostatic interactions with the β-sheets in 5D-PYP.

To further investigate the interaction between the N-terminus and β-sheets, we performed MD simulations for the following four constructs (details are described in Methods): (i) pG state of wt-PYP, (ii) pB_2 state of wt-PYP, (iii) pG state of 5D-PYP and (iv) pB_2 state of 5D-PYP. Using MD snapshots from the simulation, we calculated root mean square fluctuation (RMSF) of protein moiety and the distribution for the center of mass (COM) distance between the N-terminus and the β-sheets for each construct (Figure S5). For the pG state, the RMSF plot of 5D-PYP shows decreased values compared to those of wt-PYP in the N-terminal region, while wt-PYP and 5D-PYP exhibit similar distribution of the distances between the N-terminus and the β-sheets. The comparison of these distributions only shows a slight increase of 0.03 nm in the distribution of 5D-PYP compared to that of wt-PYP, relative to the mean value of the distributions (Figures S5a and S5b). These results suggest that the addition of the five Asp residues to PYP in the pG state reduces the structural flexibility of the N-terminus without mediating further interactions between the N-terminus and the β-sheets. For the pB_2 state, the RMSF plot of 5D-PYP displays decreased values compared to those of wt-PYP in the N-terminal region, similar to the case in the pG state (Figure S5c). On the other hand, the distribution of pB_2 significantly decreased than that of wt-PYP (Figure S5d). This change of the distribution is approximately ten times larger than that in the pG state, with a magnitude of 0.33 nm, relative to the mean value of the distributions (Figures S5b and S5d). These results demonstrate that the addition of five Asp residues to PYP in the pB_2 state reduces the structural flexibility of the N-terminus and mediates further interactions between the N-terminus and the β-sheets, unlike in the case of pG state.

Together with the surface charge distributions, the results of MD simulations suggest the following mechanistic insight into the interaction between the N-terminus and the β-sheets in 5D-PYP. In the pG state, the N-terminus blocks the β-sheets from interacting with other residues as observed in the crystal structure of the state (PDB ID: 2PHY). When 5D-PYP is illuminated and the pB_2 state is formed, the N-terminus of the protein becomes partially unfolded and protruded as observed in the previous

structural studies (12, 18), enabling the β-sheets to form interactions with other residues of the protein. Considering that the side chain of the five aspartic acid residues are negatively charged and the protein in the pB_2 state carries positive charge within the β-sheets in our experimental conditions, the significant decrease in the mean value, as observed when comparing the distributions of wt-PYP and 5D-PYP in the pB_2 state (Figure S5d), indicates the possibility that the extended N-terminus could potentially form additional electrostatic interactions with the β-sheets in 5D-PYP. Such interactions in 5D-PYP are expected to interfere with the transition from pB_2 to pG, potentially delaying the dark recovery of the protein as observed in our TG experiments.

Table S1. The calculated contents of the secondary structure element of wt-PYP, 5aa-PYP, and 17aa-PYP. The CDSSTR method in the DichroWeb program[17] was used to perform the calculations, with reference data set 4 included in the program. As the length of the N-terminal expansion increases, a slight variation can be observed in the content of the secondary structure, which can be attributed to the expansion of the N-terminus. Nevertheless, the overall secondary structural content appears to remain unchanged.

	Helix1	**Helix2**	**Strand1**	**Strand2**	**Turns**	**Unordered**	**Total**
wt-PYP	0.17	0.10	0.14	0.10	0.21	0.28	1
5aa-pyp	0.17	0.11	0.14	0.09	0.19	0.3	1
17aa-PYP	0.18	0.14	0.08	0.07	0.22	0.31	1

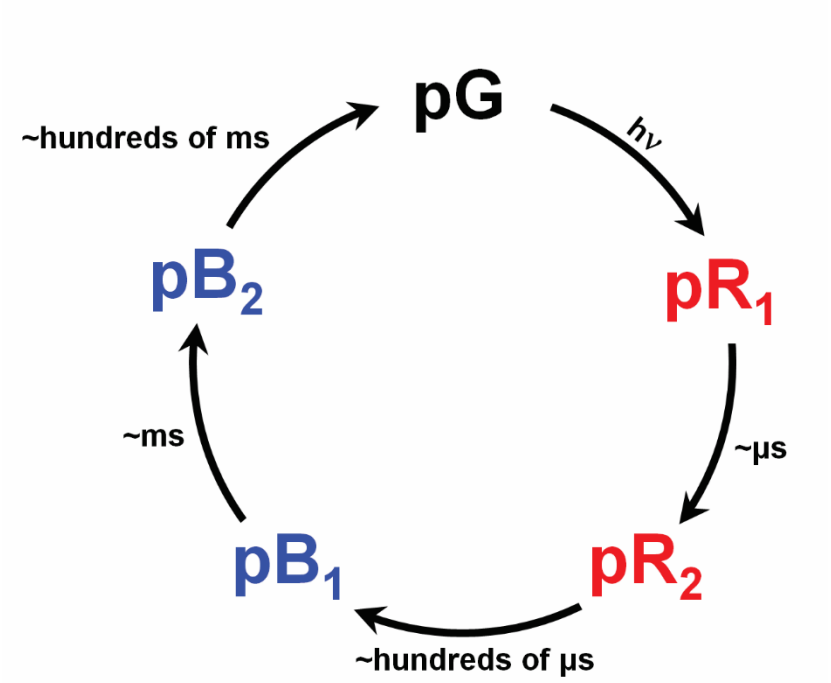

Figure S1. Schematic illustration of the photocycle of wt-PYP. Unlike the parallel model shown in Figure 1, the formation of pR_1 and pR_2 is described as the sequential model.

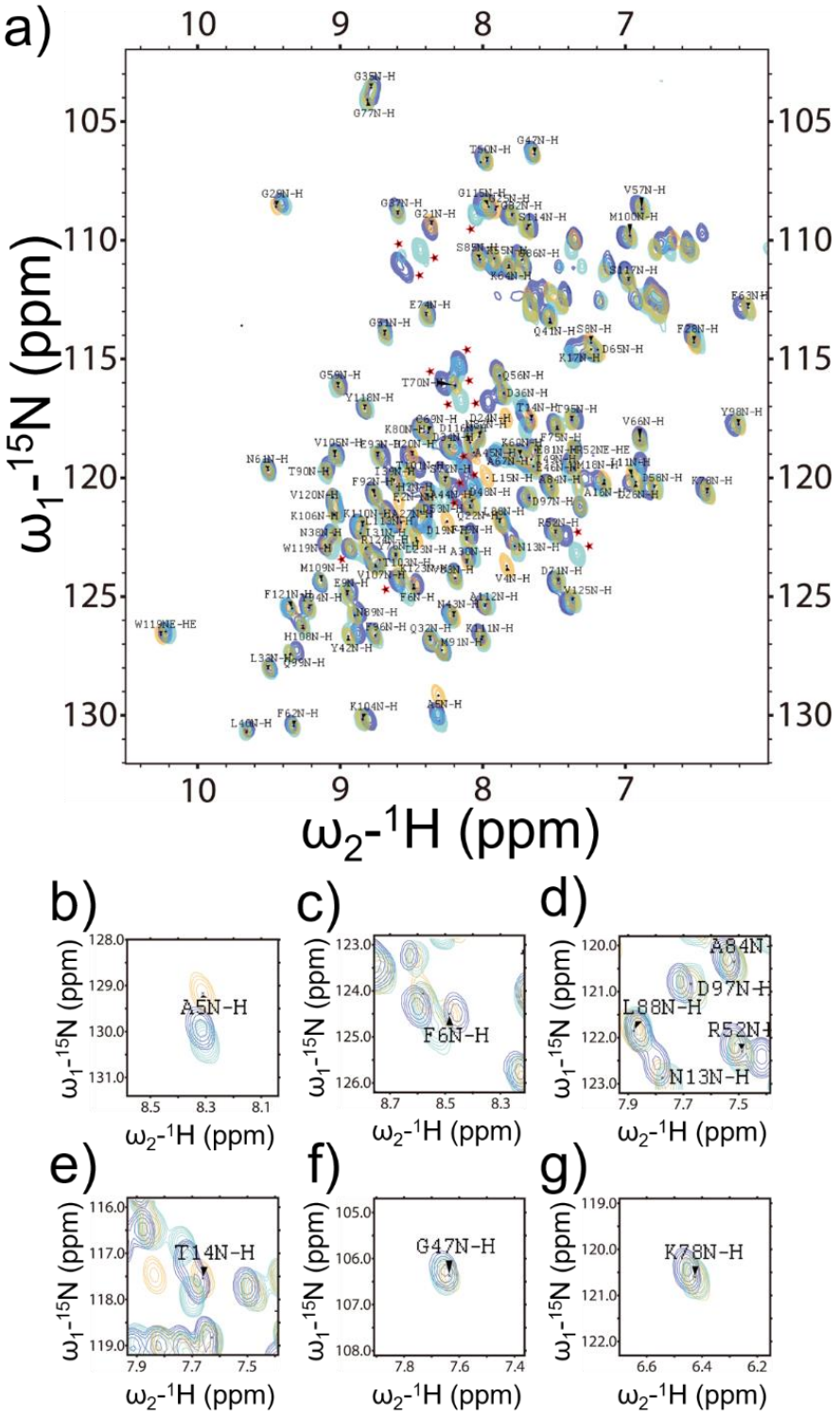

Figure S2. Overlaid ^{1}H, ^{15}N- HSQC spectra of wt-PYP (orange), 5aa-PYP (green), and 17aa-PYP (blue). (a) Spectra in the overall range. The red star symbols indicate additional peaks found in the spectra of 5aa- and 17aa-PYPs that are not observed in that of wt-PYP. (b-g) Enlarged spectra of the selective residues (A5, F6, R52-A84-L88-D97-N13, T14, G47, and K78), respectively.

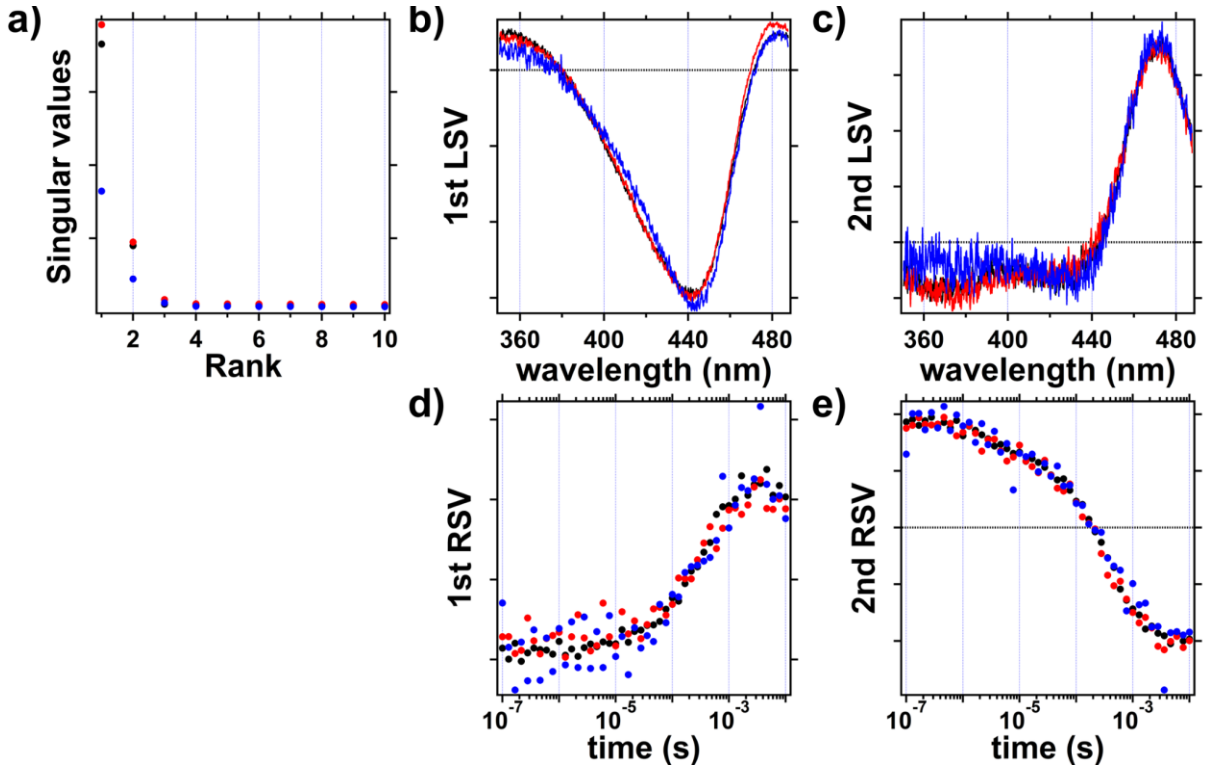

Figure S3. SVD analysis for TA spectra of wt-PYP, 5aa-PYP, and 17aa-PYP. (a) Singular values, (b) 1st left singular vectors (LSV), (c) 2nd LSV, (d) 1st right singular vectors (RSV) and (e) 2nd RSV. In each panel, black, red and blue correspond to wt-PYP, 5aa-PYP and 17aa-PYP, respectively.

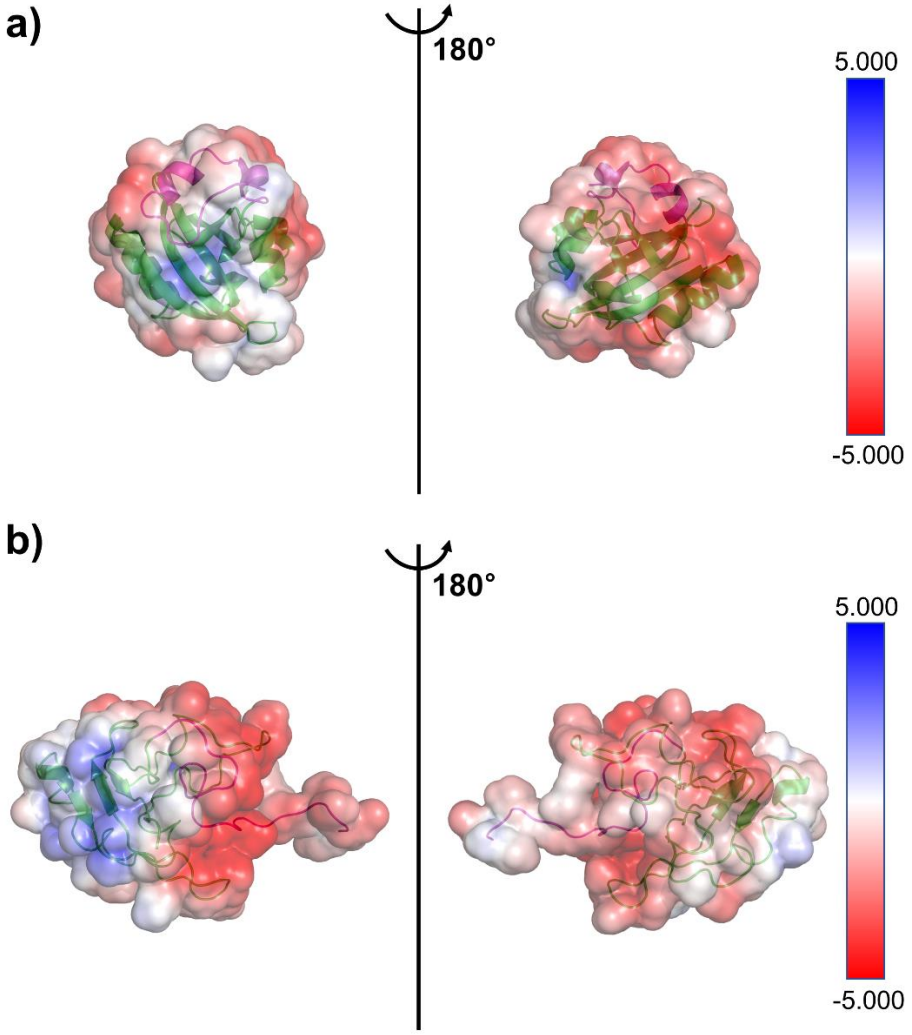

Figure S4. The surface charge distribution of PYP computed by the pdb2pqr and APBS programs. (a) The computed surface charge distribution for the pG state of PYP. The distribution, visualized in a transparent manner, is superimposed onto the crystal structure of the pG state (PDB ID: 2PHY). (b) The computed surface charge distribution for the pB_2 state of PYP. The surface charge distribution, visualized in a transparent manner, is superimposed onto the NMR structure of the pB_2 state (PDB ID: 2KX6). In (a) and (b), the N-terminal regions of crystal and NMR structures are highlighted in magenta. The surface colors are assigned red (-) or blue (+), indicating the desired values of ±5 kT/e. Here, k represents Boltzmann's constant, T is the temperature, and e is equal to $1.602176634 \times 10^{-19}$ C. During the calculations, the pdb2pqr program determines the protonation state of amino acids at pH 7.0, while APBS calculates the electrostatics.

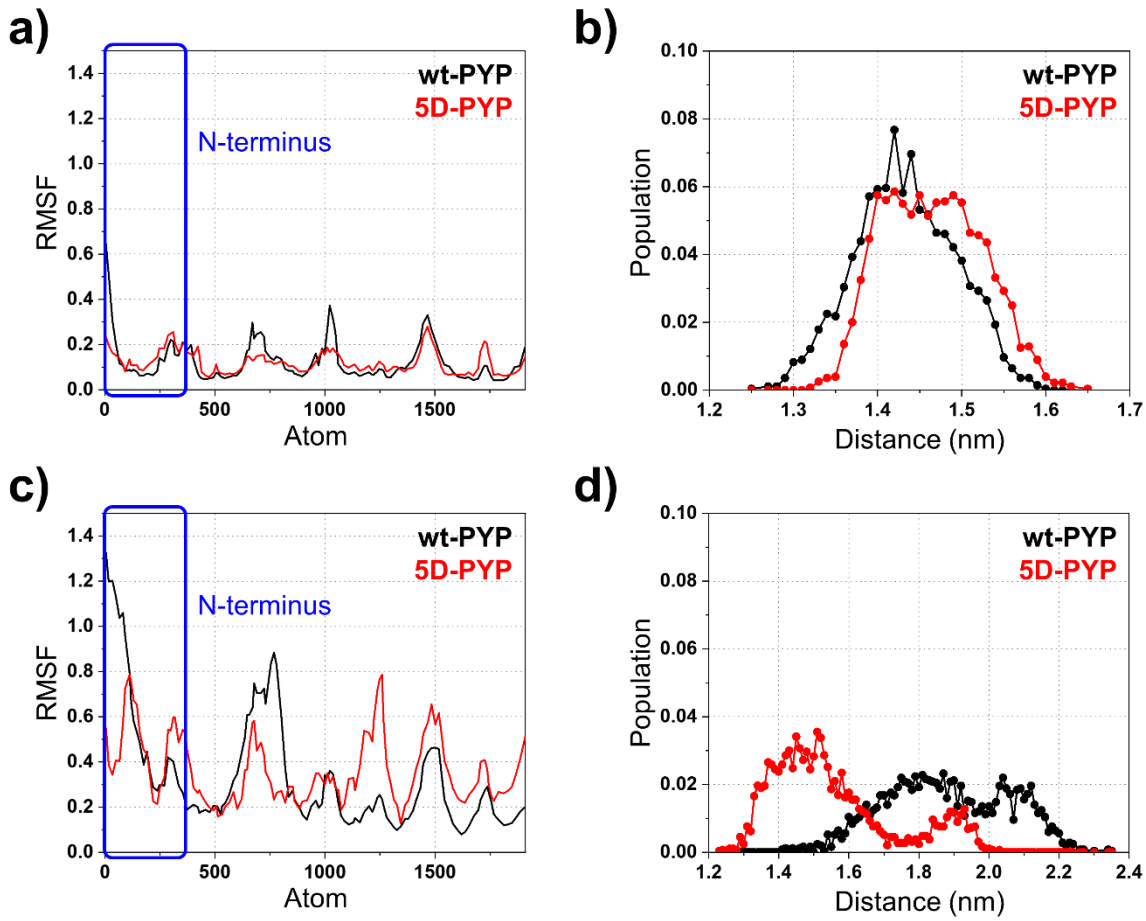

Figure S5. The results from MD simulations for the PYP constructs. (a) RMSF plot of PYP in pG state (black) and 5D-PYP in pG state (red). (b) Distributions of the distance between N-terminus and β-sheets of PYP in pG state (black) and 5D-PYP in pG state (red). The mean values of the distributions for wt-PYP and 5D-PYP are 1.43 nm and 1.46 nm, respectively. (c) RMSF plot of PYP in pB_2 state (black) and 5D-PYP in pB_2 state (red). (d) Distributions of the distance between N-terminus and β-sheets of PYP in pB_2 state (black) and 5D-PYP in pB_2 state (red). The mean values of the distributions for wt-PYP and 5D-PYP are 1.88 nm and 1.55 nm, respectively. In each RMSF plot, the region of N-terminus is indicated by a blue rectangle.

Reference

(1) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR* **1995**, *6* (3), 277-293.

(2) Goddard, T. D.; Kneller, D. G. *SPARKY 3*, University of California, San Francisco.

(3) Choi, J.; Jung, Y. O.; Lee, J. H.; Yang, C.; Kim, B.; Ihee, H. Folding Dynamics of Ferrocytochrome c in a Denaturant-Free Environment Probed by Transient Grating Spectroscopy. *ChemPhysChem* **2008**, *9* (18), 2708-2714.

(4) Choi, J.; Muniyappan, S.; Wallis, J. T.; Royer, W. E.; Ihee, H. Protein Conformational Dynamics of Homodimeric Hemoglobin Revealed by Combined Time-Resolved Spectroscopic Probes. *ChemPhysChem* **2010**, *11* (1), 109-114.

(5) Choi, J.; Yang, C.; Kim, J.; Ihee, H. Protein Folding Dynamics of Cytochrome c Seen by Transient Grating and Transient Absorption Spectroscopies. *J. Phys. Chem. B* **2011**, *115* (12), 3127-3135.

(6) Yang, C.; Choi, J.; Ihee, H. The time scale of the quaternary structural changes in hemoglobin revealed using the transient grating technique. *Phys. Chem. Chem. Phys.* **2015**, *17* (35), 22571-22575.

(7) Chen, E. F.; Gensch, T.; Gross, A. B.; Hendriks, J.; Hellingwerf, K. J.; Kliger, D. S. Dynamics of protein and chromophore structural changes in the photocycle of photoactive yellow protein monitored by time-resolved optical rotatory dispersion. *Biochemistry* **2003**, *42* (7), 2062-2071.

(8) Losi, A.; Gensch, T.; van der Horst, M. A.; Hellingwerf, K. J.; Braslavsky, S. E. Hydrogen-bond network probed by time-resolved optoacoustic spectroscopy: photoactive yellow protein and the effect of E46Q and E46A mutations. *Phys. Chem. Chem. Phys.* **2005**, *7* (10), 2229-2236.

(9) Anderson, S.; Srajer, V.; Moffat, K. Structural heterogeneity of cryotrapped intermediates in the bacterial blue light photoreceptor, photoactive yellow protein. *Photochem Photobiol* **2004**, *80* (1), 7-14.

(10) Ihee, H.; Rajagopal, S.; Srajer, V.; Pahl, R.; Anderson, S.; Schmidt, M.; Schotte, F.; Anfinrud, P. A.; Wulff, M.; Moffat, K. Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (20), 7145-7150.

(11) Rajagopal, S.; Anderson, S.; Srajer, V.; Schmidt, M.; Pahl, R.; Moffat, K. A structural pathway for signaling in the E46Q mutant of photoactive yellow protein. *Structure* **2005**, *13* (1), 55-63.

(12) Kim, T. W.; Lee, J. H.; Choi, J.; Kim, K. H.; van Wilderen, L. J.; Guerin, L.; Kim, Y.; Jung, Y. O.; Yang, C.; Kim, J.; Wulff, M.; van Thor, J. J.; Ihee, H. Protein Structural Dynamics of Photoactive Yellow Protein in Solution Revealed by Pump-Probe X-ray Solution Scattering. *J. Am. Chem. Soc.* **2012**, *134* (6), 3145-3153.

(13) Jung, Y. O.; Lee, J. H.; Kim, J.; Schmidt, M.; Moffat, K.; Srajer, V.; Ihee, H. Volume-conserving trans-cis isomerization pathways in photoactive yellow protein visualized by picosecond X-ray crystallography. *Nat. Chem.* **2013**, *5* (3), 212-220.

(14) Yeremenko, S.; van Stokkum, I. H. M.; Moffat, K.; Hellingwerf, K. J. Influence of the Crystalline State on Photoinduced Dynamics of Photoactive Yellow Protein Studied by Ultraviolet-Visible Transient Absorption Spectroscopy. *Biophys. J.* **2006**, *90* (11), 4224-4235.

(15) Terazima, M. Translational diffusion of intermediate species in solutions. *Res. Chem. Intermed.* **1997**,

23 (9), 853-901.

(16) Kuwajima, K. Water and biomolecules : physical chemistry of life phenomena. In *Biological and medical physics, biomedical engineering*, Springer: Berlin, **2009**, pp 149-172.

(17) Miles, A.J.; Ramalli, S.G.; Wallace, B.A. DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data. *Protein Sci.* **2022**, 31 (1), 37-46.

(18) Ramachandran, P. L.; Lovett, J. E.; Carl, P. J.; Cammarata, M.; Lee, J. H.; Jung, Y. O.; Ihee, H.; Timmel, C. R.; van Thor, J. J. The Short-Lived Signaling State of the Photoactive Yellow Protein Photoreceptor Revealed by Combined Structural Probes. *J. Am. Chem. Soc.* **2011,** *133* (24), 9395-9404.