Cell Reports Physical Science, Volume *2*

Supplemental information

Reversible molecular motional switch based on circular photoactive protein oligomers exhibits unexpected photo-induced contraction

Sang Jin Lee, Youngmin Kim, Tae Wu Kim, Cheolhee Yang, Kamatchi Thamilselvan, Hyeongseop Jeong, Jaekyung Hyun, and Hyotcherl Ihee

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Construction and preparation of coPYPs

To construct coPYPs, which were used to develop the molecular switch, we used two different methods. In the first method, we first generated the PYP oligomer sequences (three, four, and five PYP units). Subsequently, the split inteins were inserted in the N- and C-termini of the sequences to allow each PYP oligomer to form a circular shape. In the second method, we first generated the sequence of the circularly permuted PYP (cPYP), which is a modified PYP for inserting a target protein between the N- and C-termini of PYP[1]. For cPYP, the N- and C-termini of monomer PYP are connected by the flexible GS (glycine-serine) linkers (GGSGGSGG), and new N- and C-termini are introduced to G115 and S114[1]. For the construction of coPYP-4, we made linear trimer PYP (loPYP-3) in which three monomer PYP units are linearly connected by two flexible GS (glycine-serine) linkers (GGSGGSGG) as depicted in Figure S2B. Finally, cPYP and loPYP-3 were genetically linked by two GS linkers to generate coPYP-4 (Figure S2C). For the construction of coPYP-3 and coPYP-5, we generated dimer PYP (PYP2) and linear tetramer PYP (loPYP-4), and each protein was genetically linked to cPYP by the two GS linkers as in coPYP-4 (Figures S2B and S2C). Here, we note that the flexible GS linkers were used rather than rigid linkers to prevent steric hindrance between the N-terminals in the PYP units and the linkers. In this way, the C- and N-termini of each inserted PYP oligomer are connected by cPYP, forming the circular-shaped PYP oligomers. As a result, the circular oligomers of PYP trimers, tetramers, and pentamers were produced (Figure S1). Also, a method was devised for connecting more than two PYP oligomers by making a cPYP and split intein simultaneously, enabling the production of coPYPs of various sizes, including those that include higher-order PYP units. Since the yield of coPYPs produced by the split intein method was too low, all subsequent experiments were performed using cPYP-produced samples.

In the cPYP-based method, the PYP oligomers were generated based on the cPYP gene in the PQE80L vector. A circular PYP oligomer was constructed by inserting a PYP gene with a linker sequentially into a plasmid expressing a single cPYP gene[1] (Figure S1). The amino acid sequences of coPYPs are as follows. The dimer PYP was not circular but was in a leaner form.

Monomer PYP

MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQ
VIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKK
ALSGDSYWVFVKRV

Dimer PYP

MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQ
VIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKK
ALSGDSYWVFVKRV**GGSGGSGG**MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQL
DGDGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNT
MFEYTFDYQMTPTKVKVHMKKALSGDSYWVFVKRV

coPYP-3

GDSYWVFVKRV**GGSGGSGG**MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDG
DGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMF
EYTFDYQMTPTKVKVHMKKALSGDSYWVFVKRV**GGSGGSGG**MEHVAFGSEDIEN
TLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPC
TDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKKALSGDSYWVFVKR
V**GGSGGSGG**MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAE
GDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQETPT
KVKVHMKKAL**S**

coPYP-4

GDSYWVFVKRV**GGSGGSGG**MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDG
DGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMF
EYTFDYQMTPTKVKVHMKKALSGDSYWVFVKRV**GGSGGSGG**MEHVAFGSEDIEN
TLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPC
TDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKKALSGDSYWVFVKR
V**GGSGGSGG**MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAE
GDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPT
KVKVHMKKALSGDSYWVFVKRV**GGSGGSGG**MEHVAFGSEDIENTLAKMDDGQL
DGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFK
EGVASGNLNTMFEYTFDYQETPTKVKVHMKKAL**S**

coPYP-5

GDSYWVFVKRV**GGSGGSGG**MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDG
DGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMF
EYTFDYQMTPTKVKVHMKKALSGDSYWVFVKRV**GGSGGSGG**MEHVAFGSEDIEN
TLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPC
TDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKKALSGDSYWVFVKR
V**GGSGGSGG**MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAE
GDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPT
KVKVHMKKALSGDSYWVFVKRV**GGSGGSGG**MEHVAFGSEDIENTLAKMDDGQL
DGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFK
EGVASGNLNTMFEYTFDYQMTPTKVKVHMKKALSGDSYWVFVKRV**GGSGGSGG**
MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQ
VIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQETPTKVKVHMKKA
L**S**

Higher-order coPYP-6 (using both cPYP and split intein methods) requiring co-expression of both 1 and 2

1. IntC-N-S-His6-PYPwtx2_cPYP1

MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASNCFNGT NS
HHHHHHMEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDIT
GRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVK
VHMKKALSGDSYWVFVKRV**GGSGGSGGSGSG**MEHVAFGSEDIENTLAKMDDGQL
DGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFK
EGVASGNLNTMFEYTFDYQMTPTKVKVHMKKALSGDSYWVFVKRV**GGSGGSGGS**
GSGMEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRD
PKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQETPTKVKVHM
KKALS.

2. cPYP2(GDSYWVFVKRV)-PYPwtx3-S-IntN

GDSYWVFVKRV

MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQ
VIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKK
ALSGDSYWVFVKRV**GGSGGSGGSGSG**MEHVAFGSEDIENTLAKMDDGQLDGLAF
GAIQLDGDGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVAS
GNLNTMFEYTFDYQMTPTKVKVHMKKALSGDSYWVFVKRV**GGSGGSGGSGSG**M
EHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQVI
GKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKKAL
SGDSYWVFVKRV S
CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEY
CLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

After construction, coPYPs were over-expressed in Escherichia coli BL21 (DE3) by IPTG induction, reconstituted with p-coumaric anhydride, and purified by Ni affinity chromatography and ion-exchange chromatography. The purified proteins were dialyzed with 20 mM Tris, pH 7.0, and 20 mM NaCl. The concentrations of coPYPs were determined by UV-vis spectroscopy (UV-2550; Shimadzu, Kyoto, Japan).

MALDI/TOF, MALS, and DLS

We implemented several spectroscopic methods, matrix-assisted laser desorption ionization time of flight mass spectroscopy (MALDI/TOF), multiangle light scattering (MALS), and dynamic light scattering (DLS), to determine the molecular weight and size of each protein sample

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF). Molecular weights of coPYPs, monomer PYP, and dimer PYP were measured using a MALDI-TOF mass spectrometer (Bruker Autoflex III; Bruker Daltonics Inc., Bremen, Germany). Sinapinic acid was used as a matrix material for the measurement. The laser source for MALDI-TOF was the third harmonic of an Nd:YAG laser (355 nm), and the repetition rate was 1 Hz.

Multi-angle light scattering (MALS). SEC-MALS measurements were performed for coPYPs, monomer PYP, and dimer PYP using high-performance liquid chromatography (HPLC) (Shimadzu, Japan) connected in series to a Tosoh TSKgel GMPWxL (Tosoh Bioscience, King of Prussia, PA, USA), followed by a Photodiode Array (2996 Detector; Waters, Milford, MA, USA), a Dawn Heleos II MALS (Wyatt Technologies, Santa Barbara, CA, USA), and an Optilab T-rEX refractometer (Wyatt Technologies). For each protein, approximately 100 μg of

the sample was injected into an SEC column operating at a flow rate of 0.5 mL/min. SEC-MALS data were analyzed using ASTRA (Wyatt Technologies) to determine the molar mass (M_w) across the SEC chromatogram.

***Dynamic light scattering*.** DLS was performed for coPYPs, monomer PYP, and dimer PYP using a Zetasizer Nano ZS (Malvern Instruments, Malvern, UK). Particle size distributions were determined by DLS using Dynamics (Wyatt Technologies) or Zetasizer Nano ZS (Malvern Instruments). DLS utilizes a regularization algorithm along with the Stokes-Einstein relation to determine the particle size distribution.

Size-exclusion chromatography

Protein samples were injected into a Superdex 200 column (10/300 GL; GE Healthcare, Little Chalfont, UK) with FPLC (Medium Pressure Liquid Chromatography System, NGC Scout 10 Plus System) and run at 0.5 mL/min with an elution buffer (500 mM NaCl, 20 mM Tris, pH 7.0, or PBS). The elution profiles for the protein samples were monitored at 280 nm and 465 nm.

Transmission electron microscopy

The sample solution (5 μL) was loaded onto an EM-grid with a continuous carbon film, which was rendered hydrophilic by glow discharge. After 60 s of sample adsorption, the grid was washed three times with droplets of deionized water, followed by negative staining with 5 μL of 1% uranyl acetate solution. The excess staining solution was blotted using a piece of filter paper. Images were obtained using a TecnaiTM G2 Spirit TWIN (FEI Co., Hillsboro, OR, USA) transmission electron microscope operated at an acceleration voltage of 120 kV and were recorded using an Ultrascan 4000 CCD camera (Gatan Inc., Pleasanton, CA, USA).

Transient absorption (TA) spectroscopy

The method for TA spectroscopy was similar to those of the previous studies[2-4]. Briefly, an optical parametric oscillator (LT-2214-PC; LOTIS), which was pumped by the third harmonic wavelength of an Nd:YAG laser (NL 301G; EKSPLA), produced a 460 nm laser as a pump beam. As a probe source, a continuous wave 250 W Xe-lamp was used. The probe light and the pump laser beam were intersected at the sample position at an acute angle. The TA signal was detected using a combination of a spectrometer (SpectraPro 2300i; Princeton Instruments, Trenton, NJ, USA) and a PMT. The TA signal measured using the PMT was digitized using a

digital oscilloscope (TDS 3052B; Tektronix). To improve the signal-to-noise ratio, the average signal from 320 estimates was obtained. The concentrations of the chromophores of PYPs were adjusted to approximately 70 μM, which was calculated using an extinction coefficient of 45.5 $mM^{-1}cm^{-1}$ at 446 nm. To avoid photosaturation, the sample solution was flowed using a syringe pump and a flow cell (2 mm path length). The repetition rate of the pump beam was 1 Hz. To obtain measurements after a 1s time delay, the flow was stopped, and the repetition rate of the pump beam was adjusted to 0.5 Hz to reduce photosaturation as much as possible.

Static x-ray scattering and preliminary analysis

Static x-ray scattering data [$I(q)$ vs. q (= $(4\pi/\lambda) \sin \theta$), where q and 2θ are the scattering vector and scattering angle, respectively] were obtained at the 4C SAXS II beamline of Pohang Light Source (PLS-II) in Korea. The x-ray wavelength was 0.7336 Å (16.90 KeV). A two-dimensional charge-coupled detector (Rayonix, Evanston, IL, USA) was used, and the sample-to-detector distance was set at 1 m. The q range of the obtained data was 0.03 – 0.6 $Å^{-1}$. The concentration of coPYP-4 was adjusted to 4 mg/mL in the buffer, and the resulting solution was centrifuged to remove aggregates (at 10,000 × g, 10 min). The supernatant was used for static x-ray scattering experiments.

During the static x-ray scattering measurements for coPYP-4, the scattering profiles were collected as the protein sample solution flowed into the capillary system at a speed of 100 μL per min. For the ground state, the scattering profiles were obtained after the sample was incubated in the dark for 24 hours. For the light-induced state, the scattering profiles were obtained after the sample was incubated under light irradiation at 460 nm for 30 minutes. The capillary was 1.0 mm thick, and the x-ray exposure time was 10 or 60 s. The solution buffer contained 20 mM Tris (pH 7.0) and 20 mM NaCl.

The static x-ray scattering data were processed and analyzed using the ATSAS package (http://www.embl-hamburg.de/biosaxs/software.html)[5,6]. One-dimensional scattering data $I(q)$ as a function of q were obtained by radial averaging. The scattering intensity from the buffer solution were measured for the background and used to eliminate the contribution of the buffer in the static scattering curve of the protein sample.

Time-resolved x-ray solution scattering

Time-resolved x-ray solution scattering (TRXSS) data were collected at the ID14B beamline at Advanced Photon Source (Argonne, IL, USA) following an established experimental

protocol[7-9]. Before the experiment using coPYP-4, a control experiment was performed using monomer PYP to optimize the experimental conditions. The control experiment effectively reproduced the time constants previously reported for monomer PYP. For the experiment with coPYP-4, a 1.1 mM coPYP-4 solution dissolved in 20 mM sodium phosphate buffer with 20 mM NaCl at pH 7.0 was enclosed in a quartz capillary (Hampton Research, Aliso Viejo, CA, USA). The capillary was mounted on a linear translational stage (Parker, Charlotte, NC, USA) to move the sealed capillary back and forth periodically. During the measurement, the temperature of the sample was kept at 293 K using a Cryostream Temperature Controller (Oxford, Abingdon, UK). The coPYP-4 solution was irradiated by circularly polarized nanosecond (ns) laser pulses (1 mJ/mm^2 fluence at 460 nm) and was probed by x-ray pulses incident at well-defined time delays. A polychromatic x-ray probe pulse (15 keV) with a pulse duration of 1.5 μs was generated from the 324-bunch mode operation at Advanced Photon Source. In the TRXSS measurement, the pump laser was focused onto the sample capillary with a size much larger than that of the probe x-ray. The probe x-ray was aligned at an angle of an incidence of 90 degrees to the propagation direction of the pump laser and was focused to penetrate approximately 150 microns through the sample capillary. The pump laser and probe x-ray were precisely overlapped. Considering the fluence of the pump laser and the size of the focal point (0.5 × 0.15 mm^2), the number of photons (2.31×10^{15} photons/mm^2 × 0.5 × 0.15 mm^2) and the number of chromophores in coPYP-4 (4 × 1.1 mM × 0.5 × 0.15 × 0.15 mm^3 × 6.022×10^{23} molecules/mol) in the focal point were calculated. The ratio of the number densities between the photons and the chromophores is about 5.8 photons/chromophore, indicating that the experimental condition lies in the photo-saturated regime. Time-dependent x-ray scattering patterns were recorded using a Mar165 CCD detector (Rayonix, Evanston, IL, USA) covering a wide-angle region. The scattering pattern at a negative time-delay (–5 μs) was also collected and used as a reference for calculating time-resolved difference scattering curves. The scattering pattern at a time delay of –5 μs contained structural information for the ground state, while the data for positive time delays represented a mixture of the ground state and photoproducts. The time-resolved difference scattering curves, $\Delta S(q, t) = S(q, t) - S(q, -5\ \mu s)$, were generated by subtracting the scattering curve at –5 μs from the curves at positive time delays. The $\Delta S(q, t)$ curves provide information on the change caused by laser photoexcitation. Data acquisition was implemented at a repetition rate of 2 Hz while translating the quartz capillary containing the sample. The TRXSS curves were measured at 16 temporal delays over a wide time range from 3.16 μs to 178 ms. At each time delay, >150 scattering curves were

averaged to achieve a high signal-to-noise ratio. The TRXSS data were analyzed by applying singular value decomposition (SVD) and principal component analysis (PCA) to extract the kinetics and structural dynamics of light-activated coPYP-4.

Reversibility of coPYP-4. A Photoswitchable protein refers to a protein that can control the transition between two stable states, ground and final photoproduct, using light. In order to effectively apply such a photoswitchable protein to bio-engineering, its transition must be reversibly controlled depending on the presence or absence of light. In TRXSS data, the standard deviation is significantly smaller than the amplitude of signal intensity at each *q* point (Figure S4). These low standard deviations indicate that the repeatability and reproducibility of the TRXSS experiment was maintained high even though the TRXSS data were obtained from more than 150 repeated pump-probe measurements at the same focal points of the sample capillary. Such high repeatability and reproducibility demonstrate that during the TRXSS experiment, coPYP-4 reversibly regulates the structural changes between the ground and final photoproduct, implying that the reversibility of coPYP-4 was maintained more than 150 cycles.

Elimination of solvent heating contribution from time-resolved difference scattering curves. Time-resolved difference scattering curves involve both the contribution of the structural changes for the target protein and the contribution from the solvent heating induced by the energy transferred from the excited proteins into the solvent (Figure S5). Therefore, the contribution of the solvent heating needs to be removed from the time-resolved difference scattering curves to extract only the contributions from the structural change of the protein. To this end, we reconstructed the contribution of solvent heating as a linear combination of temperature change at constant pressure and density change at a constant temperature using a well-established method[8,9]. Subsequently, the contribution of solvent heating is eliminated from the difference scattering curves, and only the contribution for the structural changes of the protein is extracted.

Kinetic analysis of coPYP-4 using SVD and PCA

The kinetics of structural changes in coPYP-4 were determined by applying singular value decomposition (SVD). During the SVD analysis, the data matrix of TRXSS, **A**, was decomposed into left- and right singular vectors satisfying the relationship of $\mathbf{A = USV^T}$, where **U** is a time-independent q spectra called the left singular vectors (**lSVs**), **V** is a time-dependent amplitude change of **U** called the right singular vectors (**rSVs**), **S**, singular values, contains the weights of the singular vectors. The information on the scattering curves of the transient

intermediates is included in lSVs, while the population dynamics of those transient intermediates are included in rSV. Thus, we can obtain the number of structurally distinct species and the dynamics of each species regardless of the kinetic model using SVD analysis. From the SVD analysis in the time range from 3.16 μs to 178 ms, three significant singular vectors were determined from the singular values (S) and the autocorrelation factors of the corresponding singular vectors (Figures S6A and S6B). Figure S6D shows the significant three rSVs scaled with singular values. By globally fitting the rSVs by two exponential sharing common relaxation time, the time constants 298 (± 48) μs and 2.55 (± 0.12) ms. The kinetic components for the transitions from first species to third species were used to extract three species-associated difference scattering curves (SADS) corresponding to the three intermediates by applying principal component analysis (PCA) from the experimental scattering curves as follows:

$$\Delta S_{theory}(q_i, t_j) = \sum_{k=1}^{3} \left[C_k(t_j) \right] \Delta S_{C_k}(q_i) \quad (1)$$

where $\Delta S_{theory}(q_i, t_j)$ is the theoretical difference scattering curve at given q and t values, $\Delta S_{Ck}(q_i)$ is the SADS corresponding to the k-th intermediate species at a given q value, $C_k(t_j)$ is the instantaneous population of the k-th intermediate at a given t value and can be calculated using the time constants obtained from the SVD analysis. Then, we applied the Nelder-Mead simplex algorithm[10] to minimize the discrepancy between the theoretical and experimental curves.

Structural analysis using molecular dynamic simulations and ensemble optimization method

For the structural analysis using the ensemble optimization method (EOM), the species associated-scattering curves (SASs) for each intermediate were first reconstructed[7]. To reconstruct the SASs from the TRXSS curves, we implemented additional static measurements by using dark-incubated state, so-called the ground state, and the light-activated state generated from the continuous illumination of LED light. The difference between the scattering curves of light-activated state and ground state was compared with the third species-associated difference scattering curve (SADS3), named as the photoproduct state. From this comparison, we could find the scaling factor between the static scattering curve and the difference scattering curves from the time-resolved measurement. Here, we note that the scaling factor depends on both the photoconversion yield and the static scattering-to-difference scattering scale factor.

After determination of the scaling factor, SASs for the first and second intermediates were constructed by adding the scaled SADS for each intermediate to the static scattering curve for the ground state[7]. The experimental static curves of the ground and light-activated states and the SASs of the intermediates were used in the EOM analysis to extract the structural parameters involved in the light-induced transition of coPYP-4.

To extract the structural parameter of coPYP-4 such as the radii of gyration (R_g) and three-dimensional protein conformations of the ground and light-induced states, we employed the structural analysis aided by molecular dynamics (MD) simulations[9,11] and ensemble optimization method (EOM)[9,12-14]. This analysis consists of two steps: (i) generating candidate structures to describe experimental scattering curves from the ground and light-induced states of coPYP-4 by performing MD simulations and (ii) extracting the structural parameter and the three-dimensional conformations for coPYP-4 in the two states using the ensemble of protein conformations that reproduces the experimental scattering curves. In the first step, different starting conformations were simulated for each state. The structure of coPYP-4 obtained from the experiment-restrained rigid-body MD simulation was used as (a) the starting conformation of MD simulation for the ground state. For (b) the starting conformation of MD simulation for the light-induced state, each monomer PYP structure constituting (a) was replaced with the structure of monomer PYP in the signaling state (PDB ID: 2KX6). The 2KX6 structure is an ensemble of 14 structures, and (b) is constructed based on model 1 among the 14 structures. The second step was achieved by applying the ensemble optimization method (EOM)[9,12-14] on the sampled structures from the MD simulations. The details of these two steps are as follows. ***MD simulation.*** GROMACS 5.1.4 and the Charmm36 force field were used to perform the simulations[15-18]. In the starting conformations, (a) and (b), all bonds between the linkers and the PYP units were linked by modifying topology. In each simulation, the starting conformation was first energy minimized to a maximum force of 500 KJ/mol/nm with a cubic simulation box. The conformation was solvated with TIP3P water and neutralized by ions corresponding to the experimental condition of TRXSS (20 mM NaCl). The system was equilibrated under NVT condition for 100 ps with velocity-rescale thermostat (τ_T = 0.1 ps, T = 300 K) after another energy minimization. And it was subsequently equilibrated under NPT condition for 150 ps with velocity-rescale thermostat (τ_T = 0.1 ps, T = 300 K) and Parrinello-Rahman barostat (τ_T = 0.5 ps, P = 1 bar). After equilibration, production simulation was conducted with the equilibrated system for 500 ns. The candidate structures were sampled at 10 ps intervals for (ii). And the theoretical x-ray scattering curves of the candidate structures were calculated

using CRYSOL[6]. The theoretical scattering curves were calculated using the values for the electron density of the solvent = 0.334 e/Å^3, the contrast of the hydration shell = 0.03 e/Å^3, and the average atomic radius = 1.611 Å.

Ensemble optimization method. For the candidate structures from the MD simulations, we implemented EOM analysis using a well-established method[9,11]. Each candidate structure from the MD simulations is considered the potential conformations to be included in the ensembles for the ground and the light states of coPYP-4. The theoretical x-ray scattering curves of the candidate structures were calculated using CRYSOL[6]. Assuming that all conformers exist with the same probability, the experimental data can be expressed by an ensemble containing *M* different conformers. The average curves for target protein can be obtained from the structures included in the ensemble and compared with the experimental x-ray scattering data:

$$S_{avr}(q) = \frac{1}{M}\sum_{n=1}^{M} S_n(q) \tag{2}$$

where $S_{avr}(q)$ is the averaged scattering curve for an ensemble, $S_n(q)$ is the theoretical X-ray scattering curve for the *n-th* conformer, and *M* is the total number of conformers in an ensemble. In the general algorithm, 50 ensembles (chromosomes) containing 20 different structures randomly selected from the pool of candidate structures were first generated. For these 50 chromosomes, mutation and cross operations were performed in each generation. During the mutation, 50% of the components of each chromosome were exchanged for unselected structures of the pool or structures present in other chromosomes of the same generation. During the crossing, 40% of the components of two randomly selected chromosomes of the same generation were exchanged while maintaining the number of components. After mutation and crossing operation, the number of ensembles in a generation tripled from 50 to 150. The averaged theoretical scattering curve, $S_{avr}(q)$, for each chromosome was compared with the experimental scattering curve to yield the parameter χ^2_{red}:

$$\chi^2_{red} = \frac{1}{K-1}\sum_{j=1}^{K}\left[\frac{\alpha \cdot S_{avr}(q_j) - S_{exp}(q_j)}{\sigma(q_j)}\right]^2 \tag{3}$$

where, $S_{exp}(q)$ is the experimental scattering curve, $S_{avr}(q)$ is the averaged theoretical scattering curve for each ensemble (chromosome), α is the scaling factor between the theoretical and experimental curves, *K* is the number of experimental q-points, and $\sigma(q)$ is the experimental standard deviation. The 50 chromosomes with the smallest χ^2_{red} were selected from the 150 chromosomes produced by mutation and crossing of the initial chromosomes and propagated

to the next generation (evolution of chromosome). The three processes of mutation, crossing, and evolution of chromosomes were repeated over 1000 generations. At the end of evolution, the most suitable chromosome for experimental data was finally collected to extract the optimal structural parameters of R_g and three-dimensional protein conformations. In the case of the analysis process using the genetic algorithm, ground and light-induced states were separately performed. The EOM analysis process was performed five times for the pool of candidate structures in each state, and the resulting R_g distribution was averaged by calculating the standard deviation. Among the protein conformations in the optimized ensemble, the structure observed with the highest frequency was chosen as the representative structure. In the photoproduct, the newly formed populations were observed around the R_g value of 28.5 Å, indicating that the populations involve the protein conformations corresponding to the final product originating from the light-induced activation of coPYP-4 (Figure 4). Therefore, in the case of photoproduct, the structure observed with the highest frequency was chosen as the representative structure among the protein conformations in the newly formed populations. After the EOM analysis, for each state, we calculated the weighted R factor (*wR*) using the following equation to quantify the agreement between the theoretical scattering curve of the optimized ensemble and the experimental scattering curve.

$$wR = \sqrt{\frac{\sum_i \frac{(S_{opt}(q_i) - S_{exp}(q_i))^2}{(\sigma(q_i))^2}}{\sum_i \frac{(S_{exp}(q_i))^2}{(\sigma(q_i))^2}}} \tag{4}$$

where, $S_{opt}(q)$ is the theoretical scattering curve of the optimized ensemble, $S_{exp}(q)$ is the experimental scattering curve, $\sigma(q)$ is the experimental standard deviation.

Applying a specific constraint to the MD simulation can cause the structure pool obtained from the simulation to be biased toward a specific reaction coordinate. For this reason, we did not enforce motions such as expansion or contraction for the MD simulations. Accordingly, the candidate structures obtained from the unconstrained simulations performed for the two different starting points cover diverse structural information related to the conformational changes of coPYP-4. Meanwhile, considering that the force field applied to the simulations is mainly defined by well-structured proteins, it is plausible that our simulations may have limitations in describing subtle structural changes of coPYP-4 caused by the partial unfolding

of N-terminals in the PYP units. Nevertheless, the main purpose of the structural analysis is to investigate global conformational changes such as expansion or contraction motions associated with light-induced transitions of coPYP-4. In this regard, the results from the structural analysis are not likely to be highly affected by the limitations. Therefore, the diversity of structural information allows that the conformational parameters and ensemble structures optimized for x-ray scattering through EOM analysis from the candidate structures can provide a reliable solution to the global structural changes of coPYP-4.

We also reconstructed the SAS curves of the third species, so-called the photoproduct, by varying the photoconversion yield from 0.1 to 0.5 and implemented the EOM analysis using these static curves to investigate the sensitivity of the scaling factor to the structural analysis. As expected, the optimal distributions of R_g show that the relative ratio of the populations around 28.5 Å corresponding to the final photoproduct gradually becomes larger than that of the populations around 32 Å corresponding to the ground as the photoconversion yield increases (Figure S10). Nevertheless, even though the photoconversion yield changes, the populations of the final photoproduct and ground are constantly observed around 28.5 Å and 32 Å, respectively (Figure S10), implying that the same structural changes of the light-activated coPYP-4 are obtained regardless of the changes in the photoconversion yield. These results indicate that the changes of the photoconversion yield affect the relative ratio between the final photoproduct and ground, but not the extracted structural parameters of the intermediates involved in the light-induced transitions of coPYP-4.

Experiment-restrained rigid-body MD simulation

Prior to the MD simulations of coPYP-4 which were used for the EOM analysis, two different starting conformations of coPYP-4 were obtained. The starting conformation for the ground state of coPYP-4 was constructed from the crystal structure of monomer PYP in the ground state (PDB ID: 2PHY) using the experiment-restrained rigid-body (ERRB) MD simulation. After the construction, each monomer PYP constituting the starting structure for the ground state constructed from the ERRB MD simulation was replaced with the structure of the monomer PYP in the signaling state (PDB ID: 2KX6). The ERRB MD simulation method is nearly identical to that described in a previous study[8,19-23]. In the ERRB MD simulation, the crystal structure of monomer PYP in the ground state (PDB ID: 2PHY) was used as a starting point. The structures were divided into a number of rigid bodies composed of several amino acids. The rigid bodies were allowed to move under the influence of the chemical and χ^2 force

fields. Because the atomic structure within a rigid body was constrained to be the same as that of the crystal structure, the force field within the rigid body was not considered. However, the van der Waals interactions between rigid bodies and the N-C bond-length corrections between rigid bodies were included in the chemical force field. The χ^2 force field was introduced to drive the molecular structure generated by MD simulations toward a structure that yields a difference scattering curve that matches the experimental difference scattering curve[6,24].

The total potential (U) on the rigid bodies has the van der Waals term (U_{LJ}), the χ^2 term, and the bond correction term (B) as follows:

$$U = c_1 U_{LJ} + c_2 \chi^2 + c_3 B \tag{5}$$

where c_1, c_2, and c_3 are the weighting parameters to scale the magnitude of the three terms appropriately. U_{LJ} and χ^2, which define the agreement between the experimental data and theoretical value, are as follows (the scaling factor c_s and reduced units are used):

$$U_{LJ} = \sum_{i=1}^{N} \sum_{j \neq i}^{N} 4 \left(\frac{\sigma^{12}}{r_{ij}^{12}} - \frac{\sigma^{6}}{r_{ij}^{6}} \right) \tag{6}$$

$$\chi^2 = \sum_{q=1}^{N_q} \left(\frac{I_{\exp}(q) - c_s I_{cal}(q)}{\sigma_q} \right)^2 \tag{7}$$

where, r_{ij} is the distance between atom i and atom j, $I_{exp}(q)$ is the experimental scattering curve of coPYP-4, $I_{cal}(q)$ is the theoretical scattering curve of the model structure generated from the crystal structure of monomer PYP (PDB ID: 2PHY), and σ_q is the experimental standard deviation. The force field is the gradient of the total potential; therefore, the total force acting on the i-th particle among the total number of atoms, N, is as follows:

$$\begin{aligned} \mathbf{f}_i &= -\nabla_i U \\ &= 24 c_1 \sum_{j \neq i}^{N} \left(2 \frac{\sigma^{12}}{r_{ij}^{13}} - \frac{\sigma^{6}}{r_{ij}^{7}} \right) \\ &+ 2 c_2 c_s \sum_{q=1}^{N_q} \left(\frac{I_{\exp}(q) - c_s I_{cal}(q)}{\sigma_q^2} \right) \sum_{j \neq i}^{N} f_i^{s}(q) f_j^{s}(q) \left(\cos(q r_{ij}) - \frac{\sin(q r_{ij})}{q r_{ij}} \right) \frac{\mathbf{r}_{ij}}{r_{ij}^2} \end{aligned} \tag{8}$$

In the L-J potential, σ is defined as the r value where the corresponding potential is zero and $\mathbf{r}_{ij}$ (= $\mathbf{r}_i - \mathbf{r}_j$) is the relative position vector between atom i and atom j. If the distance between the two atoms is smaller than σ, the repulsion increases steeply. The following two

types of σ values were defined for the L-J potential between rigid bodies. For the N-C bond between two rigid bodies consisting of helices, $\sigma_{N\text{-}C}$ = 1.28 Å, and for the atoms between two rigid bodies, $\sigma_{a\text{-}a}$ = 1.2 Å. The bond correction term, *B*, is simply the sum of $r(\text{N-C})_k - 1.28$ Å, where *r*(N-C) is the bond length for the N-C bond between two connected rigid bodies and k covers all relevant N-C bonds between connected rigid bodies, introduced to maintain the distance between rigid bodies at the initial value (1.28 Å). Once the total forces between rigid bodies were determined, MD simulations were run based on Newtonian equations. The coordinate of the center of mass (COM) of the rigid body was determined as follows:

$$\mathbf{R} = \frac{\sum m_i \mathbf{r}_i}{M} \tag{9}$$

where m_i and $\mathbf{r}_i$ are the mass and position of the *i*-th atom of the rigid body and *M* is the mass of a rigid body a rigid body ($M = \sum_i m_i$).

The COM moved translationally, and the relative rotational motion with respect to the COM determined the relative position ($\mathbf{r}_i' = \mathbf{r}_i - \mathbf{R}$) of the *i*-th atom referenced to **R**, $\mathbf{r}_i'$. At each step of the MD simulation, **R,** $\mathbf{r}_i$, and $\mathbf{r}_i'$ were updated. The equation for the translational motion of the COM was as follows:

$$\ddot{\mathbf{R}} = \frac{\mathbf{F}}{M} \tag{10}$$

where $\mathbf{F} = \sum_i \mathbf{f}_i$ is the total force on the rigid body defined as the sum of $\mathbf{f}_i$ acting on the *i*-th atom within the rigid body. The total torque acting on each rigid body with respect to the COM was as follows:

$$\mathbf{N} = \sum_i \mathbf{r}_i' \times \mathbf{f}_i \tag{11}$$

The velocity of $\mathbf{r}_i'$ relative to **R** was as follows:

$$\dot{\mathbf{r}}_i' = \omega \times \mathbf{r}_i' \tag{12}$$

where **ω** is the angular velocity with respect to the COM of the rigid body. The moment of inertia with respect to the COM was as follows:

$$\mathbf{I} = \begin{bmatrix} \sum m_i(y_i'^2 + z_i'^2) - \sum m_i x_i'^2 y_i'^2 - \sum m_i x_i'^2 z_i'^2 \\ -\sum m_i x_i'^2 y_i'^2 \sum m_i(x_i'^2 + z_i'^2) - \sum m_i y_i'^2 z_i'^2 \\ -\sum m_i x_i'^2 z_i'^2 - \sum m_i y_i'^2 z_i'^2 \sum m_i(x_i'^2 + y_i'^2) \end{bmatrix} \tag{13}$$

The rotational equation of motion of a rigid body around **R** can be expressed as:

$$\dot{\omega} = \mathrm{I}^{-1} \cdot (\mathrm{N} - \omega \times (\mathrm{I} \cdot \omega)) \tag{14}$$

Eq. (5) updates **R** and the combination of eqns. (11) – (14) updates $\mathbf{r}_i'$.

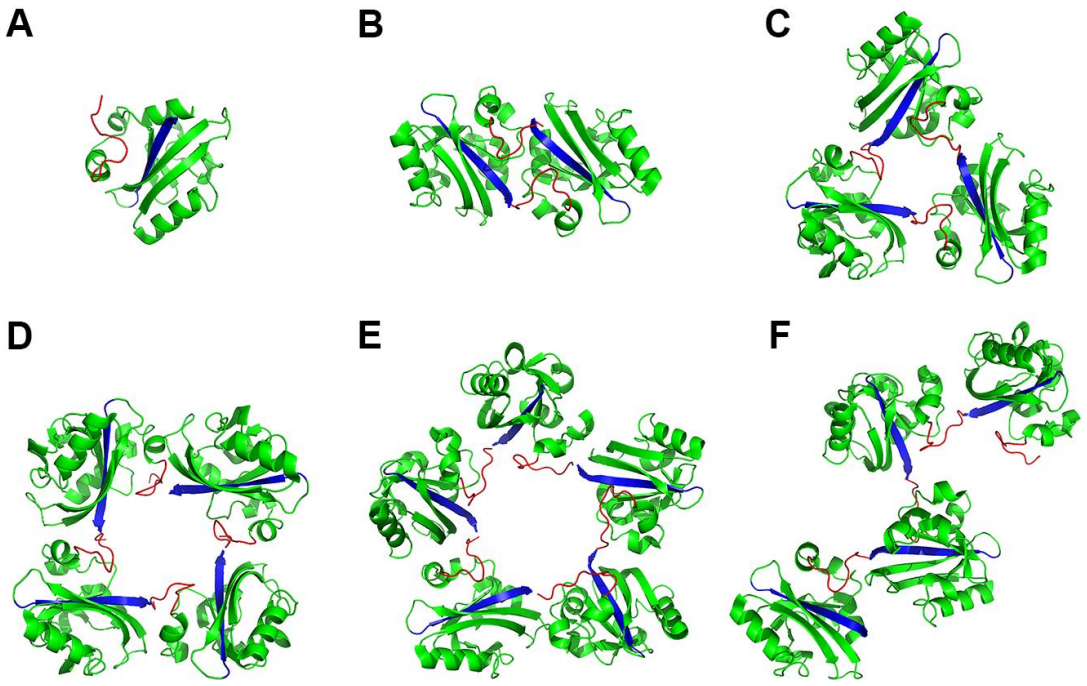

Figure S1. Schematic diagram of coPYPs, loPYP-4, dimer, and monomer. (A) Monomer. (B) Dimer. (C) Circular trimer (coPYP-3). (D) Circular tetramer (coPYP-4). (E) Circular pentamer (coPYP-5). (F) Linear tetramer PYP (loPYP-4). The ten residues at the N-terminal of PYP units and the ten residues at the C-terminal of PYP units are indicated in red and blue, respectively.

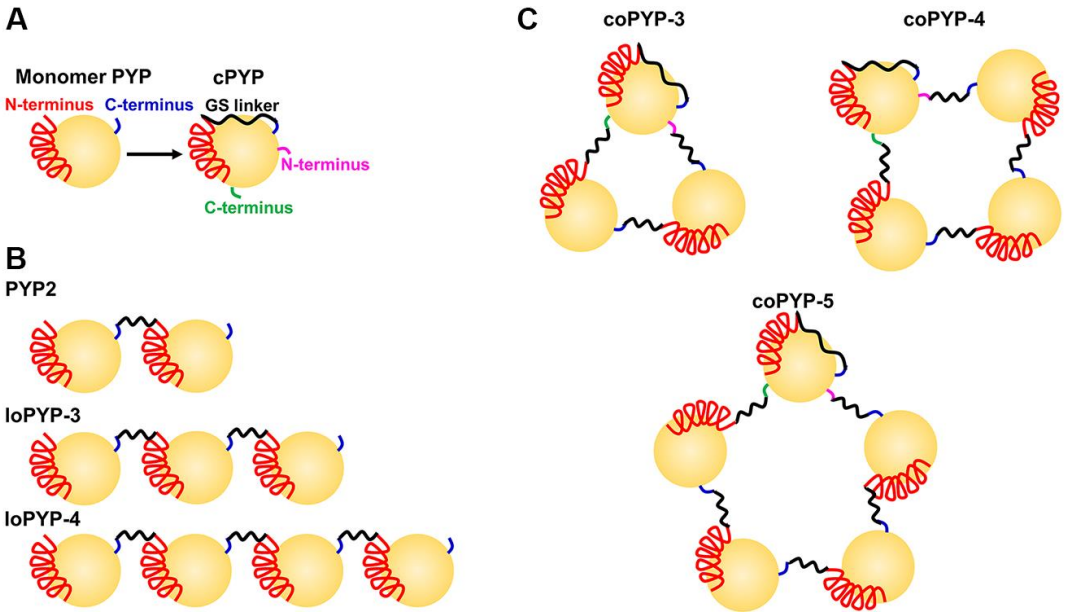

Figure S2. Overview of constructing coPYPs. (A) The process of generating cPYP from monomer PYP. In this process, the C- (blue) and N-terminus (red) of monomer PYP are connected by the GS linker (black), and new N- (Magenta) and C-terminus (green) are introduced to G115 and S114. (B) Schematics for PYP2, loPYP-3, and loPYP-4. (C) Schematics for coPYP-3, coPYP-4, and coPYP-5. Each linear oligomer PYP (B) is genetically linked to cPYP to form coPYPs. In (B) and (C), the C-terminals, N-terminals, and the GS linkers are represented by the red, blue, and black lines, respectively.

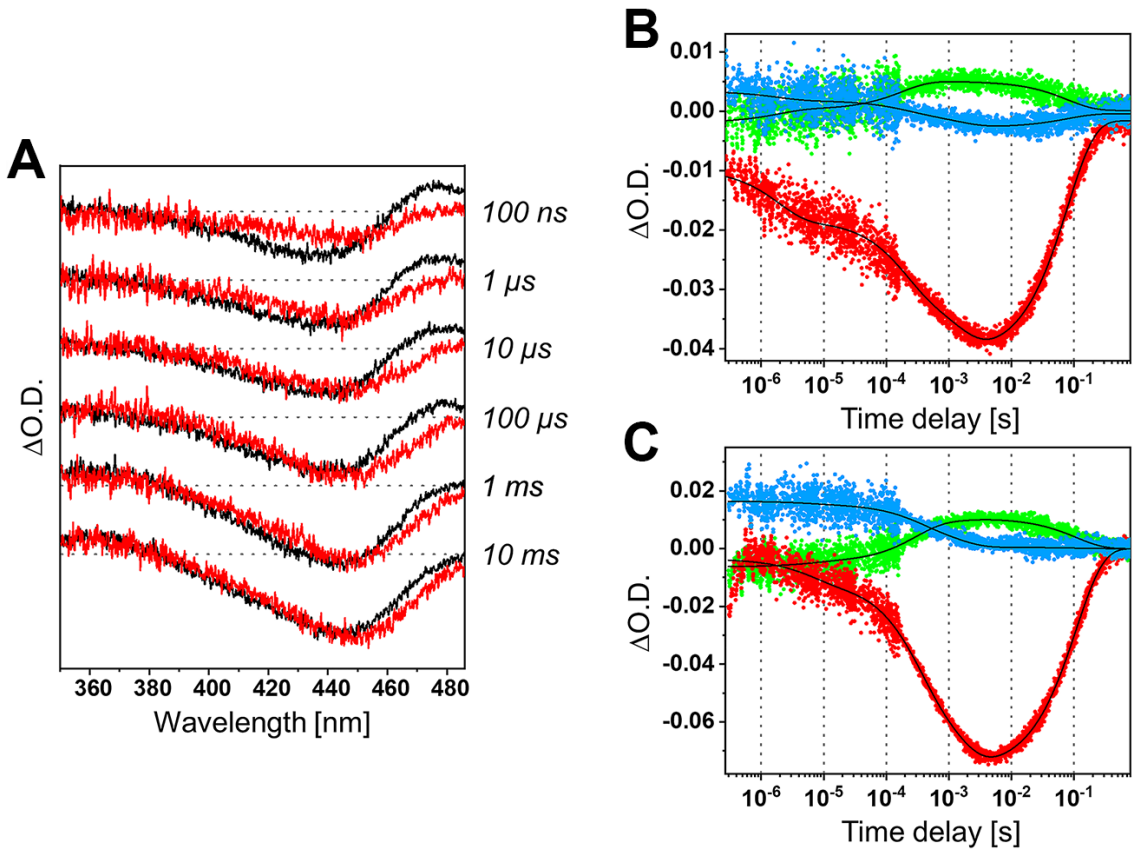

Figure S3. TA result of coPYP-4 compared with monomer PYP. (A) Difference TA spectra of coPYP-4 (red) and monomer PYP (black). For comparison, TA spectra were scaled with an arbitrary scaling factor. From top to bottom, time delays are 100 ns, 1 μs, 10 μs, 100 μs, 1 ms, and 100 ms, respectively. (B, C) TA decay profiles at 380 nm (green), 465 nm (red), and 494 nm (blue) of (B) coPYP-4 and (C) monomer PYP, respectively. Black dotted lines indicate the corresponding fitting curves for each profile. A sum of four exponentials reproduces the experimental curves well. The time constants determined by the fitting are 1.9 (±1.6) μs, 180 (±60) μs, 1.4 (±0.3) ms, and 79 (±3) ms for coPYP-4 and 6.2 (±6.4) μs, 280 (±110) μs, 1.3 (±0.6) ms, and 110 (±3) ms for monomer PYP.

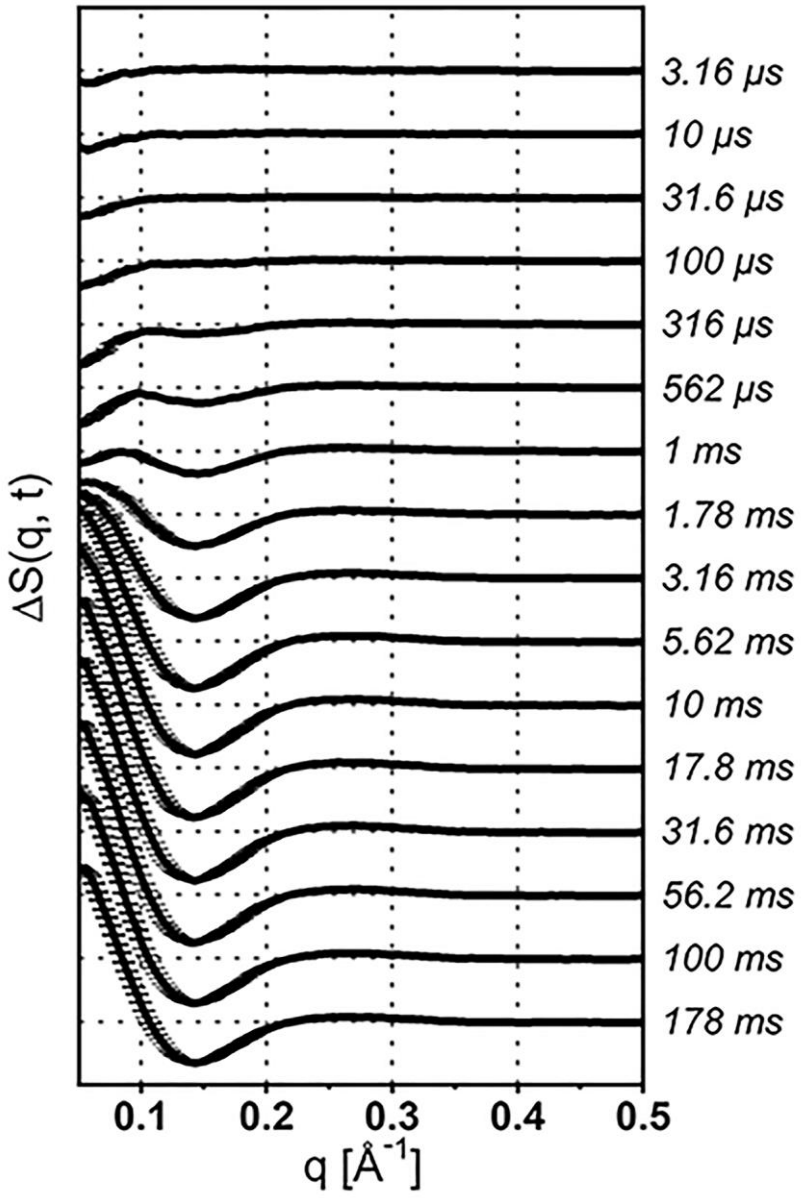

Figure S4. Difference scattering curves with their standard deviations from TRXSS measurement. Each difference scattering curve represents a low value of standard deviation.

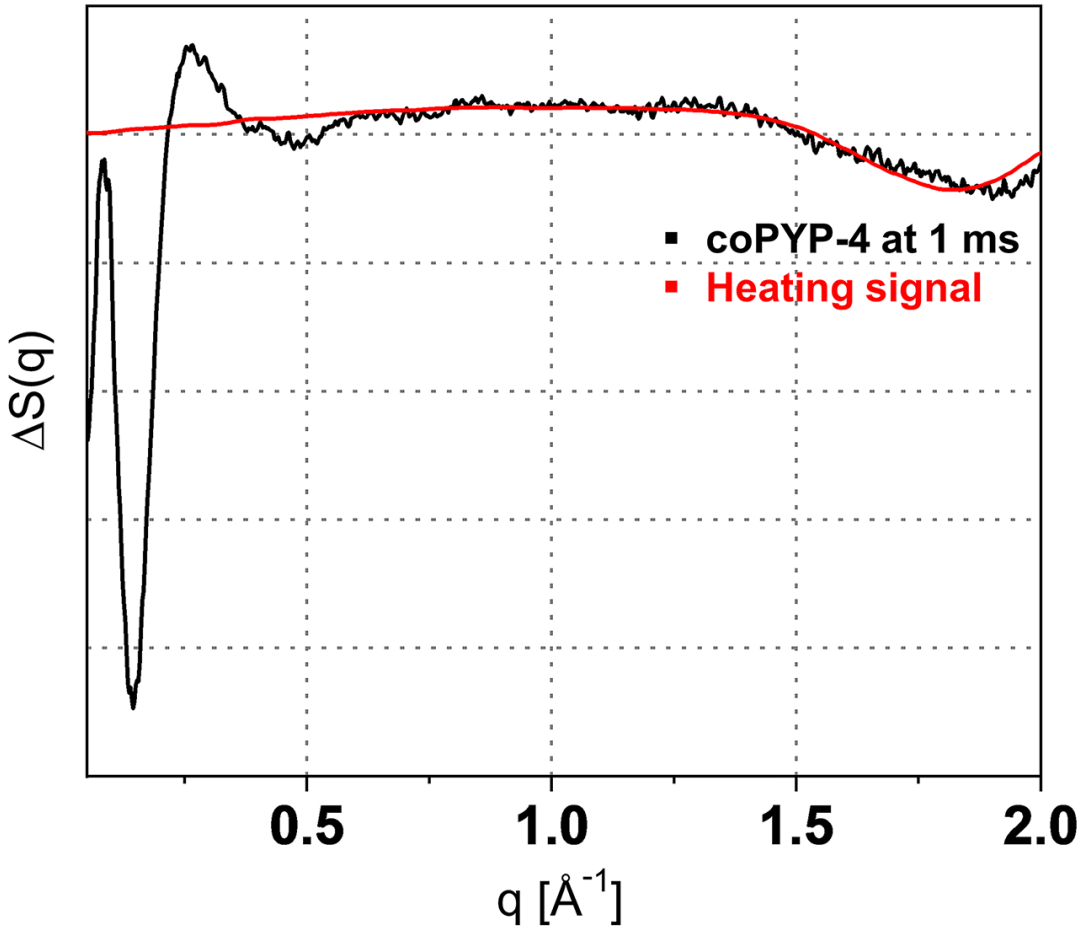

Figure S5. Comparison of a time-resolved x-ray solution scattering curve of coPYP-4 with a solvent heating signal. The difference scattering curve of coPYP-4 at 1 ms time delay (black line) is compared with the solvent heating signal (red line), reconstructed from a linear combination of temperature change at constant pressure and density change at a constant temperature. The solvent heating signal does not show any significant features in the *q*-range smaller than 0.6 $Å^{-1}$.

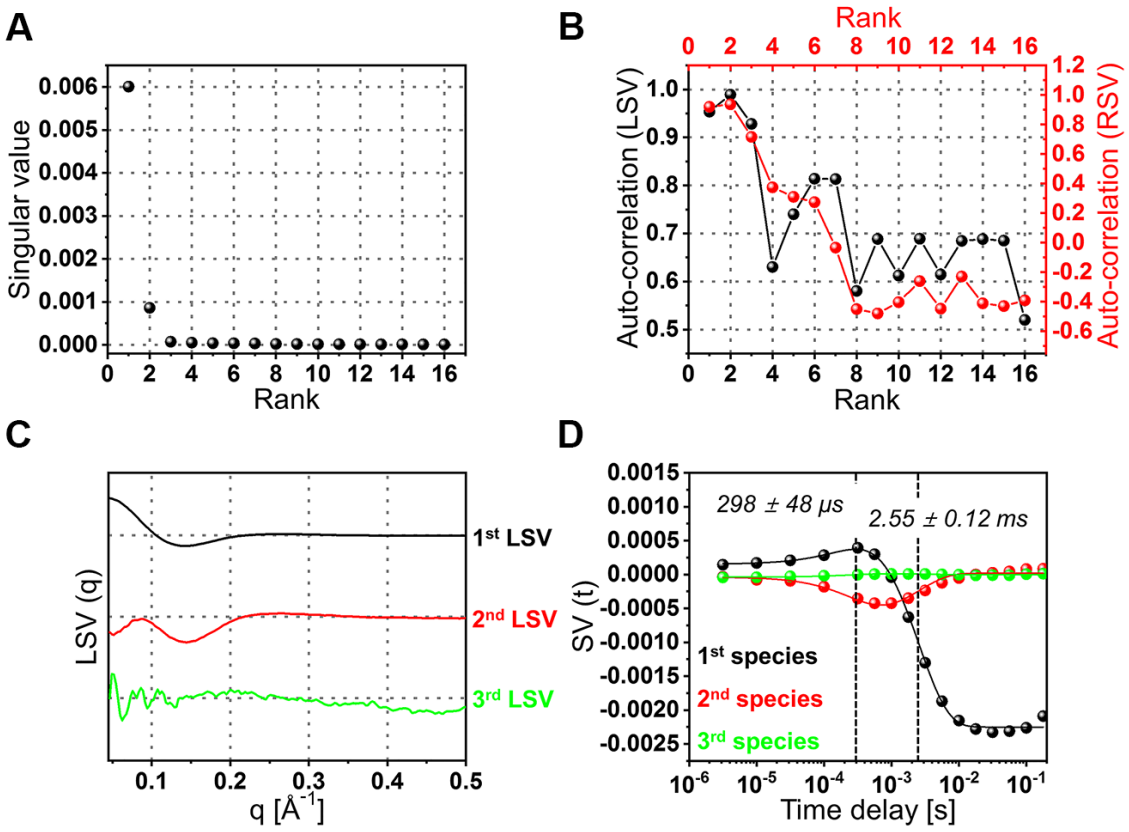

Figure S6. SVD analysis of the time-resolved x-ray scattering data. (A-D) SVD analysis of the TRXSS data. (A) Singular values (S). (B) Autocorrelation values of lSVs (black dots) and rSVs (red dots). (C) The most significant lSVs. (D) The most significant rSVs multiplied by the corresponding singular values (S). The two common time-constants of 298 μs and 2.55 ms were obtained from global exponential fitting against the rSVs (solid lines). Considering the singular values and features of singular vectors and their autocorrelation values, the first three lSVs and rSVs are significant for the data from 3.16 μs to 178 ms. The exact recovery time could not be obtained because it was larger than the maximal time delay of 178 ms.

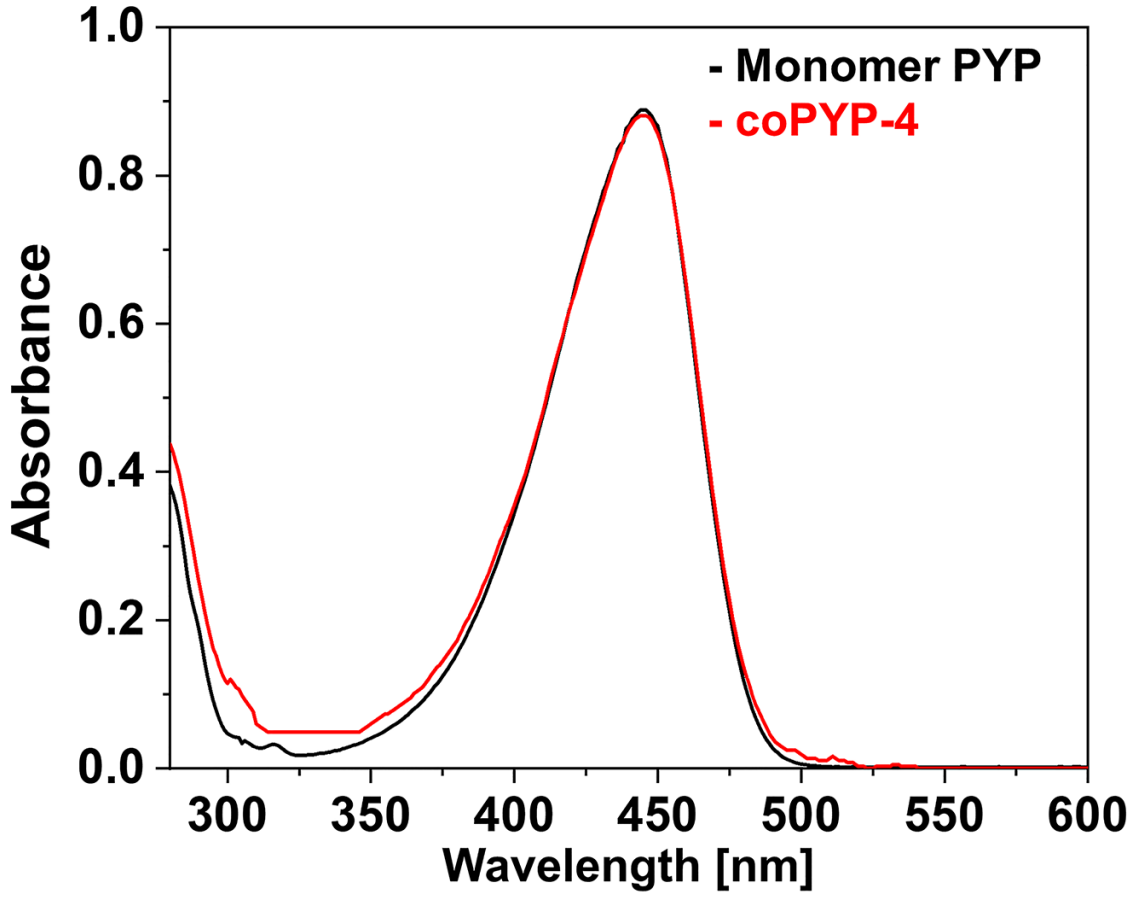

Figure S7. UV-VIS spectra of monomer PYP and coPYP-4. Spectra for monomer PYP and coPYP-4 are indicated by black and red lines, respectively. Between 280 nm and 350 nm, the spectrum of coPYP-4 has a relatively higher intensity than monomer PYP. The higher intensity is due to the absorbance of glycine and serine that constitute the linkers of coPYP-4.

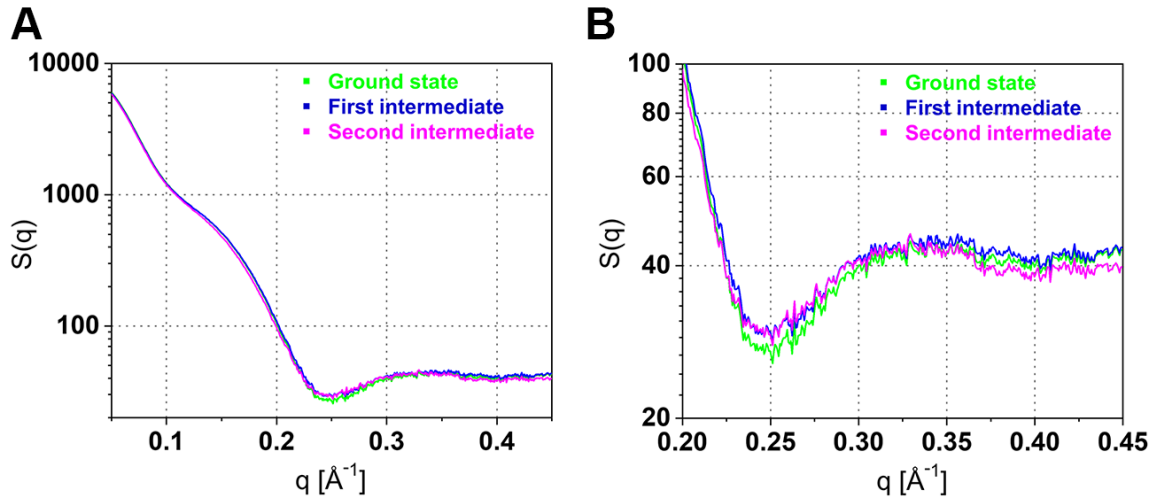

Figure S8. Comparison of the static curves for the ground, the first intermediate and the second intermediate. (A) Comparison of the static curves for the ground (green), the first intermediate (blue) and the second intermediate (magenta) in the q-range of 0.05 Å^{-1} < q < 0.45 Å^{-1}. (B) Comparison of the static curves in the q-range of 0.2 Å^{-1} < q < 0.45 Å^{-1}.

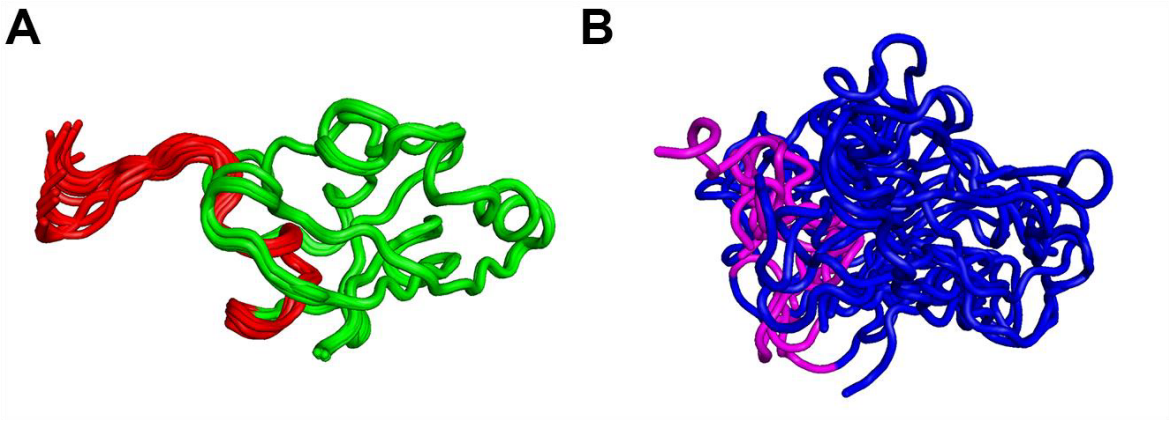

Figure S9. Comparison between the ensemble structure of monomer PYP in the signaling state and PYP units in coPYP-4 in the photoproduct. (A) The 14 structures in the ensemble of monomer PYP in the signaling state (PDB ID: 2KX6). The N-terminus and the rest of the protein body are represented by red and green, respectively. (B) The 4 structures of PYP units in coPYP-4 of the photoproduct. The N-terminus and the rest of the protein body are represented by magenta and blue, respectively. The structures are expressed in the tube style using the PyMOL visualization tool.

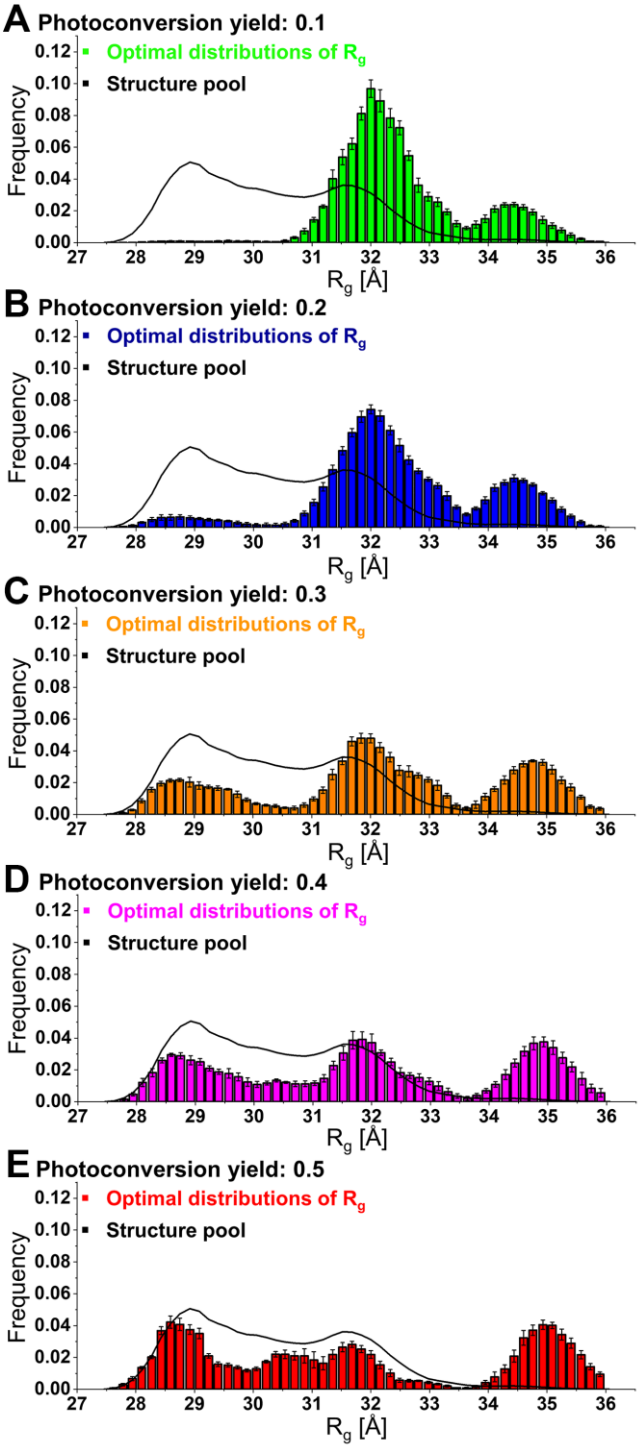

Figure S10. The optimal distributions of R_g from the structural analysis using the condition of photoconversion yield from 0.1 to 0.5. (A) Result from the EOM analysis for the photoconversion yield of 0.1, (B) the photoconversion yield of 0.2, (C) the photoconversion yield of 0.3, (D) the photoconversion yield of 0.4, and (E) the photoconversion yield of 0.5.

Supplemental References

1. Kumar, A., Burns, D.C., Al-Abdul-Wahid, M.S., and Woolley, G.A. (2013). A circularly permuted photoactive yellow protein as a scaffold for photoswitch design. Biochemistry *52*, 3320-3331.
2. Yeremenko, S., Van Stokkum, I.H., Moffat, K., and Hellingwerf, K.J. (2006). Influence of the crystalline state on photoinduced dynamics of photoactive yellow protein studied by ultraviolet-visible transient absorption spectroscopy. Biophys. J. *90*, 4224-4235.
3. Yang, C., Kim, S.O., Kim, Y., Yun, S.R., Choi, J., and Ihee, H. (2017). Photocycle of photoactive yellow protein in cell-mimetic environments: molecular volume changes and kinetics. J. Phys. Chem. B *121*, 769-779.
4. Imamoto, Y., Kamikubo, H., Harigai, M., Shimizu, N., and Kataoka, M. (2002). Light-Induced Global Conformational Change of Photoactive Yellow Protein in Solution. Biochemistry *41*, 13595-13601.
5. Svergun, D. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr *25*, 495-503.
6. Svergun, D., Barberato, C., and Koch, M.H. (1995). CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr *28*, 768-773.
7. 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., et al. (2012). Protein structural dynamics of photoactive yellow protein in solution revealed by pump–probe X-ray solution scattering. J. Am. Chem. Soc. *134*, 3145-3153.
8. Kim, K.H., Muniyappan, S., Oang, K.Y., Kim, J.G., Nozawa, S., Sato, T., Koshihara, S.-y., Henning, R., Kosheleva, I., Ki, H., et al. (2012). Direct observation of cooperative protein structural dynamics of homodimeric hemoglobin from 100 ps to 10 ms with pump–probe X-ray solution scattering. J. Am. Chem. Soc. *134*, 7001-7008.
9. Kim, T.W., Lee, S.J., Jo, J., Kim, J.G., Ki, H., Kim, C.W., Cho, K.H., Choi, J., Lee, J.H., Wulff, M., et al. (2020). Protein folding from heterogeneous unfolded state revealed by time-resolved X-ray solution scattering. Proc. Natl. Acad. Sci. U.S.A. *117*, 14996-15005.
10. Lagarias, J.C., Reeds, J.A., Wright, M.H., and Wright, P.E. (1998). Convergence properties of the Nelder--Mead simplex method in low dimensions. SIAM J OPTIMIZ *9*, 112-147.
11. Zoete, V., Cuendet, M.A., Grosdidier, A., and Michielin, O. (2011). SwissParam: a fast force field generation tool for small organic molecules. J. Comput. Chem *32*, 2359-2368.
12. Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. J. Am. Chem. Soc. *129*, 5656-5664.
13. Bernadó, P., Svergun, D.I., and Bernadó, P. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Mol Biosyst *8*, 151-167.
14. Jones, G. (1998). Genetic and evolutionary algorithms. Encyclopedia of Computational Chemistry *2*, 1127-1136.
15. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. Softwarex *1*, 19-25.
16. Berendsen, H.J., van der Spoel, D., and van Drunen, R. (1995). GROMACS: a message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. *91*, 43-56.
17. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics *29*, 845-854.

18. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H., and MacKerell, A.D. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods *14*, 71-73.
19. Kim, Y., Ganesan, P., Jo, J., Kim, S.O., Thamilselvan, K., and Ihee, H. (2018). Chromophore-removal-induced conformational change in photoactive yellow protein determined through spectroscopic and X-ray solution scattering studies. J. Phys. Chem. B *122*, 4513-4520.
20. Kim, Y., Yang, C., Kim, T.W., Thamilselvan, K., Kim, Y., and Ihee, H. (2018). Regulation of Protein Structural Changes by Incorporation of a Small-Molecule Linker. Int. J. Mol. Sci. *19*, 3714.
21. Ahn, S., Kim, K.H., Kim, Y., Kim, J., and Ihee, H. (2009). Protein Tertiary Structural Changes Visualized by Time-Resolved X-ray Solution Scattering. J. Phys. Chem. B *113*, 13131-13133.
22. Lee, Y., Kim, J.G., Lee, S.J., Muniyappan, S., Kim, T.W., Ki, H., Kim, H., Jo, J., Yun, S.R., Lee, H., et al. (2021). Ultrafast coherent motion and helix rearrangement of homodimeric hemoglobin visualized with femtosecond X-ray solution scattering. Nat. Commun. *12*, 3677.
23. Kim, J.G., Kim, T.W., Kim, J., and Ihee, H. (2015). Protein structural dynamics revealed by time-resolved x-ray solution scattering. Acc. Chem. Res. *48*, 2200-2208.
24. James, F., and Roos, M. (1975). MINUIT: a system for function minimization and analysis of the parameter errors and corrections. Comput. Phys. Commun. *10*, 343-367.