SUPPLEMENTARY INFORMATION

Sample Delivery at EuXFEL SPB/SFX

The sample was delivered to the X-ray interaction region using liquid jets generated from a 3D-printed type C GDVN provided by the EuXFEL Sample Environment Group. The nozzles used are part of the standard 3D-printed suite offered by the EuXFEL[1]. Fused silica capillaries (0.360 mm OD and 0.150 mm ID, Polymicro) were fastened to the liquid and gas inlets using a small dab of epoxy glue (Devcon). The standard SPB/SFX liquid injector rod was used to mount the GDVNs. The rod was assembled as follows: a 1/8 inch OD stainless steel tube was first glued to the capillaries ~5 mm above the nozzle. This tube was fastened into a stainless steel nozzle adaptor using a 10-32 PEEK fitting (Idex). The capillaries were then fed through the entire length of the rod and the end piece was screwed into the tip.

Liquid and gas were delivered to the nozzle as previously described[1]. In short, liquid reservoirs were connected to the nozzle inlets via PEEK tubing (Idex, 0.250 µm ID). Multiple reservoirs were connected in parallel to facilitate fast switching between sample, buffer, and wash solutions using a high speed electronic valve (Rheodyne). Liquid flow was regulated using an HPLC pump (Shimadzu LC-20AD), while helium gas flow was regulated with an electronic pressure regulator (Proportion Air GP1). Gas and liquid flow rates of 23 mg/min and 30 µL/min respectively were typical during the measurement and monitored with in-line flow meters (Bronkhorst F-111B-2K0-TGD-33-V, 0-700 mg/min and Bronkhorst ML120V00-TGD-CC-0-S, 0-100 µL/min respectively), which resulted in jets velocities on the order of 10s of meters per second, which is sufficient to outrun the radiation-induced explosion caused by the ultrafast X-ray pulse and to replenish the sample for each X-ray exposure.

Alignment of the nozzle tip with respect to the interaction region was carried out by manipulating the position of the injector rod using motorized stages. This placement was aided by visualization with the side-view microscope camera illuminated with the EuXFEL femtosecond laser coupled into the sample chamber via a fibre bundle laser synchronized with the X-ray pulse[2,3].

Optical Excitation Scheme

Actinic excitation of the sample was carried out with an optical parametric oscillator (Opolette 355, Opotek) tuned to 475 nm with pulse duration of ~5 nanoseconds and pulse energy of 2

mJ/mm^2 [2]. The output beam (~325 x 338 μm FWHM) was aligned to overlap with the lower half of the GDVN to facilitate excitation of a sufficient sample volume to span the entire X-ray pulse train (~1 nL). The optical pump laser was modulated at half the XFEL intertrain repetition rate (5 Hz) yielding alternating light and dark trains and enables robust extraction of the light-induced scattering response. In this way, the experimental time range and resolution were directly defined by the XFEL pulse bunch length and intra-train repetition rate (~300 μs and 564 MHz respectively). This strategy represents a convenient means to access dense temporal sampling on the micro- to millisecond scale that does not require complex electronic triggering or changes to optical beam alignment.

EuXFEL SPB/SFX Beamline configuration

The data were collected at the SPB/SFX instrument of the EuXFEL in September 2022, under the proposal p3046. The EuXFEL delivered bunch trains at 10 Hz with an intra-train pulse repetition rate of 564 kHz. The photon energy was 8000 eV, which corresponds to a wavelength of 1.55 Å. From previous measurements, the focal spot was estimated at around 300 nm x 300 nm FWHM. The energy of every X-ray pulse was measured by a gas monitor detector upstream and was close to 2 mJ. With this beamline configuration and photon energy the beamline transmission between the gas monitor detector and the interaction region is estimated to be 65%. The AGIPD 1M detector was placed 0.281 m downstream from the interaction region[4]. The experiment was monitored online with Hummingbird[5].

Computation of time-resolved difference X-ray scattering

Data was collected at two different concentrations: 15 and 11 mg/mL. Data filtering was performed by comparing the correlation between individual trains within the run against the train-average for a run; trains below a threshold 0.99995 were considered of low quality and omitted from the averaging of the scattering curves. The averaging of the scattering curves were done for the two concentrations separately over repeats and runs of light and dark resulting in $S_{light}(q,t)$ and $S_{dark}(q,t)$. If not indicated otherwise, i.e Figure 1c, and Supplementary Figure 3, the filtered data was then normalized over the entire q-range ($2.1Å^{-1} > q > 0.08Å^{-1}$) by dividing each scattering point by the sum of the total scattering within the selected q-range. Once normalized, difference scattering curves were calculated, $\Delta S = S_{light}(q,t) - S_{dark}(q,t)$. Subsequently, the low concentration was scaled to match the high concentration, a small offset was also applied to account for systematic detector errors, and the difference curves (ΔS) of the two concentrations were then merged using a weighted average, where the weights correspond to the number of light frames in each data set. The

data displayed in Figure 1c was recorded during 8h of total experiment time and the duration of pure data collection was 3h 15min.

Kinetic modeling

Kinetic decomposition of the experimental data was performed to better understand the reaction dynamics of the AsLOV2 photocycle. Global fitting was carried out assuming a sequential reaction scheme with a variable number of states. The TR-WAXS scattering data can be expressed as a linear combination of time independent basis spectra:

$$\Delta I(q,t) = \Sigma_i (BS_i(q) * C_i(t)) \quad (1)$$

Where $\Delta I(q,t)$ is the measured transient intensity, $BS(q)$ are the time-independent basis spectra, and $C(t)$ are the time-dependent concentrations of the components *i*. In the fitting procedure, the time-dependence of a two-state (A→B) and three-state (A→B→C) model were expressed as exponential functions as:

$$C_A(t) = exp(-k_A t)\ , \quad (2)$$

$$C_B = 1 - C_A\ , \quad (3)$$

and

$$C_A(t) = exp(-k_A t)\ , \quad (4)$$

$$C_B(t) = \frac{k_A}{(k_B - k_A)} \cdot \left[exp(-k_A t) - exp(-k_B t)\right]\ , \quad (5)$$

$$C_C(t) = 1 - \frac{1}{(k_B - k_A)} \cdot [k_B exp(k_A t) - k_A exp(k_B t)]\ . \quad (6)$$

The constants k_i where optimized using simplex minimization of the target function *r* using "fminsearch.m" of Matlab v 2019:

$$r = \left[\Sigma_{q,t} \Delta I(q,t) - \Sigma_i (BS_i(q) * C_i(k))\right]^2 \quad (7)$$

whereby $BS_i(q)$ was determined on each iteration by the least-square solution of Eq. 1 using the backslash operator ("mldivide") in Matlab. The goodness of the fits was judged by plotting the refined kinetics against time-slices of the data in (Supplementary Figure 4). The three-state model gave a lower r compared to the two-state model and a better agreement in terms of shape of the fit.

AlphaFold models

In order to simulate unfolding of the Jα helix, 2000 initial AlphaFold models were created using AFsample with dropout enabled, to determine whether the aggressive sampling could capture the unfolding[6,7]. The initial two thousand models did not display any unfolding and showed only small differences in the last two residues of the Jα helix.

A second approach was taken to ensure the models would have the unfolded helix. By substituting every second amino acid in the helix it can be destabilized artificially, making AlphaFold unable to find any similar sequence in its database. Therefore, it classifies it as a disordered loop instead. The input sequence was modified by introducing 3, 5, 7, 9 and 11 glycine mutations starting from the second-to-last residue and substituting every second until the desired number of mutations was reached. These mutated sequences were then used to run AFsample again, generating 2000 models for each sequence, resulting in a total of 10000 models. In addition, we also introduced mutation in the N-terminal helix to investigate its possible effects on the scattering, five glycine inserts were introduced on top of the Jα inserts resulting in an additional 10000 models.

AFsample was run using the following settings: 1000 models with dropout templates enabled, 500 models with dropout enabled and no templates, each with a maximum of 21 recycles, and 500 models with dropout enabled and no templates, each with a maximum of 9 recycles. The mutated amino acids were then reverted back to the original residue using coot[8]. From the initial 2000 models, which showed very small variation in structure, one was chosen to represent the native state of the protein.

Computation of scattering profiles

Theoretical scattering profiles were calculated using Pepsi-SAXS from the AlphaFold models[9]. We used Pepsi-SAXS, because it computes the scattering of the solvation shell from a grid, which we expect to lead to accurate results for partially unfolded proteins and

because the software is very efficient in computing the scattering of the candidate compounds. The scattering was computed for 170 points, in the range between 0.08 $Å^{-1}$ < q < 1.5 $Å^{-1}$. The theoretical difference scattering (ΔS_{model}) was calculated by subtracting scattering of a native predicted model from each of the unfolded models.

Structural fitting

To compare the theoretical and experimental scattering, the experimental scattering was put on an absolute scale by scaling the experimental dark scattering to the theoretical dark scattering (eq. 8), this scale was then applied to the difference scattering (see Supplementary Figure 5). For comparing the models to the experiment we optimized a projected photoactivation yield *c* for each candidate structural pair according to:

$$SS_{res\ dark} = \Sigma_q (S_{dark.exp.} * c_{abs} - S_{dark.theory})^2 \quad (8)$$

$$R^2 = \frac{\Sigma_q (c\Delta S_{exp.\ scaled} - \Delta S_{model})^2}{\Sigma_q \Delta S^2_{exp.\ scaled}} \quad (9)$$

c corresponds to the photoactivation yield, which a certain structural pair would require the difference X-ray scattering to have for an optimal fit. By performing a least square optimization on the numerator in Eq. 9, *c* could be estimated for each structural pair. The q-scale used for fitting was q < 0.16 $Å^{-1}$. We used it to discriminate good fits. We selected 6032 models with the highest R^2 and which had a projected activation yield of 15 ± 5%.

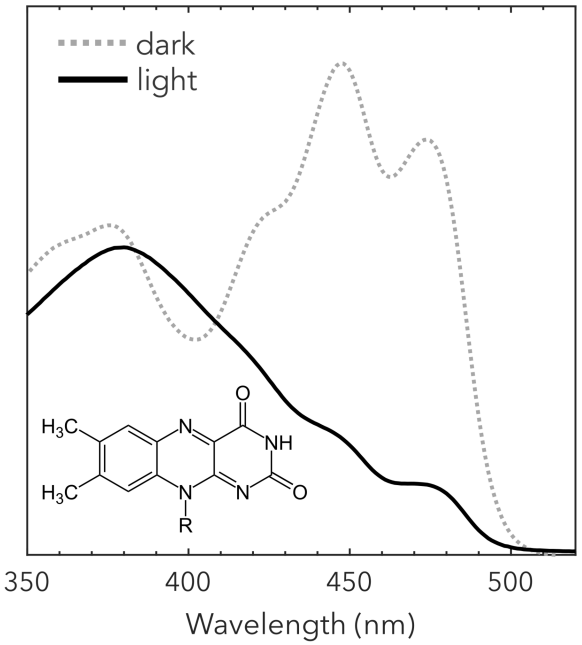
Determination of the photoexcitation yield by auxiliary SAXS experiments

In order to determine the photoexcitation (activation) yield, we performed a separate SAXS experiment of AsLOV2. The data was recorded at Diamond Light Source (beamline B21) at room temperature (20℃). First, we recorded SAXS data in complete darkness followed by illumination of the sample for one second using a laser diode (wavelength: 470 nm, average power: 68 W/m^2, spot size was an ellipse of 4.5 mm x 1.9 mm with an area of 6.7 mm^2). The protein was allowed to dark-revert for five minutes and the procedure was repeated four more times with increasing illumination time for each cycle (2, 5, 10, and 20 seconds). Saturation of the difference signal (light-dark) was observed from 5 second illumination onwards (Supplementary Figure 6a). By comparing the signal height of this SAXS difference

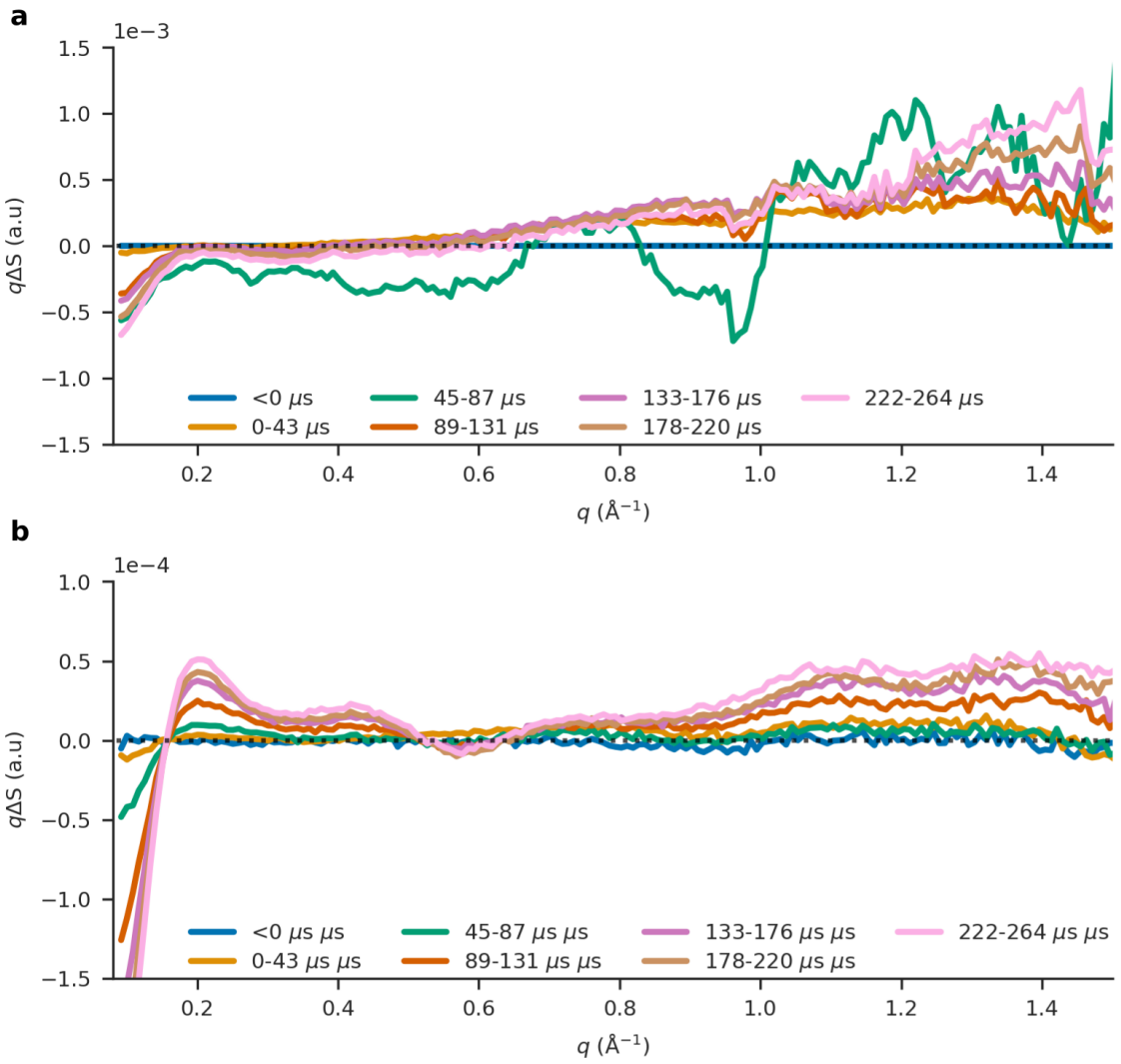
data for full photoconversion to ΔS_{exp} from the XFEL, we determined the excitation yield at the EuXFEL to be 15 ± 5% (Supplementary Figure 6c).

Protein Expression and Purification

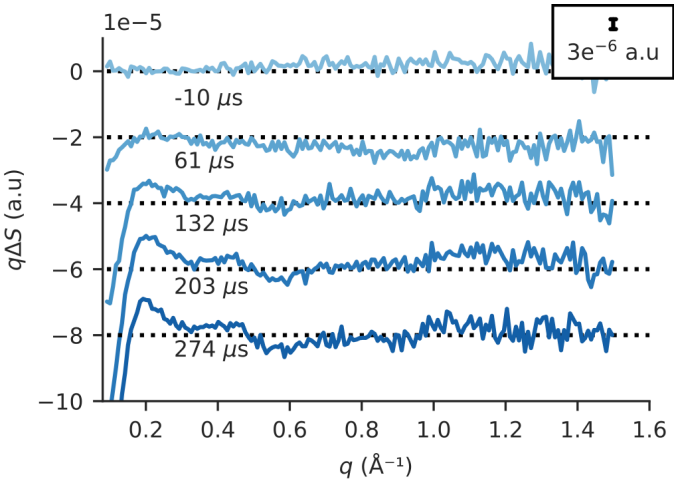
AsLOV2 expression and purification followed from previously reported protocol[10]. The expression plasmid (6His-Gb1-AsLOV2) was obtained from the group of Kevin Gardner at CUNY. The AsLOV2 was expressed in *E coli* BL21 (DE3) STAR (Thermofisher Scientific). This culture was propagated in 11 L of the LB medium, induced with 1 mM IPTG at OD_{600} = 0.8-0.9, and then incubated at 18°C, 180 rpm for 16 h. The cells were centrifuged at 6,000 x*g*, 4°C for 20 min. The cell pellet was washed with 30 mL of Tris buffer (50 mM Tris-HCl, 500 mM NaCl, 0.5 mM DTT, 5% (v/v) glycerol, pH 8). The washed pellet was then resuspended in 60 mL of Tris buffer and sonicated for 2 min with cycles of 15 seconds of sonication separated by 45 second intervals at 50% pulse amplitude using a Branson 450 Digital Sonifier (Branson, BBU13119802A). The sonicated lysate was then cleared at 15,000 x*g*, 4°C for 35 minutes and filtered with a 0.45-µm filter. This was then equilibrated within a Ni-NTA resin column (88222, ThermoScientific) for the further purification. The resin was washed with Tris buffer / 50 mM imidazole. Then, elution was performed in steps up to 500 mM imidazole. The 6xHis tag was cleaved with TEV protease (T4455, Sigma Aldrich) in a 1:50 molar ratio of TEV:AsLOV2, and the mixture was dialyzed twice in Tris buffer at 4°C. The dialyzed AsLOV2 sample was applied to another Ni-NTA resin column to remove the cleaved 6xHis tag and residual TEV. The AsLOV2 sample was further purified using a HiPrep™ 26/60 Sephacryl™ S-100 HR (17119401, Cytiva) size exclusion column. The final yield was 325 mg (from 11 L of culture) and it was stored at -80°C before the experiment at 13 mg/mL.



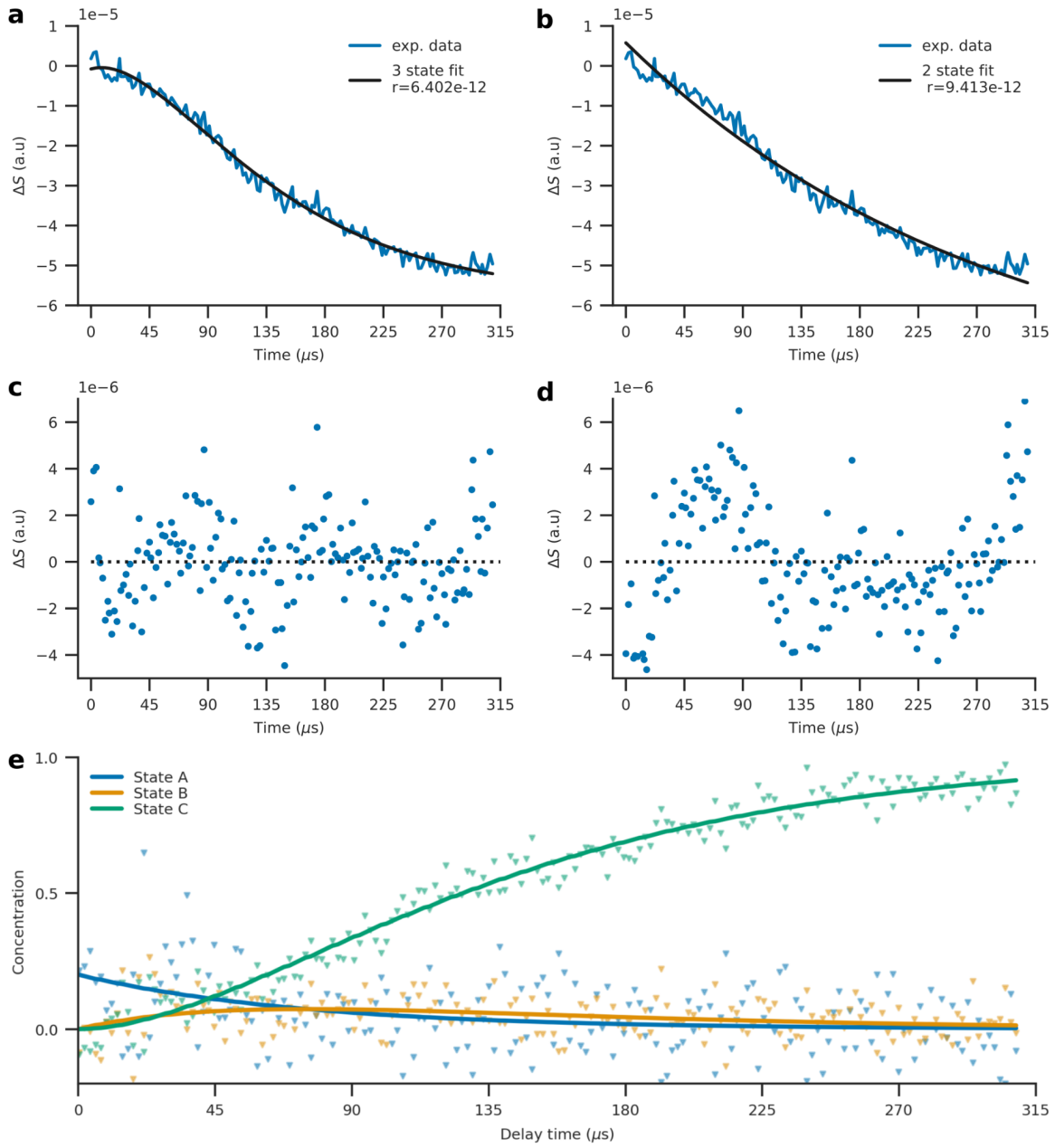
Supplementary Figure 1. AsLOV2 UV-VIS absorption spectrum in dark and light states together with FMN cofactor chemical structure.



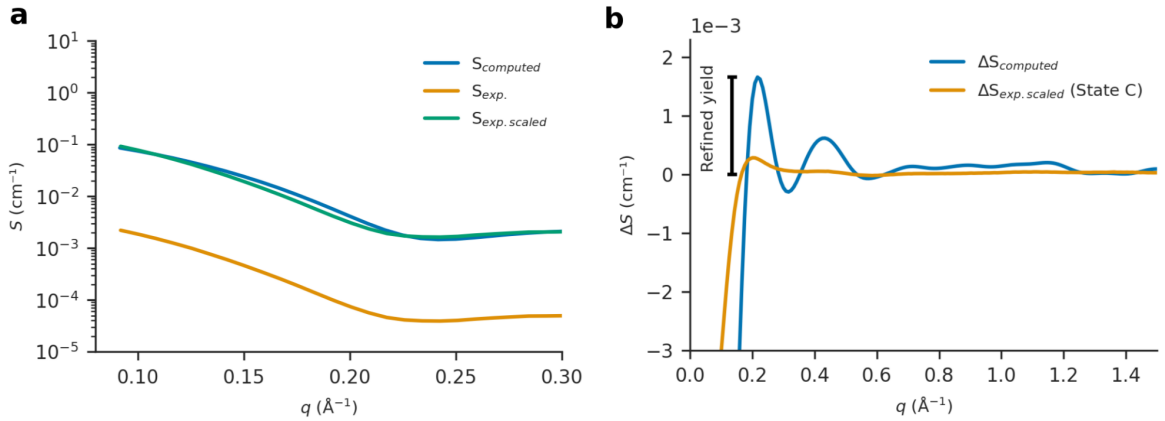
Supplementary Figure 2. Comparison of different ways to compute difference scattering. a. Here, we present an alternative approach that relies solely on the dark frames recorded during laser induced-trains. The difference scattering is calculated by averaging the dark measurements in each of the light-induced trains and using this average as the dark reference. **b.** Our more reliable approach of calculating the difference scattering is to subtract an entire dark train from a light train. This removes systematic errors from the data acquisition in the detector. Each time point in the plots is composed of 50 pulses binned together, and the averaged dark scattering is subtracted to produce the difference scattering. The difference scattering has been scaled to match the theoretical scattering from Pepsi-SAXS [9].



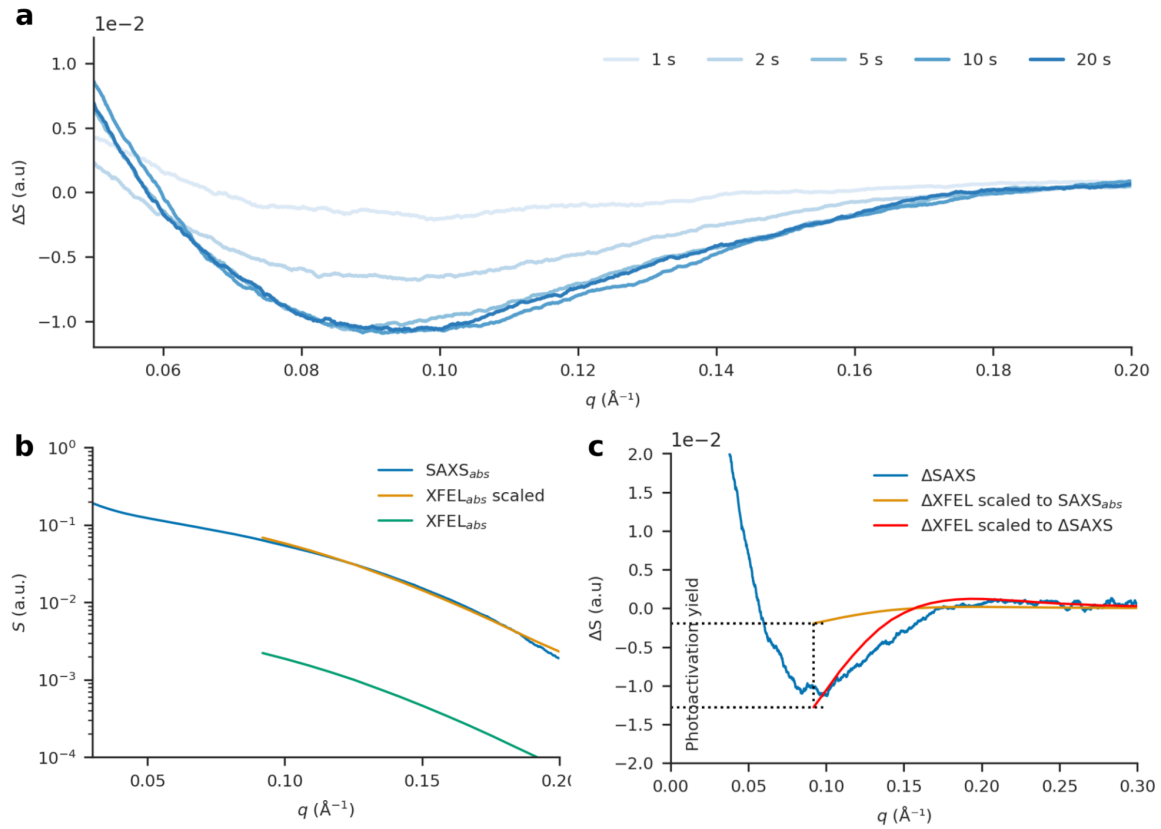
Supplementary Figure 3. Time-resolved WAXS data collected with low noise levels recorded at the SPB/SFX instrument of European XFEL. Our data shows exceptional signal to noise ratio that allows for difference signals of ~10^{-6} to be recorded. As the normalization conditions alters the scale of the difference signal, the experimental data was normalized to the scattering in the range of 1.6 Å^{-1} > q > 1.4 Å^{-1} to match the data treatment of previously published time-resolved solution scattering studies performed at a synchrotron, where best-case noise levels are ~10^{-4} and ~5*10^{-5} in pump-probe and sequential acquisition schemes[11].



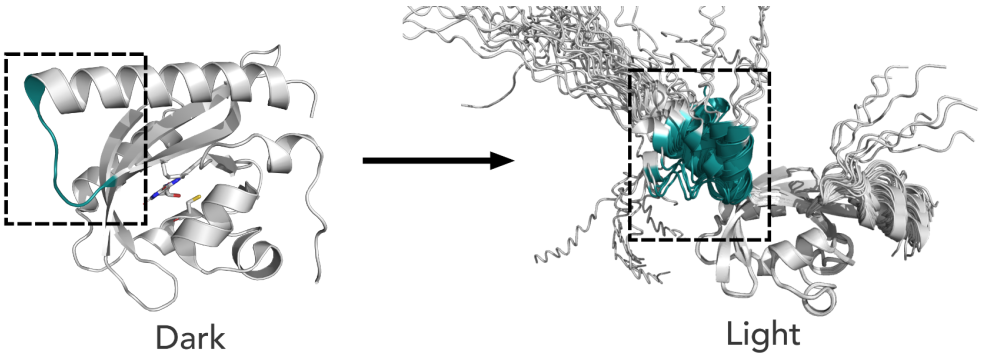
Supplementary Figure 4. A three-state model is required for kinetic fitting of the difference WAXS data. (a-d) The reconstructed kinetics and residuals (*r,* see Eq. 7) of the AsLOV2 difference scattering data for a decomposition with a three-state model (**a,c**) are compared to those from a two-state kinetic model (**b,d**)**.** The scattering data and residuals are shown as an average over the q range (0.08 $Å^{-1}$ < q < 1.2 $Å^{-1}$). The three-state model is preferred over the two-state model, because the fits and residuals of the three-state model show a very good agreement over the entire time-series (panels **a** and **c**), whereas the two-state model shows systematic deviations at early and late times (panels **b** and **d**). Moreover, the *r* value was better for the three-state model ($6.4*10^{-12}$) compared to the two-state model ($9.413*10^{-12}$). (**e**) The time evolution of the constituent states derived from spectral decomposition are shown including experimental data points. States A and B exhibit high noise because of the low signal amplitude (Figure 2b), however, as demonstrated, a three-state model is needed to correctly capture the kinetics. State A and B are scaled by a factor of 0.2 for better visualization.



Supplementary Figure 5. Scaling procedure to assign units to the experimental difference X-ray scattering. (**a**.) The buffer-subtracted and averaged experimental X-ray scattering in dark is scaled to the scattering computed from the dark model of AsLOV2 ($\Delta S_{computed}$). (**b**.) The difference scattering (ΔS_{exp}) was scaled with the same scaling factor. Now, the scaling factor between the experimental and theoretical difference scattering, which is shown for one candidate structural pair in the panel, will correspond to the refined photoactivation yield for a candidate structural pair. The factor c was determined by Eq. 8 in the structural fit procedure for each difference scattering curve from candidate structural pairs.



Supplementary Figure 6. Determination of the photoactivation yield from steady state difference SAXS. a. Difference SAXS curves (light - dark) of AsLOV2 with different illumination times recorded at Diamond Light Source, beamline B21. The difference signal saturated for illuminations of more than five seconds, indicating that the maximum photoactivation yield of 100% had been achieved. **b.** The XFEL scattering was then scaled to the SAXS scattering from Diamond to facilitate direct comparison between them. **c.** The photoactivation of the XFEL was subsequently determined by comparing the ratio between the scaled XFEL and the saturated SAXS difference scattering. From this analysis, we determined that the photoexcitation yield at the XFEL was approximately 15%. The difference scattering obtained from the SAXS experiment was smoothed using a Savitzky–Golay filter with a window length of 120 and 1st degree polynomial.



Supplementary Figure 7. The loop region (shown in green) in the dark structure (PDB 7GPX) converts to a helix in the light following from AlphaFold modeling of State C.

References

1. Vakili, M. *et al.* 3D printed devices and infrastructure for liquid sample delivery at the European XFEL. *J. Synchrotron Radiat.* 29, 331–346 (2022).
2. Koliyadu, J. C. P. *et al.* Pump-probe capabilities at the SPB/SFX instrument of the European XFEL. *J. Synchrotron Radiat.* 29, 1273–1283 (2022).
3. Palmer, G. *et al.* Pump-probe laser system at the FXE and SPB/SFX instruments of the European X-ray Free-Electron Laser Facility. *J. Synchrotron Radiat.* 26, 328–332 (2019).
4. Allahgholi, A. *et al.* The Adaptive Gain Integrating Pixel Detector at the European XFEL. *J. Synchrotron Radiat.* 26, 74–82 (2019).
5. Daurer, B. J., Hantke, M. F., Nettelblad, C. & Maia, F. R. N. C. Hummingbird: monitoring and analyzing flash X-ray imaging experiments in real time. *J. Appl. Crystallogr.* 49, 1042–1047 (2016).
6. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021).
7. Wallner, B. AFsample: Improving Multimer Prediction with AlphaFold using Aggressive Sampling. *BioRxiv* (2022) doi:10.1101/2022.12.20.521205.
8. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501 (2010).
9. Grudinin, S., Garkavenko, M. & Kazennov, A. Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. *Acta Crystallogr. D Struct. Biol.* 73, 449–464 (2017).
10. Harper, S. M., Neil, L. C. & Gardner, K. H. Structural basis of a phototropin light switch. *Science* 301, 1541–1544 (2003).
11. Björling, A. *et al.* Structural photoactivation of a full-length bacterial phytochrome. *Sci. Adv.* 2, e1600920 (2016).