The short lived signalling state of the photoactive yellow protein photoreceptor revealed by combined structural probes

Pradeep L. Ramachandran[†], *Janet E. Lovett*[‡¥], *Patrick J. Carl*[⊥], *Marco Cammarata*[±], *Jae Hyuk Lee*[§], *Yang Ouk Jung*[§], *Hyotcherl Ihee*[§*], *Christiane R. Timmel*[‡], *Jasper J. van Thor*[†*]

[†]Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK.

[‡]Inorganic Chemistry Laboratory, Centre for Advanced Electron Spin Resonance, Department of Chemistry, University of Oxford, South Parks Road, Oxford, OX1 3QR, UK.

[¥]EaStCHEM School of Chemistry, University of Edinburgh, EH9 3JJ, UK

[⊥]Bruker BioSpin GmbH, Silberstreifen 4,76287 Rheinstetten, Germany.

[±] European Synchrotron Radiation Facility, Grenoble Cedex 38043, BP 220, France

S1) [§]Center for Time-Resolved Diffraction, Department of Chemistry, Graduate School of Nanosecience & Technology (WCU), KAIST, Daejeon 305-701, Republic of Korea

[†]Division of Molecular Biosciences, South Kensington Campus, Imperial College London, London SW7 2AZ, UK.

* corresponding author: j.vanthor@imperial.ac.uk , hyotcherl.ihee@kaist.ac.kr

Table of Contents for Supplementary Materials

Table S1: Percentage of I_2' formation calculated from the corresponding kinetic data.

Table S2: C_α-C_α distances for selected pairs from the ensemble PDB 3PHY and reported ground state distance distributions from DEER measurements.

Table S3: Values for NO·-NO· distances and their upper and lower bound used as restraints in structure calculations.

Table S4: Peak-list of observable resonances for the ground and I_2' ^{1}H-^{15}N HSQC spectra.

Table S5: Structure statistics for the ensemble of the ground state.

Supplementary Methods:

Site Directed Mutagenesis and Spin Labelling.

The *pyp* gene encoding PYP [1] was engineered to include an N-terminal His_6-tag and a 3C protease cleavage site (amino acid sequence: LEVLFQ↓GP). Single A5C and Q56C and double cysteine mutant constructs D48C/H3C, D48C/S8C, Q56C/E9C, Q99C/E9C, and Q99C/A5C were prepared, reconstituted with p-coumaric anhydride as described [2] and after removal of the His_6-tag with 5% (w/w) 3C protease (Novagen) in 1.5 M NaCl, 0.5 M Tris-HCl, pH 7.5 at 4 ºC for 16 hours, was purified using Sepharose fast flow Mono-Q (Amersham). Mass spectrometry confirmed samples to be singly reconstituted mutants PYP. After 3C protease cleavage, the polypeptide sequence has GPLGS N-terminal of the native translation start for samples that was used for DEER and NMR spectroscopy (but not those used for TR-SAXS/WAXS, which were native). 1-oxyl- 2,2,5,5-tetramethylpyrroline-3-methyl methanethiosulfonate (MTSSL, Toronto Research Chemicals) [3], was dissolved in 100% dimethyl sulfoxide (Sigma) and reacted at 75 μM in 10 ml with 15 μM double cysteine mutant PYP containing 50 mM Tris-HCl, pH 7.0, for >3 hrs at room temperature, after pre-reduction with 75 μM dithiolthreitol (DTT, sigma) in 50 mM Tris-HCl, pH 7.0 for 2 hrs.

The ground state visible absorption spectra for all single and double mutants were similar to that of wild type PYP with maxima at 446 nm, whereas mutants Q56C/E9C and Q56C showed minor additional absorption at 350 nm indicative of the presence of small populations of I_2'-like species in the dark resting state. Double mutants D48C/H3C, D48C/ S8C, Q56C/ E9C, Q99C/E9C, and Q99C/A5C and single mutants Q56C and A5C showed biphasic and delayed I_2' to ground state recovery kinetics compared to that of the wild type PYP having a monophasic recovery of τ= 0.46 s, at pH 7.0 and 293 K (Table S1).

Four-pulse DEER EPR spectroscopy

Double electron-electron resonance (DEER) measurements on doubly spin labelled mutants D48C/H3C, D48C/S8C, Q56C/E9C, Q99C/E9C, and Q99C/A5C were performed on a Bruker ELEXSYS E580 spectrometer operating at X-band frequency (9.5GHz) and fitted with a 3 mm ER4118X-MS3 split-ring resonator cooled to 60 K with helium . A microwave bridge equipped with a E580-400 ELDOR unit produced the microwave frequency. Samples contained 50 mM Tris-HCl, pH 7.0, 20 % (v/v) glycerol and 75 μM double labelled protein. Samples were loaded into a 3 mm quartz capillary (Bruker) and frozen in liquid nitrogen before insertion into the resonator. Preparation for measurement of the I_2' intermediate state included slow freezing in liquid nitrogen before, after and during illumination with a SCHOTT KL 2500 light source (250 W) coupled to a 15-mm diameter optic fibre bundle. The process of trapping the I_2' state consisted of several seconds of continuous illumination prior to and during freezing, that was determined to reach photostationary state from prior optical spectroscopy measurements.

Instruments were set up to perform the following pulse sequence: $\pi/2(\upsilon_{obs}) - \tau_1 - \pi(\upsilon_{obs}) - \tau_0 - \pi(\upsilon_{pump}) - (\tau_1+\tau_2-\tau_0) - \pi(\upsilon_{obs}) - \tau_2 -$ echo [4]. The $\pi/2$-pulse was set to 32 ns with π-pulses of 32 ns having twice the amplitude of the $\pi/2$-pulse. The ELDOR π-pulse was set to 12 ns. The time delays, τ_1 (200 ns) and τ_2 (2500 ns) were set to maximise signal to noise and to ensure proper collection of the dipolar time domain signal for the expected distances. Nuclear modulations were suppressed by averaging t_1 over 8 values starting at 200 ns using either 8 ns or 16 ns increments. A two-step phase cycle of the first pulse (+x-x) with signal subtraction (+s-s) was employed to correct for the receiver offset. The pump pulse was set to the central feature of the field sweep echo detected (FSED) spectrum and the observe pulse was set between -62 and -70 MHz from the pump pulse frequency. The entire pulse sequence was repeated with a repetition time of 2040 μs.

Analysis of the DEER data was performed using two different approaches. Firstly, Tikhonov regularisation suitable for solving the ill-posed problem of determining a distance distribution directly from the DEER time traces without any prior assumptions on the form of the distribution was implemented in DeerAnalysis2006 [5]. With the available data quality, the accuracy of the mean distance distributions was estimated to be ~1 Å, and the Gaussian widths at distances near 2-3 nm are determined with greater accuracy than those near 4 nm with measurements to 2 μs [5].

In the second approach (for short, the χ^2 method) the echo decay traces were fitted in the time domain with functions derived from Gaussian distances and distributions. As outlined in the methods section of the main manuscript, (typically) two Gaussians (labelled 1 and 2) were fitted to the time-domain DEER traces using a 5-dimensional parameter space (fwhh(1); mean distance (1); fwhh (2); mean distance (2) and weight of 2 (normalized by area)). The fit with the minimal (best) χ^2-per degree of freedom, $\chi^2_{pdf,min}$, was obtained. Next, all fits with a ($\chi^2_{pdf,min}$+1) or smaller were chosen for further analysis. Finally, from this set of data we selected the fits with the smallest and largest mean distance, respectively(in both Gaussians) keeping χ^2_{pdf} as close to ($\chi^2_{pdf,min}$+1) as possible. This analysis should provide a test for the distinguishibility of illuminated vs dark fits. The results are plotted for both the time domain DEER traces (A, C, E, G) and the distance domain distributions (B, C, F, H) for the 5 systems. The experimental DEER data are given for the irradiated and non-irradiated (dark) samples in black and blue, respectively. The best fits are given by red and green lines for irradiated and dark samples, respectively. Yellow lines show the Gaussians fits of shortest and longest distance, respectively (for both Gaussians) for the irradiated samples and cyan lines correspond to the results for the non-irradiated samples. The legend gives the following fitting parameters in order: Gaussian 1 mean (nm); Gaussian 1 fwhh (nm); Gaussian 2 mean (nm); Gaussian 2 fwhh (nm); Weight of Gaussian 2 (normalized by area). Two Gaussians

always fitted the experimental data with a smaller χ^2-per degree of freedom than that given by a single Gaussian. Each Gaussian function is defined by a mean and a full-width at half-height (fwhh). All Gaussians are normalised so that the total integrated area of the distance distribution is identical for all simulations.

The parameters were therefore mean and Full Width at Half Maximum (fwhm) for each Gaussian and a weighting factor. In general one or two Gaussian lineshapes were used, except D48C/S8C which required three Gaussians to account for the longer undefined distribution. This longer distribution is thought to be an artefact resulting from an uncleaved $(His)_6$-tag and is supported by evidence from preliminary experiments with $(His)_6$-tag inclusive Q99C/A5C samples where a longer undefined distance was also observed. However, the longer distance for Q99C/A5C was not observed in experimental repeats where the $(His)_6$-tag was proteolytically cleaved and removed by purification. In addition, modulation depth scaling of dark and illuminated data for $(His)_6$-tagged Q99C/A5C and D48C/S8C resulted in identical distance distributions for each sample, indicative of separate population unresponsive to light illumination (Fig 2G). Comparison of P and I_2' distributions between the $(His)_6$-tagged and untagged Q99C/A5C data did not indicate significant differences in distributions assigned to each population.

The best fit to the data was found by minimizing the χ^2 [6]. The χ^2 was calculated using a variance derived from the random noise in the data, as implemented in earlier work[7]. Due to some experimental artefacts in the DEER time traces towards the end of time some DEER data needed to be cut to stop a linear fit to the variance measure being a steep gradient such that it is negative at short times.

The average χ^2 per degree of freedom for the 10 data sets was 7.9 ± 2.3 indicating that the model fitting procedure was not perfect. This is not to be expected if only two Gaussians

are used, yet the simplicity of this two-Gaussian model allows physical insights not afforded by more complicated modelling (such as the Tikhonov regularisation procedure discussed above).As demonstrated in the SI, it was pleasing to note that within the constraints of the applied model the data fitting appeared not to yield significant residuals (Fig. S11).

By collating all the fits to the data that fell between the best χ^2 per degree of freedom and the value plus one found by exploring all fitting-parameter space we obtained an (over) estimate of the distribution in the fits and parameters. We used the fits from the DEER traces obtained with dark samples to attempt to fit the traces from irradiated samples and *vice-versa*. If there was any overlap, the analysis was deemed to have shown that there was not a significant difference between the dark and irradiated samples. The trend in the difference in the fitting parameters between the dark and the irradiated samples could then be used to determine whether the distance between the nitroxide spin labels increased or decreased allowing for the discussion of true differences between the P and I_2' states. These distance differences were then compared to the distances and distributions given by Tikhonov Regularisation analysis of the DEER time traces using DeerAnalysis2006 [5]. The χ^2 analysis provides a measure of the statistical significance of the amplitude of each of the components. This may be compared to Table S1 which indicates the concentration of the I_2' intermediate in the photostationay state as determined from spectroscopy. We find that the DEER determined amplitudes agree relatively well, or are less than the photostationary concentration possibly due to some ground state recovery during the freezing of the illuminated samples for DEER spectroscopy.

While the relative amplitude analysis (Table S1, Fig S11) provides strong indication for the assignment of distances to I_2' and P respectively, we extract no detailed information regarding the precise concentration of I_2' and P in the cryo-trapped photostationary mixture

from the χ^2 analysis, which in addition is affected by partial thermal recovery to P during the cryo-trapping, which cannot be precisely controlled under the conditions used

^{1}H-^{15}N Heteronuclear NMR spectroscopy of the P and I_2' states of PYP.

The ^{15}N-labelled, single cysteine mutants A5C and Q56C of PYP were overproduced in Minimal Media M9 containing 1 gram per litre $^{15}NH_4Cl$ (Sigma). Purified mutant PYP samples A5C and Q56C at a concentration of 25 μM and 50 μM, respectively, were exchanged into 50 mM Phosphate buffer, pH 5.75, containing 5% D_2O and a 5-fold molar excess of deuterated DTT.

Output from a 100 W mercury arc lamp (Nikon) was first passed through an infrared-absorbing filter, an infrared-reflector, then through a band pass filter (λ_{max}=455 nm, fwhm=30 nm) and coupled into an optic fibre with a length of 10 metres and diameter of 3 millimetres (Thor Labs). The output power at the end of the optic fibre was measured to be 14 mW, and 17.5 mW for samples Q56C and A5C, respectively. A Wilmad NMR tube (Sigma) containing the PYP sample was fitted with an adaptor connected to a 3 metre sleeve made of PTFE. After lowering the sample and adaptor into the spectrometer the optic fibre was fed into the NMR tube guided by the PTFE sleeve and submerged into the sample containing PYP. The percentage of I_2' state formed for each mutant was calculated using the measured initial rate of I_2' formation and rate of ground state recovery and modelled the optical penetration profiles of absorbing samples across the optical path, neglecting contribution from diffusion (Table S1). We note that for the conditions used for DEER and NMR, an equilibrium exists between the I_2 and I_2' states, which in the wild type is associated with a pK(a) of 6.4 [8]. The I_2' state is associated with conformational changes[9] and our analysis of NMR and EPR measurements in the photostationary state addresses two structural species where any possible accumulated I_2 state is interpreted to structurally resemble the ground state and the new population is assigned to the I_2' state.

All spectra were recorded on a Bruker Avance 500, equipped with a cryoprobe. The sample temperature was maintained at 293 K for the duration of all experiments. After insertion of the optic fibre into the sample a complete HSQC spectra of the dark state was collected followed by a complete spectra under continuous illumination, followed a final dark spectrum, agreeing closely with the first result. Manipulation and peak listing of HSQC spectra was carried out using the program SPARKY [10], using previously reported assignments for wild type PYP [11].

Due to the experimental set-up of EPR experiments, and to maximise signal-to-noise for NMR observables, the EPR and NMR experiments were conducted under different experimental conditions. Experiments EPR samples were conducted at a temperature of 50 K, and pH 7.00. This is in contrast to the experimental conditions of the NMR samples, which were pH 5.8, and temperature 293 K. For NMR, the experiment is conducted at pH 5.8 for optimal steady state accumulation of I_2'. For EPR, the experiment is conducted at pH 7.00 to minimise the proton modulations that would otherwise be largely apparent at lower pH. These conditions reduce artefacts and improves the confidence of data interpretation.

Time-resolved solution X-ray scattering

Protocols for TR-WAXS data collection and data processing have been reported in detail elsewhere [12]. The experiment was performed in beamline ID09B of ESRF. A 4.4 mM PYP solution in a Na-phosphate buffer with 20 mM NaCl at pH 7.00 was mounted in a quartz capillary with 1 mm diameter (Hampton Research), and temperature maintained at 293 K. The capillary was mounted in a goniometer so that its axis is perpendicular to the directions of the laser and X-ray pulses, which are also perpendicular to each other. To initiate a photoreaction, we used nanosecond laser pulses at 460 nm generated from a solid-sate

nanosecond laser (Vibrant, Opotek) utilizing an OPO setup pumped by a Nd:YAG laser. The laser beam was sent to the sample vertically (from above to down) and was focused to an elliptical spot of 1.28 × 0.5 mm^2 at the top part of the capillary and the energy density was 1 mJ/mm^2. Two seconds were given between successive laser excitation so that the PYP photocycle can be fully completely and fresh ground state can be recovered for each excitation. We have addressed the resulting level of photoconversion under these conditions by parameter analysis of the scattering differences, which were also supported by calculation, as discussed below (page 15, section Structure calculations). Polychromatic (3.5% band width) X-ray pulses at 18 KeV maximal in 16-bunch mode were sent to the sample in a direction perpendicular to both the capillary axis and the laser excitation and the scattered X-rays were collected in a CCD detector (MarCCD) and radially integrated from q=0.05647 to 2.0 $Å^{-1}$ and subsequently processed from q=0.05647 to 0.8 $Å^{-1}$ (where $q=4\pi\sin\theta/\lambda$) . The X-ray spot size at the sample was 0.06 × 0.1 mm^2 and its position at the sample was 150 μm down from the top surface of the capillary. Time-dependent WAXS images were collected at the time delays of -50 μs and 10 ms. The data at -50 μs contains structural information of ground state PYP and the data at 10 ms contributed by a mixture of ground state and the signalling state. A difference of these two signals brings out the change caused by photoexcitation and can be modelled by the signalling state minus the ground state. A total of 100 images were collected for each time delay and averaged to provide a sufficient signal to noise ratio. Two-dimensional images were converted to one-dimensional curve, I(q), as a function of momentum transfer q according to the established protocol [12]. To subtract contribution from heating, we conducted a separate control experiment with a buffer solution with 10 mM $KMnO_4$. In this case, the photon energy absorbed by MnO_4^- is eventually transferred to the buffer, and thus the difference curve contains features due only to the

heating of the buffer, and this was subtracted from the difference curve obtained from the protein excitation to yield a heat-subtracted difference curve, $\Delta I(q)$.

Structure calculations

Structure calculations were performed with CNS v1.1 [13] using the RECOORDscripts. A torsion angel dynamics (TAD) annealing procedure was used which included: 10,000 steps at 10,000 K, followed by an 8,000 step TAD cooling stage to a temperature of 2,000 K. This was followed by a 10,000 step first Cartesian cooling stage to 1,000 K and finally a 10,000 step second Cartesian cooling stage to 50 K. 300 structures were calculated for both P and I_2' states. Calculated structures were subjected to a water refinement stage, after which 20 of the lowest energy structures were obtained and examined for violations using WHAT-IF web server (http://swift.cmbi.kun.nl/WIWWWI/). Ramachandran analysis conducted used Molprobity (http://molprobity.biochem.duke.edu/) [14]. Chromophore coordinates were obtained from PDB 1XFQ and PDB 1XFN for the P and I_2' structures, respectively. average structures used for analysis were calculated in CNS using the accept.inp script.

Structural ensembles calculated using only NO-NO restraint for fully modelled spin labels were conducted by replacing the sites measured by DEER with a new residue called R1A. Spin labels coordinates were obtained from [15] and energy minimised using the PRODRG2 server [16]. The average Cα-NO. distance in the MTSSL label is approximately 7 Å[17]. Paramerisation of the MTSSL group was taken from [18]. The van der Waals radii of all atoms subsequent to S_δ on the spin label were set to zero in the non-bonded contact section within the CNS parameter file. This ensured that neighbouring spin labels would not sterically clash with one another.

Additional ab initio dynamical annealing calculations starting from linear popypeptide configuration that included all NO·-NO· pair distance restraints of the force field

parameterised MTSSL labels at positions H3C, A5C, S8C, E9C, D48C, Q56C and Q99C with their van der Waals radii set to zero, and did not include Cα-Cα pair distance restraints, generated structures with the general fold. 400 structures were calculated from an extended structure, and from these the lowest energy structure lacking violations was selected and used as the starting structure for the final round of structure calculations. 800 structures were calculated, and the 300 lowest energy structures were water refined and binned in to groups determined by the total number of NO-NO distances that were greater than the Cα-Cα distances for the five site pairs for each structure. Structures which had all five NO·-NO· distances greater than Ca-Ca distance for the label were analysed in terms of violations, and energies. Rejected structures from those calculations included those with clashes, restraint violations and a population with an extended N-terminal region. The latter population was rejected because of unphysical preferred MTSSL orientations towards the protein core leading to NO·-NO· pair distances smaller than Cα-Cα pair distances, and would not be supported by the measured H/D protection factors for these residues [19] and widths of distance distributions observed from DEER measurements (Fig. 2). Importantly the relatively narrow widths of the NO·-NO· pair distance distributions in the I_2' state argue against an extended and disordered N-terminal domain and in favour of a more compact and well ordered structure, that in general corresponds with solvent-oriented MTSSL labels and NO·-NO· pair distances greater than corresponding Cα-Cα pair distances for the sites selected.

In initial dynamical annealing calculations, no changes in preferred orientation of the spin label pairs were assumed for the I_2' state and Cα-Cα distance restraints were derived by adding to the ground state Cα-Cα distance for the light-induced difference of the experimental NO·-NO· pair distances. The use of these restraints in the calculations of the I_2' state for Cα-Cα coordinates led to one structural population that places the N-terminal domain in contact with the exposed chromophore binding cleft and additionally reduced

disorder in the ensemble (Fig. S5). Including force field parameterised MTSSL pairs, but placing no distance restraints on those, showed the possible spin label distributions for these ensembles (Fig. S2) which resulted in very similar systematic differences of the mean distances for both P and I_2' ensembles relative to the experimental inter-spin distance measurements. This indicates that for the I_2' state, a general protocol is allowed that combines DEER spectroscopy derived long range Cα-Cα distances with NMR derives short distance restraints.

Comparison with coordinates for the ground state structure known from NMR spectroscopy and X-ray crystallography showed that the MTSSL pairs added an average 9.9 +/- 6.5 Å to the mean Cα-Cα distances (Table S2). This variation is in agreement with a recent systematic study comparing pulse echo measurements of spin labels with C_β-C_β distances in proteins, that showed a mean variation of 6 Å relative to the C_β-C_β distances for a large number of individual measurements [20]. Specific spin label orientation was further addressed by structure calculations using dynamic annealing [13] of the ground state using previously reported restraints [11] (PDB 3PHY), which included force field parameterised MTSSL pairs in individual runs. Including no explicit NO·-NO· pair distance restraints on the MTSSL labels, the calculations provided a comparison of the mean statistical distribution of MTSSL NO·-NO· pair distances, that in general showed satisfactory correspondence with the experimental measurements with the exception of the Q56/E9 pair distance where the experimental distance was reproduced but not its statistical distribution which was found for shorter distance, which was also observed for this site pair in the I_2' state as well (Figs. 2, S2). Including no explicit NO·-NO· pair distance restraints for the spin label pairs in the structure calculations reduced the differences, generating increased number of structures in the ensembles which matched the experimentally determined distances more closely (Figs. 2,

S3). Moreover, the calculations indicate that the underlying Cα-Cα distance distributions are narrower then the measured inter-nitroxide distributions

Dynamical annealing calculations that included SAXS/WAXS data were done using methods and code described [21]. The code uses an addition to the CNS package to include X-ray scattering scattering data in the gradient that is applied during the dynamical annealing calculation. Thus, the dynamical annealing calculations that were done included energy terms from DEER-derived restraints, NMR derived restraints together with a constructed SAXS/WAXS curve for the I_2' state, and the resulting refined structural ensemble for I_2' was shown to simultaneously satisfy all three types of experimental restraints.

The experimental determination of the I_2' minus P difference scattering (Fig. 3) results from the flash induced formation of transient population of the intermediate in the interaction region with the X-ray beam. We addressed the photolysis level by calculation from known lifetime data and the photon flux density, which showed that these were in the saturation regime and resulted in full photoconversion to I_2'. We verified the resulting constructed species associated data and additionally investigated the sensitivity of the structure calculations to changes in the concentration of the intermediate, which was not found to be critical (see below).

The experimental geometry was with optical pumping using an imaged stripe onto the capillary that matches the diameter of the X-ray beam, which was at 90 degrees incidence. The beamline used provides a very good focus such that the X-ray beam probes a penetration depth of about 150 microns into the 1mm capillary. The precision of this arrangement is such that full overlap is guaranteed. Therefore, the phototransformation to the I_2' state is fully determined by the photochemical dynamics. The excited state decay is known to be 1.4 ps [22].The optical spot was elliptical with 1.28 — 0.5 mm^2 dimensions with an energy density 1 mJ/mm^2 at 460 nm. The power density is $1.14E^{15}$ photons / $6.4E^{-3}$ cm^2. The optical cross

section at 460 nm is $2.5E^{-17}$ cm^2, so the excitation probability in the dilute limit equals the cross section divided by the area multiplied with the number of photons = 4.4 / pulse. We obtained identical results with small spot size and power density up to an excitation probability of 7 per pulse. Therefore with a 7 ns pulse (5000 times the excited state lifetime), the interaction volume will be re-pumped to reach up to 7 times the primary quantum yield, which is 0.35 [23].

At 4.4 mM concentration the optical absorption is 2.2 for the 150 micron penetration depth. The interaction volume is $9.6E^{-5}$ cm^3, which contains $2.5E^{14}$ molecules. This is 4.6 photons per molecule. Therefore the nanosecond pulse will interact to the limit given for the dilute regime up to 4.6 times the primary quantum yield (4.6x0.35=1.6 > 1)

This shows that we likely operated in the saturation regime, with significant or full conversion to the I_2' state. Taking a lower limit of 0.1 for the primary quantum yield, which has been suggested recently, this would indicate at least 50% conversion. However, analysis of the TR-SAXS/WAXS data showed this to be an underestimate, and our results are consistent with the conventional value for the primary quantum yield of 0.35. For instance, using 50% conversion, the resulting pure scattering indicated an increase of the radius of gyration by 3-4 Å, which is unrealistic and cannot be supported by the protein molecular weight. In addition, the light induced changes of the radius of gyration of the M100L mutant of the full length PYP as observed by small angle X-ray scattering are 5%[24] and are also small for truncated variants[25]. Our results were therefore consistent by taking the photoconversion to be complete, with the I_2' having an Rg value of 14.6 Å, corresponding to an Rg of 13.3 Å for the dark ground state. We subsequently investigated the sensitivity of the resulting structure calculations, by varying the scaling to constructed curves that corresponded to Rg values between 0.9 and 1.7 Å more than the Rg for the ground P state that the difference curve was added to. The refined structures that used a larger fraction of the

experimental difference curve, which correspond to higher Rg values, typically resulted in increased maximal particle diameter, Dmax, in the region of 60 Å. The resulting structures were similar to the result that is presented in Figures 3 and 4, which had an Rg value 1.3 Å higher than that of the dark P state, but typically differed by further extension of the N-terminal region outside of the main core or the protein. Having determined that the sensitivity of the scaling procedure used is not unacceptably large, we selected a constructed curve for I_2' that corresponded to an Rg value of 14.6 Å, using a theoretical curve calculated from the average NMR structure of the P ground state (PDB) using CRYSOL that corresponded to an Rg of 13.3 Å. In summary, placing a limit on the Rg value of the I_2' state provides the most realistic approach considering the fixed molecular weight, which was independently supported by estimation of the photoconverted fraction from calculations described above.

The polychromicity of the X-ray spectrum used to record the difference signal was found to have negligible effect on the data processing, because of the relatively small band width, 3.5%, and its small proportional presence in the constructed curve for I_2'. Furthermore, from data processing and analysis of pure polychromatic (at 3.5% band width) and monochromatic scattering data from the P dark state, it was found that there is a relatively small difference between the resulting Rg values, the Tikhonov regularisation procedure and distance distribution functions. Therefore, we concluded that the small wavelength 'smearing' present in the constructed I_2' curve is not important or prohibitive for its effective use in dynamical annealing calculations. Processing of the curves used for the P dark state and I_2' state, available between 0.05647 to 0.8 $Å^{-1}$ indicate in the Guinier region an increased Rg value fitted for I_2' and an increased projected I_0 value, that is also visually seen in the data trend at low angle in the difference curve (Fig. 3, main manuscript). A deviation from the extrapolated curve between 0 and 0.05647 $Å^{-1}$ that might not be detected in these experiments

would arise from low resolution terms and specifically from aggregation, which we believe has not occurred under our conditions and concentration used. Therefore the increased I_0 value for I_2' is well supported by the data.

Calculations included force fields with energy terms from DEER and NMR derived restraints and a repulsive van der Waals term was scaled down by a factor 0.85[21]. The temperature was decreased from 1001 K to 1 K in 200 stages of 250 steps using a 1 fs time-step. SAXS/WAXS data was included in the range of 0.05647 to 0.8 $Å^{-1}$, using 100 data points and an 0.5 force constant. 37 glob types were defined, ignoring the chromophore atoms, and globic correction factors were calculated as described [21] in multiple rounds until convergence. Similarly, corrections for the surface hydration layer were refined in multiple cycles until converges, using CRYSOL and the procedure described [21]. A fully refined minimum structure was generated after 20 successive rounds of thermal cooling cycles, each set of rounds representing one cycle of overall refinement for the globic correction factors and the surface scattering correction. The final refined structure for the I_2' state, after full convergence, showed very close correspondence between the calculated curve together with both globic and surface scattering corrections, and the theoretical curve calculated by CRYSOL, that included refinement of the excluded volume and the contrast value of the surface hydration layer. Thus, after convergence the calculated scattering using the globic approximation method of the NIH CNS-SAXS method and the all-atom calculation implemented by CRYSOL, fully agreed. Coordinates were deposited to the pdb under accession number 2KX6.

Since the NIH CNS-SAXS method normalises the value of the zero angle cut off value, a theoretical calculation of the difference curve for comparison with the data shown in Figure 3 that input the coordinates of the P and I_2' states must account for the I_0 intensity and the integrated scattering intensity differences of the two states by adjusting the values for the

excluded volume the increased scattering intensity of I_2' at small angle. In order to reproduce the difference curve without fitting and applying an overall scaling factor, as is done in CRYSOL in fitting mode, the scattering of the dark P state is calculated with default values, using for the surface contrast $\delta\rho$=0.03 e/ Å^3 and the excluded volume 17470 Å^3 (average atomic radius r_0= 1.617 Å). The I_2' scattering is calculated with inputting the coordinates as deposited in the PDB database with accession code (2KX6), with using values for the surface contrast $\delta\rho$=0.014 e/ Å^3 and the excluded volume 16630 Å^3 and average atomic radius r_0= 1.880 Å. These values for the I_2' state were obtained from a fitting procedure using CRYSOL by reduction of the excluded volume until the I_0 matched the fitted value for the constructed I_2' curve, which was determined using GNOM [26-27].

The presented calculations are based on addition of the experimental difference measurement to the theoretical P ground state scattering. We additionally explored using available experimental data, published previously. The data were used to refine the P ground state structure together with known NMR derived energy terms[28], which will be reported elsewhere. It was previously concluded that based on this ground state scattering data, the best fit to the NMR structure [28] disregarded the contribution of surface hydration shell scattering [29]. There were small but no essential differences in the resulting structures for the I_2' state whether experimental or calculated scattering was used for the P dark state when constructing the I_2' scattering curve. Additionally when an I_2' curve was constructed by using a theoretical curve that neglected contrasts for the surface hydration layer, which was thus a fit to experimental data, this led to an I_2' structure that was also very similar to the structure refined on the basis of a theoretical curve that used a surface hydration layer contrast of $\delta\rho$=0.03 e/ Å^3 rather than zero. However in this case also the calculations of the I_2' structure refined and converged the globic correction factors but not the surface layer correction, which was confirmed by a data fit using CRYSOL which resulted in a zero value for the surface

layer contrast $\delta\rho$. Therefore, there is no strong effect of the weighting of the surface layer scattering on the final refined structure for I_2'.

Figure S1

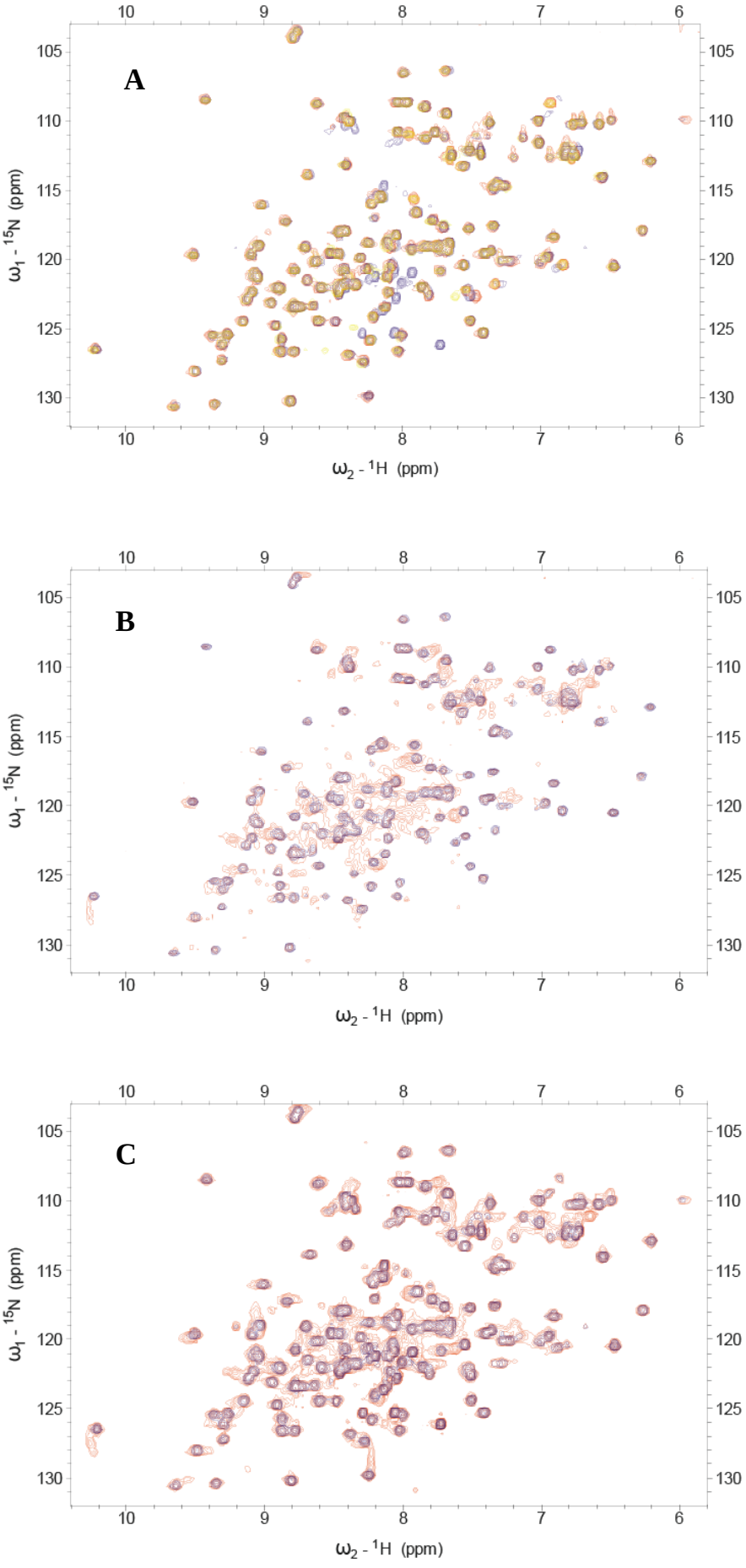

A
B
C
10
9
8
7
6
105
110
115
120
125
130
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)

Figure S1: ^{1}H-^{15}N HSQC spectra of the ground state P and and I_2' state. (A) Overlay of the ^{1}H-^{15}N HSQC spectra of ground state wild-type (red), and mutants A5C (gold) and Q56C (blue) PYP. Additional cross-peaks are observed for Q56C, attributed to a population of I_2'-like dark species that has maximal absorption at 350 nm. (B) Overlay of the ^{1}H-^{15}N HSQC spectra of the ground state (blue) and I_2' state of mutant A5C (red) PYP. (C) Overlay of the ^{1}H-^{15}N HSQC spectra of ground state (blue) and I_2' state of mutant Q56C (red) PYP. Native contact restraints and dihedral angle restraints were derived from the data (see also Table S4). The results show that the mutants used allow HSQC spectroscopy of the photoaccumulated I_2' state and that chemical shift changes and exchange broadening of I_2' relative to P dark state are similar to those of the wild type.

Figure S2

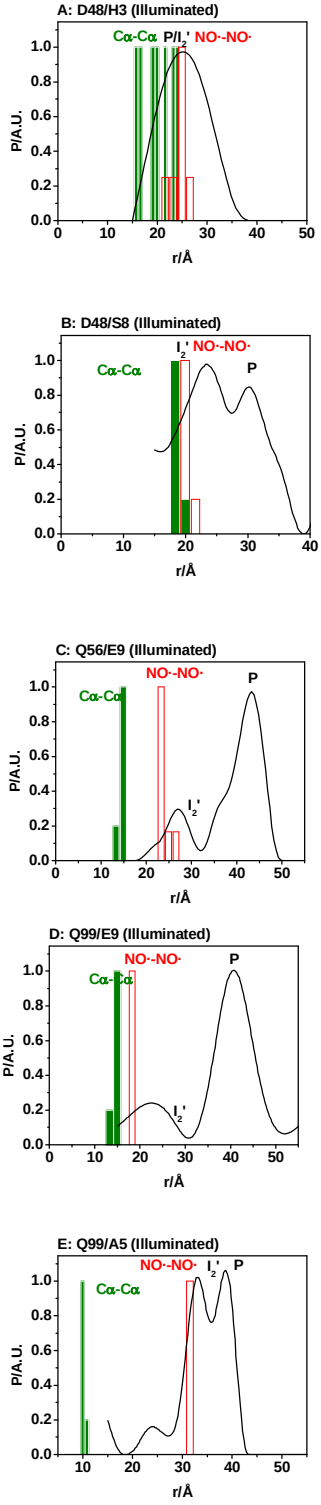

A: D48/H3 (Illuminated)
Cα-Cα
P/I2'
NO·-NO·
P/A.U.
r/Å
B: D48/S8 (Illuminated)
I2'
NO·-NO·
Cα-Cα
P
P/A.U.
r/Å
C: Q56/E9 (Illuminated)
NO·-NO·
P
Cα-Cα
I2'
P/A.U.
r/Å
D: Q99/E9 (Illuminated)
NO·-NO·
P
Cα-Cα
I2'
P/A.U.
r/Å
E: Q99/A5 (Illuminated)
NO·-NO·
I2'
P
Cα-Cα
P/A.U.
r/Å

Figure S2: Results from the structure calculations that used NMR derived energy terms, DEER derived NO·-NO· distance restraints for tyrosine residues in place of fully modelled spin labels, and TR-SAXS/WAXS data (see Figure S11 for corresponding structure). Histograms of C_α-C_α distances (green), and NO·-NO· distances (red) resulting from structure calculations are compared to the measured DEER distributions (black) for the site pairs D48/H3, D48/S8, Q56/E9, Q99/E9, and Q99/A5 in the I_2' state (A-E, respectively). Bin widths of histogram plots vary according to the number of accepted structures for each set of structure calculations on a site pairs.

Figure S3

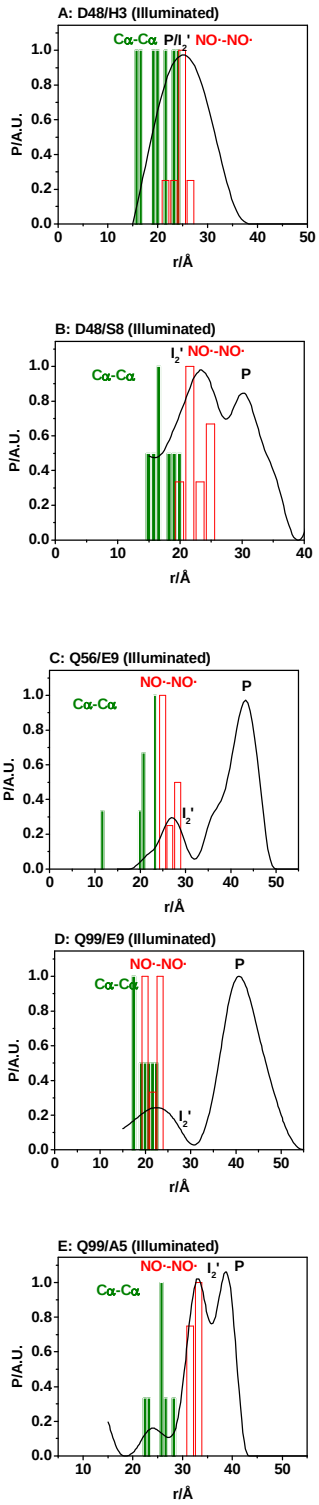

A: D48/H3 (Illuminated)
Cα-Cα
P/I2'
NO·-NO·
B: D48/S8 (Illuminated)
I2'
NO·-NO·
Cα-Cα
P
C: Q56/E9 (Illuminated)
NO·-NO·
P
Cα-Cα
I2'
D: Q99/E9 (Illuminated)
NO·-NO·
P
Cα-Cα
I2'
E: Q99/A5 (Illuminated)
NO·-NO·
I2'
P
Cα-Cα
P/A.U.
r/Å

Figure S3: Results from the structure calculations that used NMR derived energy terms and DEER derived NO·-NO· distance restraints for fully modelled MTSSL spin labels having atoms with zero Van der Waals radii. Histograms of C_α-C_α distances (green), NO·-NO· distances (red) resulting from structure calculations are compared to the measured DEER distributions (black) for the site pairs D48/H3, D48/S8, Q56/E9, Q99/E9, and Q99/A5 in the I_2' state (A-E, respectively). Bin widths of histogram plots vary according to the number of accepted structures for each set of structure calculations on a site pairs. Structures from ensembles that had any MTSSL spin labels with unphysical orientations were rejected. In general, the accepted structures had a higher RMSD overall, with the minority of the structures having multiple orientations for residues 1-8 and the majority having orientations similar to the SAXS and C_α-C_α restraint case.

Figure S4

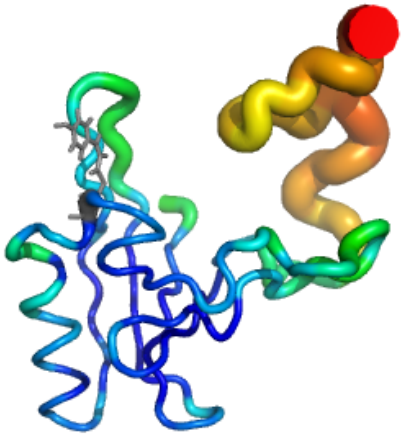


Figure S4: Representation of the average structure generated from ensembles which used only NMR derived energy terms, and NO·-NO· distance restraints from DEER measurements. Ensembles for fully force-field modelled spin labels having reduced VdW radii. Average structures were generated from an ensemble of 13 structures. Both the colour and the thickness of the tube visually represent the local RMSD values in the ensembles.

Figure S5

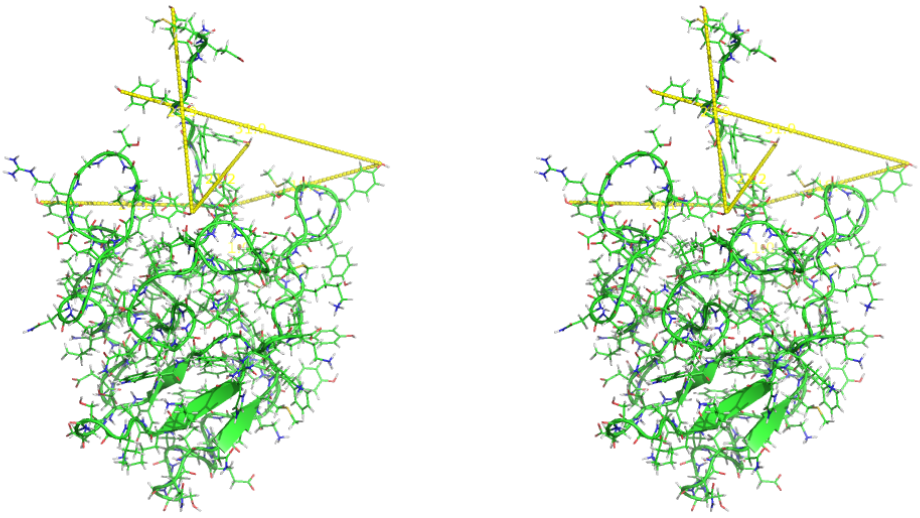


Figure S5: Internitroxide distances modelled as inter-tyrosine distances for a combined SAXS/WAXS, DEER and NMR refinement. The refinements were done by replacing spin labels with tyrosines and placing the distance restraints between the tyrosine OH groups. The refinements satisfied SAXS/WAXS, NMR and DEER restraints and the resulting structures were very close to the ensemble that used corresponding C_α-C_α distances, presented in Figures 3 and 4 of the main manuscript (PDB accession code: 2KX6.)

Figure S6

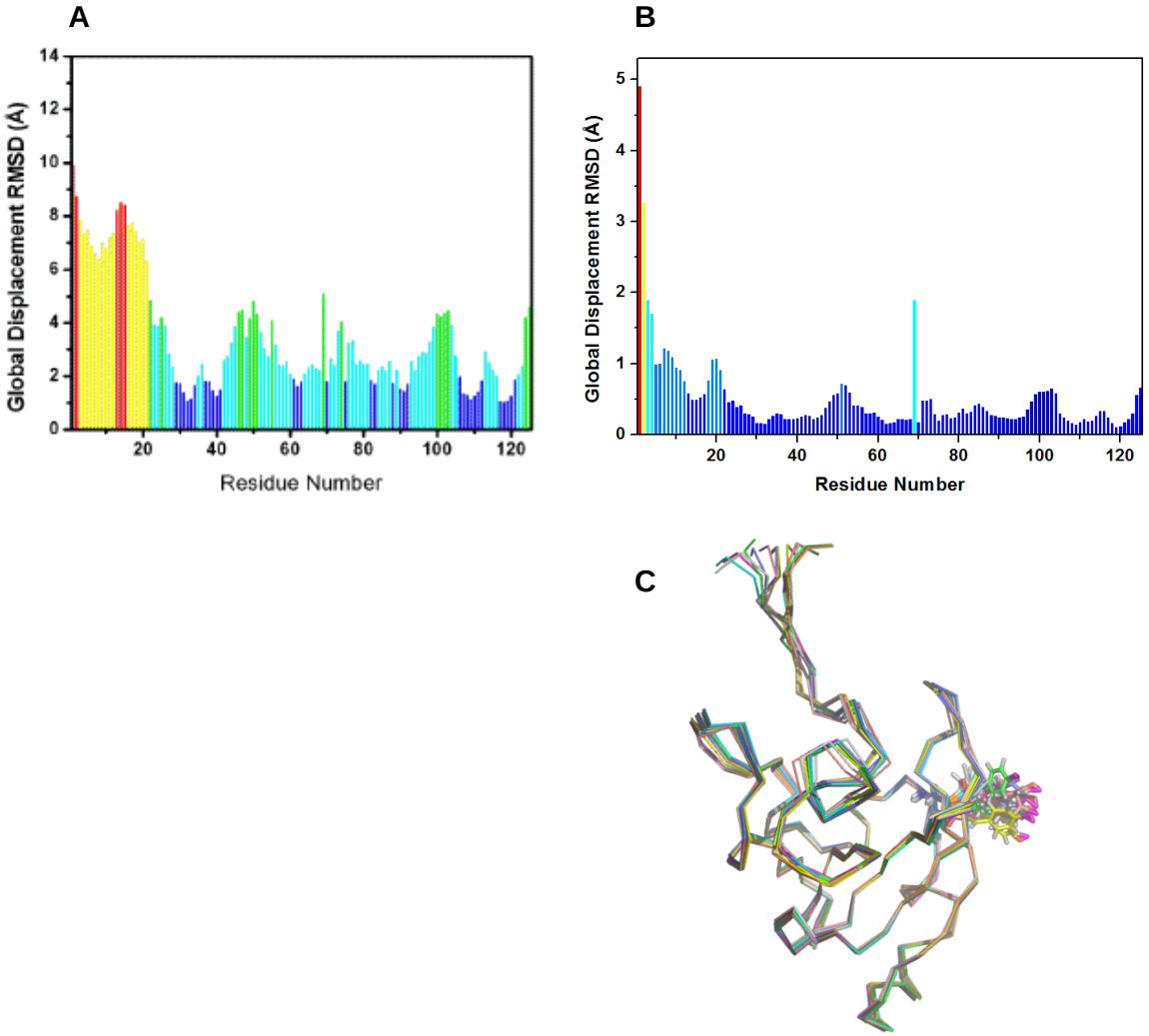

A
Global Displacement RMSD (Å)
Residue Number
B
Global Displacement RMSD (Å)
Residue Number
C

Figure S6: Backbone RMSD plot of the combined DEER, and NMR refined structural ensemble and of the combined SAXS/WAXS, DEER and NMR refined I_2' structural ensemble. (A) RMSD for the 10 lowest energy and accepted structures calculated using DEER and NMR restraints only. (B & C) The resulting ensemble RMSD for 14 lowest energy and accepted structures refined using SAXS/WAXS, DEER and NMR data is significantly lowered compared to the DEER and NMR derived structure.

Figure S7

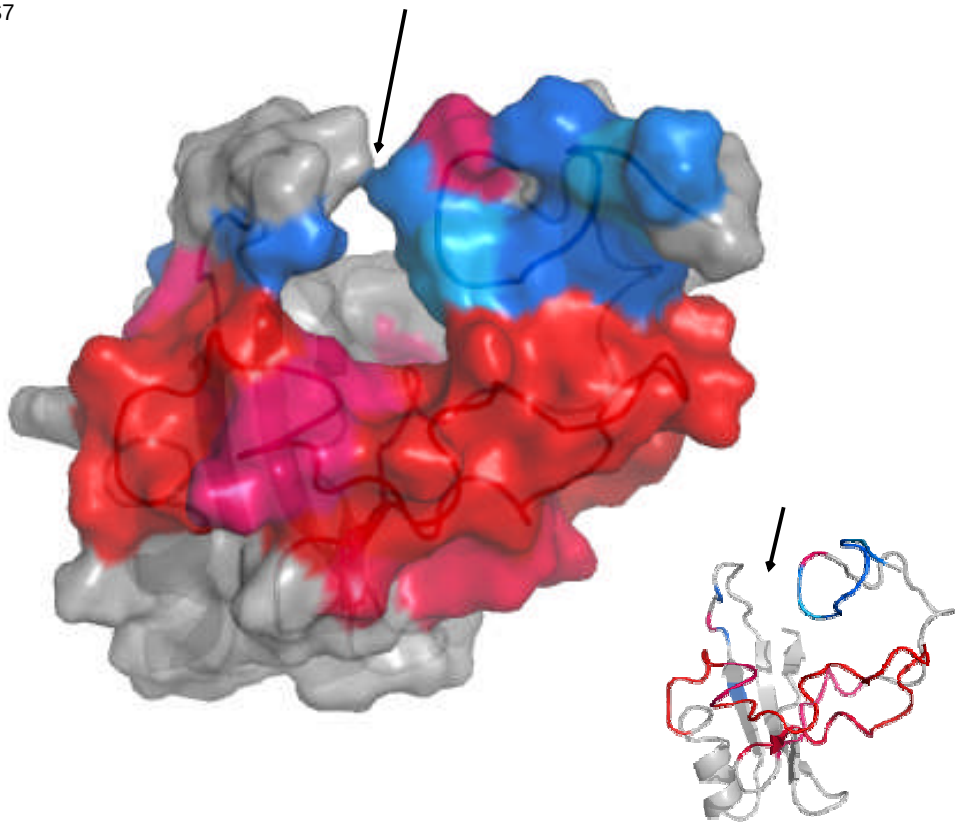


Figure S7: Analysis of light-induced changes (- Red / + Blue) of H/D protection factors determined by mass spectrometry reported by[19] mapped onto the molecular surface of the ensemble average structure of the NMR and DEER derived I_2' state. The exposed cavity (location indicated by arrow) correlates with substantially decreased protection factors, whereas the increased protection factors of the N-terminal region are in agreement with the transient interactions with the chromophore binding cleft.

Figure S8

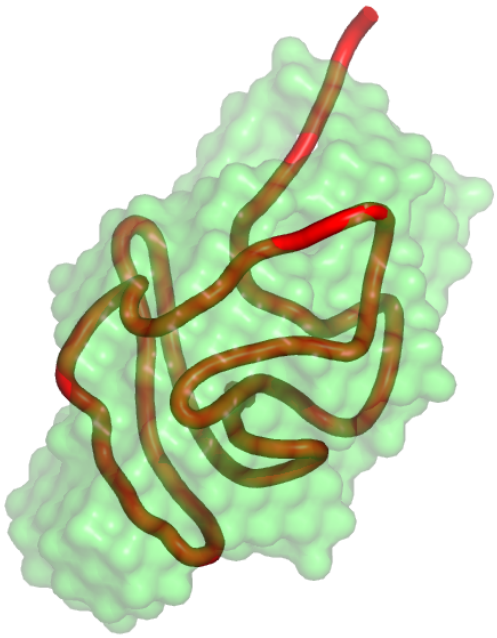


Figure S8: Ab-initio shape reconstruction of the I_2' scattering curve using DAMMIN. The combined DEER, SAXS/WAXS and NMR refined average structure is superimposed into the shape.

Figure S9

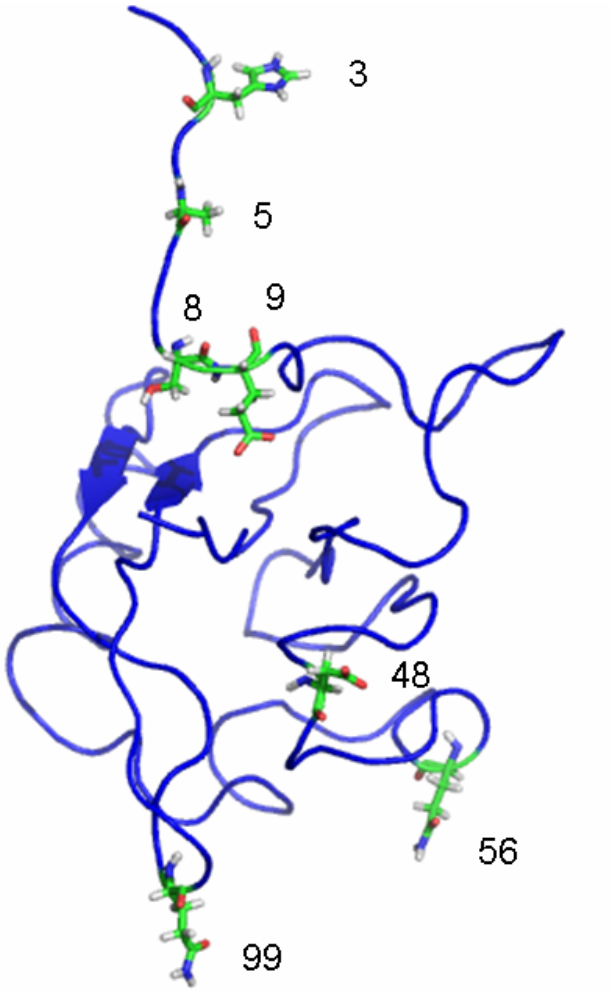


Figure S9: An example of an I_2' structure that satisfies SAXS/WAXS and NMR restraints but leaving out DEER derived restraints.

This test examined the case when SAXS/WAXS refinement was executed of the I_2' state that had a ground state P input structure and I_2' NMR derived restraints. The local search and refinement inhibited the opening of the chromophore binding cleft as a result of the SAXS/WAXS derived force constant, and extended the N-terminus to fit the SAXS/WAXS gradient with regard to the increased Dmax value for I_2' relative to P. The resulting inter-nitroxide distances, are far in excess of the experimental measurements – compare with structure in Fig. S5 which shows that spin label sites need to be present on one side of the molecule in contrast to this result. Calculations that attempted to model spin labels onto this conformation had unphysical orientation, with spin labels pointing towards each other, and could not match the experimental DEER distances. Additionally, the expected distance distributions would be expected to be increased relative to the P dark state for an extended and disordered N-terminus, which was also not observed. The protein structure and spin label orientation were similar to that shown in Fig. S10. This SAXS/WAXS refinement was also unstable and generated unfolded polypeptide further in the trajectory. For these reasons, these structures were therefore rejected.

Figure S10

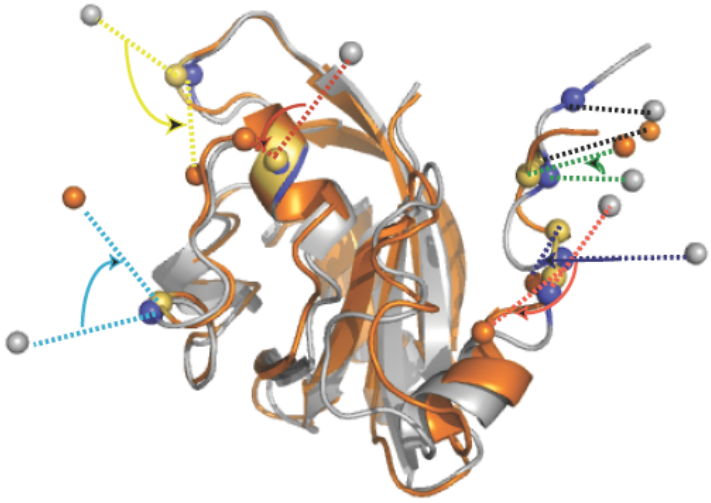
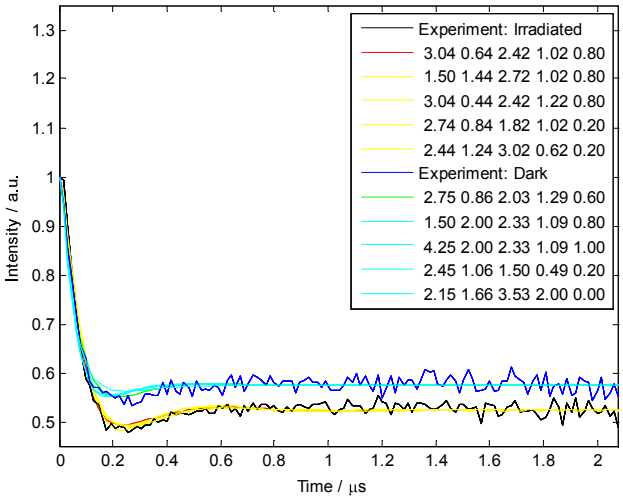


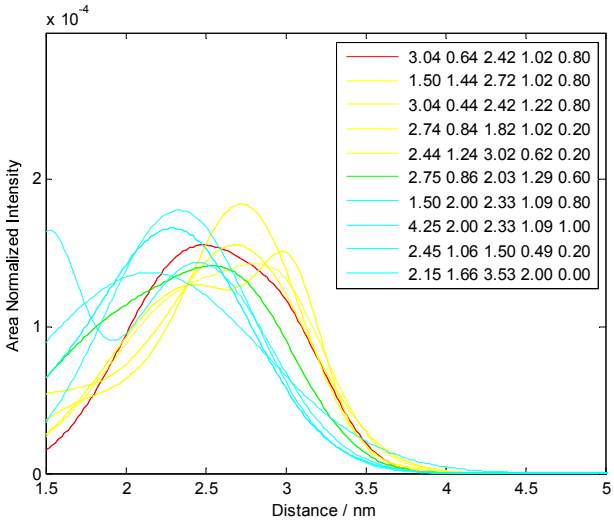
Figure S10: Structures from simulated annealing calculations that force a ground state P protein structure with NO·- NO· distances from I_2'. This scenario probes the unlikely case if light absorption would lead to spin label orientation changes exclusively. The I_2' NO·- NO· restraints were combined in simulated annealing calculation with NMR restraints for P with fully modelled and parameterised spin labels resulted in the labels assuming unphysical orientations, where the NO·- NO· distances were smaller than the corresponding C_α-C_α distances. In addition, the spin label distribution resulting from these calculations were very isotropic, and would be reflected by significantly narrower distribution widths in the experimental EPR data for I_2'. The arrows show the orientational changes in the calculation relative to the refined orientations of the spin labels in the ground state P together with internitroxide distances from the dark state. The resulting structure does not satisfy SAXS/WAXS data and is additionally rejected because light-induced exclusive spin-label reorientation is unlikely.

D48C/H3C

A

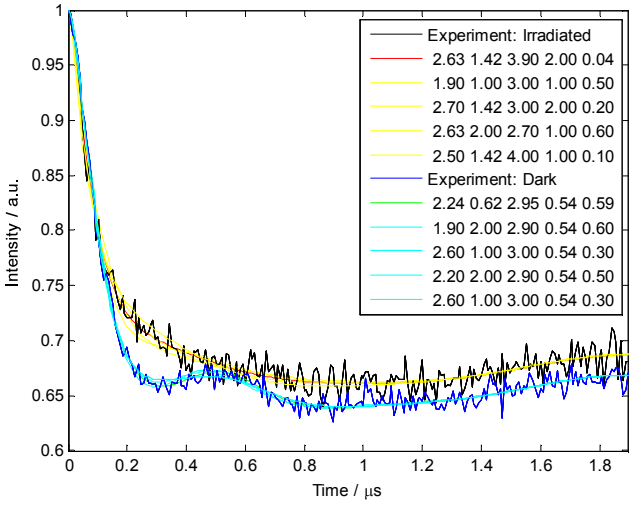


B



D48C/S8C

C



D

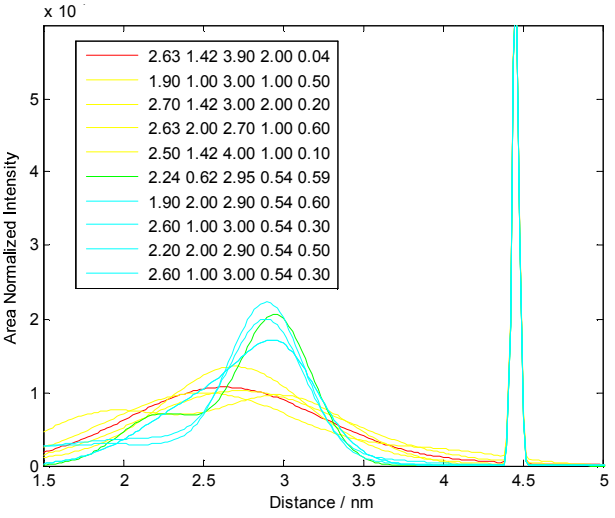
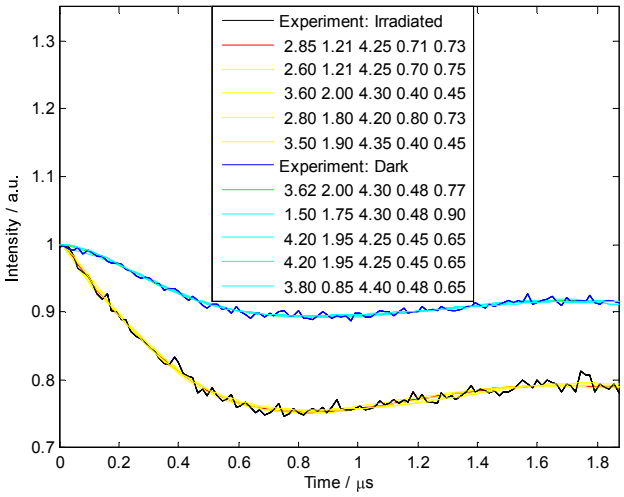


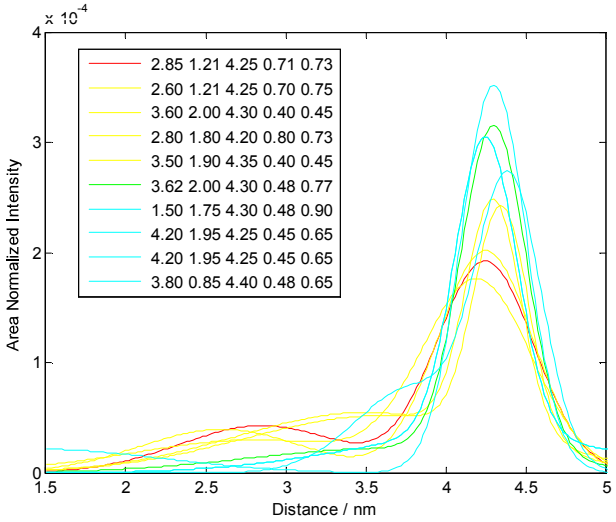
Figure S11 (continues on next page...)

Q56C/E9C

E

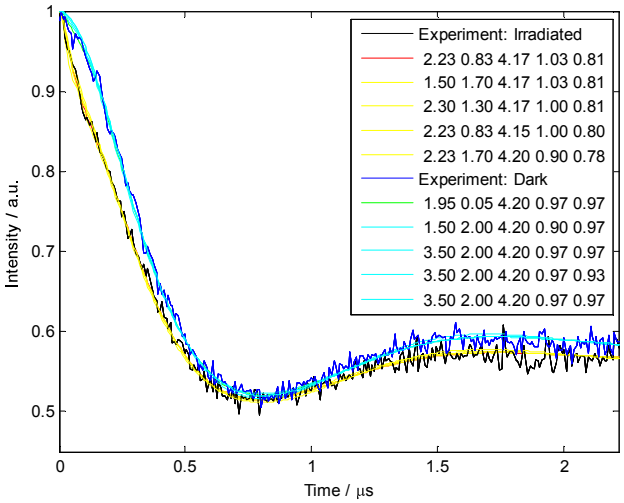


F



Q99C/E9C

G



H

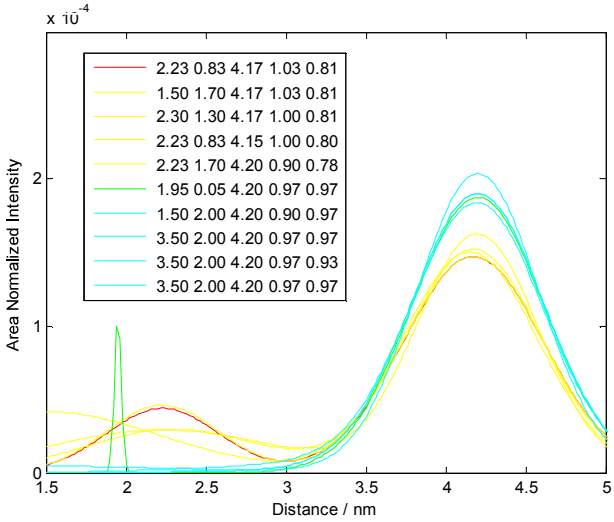
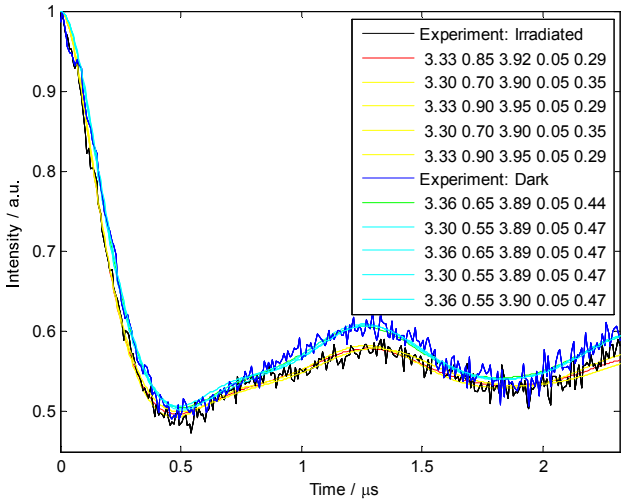


Figure S11 (continues on next page...)

Q99C/A5C

I


Experiment: Irradiated
3.33 0.85 3.92 0.05 0.29
3.30 0.70 3.90 0.05 0.35
3.33 0.90 3.95 0.05 0.29
3.30 0.70 3.90 0.05 0.35
3.33 0.90 3.95 0.05 0.29
Experiment: Dark
3.36 0.65 3.89 0.05 0.44
3.30 0.55 3.89 0.05 0.47
3.36 0.65 3.89 0.05 0.47
3.30 0.55 3.89 0.05 0.47
3.36 0.55 3.90 0.05 0.47
Intensity / a.u.
Time / µs

J

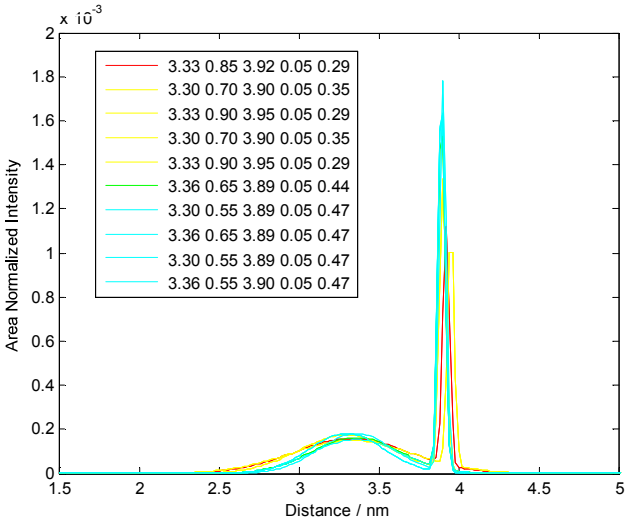

x 10^-3
3.33 0.85 3.92 0.05 0.29
3.30 0.70 3.90 0.05 0.35
3.33 0.90 3.95 0.05 0.29
3.30 0.70 3.90 0.05 0.35
3.33 0.90 3.95 0.05 0.29
3.36 0.65 3.89 0.05 0.44
3.30 0.55 3.89 0.05 0.47
3.36 0.65 3.89 0.05 0.47
3.30 0.55 3.89 0.05 0.47
3.36 0.55 3.90 0.05 0.47
Area Normalized Intensity
Distance / nm

Figure S11: Examples for plots of best fits and the fits with best χ^2-per degree of freedom + 1 sorted by change in the mean of the Gaussians.

The distribution in the distances from the Gaussian- χ^2 results was investigated. The variables for the Gaussians in the distance distribution which correspond to the region between the smallest (best) χ^2-per degree of freedom and the value plus one for the fit to the DEER time domain data were sorted from the total results set. The parameters corresponding to the smallest and largest mean distance for each Gaussian fitted[1] with the χ^2-per degree of freedom value closest to the best fit plus one were found. The results are plotted for both the time domain DEER traces (left column) and the distance domain distributions (right column). The experimental DEER data are given for the irradiated and non-irradiated (dark) samples in black and blue respectively. The best fits are given by red and green lines for irradiated and dark samples respectively. Yellow lines are used for the variation in the Gaussian means for the irradiated samples and cyan lines correspond to the results for the non-irradiated samples. The legend gives the following fitting parameters in order: Gaussian 1 mean (nm); Gaussian 1 fwhh (nm); Gaussian 2 mean (nm); Gaussian 2 fwhh (nm); Weight of Gaussian 2 (normalized by area). Two Gaussians always fitted the experimental data with a smaller χ^2-per degree of freedom than that given by a single Gaussian. Each Gaussian function is defined by a mean and a full-width at half-height (fwhh). The second (longer distance) Gaussian is multiplied by a weighting factor such that the total integrated area of the distance distribution is unity.

Figure S12

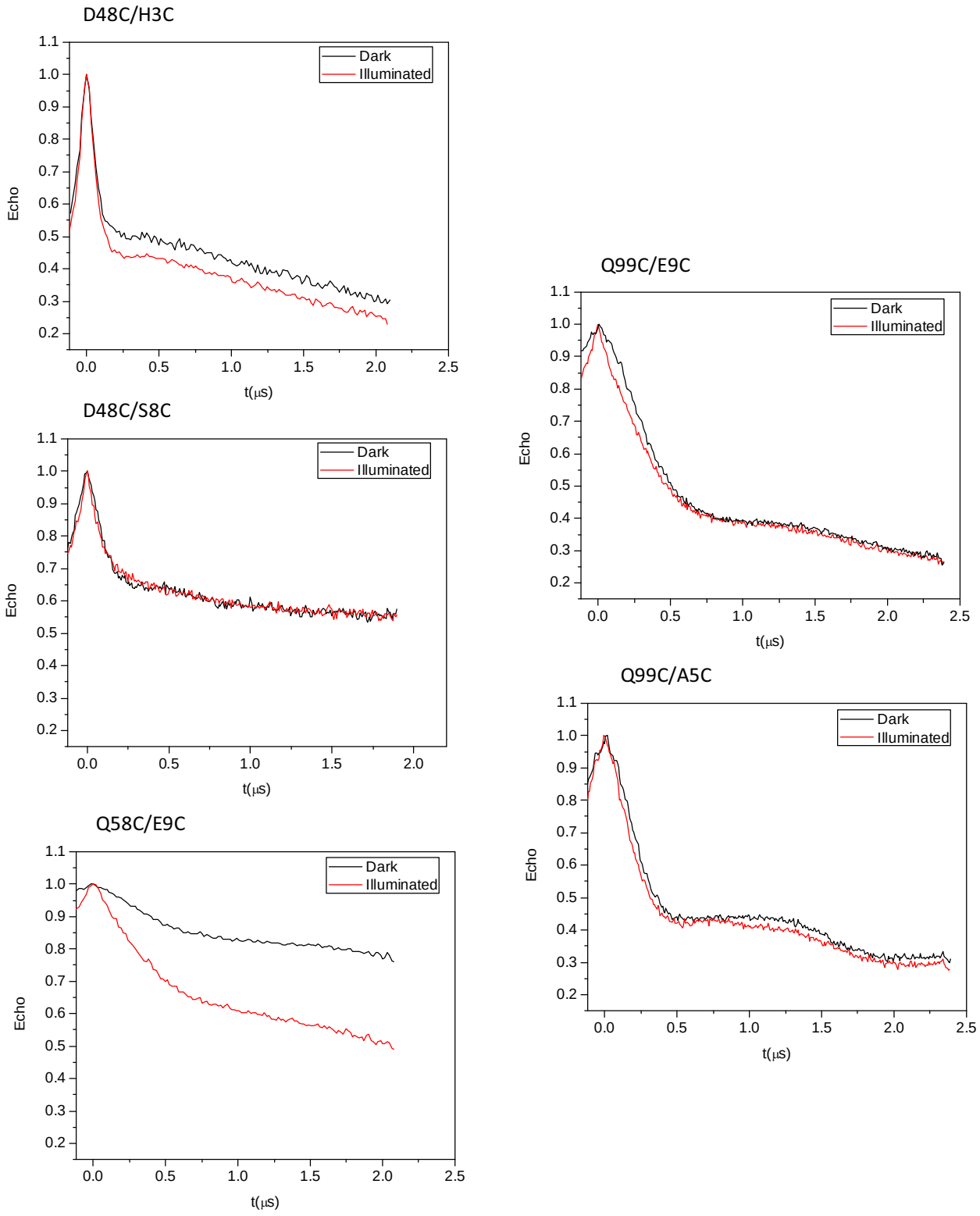

D48C/H3C
Dark
Illuminated
Echo
t(μs)
D48C/S8C
Dark
Illuminated
Echo
t(μs)
Q58C/E9C
Dark
Illuminated
Echo
t(μs)
Q99C/E9C
Dark
Illuminated
Echo
t(μs)
Q99C/A5C
Dark
Illuminated
Echo
t(μs)

Figure S12: Plots of raw DEER data.

Plots of the raw DEER data for D48C/H3C, D48C/S8C, Q56C/E9C, Q99C/E9C, and Q99C/A5C (A-F, respectively), are presented. In all plots, Black curves represent echo modulations of dark samples, and Red curves represent echo modulation of illuminated samples

Figure S13

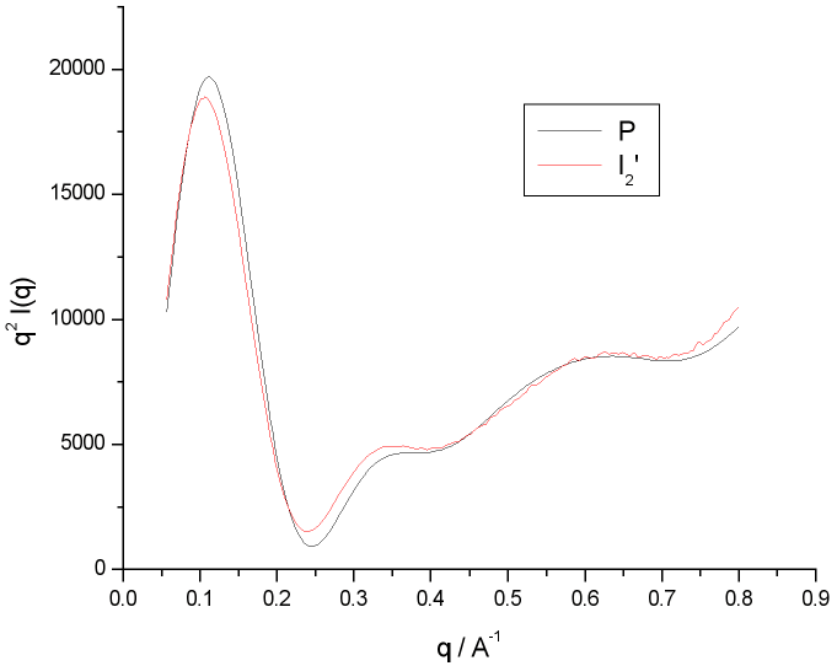

20000
15000
10000
5000
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
q^2 I(q)
q / A^-1
P
I2'

Figure S13: Kratky analysis of the experimental difference signal I_2' minus P.

Kratky analysis can identify unfolded samples [30]. In a Kratky plot q^2 I(q) is plotted for the q range. Globular proteins follow Porod's law and have bell shaped curves. Extended molecules lack this peak and have a plateau, typically at values of q=0.1 A^{-1} and higher [30]. Partially folded proteins are intermediate at high q range. The measurements for I_2' clearly show that for q values as high as q=0.8, the Kratky plot shows no increased disorder for the I_2' state. The Kratky plot for the P ground state is constructed from the calculated curve for the ground state coordinates, and shows amplitude at high q range in the Kratky plot resulting from the effective resolution of the reconstruction. However, the experimental difference curve for the I_2' state shows no increased signal even at high q. Furthermore, the width of the bell shaped curve is comparable, if not narrower, than the P ground state. This is experimental evidence for the ordered structure of the I_2' intermediate.

Table S1

Mutants used for DEER	**τ_1 /s (Amplitude)**	**τ_2 /s (Amplitude)**	**% I_2' ¶**
D48C/H3C	0.33 +/- 0.01 (0.68)	5.75 +/- 0.12 (0.32)	54
D48C/S8C	0.24 +/- 0.02 (0.85)	3.18 +/- 0.60 (0.15)	31
Q56C/E9C	0.44 +/- 0.01 (0.88)	5.90 +/- 0.48 (0.12)	63
Q99C/E9C	0.18 +/- 0.01 (0.88)	3.32 +/- 0.69 (0.12)	42
Q99C/A5C	0.32 +/- 0.01 (0.93)	4.22 +/- 0.61 (0.07)	62
Mutants used for NMR			
Q56C	0.98 +/- 0.05 (0.70)	4.7 +/- 0.60 (0.30)	30
A5C	0.38 +/- 0.1 (0.89)	2.01 +/- 0.43 (0.11)	17

Table S1: Calculated percentages of photo-accumulated I_2'.

Photocycle kinetics, calculated percentage of I_2' photo-accumulated with illumination at pH 7.0 and 293 K. Time constants are given for the biphasic ground state recovery $I_2' \rightarrow P$ in addition to their amplitudes. Conditions for obtaining and characterisation of photostationary states and were pH 7.0 and 293 K for DEER experiments and pH 5.8 and 293 K for NMR spectroscopy. Wild-type displayed a monophasic recovery with a τ_1 of 0.46 s, at pH 7.0 and 293 K. The final column presents the weighted Gaussian of I_2' normalized by area is included for comparison.

¶ Calculated equilibrium concentrations of I_2' used the rates of I_2' formation and ground state recovery and modelled the optical penetration profiles of absorbing samples across the optical path, neglecting contribution from diffusion. Rates of formation of I_2' were determined under representative illumination conditions using the initial rate method and were for double mutants D48C/H3C, D48C/S8C, Q56C/E9C, Q99C/E9C, Q99C/A5C 4.02, 1.66, 2.30, 4.00, 3.20 s^{-1}, respectively, and for the single mutants Q56C, and A5C were 1.46 and 1.26 s^{-1}, respectively.

Table S2

Site pairs	Mean Cα-Cα Distance (Å) (3PHY) [9]	DEER NO·-NO· Mean distance / fwhm (Å) – Dark	DEER NO·-NO· distance *minus* Cα-Cα Distance (Å)
	P		
D48/H3	24.2	23.8 / 11.6	-0.4
D48/S8	17.8	29.8 / 3.3	+12.0
Q56/E9	25.5	43.0 / 5.5	+17.5
Q99/E9	30.6	40.9 / 6.9	+10.3
Q99/A5	27.9	38.0 / 7.1	+10.1

Table S2. Results from DEER measurements of the trapped ground state and comparison to the mean C_α-C_α coordinates from the 3PHY ensemble.

Distances are reported as the average distance between pairs of C_α-C_α coordinates from the ground state ensemble, as determined previously from NMR spectroscopy (PDB 3PHY) [11]. The mean distance between MTSSL labels in the doubly labelled cysteine mutants in the ground state including their fitted widths, recognising reduced accuracy for the widths determined near 40 Å with DEER data extending to 2 μs.

Table S3

Site pairs	NO··NO·					
	Dark			Light		
	Mean distance (Å)	Upper bound (relative to mean distance) (Å)	Lower bound (relative to mean distance) (Å)	Mean distance (Å)	Upper bound (relative to mean distance) (Å)	Lower bound (relative to mean distance) (Å)
D48/H3	23.8	7.5	-7.5	25.2	10.2	-10.8
D48/S8	28.8	10.0	-10.0	23.6	13.6	-11.5
Q56/E9	43.3	11.3	-6.7	27.2	7.2	-5.0
Q99/E9	40.8	8.8	-9.2	21.8	9.8	-8.2
Q99/A5	38.9	9.4	-8.1	34.9	5.4	-7.1

Table S3: Long-distance restraints of NO·-NO· pairs with upper and lower bounds. These restraints were used in structure calculations of restrained spin label NO·-NO· pairs for the ground and I_2' states of PYP. Upper and lower bounds that were used resulted in distance distributions for spin label pairs that approximated the mean and width of the Gaussian distance distributions resulting from the fits of the dipolar evolution curves of the ground and I_2' state populations (Figs. 2, main manuscript).

Table S4: List of assignable cross-peaks for ground and I_2' states of mutants A5C and Q56C. Cross-peaks correspond to those reported for wild type PYP (10). Proton and Nitrogen chemical shifts are referenced relative to the water resonance at 4.76 ppm at 298 Kelvin.

Dark				Illuminated			
Residue	Number	N ppm	H ppm	Residue	Number	N ppm	H ppm
His	3	--	--	His	3	--	--
Val	4	--	--	Val	4	--	--
Ala	5[(1)]	129.97	8.242	Ala	5[(1)]	--	--
Phe	6	124.915	8.353	Phe	6	--	--
Gly	7	--	--	Gly	7	--	--
Ser	8	--	--	Ser	8	--	--
Glu	9	124.849	8.904	Glu	9	--	--
Asp	10	115.534	8.154	Asp	10	--	--
Ile	11	120.204	7.079	Ile	11	--	--
Glu	12	126.54	10.231	Glu	12	--	--
Asn	13	121.945	7.863	Asn	13	--	--
Thr	14	117.437	7.702	Thr	14	--	--
Lue	15	--	--	Lue	15	--	--
Ala	16	120	7.293	Ala	16	--	--
Lys	17	114.803	7.249	Lys	17	--	--
Met	18	119.4	7.382	Met	18	--	--
Asp	19	120.796	8.235	Asp	19	--	--
Asp	20	119.371	8.52	Asp	20	--	--
Gly	21	108.696	8.611	Gly	21	--	--
Gln	22	120.615	8.094	Gln	22	--	--
Leu	23			Leu	23	--	--

Asp	24	117.238	7.798	Asp	24	--	--
Gly	25	108.621	8.029	Gly	25	108.3	8.103
Leu	26	119.777	6.969	Leu	26	--	--
Ala	27	--	--	Ala	27	--	--
Phe	28	13.938	6.572	Phe	28	--	--
Gly	29	108.5	9.424	Gly	29	--	--
Ala	30	123.5	8.133	Ala	30	--	--
Ile	31	--	--	Ile	31	--	--
Gln	32	126.833	8.378	Gln	32	126.1	8.2
Leu	33	128	9.501	Leu	33	128.2	9.413
Asp	34	118.9	8.264	Asp	34	118.7	8.3
Gly	35	103.5	8.767	Gly	35	103.3	8.715
Asp	36	116.7	7.902	Asp	36	116.7	7.966
Gly	37	108.8	8.614	Gly	37	108.6	8.614
Asn	38	122.4	9.085	Asn	38	122.1	8.961
Ile	39	12.2	8.634	Ile	39	120.2	8.657
Leu	40	13.6	9.659	Leu	40	129.7	9.476
Gln	41	113.2	7.569	Gln	41	114	7.67
Tyr	42	126.6	8.892	Tyr	42	--	--
Asn	43	125.8	8.23	Asn	43	--	--
Ala	44	125.8	8.23	Ala	44	--	--
Ala	45	118.2	8.055	Ala	45	--	--
Glu	46	--	--	Glu	46	--	--
Gly	47	106.341	7.692	Gly	47	--	--
Asp	48	121.3	8.122	Asp	48	--	--
Ile	49	119.1	7.836	Ile	49	--	--
Thr	50	106.535	7.994	Thr	50	--	--
Gly	51	113.9	8.69	Gly	51	--	--
Arg	52	122.163	7.542	Arg	52	--	--

Asp	53	--	--	Asp	53	--	--
Pro	54	--	--	Pro	54	--	--
Lys	55	110.948	7.954	Lys	55	--	--
Gln	56[(2)]	115.59	7.919	Gln	56[(2)]	114.9	7.91
Val	57	108.728	6935	Val	57	--	--
Ile	58	120.5	6.88	Ile	58	120.4	6.817
Gly	59	116.076	9.021	Gly	59	115.935	9.09
Lys	60	119.117	7.748	Lys	60	119.09	7.752
Asn	61	119.678	9.515	Asn	61	119.705	9.508
Phe	62	130.385	9.354	Phe	62	129.51	9.148
Phe	63	112.861	6.212	Phe	63	--	--
Lys	64	111.227	7.838	Lys	64	--	--
Asp	65	114.48	7.315	Asp	65	--	--
Val	66	118.345	6.909	Val	66	--	--
Ala	67	119.364	7.936	Ala	67	--	--
Pro	68	--	--	Pro	68	--	--
Cys	69	117.998	8.413	Cys	69	--	--
Thr	70	115.919	8.227	Thr	70	--	--
Asp	71	124.384	7.513	Asp	71	--	--
Ser	72	119.9	8.316	Ser	72	--	--
Pro	73	--	--	Pro	73	--	--
Glu	74	113.159	8.426	Glu	74	--	--
Phe	75	117.787	7.524	Phe	75	117.8	7.612
Tyr	76	123.26	8.655	Tyr	76	--	--
Gly	77	104.005	8.796	Gly	77	--	--
Lys	78	120.493	6.477	Lys	78	--	--
Phe	79	122.352	8.105	Phe	79	--	--
Lys	80	118.1	8.462	Lys	80	--	--
Glu	81	118.9	7.685	Glu	81	--	--

Gly	82	109.004	7.851	Gly	82	108.7	7.781
Val	83	124.077	8.213	Val	83	124.317	8.179
Ala	84	120.4	7.561	Ala	84	120.7	7.629
Ser	85	110.732	8.039	Ser	85	111	8.008
Gly	86	110.79	7.771	Gly	86	110.45	7.7
Asn	87	117.546	8.111	Asn	87	117.461	8.116
Leu	88	122.408	7.821	Leu	88	121.642	7.773
Asn	89	125.727	8.888	Asn	89	125.727	8.888
Thr	90	119.647	9.097	Thr	90	119.706	9.051
Met	91	127.427	8.214	Met	91	126.925	8.355
Phe	92	120.757	8.783	Phe	92	--	--
Glu	93	119.165	8.713	Glu	93	119.67	8.746
Tyr	94	125.429	9.267	Tyr	94	--	--
Thr	95	117.521	7.348	Thr	95	--	--
Phe	96	126.597	8.787	Phe	96	--	--
Asp	97	120.8	7.731	Asp	97	120.746	7.642
Tyr	98	117.836	6.274	Tyr	98	--	--
Gln	99	127.305	9.31	Gln	99	--	--
Met	100	109.903	7.022	Met	100	--	--
Thr	101	119.635	8.466	Thr	101	--	--
Pro	102	--	--	Pro	102	--	--
Thr	103	--	--	Thr	103	--	--
Lys	104	130.21	8.817	Lys	104	--	--
Val	105	118.923	9.047	Val	105	--	--
Lys	106	121.361	9.047	Lys	106	122.8	9.055
Val	107	123.403	8.779	Val	107	122.13	8.805
His	108	126.004	9.303	His	108	126.116	9.415
Met	109	124.555	9.156	Met	109	124.645	9.222
Lys	110	122.094	8.892	Lys	110	121.3	9.302

Lys	111	126.547	8.038	Lys	111	126.764	7.974
Ala	112	125.499	8.02	Ala	112	125.324	8.059
Leu	113	121.53	8.704	Leu	113	--	--
Ser	114	109.45	7.679	Ser	114	110.365	7.673
Gly	115	108.664	7.972	Gly	115	108.6	7.79
Asp	116	118.834	8.114	Asp	116	118.8	8.1
Ser	117	111.3	7.027	Ser	117	111.2	7.076
Tyr	118	117.274	8.845	Tyr	118	117.028	8.743
Trp	119	122.9	9.13	Trp	119	122.421	9.23
Val	120	121.16	9.052	Val	120	121.102	9.182
Phe	121	125.432	9.364	Phe	121	126.158	9.414
Val	122	123.063	8.954	Val	122	--	--
Lys	123	124.345	8.607	Lys	123	--	--
Arg	124	--	--	Arg	124	--	--
Val	125	125.218	7.417	Val	125	--	--
Other chemical shifts							
		Nδ/Nε	Hδ/Hε				
Asn	13 $^{\delta}$	112.606	6.774				
Asn	13 $^{\delta}$	--	--				
Gln	22 $^{\varepsilon}$	112.03	6.8				
Gln	22 $^{\varepsilon}$	112.034	7.522				
Gln	32 $^{\varepsilon}$	109.825	6.527				
Gln	32 $^{\varepsilon}$	--	--				
Asn	38 $^{\delta}$	--	--				
Asn	38 $^{\delta}$	--	--				
Gln	41 $^{\varepsilon}$	110.032	7.382				
Gln	41 $^{\varepsilon}$	110.122	6.712				

Asn	43 $^{\delta}$	111.022	7.482				
Asn	43 $^{\delta}$	--	--				
Gln	46 $^{\varepsilon}$	--	--				
Arg	52	--	--				
Arg	52						
Gln	56 $^{\varepsilon(2)}$	113.239	7.568				
Asn	61 $^{\delta}$	112.555	6.935				
Asn	61 $^{\delta}$	--	--				
Asn	87 $^{\delta}$	112.39	6.837				
Asn	87 $^{\delta}$	112.371	7.44				
Asn	89 $^{\delta}$	111.23	7.14				
Asn	89 $^{\delta}$	--	--				
Gln	99 $^{\varepsilon}$	110.135	6.775				
Gln	99 $^{\varepsilon}$	110.178	6.587				
Arg	124	--	--				
Arg	124	--	--				

Table S4: List of assignable cross-peaks for ground and I_2' states of mutants A5C and Q56C. Cross-peaks correspond to those reported for wild type PYP (10). Proton and Nitrogen chemical shifts are referenced relative to the water resonance at 4.76 ppm at 298 Kelvin.

(1) Not present for mutant A5C.
(2) Not present for mutant Q56C.

Table S5 Structure statistics for ground state ensemble	
Energies (KJ/mol)	
Overall	-21373.88 +/- 341.19
VdW	-2019.99 +/- 92.47
Electrostatic	-23140.02 +/- 329.80
Distance Restraints **	
Experimental distance restraints	1352
Experimental dihedral restraints	75
Violations (mean and s.d.)	
Distance constraints (Å)	0
Dihedral angle constraints (º)	0
Max. dihedral angle violation (º)	0.40 +/- 0.10
Max. distance constraint violation (Å)	0.037 +/- 0.02
Deviations from idealized geometry	
Bond lengths (Å)	$0.01 +/- 3.40 \times 10^{-4}$
Bond angles (º)	$1.22 +/- 3.51 \times 10^{-2}$
Impropers (º)	$1.71 +/- 7.83 \times 10^{-2}$
Average pairwise r.m.s.d. (Å)	
Heavy	1.03 +/- 0.12
Backbone	1.70 +/- 0.13
Ramachandran Analysis ††	
Favoured regions	80.33%
Additionally allowed regions	97.5%
Disallowed regions	2.5%

Table S5: Statistics of the ground state ensemble calculations. These are based on previously reported restraints[11], and used the dynamical annealing protocol also used for calculations of I_2' ensembles. Results were also representative for calculations of the ground state that included parameterised MTSSL labels.

[**] Restraints were obtained from the PDB 3PHY.

[††] Ramachandran analysis was conducted on the calculated average structure from the ensemble using Molprobity (http://molprobity.biochem.duke.edu/).

References

(1) Hendriks, J.; Gensch, T.; Hviid, L.; van Der Horst, M. A.; Hellingwerf, K. J.; van Thor, J. J. *Biophys J* **2002**, *82*, 1632.

(2) Imamoto, Y.; Ito, T.; Kataoka, M.; Tokunaga, F. *FEBS Lett* **1995**, *374*, 157.

(3) Hubbell, W. L.; Cafiso, D. S.; Altenbach, C. *Nat Struct Biol* **2000**, *7*, 735.

(4) Pannier, M.; Veit, S.; Godt, A.; Jeschke, G.; Spiess, H. W. *J Magn Reson* **2000**, *142*, 331.

(5) Jeschke, G.; Chechik, V.; Ionita, P.; Godt, A.; Zimmermann, H.; Banham, J.; Timmel, C. R.; Hilger, D.; Jung, H. *Appl. Magn. Res.* **2006**, *30*, 473.

(6) Bevington, P. R.; Robinson, D. K. *Data reduction and error analysis for the physical sciences*; 2nd ed.; McGraw-Hill: Boston, 1992.

(7) Lovett, J. E.; Hoffmann, M.; Cnossen, A.; Shutter, A. T.; Hogben, H. J.; Warren, J. E.; Pascu, S. I.; Kay, C. W.; Timmel, C. R.; Anderson, H. L. *J Am Chem Soc* **2009**, *131*, 13852.

(8) Borucki, B. *Photochem Photobiol Sci* **2006**, *5*, 553.

(9) Brudler, R.; Rammelsberg, R.; Woo, T. T.; Getzoff, E. D.; Gerwert, K. *Nat Struct Biol* **2001**, *8*, 265.

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

(11) Dux, P.; Rubinstenn, G.; Vuister, G. W.; Boelens, R.; Mulder, F. A.; Hard, K.; Hoff, W. D.; Kroon, A. R.; Crielaard, W.; Hellingwerf, K. J.; Kaptein, R. *Biochemistry* **1998**, *37*, 12689.

(12) Cammarata, M.; Levantino, M.; Schotte, F.; Anfinrud, P. A.; Ewald, F.; Choi, J.; Cupane, A.; Wulff, M.; Ihee, H. *Nat Methods* **2008**, *5*, 881.

(13) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr D Biol Crystallogr* **1998**, *54*, 905.

(14) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. *Nucleic Acids Res* **2007**, *35*, W375.

(15) Langen, R.; Oh, K. J.; Cascio, D.; Hubbell, W. L. *Biochemistry* **2000**, *39*, 8396.

(16) Schuttelkopf, A. W.; van Aalten, D. M. *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 1355.

(17) DeSensi, S. C.; Rangel, D. P.; Beth, A. H.; Lybrand, T. P.; Hustedt, E. J. *Biophys J* **2008**, *94*, 3798.

(18) Murzyn, K.; Rog, T.; Blicharski, W.; Dutka, M.; Pyka, J.; Szytula, S.; Froncisz, W. *Proteins* **2006**, *62*, 1088.

(19) Brudler, R.; Gessner, C. R.; Li, S.; Tyndall, S.; Getzoff, E. D.; Woods, V. L., Jr. *J Mol Biol* **2006**, *363*, 148.

(20) Sale, K.; Song, L.; Liu, Y. S.; Perozo, E.; Fajer, P. *J Am Chem Soc* **2005**, *127*, 9334.

(21) Grishaev, A.; Wu, J.; Trewhella, J.; Bax, A. *J Am Chem Soc* **2005**, *127*, 16621.

(22) van Wilderen, L. J.; van der Horst, M. A.; van Stokkum, I. H.; Hellingwerf, K. J.; van Grondelle, R.; Groot, M. L. *Proc Natl Acad Sci U S A* **2006**, *103*, 15050.

(23) Groot, M. L.; van Wilderen, L. J.; Larsen, D. S.; van der Horst, M. A.; van Stokkum, I. H.; Hellingwerf, K. J.; van Grondelle, R. *Biochemistry* **2003**, *42*, 10054.

(24) Sasaki, J.; Kumauchi, M.; Hamada, N.; Oka, T.; Tokunaga, F. *Biochemistry* **2002**, *41*, 1915.

(25) Imamoto, Y.; Kamikubo, H.; Harigai, M.; Shimizu, N.; Kataoka, M. *Biochemistry* **2002**, *41*, 13595.

(26) Svergun, D. I.; Barberato, C.; Koch, M. H. J. *J. Appl. Cryst.* **1995**, *28*, 768.

(27) Svergun, D. I.; Pedersen, J. S.; Serdyuk, I. N.; Koch, M. H. *Proc Natl Acad Sci U S A* **1994**, *91*, 11826.

(28) Bernard, C.; Houben, K.; Derix, N. M.; Marks, D.; van der Horst, M. A.; Hellingwerf, K. J.; Boelens, R.; Kaptein, R.; van Nuland, N. A. *Structure* **2005**, *13*, 953.

(29) Kim, K. T., Zuo, Xiaobing, Tiede, M. David, Ihee, Hyotcherl *Bull. Korean Chem. Soc.* **2004**, *25*.

(30) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. *Q Rev Biophys* **2007**, *40*, 191.