

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Takala, Heikki; Björling, Alexander; Berntsson, Oskar; Lehtivuori, Heli; Niebling, Stephan; Hoernke, Maria; Kosheleva, Irina; Henning, Robert; Menzel, Andreas; Ihalainen, Janne; Westenhoff, Sebastian

Title: Signal amplification and transduction in phytochrome photosensors

Year: 2014

Version:

Please cite the original version:

Takala, H., Björling, A., Berntsson, O., Lehtivuori, H., Niebling, S., Hoernke, M., Kosheleva, I., Henning, R., Menzel, A., Ihalainen, J., & Westenhoff, S. (2014). Signal amplification and transduction in phytochrome photosensors. *Nature*, 509(8 May), 245-248. <https://doi.org/10.1038/nature13310>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Signal amplification and transduction in phytochrome

1 Source Data

The Source_Data.zip file contains the graphical data in ASCII format, the Charm force-field parameters for the biliverdin chromophore in Pr and Pfr, and pdb files of the proposed solution structures

2 Materials and Methods

2.1 Protein Preparation and Spectroscopy

2.1.1 Production and Purification

Photosensory core module (PAS-GAF-PHY) and chromophore-binding domain (PAS-GAF) from *Deinococcus radiodurans* was produced as (His)₆-tagged apoprotein in BL21 DE3 as described elsewhere.³⁷ After lysis, the sample was incubated in at least 10x molar excess biliverdin (Frontier Scientific) on ice overnight, and the holoprotein was Ni²⁺-affinity purified (HisTrap, GE Healthcare), followed by size-exclusion chromatography over a 26/60 Superdex 200 prep grade column (GE Healthcare) in buffers presented in Extended Data Table 2b. The purified protein was concentrated to 20–40 mg/ml and flash-frozen in liquid nitrogen.

2.1.2 Absorption Spectroscopy of Solutions

The visible light absorption spectra of the protein solutions were measured with a Perkin Elmer LAMBDA 850 UV-Vis spectrophotometer, from samples diluted with buffer (30 mM Tris·HCl, pH 8.0) to obtain an absorbance at 280 nm close to 0.1. Before measuring the absorption spectra, samples were irradiated for 15 min with red light (655 nm, 7 mW LED) to convert the protein to the Pfr state, or with far-red light (780 nm, 9 mW LED) to convert the protein to the Pr state. All spectra were measured at ambient conditions.

2.1.3 Absorption Spectroscopy of Crystals

Protein crystals were mounted in cryoloops in reservoir solution (see section 1.3.1) with 15 % glycerol and the cryo-loops were flash-frozen before acquisition. Absorption spectra were collected using a home-built micro-focusing device connected by optical fibre to a standard deuterium lamp and a diode array spectrometer (Ocean Optics). All measurements were conducted under cryo conditions (123 K). At this temperature, no photoconversion due to the probe light was detected for exposure times of at least 15 min. Crystals were solubilized by mixing them with small aliquots of buffer (30 mM Tris pH 8.0), followed by filtering with 0.22 µm centrifugal filters (Amicon Ultrafree, Millipore).

To test if room temperature illumination leads to photoswitching in the *illuminated* crystals, the cryo-stream was temporarily pointed away from the crystal loops. At the same time the detection lamp was covered in order to exclude unwanted photoconversion. Crystals in the cryo-loop were then

illuminated with LED lights (as described in the section 1.1.2) for <1 min, until the cryo-stream position was restored and the optical absorption was measured.

2.1.4 Transient Absorption Spectroscopy

Optical absorption kinetics were measured using a modified Luzchem laser flash-photolysis system (mLFP111 prototype from Luzchem Co.).³⁸ The excitation pulses were at 700 nm, had a power density of 1.2 mJ cm^{-2} , a duration of 10 ns, and a repetition rate of 0.25 Hz. The pulses were generated by a tuneable Ti:sapphire laser (Solar TII CF125 with built-in SHG), pumped by the second harmonic of a Nd:YAG laser (Solar TII LF-117). The continuous wave probe light was from a tungsten halogen lamp (AvaLight-HAL, Avantes). Interference filters were inserted before and after the sample and 100 measurements were averaged at each wavelength. The optical response were recorded by a PMT (Hamamatsu, R7400U-20) and digitized in an oscilloscope (Tektronix, TDS3032B, 300 MHz). The sample resided in a vertically mounted glass capillary with an inner diameter close to 1.1 mm (VITREX, micro-haematocrit) with OD_{700} of about 0.2–0.4 1/mm. To avoid excessive sample degradation, the sample solution (volume 400 μl) was cycled using a peristaltic pump (Ismatec, Reglo Digital) at a flow rate of 0.5 ml/min through a glass reservoir, the capillary, and connecting Teflon tubing (1 mm inner diameter). A far-red diode (750 nm, 3 mW, Leading-Tech Laser Co.) was used to transform the sample to the Pr state by constantly illuminating the sample through the Teflon tubing. A typical measurement cycle consisted of 3 s of optical measurements time without pumping and 1 s of pumping.

2.2 Solution X-ray Scattering

2.2.1 Millisecond time-resolved WAXS at cSAXS:

Sample Preparation: Phytochrome samples (~20–40 mg/ml) were thawed just before measurement, and filtered with 0.22 μm centrifugal filters (Amicon Ultrafree, Millipore). The samples' steady-state absorption spectra were recorded with a Nanodrop (ND-1000) spectrophotometer before and after measurements to verify sample quality. See sample details in Extended Data Table 2b.

Data acquisition: Real-time measurements were performed at the beamline cSAXS-X12SA of the Swiss Light Source, using the rapid-readout method.²⁸ A two-dimensional Pilatus 3K-W pixel detector placed 190.5 mm from the sample position was read out at 100 Hz, allowing for 7 ms of photon counting and 3 ms readout times. Each series of X-ray exposures with a monochromatic X-ray beam at 11.0 keV lasted for 6 s and thus comprised 600 images. During measurement, the sample was alternately exposed to two flashes from a red continuous wave (CW) DPSS laser (5 ms, $\sim 4.4 \text{ mJ/mm}^2$, 671 nm, Altechna) and three flashes from a CW far-red diode laser (10 ms, $\sim 8.4 \text{ mJ/mm}^2$, 750 nm, Leading-tech laser co.), starting and ending with the latter. The length of the flashes was controlled by fast shutters (Uniblitz) and attenuated with variable neutral-density filters. Flashes were separated in time by at least 400 ms, more than enough for the photocycle to complete. All experiments were carried out in “toggled laser mode” with every other measurement series “dark” (only first recovering flash issued), and the rest “light” (pulse sequence as just described).

Independent measurements of the X-ray scattering response to heating were carried out by applying infrared pulses from a fibre-coupled gated diode laser (1470 nm, 20 W, 10 ms, Lumics). The fibre output was very divergent and the energy density at the sample position was not measured, but the

exposure time was adjusted to give water heating response (visible at $q > 1.5 \text{ \AA}^{-1}$) of the same order as that observed with optical excitation. These experiments were also done in toggled laser mode.

The sample environment was a quartz capillary with a diameter close to 1 mm (Hampton Research), vertically mounted in the X-ray beam path and connected to PEEK tubing (Vici) using inert Kynar heat-shrink tubing (Elfa). To dilute the effect of radiation damage, the sample was pumped in between exposure series using a syringe pump (Cetoni) connected with Teflon and PEEK fittings. Several hundred microliters of sample were used per batch. The solution was stationary during X-ray exposure. The laser spots (horizontal x vertical: 1000 x 400 μm and 2500 x 300 μm FWHM for red and far-red lasers, respectively) were overlapped at the sample position together with the X-ray beam. The X-ray beam probed the capillary at an offset from its centre so as to find an empirical compromise between X-ray path length and the limited penetration depth of the visible laser light.

Data reduction: All detector images were radially integrated to one-dimensional scattering curves using scripts available at the beamline (Matlab). Occasional air bubbles in the sample caused outliers that were removed by manually inspecting each data series. As a rejection criterion we used the absolute scattering signal in the water region ($q > 1.5 \text{ \AA}^{-1}$). Around 18% of the “light” images had to be discarded because of air bubbles in them or the associated “dark” images.

The difference signal $\Delta S(q, t, r)$ (where $q = 4\pi \sin(\theta)/\lambda$ with 2θ the scattering angle, t is time, and r indexes the “light” exposure series from 1 to R) was calculated by subtracting from each “light” image the average of the corresponding “dark” images in the series before and after. In this way, drifts due to radiation damage and build-up of degradation products on the capillary wall effectively cancelled out together with any heating effects caused by X-ray radiation. The response due to laser-induced heating of the solution was removed by identifying a fingerprint response $\Delta S_{heat}(q)$ from the independent heating experiments. The time trace of the heating response from the optically excited experiments was found from the q region ($1.5 \text{ \AA}^{-1} < q < 2.2 \text{ \AA}^{-1}$). In this region the signal can be assumed only to originate from the solvent. $\Delta S_{heat}(q)$ was then subtracted from the data to yield the heating corrected protein response.

The resulting outlier- and heating-corrected data were then averaged for each construct to yield $\Delta S(q, t)$, with t consisting of 600 time points 10 ms apart. The PAS-GAF construct was averaged over $n = 1020$ “light” image series, for PAS-GAF-PHY the $n = 734$.

2.2.2 Microsecond time-resolved WAXS at BioCARS

Sample preparation: As above and in Extended Data Table 2b.

Data acquisition: Time-resolved solution scattering was measured using the pump-probe method as implemented at beamline 14-ID-B of the Advanced Photon Source.³⁹ A pink beam with a peak energy of 12.0 keV (FWHM 0.32 keV) was used, in 11-bunch mode when time delays allowed it. The experiment was repeated at 10 Hz, which is slow enough for the photocycle to complete, and the X-ray scattering was integrated on the detector so as to achieve maximum readings within its dynamic range. Activation of the sample was achieved with the picosecond laser pulse (680 nm, 2 mJ/mm², 2 ps). A recovering laser pulse (750 nm, <2.6 mJ/mm², typically 10-20 ms) was delivered 40 ms after the excitation pulse. The experiment was carried out in toggled laser mode as described above, however, “dark” images were recorded as negative time delays (-10 μs) between X-ray and laser pulse. At least every fourth image was a dark.

The sample environment was identical to that used at cSAXS (above) but with the capillary mounted horizontally. The sample was continuously pumped (typically at 0.5 $\mu\text{l/s}$) during acquisition. To avoid drifts it was found to be necessary to have a recovering laser spot of >1.5 mm diameter, much larger than the ps excitation laser spot (500 x 100 μm , long in the X-ray beam direction) and the X-ray beam size (95 x 30 μm).

No independent heating measurements were made as the recorded difference data contained no sign of water heating response (as judged from ΔS in the range $1.5 \text{ \AA}^{-1} < q < 2.2 \text{ \AA}^{-1}$). This is because the “dark” images contain the same amount of heating as the “light” images.

Data reduction: Occasional air bubbles rendered some detector exposures useless. For each series of images (where a series corresponds to a certain capillary position and sample) a good indication of air bubbles is deviation of the absolute scattering intensity at $q > 1.5 \text{ \AA}^{-1}$. Thus all images where the average scattering over the range $1.64 \text{ \AA}^{-1} < q < 1.97 \text{ \AA}^{-1}$ deviated from the series median by more than 2% were rejected. Further, since pressure artefacts were occasionally introduced by reversing the pump, light images which lay between a pump reversal event and a dark image were rejected. In total, 1528 images were included in the analysis.

Difference scattering was calculated by subtracting from each light image an average of the adjacent dark images, with no normalization of the absolute data. This average was weighted to reflect the time passed between the dark image before the light image of interest and the dark image after it. This effectively cancels out drifts due to radiation damage and build-up of degradation products.

Difference scattering curves corresponding to the same time delays were directly averaged together, resulting in the time-resolved scattering data shown in Figure 1 and Extended Data Figure 1. These curves were subject to analyses (Extended Data Figure 1) using Python and its NumPy extension.

2.2.3 Static SAXS at BM29

Sample preparation: As above and in Extended Data Table 2. Samples were either kept in the dark or illuminated for 15 minutes using a 655 nm light-emitting diode (5 mW), giving samples rich in the Pr and Pfr states respectively. Sample loading and acquisition was conducted in dark, and illuminated samples were measured immediately (delay between illumination and data acquisition about 2 min). All samples were diluted in a series as described in Extended Data Table 2 and Extended Data Figure 5b. Bovine serum albumin (Sigma-Aldrich) in 3 mg/ml concentration was used as a standard protein for calibration.

Data acquisition: Static Small-Angle X-ray Scattering (SAXS) data were collected at the recently upgraded beamline BM29 (BioSAXS) at the European Synchrotron Radiation Source.⁴⁰ A Pilatus 1M detector was positioned at 2.847 m from the sample which was contained in a quartz capillary (diameter 1.8 mm). A monochromatic X-ray beam at 12.5 keV was used and data was collected up to 4.5 nm^{-1} .

Data reduction: Radial integration of detector images was done using the automatic beamline software. The resulting scattering profiles were merged and the buffer background subtracted manually. The resulting concentration series of one-dimensional scattering curves for each sample were manually merged and analysed (Extended Data Table 2). All data reduction and analysis was done with the PRIMUS tool,³⁴ except that shown in Extended Data Figure 5 which was done manually with Python and NumPy.

The quality of the SAXS data were estimated from the statistics presented in Extended Data Table 2a. These values were calculated using a 1 mg/ml sample concentration in which concentration-dependent scattering was negligible as shown in Extended Data Figure 5b. The data show that the PAS-GAF-PHY protein occurs as a dimer of approximately 100 kDa with similar volumes in both Pr and Pfr states. The data also indicate that the Pr-to-Pfr transition gives a slight increase in radius of gyration (R_g) and maximal dimension (D_{\max}), consistent with the opening of the dimer found in this paper.

2.2.4 Structural Fits

We generated sets of likely Pr and Pfr solution structures using a three-step analysis. First, MD simulations (details below) were run in order to sample conformational space in the vicinity of the models obtained from crystallization studies (section 1.3 below). The solution X-ray scattering profile was calculated for each snapshot. Secondly, these scattering profiles were pairwise subtracted (Pfr - Pr) and compared to difference scattering data from the cSAXS experiment. Solution structures were then proposed based on the best fits. Thirdly, these structures were verified by comparison to absolute SAXS data.

MD snapshots: Molecular simulations were carried out under various conditions as described below. In all, 3002 and 4003 conformations representing the Pr and Pfr states, respectively, were considered. The theoretical X-ray scattering curve of each snapshot was calculated on the interval $0 \leq q \leq 5 \text{ nm}^{-1}$ using SASTBX and Zernike expansion (options: `znk_nmax=40`, `n_step=100`).³⁰ As the position of hydrogen atoms was specified in the MD simulations these were also explicitly accounted for in the calculations.

Fitting to difference X-ray scattering: The strategy was to find the difference curves of individual pairs of snapshots which reproduced experimental data (Fig. 1), and to accept the conformations participating in the N best pair's curves as solution structures.

Each theoretical difference curve was scored against the experiment by calculating the sum of squares of the error (SSE) after scaling,

$$SSE = \min_k \sum_q \left(\Delta S(q) - k \cdot \Delta \hat{S}(q) \right)^2,$$

where $\Delta S(q)$ is the experimental difference scattering and $\Delta \hat{S}(q)$ is the calculated difference scattering. The scaling coefficient k is necessary in solution X-ray scattering experiments and accounts for, for example, variations in X-ray path length in quartz capillaries, concentration variations due to progressive sample degradation, and that the conversion efficiency of the excitation ratio is a complex and unknown function of the experimental geometry and the photocycle of the sample.

Of the M theoretical difference scattering curves, where M was on the order of 10^7 , the N best-matching ones were found based on the SSE score. Two sets of conformations, A and B , participate in these curves, where the number of members N_A and N_B are less than or equal to N since each conformation can participate in more than one of the N pairs. Considering all pairwise differences between the structures in A and B produces $N_{AB} = N_A N_B$ curves, where typically $N_{AB} \gg N$ (see Extended Data Figure 7f).

Once the sets of likely solution structures A and B were determined, two statistics were used to evaluate their quality. First, the sets were treated as ensembles and their average absolute scattering profiles were calculated before subtraction and scaling. This produces the average error,

$$SSE_{av} = \min_k \sum_q (\Delta S(q) - k \cdot [\langle \hat{S}^{Pfr}(q) \rangle_B - \langle \hat{S}^{Pr}(q) \rangle_A])^2 .$$

$\hat{S}^{Pr}(q)$ and $\hat{S}^{Pfr}(q)$ are the theoretical Pr and Pfr scattering profiles, respectively, $\langle \rangle_{A,B}$ denotes the ensemble average over groups A or B . Secondly, for two internally consistent sets A and B , all N_{AB} pairwise difference curves should agree reasonably with experiment even though most of them were not among the N original best-fitting curves. The total error describes the average deviation of these differences from experiment,

$$SSE_{tot} = \frac{1}{N_{AB}} \sum_{a \in A} \sum_{b \in B} \min_k \sum_q (\Delta S(q) - k \cdot [\hat{S}^{Pfr}(q) - \hat{S}^{Pr}(q)])^2 .$$

Structural analysis of absolute SAXS data usually relies on the error-weighted scoring function χ^2 . Its value is modified by experimental noise. If the noise is q -dependent it may also influence the best fit chosen.⁴¹ Our difference SAXS measurement does not contain any apparent noise (red, Fig 1b), and as such q -dependent noise is also absent. Therefore, we use the simpler R^2 to assess how well solution-structural models or crystal structures fit the experimental difference scattering data. Higher R^2 means better agreement. We define it as

$$R^2 = 1 - \frac{SSE_{av}}{SST} ,$$

where SST is the total sum of squares,

$$SST = \sum_q [y(q) - \bar{y}]^2 ,$$

$y(q)$ is some experimental curve (*e. g.* absolute or difference scattering) and \bar{y} is its average $\bar{y} = \langle y(q) \rangle_q$.

Guided by the crystal structures, we adopted the centre of mass distance between the C-terminal helices (residues 484-503) on opposing monomers as a structural descriptor. This simplification, which we call the PHY domain separation, R_{PP} , allows us to compare structures on a single axis.

The structures participating in the $N = 100$ best pairwise fits (sets of 83 Pr and 9 Pfr conformations) were accepted as solution-structural models and the Pr and Pfr scattering from these were averaged separately to obtain the theoretical curve of Figure 3a. Representative Pr and Pfr structural models were chosen using the GROMACS tool `g_cluster`. Each set was treated as a single cluster and the central member was identified using the single linkage method.

Extended Data Figure 7 lists all parameters for the fitting procedure and illustrates the details of the structural refinement procedure.

Verification by comparison to absolute X-ray scattering data: The absolute theoretical scattering curves were compared to SAXS data recorded at BM29 of the ESRF. The experimental curves were considered to represent Pr/Pfr mixtures with fractions ν the Pr population fraction, of each state between 0 and 1. We used the sum of squares of the error

$$SSE_{abs} = \min_k \sum_q \left(S(q) - k \cdot (\nu \cdot \hat{S}^{Pr}(q) + (1 - \nu) \cdot \hat{S}^{Pfr}(q)) \right)^2 ,$$

where $S(q)$ is the experimental scattering, and k the scaling variable, which was optimized individually to yield the best fit to experimental data for each comparison. For each Pr conformation, the average $\hat{S}^{Pfr}(q)$ for all the Pfr solution structures was used when evaluating SSE_{abs} , and *vice versa*. Extended Data Figure 5d shows how the relative populations were determined by minimizing SSE_{abs} on the interval with respect to ν .

For the fit to difference scattering the q range considered was $0.7 \text{ nm}^{-1} \leq q \leq 2.5 \text{ nm}^{-1}$ and for the absolute scattering $0.25 \text{ nm}^{-1} \leq q \leq 2.5 \text{ nm}^{-1}$. These ranges were chosen due to artefacts caused by the beam stop below the lower limits and because of low signal to noise ratio and increased uncertainty of the Zernike expansion for calculation of the X-ray scattering above the upper limits.

Solution structure refinement summary: The *MD trajectories* sampled a reasonable and sufficient space along the PHY-PHY separation (Extended Data Figure 7). By *fitting to difference X-ray scattering data*, 100 pairs of solution structural candidates were identified with an R^2 of 0.987. The pairs consisted of 9 Pfr and 83 Pr structures. This is a significant improvement compared to the pair of crystal structures, which has an R^2 of 0.103 (see Figure 3a and Extended Data Figure 7f). The representative solution structures are presented in Fig. 4.

Comparison of the averaged theoretical scattering of the Pr and Pfr candidate structures to *absolute X-ray scattering data* yielded $\nu = 1.0$ for the Pr data and $\nu = 0.36$ for the Pfr data (Extended Data Figure 5d), meaning that the protein is to 100 % in Pr in the SAXS measurements, while photoconversion into Pfr yielded 64 % Pfr. These figures are reasonable considering that the absorption spectra overlap and that illumination at 655 nm also excites molecules in the Pfr state (see Fig. 1a). The best fits to absolute data gave SSE_{abs} of 4.6 and 6.1 for Pr and Pfr, respectively, which is significantly lower than the corresponding SSE_{abs} for the crystal structures (13.7 and 12.8 for Pr and Pfr, respectively). The R^2 for all these cases were above 0.99.

2.3 Crystallography

2.3.1 Crystallization and Data Collection

The *dark form* of the PAS-GAF-PHY construct from *D. radiodurans* phytochrome was crystallized at 298 K by hanging-drop vapour diffusion with 11 mg/ml protein concentration. 2 μ l droplets containing equal volumes of protein solution (20 mM Tris·HCl pH 7.0, 150 mM NaCl) and reservoir solution (30 % PEG1500) were equilibrated against 1 ml of reservoir solution. The protein formed rod-like crystals of 200 μ m approximate length (see Extended Data Figure 3a), and grew in complete darkness at room temperature within a week. The protein did not crystallize in this condition after illumination (*i.e.*, when mostly in the Pfr state). The *illuminated form* of the PAS-GAF-PHY construct from *D. radiodurans* phytochrome was crystallized with the same hanging-drop vapour diffusion method as above, but in a reservoir solution of (100 mM Tris·HCl pH 8.5, 200 mM NaCl, 25 % PEG3350). The protein was converted to the Pfr state by illumination with 655 nm for 15 min using a LED (5 mW) before crystallization, and thereafter twice a day every morning and afternoon. The slow dark reversion rate of the PAS-GAF-PHY protein ensured that the concentration of Pfr did not drop significantly during the crystal formation. First rod-like crystals of approximate dimensions of 200x40x40 μ m (see Extended Data Figure 3a) appeared readily within a few hours and were mounted after two days. The protein did not form crystals in this condition while kept in dark. Both crystal forms (dark and illuminated) were mounted under minimal unfiltered microscope light. Mounting was

conducted as quickly as possible by transferring the crystal to the cryoprotectant solution (reservoir solution with 15 % glycerol), and flash-frozen with liquid nitrogen.

Diffraction data were collected at 100 K at beamline ID23-1 of the European Synchrotron Radiation Facility (ESRF), Grenoble with an X-ray wavelength of 0.980 Å.⁴² Data was processed using the XDS program package.⁴³ The dark crystals belong to space group $P6_4$ in a twinned hexagonal crystal system with two monomers in asymmetric unit. The illuminated crystals belong to space group $P2_12_12_1$ with four monomers in an asymmetric unit. In order to maximize useful crystallographic information, the data was cut at a cross-correlation value of $CC_{1/2} = 20.9\%$ (dark) and $CC_{1/2} = 35.8\%$ (illuminated).⁴⁴

2.3.2 Structure Determination and Refinement

The crystal structure of the PAS-GAF-PHY domains were solved by molecular replacement with Phaser⁴⁵ using the structure of the PAS-GAF fragment of *D. radiodurans* phytochrome (PDB code 2O9C) as a search model.³² Model building was then carried out manually using the program Coot.⁴⁶ Structural information from homologous PAS-GAF-PHY structures (PDB codes 2VEA and 3NHQ) were utilized in model building,^{13,20} and due to low resolution of the data the building process was kept conservative. The structure was further refined using REFMAC5.⁴⁷ Tight main-chain and side-chain non-crystallographic symmetry restraints were applied in the refinement, and the geometry was restrained tightly by applying an X-ray matrix-weighting term of 0.005. Amplitude-based twin refinement was applied for the dark crystal form data. The final R_{work} and R_{free} values of dark model were 25.29 % and 26.57 %, respectively; for the illuminated form, the final R_{work} and R_{free} values were 23.27 % and 25.97 %, respectively. The Ramachandran statistics for the dark (and illuminated) crystal structures were 92.6 % (95.2 %) in preferred, 7.2 % (4.4 %) in allowed, and 0.2 % (0.4 %) in outlier Ramachandran regions. The statistics of data collection, structure determination and refinement are summarized in Extended Data Table 1. All figures of crystal structures were generated using *PyMOL* software (DeLano Scientific, San Carlos, California, USA).

2.4 Molecular Dynamics Simulations

2.4.1 Software, Force Fields and Parameters

Simulations were performed using GROMACS 4.5.5 (Ref.²⁹) and the Charmm27 force field.^{48,49} Force-field parameters for the biliverdin chromophore in the Pr state⁵⁰ were manually adapted to GROMACS format. As a first approximation for the Pfr state of the chromophore, the D-ring was rotated manually in the force-field, leaving all other parameters including partial charges unchanged. The D-ring rotation represents the first reaction coordinate in a recent description of the details of the chromophore in the Pr and Pfr states.⁵¹ While a rigorous treatment of the Pfr state should certainly start from quantum-chemical calculations, the simple approach taken here does account for the change in the ability of the D ring to act as a hydrogen-bond donor against residues such as Asp207.⁵¹

Periodic boundary conditions were used in three dimensions. All bonds were constrained using the LINCS algorithm and a time step of 2 fs was used throughout. Simulations employed Particle Mesh Ewald (PME) electrostatics with fourth order interpolation and 0.16 nm grid spacing. Cutoffs for short-range electrostatic and van der Waals interactions were 1.0 nm and the neighbour list was updated every 5 steps with a 1.0 nm cutoff. Production runs were carried out under an isothermal-isobaric (NPT) ensemble using the Velocity-rescale thermostat ($\tau_T = 0.5$ ps, $T = 300$ K) and the Parrinello-Rahman barostat ($\tau_P = 2$ ps, $P = 1$ bar).

2.4.2 Equilibration

The crystal structures for the dark and illuminated forms of the PAS-GAF-PHY domains (Fig. 2) were used as starting points for simulation. All histidines were protonated on C ϵ only, giving a total charge of -48 e per dimer. Using cubic simulation boxes initially 2 nm larger than the protein in each direction, the starting structures were first energy-minimized to a maximum force of 2000 kJ/mol/nm. The systems were energy-minimized again after addition of TIP3P water and ions corresponding to approximately 150 mM NaCl (in addition to 48 cations to neutralize the protein). The simulation boxes were equilibrated under NVT conditions for 50 ps using the velocity-rescale thermostat ($\tau_T = 0.1$ ps, $T = 300$ K). They were subsequently equilibrated under NPT conditions for 500 ps using the same thermostat ($\tau_T = 0.5$ ps, $T = 300$ K) and the Parrinello-Rahman barostat ($\tau_P = 2$ ps, $P = 1$ bar). All non-hydrogen atoms were position-restrained to their initial positions during equilibration (force constants 1000 kJ/mol/nm²).

2.4.3 Production Runs

Four simulations were included in the final analysis.

1. The equilibrated Pr crystal structure was simulated with no restraints for 100 ns. The PHY domain separation (the centre-of-mass separation of residues 494-503 of the opposing monomers) flickered between 2 nm and 3 nm, which corresponds to a closed and half-opened state, respectively.
2. To test for artefacts due to details in the initial structure, a salt bridge between Asp496 of one of the dimer's chains and Arg495 of the other was manually imposed during a 10 ns run. The simulations were then continued for another 100 ns with no restraints, resulting in a PHY domain separation fluctuating about 2 nm (Extended Data Figure 7a), representing a closed state.
3. The equilibrated illuminated crystal structure was simulated under the modified force-field with no restraints for 100 ns. The trajectory remains in the open state PHY domain separation of around 3.5 nm (Fig. 3b and Extended Data Figure 7) although a weak drift towards smaller separation is observed.
4. To increase sampling at and above a PHY domain separation of 4 nm, this separation was artificially scanned using the GROMACS pull code. An umbrella potential with a force constant of 1000 kJ/mol/nm² was used.
 - a. After initially holding the separation at 3 nm for 10 ns, it was scanned ± 5 nm at a rate of 0.1 nm/ns, in total adding 100 ns of simulation to the analysis.
 - b. Sampling between 5 nm and 6 nm was achieved by starting at a snapshot of run 4a with a PHY domain separation of 5.0 nm and scanning once to 6.0 nm and back at a rate of 0.1 nm/ns, adding an additional 20 ns of simulation.

These four production runs are represented in Figure 3b in blue, dark blue, red and dark red, respectively.

The simulations are expected to sample conformational space in the vicinity of the crystal structures. The trajectories are used for direct comparison with experimental solution X-ray scattering data as described above. Simulation times of 100 ns gave plenty of time for the models to adopt solution structures free of crystal packing artefacts. We note however that it is not certain that the slight and approximate difference in force-field of the biliverdin for Pr and Pfr used here would result in very

different structural ensembles in the long-time limit. Figure 4 was rendered using *PyMOL* software (DeLano Scientific, San Carlos, California, USA).

2.5 References

- 37 Lehtivuori, H. *et al.* Fluorescence properties of the chromophore-binding domain of bacteriophytochrome from *Deinococcus radiodurans*. *J Phys Chem B* **117**, 11049-11057, doi:10.1021/jp312061b (2013).
- 38 Tkachenko, N. V. *Optical Spectroscopy - Methods and Instrumentations*. (Elsevier, 2006).
- 39 Cho, H. S. *et al.* Protein structural dynamics in solution unveiled via 100-ps time-resolved x-ray scattering. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7281-7286, doi:DOI 10.1073/pnas.1002951107 (2010).
- 40 Pernot, P. *et al.* Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J Synchrotron Radiat* **20**, 660-664, doi:Doi 10.1107/S0909049513010431 (2013).
- 41 Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477-+, doi:Doi 10.1038/Nature12070 (2013).
- 42 Nurizzo, D. *et al.* The ID23-1 structural biology beamline at the ESRF. *Journal of Synchrotron Radiation* %@ 0909-0495 **13**, 227-238 (2006).
- 43 Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* %@ 0021-8898 **26**, 795-800 (1993).
- 44 Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* **336**, 1030-1033, doi:DOI 10.1126/science.1218231 (2012).
- 45 Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallographica Section D* %@ 0907-4449 **60**, 432-438 (2004).
- 46 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* %@ 0907-4449 **60**, 2126-2132 (2004).
- 47 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D* %@ 0907-4449 **67**, 355-367 (2011).
- 48 Bjelkmar, P. r., Larsson, P., Cuendet, M. A., Hess, B. & Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *Journal of Chemical Theory and Computation* **6**, 459-466 (2010).
- 49 Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry* **25**, 1400-1415 (2004).
- 50 Kaminski, S., Daminelli, G. & Mroginiski, M. A. Molecular Dynamics Simulations of the Chromophore Binding Site of *Deinococcus radiodurans* Bacteriophytochrome Using New Force Field Parameters for the Phytychromobilin Chromophore. *The Journal of Physical Chemistry B* **113**, 945-958 (2009).
- 51 Song, C. *et al.* Solid-State NMR Spectroscopy to Probe Photoactivation in Canonical Phytychromes. *Photochemistry and Photobiology* **89**, 259-273, doi:Doi 10.1111/Php.12029 (2013).