Serial Femtosecond Crystallography Reveals that Photoactivation in a Fluorescent Protein Proceeds via the Hula Twist Mechanism

Alisia Fadini, Christopher D.M. Hutchison, Dmitry Morozov, Jeffrey Chang, Karim Maghlouhi, Samuel Perrett, Fangjia Luo, Jeslyn C.X. Kho, Matthew G. Romei, R. Marc L. Morgan, Christian M. Orr, Violeta Cordon-Preciado, Takaaki Fujiwara, Nipawan Nuemket, Takehiko Tosha, Rie Tanaka, Shigeki Owada, Kensuke Tono, So Iwata, Steven G. Boxer, Gerrit Groenhof, Eriko Nango,* and Jasper J. van Thor*

Cite This: J. Am. Chem. Soc. 2023, 145, 15796−15808

ACCESS

ABSTRACT: Chromophore cis/trans photoisomerization is a fundamental process in chemistry and in the activation of many photosensitive proteins. A major task is understanding the effect of the protein environment on the efficiency and direction of this reaction compared to what is observed in the gas and solution phases. In this study, we set out to visualize the hula twist (HT) mechanism in a fluorescent protein, which is hypothesized to be the preferred mechanism in a spatially constrained binding pocket. We use a chlorine substituent to break the twofold symmetry of the embedded phenolic group of the chromophore and unambiguously identify the HT primary photoproduct. Through serial femtosecond crystallography, we then track the photoreaction from femtoseconds to the microsecond regime. We observe signals for the photoisomerization of the chromophore as early as 300 fs, obtaining the first experimental structural evidence of the HT mechanism in a protein on its femtosecond-to-picosecond timescale. We are then able to follow how chromophore isomerization and twisting lead to secondary structure rearrangements of the protein β-barrel across the time window of our measurements.

INTRODUCTION

The light-induced cis/trans isomerization of a chromophore double bond is a key reaction in photochemistry. In the dynamical response that follows photon absorption, photoisomerization has been shown to be the primary event for a variety of photoreceptors, such as visual pigments, and for the toolbox of photoactivatable proteins used for super-resolution microscopy,1−4 A detailed understanding of the reaction pathway and of how it is steered by the protein environment is key for the rational design of more effective photosystems to employ in nanoscopy,5−8 optogenetics,9−13 and fluorescence biosensing.14−15

Despite the ubiquity of this reaction, the precise mechanism of photoisomerization in conjugated systems is hard to determine. Two conceivable pathways that a protein chromophore can follow for cis/trans photoisomerization are the one bond flip (OBF) and the hula twist (HT).16 In the conventional OBF, the only bond to rotate is the one undergoing isomerization (τ in Figure 1—top);17 half of the molecule needs to turn over, indicating that the OBF mechanism is expected to require a considerable amount of space to be available. In the 1980s, Liu and Asato reasoned that, in proteins, this volume-demanding transition seemed to be in contrast with the observed picosecond formation of an isomerized photoproduct: significant chromophore pocket residue rearrangements that might accompany a large volume sweep by the chromophore are unlikely to occur within such a timescale. As a volume-conserving alternative to the OBF for isomerization in bathorhodopsin, they proposed the HT mechanism, where both τ and the neighboring bond ϕ rotate simultaneously18 (Figure 1—bottom). Since it has become generally accepted that photoisomerization in rhodopsins occurs through a bicycle-pedal mechanism, in which the cis conformation is propagated along the chromophore by a concerted rotation about parallel pairs of double bonds and not through a HT,19−21 A recent time-resolved crystallographic study has in fact obtained structural evidence for an aborted

Received: March 3, 2023
Published: July 7, 2023
bicycle-pedal mechanism in a bovine rhodopsin starting at 1 ps. Nonetheless, crystallographic data supporting the presence of two volume-conserving isomerization pathways, including the HT, were obtained for nanosecond intermediates of the photoactive yellow protein (PYP) photocycle. Conclusive experimental evidence or consensus for whether GFP-like chromophores that are functionally embedded in proteins undergo OBF or HT as the primary photoactivation pathway has not yet been obtained. Two reasons for this are, first, that the primary photoproducts are formed on the ultrafast timescale and are very short-lived, requiring picosecond, or preferably femtosecond, time resolution to be distinguished. Second, the ground-state species formed immediately after photoisomerization by OBF and HT are indistinguishable if the part of the molecule that flips is symmetric.

Figure 1. Schematic of potential chromophore photoisomerization pathways. Two possible pathways for the cis/trans photoisomerization of a chromophore are the OBF pathway or the HT pathway, both shown here schematically. The OBF mechanism involves the rotation of only the isomerizing bond, $\tau$, and is expected to sweep a large volume, as half of the molecule is flipped in the process. In the HT pathway, on the other hand, both $\tau$ and its neighboring bond $\phi$ rotate simultaneously. The products formed by these two pathways are indistinguishable if the part of the molecule that flips is symmetric.

Figure 2. Cl-rsEGFP2 photoswitching. (a) Reversibly photoswitchable protein rsEGFP2 can be converted between a dark OFF state to a fluorescent ON state by specific frequencies of light. These changes are caused by a trans-to-cis isomerization and subsequent deprotonation of its embedded chromophore. (b) Addition of a chlorine substituent to the phenolate ring of the chromophore can break its C2 point group symmetry and allow us to distinguish between the HT and OBF mechanisms of chromophore photoisomerization in the first photoproduct of the OFF-to-ON reaction. The normalized absorption spectrum and structure of Cl-rsEGFP2 (c) possess very similar properties as the non-chlorinated protein: the OFF state absorbing predominantly at 400 nm and the ON state at around 480 nm, while the protein tertiary structure exhibits the $\beta$-barrel fold typical of GFP-like constructs. A 2 nm shift can be observed in the absorption profile between chlorinated and non-chlorinated constructs caused by the electron-withdrawing nature of chlorine.
are also resolved and matched to the temperature serial femtosecond crystallography (PDB 8A6G). The predominant chromophore (OHD) conformation is the "planar" trans anti (trans-PL). Minor populations of a "twisted" trans syn (trans-TW) and of cis anti configurations are also modeled (refined occupancies of 12 and 14%, respectively). The three chromophore configurations are accompanied by three alternate His149 conformations. Two Val151 conformations are also resolved and matched to the trans-PL and trans-TW species. (b) OFF-state equilibrium trans structures of Cl-rsEGFP2 and (c) rsEGFP2. The major configuration found in the OFF-state chlorinated structure is trans-PL, while it is trans-TW when the chlorine substituent is not present. We presume that the heavy atom substitution results in a higher energy for the trans anti state through increased steric hindrance. See also Figures S4–S6, Table S3, and Note S1.

Figure 3. OFF-state chromophore configurations in Cl-rsEGFP2. (a) Refined Cl-rsEGFP2 dark OFF-state structure obtained from room-temperature serial femtosecond crystallography (PDB 8A6G). The predominant chromophore (OHD) conformation is the "planar" trans anti (trans-PL). Minor populations of a "twisted" trans syn (trans-TW) and of cis anti configurations are also modeled (refined occupancies of 12 and 14%, respectively). The three chromophore configurations are accompanied by three alternate His149 conformations. Two Val151 conformations are also resolved and matched to the trans-PL and trans-TW species. (b) OFF-state equilibrium trans structures of Cl-rsEGFP2 and (c) rsEGFP2. The major configuration found in the OFF-state chlorinated structure is trans-PL, while it is trans-TW when the chlorine substituent is not present. We presume that the heavy atom substitution results in a higher energy for the trans anti state through increased steric hindrance. See also Figures S4–S6, Table S3, and Note S1.

construct of the reversibly switchable fluorescent protein rsEGFP2.33

Reversibly switchable fluorescent proteins (RSFPs), such as rsEGFP2, can typically be converted between a fluorescent ON state and a non-fluorescent OFF state using specific photon wavelengths.42,43 They have been widely employed in super-resolution microscopy and imaging,46–52 rsEGFP2 in particular is a β-barrel protein that can be switched from its equilibrium ON state to the non-fluorescent state by 488 nm light, whereas the OFF-to-ON transition is achieved by illumination at 405 nm. In the fluorescent ON state, the chromophore is found as the anionic cis isomer, while the neutral trans isomer is responsible for the non-fluorescent OFF state (Figure 2a).33,43 Time-resolved structural and spectroscopic data previously obtained for the OFF-to-ON reaction of rsEGFP244,45 identified the formation of a twisted chromophore intermediate within a couple of picoseconds and the presence of a ground-state protonated cis conformer 10 ns after photoexcitation, attributing deprotonation to a later, ground-state process. An experimental structural insight on the fundamental isomerization pathway and on the involvement of the speculated chromophore HT, however, remains lacking. As with previous studies, we focus here on the OFF-to-ON reaction, which has a higher quantum yield than the ON-to-OFF reaction.33,46 The 4-hydroxybenzylidene-1,2-dimethylimidazolone chromophore’s phenol ring in rsEGFP2 has a C2 point group symmetry, leading to equivalent cis products regardless of the photoisomerization pathway. We exploit the introduction of a chlorine atom substituent47 to break this symmetry and distinguish between the products formed via the OBF and HT pathways, as previously suggested.48 On the basis of IUPAC recommendations,49–51 starting from a trans chromophore with a substituent anti to the double-bonded imidazolone nitrogen, HT leads to a cis anti product, while the configuration formed via OBF is cis syn (Figure 2b). The motivation behind incorporation of chlorine in rsEGFP2 is therefore to obtain a construct with photoswitching capabilities that resemble those of rsEGFP248 but that can be investigated to confirm either HT or OBF. We note below that there are structural and spectroscopic differences between the chlorinated and unmodified constructs and the observation of the HT pathway in the chlorinated construct we consider here does necessarily imply such process in other constructs.

In a prior study,48 the same rsEGFP2 construct containing a monochlorinated chromophore (Cl-rsEGFP2) was used to study the ON-to-OFF reaction and compare protein structures before and after irradiation. Structures were obtained by rapidly cryocooling crystals after a specific irradiation period and subsequently collecting diffraction data using synchrotron X-ray radiation. This method relied on the assumption that the cryocooling irradiated structure reflected the chromophore configuration immediately after photoisomerization but was not able to account for thermal bond rotations, which can interconvert anti and syn species, or any change in chromophore configuration induced by the freezing process itself. Here, by performing a time-resolved serial femtosecond crystallography (TR-SFX) experiment with sub-picosecond resolution, we have been able to unequivocally identify the primary photoproduct of the OFF-to-ON photoisomerization reaction: the trans anti chromophore in the OFF state of Cl-rsEGFP2 is photoexcited with a 400 nm photon, and clear signals showing the formation of a cis anti photoproduce are reported starting at delays of 300 fs. This, to our knowledge, is the first direct experimental structural evidence that unequivocally supports HT chromophore photoisomerization in a protein on its ultrafast timescale, more than 30 years after it was first hypothesized.
RESULTS AND DISCUSSION

OFF-State SFX Structure for Cl-rsEGFP2. Cl-rsEGFP2 maintains very similar switching properties to rsEGFP2 (Figures 2c and S1) and is expressed using amber suppression as described by Chang et al.\textsuperscript{48} Cl-rsEGFP2 microcrystals diffracted to 1.63 Å at the SACLA X-ray Free Electron Laser (XFEL) in Japan (Figure S2). Crystals were pre-illuminated with a 488 nm CW laser in order to photoaccumulate the OFF state. The OFF-to-ON transition was then initiated with a 75 fs 400 nm pump laser pulse and probed with the XFEL pulse at pump–probe delays between 300 fs and 1 μs. To obtain a ground-state room-temperature structure of the Cl-rsEGFP2 OFF state, dark data was collected in an interleaved manner (1:5 ratio) with the pump–probe data throughout our TR-SFX experiment. To account for pump laser scatter during sample delivery and to model a more accurate dark structure, a dataset was also collected with a negative pump delay (−5 ps).

A dark structure (PDB 8A6G) was refined to the interleaved dataset was also collected with a negative pump delay (−5 ps). Positive (blue) and negative (red) DED are shown at ±3.5 σ. The refined trans anti and cis anti species are shown in light and dark gray, respectively. In the 600 fs, 900 fs, and 5 ps maps, the feature “Peak 1” is highlighted: this is the main indicator of the presence of a femtosecond intermediate, which we have called trans-FS and discussed with further details in the main text. PDBs (in order) 8A6N, 8A6O, 8A6P, 8A6Q, 8A6R, and 8A6S. See also Figures S7–S13.

Trans syn (named here trans-TW) and of a cis anti species were also manually refined to occupancies of 12 and 14%, respectively, as well as the accompanying alternate conformations of residues H149 and V151 (Figures 3 and S4). The trans-PL and trans-TW species we observe here have been previously reported in room-temperature crystal structures of rsEGFP2\textsuperscript{24,52} and in cryostructures of Cl-rsEGFP2\textsuperscript{24} (Table S3). Absorption spectra and quantum chemical calculations also point to the presence of these two forms in rsEGFP2 solutions.\textsuperscript{52} We expect that the ON-to-OFF reaction starting from the cis chromophore can therefore eventually result in either of these two trans conformations.

Adam et al.\textsuperscript{52} postulate a fast exchange between trans-PL and trans-TW in solution (on the order of ≈0.1 s) and propose that the two states are separated by a low energy barrier in the protein conformational energy landscape. In room-temperature rsEGFP2 SFX structures, the prevalent trans configuration is the twisted trans-TW (Figure S5 and ref \textsuperscript{52}), suggesting that it is a lower energy state than trans-PL. In Cl-rsEGFP2, we observe the prevalent configuration to be trans-PL and presume that the transition from the planar to the twisted form is no longer energetically favorable due to the steric hindrance introduced by the bulk of the chlorine atom or to an electronic substituent effect (see also Note S1). This is also in line with the finding that an enlargement of a chromophore pocket residue (V151L) can shift the OFF state equilibrium almost completely to trans-PL in non-chlorinated rsEGFP2.\textsuperscript{52} The states observed in our dark OFF-state structure are shown in Figure 3b. For the scope of this time-resolved study, we expect that any observed photoinduced species result from the trans-PL-to-cis reaction; since the starting population of trans-TW in our dark-state structure is below 15%, we expect that any photoprodut resulting from photoexcitation of this state will not be detectable in our data.

TR-SFX Data for Cl-rsEGFP2. Chromophore-Specific Changes. Pump–probe TR-SFX data was collected for six delays: 300 fs, 600 fs, 900 fs, 5ps, 100 ps, and 1 μs. The pump laser power density used was 0.5 mJ/mm\textsuperscript{2}. We generated Q-weighted difference electron density (DED) maps for these delays and improved signal-to-noise through principal component analysis using a python pipeline (Figures S7 and S8 and Supporting Information Procedures Section 1). The resulting maps are shown in Figure 4 at ±3.5σ. The initial negative density signals on the trans-PL imidazolinone ring oxygen are already visible at 300 fs and are accompanied by respective positive density on the same oxygen and on the cis methylene bridge. At 600 fs, there is a large negative density on the trans-PL Cl atom, and negative signals also on the two chromophore oxygens and methylene bridge. Positive density signals appear on both cis anti rings. In the 600 and 900 fs data, there is also a strong (+3.5σ) positive feature that we have labeled Peak 1 and suggests an upward shift of chlorine position (approximately 1 Å—Figure 4b,c). This peak is still weakly present in the 5 ps data (Figure 4d) and effectively

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4}
\caption{Time-resolved DED maps of Cl-rsEGFP2. Q-weighted DED maps for the collected pump–probe delays. Positive (blue) and negative (red) DED is shown at ±3.5 σ. The refined trans anti and cis anti species are shown in light and dark gray, respectively. In the 600 fs, 900 fs, and 5 ps maps, the feature “Peak 1” is highlighted: this is the main indicator of the presence of a femtosecond intermediate, which we have called trans-FS and discussed with further details in the main text. PDBs (in order) 8A6N, 8A6O, 8A6P, 8A6Q, 8A6R, and 8A6S. See also Figures S7–S13.}
\end{figure}
absent in the 100 ps and 1 μs maps. We assign Peak 1 to an intermediate state, which we call trans-FS (Figure 4b), discussed in more detail below. At 100 ps and 1 μs, the trans-PL to cis anti features are very clear and are the prevalent signals at ±3.5σ: there are strong negative and positive electron density on the chromophore chlorine substituent and on the imidazolinone ring carbonyl, as well as positive density on the cis anti methylene bridge and the phenol ring (Figure 4e,f). These difference density signals thus point to increased population of the cis anti species and decreased population of trans anti starting from 300 fs, as well as possible formation of a femtosecond intermediate. In the analysis below, we investigate this further to refine the structure and occupancy of the cis anti and trans-FS conformations for each time point.

In TR-SFX, photoinduced differences in structure factor amplitudes, caused by population changes prepared by the pump laser pulse, are usually small. A further analysis is therefore implemented to extract these minor light-induced population changes from the structure factors obtained for each time point, once key regions of interest are identified from the DED maps above (e.g., the chromophore region). This method is based on the background subtraction estimation implemented by Pearce and colleagues and consists of generating maps, called here $WΔF_{max}$ maps, from the measured illuminated and dark structure factors and reference phases. These maps differ from the DED maps above in that a background subtraction factor (0 < $N_{bg}$ < 1) is estimated to generate a weighted map $W(ΔF_{illuminated} − N_{bg} \times ΔF_{dark})$ (Figures S9–S12 and Supporting Information Procedures Section 1). The $WΔF_{max}$ map for the 100 ps time point is shown in Figure 5a and illustrates the efficacy of the background estimation: the cis anti photoproduct is clearly present, with electron density at +1.8σ outlining both rings and a large portion of the isomerized double bond. $WΔF_{max}$ maps for 300 fs, 600 fs, 900 fs, and 1 μs also confirm the formation of the cis anti species (Figure S11). The 600 and 900 fs $WΔF_{max}$ maps support the assignment of Peak 1 to a trans intermediate. The electron densities of these two maps are very similar and present four key characteristics: (i) the presence of peak 1, (ii) an elongated and uncentered peak where the cis anti chlorine is positioned, which is in contrast with the round, centered features visible in the $WΔF_{max}$ maps from the later time points, (iii) electron density that fills the cis anti chromophore phenol ring, again in contrast with the 100 ps and 1 μs maps that show instead a clear outline of the chromophore phenol ring where its center is empty, and (iv) features that suggest a tilt of the imidazolinone ring oxygen toward the phenol ring (Figures S11 and 12). Though assignment is complicated by the likely

![Figure 5](https://doi.org/10.1021/jacs.3c02313)
presence on both trans-FS and cis anti species, refinement of coordinates to 600 and 900 fs $\Delta F_{\text{max}}$ maps leads to two chromophore configurations that are almost superimposable (Figure S13). We therefore suggest that an intermediate state, trans-FS, accumulates with a 600–900 fs timescale. Because the 900 fs data appears to have less contamination from the cis anti species than the 600 fs data, we take the coordinates refined from the 900 fs $\Delta F_{\text{max}}$ map as the proposed trans-FS structure. This chromophore configuration is still trans but more twisted (the torsion angles for the refined structure are $\phi \approx 33^\circ$ and $\tau \approx 178^\circ$); such change is accompanied by a movement of the imidazolinone ring by about 0.5 Å, which is particularly noticeable from the shift in the electron density of its oxygen (Figure 4). Comparable ultrafast formation of a twisted trans chromophore in protein photoisomerization has been observed before in rsEGFP2 and PYP and has been assigned to the excited state. Similarly, we also suggest assignment of trans-FS to an excited-state species on the basis of QM-MM calculations (see below). $\Delta F_{\text{max}}$ maps thus allow us to model photoinduced species more effectively than the DED maps in Figure 4. Through them, we confirm the cis anti configuration as well as the structure of the trans-FS intermediate (Figures S10–S13). After establishing coordinates through $\Delta F_{\text{max}}$ maps, we use the raw data once again to refine occupancies of new chromophore conformations for each time point (Figure S14). We note that clear identification of photoprodut and intermediate species here is greatly helped by the higher X-ray scattering cross-section of chlorine, both in DED and in $\Delta F_{\text{max}}$ maps, highlighting a further advantage of introducing a heavy atom substituent in the chromophore structure. The confident assignment of cis anti formation at ultrafast time delays provides strong evidence for the HT pathway in this rsEGFP2 construct. The observation of photoisomerization at femtosecond delays is likely associated with the relatively high primary quantum yield of this protein. The significant transient concentration and large displacement of the chromophore in the reactive pathway of CI-rsEGFP2 strongly contribute to the light-induced differences, while additional contributions, such as vibrational coherences, will be below this level.

**Protein Residue/Structure Changes.** Strong DED features in our data are almost entirely concentrated on the chromophore (Figure 5b). To investigate the presence of any other motions across the protein structure that are associated with chromophore twisting and photoisomerization, we use the analysis and scripts from Wickstrand et al. and plot DED signals along the protein sequence for every time point collected (Figure 5c). Data from 600 fs, 900 fs, 100 ps, and 1 μs delays display the best crystallographic statistics (Table S1), so we concentrate on these to identify the areas of further protein structural change. Aside from the chromophore (the OHD ligand in the structures) that, as expected, presents the largest light-induced differences, there are two regions of notice: the central $\alpha$-helix (residues 58–67) and $\beta$-strand 7 (residues 146–152), which interfaces with the chromophore phenol ring. Figures S15 and 16 show the DED on specific residues of these two secondary structure elements for different time points. Signals on the central $\alpha$-helix suggest an overall downward motion of the region by around 1.4 Å if the distance between the strong positive/negative features on Leu65 and Thr63 is measured. A downward movement of this helix was previously observed in 1 ps DED maps of non-chlorinated rsEGFP2 and was attributed to a tilt of the imidazolinone ring upon chromophore twisting. Similarly, here the prevalence of these signals in the sub-picosecond time points indicates that this motion is closely associated with the ultrafast photoisomerization of the bound chromophore or formation of the twisted trans-FS intermediate.

Negative DED signals on the side chains and backbone of $\beta$-strand 7 (β7) residues Asn147, Asn150, and Val151 are superimposable between the 100 ps and 1 μs datasets and hint at increased structural flexibility of this secondary structure element at later time points. Conformational fluctuations involving mainly β7, as well as that side of the β-barrel (β7–10), have been reported on the nanosecond–millisecond timescale for multiple GFP-like proteins, including rsEGFP2. This structural relaxation and plasticity that arises...
after photoactivation has been hypothesized to be important for efficient photoswitching between distinct protein conformations based on NMR, spectroscopy, and MD results. We have now obtained time-resolved crystallographic evidence that supports the development of \( \beta_7 \) dynamics starting from 100 ps and peaking in the microsecond regime (Figures 5 and S16) in Cl-rsEGFP2. In our maps, \( \beta_7 \) motions are accompanied by a flip of His149 from its characteristic OFF conformation to its ON conformation (Figure S16). The negative density on the side chain of His149 that indicates the flip to its ON conformation becomes clear at 1 \( \mu s \), agreeing with recent time-resolved multiple-probe infrared spectroscopy (TRMPS-IR) data that assign His149 movement to a 42 ns time constant following ultrafast chromophore photoisomerization.

**Ultrafast Visible Transient Absorption Spectroscopy of Cl-rsEGFP2 Solutions.** We monitored the spectral changes induced by 400 nm laser excitation in solutions of Cl-rsEGFP2 in the OFF state (Figure 6a). The first photoinduced spectral changes we observe include an increase in absorption at 450 nm that peaks at around 500 fs (attributed to excited-state absorption—ESA), accompanied by ground-state bleach (GSB) in the 400 nm region and a broad stimulated emission (SE) signal around 530–540 nm. Global analysis groups these features in a first component that has a decay rate constant of 2 ps. Assuming a sequential model (see Experimental Methods Section), this first component decays into a second and then a third component that both present a new positive peak at 420 nm and respective decay rate constants of 57 and 500 ps. In the final 3.5 ns component, the absorption at 420 nm is absent, and the main positive peak present is once again at 450–460 nm (Figure 6b,c). While the wavelengths for GSB, 450 nm ESA, and SE present in Cl-rsEGFP2 time-resolved spectra closely match those of rsEGFP2 (Figures S17 and 18 and refs 44 and 61), the transient picosecond positive peak at 420 nm is unique to the chlorinated construct (Figure S18). We also note that Cl-rsEGFP2 presents an interesting property of the stimulated emission band, which actually increases on the 2 ps timescale and then begins its decay with a 57 ps time constant. Such behavior is not observed in rsEGFP2 (Figures S17 and 18 and refs 44 and 61). We discuss this and the 420 nm positive feature in the following sections.

---

**Figure 7.** Comparison between experimental and calculated chromophore structures. (a) Molecular structures obtained through QM-MM modeling before and after CI crossing are shown with their energy arrangement relative to the Franck–Condon point and their respective torsion angles. Relative energy values computed at a lower (SA2-CASSCF(12,11)/3-21G//Amber03) and higher (XMCQDPT2/SA6-CASSCF(12,11)/cc-pVDZ//Amber03—in parenthesis) level of theory are stated (see also Section 4 in the Supporting Information Procedures). The photoexcited trans-PL species isomerizes in the protein binding pocket to the planar cis anti via HT, passing through a highly twisted conformation near the CI (trans-TWCI). An excited-state twisted minimum structure (trans-TWMIN) and a fluorescent trans minimum (trans-FL) are also found. (b) The calculated conformations are compared to the planar trans anti and cis anti species, as well as to the twisted trans-FS intermediate, refined from TR-SFX data. See also Figure S19.
Quantum Chemical Analysis of HT Photoisomerization. To provide additional support for a HT photoisomerization pathway, we performed QM/MM optimizations at the SA2-CASSCF(12,11)/3-21G//Amber03 level of theory, in which we searched for conical intersections (CIs) associated with the OBF and HT isomerization mechanisms. At first, we optimized a planar trans $S_i$ fluorescent minimum structure (trans-$FL$, Figure 7), which is further discussed below. Whereas no CI could be located along the OBF pathway, we could identify a minimum energy conical intersection (MECI) for a HT isomerization. Additionally, we performed re-computation of the energies in the optimized structures at the correlated XMCQDPT2/SA6-CASSCF(12,11)/cc-pVDZ//Amber03 level of theory. For a detailed description of these calculations, see the Supporting Information Procedure Section 4. It should be noted that the MECI is not a stationary point on the molecular potential energy surface but rather a geometry through which radiationless deactivation from the electronic excited to ground state occurs with high probability. The geometry at this CI features a $\phi$ torsion angle of 20° and a $\tau$ torsion angle of 110° (Figure 7b, trans-TWCI). Geometry optimizations in the electronic ground state confirm that this CI connects the neutral trans and cis chromophore conformers (Supporting Information Procedures Section 4). The value of $\tau$ in this trans-TWCI species is close to orthogonality and could suggest a stepwise (rather than concerted) nature to the HT pathway, though this additional detail is not resolved in our crystallographic data. At the levels of theory employed here, the trans-TWCI geometry is lower in energy than the Franck-Condon point (Figure 7a). Thus, the results of our computations suggest that the HT dominates the isomerization process, while the OBF mechanism is suppressed.

In addition to the HT MECI geometry (trans-TWCI), we also identified a twisted $S_i$ minimum geometry, with a $\phi$ torsion angle of 12° and a $\tau$ torsion angle of 114° (Figure 7b, trans-TWMIN). This resembles the trans-$FS$ structure refined from the 600–900 fs datasets. Because the energy required for photon absorption from this minimum into the $S_i$ state is 434 nm and hence close to the transient absorption (TA) peak at 420 nm in the time-resolved TA spectroscopy measurements (Figure 6), the twisted $S_i$ minimum could tentatively be assigned to the transient intermediate responsible for this absorption. A more confident assignment, however, would require TA measurements on crystalline samples to determine if the kinetics of the 420 nm absorption confirm accumulation of the trans-$FS$ species in 600–900 fs.

CONCLUSIONS

This study exemplifies the capability of the temporal and structural resolution available with serial crystallography at XFEL facilities. Through the combination of the SFX technique and the breaking of the rsEGFP2 chromophore symmetry via a heavy atom substituent, we resolve the primary photoproduct of the trans-to-cis photoisomerization and confidently assign it to be the outcome of a HT pathway. To our knowledge, this is the first experimental observation of HT photoisomerization in a protein on the femtosecond-to-picosecond timescale.

We additionally identify a twisted intermediate state at 600–900 fs that can potentially be assigned to an excited-state minimum on the basis of calculations. While we observe a possible correspondence between the calculated absorption spectrum of this intermediate and the 420 nm feature in our time-resolved TA spectra, we note that the sub-picosecond DED signal is dominated by the strongest displacement and highest ordering, and this may contribute only to the first decay component of the ESA and SE heterogeneous decay kinetics. The structural data obtained from TR-SFX reports on the thermodynamic distribution, displacements, cross-sections, and concentrations of ground and excited states in the protein at a specific point in time, while TA data is a kinetic description of the often heterogeneous electronic dynamics that depend on the spectroscopic selection rules. Because of this, the signal strengths and kinetics of TR-SFX and TA transients are not necessarily expected to be the same. In this case, we believe that further investigation, such as a pump–dump scheme, would be required for a confident connection between the crystallographic and spectroscopic signals we report. The current experimental evidence does not unequivocally support the involvement of the twisted trans-$FS$ minimum in the reaction mechanism. In fact, given the observed sub-picosecond formation of the cis anti photoproduct, the presence of a twisted structure with a picosecond lifetime, and the predicted planar trans-$S_i$ fluorescent minimum from calculation (trans-$FL$ in Figure 7), we speculate that there are (at least) three major decay channels in CI-rsEGFP2: (i) an ultrafast, energetically down-hill access to the MECI to form the HT photoproduct, (ii) relaxation into the trans-$FL$ minimum, and (iii) twisted trans $S_i$ minimum formation (trans-$FS$/trans-TWMIN). The last two cases can both promote fluorescence by transiently trapping the excited-state population and could provide an explanation for the long-lived fluorescent state observed in solution TA spectroscopy data (Figure 6).

The ability to track atomic displacements immediately following photon absorption is shown here to be essential for resolving the reaction pathway without having to rely on assumptions on the timescales of thermal bond rotations that can occur after photoactivation. The calculations presented and the use of room-temperature SFX suggest, in fact, that the HT is the only accessible photoisomerization coordinate in the Cl-rsEGFP2 chromophore pocket and that previously proposed crystal-packing constraints are most likely artifacts resulting from the cryo-trapping procedures used (Note S1 and ref 48). Our results thus support the reasoning that the protein cage in Cl-rsEGFP2 hinders the OBF mechanism in favor of the HT.

What is the significance of this pathway? The presence of cis anti photoproduct signals as early as 300 fs introduces the possibility of this photoisomerization reaction being a vibrationally coherent process, as observed for rhodopsin. Vibrational dephasing times in GFP proteins have typically been reported to be within 1–2 ps, which does not exclude the 300 fs cis anti species to be the product of a vibrationally coherent photoisomerization. Further work, such as coherently controlling the photoisomerization and tracking its yield, would be required to investigate this. Interestingly, Adam and colleagues report that rsEGFP2 mutants in which the available binding pocket volume is decreased switch to their ON state much more efficiently than mutants where the available binding pocket volume is increased. It will be valuable to investigate whether a HT reaction pathway dictated by a volume-constraining protein cage is responsible for this increased switching capability and whether coherent CI
crossing is important for the efficiency of photoactivation. Understanding the details of the photoactivation process such as these holds significant promise for the rational engineering of fluorescent proteins like CI-rsEGFP2.14,15,73−75 Serial femtosecond crystallography has also allowed us here to trace the effects of the early photoisomerization event to secondary structure rearrangements that occur later in time. This data provides time-stamped structural insight that integrates and reconciles previous observations from other experimental and computational techniques; it furthers our understanding of the dynamics that enable protein function through photoactivation and develops our ability to intelligently manipulate them for improved performance.

■ EXPERIMENTAL METHODS

2021 SFX Sample Preparation, Data Collection, and Setup. Protein Expression, Purification, and Crystallization. Recombinant expression was performed in BL21 (DE3) cells with a PET-15b vector for the rsEGFP2 gene and a pULtra vector containing the 3-chlorotyrosine amber suppression codon. The latter vector was used instead of the pEVOl vector described by Chang et al.19 Large-scale bioreactor fermentation proceeded with a 50 L culture in Terrific Broth Modified (46.7 g/L) containing 100 mg/L ampicillin, 50 mg/L spectinomycin, 10.1 g/L glycerol, and 0.1 g/L 3-chlorotyrosine at 37 °C. After reaching an OD 600 nm of ∼2, the culture was cooled to 20 °C, and induction was carried out by the addition of 0.24 g/L isopropyl β-D-1-thiogalactopyranoside. The culture was grown overnight for 17 h, harvested by pelleting, and stored at −20 °C.

Thawed cell pellets were washed with 50 mM Tris–HCl pH 8.0, 100 mM NaCl, 0.1 mg/mL DNase I, and 1 tablet of EDTA-free protease inhibitors per 100 mL of cells. Lysis was carried out at 4 °C with two passes at 35 kpsi in a T5 Cell Disruptor (Constant Systems, UK). The lysate was centrifuged at 142 000 × g and 4 °C for 30 min before incubating the cell free extract with Ni-NTA resin for 45 min while shaking. After loading the slurry onto a glass column, the bound protein was washed with 2 column volumes of the above buffer containing 20 mM imidazole, followed by 2 column volumes containing 30 mM imidazole. The protein was then eluted with 200 mM imidazole, concentrated, and buffer-exchanged into 50 mM Hepes pH 7.5 and 100 mM NaCl using Vivaspin 10 kDa MWCO filters at 4 °C. The final purification step was carried out with a Superdex S75 gel filtration resin, and the main fractions were concentrated and finally buffer-exchanged into 20 mM Hepes pH 7.5 and 200 mM NaCl. The protein was either stored at −20 °C or used for crystallization.

Seeded batch crystallization proceeded with a final protein concentration of 20−24 mg/mL in a buffer consisting of 75 mM Hepes pH 8.2, 20 mM NaCl, and 1.1−1.3 M ammonium sulfate at 20 °C. Cuboid crystals with typical dimensions of 4−8 × 4−8 × 20−80 μm were obtained after 3 days.

Sample Preparation and Injection. Before injection, the crystals were buffer-exchanged into 25% PEG buffer and size-selected using a 50 μm filter and subsequently a 30 μm filter. Crystals were then delivered to the interaction region using a SACLA’s droplet-on-demand injector (MICROJET)82 and an 80 μm nozzle. Volumes of around 250 μl of crystal slurry were reverse-loaded through the jetting aperture to reduce the risk of blockages. The crystal slurry density was between 2 and 5 × 106 crystal/mL. The driving voltage for the injector piezo was normally 110 V, and current pulses were 100 μs long. Throughout the experiment, a hit rate between 30 and 60% was maintained.

SFX Data Collection and Optical Setup. 2021 data at SACLA77 was collected at Experimental Hutch 2 of BL3. The detector used was the MPCCD-phase III detector,77 and the detector distance was refined to 49.2 mm using unit cell distributions. The Xfel was operated with a repetition rate of 30 Hz, a photon energy of 10.5 keV, and a focal spot of ∼1.5 μm in full width at half-maximum (FWHM).

Crystals were pre-illuminated by direct illumination of the glass tip of a droplet injector with an unfocused (∼2 mm beam size) 100 mW 488 nm CW laser. Offline testing with the same laser and crystal concentration demonstrated maximum conversion after 2 s of illumination well within the ∼8 s transit time of the illuminated part of the jet during normal operation. 400 nm optical pump pulses were created by second harmonic generation of the SACLA synchronized femtosecond laser system using a 100 μm BBO crystal. The pulse length of 800 nm fundamental was measured to be 75 fs by autocorrelation. 7.5 μj pump pulses were focused onto the interaction region using a 300 mm lens to give a spot size of 130 μm (FWHM) and a corresponding energy density of 0.5 μJ/mm². Light and dark data were interleaved in a 1:1 ratio. A spatial overlap between the optical and XFEL beams was confirmed using a 50 μm Ce:YAG crystal. The same crystal was used to perform a cross-correlation and find a temporal overlap; the cross-correlation was fitted using the same methods described in 79. A temporal jitter between the optical and X-ray pulses was monitored using the SACLA timing tool system;17 however, recent upgrades to the synchronization of the optical laser and XFEL15 showed that the measured jitter measured was ∼50fs (FWHM) less than the instrument response of the measurement (∼100 fs), indicating that post-processing sorting of the TR-SFX data was unnecessary. The new system also actively corrects for slow timing drift between optical and XFEL beams, which was confirmed by cross-correlation measurements at the start and end of data collection (see SI for more details). Following data analysis methods are detailed in Section 1 of the Supporting Information Procedures.

Femtosecond TA Spectroscopy. Femtosecond TA data was measured using the system described in Lincoln et al.17 White light probe pulses were generated using filamentation in a CaF glass window. 400 nm pump pulses were generated by doubling the fundamental 800 nm (Hurricane, Spectra Physics) in a 100 μm-thick SHG-BBO (Eksma optics). The pump pulses were focused onto the sample using an AR-coated f = 500 mm lens (Thorlabs). The pump energy density was ∼0.02 μJ/mm². A 20 μl solution of Cl-rsEGFP2 was mounted in a liquid flow cell (Harrick Scientific Products Inc) between a 1 mm (front) and a 2 mm (back) sapphire windows (Crystlan Ltd) using a 25.6 μm spacer. The sample concentration was chosen to give an OD of ∼0.1 at the 400 nm peak of the OFF state (Figure 2c). The sample was pre-illuminated using an unfocused 50 mW 488 nm diode laser to ensure presence of the target OFF state and was continuously translated during data collection. Using the Ultrafast Spectroscopy Modeling Toolbox,83 a second-order polynomial was fitted to the pump coherent artefact and was sufficient to correct the majority of the inherent spectral chirp of the white light probe. A sequential model was then applied in the Toolbox for SVD analysis and global fitting. Application of a parallel model leads to overfitting, which is made clear by the presence of compensating amplitudes in the fitted components (not shown).

■ ASSOCIATED CONTENT

* Supporting Information. The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.3c02313.

Frame data processing and crystallographic analysis; crystallographic map generation procedures; 2019 rsEGFP2 OFF state and cryo-trapping experiment; quantum chemical modeling procedures; anomalous crystallographic maps for confirmation of chlorine atom; and quantum yield calculations for CI-rsEGFP2 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Eriko Nango — RIKEN Spring-8 Center, Sayo, Hyogo 679-5148, Japan; Institute of Multidisciplinary Research for
The XFEL experiments were performed at SACLA BL3 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (proposal nos. 2021A8006 and 22019B8021). JyVT and AF acknowledge funding from the Imperial College President’s PhD Scholarship and the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/P00752X/1). GG and DM acknowledge funding from the academy of Finland (grant no. 332743). This work was supported in part by a National Institutes of Health Grant (no. R35GM118044 to SGB). This work was also supported by the Japan Society for the Promotion of Science KAKENHI grants no. 19H05781 (E.N.) and 19H05776 (S.I.). The Platform Project for Supporting Drug Discovery and Life Science Research from the Japan Agency for Medical Research and Development under grant no. JP21am0101070 (S.I.). Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (contract no. DE-AC02-76SF00515). The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health (NIH), National Institute of General Medical Sciences (NIGMS) (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS or NIH. We acknowledge TJ Lane for suggestion and discussion on the background subtraction estimation. We also thank Irimpan Matthews for assistance in the data collection at SSRL and acknowledge Diamond Light Source for time on I23 under proposal 23620. The crystallization facility at Imperial College was funded by BBSRC (BB/D524840/1) and the Wellcome Trust (202926/Z/16/Z).

ACKNOWLEDGMENTS

The authors declare no competing financial interest.

REFERENCES


Serial Femtosecond Crystallography Reveals that Photoactivation in a Fluorescent Protein Proceeds via the Hula Twist Mechanism - Supporting Information

Authors:
Alisia Fadini\(^1\), Christopher D.M. Hutchison\(^1\), Dmitry Morozov\(^2\), Jeffrey Chang\(^3\), Karim Maghlaoui\(^1\), Samuel Perrett\(^1\), Fangjia Luo\(^4\), Jeslyn C.X. Kho\(^1\), Matthew G. Romei\(^5\), R. Marc L. Morgan\(^1\), Christian M. Orr\(^6\), Violeta Cordon-Preciado\(^1\), Takaaki Fujiwara\(^7\), Nipawan Nuemket\(^4,8\), Takehiko Tosh\(^4\), Rie Tanaka\(^4,8\), Shigeki Owada\(^4,9\), Kensuke Tono\(^4,9\), So Iwata\(^4,8\), Steven G. Boxer\(^5\), Gerrit Groenhof\(^2\), Eriko Nango\(^4,7\), Jasper J. van Thor\(^1\)

Affiliations:
\(^1\) Department of Life Sciences, Faculty of Natural Sciences, Imperial College London, London, SW7 2AZ, UK
\(^2\) Nanoscience Center and Department of Chemistry, University of Jyväskylä, Jyväskylä, 40014, Finland
\(^3\) Department of Physics, Stanford University, Stanford, CA 94305, USA
\(^4\) RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo, Sayo, Hyogo, 679-5148, Japan
\(^5\) Department of Chemistry, Stanford University, Stanford, CA 94305, USA
\(^6\) Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot, OX11 0DE, UK
\(^7\) Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, 2-1-1 Katahira, Aoba, Sendai, Miyagi, 980-8577, Japan
\(^8\) Department of Cell Biology, Graduate School of Medicine, Kyoto University, Yoshidakonoe, Sakyō, Kyoto, 606-8501, Japan
\(^9\) Japan Synchrotron Radiation Research Institute, 1-1-1 Kouto, Sayo, Sayo, Hyogo, 679-5198, Japan
**Supporting Figures and Tables**

![Graph showing absorbance decay over time for different constructs.](image)

**Figure S1.** OFF-to-ON switching rates for chlorinated and unchlorinated rsEGFP2 constructs measured under the same conditions. An exponential decay of the form \( y = Ae^{-kt} \) was fitted to obtain the rate \( k \) for each construct. Related to Figure 2.

![Graph showing resolution cutoff choices for interleaved dark and Neg5ps datasets.](image)

**Figure S2.** Choice of resolution cutoff from the dark SFX data. A CC\(_{1/2}\) of 0.5 was chosen, corresponding to \( \approx 1.63 \) Å for both the interleaved dark and -5ps datasets.
Figure S3. (a) Cross-correlation between the laser pump and XFEL probe for time zero determination and (b) jitter analysis for the SACLA 2021 experiment. Related to Figure 4.

Figure S4. 2mFo-DFc maps, shown at +1.8σ, computed from the -5 ps (a) and interleaved dark (b) datasets and the refined dark structure (PDB 8A6G). mFo-DFc maps are shown in red and green at -3σ and +3σ respectively. The top and bottom panels show two different views of the chromophore (OHD) region. The three chromophore populations modeled in the dark state, planar trans anti (trans-PL), twisted trans syn (trans-TW), and cis anti, are shown in distinct colors with the corresponding interacting residue conformations. Related to Figure 3.
**Figure S5.** (a) 2mFo-DFc map, shown at 1.8σ, computed with the reflections from dark data collected on rsEGFP2 (unchlorinated) at SACLA in 2019 and the refined dark structure (PDB 8A7V). The prevalent chromophore conformation is the trans, though a small population of cis is also refined. (b) Superposition of the refined dark room temperature (RT) structures for the unchlorinated rsEGFP2 (SACLA 2019, green) and Cl-rsEGFP2 (SACLA 2021, gray). It is highlighted how the major chromophore trans configuration differs between the two: rsEGFP2 is found in the twisted trans configuration (trans-TW), while Cl-rsEGFP2 is mainly in the planar trans configuration (trans-PL). Both datasets present minor cis chromophore conformations. Related to Figure 3.
Figure S6. 2mFo-DFc map, shown at 1.8σ, computed with the reflections from data collected on Cl-rsEGFP2 at SSRL after illumination and dehydration and the refined structure (PDB 8A83). The prevalent chromophore conformation is trans-PL. Related to Figure 3.
\[ w = (1 + \frac{\sigma^2 \Delta F}{\Delta F} + \alpha \frac{|\Delta F|^2}{|\Delta F|^2})^{-1} \]

\[ \text{WDF} = w \times \Delta F \]

Figure S7. (a) Example of scale factor determination. The structure factors from the -5 ps dataset (\( F_{\text{obs}}^D \)) are plotted against structure factors from the refined dark model (\( F_{\text{calc}} \)) and a line of the form \( y = mx \) is fitted to the data. For the dark structure factors, \( m = 0.75 \). In the same manner, the light structure factors for every time point are scaled to \( F_{\text{calc}} \). (b) Example of the result of scaling on dark (\( F_{\text{obs}}^D \)) and light (\( F_{\text{obs}}^L \)) structure factors (here the 100 ps dataset is shown). For the data presented in this work, this simple scaling procedure is sufficient to yield highly correlated light and dark structure factors. (c) Mathematical procedure to obtain weighted difference structure factors (WDF) from dark and light structure factors. \( \sigma \Delta F \) is the structure factor standard deviation and the term \( \alpha \) reduces the contribution of any single \( \Delta F \). (d) Weights applied to the 100 ps dataset for \( \alpha = 0.2 \), shown as an example. Difference structure factors that are large or measured with a higher error are given progressively lower weight. Related to Figure 4.
Compose data matrix $A$ of Q-weighted maps for every time point (masked around chromophore):

- Perform PCA on matrix $A$ in python
- Reconstruct maps without last, noisy component
- Visually inspect components in Coot

**Figure S8.** The pipeline implemented to denoise Q-weighted maps through PCA is shown. Each map from a collected time point is a row in the data matrix $A$ (6 time points total: 300 fs, 600 fs, 900 fs, 5 ps, 100 ps, 1 µs) and PCA is performed using the sci-kit learn package in python. The last (C6) component found explains less than 5% of the total variance in the data and, upon visual inspection, appears as noise. The first five components are shown in the bottom panels at +/-3.5σ. Related to Figure 4.

Generate q-weighted maps of the form:

$$W \Delta F_{corr} = w \times (|F_{obs_L}| - N_{bg} \times |F_{obs_D}|)$$

for $0 \leq N_{bg} \leq 1$

Save $W \Delta F_{corr}$ as CCP4 maps. In real space, identify the chromophore region ($R_{loc}$) and the rest of the protein ($R_{glob}$). ($R_{loc}$ is defined by a 5 Å sphere centered on the chromophore)

**Figure S9.** Schematic of the steps involved in the generation of background subtracted maps ($W \Delta F_{max}$ maps). Related to Figure 5.
Figure S10. Identification of the appropriate background subtraction factor ($N_{bg}$) for the $W\Delta F_{\text{max}}$ map of each timepoint (a-f). The top panels plot the correlation coefficient between the original and the background subtracted map for different $N_{bg}$ values calculated across the entire protein structure (global) and a specific 5 Å sphere centered on the chromophore (local). The bottom panels plot the difference between the global and local lines and highlight the value of $N_{bg}$ for which the difference in correlation coefficients is maximized. Related to Figures 4 and 5.

Figure S11. $W\Delta F_{\text{max}}$ maps generated using the method described in this work and illustrated in Figure S9. The following four characteristics, which we attribute to the presence of a femtosecond intermediate, are marked in the 600 fs and 900 fs maps: (i) the presence of Peak 1 (ii) an elongated and uncentered peak where the cis anti chlorine is positioned, which is in contrast with the round, centered features visible in the $W\Delta F_{\text{max}}$ maps from the later time points (iii) electron density that fills the cis anti chromophore phenol ring (iv) features that suggest a tilt of the imidazolidine ring oxygen towards the phenol ring. Though less pronounced, these features are also identifiable in the 5 ps map. Related to Figures 4 and 5.
Figure S12. WΔF\textsubscript{max} maps generated by subtraction of dark and light 2mFo-DFc maps, as described in Pearce et al.\textsuperscript{1} (in green), overlaid with the WΔF\textsubscript{max} maps from Figure S11 (in blue), both shown at +2.1σ. The two sets of maps largely overlap. The following four characteristics, which we attribute to the presence of a femtosecond intermediate, are marked in the 600 fs and 900 fs maps: (i) the presence of Peak 1 (ii) an elongated and uncentered peak where the cis anti chlorine is positioned, which is in contrast with the round, centered features visible in the WΔF\textsubscript{max} maps from the later time points (iii) electron density that fills the cis anti chromophore phenol ring (iv) features that suggest a tilt of the imidazolidone ring oxygen towards the phenol ring. Related to Figures 4 and 5.
Figure S13. The $\Delta F_{\text{max}}$ maps for the (a) 600 fs and (b) 900 fs datasets are shown together at +2.1$\sigma$ with the respective refined chromophore coordinates. The following four characteristics of these maps are marked: (i) the presence of Peak 1 (ii) a peak where the phenol ring oxygen is positioned by refinement (iii) electron density for two further carbons of the phenol ring (iv) features that suggest a tilt of the imidazolidone ring oxygen towards the phenol ring. (c-d) show the superposition of the two femtosecond refined structures, which we attribute to the intermediate trans-FS, and their relationship to the dark trans-PL configuration. Related to Figures 4 and 5.
Figure S14. Occupancy refinement for light-induced chromophore conformations. These values were used to establish occupancies for the light coordinates (in order: PDBs 8A6N, 8A6O, 8A6P, 8A6Q, 8A6R, 8A6S). Related to Figure 4.

Figure S15. Difference electron density (DED) signals in the region of the central α-helix. $\Delta F_{\text{max}}$ maps are shown for each timepoint at +/-3σ (red=negative, blue=positive). The labeled residues present the most significant DED features and suggest an overall downward movement of the helix on a sub-picosecond timescale, which is particularly evident looking at residues Leu65, Thr63, and Pro59. Similar signals persist at the longer time points (c-d) but are weaker. Related to Figure 5.
Figure S16. Difference electron density (DED) signals in the region of the β7 sheet. WΔF_{max} maps are shown for each timepoint at +/-3σ (red=negative, blue=positive). The femtosecond maps (a-b) have strong negative features on one of the His149 conformations. At 100 ps and 1 µs (c-d), there are negative signals on the side chains and backbone of residues of Asn147, Asn150, and Val151. These are superimposable between the two datasets and suggest a downward movement of the β-sheet away from the chromophore. The His149 flip that accompanies the trans-to-cis isomerization of the chromophore is marked by negative density on the trans-PL-associated conformation of this residue. Related to Figure 5.

Figure S17. (a) Transient difference absorption spectra recorded at different pump-probe time delays after a femtosecond laser excitation (400 nm) starting from the rsEGFP2 (non-chlorinated) OFF state. (b) Components fitted through global analysis of the data shown in (a), using a sequential model. Related to Figure 6.
Figure S18. UV-Vis TA global analysis results comparison for rsEGFP2 (non-chlorinated) and Cl-rsEGFP2. Respective components are overlapped in (a-d). Related to Figure 6.
Figure S19. Geometry of the QM atoms at minimum energy conical intersection point (MECI) optimized with QM-MM. Red arrows indicate movement of the bridge hydrogen required to reach trans (left arrow) or cis (right arrow) isomers of the chromophore upon deactivation. Related to Figure 7.
Figure S20. (a) Cartoon overview of Cl-rsEGFP2 with anomalous scatterer-containing residues and chromophore locations shown as sticks. (b) Anomalous difference Fourier maps for Met100, Cys82 and the chromophore above (4000 eV) and below (2800 eV) the chlorine K edge. Density presence in both maps indicates a sulphur atom, presence in the 4000 eV map combined with absence in the 2800 eV map confirms chlorine location. (c) $f'$ and $f''$ plot for Cl over the wavelength range used. Data collection wavelengths indicated as dashed lines, green for 4000eV above chlorine K edge, red for 2800 eV below. (d) Anomalous difference Fourier maps, both at 2800 eV and 4000 eV, for all scatterers in the structure shown with respective residues and chromophore. Related to Figure 3.
<table>
<thead>
<tr>
<th>Dataset</th>
<th>2.000 Å</th>
<th>1.500 Å</th>
<th>1.000 Å</th>
<th>0.500 Å</th>
<th>0.250 Å</th>
<th>0.125 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMBL</td>
<td>213992</td>
<td>18196</td>
<td>13883</td>
<td>11470</td>
<td>10052</td>
<td>7782</td>
</tr>
<tr>
<td>EMBL</td>
<td>213992</td>
<td>18196</td>
<td>13883</td>
<td>11470</td>
<td>10052</td>
<td>7782</td>
</tr>
<tr>
<td>Never</td>
<td>239516</td>
<td>22644</td>
<td>22305</td>
<td>20680</td>
<td>20680</td>
<td>20680</td>
</tr>
<tr>
<td>Never</td>
<td>239516</td>
<td>22644</td>
<td>22305</td>
<td>20680</td>
<td>20680</td>
<td>20680</td>
</tr>
<tr>
<td>X-ray</td>
<td>22093</td>
<td>22093</td>
<td>22093</td>
<td>22093</td>
<td>22093</td>
<td>22093</td>
</tr>
<tr>
<td>X-ray</td>
<td>22093</td>
<td>22093</td>
<td>22093</td>
<td>22093</td>
<td>22093</td>
<td>22093</td>
</tr>
<tr>
<td>20680</td>
<td>21000</td>
<td>10052</td>
<td>10052</td>
<td>10052</td>
<td>10052</td>
<td>10052</td>
</tr>
<tr>
<td>20680</td>
<td>21000</td>
<td>10052</td>
<td>10052</td>
<td>10052</td>
<td>10052</td>
<td>10052</td>
</tr>
</tbody>
</table>

Table S1: CrystFEL merging statistics for the eight datasets collected, reported at three different high resolution cutoffs.
Table S2. Refinement statistics for the refined dark model from the SACLA 2021 experiment, computed using the interleaved dark (dark-CW) and negative 5 ps datasets, the refined dark model from the SACLA 2019 experiment, the SSRL cryo-dataset. Related to Figures 3-5.

<table>
<thead>
<tr>
<th>Dataset:</th>
<th>SACLA 2021 interleaved dark Cl-rsEGFP2</th>
<th>SACLA 2021 interleaved dark Cl-rsEGFP2</th>
<th>SACLA 2019 interleaved dark nsEGFP2</th>
<th>SSRL Cryo+ dehydrated Cl-rsEGFP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range:</td>
<td>1.63 (1.688 - 1.63)</td>
<td>1.63 (1.688 - 1.63)</td>
<td>19.1 - 1.463 (1.515 - 1.46)</td>
<td>38.79 - 1.81 (1.875 - 1.81)</td>
</tr>
<tr>
<td>Space group:</td>
<td>P 21 21 21</td>
<td>P 21 21 21</td>
<td>P 21 21 21</td>
<td>P 21 21 21</td>
</tr>
<tr>
<td>Unit cell:</td>
<td>51.99 62.91 72.03 90 90 90</td>
<td>51.99 62.91 72.03 90 90 90</td>
<td>51.97 63.04 72.03 90 90 90</td>
<td>51.48 59.02 64.48 90 90 90</td>
</tr>
<tr>
<td>Reflections used in refinement:</td>
<td>30141 (2952)</td>
<td>30141 (2952)</td>
<td>36821 (340)</td>
<td>18407 (1804)</td>
</tr>
<tr>
<td>Reflections used for R-free:</td>
<td>1510 (146)</td>
<td>1510 (146)</td>
<td>1790 (19)</td>
<td>921 (91)</td>
</tr>
<tr>
<td>R-work:</td>
<td>0.1635 (0.3102)</td>
<td>0.1634 (0.2931)</td>
<td>0.1696 (0.4389)</td>
<td>0.2027 (0.3088)</td>
</tr>
<tr>
<td>R-free:</td>
<td>0.2013 (0.3369)</td>
<td>0.2000 (0.3428)</td>
<td>0.1952 (0.4783)</td>
<td>0.2266 (0.3376)</td>
</tr>
<tr>
<td>Number of non-hydrogen atoms:</td>
<td>2277 macromolecules: 2099</td>
<td>2277 macromolecules: 2099</td>
<td>2802 solvents: 125</td>
<td>2097 macromolecules: 1883</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1756 ligands: 63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40 solvent: 125</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>174 protein residues: 235</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>235 RMS(bonds): 0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.017 RMS(angles): 1.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.46 Ramachandran favored (%): 99.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>99.13 Ramachandran allowed (%): 0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.87 Ramachandran outliers (%): 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 Rotamer outliers (%): 2.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.09 Clashscore: 6.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.16 Average B-factor: 28.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28.44 macromolecules: 28.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28.23 ligands: 19.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19.27 solvent: 36.56</td>
</tr>
</tbody>
</table>
| | | | | 36.56
### Supporting Notes

**1. OFF State SFX Structure for Cl-rsEGFP2**

Our room temperature dark structure for Cl-rsEGFP2 presents predominantly the planar \textit{trans} \textit{anti} configuration (\textit{trans-PL}) and minor populations of \textit{trans} \textit{syn} (\textit{trans-TW}), and \textit{cis} \textit{anti} configurations (Figure S4). A previous cryotrapping crystallographic study\(^2\) identified the \textit{trans-PL} and \textit{trans-TW} configurations as photoproducts of the HT and OBF pathways respectively. The \textit{trans-PL} photoproduct was found in the structure with a contracted unit cell, while \textit{trans-TW} was the primary configuration in the structure from the larger unit cell. This led to the conclusion that the choice of pathway is dependent on the crystal packing, where tighter packing favors the volume-conserving HT. Interestingly, our dark room temperature SFX structure presents mainly the \textit{trans-PL} form, despite its large unit cell, when the chlorine substituent is present (Table S3 and Figure S5). This suggests that the \textit{cis}-to-\textit{trans} isomerization pathway at room temperature is a hula-twist, independent of unit cell dimensions. Table S3 lists the predominant conformations for the Cl-rsEGFP2 structures in this paper and in the cryotrapping study\(^2\), with the corresponding treatments, as well as other published rsEGFP2 SFX structures. Room temperature SFX data indicate that the difference in chromophore conformation observed by Chang \textit{et al.} \(^2\) was induced by the dehydration protocol or the freezing process itself, rather than dictated by a lattice-dependent change in pathway. This interpretation is further supported by a separate cryotrapping experiment (row 3 of Table S3 and Figure S6, PDB 8A83) where the ON state crystal is first irradiated with 488 nm light, dehydrated into the smaller unit cell size, and then subsequently cryocooled; the resulting structure shows a \textit{trans-PL} chromophore.

### Supporting Procedures

**1. 2021 SFX Data Analysis**

#### 1.1. Frame Data Processing and Crystallographic Analysis

A bad/damaged pixel mask for the detector was generated from dark scans recorded before data collection and was applied across all image data processing. Peak and hit finding were performed using Cheetah\(^3,4\). Identified hits were then indexed and integrated in CrystFEL\(^5,6\) using the XGANDALF algorithm\(^7\). The achieved indexing rate was around 80% for all data. Indexed crystals from CrystFEL streamfiles were binned into different time delays and separated as “light” (pump laser on) and “dark” (pump laser off) to generate respective reflection files using a custom python script. Merging statistics for the datasets collected are reported in Table S1. The PHENIX\(^8\) reflection conversion function was used to convert intensities to structure factors and model refinement was done using the phenix.refine function. Minor chromophore...
populations in the dark structure were manually refined (PDB 8A6G). Low resolution (30 Å) and high resolution (1.63 Å) cutoffs were applied. Q-weighted difference electron density maps and background subtracted maps were then generated as explained in Section 2 of the Supporting Procedures.

For the analysis of protein-wide pump-induced changes, we used the scripts by Wickstrand et al. 9. In short, a spherical volume of radius 2.0 Å was walked across every atom in the protein and the positive and negative difference electron densities from each dataset’s Q-weighted map were averaged separately within each volume. A cut-off of +/-3σ and a grid spacing of 0.5 Å were used in the scripts.

1.2. Time Zero Determination by Cross Correlation

The cross-correlation of the optical and X-rays pulses was measured via the standard technique, previously described 10. Briefly, a 50 µm thick, semi-conducting crystal (Ce:YAG) was placed in the interaction region. The transmission intensity of the optical pulse through the crystal was measured by a photo-diode. Exposure to hard X-rays causes modulation of the crystal’s reflective and transmissive properties. The amplitude of the optical transmission was monitored while the temporal delay between the two pulses was scanned. With appropriate volume of data and averaging, the resulting signal exhibits a step-like function, centred approximately when the two pulse are overlapped in the temporal delay. The signal measured on the photo-diode (Figure S3(a)) as a function of delay was fitted as previously described 11. The averaged time zero over a number of runs was used to correct the delay stage, to give an accurate time zero and, as such, an accurate binning for each data set (300 fs, 600 fs, 900 fs, 5 ps).

SACLA has recently implemented a feedback system between the optical and X-ray pulse, using a balanced optical-microwave phase detector (BOMPD) 12. This corrects for long-term temporal drift over several hours or days which previously caused movement on the order of 60 - 100 fs 11. The BOMPD feedback loop, additionally, reduces X-ray jitter stemming from an intrinsic instability in the Self Amplified Spontaneous Emission (SASE) process to sub 50 fs. Since it was difficult to suppress the actual timing drift using only a BOMPD, the drift here was further compensated using a phase shifter and the timing tool data. The jitter measured from the cross-correlation fittings is shown in (Figure S3(b)). As an additional precaution to mitigate drift, the cross-correlation was measured before the start of each day. Furthermore, the reflection intensity (transmission −1) was measured and used as a comparative cross-correlation.

With the various steps taken described above, as well as the new feedback system implemented at SACLA, we were able to accurately bin the data into sub picosecond bins (300, 600 and 900 fs) and confidently ascribe the ultrafast dynamics.

1.3. Generation of Q-weighted Maps

Through the reciprocalspaceship library 13, dark and light structure factors (F_{obs}D and F_{obs}L respectively) are in turn scaled to the structure factors calculated from the refined dark structure (F_{calc}). A simple scale factor \(m\) is applied to an entire dataset (Figure S7(a-b)). Q-weighted difference electron density maps are then calculated using weighted structure factor amplitudes, where the Bayesian weight applied is based on the work of Ursby et al. 14,15 and implemented in python using reciprocalspaceship (Figure S7(c-d)). A value of \(α = 0.2\) is used here. Maps where the scaling is done using the SCALEIT program within the CCP4 suite 16 and the Q-weighting is implemented as in 17-19 present the same key features as those generated by the method described here (not shown). The easy implementation in python, however, allows for easy manipulation and visualization of the structure factor distributions and screening of improved map generation parameters (such as scale factor and \(α\) values).

1.4. Principal Component Analysis (PCA) of Electron Density Maps

Principal component analysis (PCA) of the Q-weighted maps described above is performed in python by loading each map as a Numpy array through the GEMMI library for structural biology (GEMMI version 0.4.5). A solvent mask and a mask that only retains grid points within 7 Å of the chromophore are applied. PCA is then sklearn.decomposition.PCA function (Scikit-learn version 1.0.1). Figure S8 presents the pipeline and results for PCA on the six DED maps. The first 5 components (C1-C5) are shown (noting that the
component sign is arbitrary in the analysis). PCA here is used primarily as a method to improve the data signal-to-noise ratio: the 6th component (C6) is identified as noise and the final difference maps shown in Figure 4 are reconstructed without it.

1.5. Generation of Background Subtracted Maps (WΔF<sub>max</sub> maps)

In order to refine light-induced coordinates, one ideally wants to separate the signal that belongs to the photoproduct population from that of the dark state. Because only a small percentage of protein molecules actually undergo photoactivation, the problem of extracting the electron density of the photoproduct is a very similar issue to the one Pearce and colleagues describe of discerning minor states in macromolecular crystallography. Pearce et al. generate background corrected maps by subtracting the "ground state" 2mF<sub>o</sub>-DF<sub>c</sub> map from the "dataset" 2mF<sub>o</sub>-DF<sub>c</sub> map as follows:

\[
[\text{corrected map}] = [\text{dataset map}] - \text{N}_{bg} \times [\text{ground state map}]
\]

where the background correction factor (we refer to it here as N<sub>bg</sub>, in it is called BDC – Background Density Correction factor) is the value that maximizes the difference in correlation (calculated from the ground state and the corrected map) between the entire protein and a specific area of change.

We have extended this here to Q-weighted maps. Figure S9 describes the steps involved in generating what we have called WΔF<sub>max</sub> maps. Q-weighted maps of the form:

\[
\text{WΔF}_{\text{corr}} = w \times (|F_{\text{obs}}| - \text{N}_{bg} \times |F_{\text{obs}}|)
\]

are generated for a range of N<sub>bg</sub> between 0 and 1 and saved as electron density maps in CCP4 format. In real space, a local region, R<sub>loc</sub>, is defined as a sphere of 5 Å centered around the chromophore double bond. The entire protein is defined as R<sub>glob</sub> after a solvent mask is applied. For each value of N<sub>bg</sub>, we compute the Pearson correlation coefficient between the respective WΔF<sub>corr</sub> map and the map obtained from the dark model calculated structure factors (F<sub>calc</sub>). This is done for both R<sub>loc</sub> and R<sub>glob</sub>. We then choose the N<sub>bg</sub> value that maximizes the difference between these two correlation coefficients and save the corresponding map (WΔF<sub>max</sub>) for that timepoint. The determination of the appropriate background subtraction value for each timepoint is shown in Figure S10 and the corresponding WΔF<sub>max</sub> maps are shown in Figure S11. The cis anti photoproduct is discernible already at 300 fs and is very clear in the 100 ps and 1 μs maps. The maps from the 600 fs and 900 fs datasets contain additional features (numbered (i-iv)) that support the presence of the femtosecond intermediate trans-FS (see main text and Figure S13). The 5 ps WΔF<sub>max</sub> map, just as the respective Q-weighted map, is noisier than the other datasets, though it still contains weak trans-FS features and some cis anti population. Figure S12 displays background corrected maps generated by subtraction of light and dark 2mF<sub>o</sub>-DF<sub>c</sub> as done in 1. These are very similar to the ones calculated directly from reciprocal space and support the conclusions drawn above.

Refinement of Cl-rsEGFP2 coordinates to WΔF<sub>max</sub> maps is done, starting from the dark coordinates, with phenix.refine by allowing only atom positions from the chromophore and the residues immediately next to it in the sequence (residues 63-65 and 69-71) to vary. This is because the background subtracted structure factors are much lower signal-to-noise than the ones used to refine the dark structure, and we therefore limit their use to the chromophore region where light-induced changes are strong enough to clearly reveal the presence of minor populations. Occupancies for the cis anti and trans-FS species in light datasets were refined so as to minimize R<sub>work</sub> (Figure S14). Despite the weak presence of trans-FS features in the 5 ps maps, trans-FS occupancy refinement (not shown) for this timepoint did not yield a significant population, so this conformation was not included in the final coordinates. Final structures for each time point are deposited (in order from 300 fs to 1 μs) as PDBs 8A6N, 8A6O, 8A6P, 8A6Q, 8A6R, 8A6S.

2. 2019 SFX for OFF-state rsEGFP2 structure

Dark data was collected at SACLA at the BL3 EH2 beamline with setup equivalent to the one described above (MPCCD-phase III detector, 10.5 keV), but a refined detector distance of 50.6 mm. Data was then
processed as described for the 2021 SFX experiment, with the difference that structure refinement was
done in REFMAC\textsuperscript{20} (PDB 8A7V).

3. Extra-Cryotrapping Experiment

Crystals were grown as described in\textsuperscript{2}. For the irradiation/dehydration protocol, the crystal was fished
onto a loop and transferred to a separate droplet of mother liquor on a cover slip. The entire cover slip was
picked up with tweezers and placed in the beam of a 488nm laser (specifics as in\textsuperscript{3}). Illumination lasted
between 20-30 seconds. Following this, the crystal was transferred to a drop of dehydrating cryoprotectant\textsuperscript{2}
and placed long enough to contract into the smaller unit cell size (approximately 5 seconds). It was then
fished out onto a loop again and plunged into liquid nitrogen. Data was collected at the SSRL 7-1 beamline.
Data reduction and refinement were then done using XDS\textsuperscript{21} and Phenix\textsuperscript{8} (PDB 8A83). Refinement statistics
for this cryotrapped structure and for the dark structures collected at SACL\textsuperscript{A} are reported in Table S2.

4. Quantum Chemical Modeling Details

QM-MM simulations in the protein were performed using our own version of GROMACS 4.6.5\textsuperscript{22,23}
coupled to the Terachem\textsuperscript{24,25} (for ground state) and GAMESS(US)\textsuperscript{26} (for excited-state) quantum chemistry
packages. We searched for minimum-energy geometries in all systems using the limited-memory
Broyden-Fletcher-Goldfarb-Shannon (L-BFGS) quasi-Newton optimization algorithm, without PBC and
with infinite cut-offs for the Coulomb and Lennard-Jones interactions. All optimizations in the work were
performed until the maximum component of the force was lower than 10 kJ/mol/nm. Vertical absorption
and emission spectra calculations were performed with xMCQDPT2 method implemented in the Firefly
package\textsuperscript{27}.

4.1. Planar Ground and Excited State trans-PL Structures

Initially, we performed optimization of the Cl-rsEGFP2 on ground and excited states, starting from the
dark trans-PL crystal structure. On the ground state, we performed optimization with density functional theory
(DFT) at PBE0/cc-pVDZ//Amber03\textsuperscript{28} with empirical corrections to dispersion energies and interactions
introduced with Grimme’s DFT-D3 model\textsuperscript{31}. On the excited state, a double optimization scheme was
employed. At first, SA2-CASSCF(2,2)/3-21G//Amber03\textsuperscript{32} was used to minimize the S1 state. We used
a small active space in this optimization to prevent the interchange between the S1 and S2 states, which
typically happens for the neutral GFP chromophores when using larger active spaces without electron
correlation. The structure optimised with the small active space, was subject to a second optimisation at the
SA2-CASSCF(12,11)/3-21G//Amber03 level of theory.

Both S0 and S1 optimized structures are planar. Next, we computed vertical excitation and emission
energies at the xMCQDPT2/SA6-CASSCF(12,11)/cc-pVDZ//Amber03 level of theory. All 6 states were
included into averaging as well as into the effective Hamiltonian. Results suggest absorption at 405 nm
(3.06 eV) and emission at 518 nm (2.39 eV). In addition for the S1 minimum energy geometry, we computed
an excited-state absorption (ESA) from the S1 into the S5 of 425 nm (2.92 eV). These excitation energies
are in line with the experimental data (Figure 6), and hence suggest that the model provides an adequate
qualitative description for our system.

4.2. Identification of the Photoisomerization Pathway

At first, we performed a search for the isomerization pathway connected to the TR-SFX-resolved
trans-FS structure at 900 fs. Optimization on the ground (PBE0/cc-pVDZ//Amber03) and excited
(SA2-CASSCF(12,11)/3-21G//Amber03) states resulted in the planar trans-PL structure, suggesting that
this twisted structure is not connected to any minima on the excited state. To identify minimum energy
conical intersection points(MECI), we implemented a penalty function MECI search algorithm\textsuperscript{33} with $\sigma$=0.02
and $\alpha$=16 as the parameters for the penalty function. When starting the optimization from the the trans-FS
structure, no MECI could be located, in line with previous computational studies\textsuperscript{34,35} that suggest that in the
neutral GFP-chromophores there is no conical intersection associated with the rotation of the phenol ring.
We next searched for a hula-twist MECI by manually increasing $\tau$ torsion angle by 45° in the trans-FS structure and optimising its geometry at the SA2-CASSCF(12,11)/3-21G//Amber03 level of theory. This optimisation resulted in a twisted structure with an S1/S0 energy gap of $\approx 0.035$ a.u. Starting a MECI optimization from this geometry leads to the MECI structure shown in Figure S19 with an S1/S0 energy gap of $\approx 0.0005$ a.u. After the optimizations we also recomputed energies for both twisted minima and MECI points with xMCQDPT2/SA6-CASSCF(12,11)/cc-pVDZ//Amber03. In addition, to determine whether the MECI point connects the anti-trans conformer to the anti-cis conformer, we performed a ground state optimisation at PBE0/cc-pVDZ//Amber03 level starting from the MECI. While a direct optimization leads to the original planar anti-trans-PL chromophore conformer, a slight perturbation of the position of the methylene hydrogen atom by $\approx 0.3$ Å towards a more cis-like position, leads to the anti-cis conformer.

5. Confirming the Presence of the Chlorine Substituent

To confirm incorporation of the chlorine substituent into the chromophore, we analyzed anomalous difference Fourier maps from anomalous data taken at different wavelengths. Single crystals were obtained by the hanging-drop vapor diffusion method at 20 °C described previously. Briefly, the protein solution (12 mg/mL, 50mM Hepes pH 7.5, 20 mM NaCl) was mixed 1:1 with the precipitant solution (100 mM Hepes pH 8.0, 1.80 M ammonium sulphate, 20 mM NaCl) to yield 2-4 $\mu$L drops that were placed over a well containing 1 ml of the precipitant solution. Mature crystals with dimensions up to 500 $\mu$L x 200 $\mu$L x 200 $\mu$L appeared after three weeks. Prior to flash-freezing in liquid nitrogen, crystals were cryoprotected by passage through three solutions of incrementally increasing amounts of sucrose to a final concentration of 1.2M in 75 mM Hepes pH 8.1, 20 mM NaCl, 0.9 M ammonium sulphate.

Data was collected on beamline I23, Diamond Light Source, using the Pilatus 12M semi-cylindrical detector at a temperature of 50K. Data were collected above and below the chlorine K absorption edge at 4.0 keV ($\lambda=3.1$ Å) and 2.8 keV ($\lambda=4.4$Å) respectively. Each dataset consisted of 3600 images at 0.1°irc oscillation with 0.1s exposure. 3 x 360°irc datasets at 4.0 keV were used to phase the structure using CRANK2. Further structure refinement was performed in Phenix and Coot (PDB 8AM4). Anomalous difference Fourier maps were generated using ANODE. Figure S20 shows the resulting anomalous difference Fourier maps from 4.0 keV and 2.8 keV. The presence of density in the 4.0 keV dataset (green mesh, 5σ) in conjunction with the absence of density in the 2.8 keV dataset (red mesh, 4σ) is conclusive of Cl, confirming the identity of the chromophore heavy atom substitution.

6. OFF-to-ON Quantum Yield for the Cl-rsEGFP2 construct

We carried out a comparative measurement to estimate the reaction quantum yield (QY) for our chlorinated rsEGFP2 construct. The OFF-to-ON QY for unchlorinated rsEGFP2 was most recently estimated to 0.23. rsEGFP2 and Cl-rsEGFP2 solutions were fully converted to the OFF state using through illumination with a 488 nm LED. The final OD at 405 nm for both was $\approx 0.25$. The OFF-to-ON conversion was then driven with a 405 nm LED at 4.4 mW power and the absorption at 488 nm and 405 nm monitored with an Agilent 8453 spectrophotometer.

Triplicate measurements were averaged and normalized. Figure S1 shows the 405 nm absorption decay for both constructs with the respective fitted decay rates ($k$). From the ratio of the two fitted rates we estimate the OFF-to-ON QY for Cl-rsEGFP2:

$$\phi_{CI} = \phi_{noCI} \cdot \frac{k_{CI}}{k_{noCI}} = 0.2$$ (3)
Supporting References


