Structural Evidence of Photoisomerization Pathways in Fluorescent Proteins

Jeffrey Chang, Matthew G. Romei, and Steven G. Boxer

Department of Physics, Stanford University, Stanford, California 94305, United States
Department of Chemistry, Stanford University, Stanford, California 94305, United States

Supporting Information

ABSTRACT: Double-bond photoisomerization in molecules such as the green fluorescent protein (GFP) chromophore can occur either via a volume-demanding one-bond-flip pathway or via a volume-conserving hula-twist pathway. Understanding the factors that determine the pathway of photoisomerization would inform the rational design of photoswitchable GFPs as improved tools for super-resolution microscopy. In this communication, we reveal the photoisomerization pathway of a photoswitchable GFP, rsEGFP2, by solving crystal structures of cis and trans rsEGFP2 containing a monochlorinated chromophore. The position of the chlorine substituent in the trans state breaks the symmetry of the phenolate ring of the chromophore and allows us to distinguish the two pathways. Surprisingly, we find that the pathway depends on the arrangement of protein monomers within the crystal lattice: in a looser packing, the one-bond-flip occurs, whereas, in a tighter packing (7% smaller unit cell size), the hula-twist occurs.

Light-driven chemistry often involves the cis-trans photoisomerization of a double bond. Photoisomerization plays a central role in photoreceptive biological functions such as vision and phototaxis, as well as in broad areas of chemical and industrial importance including optogenetics, optical data storage, molecular switches, and molecular motors. In conjugated systems such as the green fluorescent protein (GFP) chromophore, double-bond isomerization can proceed either via a one-bond-flip (OBF) pathway, in which only the isomerizing τ-bond rotates, or via a hula-twist (HT) pathway, depending on which chemical bonds rotate during the process (Figure 1). Since the OBF atomic movement sweeps out a greater volume, the OBF pathway tends to occur in fluid solution where the surroundings allow for free movement, whereas the volume-conserving HT pathway is favored in sterically constrained environments such as vitriﬁed solvents. Currently, it is not deﬁnitively known whether the protein environment surrounding the GFP chromophore favors isomerization via the OBF or HT pathway.

To address this question, we adopt the model system of rsEGFP2 (reversibly switchable enhanced green fluorescent protein 2), a photoswitchable fluorescent protein which is readily crystallizable and has a well-characterized photocycle. The chromophore within the rsEGFP2 β-barrel naturally adopts the cis isomer and readily photoisomerizes to the less fluorescent trans isomer upon 488 nm light irradiation; the trans isomer, in turn, photoisomerizes back to the cis isomer upon 405 nm light irradiation (Figure 2A). The higher-energy trans isomer is relatively stable at room temperature and thermally relaxes to the cis isomer with a rate constant of 0.2 h−1. Since rsEGFP2 ﬂuorescence can easily be toggled by laser excitation, its photoswitching property is often exploited for cellular imaging techniques such as super-resolution microscopy. Understanding the pathway of the photochemistry would help inform the rational design of fluorescent proteins as imaging tools.

An additional advantage of rsEGFP2 is that substituents can be readily incorporated into the chromophore to distinguish the OBF and HT pathways. In the wild-type chromophore, both pathways yield the same trans product, but if a substituent breaks the symmetry of the chromophore’s phenolate ring, then the two pathways yield distinct products. Specifically, as shown in Figure 2B, when the cis chromophore with a substituent anti to the double-bonded imidazolinone nitrogen undergoes OBF isomerization, the resulting product is trans syn, whereas if the initial cis anti state undergoes HT isomerization, then the product is trans anti. Thus, placing a
maintained similar photoswitching properties as the wild-type protein variant with this substitution, denoted Cl-rsEGFP2, be in and more broadly suggest that a protein chromophore dynamics are sensitive to their surroundings, into a crystal lattice. Our structures demonstrate that the pathway depends on how the rsEGFP2 monomers pack and after 488 nm light irradiation. Remarkably, we found that isomerization pathway of an rsEGFP2 variant with a chlorotyrosine, the autocatalytic chromophore formation reaction produces a chromophore with a chlorine substituent on the chromophore allows us to deduce the electron density of the nonirradiated crystals was well modeled by a chromophore in the cis state; the irradiated crystals required a mixture of cis and trans structural models to explain the electron density. Crystallographic data collection and refinement statistics are listed in Table S2, and the minor conformations of the irradiated crystals, such as the residual cis population from incomplete isomerization, are discussed in Section S2.

As expressed with the chromophore in the cis state, the chlorine substituent is incorporated completely in the anti orientation (Figure 3A). Our observation of a single substituent orientation is consistent with previous protein crystal structures containing chlorinated tyrosines. When the chromophore is photosomerized to the trans state, the chlorine substituent ends up in the syn orientation, and the two rings of the chromophore are no longer planar with respect to one another (τ = 12°, ϕ = -13° in cis; τ = 187°, ϕ = -32° in trans, where τ is the N=C=C=C1 dihedral angle and ϕ is the C=C−C1−C6 dihedral angle (Figure 2B)). In addition, the surrounding pocket distorts slightly to accommodate the new trans geometry: the phenolic oxygen loses its hydrogen bond to His149 and instead forms a hydrogen bond with a new water molecule in the trans structure; in turn, the Tyr146 residue adopts a different rotamer and acts as a surrogate hydrogen-bonding partner to His149 (Figure 3A).

While collecting data on several Cl-rsEGFP2 crystals, we observed that some crystals had a unit cell size smaller than that of previously reported crystals and the ones discussed above by up to 7% (Table S2). It is well-known that protein crystals can expand or contract when water enters or exits the solvent channels between individual proteins, and in our case, slight variations in experimental handling caused the solvent content of Cl-rsEGFP2 to vary. These variations included the incubation time in cryoprotectant and the exposure time of cryoprotectant to ambient air during the looping and irradiation process. Upon further investigation, we discovered that we could controllably diminish the lattice spacing of Cl-rsEGFP2 crystals by soaking them in a cryoprotectant of low relative humidity (see methods in Section S1), which pulled water out of the crystal and forced the proteins to pack more closely together. The resulting “contracted” crystal consistently had a smaller unit cell volume (19 900 Å³ vs 21 400 Å³) and a lower solvent content (29% vs 34%) than the “expanded” crystal (Table S2).

When the Cl-rsEGFP2 unit cell contracts into the tighter lattice, the protein retains its β-barrel fold but develops slight deviations in tail and loop regions compared with the more expanded lattice (Figure S4). The interior of the protein around the cis state of the chromophore maintains the same geometry as in the expanded lattice, but the chromophore pocket is significantly distorted around the trans state of the chromophore when the lattice contracts (Figure 3B). Interestingly, the trans structure in the contracted lattice has the chlorine substituent in an anti orientation, in contrast to the expanded lattice. Furthermore, the two rings of the

rsEGFP2 (Figure S1) and formed crystals suitable for X-ray diffraction. To capture the photosomerized trans state, the Cl-rsEGFP2 crystals were looped and irradiated with 488 nm light at room temperature, and then immediately cryocooled (see Section S1 for detailed methods). We collected X-ray diffraction data sets of cryocooled Cl-rsEGFP2 crystals before and after irradiation and found that the crystals had lattice constants similar to the previously reported wild-type structures (Figures S2 and S3). The electron density of the nonirradiated crystals was well modeled by a chromophore in the cis state; the irradiated crystals required a mixture of cis and trans structural models to explain the electron density. Crystallographic data collection and refinement statistics are listed in Table S2, and the minor conformations of the irradiated crystals, such as the residual cis population from incomplete isomerization, are discussed in Section S2.

In this study, we investigated the cis-to-trans photosomerization pathway of an rsEGFP2 monomers pack into a crystal lattice. Our structures demonstrate that chromophore dynamics are sensitive to their surroundings, and more broadly suggest that a protein’s internal structure can be influenced by crystallization conditions.

rsEGFP2 was expressed with a monochlorinated chromophore using amber suppression. Upon replacing the tyrosine at residue 67 with the noncanonical amino acid 3-chlorotyrosine, the autocatalytic chromophore formation reaction produces a chromophore with a chlorine substituent ortho to the phenolic oxygen (Figure 2B). The resulting protein variant with this substitution, denoted Cl-rsEGFP2, maintained similar photoswitching properties as the wild-type
The Cl-rsEGFP2 structures also present a striking example of how crystallization conditions may perturb the structure of buried portions of proteins, and, in this case, with functional consequences. It is well-accepted that the outward-facing residues and flexible loops of proteins are distorted by interactions between proteins in a crystal, but it is less common for the interior of the protein matrix to be affected by packing interactions. Nevertheless, the crystal structures of Cl-rsEGFP2 exhibit a significant distortion of the internal chromophore pocket upon 7% unit cell contraction (Figure 3, Figure S5), indicating that crystal packing affects the conformation of internal residues in Cl-rsEGFP2.

It is important to note that the Cl-rsEGFP2 structures reported here only provide snapshots before and after photoisomerization rather than a time-resolved depiction of the process. The mechanistic conclusions drawn in this paper rely on the assumption that the trans crystal structures reflect the chlorine orientation immediately following photoisomerization, which presumes that the trans and trans anti states do not interconvert on the experimental time scale (that is, in the few seconds between photoisomerization and cryocooling). The interconversion between syn and anti conformers (a phenolate ring-flip) requires rotation about the $\omega$-bond (Figure 2B), which has substantial double-bond character since it is conjugated to a large $\pi$-system. In Section S3, we argue that the energetic barrier is likely too costly for this process to occur freely at room temperature within the relevant time scale. More definitive proof of the photoisomerization mechanism would come from time-resolved serial femtosecond crystallography of rsEGFP2 with a substituted chromophore, which would reveal snapshots of the atomic trajectory immediately following photon absorption. The atomic motion of isomerization is sensitive to the steric properties of the surrounding medium, it is reasonable that the arrangement of the trans chromophore and its environment between the contracted and expanded lattices (Figure S5) were reproducibly observed in multiple crystals.

The crystal structures of Cl-rsEGFP2 indicate that a subtle rearrangement of the crystal lattice is sufficient to alter the photoisomerization pathway. As illustrated in Figure 3A and 3B, photoisomerization occurs via the OBF pathway (cis anti $\rightarrow$ trans syn) in the expanded crystal and via the HT pathway (cis anti $\rightarrow$ trans anti) in the contracted crystal. Since cis$\rightarrow$trans isomerization is sensitive to the steric properties of the surrounding medium, it is reasonable that the arrangement of rsEGFP2 monomers within the crystal lattice affects the photoisomerization pathway. We speculate that the tighter packing and lower solvent content of the contracted crystal create a more rigid chromophore pocket that restricts atomic motion during isomerization, favoring the volume-conserving HT trajectory.

Our structures demonstrate that the dynamics of the chromophore pocket inside Cl-rsEGFP2 are affected by changes in the external environment around the protein. This result is consistent with a recent study by Kao et al. on the reversibly photoswitchable fluorescent protein Dronpa, in which the authors observe that increasing the viscosity of the surrounding medium slows down the photoswitching rate. The fact that both solvent viscosity and crystal packing affect chromophore isomerization suggests that photoisomerization of the internal chromophore is coupled to motion of the protein $\beta$-barrel.
ization occurs within picoseconds of excitation, as shown from pump–probe studies, so the time resolution of a serial femtosecond crystallography experiment should be sufficient to capture the isomerization process as it occurs.

In summary, we have observed mechanistic details of cis–trans photoisomerization in crystals of Cl-rsEGFP2. Remarkably, the pathway is not intrinsic to the protein but depends on the packing of the protein within the crystal: an expanded crystal lattice favors the more volume-demanding OBF pathway, while a 7% reduction in unit cell size instead selects for the HT pathway. This finding suggests chromophore isomerization in fluorescent proteins is coupled to motion of the surrounding β-barrel, and that the chromophore’s photoisomerization mechanism may not be conserved across different fluorescent protein variants. Our results also indicate that crystallization conditions can cause structural rearrangements of interior-facing residues in proteins, suggesting that crystal structures should be interpreted with care.

**REFERENCES**


**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.9b08356.

Protein and DNA sequences, detailed experimental methods, discussion of minor populations, discussion of ring-flip time scale, comparison of crystal structures, and crystallographic data and refinement statistics (PDF)

**AUTHOR INFORMATION**

**Corresponding Author**

sboxer@stanford.edu

Jeffrey Chang: 0000-0001-9342-5072
Matthew G. Romei: 0000-0001-6798-7493
Steven G. Boxer: 0000-0001-9167-4286

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under 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, National Institute of General Medical Sciences. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. Part of this work was performed at the Stanford Nano Shared Facilities (SNSF)/Stanford Nanofabrication Facility (SNF), supported by the National Science Foundation under Award ECCS-1542152. We thank Dr. Irimpan Mathews at the SSRL for assistance with data collection, Dr. Jacek Kozuch for conceiving the project idea, and Chi-Yun Lin for many helpful discussions. Support from the Stanford Bio-X Undergraduate Fellowship (to J.C.) and the Center for Molecular Analysis and Design Graduate Fellowship (to M.G.R.) is gratefully acknowledged. This work was supported, in part, by NIH Grant GM118044 (to S.G.B.).

Supporting Information for
“Structural Evidence of Photoisomerization Pathways in Fluorescent Proteins”

Jeffrey Chang†, Matthew G. Romei‡, and Steven G. Boxer*†

†Department of Physics, Stanford University, Stanford, California 94305, United States
‡Department of Chemistry, Stanford University, Stanford, California 94305, United States
*Correspondence to: sboxer@stanford.edu

Table of Contents

Section S1 Methods.........................................................................................S2
  Plasmid Construction....................................................................................S2
  Protein Sequence.........................................................................................S2
  DNA Sequence..............................................................................................S2
  Protein Expression.......................................................................................S2
  Protein Purification......................................................................................S3
  Protein Crystallization and Cryoprotection.................................................S4
  X-Ray Data Collection and Structure Refinement ..........................................S5
Section S2 Minor populations in the irradiated crystal structures..................S7
Section S3 Discussion of the timescale of ring-flipping between anti and syn conformers..S8
Section S4 Supplementary Figures ..................................................................S10
Section S5 References .....................................................................................S15

S1
Section S1  Methods

Plasmid Construction

The rsEGFP2 gene in the pQE-31 plasmid was generously provided by Stefan Jakobs at the Max Planck Institute for Biophysical Chemistry and inserted into the pET-15b vector (Novagen) using NEBuilder HiFi DNA assembly (New England Biolabs), retaining the start codon and polyhistidine affinity tag of the pQE-31 plasmid. The amber stop codon (TAG) was inserted at Y67 using the QuikChange Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer’s instructions. The protein and amino acid sequences are shown below. The chromophore-forming residues are underlined; the amber suppression site is bolded.

Protein Sequence

MRGSHHHHHHTDPMVSKGEELEFTGVVPILVELGDVNGHKFSVSXEGEGDATYKLTTLKFACTTGGKLPVPW
PTLVTLALLYGVCFSRYPDHKQHDFKSAMPEGYVQERTIFFKDDGNYKTRAEEKFEGDTLNVRIELKGI
KEDGNILGHKLEYNNSHNYIMADPKNGIKSKNFKIRHNIEDGVSQADHYQQNTPIGDPVLLPDNYLS
TQSKLSKDPNEKRHDMLVLEFVTAAGRITLMDELYK

DNA Sequence

ATGAGAGGATCTCACCATCACCATCACCATCACCATACGGATCCGATGGTGAGCAAGGGCGAGGAGCTGTTCACC
GGGTTGGTGGCCCATCTGTCGGAGCTGAGCGACGACGTAACCGGCAAGATCCGCTGTCGCCGCGA
GGGCCAGGCGGATGCCCACCTACGGGCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCTGCCGCCAAGCT GCC

Protein Expression

Chemically competent BL21 Escherichia coli cells (Invitrogen) were co-transformed with (1) a pET-15b expression vector containing the Y67TAG rsEGFP2 gene and (2) a pEVOL vector containing the 3-chlorotyrosine (3-ClY) amber suppression machinery. Single colonies of these cells were inoculated into 10 mL of LB Miller (25 g/L) containing 100 mg/mL ampicillin (Sigma-
Aldrich, CAS 69-52-3) and 35 mg/L chloramphenicol (Sigma-Aldrich, CAS 56-75-7) and grown overnight to saturation at 37°C. The starter cultures were used to inoculate 1 L of 46.7 g/L Terrific Broth Modified (Fisher), 8 mL glycerol (Fisher, CAS 56-81-5), 0.1 mg/L 3-chlorotyrosine (Sigma-Aldrich, CAS 7423-93-0), 100 mg/L ampicillin, and 35 mg/L chloramphenicol. Cells were grown in baffled 3-L flasks at 37°C with 200 rpm shaking until an O.D. of 0.6-0.7 at 600 nm was reached. At this point, the flasks were placed on ice, and 0.5 g/L of L-arabinose (Sigma-Aldrich, CAS 5328-37-0) was added to the cell culture to induce expression of the amber suppression machinery. Once the flasks reached 20 °C, 0.24 g/L of isopropyl β-D-1-thiogalactopyranoside (Fisher, CAS 367-93-1) was added to induce rsEGFP2 expression. The cells were then incubated overnight (~17 hours) at 20 °C while shaking at 200 rpm.

**Protein Purification**

*E. coli* containing the protein of interest were separated from the media via centrifugation at 6500 rcf for 30 min. The cell pellet was resuspended in lysis buffer, an aqueous buffer at pH 8.0 containing 50 mM Tris-HCl (Fisher, CAS 1185-53-1) and 300 mM NaCl (Fisher, CAS 7647-14-5). The cells were then lysed by a high-pressure homogenizer (Avestin EmulsiFlex-C3). The lysate was centrifuged at 25000 rcf for 90 minutes, and the supernatant was centrifuged again at 25000 rcf for another 90 minutes to pellet any remaining cell debris. The resulting supernatant was passed through a 0.45 µm filter (Millex-HV Syringe Filter Unit, PVDF; Millipore) and then added to a column of Ni-NTA Agarose resin (QIAGEN) that was pre-equilibrated with lysis buffer. The column was rinsed with 3 column volumes of lysis buffer containing 20 mM imidazole (Aldrich, CAS 288-32-4) and then eluted with 3 column volumes of lysis buffer containing 200 mM imidazole. All the eluting fractions were inspected by eye, and those that were judged to contain rsEGFP2 were pooled together and exchanged into Buffer A, an aqueous buffer at pH 8.0 containing 10 mM NaCl and 50 mM Tris-HCl. The protein was then purified by anion-exchange chromatography (HiTrap 5 mL Q HP; GE Healthcare) with a gradient of Buffer A and Buffer B (1 M NaCl and 50 mM Tris-HCl pH 8.0 aqueous buffer).

The identity of the protein was confirmed with electrospray ionization mass spectrometry (ESIMS) measured with LC-MS (Waters 2795 HPLC with ZQ single quadrupole MS in Stanford University Mass Spectrometry (SUMS) facility). As seen in Table S1, we observed a systematic discrepancy of about 145 Da between the observed and expected masses of the protein. This discrepancy was puzzling at first because the residues resolved in the crystal structure showed no deviation in the amino acid identities, and an MS/MS analysis also showed no mass discrepancies. However, the N-terminal tail is not completely resolved in the crystal structures, and after digestion with trypsin and LysC, the remaining N-terminal peptide was too short to resolve by MS/MS. Hence any possible modifications must be localized to the N-terminal residue. In agreement with an N-terminal modification, the N-terminal peptide formed from formic acid cleavage of the Asp-Pro peptide bond had a mass of 146 Da greater than the
expected peptide (RGSHHHHHHTD). This mass difference can be attributed to an N-terminal methylated methionine. N-terminal methionine cleavage often occurs during recombinant protein expression in *E. coli*, but the native enzyme responsible for this cleavage, methionyl-aminopeptidase, is hindered by a bulky neighboring residue. The arginine in the second position of the rsEGFP2 sequence likely prevents N-terminal methionine cleavage. Methylation at the N-terminus of proteins in *E. coli* has been observed previously.

Table S1. Expected and observed mass for the proteins in this study.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Expected massa (Da)</th>
<th>Observed mass (Da)</th>
<th>Expected mass of N-terminal peptide after formic acid digestionb (Da)</th>
<th>Observed mass of N-terminal peptide after formic acid digestion (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rsEGFP2</td>
<td>28358</td>
<td>28503</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Cl-rsEGFP2</td>
<td>28392</td>
<td>28536 (+33 from wild-type)</td>
<td>1357</td>
<td>1503 (+146 from expected)</td>
</tr>
</tbody>
</table>

a Predicted from the primary sequence of the protein with N-terminal methionine removed, and subtracting 20 Da for the dehydration and reduction of chromophore formation.

b Predicted from the primary sequence of the N-terminal peptide after cleavage of the Asp-Pro bond, with the N-terminal methionine removed.

**Protein Crystallization and Cryoprotection**

Y67(3-CiY) rsEGFP2, denoted Cl-rsEGFP2, was crystallized by hanging-drop vapor diffusion in a VDX plate with sealant (Hampton Research). The protein was first exchanged into an aqueous buffer at pH 7.5 of 50 mM HEPES (Sigma-Aldrich, CAS 7365-45-9) at a protein concentration of ~12 mg/mL. This protein solution was mixed 1:1 with the mother liquor (0.1 M HEPES pH 8.1, 1.76 M ammonium sulfate (Sigma-Aldrich, CAS 7783-20-2)) to form the initial 2 μL drop on a plastic cover slip, which was inverted over a well with 0.5 mL of mother liquor. Orthorhombic crystals of ~50-300 μm in size appeared after 3-7 days of incubation at room temperature in the dark.

Once the Cl-rsEGFP2 crystals were grown, they were looped from the mother liquor drop, placed briefly (~3-5 s) in solution of cryoprotectant (see below), optionally irradiated (see below), and immediately plunged into liquid nitrogen to trap the species formed by irradiation. The expanded Cl-rsEGFP2 crystals were soaked in a cryoprotectant containing 0.1 M HEPES pH 8.1, 1.7 M ammonium sulfate, and 1 M sucrose (Sigma-Aldrich, CAS 57-50-1), whereas the contracted crystals were soaked in a more dehydrating cryoprotectant containing 0.16 M HEPES pH 8.1, 2.72 M ammonium sulfate, and 1.6 M sucrose. The non-irradiated Cl-rsEGFP2 crystals were cryocooled immediately after cryoprotection, whereas the irradiated Cl-rsEGFP2 crystals were obtained by placing the loop containing the crystal in the path of a 27 mW 488 nm diode laser (85-BCD-030-115, Melles Griot) after incubation in cryoprotectant and prior to cryocooling.

We observed that the exposure time to the more dehydrating cryoprotectant was an important variable – too short an exposure did not allow the unit cell to contract sufficiently to
distort the chromophore pocket, whereas too long an exposure caused the diffraction quality to degrade. We also observed that \( \sim 3 \) s of irradiation with our 27 mW laser was sufficient to minimize the residual cis population visible in the electron density from subsequent x-ray diffraction analysis. Since the irradiated, contracted Cl-rsEGFP2 crystals are subjected both to osmotic shock from the dehydrating cryoprotectant and to mechanical stresses from photoisomerization, their diffraction quality degrades especially rapidly, but by limiting the irradiation time to \( \sim 1 \) second and moving quickly during the experimental procedure, we were able to preserve the crystal’s high-quality diffraction. Subsequent modeling of the electron density showed that this brief irradiation photoswitches about 50% of the monomers in the crystal lattice to the trans state. The contracted irradiated crystal was modeled by a mixture of cis and trans states; the trans state is the one discussed throughout this paper.

**X-Ray Data Collection and Structure Refinement**

X-ray diffraction data were collected at 100 K at the Stanford Synchrotron Radiation Lightsource on beamlines 12-2 for expanded cis, 14-1 for expanded trans, and 9-2 for contracted cis and trans. The data were processed (indexing, integration, scaling, and merging) with X-ray Detector Software (XDS)\(^8\) using the autoxds script.\(^9\) Molecular replacement was performed with phenix.phaser\(^10\) using the wild-type rsEGFP2 structures (Protein Data Bank entries: 5dtx and 5dty\(^11\)) as search models for the cis and trans states, respectively. Cycles of model building and refinement were done using Coot\(^12\) and phenix.refine. During the initial rounds of refinement, the occupancy of the chromophore was set to 0 to avoid biasing its isomerization state or substituent orientation. Once the \( 2mF_{\text{obs}} - DF_{\text{calc}} \) electron density of the chromophore became unambiguous, the chromophore was added to the atomic model, and the structures were refined until the R-free score converged. Data collection and refinement statistics are listed in Table S2.

**Table S2. Crystallographic data collection and refinement statistics.**

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Expanded cis</th>
<th>Expanded trans</th>
<th>Contracted cis</th>
<th>Contracted trans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>6PFR</td>
<td>6PFT</td>
<td>6PFU</td>
<td></td>
</tr>
<tr>
<td>Resolution range</td>
<td>0.97946</td>
<td>1.305053</td>
<td>0.97946</td>
<td>0.97946</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 volume (Å(^3))</td>
<td>214140</td>
<td>219540</td>
<td>198950</td>
<td>199660</td>
</tr>
<tr>
<td>a (Å)</td>
<td>50.699</td>
<td>51.014</td>
<td>51.422</td>
<td>51.320</td>
</tr>
<tr>
<td>b (Å)</td>
<td>62.489</td>
<td>62.716</td>
<td>59.246</td>
<td>59.479</td>
</tr>
<tr>
<td>c (Å)</td>
<td>67.591</td>
<td>68.619</td>
<td>65.302</td>
<td>65.408</td>
</tr>
<tr>
<td>α</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>β</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Total reflections</td>
<td>64376 (6726)</td>
<td>77790 (7299)</td>
<td>40521 (3886)</td>
<td>51967 (5060)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>32194 (3364)</td>
<td>38992 (3674)</td>
<td>20330 (1968)</td>
<td>25996 (2533)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>2.0 (2.0)</td>
<td>2.0 (2.0)</td>
<td>2.0 (2.0)</td>
<td>2.0 (2.0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>93.52 (99.61)</td>
<td>98.16 (93.61)</td>
<td>99.60 (98.35)</td>
<td>99.38 (98.37)</td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>19.50 (1.86)</td>
<td>21.46 (5.52)</td>
<td>8.52 (1.68)</td>
<td>14.19 (1.67)</td>
</tr>
<tr>
<td>Wilson B-factor</td>
<td>20.60</td>
<td>17.03</td>
<td>21.74</td>
<td>20.13</td>
</tr>
<tr>
<td>R-merge</td>
<td>0.01152 (0.3537)</td>
<td>0.008496 (0.08949)</td>
<td>0.03846 (0.4464)</td>
<td>0.02333 (0.4409)</td>
</tr>
<tr>
<td>R-meas</td>
<td>0.0163 (0.5002)</td>
<td>0.01201 (0.1266)</td>
<td>0.05439 (0.6313)</td>
<td>0.03299 (0.6235)</td>
</tr>
<tr>
<td>R-pim</td>
<td>0.01152 (0.3537)</td>
<td>0.008496 (0.08949)</td>
<td>0.03846 (0.4464)</td>
<td>0.02333 (0.4409)</td>
</tr>
<tr>
<td>CC1/2</td>
<td>1 (0.854)</td>
<td>1 (0.982)</td>
<td>0.999 (0.703)</td>
<td>1 (0.703)</td>
</tr>
<tr>
<td>CC*</td>
<td>1 (0.96)</td>
<td>1 (0.995)</td>
<td>1 (0.909)</td>
<td>1 (0.909)</td>
</tr>
<tr>
<td>Reflections used in refinement</td>
<td>32167 (3358)</td>
<td>38991 (3674)</td>
<td>20325 (1968)</td>
<td>25986 (2533)</td>
</tr>
<tr>
<td>Reflections used for R-free</td>
<td>1610 (161)</td>
<td>1562 (147)</td>
<td>1998 (194)</td>
<td>1998 (195)</td>
</tr>
<tr>
<td>R-work</td>
<td>0.1876 (0.3330)</td>
<td>0.1503 (0.1999)</td>
<td>0.1734 (0.3029)</td>
<td>0.1789 (0.3012)</td>
</tr>
<tr>
<td>R-free</td>
<td>0.2116 (0.3565)</td>
<td>0.1757 (0.2449)</td>
<td>0.2105 (0.3464)</td>
<td>0.2076 (0.3597)</td>
</tr>
<tr>
<td>CC(work)</td>
<td>0.959 (0.863)</td>
<td>0.964 (0.931)</td>
<td>0.968 (0.855)</td>
<td>0.962 (0.817)</td>
</tr>
<tr>
<td>CC(free)</td>
<td>0.960 (0.861)</td>
<td>0.965 (0.905)</td>
<td>0.965 (0.830)</td>
<td>0.960 (0.776)</td>
</tr>
<tr>
<td>Number of non-hydrogen atoms</td>
<td>2153</td>
<td>2316</td>
<td>2095</td>
<td>2073</td>
</tr>
<tr>
<td>macromolecules</td>
<td>1915</td>
<td>1981</td>
<td>1887</td>
<td>1854</td>
</tr>
<tr>
<td>ligands</td>
<td>21</td>
<td>26</td>
<td>46</td>
<td>52</td>
</tr>
<tr>
<td>solvent</td>
<td>217</td>
<td>309</td>
<td>162</td>
<td>167</td>
</tr>
<tr>
<td>Protein residues</td>
<td>240</td>
<td>241</td>
<td>232</td>
<td>229</td>
</tr>
<tr>
<td>RMS(bonds)</td>
<td>0.008</td>
<td>0.021</td>
<td>0.007</td>
<td>0.008</td>
</tr>
<tr>
<td>RMS(angles)</td>
<td>1.18</td>
<td>1.90</td>
<td>1.21</td>
<td>1.25</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
<td>97.87</td>
<td>98.31</td>
<td>96.92</td>
<td>98.21</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
<td>2.13</td>
<td>1.69</td>
<td>2.64</td>
<td>1.34</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.44</td>
<td>0.45</td>
</tr>
<tr>
<td>Rotamer outliers (%)</td>
<td>1.42</td>
<td>0.92</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Clashscore</td>
<td>3.92</td>
<td>3.80</td>
<td>1.84</td>
<td>3.22</td>
</tr>
<tr>
<td>Average B-factor</td>
<td>32.57</td>
<td>24.04</td>
<td>28.82</td>
<td>25.87</td>
</tr>
<tr>
<td>macromolecules</td>
<td>32.10</td>
<td>22.60</td>
<td>27.88</td>
<td>25.16</td>
</tr>
<tr>
<td>ligands</td>
<td>23.33</td>
<td>19.92</td>
<td>42.28</td>
<td>20.42</td>
</tr>
<tr>
<td>solvent</td>
<td>37.65</td>
<td>33.61</td>
<td>36.00</td>
<td>35.48</td>
</tr>
<tr>
<td>Number of TLS groups</td>
<td>9</td>
<td>12</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>
Section S2  Minor populations in the irradiated crystal structures

In the main text we discussed the trans structural models in the context of the irradiated crystals. For the contracted irradiated crystal, the electron density is well explained by a 50-50 mixture of cis and trans chromophores. For the expanded irradiated crystal, when we model the electron density with just a trans structural model, the difference density map \((mF_{obs} - DF_{calc})\) displays small peaks suggestive of other minor populations of the chromophore. These peaks likely show the position of the chlorine atom in minor populations because Cl is electron-rich and scatters x-rays most strongly.

One peak of the difference map is found at the position of the chlorine atom in the cis chromophore. This peak, along with another peak on the cis state rotamer of Tyr146, suggest that a fraction of the monomers in the crystal lattice remain in the cis state and were not photoswitched. (The residual cis population was also detected in a previous serial femtosecond crystallography study,\(^{13}\) where the irradiated crystal of wild-type rsEGFP2 was modeled as a mixture of 90% trans and 10% cis states.) Note that the rest of the cis chromophore (the non-chlorine atoms) is undetectable in the difference density map.

Another peak of the difference map is found near carbons 4 and 5 (Figure 2) of the trans structural model. This peak could be explained by the chlorine atom of a trans anti chromophore, suggesting that a small fraction of the Cl-rsEGFP2 molecules in the expanded irradiated crystal may undergo a hula-twist and end up in the trans anti state. Oddly, this peak is located over the C4-C5 bond, rather than a C-Cl bond length’s distance away from C5, so the ring geometry of this putative trans anti minor conformer would have to be slightly distorted compared to the predominant trans syn conformer. Since the non-chlorine atoms of the chromophore are undetectable in the difference density map, it is not possible to build a reliable atomic model of this minor conformer.
Discussion of the timescale of ring-flipping between anti and syn conformers

In reaching our conclusion on the photoisomerization pathway, we assumed that the trans state structure observed in the cryocooled crystal reflects the configuration of the chromophore immediately after photoisomerization. That is, we assumed that the chlorine substituent is locked in a particular orientation (either syn or anti). The trans syn and trans anti conformers could interconvert upon a 180-degree torsion of the \(\varphi\)-bond, which we will call a ring-flip. For the assumption to hold, the kinetic barrier to ring-flipping must be sufficiently high to prevent any interconversion during the experiment. In our procedure, once the crystal is photoswitched, it spends a few seconds at room temperature, and then it is stored for weeks to months at liquid nitrogen temperatures. We assume that the ring-flip occurs too slowly at cryogenic temperatures to be relevant. Thus, if ring-flipping occurs at all in our experiment, it must happen during the few seconds between photoisomerization and cryocooling. It is important to note that the observed fractional occupancies of trans syn and trans anti in the expanded and contracted lattices, respectively, are predominantly one conformer rather than a mixture. If the two conformers were freely exchanging, this observation could only be explained by an energetic bias of many kT for one conformer over the other between the two lattices, which seems unlikely.

In the following we consider whether the \(\varphi\)-bond rotation barrier is high enough to prevent thermal ring-flipping on the second timescale at room temperature. There are few experiments (short of full serial crystallography) which can probe this process directly, so we present a qualitative argument based on what is known about the rate of related processes. The relevant chemical process is a 180-degree ring-flip of the 3-chlorophenolate moiety about the \(\varphi\)-bond, at room temperature, in the confined environment of a protein cavity. A similar process is the ring-flipping of buried tyrosine residues in the interiors of proteins. NMR studies of chemical exchange have shown that this process occurs with typical rates of \(\sim 10-1000 \text{ s}^{-1}\), \(14-16\) and compared to the Cl-rsEGFP2 chromophore ring-flip, the ring-flip of a buried tyrosine involves a similar steric size and occurs in a similar cavity environment. The key difference for the Cl-rsEGFP2 chromophore is that the rotating \(\varphi\)-bond lies in a conjugated system and has significant double bond character. This effect hinders rotation about the \(\varphi\)-bond and hence is expected to slow down the ring-flipping process compared to buried tyrosines.

The partial double bond character of the \(\varphi\)-bond can be estimated using a resonance picture. The Cl-rsEGFP2 chromophore has an additional resonance form where the locations of the single and double bonds are swapped; in this case the \(\varphi\)-bond is a double bond and the \(\tau\)-bond is a single-bond. The amount of double bond character of the \(\varphi\)-bond is determined by the size of the contribution from the secondary resonance form. If there were no contribution, then the \(\varphi\)-bond and the \(\tau\)-bond would be “pure” C-C single and double bonds, respectively, and their bond lengths would differ by roughly 0.2 Å; \(17\) if there were equal weightings of the two resonance forms (as in the allyl anion), then the \(\varphi\)-bond and the \(\tau\)-bond would be roughly...
equal in length. The measured difference in bond lengths between the $\varphi$-bond and the $\tau$-bond is 0.10 Å in small-molecule crystallography of the model GFP chromophore and 0.06 Å in protein crystallography of GFP at subatomic resolution. (Both these bond lengths are of the protonated cis state which is more experimentally accessible; while high resolution structural data is not available for the protonated trans state because of its transient nature, it is reasonable to expect similar bond length differences in the trans state.) These crystallographic bond lengths indicate that the $\varphi$-bond has double bond character.

Since double bonds rotate less freely than single bonds, the partial double bond nature of the $\varphi$-bond is expected to slow down ring-flipping in the Cl-rsEGFP2 compared to the analogous process in typical buried tyrosine residues. One way to estimate the rate of $\varphi$-bond rotation is to start with the ~10-1000 s$^{-1}$ of buried tyrosines and then account for how much the double bond character increases the rotation barrier. Each kcal/mol increase in barrier height causes the rate to decrease by a factor of $\exp(1/RT) = 5.4$ at room temperature. A typical C=C double bond has a rotation barrier on the order of 50-65 kcal/mol, but since the $\varphi$-bond is only partially a double bond, its barrier is likely only a fraction of this. Even a modest increase of 5 kcal/mol in the barrier height would slow down the rate 5000-fold to a range of ~0.002-0.2 s$^{-1}$, which would lock the chlorine substituent in place for the duration of our experiment. While this analysis suggests that ring-flipping is likely not occurring faster than the rate of freezing following irradiation, a definitive result could be obtained using time-resolved crystallographic approaches.
Section S4  Supplementary Figures

Figure S1. Cl-rsEGFP2 maintains the photoswitching property of wild-type rsEGFP2. Absorption spectra of (A) wild-type rsEGFP2 and (B) Cl-rsEGFP2 upon 488 nm laser irradiation. The emerging peaks around 410 nm correspond to the trans states.
Figure S2. Comparison of Cl-rsEGFP2 to wild-type rsEGFP2 cis structure. (A) The structure of cis Cl-rsEGFP2 (light blue) in the expanded unit cell overlays onto the previously reported wild-type cis rsEGFP2 structure (gray; PDB ID: 5DTX) with an α-carbon root-mean-square-deviation of 0.17 Å. The wild-type cis rsEGFP2 has lattice constants $a = 51.0$ Å, $b = 62.9$ Å, $c = 70.8$ Å, and the expanded cis Cl-rsEGFP2 has lattice constants $a = 50.7$ Å, $b = 62.5$ Å, $c = 67.6$ Å. The proteins are shown in cartoon representation. (B) Inset of chromophore pocket with atoms shown in stick representation.
Figure S3. Comparison of Cl-rsEGFP2 to wild-type rsEGFP2 trans structure. (A) The structure of trans Cl-rsEGFP2 (blue) in the expanded unit cell overlays onto the previously reported wild-type trans rsEGFP2 structure (gray; PDB ID: 5DTY) with an α-carbon root-mean-square-deviation of 0.11 Å. The wild-type trans rsEGFP2 has lattice constants $a = 51.2$ Å, $b = 62.9$ Å, $c = 70.6$ Å, and the expanded trans Cl-rsEGFP2 has lattice constants $a = 51.0$ Å, $b = 62.7$ Å, $c = 68.6$ Å. The proteins are shown in cartoon representation. (B) Inset of the chromophore pocket with atoms shown in stick representation.
Figure S4. Comparison between expanded and contracted trans crystals. CI-rsEGFP2 maintains the same overall β-barrel fold in the expanded (above) and contracted (below) crystals, but the packing interactions between neighboring units affect the structure of tails, loops, and certain outward-facing residues. (A) A CI-rsEGFP2 monomer is shown in the orthorhombic unit cell with lattice constants labeled (Supp. Table 2). The N-terminal and C-terminal tails are circled in purple and red, respectively. (B) In the contracted unit cell (below), the N-terminal helix (circled in purple) is pressed closer to the β-barrel. The tighter packing also distorts the loop following strand 4 (loop s4) and forces Tyr183 to adopt a different rotamer. (C) The C-terminal loop (circled in red) is resolved in the expanded crystal (above) but unresolved in the contracted crystal (below). The protein and its neighboring crystallographic monomers are shown in cartoon representation, and the chromophore, sulfates, and Tyr183 are shown in stick representation.
Figure S5. Effect of unit cell contraction on the trans Cl-rsEGFP2 chromophore pocket. Shown above are the chromophore pockets of the expanded (blue) and contracted (green) trans Cl-rsEGFP2 crystal structures, with water molecules shown as spheres and hydrogen bonds (≤3.1Å) shown as dotted lines. A 7% reduction in unit cell size (Supp. Table 2) results in the internal structural changes shown above.
Section S5 References


(9) A Quick XDS Tutorial for SSRL


(14) Wagner, G. Characterization of the distribution of internal motions in the basic pancreatic trypsin inhibitor using a large number of internal NMR probes. Q. Rev. Biophys. 1983, 16 (1), 1–57.


