The FEBS Journal

FEBS PRESS science publishing by scientists

Crystal structure of a domain-swapped photoactivatable sfGFP variant provides evidence for GFP folding pathway

Stephanie Kesgin-Schaefer[1,2], Johannes Heidemann[3], Anke Puchert[1,4], Knut Koelbel[3], Briony A. Yorke[1,4], Nils Huse[1,4], Arwen R. Pearson[1,4], Charlotte Uetrecht[3,5] and Henning Tidow[1,2]

1 The Hamburg Centre for Ultrafast Imaging, Germany
2 Department of Chemistry, Institute for Biochemistry and Molecular Biology, University of Hamburg, Germany
3 Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany
4 Department of Physics, Center for Free-Electron Laser Science, Institute for Nanostructure and Solid State Physics, University of Hamburg, Germany
5 European XFEL GmbH, Schenefeld, Germany

Keywords
crystal structure; domain-swap; GFP; *ortho*-nitrobenzyl-tyrosine; unnatural amino acid

Correspondence
H. Tidow, Department of Chemistry, Institute for Biochemistry and Molecular Biology, University of Hamburg, Martin-Luther-King-Platz 6, Hamburg D-20146, Germany
Tel: +49 40428388984
E-mail: tidow@chemie.uni-hamburg.de

(Received 7 September 2018, revised 7 February 2019, accepted 27 February 2019)

doi:10.1111/febs.14797

Photoactivatable fluorescent proteins (PA-FPs) are a powerful non-invasive tool in high-resolution live-cell imaging. They can be converted from an inactive to an active form by light, enabling the spatial and temporal trafficking of proteins and cell dynamics. PA-FPs have been previously generated by mutating selected residues in the chromophore or in its close proximity. A new strategy to generate PA-FPs is the genetic incorporation of unnatural amino acids (UAAs) containing photocaged groups using unique suppressor tRNA/aminoacyl-tRNA synthetase pairs. We set out to develop a photoactivatable GFP variant suitable for time-resolved structural studies. Here, we report the crystal structure of superfolder GFP (sfGFP) containing the UAA *ortho*-nitrobenzyl-tyrosine (ONBY) at position 66 and its spectroscopic characterization. Surprisingly, the crystal structure (to 2.7 Å resolution) reveals a dimeric domain-swapped arrangement of sfGFP66ONBY with residues 1–142 of one molecule associating with residues 148–234 from another molecule. This unusual domain-swapped structure supports a previously postulated GFP folding pathway that proceeds via an equilibrium intermediate.

Introduction

Photoactivatable fluorescent proteins (PA-FPs) are molecular probes which can be controlled by light. Their spectral characteristics can be controlled and modified by irradiating them with light of a specific wavelength, intensity and duration, leading to a fluorescent state [1]. Inactivated PA-FPs are non-fluorescent, whereas activated PA-FPs yield bright signals in contrast to a dark background. These properties allow the spatial and temporal labelling of specific subcellular structures and thus enable the visualization of protein, organelle and cell dynamics. This makes PA-FPs a powerful non-invasive tool for high-resolution live-cell imaging [2,3]. Various PA-FPs have been engineered, especially proteins of the GFP family, with altered properties such as shifted fluorescence spectra, enhanced fluorescence and higher quantum yields [2,4–8].

GFP, discovered in the jellyfish *Aequorea victoria*, exhibits bright green fluorescence when exposed to

Abbreviations
aaRS, aminoacyl-tRNA synthetase; EQL, equilibrium; FWHM, full width half maximum; MS, mass spectrometry; NMS, native mass spectrometry; ONBY, *ortho*-nitrobenzyl-tyrosine; PA-FP, photoactivatable fluorescent proteins; SEC, size-exclusion chromatography; sfGFP, superfolder GFP; UAA, unnatural amino acid.

light in the blue to ultraviolet range [9]. The 11 β-strands form a cylindrical β-barrel with the α-helix harbouring the chromophore spanning straight through the barrel centre [10,11]. Due to its ability to auto-catalytically form its highly visible, efficiently emitting internal chromophore [12–14], GFP has become a well-established and powerful tool for many applications in cell and molecular biology [15].

Time-resolved (TR) structural biology is the key to understand the mechanisms of biological processes [16]. In order to perform TR experiments – either time-resolved X-ray diffraction (TR-XRD) or time-resolved small-angle X-ray scattering (TR-SAXS) – an ensemble of molecules first needs to be populated/trapped in a defined state. Conformational changes can then be triggered by means of rapid mixing or light irradiation [16–19]. The use of light in principle allows observations in the ps-regime if the decaging reaction is fast enough.

The incorporation of unnatural amino acids (UAAs) into the chromophore of GFP has previously been accomplished by introducing boronate, azido, keto, and nitro substituents [20], phenyl azide [21] and azobenzene [22], thus establishing photosensitive GFP variants. In order to develop and apply a GFP-based system that can be used for simultaneous TR structural and spectroscopic studies, we further investigated a photoactivatable GFP variant, originally designed by Schultz and coworkers [23], that contained the photocaged tyrosine analogue *ortho*-nitrobenzyl-tyrosine (ONBY) [24,25]. Photocaged UAAs are synthetically modified amino acids which can be controlled by light, usually by photolytic conversion from an inactive to an active form [26]. The most common photocaging groups are *ortho*-nitrobenzyl-groups and their derivatives, which can be readily cleaved upon irradiation with 355 nm light [27]. The incorporation of photocaged ONBY in *Escherichia coli* is achieved by genetically encoding it using an amber stop codon (TAG) in combination with an orthogonal suppressor tRNA/aminoacyl-tRNA synthetase (aaRS) pair [24]. Incorporation of ONBY at residue 66, a key residue of the chromophore, leads to a non-fluorescent ('dark-state') GFP molecule. Upon irradiation with light (355 nm), the photocaging *ortho*-nitrobenzyl-group dissociates and the fluorescent state is re-established. However, the structure of this photocaged GFP variant, the decaging mechanism, as well as its suitability for TR structural studies are still not understood.

In this study, we determined the crystal structure of the dark-state sfGFP66ONBY variant. The structure reveals an unusual and unexpected dimeric domain-swapped arrangement, which in combination with spectroscopic and native mass spectrometry (NMS) experiments provides experimental evidence for a GFP folding pathway.

Results and Discussion

Spectroscopic and hydrodynamic properties of sfGFP66ONBY

In order to investigate the suitability of GFP variants for TR structural and spectroscopic studies, we expressed and purified a superfolder GFP (sfGFP) variant, which had the genetically encoded UAA ONBY incorporated at position 66, replacing the central Tyr residue of the chromophore. The protein predominantly adopts a monomeric state as judged from size-exclusion chromatography (SEC; Fig. 1A). It also migrates as monomer on an SDS/PAGE gel when samples were fully denatured (Fig. 1B).

The wt-GFP chromophore is formed by the auto-catalytic cyclization of the Ser/Thr65-Tyr66-Gly67 tripeptide [10,11], which is associated to an intramolecular hydrogen-bonding network linking the two ends of the chromophore in the protein pocket. Upon excitation with light of wavelength 355 nm, an excited-state proton transfer is induced which enables structural transitions and involves multiple proton transfers through the hydrogen-bonding network resulting in fluorescence [28,29]. In contrast to wt-GFP, the sfGFP66ONBY variant protein is non-fluorescent. The bulky *ortho*-nitrobenzyl-group most likely blocks the proton-wire, resulting in a non-functional chromophore [30]. Moreover, it has been reported that *ortho*-nitrobenzyl-groups quench excited fluorophores through a photo-induced electron transfer from the S1 excited chromophore to the *ortho*-nitrobenzyl-group upon irradiation [31,32].

The decaging characteristics of sfGFP66ONBY were investigated by absorption spectroscopy. Upon photolysis with 355 nm pulsed light, the UV/Vis absorption spectrum exhibits a loss of the absorption peak at 390 nm for the caged protein, while a new absorption peak at 488 nm appears for the photolysed protein (Fig. 2A,B). Photolysed sfGFP66ONBY shows an absorption spectrum similar to wild-type sfGFP (wt-sfGFP). Additionally, fluorescence spectra of photolysed sfGFP66ONBY and wt-sfGFP using 355 nm excitation are similar with fluorescence maxima at 509 and 517 nm, respectively (Fig. 2C). A pronounced spectral shoulder at 541 nm is also present in both fluorescence spectra. The similarity of absorption and fluorescence spectra of photolysed sfGFP66ONBY and wt-sfGFP confirm previous measurements and dissociation of the ONBY group from the mutant [20].

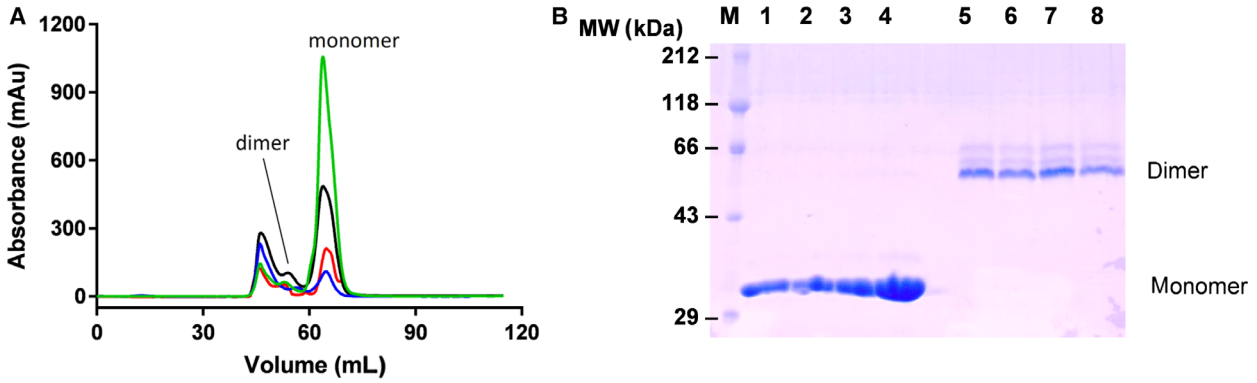


Fig. 1. SDS/PAGE and SEC analysis of sfGFP variants. (A) wt-sfGFP (green), sfGFP66ONBY (blue), sfGFP66ONBY-F145A (red) and decaged sfGFP66ONBY (black) predominantly elute as monomers from SEC. (B) Without heat treatment wt-sfGFP, sfGFP66ONBY, sfGFP66ONBY-F145A and decaged sfGFP66ONBY migrate as dimers (54 kDa) in the SDS/PAGE gel (5–8). In contrast, after heat treatment (1–4) proteins migrate as monomer (27 kDa) (M = molecular weight marker, 1 + 5 = wt-sfGFP, 2 + 6 = sfGFP66ONBY, 3 + 7 = sfGFP66ONBY-F145A, 4 + 8 = decaged sfGFP66ONBY).

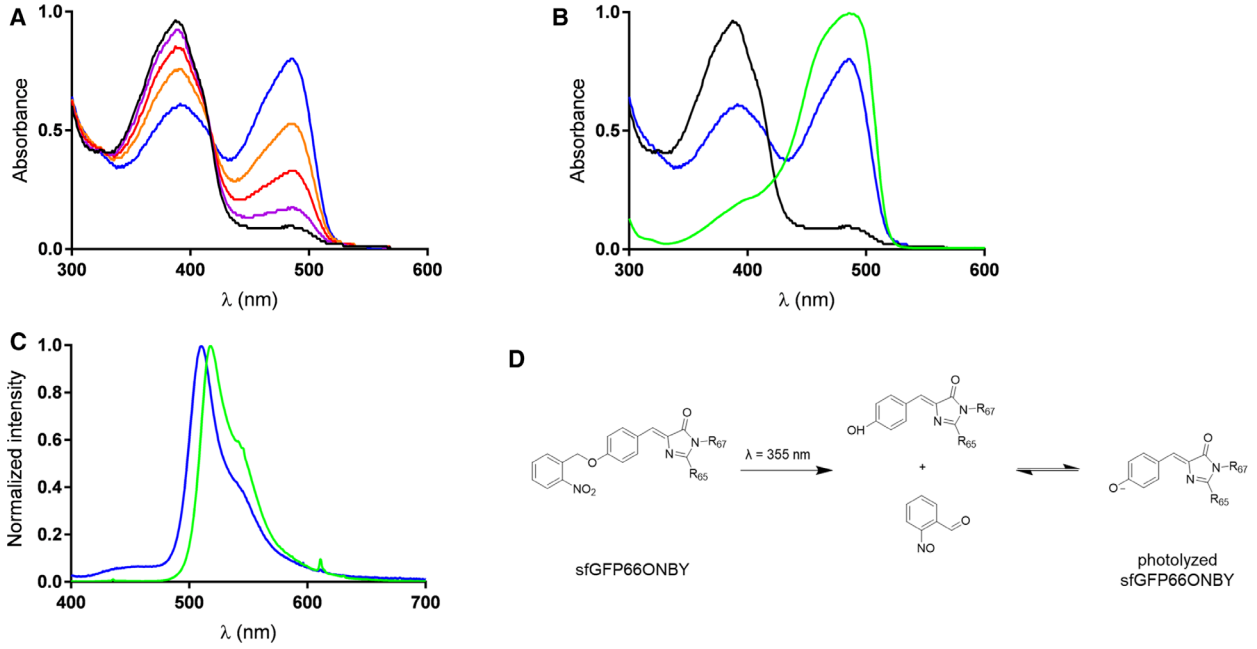


Fig. 2. Spectroscopic characterization of sfGFP66ONBY. (A) Evolution of UV/Vis absorption spectra of sfGFP66ONBY during 355-nm pulsed irradiation. Samples were irradiated for 0 min (black), 1 min (purple), 4 min (red), 9 min (orange), 30 min (blue). (B) UV/Vis absorption spectra of sfGFP66ONBY before (black) and after photolysis (blue) and wt-sfGFP (green). Spectrum of sfGFP (green) was rescaled by a factor of 1.8 for better comparison. (C) Fluorescence spectra of photolysed sfGFP66ONBY (blue) and wt-sfGFP (green) after pulsed excitation at 355 nm. For (B) and (C) samples were irradiated with 355-nm pulses at 1 kHz and 35 mW average power for 30 min. (D) Photochemical reaction of the decaging of the non-fluorescent sfGFP66ONBY by UV photolysis.

Crystal structure of dark-state sfGFP66ONBY shows (dimeric) domain-swapped arrangement

To visualize the chemical environment of the caged chromophore we crystallized dark-state (photocaged) sfGFP66ONBY and determined its structure to 2.7 Å resolution (Table 1). To our great surprise, the sfGFP66ONBY structure revealed a domain-swapped dimeric arrangement with residues 1–143 (domain 1)

of the first molecule associating with the residues 148–235 (domain 2) of the second molecule (Fig. 3A). Clear electron density identified the connecting loops ranging from residues Tyr143 to His148 (Fig. 3C), while no electron density was present in the position corresponding to β-strand β_7 in the search model (as previously observed by Groff *et al*. [23]). The electron density extends away from the search model, leading to the second half-subunit where the protein backbone is also shifted due to the steric interference of the bulky *ortho*-nitrobenzyl-group with the normal tight packing of the β_7-strand against the chromophore. The peptide side chains of the β_7-strand are reoriented and the bulky side chain of Phe145 is rotated by 180° to the outside of the barrel. These backbone rearrangements lead to the exchange of half of the structure of one subunit with the other half of the other subunit, as the backbone folds back on itself between residues 143 and 148 to form intertwined domain-swapped dimers (Fig. 3A).

The domain-swapped dimer structure of sfGFP66ONBY shows a significantly different dimer interface compared to the wt-sfGFP structure, which also crystallizes as a dimer (Fig. 4A). In the wt-sfGFP structure the dimer interface is formed by β-strands β_7 and β_{10}. In the domain-swapped sfGFP66ONBY structure the β_{10} strands are far apart and the only connection between the two barrels are the loops formed by residues 143–148 (Fig. 4A). While the loops (res. 143–148) connecting the swapped domains are clearly distinct from the wt-sfGFP structure, the overall structures of the individual barrel moieties are very similar with an all-atom RMSD of 1.27 Å (0.35 Å for Cα atoms). Both structures have an 11-strand β-barrel as well as a mature cofactor derived from cyclization of the Thr65 – Tyr66/ONBY66 – Gly67 tripeptide that differs only in the presence of the photocaged tyrosine (ONBY) in the sfGFP66ONBY structure, which is clearly visible in the electron density (Fig. 3B).

Table 1. Data collection and refinement statistics.

	sfGFP66ONBY (pdb: 6H01)
Data collection	
Space group	P 41212
Cell dimensions	
a, *b*, *c* (Å)	155.3, 155.3, 162.0
α, β, γ (°)	90, 90, 90
Resolution (Å)	90.93–2.70 (2.80–2.70)
R_{merge}	0.055 (0.609)
$<I>/\sigma I$	9.7 (1.4)
CC(1/2)	0.997 (0.546)
Completeness (%)	99.90 (99.87)
Multiplicity	2.0 (2.0)
Refinement	
Resolution (Å)	49.32–2.70 (2.80–2.70)
No. reflections	54 873 (5387)
R_{work}/R_{free}	0.201/0.239 (0.35/0.37)
No. atoms	
Protein	7273
Ligand	128
Water	133
Average *B*-factors (Å^2)	
Protein	47.14
Chromophore	44.28
Water	48.20
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.37

Values in parentheses are for the highest-resolution shell.

The phenomenon of domain-swapped oligomers has been described in several proteins [33,34], and in 3DSWAP, the curated knowledgebase of proteins involved in 3D domain swapping, nearly 300 depositions can be found [35]. Many domain-swapped dimers are only observed in crystals due to their low affinity resulting from relatively small differences in free energy of monomers and dimers [36,37]. Hence, clear evidence for their functional role *in vivo* is lacking. However, three possible *in vivo* functions have been suggested: regulation of protein function, a mechanism for protein misfolding and aggregation, and structural diversification during evolution [33]. The sfGFP66ONBY structure presented here constitutes a *bona fide* domain-swapped structure [38], an arrangement where the dimer adopts a domain-swapped conformation and the monomer adopts a closed conformation [35], as previously observed for diphtheria toxin or RNaseA [39,40].

Implications for the folding pathway of GFP

The surprising observation of a domain-swapped GFP structure prompted us to speculate about its assembly mechanism. In general, the native and the domain-swapped state are separated by a large kinetic barrier [41,42], while the mechanism by which a monomeric protein transforms into a domain-swapped dimer is highly dependent on the folding mechanism of the particular protein [33]. Due to the huge thermodynamic stability of folded GFP [43–45] it is very unlikely that the domain-swapped structure forms via unfolding of properly folded monomeric GFP, for example during crystallization. Therefore, we investigated whether we could infer anything about folding intermediates from this domain-swapped structure.

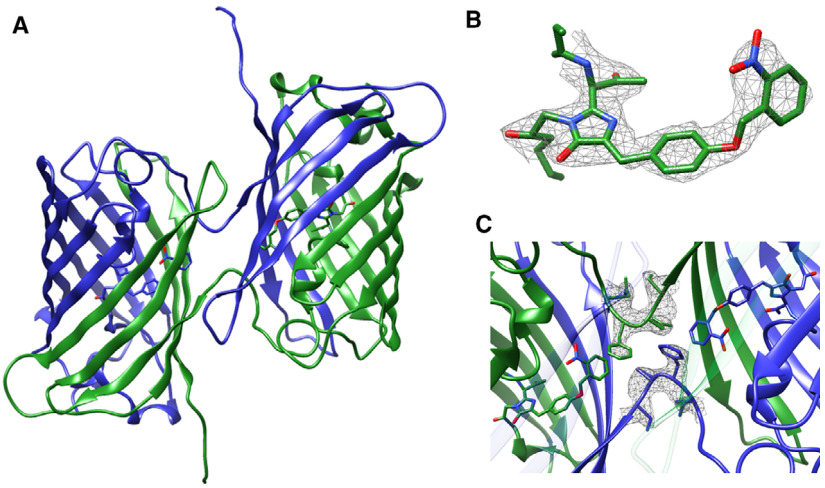


Fig. 3. Crystal structure of the domain-swapped sfGFP66ONBY. (A) Structure of domain-swapped sfGFP66ONBY dimer with each chain coloured in blue and green, respectively (pdb: 6H01). (B) sfGFP66ONBY was purified and crystallized to yield crystals that diffracted to 2.7 Å. The 2Fo-Fc map contoured at 1 r.m.s.d. clearly shows the presence of the *ortho*-nitrobenzyl group and the cyclized backbone forming the chromophore. (C) Omit map of the open interface and hinge loop region of the domain-swapped dimer contoured at 1 r.m.s.d showing the continuous electron density along the loop connecting the two domains of each chain. This open interface exists only in the domain-swapped dimer, but not in the wild-type monomer or dimer.

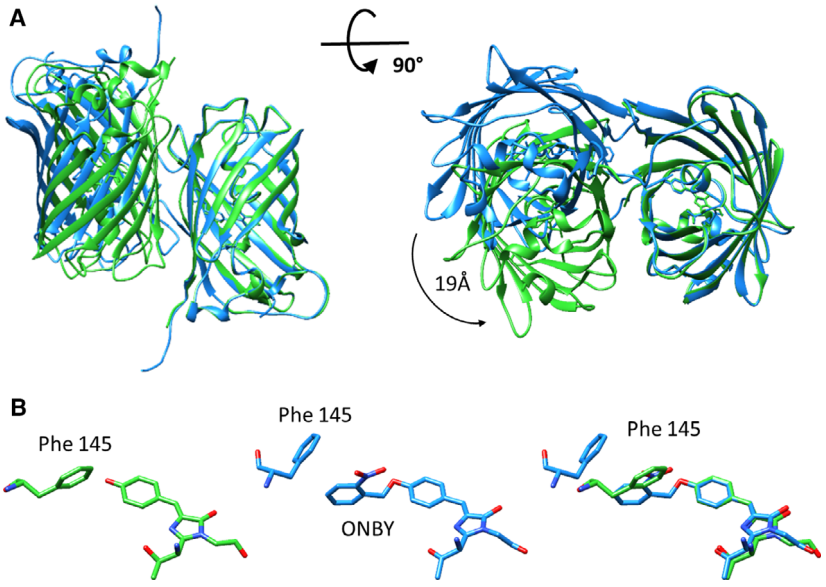


Fig. 4. Structural comparison of sfGFP66ONBY domain-swapped structure (cyan, pdb: 6H01) with wt-sfGFP dimer structure (green, pdb: 2B3Q). (A) One monomer of the dimer is superimposed. The different orientation of the second monomer in the domain-swapped structure and the corresponding lack of dimer interface are clearly visible. The second dimer is shifted by 19 Å. (B) Chromophores are shown as sticks. Tyr66 of the chromophore is superimposed. Comparison of the orientation of Phe145 in the sfGFP66ONBY structure (cyan) with Phe145 in the wt-sfGFP structure (green) indicates that a clear steric clash between Phe145 in its wt-sfGFP conformation and the *ortho*-nitrobenzyl-group of ONBY at position 66 would occur, thus causing the structural rearrangement.

The folding of GFP and its variants has been excessively investigated experimentally and theoretically [43,46–52]. Thirumalai and coworkers used molecular simulations to map the folding landscape of GFP and, consistently together with previous studies, suggest that GFP folds via equilibrium or kinetic intermediates along one of four possible pathways (EQL, KIN1, KIN2, KIN3) with the EQL pathway characterized by

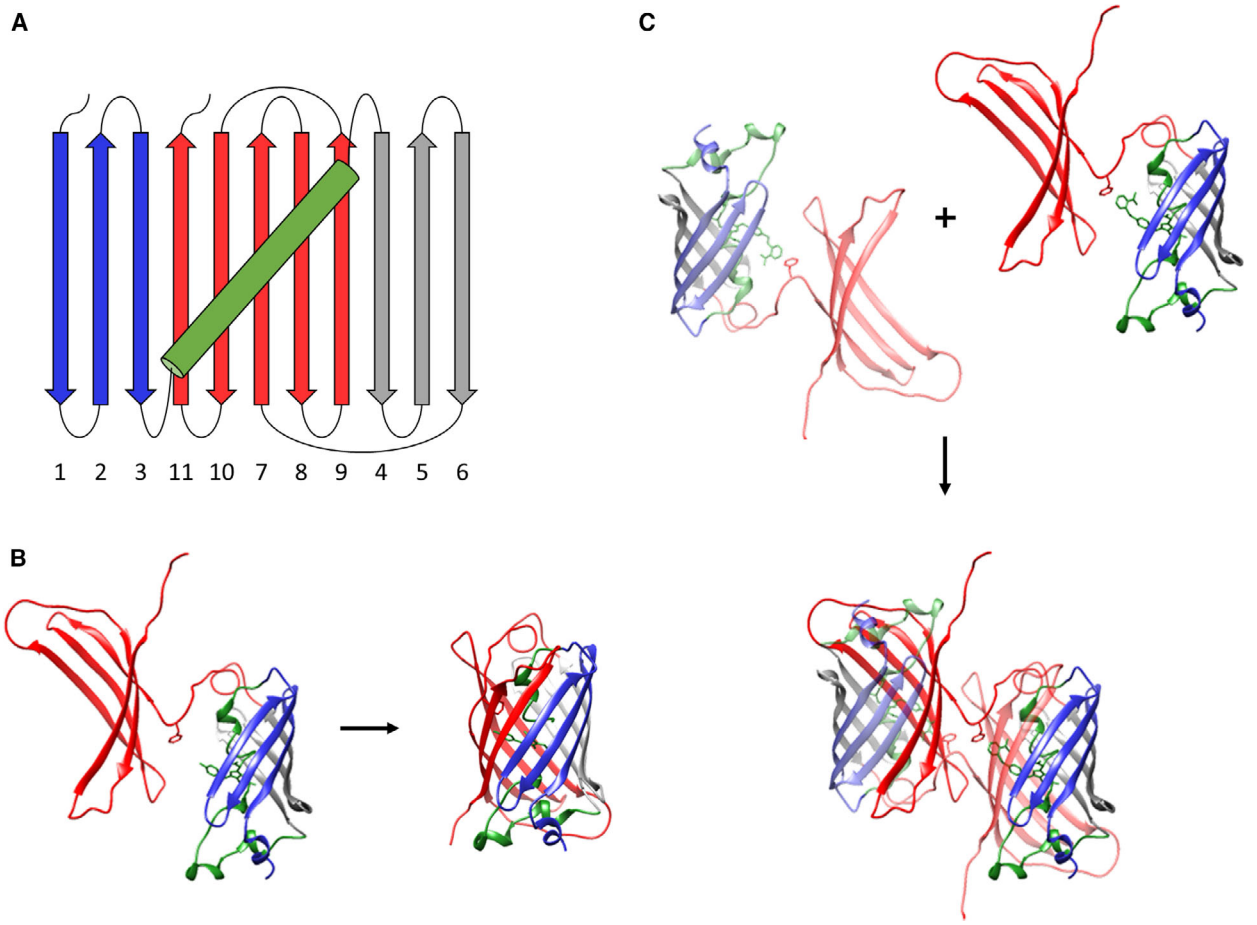


Fig. 5. Postulated folding mechanism of wt-GFP and its implications for the formation of domain-swapped sfGFP66ONBY (pdb: 6H01). (A) Topological depiction of wt-GFP. The N-terminal β-strands are represented in blue, the helix containing the chromophore in the centre of the β-strand barrel is in green, the central three β-strands are in grey, and the C-terminal β-strands are in red. (B) The postulated folding of wt-GFP in the EQL pathway occurs in a two-state manner via equilibrium intermediates. The EQL intermediate comprises two domains: a structured N-terminus (β_1–β_6) and structured C-terminus (β_7–β_{11} in red) connected by a loop. The EQL intermediate is able to fold into the correct monomeric species. (C) In the sfGFP66ONBY protein, steric clashes of Phe145 with the *ortho*-nitrobenzyl group interferes with folding and joining of C-terminal β_7–β_{11} (red) to the structured β_1–β_6 at the N-terminus and prevent folding of the correct monomer. Hence, the intermediate of the EQL pathway can only fold into the unusual domain-swapped sfGFP66ONBY dimer.

an equilibrium intermediate [53]. The sfGFP protein chain is built by eleven β-strands forming a cylindrical barrel, with one α-helix threading straight through the β-barrel. The β-strands can be divided into four folding units: the N-terminal β-strands (strand 1–3), the chromophore-containing helix at the centre of the β-strand barrel, the three β-strands in the centre (strands 4–6), which form local contacts, and the five C-terminal β-strands (strands 7–11) (Fig. 5). In the EQL folding pathway, the N-terminal β-sheets β_1–β_6 and the central helix fold and join to build half of the β-barrel structure. The C-terminal β-sheets β_7–β_{11} are unstructured and flexible and are not part of the β-barrel structure. Hence, the C-terminal β-sheets do not interact with the ordered N-terminal strands β_1–β_6 in this equilibrium intermediate state. Finally, the C-terminal β-sheets β_7–β_{11} fold and join with the rest of the structure folding into the β-barrel structure. This EQL folding pathway agrees well with experimental HDX/NMR data that fit a three-state model and indicate higher flexibility for β-strands β_7–β_{10} as well as stable intermediate states along the folding pathway of GFP. These folding intermediates retain considerable secondary and tertiary structure [47].

The domain-swapped sfGFP66ONBY structure presented here contains two domains, β_1–β_6 (res. 1–143) and β_7–β_{11} (res. 148–235), which makes it tempting to speculate that the domain-swapped structure is formed from an intermediate resembling the EQL intermediate postulated by Reddy *et al.* [53]. We identified Phe145

as a key residue preventing folding of sfGFP66ONBY into a monomeric structure due to its steric clashes with the bulky *ortho*-nitrobenzyl moiety of the ONBY chromophore if it were to adopt the conformation observed in wild-type and sfGFP. These steric clashes would prevent the folding and joining of β_7–β_{11} to the structured β_1–β_6 and inhibit folding into the correct monomeric form. Hence, it is not surprising that sfGFP66ONBY is able to access an alternate domain-swapped folded structure (Figs 4 and 5C) under certain conditions.

Therefore, we expressed and purified the sfGFP66ONBY-F145A point mutant and compared its oligomeric status to wt-sfGFP and sfGFP66ONBY. All investigated sfGFP variants migrate as monomers using denaturing SDS/PAGE analysis and predominantly elute as monomers from SEC [54] (Fig. 1). In order to further verify the oligomeric state of the sfGFP variants, we also employed NMS [55]. For all sfGFP variants (wt-sfGFP/sfGFP66ONBY before and after decaging/sfGFP66ONBY-F145A), monomeric and dimeric species were present at low μM concentration with the dimeric fraction representing the minor population (Fig. 6). For wt-sfGFP, we assigned the monomeric and dimeric species to the properly folded monomer and dimer structures as observed in the crystal structure (pdb: 2B3Q [54]). For sfGFP66ONBY, we assigned the dimer mass to the domain-swapped dimer, while the monomer mass could correspond to the partly unfolded intermediate (Fig. 5C). Indeed, we have observed that sfGFP66ONBY protein samples are less stable compared to sfGFP and prone to aggregation within days. This indicates that partly unfolded barrel structures exist for sfGFP66ONBY in solution, probably because of its inability to fold into a proper barrel structure due to steric clashes between strand β_7 and the ONBY moiety. We conclude that the samples used for spectroscopy and crystallography contained a mixture of monomeric and dimeric domain-swapped sfGFP66ONBY and that only decaging of the monomeric sfGFP66ONBY (folding intermediate) population is able to restore the wt-sfGFP phenotype and fluorescence, while the crystallization process may have forced the monomeric folding intermediate into a domain-swapped dimer structure. The sfGFP66ONBY-F145A variant showed the lowest dimer population in NMS analysis (Fig. 6) suggesting that this mutant indeed allows proper folding of sfGFP66ONBY and thus might be a suitable variant for future TR spectroscopic and structural studies.

In summary, the crystal structure of sfGFP66ONBY containing a genetically incorporated UAA as part of the chromophore revealed a novel and unexpected domain-swapped arrangement. It is likely that the domain-swapped form of sfGFP66ONBY derives from the same folding intermediate as that postulated for the EQL folding pathway.

Materials and methods

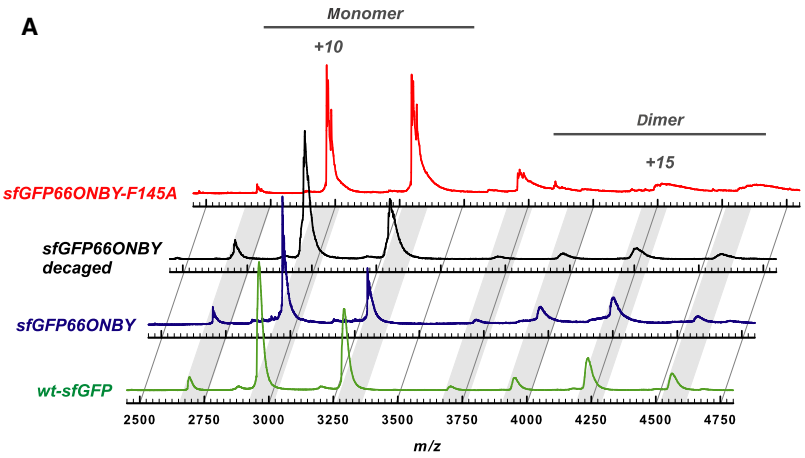
Ortho-nitrobenzyl-tyrosine synthesis

L-Tyrosine (2.0 g, 11.0 mmol) was dissolved in 2 M NaOH aq. (10 mL), and $CuSO_4 \cdot 5H_2O$ (1.9 g, 7.28 mmol), dissolved in a minimal amount of water, was added slowly at room temperature (RT). The solution was heated to 60 °C and stirred for 20 min and then allowed to cool to RT before adjusting to pH 7 using 1 M HCl. The light-blue solid was filtered and washed three times with 25 mL water, before it was suspended in 75% aqueous dimethylformamide (60 mL). K_2CO_3 (1.5 g, 11.04 mmol) and *ortho*-nitrobenzyl bromide (1.8 g, 8.49 mmol) were added and the reaction was allowed to proceed for 72 h at RT while kept in the dark. The solid was filtered, washed with 75% aqueous dimethylformamide (40 mL × 2), water (40 mL × 2), 75% aqueous acetone (40 mL) and ice-cold acetone (10 mL), and then suspended in 1 M HCl (100 mL) to stir for 2 h at RT. The white solid was filtered and stirred once more with fresh 1 M HCl (100 mL) for another 30 min. The solid was finally filtered, washed with water (40 mL × 2) and ice-cold acetone (10 mL) and dried to give *ortho*-nitrobenzyl-tyrosine as an off-white solid (1.85 g, 68%).

Cloning, protein expression and purification

To produce photoactivatable *ortho*-nitrobenzyl-photocaged sfGFP (sfGFP66ONBY), the stabilized GFP variant 'superfolder GFP' [54] was cloned into a pET28a vector with an N-terminal fusion consisting of a hexahistidine-tag and a TEV protease cleavage site. For incorporation of ONBY at residue position 66, tyrosine 66 was mutated to the amber codon, TAG, via site-directed mutagenesis, to generate pET-sfGFP66TAG-His6. This plasmid was cotransformed into *E. coli* BL21 (DE3) cells together with the plasmid pEVOL-ONBY, containing the orthogonal aaRS and an amber suppressor tRNA, both derived from the *Methanococcus jannaschi* tyrosyl tRNA/aaRS pair [24,56].

Cells were grown at 37 °C to an OD_{600} = 0.9 in Terrific Broth and shifted to 20 °C. ONBY was added to a final concentration of 1 mM, and protein production was induced by the addition of 1 mM IPTG with a delay of 60 min. After 20 h, the cells were harvested by centrifugation at 4000 ***g***, lysed by sonication and GFP66ONBY was purified using standard His_6-tag purification protocols followed by TEV protease digestion and a second Ni-affinity



B

		M_{theor}	M_{exp}	St. dev.	FWHM	Spectra
wt-sfGFP	Monomer	29 621	29 604	1.6	13	MS/MS
	Dimer	59 242	59 230	9	364	MS
wt-sfGFP66ONBY	Monomer	29 756	29 737	2.4	14	MS/MS
	Dimer	59 512	59 500	26	523	MS
wt-sfGFP66ONBY (decaged)	Monomer	29 756	29 766	33	280	MS
	Dimer	59 512	59 671	33	588	MS
wt-sfGFP66ONBY-F145A	Monomer	29 680	29 659	1.9	32	MS/MS
	Dimer	59 360	59 340	17	873	MS

Fig. 6. Native mass spectrometry analysis of sfGFP variants. (A) Native mass spectra indicate the presence of monomeric and dimeric species for wt-sfGFP (green), sfGFP66ONBY (blue), decaged sfGFP66ONBY (black), and sfGFP66ONBY-F145A (red). Main peaks are annotated with their charge states; positions of equal charge states are depicted in grey. (B) Masses from GFP monomers and dimers were determined from at least three MS or MS/MS measurements and listed with standard deviations and the average FWHM value as a measure of protein heterogeneity and desolvation (all in Da).

chromatography step to separate the His_6-tag. The protein was kept in 40 mM Tris, pH 8.0. Protein identity and purity were assessed using mass spectrometry (MS) and SDS/PAGE.

Size-exclusion chromatography and SDS/PAGE analysis

Size-exclusion chromatography was used to separate the monomer and dimer species from aggregates. For the SEC, a Superdex75 (GE Healthcare, Freiburg, Germany) column connected to a ÄKTA*pure* (GE Healthcare, Freiburg, Germany) was used at 4 °C and equilibrated with 40 mM Tris, pH 8.0. Loaded protein concentrations were 5–20 $mg{\cdot}mL^{-1}$. The protein was detected by measuring the UV absorbance at 280 nm. The resulting fractions of the purified protein were analysed by SDS/PAGE. Prior to being applied to the SDS/PAGE gel, protein samples were mixed with an appropriate amount of 5× sample buffer [50 mM Tris/HCl, pH 6.8, 10% (w/v) SDS, 50% (v/v) glycerol, 125 mM DTT, 0.1% (w/v) bromophenol blue]. Samples were directly loaded onto the gel or heated for 5 min to 96 °C.

Crystallization and data collection

Initial crystallization conditions were screened for using commercially available matrices from Hampton Research and Molecular Dimensions using the vapour-diffusion technique in sitting drops. Crystallization of sfGFP66ONBY was accomplished by the sitting-drop vapour-diffusion method. Sitting drops containing 2 μL of protein solution at 12 $mg{\cdot}mL^{-1}$ and 2 μL of buffer (0.1 M Na-HEPES, pH 7.5, 1.1 M tri-Na-citrate) were equilibrated at 20 °C against 1 mL of the same buffer in the reservoir. Small rod-like crystals formed after 8 weeks. Crystals were mounted and flash-cooled in liquid nitrogen. Diffraction data were collected at 100 K at the MASSIF1 beamline at ESRF, Grenoble, France, at a fixed energy of 12.835 keV (0.966 Å). Full datasets with an oscillation angle of 0.1° per image were collected to a resolution limit of 2.7 Å. All datasets were processed with XDS [57] and merged with AIMLESS [58]. GFP66ONBY crystals belonged to space group $P4_12_12$ with unit-cell dimensions $a = 155.3$, $b = 155.3$, $c = 162.0$, $\alpha = \beta = \gamma = 90°$. A summary of the data statistics is given in Table 1.

Structure determination, refinement and analysis

The structures were solved by molecular replacement with PHASER [59] using pdb: 1EMA [10] as search model. Building of the photocaged chromophore was achieved using ELBOW [60]. Subsequent rounds of manual building using COOT [61] and refinement using PHENIX.REFINE [62] allowed complete model building in space group $P\ 4_12_12$. The final model yielded crystallographic R factors of 0.20/0.24

(R_{work}/R_{free}) with 97% of residues falling within the Ramachandran favoured region and no outliers in disallowed regions, and a MOLPROBITY [63] clashscore of 4.34. Molecular graphics images were produced using the UCSF CHIMERA package [64]. The structural data have been deposited in the Protein Data Bank with accession code pdb: 6H01.

Spectroscopy

UV/Vis absorption spectra of sfGFP66ONBY during pulsed 355-nm irradiation were measured using a QE Pro spectrometer and a DH-2000 light source (both from Ocean Optics). A Q-switched diode-pumped Nd:YAG laser (Standa-Q1-TH), operating at 1 kHz and producing 35 mW average power of 355 nm light, was used to initiate photo-dissociation of the ONBY group. The laser was continuously pulsed with pulse durations of 750 ps and a beam size of 600 μm full width half maximum (FWHM). 1 mL protein solution (30 μM protein in 40 mM Tris, pH 8.0) was used in a rectangular quartz cuvette (4 mm × 10 mm lateral dimensions) without stirring with 355 nm excitation along the short cuvette axis and absorption/fluorescence measurements in orthogonal direction. The integration time of each absorption spectrum was 22 ms. Fluorescence spectra of photolysed sfGFP66ONBY were recorded with an integration time of 100 ms using an Ocean Optics QE Pro spectrometer, Ocean Optics, Inc., Largo, Florida, USA via a quartz lens and a UV grade optical fibre (400 μm diameter). Samples were photolysed for 30 min before fluorescence spectra were recorded. The Standa-Q1-TH laser was used in the operating mode described above for both, sfGFP66ONBY photolysis and fluorescence spectroscopy.

Native mass spectrometry

Proteins were exchanged into 50 mM ammonium acetate pH 8.0 using centrifugal filter units (Vivaspin 500, 10k MWCO; Sartorius, Göttingen, Germany). Samples were filled into ESI capillaries that were prepared as follows: Borosilicate capillaries (1.2 mm OD, 0.68 mm ID, with filament; World Precision Instruments, Sarasota, USA) were produced with a micropipette puller (P-1000; Sutter Instruments, Novato, USA), equipped with a squared box filament (2.5 × 2.5 mm; Sutter Instruments). Capillaries were gold-coated using a sputter coater (Q150R; Quorum Technologies Laughton, UK, 40 mA, 200 s, tooling factor 2.3, end bleed vacuum 8×10^{-2} mbar). Samples were analysed in positive ion mode on a QToF2 (Waters and MS Vision) modified for high mass experiments [65]. Data were acquired with 7 mbar source pressure and 1.5×10^{-2} mbar argon as collision gas, 1.3 kV capillary voltage and 100 V sample cone voltage. MS measurements were performed with 2 μM protein concentration and spectra were acquired with 30 V collision energy.

Acknowledgements

We thank K. Veith for technical assistance and members of the Tidow lab for helpful discussions. We are grateful to Matthew Bowler for support at the MASSIF1 beamline (ESRF) and the Radiation Damage BAG (mx1932) for beamtime, Henning Mootz for the ONBY synthesis protocol, and John Joseph Zaitsev-Doyle for help with the ONBY synthesis. We acknowledge access to the Sample Preparation and Characterization (SPC) Facility of EMBL, Hamburg. The pET-His6-GFP-TEV-LIC cloning vector (1GFP) was a gift from Scott Gradia (Addgene plasmid # 29663). We thank Peter Schultz (TSRI) for the pEVOL plasmid. The Heinrich Pette Institute, Leibniz Institute for Experimental Virology is supported by the Free and Hanseatic City of Hamburg and the Federal Ministry of Health. JH and CU acknowledge funding by BMBF VISAVIX 05K16BH1 and the Leibniz Association through SAW-2014-HPI-4 grants. CU and KK acknowledge funding through EU Horizon 2020 ERC StG-2017 759661. BY acknowledges funding by the Wellcome Trust (grant code: 110296/Z/15/Z). This work has been supported by the Cluster of Excellence 'The Hamburg Centre for Ultrafast Imaging' of the Deutsche Forschungsgemeinschaft (DFG) – EXC 1074 – project ID 194651731.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

SK-S and HT conceptualized the study; SK-S, JH, AP, BAY and KK were involved in the study design and data acquisition; SK-S and HT carried out the analysis and interpretation; SK-S and HT wrote the manuscript; HT acquired funding; NH, ARP, CU and HT supervised the experiments.

References

1 Lippincott-Schwartz J & Patterson GH (2009) Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging. *Trends Cell Biol* **19**, 555–565.

2 Lukyanov KA, Chudakov DM, Lukyanov S & Verkhusha VV (2005) Innovation: photoactivatable fluorescent proteins. *Nat Rev Mol Cell Biol* **6**, 885–891.

3 Fernandez-Suarez M & Ting AY (2008) Fluorescent probes for super-resolution imaging in living cells. *Nat Rev Mol Cell Biol* **9**, 929–943.

4 Patterson GH & Lippincott-Schwartz J (2002) A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* **297**, 1873–1877.
5 Ando R, Mizuno H & Miyawaki A (2004) Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* **306**, 1370–1373.
6 Wiedenmann J, Ivanchenko S, Oswald F, Schmitt F, Rocker C, Salih A, Spindler KD & Nienhaus GU (2004) EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proc Natl Acad Sci USA* **101**, 15905–15910.
7 Tsutsui H, Karasawa S, Shimizu H, Nukina N & Miyawaki A (2005) Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. *EMBO Rep* **6**, 233–238.
8 Ando R, Hama H, Yamamoto-Hino M, Mizuno H & Miyawaki A (2002) An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc Natl Acad Sci USA* **99**, 12651–12656.
9 Shimomura O, Johnson FH & Saiga Y (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J Cell Comp Physiol* **59**, 223–239.
10 Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY & Remington SJ (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395.
11 Yang F, Moss LG & Phillips GN Jr (1996) The molecular structure of green fluorescent protein. *Nat Biotechnol* **14**, 1246–1251.
12 Reid BG & Flynn GC (1997) Chromophore formation in green fluorescent protein. *Biochemistry* **36**, 6786–6791.
13 Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA & Tsien RY (1995) Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci* **20**, 448–455.
14 Heim R, Prasher DC & Tsien RY (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci USA* **91**, 12501–12504.
15 Tsien RY (1998) The green fluorescent protein. *Annu Rev Biochem* **67**, 509–544.
16 Levantino M, Yorke BA, Monteiro DC, Cammarata M & Pearson AR (2015) Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules. *Curr Opin Struct Biol* **35**, 41–48.
17 Kubelka J (2009) Time-resolved methods in biophysics. 9. Laser temperature-jump methods for investigating biomolecular dynamics. *Photochem Photobiol Sci* **8**, 499–512.
18 Lee HM, Larson DR & Lawrence DS (2009) Illuminating the chemistry of life: design, synthesis, and applications of "caged" and related photoresponsive compounds. *ACS Chem Biol* **4**, 409–427.
19 Trincao J, Hamilton ML, Christensen J & Pearson AR (2013) Dynamic structural science: recent developments in time-resolved spectroscopy and X-ray crystallography. *Biochem Soc Trans* **41**, 1260–1264.
20 Wang F, Niu W, Guo J & Schultz PG (2012) Unnatural amino acid mutagenesis of fluorescent proteins. *Angew Chem Int Ed Engl* **51**, 10132–10135.
21 Reddington SC, Rizkallah PJ, Watson PD, Pearson R, Tippmann EM & Jones DD (2013) Different photochemical events of a genetically encoded phenyl azide define and modulate GFP fluorescence. *Angew Chem Int Ed Engl* **52**, 5974–5977.
22 Padilla MS & Young DD (2015) Photosensitive GFP mutants containing an azobenzene unnatural amino acid. *Bioorg Med Chem Lett* **25**, 470–473.
23 Groff D, Wang F, Jockusch S, Turro NJ & Schultz PG (2010) A new strategy to photoactivate green fluorescent protein. *Angew Chem Int Ed Engl* **49**, 7677–7679.
24 Deiters A, Groff D, Ryu Y, Xie J & Schultz PG (2006) A genetically encoded photocaged tyrosine. *Angew Chem Int Ed Engl* **45**, 2728–2731.
25 Krafft GA, Sutton WR & Cummings RT (1988) Photoactivable fluorophores. 3. Synthesis and photoactivation of fluorogenic difunctionalized fluoresceins. *J Am Chem Soc* **110**, 301–303.
26 Adams SR & Tsien RY (1993) Controlling cell chemistry with caged compounds. *Annu Rev Physiol* **55**, 755–784.
27 Givens RS, Weber JF, Jung AH & Park CH (1998) New photoprotecting groups: desyl and p-hydroxyphenacyl phosphate and carboxylate esters. *Methods Enzymol* **291**, 1–29.
28 Brejc K, Sixma TK, Kitts PA, Kain SR, Tsien RY, Ormo M & Remington SJ (1997) Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc Natl Acad Sci USA* **94**, 2306–2311.
29 Chattoraj M, King BA, Bublitz GU & Boxer SG (1996) Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer. *Proc Natl Acad Sci USA* **93**, 8362–8367.
30 Fang C, Frontiera RR, Tran R & Mathies RA (2009) Mapping GFP structure evolution during proton transfer with femtosecond Raman spectroscopy. *Nature* **462**, 200–204.
31 Ueno T, Urano Y, Setsukinai K, Takakusa H, Kojima H, Kikuchi K, Ohkubo K, Fukuzumi S & Nagano T (2004) Rational principles for modulating fluorescence properties of fluorescein. *J Am Chem Soc* **126**, 14079–14085.
32 Kobayashi T, Urano Y, Kamiya M, Ueno T, Kojima H & Nagano T (2007) Highly activatable and rapidly

releasable caged fluorescein derivatives. *J Am Chem Soc* **129**, 6696–6697.
33 Rousseau F, Schymkowitz JW & Itzhaki LS (2003) The unfolding story of three-dimensional domain swapping. *Structure* **11**, 243–251.
34 Liu Y & Eisenberg D (2002) 3D domain swapping: as domains continue to swap. *Protein Sci* **11**, 1285–1299.
35 Shameer K, Shingate PN, Manjunath SC, Karthika M, Pugalenthi G & Sowdhamini R (2011) 3DSwap: curated knowledgebase of proteins involved in 3D domain swapping. *Database (Oxford)* **2011**, bar042.
36 Carey J, Lindman S, Bauer M & Linse S (2007) Protein reconstitution and three-dimensional domain swapping: benefits and constraints of covalency. *Protein Sci* **16**, 2317–2333.
37 Lee SJ, Cobb MH & Goldsmith EJ (2009) Crystal structure of domain-swapped STE20 OSR1 kinase domain. *Protein Sci* **18**, 304–313.
38 Bennett MJ, Schlunegger MP & Eisenberg D (1995) 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci* **4**, 2455–2468.
39 Bennett MJ, Choe S & Eisenberg D (1994) Domain swapping: entangling alliances between proteins. *Proc Natl Acad Sci USA* **91**, 3127–3131.
40 Liu Y, Gotte G, Libonati M & Eisenberg D (2002) Structures of the two 3D domain-swapped RNase A trimers. *Protein Sci* **11**, 371–380.
41 Rousseau F, Schymkowitz JW, Wilkinson HR & Itzhaki LS (2001) Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proc Natl Acad Sci USA* **98**, 5596–5601.
42 Schlunegger MP, Bennett MJ & Eisenberg D (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv Protein Chem* **50**, 61–122.
43 Jackson SE, Craggs TD & Huang JR (2006) Understanding the folding of GFP using biophysical techniques. *Expert Rev Proteomics* **3**, 545–559.
44 Penna TC, Ishii M, Junior AP & Cholewa O (2004) Thermal stability of recombinant green fluorescent protein (GFPuv) at various pH values. *Appl Biochem Biotechnol* **113–116**, 469–483.
45 Pitman DJ, Banerjee S, Macari SJ, Castaldi CA, Crone DE & Bystroff C (2015) Exploring the folding pathway of green fluorescent protein through disulfide engineering. *Protein Sci* **24**, 341–353.
46 Hsu ST, Blaser G & Jackson SE (2009) The folding, stability and conformational dynamics of beta-barrel fluorescent proteins. *Chem Soc Rev* **38**, 2951–2965.
47 Huang JR, Craggs TD, Christodoulou J & Jackson SE (2007) Stable intermediate states and high energy barriers in the unfolding of GFP. *J Mol Biol* **370**, 356–371.
48 Andrews BT, Schoenfish AR, Roy M, Waldo G & Jennings PA (2007) The rough energy landscape of superfolder GFP is linked to the chromophore. *J Mol Biol* **373**, 476–490.
49 Andrews BT, Gosavi S, Finke JM, Onuchic JN & Jennings PA (2008) The dual-basin landscape in GFP folding. *Proc Natl Acad Sci USA* **105**, 12283–12288.
50 Orte A, Craggs TD, White SS, Jackson SE & Klenerman D (2008) Evidence of an intermediate and parallel pathways in protein unfolding from single-molecule fluorescence. *J Am Chem Soc* **130**, 7898–7907.
51 Enoki S, Maki K, Inobe T, Takahashi K, Kamagata K, Oroguchi T, Nakatani H, Tomoyori K & Kuwajima K (2006) The equilibrium unfolding intermediate observed at pH 4 and its relationship with the kinetic folding intermediates in green fluorescent protein. *J Mol Biol* **361**, 969–982.
52 Dietz H & Rief M (2004) Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc Natl Acad Sci USA* **101**, 16192–16197.
53 Reddy G, Liu Z & Thirumalai D (2012) Denaturant-dependent folding of GFP. *Proc Natl Acad Sci USA* **109**, 17832–17838.
54 Pedelacq JD, Cabantous S, Tran T, Terwilliger TC & Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79–88.
55 Chandler SA & Benesch JL (2018) Mass spectrometry beyond the native state. *Curr Opin Chem Biol* **42**, 130–137.
56 Young TS, Ahmad I, Yin JA & Schultz PG (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J Mol Biol* **395**, 361–374.
57 Kabsch W (2010) Xds. *Acta Crystallogr D* **66**, 125–132.
58 Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D* **67**, 282–292.
59 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.
60 Moriarty NW, Grosse-Kunstleve RW & Adams PD (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D* **65**, 1074–1080.
61 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126–2132.
62 Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D* **68**, 352–367.

63 Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* **66**, 12–21.
64 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.
65 van den Heuvel RH, van Duijn E, Mazon H, Synowsky SA, Lorenzen K, Versluis C, Brouns SJ, Langridge D, van der Oost J, Hoyes J *et al.* (2006) Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* **78**, 7473–7483.