JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
Cite This: J. Am. Chem. Soc. 2018, 140, 375−381
pubs.acs.org/JACS

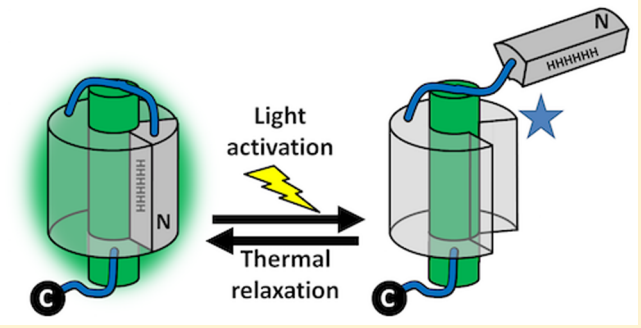
Structural Insight into the Photochemistry of Split Green Fluorescent Proteins: A Unique Role for a His-Tag

Alan Deng and Steven G. Boxer*

Department of Chemistry, Stanford University, Stanford, California 94305-5012, United States

Supporting Information

ABSTRACT: Oligohistidine affinity tags (His-tags) are commonly fused to proteins to aid in their purification via metal affinity chromatography. These His-tags are generally assumed to have minimal impact on the properties of the fusion protein, as they have no propensity to form ordered elements, and are small enough not to significantly affect the solubility or size. Here we report structures of two variants of truncated green fluorescent protein (GFP), i.e., split GFP with a β-strand removed, that were found to behave differently in the presence of light. In these structures, the N-terminal His-tag and several neighboring residues play a highly unusual structural and functional role in stabilizing the truncated GFP by substituting as a surrogate β-strand in the groove vacated by the native strand. This finding provides an explanation for the seemingly very different peptide binding and photodissociation properties of split proteins involving β-strands 10 and 11. We show that these truncated GFPs can bind other non-native sequences, and this promiscuity invites the possibility for rational design of sequences optimized for strand binding and photodissociation, both useful for optogenetic applications.



INTRODUCTION

Oligohistidine affinity tags (His-tags) are widely used as fusion tags to aid in the purification of proteins of interest via immobilized metal affinity chromatography.[1] Typically placed at the C- or N-terminal ends of these fusion proteins, His-tags are generally assumed to be functionally benign, as they have no propensity to form ordered elements, and are small enough not to significantly affect the solubility or size of the attached protein.[2] A PDB-wide survey[3] found over 16 000 structures with His-tags reported in their SEQRES records. Most His-tags are unstructured, while in about 6% of these structures, all or part of the His-tag is found in the final refinement model. The presence of the His-tag yielded no consistent impact on the structure of the fused protein. Here we report structures of two variants of truncated green fluorescent protein (GFP),[4,5] i.e., split GFPs with a β-strand removed, that were thought to behave distinctly in the presence of light. In these structures, the N-terminal His-tag and several adjacent residues play the highly unusual structural and functional role of stabilizing the truncated GFP by substituting as a surrogate β-strand in the groove vacated by the native strand.

In contrast to the more common method of forming split GFP, where the complementary pieces (typically fused to proteins whose interaction is of interest) need to assemble before chromophore maturation takes place, Kent and co-workers[6,7] generated an intact GFP with a mature chromophore, along with a sacrificial loop that can be cleaved, allowing for separation of the structural element of interest (e.g., a β-strand or the central helix). Utilizing this method, Kent and Boxer prepared and characterized a variant of Superfolder GFP[8,9] with its C-terminal 11th β-strand removed[10] (Figures 1A and 2A, hereafter the "strand 11 system"). Owing to the proximity of the N-terminus to the C-terminus, GFP can be circularly permuted[11] by genetic fusion of the native termini and formation of new termini at another point in the primary sequence. Do and Boxer prepared and characterized a circularly permuted variant of Superfolder GFP with its 10th β-strand removed[12,13] (Figures 1B and 2B, hereafter the "strand 10 system"). Topological representations for these proteins are given in Figure 1 along with a description of the general and compact notation used to describe this family of split GFPs.

Both the split strand 10 and 11 systems exhibit novel photochemical properties that appear to be very different, a surprising result given the overall similarity of their topology (Figure 1). Preparation of H6:GFP:~~loop:s11~~ (Figure 2A, see Figure 1 for nomenclature) by loop cleavage, denaturation, separation by size, refolding, and thermal relaxation results in a stable truncated protein which remains soluble up to low millimolar concentrations. This truncated protein has its own characteristic absorption spectrum distinct from that of intact H6:GFP:loop:s11,[10] and a relatively high fluorescence quantum yield (lower by a factor of 3 compared to intact H6:GFP:loop:s11). The isomerization state of the chromophore in this truncated protein was proposed to be the uncommon *trans* state (usually the more stable state is *cis*),

Received: October 6, 2017
Published: December 1, 2017

DOI: 10.1021/jacs.7b10680
J. Am. Chem. Soc. 2018, 140, 375−381

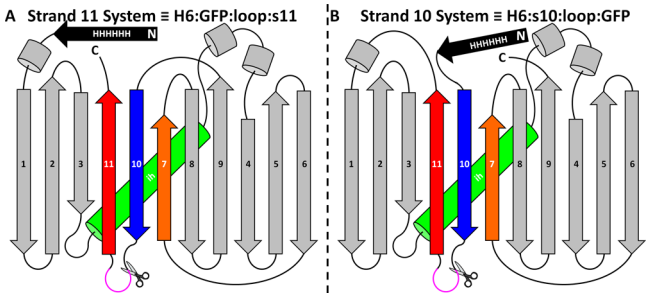


Figure 1. Strand topology of (A) the strand 11 system[10] and (B) the strand 10 system.[12,13] The colored strands 7 (orange), 10 (blue), and 11 (red) have amino acids that make important contacts with the phenol moiety of the chromophore, which is on the central helix (ih, green). The loop containing a proteolytic cleavage site for splitting the protein is shown in magenta (and the cut site shown with scissors). Note that the His-tag (black) is retained in (A), but removed along with strand 10 in (B) using the procedure shown in Figure 2; additional cleavage sites that selectively remove the His-tag are discussed in the text. The general compact notation is as follows: labels describe elements (delineated by colons) progressing from N-terminus to C-terminus when read from left to right; H6 refers to the N-terminal hexa-histidine affinity tag along with several adjacent residues that make up the thrombin recognition sequence; specific β-strands in the GFP β-barrel are denoted sX, where "X" is the number of the strand in question while "loop" refers to a loop (magenta) with proteolytic cleavage sites; "GFP" refers to the remainder of the protein; a strike through an element indicates that element has been removed; added synthetic elements are underlined; a dot (•) is used to indicate a noncovalent interaction; a superscript is used to highlight the numbered residue, and, if applicable, a subsequent letter indicates the mutation of said residue to the indicated amino acid. For example, $\underline{s11}^{222Q}$•H6:GFP:~~loop:s11~~ denotes a synthetic β-strand 11 peptide carrying the mutation at residue 222 (from E) to Q, noncovalently bound to GFP with an N-terminal His-tag, and with its original C-terminal strand 11 and loop removed. Full DNA and protein sequences are in Supporting Information S.1.

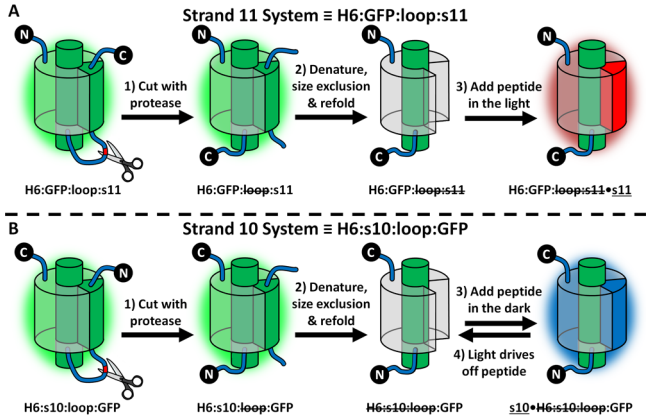


Figure 2. Preparation and photochemistry of the strand 11 and strand 10 systems. (A) H6:GFP:~~loop:s11~~ is prepared by proteolysis of a sacrificial loop inserted between strand 11 and the rest of GFP (step 1), and removal of strand 11 by denaturation in 6 M guanidinium-HCl, separation of H6:GFP:~~loop:s11~~ from the strand 11 peptide by size exclusion chromatography, and refolding via rapid dilution into nondenaturing buffer (step 2). This construct does not bind to strand 11 peptide in the dark, but does rebind strand 11 after light activation (step 3). (B) ~~H6:s10:loop~~:GFP prepared analogously (steps 1 and 2) spontaneously reassembles with strand 10 peptide in the dark (step 3), and this reassembled complex can undergo strand exchange upon light activation (step 4). Colors reflect strand colors in Figure 1, not particular spectral features, which are discussed in the text.

assigned by comparison of preresonance Raman spectra to that of model chromophores.[14] This truncated construct does not rebind added synthetic strand 11 (s11) spontaneously in the dark; however, exposure to light generates a species that is receptive to exogenous strand 11 binding from solution. In the absence of added peptide, light activated H6:GFP:~~loop:s11~~ is prone to aggregation above ~10 μM concentrations.[10]

Prepared in the same manner, ~~H6:s10:loop~~:GFP (Figure 2B) results in a truncated protein which is prone to aggregation above ~10 μM. This truncated protein also has a characteristic absorption spectrum different from intact H6:s10:loop:GFP, and a much lower fluorescence quantum yield (lower by a factor of 30 compared to intact H6:s10:loop:GFP).[12] This construct readily reassembles with added synthetic strand 10 (s10) in the dark to form s10•~~H6:s10:loop~~:GFP. Furthermore, the reassembled s10•~~H6:s10:loop~~:GFP was shown to undergo light driven strand exchange in the presence of excess replacement peptide ($s10^{203Y}$ was used to generate yellow fluorescent protein).[12] This process also occurs very slowly in the dark, because the spontaneous off-rate of strand 10 is very slow. In comparison, the spontaneous off-rate of strand 11 from reassembled H6:GFP:~~loop:s11~~•s11 in the presence of excess replacement peptide ($s11^{222Q}$ was used to generate a color shifted mutant) is effectively zero.[10] In summary, the strand 11 system, H6:GFP:~~loop:s11~~, undergoes light-driven photo*association*, while the strand 10 system, ~~H6:s10:loop~~:GFP, undergoes light-driven photo*dissociation*, but the origin of this difference was entirely unclear despite detailed kinetic analysis for both species.

RESULTS AND DISCUSSION

His-Tag Stabilization of Truncated GFPs. In order to understand the origin(s) of the differences between these two seemingly similar systems, we attempted to investigate their structures directly using X-ray crystallography. We focused first on the strand 11 system, H6:GFP:~~loop:s11~~, due to its favorable solubility. Since it is often the case that extraneous residues may hinder crystallization,[15] we cleaved off the H6 by restoring the cleavage site between the H6 and the start of GFP (Figure 1A; see sequence in Supporting Information S.1), but found that the resultant ~~H6~~:GFP:~~loop:s11~~ was prone to aggregation above ~10 μM and displayed a red-shifted UV−vis absorption spectrum compared to H6:GFP:~~loop:s11~~ (Figure S4). This suggests that the His-tag is not a neutral actor as previously thought, and we hypothesized that it may have a functional role in stabilizing the truncated protein, H6:GFP:~~loop:s11~~. Extending this to the strand 10 system, an alternative circular permutant was made with strand 10 at the other terminus (Figure S1), such that the His-tag is retained when strand 10 is removed. We found that H6:GFP:~~loop:s10~~ is a truncated protein that is soluble to low mM without aggregation, has a different UV−vis absorption spectrum, higher fluorescence quantum yield, and displays slower peptide rebinding kinetics in the dark compared to ~~H6:s10:loop~~:GFP (Figures S2 and S3). These observations suggest that the His-tag is playing a key role in stabilizing these truncated proteins and slowing or preventing peptide binding.

X-ray Structures. The crystal structure of H6:GFP:~~loop:s11~~ was solved by molecular replacement using the original Superfolder GFP structure (PDB ID: 2B3P[9]) as a search model. During model building, the replacement model

DOI: 10.1021/jacs.7b10680
J. Am. Chem. Soc. 2018, 140, 375−381

was truncated to remove strand 11; however, the electron density along the groove was consistent with the presence of another β-strand coming in from the opposite side. Building backward from the beginning of GFP, we could place nearly all 25 residues that made up the N-terminal His-tag and thrombin cleavage loop that was fused to GFP. Since these residues are upstream of the main GFP sequence, we extend the numbering convention to count backward and define M^{-23} as the true beginning of this protein. Solving the structure of H6:GFP:~~loop:s10~~ in a similar manner, we were also able to build residues from the His-tag and thrombin cleavage loop into the groove that was occupied by strand 10 in Superfolder GFP. Although the N-terminal sequence of this construct is shorter by 4 residues (i.e., residue −2 through 1 are missing in this construct, see Supporting Information S.1), we will still define M^{-23} as the beginning of this protein for ease of comparison. The final refinement models are shown in Figure 3 (collection and refinement statistics in Table S1).

In both crystal structures, the His-tag sequence forms a surrogate β-strand that slots into the groove vacated by the native strands. In H6:GFP:~~loop:s11~~ (Figure 3A), in place of the native HMVLLEFVTAA of strand 11, the main chain of the His-tag and five neighboring amino acids SHHHHHHSSGL form a canonical β-strand with hydrogen bonds to the main chain of residues on flanking strands 3 and 10, and histidine side chains alternating pointing out from and into the β-barrel (see Video S1 where this is highlighted). Similarly, in H6:GFP:~~loop:s10~~ (Figure 3C), in place of the native NHYLSTQSVLS of strand 10, hydrogen bonds are formed between the His-tag and five adjacent amino acids HHHH-HHSSGLV and flanking strands 11 and 7. In both variants, we find that the inward-facing residues on the His-tag replace key components of the excited state proton transfer (ESPT) pathway.[16] E^{222} is replaced by H^{-15} in H6:GFP:~~loop:s11~~, and T^{203} and T^{205} are replaced by H^{-14} and S^{-12}, respectively, in H6:GFP:~~loop:s10~~. In both cases, the overall β-barrel shape of GFP is retained, and the remainder of the protein environment near the chromophore remains largely unchanged.

Although both structures are very similar, there are some subtle differences to note. Even though the length of the surrogate β-strand is the same, the residues involved on the H6 are off register by one in that H6:GFP:~~loop:s11~~ starts with S^{-20}, while H6:GFP:~~loop:s10~~ starts with the next residue, H^{-19}. One possible explanation for this is the relative amount of sequence homology between the H6 and the native strands: there is no homology with s11 in either register along the length of the β-sheet, while this particular registry has three common inward-facing residues with the native strand 10, namely, H^{199}, S^{205}, and L^{207}, out of a total of five possible inward-facing residues, likely biasing the way this strand slots into the vacated groove. The H6 in H6:GFP:~~loop:s11~~ connects to strand 1 via an over-the-barrel ~20 residue linker that most closely resembles the linking region between strand 6 and strand 7 (see topology in Figure 1), while the H6 in H6:GFP:~~loop:s10~~ connects to strand 11 via a short ~4 residue linker similar to that between most other strands in the β-barrel fold that are spatially adjacent. In both cases, the chromophore appears to be in the *cis* configuration, and not *trans*, as previously posited for H6:GFP:~~loop:s11~~ based on preresonance Raman data.[10] However, the positioning of the −OH group (originally from S^{65}) where the chromophore connects back to the main protein chain is different between the two structures. This is due to the formation of a hydrogen bond to H^{-17} a distance of 2.7 Å away

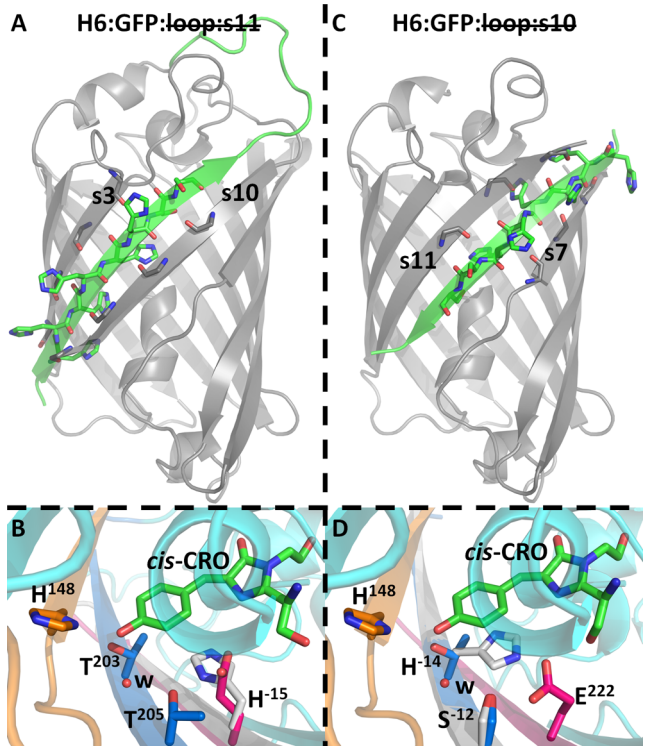


Figure 3. Structural models of H6:GFP:~~loop:s11~~ and H6:GFP:~~loop:s10~~. (A) Crystal structure of H6:GFP:~~loop:s11~~ (PDB ID: 6B7R). The H6, shown in green, forms a β-sheet and fits into the groove vacated by strand 11. Hydrogen bonding partners along the backbone of neighboring strands 3 and 10 are displayed in stick representation. (B) Excited state proton transfer (ESPT)[16] relay in H6:GFP:~~loop:s11~~ overlaid with strand 11 from Superfolder GFP (PDB ID: 2B3P[9]). The ordered water that forms a part of this relay is shown as a red sphere, labeled "w". Strands 7, 10, 11, and the H6 are in orange, blue, magenta, and gray, respectively. H^{-15} takes the place of E^{222} on strand 11, the terminal proton acceptor along this ESPT pathway. Note that H6:GFP:~~loop:s11~~ has S^{65},[10] compared to S^{65T} in Superfolder GFP, which causes E^{222} to be oriented differently.[17] (C) Crystal structure of H6:GFP:~~loop:s10~~ (PDB ID: 6B7T). The H6 forms a β-sheet and fits into the groove vacated by strand 10. Hydrogen bonding partners along the backbone of neighboring strand 11 and 7 are displayed in stick representation. (D) The ESPT relay in H6:GFP:~~loop:s10~~ overlaid with strand 10 from superfolder GFP. Strands 7, 10, 11, and the H6 are in orange, blue, magenta, and gray, respectively. H^{-14} and S^{-12} take the place of T^{203} and T^{205}, respectively, two key components along this relay. See Video S1 for a closer view of the His-tag placement.

from this S^{65} oxygen atom in H6:GFP:~~loop:s11~~. This interaction, which is not present in H6:GFP:~~loop:s10~~, may be the source of the robust gatekeeping behavior observed in the strand 11 system (see below), where even added native peptide is unable to displace the H6 slotted into this groove until it is vacated following light activation.

Incorporation of the His-tag as a well-defined structural element is highly unusual. While there are examples of structured His-tags appearing in the electron density for structures in the PDB, we have not been able to find any previous case where it is incorporated into the fused protein as a functional element. Previous studies have been conducted in which E^{222} on strand 11[18] and T^{203} on strand 10[19] have been mutated to histidine. The absorbance spectra of these mutants are qualitatively similar to those of H6:GFP:~~loop:s11~~[10] and H6:GFP:~~loop:s10~~ (Figure S2). In the former, the chromophore

DOI: 10.1021/jacs.7b10680
J. Am. Chem. Soc. 2018, 140, 375−381

population is almost exclusively in the deprotonated B state, and in the latter, the population is almost exclusively in the protonated A state.[20] To further verify that the His-tag is indeed interacting with the chromophore in solution and not an artifact of crystallization, we mutated H^{-15} in H6:GFP:~~loop:s11~~ (positioned where E^{222} is in the native fold) to Glu and Gln, and observed subtle changes to the absorbance spectrum (Figure S5). Were the His-tag truly unstructured and not interacting with the chromophore, there is no reason to expect that mutations on this tag would have any impact on the spectral properties of the protein, however subtle. With the His-tag acting as a surrogate β-strand, the overall β-barrel of GFP is preserved in H6:GFP:~~loop:s11~~, and H6:GFP:~~loop:s10~~, leading to their increased stability and solubility compared to their respective counterparts without the His-tag.

Characterization of Truly Truncated GFP. In light of these findings, the protein H6:GFP:~~loop:s11~~, while truncated in the sense that the s11 sequence is no longer present, is not truly truncated because an s11 surrogate is present in the form of the His-tag. Therefore, comparing its strand binding and photochemistry to that of ~~H6:s10:loop~~:GFP is not appropriate. To make a more relevant comparison, we purified and characterized the truly truncated ~~H6~~:GFP:~~loop:s11~~. We found that this protein has a tendency to aggregate, as it comes off earlier in the size exclusion chromatography run compared to H6:GFP:~~loop:s11~~ (Figure S6) or H6:GFP:~~loop:s11~~•s11 (data not shown), suggesting that the truncated protein likely exists as a higher order oligomeric species when there is nothing to occupy the groove vacated by strand 11. The fluorescence of ~~H6~~:GFP:~~loop:s11~~ provides another interesting observation: the fluorescence quantum yield of the oligomeric truncated species is initially comparable to that of ~~H6~~:GFP:~~loop:s11~~•s11, while the fluorescence quantum yield of H6:GFP:~~loop:s11~~ is lowered by a factor of 3 compared to ~~H6~~:GFP:~~loop:s11~~•s11 (Figure S7) (in contrast, the fluorescence quantum yield of ~~H6:s10:loop~~:GFP is 30-fold lower than that of s10•~~H6:s10:loop~~:GFP), indicating that internal strand binding of the H6 leads to monomerization and a reduction of fluorescence. These observations are consistent with previous studies where highly fluorescent oligomeric GFP complexes lose fluorescence upon monomerization.[21,22]

Strand Binding and Photochemistry. We found that when mixed in solution with excess s11 in the dark, ~~H6~~:GFP:~~loop:s11~~ does spontaneously reassemble, with a rate constant of $k = 1.2 \times 10^0$ M^{-1} s^{-1} (Figure 4A). This rebinding process is greatly accelerated by irradiation with 15 mW 488 nm light, with $k = 2.0 \times 10^2$ M^{-1} s^{-1} (Figure 4B). This light-driven photoassociation rate is very similar to the $k = 1.5 \times 10^2$ M^{-1} s^{-1} binding rate reported for s11 binding to light-activated H6:GFP:~~loop:s11~~,[10] hinting that the covalently attached but photodissociated H6 is not significantly affecting the binding process in the light-activated strand 11 system. The reassembled ~~H6~~:GFP:~~loop:s11~~•s11 is then able to undergo photodissociation, readily observed by strand exchange in the presence of excess replacement peptide ($s11^{222Q}$) (Figure 4C). Additionally, irradiation of ~~H6~~:GFP:~~loop:s11~~ with 488 nm light in the absence of an added strand generates a photostationary mixture of the light-activated species and the starting thermally stable species, with the former relaxing back to the latter with $k = 6.3 \times 10^2$ s^{-1} (Figure 4D). These findings are consistent with the photoassociation phenomenon reported previously in H6:GFP:~~loop:s11~~,[10] and also with the photodissociation phenomenon previously reported in ~~H6:s10:loop~~:GFP,[12]

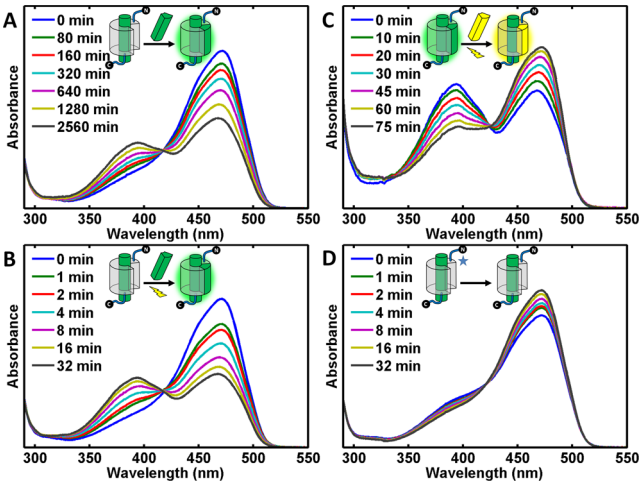


Figure 4. Characterization of truly truncated ~~H6~~:GFP:~~loop:s11~~. (A) 4 μM ~~H6~~:GFP:~~loop:s11~~ spontaneously reassembles with 20 μM s11, with $k = 1.2 \times 10^0$ M^{-1} s^{-1}. (B) Irradiation of the mixture in (A) with 15 mW 488 nm light greatly accelerates the rebinding process, $k = 2.0 \times 10^2$ M^{-1} s^{-1}. This light driven photoassociation of s11 to ~~H6~~:GFP:~~loop:s11~~ proceeds with a similar rate to the light driven photoassociation of s11 to H6:GFP:~~loop:s11~~.[10] (C) 300 mW of 488 nm light was used to irradiate a sample of 2 μM of ~~H6~~:GFP:~~loop:s11~~•s11 under pseudo-first order conditions with 30-fold excess of $s11^{222Q}$ which generates a color-shifted complex, with $k = 6.02 \times 10^3$ s^{-1}. This light driven photodissociation of s11 from ~~H6~~:GFP:~~loop:s11~~•s11 via strand exchange to $s11^{222Q}$ is analogous to the light driven photodissociation of s10 from ~~H6:s10:loop~~:GFP via strand exchange to $s10^{203T}$ (ref 12). (D) Irradiation of 2 μM ~~H6~~:GFP:~~loop:s11~~ with 15 mW 488 nm light in the absence of added peptides creates a photostationary mixture of the thermally stable and the light-activated species (denoted with a star), which then relaxes back to the former species, with $k = 6.3 \times 10^2$ s^{-1}.

suggesting that photochemistry proceeds with the same mechanism in both the strand 11 and strand 10 systems.

Previous characterization of the photochemistry of H6:GFP:~~loop:s11~~, found that light activation of the stable, unreceptive truncated protein generated a photostationary state containing a less stable species that is receptive to peptide binding.[10] The thermal relaxation of the light-activated species back to the dark species follows temperature-dependent Arrhenius behavior with an activation energy of 130 kJ/mol,[10] which is comparable to model chromophore isomerization activation energies.[14] Due to lack of structural evidence at the time, the previous model proposed that light-activation caused isomerization of the chromophore which is reversed by thermal relaxation, and that properties such as protein stability and peptide receptivity are a function of the isomerization state of the chromophore (i.e., *cis* is less stable and receptive to peptide, while *trans* is more stable and not receptive to peptide). In light of the discovery of the His-tag surrogate strand and that the thermally stable form of the H6:GFP:~~loop:s11~~ is in fact *cis*, the previous model must be updated accordingly: what was thought to be the photo-*association* of strand 11 was in fact a convolution of the photo*dissociation* of the His-tag strand and the binding of strand 11. A later, in-depth study of the mechanism and bottlenecks of the photodissociation process in similar split GFP systems found that light activation causes *cis-trans* isomerization that leads to strand displacement,[23] and also demonstrated that a light-activated truncated GFP with a chromophore in the *trans*

DOI: 10.1021/jacs.7b10680
J. Am. Chem. Soc. 2018, 140, 375−381

state is capable of binding peptide. However, while the premise that receptivity to peptide is dependent on the isomerization state of the chromophore is refuted, many of the previous findings still fit within the phenomenological framework of the old model, updated in light of the new structural information.

Combining the structural evidence with the aforementioned phenomenology, we find interesting parallels between truly truncated ~~H6~~:GFP:~~loop:s11~~ and H6:GFP:~~loop:s11~~ and can propose a revised model that is consistent with these observations, summarized in Figure 5. In this model, the thermally stable form of H6:GFP:~~loop:s11~~ is the monomeric complex with the internal H6 occupying the slot left vacant by removal of s11 as shown by X-ray crystallography (Figure 5A), while the thermally stable form of truly truncated ~~H6~~:GFP:~~loop:s11~~ is the oligomeric complex which appears to block strand binding (Figure 5B, where a dimer is shown for simplicity). Light activation of these two species generates a similar monomeric species that is able to bind s11 quickly, regardless of the presence of the internal His-tag, (200 M^{-1} s^{-1} for ~~H6~~:GFP:~~loop:s11~~ vs 150 M^{-1} s^{-1} for H6:GFP:~~loop:s11~~).[10] This suggests that once the H6 is photodissociated, it does not significantly influence the properties of the truncated protein until it rebinds. In the absence of peptide substrate, this light-activated species will relax back to the dark species. The remaining mystery lies in the large discrepancy between the relaxation time scales of the light-activated species. However, the fact that the photostationary state of the light activated ~~H6~~:GFP:~~loop:s11~~ (Figure 4D, initial spectrum) is different than photostationary state of H6:GFP:~~loop:s11~~[10] is likely due to the His-tag strand being bound to the light-activated species, which results in a different, much slower relaxation process. That the relaxation process of light activated H6:GFP:~~loop:s11~~ follows a clean isosbestic point[10] is consistent with this observation: the two species involved both have the His-tag bound only differ by isomerization state.

Truncated GFPs and Peptide Promiscuity. It is remarkable that a His-tag can substitute for a β-strand in two different grooves in GFP, with little to no sequence homology and very different chemical environments, so we tested whether the His-tag sequence itself had any propensity to bind to these truncated proteins without being covalently tethered where it has a high effective molarity. The split GFP system offers an unusual opportunity to study peptide−protein interactions with the chromophore as a sensitive, built-in reporter. We quantified the binding affinity of a His-tag-like peptide (denoted H6, see Supporting Information S.1 for sequence) in solution to ~~H6~~:GFP:~~loop:s11~~ by measuring a binding isotherm (Figure S8). Here, 2 μM of ~~H6~~:GFP:~~loop:s11~~ was mixed with different concentrations of H6 at 298 K and allowed to reach equilibrium. The equilibrium absorbance at 464 nm was plotted as a function of peptide concentration, and fit to a single exponential. From this, the calculated K_D = 2.4 μM for the His-tag peptide binding to ~~H6~~:GFP:~~loop:s11~~, in contrast to the subpicomolar affinity for native strand 11 peptide binding.

In the presence of excess His-tag-like-peptide (denoted H6, see Supporting Information S.1 for sequence) in solution, we found that both truly truncated proteins (~~H6~~:GFP:~~loop:s11~~ and ~~H6:s10:loop~~:GFP) were able to spontaneously bind H6 peptide in the dark, shown by a change to the absorbance spectrum of the complexed species (Figure 6A and B). This binding occurs on a similar time scale to that of native strand binding. We attempted to separate the bound complexes (H6•~~H6~~:GFP:~~loop:s11~~ and H6•~~H6:s10:loop~~:GFP) from the

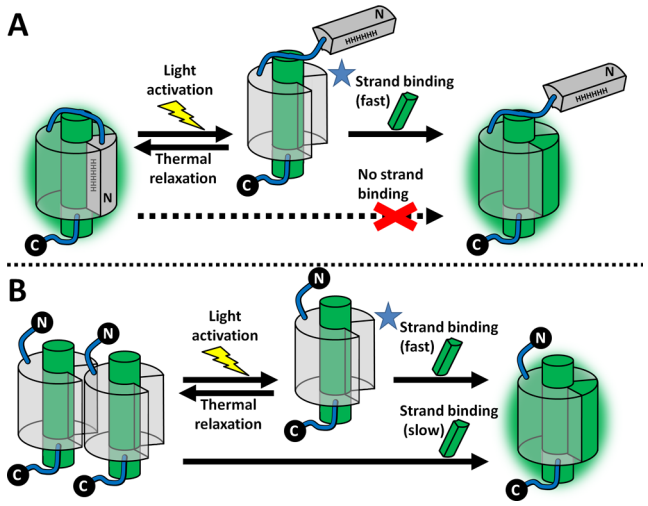


Figure 5. Revised model of the strand 11 system and comparison to the truly truncated protein. (A) The thermally stable strand 11 system, H6:GFP:~~loop:s11~~, is unable to bind s11 in the dark but light activation with 488 nm light displaces the His-tag to generate a species with a different chromophore isomerization state (denoted with star) that is able to rapidly bind free s11.[10] If no added peptide is present, this activated and H6-displaced species relaxes back to the thermally stable species which involves a convolution of the reverse isomerization process and the internal strand rebinding. (B) The thermally stable truly truncated ~~H6~~:GFP:~~loop:s11~~ exists as an oligomeric complex that is able to bind various peptides slowly, presumably because complex dissociation is required. As in panel (A), light activation generates a species with a different chromophore isomerization state (denoted with star) that is able to rapidly bind s11. If there is no peptide present, relaxation back to the thermally stable species involves a convolution of the reverse isomerization process and oligomerization.

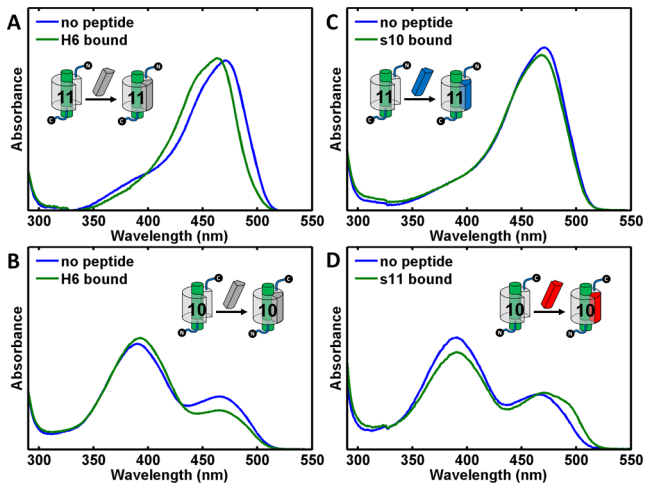


Figure 6. Receptivity of truly truncated GFP species to non-native strand binding. (A) 2 μM ~~H6~~:GFP:~~loop:s11~~ (cartoon diagram with "11") was mixed with 30-fold excess H6 in solution. (B) 2 μM ~~H6:s10:loop~~:GFP (cartoon diagram with "10") was mixed with 30-fold excess H6 in solution. (C) 2 μM ~~H6~~:GFP:~~loop:s11~~ was mixed with 30-fold excess s10. (D) 2 μM ~~H6:s10:loop~~:GFP was mixed with 30-fold excess $s11^{222Q}$ in solution,. Peptide colors are in accordance with the color scheme in Figure 1, i.e., gray, blue, and red for H6, s10, and s11 respectively. As binding occurs, the absorbance spectrum shifts from the free truncated protein in the blue trace toward the bound spectrum shown in the green trace. We note that the exact registry of the strand in the structure of the bound state is unknown.

DOI: 10.1021/jacs.7b10680
J. Am. Chem. Soc. 2018, 140, 375−381

excess H6 peptide via size exclusion chromatography. However, these complexes dissociated back into their constituents over time scales similar that of strand binding; in the case of H6•~~H6:s10:loop~~:GFP, the complex was not observed as it had dissociated before its absorbance spectrum could be taken. This long time scale for the H6 strand dissociation is in contrast to the very slow (days) time scale of the native strand exchange, suggesting that H6 binding to ~~H6~~:GFP:~~loop:s11~~ and ~~H6:s10:loop~~:GFP reflects the high concentration of the added strand and not anything unique about the sequence itself.

As a further demonstration of the promiscuity of truly truncated GFP species to strand binding, we were able to bind each truly truncated protein to the peptide complement of the construct; i.e., we bound s11 to ~~H6:s10:loop~~:GFP (Figure 6C) and s10 to H6:GFP:~~loop:s11~~ (Figure 6D) under pseudo-first-order conditions with excess peptide. We monitored the binding by observing changes to the absorbance spectra over time, and separated the complex away from the excess peptide via size exclusion chromatography. The subtle changes to the absorption spectrum are an indicator that the chromophore environment in GFP is being perturbed by the addition of this non-native peptide, and mass spectrometry of the monodisperse protein peak that elutes off during size-exclusion chromatography contains mass fragments that match with the truncated protein and that of the non-native complementary peptide. In the former case, we created s11•~~H6:s10:loop~~:GFP on a similar time scale to H6 and s10 binding. As with H6 binding previously, this complex falls apart during its purification by size exclusion chromatography. In the latter case, we created s10•~~H6~~:GFP:~~loop:s11~~ on a similar time scale to H6 and s11 binding; this complex can be successfully isolated via size exclusion chromatography (the absorbance spectrum does not revert back to that of ~~H6~~:GFP:~~loop:s11~~ over the same time scale as peptide binding). Excess s11 added to this isolated complex is able to displace the bound s10, and light accelerates this process (data not shown). Both of these experiments further reinforce the conclusion that other strands can be used to intramolecularly complement a truly truncated GFP, and that histidine residues are not necessary for this interaction.

The tolerance of these truncated GFP species to non-native sequences may be a feature unique to the β-barrel structure of GFP. Our added peptides do not contain proline residues or very bulky inward-facing substitutions in the region of interest, allowing for the robust hydrogen bonding template of the flanking strands to dictate the binding for sequences that can have no sequence homology to the native strand (e.g., H6 and s11). This is also consistent with the variety of fluorescent proteins that adopt the same overall fold as GFP, but have low sequence homology. For example, Azami Green from the stony coral only has about 27% sequence homology to GFP, despite having similar spectral properties.[24] We do note that absent structural evidence, we can only determine spectrally and via mass spectrometry (Supporting Information S.3) that a complex is being formed, but cannot comment on the exact registry of the binding.

Revised Model Unifies Properties of Split Proteins. Our analysis began with a summary of the apparent differences between the behaviors of the strand 10 and strand 11 systems as defined in Figures 1 and 2. While all of the kinetic and thermodynamic data in ref[10] and ref[12] are established, the model developed to explain the data for the strand 11 system was quite different from that of the strand 10 system because the unique role of the His-tag was not realized. These crystal structures help to demystify many of the phenomena previously observed and thought to differentiate the two systems. Removal of the His-tag from H6:GFP:~~loop:s11~~ generates a truly truncated protein (~~H6~~:GFP:~~loop:s11~~) that is receptive to peptide binding in the dark, and whose reassembled complex is capable of undergoing photodissociation, much like that of truly truncated ~~H6:s10:loop~~:GFP. Because H6:GFP:~~loop:s11~~ was found to have a surrogate β-strand in place of strand 11, it is not a truly truncated protein, but instead a photoreactive caged protein where the His-tag acts as a competitive inhibitor[25] to exogenous strand binding. Light-driven photodissociation is required to expel the His-tag and free up the groove before the protein can rebind exogenous s11. This behavior is consistent with the photodissociation phenomena observed in both ~~H6~~:GFP:~~loop:s11~~•s11 and s10•~~H6:s10:loop~~:GFP. Overall, the two systems that were thought to be distinct now converge to the same model.

Furthermore, it is an interesting and potentially useful property that an internally bound sequence can be photodissociated without being cleaved. For example, fusion of the truncated protein containing the His-tag to a target of interest and the complementary peptide to cargo would grant one-time optogenetic control over the delivery of the cargo to the target. This finding is a step toward the goal of generating a multiple turnover bistable switching GFP where light-driven strand exchange can toggle between two binding modes of a bistable switching GFP, which was first proposed by Do and Boxer.[13] Previous efforts to generate this two-tailed split GFP system required cleavage of the attached strand before strand exchange could occur; this cleavage requirement greatly limits the utility of such a system in biological applications since a protease coenzyme must be present to enable activity. The promiscuity of truncated GFPs to peptide binding suggests that directed evolution may be used to select for sequences suitable for this application, i.e., sequences that can associate to the truncated protein while covalently tethered, and photodissociate at different irradiation wavelengths to allow for switching between the two binding modes.

CONCLUSION

To summarize, we report a highly unusual configuration for a His-tag, which adopts a β-strand configuration in two truncated GFPs. The discovery of this strand resolves many of the mysteries about the differences between two seemingly similar systems of split GFPs, although some mysteries still remain. The His-tag sequence has the useful property of being able to photodissociate despite being covalently attached, and the promiscuity of truly truncated GFPs to similar non-native sequences of varying levels of homology invites the possibility of rational design or directed evolution of sequences optimized for strand binding and dissociation, with potential applications in optogenetics.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b10680.

Complete protein and DNA sequences for all constructs, protein purification methods, instrumentation, characterization of the His-tagged circular permutant of the strand

DOI: 10.1021/jacs.7b10680
J. Am. Chem. Soc. 2018, 140, 375−381

10 system (H6:GFP:~~loop:s10~~), characterization of truly truncated strand 11 system (~~H6~~:GFP:~~loop:s11~~), and crystallography data and refinement statistics (PDF)
Structure animation of H6:GFP:~~loop:s11~~ (AVI)
Coordinates file for H6:GFP:~~loop:s10~~ (CIF)
Structure factors file for H6:GFP:~~loop:s10~~ (CIF)
Coordinates file for H6:GFP:~~loop:s11~~ (CIF)
Structure factors file for H6:GFP:~~loop:s11~~ (CIF)

AUTHOR INFORMATION

Corresponding Author

*sboxer@stanford.edu

ORCID

Alan Deng: 0000-0002-3893-7128
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 (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. We thank Dr. Marc Deller from the Stanford ChEM-H Macromolecular Structure Knowledge Center, where part of this work was performed, for assistance with crystallography, and Dr. Silvia Russi and Dr. Irimpan Mathews from SSRL for assistance with the data collection. We also thank Johan Both, Dr. Keunbong Do, Dr. Kevin Kent, Chi-Yun Lin, Matt Romei, Dr. Luke Oltrogge, and Samuel Schneider for many helpful discussions and comments. We thank Dr. Wladek Minor for bringing ref 25 to our attention. This research was supported by grants from the NIH (GM27738 and GM118044).

REFERENCES

(1) Structural Genomics Consortium, Architecture et Fonction des Macromolécules Biologiques, Berkeley Structural Genomics Center, China Structural Genomics Consortium, Integrated Center for Structure and Function Innovation, Israel Structural Proteomics Center, Joint Center for Structural Genomics, Midwest Center for Structural Genomics, New York Structural GenomiX Research Center for Structural Genomics, Northeast Structural Genomics Consortium, Oxford Protein Production Facility, Protein Sample Production Facility, Max Delbrück Center for Molecular Medicine, RIKEN Structural Genomics/Proteomics Initiative & SPINE2-Complexes. *Nat. Methods* **2008**, *5*, 135−146.
(2) Porath, J. *Protein Expression Purif.* **1992**, *3*, 263−281.
(3) Carson, M.; Johnson, D. H.; McDonald, H.; Brouillette, C.; DeLucas, L. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *D63*, 295−301.
(4) Ghosh, I.; Hamilton, A. D.; Regan, L. *J. Am. Chem. Soc.* **2000**, *122*, 5658−5659.
(5) Huang, Y.-M.; Bystroff, C. *Biochemistry* **2009**, *48*, 929−940.
(6) Kent, K. P.; Oltrogge, L. M.; Boxer, S. G. *J. Am. Chem. Soc.* **2009**, *131*, 15988−15989.
(7) Kent, K. P.; Childs, W.; Boxer, S. G. *J. Am. Chem. Soc.* **2008**, *130*, 9664−9665.
(8) Cabantous, S.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2005**, *23*, 102−107.
(9) Pedelacq, J.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2006**, *24*, 79−88.
(10) Kent, K. P.; Boxer, S. G. *J. Am. Chem. Soc.* **2011**, *133*, 4046−4052.
(11) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11241−11246.
(12) Do, K.; Boxer, S. G. *J. Am. Chem. Soc.* **2011**, *133*, 18078−18081.
(13) Do, K.; Boxer, S. G. *J. Am. Chem. Soc.* **2013**, *135*, 10226−10229.
(14) Luin, S.; Voliani, V.; Lanza, G.; Bizzarri, R.; Amat, P.; Tozzini, V.; Serresi, M.; Beltram, F. *J. Am. Chem. Soc.* **2009**, *131*, 96−103.
(15) Waugh, D. S. *Trends Biotechnol.* **2005**, *23*, 316−320.
(16) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509−44.
(17) Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormo, M.; Remington, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2306−2311.
(18) Auerbach, D.; Klein, M.; Franz, S.; Carius, Y.; Lancaster, C. R. D.; Jung, G. *ChemBioChem* **2014**, *15*, 1404−1408.
(19) Kummer, A. D.; Wiehler, J.; Rehaber, H.; Kompa, C.; Steipe, B.; Michel-Beyerle, M. E. *J. Phys. Chem. B* **2000**, *104*, 4791−4798.
(20) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8362−8367.
(21) Zhou, X. X.; Chung, H. K.; Lam, A. J.; Lin, M. Z. *Science* **2012**, *338*, 810−814.
(22) Huang, Y.-M.; Banerjee, S.; Crone, D. E.; Schenkelberg, C. D.; Pitman, D. J.; Buck, P. M.; Bystroff, C. *Biochemistry* **2015**, *54*, 6263−6273.
(23) Lin, C. Y.; Both, J. H.; Do, K. B.; Boxer, S. G. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E2146−E2155.
(24) Karasawa, S.; Araki, T.; Yamamoto-Hino, M.; Miyawaki, A. *J. Biol. Chem.* **2003**, *278*, 34167−34171.
(25) There is precedent in the literature: Majorek, K. A.; Kuhn, M. L.; Chruszcz, M.; Anderson, W. F.; Minor, W. *Protein Sci.* **2014**, *23*, 1359−1368 ,where the presence of a His-tag can slow down kinetic activity in several *N*-acetyltransferases via a binding occlusion mechanism, somewhat similar to the case described here.

DOI: 10.1021/jacs.7b10680
J. Am. Chem. Soc. 2018, 140, 375−381

Supporting Information for: *Structural Insight Into the Photochemistry of Split Green Fluorescent Proteins: A Unique Role for a His-tag*

Alan Deng and Steven G. Boxer

Table of Contents

S.1 Sequence Information 2

Protein Sequences 2

DNA Sequences 3

S.2 Protein Purification 4

S.3 Instrumentation and Methods 5

S.4 Characterization of H6:GFP:~~loop:s10~~ 6

Figure S1: Strand Topology of H6:GFP:~~loop:s10~~ 6

Figure S2: Spectral Characterization of H6:GFP:~~loop:s10~~ 7

Figure S3: Binding Kinetics of H6:GFP:~~loop:s10~~ 7

S.5 Characterization of ~~H6~~:GFP:~~loop:s11~~ 7

Figure S4: Differences Resulting from Removal of His Tag 8

Figure S5: Absorption Spectra of H6:GFP:~~loop:s11~~ H^{-15} Mutants 8

Figure S6: Size exclusion chromatography of truncated vs truly truncated GFP 9

Figure S7: Fluorescence emission spectra of strand 11 system variants 10

Figure S8: H6 peptide binding isotherm for ~~H6~~:GFP:~~loop:s11~~ 11

S.6 Crystallography 12

Table S1: X-ray data collection and refinement statistics 13

Video S1: Structure Animation of H6:GFP:~~loop:s11~~ 14

S.7 References 14

S.1 Sequence Information

Protein Sequences

Proteins used in this study are adapted from Superfolder GFP[1]. The strand 11 system is based on the native sequence and the strand 10 system is based on a circular permutant where the N- and C-termini are fused by a GGTGGS linker sequence. Protease cleavage sites (cleaved by trypsin unless otherwise stated) are denoted by ▼. GSSHHHHHHSSGLVPG (highlighted in gray, and represented in short by H6) comprises the N-terminal hexahistidine affinity tag along with several adjacent residues endogenous to the cloning and expressing region of the pET-15b vector. The corresponding H6 peptide was synthesized with R as the final residue to replicate the polypeptide generated by proteolytic cleavage (LVPR▼GS). GTRGSGSIEGRHSGSGS (highlighted in magenta and represented in short by loop) is the sacrificial loop inserted between the structural element of interest and the rest of the GFP sequence. LPDNHYLSTQTVLSKDPNE (highlighted in yellow, and represented in short by s10) and RDHMVHEYVNAAGIT (highlighted in red, and represented in short by s11) are the respective sequences of strand 10 and strand 11 in the native sequence of GFP. DNA sequences are also given.

Proteins

H6:GFP:loop:s11

MGSSHHHHHHSSGLVPGGSHMGGTSSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFISTTGKLPVPWPTLV
TTLSYGVQAFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFN
SHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEK▼GTRGSGSIEGRHSGSG
SRDHMVLHEYVNAAGITHGMDELYKGSGGT

H6:s10:loop:GFP

MGSSHHHHHHSSGLVPGGSHMLPDNHYLSTQTVLSKDPNEGTRGSGSIEGR▼HSGSGSKRDHMVLHEYVNAAGITHGMDELYGGT
GGSASQGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFISTTGKLPVPWPTLVTTLSYGVQAFSRYPDHMKRHD
FFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTV
RHNVEDGSVQLADHYQQNTPIGDGPVL

H6:GFP:loop:s10

MGSSHHHHHHSSGLVPGGSHMKRDHMVLHEYVNAAGITHGMDELYGGTGGSASQGEELFTGVVPILVELDGDVNGHKFSVRGEGE
GDATIGKLTLKFISTTGKLPVPWPTLVTTLSYGVQAFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTL

VNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLGTR▼GSGSIE
GRHSGSGSLPDNHYLSTQTVLSKDPNE

Peptides

H6 = GSSHHHHHHSSGLVPR

s10 = LPDNHYLSTQTVLSKDPNE

s11 = RDHMVLHEYVNAAGIT

DNA Sequences

H6:GFP:loop:s11

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGGGCGGCAGCCATATGGGTGGTACCAGCAGCAAAGGTG
AAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAACTGGATGGCGATGTGAACGGTCACAAATTCAGCGTGCGTGGTGAAGG
TGAAGGCGATGCCACGATTGGCAAACTGACGCTGAAATTTATCAGCACCACCGGCAAACTGCCGGTGCCGTGGCCGACGCTGGTG
ACCACCCTGAGCTATGGCGTTCAGGCGTTTAGTCGCTATCCGGATCACATGAAACGTCACGATTTCTTTAAATCTGCAATGCCGG
AAGGCTATGTGCAGGAACGTACGATTAGCTTTAAAGATGATGGCAAATATAAAACGCGCGCCGTTGTGAAATTTGAAGGCGATAC
CCTGGTGAACCGCATTGAACTGAAAGGCACGGATTTTAAAGAAGATGGCAATATCCTGGGCCATAAACTGGAATACAACTTTAAT
AGCCATAATGTTTATATTACGGCGGATAAACAGAAAAATGGCATCAAAGCGAATTTTACCGTTCGCCATAACGTTGAAGATGGCA
GTGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGATGGTCCGGTGCTGCTGCCGGATAATCATTATCTGAGCAC
GCAGACCGTTCTGTCTAAAGATCCGAACGAAAAAGGCACCCGGGGTTCTGGTTCTATCGAAGGTCGTCACTCTGGTTCTGGTTCT
CGTGACCACATGGTGCTGCATGAATATGTGAACGCGGCGGGCATCACGCATGGTATGGATGAACTGTACAAAGGTAGCGGCGGTA
CCTAA

H6:s10: loop:GFP

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGGGCGGCAGCCATATGCTGCCGGATAACCATTATCTGA
GCACCCAGACCGTGCTGAGCAAAGATCCGAACGAAGGCACCCGCGGCAGCGGCAGCATTGAAGGCCGCCATAGCGGCAGCGGCAG
CAAACGCGATCACATGGTGCTGCATGAATATGTGAACGCGGCGGGCATTACCCATGGCATGGATGAACTGTATGGCGGCACCGGC
GGCAGCGCGAGCCAGGGCGAAGAACTGTTTACCGGCGTGGTGCCGATTCTGGTGGAACTGGATGGCGATGTGAACGGCCATAAAT
TTAGCGTGCGCGGCGAAGGCGAAGGCGATGCGACCATTGGCAAACTGACCCTGAAATTTATTTCCACCACCGGCAAACTGCCGGT
GCCGTGGCCGACCCTGGTGACCACCCTGAGCTATGGCGTGCAGGCCTTTAGCCGCTATCCGGATCACATGAAACGCCATGATTTT
TTTAAAAGCGCGATGCCGGAAGGCTATGTGCAGGAACGCACCATTAGCTTTAAAGATGATGGCAAATATAAAACCCGCGCGGTGG
TGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAAGGCACCGATTTTAAAGAAGATGGCAACATTCTGGGCCATAA
ACTGGAATATAACTTTAACAGCCATAACGTGTATATTACCGCGGATAAACAGAAAAACGGCATTAAAGCGAACTTTACCGTGCGC
CATAACGTGGAAGATGGCAGCGTGCAGCTGGCGGATCATTATCAGCAGAACACCCCGATTGGCGATGGCCCGGTGCTGTAA

H6:GFP:loop:s10

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGGGCGGCAGCCATATGAAACGCGATCACATGGTGCTGC
ATGAATATGTGAACGCGGCGGGCATTACCCATGGCATGGATGAACTGTATGGCGGCACCGGCGGCAGCGCGAGCCAGGGCGAAGA
ACTGTTTACCGGCGTGGTGCCGATTCTGGTGGAACTGGATGGCGATGTGAACGGCCATAAATTTAGCGTGCGCGGCGAAGGCGAA

GGCGATGCGACCATTGGCAAACTGACCCTGAAATTTATTTCCACCACCGGCAAACTGCCGGTGCCGTGGCCGACCCTGGTGACCA
CCCTGAGCTATGGCGTGCAGGCCTTTAGCCGCTATCCGGATCACATGAAACGCCATGATTTTTTTAAAAGCGCGATGCCGGAAGG
CTATGTGCAGGAACGCACCATTAGCTTTAAAGATGATGGCAAATATAAAACCCGCGCGGTGGTGAAATTTGAAGGCGATACCCTG
GTGAACCGCATTGAACTGAAAGGCACCGATTTTAAAGAAGATGGCAACATTCTGGGCCATAAACTGGAATATAACTTTAACAGCC
ATAACGTGTATATTACCGCGGATAAACAGAAAAACGGCATTAAAGCGAACTTTACCGTGCGCCATAACGTGGAAGATGGCAGCGT
GCAGCTAGCGGATCATTATCAGCAGAACACCCCGATTGGCGATGGACCGGTGCTGGGCACCCGCGGCAGCGGCAGCATTGAAGGC
CGCCATAGCGGCAGCGGCAGCCTGCCGGATAACCATTATCTGAGCACCCAGACCGTGCTGAGCAAAGATCCGAACGAATAA

S.2 Protein Purification

Protein Expression

Proteins were expressed from pET-15b vectors in BL21(DE3) cells (Invitrogen). The cells were incubated in LB Broth, Miller (EMD), and induced with IPTG (1 mM) at OD 0.6 and incubated over night (~20 hours) at 23°C in the case of H6:GFP:loop:s11 and 16°C in the cases of H6:GFP:loop:s10 and H6:s10:loop:GFP (note: these temperatures were required to obtain high yields of these proteins). The cultures were spun down, and the resulting pellet resuspended in lysis buffer (50 mM Tris, 300 mM NaCl and 10 v/v% glycerol at pH 8.0), and then lysed with a homogenizer (Avestin). The cell lysate was spun down, and the supernatant poured onto a Ni-NTA column equilibrated with lysis buffer. Two column volumes of lysis buffer containing 20 mM imidazole were used for washing, and lysis buffer with 200 mM imidazole was used to elute the protein from the column. The eluate was further purified with anion exchange chromatography (HiTrap^{TM} 5 mL Q HP, GE) in 20 mM Tris buffer at pH 8 in a gradient from 10 mM NaCl to 1 M NaCl without glycerol.

Protein Cleavage

Lyophilized trypsin (from bovine pancreas, Sigma) was dissolved in 1 mM HCl to make 10 units/μL trypsin solution and this was mixed with purified whole protein at 100 units per 1 mg of protein in solution such that the overall protein concentration was in excess of 1 mg/mL in lysis buffer. These solutions were incubated for 30 minute at 296K, after which point the reaction was quenched with a protease inhibitor cocktail. In the case of H6:GFP:loop:s11, we extended the incubation period to 180 minutes at 296K to ensure that the cleavage product is uniformly cut to K^{214} instead of producing a mixture with the product that contains the subsequent GTR residues. The cleaved proteins were separated from the protease by

anion exchange chromatography in 20 mM Tris pH 8.0 using a salt gradient from 10 mM NaCl to 1 M NaCl.

Isolation of the Truncated Proteins

Digested and purified protein was unfolded in denaturing buffer (6M Guanidinium HCl, 50 mM Hepes and 300 mM NaCl at pH 8.0) The denatured sample was applied to a size exclusion column (Superdex^{TM} 75 10/300) equilibrated with denaturing buffer to separate the truncated proteins from their complementary peptides. The truncated proteins are partly refolded in lysis buffer by diluting the denaturing buffer more than 100 fold to reach a final protein concentrations of ~10 μM. To completely remove the guanidinium hydrochloride and isolate the truncated protein monomer from aggregates, the partially refolded sample was applied to the size exclusion column equilibrated with lysis buffer. The monomer fraction was collected for further experiments.

S.3: Instrumentation and Methods

Optical Spectroscopy

A Lambda 25 UV/Vis Spectrophotometer (Perkin Elmer) was used for all absorbance measurements. All spectra were recorded at a resolution of 1 nm, and protein concentrations were in the low μM range. Excitation and emission spectra were measured using a LS55 Fluorescence Spectrometer (Perkin Elmer). All spectra were recorded at a resolution of 1 nm, and protein concentrations were in the 100 nM range.

Mass spectroscopy

A Micromass nanoESI APIUS Quadrupole Time-of-Flight mass spectrometer was used for all protein mass measurements. All samples had the buffer exchanged into double deionized water before being injected into the spectrometer. Mass values agree with previously published results.

Concentration determination

The relative absorbance at 447 nm before and after denaturing the protein in 0.1M NaOH was used to estimate the concentration of the protein through the known 447 nm extinction coefficient of the GFP chromophore[2] (44,100 $M^{-1}cm^{-1}$) in 0.1M NaOH

S.4: Characterization of H6:GFP:~~loop:s10~~

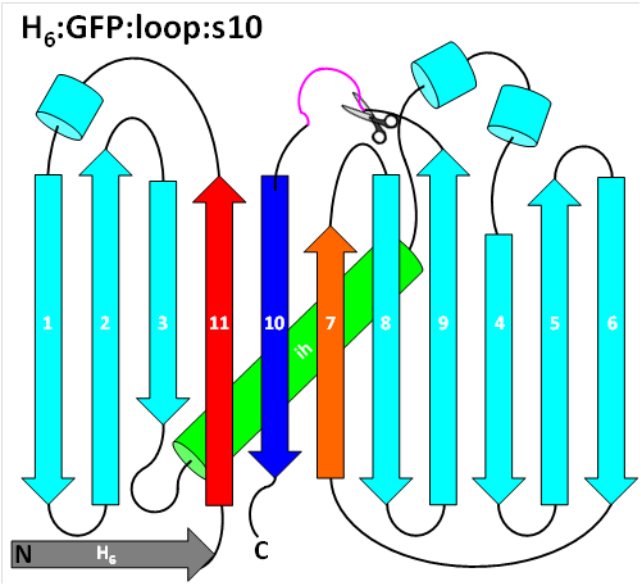


Figure S1: Strand topology of a circular permutant of the strand 10 system; detailed sequences are in Supporting Information S.1. The colored strands 7 (orange), 10 (blue), and 11 (red) have amino acids that make important contacts with the phenol moiety of the chromophore, which is on the central helix (ih, green). The loop containing a proteolytic cleavage site for splitting the protein is shown in magenta. Note that the His-tag is retained in this case, but removed along with strand 10 in H6:s10:loop:GFP (Figure 1B) using the procedure shown in Fig. 2.

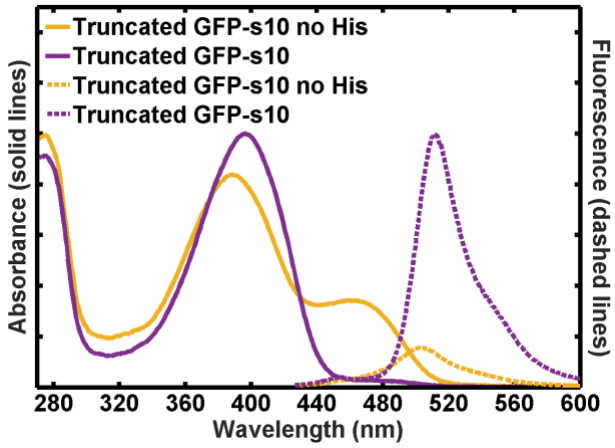


Figure S2: Spectral Characterization of H6:GFP:~~loop:s10~~. Absorption (solid) and fluorescence (dashed) spectra of equal concentrations of H6:GFP:~~loop:s10~~ (purple) vs ~~H6:s10:loop~~:GFP (orange) in lysis buffer at pH 8, normalized to the same scale. Under these conditions, the former species (which retains the His-tag) has a higher pKa than the latter species (which lacks the His-tag), shown by the relatively smaller fraction of chromophore in the deprotonated B-state.

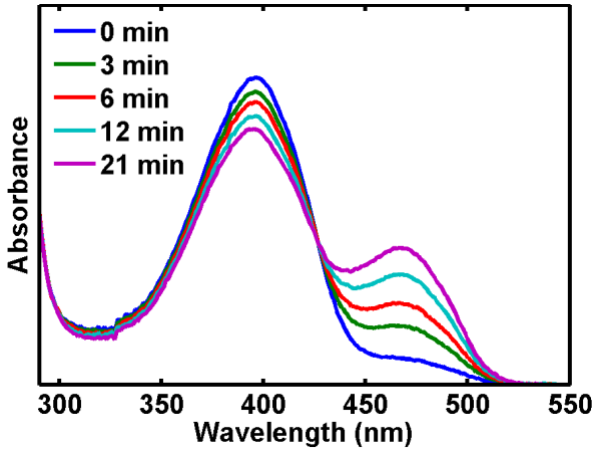


Figure S3: Binding Kinetics of H6:GFP:~~loop:s10~~. 2 µM H6:GFP:~~loop:s10~~ spontaneously reassembles with 60 µM of <u>s10</u>, with k = 2.2 x 10^1 $M^{-1}s^{-1}$. In comparison, the spontaneous association rate of <u>s10</u> with ~~H6:s10:loop~~:GFP is 4.23 x 10^3 $M^{-1}s^{-1}$, which is approximately 190 times faster.

S.5: Characterization of H6:GFP:~~loop:s11~~

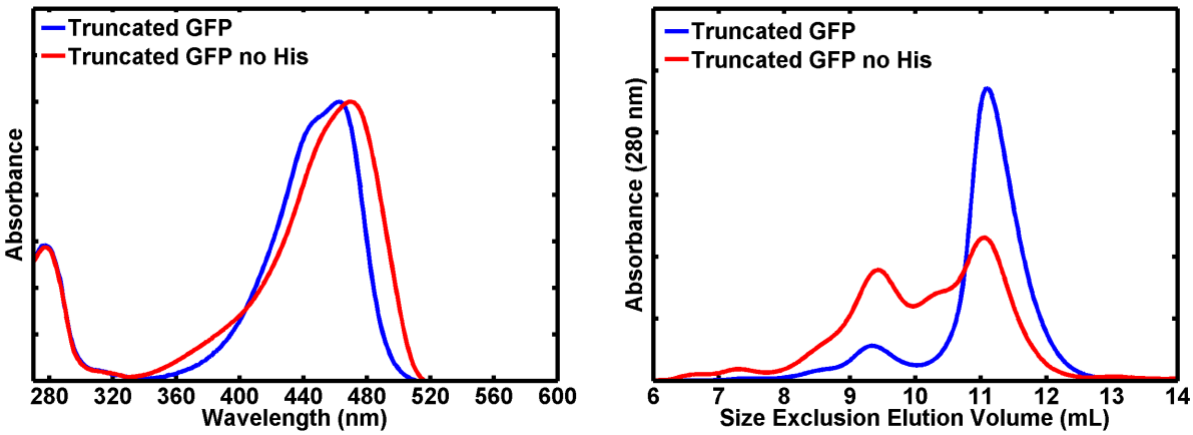


Figure S4: Differences Resulting from Removal of His Tag. Proteolytic cleavage of the His-tag results in a truncated protein (~~H6~~:GFP:~~loop:s11~~ in red) with different properties compared to the truncated protein that retains this sequence (H6:GFP:~~loop:s11~~ in blue). Left: The absorbance spectrum of the species without the His-tag is red shifted with a new λ_{max} of 470 nm compared to 464 nm with the His-tag. Right: Following the truncated protein isolation protocol described in S.2, the size exclusion chromatography trace shows that the protein without the His-tag forms higher order aggregates above ~10 μM. The monomer peak elutes around 11 mL, and the higher order species elutes earlier around 9 mL.

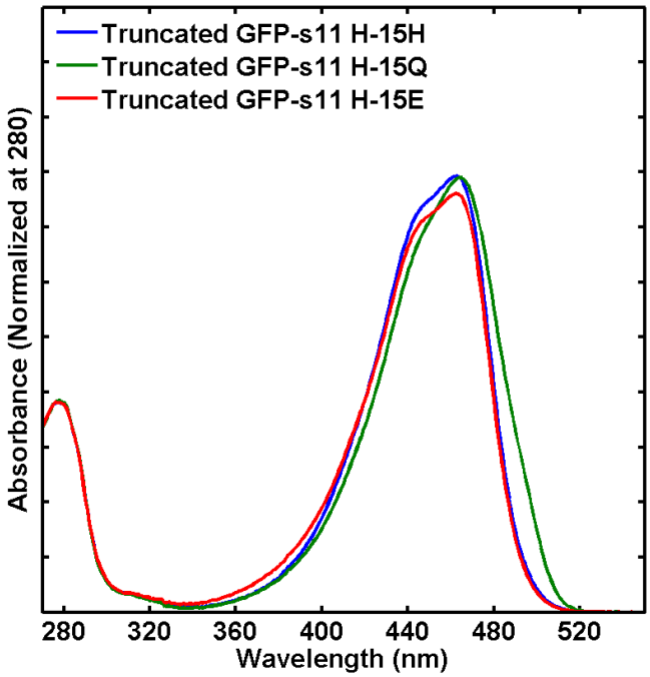


Figure S5: Absorption Spectra of H6:GFP:~~loop:s11~~ H^{-15} Mutants, normalized at 280 nm. Mutations at this residue in the His-tag causes subtle spectral differences in the resulting truncated protein, prepared by the method described in Figure 2.

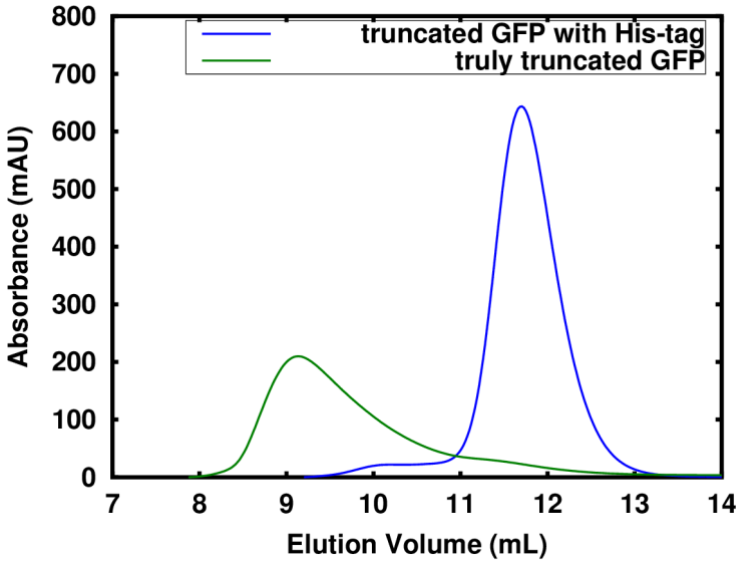


Figure S6: Size exclusion chromatography of truncated vs truly truncated GFP. Following the truncated protein isolation protocol described in S.2, the size exclusion chromatography trace shows that the protein without the His-tag forms higher order aggregates above ~10 μM. The monomer peak elutes around 11 mL, and the higher order species elutes earlier around 9 mL. Note the comparison of the green trace with Figure S4 right, which contains the 206K variant.

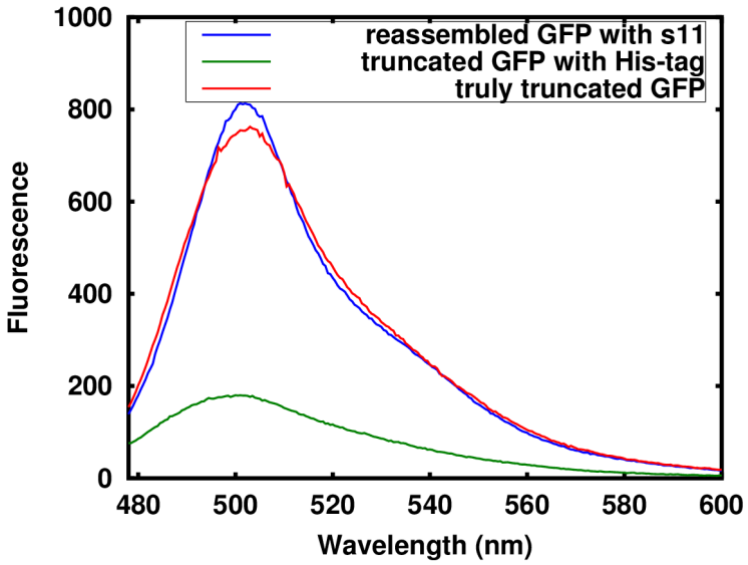


Figure S7: Fluorescence emission spectra of strand 11 system variants. Equal concentrations (100 nM) of each sample were prepared and the fluorescence emission spectra obtained with excitation at 468 nm. The truly truncated ~~H6~~:GFP:~~loop:s11~~ has a fluorescence quantum yield slightly lower than that of the reassembled ~~H6~~:GFP:~~loop:s11~~•s11, while the fluorescence quantum yield of the truncated protein with the His-tag H6:GFP:~~loop:s11~~ is lower by a factor of 3.

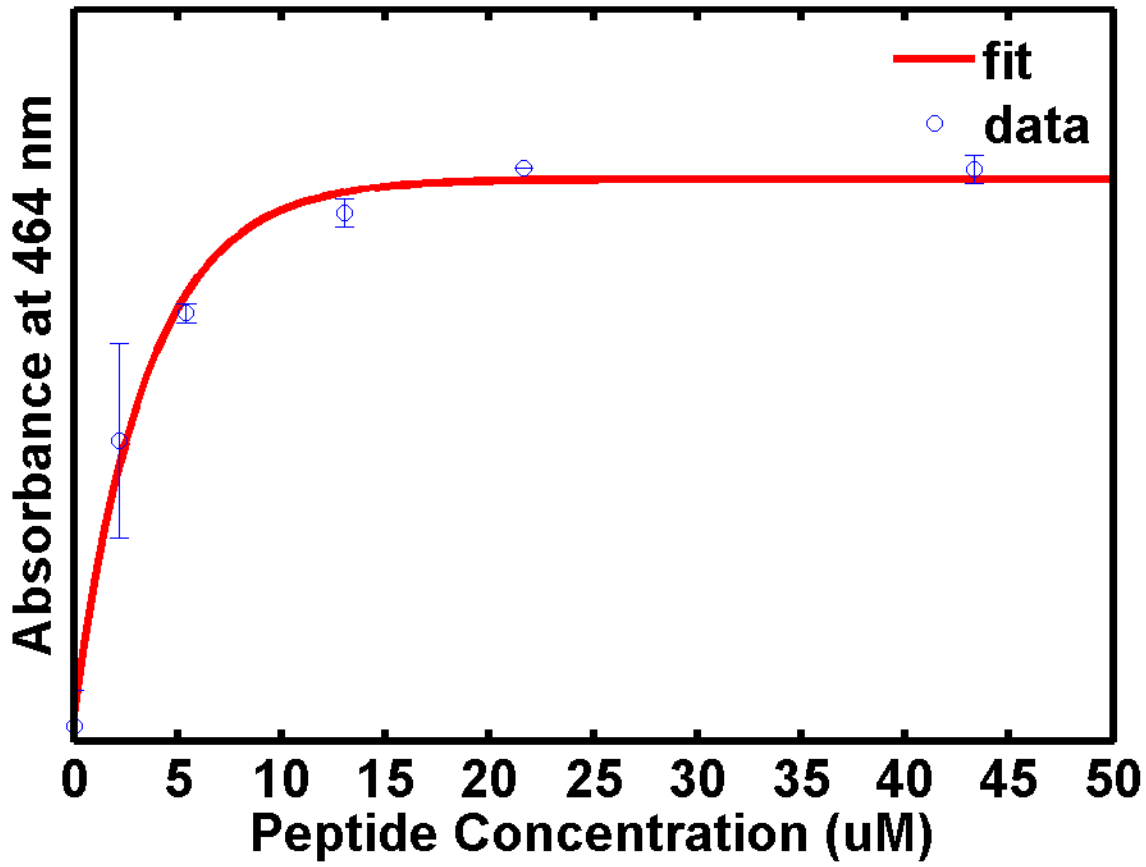


Figure S8: H6 binding isotherm for ~~H6~~:GFP:~~loop:s11~~. 2 µM of H6:GFP:loop:s11 was mixed with different concentrations of H6 at 298K and allowed to reach equilibrium. The equilibrium absorbance at 464nm was plotted as a function of peptide concentration, and fit to a single exponential of the form A*exp(-B*concentration)+C, where A = -.005126, B = .2849, and C = .07237. The K_D for the His-tag peptide is calculated to be 2.4 micromolar, in contrast to the sub-picomolar affinity of the truncated protein to the native strand 11.

S.6 Crystallography

Protein Crystallization

H6:GFP:~~loop:s11~~ and H6:GFP:~~loop:s10~~ were crystallized by hanging drop vapor diffusion. In both cases, protein of concentration ~20 mg/mL was mixed 1:1 with the mother liquor in the initial drop. For H6:GFP:~~loop:s11~~, mixing with mother liquor consisting of 0.1M CHES buffer at pH 9.5 and 1M trisodium citrate generated octahedral crystals after 72 hours of incubation at 277K. The cryoprotectant was the mother liquor with glycerol added to 10 v% and ethylene glycol added to 10 v%. For H6:GFP:~~loop:s10~~, mixing with mother liquor consisting of 0.1M Hepes buffer at pH 7.0 and 0.2M ammonium chloride and 22.5 v% PEG 6000 generated thin rectangular plates after 24 hours of incubation at 298K. The cryoprotectant was the mother liquor with glycerol added to 30 v/v.

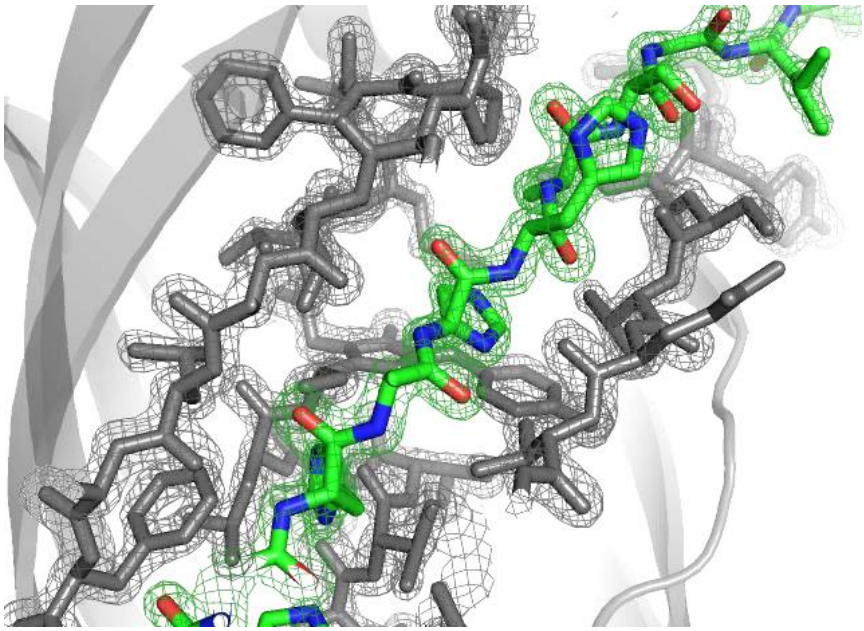
Data Collection and Structure Refinement

The x-ray diffraction data for H6:GFP:~~loop:s11~~ and H6:GFP:~~loop:s10~~ were collected on BL 14-1 and BL 12-2, respectively, at the Stanford Synchrotron Radiation Lightsource[3] (Menlo Park, CA). All data were obtained at 100 K. In the former case, the data processing (indexing, integration, scaling, and merging) was done with HKL2000[4] and converted to .mtz format using CCP4[5]. In the latter case, the data processing was done with XDS[6,7] using the autoxds script[8]. Molecular replacement was carried out in PHENIX[9] using the original Superfolder GFP structure (PDB entry: 2B3P) as the search model. Mutations and cycles of model building and refinement were performed with Coot[10] and PHENIX. The overall fold of the proteins, despite missing their native respective β-strands. were nearly identical to those of Superfolder GFP.

Table S1: X-ray data-collection and refinement statistics

Crystal	H6:GFP:~~loop:s11~~	H6:GFP:~~loop:s10~~
Space group	$P4_3$	$P12_11$
Unit-cell parameters		
a (Å)	55.548	49.369
b (Å)	55.548	67.212
c (Å)	139.935	59.991
α, β, γ (°)	90, 90, 90	90, 98.812, 90
Wavelength (Å)	1.22	0.98
Temperature (K)	100	100
Unique reflections	44426	42227
Resolution range (Å)	29.61-1.72	35.13-1.91
Redundancy	7.3	3.6
Completeness (%)	99.6	98.2
<I/σ(I)>	1.87	2.23
R_{merge}	0.098	0.061
No. of GFP copies per asymmetric unit	2	2
R_{work}/R_{free} (%)	15.79 / 18.39	17.91 / 20.86
R.m.s. Deviations		
Bond Lengths (Å)	0.007	0.011
Bond Angles (°)	0.851	1.107
Ramachandran plot (%)		
Preferred	98.9	98.5
Allowed	1.1	0.9
Outliers	0	0.6
PDB code	6B7R	6B7T

Video S1: Structure Animation of H6:GFP:~~loop:s11~~



S.7: References

(1) Pedelacq, J.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006**, 24, 79–88.

(2) Ward, W. W. *Bioluminescence and Chemiluminescence*, De Luca, M.; McElroy, D.W., eds; Academic, New York, **1981**, 235-242.

(3) 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 (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

(4) Otwinowski, Z., Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276*, 307-326. http://www.hkl-xray.com/

(5) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M.,

Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Cryst.* **2011**, *D67*, 235-242.

(6) Kabsch, W. XDS. *Acta Cryst.* **2010**, *D66*, 125-132.

(7) Kabsch, W. Integration, scaling, space-group assignment and post refinement. *Acta Cryst.* **2010**, *D66*, 133-144.

(8) Gonzalez, A and Tsai, Y. **2010** http://smb.slac.stanford.edu/facilities/software/xds/#autoxds_script

(9) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallographica* **2010**, *D66*, 213-221.

(10) Emsley, P. and Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Cryst.* **2004**, *D60*, 2126-2132.