J|A|C|S COMMUNICATIONS

Published on Web 10/19/2009

Synthetic Control of Green Fluorescent Protein

Kevin P. Kent, Luke M. Oltrogge, and Steven G. Boxer*

Department of Chemistry, Stanford University, Stanford, California 94305-5080

Received July 27, 2009; E-mail: sboxer@stanford.edu

We recently demonstrated that semisynthetic green fluorescent protein (GFP) can be assembled by adding a *synthetic* peptide to a truncated protein that is produced recombinantly.[1] This *in vitro* assembly mimics the *in vivo* assembly of "split GFP" used for protein solubility assays[2] and protein colocalization.[3,4] Unfortunately, recombinant expression of truncated GFPs often results in low sample yield[5] with the chromophore either not formed[1,6] or partially formed.[7] Here we present a method that overcomes both of these limitations and enables efficient synthetic control of all residues, *including* those in the chromophore-containing interior α-helix. This method can be used to produce samples of split GFP for *in vitro* biophysical characterization and should inform efforts for improvement or interpretation of *in vivo* experiments.

Figure 1 displays the 11 β-sheets and the interior α-helix (ih) along with the modifications made to the GFP primary sequence to introduce loops that can be selectively digested with proteases and circular permutation of the C- and N-termini.[8] Figure 2 shows our synthetic strategy beginning with the full-length GFP which is cleaved at the loop insertion, stripped of the small terminal peptide, and reassembled with a synthetic peptide. This method is reminiscent of the preparation of split ribonuclease S, which is generated by proteolysis of ribonuclease with subtilisin,[9] except that in this case a trypsin cleavage site is specifically engineered into the normally trypsin-resistant GFP[10] with loop insertions. After digesting the loop the GFP remains intact and spectrally indistinguishable from the uncut protein. The original strand is removed by denaturation and then replaced by a fully synthetic strand with any desired sequence containing natural or unnatural amino acids.

Due to the manifold distinct protein constructs, we developed the systematic notation illustrated in Figure 2. Anything to the left of the term 'GFP' is on the N-terminal side of the protein, and anything to the right of the term 'GFP' is on the C-terminal side. 'Loop' refers to the sacrificial loop insertions (Figure 1), 's11' refers to the 11th stave of the β-barrel,[11] and 'ih' refers to the interior α-helix. A strike through 'loop' implies the loop was removed with trypsin, and a strike though 'ih' or 's11' implies the original peptide was removed by denaturation and size exclusion.[12] Any synthetic peptide is underlined, and the dot (•) implies that a noncovalent complex has been formed

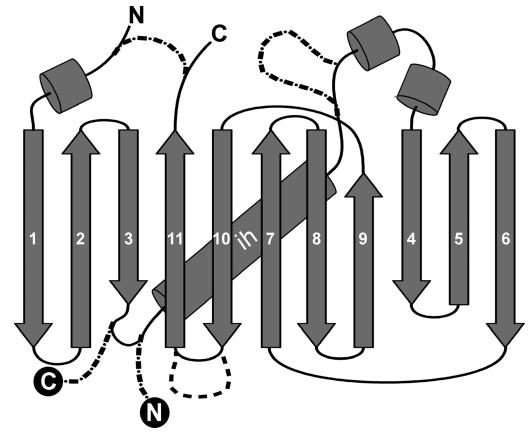

Figure 1. The 11 β-strands and α-helices of GFP represented as arrows and cylinders, respectively. The dashed line shows the loop insertion used to make GFP:loop:s11 (Figure 2A), and the dash-dotted line shows both the circular permutation and the loop insertion to make ih:loop:GFP (Figure 2B). The N and C in the black circles show the new N and C termini after making the circular permutation to get ih:loop:GFP. The loop insertion sequence for both proteins is GTRGSGSIEGRHSGSGS, and the linker added between the native N- and C-termini in ih:loop:GFP is GGTGGS.

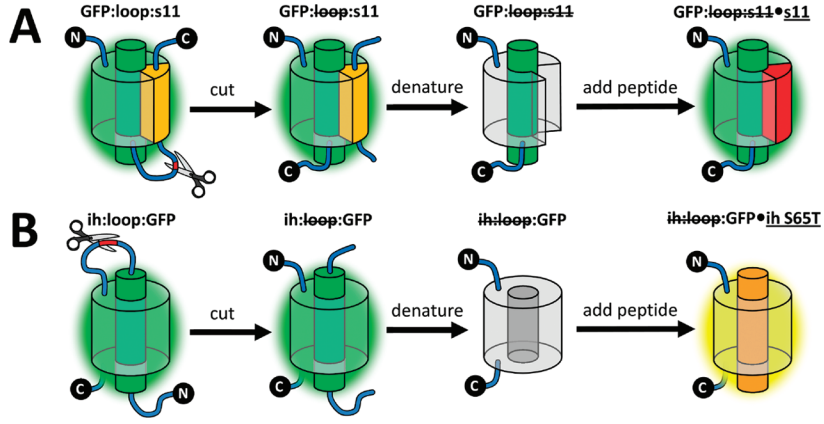

Figure 2. Methods for replacing (A) β-strand 11 (or any of the 11 β-strands by circular permutation, cf. Figure 1) or (B) the interior helix (ih) that contains the 3 amino acids that become the GFP chromophore. In each case, a GFP with a loop insertion containing trypsin cleavage sites before the C or N terminal element that will be removed is formed recombinantly in high yield as a full-length folded protein with the chromophore formed. The steps shown are (1) digest with trypsin; (2) denature with guanidine hydrochloride and isolate the larger piece of GFP by size exclusion; and (3) dilute the larger piece of GFP out of denaturant into a solution containing a synthetic peptide.

10.1021/ja906303f CCC: $40.75

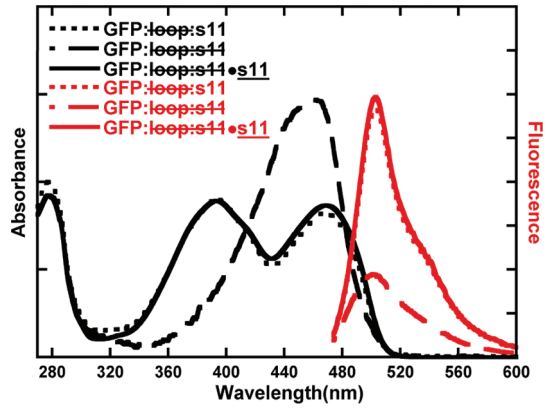

Figure 3. Absorbance and fluorescence spectra of elements in method described in Figure 2A for the removal and replacement of strand 11. GFP:loop:s11, GFP:loop:s11, and GFP:loop:s11•s11 spectra are shown by dotted, dashed, and solid lines, respectively. The spectra of GFP:loop:s11•s11 and GFP with s11 covalently attached are nearly identical[1] (data not shown). The absorbance spectrum of GFP:loop:s11 has a single band in the visible region unlike the parent protein with s11 covalently or noncovalently attached. All spectra are normalized by concentration so that the relative intensities of the absorbance spectra reflect differences in extinction coefficients and the emission spectra relative intensities reflect the difference in extinction coefficient and quantum yield upon excitation at 468 nm.

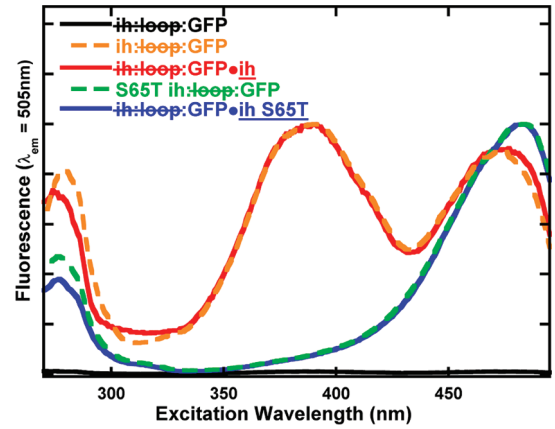

Figure 4. Fluorescence excitation spectra of the elements shown in Figure 2B. The native ih containing the chromophore incorporating serine 65 is removed, yielding a colorless protein (ih:loop:GFP), followed by addition of a synthetic ih or a synthetic ih with the S65T mutation (ih S65T). When the original ih is removed and then reconstituted with a synthetic ih of the same sequence (ih:loop:GFP•ih, solid red), the original spectrum is recovered (ih:loop:GFP, dashed orange). In contrast, if instead it is reconstituted with ih S65T (ih:loop:GFP•ih S65T, solid blue), the resulting excitation spectrum is identical to that of the protein in which the S65T mutation was introduced through mutagenesis (S65T ih:loop:GFP, dashed green). There is minimal fluorescence upon refolding ih:loop:GFP (solid black) in the absence of synthetic peptides.

between the protein and the synthetic peptide preceding and following the dot.

Following removal of s11 and the denaturant (Figure 2A), the absorption of GFP:loop:s11 is quite different from that of the native protein (Figure 3, dashed). Interestingly, this protein is much more fluorescent than expected (Figure 3, dashed red), since the chromophore itself, when outside the protein environment, is nonfluorescent in aqueous solution.[13] This suggests that there is residual structure in GFP:loop:s11[14] that prevents nonradiative decay pathways available to the free chromophore. Consistent with this, the UV circular dichroism spectrum of GFP:loop:s11 is very similar to that of GFP itself (Supporting Information). Upon addition of s11 with the identical sequence, a protein whose absorbance and fluorescence spectra are indistinguishable from the original protein is formed (Figure 3, solid).[15] The method described above can be used to gain synthetic access to any strand in GFP.[11]

Remarkably this strategy also works for the interior helix (Figure 4). ih:loop:GFP has the entire chromophore-containing helix removed. To determine if this "empty barrel"[14] can catalyze chromophore formation in a *synthetic* ih upon reconstitution, ih with the S65T mutation (ih S65T) was introduced to ih:loop:GFP using the scheme in Figure 2B. The fluorescence excitation spectrum of ih:loop:GFP•ih S65T overlays with that of S65T ih:loop:GFP but not with that of ih:loop:GFP, the starting material from which ih:loop:GFP was derived (Figure 4). The conversion of the characteristic absorption and fluorescence from that of the native chromophore with Ser at position 65 to that characteristic of the S65T mutation unambiguously demonstrates that ih:loop:GFP induces chromophore formation in ih S65T. Following the methods described in the Supporting Information, the reconstituted fluorescence develops with a half-life of roughly 1 day, and the maximum fraction reconstituted has thus far been ~20%.[15,16]

This strategy has the advantage of producing high-yield samples with a mature chromophore, unlike previous methods of producing semisynthetic GFPs.[1,6,7] Moreover, we have shown that perhaps the largest perturbation, removing and replacing the interior α-helix, is possible. Replacement of the interior α- helix with a synthetic peptide allows facile introduction of modifications to the chromophore structure using unnatural amino acids or other synthetic elements.

Acknowledgment. This work was supported in part by a grant from the NIH (Grant GM27738). L.M.O. is supported by an ARCS Foundation Stanford Graduate Fellowship.

Supporting Information Available: Protein preparation, mass spectra, and amino acid sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kent, K. P.; Childs, W.; Boxer, S. G. *J. Am. Chem. Soc.* **2008**, *130*, 9664–65.
(2) Cabantous, S.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2005**, *23*, 102–7.
(3) Shyu, Y. J.; Hu, C. D. *Trends Biotechnol.* **2008**, *26*, 622–30.
(4) Ozawa, T. *Anal. Chim. Acta* **2006**, *556*, 58–8.
(5) Degradation or precipitation of truncated GFPs likely causes low yields.
(6) Sakamoto, S.; Kudo, K. *J. Am. Chem. Soc.* **2008**, *130*, 9574–82.
(7) Huang, Y.-M.; Bystroff, C. *Biochemistry* **2009**, *48*, 929–40.
(8) Despite these modifications, the *E. coli* expression yields of the full-length protein containing the chromophore remain high (>100 mg/L). The parent GFP (see Supporting Information for full sequences) is derived from a split GFP developed for solubility assays,[2] except that position 65 is reverted back to serine, which provides an optical probe for the integrity and photochemistry of GFP[17,18] (see Figure 4).
(9) Richards, F. M.; Vithayathil, P. J. *J. Biol. Chem.* **1959**, *234*, 1459.
(10) Sniegowski, J. A.; Phail, M. E.; Wachter, R. M. *Biochem. Biophys. Res. Commun.* **2005**, *332*, 657–63.
(11) Any of the 11 β-strands can be removed by circular permutation, but we use the natural C-terminal 11th strand for illustration. We used this method to remove s7; this will be reported in subsequent papers.
(12) GFP $1-10^{mat}$ referred to previously[1] is GFP:loop:s11 in this scheme.
(13) Niwa, H.; Inouye, S.; Hirano, T.; Matsuno, T.; Kojima, S.; Kubota, M.; Ohashi, M.; Tsuji, F. I. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13617–22.
(14) The structures of GFP:loop:s11 and ih:loop:GFP are not yet known.
(15) Kinetics of reassembly of ih and s11 with ih:loop:GFP and GFP:loop:s11, incorporation of unnatural amino acids, and rate of chromophore formation in ih will be described in subsequent work.
(16) We note that native chemical ligation could be used to fuse ih to the N-terminus of ih:loop:GFP.
(17) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509–44.
(18) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8362–7.

JA906303F