Supporting Information for:
Synthetic Control of Green Fluorescent Protein (GFP)

Kevin P. Kent, Luke M. Oltrogge, and Steven G. Boxer*
Department of Chemistry, Stanford University, Stanford, California, 94305-5080

Contents:

Protein Preparation:

Gene Construction
The genes for GFP:s11 and ih:GFP (the sequences were designed without the loop inserted) were ordered from Genscript and received in the plasmid pUC57. The genes were removed from pUC57 with restriction enzymes (GFP:s11 was removed with NdeI and XhoI, and ih:GFP was removed with NcoI and XhoI) and ligated into pET-15b. The loop insertion was performed to make GFP:loop:s11 and ih:loop:GFP by digesting the genes in pET-15b with restriction enzymes that cut the DNA close to the desired loop insertion site, and then using DNA ligase to insert synthetic oligonucleotides containing the loop sequence were inserted using DNA ligase (Invitrogen). DNA sequencing (Elim Biopharmaceuticals) confirmed all constructs.

Protein Purification
GFP:s11, ih:GFP, GFP:loop:s11, and ih:loop:GFP were expressed from the pET-15b vectors in BL21(DE3) cells (Stratagene). The cells were induced with IPTG (0.25g/L) at OD 0.6 and then incubated for 4 hours at 37°C. The cells were spun down, resuspended in lysis buffer (50mM HEPES, 300mM NaCl, and 10% glycerol at pH 8), and lysed with a homogenizer. The cell lysate was spun down, and the supernatant was poured onto a Ni:NTA column (Qiagen). The column was washed with the lysis buffer with 20mM imidazole, and then the proteins were eluted with the lysis buffer with 200mM imidazole. The proteins were then dialyzed into the anion exchange A buffer (20mM bis-tris propane, 20mM NaCl at pH 9), and bound to an anion exchange column. The proteins were eluted by running a gradient from the A buffer to a buffer with 200mM salt (20mM bis-tris propane, 200mM NaCl at pH 9). Liquid chromatography in tandem with electrospray mass spectrometry (Waters 2795 HPLC and ZQ single quadrupole MS) was used to check the purity and identity of each of the proteins. Protein sequences inferred from DNA sequencing, mass spectrometry, and known trypsin cut sites of GFP:s11, ih:GFP, GFP:loop:s11, ih:loop:GFP, GFP:~~loop:s11~~, and ~~ih:loop~~:GFP are shown below.

Loop Cleavage
The lyophilized trypsin (bovine pancreas trypsin, Sigma) used for digestions was weighed out and dissolved in 1 mM HCl immediately prior to use. GFP:loop:s11 and ih:loop:GFP were digested in trypsin in different conditions. 40µM GFP:loop:s11 was digested with 100 units of trypsin per mL solution in trypsin buffer (50mM Tris, 20mM CaCl, pH 8.0) for three hours, then 1mM phenylmethanesulphonylfluoride (Amersham) was added to quench the reaction, and subsequent anion exchange removed the deactivated trypsin. ih:loop:GFP was bound to a Ni:NTA column and digested on-column with 100 units of trypsin per mg of ih:loop:GFP for 10 minutes with periodic mixing. The column was washed with lysis buffer plus 20mM imidazole, and then ih:~~loop~~:GFP was eluted with lysis buffer plus 200mM imidazole. Both cut proteins were denatured in 6M guanidine hydrochloride with 5mM DTT, and size exclusion in lysis buffer plus 6M guanidine hydrochloride was performed to separate GFP:~~loop:s11~~ and ~~ih:loop~~:GFP from s11 and ih, respectively.

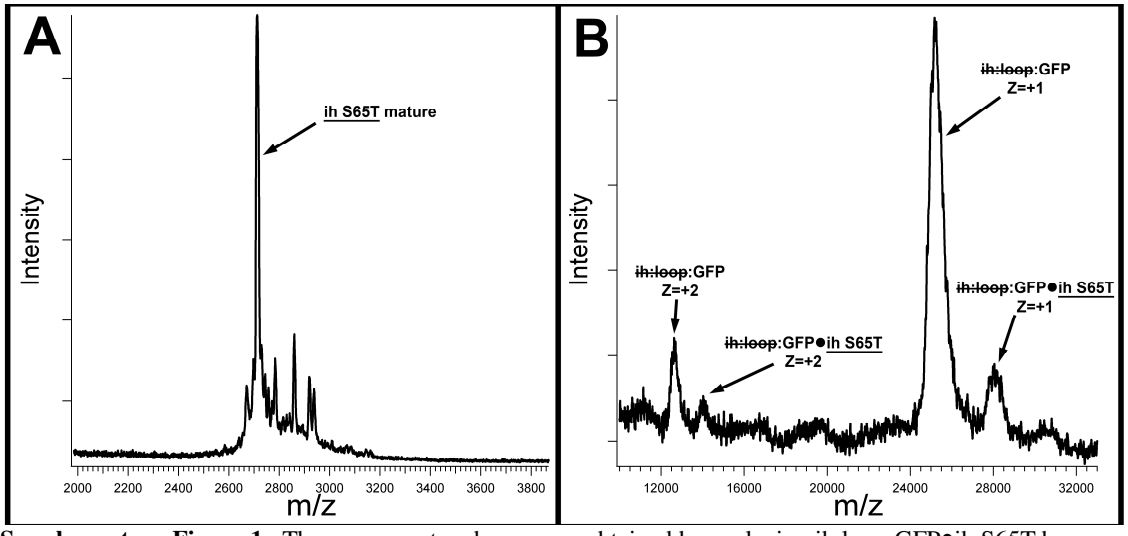
Synthetic Peptides
The peptides s11 and ih were synthesized using Fmoc chemistry on an 431A Applied Biosystems Peptide Synthesizer. ih S65T was purchased commercially (Elim Biopharmaceuticals). All peptides were purified by HPLC (Shimadzu) and the mass and purity were verified by liquid chromatography in tandem with electrospray mass spectrometry (Waters 2795 HPLC and ZQ single quadrupole MS) and/or MALDI (Perseptive Voyager-DE RP Biospectrometry MALDI-TOF mass spectrometer).

Mass Spectrometry Instrumentation and Data:

The masses of GFP:~~loop~~:s11 and ih:~~loop~~:GFP were obtained by direct injection on a Micromass nanoESI API-US Quadrupole Time-of-Flight mass spectrometer (supplementary Table 1), the reconstituted ~~ih:loop~~:GFP●ih S65T was analyzed with a Perseptive Voyager-DE RP Biospectrometry MALDI-TOF mass spectrometer (supplementary Table 1, supplementary Figure 1) and a Waters 2795 HPLC and ZQ single quadrupole MS (supplementary Table 1). ih S65T mature in supplementary table 1, and supplementary figure 1 shows that the peptide has been exposed to ~~ih:loop~~:GFP. The masses are consistent with the amino acid compositions shown on page S4 and the mass loss of approximately 20 g/mol due to chromophore formation[1]. All of the samples for mass spectrometry exchanged were into double deionized water.

Protein	*Expected Mass (g/mol)*	*Observed Mass (g/mol)*
GFP:~~loop:s11~~ (ESI)	26210.3	26209.3
~~ih:loop~~:GFP (ESI)	24769.5	24768.5
ih S65T mature (MALDI)	2707.2	2712.7
ih S65T (MALDI)	2727.2	2726.7
ih S65T mature (ESI)	2707.2	2707.5

Supplementary Table 1. This table shows a comparison between the observed and expected masses of proteins and peptides described in this work.

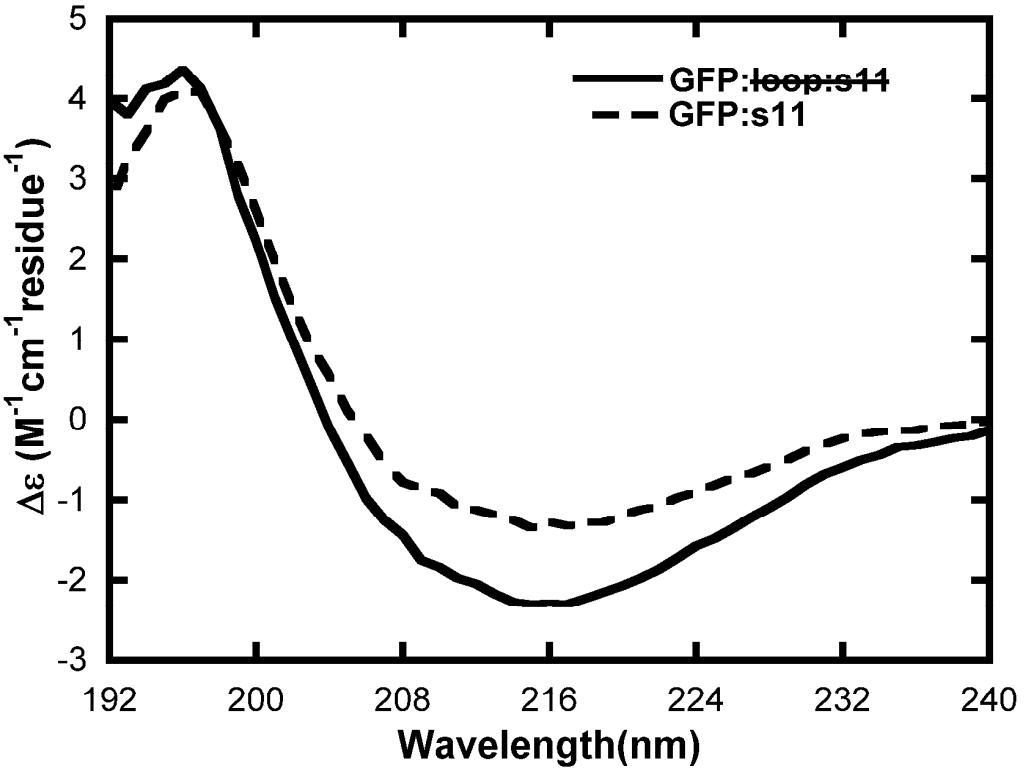


Supplementary Figure 1. The mass spectra above were obtained by analyzing ih:loop:GFP●ih S65T by MALDI. Panel A shows the ih S65T peptide, with a mass loss approximately consistent with chromophore formation. Panel B shows the mass for ih:loop:GFP, and, surprisingly, for ih:loop:GFP●ih S65T by MALDI, underscoring the high stability of this reassembled complex with the interior helix.

Spectroscopy Instrumentation, Methods, and Spectra:

All fluorescence data were obtained on a Fluorolog 3 from Jobin Yvon Horiba fluorimeter, and all absorbance data were obtained on either a Perkin-Elmer Lambda 25 or a Cary 6000i UV/Vis spectrometer. All concentrations were determined by denaturing the protein in 0.1M NaOH and comparing to the known extinction coefficient of the chromophore ($44100 M^{-1}cm^{-1}$ at 448nm)[2]. To obtain the emission spectra in Figure 3, GFP:~~loop:s11~~ was diluted out of denaturant to a final concentration of 1μM GFP:~~loop:s11~~ both into lysis buffer, and into lysis buffer with a ten-fold excess of s11 and incubated at room temperature for two days (all samples had fresh 1mM DTT). The GFP:~~loop:s11~~●s11 sample eluted as a single peak on subsequent anion exchange and in a different position than GFP:~~loop:s11~~, suggesting that the all of the GFP:~~loop:s11~~ in solution was in a non-covalent complex with s11. The absorbance spectrum in Figure 3 was obtained after anion exchange and concentration of the GFP:~~loop:s11~~ and GFP:~~loop:s11~~●s11 samples. The excitation spectra of the reconstituted proteins in Figure 4 were obtained by diluting ~~ih:loop~~:GFP out of denaturant to a final concentration of 1μM into lysis buffer, 10μM ih S65S in lysis buffer, or 10μM ih S65T in lysis buffer, respectively (all samples had fresh

0.5mM DTT). Fluorescence excitation spectra taken periodically after mixing were used to estimate the half-life, and end point of reassembly of ~~ih:loop~~:GFP with ih and ih S65T. The CD spectra in Supplementary Figure 2 were obtained on a CD Spectrometer Model 202-01 from Aviv. The CD spectra were measured in phosphate buffer (10mM NaH_2PO_4 and 10 mM NaCl , pH 8.0) for GFP:s11 (GFP:~~loop:s11~~ fused with strand 11 without a loop inserted) and GFP:~~loop:s11~~ with the two endogenous cysteines mutated to C48S and C70A.The cysteine free mutants were used for this experiment because reducing agents necessary for keeping GFP:~~loop:s11~~ in the active state (meaning able to reassemble with s11) interfered with UV CD. The absorbance and fluorescence spectra of GFP:~~loop:s11~~ without cysteine and GFP:~~loop:s11~~ are nearly identical (data not shown). Concentrations for normalization were calculated by denaturing the proteins in base and using the known extinction coefficient at 448nm[2].



Supplementary Figure 2. The CD spectrum of GFP:s11 (GFP:~~loop:s11~~ fused GFP 11 without a loop inserted between them) and GFP:~~loop:s11~~ (with C48S and C70A mutations). The fitting program K2D2[3] predicted 14% α helix and 40% β strand structure for whole GFP, and 21% α helix and 30% β strand structure for GFP:~~loop:s11~~.

Amino Acid Sequences:

GFP:s11

MGSSHHHHHHSSGLVPGGSHMGGTSSKGEELFTGVVPILVELDGD
VNGHKFSVRGEGEGDATIGKLTLKFICTTGKLPVPWPTLVTTLSYGVQCFSRYPDHMKRH
DFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKL
EYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYL
STQTVLSKDPNEKGTRDHMVLHEYVNAAGIT

GFP:loop:s11 and GFP:~~loop:s11~~*

MGSSHHHHHHSSGLVPGGSHMGGTSSKGEELFTGVVPILVELDGD
VNGHKFSVRGEGEGDATIGKLTLKFICTTGKLPVPWPTLVTTLSYGVQCFSRYPDHMKRH
DFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKL

EYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYL
STQTVLSKDPNEK▼GTRGSGSIEGRHSGSGSRDHMVLHEYVNAAGITHGMDELYKGSGGT

*the observed proteolytic cleavage site is indicated by "▼"

s11

RDHMVLHEYVNAAGIT

ih:GFP

MGHHHHHHSSGGKLPVPWPTLVTTLSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKY
KTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDG
SVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYGGTGG
SASQGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFICTT

ih:loop:GFP and ~~ih:loop~~:GFP*

MGHHHHHHSSGGKLPVPWPTLVTTLSYGVQCFSRYGTRGSGSIEGR▼HSGSGSPDHMKRHDFFKSAMP
EGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQK
NGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMVLLEFVTA
AGITHGMDELYGGTGGSASQGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFICTT

*the observed proteolytic cleavage site is indicated by "▼"

ih

GKLPVPWPTLVTTLSYGVQCFSRY

ih S65T

GKLPVPWPTLVTTLTYGVQCFSRY

References:

(1) Wachter, R. M. *Acc. Chem. Res.* **2007**, *40*, 120–7.

(2) Ward, W. W. *Bioluminescence and chemiluminescence* **1981**, (DeLuca, M. A., and McElroy, W. D., eds), 235-42, Academic Press, New York.

(3) Perez-Iratxeta, C. and Andrade-Navarro, M. A. *BMC Structural Biology* **2008**, *8*, 25-30.