Supporting Information for:

Deconstructing Green Fluorescent Protein (GFP)

Kevin P. Kent, William Childs, and Steven G. Boxer*

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

Contents

Protein Preparation S2

Mass Spectrometry Instrumentation and Data S3

Spectroscopy Instrumentation and CD Spectra S5

Amino Acid Sequences S6

References S6

Protein Preparation:

The gene for GFP 1-10 from Geoff Waldo and coworkers[1] was placed in a pET-15b plasmid, and position 65 was mutated from threonine to serine using a QuikChange mutagenesis kit (Stratagene). This mutated, N-terminal histidine tagged GFP 1-10 will now be referred to as GFP 1-10; full sequences for all proteins and peptides are given at the end of this document. GFP 1-10 was expressed in *E. coli* on a pET-15b vector (Novagen), isolated from inclusion bodies[1], crudely purified using a Ni-NTA affinity column (Qiagen), and purified by anion exchange chromatography using a P900 FPLC (Amersham Biosciences). The Ni:NTA, and anion exchange purification steps were performed in denaturing conditions (8 M urea and 5 mM DTT). GFP 11 and GFP 11 E222Q were synthesized using FastMoc Chemistry (Applied Biosystems) and purified by HPLC (Shimadzu). GFP 1-10 was diluted out of denaturing buffer with a 25 times volume of 50 mM Tris HCl, 100mM NaCl, and 10% glycerol at pH 7.4, and then mixed with a two-fold excess of synthetic GFP 11. The maturation process was performed open to the atmosphere at room temperature for two days. Reassembled GFP (GFP 1-10 + GFP 11) was then purified by anion exchange chromatography. The synthetic gene for whole GFP (Genscript) was also expressed in *E. coli* on a pET-15b vector. Whole GFP was isolated from the soluble lysate, and purified by Ni:NTA chromatography and anion exchange chromatography in non-denaturing conditions. GFP $1\text{-}10^{mat}$ was isolated by denaturing reassembled GFP in 0.1 M NaOH, exchanging the proteins into a solution containing 8 M urea, separating GFP 11 from GFP $1\text{-}10^{mat}$ using Ni^{2+} affinity chromatography, and diluting GFP $1\text{-}10^{mat}$ into a 25 times volume of 50 mM Tris HCl, 100 mM NaCl, and 10% glycerol at pH 7.4. To obtain reassembled GFP E222Q, a 2 fold

excess of GFP 11 E222Q was added to GFP 1-10^{mat}, and the protein was purified by anion exchange chromatography. All plasmids were sequenced by Elim Biopharmaceuticals, and the purity of GFP 11, GFP 11 E222Q, GFP 1-10, GFP 1-10^{mat}, reassembled GFP, reassembled GFP E222Q, and whole GFP were verified by LCMS (mass spectrometer was a ZQ single quadropole MS). All solvent exchanges for experiments in the text and supplementary materials were performed with 10000MW Amicon centrifugal filters.

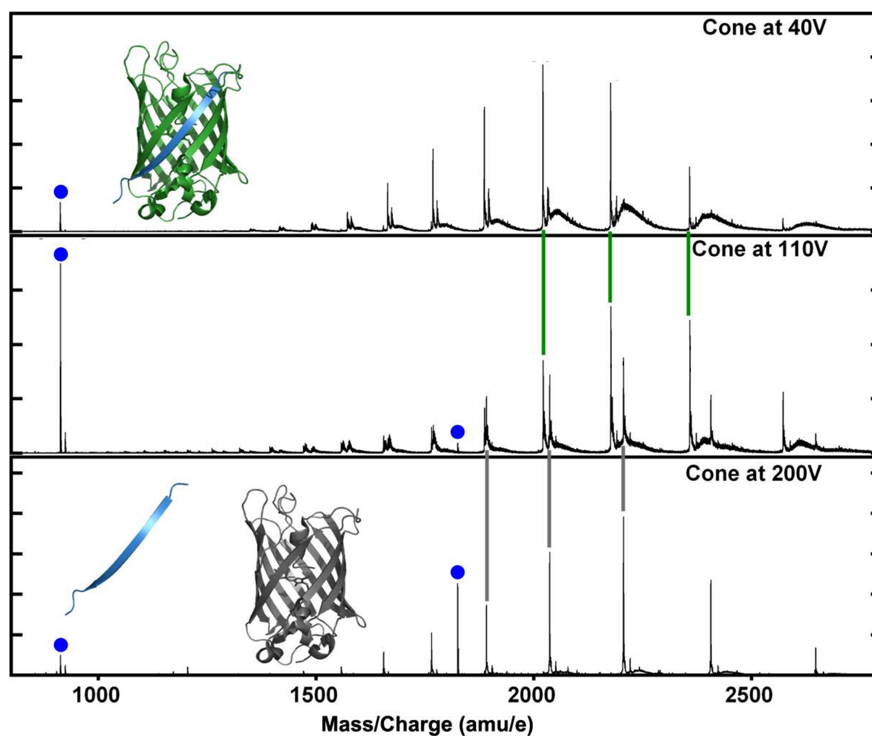
Mass Spectrometry Instrumentation and Data:

The masses of GFP 1-10, GFP 11, and whole GFP were obtained by a ZQ single quadropole MS, and the mass of reassembled GFP was obtained by directly electrospraying the sample into a Micromass Q-TOF hybrid quadropole-time of flight mass spectrometer with an Advion NanoMate nano-electrospray robot (supplementary Table 1). The mass of the complex (reassembled GFP) was obtained by keeping the cone voltage low enough (40 V) to avoid dissociation of GFP 1-10^{mat} and GFP 11, while the mass of GFP 1-10 ^{mat} from dissociated reassembled GFP was obtained by raising the cone voltage to 200 V (see Supplementary Figure 1). The masses are consistent with the amino acid compositions; the mass loss of approximately 20 g/mol in GFP 1-10^{mat} following its interaction with GFP 11 is consistent with the mass loss observed for chromophore formation in wtGFP (see supplementary table 1). The observation that this mass loss occurs only after addition of GFP 11 demonstrates that the chromophore does

not form in GFP 1-10. All of the samples for mass spectrometry were dissolved or exchanged into double deionized water.

Sample	Mass (g/mol)
GFP 11	1826 (± 1)
GFP 1-10	26486 (± 2)
Reassembled GFP	28292 (± 1)
GFP 1-10^{mat}	26467 (± 1)
Whole GFP	28275 (± 2)

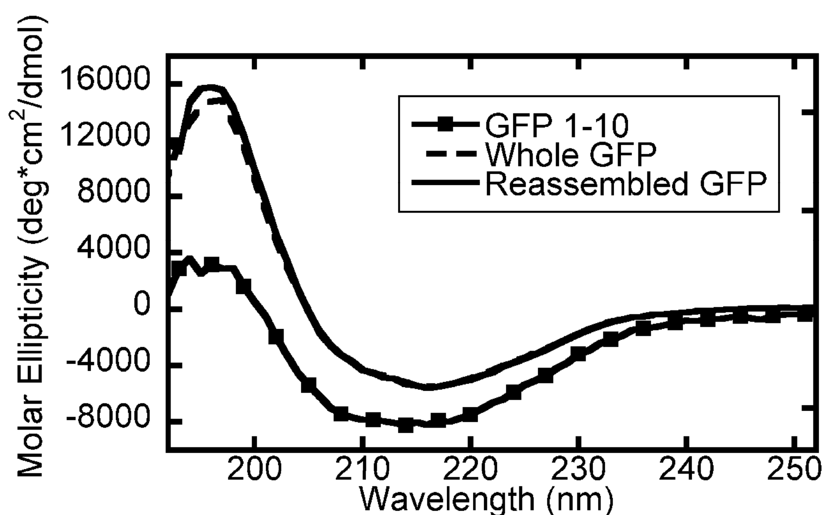
Supplementary Table 1. Mass spectral data for GFP 1-10 before and after interacting with GFP 11 (GFP 1-10 and GFP 1-10^{mat}, respectively), and the complex of GFP 1-10^{mat} + GFP 11 (reassembled GFP). The mass difference of approximately 20 g/mol between GFP 1-10^{mat} and GFP 1-10 is consistent with chromophore formation. The mass difference of approximately 18 g/mol between whole GFP and reassembled GFP is due formation of a peptide bond between GFP 1-10 and GFP 11 by condensation.



Supplementary Figure 1. Mass spectra of reassembled GFP at varying cone voltages. The masses of reassembled, GFP 1-10^{mat}, and GFP 11 are shown in Supplementary Table 1. The green lines indicate mass/charge peaks of reassembled GFP, the grey lines indicate mass/charge peaks of GFP 1-10^{mat}, and the blue dots represent mass/charge peaks of GFP 11.

Spectroscopy Instrumentation and CD Spectra:

Absorbance spectra were obtained on a Perkin-Elmer Lambda 25 UV/Vis spectrometer, fluorescence spectra on a Fluorolog 3 from Jobin Yvon Horiba, and CD spectra (Supplementary Figure 2) on a CD Spectrometer Model 202-01 from Aviv. The time-resolved fluorescent decays were obtained by fluorescence upconversion, performed largely as described previously[3,4]. The spectra shown in Figures 2 and 3 were obtained with the proteins in pH 6.5, 50 mM MES, and 100 mM NaCl. The absorbance spectra in Figure 4 were obtained with the proteins in pH 8, 50 mM HEPES, 300 mM NaCl, and 10% glycerol. The CD spectra for whole and reassembled GFP in Supplementary Figure 2 were obtained with the proteins in 10 mM NaH_2PO_4, 10 mM NaCl at pH 6.5, and the GFP 1-10 spectrum was obtained in the same solvent supplemented with 10% glycerol.



Supplementary Figure 2. Comparison of UV CD spectra for whole GFP, GFP 1-10 prior to addition of GFP 11 and reassembled GFP. The CD spectra of whole and reassembled GFP are indistinguishable, consistent with absorption, fluorescence and time-resolved fluorescence data presented in the main body of the paper. Substantial differences are observed for GFP 1-10 suggesting that it is considerably less structured than whole or reassembled GFP as expected.

Amino Acid Sequences:

Bold red lettering denotes position 222.
Bold green lettering denotes the three amino acids that make up the chromophore.

GFP 11 (identical to the amino acid sequence of GFP 11 M3 in previous work)[5]

RDHMVLH**E**YVNAAGIT

GFP 11 E222Q:

RDHMVLH**Q**YVNAAGIT

GFP 1-10:

GSSHHHHHHSSGLVPRGSHMGGTSSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKL
TLKFICTTGKLPVPWPTLVTTL**SYG**VQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYK
TRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKGT

Whole GFP:

GSSHHHHHHSSGLVPRGSHMGGTSSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKL
TLKFICTTGKLPVPWPTLVTTL**SYG**VQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYK
TRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKGTRDHMVLH**E**YVNAAGIT

References:

(1) Cabantous, S.; Waldo, G., S. *Nat. Meth.* **2006**, *3*, 845-54.
(2) Wachter, R., M. *Acc. Chem. Res* **2007**, *40*, 120 -7.
(3) Chattoraj, M.; King, B., A.; Bublitz, G., U.; Boxer, S., G. *PNAS* **1996**, *93*, 8362-7.
(4) Stanley, R., J.; Boxer, S., G. *J. Phys. Chem.* **1995,** *99*, 859–63.
(5) Cabantous, S.; Terwilliger, T., C.; Waldo, G., S. *Nat. Biotech.* **2005**, *23*, 102-7.