CHAPTER 2

Evolution of a Fluorinated Green Fluorescent Protein

This chapter appeared as a paper in Tae Hyeon Yoo, A. James Link, and David A. Tirrell, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13887-13890.

Abstract

The fluorescence of bacterial cells expressing a variant (GFPm) of the green fluorescent protein (GFP) was reduced to background levels by global replacement of the Leu residues of GFPm by 5,5,5-trifluoroleucine. Eleven rounds of random mutagenesis and screening via fluorescence-activated cell sorting yielded a GFP mutant containing 20 amino acid substitutions. The mutant protein in fluorinated form showed improved folding efficiency both *in vivo* and *in vitro*, and the median fluorescence of cells expressing the fluorinated protein was improved approximately 650-fold in comparison to that of cells expressing fluorinated GFPm. The success of this approach demonstrates the feasibility of engineering functional proteins containing many copies of abiological amino acid constituents.

Introduction

The properties of proteins are strikingly susceptible to engineering via rational and combinatorial strategies.[1-3] In chemical terms, however, the engineering strategies employed to date are remarkably conservative, in that they employ a common set of twenty "canonical" amino acid constituents. Recent work from several laboratories has shown that a substantially expanded repertoire of non-canonical amino acids can be

incorporated into recombinant proteins, either by amino acid replacement or by nonsense or frameshift suppression.[4-6] This work raises intriguing questions about the prospects for creating novel proteins by recruitment of new amino acid constituents. In particular, global replacement of one of the canonical amino acids by a non-canonical analog would be expected to cause marked changes in protein structure, dynamics, and function. In principle, such changes might be advantageous, and might lead to proteins of enhanced stability or reactivity.[7-11] It seems likely, however, that global replacement will in many cases compromise protein folding and function,[11-13] because both protein structure and protein folding pathways have evolved in a context defined by the canonical amino acid side chains.

Here we examine "re-evolution" of GFP after global replacement of its Leu residues by 5,5,5-trifluoroleucine (Tfl). Introduction of fluorinated amino acids into coiled-coil and four-helix bundle proteins has been shown to cause substantial increases in the stability of such proteins with respect to thermal and chemical denaturation, as anticipated from the hyper-hydrophobic character of fluorocarbons and the critical role of hydrophobic interactions in stabilizing native protein folds. [7-9, 14, 15] On the other hand, global replacement of Leu by Tfl destabilizes the more complex globular protein chloramphenicol acetyltransferase.[13]

GFP has been thoroughly studied in terms of structure, spectral properties, stability, and folding behavior[15-28] and is an attractive target for evolutionary studies, because its inherent fluorescence permits high-throughput analysis and screening of mutant libraries via fluorescence-activated cell sorting. GFP contains 238 amino acid residues, 19 of which are Leu. Global replacement of Leu by Tfl causes misfolding of GFP and loss of fluorescence emission. Here we show that re-evolution of the fluorinated protein restores proper folding and yields a fluorinated variant that exhibits physical and spectroscopic behavior nearly identical to that of wild-type GFP.

Results and Discussion

Two mutations (S65G and S72A)[19] were introduced into cycle-3 GFP (Q80R, F99S, M153T, and V163A)[20] to render its spectral properties appropriate for cell sorting (Fig. 2.1). The resulting GFP variant (GFPm) was expressed in media depleted of Leu and supplemented with 1 mM Tfl.[7] The protein was purified under denaturing conditions, and the extent of replacement of Leu by Tfl was determined to be 78% by matrix-assisted laser desorption-ionization mass spectrometry (MALDI-MS) (Fig. 2.2) and 77% by amino acid analysis (AAA). Replacement of Leu by Tfl caused a reduction of more than 500-fold in the median fluorescence of cells expressing the fluorinated protein (Fig. 2.3

and Table 2.1), such that the observed fluorescence was comparable to that characteristic of cells lacking the gene coding for GFP (data not shown).

GFP fluorescence is produced by a chromophore formed by a spontaneous reaction involving residues 65-67 in the interior of the β-barrel structure; the chromophore forms only after the protein folds into a nearly native conformation.[17] To determine whether the reduction in fluorescence observed upon introduction of Tfl into GFPm is caused by aberrant protein folding or by inhibition of chromophore formation within a properly folded protein framework, the solubility of the expressed proteins was examined. Although nearly all of the GFPm made with Leu in the medium (GFPm-L) was present in the cell lysate in soluble form, that made with Tfl (GFPm-T) was almost entirely insoluble (Fig. 2.4B and 2.4C). This result suggests that the observed reduction in fluorescence is most probably a consequence of misfolding and aggregation of the fluorinated form of GFPm.

To determine whether it is possible to recover properly folded, fluorinated variants of GFP, we applied a directed evolution approach involving randomization of the sequence of GFPm by error-prone PCR, expression of mutants in media depleted of Leu and supplemented with Tfl, and selection of highly fluorescent mutants by cell sorting. During the initial rounds of screening under these conditions, however, no improvement

of fluorescence was observed; clones tested after several rounds of sorting exhibited only background fluorescence. The low fluorescence of cells expressing fluorinated proteins (which was comparable to the background autofluorescence of *Escherichia coli*) hindered the use of media devoid of Leu as the starting point for directed evolution. Instead, we adopted a procedure in which, for the first few rounds of evolution, a small amount of Leu was added to the medium during protein expression. Several different concentrations of Leu were tested for the first library (Fig. 2.5); 45 μM Leu was used for protein expression. After two rounds of sorting, the cell population was enriched in fluorescent variants (Fig. 2.6), and four of seven randomly selected clones showed enhanced fluorescence (data not shown). Using the plasmid DNA from this pool of sorted cells, we prepared a second randomized library. The Leu concentration was reduced during expression of the second-generation library (Table 2.2), and from the fourth generation forward, we added no Leu to the expression medium (Table 2.2).

Eleven rounds of directed evolution yielded a variant of GFP (11.3.3) containing 20 amino acid substitutions (Table 2.3. Clones are designated by three digits. The first is the library generation number, the second is the number of times the library was sorted, and the third is the number of the clone isolated from the library). The median fluorescence of cells that express this variant in fluorinated form (11.3.3-T) is increased

≈650-fold over that of cells expressing GFPm-T and higher than that of cells expressing GFPm-L (Fig. 2.3 and Table 2.1). Two more mutants, 4.2.2 from the fourth-generation library and 8.3.3 from the eighth-generation library, were chosen for comparison (Fig. 2.3 and Table 2.3). Although we found a significant increase in the fluorescence of cells between generations four and eight, only minor improvements were observed in the last three rounds of evolution (compare 8.3.3 and 11.3.3). The fluorescence of cells that express GFP variants in media containing the twenty canonical amino acids also increased, as shown by comparing the median fluorescence of cells expressing 11.3.3-L with that of cells expressing GFPm-L (Fig. 2.3B).

The increase in cell fluorescence can be attributed to an increase in the amount of properly folded, soluble, functional protein in the cell as well as to changes in the spectral properties of the protein. The amount of soluble protein depends both on the expression level and on the fraction of the expressed protein that remains soluble. Both the expression level and the soluble fraction increased as the evolution progressed (Fig. 2.4). As mentioned previously, the soluble fraction of GFPm-T was negligible; in contrast, the majority of 11.3.3-T was present in soluble form. The photophysical properties of the purified proteins were also improved by evolution (Table 2.4). GFPm-T was not characterized because it formed inclusion bodies, and our attempts to refold the denatured

protein were unsuccessful. 11.3.3-T exhibited significant increases in extinction coefficient and quantum yield as compared to 4.2.2-T, and its brightness (defined as the product of extinction coefficient and quantum yield) is 1.6 fold higher than that of GFPm-L. The values of the absorption and emission maxima, extinction coefficient, quantum yield, and pK_a differ slightly for the Leu and Tfl forms of each protein variant, suggesting small changes in the environment around the chromophore.[16-18] The oligomerization states of GFPm, 11.3.3-L, and 11.3.3-T were determined by sedimentation velocity analysis (Fig. 2.7). Although GFPm-L was characterized by monomer-dimer self-association with K_d = 23 μM, both 11.3.3-L and 11.3.3-T were primarily dimeric in the concentration range studied. We found no evidence that Tfl stimulates aggregation of the protein. Taken together, these results suggest that the primary consequence of the evolutionary process is an enhancement in the folding behavior of the fluorinated protein.

To explore this conjecture more directly, we examined the refolding kinetics of each protein after acid denaturation.[23, 24] Each variant was denatured in 100 mM citrate buffer (pH 1.9, 2 mM DTT), and the recovery of fluorescence was monitored upon 100-fold dilution into 20 mM Tris buffer (pH 8.0, 100 mM NaCl/1 mM DTT) (Fig. 2.8). The refolding rate was markedly enhanced by evolution (Table 2.4 and Fig. 2.9B), to the extent that 11.3.3-T and GFPm-L refold at comparable rates. The refolding rates of the

Leu forms of the protein are also enhanced; 11.3.3-L folds faster than both GFPm-L and 11.3.3-T.

The directed evolution strategy adopted here yielded a fluorinated GFP variant (11.3.3-T) with physical and spectroscopic properties comparable to those of GFPm-L. It is interesting, however, that although the screening was done in Leu-depleted media, the properties of the Leu form of the protein also improved (albeit far less than those of the Tfl form). This fact suggests that the acquired mutations not only accommodate the replacement of Leu by Tfl, but also improve the folding properties of GFP irrespective of Tfl incorporation. Indeed, four of the mutations found here (S30R, Y145F, I171V, and A206V) were recently reported to improve the folding efficiency of GFP, and to contribute to the development of a "superfolder" variant of the protein.[25] Pédelacq and coworkers developed the superfolder variant by fusing GFP to a poorly folded protein (bullfrog red-cell H-subunit ferritin), and then evolving the fusion construct to enhance folding behavior.[25] Although our objectives and approach differ from those of Pedelacq et al., Tfl also inhibits GFP folding, and it is perhaps not surprising that we would acquire mutations that enhance folding behavior in other contexts. 11.3.3-L and superfolder GFP both exhibit increased expression levels (Fig. 2.4C), improved folding kinetics (Table 2.4 and Fig. 2.9), and enhanced resistance to chemical denaturation (Fig. 2.10). Noncanonical

amino acids may prove to be generally useful tools for improving the folding behavior of proteins.

The replacement of Leu by other canonical amino acids at six positions in 11.3.3 (L42I, L53I, L119I, L194R, L207I, and L221H) suggests that there may be sites at which Tfl cannot be accommodated through compensating mutations. Four Leu residues (L42, L53, L119, and L207) that point toward the inside of the β-barrel (in the structure of wild-type GFP[21]) were mutated to Ile, whereas two surface-exposed Leu positions (L194 and L221) [21] are occupied by hydrophilic residues in the evolved mutant.

There are also two sites at which new Leu residues were added to the sequence (F46L and S202L). F46L has been reported to increase the maturation rate of the chromophore of yellow fluorescent protein[26]; Tfl at position 46 may play a similar role. Among the remaining mutations, only two (F64C and Y106F) are located in the hydrophobic core; the rest are exposed on the outer surface of the β-barrel. The F64C mutation immediately precedes the chromophore, and it is known that an alternative mutation at this position (F64L) improves the maturation of GFP at 37^{o}C.[27] Intentional introduction of a Leu codon at position 64 of 11.3.3 led to a significant decrease in the brightness of the fluorinated form of the protein (Fig. 2.11). Evolution of the fluorinated protein selects Cys rather than Tfl at position 64.

Conclusions

The results reported here show that simple directed evolution strategies, combined with global incorporation of non-canonical amino acids, can produce properly folded, functional proteins of novel composition. Replacement of the Leu residues of GFP by Tfl causes misfolding of the protein in the bacterial cell, probably as a consequence of the increased hydrophobic character of the fluorinated Tfl side chains. Directed evolution of the fluorinated protein yields twenty mutations, eliminating six Leu sites but introducing two others; the evolved mutant carries fluorinated side chains at fifteen positions and exhibits physical and spectroscopic properties nearly identical to those of the parent GFPm. The evolutionary process also improved the folding behavior of the Leu form of the protein, suggesting that noncanonical amino acids may be useful in evolving folding behavior more generally.

Materials and Methods

Preparation of GFPm expression plasmid

Oligonucleotides for site-directed mutagenesis. All oligonucleotides were synthesized at Qiagen (Valencia, CA) on a scale of 10 nmol.

GFP-1: 5’-CATCACGGATCCATGAGTAAAGGAGAAGAACTTTTCACTGG-3’

GFP-2: 5’-CTAATTAAGCTTCTATTTGTAGAGCTCATCCATGCCATG-3’

GFP-3: 5’-CAACATTGAAGATGGTTCCGTTCAACTAGCAG-3’

GFP-4: 5’-CTGCTAGTTGAACGGAACCATCTTCAATGTTG-3’

GFP-5: 5’-CTTGTCACTACTTTCGGTTATGGTGTTCAATGC-3’

GFP-6: 5’- GCATTGAACACCATAACCGAAAGTAGTGACAAG -3’

GFP-7: 5’-GGTGTTCAATGCTTTGCGCGTTATCCGGATC-3’

GFP-8: 5’-GATCCGGATAACGCGCAAAGCATTGAACACC-3’

Addition of two restriction sites, BamHI and HindIII, and removal of the internal BamHI restriction site. CLONTECH GFP-UV vector, containing the “cycle-3” variant, was used as a template for two PCR steps using *Pfu* DNA polymerase (Strategene). The first fragment was obtained by PCR with GFP-1 and GFP-4. The second fragment was obtained by PCR with GFP-2 and GFP-3. The two fragments were purified on a 2% agarose gel (QIAquick gel extraction kit, Qiagen) and assembled by PCR. The PCR products were purified on a 2% agarose gel.

S65G mutation. The purified PCR product was used as a template to mutate amino acid residue 65 from Ser to Gly. Following the same procedure as above, two fragments were generated by PCR with GFP-1 and GFP-6, or with GFP-2 and GFP-5. The fragments

were purified on a 2% agarose gel and assembled by PCR. The PCR products were purified on a 2% agarose gel.

S72A mutation. A DNA fragment bearing terminal BamHI and HindIII sites and the S65G mutation was used as a template for PCR to mutate position 72 to Ala. One fragment was generated by PCR with GFP-1 and GFP-8; the other with GFP-2 and GFP-7. The two fragments were assembled by PCR and purified on a 2% agarose gel.

Construction of the expression plasmid. The DNA fragment encoding GFPm was digested with BamHI and HindIII and ligated into expression plasmid pQE-80L (Qiagen) using T4 DNA ligase (New England Biolab). The resulting plasmid was designated pQE-80L/GFPm.

Construction of randomly mutated GFPm libraries

A Stratagene GeneMorph® random mutagenesis kit was used to generate GFPm libraries according to the manufacturer's protocol by using primers FP (5'-CTTTCGTCTTCACCTCGAG-3') and RP (5'-CTCCATTTTAGCTTCCTTAGCTC-3') with pQE-80L/GFPm as template. The targeted mutation rate was 2-4 mutations per gene in each library. The PCR product was purified on a 1% agarose gel and digested with BamHI and HindIII. The resulting DNA fragment was purified on a 1% agarose gel and

ligated into BamHI/HindIII-digested pQE-80L using T4 DNA ligase. The ligated product was transformed into *E. coli* strain DH10B by electroporation. After overnight culture in LB medium supplemented with 200 mg/L ampicillin, the cells were stored in 25 % glycerol at -80^{o}C.

Protein expression

M9 medium supplemented with 0.4% glucose/3.5 mg/L thiamine/1 mM $MgSO_4$/0.1 mM $CaCl_2$/20 amino acids (40 mg/L)/200 mg/L ampicillin was inoculated 1/100 with an overnight culture (M9) of *E. coli* strain DH10B transformed with mutants or libraries. After each culture reached OD_{600} = 0.9 – 1.0, the cells were harvested by centrifugation (6,000×g, 4 °C, 6 min) and washed twice with cold 0.9% NaCl. The cell pellets were resuspended in supplemented M9 medium containing 19 amino acids (no Leu) and 1 mM Tfl. As discussed in the text, a small amount of Leu was added to the media in early rounds of evolution (Table 2) to control the level of incorporation of Tfl. Protein expression was induced 10 min after the medium shift by addition of isopropyl-β-D-thiogalactoside (IPTG) to a final concentration of 1 mM. After 3 h, the cells were harvested by centrifugation (6,000×g, 4 °C, 10 min). The harvested cells were washed twice with 10 mM phosphate buffer (pH 7.4) containing 100 mM NaCl (PBS) and

resuspended in PBS to ≈0.05 OD_{600} for flow cytometric analysis and cell sorting. Alternatively, the cells were stored at -80°C.

Flow cytometric analysis and cell sorting

All flow cytometric analyses and cell sorting were carried out on a DakoCytomation MoFlo cell sorter (DakoCytomation, Ft. Collins, CO) equipped with an argon ion laser emitting at 488 nm and a 530/40 bandpass filter. The throughput rate of cells was adjusted to 20,000 – 30,000 events/sec, and ≈10^8 cells were sorted in single mode. A gate in the fluorescence channel was set to recover 0.3 – 0.5 % of the most highly fluorescent cells. Additional gates were set in the forward- and side-scatter channels to exclude events arising from large particles. The sorted cells were incubated with 2 mL SOC medium at 37°C for 1 h and then diluted with 20 mL LB medium supplemented with 200 mg/L ampicillin. After overnight growth at 37°C, the cells were stored in 25 % glycerol at -80°C.

Purification of GFPm and GFP variants

N-terminally histidine-tagged GFP mutants were purified by affinity chromatography using Ni-NTA resin (Qiage, Chatsworth, CA) according to the manufacturer's protocols.

Proteins purified under native conditions were precipitated by adding ammonium sulfate to 80 % saturation. The pellets were resuspended with PBS and desalted on PD-10 columns (GE Healthcare, Piscataway, NJ) as recommended by the manufacturer.

Amino acid analysis, molar extinction coefficients, quantum yield, and pK_a

AAA was performed at the W. M. Keck Facility at Yale University (New Haven, CT) on a Beckman Model 7300 ion-exchange instrument (Beckman, Fullerton, CA) after 16 h hydrolysis at 115 ^{o}C in HCl (100 mL, 6 N, 0.2 % phenol) that also contained norleucine (2 nmol). The concentration of purified proteins was measured by using the BCA assay kit (Pierce, Rockford, IL). Each fluorescent protein was diluted to four different concentrations, and the absorbance was measured at the excitation maximum. Beer's law was used to determine the extinction coefficient from the slope of the plots of absorbance vs. protein concentration. The molar mass of fluorinated protein was calculated by using the extent of incorporation of Tfl measured by AAA. Quantum yields were determined by comparison of the integrated spectrum of each protein (480 – 600 nm emission; excitation at 470 nm) to that of fluorescein in 100 mM sodium borate (pH 9.5) (Molecular Probes, Eugene, OR; quantum yield = 0.93). The pK_a values of the proteins were determined as described. [28]

Refolding kinetics

Proteins were unfolded in 100 mM citrate (pH 1.9) containing 2 mM DTT at room temperature for 1 h. The denatured protein solution was diluted 100-fold into 20 mM Tris (pH 8.0)/100 mM NaCl/1 mM DTT. Refolding was monitored by recovery of chromophore fluorescence at 510 nm with excitation at 490 nm. The data were fitted to double-exponential functions by using Origin software (OriginLab, Northampton, MA). We also examined refolding after denaturation in urea solutions; in our hands, however, acid denaturation afforded the most reproducible refolding kinetics.

Equilibrium acid denaturation

Tryptophan fluorescence was measured at 320 nm with excitation at 295 nm after protein samples were equilibrated in 100 mM buffer containing 100 mM NaCl at room temperature. The buffers used were HEPES (pH 8.0, 7.5, and 7.0), MES (pH 6.5, 6.0, and 5.5), and acetate (pH 5.0, 4.5, and 4.0).

Sedimentation velocity analysis

Three protein samples (GFPm-L, 11.3.3-L, and 11.3.3-T) were equilibrated with PBS by

passage through PD-10 columns. Sedimentation velocity analysis was performed at the National Analytical Ultracentrifugation Facility at the University of Connecticut by using a Beckman XL-I Analytical Ultracentrifuge at 20°C. The concentrations of protein samples were 0.125, 0.25, and 0.5 mg/mL. The rotor was accelerated to 55,000 rpm, and interference scans were acquired at one minute intervals for 4 h. The data were analyzed by using the program DcDt+[29] to obtain normalized g(s*) versus s* plots (Fig. 7). Analysis of the sedimentation velocity runs on GFPm-L showed that it exists in a reversible monomer-dimer self-association; K_d was calculated as 23 μM by global fitting using Sedphat.[30] 11.3.3-L and 11.3.3-T are present largely as dimers in the concentration range studied. The program Sedfit (version 9.3)[31] indicated small concentrations of tetramer (≈8 % and ≈3 % for 11.3.3-L and 11.3.3-T, respectively).

Acknowledgments

We thank Jin K. Montclare and Frances H. Arnold for helpful discussion. We also thank Mona Shahgholi for assistance with MALDI-MS. This work was supported by NIH grant GM62523 and Office of Naval Research Grant N00014-03-1-0793 (to D.A.T) and by a Samsung Scholarship (to T.H.Y.).

References

1. Arnold FH (2001) *Nature* 409:253-257.

2. Roodveldt C, Aharoni A, Tawfik DS (2005) *Curr Opin Struct Biol* 15:50-56.

3. Schueler-Furman O, Wang C, Bradely P, Misura K, Baker D (2005) *Science* 310:638-642.

4. Link AJ, Mock ML, Tirrell DA (2003) *Curr Opin Biotechnol* 14:603-609.

5. Budisa N (2004) *Angew Chem Int Ed* 43:6426-6463.

6. Wang L, Schultz PG (2005) *Angew Chem Int Ed* 44:34-66.

7. Tang Y, Ghirlanda G, Petka WA, Nakajima T, DeGrado WF, Tirrell DA (2001) *Angew Chem Int Ed* 40:1494-1496.

8. Tang Y, Tirrell DA, *J Am Chem Soc* 123:11089-11090.

9. Son S, Tanrikulu IC, Tirrell DA (2006) *ChemBioChem* 7:1251-1257.

10. Link AJ, Tirrell DA (2003) *J Am Chem Soc* 125:11164-11165.

11. Cirino PC, Tang Y, Takahashi K, Tirrell DA, Arnold FH (2003) *Biotechnol Bioeng* 83:729-734.

12. Rubini M, Lepthien S, Golbik R, Budisa N (2006) *Biochim Biophys Acta* 1764:1147-1158.

13. Montclare JK, Tirrell DA (2006) *Angew Chem Int Ed* 45:4518-4521.

14. Bilgiçer B, Kumar K (2004) *Proc Natl Acad Sci USA* 101:15324-15329.

15. Lee H-Y, Lee K-H, Al-Hashimi HM, E. Marsh ENG (2006) *J Am Chem Soc* 128:337-343.

16. Phillips GN, Jr (1997) *Curr Opin Struct Biol* 7:821-827.

17. Tsien RY (1998) *Annu Rev Biochem* 67:509-544.

18. Zimmer M (2002) *Chem Rev* 102:759-781.

19. Cormack BP, Valdivia RH, Falkow S (1996) *Gene* 173:33-38.

20. Crameri A, Whitehorn EA, Tate E, Stemmer WPC (1996) *Nat Biotech* 14:315-319.

21. Yang F, Moss LG, Phillips GN, Jr (1996) *Nat Biotech* 14:1246-1251.

22. Reid BG, Flynn GC (1997) *Biochemistry* 36:6786-6791.

23. Merkel JS, Regan L (2000) *J Biol Chem* 38:29200-29206.

24. Enoki S, Saeki K, Maki K, Kuwajima K (2004) *Biochemistry* 43:14238-14248.

25. Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) *Nat Biotech* 24:79-88.

26. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A (2002) *Nat Biotech* 20:87-90.

27. Helm R, Cubitt AB, Tsien RY (1995) *Nature* 373:663-664.

28. Rizzo MA, Springer GH, Granada B, Piston DW (2004) *Nat Biotech* 22:445-449.

29. Philo JS (2000) *Anal Biochem* 279:151-163.

30. Schuck P (2003) *Anal Biochem* 320:104-124.

31. Schuck P (2000) *Biophys J* 78:1606-1619.

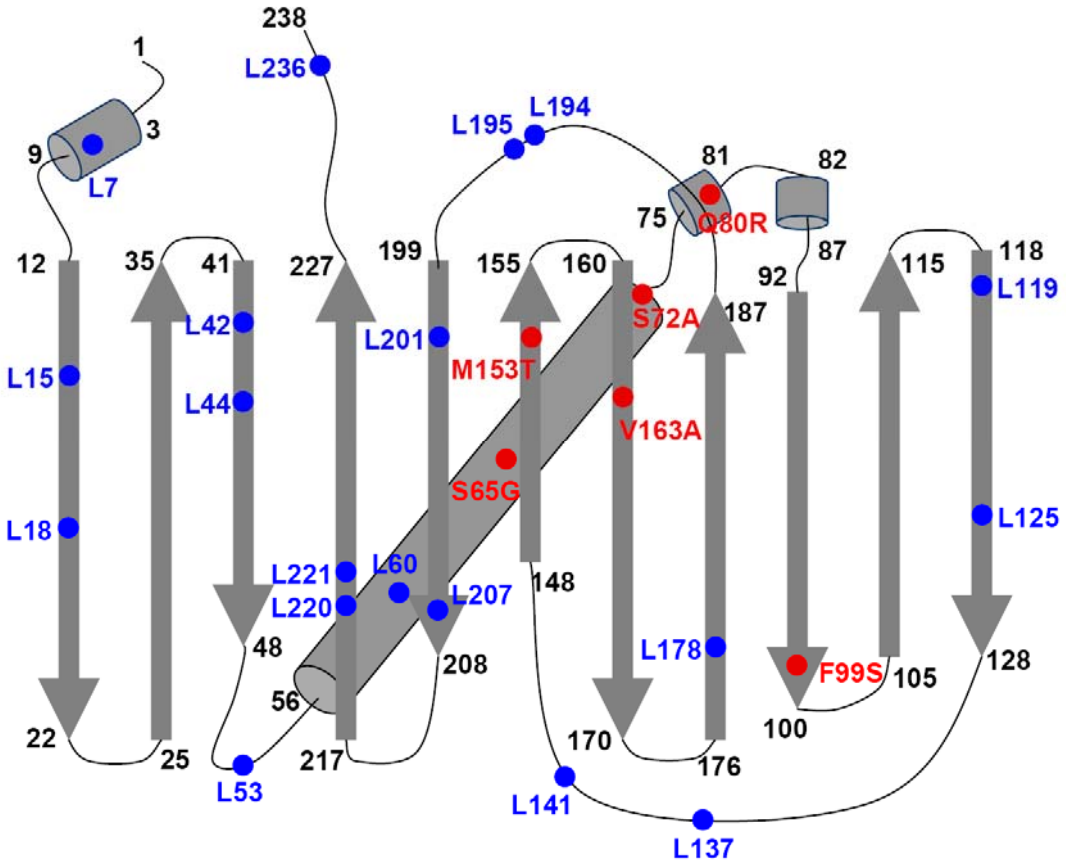

Fig. 2.1. Schematic representation of the GFPm scaffold. Two mutations (S65G and S72A, red circles) were introduced into the cycle-3 GFP mutant (Q80R, F99S, M153T, and V163A, red circles) to change the spectral properties. There are 19 Leu residues (blue circles).

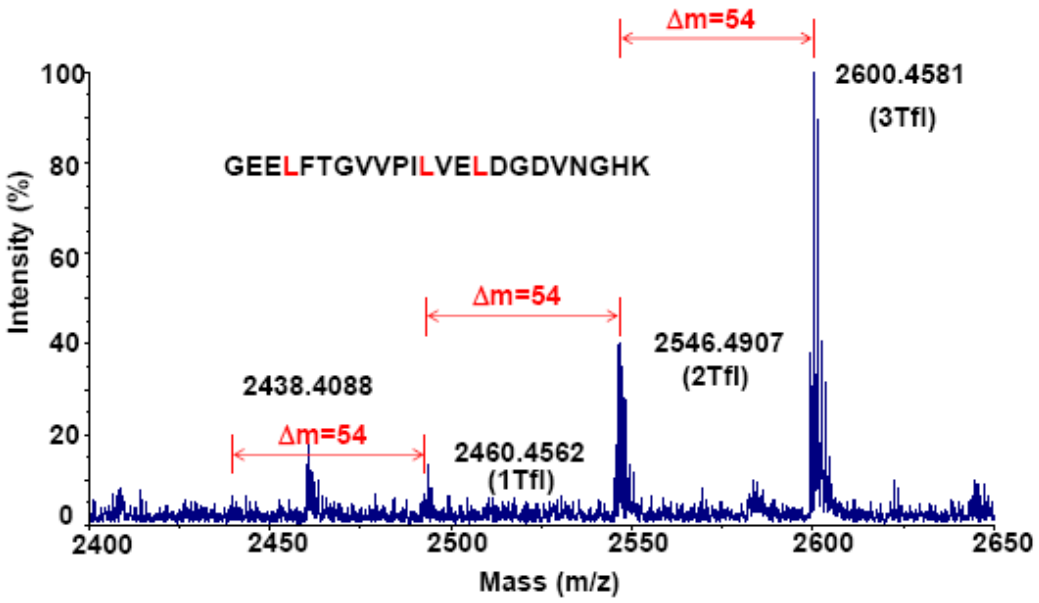

Fig. 2.2. MALDI-MS analysis of GFPm-T after trypsin digestion. A peptide fragment of sequence GEELFTGVVPILVELDGDVNGHK yields the spectrum shown. Replacement of Leu with Tfl results in a 54-amu mass increase per site.

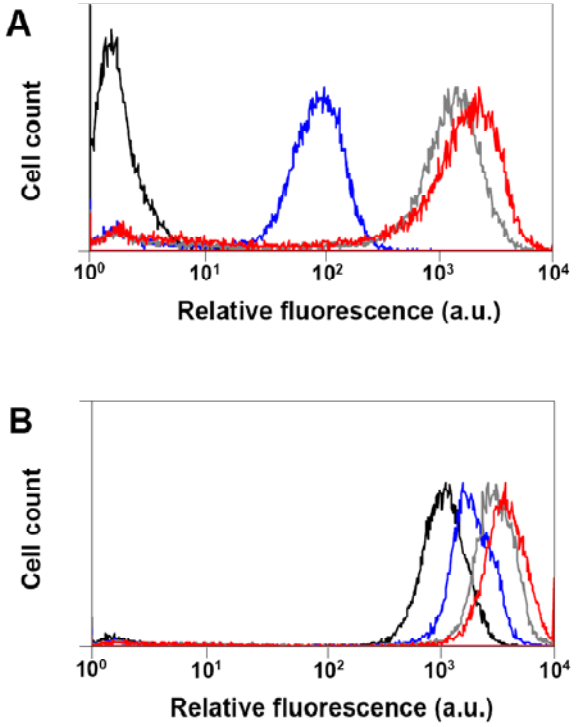

Fig. 2.3. Flow cytometric analysis of GFPm and GFP variants. Proteins were expressed in media depleted of Leu and supplemented with Tfl (A) or in media containing all 20 canonical amino acids (B). Black line: GFPm; blue line: 4.2.2; gray line: 8.3.3; red line: 11.3.3.

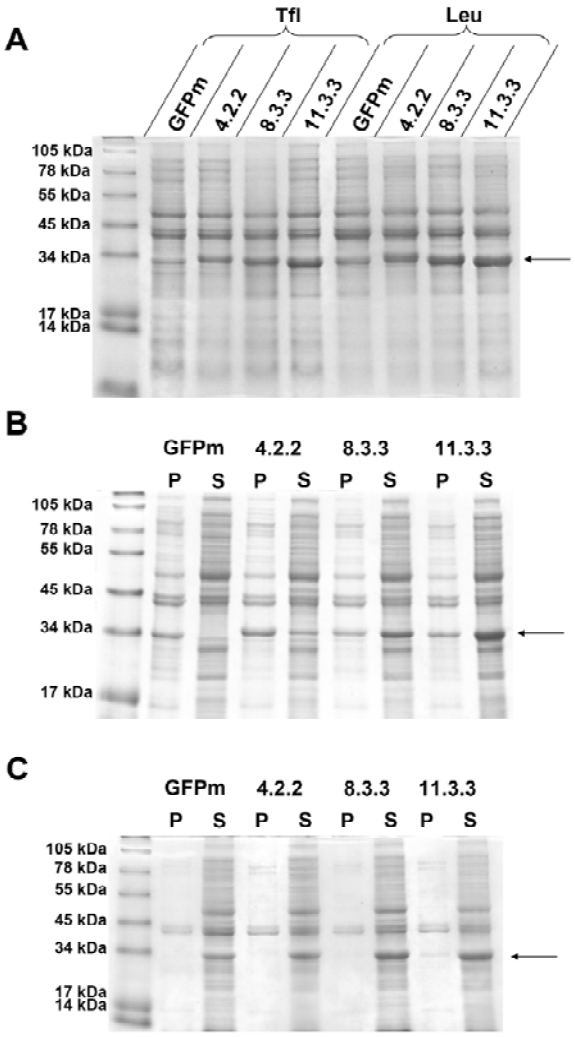

Fig. 2.4. Expression level and solubility *in vivo* of GFPm and GFP variants. (A) SDS-PAGE of whole cell lysates after IPTG induction for 3 h. (B and C) SDS-PAGE of crude cell lysate pellets (P) and supernatants (S). The proteins were expressed in media depleted of Leu and supplemented with Tfl (B) or in media containing all 20 canonical amino acids (C). The arrows indicate the expected size of GFPm or GFP variants.

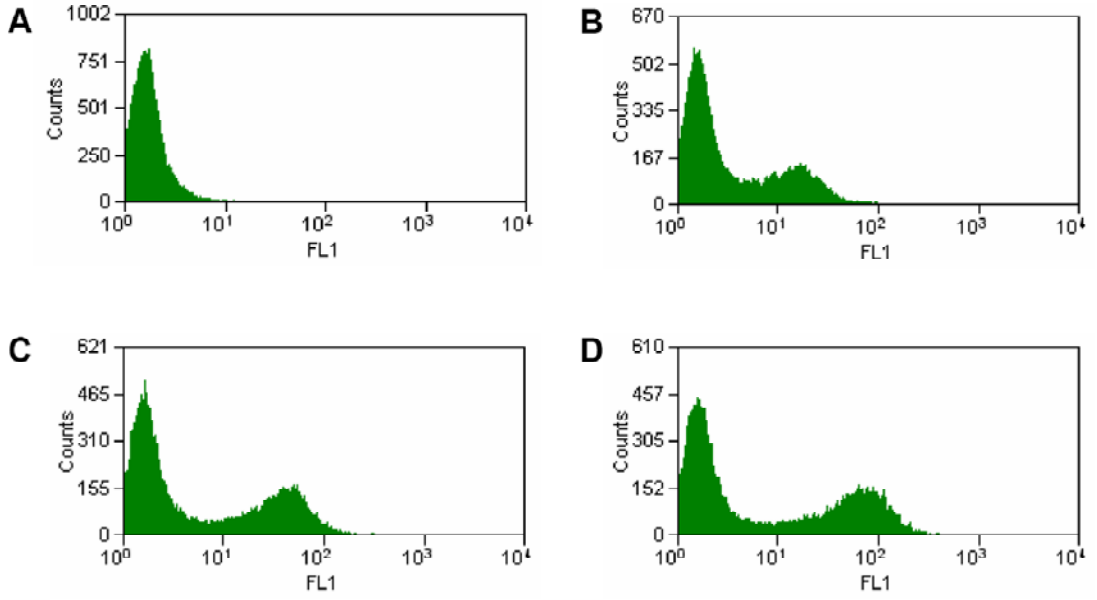

Fig. 2.5. Flow cytometric analysis of cells expressing the first library in media supplemented with 1 mM Tfl and no Leu (A), 15 μM Leu (B), 30 μM Leu (C), or 45 μM Leu (D).

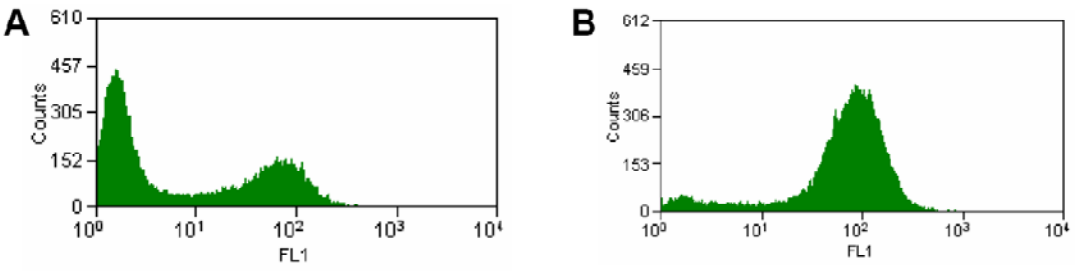

Fig. 2.6. Flow cytometric analysis of cells expressing proteins in media supplemented with 1 mM Tfl and 45 μM Leu. (A) Cells harboring the first library. (B) Cells after two rounds of sorting starting from A.

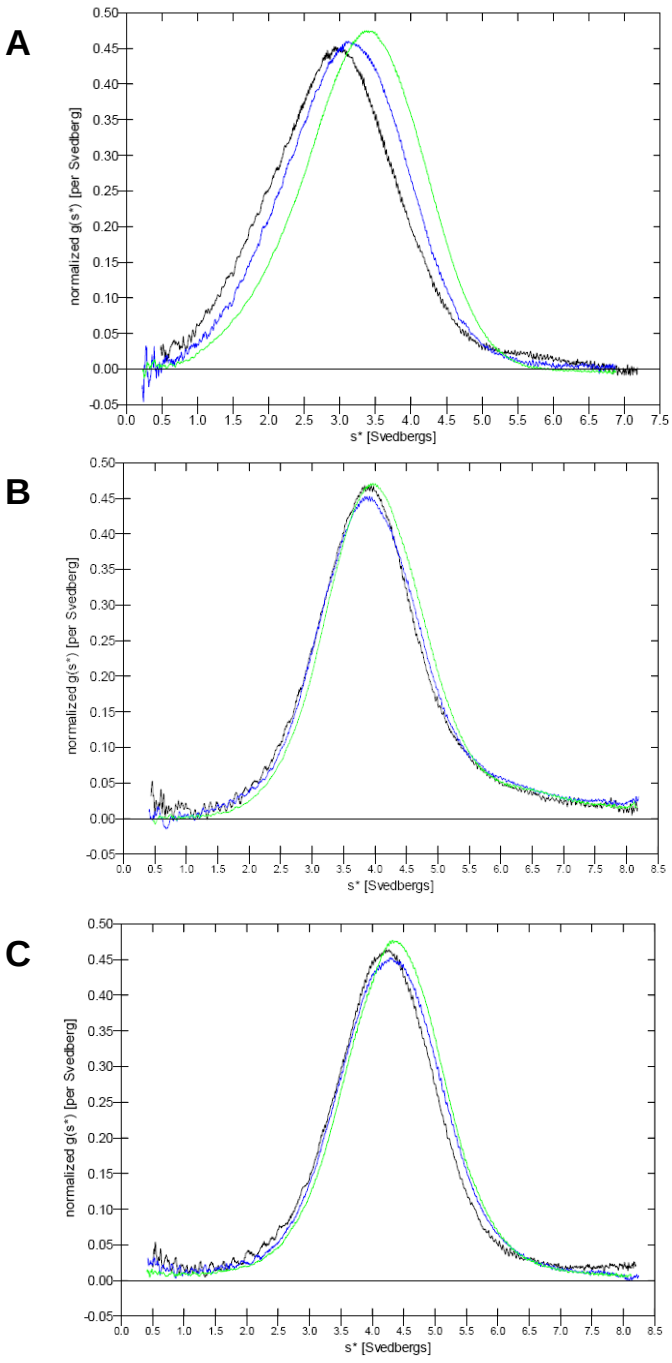

Fig. 2.7. Overlay of the normalized g(s*) plots from DcD+ analysis for GFPm-L(A), 11.3.3-L (B), and 11.3.3-T (C). Black line: 0.125 mg/mL; blue line: 0.25 mg/mL; pale green line: 0.5 mg/mL

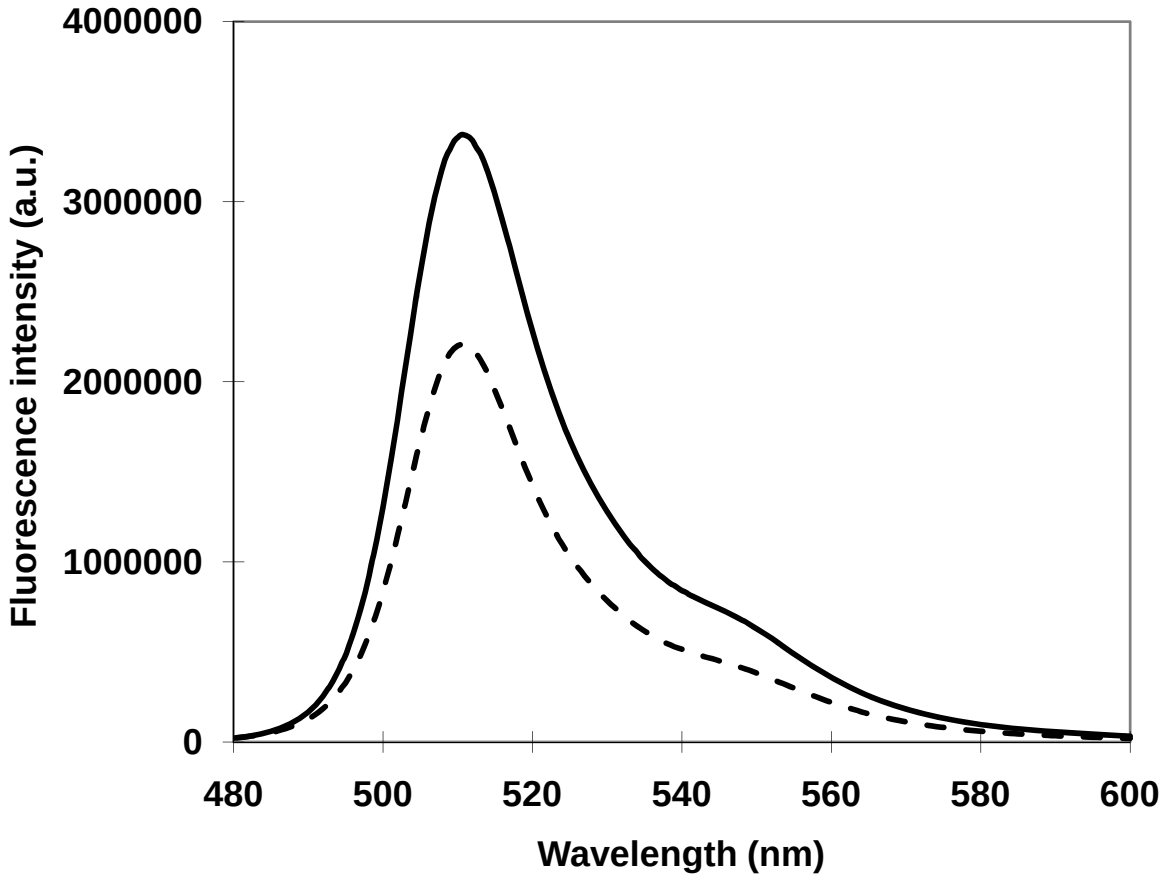

Fig. 2.8. Fluorescence spectra of 11.3.3-L before (solid line) and after (dashed line) acid denaturation. Purified protein was unfolded in 100 mM citrate (pH 1.9) containing 2 mM DTT at room temperature for 1 h. The denatured protein solution was diluted 100-fold into 20 mM Tris (pH 8.0)/100 mM NaCl/1 mM DTT. Incubation of GFP at pH 1.9 is known to cause some irreversible denaturation[24]; however, the kinetics of refolding are independent of the extent of irreversible denaturation and the spectrum of the refolded protein is consistent with refolding to an authentic native structure.

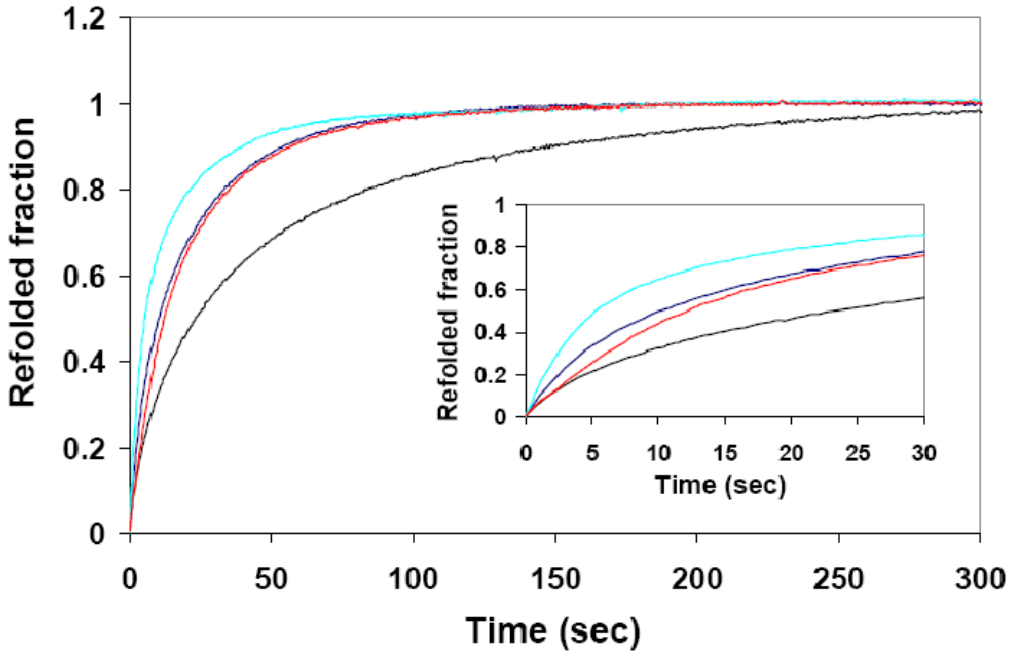

Fig. 2.9. Representative plots of refolding kinetics after acid denaturation for GFPm-L (blue), 4.2.2-T (black), 11.3.3-L (aqua), and 11.3.3-T (red). The inset plots show the early stages of the refolding process.

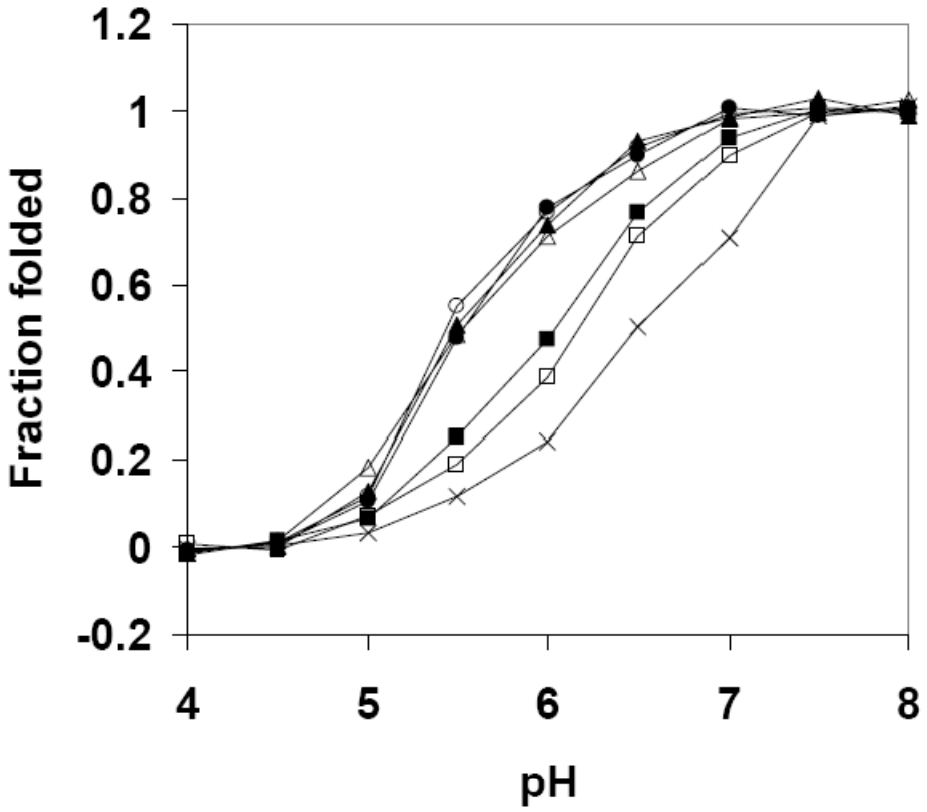

Fig. 2.10. Equilibrium acid denaturation plots. The purified proteins were equilibrated with 100 mM buffer (HEPES for pH 8, 7.5, and 7, MES for pH 6.5, 6, and 5.5, Acetate for pH 5, 4.5, 4) containing 100 mM NaCl at room temperature. Trp fluorescence was measured at 320 nm with excitation at 295 nm. GFPm-L (×), 4.2.2-L (■), 4.2.2-T (□), 8.3.3-L (▲), 8.3.3-T (△), 11.3.3-L (●), and 11.3.3-T (○).

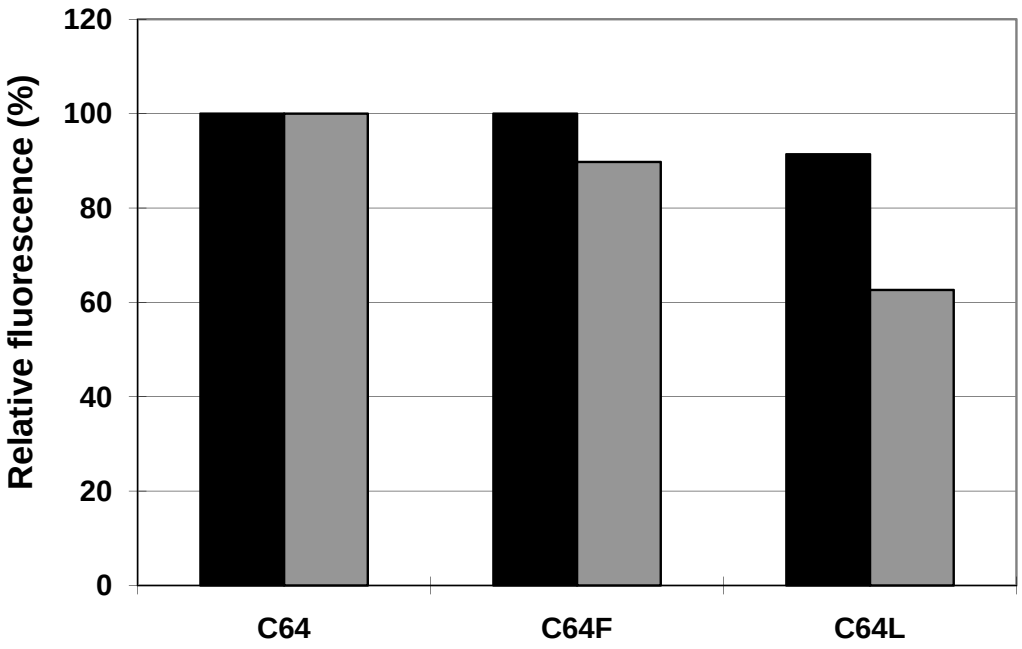

Fig. 2.11. Effects of mutation at position 64 of 11.3.3. Position 64 was mutated into Phe or Leu, and the protein was expressed in media supplemented with the 20 canonical amino acids (black bars) or in media depleted of Leu and supplemented with Tfl (gray bars). Median cell fluorescence was measured by a cell sorter equipped with an argon ion laser emitting at 488 nm and a 530/40 bandpass filter. Relative fluorescence is defined as the median cell fluorescence compared to that of cells expressing 11.3.3 in media supplemented with Leu or with Tfl, respectively.

Table 2.1. Fluorescence of cells expressing GFPm and GFP variants.

	Relative fluorescence[a]
GFPm-L	537
GFPm-T	1
4.2.2-L	938
4.2.2-T	43
8.3.3-L	1498
8.3.3-T	620
11.3.3-L	1826
11.3.3-T	662

[a]Relative median fluorescence with respect to fluorescence of cells expressing GFP in media supplemented with 19 amino acids (-Leu) plus Tfl. The median cell fluorescence was measured by a cell sorter equipped with an argon ion laser emitting at 488 nm and a 530/40 bandpass filter.

Table 2.2. Libraries and screening conditions.

Library	1[a]	2	3	4	5	6	7	8	9[a,b]	10	11
Library size[c]	1×10^7	1×10^7	5×10^6	2×10^7	6×10^7	5×10^7	6×10^7	5×10^7	1×10^8	1×10^8	1×10^8
Leu (μM)[d]	45	30, 15[e]	7.5, 0[e]	0	0	0	0	0	0	0	0
No. of sorting	2	2	2	2	2	2	2	3[f]	2	2	3[f]

[a]Measured mutation rates in libraries 1 and 9 were 2.4 and 1.9 mutations per gene, respectively.

[b]Clone 8.3.3 was used for library generation.

[c]Library size was determined by counting colonies after spreading on agar plates.

[d]Leu concentration in medium during protein expression.

[e]Leu concentration for the first and second sortings, respectively.

[f]After sorting twice, the genes for the GFP mutants were transferred into fresh pQE-80L plasmids.

Table 2.3. Amino acid sequences of wild-type GFP (wtGFP), GFPm, and GFP variants.

	5	10	15	20	25	30
wtGFP	M S K G E E L F T G V V P I L V E L D G D V N G H K F S V S					
GFPm						
4.2.2						T
8.3.3						R
11.3.3						R

	35	40	45	50	55	60
wtGFP	G E G E G D A T Y G K L T L K F I C T T G K L P V P W P T L					
GFPm						
4.2.2	N	I	L			
8.3.3		I	L			
11.3.3		I	L		I	

	65	70	75	80	85	90
wtGFP	V T T F S Y G V Q C F S R Y P D H M K Q H D F F K S A M P E					
GFPm	G	A		R		
4.2.2	C G	A		R		
8.3.3	C G	A		R		
11.3.3	C G	A		R		

	95	100	105	110	115	120
wtGFP	G Y V Q E R T I F F K D D G N Y K T R A E V K F E G D T L V					
GFPm		S				
4.2.2		S				
8.3.3		S	K F			I
11.3.3		S	K F			I

	125									130								135								140								145								150
wtGFP	N	R	I	E	L	K	G	I	D	F	K	E	D	G	N	I	L	G	H	K	L	E	Y	N	Y	N	S	H	N	V										
GFPm																																									
4.2.2																				R																				
8.3.3				K																									D										
11.3.3				K																					F				D										

	155									160								165								170								175								180
wtGFP	Y	I	M	A	D	K	Q	K	N	G	I	K	V	N	F	K	I	R	H	N	I	E	D	G	S	V	Q	L	A	D										
GFPm			T										A																												
4.2.2			T										A								V																			
8.3.3			T						T				A								V																			
11.3.3									T				A								V																			

	185									190								195								200								205								210
wtGFP	H	Y	Q	Q	N	T	P	I	G	D	G	P	V	L	L	P	D	N	H	Y	L	S	T	Q	S	A	L	S	K	D										
GFPm																																									
4.2.2																																								
8.3.3														R								L				V	I													
11.3.3														R								L				V	I													

	215									220								225								230								235			238
wtGFP	P	N	E	K	R	D	H	M	V	L	L	E	F	V	T	A	A	G	I	T	H	G	M	D	E	L	Y	K									
GFPm																																							
4.2.2											H											D																
8.3.3											H																											
11.3.3											H												I															

Table 2.4. Photophysical properties and refolding kinetitcs of GFPm and GFP variants.

Clone	λ_{abs} (ε)[a]	λ_{em} (QY)[a]	Relative brightness[b]	pK_a	Refolding kinetic parameters				Percentage Tfl by AAA[e]
					k_1 (10^{-1} s^{-1})[c]	k_2 (10^{-2} s^{-1})[c]	A_1[d]	A_2[d]	
GFPm-L	503 (84)	509 (0.42)	1	6.7	1.5	3.4	0.36	0.64	-
4.2.2-L	507 (127)	511 (0.60)	2.2	6.2	1.9	3.3	0.37	0.63	-
4.2.2-T	500 (56)	509 (0.28)	0.44	6.9	0.55	1.1	0.47	0.53	78
8.3.3-L	507 (133)	512 (0.67)	2.5	5.7	2.1	4.1	0.56	0.44	-
8.3.3-T	503 (103)	510 (0.53)	1.6	6.0	0.86	3.1	0.47	0.53	85
11.3.3-L	505 (147)	510 (0.61)	2.5	5.6	1.8	3.3	0.58	0.42	-
11.3.3-T	501 (114)	509 (0.48)	1.6	6.0	0.94	2.6	0.57	0.43	87

[a]Absorption and emission maxima are expressed in nm; the extinction coefficient (ε) is expressed in $10^3 M^{-1}cm^{-1}$. QY is the fluorescence quantum yield. [b]Relative brightness was calculated as the product of ε and QY divided by the brightness of GFPm-L. [c]Rate constants for double-exponential fits of the refolding curves. [d]Relative magnitude of each component of the refolding process. [e]Percentage of replacement of Leu by Tfl as determined by AAA.