Gene, 173 (1996) 33–38
© 1996 Elsevier Science B.V. All rights reserved. 0378-1119/96/$15.00

GENE 09389

FACS-optimized mutants of the green fluorescent protein (GFP)

(GFP mutation; FITC; fluorescence-activated cell sorter; fluorescence intensity)

Brendan P. Cormack[a], Raphael H. Valdivia[a] and Stanley Falkow[a,b]

[a]*Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305-5402, USA; and*
[b]*Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, Hamilton, MT 59840, USA*

Received by D.C. Youvan: 22 May 1995; Revised/Accepted: 3 August/23 August 1995; Received at publishers: 25 September 1995

SUMMARY

We have constructed a library in *Escherichia coli* of mutant *gfp* genes (encoding green fluorescent protein, GFP) expressed from a tightly regulated inducible promoter. We introduced random amino acid (aa) substitutions in the twenty aa flanking the chromophore Ser-Tyr-Gly sequence at aa 65–67. We then used fluorescence-activated cell sorting (FACS) to select variants of GFP that fluoresce between 20-and 35-fold more intensely than wild type (wt), when excited at 488 nm. Sequence analysis reveals three classes of aa substitutions in GFP. All three classes of mutant proteins have highly shifted excitation maxima. In addition, when produced in *E. coli*, the folding of the mutant proteins is more efficient than folding of wt GFP. These two properties contribute to a greatly increased (100-fold) fluorescence intensity, making the mutants useful for a number of applications.

INTRODUCTION

The green fluorescent protein (GFP) of the jellyfish *Aequorea victoria* absorbs light with an excitation maximum of 395 nm, and fluoresces with an emission maxima of 510 nm (Morise et al., 1974; Ward et al., 1980). Since this can occur in the absence of any cofactors, GFP is extremely useful as a marker for gene expression and as a tag in studying protein localization in a variety of organisms (Chalfie et al., 1994; Heim et al., 1994; Wang and Hazelrigg, 1994). The presence of GFP can be monitored using standard fluorescein isothiocyanate (FITC) excitation-emission filter sets by virtue of a minor absorption peak at 470 nm (see Fig. 4). This absorption peak also permits excitation of GFP with the 488 nm line of an argon laser which can be used in fluorescence-activated cell sorting (FACS). The resulting fluorescence, however, is less intense than that resulting from optimal excitation, and therefore mutants that fluoresce more intensely when excited at 488 would be of clear value in many GFP applications. Indeed, mutations in GFP which shift the excitation maxima from 395 nm to around 490 nm have been reported and these proteins do fluoresce more intensely when excited at 488 nm (Delagrave et al., 1995; Heim et al., 1995).

In this study, we set out to isolate mutants of GFP optimized for use in FACS analysis. We constructed a library of mutant GFP molecules using an oligo-directed, codon-based mutagenesis method (Glaser et al., 1992; Cormack and Struhl, 1993). The mutagenesis strategy allowed for simultaneous mutagenesis of a 20-aa region surrounding the chromophore. We then used FACS with a standard FITC filter set to screen this library for GFP mutants with increased fluorescence when excited at 488 nm. We isolated three distinct classes of mutant GFP,

Correspondence to: Dr. B.P. Cormack, Fairchild D035, Department of Microbiology and Immunology, Stanford Medical School, Stanford, CA 94305-5402, USA. Tel. (1-415) 723-2671; Fax (1-415) 723-1837; e-mail: cormack@cmgm.stanford.edu

Abbreviations: *A*, absorbance (1 cm); aa, amino acid(s); Ap, ampicillin; FACS, fluorescence-activated cell sorter; FITC, fluoroisothiocyanate; GFP, green fluorescent protein; *gfp*, gene encoding GFP; IPTG, isopropyl-β-D-thiogalactopyranoside; K, G or T; kb, kilobase(s) or 1000 bp; N, A or C or G or T; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; PCR, polymerase chain reaction; wt, wild type; 2 × YT, see legend to Fig. 2.

SSDI 0378-1119(95)00685-0

all of which have markedly shifted excitation maxima. In addition, the mutant proteins fold more efficiently than wt GFP in *E. coli*.

EXPERIMENTAL AND DISCUSSION

(a) Mutagenesis of *gfp*

In our mutagenesis of *gfp*, we targeted aa 55–74, the 20 aa immediately surrounding aa 65–67 which ultimately form the chromophore. In standard oligo-directed mutagenesis schemes, defined amounts of the three non-wt nt are included at each step of oligo synthesis; this doping results in a defined mutation rate per nt position, and an equal frequency of the three possible nt substitutions. There is, however, a strong bias at the aa level in favor of those substitutions corresponding to codons that differ by a single nt from the wt codon. We have therefore used the codon-based mutagenesis scheme shown in Fig. 1 (Glaser et al., 1992; Cormack and Struhl, 1993). This method yields a mutant library in which there is a certain probability that a particular codon is mutated, but if so then an equal probability of substituting any of the 32 possible NNK (where K = G or T) codons. As a result, the method results in a compact yet highly representative library of mutations. For example, in the region mutagenized in this study, all possible single aa changes are covered in approx. 2400 clones.

(b) GFP mutant selection

Using this method, we mutagenized aa 55–74 of GFP at a frequency of 10% per codon. We constructed a library in *E. coli* of 6×10^6 mutant genes transcribed from a tightly regulated IPTG-inducible promoter. We used a FACS machine to analyze the mutant library: in log phase, in the absence of inducer, the pool of mutants showed negligible fluorescence (data not shown). After IPTG induction of log phase cells for 2.5 h, there was a clear increase in fluorescence for a sub population of the library (Fig. 2A). The non fluorescing portion of the library is likely to contain non-fluorescent mutants of GFP. After 2.5 h of induction with IPTG, we sorted the most fluorescent members of the induced population. As expected, when this population was amplified in broth and induced with IPTG, it showed a much higher average fluorescence than the starting pool (Fig. 2B). We recovered the most intensely fluorescing 0.5% of this already enriched population. From this pool, we analyzed 50 strains in detail. After induction, individual bacterial strains fluoresced between 10- and 110-fold more intensely than a control strain expressing wt GFP (e.g., Fig. 2C). This increase in fluorescence intensity does not depend on the bacterial strain used, since identical

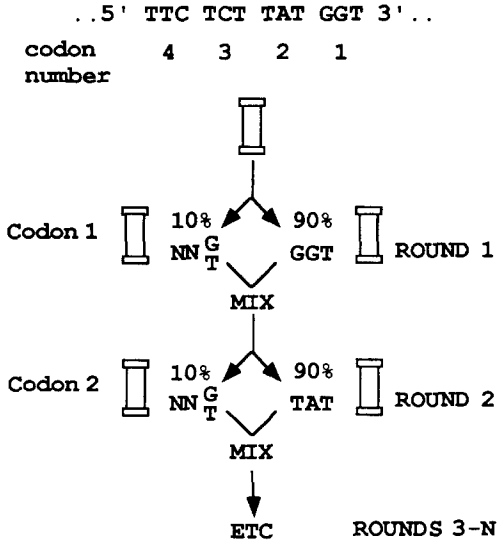

Fig. 1. Construction of the GFP mutant library. Details of oligo synthesis method are described elsewhere (Glaser et al., 1992; Cormack and Struhl, 1993). Briefly, at the step in oligo synthesis corresponding to a given codon, the synthesis column is dismantled, and the silica matrix is split into two portions and repacked into two synthesis columns. One column is subjected to three rounds of synthesis with the wt nt; the second column is subjected to three rounds of synthesis yielding the codon NNK, where N corresponds to an equimolar mix of the four nt, and K to an equimolar mix of G and T. The matrix in the columns is combined and the process repeated for each codon being mutated. For ease of presentation, the example synthesis steps shown in the figure correspond to the coding strand of codons 64–67, even though the oligo actually synthesized is complementary to the coding strand. For each codon, 10% of the total silica matrix received the NNK codon, yielding an average of two mutated codons per molecule. **Methods:** For these experiments, *gfp* was expressed under control of the *tac* promoter in pKEN2, a high copy phagemid (gift of G. Verdine). The oligo GF1, shown below, was used to introduce the ribosome-binding site of the phage T7 gene*10* upstream from the *gfp* gene. GF1: 5′ - GATTTCTAGA*ATTTAAGAAGGAGATATACAT***ATGAGTAAAGGAGAAG** (in bold is the coding region of the *gfp* gene, in italics, the sequence from T7 gene *10*). The *Xba*I site at the 5′ end of the oligo was used for cloning. The sequence of the mutagenic oligo GF2 (the mutagenized codons are in bold; the *Bgl*II restriction site used for cloning is in italics) is as follows: TG.CTG.TTT.CAT.*AAG.ATC.T*GG.**GTA.TCT.TGA.AAA.GCA.TTG.AAC.ACC.ATA.AGA.GAA.AGT.AGT.GAC.AAG.TGT.TGG.CCA.TGG.AAC**.AGG.TAG.TTT.TCC.AGT.AGT.GC. Oligos GF1 and GF2 were used in a PCR amplification of wt *gfp*. PCR was carried out for 20 cycles of 1 min at 93°C, 3 min at 50°C and 1 min at 72°C. The amplified fragment was cloned, using the *Xha*I and *Bgl*II sites, into a *gfp* derivative (into which we had introduced a *Bcl*I site at nt 225, numbering from the ATG). The library was transformed into *E. coli* strain XA90[F′*lacI*^{Q1} *pro-A*$^+$*B*$^+$] Δ(*lac-pro*)XIII *ara nal argE*(am) *thi rif*^{R}), an overproducer of *lac* repressor. In this strain there is very little expression from the *tac* promoter in the absence of inducer.

increases in fluorescence intensity are seen when mutant GFP is expressed in *E. coli* strain DH12S, *Yersinia pseudotuberculosis*, or *Salmonella typhimurium* (data not shown). Furthermore, the observed increase in fluorescence is not an artifact of the particular expression system used, since the mutant GFPs all resulted in more intense

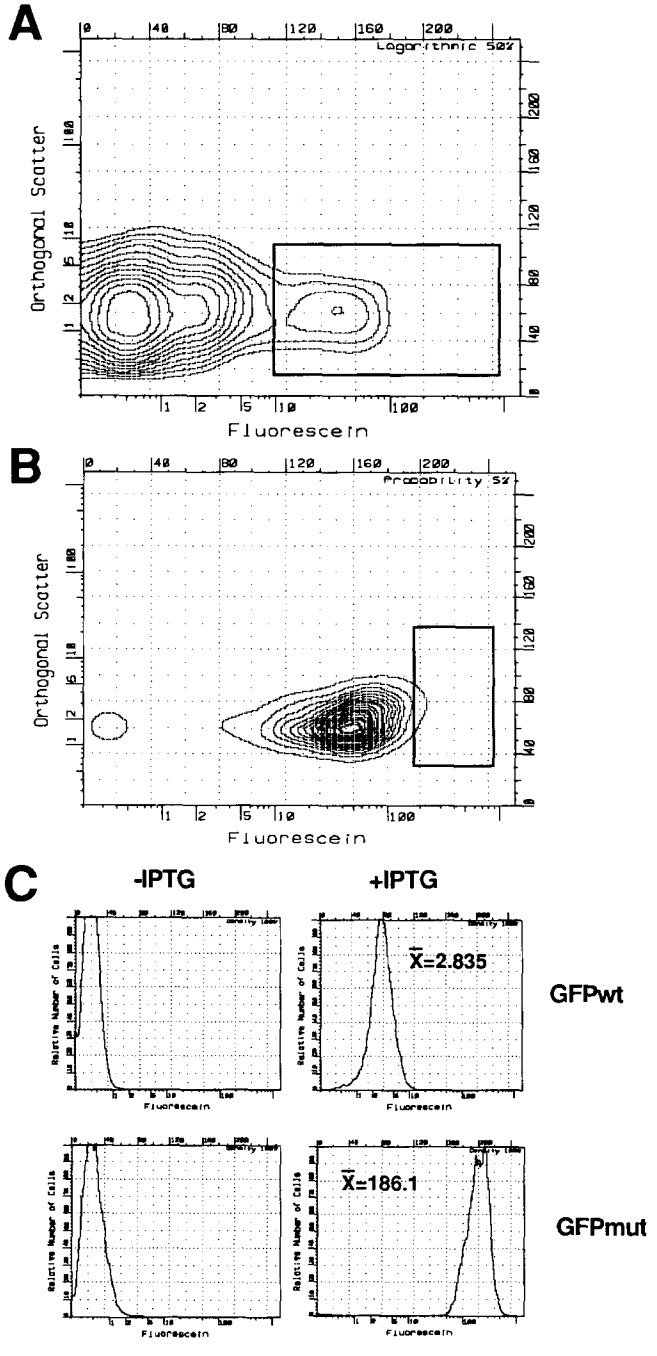

Fig. 2. Selection of fluorescence-enhanced GFP mutants. (**A**) FACS scan of the total mutagenized *gfp* pool after 2.5 h of induction. A pool representing 10^6 independent mutations was diluted 1:100 in 2 × YT broth (16 g tryptone/10 g yeast extract/5 g NaCl per liter) supplemented with 100 μg Ap/ml and 0.2 mM IPTG and grown at 37°C. After 2.5 h, the bacterial population was diluted 1:10 in phosphate-buffered saline (137 mM NaCl/2.7 mM KCl/4.3 mM $Na_2HPO_4 \cdot 7H_2O$/1.4 mM KH_2PO_4) and analysed in a FACStar^{Plus} (Beckton Dickinson) machine, essentially as described in an accompanying paper (Valdivia et al., 1996). Fluorescence emission was read with a 515/40 band pass filter. Fluorescein and side scatter data were collected with logarithmic amplifiers. The relatively rare population of highly fluorescent GFP mutants can be seen by logarithmic contour displays of a scan of 5×10^4 bacteria. The inner boundaries (box) in the FACS scan represent the fluorescence channel boundaries (gates) that were set to sort the mutant population with highest fluorescence intensity. Out of a total of 4×10^6 events, approx. 10^4 that fell within the imposed gates were collected and amplified for the second round of selection. (**B**) FACS scan of high-fluorescence GFP mutants. The sorted population from panel **A** was amplified in 2 × YT broth supplemented with 100 μg Ap/ml. For the second round of FACS selection, the high fluorescent GFP mutant pool was prepared and scanned as in **A**. The gates placed within this first population (box) were designed to sort the top 0.5% of the high-fluorescent population. Approx. 500 individual events were sorted and amplified in 2 × YT broth. (**C**) Scan of an individual enhanced-GFP mutant compared to wt GFP. Overnight cultures of a single colony from the population sorted in **B** and of a strain expressing wt GFP were subcultured in 2 × YT supplemented with 100 μg Ap/ml or 2 × YT with Ap + IPTG (0.2 mM) for 2.5 h. FACS scan were run on the induced and uninduced cultures as described in the legend to panel **A**. The integral mean fluorescence intensities ($\bar{X}$) of the induced and uninduced cultures were calculated using the FACS-DESK program.

fluorescence than wt GFP when expressed from a T7 promoter (plasmid pT7.7, a gift from Stan Tabor).

(c) Characterization of GFP mutants

We sequenced the genes corresponding to the 12 most fluorescent isolates (Table I), and could distinguish three classes of mutants. GFPmut1 has a double substitution: F64L,S65T. GFPmut2 has a triple substitution: S65A, V68L, S72A. GFPmut3, represented by 10 of the 12 mutants, has the same double aa substitution: S65G S72A. Thus, in all three mutant classes Ser^{65} is mutated and always in the context of at least one other mutation. After induction with IPTG, the bacterial strains expressing the three mutant GFPs already show a substantial (3 ×) increase in fluorescence by the first time point (8 min) after adding inducer (Fig. 3). Half-maximal fluorescence appears by 25 min, and within 1 h the bacteria fluoresce approx. 100-fold more intensely than those expressing wt protein. Sequence of the promoter revealed no mutations for any of the plasmids carrying the mutant GFP. For each of the three mutants, we verified that there were no mutations outside the targeted region by sequencing the entire GFP coding region. We then recloned the mutant GFP coding regions into the original pKEN expression vector. These reconstructed vectors behaved identically to the original mutant isolates, demonstrating that the novel properties of the mutated GFPs were the result of the identified mutations. The difference in fluorescence intensity between strains expressing wt and mutant GFP could in principle be due to any of a number of factors: increased protein expression, more efficient protein folding, increased $A_{488\ nm}$ or faster chromophore formation. After induction with IPTG, the amount of protein produced by the wt and mutant strains is the same. There is, however, a significant effect of the mutations on protein folding. Consistent with what has been previously observed (Heim et al., 1994), we observed that much of the wt GFP is found in inclusion bodies as

TABLE I

Sequence of GFP mutants

	Position[a]								
	64	65	66	67	68	69	70	71	72
wt GFP	TTC	TCT	TAT	GGT	GTT	CAA	TGC	TTT	TCA
	Phe	Ser	Tyr	Gly	Val	Gln	Cys	Phe	Ser
GFPmut 1	CTG	ACT							
	Leu	Thr							
GFPmut 2		GCG			CTT				GCG
		Ala			Leu				Ala
GFPmut 3a		GGG							GCG
		Gly							Ala
GFPmut 3b		GGT							GCG
		Gly							Ala

[a] Sequencing was performed using Sequenase 2.0 (US Biochemical, Cleveland, OH, USA) Only the sequence corresponding to aa 64–72 is shown, since there are no mutations outside of this region.

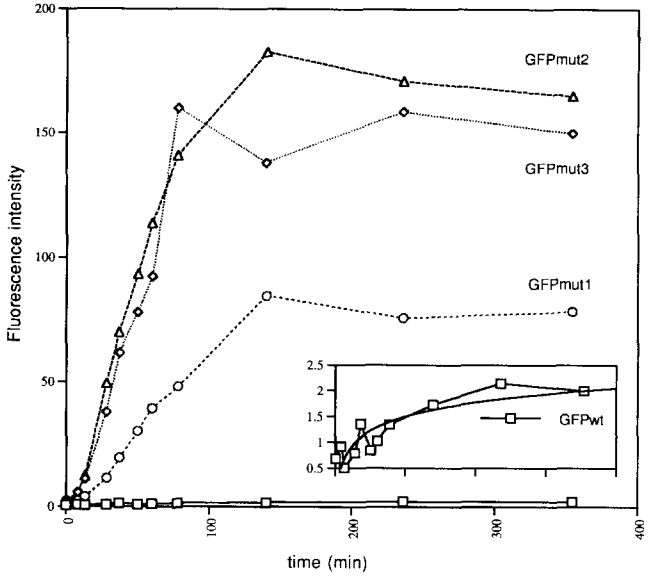

Fig. 3. Fluorescence of bacterial strains expressing GFP mutants. Strain XA90 expressing wt or mutant GFP were grown in 2 × YT media to early log phase; cultures were diluted 1:100 into 2 × YT broth supplemented with 100 μg Ap/ml and 0.2 mM IPTG. Dilutions were chosen so that for the entire course of the experiment, bacteria were in log phase. At various times after addition of inducer, FACS scans were run on the induced cultures as described in the legend to Fig. 2A. The mean fluorescence of the population was calculated for each time point, using the FACS-DESK program and the results plotted. The inset shows the average fluorescence of a strain expressing wt GFP using an expanded *y*-axis scale.

non-fluorescent insoluble protein (data not shown). By contrast, when expressed under identical conditions, 90% of GFPmut1 and virtually all of GFPmut2 or GFPmut3 is soluble (data not shown). This contributes in part to the increased fluorescence of the bacteria expressing the mutant GFP proteins.

To assess the effect of the mutations on the fluorescence characteristics of the protein itself, we isolated soluble GFP from the three mutant strains and analyzed their excitation and emission spectra using fluorescence spectroscopy (Fig. 4). Not unexpectedly, all three mutant proteins show a severe shift in absorption frequency from 395 nm for parental GFP to between 480 nm and 501 nm for the three mutant proteins. The emission wavelengths for the three mutants are essentially unchanged (ranging from 507–511 nm). This gross shift in absorption is responsible for most of the increased fluorescence: spectral measurements with equal amounts of soluble GFP show that with excitation at 488 nm, fluorescence per unit soluble GFP is between 19- and 35-fold higher for the three mutants than for wt GFP (Table II). Interestingly, while the fluorescence per unit of soluble protein for mutants 2 and 3 is approx. half that for mutant 1, bacteria expressing mutants 2 and 3 fluoresce with more than twice the intensity of those expressing mutant 1 (Fig. 4). The difference is most likely due to the folding characteristics of the three mutant proteins. This interpretation is consistent with the fact that some GFPmut1 protein is found in inclusion bodies, whereas virtually all of GFPmut2 or GFPmut3 is soluble. The increase in fluorescence seen in bacteria expressing GFP mutants 1–3 is thus due to a shift in absorption spectra, and in addition to an increase in soluble protein due to more efficient protein folding.

(d) Advantages of the GFP mutants

Previous mutagenesis of GFP has revealed two broad classes of mutants: those that fluoresce blue (emission wavelength 458 nm) when excited by UV light (Heim et al., 1994), and those with shifts in excitation maximum (to a maximum of 490 nm). Mutation T66H results in a shift in emission maxima to 448 nm (Heim et al., 1994). Mutation of Ser^{65} to Ala, Gly, Ile, Thr, or Cys results in

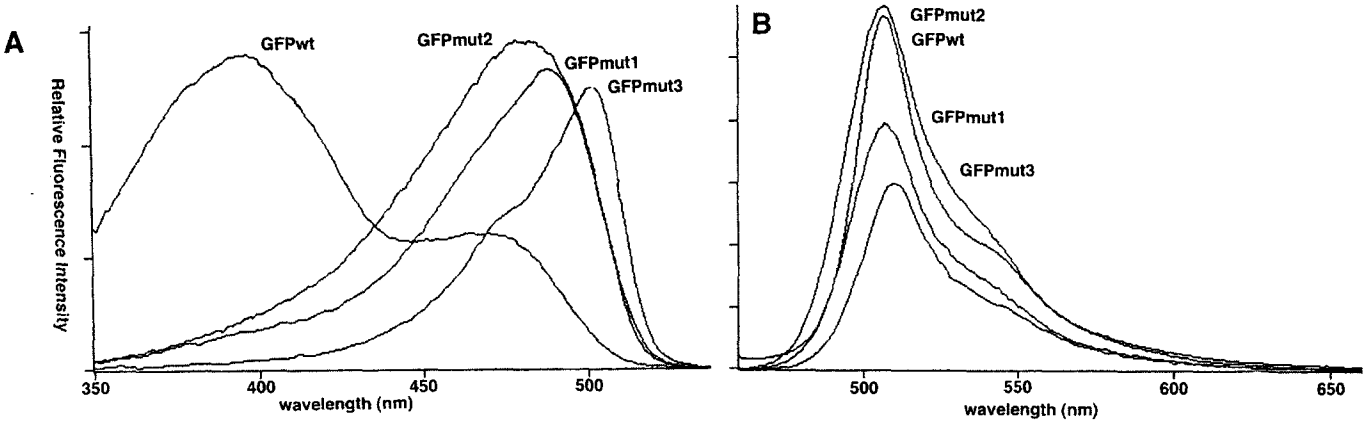

Fig. 4. Excitation and emission spectra for wt and mutant GFPs. Spectra were measured on a SPEX fluorolog fluorimeter using 1.0 nm band widths. Emission and excitation spectra were corrected using standard correction files supplied by the manufacturer. The scales for individual mutants have been arbitrarily chosen for purposes of display; comparisons of amplitudes should not be made, therefore, between different proteins. For comparison between proteins, see Table II. (**A**) Excitation spectra, with emission recorded at 540 nm. Excitation maxima: wt GFP, 395 nm; GFPmut1, 488 nm; GFPmut2, 481 nm; GFPmut3, 501 nm. (**B**) Emission spectra, recorded at excitation wavelength of 450 nm for the mutant proteins and 395 nm for the wt protein. Emission maxima: wt GFP, 508 nm; GFPmut1, 507 nm; GFPmut2, 507 nm; GFPmut3, 511 nm.

TABLE II

Fluorescence emission intensities of GFP mutants

	Fluorescence intensity (relative to wt)	
	395[a]	488[b]
wt GFP	1	1
GFPmut1	0.8	35
GFPmut2	0.5	19
GFPmut3	0.2	21

[a] Ratio of emission maxima for each GFP variant to wt. Excitation at 395 nm.

[b] Ratio of emission maxima for each GFP variant to wt. Excitation at 488 nm.

Methods: Strains expressing the GFP variants were grown to saturation at 30°C in 2 × YT supplemented with 100 μg Ap/ml and 0.2mM IPTG. At this temperature, all four GFP variants are expressed at approximately the same level, and the variant proteins are all more than 90% soluble. Equal numbers of cells for each strain were lysed using a French press. The lysates were clarified by centrifugation at 17000 × *g* for 30 min. We took measurements on equal amounts of GFP, as determined by densitometry of 0.1% SDS-12% PA gels stained with Coomassie blue.

large shifts in excitation maxima (Heim et al., 1994; Delagrave et al., 1995). The mutations isolated in our study are clearly related to those in this second class. We isolated at least three independent substitutions of Ser^{65} (S65A and 2 codons for S65G; see Table I). These changes were invariably accompanied by a second substitution, S72A. The double aa combination of S65G, S72A is present in the library at a theoretical frequency of 1 in 3.6×10^5 clones. The fact that we isolated the same double aa substitution independently at least twice suggests strongly that the substitutions at both aa positions are important for maximal fluorescence. Likewise, while S65T results in a fluorescence intensity increase of six-fold (Heim et al., 1995), S65T, F64L results in a 30-fold increase over wt. This also suggests that the presence of the second mutation is critical for maximal fluorescence. The additional mutations may be necessary for efficient folding or chromophore formation; alternately they may have a more direct effect on absorption efficiency.

The mutant GFPs described here should be useful in a wide variety of applications These mutants have a clear advantage over wt GFP for studies using excitation in the range of 489 nm. In bacteria growing in log phase at 37°C, wt GFP is not easily detectable until 1–2 h after induction, limiting its use as a reporter in gene expression studies. Under identical conditions, fluorescence from the mutant proteins is detectable within 8 min. Should these mutants show similar effects in other systems, they will be of clear utility as reporters of gene expression, even at early time points after induction. As has been suggested by others (Heim et al., 1994; Delagrave et al., 1995), pairing these mutants (particularly GFPmut3 which has very low excitation in the UV range) with complementary GFP mutants having a specifically reduced 470 nm absorption peak could allow efficient spectral separation of two simultaneously expressed fluorescent tags. Lastly, because of their novel properties in protein folding, and potentially in chromophore formation, the mutant GFPs may be fluorescent in systems where wt GFP has proven to be non-or only weakly fluorescent.

(e) Conclusions

(*1*) We have isolated three GFP mutants. When expressed in *E. coli*, they fluoresce with approx. 100-fold higher intensity than bacteria expressing wt protein.

(*2*) The excitation maxima of the three mutants are

red-shifted by about 100 nm, permitting efficient excitation at 488 nm.

(*3*) When excited at 488 nm and equalized for total amount of soluble GFP, the mutants fluoresce between 20- and 35-fold more intensely than wt.

(*4*) These mutants should have a wide applicability in any GFP studies using standard FITC excitation-emission filters.

ACKNOWLEDGEMENTS

We gratefully acknowledge Mark Troll for fluorimetric analysis, Harry Noller for use of his oligo synthesizer and Rachel Green for help with oligo synthesis and for a careful reading of the manuscript; we thank Greg Verdine for supplying pKEN2 and Stan Tabor for pT7.7. This work was supported by a grant from the NIH (AI 36396), and by an unrestricted gift from Praxis Pharmaceuticals to S.F. B.P.C. is supported by a Helen Hay Whitney postdoctoral fellowship.

REFERENCES

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.: Green fluorescent protein as a marker for gene expression. Science 263 (1994) 802–805.

Cormack, B.P. and Struhl, K.: Regional codon randomization: defining a TATA-binding protein surface required for RNA polymerase III transcription. Science 262 (1993) 244–248.

Delagrave, S., Hawtin, R.E., Silva, C.M., Yang, M.M. and Youvan, D.C.: Red-shifted excitation mutants of the green fluorescent protein. Biotechnology 13 (1995) 151–154.

Glaser, S.M., Yelton, D.E. and Huse, W.D.: Antibody engineering by codon-based mutagenesis in a filamentous phage vector system. J. Immunol. 149 (1992) 3903–3913.

Heim, R., Cubitt, A.B. and Tsien, R.Y.: Improved green fluorescence. Nature 373 (1995) 663–664.

Heim, R., Prasher, D.C. and Tsien, R.Y.: Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc. Natl. Acad. Sci. USA 91 (1994) 12501–12504.

Morise, J.G., Shimomura, O., Johnson, F.H. and Winant, J.: Intermolecular energy transfer in the bioluminescent system of *Aequorea*. Biochemistry 13 (1974) 2656–2662.

Valdivia, R.H., Hromockyj, A.E., Monack, D., Ramakrishnan, L. and Falkow, S.: Applications for the green fluorescent protein (GFP) in the study of host-pathogen interactions. Gene 173 (1996) 47–52.

Wang, S. and Hazelrigg, T.: Implications for *bcd* mRNA lacalization from spatial distribution of *exu* protein in *Drosophila* oogenesis. Nature 369 (1994) 400–403.

Ward, W.W., Cody, C.W., Hart, R.C. and Cormier, M.J.: Spectrophotometric identity of the energy-transfer chromophores in *Renilla* and *Aequorea* green fluorescent proteins. Photochem. Photobiol. 31 (1980) 611–615.