The molecular structure of green fluorescent protein

Fan Yang, Larry G. Moss[1], and George N. Phillips, Jr.*

Department of Biochemistry and Cell Biology and the W.M. Keck Center for Computational Biology, Rice University, 6100 S. Main St., Houston, TX 77005-1892 and [1]Division of Endocrinology, Department of Medicine, Tufts University School of Medicine and the New England Medical Center, Boston, MA 02111.
**Corresponding author (e-mail:georgep@rice.edu).*

Received 30 July 1996; accepted 16 August 1996

The crystal structure of recombinant wild-type green fluorescent protein (GFP) has been solved to a resolution of 1.9 Å by multiwavelength anomalous dispersion phasing methods. The protein is in the shape of a cylinder, comprising 11 strands of β-sheet with an α-helix inside and short helical segments on the ends of the cylinder. This motif, with β-structure on the outside and α-helix on the inside, represents a new protein fold, which we have named the β-can. Two protomers pack closely together to form a dimer in the crystal. The fluorophores are protected inside the cylinders, and their structures are consistent with the formation of aromatic systems made up of Tyr^{66} with reduction of its Cα–Cβ bond coupled with cyclization of the neighboring glycine and serine residues. The environment inside the cylinder explains the effects of many existing mutants of GFP and suggests specific side chains that could be modified to change the spectral properties of GFP. Furthermore, the identification of the dimer contacts may allow mutagenic control of the state of assembly of the protein.

Keywords: fluorescence, protein structure, β-can motif, GFP

Green fluorescent protein (GFP) is a spontaneously fluorescent protein isolated from coelenterates, such as the Pacific jellyfish, *Aequorea victoria*[1]. Its role is to transduce, by energy transfer, the blue chemiluminescence of another protein, aequorin, into green fluorescent light[2]. The molecular cloning of GFP cDNA[3] and the demonstration by Chalfie and colleagues[4] that GFP can be expressed as a functional transgene have opened exciting new avenues of investigation in cell, developmental, and molecular biology. Fluorescent GFP has been expressed in bacteria[4], yeast[5], slime mold[6], plants[7,8], *Drosophila*[9], zebrafish[10], and in mammalian cells[11,12]. GFP can function as a protein tag, because it tolerates N-terminal and C-terminal fusion to a broad variety of proteins many of which have been shown to retain native function[6,9,13,14]. When expressed in mammalian cells fluorescence from wild-type GFP is typically distributed throughout the cytoplasm and nucleus, but excluded from the nucleolus and vesicular organelles[13]. However, highly specific intracellular localization including the nucleus, mitochondria[15], secretory pathway[16], plasma membrane[17], and cytoskeleton[5] can be achieved via fusions both to whole proteins and individual targeting sequences. The enormous flexibility as a noninvasive marker in living cells allows for numerous other applications such as a cell lineage tracer, reporter of gene expression, and as a potential measure of protein-protein interactions[18].

GFP is composed of 238 amino acids. Its wild-type absorbance/excitation peak is at 395 nm with a minor peak at 475 nm with extinction coefficients of roughly 30,000 and 7,000/M/cm, respectively[19]. The emission peak is at 508 nm. Interestingly, excitation at 395 nm leads to decrease over time of the 395 nm excitation peak and a reciprocal increase in the 475 nm excitation band[13]. This presumed photoisomerization effect is especially evident with irradiation of GFP by ultraviolet (UV) light. Analysis of a hexapeptide derived by proteolysis of purified GFP led to the prediction that the fluorophore originates from an internal Ser-Tyr-Gly sequence, which is post-translationally modified to a 4-(*p*-hydroxybenzylidene)-imidazolidin-5-one structure[20]. Studies of recombinant GFP expression in *Escherichia coli* led to a proposed sequential mechanism initiated by a rapid cyclization between Ser 65 and Gly 67 to form an imidazolin-5-one intermediate followed by a much slower (hours) rate-limiting oxygenation of the Tyr 66 side chain by O_2 (ref. 21). Combinatorial mutagenesis suggests that the Gly 67 is required for formation of the fluorophore[22]. While no known cofactors or enzymatic components are required for this apparently autocatalytic process, it is rather thermosensitive with the yield of fluorescently active to total GFP protein decreasing at temperatures greater than 30°C[23]. However, once produced, GFP is quite thermostable.

Physical and chemical studies of purified GFP have identified several important characteristics. It is very resistant to denaturation requiring treatment with 6 M guanidine hydrochloride at 90°C or pH of <4.0 or >12.0. Partial to near total renaturation occurs within minutes following reversal of denaturing conditions by dialysis or neutralization[24]. Circular dichroism predicts significant amounts of β-sheet structure that is subsequently lost on denaturation[24]. Over a nondenaturing range of pH, increasing pH leads to a reduction in fluorescence by 395 nm excitation and an increased sensitivity to 475 nm excitation[25]. Reduction of purified GFP by sodium dithionite results in a rapid loss of fluorescence that slowly recovers in the presence of room air. While insensitive to sulfhydryl reagents such as 2-mercaptoethanol, fluorescence is irreversibly eliminated by treatment with the sulfhydral reagent dithiobisnitrobenzoic acid[26].

The availability of *E. coli* clones expressing GFP has led to extensive mutational analysis of GFP function. Truncation of more than seven amino acids from the C-terminus, or more than the N-terminal Met, led to total loss of fluorescence[27]. All nonfluorescent mutants also failed to exhibit absorption spectra characteristic of the intact fluorophore, implying a possible defect in post-translational processing. Screens of random and directed point mutations for changes in fluorescent behavior have uncovered a number of informative amino acid substitutions. Mutation of Tyr 66 in the fluo-

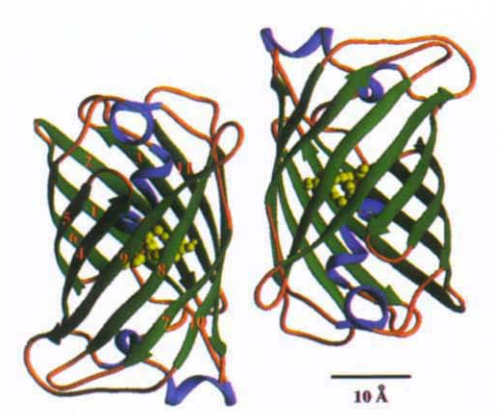

Figure 1. The overall shape of the protein and its association into dimers. Eleven strands of β-sheet (green) form the walls of a cylinder. Short segments of α-helices (blue) cap the top and bottom of the 'β-can' and also provide a scaffold for the fluorophore, which is near geometric center of the can. This folding motif, with β-sheet outside and helix inside, represents a new class of proteins. Two monomers are associated into a dimer in the crystal and in solution at low ionic strengths. This view is directly down the twofold axis of the noncrystallographic symmetry. Figures 1 and 6 were produced with Ribbons[46].

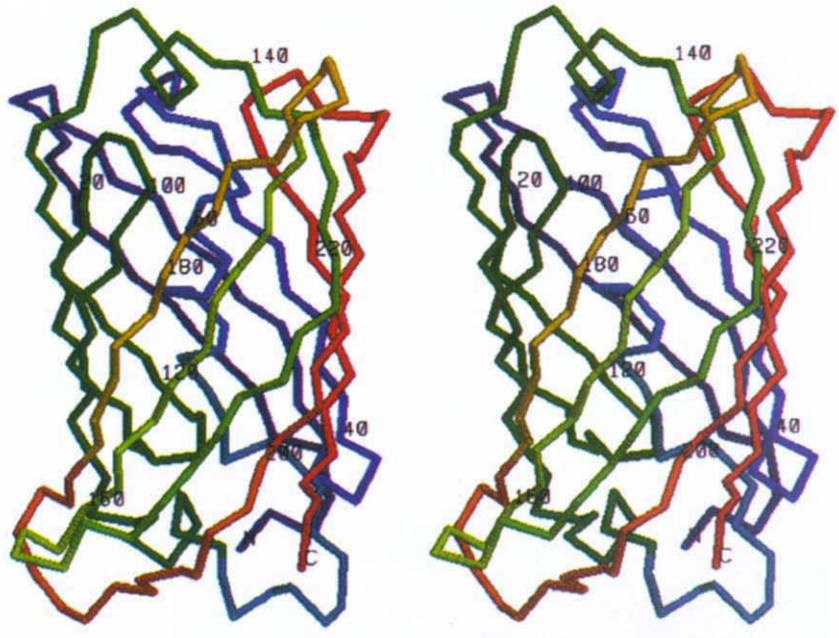

Figure 2. Stereo view of a monomer, with colors that vary slowly as a function of the distance along the polypeptide chain. The termini and Cα atoms of every twentieth amino acid are marked just to the upper right of each atom. Figure produced by RasMol46.

rophore to His results in a shift of the excitation maximum to the UV (383 nm) with emission now in the blue at 448 nm (ref. 21). A Tyr 66-Trp mutant is blue-shifted, albeit to a lesser degree. Both changes are associated with a severe weakening of fluorescence intensity compared with wild-type GFP. Mutation of Ser 65 to Thr, Ala, Cys, or Leu causes a loss of the 395-nm excitation peak with a major increase in blue excitation[22,28]. When combined with Ser 65 mutants, mutations at other sites near the fluorophore such as Val 68-Leu and Ser 72-Ala can further enhance the intensity of green fluorescence produced by excitation at 488 nm (refs. 22 and 29). However, amino acid substitutions significantly outside this region also affect the protein's spectral character. For example, Ser 202-Phe and Thr 203-Ile both cause the loss of excitation in the 475 nm region with preservation of 395 nm excitation[4,21,30]. Ile 167-Thr results in a reversed ratio of 395 to 475 nm sensitivity[13], while Glu 222-Gly is associated with the elimination of only the 395 nm excitation[30]. Another change, Val 163-Arg, not only enhances the magnitude of the Ser 65-Thr mutant, but also increases the temperature tolerance for functional GFP expression[19]. Molecular evolution techniques have been reported to improve GFP fluorescence[31]. Unfortunately, a roster of substitutions associated with complete loss of function has not been published.

Because GFP *in crystallum* exhibits a fluorescence spectrum and life span nearly identical to that of GFP in aqueous solution[32] and fluorescence is not an inherent property of the isolated fluorophore, the elucidation of its three-dimensional structure will help provide an explanation for the generation of fluorescence in the mature protein, as well as the mechanism of autocatalytic fluorophore formation. Furthermore, the development of fluorescent proteins with additional emission and excitation characteristics would dramatically expand their biological applications. Color vision is based on the fact that spectral properties of a common fluorophore, *cis*-retinal, are altered as a function of protein environment within red, blue, or green opsins[33]. The GFP from the sea pansy, *Renilla reniformis*, which exhibits a single major excitation peak at 498 nm, apparently uses an identical core fluorophore to that of *A. victoria* GFP. These findings, taken together with the spectral changes exerted by substitutions in amino acids over 100 residues from the GFP fluorophore, suggest that a rational strategy to modify and expand the fluorescence behavior of GFP based on protein structure may be possible. We report the X-ray diffraction structure derived from a crystal of wild-type, recombinant *A. victoria* green fluorescent protein.

Results

The structure of GFP has been solved using seleniomethionyl-substituted protein and multiwavelength anomalous dispersion (MAD) phasing methods. The electron density maps produced by the MAD phasing were very clear, revealing a dimer comprised of two quite regular β-barrels with 11 strands on the outside of cylinders (Figs. 1–3). These cylinders have a diameter of about

Table 1. List of amino acid side chains with close contacts (less than 5 Å) to the fluorophore.

Protein residue	Protein atom	Fluorophore residue	Fluorophore atom	Distance (Å)
Arg 96	NH2	Tyr 66	O	2.7
Gln 94	NE2	Tyr 66	O	3.0
His 148	ND1	Tyr 66	OH	3.3
Gln 69	CD	Tyr 66	O	3.4
Glu 222	OE2	Tyr 66	CE2	3.5
Val 150	CG2	Tyr 66	CE1	3.6
Phe 165	CE1	Tyr 66	CD1	3.6
Thr 203	CG2	Tyr 66	CE2	3.6
Ile 167	CD1	Tyr 66	OH	3.7
Thr 62	CG2	Tyr 66	CG	3.7
Tyr 145	CE2	Tyr 66	OH	3.7
Ser 205	OG	Tyr 66	CE2	4.0
Val 61	CG1	Tyr 66	CE2	4.4
Gln 183	NE2	Tyr 66	O	4.8
Val 68	CG2	Ser 65	C	4.9

The fluorophore is defined as the seven atoms of the phenol of Tyr 66, the six atoms of the imidazolidone, and the bridging methylene between the rings. The amino acid side chains listed here would be expected to have the most direct effects on fluorescence and perhaps fluorophore formation. The atom names are taken from the Brookhaven Protein Data Bank nomenclature.

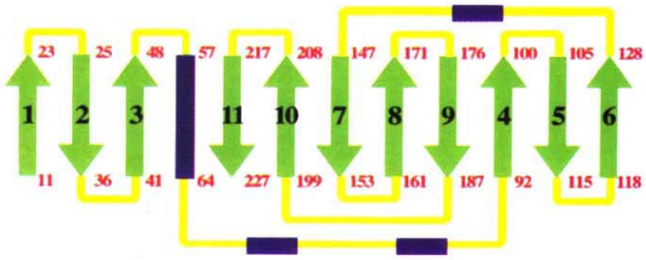

Figure 3. A topology diagram of the folding pattern in GFP. The ß-sheet strands are shown in light green, α-helices in blue, and connecting loops in yellow. The positions in the sequence that begin and end each major secondary structure element are also given. The antiparallel strands (except for the interactions between stands 1 and 6) make a tightly formed barrel.

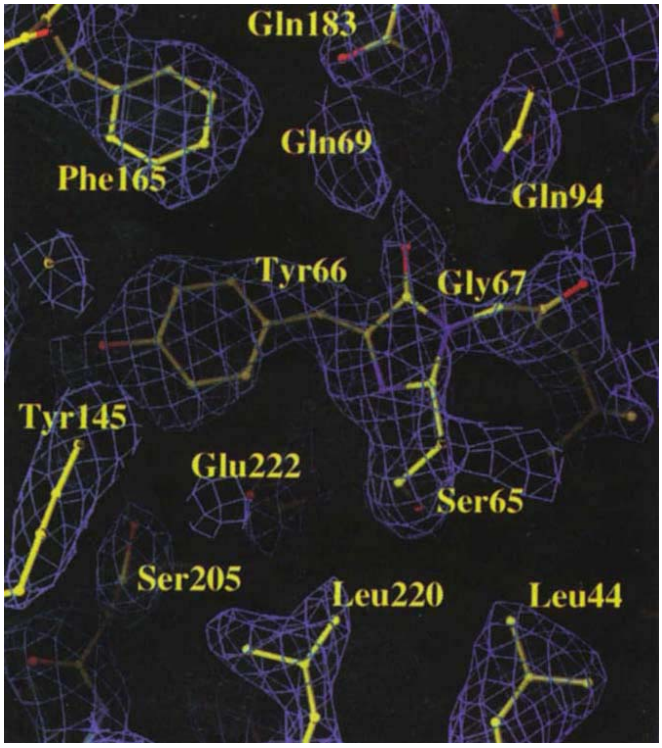

Figure 4. Model of the fluorophore and its environment superimposed on the MAD-phased electron density map at 2.2 Å resolution. The clear definition throughout the map allowed the chain to be traced and side chains to be well placed. The density for Ser 65, Tyr 66 and Gly 67 is quite consistent with the dehydrotyrosine–imidazolidone structure proposed for the fluorophore. Many of the side chains adjacent to the fluorophore are labeled. Figures 4 and 5 were produced with O (ref. 43).

30 Å and a length of about 40 Å. Inspection of the density within the cylinders reveals modified tyrosine side chains as a part of an irregular α-helical segment (Fig. 4). Small sections of α helix also form caps on the ends of the cylinders. This motif, with a single α-helix inside a very uniform cylinder of β-sheet structure, represents a new protein class, as it is not similar to any other known protein structure.

The fluorophore is highly protected, located on the central helix within a couple of angstroms of the geometric center of the cylinder. The pocket containing the fluorophore has a surprising number of charged residues in the immediate environment (Fig. 5, Table 1). The environment around the fluorophore includes both apolar and polar amino acid side chains. Phe 64 and Phe 46 are near the fluorophore and separate the single tryptophan, Trp 63 from direct contact with fluorophore (closest distance of 13 Å). A table of all atoms that come in contact with the fluorophore and their distances to the fluorophore is provided (Table 1).

The crystallographic contacts are all rather tenuous, consisting of a few amino acid side chains for each. The noncrystallographic symmetry is maintained by extensive contacts and thus is likely to be the source of the dimerization seen in solution studies (Fig. 6). The dimer contacts are fairly tight and consist of a core of hydrophobic side chains from Ala 206, Leu 221, and Phe 223 from each of the two monomers and a wealth of hydrophilic contacts (Fig. 6), including Tyr 39, Glu 142, Asn 144, Asn 146, Ser 147, Asn 149, Tyr 151, Arg 168, Asn 170, Glu 172, Tyr 200, Ser 202, Gln 204, and Ser 208. Contacts with other crystallographic molecules are not extensive, and the salt-dependence of this dimer interface and/or the loose contacts with neighboring molecules may explain the difficulties with isomorphism in initial heavy atom phasing studies.

Mass spectrometry studies of the bacterially expressed wild-type and selenio-methionyl protein show masses of 26836.1 (±0.9) and 27069.3 (±1.4) g/mol, respectively. The masses calculated for the known TU#58 gene sequence, including the original inadvertent Gln 80-Arg PCR error during the cloning of the gene for GFP[4] and the cyclization and oxidation of the tyrosine are 26835.5 and 27070.0 for the seleniomethionine, respectively. The differences of 0.6 and –0.7 g/mol are small and the results are therefore consistent with essentially complete fluorophore formation, including the loss of water after cyclization. The error limits do not allow accurate determination of the degree of oxidation of the dehydrotyrosine, however. These results indicate that the starting material for the crystallization was essentially fully formed GFP, and the lack of difference density in Fo-Fc maps in this region shows that the crystal contains fully cyclized GFP.

Discussion

The remarkable cylindrical fold of the protein seems ideally suited for its function. The strands of β-sheet are tightly fitted to each other like staves in a barrel and form a regular pattern of hydrogen bonds. Together with the short α-helices and loops on the ends, the 'can' structure forms a single compact domain and does not have obvious clefts for easy access of diffusable ligands to the fluorophore. This fold, taken with the observation that the fluorophore is near the geometric center of the molecule, explains the observed protection of the fluorophore from collisional quenching by oxygen (K_{bm} < 0.004/M/sec)[34] and hence reduction of the quantum yield. Perhaps more seriously, photochemical damage by the formation of singlet oxygen through intersystem crossing is reduced by the structure. The tightly constructed β-can would appear to serve this role nicely, as well as provide overall stability and resistance to unfolding by heat and denaturants.

The location of certain amino acid side chains in the vicinity of the fluorophore also begins to explain the fluorescence and the behavior of certain mutants of the protein. At least two resonant forms of the fluorophore can be drawn, one with a partial negative charge on the benzyl oxygen of Tyr 66, and one with the charge on the carbonyl oxygen of the imidazolidone ring. Interestingly, basic residues appear to form hydrogen bonds with each of these oxygen atoms, His 148 with Tyr 66 and Gln 94, and Arg 96 with the imidazolidone. These bases presumably act to stabilize and possibly further delocalize the charge on the fluorophore. Most of the other polar residues in the pocket form an apparent hydrogen-bonding network on the side of Tyr 66 that requires abstraction of protons in the oxidation process. It is tempting to speculate that these residues help abstract the protons. As for the mutants, atoms in the

side chains of Thr 203, Glu 222, and Ile 167 are in van der Waals contact with Tyr 66, so their mutation would have direct steric effects on the fluorophore and would also change its electrostatic environment if the charge were changed, as suggested[30]. A quantitative explanation will require further examination. It seems likely that other mutations of the residues near the fluorophore would also have effects on the absorption and/or emission spectra, and such experiments to change the electrostatic environment around the fluorophore are in progress. By virtue of their varied fluorophore environments and hence altered spectra, these mutants should lead to expanded uses of green fluorescent protein as gene markers and cell lineage markers, and should encourage other uses in biotechnology.

Mutations in regions of the sequence adjacent to the fluorophore, i.e. in the range of positions 65–67, have been systematically explored[22]; some cause significant wavelength shifts and most cause a loss of fluorescence intensity. For example, mutation of the central Tyr to Phe or His shifts the excitation bands, but there is an overall loss of intensity. Secondary mutations to compensate for the deleterious intensity effects should also now be possible. The Ser 65-Thr mutant is particularly interesting because of its reported increase in fluorescence intensity[21,28]. The mechanism for increased fluorescence may be reduced collisional quenching, because the additional methyl group may make for better packing in the interior of the protein. On the other hand, the effect has also been suggested to be through improved conversion of the tyrosine to dehydrotyrosine. However, the fact that we see significantly altered structure relative to standard protein conformations in the wild-type argues against a dramatic increase in cyclization and/or oxidation. This effect is most likely produced by increased expression and/or folding of the protein. The crystal structure of the Ser 65-Thr mutant has also been solved[35], and it will be interesting to compare the two structures for clues about the fluorescence and other differences. The report of improvements in GFP by DNA shuffling[31], comprising mutations Phe 99-Ser, Met 153-Thr and Val 163-Ala are difficult to explain based on the structure. Positions 153 and 163 are on the surface of the protein and may exert their effects through improved solubility and/or reduced aggregation as proposed. The Phe-Ser mutation at first glance would appear to destabilize the core of the protein and thus we do not know how it would improve the system.

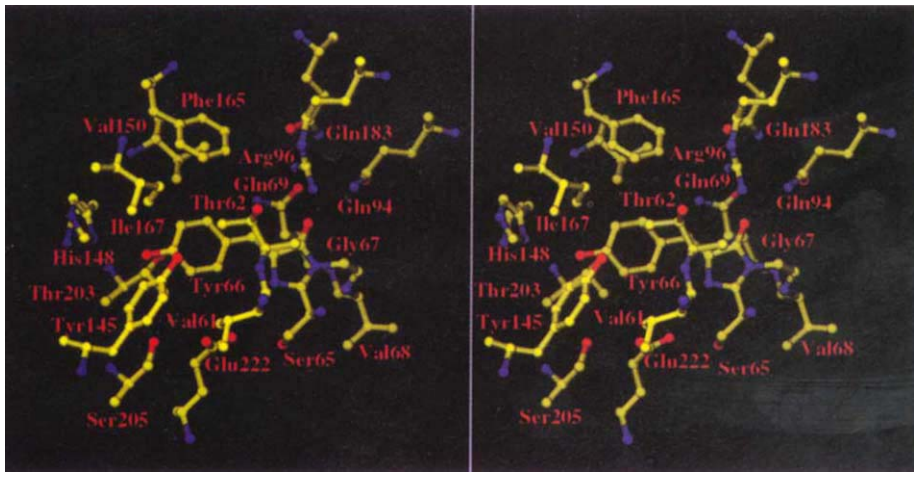

Figure 5. Stereo view of the fluorophore and its environment. His 148, Gln 94 and Arg 96 can be seen on opposite ends of the fluorophore and probably stabilize resonant forms of the fluorophore. Charged, polar, and nonpolar side chains all contact the fluorophore in some way.

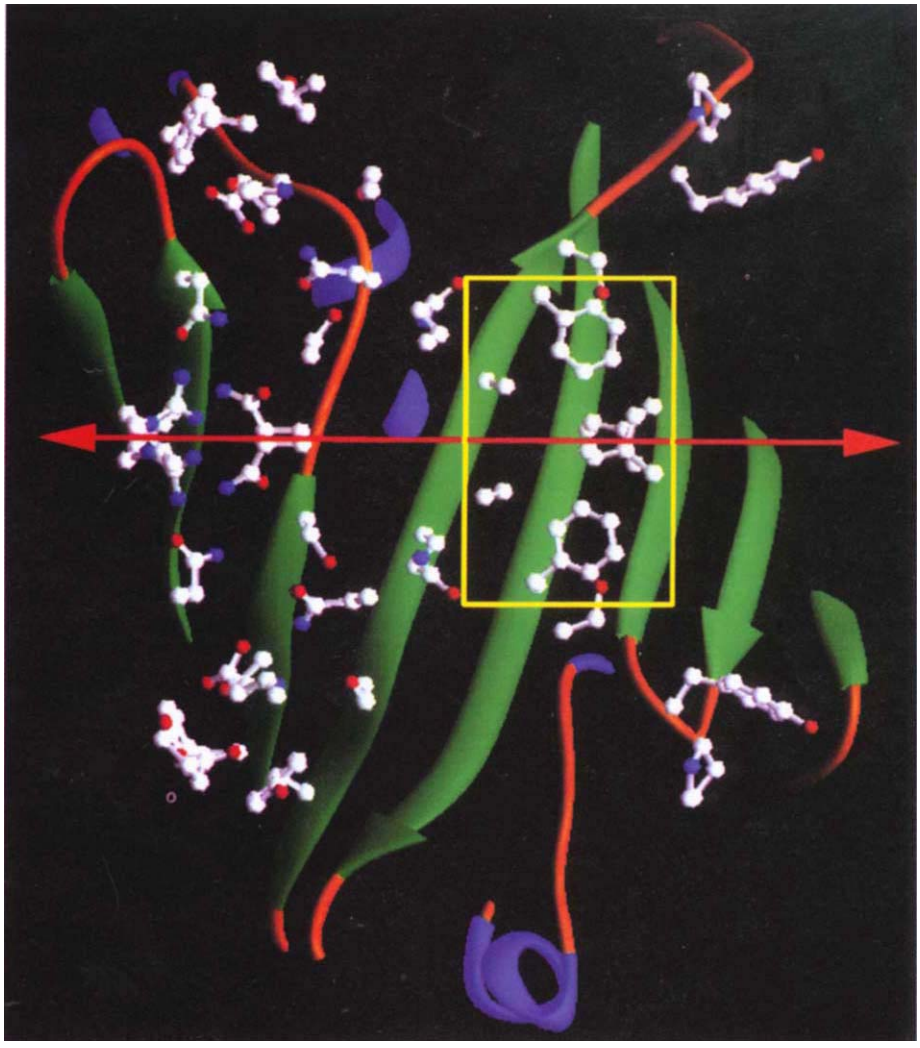

Figure 6. The dimer contact region. The two polypeptide chains associate over a broad area, with a small hydrophobic patch (in the yellow box) and numerous hydrophilic contacts. The twofold symmetry axis is in the plane of the page, and is marked by the red arrow. The polar residues are marked with red for oxygen atoms and blue for nitrogen.

Table 2. Anomalous diffraction differences and scattering factors for seleniomethionyl GFP at the four wavelengths used.

Wavelength (Å)	30 > d > 3.4 (Å)				3.4 > d > 2.7 (Å)				2.7 > d > 2.2 (Å)				scattering factors (e)	
	0.9879	0.9794	0.9792	0.9686	0.9879	0.9794	0.9792	0.9686	0.9879	0.9794	0.9792	0.9686	f′	f″
0.9879	0.026 (0.026)	0.042	0.030	0.020	0.037 (0.035)	0.046	0.037	0.032	0.064 (0.052)	0.071	0.068	0.064	−4.0	1.1
0.9794		0.050 (0.029)	0.026	0.045		0.058 (0.039)	0.035	0.049		0.088 (0.060)	0.065	0.077	−10.5	3.9
0.9792			0.067 (0.031)	0.032			0.074 (0.042)	0.040			0.101 (0.062)	0.071	−7.9	5.5
0.9686				0.049 (0.027)				0.059 (0.039)				0.089 (0.064)	−3.4	3.9

Following the format used by Yang et al.[41], Bijvoet differences ratios are given in diagonal elements with centric values in parentheses, and dispersive differences are given in off-diagonal elements. Scattering factors were chosen to minimize the cumulative errors in Bijvoet and dispersive terms.

The mechanism of activation of the fluorophore from ordinary protein structure is consistent with a nonenzymatic cyclization mechanism like that of Asn-Gly deamidation[36] followed by oxidation of the tyrosine to dehydrotyrosine, as previously suggested. The role of molecular oxygen in this mechanism and in GFP fluorescence is paradoxical, however. Molecular oxygen is proposed to be needed for formation of the double bond between Cα and Cβ on the tyrosine to form an extended aromatic system, but oxygen must also be excluded from regular interactions with the fluorophore or else collisional quenching of the fluorescence or damaging photochemistry will occur. The low bimolecular quenching rate suggests that the protein's design sacrifices efficient fluorophore formation for stability and higher quantum yields once fully formed.

The excited state dynamics of GFP have been studied using Stark, steady state, and time-resolved fluorescence spectroscopies[37] (and Youvan and Michel-Beyerle, personal communication). The results suggest that proton transfer is involved in interconversions within two grounded and two excited states. The extended set of polar interactions around the flourophore could easily accomodate proton rearrangements, with the most likely direct effects being between the His 148 with the hydroxyl of Tyr 66, Arg 96 interactions with the imidazolidone, or Glu 222 interactions with the hydroxyl of Ser 65. Since the Ser 65/Glu 222 mutants have both lost their native interactions together with their approximately 400 nm absorption bands, one possibility is that the 400-nm band arises from the abstraction of the Ser 65 hydroxyl proton by Glu 222. This is speculation, but similar spectroscopic studies on mutants at these positions may be able to differentiate the roles of these sites in excited state dynamics.

The N-termini and C-termini truncation studies[27] and the fluorescent fusion products[6,13,14] are now understandable, given the structure of the protein. Since the C-terminus loops back outside the cylinder and the last seven or so amino acids are disordered it should not be critical to have them present, and further addition would seem to be easily tolerated. These residues do not form a stave of the barrel. The role of the N-terminus is a little less clear, because the first strand in the barrel does not begin until amino acid 10 or 11. Thus, barrel formation does not require the N-terminal region. The N-terminal segment is, however, an integral part of the 'cap' on one end of the protein and may be essential in folding events or in protecting the fluorophore. Again, extensions at the N-terminus would not disrupt the motif structure of the protein.

The chemical modification studies[26] using sulfhydral reagents can be partially explained. Reaction of one of the cysteines near one end of the cylinder, Cys 70, would appear to disrupt the packing of the cap on that end, and therefore allow quenching of the fluorophore. Significant fluorescence intensity effects by the modification of Cys 48 on the exterior of the protein would not be expected, *a priori*. The structure determination of the dithionite-reduced, nonfluorescent species has not yet been studied, but should provide additional data on the nature of the fluorophore. The pH dependence of the excitation bands at 395 nm and 475 nm (ref. 25) is almost certainly due to His 148, whose Nδ atom is 3.3 Å from the Tyr 66 hydroxyl oxygen atom of the fluorophore, although nuclear magnetic resonance pKa measurements or mutagenesis studies would be needed for confirmation.

The dimer we see as the asymmetric unit in the crystal is likely to be the same one formed in solution, because the ionic strength of the crystallization buffer is low, and we see dimers at low (<100 mM) ionic strengths in solution. Thus, it is not surprising to see the large number of hydrophilic dimer contacts. The smaller hydrophobic patch could conceivably be involved in physiological interactions with aequorin, as there would be a natural advantage to close proximity for efficient energy transfer. It is not known at present whether dimers form in physiological circumstances, or what the effect of dimerization is on energy transfer, aside from the circumstantial inferences on the excitation spectra previously reported on the native protein[25]. The dimer contacts should now be able to be modified in such a way as to disrupt the formation of dimers without affecting stability and folding. Other nearby residues could also be converted to hydrophobic residues to enhance dimer stability if desired. Control of the dimerization will be important for fluorescence resonance energy transfer studies of protein-protein interactions using GFP, as one would not want to induce association and therefore resonance energy transfer between the differently colored GFP proteins by mechanisms other than that of the target protein interactions. Mutants may also be developed for reduction of aggregation during expression and hence fewer problems with inclusion bodies.

The three-dimensional structure of GFP has provided a physicochemical basis for many observed features of the protein, including its stability, protection of its fluorophore, behavior of mutants, and dimerization properties. The structure will also allow directed mutation studies to complement random and combinatorial approaches.

Experimental protocol

Purification. GFP was purified from *E. coli* strain BL21(DE3)pLysS (Novagen, Madison, WI) containing plasmid TU#58, bearing the wild-type *A. victoria* green fluorescent protein[4]. For the seleniomethionine protein, the plasmid was moved to *E. coli* methionine auxotroph strain B834(DE3)pLysS (Novagen). The purification involved cell lysis, centrifugation of cell debris, and four column chromatography steps: DEAE anion exchange column (Sigma, St. Louis, MO, CL-6B) with a zero- to 1-M NaCl gradient in 10 mM phosphate, 2 mM EDTA, 2 mM DTT, pH 7; a hydrophobic interaction column (Sigma, CL-4B) with a 0.1- to zero-M phosphate gradient in 2 mM

EDTA, 2 mM DTT, pH 7; an HPLC anion exchange column (Bio-Rad, Hercules, CA, Bio-Gel DEAE-5PW) with a zero- to 1-M NaCl gradient in 10 mM phosphate, 2 mM EDTA, 2 mM DTT, pH 7; and an HPLC gel filtration column (Bio-Rad, Bio-Gel SEC-125) with 0.1 M phosphate, 2 mM EDTA, 2 mM DTT, pH 7. Gel filtration columns run at 10 mM phosphate showed predominately a twofold higher molecular weight species. Matrix-assisted laser desorption ionization mass spectrometry was performed by the University of Texas Health Sciences Center analytical chemistry service.

Crystallization. The protein was crystallized in sitting drop vapor diffusion wells (Hampton Research) at room temperature using 58% 2-methyl-2,4-pentanediol (Aldrich), 50 mM morpholino ethane sulfonic acid, 0.1% sodium azide at pH 6.8. The protein concentration varied, but was typically 20 to 30 mg/ml. Crystals grew as green fluorescent square bipyramids up to 0.5 mm on a side. The space group was determined to be $P4_12_12$ or its enantiomorph, with a=b=87.15 Å and c=119.85 Å at cryogenic temperatures, and a=b=89.23 Å and c=119.78 Å at room temperature. The unit cell also varies with changes in ionic strength, and this effect thwarted solution by multiple isomorphous replacement. Packing density calculations suggested that there were probably two molecules per asymmetric unit.

Multiwavelength anomalous dispersion. Data were taken at Brookhaven beam line X4A at four wavelengths. The wavelengths to be used were determined by reference and crystal absorption scans. The data were taken at liquid nitrogen temperature using inverse-beam geometry in wedges of 4° and processed using DENZO[38]. Native and selenio-methionine data sets were also taken in the laboratory on an R-AXIS IIC detector with CuKα radiation. The native data set used in the refinement had an R_{merge} of 7.7% to 1.9 Å resolution (99+% complete in all shells with fivefold average redundancy). Selenium atoms were located initially by standard difference Patterson maps between selenium-substituted and native protein using SHELX96 (ref. 39) and HEAVY (ref. 40) and confirmed by Patterson maps using the MAD data. MADSYS software[41] was used to give the anomalous diffraction differences shown in Table 2 and to extract Fa, Ft, and phase information. The resulting MAD-phased map was solvent flattened and twofold averaged based on the selenium sites using CCP4 (ref. 42) and skeletonized using the program O (ref. 43); it immediately revealed two 11-stranded cylindrical β-barrels. The polypeptide chain was traced for one of the barrels beginning from the seleniomethionines and extending the structure in each direction, helped by the recognition of the modified tyrosine in the middle of the barrel as Tyr 66, the nucleus of the fluorophore. The correct enantiomorph is space group $P4_12_12$, as confirmed by the handedness of the β-barrel and the α-helices. Refinement has been started using the program X-Plor (ref. 44) using the native data collected at room temperature; the current R-factor at 1.9 Å is 0.21 with an R-free of 0.26, with good geometry (rms bond and angle deviations from ideality of 0.013 Å and 1.8°, respectively) and tight restraint of the noncrystallographic symmetry. All measured data were included in the refinement. Coordinates and structure factors have been deposited at the Brookhaven Protein Data Bank under accession numbers 1GFL and R1GFLSF, respectively.

Acknowledgments

We would like to thank J. Sobelewski and L. Moitoso-deVargas for initial purification of GFP; Dan Leahy for procedures for growing the seleniomethionine auxotroph and helpful suggestions; Mike Berry, Frank Whitby, Mike Soltis, Henry Bellamy, Michael Stowell, and Craig Ogata for help with MAD data collection; Frank Prendergast for suggesting the name β-can; the Howard Hughes Medical Institute and Brookhaven National Laboratory (beamline X4A) and Stanford's SSRL (beamline X1-5) for synchrotron time; and the W.M. Keck Foundation, Robert A. Welch Foundation, and NIH AR40252 (GNP), GRASP Center (DK34928) and DK34447 (LGM) for financial support.

1. Morin, J. and Hastings, J. 1971. Energy transfer in a bioluminescent system. *J. Cell Physiol.* **77**:313–318.
2. Ward, W. 1979. pp. 1–57 in *Photochemical and photobiological reviews*. K. Smith (ed.). Plenum Press, New York.
3. Prasher, D., Eckenrode, V., Ward, W., Prendergast, F., and Cormier, M. 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* **111**:229–233.
4. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W., and Prasher, D. 1994. Green fluorescent protein as a marker for gene expression. *Science* **263**:802–805.
5. Kahana, J., Schapp, B., and Silver, P. 1995. Kinetics of spindle pole body separation in budding yeast. *Proc. Natl. Acad. Sci. USA* **92**:9707–9711.
6. Moores, S., Sabry, J., and Spudich, J. 1996. Myosin dynamics in live *Dictyostelium* cells. *Proc. Natl. Acad. Sci. USA* **93**:443–446.
7. Casper, S. and Holt, C. 1996. Expression of the green fluorescent protein-encoding gene from a tobacco mosaic virus-based vector. *Gene* **173**:69–7
8. Epel, B., Padgett, H., Heinlein, M., and Beachy, R. 1996. Plant virus movement protein dynamics probed with a GFP-protein fusion. *Gene* **173**:75–79.
9. Wang, S. and Hazelrigg, T. 1994. Implications for bcd mRNA localization from spatial distribution of exu protein in *Drosophila* oogenesis. *Nature* **369**:400–403.
10. Amsterdam, A., Lin, S., Moss, L., and Hopkins, N. 1996. Requirements for green fluorescent protein detection in transgenic zebrafish embryos. *Gene* **173**:99–103.
11. Ludin, B., Doll, T., Meill, R., Kaech, S., and Matus, A. 1996. Application of novel vectors for GFP-tagging of proteins to study microtubule-associated proteins. *Gene* **173**:107–111.
12. DeGiorgi, F., Brini, M., Bastianutto, C., Marsault, R., Montero, M., Pizzo, P, et al. 1996. Targeting aequorin and green fluorescent protein to intracellular organelles. *Gene* **173**:113–117.
13. Cubitt, A., Heim, R., Adams, S., Boyd, A., Gross, L., and Tsien, R. 1995. Understanding, improving and using green fluorescent proteins. *TIBS* **20**:448–455.
14. Olsen, K., McIntosh, J., and Olmstead, J. 1995. Analysis of MAP4 function in living cells using green fluorescent protein (GFP) chimeras. *J. Cell Biol.* **130**:639–650.
15. Rizzuto, R., Brini, M., De Giorgi, F., Rossi, R., Heim, R., Tsien, R., et al. 1996. Double labeling of subcellular structures with organelle-targeted GFP mutants *in vivo*. *Curr. Biol.* **6**:183–188.
16. Kaether, C. and Gerdes, H. 1995. Visualization of protein transport along the secretory pathway using green fluorescent protein. *FEBS Lett.* **369**:267–271.
17. Marshall, J., Molloy, R., Moss, G., Howe, J., and Hughes, T. 1995. The jellyfish green fluorescent protein:a new tool for studying ion channel expression and function. *Neuron* **14**:211–215.
18. Mitra, R., Silva, C., and Youvan, D. 1996. Fluorescence resonance energy transfer between blue-emitting and red-shifted excitation derivatives of the green fluorescent protein. *Gene* **173**:13–7.
19. Kahana, J. and Silver, P. 1996. pp. 9.7.22–9.7-28. in *Current protocols in molecular biology*. Ausabel F., et al. (eds.). Green and Wiley, New York.
20. Cody, C.W., Prasher, D.C., Westler, W.M., Prendergast, F.G., and Ward, W.W. 1993. Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein. *Biochemistry.* **32**:1212–1218.
21. Heim, R., Prasher, D.C., and Tsien, R.Y. 1994. Wavelength mutations and post-translational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **91**:12501–12504.
22. Delagrave, S., Hawtin, R., Silva, C., Yang, M., and Youvan, D. 1995. Red-shifted excitation mutants of the green fluorescent protein. *Bio/Technology* **13**:151–154.
23. Lim, C., Kimata, K., Oka, M., Nomaguchi, K., and Kohno, K., 1995. Thermosensitivity of a green fluorescent protein utilized to reveal novel nuclear-like compartments. *J. Biochem. (Tokyo)* **118**:13–17.
24. Ward, W.W. and Bokman, S.H. 1982. Reversible denaturation of *Aequorea* green-fluorescent protein: physical separation and characterization of the renatured protein. *Biochemistry* **21**:4535–4540.
25. Ward, W., Prentice, H., Roth, A., Cody, C. and Reeves, S. 1982. Spectral perturbations of the *Aequoria* green fluorescent protein. *Photochem. Photobiol.* **35**:803–808.
26. Inouye, S. and Tsuji, F.I. 1994. Evidence for redox forms of the *Aequorea* green fluorescent protein. *FEBS Lett.* **351**:211–214.
27. Dopf, J. and Horiagan, T. 1996. Deletion mapping of the *Aequoria victoria* green fluorescent protein. *Gene* **173**:39–44.
28. Heim, R., Cubitt, A., and Tsien, R. 1995. Improved green fluorescence. *Nature* **373**:663–664.
29. Cormack, B., Valdivia, R., and Falkow, S. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**:33–38.
30. Ehrig, T., O'Kane, D., and Prendergast, F. 1995. Green-fluorescent protein mutants with altered fluorescence excitation spectra. *FEBS Lett.* **367**:163–166.
31. Crameri, A., Whitehorn, E., Tate, E., and Stemmer, W. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotech.* **14**:315–319.
32. Perozzo, M., Ward, K., Thompson, R., and Ward, W. 1988. X-ray diffraction and time-resolved fluorescence analyses of *Aequorea* green fluorescent protein crystals. *J. Biol. Chem.* **263**:7713–7716.
33. Merbs, S. and Nathans, J. 1992. Absorption spectra of the hybrid pigments responsible for anomalous color vision. *Science* **258**:464–466.
34. Rao, B., Kemple, M., and Prendergast, F. 1980. Proton nuclear magnetic resonance and fluorescence spectroscopic studies of segmental mobility in aequorin and a green fluorescent protein from *Aequorea forskalea*. *Biophys. J.* **32**:630–632.
35. Ormö, M., Cubitt, A., Kallio, K., Gross, L., Tsien, R., and Remington, S. 1996. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* In press.
36. Wright, H. 1991 Nonenzymatic deamidation of asparaginyl and glutaminyl residues in proteins. *Crit. Rev. Biochem. Mol. Biol.* **26**:1–52.
37. Chattoraj, M., King, B., Bublitz, G., and Boxer, S. 1996. Ultra-fast excited state dynamics in green fluorescent protein: Multiple states and proton transfer. *Proc. Natl. Acad. Sci. USA* **93**:8362–8367.
38. Otwinowski, Z. 1993. Data collection and processing, in *Proceedings of the CCP4 study weekend*. Warrington, England: Daresbury Laboratory.
39. Sheldrick, G., Dauter, Z., Wilson, K., Hope, H., and Sieker, L. 1993. The application of direct methods and Patterson interpretation to high-resolution native protein data. *Acta Cryst.* **D49**:18–23.
40. Terwilliger, T., Kim, S.-H., and Eisenberg, D. 1987. Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Cryst.* **A43**:1–5.
41. Yang, W., Hendrickson, W., Crouch, R., and Satow, Y. 1990. Structure of ribonuclease H phased at 2 A by MAD analysis of the seleniomethionyl protein. *Science* **249**:1398–1405.
42. Collaborative Computational Project, N. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Cryst.* **D50**:760–763.
43. Jones, T., Zou, J., Cowan, S., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **47**:110–119.
44. Brunger, A. 1992. *X-PLOR Version 3.1: A system for X-ray crystallography and NMR*. Yale University Press, New Haven, CT.
45. Carson, M. 1987. Ribbon models of macromolecules. *J. Mol. Graphics* **5**:103–106.
46. Sayle, R. and Milner-White, E. 1995. RasMol: Biomolecular graphics for all. *TIBS* **20**:374–375.