PROTEINS: Structure, Function, and Genetics 41:429–437 (2000)

Crystal Structure and Refolding Properties of the Mutant F99S/M153T/V163A of the Green Fluorescent Protein

Roberto Battistutta,[1*] Alessandro Negro,[2] and Giuseppe Zanotti[1]
[1]*Dipartimento di Chimica Organica and Centro Studi sui Biopolimeri, Università di Padova, Padova, Italy*
[2]*Dipartimento di Chimica Biologica, Università di Padova, Padova, Italy*

ABSTRACT **The mutant F99S/M153T/V163A of the Green Fluorescent Protein (c3-GFP) has spectral characteristics similar to the wild-type GFP, but it is 42-fold more fluorescent *in vivo*. Here, we report the crystal structure and the refolding properties of c3-GFP and compare them with those of the less fluorescent wt-GFP and S65T mutant. The topology and the overall structure of c3-GFP is similar to the wild-type GFP. The three mutated residues, Ser99, Thr153, and Ala163, lie on the surface of the protein in three different β-strands. The side chains of Ser99 and Thr153 are exposed to the solvent, whereas that of Ala163 points toward the interior of the protein. No significant deviation from the structure of the wild-type molecule is found around these positions, and there is not clear evidence of any distortion in the position of the chromophore or of the surrounding residues induced by the mutated amino acids. In vitro refolding experiments on urea-denatured c3-GFP reveal a renaturation behavior similar to that of the S65T molecule, with kinetic constants of the same order of magnitude. We conclude that the higher fluorescence activity of c3-GFP can be attributed neither to particular structural features nor to a faster folding process, as previously proposed. Proteins 2000;41:429–437.**

Key words: GFP; hydrogen bond pattern; x-ray crystallography; folding; renaturation

INTRODUCTION

The Green Fluorescent Protein (GFP) from the jellyfish *Aequorea victoria* is a single domain protein of 238 amino acids,[1] that converts blue radiation into green light (for a recent review, see Tsien, 1998[2]). In vivo, a primary chemiluminescent photoprotein, aequorin, upon binding of Ca^{2+} ions undergoes an intramolecular reaction with formation of a Blue Fluorescent Protein in a singlet excited state (BFP*); the energy transfer process from BFP* to GFP is then followed by the emission of green light, probably for better penetration in the ocean.

GFP is characterized by the presence of a chromophore consisting of a 4-(4-hydroxyphenyl) methylideneimidazol-5-one group.[3] The chromophore originates from a posttranslational autocatalytic cyclization of the tripeptide segment -Ser65-Tyr66-Gly67-, with successive dehydrogenation of the α-β bond of the tyrosine in the presence of molecular oxygen.[4–6]

In the mature protein, the chromophore exhibits two characteristic absorption bands, the major one at 395 nm and the minor one at 475 nm,[7,8] corresponding to the protonated and anionic state of the hydroxyl group of the chromophoric phenol, respectively.[9–11] Upon excitation at 395 or 475 nm, the chromophore emits green light with an emission maximum around 508 nm and a quantum yield of approximately 0.8. Fluorescence is lost upon protein denaturation[12,13] and in the isolated chromophore; synthetic model compounds containing the chromophore show fluorescence activity only at a temperature as low as 77°K,[14] suggesting that a restriction in the chromophore movement is necessary for the fluorescence activity.[15,16]

Due to its intrinsic fluorescence properties, its ability to be efficiently fused to many other proteins and expressed in a variety of organisms, both prokaryotes and eukaryotes, GFP is widely used as fluorescent probe in several biotechnological applications.

In the last years, the structure of wild-type GFP and of several mutants have been solved.[11,17–21] The protein adopts a so called "β-can" structure, in which 11 β-strands form an almost perfect cylinder, with a diameter of approximately 30 Å and a length of approximately 40 Å; the β-barrel is capped at the extremities by short helices and loops. The fluorophore, that belongs to a distorted central helix that spans the inner portion of the cylinder, is completely buried in the interior of the protein and isolated from the bulk solvent. Interestingly, several polar residues and water molecules are located around the chromophore, forming an extended hydrogen-bond network. The fluorescence mechanism involves a fast proton transfer in the excited state and a slow conformational rearrangement attributed to Thr203, whose side-chain can exist in two different conformations.

To improve or modify the spectral characteristics of the Green Fluorescent Protein, several mutants have been

Abbreviations: GFP, Green Fluorescent Protein; wt-GFP, wild-type Green Fluorescent Protein; c3-GFP, cycle 3 GFP, i.e. mutant F99S/M153T/V163A of the Green Fluorescent Protein; PEG, poly(ethylene glycol); RMS, root-mean-square deviation; HEPES, N-[2-hydroxyethyl]piperazine-N′-[2-ethanesulfonic acid]; DTT, dithiothreitol; NCS: noncrystallographic symmetry.

Grant sponsor: Ministero della Ricerca Scietifica (MURST); Grant sponsor: Consiglio Nazionale delle Ricerche (CNR), Rome, Italy.

*Correspondence to: Roberto Battistutta, Dipartimento di Chimica Organica, via Marzolo 1, 35131 Padova, Italy.
E-mail: roberto@chor.unipd.it

Received 3 February 2000; Accepted 18 July 2000

produced and studied, mainly with variations in the chromophore itself or in the residues located in its immediate proximity.[8,22] The mutant F99S/M153T/V163A (c3-GFP), obtained by DNA shuffling techniques, was found to be around 42-fold more fluorescent *in vivo* than the wt-GFP, with almost identical extinction coefficients, quantum yield and peak wavelengths of the excitation and emission spectra.[2,23] Unlike other mutants, in c3-GFP the substitutions are not directly in contact with the chromophore and it is unclear how they can cause such increase in fluorescence. We have purified and crystallized the c3-GFP mutant, and here we present and discuss the crystal structure. The renaturation behavior of this protein was also investigated in vitro and compared with that of the lower fluorescent mutant, S65T.

MATERIALS AND METHODS

Protein Expression and Crystallization

His-tagged c3-GFP was expressed and purified according to the previously described procedure.[24] During storage, 13 NH_2-terminal residues, including the hexahistidine sequence, were cleaved off by proteases. After a gel permeation purification step (on HiLoad Sephadex 75 column, Pharmacia-Biothec), reverse phase chromatography and N-terminal sequence analysis revealed the presence of homogeneous c3-GFP, with the NH_2-terminal sequence starting with the alanine in position 2.

Crystals of c3-GFP were obtained in 4–5 days at 20°C in hanging drops made up mixing 2 μl of protein at a concentration of 23 mg/ml in 10 mM HEPES pH 7.4 and 2 μl of precipitant solution, composed by 22% PEG 4000, 50 mM HEPES pH 8.5, 50 mM $MgCl_2$, and 10 mM 2-mercaptoethanol. Drops were equilibrated against 0.75 ml of the same precipitant solution. The presence of a thin layer of 50% paraffin in mineral oil allowed the growth of crystals of better quality.[25] Flat crystals had approximate dimensions of $0.05 \times 0.3 \times 0.7$ mm^3.

Spectroscopic and Refolding Analysis

Absorbance spectra were determined on a Perkin-Elmer Lambda 16 Spectrophotometer for GFP in solution and on a Zeiss MPM 800 microspectrophotometer for GFP crystals. All fluorescence spectra and refolding kinetics were carried out with a Perkin-Elmer LS50B Luminescence Spectrometer. Refolding studies were performed according to the procedure described for S65T-GFP.[6] Briefly, denatured GFP was obtained by heating samples for 5 minutes at 95°C in 8 M urea and 100 mM DTT (no more fluorescence was detectable). In refolding experiments, denatured GFP was diluted 100-fold in a renaturation buffer composed of 50 mM Tris-HCl, pH 7.5, 35 mM KCl and 2 mM $MgCl_2$, in presence of 100 mM DTT. The refolding kinetic was followed by monitoring the fluorescence signal at 512 nm, with an excitation wavelength of 488 nm. Kinetic constants were derived with MicroMath Scientist software (MicroMath Inc.) by fitting experimental data to differential equations derived both for parallel and for sequential kinetic refolding models, with k_1 and k_2 first-order rate constants for the initial fast phase and the second slow phase, respectively.[6]

TABLE I. Statistics on Data Collection and Processing[†]

Number of crystals	2
Total reflections	165,357
Unique reflections	36,746 (5,445)
High resolution limit (Å)	2.37
Completeness to 2.37 Å (%)	84.0 (45.0)
Completeness to 2.64 Å (%)	99.3
R_{merge}[a]	0.062 (0.146)
$\langle I/\sigma(I)\rangle$	6.9 (4.2)

[†]Numbers in parentheses refer to the highest resolution shell.
[a]$R_{merge} = \Sigma|I_{hkl} - \langle I\rangle|/\Sigma\langle I\rangle$ where $\langle I\rangle$ = average of individual measurements of I_{hkl}.

Diffraction Data Collection and Structure Refinement

Because crystals diffracted only at approximately 7 Å under a rotating anode source, diffraction data from two different crystals were measured at the ELECTRA synchrotron in Trieste (Italy), with an imaging plate detector system (MAR Research). At 100°K the maximum resolution limit was 2.37 Å. A cryoprotectant solution of 26% (v/v) glycerol in the precipitant solution was used to preserve the crystal integrity. Data were processed with the MOSFLM software,[26] and statistics on data collection and processing are reported in Table I. Crystals of c3-GFP belong to the orthorhombic space group $P2_12_12_1$, with unit cell parameters **a** = 83.48, **b** = 85.37, and **c** = 140.06 Å. Sixteen molecules are present in the unit cell, corresponding to a V_M of 2.3 Å^3/Da and a water content of 50%.

The crystal structure was solved with the molecular replacement method, by using the software AMoRe[27] from the CCP4 suite.[28] The coordinates of the wt-GFP dimer[18] were used as template. After application of rotation and translation functions, two dimers were placed in the asymmetric unit; a rigid body refinement gave a crystallographic R factor of 49.5. The initial model was refined alternating automatic minimization protocols performed with the X-PLOR package[29,30] with visual inspection of the electron density map and manual adjustment by using the program O.[31] Restraints on noncrystallographic symmetries were imposed to main-chain and side-chain atoms, with different weights. Initially, the four molecules in the asymmetric unit (named A, B, C, and D) were considered equal, but during the refinement, it became clear that one dimer (AB) was slightly different from the other (CD). Consequently, in the last refinement cycles, the noncrystallographic symmetry imposed was A = B and C = D. In all cases, zones with poor density, usually in exposed loops, the chromophore and its surroundings were excluded from noncrystallographic symmetries. During refinement, water molecules were added to the model, both automatically and manually, and those with B factors higher then 60 were manually excluded. In the last few refinement cycles, very weak reflections with a ratio F_{calc}/F_{obs} greater than 3.5 (3,662 on a total of 36,746) were excluded from the calculations. As far as we can judge from their large F_{calc}/F_{obs} ratio, these weak reflections were clearly underestimated, and consequently they had to be considered

TABLE II. Statistics on Refinement and Final Model

Unique reflections used (high resolution limit (Å)	33,030 2.37)
Protein atoms	7184
Water molecules	286
Noncrystallographic symmetry weights (kcal/mol · Å^2)[a]	
On main chain	100
On side chains	50
Crystallographic R (%)	20.4
R_{free}[b] (%)	28.0
Root mean square deviation for	
Bond length (Å)	0.015
Bond angle (deg)	3.3

[a]X-PLOR force constant for positional restraints.[29,30]
[b]10% of the total unique reflections were used for R_{free} cross-validation.

TABLE III. Root Mean Square Values (Å) Comparing c3-GFP and Monomeric[17] and Dimeric[18] wt-GFP[†]

Parameter	B	C	D	1GFL A	1GFL B	1EMA	(CD)	1GFL (AB)
A	0.24	0.28	0.31	0.49	0.47	0.55		
B		0.26	0.27					
C			0.17					
D				0.47	0.46	0.55		
(AB)							0.59	0.86

[†]Capital letters A, B, C, and D indicate the single-chains present in the asymmetric unit. (AB) and (CD) are the two dimers of c3-GFP in the asymmetric unit. GFP, Green Fluorescent Protein; wt, wild-type.

affected by systematic measurements errors, possibly owning to the low quality of the crystals. The exclusion of these reflections improved both the R and R_{free} values of approximately 3%. Although the above-mentioned data cut-off decreased the number of experimental data, the final electron density and model have a satisfactory quality, also because there are four molecules in the asymmetric unit and NCS restrains could be applied. The Ramachandran plot analysis of the final model shown only one residue in the generously allowed regions and none in the disallowed regions; 87.1% of the residues were found in the most favored and 12.8% in the additionally allowed regions.

Based on the electron density, residues from position 4 to 230 of the amino acid sequence could be located in the final model, that includes also 286 water molecules. Refinement protocol and final model statistics are shown in Table II.

RESULTS

Crystal Structure

On the whole, the single molecule of c3-GFP closely resembles the wild-type molecule: 11 β-strands form the classic "β-can" structure with the chromophore belonging to the central distorted helix well-buried inside. In all four chains, the chromophores are planar. Although crystallization conditions are similar to that used for the S65T mutant,[17] which crystallized as monomer, in our case, there are two dimers in the asymmetric unit.

RMS values (Table III) reveal that, in our model, single chains are more similar if in the same dimer. As mentioned in the experimental session, noncrystallographic symmetries relating A with B and C with D were applied only at the final stages of refinement, when the two dimers in the asymmetric units came out to be apparently different. Nevertheless, one must be aware that the final RMS values between the four monomers reported in Table III are biased by the use of that type of NCS restrains. The main differences are located in flexible exposed loops and in the region of crystallographic contacts between dimers; with these regions excluded, RMS values for chains A/B and C/D decrease below 0.1 Å. The comparison of a single chain with that of the wild-type molecule gives an RMS deviation for equivalent α carbons of 0.46–0.49 Å (Table III), depending on the single chains considered, whereas for the dimers, the RMS increases to 0.86 Å; the superimposition of the two dimers (AB) and (CD) of our model gave an RMS of 0.59 Å.

In the asymmetric unit, the four molecules are arranged in such a way that the axes of the cylinders of chains A and B are roughly perpendicular to that of chains D and C, respectively (A on D = 93.7 degree rotation and B on C = 93.9 degree rotation), giving rise to a pseudo-tetragonal packing (Fig. 1).

The single chains in a dimer are related by a twofold axis (A on B = 179.9 degree rotation, C on D = 179.3 degree rotation) in a way similar to that of the wt-GFP dimeric structure (Fig. 2). Also the portions of the molecules in contact within each dimer are essentially the same as in the wild-type molecule, with extended interactions encompassing the hydrophobic residues Ala206, Leu221, and Phe223, and many hydrophilic residues, responsible for the presence of as well as 17 intermolecular hydrogen bonds. The solvent excluded interfaces between two chains of the same dimer, i.e., the accessible surface area buried on dimer formation (calculated with a probe radius of 1.4 Å) is around 910 Å^2.

The weak interactions between the two dimers of the c3-GFP are probably due to solely crystallographic packing reasons and do not reflect a similar aggregation state in solution. The contacts between the two dimers involve essentially residues of chains B and D; while the solvent excluded interfaces between the two dimers is quite large (around 650 Å^2 by using a probe of radius 1.4 Å), the direct interactions are limited. There are only three hydrogen bonds, between Glu B17 and Tyr D182, Asp B117 and Lys D158, and Asp B117 and Glu D184, and one electrostatic interaction between the positively charged Arg B215 and the π electron density of the ring of Tyr D151. Most of the space between the two dimers is indeed filled with water molecules.

In the final model, the average B factor for the backbone of the (AB) dimer is 40 Å^2, significantly higher than that of the (CD) dimer, 23 Å^2, suggesting an higher mobility for the former. Thermal B factors of the chromophore and surrounding residues are slightly lower than the average

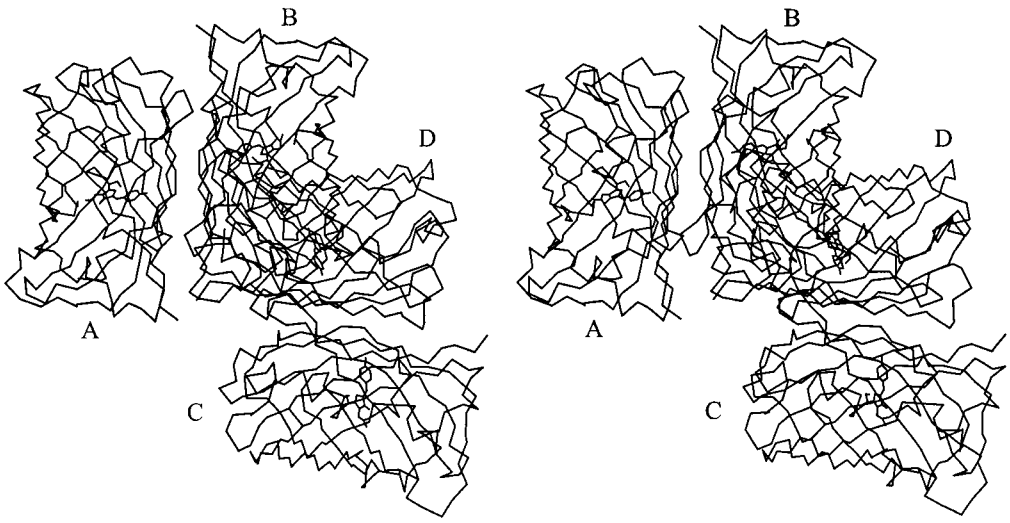

Fig. 1. Stereo drawing of the Cα trace of the four molecules of Green Fluorescent Protein (c3-GFP) in the asymmetric unit. The four chain are named A, B, C, and D and form two dimers, i.e., (AB) and (CD).

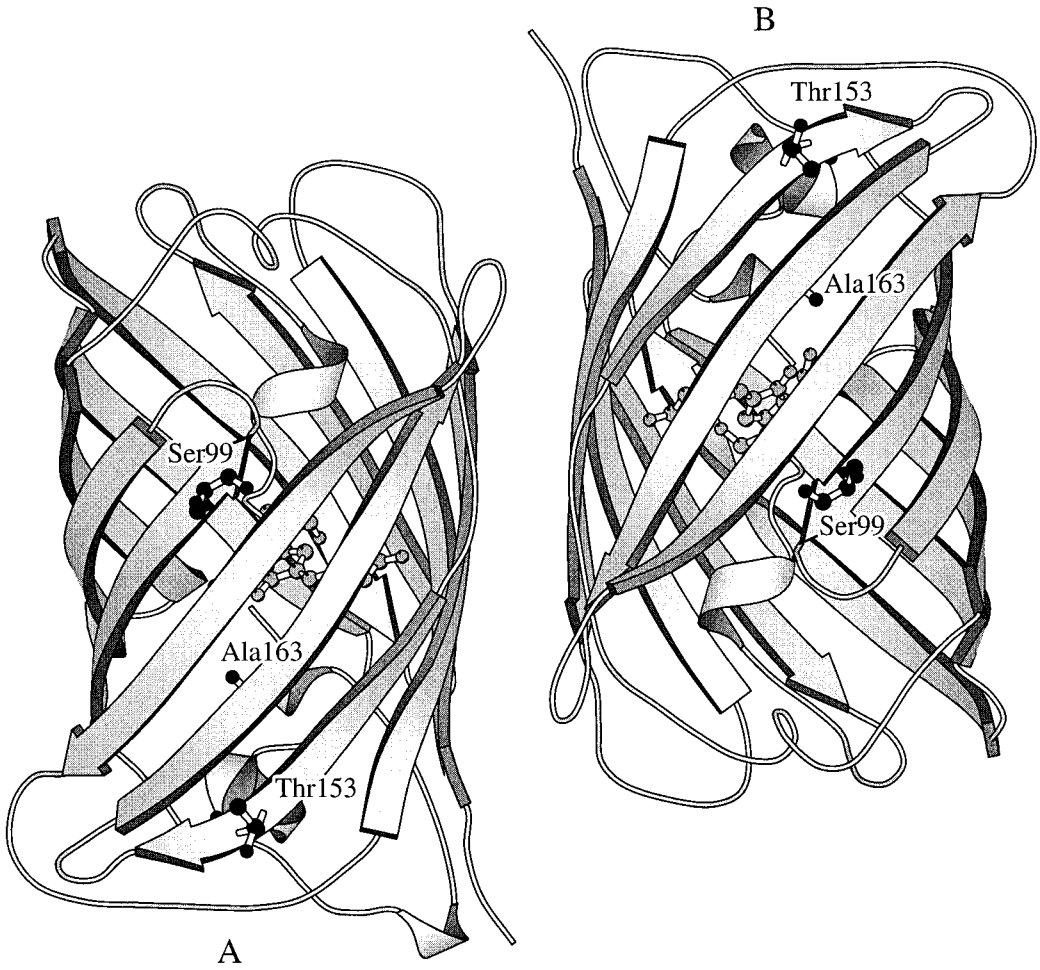

Fig. 2. Schematic drawing of the Green Fluorescent Protein (c3-GFP) dimer, viewed along the twofold axis. The chromophore (in gray) and the three mutated residues (in black) are represented as ball-and-stick.

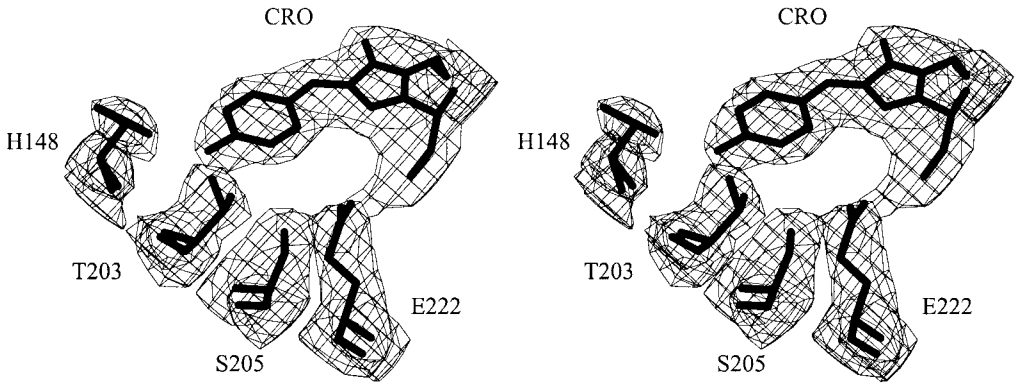

Fig. 3. Stereo view of the 2Fo-Fc electron density map contoured at 1 σ around the chromophore (CRO) and the main interacting residues His148 (H148), Thr203 (T203), Ser205 (S205), and Glu222 (E222). Water molecules around the chromophore are not shown.

values, but they do not differ significantly from that of other residues in the interior of the protein.

The three residues mutated from the wild-type molecule, i.e., Ser99, Thr153, and Ala163, lie on the protein surface, parallel to the cylinder axis, in three different β-strands (Fig. 2). The side chains of Ser99 and Thr153 are exposed to the solvent, whereas that of Ala163 points toward the interior of the protein, and interacts with Gln183, Tyr151, Lys162, and Asn164. None of the mutated residues belongs to the contact surfaces between two chains of the same dimer or between the two dimers or are involved in crystallographic contacts. No significant deviations from the structure of the wt molecule is found around these positions, and there is no clear evidence of any distortion of the position of the chromophore or of the surrounding residues induced by mutated amino acids.

Since the beginning of the refinement, it was evident that near the chromophores of chains A and B the electron density differed from that in the proximity of chromophores in chains C and D. An example of the quality of the final electron density map is illustrated in Figure 3, showing the chromophore and some of the surrounding residues. Four water molecules were positioned in the electron density for chains C and D and only two for chain A and one for chain B. Two hydrogen bonded water molecules (W1 and W2) are placed near the imidazolone ring of chain A, making hydrogen bonds with the N2 nitrogen of the chromophore, the OH of Ser65, the carboxylate of Glu222 and the amidic nitrogen of Val68 (Figs. 4a and 5a). The hydrogen bond network includes also the side chain of Ser205 that interacts with the carboxylate of Glu222; the OH of the chromophore is connected to Thr203 and His148 by means of two hydrogen bonds. A similar arrangement is found in chain B, except for the lack of a water molecule corresponding to W1.

A different situation is found in chains C and D; here, four solvent molecules surround the chromophore (Figs. 4b and 5b). Now a water molecule (W4) makes an hydrogen bond with the hydroxyl group of Tyr66 and, through Ser205 and Glu222, connects this OH with the Ser65 side chain. The side-chain of Thr203 is rotated in such a way to

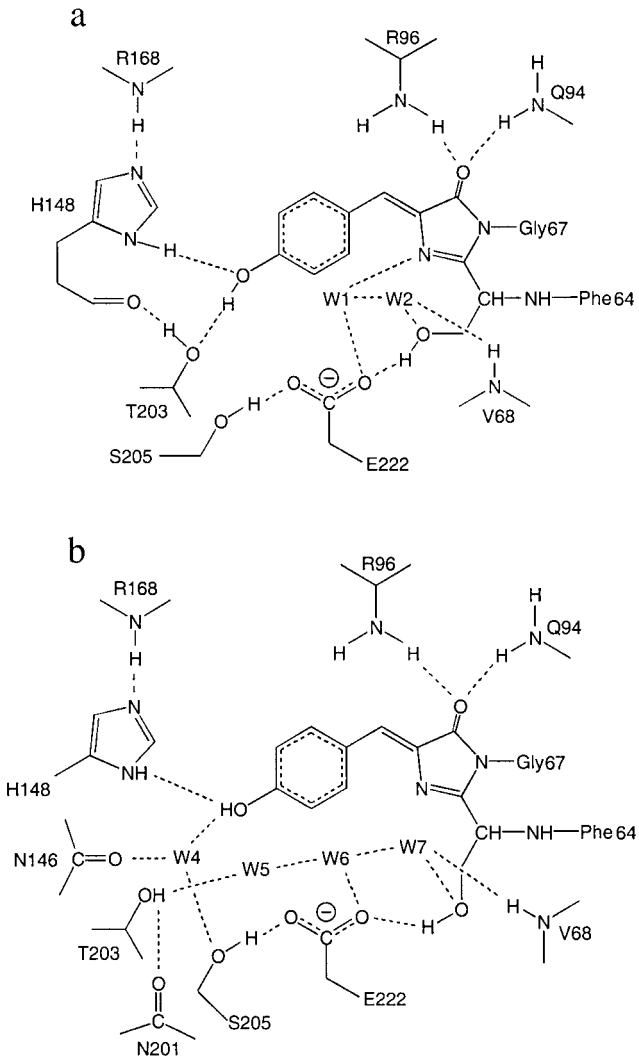

Fig. 4. Schematic diagram showing all hydrogen bonds (dashed line) involving the chromophore in dimer (AB) (**a**) and dimer (CD) (**b**). W1 e W2 are the two water molecules present in chain A, whereas W4, W5, W6, and W7 are the four water molecules present in chain C. Amino acids are indicated in single letter code.

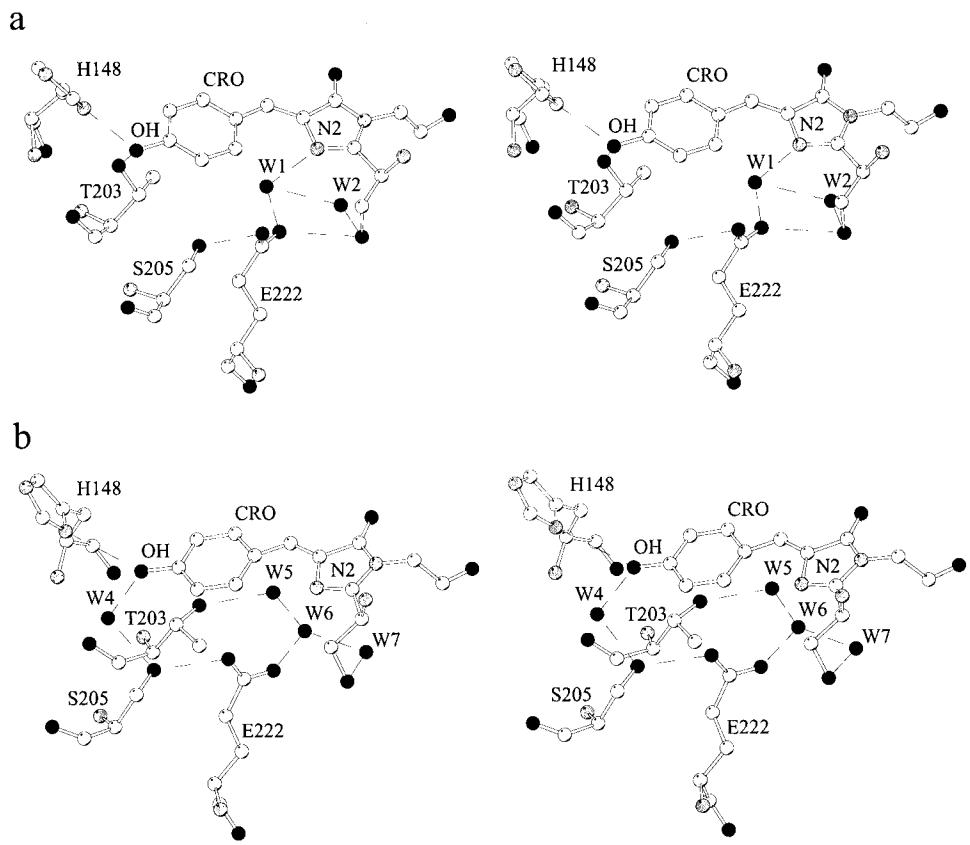

Fig. 5. Stereo view of the main interactions (dashed line) of chromophore CRO with His148 (H148), Thr203 (T203), Ser205 (S205), Glu222 (E222), and the water molecules W1 and W2 in chain A (**a**), and W4, W5, W6, and W7 in chain C (**b**). Note the different orientation of the side chain of Thr203 in the two dimers. Carbon atoms are represented in white, nitrogen in gray, and oxygen in black.

interact with water molecule W5 and not with the chromophoric OH as in chains A and B; in this arrangement, it is further stabilized by a second hydrogen bond with the carboxyl group of Asn201. As confirmed by accessible surface area calculations, the chromophores and the surrounding water molecules in all four chains are entirely buried inside and isolated from the bulk solvent.

Absorbance and Fluorescence Spectroscopy

To verify whether the spectral characteristics of c3-GFP change upon crystallization, we measured absorbance and fluorescence spectra of the protein, both in solution and in the crystal. As illustrated by Figure 6, the protein in the crystal displays essentially the same spectral properties of the protein in solution, indicating that the chromophore environment does not differ substantially in the two states. The high ratio between the absorption band around 395 nm, corresponding to the protonated form, and that at longer wavelength, corresponding to the deprotonated form of the chromophore, indicates the large predominance of the former both in the crystal and in solution.

Renaturation Properties

To assess whether the higher fluorescence of the c3-GFP mutant can be attributed to a faster kinetic of folding,[32] we performed renaturation kinetic experiments on denatured GFP according to the procedure used for the less fluorescent S65T-GFP mutant.[6] For the complete denaturation of c3-GFP, aliquots of the protein in 8 M urea and 100 mM

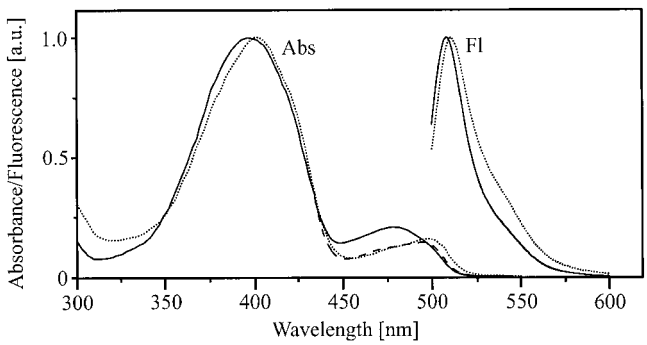

Fig. 6. Normalized absorbance (Abs) and fluorescence (Fl) spectra of c3-GFP in diluted solution (———) and in the crystal (· · · · · ·). The small shift in the fluorescence emission maximum is similar to that reported for wild-type GFP hexagonal crystals.[38] In solution, the ratio between the peaks at 395 and 480 nm increases with protein concentration[2] and, for comparison, a portion of the absorbance spectrum of concentrated c3-GFP in the crystallization solution is also reported (--------).

DTT were heated at 95°C for 5 minutes; this procedure results in a complete loss of the fluorescent activity (verified by fluorescence measurements), because of the exposure of the chromophore to the bulk solution due to protein unfolding. The renaturation kinetic was followed by monitoring the recovery of the fluorescence signal at 512 nm upon 100-fold dilution of denatured GFP in a renaturation buffer (50 mM Tris-HCl, pH 7.5, 35 mM KCl, and 2 mM $MgCl_2$), in presence of 100 mM DTT. The

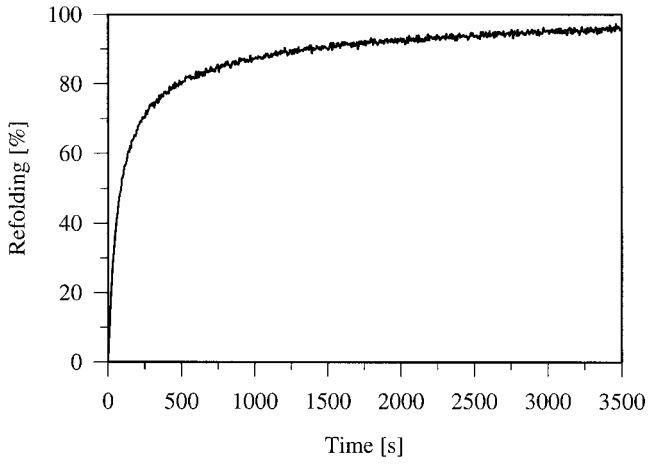

Fig. 7. Renaturation time-course of Green Fluorescent Protein (c3-GFP) followed by monitoring the recovery of the fluorescence emission at 512 nm. The percentage of refolding was calculated on the basis of the final constant amount of fluorescence, corresponding to 100% of refolding.

addition of a reducing agent as DTT is necessary to prevent nonproductive refolding due to the incorrect formation of disulphide bonds. The time-course (a typical renaturation curve is reported in Fig. 7) and the efficiency (around 60%) of renaturation of c3-GFP are similar to that of S65T mutant. The renaturation kinetic exhibits two distinct first-order kinetic phases, probably due to a heterogeneity of the starting urea-denatured material in the backbone conformation of prolines.[6] Experimental data were fitted to differential equations derived both for parallel and for sequential kinetic refolding models, with k_1 and k_2 as first-order rate constants for the initial fast phase and the second slow phase, respectively. In both models, two initial denatured states, U1 and U2, are postulated; in the parallel scheme, U1 and U2 refolds independently to the final native state N, with rate constant k_1 and k_2, respectively; in the sequential scheme, U2 converts slowly to U1 with rate constant k_2, whereas U1 undergoes a fast refolding (k_1) to the native state. The renaturation time-course could be fitted equally well with both models, giving essentially the same values for k_1 and k_2, hindering the possibility to discriminate between them. Three independent experiments gave the following averaged results: $k_1 = 1.42 \times 10^{-2}\ s^{-1}$ and $k_2 = 1.24 \times 10^{-3}\ s^{-1}$, slightly lower but in the same order of magnitude of those found for the S65T mutant ($k_1 = 2.45 \times 10^{-2}\ s^{-1}$ and $k_2 = 2.44 \times 10^{-3}\ s^{-1}$).

DISCUSSION

Although the topology and the overall structure of c3-GFP is similar to the other Green Fluorescent Proteins, nevertheless, it presents some specific and peculiar characteristics. The two dimers of c3-GFP mutant in the asymmetric unit are not equivalent. Considering the 14 structures of the Green Fluorescent Protein and its mutants available in the Protein Data Bank, only in one case (mutant F64L, I167T, and K238N, PDB code 1EMC) the asymmetric unit is composed of four molecules, but the two dimers are arranged with the axes of the cylinders parallel to each other, in a different way from c3-GFP. As well as in c3-GFP, in mutant F64L, I167T, and K238N, the number and location of the water molecules around the chromophore in each monomer are different. In all four chains of this mutant the hydroxyl group of Thr203 is turned away from the OH of the chromophore as in chains C and D of c3-GFP.

In chains A and B of c3-GFP, the conformation of the Thr203 side chain allows an hydrogen bond between the OH of this residue and that of the chromophore in the absence of any water molecule. A similar arrangement of Thr203 is found in other four mutants, i.e., S65T (PDB code 1EMA), F64L (PDB code 1EMM), F64L and K238N (PDB code 1EML), and F64L, S65C, I167T, and K238N (PDB code 1EMK), and in two Blue Fluorescent variants (PDB codes 1BFP and 2EMN) where the Tyr66 is substituted by a His, but it is always associated with the presence of a water molecule close to the phenolic oxygen. Taking into account also the different arrangement of the chromophore environment due to the substitution of Thr203 with a tyrosine in the Yellow Fluorescent variant recently published,[21] it comes out that the quite hydrophilic environment of the chromophore can adjust itself to accommodate different numbers of water molecules with minor perturbations of the overall protein structure.

Other significant differences around the chromophore are the hydrogen-bond network involving His148, Ser205, Glu222, and the side chain of Ser65, and the orientation of the side chain of Thr203 (toward or away from the OH of the chromophore). In other GFPs,[11,19,33] the phenol group of the chromophore in the deprotonated form makes a direct hydrogen-bond with the side chain of Thr203, whereas, when the phenol group is in the neutral form, the Thr203 side chain is turned away. The latter arrangement is very similar to that found in our dimer CD that, therefore, can be assumed to have a protonated phenolic oxygen. In the other dimer AB, the hydroxyl group of Thr203 makes two hydrogen bonds, one with the carbonyl oxygen of His148 (distance: 2.73 Å) and the other with the phenol of the chromophore (distance: 2.72 Å); because, in the former case, the Thr203 hydroxyl group must act as a hydrogen donor, in the latter, it must be an acceptor, and hence, the phenolic hydroxyl of the chromophore must be protonated, at least in chain A. This unique arrangement of Thr203 side chain in the presence of a protonated chromophore is feasible and, possibly, necessary because of the absence of the water molecule hydrogen-bound to the phenolic hydroxyl. The situation is less clear in chain B, where the distance between the carbonyl of His148 and the Thr203 hydroxyl is longer, 4.4 Å. In dimer (AB), the Nδ1 nitrogen of His148 is only 3.2 Å apart from the carbonyl of residue 146, whereas in the dimer (CD) the relative distance is around 3.8 Å, probably to accommodate the water molecule that bridges the carbonyl and the hydroxyl of the chromophore in the (CD) dimer. The absorption spectra of c3-GFP demonstrate that the neutral form is largely predominant both in solution and in the

crystal, confirming the crystal data that suggest that even with the Thr203 side chain hydrogen-bound to the phenolic hydroxyl group, as in our case, the chromophore is protonated.

From crystallographic studies at acidic and basic pH on S65T-GFP, it has been suggested that the titratable group is the phenolic hydroxyl, whereas the heterocyclic ring nitrogen is always deprotonated.[33] On the other hand, based on quantum chemical calculations aimed to clarify the energies involved in the excitation and emission transitions, it has been proposed that the chromophore can exist in a positive or a zwitterionic state, with the heterocyclic nitrogen always protonated.[34,35] In this respect, in all four chains of our mutant the nitrogen of the chromophore is not far from a water molecule, their distances ranging from 3.1 to 3.6 Å. This water molecule is hydrogen bonded to the carboxylate of Glu222, directly or through another water molecule; such arrangement could be a possible route for an easy exchange of a proton between the nitrogen and the carboxylate in c3-GFP.

Renaturation data presented here are in contrast with the hypothesis that the higher fluorescence of c3-GFP is due to a faster rate of folding; in fact, the rate-constants of renaturation of c3-GFP are of the same order of magnitude of the less fluorescent mutant S65T. The mechanism of folding can be of some importance, though. Possibly due to the decrease of the steric hindrance of the side chains of mutated residues (a serine, a threonine, and an alanine instead of a phenylalanine, a methionine, and a valine, respectively), the pathway of folding of our mutant can be different and more productive than that of the wt molecule, e.g., because of more soluble intermediate species. Indeed, although wt-GFP is mainly found inactive in inclusion bodies,[8,36] the majority of c3-GFP molecules are found in solution, giving rise to a larger amount of fluorescent protein.[23] Furthermore, temperature dependence studies of chromophore formation of several GFPs demonstrated that our mutant folds better at high temperature (37°C) that wt and S65T mutant.[37] All together, these results seem to indicate that c3-GFP can undergo a more productive folding process, and that the increase of *in vivo* fluorescence can be ascribed to the consequent increase in the concentration of the mature fluorescent protein inside the cell.

Accession Numbers

The coordinates for c3-GFP have been deposited in the Brookhaven Protein Data Bank with accession code 1b9c.

ACKNOWLEDGMENTS

Dr. Silvia Recchia contributed to this work during her degree thesis. We thank prof. R. Berni (University of Parma) for the spectroscopic measurements on crystals and the CNR staff at ELETTRA, Trieste, Italy, for the assistance during the measurements at the X-ray Diffraction Beamline supported by CNR and Sincrotrone Trieste.

REFERENCES

1. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the *Aequorea victoria* green-fluorescent protein. Gene 1992;111:229–233.
2. Tsien RY. The green fluorescent protein. Annu Rev Biochem 1998;67:509–544.
3. Cody CW, Prasher DC, Westler WM, Prendergast FG, Ward WW. Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein. Biochemistry 1993;32:1212–1218.
4. Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY. Understanding, improving and using green fluorescent proteins. Trends Biochem Sci 1995;20:448–455.
5. Nishiuchi Y, Inui T, Nishio H, Bodi J, Kimura T, Tsuji FI, Sakakibara S. Chemical synthesis of the precursor molecule of the *Aequorea* green fluorescent protein, subsequent folding, and development of fluorescence. Proc Natl Acad Sci USA 1998;95:13549–13554.
6. Reid BG, Flynn GC. Chromophore formation in green fluorescent protein. Biochemistry 1997;36:6786–6791.
7. Morise H, Shimomura O, Johnson FH, Winant J. Intermolecular energy transfer in the bioluminescent system of *Aequorea*. Biochemistry 1974;13:2656–2662.
8. Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc Natl Acad Sci USA 1994;91:12501–12504.
9. Ward WW, Cody CW, Hart RC, Cormier MJ. Spectrophotometric identity of the energy transfer chromophores in *Renilla* and *Aequorea* green fluorescent proteins. Photochem Photobiol 1980;31:611–615.
10. Chattoraj M, King BA, Bublitz GU, Boxer SG. Ultra fast excited state dynamics in green fluorescent protein: multiple states and proton transfer. Proc Natl Acad Sci USA 1996;93:8362–8367.
11. Brejc K, Sixma TK, Kitts PA, Kain SR, Tsien RY, Ormo M, Remington SJ. Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. Proc Natl Acad Sci USA 1997;94:2306–2311.
12. Bokman SH, Ward WW. Renaturation of *Aequorea* green fluorescent protein. Biochem Biophys Res Commun 1981;101:1372–1380.
13. Ward WW, Bokman SH. Reversible denaturation of *Aequorea* green-fluorescent protein: physical separation and characterization of the renatured protein. Biochemistry 1982;21:4535–4540.
14. Niwa H, Inouye S, Hirano T, Matsuno T, Kojima S, Kubota M, Ohashi M, Tsuji FI. Chemical nature of the light emitter of the *Aequorea* green fluorescent protein. Proc Natl Acad Sci USA 1996;93:13617–13622.
15. Kummer AD, Kompa C, Lossau H, Pollinger-Dammer F, Michel-Beyerle ME, Silva CM, Bylina EJ, Coleman WJ, Yang MM, Youvan DC. Dramatic reduction in fluorescence quantum yield in mutants of green fluorescent protein due to fast internal conversion. Chem Phys 1998;237:183–193.
16. Voityuk AA, MichelBeyerle ME, Rosch N. Structure and rotation barriers for ground and excited states of the isolated chromophore of the green fluorescent protein. Chem Phys Lett 1998;296:269–276.
17. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. Science 1996;273:1392–1395.
18. Yang F, Moss LG, Phillips GN. The molecular structure of green fluorescent protein. Nat Biotechnol 1996;14:1246–1251.
19. Palm GJ, Zdanov A, Gaitanaris GA, Stauber R, Pavlakis GN, Wlodawer A. The structural basis for spectral variations in green fluorescent protein. Nat Struct Biol 1997;4:361–365.
20. Wachter RM, King BA, Heim R, Kallio K, Tsien RY, Boxer SG, Remington SJ. Crystal structure and photodynamic behavior of the blue emission variant Y66H/Y145F of green fluorescent protein. Biochemistry 1997;36:9759–9765.
21. Wachter RM, Elsliger MA, Kallio K, Hanson GT, Remington SJ. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. Structure 1998;6:1267–1277.
22. Heim R, Tsien RY. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. Curr Biol 1996;6:178–182.
23. Crameri A, Whitehorn EA, Tate E, Stemmer WPC. Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat Biotechnol 1996;14:315–319.
24. Negro A, Grassato L, Polverino De Laureto P, Skaper SD. Genetic construction, properties and application of a green fluorescent protein-tagged ciliary neurotrophic factor. Protein Eng 1997;10:1077–1083.

25. Chayen NE. A novel technique to control the rate of vapour diffusion, giving larger protein crystals. J Appl Crystallogr 1997; 30:198–202.
26. Leslie AGW. Molecular data processing. In: Moras D, Podjarny AD, Thierry JP, editors. Crystallographic computing V. Oxford, UK: Oxford University Press; 1991. p 50–61.
27. Navaza J. AMoRe: an automated package for molecular replacement. Acta Crystallogr 1994;A50:157–163.
28. Collaborative Computational Project N. 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr 1994;D50:760–763.
29. Brünger AT, Kuryan J, Karplus M. Crystallographic R factor refinement by molecular dynamics. Science 1987;235:458–460.
30. Brünger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 1992;355:472–475.
31. Jones TA, Zou JY, Cowan SW, Kjeldgaard Y. Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr 1991;A47:110–119.
32. Prendergast FG. Biophysics of the green fluorescent protein. Methods Cell Biol 1999;58:1–18.
33. Elsliger MA, Wachter RM, Hanson GT, Kallio K, Remington SJ. Structural and spectral response of green fluorescent protein variants to changes in pH. Biochemistry 1999;38:5296–5301.
34. Voityuk AA, MichelBeyerle ME, Rosch N. Protonation effects on the chromophore of green fluorescent protein: quantum chemical study of the absorption spectrum. Chem Phys Lett 1997;272:162–167.
35. Voityuk AA, MichelBeyerle ME, Rosch N. Quantum chemical modeling of structure and absorption spectra of the chromophore in green fluorescent proteins. Chem Phys 1998;231:13–25.
36. Cormack BP, Valdivia RH, Falkow S. FACS optimized mutants of the green fluorescent protein (Gfp). Gene 1996;173:33–38.
37. Patterson GH, Knobel SM, Sharif WD, Kain SR, Piston DW. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. Biophysical J 1997;73:2782–2790.
38. Perozzo MA, Ward KB, Thompson RB, Ward WW. X-ray diffraction and time-resolved fluorescence analyses of *Aequorea* Green Fluorescent Protein crystals. J Biol Chem 1988;263:7713–7716.