Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein

Rebekka M Wachter, Marc-André Elsliger, Karen Kallio, George T Hanson and S James Remington*

Background: Because of its ability to spontaneously generate its own fluorophore, the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is used extensively as a fluorescent marker in molecular and cell biology. The yellow fluorescent proteins (YFPs) have the longest wavelength emissions of all GFP variants examined to date. This shift in the spectrum is the result of a T203Y substitution (single-letter amino acid code), a mutation rationally designed on the basis of the X-ray structure of GFP S65T.

Results: We have determined the crystal structures of YFP T203Y/S65G/V68L/S72A and YFP H148G to 2.5 and 2.6 Å resolution, respectively. Both structures show clear electron density for nearly coplanar π–π stacking between Tyr203 and the chromophore. The chromophore has been displaced by nearly 1 Å in comparison to other available structures. Although the H148G mutation results in the generation of a solvent channel to the chromophore cavity, intense fluorescence is maintained. The chromophore in the intact protein can be titrated, and the two variants have pK_a values of 7.0 (YFP) and 8.0 (YFP H148G).

Conclusions: The observed red shift of the T203Y YFP variant is proposed to be mainly due to the additional polarizability of the π-stacked Tyr203. The altered location of the chromophore suggests that the exact positions of nearby residues are not crucial for the chemistry of chromophore formation. The YFPs significantly extend the pH range over which GFPs may be employed as pH indicators in live cells.

Address: Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon 97403, USA.

*Corresponding author.
E-mail: Jim@uoxray.uoregon.edu

Key words: fluorescent tag, green fluorescent protein, pi–pi stacking interaction, spectral shift, structure-based design, X-ray crystallography

Received: **1 June 1998**
Revisions requested: **17 July 1998**
Revisions received: **27 July 1998**
Accepted: **27 July 1998**

Structure 15 October 1998, **6**:1267–1277
http://biomednet.com/elecref/0969212600601267

Introduction

The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* has been used extensively as a fluorescent label in molecular biology. Various applications in physiological studies of living cells have been described in recent reviews [1–4]. The structures of wild-type GFP [5,6] and the variant S65T were determined in 1996 [7]; the overall fold consists of an unusual eleven-stranded β barrel containing a central helix from which the chromophore is formed. The spontaneous generation of the fluorophore is achieved by cyclization of Ser65–Tyr66–Gly67, leading to the formation of a heterocyclic imidazolinone ring, followed by αβ oxidation of Tyr66 in the presence of molecular oxygen [8]. This fluorophore becomes nonfluorescent upon denaturation; the protein fold is therefore essential for function.

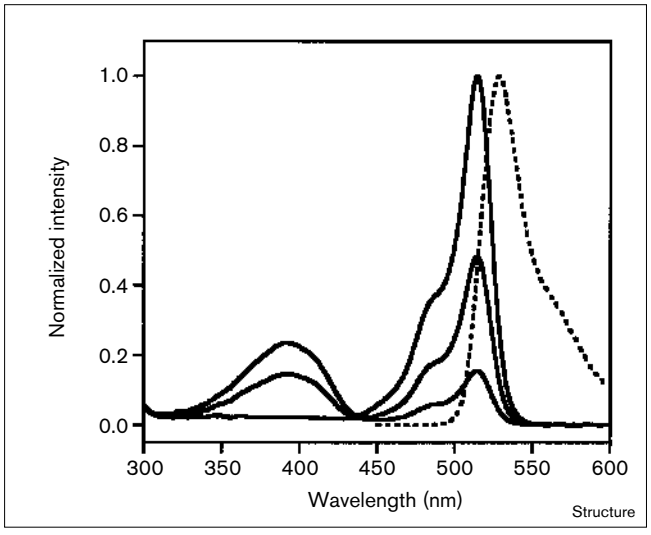
A large number of mutants have been identified that exhibit broadly varying absorption and emission maxima [8,9]. Multicolor reporting of cellular processes has become popular [10] as it allows for simultaneous observation of two or more gene products, if they are labeled with different colored GFP variants. More recently, fluorescence resonance energy transfer (FRET) measurements of different colored GFPs have been used to study protein–protein interactions *in vivo* [9,11–13]. In this type of experiment, two different colored GFPs are chosen so that energy transfer occurs from the emission of the shorter-wavelength GFP variant to the absorbance of the longer-wavelength one. The yellow fluorescent protein (YFP) variant is of particular interest because its spectrum is shifted sufficiently to render it readily distinguishable from the spectrum of cyan fluorescent protein (CFP) in FRET measurements [4,12].

Wild-type GFP exhibits two absorption maxima: the major band absorbs at 398 nm and the minor band absorbs at 475 nm [14]. Excitation of either of these bands leads to emission of green light with a maximum between 504 and 508 nm. Before a structure was available, several GFP variants with altered spectral characteristics were identified by random mutagenesis. Some of these mutants, such as Tyr66→His (Y66H) and Tyr66→Trp (Y66W) [4,8], feature covalent changes within the chromophore itself, resulting in blue-shifted absorbance and emission maxima. Others, such as S65T [15] and T203I [8], focus on changes in the

immediate environment of the chromophore π system. At physiological pH, S65T exhibits only one major absorption band at 489 nm that is red-shifted by 14 nm from that of wild-type GFP and is almost six times brighter [15]. Yet, the emission spectrum is shifted by only 3 nm to 511 nm, and so it cannot easily be distinguished from the wild-type emission. Random-mutagenesis techniques produced only one further red-shifted variant, S65T/M153A/K238E (in single-letter amino acid code), which increases the excitation and emission wavelengths of S65T by 15 and 3 nm, respectively [9].

Here, we describe crystal structures of the first set of GFP variants rationally designed on the basis of the X-ray structure of GFP S65T [7]. These variants, termed YFPs, exhibit the longest wavelength emissions of all GFPs generated by mutagenesis (Figure 1). The YFPs fluoresce around 528 nm, red-shifted by 16 nm relative to S65T and easily distinguishable from it using a fluorescence microscope. The specific YFP investigated here is the quadruple mutant T203Y/S65G/V68L/S72A, in which the substitution T203Y, introduced because of the structural considerations detailed below, is responsible for the red shift. The other three mutations have been shown to improve the brightness of YFP in live cells [16]. The T203Y mutation would have been difficult to isolate by random mutagenesis because this amino-acid substitution requires three changes at the nucleotide level.

Figure 1



Normalized absorbance and fluorescence intensities of the YFP variant in 75 mM phosphate or acetate, and in 140 mM NaCl. Solid lines, absorbance scans at pH 6.0 (top scan at 400 nm), pH 7.0 (middle scan at 400 nm), and pH 8.0 (bottom scan at 400 nm); dashed line, the YFP fluorescence-emission scan at pH 7.0 (excitation wavelength 514 nm).

As Thr203 is positioned close to the chromophore, we speculated that its replacement with a tyrosine would result in π-stacking interactions between the chromophore and the highly polarizable phenol [7], leading to red-shifted spectral properties. The structure of S65T suggested that an aromatic amino acid introduced at position 203 would extend into the water-filled cavity adjacent to the chromophore [7]. Replacement of Thr203 with any of the aromatic amino acids histidine, tryptophan, tyrosine, or phenylalanine was found to result in the desired spectral shifts [7,17]. As the most dramatic red shift was observed for the T203Y substitution, we termed this variant YFP. In order to determine the role (if any) of His148 in modulating the pK_a of the chromophore or its spectral properties, we introduced an additional mutation, H148G, in the YFP background.

The X-ray structures of YFP and YFP H148G were analyzed in order to correlate structural changes more closely with spectral properties. The results clearly reveal the predicted π-stacking interaction and show that solvent access to the chromophore is not detrimental to the general properties of the fluorophore. It appears that the YFPs constitute a particularly versatile set of GFP variants that will be useful in multicolor reporting because of the longer emission wavelength, and as indicators of solvent pH because of their high sensitivity to proton concentration.

Results

Structure determination of YFP and YFP H148G

YFP was crystallized in 2.2 M sodium/potassium phosphate at pH 6.9 in spacegroup $P2_12_12$, with two molecules per asymmetric unit (chains A and B). The GFP S65T structure was used as a search model for molecular replacement against a 3.0 Å dataset using the program AMoRe [18], and the YFP structure was refined. Later, the refined structure of YFP H148G (see below) was used for phasing and refinement against a 2.5 Å dataset. Even though the introduced Tyr203 and the chromophore itself were not modeled during early cycles of refinement, clear electron density for a planar chromophore and a stacked Tyr203 phenol was immediately apparent. Noncrystallographic-symmetry (NCS) constraints were employed throughout refinement of the model using the program TNT and maps were averaged. At the end of refinement, nonaveraged maps for the A and B chains in the asymmetric unit were calculated and compared to one another. No obvious features were identifiable that would suggest significant differences between the two chains. A test run of refinement without any NCS constraints confirmed that the differences would be smaller than the root mean square (rms) error of a 2.5 Å structure. We therefore did not relax or eliminate NCS constraints. Data-collection and atomic-model statistics are shown in Table 1. The final R factor of the YFP model was 19.2% for all data between 20 and 2.5 Å resolution.

Table 1

Data collection and atomic-model statistics for YFP and YFP H148G.

	YFP	YFP H148G
Total observations	53,039	29,904
Unique reflections	18,916	7,373
Completeness*(%)	92	99
Completeness (shell†; %)	94	97
Number of crystals	2	1
R_{merge} ‡ (%)	8.0	6.5
Resolution (Å)	2.5	2.6
Atomic model statistics		
Space group	$P2_12_12$	$P2_12_12_1$
Molecules per asymmetric unit	2	1
Crystallographic R factor	0.192	0.159
Protein atoms	1,810	1,810
Solvent atoms per asymmetric unit	130	30
Bond length deviations (Å)	0.013	0.012
Bond angle deviations (°)	1.76	2.07
Thermal parameter restraints (Å^2)	4.53	3.82

*Completeness is the ratio of the number of observed $I > 0$ divided by the theoretically possible number of intensities. †Shell is the highest resolution shell (2.56–2.50 Å for YFP, and 2.80–2.60 Å for YFP H148G). ‡$R_{merge} = \Sigma | I_{hkl} - < I > | / \Sigma < I >$, where $< I >$ is the average of individual measurements of I_{hkl}.

The refined YFP structure clearly shows that the overall fold is undisturbed, with an rms deviation from the GFP S65T structure of 0.36 Å for α carbons. Three larger contact areas with adjacent molecules were identified; the largest of these covers ~722 Å^2 of one monomer surface, includes a series of hydrophobic residues consisting of Ala206, Phe223, and Leu221, as well as a number of hydrophilic contacts. This interface is essentially identical to the dimer interface for wild-type GFP described by Yang *et al.* [5]. High salt conditions during crystallization experiments appear to favor dimerization, as has been suggested previously [19].

YFP H148G crystallized as a monomer in the presence of polyethylene glycol and acetate at pH 4.6 in spacegroup $P2_12_12_1$, isomorphous to S65T [7] and the blue-emission variant BFP [20]. Molecular replacement using the S65T structure for phasing and refinement gave a final model with an R factor of 15.9% for all data between 24.0 and 2.6 Å (Table 1). As with YFP, electron density for a stacked phenol was clearly visible even before the Tyr203 ring was modeled. The rms deviation between YFP H148G and S65T α carbons is 0.31 Å, and the deviation between YFP H148G and YFP α carbons is 0.35 Å. The β strands of the two YFP variants overlay closely in all areas except around the C_α of residue 148, where a movement of 1.1 Å is observed (see below). This movement has not been observed in other pH 4.6 structures grown under similar conditions and crystallizing in the same spacegroup, such as the BFP structure [20]. Residue 148 and adjacent residues are not involved in crystal contacts, further indicating that the observed movement is due to the H148G substitution and not due to the crystallization conditions.

The introduced phenol

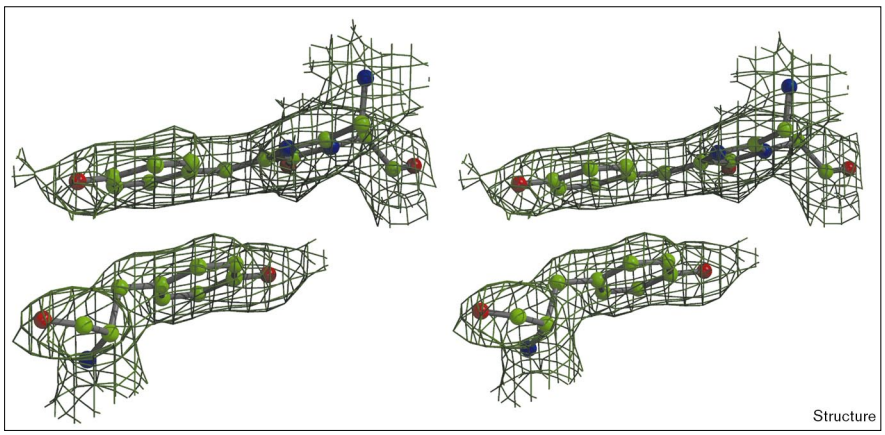
Electron densities of the chromophore and the phenol ring of Tyr203 appeared to be completely planar before the atoms for these groups were added to the model. When the tyrosine sidechain was first introduced into the model, it was modelled as being coplanar with the chromophore. Refinement consistently rotated the phenol ring by 12° with respect to the chromophore plane in both YFP and YFP H148G. Figure 2 shows the electron density of the refined YFP chromophore structure together with the phenol ring of Tyr203. The distance of closest approach between atoms of the two interacting rings is 3.3 to 3.4 Å, and occurs at the edge of the chromophore plane that is opposite the *exo*-methylene bond (Figure 2). It appears that the phenol tilts towards this area of the chromophore because it is more open, with fewer atoms to clash with sterically.

The distance of largest separation between the rings is 3.5 to 3.8 Å; it occurs at the same edge of the chromophore plane as the *exo*-methylene bond, so steric clash with the *exo*-methylene carbon is avoided. This range of plane-to-plane distances is typical for face-to-face π–π stacking interactions found in proteins, and is consistent with interaction energy calculations that show a potential-energy minimum for two horizontally stacked benzene molecules with a vertical separation of 3.3 Å [21]. A recent analysis of protein structures [22] has led to the conclusion that aromatic-ring interactions in an off-centered parallel orientation have an energetically favorable, stabilizing effect and, in fact, are the preferred interactions.

Positional shift of the chromophore

The entire chromophore ring system of YFP has moved out towards the protein surface by about 0.9 Å in comparison to that of S65T or wild-type GFP (Figure 3). The chromophore of YFP H148G has moved in the same direction, but to a lesser extent, ~0.5 Å. Overlay of all α carbons shows that this shift is very much a local effect, involving only residues 65–68 (Figure 3). The overlay suggests that this shift may be due to the compensating effects of the V68L and S65G substitutions. The Leu68 $C_{\delta 1}$ occupies the same space as the original Val68 $C_{\gamma 1}$, whereas the Leu68 backbone is displaced so that the chromophore is pushed further out towards the protein surface. As part of the same movement, the C_α of Gly65 is pushed into the position of the wild-type C_β of Ser65 (Figure 3). The V68L and S65G substitutions had been previously found to increase the brightness of GFP-expressing cells significantly in a wild-type background [16], and were suggested to either improve folding at 37° or increase the rate of chromophore formation. It is unclear at this point why the chromophore is not shifted to the same extent in the YFP

Figure 2



Stereoview of the $2F_o–F_c$ electron-density map of the YFP chromophore and the stacked Tyr203 after refinement. The 2.5 Å resolution map was contoured at +1 standard deviation.

and YFP H148G structures, even though both of these incorporate the V68L and S65G mutations.

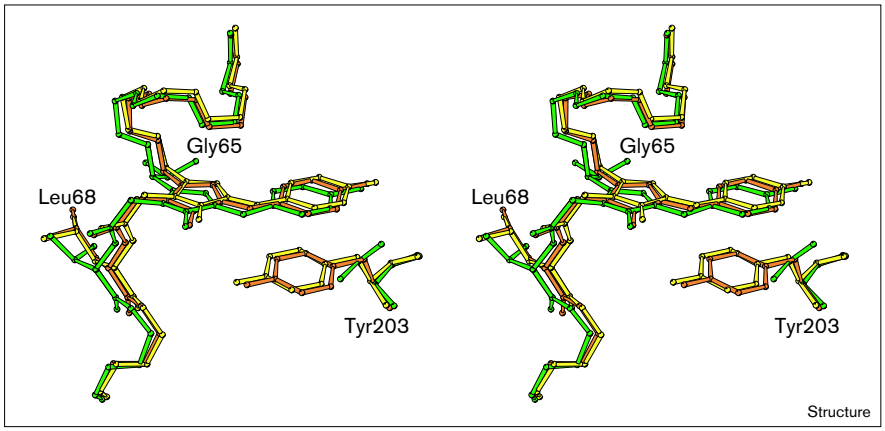
Although the imidazolinone ring of the YFPs is not in the same position as in wild-type GFP [6] , S65T [7], and BFP [20], no electron density consistent with a partially formed or unformed chromophore is observed. This indicates that the machinery generating the chromophore is not only intact, but is also more flexible than previously thought (see Discussion section). Given the nearly exact superposition of the imidazolinone rings observed in wild-type GFP, S65T, and BFP [5–7,19], the exact positions of the backbone atoms of residues 65 and 67 that undergo the cyclization reaction is apparently not as crucial as was previously suggested [19,20].

Chromophore spectral properties, charge state and hydrogen-bonding interactions

Like S65T [23], the YFPs have two absorbance maxima whose relative ratio is pH-dependent (Figure 1 and Table 2). The UV-absorption peaks at 392 (YFP) or 397 nm (YFP H148G) have been ascribed to the neutral chromophore, whereas the visible absorption peaks at 514 (YFP) or 512 nm (YFP H148G) have been ascribed to the anionic chromophore [24]. The lower-energy peak exhibits clear vibrational structure, as indicated by the pronounced shoulder at 480–490 nm, and its mirror-image relationship with the emission band is striking (Figure 1). These features are consistent with luminescence properties of large and rigid systems in condensed phases [25] and may be more pronounced in the YFPs because of decreased chromophore flexibility in the presence of the stacked phenol. Both the YFPs fluoresce intensely when excited at the longer-wavelength band, with maximum emission occurring at 528 nm (Figure 2). Fluorescence is extremely weak when the excitation occurs at the shorter-wavelength band (Table 2), even if the experiment is carried out at a pH value at which this peak dominates.

The chromophore pK_a in the intact protein was determined as being 7.00 (±0.03) for YFP and 8.02 (±0.01) for YFP

Figure 3



Stereo drawing of the chromophore showing residues 65, 68, 203, and backbone atoms of residues 61–64, 69 and 70. Green, S65T; yellow, YFP; orange, YFP H148G. The α carbons of the entire structures were overlayed (see text).

Table 2

Summary of absorption and emission maxima.

	Absorbance band 1	Absorbance band 2	Emission band 1*	Emission band 2*
Wild-type GFP	398	475	460/508	504
S65T	394	489	(weak)	511
YFP	392	514	(weak)	528
YFP H148G	397	512	(weak)	528

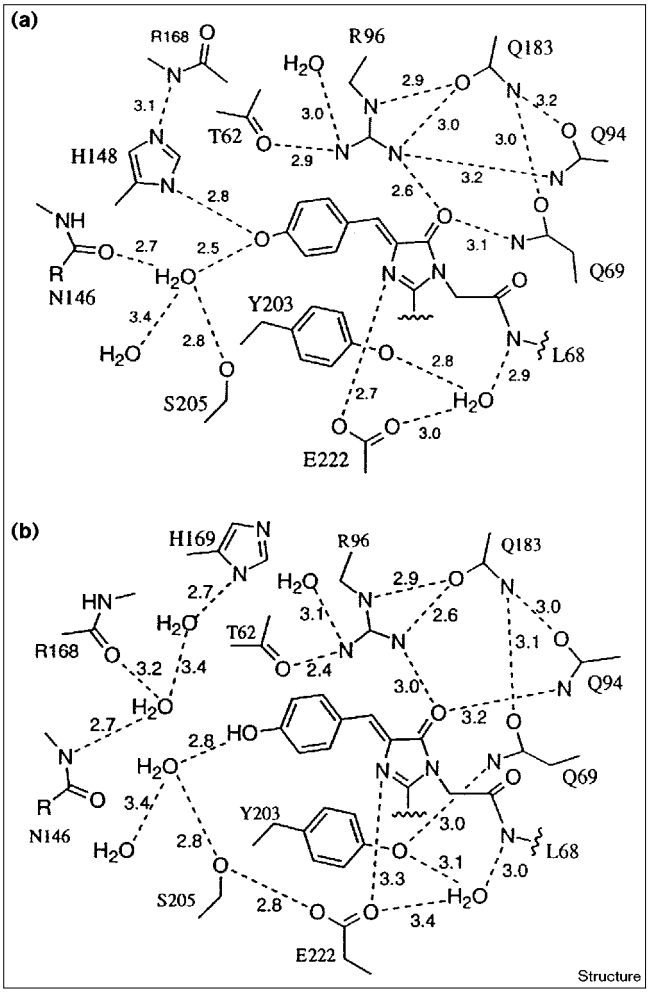
*The emission band 1 results from excitation at the absorbance peak 1, and the emission band 2 results from excitation at the absorbance peak 2.

H148G by absorbance measurements at varying pH. The pK_a values determined by fluorescence were 6.95 (±0.03) and 7.93 (±0.04), respectively, for the two variants. The YFP pK_a is remarkably similar to that of 'enhanced' EYFP Ser65→Gly/Ser72→Ala/Thr203→Tyr/His231→Leu (S65G/S72A/T203Y/H231L), a closely related variant [26]. All titration curves gave an excellent fit to a single pK_a value. The details of the pH titrations and structural changes that occur upon titration of S65T will be reported elsewhere (RMW, M-AE, KK, GTH and SJR, unpublished results).

It is likely that the charge state of the chromophore is variable in the YFP crystals that were grown at pH 7, which is the chromophore pK_a. In YFP, His148 is directly hydrogen-bonded to the phenolic end of the chromophore (Figures 4 and 5). Its electron density is well-defined, suggesting that the imidazole ring does not change position when the chromophore ionizes. It is therefore unlikely that structural rearrangements in the immediate chromophore environment occur in response to changes in the chromophore charge state. In both the YFP and YFP H148G structures, the phenolic end of the chromophore is nearly in hydrogen-bonding contact with bulk solvent via two ordered waters (Figure 4) and therefore may not be as tightly embedded in the protein as in the wild-type and S65T [6,7] structures. Structural readjustments to accommodate the anion may only affect solvent molecules.

The strong hydrogen bond to Arg96 that has been suggested to play a role in the chemistry of backbone cyclization [7] is maintained in both structures (Figures 4 and 5). The carbonyl oxygen of the chromophore imidazolinone ring interacts with two hydrogen-bond donors, Arg96 and Gln69 in YFP, and Arg96 and Gln94 in YFP H148G. This compares to similar interactions with Arg96 and Gln94 in the wild-type and S65T GFP variants. The Glu222 carboxy oxygen approaches the chromophore imidazolinone-ring nitrogen to within 2.7 (YFP) and 3.3 Å (YFP H148G), considerably closer than in the wild-type and S65T structure (4.3 and 4.0 Å, respectively). This close approach appears to be related to the chromophore positional shift described above. Distance and geometry for hydrogen bonding between Glu222 and the

Figure 4



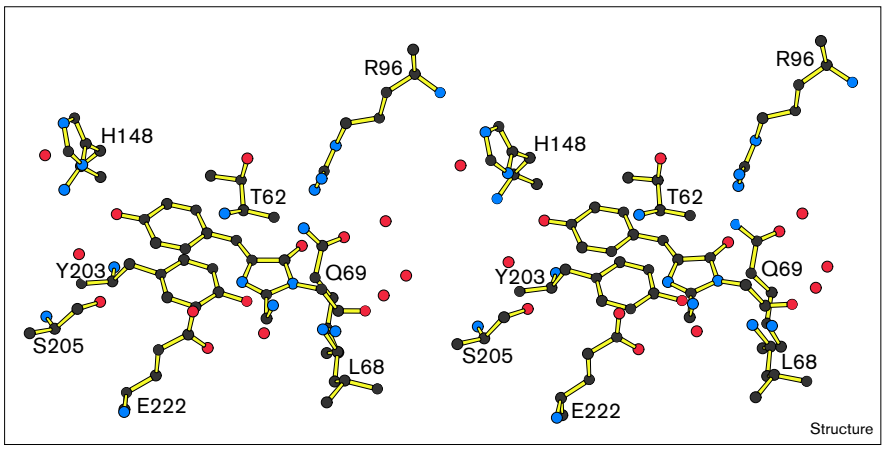
Schematic diagram showing the immediate chromophore environment of **(a)** YFP and **(b)** YFP H148G. The proposed positions of hydrogen bonds are shown as dashed lines.

chromophore-ring nitrogen are excellent in YFP and somewhat less optimal in YFP H148G, where the presumed hydrogen bond makes a roughly 45° angle with the chromophore plane. The YFP structure is the first GFP structure solved that suggests hydrogen-bonding interactions of the heterocyclic-ring nitrogen. The most likely interpretation in terms of charge states is a deprotonated-ring nitrogen and a protonated Glu222, rendering both groups neutral, but it is clear that they share a proton.

Solvent-accessible surface and cavities

To examine the effects of solvent accessibility on the fluorescent properties and the ionization constant of the chromophore, the mutation H148G was introduced into YFP (see the description of the X-ray structure above). In all the GFP structures examined to date, the β barrel is somewhat perturbed around the phenolic end of the chromophore. The β strand that covers the chromophore in that

Figure 5



Stereoview of the YFP chromophore and its environment. Oxygen atoms are shown in red, nitrogen atoms in blue, and carbon atoms in gray.

area bulges out around His148 (Figure 6), so that the backbone from residues 144 to 150 is not directly hydrogen bonded to the adjacent backbone between residues 165 and 170. Rather, they are laced together by forming hydrogen bonds with the imidazole ring of His148 (Arg168 backbone N to His148 $N_{\varepsilon 2}$ in S65T and in the wild-type GFP) and several water molecules. The phenolic end of the chromophore is located directly 'behind' the ring of His148 (Figure 6). We anticipated that the substitution of histidine with glycine would open up a solvent channel to the chromophore in the absence of other structural perturbances, or perhaps permit the bulge to close up.

The crystal structure clearly shows this anticipated solvent channel as an invagination of the protein surface within which there are no ordered solvent molecules (Figure 7). Elimination of the imidazole ring in the H148G substitution leads to only minor structural rearrangements of protein groups. The β strands do not close up to form a directly hydrogen-bonded sheet between residues 144 and 150. Instead, the C_α of residue 148 has actually moved in the opposite direction by 1.1 Å, causing an even larger strand separation between the backbones of residues 148 and 168 (Figure 7). The sidechain of Ile167 has moved by 1.1 Å towards the space previously occupied by the imidazole ring. Nevertheless, direct solvent access to the phenolic end of the chromophore is greatly improved (Figure 7). Calculation of the solvent-accessible area of the chromophore using a probe sphere of radius of 1.4 Å [27], as implemented by MidasPlus™ [28], shows that 22% of the chromophore surface is solvent-accessible. Only the phenolic end of the chromophore is exposed to exterior solvent, though, because of the opening in the protein wall (Figure 7). The phenolic oxygen of the chromophore is also hydrogen bonded to a water molecule that is near hydrogen-bonding distance to a surface water, although a 1.4 Å probe cannot access the chromophore via this path (Figure 4b). If both the solvent channel and this hydrogen bond are included, 8% of the chromophore surface, entirely at the phenolic end, is accessible to exterior solvent, and 14% is accessible to interior solvent as a result of contact with internal cavities (Figure 7).

YFP H148G contains two larger interior cavities that are in contact with the chromophore cavity and are filled with some ordered waters (Figure 7). The cavity that was largest in S65T has decreased in size from approximately 127 $Å^3$ (S65T) to 88 $Å^3$ (YFP H148G), because some of the space is now filled with the phenol of Tyr203. In YFP, this cavity is not accessible to a 1.4 Å probe at all, because several groups have moved into this space. The more significant structural adjustments are $C_{\gamma 2}$ of Val224, which has moved by 1.4 Å, and $C_{\delta 1}$ of Leu42, which has moved by 2.0 Å, essentially filling the cavity. The second larger cavity in contact with the chromophore is nearly invariant for S65T, YFP, and YFP H148G, and is between 98 and 103 $Å^3$ in size.

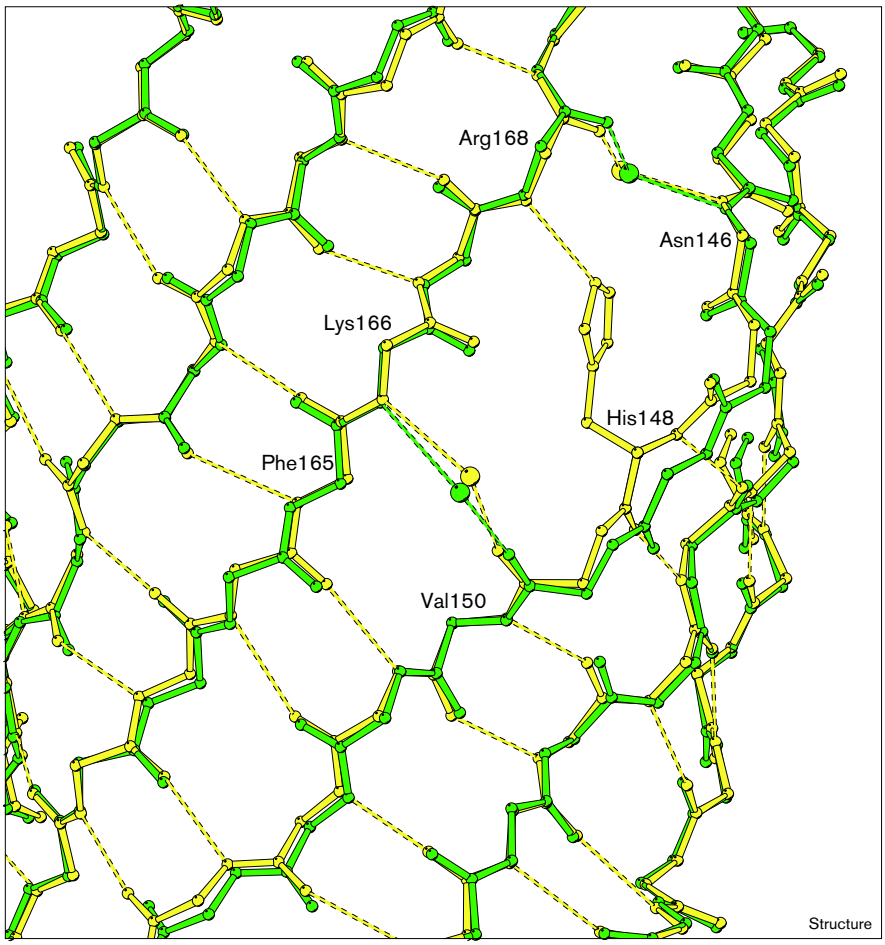
Discussion

Possible mechanisms for the observed red shift of absorption and fluorescence

The observed π-stacking interaction of the chromophore with Tyr203 in the YFPs serves as a basis for discussing possible mechanisms that could be operational in the red shift. Several different factors could be responsible for the observed shifts in absorbance and fluorescence in the YFPs (Table 2), and the structural studies alone do not necessarily allow one to discriminate between them. The photophysics is further complicated by the fact that the chromophore ground state exists as an equilibrium of two charge states that absorb at different energies [24,29] (Table 2). It is very interesting that, in the YFPs, only the low-energy absorption band that is due to the chromophore anion is red-shifted. The high-energy absorbance band shifts by only a few nm in wild-type GFP, S65T, and the YFPs, with no clear trend that could be attributed to the

Figure 6

Overlay of the backbone-atom trace of YFP (yellow) and YFP H148G (green). Part of the β barrel and the β-strand bulge around His148 are shown. The dashed lines represent possible hydrogen bonds.

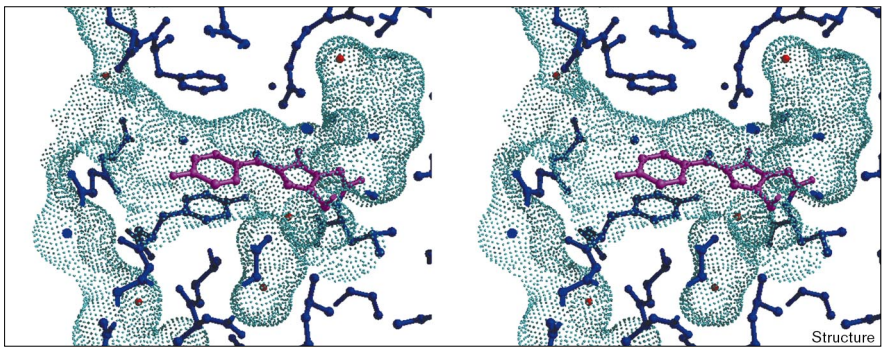


π-stacking interaction (Table 2). On the other hand, the low-energy band due to the chromophore anion (Table 2) is red-shifted by 25 nm in comparison with that of S65T. Apparently, the electronic energy levels of the anionic state are strongly affected by the stacking interaction, with little or no effect on the neutral state. This poses an interesting challenge to computational chemists, and any calculations of the spectral properties of model compounds will have to take this observation into account.

One important factor to consider is the polarizability of groups adjacent to the chromophore, such as the phenol of Tyr203. Extended π systems can respond essentially instantaneously to a change in chromophore polarity by a shift of the π-electron cloud, preferentially stabilizing the electronic state with the larger dipole moment [30]. A spectral shift to longer wavelengths is expected if the excited state has a larger dipole moment than the ground state, which is almost always the case for π–π* transitions [30]. A good example is tryptophan, where polar solvents decrease the energy separation between π and π* levels as a result of preferential stabilization of the excited state, causing a red shift in the absorption spectrum [31]. The dipole moment magnitude of the wild-type GFP chromophore in state B (475 nm absorbing) has been shown to change significantly upon excitation, and the excited state has been shown to be more acidic than the ground state [29], suggesting a decrease in negative charge on the phenolic oxygen upon excitation. This is consistent with what is known about phenols: there is a net transfer of charge from the oxygen to the phenol ring in the lowest excited singlet state, which lowers its pK_a [25]. These observations are consistent with the observed red shift in YFP in the presence of the highly polarizable phenol of Tyr203.

Replacement of Thr203 with a histidine leads to emission at 524 nm [7], and replacement with a phenylalanine results in emission at 525 nm [17]. Both maxima are close to the values between 525 and 528 nm reported for a series of variants incorporating the T203Y substitution [7] (Table 2). All three sidechains are aromatic and highly polarizable, although their permanent dipole moments vary in both magnitude and direction, with phenylalanine

Figure 7



Stereoview of the solvent-accessible surface of YFP H148G, calculated using a 1.4 Å probe radius (see text). The surface (cyan) was calculated after deleting the chromophore (magenta) and all water molecules (red). The outer surface of the protein is along the left edge of the figure. Internal cavities appear as appendages to the chromophore cavity within the right half of the figure (see text).

being essentially apolar. These results suggest that dipolar interactions between the stacked groups do not contribute significantly to the observed red shift, whereas polarizability is likely to be a major factor. In further agreement with this interpretation are the results of Niwa *et al.* [24] that show a clear red shift, which follows the polarizability, of the synthetically generated chromophore with solvents of increasing refractive index.

Another important factor to consider in light of spectral shifts is the planarity of the chromophore π system. In both the YFP and the YFP H148G structures the chromophore appears entirely planar (Figure 3), consistent with all other GFP structures published to date [5–7,18,19]. Crystals were grown at low pH (pH 4.6, YFP H148G), neutral pH (pH 6.9, YFP), and higher pH (pH 8.0, S65T) [7], suggesting that the planarity is not perturbed by the charge state of the chromophore. Even so, a small out-of-plane distortion, such as bending by a few degrees or twisting around the *exo*-methylene bond, is possible but might not be detected because of coordinate error in the refined structures. We therefore cannot exclude the possibility that the π interaction of the chromophore with Tyr203 or the hydrogen bond of the heterocyclic nitrogen to Glu222 could subtly modulate planarity.

Modulation of absorption and emission maxima by out-of-plane distortion is well-documented for porphyrin rings [32]. Ruffling of substituted, sterically crowded porphyrins has been shown to lead to a significant red shift of absorption bands that arise from π–π* transitions because the highest filled orbitals are destabilized due to a decrease in conjugation, whereas the energy of the lowest unoccupied orbitals is less affected [33]. In these compounds, the red shift becomes significant around a dihedral angle of 20°, where an 18 nm shift is observed for some of the transitions [33]. On the other hand, acetyl substitutions at the periphery of the porphyrin skeleton, with torsion angles departing significantly from planarity, break the conjugation with the macrocycle and therefore lead to blue-shifting of optical spectra [34]. By analogy, any torsion around the *exo*-cyclic double bond of the GFP chromophore would be expected to shift the spectrum to higher energy, and any bending of the chromophore plane could potentially shift the spectrum to lower energy. Although it is difficult to estimate the degree of spectral perturbation caused by a slight out-of-plane distortion of the chromophore, the effect is probably small (if present at all) in the YFPs.

Another consideration regarding optical characteristics of the GFP variants is the hydrogen-bond pattern around the chromophore. The most obvious difference in hydrogen bonding between YFP and S65T [7] is the hydrogen bond between Glu222 and the heterocyclic ring nitrogen, suggesting a neutral chromophore-ring nitrogen (Figure 4). There is no experimental evidence for the imidazolinone-ring nitrogen being protonated in any of the structures solved to date, a result that appears to be contradictory to the results of theoretical calculations [35]. Nevertheless, hydrogen bonding or protonation of the ring nitrogen would not be expected to change the spectral characteristics dramatically given that the nitrogen orbital in question is not part of the chromophore π system. Small changes in the geometries of the other hydrogen bonds to the chromophore, though beyond the resolution limit of the YFP structures, are likely to occur and may influence the positions of spectral bands to some degree.

The ability of hydrogen bonds to rearrange themselves to relax to a lower energy conformation upon chromophore excitation should modulate the observed Stokes shifts. It appears that, to a first approximation, the more tightly the protein atoms are packed against the surface area of the chromophore in GFP, the smaller the Stokes shift; with the low-energy band (Table 2), the Stokes shift is 29 nm for the wild-type, 22 nm for S65T, 16 nm for YFP H148G, and 14 nm for YFP. This series is roughly in the order of increasing number of protein atoms in van der Waals contact with the chromophore. Tighter packing may limit positional rearrangements of sidechains because of steric constraints

and may therefore allow for less stabilization of the excited state. Tighter packing may also result in less vibrational relaxation of the excited state. It might be possible to design a new GFP variant with similar π-stacking interactions and absorbance maxima to those of the YFPs, but with a larger Stokes shift due to a reduction of the number of protein atoms that are in contact with the chromophore.

The chemistry of chromophore formation

The YFP structures exhibit shifts in chromophore position (Figure 3) that have not been observed previously with other variants. This suggests that the exact positions of potential catalytic groups, if any are required, may not be crucial for backbone cyclization. It is intriguing to consider the possibility that no groups directly catalyze specific chemical steps such as protonation or deprotonation, and that the protein fold as a whole may be solely responsible for effecting the proximity of reacting atoms. Removing more than one residue at the N terminus, or more than six residues at the C terminus, renders the protein unable to form the chromophore [36]. Therefore, the β barrel of GFP has to be intact for the generation of fluorescence, and it may lock the chromophore-forming segment of the central helix into a conformation in which there is a very short distance between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67, facilitating nucleophilic attack [37]. The two most obvious candidates for catalytic groups are Arg96 and Glu222, but neither is required for chromophore formation (R Ranganathan, unpublished results) [38], consistent with our observations.

The changes in chromophore position, though striking, all occur within the same plane, and are therefore essentially two-dimensional (Figure 3). This may imply that the positions of reacting atoms in the third dimension, roughly along the long axis of the β barrel, are more strictly defined by the constraints of the protein fold. The same plane is occupied by the Tyr203 phenol in both of the YFP structures as well, with the difference that, in YFP, the two stacked rings have slid past one another to some extent, whereas in YFP H148G, they face one another more closely (Figure 3). This effect may be due to the small structural adjustments resulting from the replacement of His148.

Solvent accessibility to the chromophore

YFP H148G was found to be highly fluorescent, with a bright greenish-yellow color in ordinary daylight. The light-emitting properties of the fluorophore do not appear to be changed to any extent by the introduction of a solvent channel to the chromophore, indicating that significant quenching does not occur. As the protein fold is entirely intact in YFP H148G in spite of the generation of an opening in the β barrel, the H148G substitution may be especially useful for allowing the access of various small-molecule species to the chromophore. This substitution may be introduced into other GFP variants with a larger cavity adjacent to the chromophore, such as S65T [7] or S65G, allowing for analyte-binding studies where specific spectral shifts resulting from the interaction with small molecules or ions of interest could be monitored.

Elevated chromophore ionization constant

The spectra of several GFP mutants are highly sensitive to pH (RMW, M-AE, KK, GTH and SJR, unpublished results). These mutants may therefore be useful as probes for the measurement of pH in specific cellular compartments of live tissue, as was originally proposed for the blue-emission variant BFP [20]. That this is feasible has recently been shown for S65T, which has a pK_a of 5.9 [23] and which was demonstrated to work well for the measurement of mitochondrial and Golgi pH using the EYFP, which includes the T203Y substitution and has a pK_a of 7.1 [26]. Titrations of pH reveal that the YFPs generally have pK_a values at or above 7; potential protonation site(s) of the chromophore will be discussed elsewhere (RMW, M-AE, KK, GTH and SJR, unpublished results). The highest ionization constant of all variants examined to date is found for the YFP H148G mutant with a pK_a of 8.0. In this mutant, the chromophore is solvent-exposed, consistent with a similarly high pK_a when the protein is denatured [39]. YFP H148G could be a particularly useful research tool for studying the acidity of organelles with elevated pH. For example, mitochondrial pH has been shown to be higher than 7.5 using S65T as a pH-sensitive probe [23], but it could not be determined more accurately because S65T fluorescence is not pH-sensitive above pH 7.

Biological implications

Mutants of green fluorescent protein (GFP) form an excellent model system for the study of how the protein environment can modify the chromophore absorption and emission characteristics of biological pigments. Such studies are relevant to all aspects of photobiology, and a careful investigation of structural and photophysical properties of the GFPs should lead to a better understanding of photoreceptors in general. The YFPs represent a particularly interesting set of GFP variants because of the aromatic-ring stacking interaction and their red-shifted spectra. The availability of their structures is expected to stimulate spectroscopic investigations that are aimed at gaining a better understanding of the factors responsible for the modulation of optical properties. This information should aid in the design of other types of mutants with novel spectroscopic characteristics. For example, π-stacking mutants of particular interest may include the blue- or cyan-fluorescent variants that have an altered chromophore structure as a result of the incorporation of histidine or tryptophan at position 66.

YFP and its variants constitute extremely useful probes in molecular and cell biology, with a range of applications

from multicolor reporting and protein-interaction studies to the measurement of the acidity of cellular compartments with elevated pH. The more general aspects of the π-stacking interaction observed for the YFPs could be helpful in the design of GFP variants with novel and useful properties, such as further red-shifted spectra that can be distinguished from those of other GFPs under a fluorescent microscope. Of considerable interest to biologists and biochemists would be red-shifted GFP with a larger Stokes shift — a larger separation between excitation and emission wavelengths. This could be achieved by decreasing the number of protein atoms packed around the chromophore. Recently, it has been shown that red-fluorescent proteins (RFPs) can be generated by photoactivation of various GFPs with blue light under anaerobic conditions [40,41]. The chromophore structure of these red photoproducts is, as yet, unknown, and their usefulness is somewhat limited, because the introduction of atmospheric oxygen leads to loss of fluorescence. Understanding the oxygen-sensitivity of the RFPs in conjunction with what is known about YFP should greatly aid in the design of stable red-fluorescent variants.

Materials and methods

Mutagenesis and protein preparation

GFP variants were prepared using the polymerase chain reaction (PCR) based QuikChange™ site-directed mutagenesis kit (Stratagene, La Jolla, CA), following the manufacturer's instructions and using the coding sequence of the original YFP mutation 10c as template DNA [7]. This template incorporates the mutations T203Y/S65G/V68L/S72A, as well as the ubiquitous Q80R substitution that was accidentally introduced into the *gfp* cDNA early on [7,42]. All of the GFP variants were expressed and purified as described in [7].

Structure determination of YFP H148G

YFP H148G was concentrated to 12 mg/ml in 20 mM HEPES pH 7.9. Rod-shaped crystals with approximate dimensions of $1.8\times0.08\times0.04$ mm were grown in hanging drops containing 2 μl protein and 2 μl mother liquor at 4°C within four days. The mother liquor contained 16% PEG 4000, 50 mM sodium acetate pH 4.6, and 50 mM ammonium acetate. X-ray diffraction data were collected from a single crystal at room temperature using a Xuong-Hamlin area detector [43]. Data were collected in spacegroup $P2_12_12_1$ to 99% completeness at 2.6 Å resolution, and were reduced using the supplied software [44]. Unit cell parameters were a = 52.0, b = 62.7, and c = 69.9. The GFP S65T coordinate file [7], which served as a model for phasing, was edited to reflect the mutations, with the introduced residues Tyr203 and Leu68 initially modeled as alanines to prevent model bias. A model for the anionic chromophore was obtained by semi-empirical molecular-orbital calculations using AM1 in the program SPARTAN version 4.1 (Wavefunction Inc., Irvine, CA). The minimized structure, which was planar, compared very favorably with a related small-molecule crystallographic structure [45] and also with the model used during refinement of GFP S65T, where a simpler modeling program had been employed [7].

Using the program TNT [46], rigid-body refinement was carried out to position the isomorphous model in the unit cell of YFP H148G. Initial positional refinement was carried out using the data to 4.0 Å, then to 3.5, 3.0, and finally to 2.6 Å. Electron-density maps ($2F_o-F_c$ and F_o-F_c) were inspected using the program O [47], and solvent molecules were added if they were consistent with F_o-F_c features, and only when in proximity of hydrogen-bond partners. B factors were refined using a strong correlation between neighbouring atoms because of the relatively low resolution. As no B factor library is available for the chromophore itself, the B factors of all chromophore atoms were set to the values obtained in the 1.9 Å structure of GFP S65T [7] and then refined as a group, with identical shifts for the grouped atoms.

Structure determination of YFP

YFP was concentrated to 10 mg/ml in 50 mM HEPES pH 7.5. After two weeks crystals grew to a size of $0.03\times0.12\times0.8$ mm at 15°C in hanging drops containing 5 μl protein and 5 μl well solution, which contained 2.2 M sodium/potassium phosphate pH 6.9. These crystals belong to space group $P2_12_12$ and have the unit cell dimensions a = 77.1, b = 117.4, and c = 62.7. X-ray diffraction data were collected on two isomorphic crystals at room temperature using an Raxis-IV imaging plate mounted on a Rigaku RUH3 rotating-anode generator equipped with mirrors. The data were processed with the program Denzo and scaled using the program ScalePack [48]. The YFP structure was solved by molecular replacement using the program AMoRe [18], with the 1.9 Å GFP S65T coordinate file as the search model [7]. Two solutions were identified, consistent with two molecules per asymmetric unit.

For refinement, we chose the 2.6 Å structure of YFP H148G as the initial model, which was edited to reflect the mutations present in YFP. To avoid model bias, the occupancies of the Tyr203 sidechain atoms and all chromophore atoms were set to zero during the first few rounds of refinement. Constrained NCS averaging over the A and B chains in the asymmetric unit was applied, initial refinement was carried out to 3.5 Å only, and the electron density maps ($2F_o-F_c$ and F_o-F_c) were averaged. These maps were then inspected and the model adjusted using O [47], followed by additional positional refinement to 2.5 Å. Chromophore and Tyr203 densities were very clear and both were planar. The model was edited to include these groups in refinement, and solvent molecules were added where appropriate. B factors were refined using a strong correlation between neighbouring atoms because of the relatively low resolution.

Absorbance and fluorescence spectra

Small aliquots of protein (16 mg/ml) were diluted 48-fold into 75 mM buffer (acetate, phosphate, Tris, or CHES), 140 mM NaCl, and then scanned for absorbance between 250 and 600 nm (Shimadzu 2101 spectrophotometer at medium scan rate and room temperature). The optical density at 514 or 512 nm was plotted as a function of pH and computer-fitted to a titration curve (Kaleidagraph™, Synergy Software).

Fluorescence measurements were carried out on a Hitachi F4500 fluorescence spectrophotometer at a constant protein concentration of approximately 0.01 mg/ml, with buffer conditions identical to those of absorbance measurements. The excitation wavelength was set to the absorbance maximum of the long-wave band of the particular mutant (see Table 2). The emission was scanned between 500 and 600 nm, and peak emission intensity was plotted as a function of pH and curve-fitted as above.

Accession numbers

The coordinates for YFP and YFP H148G have been deposited in the Brookhaven Protein Data Bank with accession codes 1yfp and 2yfp, respectively.

Acknowledgements

This work was supported in part by grants from the National Science Foundation (MCB 9728162 to SJR), a National Institutes of Health postdoctoral fellowship (1 F32 GM19075-01 to RMW), and a Fonds de la recherche en sante du Quebec fellowship (980457-103 to M-AE).

References

1. Cubitt, A.B., *et al.*, & Tsien, R.Y. (1995). Understanding, improving, and using green fluorescent proteins. *Trends Biochem. Sci.* **20**, 448-455.
2. Prasher, D.C. (1995). Using GFP to see the light. *Trends Genet.* **11**, 320-323.

3. Gerdes, H.-H. & Kaether, C. (1996). Green fluorescent protein: applications in cell biology. *FEBS Lett.* **389**, 44-47.
4. Tsien, R.Y. (1998). The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509.
5. Yang, F., Moss, L.G. & Phillips, G.N. (1996). The molecular structure of green fluorescent protein. *Nat. Biotech.* **14**, 1246-1251.
6. Brejc, K., *et al*., & Remington, S.J. (1997). Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc. Natl Acad. Sci. USA.* **94**, 2306-2311.
7. Ormö, M., *et al*., & Remington, S.J. (1996). Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392-1395.
8. Heim, R., Prasher, D.C. & Tsien, R.Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl Acad. Sci. USA* **91**, 12501-12504.
9. Heim, R. & Tsien, R.Y. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.* **6**, 178-182.
10. Rizzuto, R., *et al*., & Pozzan, T. (1996). Double labelling of subcellular structures with organelle-targeted GFP mutants *in vivo*. *Curr. Biol.* **6**, 183-188.
11. Mitra, R.D., Silva, C.M. & Youvan, D.C. (1996). Fluorescence resonance energy transfer between blue-emitting and red-shifted excitation derivatives of the green fluorescent protein. *Gene* **173**, 13-17.
12. Miyawaki, A., *et al*., & Tsien, R.Y. (1997). Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* **388**, 882-887.
13. Romoser, V.A., Hinkle, P.M. & Persechini, A. (1997). Detection in living cells of Ca^{2+}-dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. *J. Biol. Chem.* **272**, 13270-13274.
14. Morise, H., Shimomura, O., Johnson, F.H. & Winant, J. (1974). Intermolecular energy transfer in the bioluminescent system of *Aequorea*. *Biochemistry* **13**, 2656-2662.
15. Heim, R., Cubitt, A.B. & Tsien, R.Y. (1995). Improved green fluorescence. *Nature* **373**, 663-664.
16. Cormack, B.P., Valdivia, R.H. & Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33.
17. Dickson, R.M., Cubitt, A.B., Tsien, R.Y. & Moerner, W.E. (1997). On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* **388**, 355-358.
18. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Cryst. A* **50**, 157-163.
19. Palm, G.J., *et al*., & Wlodawer, A. (1997). The structural basis for spectral variations in green fluorescent protein. *Nat. Struct. Biol.* **4**, 361-365.
20. Wachter, R.M., *et al*., & Remington, S.J. (1997). Crystal structure and photodynamic behavior of the blue emission variant Y66H/Y145F of green fluorescent protein. *Biochemistry* **36**, 9759-9765.
21. Burley, S.K. & Petsko, G.A. (1986). Dimerization energetics of benzene and aromatic amino acid side chains. *J. Am. Chem. Soc.* **108**, 7995-8001.
22. McGaughey, G.B., Gagné, M. & Rappé, A.K. (1998). π-Stacking interactions. *J. Biol. Chem.* **273**, 15458-15463.
23. Kneen, M., Farinas, J., Li, Y. & Verkman, A.S. (1998). Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophys. J.* **74**, 1591-1599.
24. Niwa, H., *et al*., & Tsuji, F.I. (1996). Chemical nature of the light emitter of the *Aequorea* green fluorescent protein. *Proc. Natl Acad. Sci. (USA)* **93**, 13617-13622.
25. Barltrop, J.A. & Coyle, J.D. (1978). *Principles of Photochemistry.* John Wiley and Sons, New York. pp 51-52 and 78-79.
26. Llopis, J., McCaffery, J.M., Miyawaki, A., Farquhar, M. & Tsien, R.Y. (1998). Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc. Natl Acad. Sci. USA* **95**, 6803-6808.
27. Connolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709-713.
28. UCSF MidasPlus™. Computer Graphics Laboratory, University of California at San Francisco, CA 94143, USA.
29. Chattoraj, M., King, B.A., Bublitz, G.U. & Boxer, S.G. (1996). Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer. *Proc. Natl Acad. Sci. USA* **93**, 8362-8367.
30. Cantor, C.R. & Schimmel, P.R. (1984). *Biophysical Chemistry Part II: Techniques for the Study of Biological Structure and Function.* W.H. Freeman and Company, New York. pp 386-387 and 392-398.
31. Creed, D. (1984). The photophysics and photochemistry of the near-UV absorbing amino acids–I. Tryptophan and its simple derivatives. *Photochem. Photobiol.* **39**, 537-562.
32. Gentemann, S., *et al*., & Holten, D. (1994). Photophysical properties of conformationally distorted metal-free porphyrins. Investigation into the deactivation mechanisms of the lowest excited singlet state. *J. Am. Chem. Soc.* **116**, 7363-7368.
33. Jentzen, W., *et al*., & Shelnutt, J.A. (1995). Ruffling in a series of Nickel(II) *meso*-tetrasubstituted porphyrins as a model for the conserved ruffling of the heme of cytochromes c. *J. Am. Chem. Soc.* **117**, 11085-11097.
34. Gudowska-Nowak, E., Newton, M.D. & Fajer, J. (1990). Conformational and environmental effects on bacteriochlorophyll optical spectra: Correlations of calculated spectra with structural results. *J. Phys. Chem.* **94**, 5795-5801.
35. Voityuk, A.A., Michel-Beyerle, M.-E. & Rosch, N. (1997). Protonation effects on the chromophore of green fluorescent protein. Quantum chemical study of the absorption spectrum. *Chem. Phys. Lett.* **272**, 162-167.
36. Dopf, J. & Horiagon, T.M. (1996). Deletion mapping of the *Aequorea victoria* green fluorescent protein. *Gene* **173**, 39-44.
37. Branchini, B.R., Nemser, A.R. & Zimmer, M. (1998). A computational analysis of the unique protein-induced tight turn that results in posttranslational chromophore formation in green fluorescent protein. *J. Am. Chem. Soc.* **120**, 1-6.
38. Ehrig, T., O'Kane, D.J. & Prendergast, F.G. (1995). Green-fluorescent protein mutants with altered fluorescence excitation spectra. *FEBS Lett.* **367**, 163-166.
39. Ward, W.W., Cody, C.W., Hart, R.C. & Cormier, M.J. (1980). Spectrophotometric identity of the energy transfer chromophores in *Renilla* and *Aequorea* green-fluorescent proteins. *Photochem. Photobiol.* **31**, 611-615.
40. Elowitz, M.B., Surette, M.G., Wolf, P.-E., Stock, J. & Leibler, S. (1997). Photoactivation turns green fluorescent protein red. *Curr. Biol.* **7**, 809-812.
41. Sawin, K.E. & Nurse, P. (1997). Photoactivation of green fluorescent protein. *Curr. Biol.* **7**, R606-R607.
42. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-805.
43. Hamlin, R. (1985). Multiwire area X-ray diffractometers. *Methods Enzymol.* **114**, 416-452.
44. Howard, A.J., Nielsen, C. & Xuong, N.H. (1985). Software for a diffractometer with multiwire area detector. *Methods Enzymol.* **114**, 452-471.
45. Tinant, B.T., Germain, G., Declercq, J.-P. & Van Meerssche, M. (1980). 4-Benzylidene-2-phenyl-imidazoline-5-one. *Cryst. Struct. Commun.* **9**, 671-674.
46. Tronrud, D.E., Ten Eyck, L.F. & Matthews, B.W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Cryst. A* **43**, 489.
47. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110.
48. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.