Structure and fluorescence mechanism of GFP

Douglas C. Youvan and Maria E. Michel-Beyerle

The crystallographic findings of Yang, Moss, and Phillips[1] and the work of two independent teams of femtosecond spectroscopists[2,3] now provide us with the information to understand the nature of the absorption and emission phenomena of green fluorescent protein (GFP). These phenomena relate the atomic resolution structure of GFP to the first known example of a Förster cycle within the core of a protein—another big surprise from this most unusual protein.

In 1971, Morin and Hastings noticed that an abrupt shift in the wavelength of light emission accompanied isolation of the photocomplex from the colonial hydroid *Obelia*: When a fluorescent protein present in the photocomplex dissociated from its light source, aequorin, the light emitted changed from green to blue. This early observation marks the discovery of GFP (as recently reviewed[4] in a special issue of *Gene*[5]). In the native photocomplex, GFP transduces high-energy blue light emitted by aequorin—which is essentially a luciferase-type photoprotein using a chemiluminescent mechanism—into lower energy green light.

Although fluorescent proteins are widespread in nature, GFP has turned out to be unique. Rather than binding a fluorophore formed through a complex biosynthetic pathway (e.g., tetrapyrrole as in phycobiliproteins), GFP's fluorophore forms autocatalytically on the nascent apoprotein's backbone. A structure for the fluorophore was proposed in a landmark 1993 publication[6]. The new findings[1,2] substantiate all of this earlier work.

A fairly simple model for the mechanism of GFP fluorescence, supported by recent mutagenesis data, has been suggested on the basis of the proposed fluorophore structure and the changes in the fluorescence excitation spectrum as a function of pH. In this model, at high pH or after certain mutations are introduced near the GFP fluorophore, Tyr66 is postulated to be in the phenolate form, whereas at low pH, it is in the hydroxyl form (Fig. 1). Given a small Stokes shift and a large zero point crossing (i.e., significant spectral overlap), one would envisage that the 477-nm excitation band and 510-nm emission band of GFP correspond to transitions involving the first excited singlet state of the phenolate form of Tyr66. Similarly, one would model the hydroxyl form of the fluorophore to be responsible for the 395-nm excitation band, albeit with a remarkably large 105-nm Stokes shift. In the point mutant Ser65 →Thr or the combinatorial mutant RSGFP4, the phenolate form of Tyr66 with the smaller Stokes shift would be favored. In this case, one might be tempted to assign the 395-nm band to a higher-lying S_2 state. As more recent data indicate, this last interpretation is not correct.

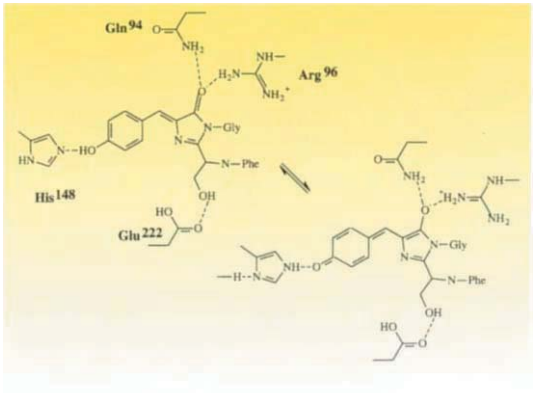

Figure 1. Atomic resolution details of the interaction of four amino acid residues with fluorophore of GFP, indicating resonant structures involved in the Förster cycle. While both forms absorb light, only the phenolate form fluoresces.

Phillips' succinct description of GFP as "paint in a can" begins to convey the basic structural motif: Eleven β-strands surround a central α-helix, which in turn bears the fluorophore. Remarkably, the electron density map of the fluorophore fits the previously proposed structure of an autocatalytically formed heterocycle with extended π conjugation through Tyr66. Some of the more interesting protein contacts with the fluorophore are shown in Figure 1, while all residues within 5 Å are listed in the original work. These structural contacts may eventually explain the autocatalytic formation of the fluorophore and help direct protein engineering projects aimed at creating a "rainbow" of different fluorescent proteins, but immediate insights are provided into other aspects of GFP of current importance to biotechnologists. For example, the successful use of GFP in protein fusions has a structural basis in the tight and compact protein fold, which does not involve either the carboxyl or amino terminus of this protein. The ends are clearly free for His-tagging or fusion to other proteins.

In recent fluorescence resonance energy transfer experiments[7], two different GFP derivatives have been successfully fused using a relatively short 20 amino acid linker between the carboxyl terminus of a green derivative and the amino terminus of a blue derivative. Again, the X-ray crystallographic structure provides insight into why these constructions were feasible. At high protein concentrations, GFP is known to form a dimer. There is a strong implication that the dimer observed in the crystal is in fact the same dimer characterized in solution. The native dimeric structure shows a 25 Å separation between the two fluorophores of the GFP homodimer. Given the twofold symmetrical axis between the dimers in this native structure, one can envisage that a 20 amino acid linker would not disrupt formation of a putative heterodimer.

Phillips' group was able to solve the structure of wild-type GFP. This is surprising, given the observation that the fluorescence of wild-type GFP is rather dim, suggesting that protein preparations would be contaminated with unmodified (nonfluorescent) apoprotein. However, given their mass spectrometry data and the good fit of the electron density around the fluorophore, not much apoprotein is present in the crystal. "Brightness" in *Escherichia coli* seems to have something to do with the absence of inclusion bodies. Wild-type GFP is not very bright compared with Ser65 →Thr or RSGFP4, two of the first mutants with improved fluorescence[8,9]. Although some of this effect is attributable to enhancing the 490-nm excitation band, many researchers believe that these mutations help to solubilize the over-expressed protein in *E. coli*, thereby circumventing the inclusion body/apoprotein problem.

In work in our laboratories, none of the original series of combinatorial mutations in the fluorophore region bears the Ser65 →Thr mutation. In fact, we found a variety of substitutions at this position, including residues as diverse as leucine, cysteine, glycine, and alanine, along with other variants that still have the wild-type serine. In remaking the Ser65 →Thr mutant through oligonucleotide-

Douglas C. Youvan is chief scientific officer at KAIROS Scientific Inc., 3350 Scott Blvd., Building 62, Santa Clara, CA 95054 (dyouvan@kairos-scientific.com) and Maria E. Michel-Beyerle is professor at the Institute for Physical and Theoretical Chemistry, Technical University of Munich, Lichtenbergstrasse 4, D-8046 Garching bei Munchen, Germany (mibe@zentrum.phys.chemie.tu-muenchen.de).

mediated, site-directed, mutagenesis, we found a small percentage of "ultra-bright"mutants, one of which sequenced as a double mutant (Ser65 →Thr; Phe64 →Leu). The second mutation in this pair is presumably a low-frequency synthetic DNA error, but this does bring into question what some researchers are actually working with if they pass Ser65 →Thr a few times and pick bright colonies. The RSGFP8 mutant (Ser65 →Thr; Phe64 →Leu) literally glows in direct sunlight, requiring no filters for imagery. It is the same variant that Falkow's group[10] isolated via an elegant cell sorter enrichment.

Perhaps the most unexpected insight gleaned from the early mutagenesis experiments is the absolute conservation of Gly68 in all known combinatorial mutants[8]. This is the only amino acid in the hexapeptide fluorophore region that cannot be substituted. Presumably, this is a biosynthetic constraint rather than a photochemical constraint on the sequence. The nonenzymatic cyclization of the GFP apoprotein to form the fluorophore may be akin to other deamidation pathways previously described for Asn-Gly or Gln-Gly sequences[11].

So where does this leave us? The fluorophore responsible for the green emission is believed to be the deprotonated (phenolate) and the protonated (hydroxyl) Tyr66 species represented in the absorption and excitation peaks at 395 nm and 477 nm. It has been suggested[9] that the green emission would come only from the deprotonated species, since the excited states of phenols are much more acidic than their ground states. Such an excited state protolytic reaction of a species RH* forming the excited anion R-* has in fact become a classic in solution photochemistry since its discovery by Förster half a century ago[12,13].

The hypothesis of a Förster cycle in GFP is strongly supported by the recent picosecond, time-resolved, fluorescence measurements on native and deuterated GFP[2,3] as well as mutants. Although there are several sites in the fluorophore that could be involved in excited-state proton transfer, the new X-ray structure provides insights to understanding the role of nearby amino acid residues such as the tyrosine on the fluorophore and the adjacent histidine with its inherent proton donor/acceptor properties. For example, the structural data[14] summarized in Figure 1 illustrate how Gln94/Arg96 and His148 could participate in a "push–pull" mechanism for the proton translocation by stabilizing resonant forms of the ionized tyrosine and imidazolidone, respectively.

As we get our first look at the X-ray crystallographic structure of the green fluorescent protein, it is instructive to think that Theodor Förster could not have imagined in 1949 that in 1996 cell biologists would be lighting up the neurons of worms[15] and flies with a protein that transduces blue light into green by a proton translocation cycle that bears his name—such is the nature of important scientific legacies.

1. Yang, F., Moss, L.G., and Phillips, G.N. 1996. *Nature Biotechnology* **14**:1246–1251.
2. Chattoraj, M., King, B.A., Bublitz, G.U., and Boxer, S.G. 1996. *Proc. Natl. Acad. Sci. USA* **93**:8362–8367.
3. Poellinger-Dammer, F., et al., unpublished data.
4. Hastings, J.W. 1996. *Gene* **173**:5–11.
5. Youvan, DC and Larrick, J.W (eds.). 1996. *Gene* **173**:1–117.
6. Cody, C.W., et al. 1993. *Biochemistry* **32**:1212–1218.
7. Mitra, R., Silva, C.M., and Youvan, DC 1996. *Gene* **173**:13–17.
8. Delagrave, S., Hawtin, R.E., Silva, C.M., Yang, M.M., and Youvan, DC 1995. *Bio/Technology* **13**:151–154.
9. Heim, R., Prasher, DC, and Tsien, R.Y. 1994. *Proc. Natl. Acad. Sci. USA* **91**:12501–12504.
10. Cormack, B.P., Valdivia, R.H., and Falkow, S. 1996. *Gene* **173**:33–38.
11. Wright, H. 1991. *Crit. Rev. Biochem. Mol. Biol.* **26**:1–52.
12. Förster, T. 1949 *Naturwiss.* **36**:186–194.
13. Förster , T.Z. 1950. *Elektrochemie* **54**:42–46.
14. Yang, F., Moss, L.G., and Phillips, G.N. 1997. *Proceedings of the 9th International Symposium on Bioluminescence and Chemiluminescence*, in press
15. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, DC 1994. *Science* **263**:802–805.