

Evolving trends in biosciences: multi-purpose proteins – GFP and GFP-like proteins

Mythili Krishna¹ and Baban Ingole^{2,*}

¹Institute of Science and Technology, JNT University, Kukatpally, Hyderabad 500 072, India

²National Institute of Oceanography, Dona Paula, Goa 403 004, India

The sea is considered as holding a clue to many known and unknown biologically active compounds. A family of protein named Green Fluorescent Proteins (GFP)-like proteins, initially isolated from marine organisms, started a trend in biotechnological research, which is expanding day-by-day. A gross review of the same is presented in this article dealing with their occurrence, chemistry, applications, phylogenic analysis and the Indian perspective. These proteins are present in a wide variety of marine organisms, from corals to jellyfish. Chromophore in these proteins is composed of three amino acid residues, Ser65–Try66–Gly67, and requires molecular oxygen for its maturation. Its applications include use in *in vivo* imaging as well as in plant biology. Thus, the protein has become one of the most important tools used in contemporary biosciences. Though the very first protein identified was from the jellyfish *Aequorea victoria* in the 1960s, the list is ever-expanding with not only the fluorescent but also the non-fluorescent proteins being connected to the same superfamily. Hence, they also offer help in the phylogenetic analysis of different organisms, to know which period of evolution has diverted a particular species. As GFP-like proteins have numerous applications under their spectrum, they can be aptly called as ‘multi-purpose proteins’.

Keywords: *Aequorea victoria*, biosciences, chromophore, green fluorescent protein.

PROTEINS are basic biomolecules which play a vital role in many biological processes, from reproduction to the formation of the structure and all through the functioning of almost all organisms. These have been classified into different families based on the similarities in their super secondary structure characteristics or motifs¹. One such family is the green fluorescent protein (GFP)-like protein family discovered from the jellyfish, *Aequorea victoria*². This jellyfish is shaped like a hemispherical umbrella and emits blue fluorescence (Figure 1). During an experiment to identify the compound responsible for the blue fluorescence, a green fluorescent substance was found. The blue fluorescent substance was named as aequorin and the green fluorescent substance was named as green protein

(Figure 2). This pioneering work was conducted by Shimomura and co-workers in the early 1960s. Later, in 1971

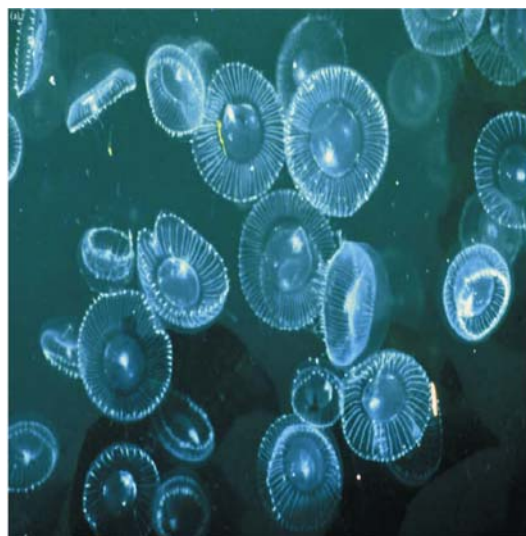


Figure 1. Jellyfish, *Aequorea victoria*. [Figure reprinted from ref. 2, with permission.]

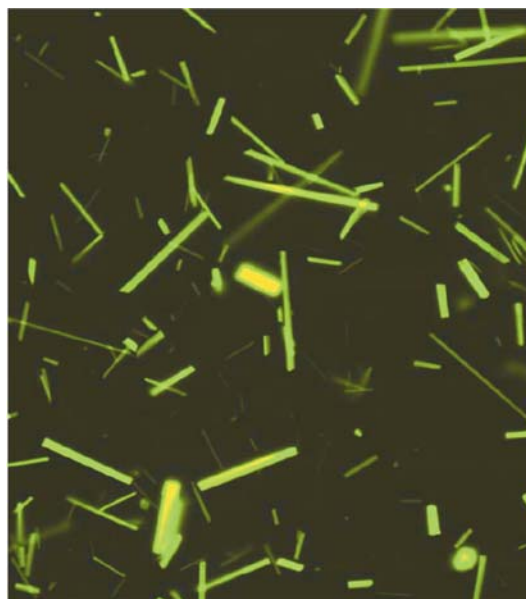


Figure 2. Green fluorescent protein. [Figure reprinted from ref. 3, with permission.]

*For correspondence. (e-mail: baban@nio.org)

Table 1. Properties of fluorescent proteins with published extinction coefficients and quantum yields, that appear bright enough to be useful in transgenic organisms. Most are untested in transgenics [reprinted from ref. 12, with permission]

Species of origin and designation(s)	Emission maximum (nm) and (extinction coefficient) (103 MK1 cm K1)	Excitation maximum (nm) and (% quantum yield)	Oligomerization (MZ monomer, DZ dimer, TZ tetramer)
<i>Aequorea victoria</i> GFP	395 (27)	504 (79)	M
<i>A. victoria</i> GFP S65T	489 (55)	510 (64)	M
<i>A. victoria</i> EGFP	488 (56)	508 (60)	M
<i>A. victoria</i> GFP Emerald	487 (58)	509 (68)	M
<i>A. victoria</i> GFP YFP Topaz	514 (94)	527 (60)	M
<i>A. victoria</i> GFP YFP Venus	515 (92)	528 (57)	M
<i>Zoanthus</i> sp. ZsGreen	497 (36)	506 (63)	M
<i>Zoanthus</i> sp. ZsYellow	528 (20)	538 (2042)	T
<i>Anemonia majano</i> AmCyan	458 (40)	486 (24)	T
<i>Heteractis crispa</i> t-HcRed1	590 (160)	637 (4)	T
<i>Discosoma</i> sp. DsRed	558 (75)	583 (79)	T
<i>Discosoma</i> sp. mRFP1	584 (50)	607 (25)	M
<i>Discosoma</i> sp. dimer2	552 (69)	579 (29)	D
<i>Discosoma</i> sp. mOrange	548 (71)	562 (69)	M
<i>Discosoma</i> sp. dTomato	554 (69)	581 (69)	D
<i>Discosoma</i> sp. tdTomato	554 (138)	581 (69)	D (tandem)
<i>Discosoma</i> sp. mStrawberry	574 (90)	596 (29)	M
<i>Discosoma</i> sp. mCherry	587 (72)	610 (22)	M
<i>Discosoma</i> sp. mPlum	590 (41)	649 (10)	M
<i>Galaxeidae</i> sp. Azumi Azami Green (AG)	492 (72)	505 (67)	T
<i>Galaxeidae</i> sp. mAG	492 (72)	505 (81)	M
<i>Cerianthus membranaceus</i> cmFP512	503 (59)	512 (66)	T
<i>Entacmaea quadricolour</i> eqFP611	559 (78)	611 (45)	T
<i>Trachyphyllia geoffroyi</i> Kaede	508 (99)	518 (80)	T
<i>Acropora</i> sp. MiCy	472 (27)	495 (90)	D
<i>Fungia concinna</i> KO	548 (110)	561 (45)	T
<i>Fungia concinna</i> mKO	548 (52)	559 (60)	M
<i>Phialidium</i> sp. phiYFP	525 (115)	537 (60)	D
<i>Pontellina plumata</i> pplu GFP2	482 (70)	502 (60)	M
<i>Pontella meadi</i> pmeaGFP1	489 (99)	504 (74)	Unknown
<i>Pontella meadi</i> pmeaGFP2	487 (98)	502 (72)	Unknown
Unidentified pdae1GFP	491 (105)	511 (68)	Unknown
<i>Lobophyllia hemprichii</i> EosFP	506 (72)	516 (70)	T
	571 (41)	581 (55)	
<i>Lobophyllia hemprichii</i> d2EosFP	506 (84)	516 (66)	D
	569 (33)	581 (60)	
<i>Lobophyllia hemprichii</i> mEosFP	505 (67)	516 (64)	M
	569 (37)	581 (66)	

Morin and Hastings termed it as GFP³. Though initially these compounds were thought to be present only in the bioluminescent species, they were later identified in many non-bioluminescent organisms, probably through the food-chain relationship². Ever since several members of this family have been identified and reported. This ever-expanding family has brought forth a biotechnological revolution, as it has become an essential tool in the production and analysis of transgenic organisms in basic and applied biology.

Occurrence

Fluorescent and non-fluorescent GFP-like proteins occur in a variety of marine organisms such as corals^{4,5} and jellyfish². Although the very first of this category has been discovered from a jellyfish, reef-building corals seem to

be offering a wide spectrum of these proteins ranging from blue and green to yellow, pink, orange and red. They were initially discovered in non-bioluminescent, zooxanthellate Anthozoa, including *Actinaria*, *Zoantharia*, *Corallimorphoria*, *Stolonifera* and *Scleractinia*, and were subsequently recognized as major colour determinants of hermatypic reef corals^{4,5}. GFPs are also present in azooxanthellate anthozoans. The reported members include proteins from sea anemone *Heteractis crispa*, giant anemone *Condylactis gigantia* and coral *Goniopera tenuidens* (which could be converted into far red fluorescent protein (λ_{max} , 615–640 nm) by an identical single amino acid substitution)⁶. Proteins that change colour from green to red upon irradiation with 400 nm light were reported from *Catalaphyllia jardinie* and *Lobophyllia hemprichii* (Eos FP) – these are localized mainly in ectodermal tissue and contribute up to 7.0% of the total solu-

ble cellular proteins in these species⁴. Green and cyan fluorescent proteins were identified from the tissues of the great star coral *Montastraea cavernosa*⁷. *Discosoma* coral was reported to contain a red-shifted protein (dsFP 593) with an emission maximum⁸ at 593 nm. A protein from *Trachyphyllia geoffroyi* was identified (kaede) which emits green, yellow and red light by irradiation with UV or violet light (350–400 nm)⁹. Sea anemone *Anemonia sulcata* contains a non-fluorescent protein, which converts to a green fluorescent form upon irradiation with green light¹⁰. A red-shifted fluorescent protein has been identified from *Actinia equina* (aeCP597)¹¹. Another protein (Rtms5), which is blue in colour, has been identified from the coral *Montipora efflorescens*⁵. Some more fluorescent proteins¹² are presented in Table 1. With this knowledge it became possible to find a natural protein with the desired properties for any biotechnological applications, rather than resorting to engineer a semi-synthetic protein of required characteristics.

Chemistry

The structure of GFP has been reported to have been solved using seleniomethionyl-substituted protein and multi-wavelength anomalous dispersion (MAD) phasing methods. The electron density maps produced by the MAD phasing are clear, revealing a dimer comprised of two quite regular barrels with 11 strands on the outside of cylinders. These cylinders have a diameter of about 30 Å

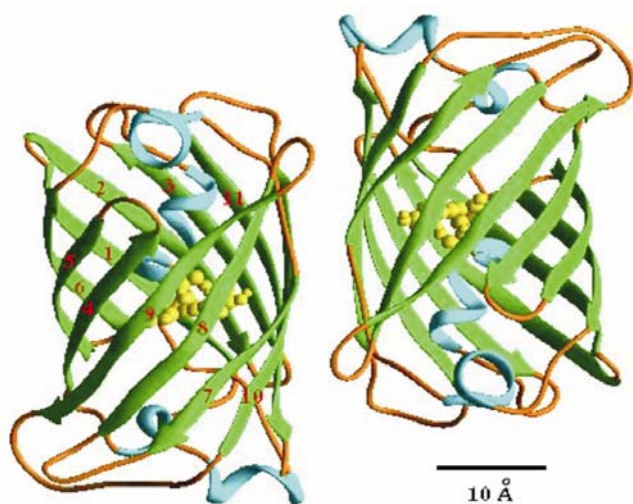


Figure 3. The overall shape of the protein and its association into dimers. Eleven strands of β -sheet (green) form the walls of a cylinder. Short segments of α -helices (blue) cap the top and bottom of the ‘ β -can’ and also provide a scaffold for the fluorophore, which is near the geometric centre of the can. This folding motif, with β -sheet outside and helix inside, represents a new class of proteins. Two monomers are associated into a dimer in the crystal and in solution at low ionic strengths. This view is directly down the twofold axis of the non-crystallographic symmetry. [Figure reprinted from ref. 13, with permission.]

and a length of about 40 Å. Inspection of the density within the cylinders revealed modified tyrosine side chains as part of an irregular helical segment. Small sections of α -helix also form caps on the end of the cylinders (Figures 3 and 4)¹³. This motif or folding arrangement, with a single α -helix inside a very uniform cylinder of β -sheet structure, represents a new protein class ($\alpha + \beta$), as it is not similar to any known protein structure. Two protomers pack closely together to form a dimer in the crystal. The protein is comprised of 238 amino acids and has a molecular weight of 26.9 kDa. Its wild type absorbance/excitation peak is at 395 nm with a minor peak at 475 nm and with extinction coefficients of roughly 30,000 and 7000 M⁻¹ cm⁻¹ respectively. The emission peak¹³ is at 508 nm.

Chromophore biosynthesis

The production of visual colour is related to the formation and maturation of a chromophore system. The part of the biomolecule responsible for the production of any sort of colour or luminescence is called a chromophore. It is a covalently unsaturated group (e.g. C=C, C=O, NO₂, etc.), responsible for absorption of energy from a radiant source and brings about electronic transitions within the molecule. When these transitions are of π - π^* or n - π^* type, the chromophore system that is formed, produces colour in the visible region of the electromagnetic spectrum. However, when these transitions are not stable and electrons drop back to the ground state from the excited state emitting energy in the form of radiation, then it is called fluorescence and the chromophore is then called the fluorophore¹⁴.

In GFPs the chromophore is generated only under conditions permissive of protein folding; that is, the polypeptide must be able to obtain its native three-dimensional structure to become visible fluorescent. It is generated in the presence of molecular oxygen. This autocatalytic



Figure 4. A topology diagram of the folding pattern in GFP. The β -sheet strands are shown in light green, α -helices in blue, and connecting loops in yellow. The positions in the sequence that begin and end each major secondary structure element are also given. The anti-parallel strands (except for the interactions between stands 1 and 6) make a tightly formed barrel. [Figure reprinted from ref. 13, with permission.]

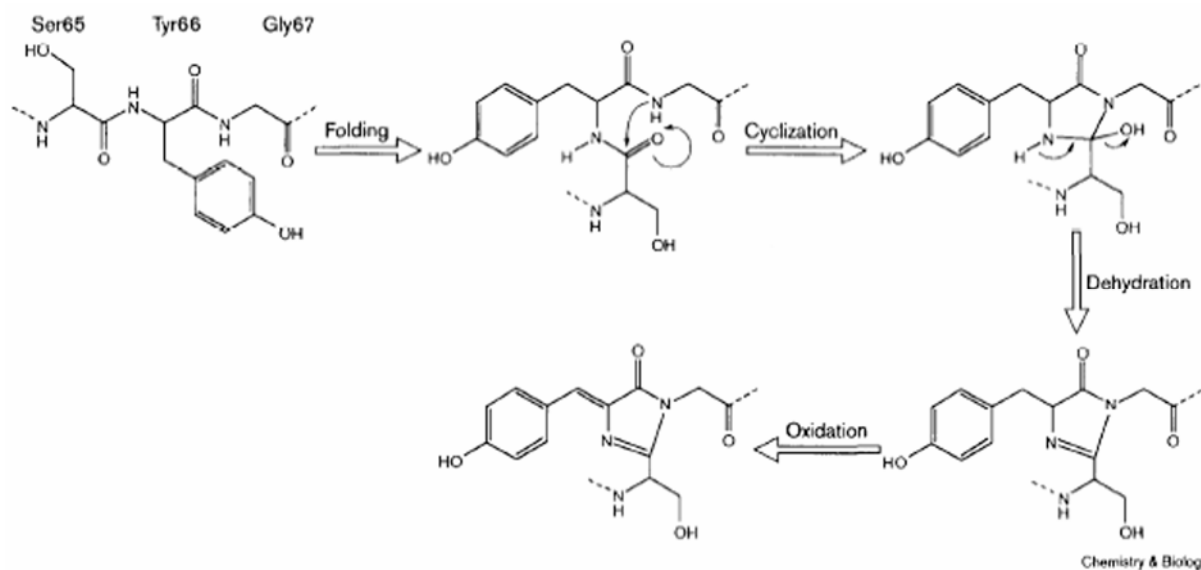


Figure 5. Fluorophore formation in GFP. Folding of proteins promotes cyclization, which is followed by dehydration of the ring and oxidation of tyrosine. [Figure reprinted from ref. 15, with permission.]

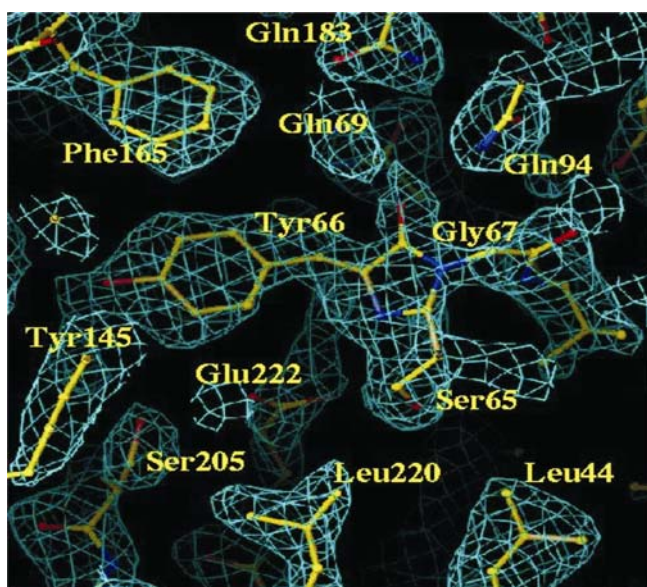


Figure 6. Model of fluorophore and its environment superposed on the MAD-phased electron density map at 2.2 Å resolution. The clear definition throughout the map allowed the chain to be traced and side chains to be well placed. The density for Ser65, Tyr66 and Gly67 is quite consistent with the dehydrotyrosine–imidazolidone structure proposed for the fluorophore. Many of the side chains adjacent to the fluorophore are labeled. [Figure reprinted from ref. 13, with permission.]

mechanism is initiated by an intrachain ring closure that leads to the formation of a cyclopentyl group from the backbone of the original peptide (Ser65, Tyr66 and Gly67). The fluorophore originates from an internal Ser–Tyr–Gly sequence which is post-translationally modified to a 4-(*p*-hydroxybenzylidene)-imidazolidin-5one struc-

ture (Figures 5 (ref. 15) and 6 (ref. 13)). Studies on recombinant GFP expression in *Escherichia coli* led to a proposed sequential mechanism initiated by a rapid cyclization between Ser65 and Gly67 to form a imidazolidin-5-one intermediate¹³, followed by a much slower (hours) rate-limiting oxygenation of the Tyr66 side chain by O₂. Combinatorial mutagenesis suggests that Gly67 is required for the formation of the fluorophore. While no known co-factors or enzymatic components are required for this apparently autocatalytic process, it is rather thermosensitive, with the yield of fluorescently active to total GFP protein decreasing at temperatures greater than 30°C. However, once produced, GFP is quite thermostable¹³.

Physical and chemical characteristics

GFP is resistant to denaturation requiring treatment with 6M guanidine hydrochloride at 90°C or pH <4.0 or >12.0. Partial to near-total renaturation occurs within minutes following reversal of denaturing conditions by dialysis or neutralization. Circular dichroism predicts significant amount of β -sheet structure that is subsequently lost on denaturation. Over a non-denaturing range of pH, increasing pH leads to a reduction in fluorescence by 395 nm excitation and an increased sensitivity to 475 nm excitation. Reduction of purified GFP by sodium dithionite results in a rapid loss of fluorescence that slowly recovers in the presence of room air. While insensitive to sulfhydryl reagents such as 2-mercaptoethanol, treatment with the sulfhydryl reagent dithiobisnitrobenzoic acid (DTNB) irreversibly eliminates fluorescence¹³.

The remarkable cylindrical fold of the protein seems ideally suited for its function. The strands of the β -sheet are tightly fitted to each other like staves in a barrel and form a regular pattern of hydrogen bonds. Together with the short helices and loops on the ends, the 'can' structure forms a single compact domain and does not have obvious clefts for easy access of diffusible ligands to the fluorophore. Photochemical damage by the formation of singlet oxygen through intersystem crossing is reduced by the structure. The tightly constructed β can would appear to serve this role well and also provide overall stability and resistance to unfolding by heat and denaturants¹³.

Physiological functions

The pigments containing these proteins display slow decay rates characterized by half-lives of ~20 days. The slow turnover of GFP-like proteins implies that the associated energetic costs for being colourful are comparatively low. Moreover, high *in vivo* stability makes GFP-like proteins suitable for functions requiring high pigment concentration, such as photoprotection¹⁶. The underlying mechanism, however, remains controversial, as some fluorescent proteins have spectral properties that appear to be inappropriate for photoprotecting tissue by modulating the intracellular light climate, while others are spectrally well suited to fluorescence energy transfer and dissipation of light energy via radiative and nonradiative pathways. An antioxidant function has recently been suggested¹⁶.

The relation between core amino acid composition and observed change in fluorescence in GFP homologues

The availability of *E. coli* clones expressing GFP has led to extensive mutational analysis of GFP function. It has been reported that removal of more than seven amino acids from the C-terminus or more than the N-terminal Met leads to total loss of fluorescence. All non-fluorescent mutants also failed to exhibit absorption spectra characteristic of the intact fluorophore, employing a possible defect in post-translational processing. Screens of random and directed point mutations for changes in fluorescent behaviour have uncovered a number of informative amino acid substitutions. Mutation of Tyr66, in the fluorophore to His, results in a shift of the excitation maximum to the UV (383 nm), with emission, in the blue at 448 nm. A Tyr 66Trp mutant is blue-shifted though to a lesser degree. Both changes are associated with a severe weakening of fluorescence intensity compared to wild-type GFP. Mutation of Ser65 to Thr, Ala, Cys or Leu causes a loss of the 395 nm excitation peak, with a major increase in blue excitation. When combined with Ser65 mutants, muta-

tions at other sites near the fluorophore such as Val68Leu and Ser72Ala can further enhance the intensity of green fluorescence produced by excitation at 488 nm. However, amino acid substitutions significantly outside this region also affect the spectral character of the protein. For example, Ser202Phe and Thr203Ile both cause loss of excitation in the 475 nm region, with preservation of 395 nm excitation. Ile167Thr results in a reversed ratio of 395 to 475 nm sensitivity, while Glu222Gly is associated with the elimination of only the 395 nm excitation. Another change, Val163Arg, not only enhances the magnitude of the Ser65Thr mutant, but also increases the temperature tolerance for functional GFP expression¹³.

The location of certain amino acid side chains in the vicinity of the fluorophore also explains the fluorescence and behaviour of certain mutants of the protein. At least two resonant forms of the fluorophore can be drawn, (Figure 7), one with a partial negative charge on the benzyl oxygen of Tyr66, and another with the charge on the carbonyl oxygen of the imidazolidine ring. Interestingly, basic residues appear to form hydrogen bonds with each of these oxygen atoms, His148 with Tyr66, and Gln94 and Arg96 with the imidazolidone. These bases presumably act to stabilize and possibly further delocalize the charge on the fluorophore. Most of the other polar residues in the pocket form an apparent hydrogen-bonding network on the side of Tyr66 that requires abstraction of the protons in the oxidation process¹³.

As for the mutants, atoms in the side chains of Thr203, Glu222 and Ile167 are in Van der Waals contact with Tyr66, so that their mutation would have direct steric effects on the fluorophore and would also change its electrostatic environment if the charge were changed¹². Molecular mechanics simulations also indicate that amino acid residues are placed around the chromophore in such a way as to favour hula-twist motions of the chromophore because of its volume-conserving nature¹⁰.

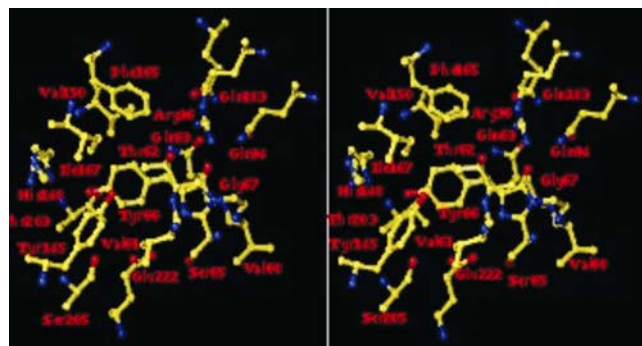


Figure 7. Stereo view of the fluorophore and its environment. His148, Gln94 and Arg96 can be seen on opposite ends of the fluorophore and probably stabilize its resonant forms. Charged, polar and non-polar side chains all contact the fluorophore in some way. [Figure reprinted from ref. 13, with permission.]

Regardless of final colour, a GFP-like chromophore appears to be an intermediate state in all chromoprotein homologues to GFP that have been studied so far. Clearly a chemical entity similar to that present in the jellyfish-derived GFP must be synthesized first, to generate the broad spectrum of chromophores¹⁰.

Applications

In vivo imaging

GFP was discovered more than 40 years ago, but it was not until 1994, after cloning and successful heterologous expression of the *gfp* gene (GenBank accession no. U17997), that this protein attracted attention¹⁷. GFP and its mutants soon became popular tools for cell and molecular biology. During the past few years, the great spectral and phylogenetic diversity of GFP-like proteins has been characterized in marine organisms. Furthermore, several useful mutant variants of fluorescent proteins have been generated. Consequently, a panel of fluorescent proteins is now reported that covers almost the whole visible spectrum, each possessing different biochemical characteristics¹⁷. The development of various sophisticated fluorescent proteins such as Photoactivable fluorescent proteins (PA-GFP, Kaede, EOS FP, PS-CFP, Dronpa, Pa-mRFP1 and KkGR, Dendra), Timer, a series of fluorescent sensors and split GFPs has opened up novel applications for *in vivo* fluorescent labelling such as studies of protein expression, interaction, activity, movement, and turnover, direct measurement of cell parameters and state organelle function, and cell motility studies¹⁷⁻¹⁹.

Among the most promising of fluorescent proteins are the PA-FPs. Invisible at the imaging wavelength until activated by irradiation at a different wavelength, PA-FPs allow controlled highlighting of distinct molecular populations within the cell²⁰. Among these PA-FPs, Dendra, derived from *Octocoral dendronephthya* species represents the first protein, which is simultaneously monomeric, efficiently matures at 37°C, demonstrates high photostability of the activated state, and can be photoactivated by a common, marginally phototoxic 488 nm laser line. It is capable of 1000 to 4500-fold photo conversion from green to red fluorescent states in response to either visible, blue or UV-Violet light¹⁸. Using photo conversion offers a strategy to precisely and non-invasively monitor protein degradation in the cell. Cells expressing one of these photoactivable proteins are irradiated with activating light at a specific time and subsequently the converted form can be monitored independent of the cellular pool of newly synthesized original proteins. Consequently, degradation of the photoconverted molecules can be followed via the decay in the fluorescence-converted form¹⁶.

Aequorea victoria GFP gene was first cloned and transformed into the nematode *Caenorhabditis elegans*. Since then fluorescent proteins have become essential tools in the production and analysis of transgenic organisms in basic and applied biology. Model organisms were the first to be transformed with GFP. After *C. elegans* came the fruitfly *Drosophila melanogaster*, followed by mammalian cell lines and yeast. Larger organisms such as plants, mouse, zebrafish and the frog *Xenopus laevis* were transformed with GFP variants yielding visible green fluorescence. The most successful commercial ornamental transgenic organism to date is 'GloFish', marketed by Yorktown technologies (Figure 8)¹².

Fluorescent proteins are neutral markers. They have properties which indicate that it is a neutral, non-toxic, universal marker when expressed in transgenic organisms. Fluorescent proteins do not confer any measurable ecological host costs when expressed in transgenic plants. They also appear to be neutral with regard to oral toxicity and allergenicity¹².

When a living cell is seen under a light microscope, only the most morphologically profound cellular events are visible. Nevertheless, underlying the events of major morphological reorganization is a complex, unseen bustle of chemical events. Cell biologists have often taken advantage of the sensitivity and versatility of fluorescent probes to explore the chemical details of the intracellular environment, using such probes as tags and positional markers to measure distance and intracellular transport and as sensors of intracellular concentrations of small molecules and ions, so that protein distribution and environment can be visualized. Chemical labelling of a purified protein *in vitro* does not lend itself to visualization of the unperturbed intracellular milieu. Introduction of fluorescently labelled antibodies raised against a protein of interest is feasible only in special cases, and antibody binding can interfere with the function of the protein. In contrast, intrinsically fluorescent proteins have found



Figure 8. Glo Fish™ which contains a red fluorescent protein (RFP) [Figure reprinted from ref. 12, with permission.]. Courtesy: glofish.com.

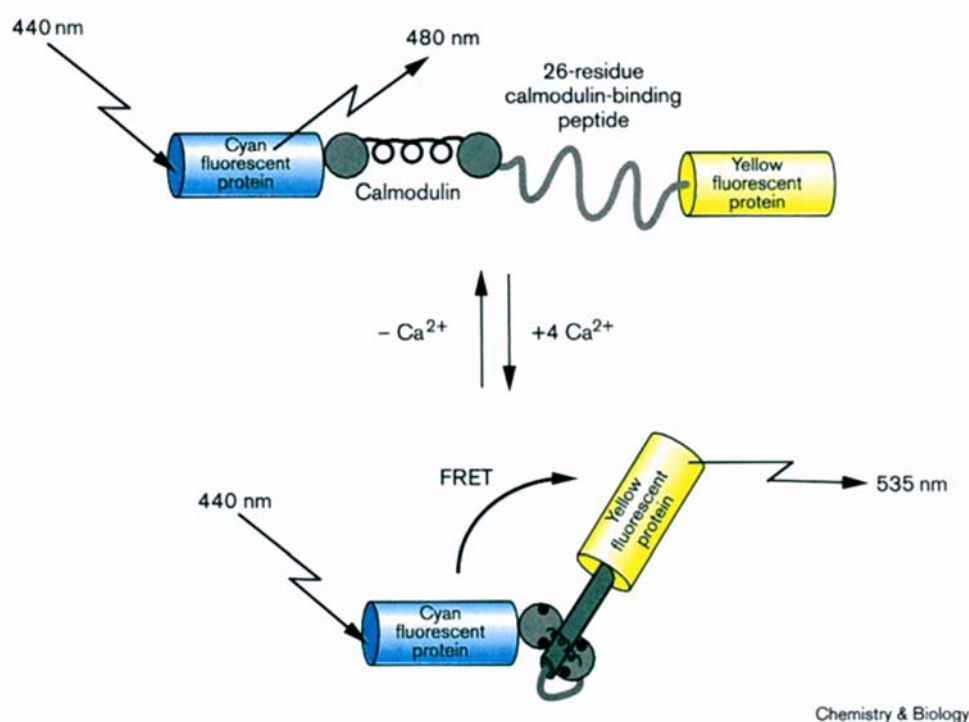


Figure 9. Design of fluorescence indicators for Ca^{2+} using fluorescent proteins and calmodulin. Variants of GFP are appended to the termini of a fusion of calmodulin with a 26-residue calmodulin-binding peptide. Upon association with Ca^{2+} , calmodulin wraps around the 26-residue peptide, bringing the two fluorescent proteins closer to each other and enhancing FRET between them. [Figure reprinted from ref. 15, with permission.]

wide applicability as genetically encodable fluorescent labels. By encoding a label into the primary sequence of a protein of interest at the level of DNA, that protein can be specifically observed in the context of the cell itself. Disruptive post-translational manipulation of the protein is avoided. Hence, fluorescent proteins are considered as reporters of gene expression and protein localization^{15,21}.

The interaction of two fluorescently labelled proteins can be monitored using fluorescence resonance energy transfer (FRET). FRET occurs between two fluorophores incorporated into two proteins that are associated with each other and diminishes precipitously upon their dissociation. Taking advantage of this phenomenon, it is applied to sensing ligand-dependent conformational changes of proteins to monitor intracellular concentrations of Ca^{2+} ; visualizing exocytosis and synaptic transmission with pH-sensitive fluorescent proteins and covalent labelling of recombinant proteins in living cells using a small molecule^{15,22} (Figure 9).

This principle is also applied in the Cameleon. The Cameleon, a FRET-based nanosensor, uses the ability of the calcium-bound form of calmodulin to interact with calmodulin-binding polypeptides to turn the corresponding dramatic conformational change in resonance energy transfer between two fluorescent proteins. This metabolite nanosensor consists of two variants of the GFP

fused to bacterial periplasmic-binding proteins. Different from the Cameleon, a conformational change in the binding properties is directly detected as a change in FRET efficiency²³.

Photosensitizers are chromophores that generate reactive oxygen species (ROS) upon light irradiation. They are used for inactivation of specific proteins by chromophore-assisted light inactivation (CALI) and for light-induced cell killing in photodynamic therapy. Whereas known photosensitizers must be added to living systems exogenously, one application reported a genetically encoded photosensitizer called Killer Red, developed from the hydrozoan chromoprotein anm2CP, a homologue of the GFP. Killer Red generates ROS upon irradiation with green light. It has been utilized for light-induced killing of *E. coli* and eukaryotic cells, and for inactivating fusions to β -galactosidase and phospholipaseC delta1 Pleckstrin homology domain²⁴.

Misfolded proteins, aggregates and inclusion bodies are hallmarks of the cytopathology of neurodegenerative disorders, including Huntington's disease, Amyotrophic lateral sclerosis, Parkinson's disease, Prion diseases and Alzheimer's disease. The appearance of proteins with altered folded states is regulated by the protein folding quality control machinery and is age-dependent. In one experiment, using fluorescent proteins an unexpected

molecular link was identified between the metabolic state, accumulation of damaged proteins, the heat-shock response and chaperones, and longevity. Nearly 200 genes were identified defining a 'protein quality control proteome'²⁵.

Mainly two drawbacks have been revealed in the use of novel fluorescent proteins, which include formation of tetramers (tetramerization) and high molecular weight aggregates (aggregation). Though oligomerization does not limit the use of fluorescent proteins as markers of gene expression, aggregation may impede all possible applications due to cellular toxicity. Elimination of basic residues located near the N-termini of fluorescent proteins results in the generation of non-aggregating versions of several fluorescent proteins²⁶, specifically drFP538 (DsRed), DsRed-Timer, ds/drFP616, zFP506, zFP538, amFP486 and asFP595.

Plant biology

Study of plant cell biology has benefitted tremendously from the use of fluorescent proteins. Development of well-established techniques in genetics, by transient expression or by *Agrobacterium*-mediated plant cell transformation, makes it possible to readily create material for imaging molecules tagged with fluorescent proteins. Confocal microscopy of fluorescent proteins is routine and in highly scattering tissues, multiphoton microscopy improves deep imaging. The abundance of autofluorescent compounds in plants in some cases potentially interferes with fluorescent protein signals, but spectral imaging is an effective tool in unmixing overlapping signals. This approach allows separate detection of DsRed and chlorophyll, DsRed and GFP, GFP and YFP²⁷.

Fluorescent proteins have been targeted to most plant organelles. Free (untargeted) fluorescent proteins in plant cells are not only cytoplasmic, but also go into the nucleus due to their small size. Fluorescent protein fluorescence is potentially unstable in acidic vacuoles in seeds. Endoplasmic reticulum (ER)-targeted GFP has identified novel inclusion bodies that are surprisingly dynamic. Fluorescent protein-tagged Rab GTPases have allowed documentation of the dynamics of membrane trafficking. Investigation of virus infection has progressed significantly with the aid of fluorescent protein-tagged virus proteins²⁷.

Fluorescent proteins can provide instantaneous data on homo or heterozygosity, and determine transgene zygosity of dominantly or semi-dominantly expressed traits. In one study, heterozygous (hemizygous) transgenic canola plants exhibited half the green fluorescence of homozygous plants. Furthermore, when GFP homozygous canola plants were hybridized with non-transgenic wild relatives, the progeny, which were heterozygous parent crop, indicated that fluorescent proteins might be a tool in analysing hybridizations and introgression status¹².

Evolution

Phylogenetic analysis

Recreating ancestral proteins is being used increasingly in the laboratory to study the evolutionary history of protein function. More efficient gene synthesis techniques and the decreasing costs of commercial oligosynthesis are making this approach both simpler and less expensive to perform. Developments in ancestral reconstruction methods, particularly more realistic likelihood models of molecular evolution, allow for the accurate reconstruction of more ancient proteins than previously possible. Recently, in a review of methods for reconstruction of ancient proteins, GFP and GFP-like proteins were employed as an example²⁸, for reconstructing ancient protein history.

Natural pigments are normally products of complex biosynthetic pathways where many different enzymes are involved. GFP-like proteins containing organisms represent the only exception: in these organisms, each host colour is essentially determined by a sequence of a single protein homologous to the GFP. This direct sequence colour linkage provides unique opportunity for colour evolution studies⁷.

It has been reported that the bilaterian animals also harbour an unmistakable GFP-like domain that comprises part of the G2F fragment ('globular fragment two') of extracellular matrix proteins called nidogens (entactins) and fibulins. This GFP-like domain of the G2F fragment (called G2FP domain) has clearly a homologous fold, but is less than 10% identical to GFPs by amino acid sequence. G2FP domains are found in all bilaterian genomes reported thus far and are neither coloured nor fluorescent. Instead they serve as protein-binding modules that participate in the control of the extracellular matrix formation during development^{29,30}.

It has been suggested that the group of structural homologues of GFP which share the GFP-like 'beta-can' fold should be regarded as a superfamily following the criteria proposed by the Protein Information Resource. The main reason for such a classification is that this group unites at least two clearly definable protein families. The first one consists of G2FP domains, which are incapable of autocatalytic chromophore syntheses and are found within multi-domain proteins of the extracellular matrix. The second one includes fluorescent and/or coloured proteins capable of synthesizing the chromophore autocatalytically and which are not found in a multi-domain context (Figure 10)^{29,30}.

The evolutionary route of the GFP superfamily could be imagined as two separate lineages, whereby one, in Cnidaria, gained fluorescent properties, while another, in Bilateria, specialized in protein binding and became part of extracellular proteins. It is also possible, however, that these two lineages originated as a result of an ancient gene duplication preceding the separation of Cnidaria and

Bilateria, so that both G2FPs and fluorescent GFPs could be found within a single genome, both in Cnidaria and Bilateria^{29,30}.

One of the basic issues regarding the evolution of anthozoan colours that can be addressed at the present

state of knowledge is the basis of intraspecific colour variation. Many coral species exhibit these variations ranging from fluorescent blue to fluorescent red and non-fluorescent purple. Two alternative mechanisms of generating diversity of colouration are possible, which are

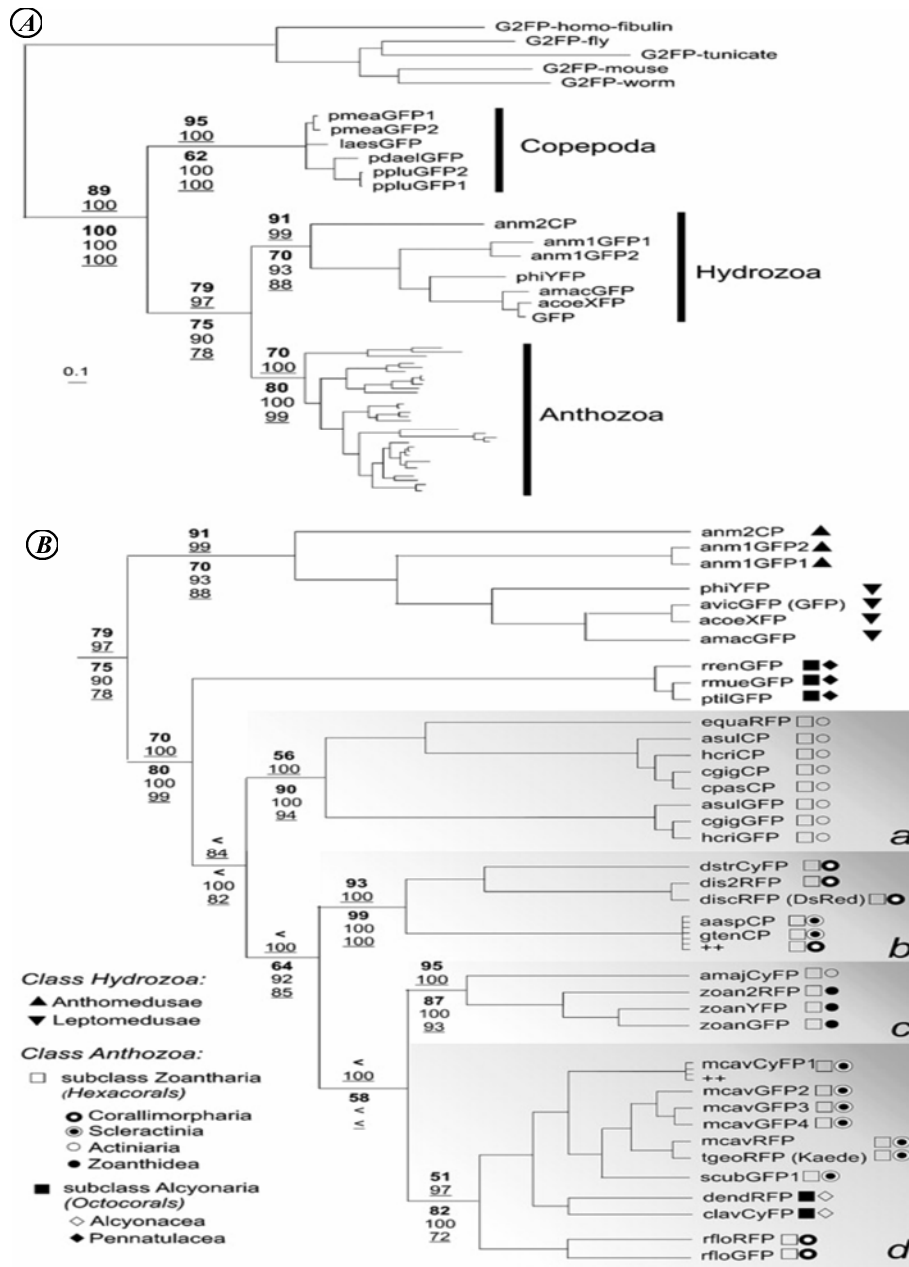


Figure 10. **A**, Phylogenetic tree of GFP-like proteins. The clade of Anthozoa proteins is represented schematically for the purpose of illustrating the characteristic branch lengths. Alignment of remotely related sequences, such as of G2FP domains and fluorescent proteins, was done following the superimposition of known 3D protein structures. The values above the branches are Tree-Puzzle support indexes obtained after analysis of protein sequences (bold) and the Bayesian analysis of cDNA (underlined). Below the branches are support values obtained by non-parametric bootstrap analysis of cDNA under the following criteria: maximum likelihood (bold), minimum evolution (normal), and maximum parsimony (underlined). Scale bar: 0.1 replacements/nucleotide. **B**, Phylogenetic tree topology of CnidarianGFP-like proteins. Support values of the four major clades (a–d) of Anthozoa proteins and problematic deep internal branches follow the legend to panel A, '<' indicating less than 50% support. All other branches are supported at better than 50% level by all DNA-based methods. Groups of very similar proteins are depicted as terminal polytomies, with only one or two representative protein names shown. Systematic position of the host is denoted. The colour class of a protein is encoded within its name: GFP, Green; YFP, Yellow; RFP, Red; CyFP, Cyan; CP, Chromoprotein³⁰. [Figure reproduced from ref. 30, with permission.]

usually called polymorphism and polyphenism. Polymorphism is the situation when differences in the genome exist, in which case the colour appearance of an organism is largely determined at the moment of zygote formation and there is little possibility for it to change afterwards, except for the intensity. This explanation of colour diversity in corals seems particularly attractive, considering the relative ease of certain types of colour conversion by random mutations. In contrast, polyphenism is recognized when the same genome gives rise to different phenotypes. In this case, the colour can be a much more flexible character during the lifetime of the organism than is allowed under the polymorphism model. The colour diversity may also be a result of a combination of these two models, in which case both allelic polymorphism and expression level variations would play significant roles⁷.

It has been finally concluded that different chromophore structures and alternative products synthesized within a similar autocatalytic environment and the phylogenetic pattern and colour diversity in reef Anthozoa are a result of a balance between selection for GFP-like proteins of particular colours and mutation pressure driving the colour conversions⁷.

Indian perspective

In India, several aspects of GFPs have been explored over the recent years. Some of the developments are the following: to begin with, a better and simpler technique of purification of GFP has been developed by the Department of Chemistry, Indian Institute of Technology, Delhi. This technique is called Three Phase Partitioning (TPP), as it is a simple, one-step non-chromatographic method compared to the other purification procedures involving at least two steps of column chromatography³¹. The Department of Microbiology, CCS Haryana Agricultural University, Hisar, has experimented on the construction of GFP-marked strains of *Bradyrhizobium* for ecological studies. The GFP-marked *Bradyrhizobium* was successfully identified on the basis of green fluorescence to study its competition and survival in the soil and in charcoal-based inoculants³². The Central Drug Research Institute (CDRI), Lucknow, in coordination with the Institute Pasteur de Cille, Lill Cedex, France, has conducted an investigation to develop an *in vivo* expression system based on GFP and kanamycin resistance to identify macrophage-specific *Mycobacterium tuberculosis* genes, which appear to be up-regulated in infected macrophages. The experiment confirmed the up-regulation of these genes and the level of up-regulation varied from 2 to 40 fold in macrophages compared to growth *in vitro*³³. The Division of Biochemistry and Division of Parasitology, CDRI, Lucknow has improved and simplified the screening procedure for monitoring *Leishmania* susceptibility to drugs using flow cytometry. The procedure introduced a

GFP marker into field strains of *Leishmania* causing Kala-azar (visceral leishmaniasis) and explored the suitability of transgenic *L. donovani* promastigotes that constitutively express GFP in their cytoplasm as target cells for *in vitro* screening of anti-leishmanial drugs³⁴. GFP has been put to use in nanotechnology as well. The Department of Chemistry and Centre for Nanotechnology, IIT Guwahati, has developed a model system for investigating the antimicrobial activities of silver nanoparticles using GFP-expressing *E. coli*³⁵. In another finding from the same department, an experiment has been performed using GFP for *in situ* synthesis of highly uniform gold nanoparticles and monitoring protein denaturation³⁶.

Summary

GFP has revolutionized research in medicine and biology, enabling scientists to get a visual fix on how organs function, spread of disease and response of infected cells to treatment. Fluorescent proteins have linked many fields of biosciences – genomics, proteomics, cell biology, microarray biochip technology, flow cytometry, fluorescence microscopy, immunoassays, fluorescent *in situ* hybridization (FISH) and imaging technologies. India started research with known fluorescent proteins. More experimentation has to be done to find proteins from native sources, so that the economical factor is also taken into account. Nanotechnology needs to be conjugated with these proteins for better prospects. The time is now right for increased multidisciplinary collaboration between biochemists, engineers and biotechnologists to tailor proteins and instruments, in tandem, for monitoring applications. Time to go with the glow.

1. Nelson, D. L. and Cox, M. M., The three dimensional structure of proteins. In *Lehninger Principles of Biochemistry*, W.H. Freeman & Co, New York, 2005, 4th edn, pp. 141–144.
2. Shimomura, O., The discovery of aequorin and green fluorescent protein. *J. Microbiol.*, 2005, **217**, 3–15.
3. Morin, J. G. and Hastings, J. W., Biochemistry of bioluminescence of colonial hydroids and other coelenterates. *J. Cell. Physiol.*, 1971, **77**, 305–311.
4. Oswald, F. *et al.*, Contribution of host and symbiont pigments to the coloration of reef corals. *FEBS J.*, 2007, **274**, 1102–1109.
5. Prescott, M. *et al.*, The 2.2 Å crystal structure of a pocilloporin pigment reveals a nonplanar chromophore conformation. *Structure*, 2003, **11**, 275–284.
6. Gurskaya, N. G. *et al.*, GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Lett.*, 2001, **507**, 16–20.
7. Kelmanson, I. V. and Matz, M. V., Molecular basis and evolutionary origins of color diversity in Great Star Coral *Montastraea cavernosa* (Scleractinia: Faviida). *Mol. Biol. Evol.*, 2003, **20**, 1125–1133.
8. Fradkov, A. F., Chen, Y., Ding, L., Barsova, E. V., Matz, M. V. and Lukyanov, S. A., Novel fluorescent protein from *Discosoma* coral and its mutants possesses a unique far-red fluorescence. *FEBS Lett.*, 2000, **479**, 127–130.
9. Hama, A. R. H., Hino, Y. M., Mizuno, H. and Miyawaki, A., An optical marker based on the UV-induced green-to-red photo-

- conversion of a fluorescent protein. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 12651–12656.
10. Watcher, R. M., Symposium in print: green fluorescent proteins and homologs. *Photochem. Photobiol.*, 2006, **82**, 339–344.
 11. Shkrob, M. A. *et al.*, Far-red fluorescent proteins evolved from a blue chromoprotein from *Actinia equina*. *Biochem. J.*, 2005, **392**, 649–654.
 12. Stewart Jr, C. N., Go with the glow: fluorescent proteins to light transgenic organisms. *Trends Biotechnol.*, 2006, **24**, 155–162.
 13. Yang, F., Moss, L. G. and Philips Jr, G. N., The molecular structure of green fluorescent protein. *Nature Biotechnol.*, 1996, **14**, 1246–1251.
 14. Silverstein, R. M., Bassler, G. C. and Morill, T. C., Ultra violet spectroscopy. In *Spectrometric Identification of Organic Compounds*, John Wiley, Singapore, 1991, 5th edn, pp. 289–294.
 15. Leubke, K. J., A flash of insight into cellular chemistry: genetically encoded labels for protein visualization *in vivo*. *Chem. Biol.*, 1998, **5**, 317–322.
 16. Verkhusha, V. V., Chudakov, D. M., Gurskaya, N. G., Lukyanov, S. and Lukyanov, K. A., Common pathway for red chromophore formation in fluorescent proteins and chromoproteins. *Chem. Biol.*, 2004, **11**, 845–854.
 17. Chudakov, D. M. *et al.*, Fluorescent proteins as a tool kit for *in vivo* imaging. *Trends Biotechnol.*, 2005, **23**, 605–613.
 18. Gurskaya, N. G. *et al.*, Engineering of a monomeric green-to-red photoactivable fluorescent protein induced by blue light. *Nature Biotechnol.*, 2006, **24**, 461–465.
 19. Neinhuis, G. U. *et al.*, Symposium in print: green fluorescent protein and homologs. *Photochem. Photobiol.*, 2006, **82**, 351–358.
 20. Lippincott-Schwartz, J. and Patterson, G. H., Fluorescent proteins for photoactivation experiments. *Methods Cell Biol.*, 2008, **85**, 45–61.
 21. Petrou, S., Bowser, D. N., Nicholls, R. A., Panchal, R. G., Smart, M. L., Reilly, A. M. and Williams, D. A., Genetically targeted calcium sensors enhance the study of organelle function in living cells. *Clin. Exp. Pharmacol. P.*, 2000, **27**, 738–744.
 22. Jakobs, S., High resolution imaging of live mitochondria. *Biochim. Biophys. Acta*, 2006, **1763**, 561–575.
 23. Fehr, M. *et al.*, Development and use of fluorescent nanosensors for metabolite imaging in living cells. *Biochem. Soc. T.*, 2005, **33**, 287–290.
 24. Bulina, M. E. *et al.*, A genetically encoded photosensitizer. *Nature Biotechnol.*, 2006, **24**, 95–99.
 25. Morimoto, R. I., Morley, J., Brignull, H., Garcia, S., Ritcher, K. and Kaiser, C., Stress and misfolded proteins: modulators of neurodegenerative diseases and longevity. *FEBS J.*, 2005, **272**, s1.
 26. Yanushevich, Y. G. *et al.*, A strategy for generation of non-aggregating mutants of Anthozoa fluorescent proteins. *FEBS Lett.*, 2002, **511**, 11–14.
 27. Berg, R. H. and Beachy, R. N., Fluorescent protein application in plants. *Methods Cell Biol.*, 2008, **85**, 153–177.
 28. Chang, B. S. W., Ugalde, J. A. and Matz, M. V., Application of ancestral protein reconstruction in understanding protein function: GFP-like proteins. *Methods Enzymol.*, 2005, **395**, 652–670.
 29. Labas, Y. A., Gurskaya, N. G., Yanushevich, Y. G., Fradkov, A. F., Lukyanov, K. A., Lukyanov, S. A. and Matz, M. V., Diversity and evolution of green fluorescent protein family. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 4256–4261.
 30. Shagin, D. A. *et al.*, GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Mol. Biol. Evol.*, 2004, **21**, 841–850.
 31. Sulakshana, J., Rajni, S. and Gupta, M. N., Purification of recombinant green fluorescent protein by three-phase partitioning. *J. Chromatogr. A*, 2004, **1035**, 83–86.
 32. Bhatia, R., Dogra, R. C. and Sharma, P. K., Construction of green fluorescent protein (GFP)-marked strains of *Bradyrhizobium* for ecological studies. *J. Appl. Microbiol.*, 2002, **93**, 835–839.
 33. Srivastava, V., Rouanet, C., Srivastava, R., Ramalingam, B., Locht, C. and Srivastava, B. S., Macrophage-specific *Mycobacterium tuberculosis* genes: identification by green fluorescent protein and kanamycin resistance selection. *Micro*, 2007, **153**, 659–666.
 34. Singh, N. and Dube, A., Short report: Fluorescent *Leishmania*: application to anti-leishmanial drug testing. *Am. J. Trop. Med. Hyg.*, 2004, **71**, 400–402.
 35. Sanpui, P., Pandey, S. B., Ghosh, S. S. and Chattopadhyay, A., Green fluorescent protein for *in situ* synthesis of highly uniform Au nanoparticles and monitoring protein denaturation. *J. Colloid Interface Sci.*, 2008, **326**, 129–137.
 36. Gogoi, K. S., Gopinath, P., Paul, A., Ramesh, A., Ghosh, S. S. and Chattopadhyay, A., Green fluorescent protein-expressing *Escherichia coli* as a model system for investigating the antimicrobial activities of silver nanoparticles. *Langmuir*, 2006, **22**, 9322–9328.

ACKNOWLEDGEMENTS. We thank the Director, National Institute of Oceanography, Goa for providing the necessary facilities. This is the contribution no. 4555 of NIO (CSIR), Goa.

Received 12 January 2009; revised accepted 1 June 2009