

8 October 2008



KUNGL.  
VETENSKAPSAKADEMIEN  
THE ROYAL SWEDISH ACADEMY OF SCIENCES



Scientific Background on the Nobel Prize in Chemistry 2008

# The green fluorescent protein: discovery, expression and development

## Background

During the 20<sup>th</sup> century the foundations of biochemistry were laid and used to explore the basal principles of the anabolic and catabolic pathways inside living cells. The 20<sup>th</sup> century also witnessed a revolution in our understanding of enzyme function and, through crystallography and nuclear magnetic resonance, of protein structures revealed at atomic resolution. During the second half of that century, classical genetics and nucleic-acid chemistry merged into modern genomics, based on whole-genome sequencing of an ever increasing number of organisms. This genetics revolution, supported by bioinformatics and other auxiliary techniques, has had an impact in many areas of the biological sciences, with practical consequences for medicine, pharmacy and ecology. However, neither the biochemical nor the genetics revolution provided the experimental tools that would allow for quantitative and experimentally well-defined monitoring at the molecular level of the spatio-temporal intra- and inter-cellular processes that define the dynamic behaviour of all living systems. To obtain such knowledge, new experimental and conceptual tools were required. Now, at the beginning of the 21st century, we are witnessing the rapid development of such tools based on the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria*, its siblings from other organisms and engineered variants of members of the “GFP family” of proteins.

These GFP-like proteins allow the monitoring in time and space of an ever-increasing number of phenomena in living cells and organisms like gene expression, protein localization and dynamics, protein-protein interactions, cell division, chromosome replication and organization, intracellular transport pathways, organelle inheritance and biogenesis, to name but a few. In addition, the fluorescence from single GFP molecules has made it feasible to image at a spatial resolution higher than the diffraction limit. Furthermore, sensors that report pH values, Ca<sup>2+</sup> concentrations and other essential features of the interior of living cells have been engineered from GFP-like proteins.

The technical revolution resulting from the discovery of GFP relates to a miraculous property of the chromophore that is responsible for its fluorescence. This chromophore is formed spontaneously from a tri-peptide motif in the primary structure of GFP, so that its fluorescence is “automatically” turned on in every organism where it is expressed. In other words, the maturation of the tri-peptide-based chromophore in GFP only requires oxygen and does not depend on the presence of enzymes or other auxiliary factors. GFP and its related variants thereby provide universal genetic tags that can be used to visualize a virtually unlimited number of spatio-temporal processes in virtually all living systems. This GFP revolution in the biological sciences has been greatly accelerated by a rapid parallel development of quantitative light microscopy, electronics, computational power and molecular modelling of intra- and inter-cellular processes with systems-biology approaches.

Outlined below are: (i) basal properties of GFP from *Aequorea victoria*; (ii) basal properties of its native relatives from other organisms and engineered variants of members of the GFP family; (iii) scientific areas in which GFP-like proteins have had a great impact. Then follow three sections of direct relevance for the choice of the Nobel Laureates in 2008: (iv) the discovery of GFP, (v) the demonstration that GFP can fluoresce when expressed in organisms other than *Aequorea victoria*, and (vi) the design of variants of members of the GFP family to create a universal tool-box for the monitoring of spatio-temporal processes in living systems.

### Properties of the GFP protein and its fluorescence chromophore

The native green fluorescent protein (GFP), first so named by Morin and Hastings (1971 ab), from the jellyfish *Aequorea victoria* (Shimomura et al., 1962) contains 238 amino acids (Prasher et al., 1992). Residues 65-67 (Ser-Tyr-Gly) in the GFP sequence spontaneously (Heim et al., 1994) form the fluorescent chromophore p-hydroxybenzylideneimidazolinone (Cody et al., 1993; Shimomura, 1979) (Fig. 1).

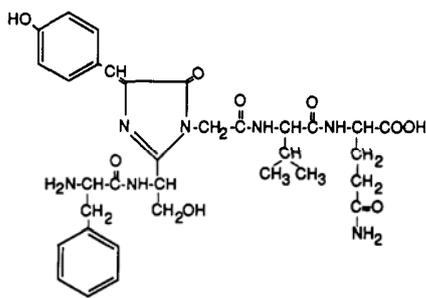


Figure 1. The fluorescence chromophore formed by amino acid residues 65-67 (Ser-Tyr-Gly) in the primary structure of GFP (From Cody et al., 1993).

The excitation spectrum of GFP fluorescence has a dominant maximum at about 400 nm and a significantly smaller maximum at about 470 nm, while the emission spectrum has a sharp maximum at about 505 nm and a shoulder around 540 nm (Tsien, 1998) (Fig. 2).

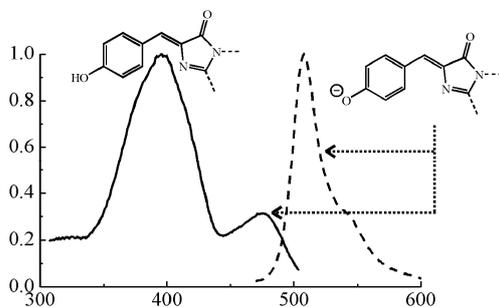


Figure 2. Fluorescence excitation (full-line curve) and emission (dashed curve) spectra of native GFP from *Aequorea victoria* (Tsien et al., 1998).

The crystal structure of GFP (Ormö et al., 1996; Yang et al., 1996) is an eleven-stranded  $\beta$ -barrel, threaded by an  $\alpha$ -helix, running up along the axis of the cylinder (Fig. 3). The chromophore (Fig. 1) is in the  $\alpha$ -helix, very close to the centre of the can-like cylinder (Fig. 3). A very large part of the primary structure of the protein is used to construct the  $\beta$ -barrel and the threading  $\alpha$ -helix. The N-terminal residue and the C-terminal residues 230-238, approximately corresponding to the maximal numbers of residues that can be removed from the N- (2 residues) and C-terminal (6 residues) respectively of GFP at retained fluorescence, are disordered and therefore unresolved in this structural image.

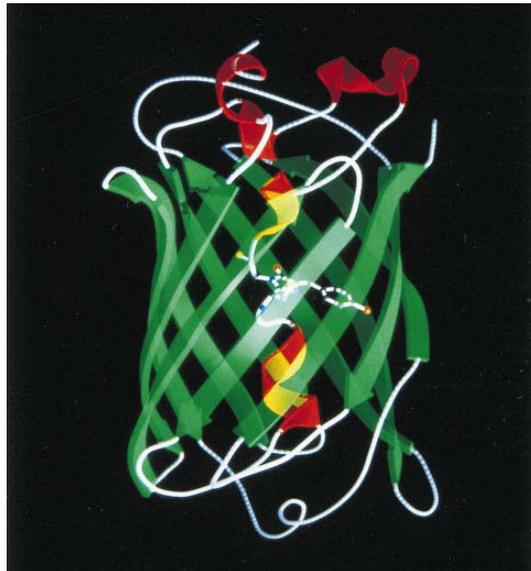


Figure 3. The tertiary structure of GFP, displaying its can-like shape with the  $\alpha$ -helix, containing the chromophore, threading up through the can (Brejc et al., 1997).

The tripeptide motif Ser65-Tyr66-Gly67- in the primary structure of unfolded or denatured GFP does not display any striking feature (Fig. 4, top). However, as the GFP protein folds into its native conformation, these three amino acids are forced into a sharp turn (Fig. 4, middle, left), greatly favouring a nucleophilic attack of the amide of Gly67 on the carbonyl of Ser65, leading to imidazolinone formation by cyclization (Fig. 4, middle, right) and dehydration (Fig. 4, bottom, left). At this point, GFP does not fluoresce (Heim et al., 1994) but, conditional on the presence of molecular oxygen, the  $\alpha$ - $\beta$  bond of residue 66 is subsequently dehydrogenated into conjugation with the imidazolinone (Fig. 4, bottom, right), which results in maturation of the GFP chromophore to its fluorescent form (Heim et al., 1994; Cubitt et al., 1995).

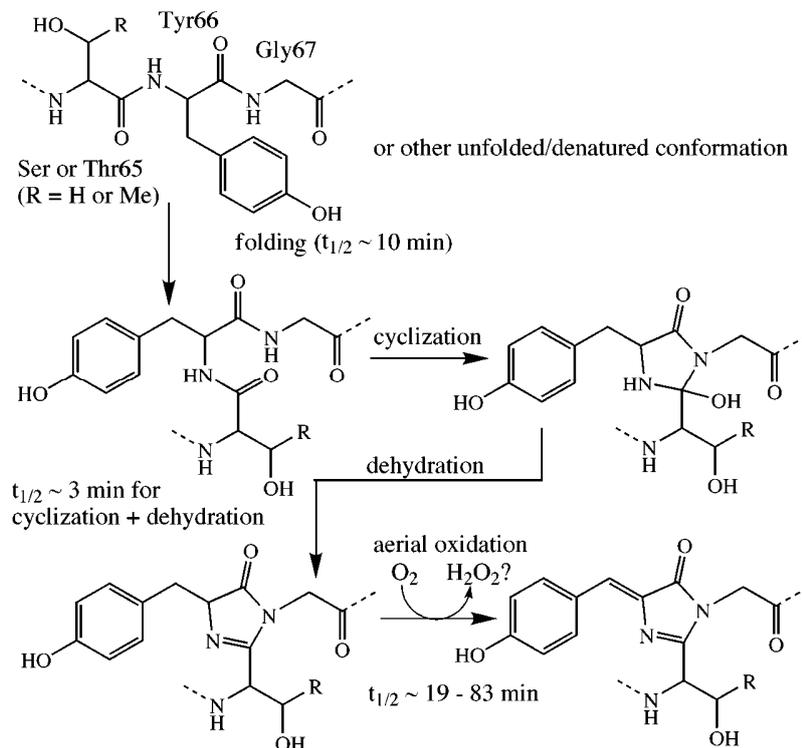


Figure 4. Chemical reaction scheme accounting for the spontaneous formation of the GFP chromophore from the Ser65-Tyr66-Gly67 motif in the native conformation of the protein in the presence of molecular oxygen (Tsien, 1998).

GFP is generally non-toxic and can be expressed to high levels in different organisms with minor effects on their physiology (Chalfie et al., 1994). Furthermore, when the gene for GFP is fused to the gene of a protein to be studied in an organism of interest, the expressed protein of interest retains its normal activity and, likewise, GFP retains its fluorescence, so that the location, movement and other activities of the studied protein can be followed by microscopic monitoring of the GFP fluorescence (Wang and Hazelrigg, 1994). Taken together, the remarkable and unexpected properties of GFP from *Aequorea victoria* summarized in this section are essential for the usefulness of GFP for studies at the molecular level of dynamic processes in living cells.

Today, there are many novel variants of GFP with improved or complementary properties in relation to those of GFP from *Aequorea victoria*, some of which will be discussed in the next section.

### The GFP family and its mutants

From the perspective of its usefulness in the biosciences, some aspects of GFP function stand out. Among these are: (i) the *brightness* of the molecule, defined as the extinction coefficient at the maximum of the excitation spectrum (see Fig. 1) multiplied by the quantum yield, *i.e.* the probability that an excitation of the electronic dipole of the chromophore leads to the

emission of a photon rather than to another, heat-generating transition to the ground state of the chromophore; (ii) the photo-stability of the molecule, *i.e.* the average number of photons that the chromophore emits before the fluorescence is lost due to chemical events emanating from the first singlet excited state and leading to photo-decomposition; (iii) the existence of GFP-like molecules with different excitation and emission spectra throughout the whole visible region; (iv) rapid and efficient folding of the molecule in the intracellular context; (v) rapid maturation of the chromophore subsequent to protein folding; and (vi) monomeric configuration of GFP-like proteins, to facilitate their fusing with proteins of interest. Great brightness (i) and photo-stability (ii) are determinants of the signal-to-noise ratio of GFP-derived intracellular signals. Variable excitation and emission properties (iii) will, in particular, allow for the monitoring of events when two GFP molecules, where the emission spectrum of one (the donor) matches the excitation spectrum of the other (the acceptor), come close to each other. The principle, usually referred to as Fluorescence (or Förster) Resonance Energy Transfer (FRET), is a radiation-less energy transfer between two electronic dipoles occurring with a first-order rate constant inversely proportional to the sixth power of the distance between them. FRET can be used to estimate inter-chromophoric distances smaller than about 100Å, and the method has been generalized to involve energy transfer between three, rather than two, GFP chromophores.

Engineering of GFP from *Aequorea victoria* has led to improved brightness and photo-stability. In addition, variants with better folding properties at temperatures higher than those in the northern Pacific Ocean, the habitat of *Aequorea victoria*, have been engineered, along with GFP molecules with varying excitation and emission spectra (Tsien, 1998; Matz et al., 2002). GFP-like molecules emitting in the red part of the visible spectrum were eventually found among Antozoans (corals) (Matz et al., 1999; 2002), and made useful for intracellular protein-tagging by elimination of their oligomerizing propensity through extensive mutagenesis (Campbell et al, 2002). The coverage of the visible spectrum of light by emission spectra from GFP-like proteins existing in the year 2002 is shown in Fig. 5.

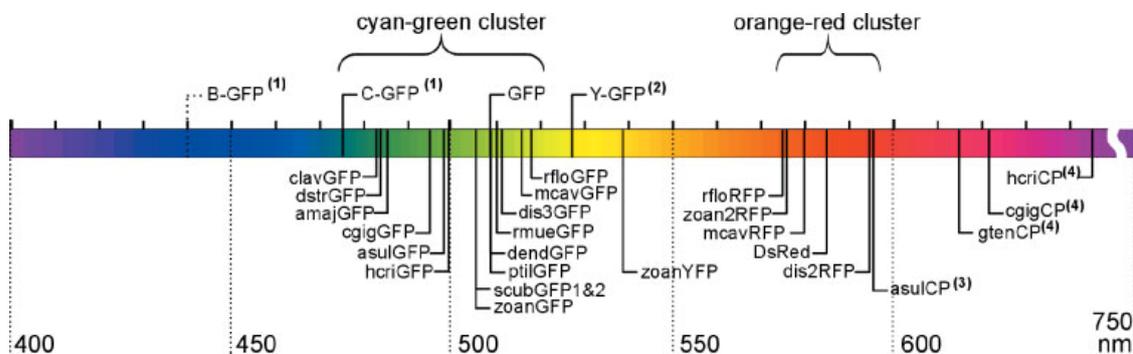


Figure 5. Spectral properties of variants of the GFP family (Matz et al., 2002).

Other important GFP-like mutants can be photo-activated from a non-fluorescent to a fluorescent state or photo-converted from one to another emission wavelength (*e.g.* Patterson and Lippincott-Schwartz, 2002; Ando et al., 2002; Tsutsui et al., 2005).

### The scientific impact of GFP

Synergisms between the suitability of GFP-like proteins as precisely targeted intracellular genetic tags, the rapid development of imaging techniques, and data analysis have boosted the use of GFP in the biological sciences. Early-developed biophysical fluorescence methods like FRET (see above), fluorescence correlation spectroscopy (FCS), fluorescence cross-correlation spectroscopy (FCCS), fluorescence recovery after photo-bleaching (FRAP) and total internal reflection microscopy are now extensively used to monitor intracellular events from signals provided by the fluorescence from GFP-like proteins. In addition, fluorescence life-time imaging (FLIM), high-resolution photo-activation localization microscopy (PALM) and other high resolution methods based on GFP have emerged. Indeed, no other recent discovery has had such a large impact on how experiments are carried out and interpreted in the biological sciences, as witnessed by the appearance of more than 20,000 publications involving GFP since 1992. Some examples of GFP usage are provided below, just to indicate how GFP-based methods have radically changed the experimental potential within essentially all branches of the biological sciences.

The most common use of GFP has been to monitor the location, movement and chemical reactions involving proteins expressed as fusion partners with GFP. The localization of GFP-fusion proteins in different parts of cells, for example during the cell cycle or during exponential growth, has been extensively studied. One example is provided by the pole-to-pole oscillations of the MinDE system that determine the midpoint of bacterial cells for septum formation and cell division, as illustrated in Fig. 6.

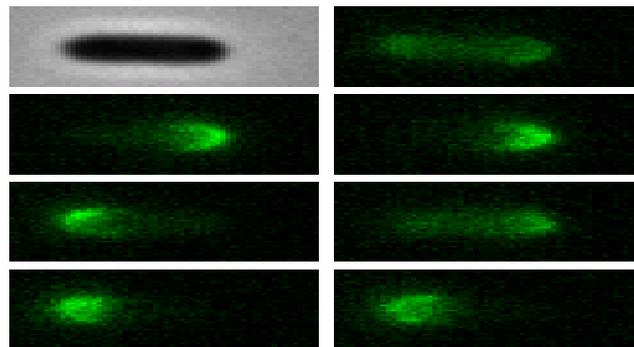


Figure 6. Pole-to-pole oscillations of GFP (green fluorescent protein) fused to MinC (following MinD) (David Fange, Uppsala, unpublished data obtained in the laboratory of J. Paulsson in Harvard, Med. School). Top row, left column: Microscope image of an *E. coli* cell. Sequence of snapshots of the fluorescence from GFP-MinC in the *E. coli* cell at five-second intervals in the order: (left column, rows 2, 3, 4), (right column, rows 1, 2, 3, 4). The figure shows the oscillation of MinC from cell-pole to cell-

pole in a time window slightly longer than the period time of the oscillation. Through these oscillations the average concentration of the septum inhibiting MinD protein is smallest at the mid-point of the cell, which marks the mid-point for septum formation preceding cell division (Raskin and de Boer, 1999ab).

Several spectral variants of GFP tagged to different protein populations in the cell have provided data on how the dynamics of these populations respond to chemical inhibitors, mutations and gene knock outs (Ellenberg et al., 1998).

Another application is provided by monitoring the temporal expression of genes, for example in the formation of large molecular machines like the flagella and their motors, that allow for the chemotactic swimming responses of bacteria (Kalir et al., 2001).

By expressing full-length GFP-tagged proteins from their endogenous chromosomal locations at natural levels, it has become feasible to monitor the intracellular location and concentration spectrum of the whole proteasome of different organisms (Ghaemmaghami et al., 2003).

Another window for detailed kinetic studies of kinetic phenomena in cells is provided by single-cell, single-molecule microscopy based on GFP fluorescence (Xie et al., 2008). For instance, a single protein molecule bound to an immobile target in the cell can be detected as a well-defined spot, provided that its fluorescence is collected during a time shorter than the lifetime of its bound state. If, in contrast, the protein diffuses freely in the cell during the time it is monitored, then its fluorescence is hidden in the background fluorescence. This principle is illustrated by recent studies (Elf et al., 2007), of the dissociation kinetics of a GFP-fused transcription factor when bound to its specific site or to non-specific sites on the chromosome. By varying the microscope monitoring time by the use of laser pulses of different lengths, the chromosomal residence time of the factor could be determined in the two cases, and thus the rate constants for dissociation of the factor from its specific and non-specific sites.

When two proteins, one fused to a GFP with an emission spectrum overlapping the excitation spectrum of the GFP fused to the other one, interact with each other, there will be a strong FRET signal to report their union. Another way to study such direct interactions is through protein complementation with GFP. Here, one protein is fused to one half of GFP, while the other protein is fused to the other half of GFP. As the two proteins form a complex, the two halves of GFP are joined and fluorescence emerges (Cabantous et al., 2004).

An indirect way to identify protein complexes is through Fluorescence Cross Correlation Spectroscopy. Here, two proteins of interest are labelled with GFP molecules emitting light at wavelengths separate from each other. When the fluorescence intensities at these two wavelengths, resulting from laser illumination of a small volume element in a living cell, are cross-

correlated, it can be decided whether the two fused proteins diffuse independently of each other or whether they diffuse as a complex.

A number of GFP-based techniques have been used to study intracellular trafficking (*e.g.* Lippincott Schwartz et al., 2001; Phair and Misteli, 2001). Among these are the highlighting of specific protein pools after photo-bleaching or photo activation of GFP molecules. These methods have made it possible to measure the kinetics, pool sizes and residence times of proteins moving between different subcellular sites.

GFP has been used to study membrane-bound organelles, and one key finding is that many of these continuously exchange protein components with each other. For instance, the Golgi apparatus, which receives secretory cargo from the Endoplasmic Reticulum (ER), constitutively recycles its components back to the ER and disassembles during mitosis.

In the area of nuclear architectures, GFP-based studies have revealed that interphase nuclear structures are both dynamic and self-organizing.

GFP fusion proteins have been successfully used to construct sensors of intracellular parameters, like pH,  $\text{Ca}^{2+}$  or various metabolite concentrations.

Finally, GFP fusions have been extensively used for imaging cells and tissues within multicellular organisms and have become a very important experimental tool in neurobiology.

### **The discovery of GFP**

The jellyfish *Aequorea victoria* is bioluminescent, *i.e.* it produces light with the help of chemical reactions that provide the energy for photon emission and emits green light as first described by Davenport and Nicol (1955). In 1960, Osamu Shimomura joined the laboratory of Frank Johnson at Princeton to clarify the molecular mechanism of the bioluminescence of *Aequorea victoria*. Shimomura came from Nagoya University, where he had completed extensive work on the bioluminescence of the small ostracod *Cypridina*, together with Prof. Y. Hirata. *Aequorea victoria* jellyfish were collected during the following summers in Friday Harbor in the Puget Sound of Washington state on the coast of the Pacific Ocean (Shimomura, 2005).

The active component of the *Aequorea* bioluminescence was identified as a protein, named aequorin, emitting blue light in a  $\text{Ca}^{2+}$ -dependent manner (Shimomura et al., 1962). That the light emission of purified aequorin peaked in the blue part of the visible spectrum came as a surprise, since the bioluminescence of *Aequorea victoria* is distinctly green. In the quest to explain this riddle, Shimomura and his colleagues isolated yet another protein, displaying

strong, green fluorescence (Shimomura et al., 1962). The following quotes from Shimomura et al. (1962) summarize their findings and ideas relating to GFP at this early time-point:

“A protein giving solutions that are slightly greenish in sunlight though only yellowish under tungsten lights, and exhibiting a very bright, greenish fluorescence in the ultraviolet of a Mineralite, has also been isolated from squeezates (*i.e.* extracted by squeezing out the light-emitting cells of *Aequorea victoria*). No indications of a luminescent reaction of this substance could be detected. Studies of the emission spectra of both this protein and aequorin are in progress.”

and

“...it is reasonable to suppose that the greenish quality (*i.e.* of *Aequorea* bioluminescence) results from a light-filtering effect and fluorescence of the green protein which is highly concentrated together with aequorin in the photogenic cells (of *Aequorea*).”

Subsequently, Shimomura and his colleagues found that the blue luminescence from aequorin ( $\lambda_{\text{max}} = 470$ ) (Shimomura and Johnson, 1969; Kohama et al., 1971) matches the long wavelength peak ( $\lambda_{\text{exmax}}=460$  nm) in the excitation spectrum of GFP, with the emission spectrum peaking at about 510 nm ( $\lambda_{\text{emmax}}$  508-515 nm) (Cormier et al., 1973). From these data, it had been suggested without direct experimental proof that the green light of *Aequorea victoria* originates in direct, radiation-less transfer of the energy of the chemically excited electronic dipole in aequorin to excite the electronic dipole of GFP, followed by green photon emission as the latter returns to the ground state (Morin and Hastings, 1971ab). In modern language, this would mean that GFP is the acceptor and aequorin the donor in an energy-transfer reaction of FRET type (see above on FRET). Experimental evidence for this hypothesis was provided by Shimomura and colleagues in a publication (1974), including estimates of the distance between aequorin and GFP in the photogenic cells of *Aequorea victoria*, along with purification protocols and detailed characterization of the absorption and emission spectra of GFP (Morise et al., 1974).

Although Shimomura's main interest was the bioluminescence of aequorin, including its use as a calcium indicator in living cells, he turned his attention to the fluorescent chromophore of GFP to clarify its chemical structure (Shimomura, 1979). He digested GFP with papain, which led to the disappearance of the fluorescence of the protein, but found one peptide fragment with the same absorption spectrum as that of the intact protein. He studied the physical-chemical properties of this peptide compared with those of a model compound, and suggested that the chromophore is a p-hydroxybenzylideneimidazolinone moiety (See Fig. 1). With access to the primary structure of GFP (Prasher et al., 1992; see below), Cody et al. (1993) re-characterized the GFP chromophore, also using Nuclear Magnetic Resonance

(NMR). They confirmed the functional portion of the chromophore as p-hydroxybenzylideneimidazolinone, as suggested by Shimomura, but re-identified the two amino acids making up the imidazolone ring as Ser and Gly (Fig. 1).

In summary, Shimomura made essential contributions to the discovery of GFP, its purification and the characterization of its physicochemical properties, including the excitation and emission spectra of its fluorescence under various conditions. He demonstrated with his colleagues that GFP can act as the acceptor in FRET from aequorin as donor, which explains why *Aequorea victoria* emits green, rather than blue, light. Finally, he correctly assigned the functional portion of the chromophore integrated in the peptide chain of GFP. Without the pioneering research of Shimomura, mainly with classical methods for protein purification and spectroscopy, it is likely that the GFP revolution would have been delayed by decades or even remained one of the hidden secrets of the Pacific Ocean.

### **The expression of GFP in organisms other than *Aequorea victoria***

Douglas Prasher, working in M.J. Cormier's laboratory, cloned the gene for the bioluminescent protein aequorin (Prasher et al., 1985). Using the same cDNA library from *Aequorea victoria*, Prasher and his colleagues made a first attempt to fish out the gene for GFP with DNA oligos derived from known peptide sequence elements of GFP (Prasher et al., 1992). In this way, they identified an open reading frame (ORF) encoding 168 amino acids, suggesting a molecular weight of the protein significantly smaller than that estimated earlier with other methods. This discrepancy suggested that their ORF was incomplete, but it could all the same be used to identify the three amino acids that make up the core of the chromophore of GFP (Fig. 1; Cody et al., 1993). With the help of a second cDNA library from *Aequorea victoria*, they obtained and cloned the complete ORF of GFP, encoding 238 amino acids, with the chromophore precursor motif Ser-Tyr-Gly positioned as residues 65-67 (Prasher et al., 1992). It was believed by many of the experts at the time that formation of a fluorescing chromophore in GFP would require an unknown enzyme system, idiosyncratic to *Aequorea victoria*. It was therefore considered likely that expression in heterologous systems would result in a non-fluorescent apo-form of GFP, rather than in its scientifically useful fluorescent form. At the end of their cloning article, Prasher et al. (1992) summarized their view of the prospects:

“These (cloning) results will enable us to construct an expression vector for the preparation of non-fluorescent apoGFP. Since no information is yet available regarding the biosynthesis of the chromophore, a recombinant form of this protein will be a valuable reagent with which to examine the biochemistry of chromophore formation in this unique class of proteins and the mechanism of energy transfer between aequorin and GFP”.

Prasher, Ward and collaborators (Code et al., 1993; see also above) used the by then known peptide sequence of GFP (Prasher et al., 1992) to identify the tri-peptide motif responsible for

eventual formation of the functional core of its chromophore. In this, they used a similar approach with initial papain degradation of the protein to that used earlier by Shimomura (1979) and, in addition, they validated their chromophore model (Fig. 1) by NMR spectroscopy (Code et al., 1993). As mentioned above, the functional core of their chromophore was the same as that suggested by Shimomura, but they could now correctly assign Ser65 and Gly67 as the precursors of the heterocyclic imidazolidinone ring. Although the GFP chromophore was now correctly assigned, its maturation from a Ser-Tyr-Gly peptide segment in “apo-GFP” to the fluorescing GFP variant remained a mystery. They summarized the situation (Code et al., 1993):

“The posttranslational events required for chromophore formation are not yet understood. It is very unlikely that the chromophore forms spontaneously, but its formation probably requires some enzymatic machinery.”

The question whether the GFP-chromophore forms spontaneously or whether its maturation requires an auxiliary enzyme system was soon to be answered, but preceding this clarification the expression of brilliantly fluorescing GFP in a selected number of model organisms was of great biological significance.

Martin Chalfie had spent post-doc time with Sidney Brenner (Nobel Laureate in Physiology or Medicine in 2002) in Cambridge, UK, working on neuron development in the small nematode (earthworm) *Caenorhabditis elegans* (*C. elegans*).

When Chalfie first heard about GFP, he was greatly excited about the prospects of using the molecular genetics worked out for *C. elegans* by S. Brenner and his colleagues to transform the organism with genes for GFP fused with a host of interesting proteins for expression under control of their authentic promoters. When affiliated at Columbia University, he obtained the GFP gene (*gfp*) clone from Prasher in 1992 after its publication (Prasher et al., 1992), and expressed it in *E. coli* the same year. The GFP protein displayed a bright, green fluorescence in this heterologous organism, suggesting that it could indeed serve as a versatile genetic marker in virtually all organisms. With the *gfp* clone obtained from Prasher, Chalfie transformed *C. elegans* with *gfp* under the control of a promoter regulating the expression of  $\beta$ -tubulin, abundant in six touch receptor neurons in *C. elegans*. The organism subsequently expressed GFP from distinct positions in its body and at distinct times in its development.

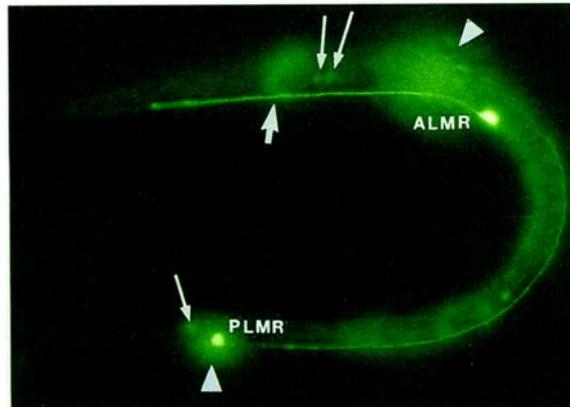


Figure 7. Expression of GFP from a first stage *C. elegans* larva (Chalfie et al., 1994). The two touch receptor neurons ALMR and PLMR are fluorescence-labelled at their cell bodies, leading to the strongly fluorescing two dots in the figure. Fluorescing processes can be seen projecting from both of these cell bodies. Halos produced from the out-of-focus of homologues of these cells on the other side of the animal are indicated by arrow heads. The thick arrow points to the nerve ring branch from the ALMR cell (out of focus); the thin arrows point to weakly fluorescing cell bodies.

Chalfie and his colleagues submitted their *gfp* expression results in *E. coli* and *C. elegans* to Science in mid-September 1993 and the article was published at the beginning of February 1994 (Chalfie et al., 1994). Their *C. elegans* results were also mentioned in an abstract on “A New Method of Looking at *C. elegans* Gene Expression” published in Worm Breeder’s Gazette in October 1993 (Chalfie et al., 1993). The latter results were summarized as follows:

“We have developed a new way to look at gene expression in *C. elegans* (and other organisms) that utilizes an inherently fluorescent protein (the green-fluorescent protein; GFP) from the jellyfish *Aequorea victoria*. GFP fluoresces bright green when illuminated with blue light. We have found that this fluorescence does not depend upon any other component specific to *Aequorea victoria*, so *gfp* can be used instead of *lacZ*, for example, to make gene expression fusions.”

The fluorescent form of GFP could be expressed not only in *E. coli* and *C. elegans* (Chalfie et al., 1993; 1994), but also in the yeast *Sacharomyces cerevisiae*, as later demonstrated by Roger Tsien and his collaborators (Chalfie et al., 1994, footnote 23) and in mammalian cells as subsequently demonstrated by Lanini and MacKeon (*Ibid*). In the same year, the possibility to use constructs of GFP fused to proteins of interest was demonstrated for yet another important model organism, namely *Drosophila melanogaster* (fruit fly) (Wang and Hazelrigg, 1994). The findings that brightly fluorescing GFP could be expressed in four very important model organisms and that GFP-fusion constructs retained both the fluorescence of GFP and the activity of the fusion partner demonstrated the virtually unlimited potential of GFP as a universal genetic tag in biological research. At the same time, the question how its chromophore matures remained unanswered (Chalfie et al., 1994):

“...chromophore formation is not species specific and occurs either through the use of ubiquitous cellular components or by autocatalysis.”

In summary, Martin Chalfie, with a *gfp* clone obtained from Douglas Prasher, was the first to demonstrate that GFP in its fluorescent form can be expressed in both *E. coli* and *C. elegans* without addition of auxiliary heterologous factors. These results meant a breakthrough in that they demonstrated the feasibility of GFP as a universal genetic marker, thereby opening the door to the universe of experimental possibilities for quantitative studies of dynamic processes in living cells that we benefit from today.

### **Development of GFP and GFP-like proteins**

Roger Tsien expressed the *gfp*-gene anaerobically in *E. coli* and found that under oxygen-deficient conditions GFP did not fluoresce, in spite of the fact that it was expressed and apparently folded. However, fluorescence slowly emerged upon addition of oxygen to such non-fluorescent GFP molecules in living cells or in very dilute cell extracts where, in both cases, there was no *de novo* synthesis of proteins (Heim et al., 1994). From these data, Tsien and his collaborators inferred that formation of the fluorescent chromophore of GFP occurs post-translationally with molecular oxygen as the only auxiliary factor. Since molecular oxygen is ubiquitous in all aerobically living cells, their discovery explained why GFP had effortlessly fluoresced in every organism in which it had so far been expressed and strongly suggested that the same would be true for every aerobically living organism where GFP- expression was attempted. They also suggested the chemical scheme for chromophore formation (Fig. 4).

In the same paper, they introduced a number of point mutations in GFP with profound effects on its spectral properties, including variants where the main peak of the excitation spectrum of wild type GFP was shifted from the UV (Fig. 2) to the blue part of the spectrum. These engineering results showed that GFP was robust vis-à-vis amino acid substitutions, including changes in the functional portion of the chromophore (Fig. 1). They gave the start signal for all those engineering steps that have improved GFP and paved the way for its many successful applications in the biological sciences.

In these early years, Tsien and his collaborators deepened the mechanistic understanding of the fluorescence of GFP and developed several new GFP variants (Tsien, 1998) with altered spectral properties (Heim et al., 1994; Heim and Tsien, 1996), improved brightness (Heim et al., 1995) and ameliorated folding (Heim and Tsien, 1996; see also Tsien, 1998, Fig. 6).

An important step forward, allowing for rational design of mutants, was the solution of the crystal structure of GFP by Remington in collaboration with Tsien (Ormö et al., 1996) and, independently, by Yang et al. (1996).

Apart from the engineering of GFP from *Aequorea victoria*, Tsien has also made a number of important contributions to the development of variants of red fluorescent proteins, often based on the DsRed fluorescing protein from the coral *Discosoma*, discovered by Lukyanov and collaborators (Matz et al., 1999; Matz et al., 2002). Tsien and his collaborators characterized the structure of the chromophore of DsRed and the chemical steps leading to its maturation (Gross et al., 2000). This work included quantum chemical computations accounting for the red-shift in the excitation and emission spectra of DsRed, compared to those of GFP from *Aequorea victoria*. He used extensive mutations of DsRed to transform it from an obligate tetramer, of limited use as a genetic tag, to a monomer with retained fluorescence properties (Baird et al., 2000; Campbell et al., 2002). Recent extensions and summaries of this work by Tsien and collaborators can be found in Shaner et al., 2004; 2008. The genealogy of different proteins developed from DsRed is shown in Fig. 8:

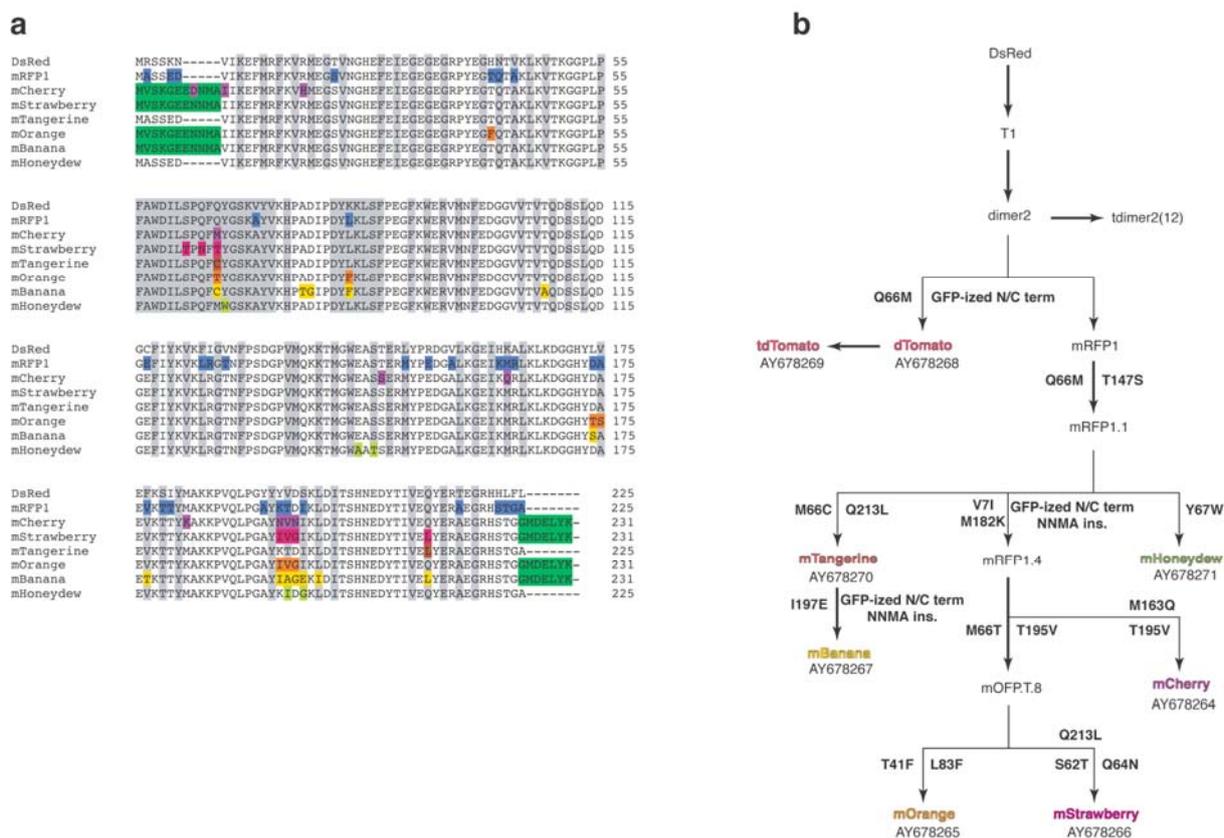


Figure 8. Genealogy of proteins derived from DsRed. 8a Sequence changes and their colours. 8b. The family tree of DsRed proteins. (From Shaner et al. 2004).

The spectroscopic qualities of these proteins are summarized in Fig. 9:

Class	Protein	Source laboratory (references)	Excitation <sup>c</sup> (nm)	Emission <sup>d</sup> (nm)	Brightness <sup>e</sup>	Photostability <sup>f</sup>	pKa	Oligomerization
Far-red	mPlum <sup>g</sup>	Tsien (5)	590	649	4.1	53	<4.5	Monomer
Red	mCherry <sup>g</sup>	Tsien (4)	587	610	16	96	<4.5	Monomer
	tdTomato <sup>g</sup>	Tsien (4)	554	581	95	98	4.7	Tandem dimer
	mStrawberry <sup>g</sup>	Tsien (4)	574	596	26	15	<4.5	Monomer
	J-Red <sup>h</sup>	Evrogen	584	610	8.8 <sup>*</sup>	13	5.0	Dimer
	DsRed-monomer <sup>h</sup>	Clontech	556	586	3.5	16	4.5	Monomer
Orange	mOrange <sup>g</sup>	Tsien (4)	548	562	49	9.0	6.5	Monomer
	mKO	MBL Intl. (10)	548	559	31 <sup>*</sup>	122	5.0	Monomer
Yellow-green	mCitrine <sup>i</sup>	Tsien (16,23)	516	529	59	49	5.7	Monomer
	Venus	Miyawaki (1)	515	528	53 <sup>*</sup>	15	6.0	Weak dimer <sup>j</sup>
	YPet <sup>g</sup>	Daugherty (2)	517	530	80 <sup>*</sup>	49	5.6	Weak dimer <sup>j</sup>
	EYFP	Invitrogen (18)	514	527	51	60	6.9	Weak dimer <sup>j</sup>
Green	Emerald <sup>g</sup>	Invitrogen (18)	487	509	39	0.69 <sup>k</sup>	6.0	Weak dimer <sup>j</sup>
	EGFP	Clontech <sup>l</sup>	488	507	34	174	6.0	Weak dimer <sup>j</sup>
Cyan	CyPet	Daugherty (2)	435	477	18 <sup>*</sup>	59	5.0	Weak dimer <sup>j</sup>
	mCFPm <sup>m</sup>	Tsien (23)	433	475	13	64	4.7	Monomer
	Cerulean <sup>g</sup>	Piston (3)	433	475	27 <sup>*</sup>	36	4.7	Weak dimer <sup>j</sup>
UV-excitable green	T-Sapphire <sup>g</sup>	Griesbeck (6)	399	511	26 <sup>*</sup>	25	4.9	Weak dimer <sup>j</sup>

<sup>a</sup>An expanded version of this table, including a list of other commercially available FPs, is available as **Supplementary Table 1**. <sup>b</sup>The mutations of all common AFPs relative to the wild-type protein are available in **Supplementary Table 3**. <sup>c</sup>Major excitation peak. <sup>d</sup>Major emission peak. <sup>e</sup>Product of extinction coefficient and quantum yield at pH 7.4 measured or confirmed (indicated by \*) in our laboratory under ideal maturation conditions, in (mM • cm)<sup>-1</sup> (for comparison, free fluorescein at pH 7.4 has a brightness of about 69 (mM • cm)<sup>-1</sup>). <sup>f</sup>Time for bleaching from an initial emission rate of 1,000 photons/s down to 500 photons/s (t<sub>1/2</sub>; for comparison, fluorescein at pH 8.4 has t<sub>1/2</sub> of 5.2 s); data are not indicative of photostability under focused laser illumination. <sup>g</sup>Brightest in spectral class. <sup>h</sup>Not recommended (dim with poor folding at 37 °C). <sup>i</sup>Citrine YFP with A206K mutation; spectroscopic properties equivalent to Citrine. <sup>j</sup>Can be made monomeric with A206K mutation. <sup>k</sup>Emerald has a pronounced fast bleaching component that leads to a very short time to 50% bleach. Its photostability after the initial few seconds, however, is comparable to that of EGFP. <sup>l</sup>Formerly sold by Clontech, no longer commercially available. mECFP with A206K mutation; spectroscopic properties equivalent to ECFP.

Figure 9. Spectroscopic and other relevant properties derived from DsRed. (From Shaner et al., 2008).

Roger Tsien has also pioneered in the development of fluorescence-based sensors of Ca<sup>2+</sup> concentrations in whole organisms, tissues, organelles and submicroscopic environments, previously out of reach for quantitative monitoring (Miyawaki et al., 1999). The monitoring principle is chimeric protein constructs, consisting of a blue or cyan mutant of GFP, calmodulin (CaM) and a glycylglycine linker, the CaM-binding domain of myosin light chain kinase (M13) and a green or yellow version of GFP. When calcium ions bind to CaM, this induces the binding of CaM to M13, which reduces the distance between the two GFP variants and thereby increases the FRET efficiency. In this way, the FRET signal can be calibrated to the intracellular concentration of Ca<sup>2+</sup>.

In summary, Roger Tsien has made seminal contributions to our understanding of the chemistry of the fluorescence properties of GFP and other members of the GFP family. He has made extensive contributions to the development of GFP variants with fluorescence emission in the whole visible spectrum, with increased brightness and photostability and improved folding properties along with rapid maturation of their chromophores. Roger Tsien's research has been instrumental for the development of GFP and GFP-like proteins to the highly efficient genetic tags that today constitute a universal "tool box" for studies of dynamic processes in all living systems.

## References

- Ando R. et al. (2002) *Proc. Natl. Acad. Sci. USA* **99** 12651-12656.
- Baird, G.S. et al. (2000) *Proc. Natl. Acad. Sci.* **97** 11984-11989.
- Brejč, K. et al. (1997) *Proc. Natl. Acad. Sci. USA* **94** 2306-2311.
- Cabantous, S. et al. (2004) *Nature Biotechniques* **23** 102-107.
- Campbell, R.E. (2002) *Proc. Natl. Acad. Sci. USA* **99** 7877-7882.
- Chalfie, M. (1993) *Worm Breeder's Gazette* **13** (1) (October 1).
- Chalfie, M. et al. (1994) *Science* **263** 802-805.
- Cody, W.C., Prasher, D.C. et al. (1993) *Biochemistry* **32** 1212-1218.
- Cormier, M.J. et al. (1973) *J. Cell Physiol.* **81** 291-297.
- Cubitt, A.B. et al. (1995) *Trends Biochem. Sci.* **20** 448-455.
- Davenport, D. and Nicol, J.A.C. (1955) *Proc. R. Soc. London, SerB* **144** 399-411.
- Elf, J. et al. (2007) *Science* **316** 1191-1194.
- Ellenberg, J. et al. (1998) *Biotechniques* **25** 838-842.
- Fradkov, A.F. (2000) *FEBS Letters* **479** 127-130.
- Ghaemmaghami et al. (2003) *Nature* **16** 737-741.
- Gross, L.A. et al. (2000) *Proc. Natl. Acad. Sci. USA* **97** 11990-11995.
- Heim, R. et al., (1994) *Proc. Natl. Acad. Sci. USA* **91** 12501-12504.
- Heim, R. et al., (1995) *Nature* **373** 663-664.
- Heim, R. and Tsien, R. (1996) *Curr. Biol.* **6** 178-182.
- Kalir, S. et al. (2001) *Science* **292** 2080-2083.
- Kohama, Y., Shimomura, O. et al. (1971) *Biochemistry* **10** 4149.
- Lippincott-Schwartz, J. et al. (2001) *Nat. Rev. Mol. Cell Biol.* **2** 444-456.
- Matz, M. et al. (2002) *BioEssays* **24** 953-959.
- Matz, M. et al. (1999) *Nature Biotechnology* **17** 969-973.
- Miyawaki, A. et al. (1999) *Proc. Natl. Acad. Sci. USA* **96** 2135.
- Morin, J.G. and Hastings, J.W. (1971a) *J. Cell Physiol.* **77** 303-312.
- Morin, J.G. and Hastings, J.W. (1971b) *J. Cell Physiol.* **77** 313-318.
- Morise H., Shimomura, O. et al. (1974) *Biochemistry* **13** 2656-2662.
- Ormö, M. et al. (1966) *Science* **273** 1392-1395.
- Patterson, G.H. and Lippincott-Schwartz, J. (2002) *Science* **297** 1873-1877.
- Phair, R.D. and Misteli, T. (2001) *Nat. Rev. Mol. Cell Biol.* **2** 898-907.
- Prasher, D. et al. (1985) *Biophys. Res. Commun.* **126** 1259-1268.
- Prasher, D. et al. (1992) *Gene* **111** 229-233.
- Raskin, D.M. and De Boer, P.A. (1999a) *Proc. Natl. Acad. Sci. U SA* **96** 4971-4976.
- Raskin, D.M. and De Boer, P.A. (1999b) *J. Bacteriol.* **181** 6419-6424.
- Shaner, N.C. et al. (2004) *Nature Biotechnology* **22** 1562-1572.
- Shaner, N.C. et al. (2008) *Nature Methods* **5** 545-551.
- Shimomura, O. (1979) *FEBS Letters* **104** 220-222.
- Shimomura, O., Johnson, F.H. and Saiga, Y. (1962) *J. Cell. Comp. Physiol.* **59** 223-240.

- Shimomura, O. and Johnson, F.H. (1969) *Biochemistry* **8** 3991-3997.
- Shimomura, O. (2005) *Journal of Microscopy* **217** 3-15.
- Tsien, R. (1998) *Annu. Rev. Biochem.* **67** 509-544.
- Tsutsui, H. et al. (2005) *EMBO Reports* **6** 233-238.
- Wang, S.X. and Hazelrigg, T. (1994) *Nature* **369** 400-403.
- Xie, X.S. et al. (2008) *Annu. Rev. Biophys.* **37** 417-444.
- Yang, F. et al. (1996) *Nature Biotechnology* **14** 1246-1251.

Uppsala 30 September 2008

Måns Ehrenberg