

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

## GFP-Based Biosensors

---

Donna E. Crone, Yao-Ming Huang, Derek J. Pitman,  
Christian Schenkelberg, Keith Fraser,  
Stephen Macari and Christopher Bystroff

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/52250>

---

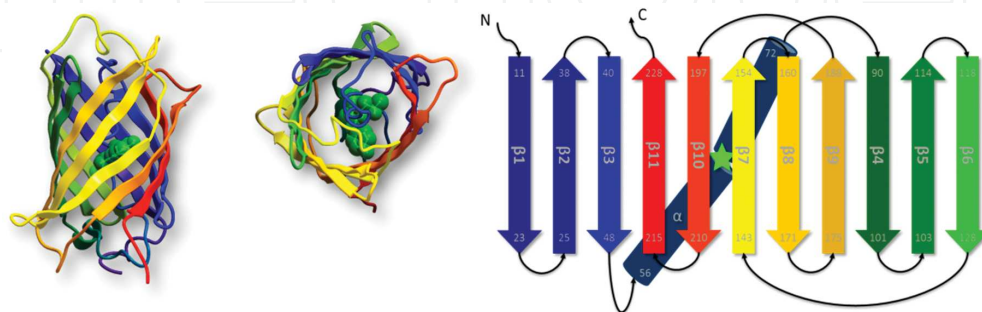
### 1. Introduction

Green fluorescent protein (GFP) is a 27 kD protein consisting of 238 amino acid residues [1]. GFP was first identified in the aquatic jellyfish *Aequorea victoria* by Osamu Shimomura *et al.* in 1961 while studying aequorin, a  $\text{Ca}^{2+}$ -activated photoprotein. Aequorin and GFP are localized in the light organs of *A. victoria* and GFP was accidentally discovered when the energy of the blue light emitted by aequorin excited GFP to emit green light. Unlike most fluorescent proteins which contain chromophores distinct from the amino acid sequence of the protein, the chromophore of GFP is internally generated by a reaction involving three amino acid residues [2]. This unique property allows GFP to be easily cloned into numerous biological systems, both prokaryotic and eukaryotic, which has paved the way for its utilisation in a variety of biological applications, most notably in biosensing.

#### 1.1. The three dimensional structure

The molecular structure of GFP was first determined in 1996 using X-ray crystallography [1]. One of the most obvious features of its tertiary structure is a beta-barrel composed of 11 mostly-antiparallel beta strands. The molecular structure of GFP is illustrated in Figure 1 along with a cartoon representation showing the organization of the secondary structure elements that compose the beta barrel. Each beta strand is 9 to 13 residues in length and hydrogen bonds with adjacent beta strands to create an enclosed structure. The bottom of the barrel contains both termini and two distorted helical crossover segments, and the top has one short crossover and one distorted helical crossover segment. The beta-barrel (sometimes referred to as a “beta can” because it contains a central alpha-helical segment) consists of three anti-parallel three-stranded beta-meander units and a two-stranded beta-hairpin

(shown in blue, green, and yellow, and red in Figure 1 respectively). The very distorted central alpha helix contains three residues which participate in an auto-catalyzed cyclization/oxidation chromophore maturation reaction which generates the *p*-hydroxybenzylidene-imidazolidone chromophore. In the unfolded state, the chromophore is non-fluorescent, presumably because water molecules and molecular oxygen can interact with and quench the fluorescent signal [3]. Therefore, the closed beta barrel structure is essential for fluorescence by shielding the chromophore from bulk solvent.



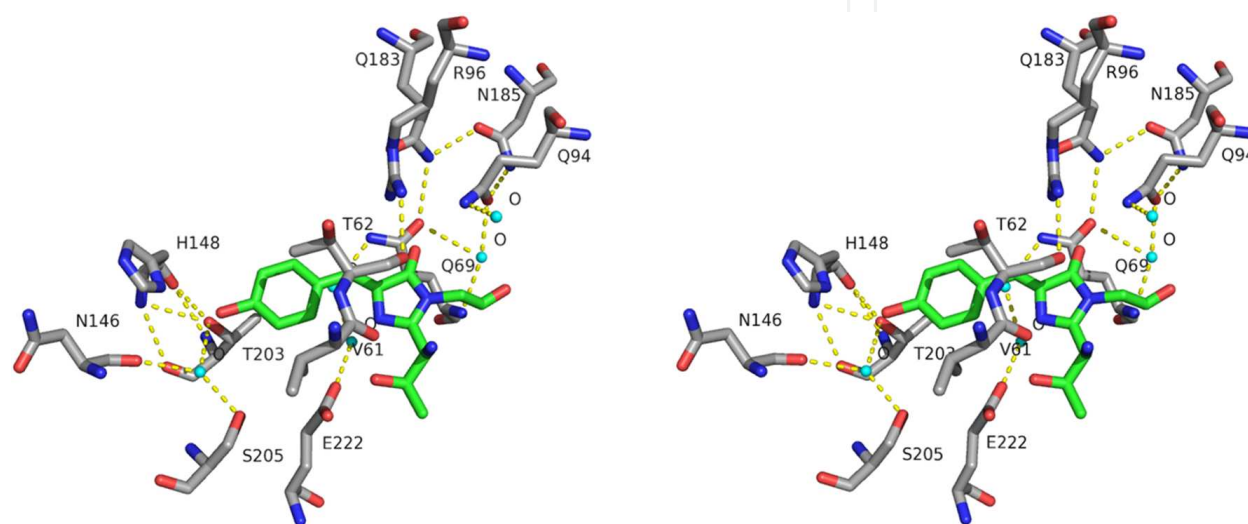
**Figure 1.** Tertiary structure of GFP as determined by x-ray crystallography (PDB code 2B3P). Shown on the bottom is a cartoon depicting the secondary structure elements, all anti-parallel beta strand pairings except  $\beta 1$  to  $\beta 6$ , which is parallel. Numbers indicate the start and end of each secondary structure element.

The interior of the GFP beta barrel is unusually polar. There is an interior cavity filled with four water molecules on one side of the central helix, while the other side contains a cluster of hydrophobic side chains which is more typical of a protein core. Several polar side chains interact with and stabilize the GFP chromophore. Three of these, His148, Thr203, and Ser205, form hydrogen bonds with the phenolic hydroxyl group of the chromophore. Arg96 and Gln94 interact with the carbonyl group of the imidazolidone ring. Figure 2 depicts these stabilizing hydrogen bonding interactions with the chromophore. Additionally, a number of internal residues interact with and stabilize Arg96, a side chain that is known to be required for the maturation of the chromophore. Specifically, Thr62 and Gln183 form hydrogen bonds with the protonated form of Arg96 stabilizing a buried positive charge within the GFP beta barrel, which in turn stabilizes a partial negative charge on the carbonyl oxygen of the imidazolidone ring.

## 1.2. Thermodynamic and kinetic properties

Wild type GFP has a number of interesting characteristics that can potentially complicate its applicability to biosensing. One is its tendency to aggregate in the cell, especially when expressed in high concentrations. Aggregation is typically caused by exposed hydrophobicity, which may be due to either the presence of hydrophobic patches on the surface of the protein, or to low thermostability, or to slow folding. Surface hydrophobic-to-hydrophilic mutations decrease the aggregation tendency of GFP [4], but some biosensing applications require surface mutations that may increase aggregation. Most likely, GFP's low *in vivo* solubility is due to its extremely slow folding and unfolding kinetics. Refolding of GFP consists of at least two observable phases, depending on the variant and the method being used to measure the kinetic

ics. Multi-phase folding kinetics indicates the existence of multiple parallel folding pathways, some fast and some slow, holding out the hope that engineered GFPs could be made to fold faster by favoring the faster folding pathway. Indeed, GFP has been engineered to eliminate the slowest phase of folding, as discussed later in this chapter. For Cycle3, a mutant whose chromophore matures correctly at 37°C, the kinetic phases range from  $10 \text{ s}^{-1}$  to  $10^{-2} \text{ s}^{-1}$  [5] (half-lives of folding ranging from 0.1 s to 100 s). Although it folds slowly, GFP unfolds *extremely* slowly, with a rate of  $10^{-6} \text{ s}^{-1}$  ( $t_{1/2}$ =8 days) in 3.0M GndHCl [6], such that when extrapolated to 0M GndHCl, the theoretical unfolding half-life is on the order of  $t_{1/2}$ = 22 years. GFP is phenomenally kinetically stable once it is folded to its native state.



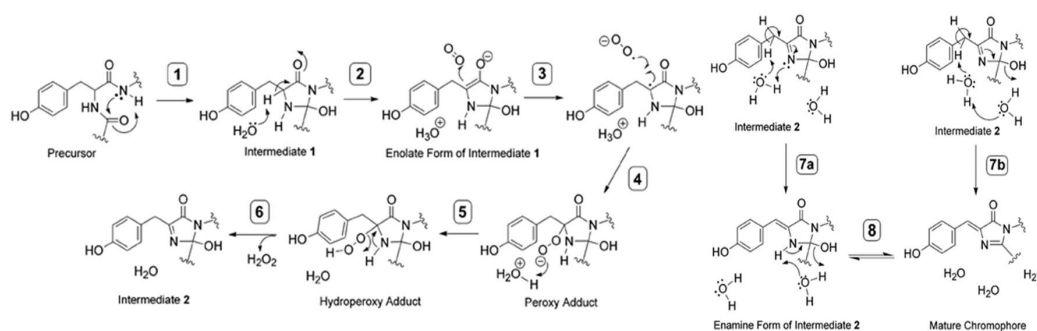
**Figure 2.** Stereo image of the hydrogen bonding patterns of the internal GFP residues with the chromophore (green), including four crystallographic waters (cyan). Drawn from superfolder GFP, PDB ID 2B3P.

### 1.3. Maturation of the chromophore

The chromophore of the native GFP structure is generated by an internal, autocatalytic reaction involving three residues on the interior alpha helix. Cyclization and oxidation of internal residues of Ser65, Tyr66, and Gly67, generate a *p*-hydroxybenzylidene-imidazolidone chromophore that maximally absorbs light at 395 nm and 475 nm [1]. Excitation at either absorption peak results in emission of green light at 508 nm. Interestingly, the sidechains of the chromophore triplet 65-SerTyrGly-67 can be mutated to other sidechains without loss of function. Tyrosine 66 can be mutated to any aromatic sidechain [7]. This allows for the synthesis of numerous variants of GFP that alter the chromophore structure or its surrounding environment to absorb and emit light at different wavelengths, producing a wide array of fluorescent protein colors [8].

The three-step mechanism for the spontaneous generation of the chromophore consists of cyclization, oxidation, and dehydration [9]. Figure 3 illustrates the mechanism, beginning with the original triplet of amino acids. The slow step in chromophore maturation is the diffusion of molecular oxygen into the active site of the closed beta barrel (step 3). The posi-

tioning of side chains surrounding the chromophore is crucial for stabilizing the intermediates in the process of chromophore maturation, especially Arg96, which stabilizes the enolate form of intermediate 1 by forming a salt bridge with the negatively-charged oxygen atom, and Glu222, which receives protons from the water molecules to cycle between the protonated and deprotonated states. The two coplanar aromatic rings of the chromophore adopt the *cis* conformation across the Tyr66 alpha-beta carbon double bond. Photo-bleaching, the light-induced loss of fluorescence, is caused by short wavelength light that causes the chromophore to isomerize to the *trans* form, accompanied by distortion of its planar geometry and surrounding side chain packing [10]. This type of photobleaching appears to be a slowly reversible process for GFP and other fluorescent proteins.



**Figure 3.** Mechanism of the maturation of the GFP chromophore. Steps 1-6 include the cyclization and deoxidation steps while step 7 indicates two possible pathways for the dehydration step. Used with permission from [9]

The two spectral absorbances of the GFP chromophore have been found to be highly sensitive to pH changes [11]. At physiological pH, GFP exhibits maximal absorption at 395 nm while absorbing lesser amounts of light at 475 nm. However, increasing the pH to about 12.0 causes the maximal absorption of light to occur around 475 nm while diminishing the absorption at 395 nm. The two absorption maxima correspond to different protonated states of the chromophore. The  $pK_a$  for the side chain hydroxyl group of Tyr66 is about 8.1 [12] and therefore, the maximal absorbance for the neutral chromophore occurs at 395 nm while maximal absorbance occurs at 470 nm for the anionic form of the chromophore. At acidic pHs lower than 6 or alkaline pHs above 12, fluorescence is diminished as GFP is denatured and the chromophore is quenched.

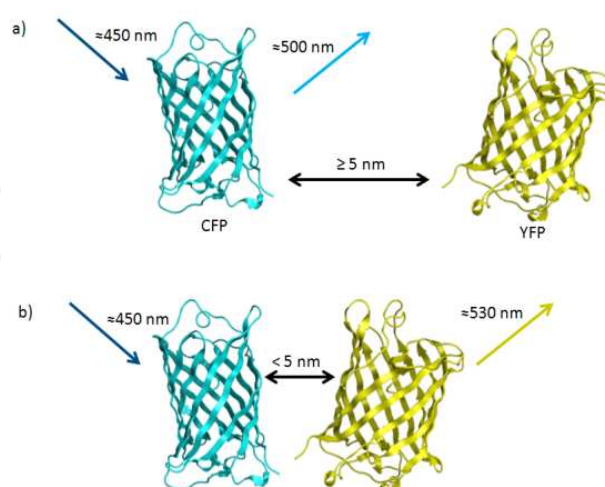
#### 1.4. Wavelength variants and FRET

Starting with homologous green and red fluorescent proteins, a rainbow of different-colored fluorescent proteins have been developed. Mutating Tyr66 of the GFP chromophore to a tryptophan produces cyan fluorescence, while a histidine mutation produces blue fluorescence. Mutating a threonine on beta strand 10 to a tyrosine introduces a pi-stacking interaction which produces yellow fluorescence. See [3] for more details. At the other end of the color spectrum, the coral-derived DsRed fluorescent protein, a structural homolog of GFP, was diversified into the mFruits library, producing eight fluorescent proteins with emission maxima ranging from 537 to 610 nm [13]. Far-red fluorescent proteins, which have potential



for use in deep tissue imaging due to the penetration of these wavelengths, have been discovered [14-16], while others have been developed in the lab [17] and even using computational approaches [18]. Further enhancement of these wavelength-shifted variants has improved their biophysical properties and made them available to more applications.

GFP and its derivatives have seen significant use as fluorescent pairs for Förster Resonance Energy Transfer (FRET) experiments. FRET emission arises when the emission spectrum of one chromophore overlaps with the excitation spectrum of another chromophore. If the two chromophores are physically close (on the order of a few nanometers) and in the correct orientation, then excitation of the first chromophore will excite the second chromophore through non-radiative energy transfer and produce fluorescence at the second chromophore's emission wavelength (Figure 4). This phenomenon can be used to detect when two fluorescent proteins (FPs) are within a certain distance, which may be induced by a ligand-dependent conformational change in a linking domain between the two fluorescent proteins, or by binding of interacting domains fused to fluorescent proteins. The canonical pairing for FRET using fluorescent proteins is cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) [19]; but this pairing has issues concerning overlapping emission spectra, stability to photobleaching, and sensitivity to the chemical environment. The study in [20] had the goal of producing a cyan fluorescent protein more suitable for use in FRET experiments. Other pairings, such as GFP and the the DsRed-based variant mCherry red fluorescent protein, have been proposed as consistent, reliable alternatives [21]. A full review of the development and usage of fluorescent proteins as tools for FRET can be found in [22]. The genetic and physical ease of use of GFP-derived fluorescent proteins, in conjunction with their wide range of colors and spectral overlaps, makes them ideal molecules for the design of FRET-based biosensors.



**Figure 4.** Illustration of the FRET phenomenon using the traditional CFP/YFP donor/acceptor pairing. a) If the two fluorescent moieties are too far apart, excitation of the donor molecule only produces observable emission from the donor. b) When in range, excitation of the donor is propagated to the acceptor molecule through non-radiative photon transfer, and emission from the acceptor is observed.

### 1.5. Mutants with improved features

Because of the aforementioned slow folding, low solubility and slow chromophore maturation, a significant effort has been put forth to improve these properties in GFP. These strategies range from specific, directed rational mutations based on structural and biophysical information to fully randomized approaches such as error-prone PCR [23] and DNA shuffling [24]. By mutating the chromophore residue serine 65 to a threonine (S65T) and phenylalanine 64 to a leucine (F64L), an “enhanced” GFP (EGFP, gi:27372525) was produced with the excitation maximum shifted from ultraviolet to blue and with better folding efficiency in *E. coli* [25]. Blue excitation is favorable because it matches up with the wavelengths of laser light used in modern cell sorting machines. Three rounds of DNA shuffling produced a mutant of GFP termed “cycle3” or GFPuv (gi:1490533) which contains three point mutations at or near the surface of the protein (F100S, M154T, V164A). This mutant has 16- to 18-fold brighter fluorescence than wild type GFP, attributed to a reduction of surface hydrophobicity and, subsequently, aggregation *in vivo* which prevents chromophore maturation [6]. Combining these sets of mutations produces a “folding reporter” GFP (gi: 83754214) which is monomeric and highly fluorescent [26], but does not fold and fluoresce strongly when fused to other poorly folded proteins. Four rounds of DNA shuffling starting with this GFP variant produced a mutant with six additional mutations, called “superfolder” GFP (gi: 391871871), which can fold even when fused to a poorly folding protein [27]. Superfolder GFP also showed increased resistance to chemical denaturation and faster refolding kinetics. This GFP variant also has exceptional tolerance to circular permutation compared to the “folding reporter” mutant of GFP (circular permutation will be discussed in Sequential rearrangements and truncations). A common theme emerges from these sets of mutations: a reduction in surface hydrophobicity leads to reduced aggregation tendency, which increases the fraction of chromophore able to mature and, consequently, the brightness of the protein *in vivo*. The hydrophobicity of the wild type GFP is hypothesized to serve as a binding site to aequorin in jellyfish [4].

Mutating surface polar residues to increase the net charge, called “supercharging”, may be one solution to the problem of aggregation. Armed with the knowledge that the net surface charge does not often affect protein folding or activity, [28] demonstrated that mutating the surface residues either to majority positive or to majority negative side chains does not significantly affect fluorescence. Furthermore, these “supercharged” variants of GFP showed increased resistance to both thermally and chemically-induced aggregation with a minimal decrease in thermal stability. The only side effects are the unwanted binding of positively supercharged GFP to DNA, and the formation of a fluorescent precipitate when oppositely supercharged variants are mixed.

Disulfide bonds have been known to confer additional stability to proteins. Two externally-placed disulfides were engineered into cycle3 GFP, one predicted to have no effect on stability, the other predicted to have a stabilizing effect [29]. The predictions, based on estimations of local disorder, were correct. Adding a disulfide where the chain is more disordered improved stability the most.

In recent, unpublished work in our lab [30], a faster-folding GFP has been made by eliminating a conserved *cis*-peptide bond. The slowest phase of folding of superfolder GFP has been known to be related to *cis*/*trans* isomerization of a peptide bond preceding a proline [5]. We targeted Pro89 for mutation, since the peptide bond is *cis* at that position in the crystal structure, but modeling studies suggested that a simple point mutation would not have worked. Instead, we added two residues creating a longer loop, and then selected new side chains for four residues based on modeling. The new variant, called “all-*trans*” or AT-GFP, folds faster, lacking the slow phase. A 2.7Å crystal structure, in progress, shows clearly that the backbone is indeed composed of all *trans* peptide bonds in the new loop region.

All of the variants discussed so far are derived from *Aequorea* GFP, but homologous fluorescent proteins from other species have also played a role in advancing the science. Rational design of a homologous GFP from the marine arthropod *Pontellina plumata* resulted in “TurboGFP” which folds and matures much faster than EGFP with reduced *in vitro* aggregation relative to its parent protein [31]. TurboGFP and its parent protein lack *cis*-peptide bonds, known to contribute to the slow phase of GFP folding [5]. The crystal structure of TurboGFP reveals a pore to the chromophore, which mutagenesis shows to be a key component to fast maturation [31]. This makes sense, since the diffusion of molecular oxygen into the core is the rate limiting step in chromophore formation. This result represents the first successful designed improvements to a non-Cnidarian fluorescent protein. Random directed mutagenesis of beta strands 7 and 8 in the cyan fluorescent protein derivative mCerulean produced a mutant with six mutations and a T65S reversion mutation in the chromophore. This construct, termed mCerulean3, has an increased quantum yield and demonstrates minimal photobleaching and photoswitching effects, making it a better FRET donor molecule [20].

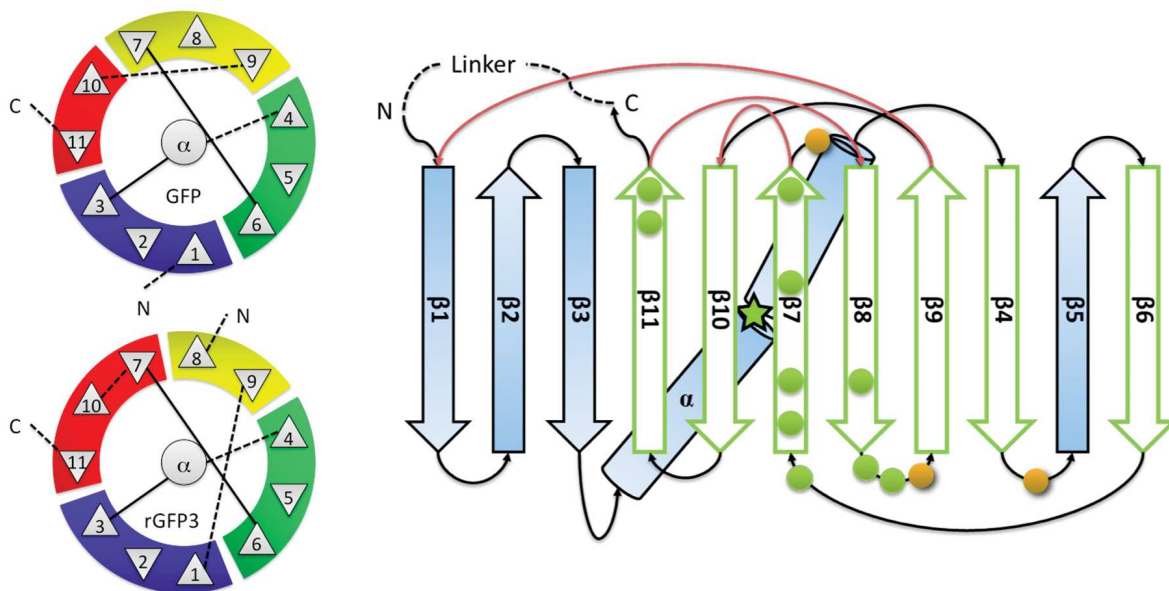
A novel fluorescent protein was developed using the consensus engineering approach, synthesizing a consensus sequence gene from 31 homologs of the monomeric Azami green protein, a distant homolog of *Aequorea* GFP. The resulting protein CGP (consensus green protein) has comparable expression to the parent protein with increased brightness and slightly decreased stability [32]. A novel directed evolution process was then carried out on CGP to stabilize it by inserting destabilizing loops into the protein, then evolving it to tolerate the insertions, then removing the destabilizing loops. After three rounds of this process, a mutant called eCGP123 demonstrated exceptional thermal stability compared to CGP and the parent Azami green protein [33]. Distantly-related fluorescent proteins have contributed much to the structural and biophysical understanding and application of the larger family.

## 1.6. Sequential rearrangements and truncations

Circular permutation is the repositioning of the N and C-termini of the protein to different regions of the sequence, connecting the original termini with a flexible peptide linker to produce a continuous, shuffled polypeptide. Many proteins retain their structure and function after permutation, provided the permutation site is not disruptive to secondary structural elements. This process demonstrates the tolerance of the protein's overall structure to significant rearrangements of primary sequence [34], enabling the design of biosensors based on split GFPs as discussed later.



GFP's rigid structure, extreme stability and unique post-translational chromophore formation reaction do not seem to suggest that it would tolerate circular permutation, and for the most part, it does not. All permutations that disrupt beta strands do not form the chromophore, and about half of the permutations in loop regions cannot form the chromophore. However, one particular permutation, starting the protein at position 145 (just before beta strand 7) expresses and fluoresces well, although it is less stable and less bright than the wild type GFP [34]. This circular permutation can also tolerate protein fusions to its new termini (positions 145 and 144 in wild type numbering), and position 145 in the wild type can accept a full protein insertion, such as calmodulin or a zinc finger binding domain [35]. The “superfolder” GFP reported in [27] was able to fluoresce after 13 of the 14 possible circular permutations, whereas the folding-reporter GFP only tolerated 3 of those 14 permutations. Figure 5 summarizes permutation and loop insertion results.



**Figure 5.** a) The wild type GFP, and (b) rewired GFP topology as drawn using the TOPS conventions [37]. Solid lines are connections at the top, dashed lines at the bottom of the barrel. (c) Green dots mark locations of the termini in viable circular permuted proteins. Orange dots mark places where long insertions have been made [38]. Green arrows mark beta strands that can be left out and added back to reconstitute fluorescence. Red lines are connections created in rGFP3, rewired GFP [36]. Topological changes and truncations are the least tolerated in the N-terminal 6 beta strands.

Circular rearrangements preserve the overall “ordering” of the secondary structural elements; however, non-circular rearrangement of the secondary structural elements is also possible. Using rational computational modeling and knowledge about GFP's folding pathway, [36] designed a “rewired” GFP with identical fluorescence properties and stability as a variant of superfolder GFP, but with the beta strands connected in a different order. These experiments demonstrate the selective robustness of GFP's structure to large-scale rearrangements in sequence, which has implications for deciphering the GFP folding pathway, as well as for design of split-GFP biosensors.

### 1.7. "Leave-One-Out" GFP

GFP can also be engineered to omit one of its secondary structural elements, either at one end or in the middle of the sequence by truncating a circular permutant. Truncation may be accomplished either at the genetic level or at the protein level, the latter by using proteolysis and gel filtration. Constructs missing one secondary structure element have been named "Leave-One-Out" or LOO, borrowing the term from a method for statistical cross-validation. When synthesized directly via the genetic approach, LOO-GFPs are non-fluorescent or weakly fluorescent. However, if co-expressed with the omitted piece, fluorescence sometimes develops *in vivo*, depending on which of the secondary structure elements was left out [39,40]. Expressing the full-length GFP and removing the beta strand by proteolysis, denaturation and gel filtration produces similar results [41]. A complete beta barrel is necessary for chromophore maturation. Once the chromophore has matured, LOO-GFP develops fluorescence rapidly upon introduction of the omitted beta strand from an external source.

That Leave-One-Out works is non-intuitive. In general, protein folding is an all-or-none process and leaving out any whole secondary structure element leads to an unfolded protein which aggregates in the cell. Yet, [40] has shown that it is possible to reconstitute LOO-GFP after truncation at several positions in the sequence. The key to understanding why LOO is sometimes possible is in the protein folding pathway. Although folding appears to be an all-or-none process by most experimental metrics, it proceeds along a loosely defined sequence of nucleation and condensation events called a folding pathway [42]. If the sequence segment that is removed is in the part of the protein that folds last, then a kinetic intermediate exists whose structure closely resembles the native state with one piece removed. This intermediate need not be the lowest energy state and may not be visible by equilibrium measurements, but its minute presence diminishes the energetic barrier of folding enough that the addition of a peptide can push the protein to the folded state. In short, Leave-One-Out uses the idea that some cyclically permuted, truncated proteins are natural sensors of the part left out.

*In vivo* solubility experiments performed on twelve LOO-GFPs (individually omitting each of 11 beta strands and the alpha helix) showed that there are significant differences in tolerance to the removal of particular secondary structural elements (SSE) as a function of solubility. The variability is best explained in terms of the order of folding of the SSEs. SSEs that are required for the early steps in folding leave a more completely unfolded polypeptide behind when they are left out. SSEs that fold late and not required for most of the folding pathway, leaving behind a mostly-folded protein which is more soluble. Leave-One-Out solubility analysis provides a unique insight into the folding pathway of GFP [40]. Omitting strand 7 (LOO7-GFP) appears to be the least detrimental to the overall structure of GFP, suggesting that strand 7 folds last. Binding kinetics data for LOO7-GFP to its missing beta strand as a synthetic peptide gives a  $K_d$  value of roughly  $0.5 \mu\text{M}$  [11]. Surprisingly, when it is omitted by circular permutation and proteolysis, the central alpha helix can be reintroduced as a synthetic peptide to the "hollow" GFP barrel and chromophore maturation proceeds and produces fluorescence [41]. However, refolding from the denatured state was required.

Some LOO-GFPs also show interesting reactions to ambient light. LOO11-GFP (beta strand 11 omitted) does not bind strand 11 when kept completely in the dark, but does bind it upon irradiation with light [43]. Raman spectroscopy showed that, in the dark, the chromophore assumes a *trans* conformation, and that light induces a switch to the native *cis* conformation. After irradiation, the chromophore relaxes back to the *trans* conformation. Following up on this result, [44] showed that using a circularly permuted LOO10-GFP construct (beta strand 10 omitted) and introducing two synthetic forms of strand 10, the wild-type strand and a strand with a mutation to cause yellow-shifted fluorescence, light irradiation increased the frequency of “peptide exchange” between the two strand 10 forms. The presence of this peptide exchange suggests that the *cis/trans* isomerization of the chromophore requires partial unfolding of the protein.

## 2. GFP-based biomarkers

The term biomarker has accumulated a variety of definitions over the years. Herein, biomarkers are defined as genetically encoded molecular indicators of state that are linked to specific genes. The utility of GFP as a biomarker was first demonstrated using GFP reporter constructs [45]. When GFP is used as a transcription reporter, a cellular promoter drives expression of the fluorescent protein resulting in fluorescent signal that temporally and locally reflects expression from the promoter *in vivo*. In the initial experiments, GFP cDNA [46] was expressed from the T7 promoter in *E. coli* or from the *mec-7* (beta tubulin) promoter in *C. elegans* [45]. *E. coli* cells fluoresced and the expression in *C. elegans* mirrored the pattern known from antibody detection of the native protein. Subsequently, GFP transcriptional reporters have been used in a wide variety of organisms; GFP expression has minimal effect on the cells and can be monitored noninvasively using techniques such as fluorescence microscopy and fluorescence assisted cell sorting (reviewed in [47]).

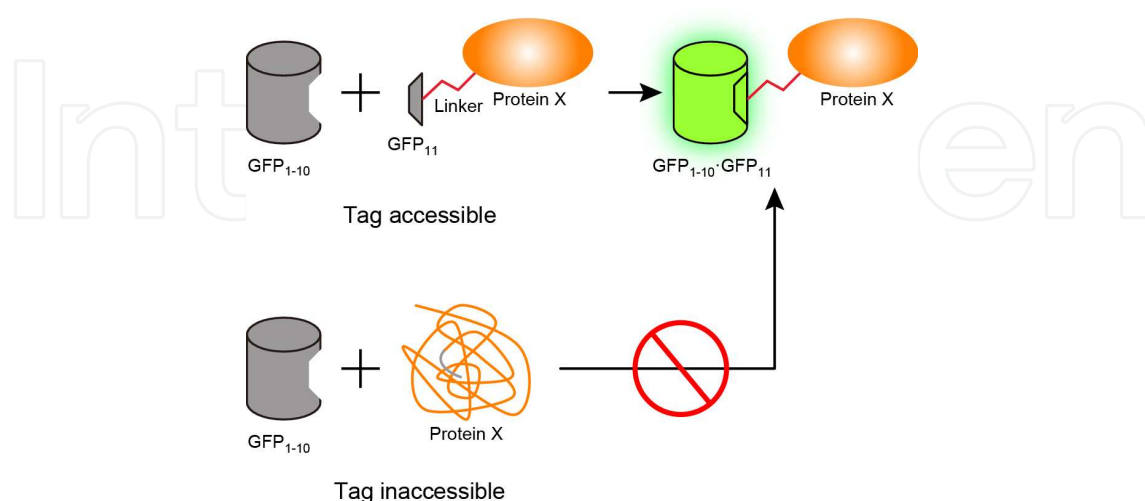
GFP fusion proteins (generated by combining the fluorescent protein coding region with the coding region of the cellular protein) are used as markers for visualization of intracellular protein tracking and interactions (reviewed in [47-49]). The GFP moiety may be N-terminal, C-terminal or even internal to the cellular protein. The availability of a color palette of fluorescent proteins allows multicolor imaging of distinct fluorescent protein fusions in the cellular environment. GFP fusion proteins are a major component of the molecular toolkit in cell biology.

### 2.1. Using GFP as an *in vivo* solubility marker

GFP has been used as a genetically encoded reporter for folding of expressed proteins. Expression of recombinant proteins in *E. coli* is a powerful tool for obtaining large quantities of purified protein; however, some overexpressed recombinant proteins improperly fold and aggregate. Manipulation of conditions to generate soluble protein can be a laborious process. Directed evolution can be employed to increase the solubility of the recombinant proteins, but detection of specific mutants with improved solubility is a challenge. However,

GFP biomarking can be utilized to address this challenge. Since GFP chromophore formation requires proper protein folding and GFP folds poorly when fused to misfolded proteins, fluorescence of a GFP fusion protein can serve as an internal signal of a specific soluble (not aggregated) protein [26]. When used as a folding reporter, GFP is fused C-terminally to the protein of interest using a short linker between the two protein domains. Detection of fluorescence indicates that GFP domain is properly folded and that the protein of interest therefore must be soluble. If the protein of interest misfolds and aggregates, the fused slow-folding GFP aggregates along with it and fluorescence is not detected. Therefore, this folding reporter assay can be used as a screening tool for soluble recombinant proteins in the context of directed evolution.

Split GFP may also be used to assay folding and solubility of a protein of interest *in vivo* by “tagging” the recombinant protein with the smaller portion of the split GFP sequence, and expressing the larger portion separately or adding it exogenously. The small size of a protein tag makes it less likely to interfere with the folding and function of the protein of interest. In the split GFP complementation assay a large fragment of GFP folding reporter (GFP<sub>1-10</sub>) is coexpressed with tagged GFP protein (GFP<sub>11</sub>-protein x) [50]. As shown in Figure 6, neither GFP<sub>1-10</sub> nor the GFP<sub>11</sub>-tagged protein fluoresce alone; however, if both components are soluble, GFP<sub>1-10</sub> and the GFP<sub>11</sub>-tagged protein reconstitute the native structure and fluorescence. For successful implementation of the assay, directed evolution of superfolder GFP<sub>1-10</sub> was required. This resulted in GFP<sub>1-10</sub> OPT which has an 80% increased solubility over the corresponding superfolder GFP<sub>1-10</sub>. GFP OPT contains 7 new mutations (N39I, T105K, E111V, I128T, K166T, I167V and S205T) in addition to the superfolder mutations [50]. Directed evolution of GFP<sub>11</sub> resulted in GFP<sub>11</sub> optima tag that had the dual properties of 1) complementation with GFP<sub>1-10</sub> OPT and 2) minimized perturbation of the protein of interest. Note that full-length GFP OPT was subsequently found to be more tolerant of circular permutation and truncation than superfolder GFP [40].



**Figure 6.** Mechanism of LOO11-GFP (GFP<sub>1-10</sub>) as an *in vivo* solubility indicator for proteins tagged with strand 11 (GFP<sub>11</sub>). Modified with permission from [50].



In addition to providing a less laborious method for detecting protein variants and reaction conditions for generating soluble recombinant protein, the split GFP complementation assay also serves as an assay of aggregation in living cells. For example, aggregates of the microtubule associated protein tau are found in neurofibrillary tangles but their role in the pathology of Alzheimer disease and Parkinson disease is not clear [51]. The split GFP complementation assay enables monitoring of the aggregation process in living mammalian cells [52,53] and was validated using GFP<sub>1-10</sub> and GFP<sub>11</sub>-tau variants. Cells containing soluble tagged protein show visible fluorescence but aggregates have little or no fluorescence. Protein aggregates of GFP<sub>11</sub>-tau sequestered the GFP<sub>11</sub> tag, leading to decreased complementation of GFP<sub>1-10</sub> and decreased fluorescence. Thus the split GFP complementation assay using tagged-GFP tau showed that it could be used as an *in vivo* model for studying factors that influence aggregation.

## 2.2. GFP biomarkers for single molecule imaging

It is also possible to utilize GFP biomarkers for single-molecule localization, a form of super-resolution microscopy. High affinity single chain camelid antibodies (nanobodies) to GFP can be used to deliver organic fluorophores to GFP tagged proteins that are in turn used in single molecule “nanoscopy.” [54, 55]. This novel approach combines the molecular specificity of genetic tagging with the high photon yield of the organic dyes. Additionally, by varying the buffer conditions used, many organic dyes can become photoswitchable. The small size of camelid antibodies and their high affinity allow for access to regions that are generally inaccessible to conventional antibodies and targets that are expressed at very low levels [56].

One should caution that the overexpression of FRET biomarkers in transgenic animals carries some concerns that this could lead to the perturbation of endogenous signaling pathways and even retardation of animal development [57]. Additionally, in compact tissue, such as the brain tissue, cell type identification is particularly tedious due the diffused expression of the biomarkers.

## 3. GFP biosensors

Biosensors are distinct from biomarkers in that they are not linked to the expression of a specific gene product. Biosensors may function *in vivo* or *in vitro*. GFP variants that exhibit analyte-sensitive properties are genetically encoded biosensors, acting *in vivo*. GFP biosensors that contain amino acid substitutions that enable detection of pH changes, specific ions (Cl<sup>-</sup> or Ca<sup>2+</sup>), reactive oxygen species, redox state, and specific peptides have been reported [39, 58-60]. In addition, modifications have been reported that enable selective activation (irreversible or reversible) of the fluorescence [61,62]. Genetically encoded GFP biosensors may be single GFP domains or FRET pairs. In the following subsections we describe selected examples of GFP-based biosensors used *in vivo* or *in vitro*, with special emphasis on computationally designed biosensors.



### 3.1. *In vivo* pH biosensors

Within the cell, pH varies from the neutral pH of the cytosol to the acidic pH of the lysosome lumen and protons may serve as cellular signals. Genetically encoded pH biosensors enable subcellular detection of pH and can provide insight into the regulation of cellular activities by pH. Addition of intracellular targeting tag directs the pH biosensor to particular subcellular compartments.

Many GFP variants show sensitivity to pH which results from protonation and deprotonation of the chromophore (see **Maturation of the GFP Chromophore**) (reviewed in [58]). The rapid and reversible response of EGFP to pH changes in the cells enabled EGFP to be used as an intracellular pH indicator [63] in place of chemical pH indicators such as fluorescein. A range of GFP based pH biosensors have been generated from modification of wtGFP and EGFP which resulted from amino acid substitutions primarily in and around the region of the chromophore.

Two classes of GFP pH indicators have been described: ratiometric and nonratiometric [58, 64]. In the ratiometric pH indicators, the chromophore environment is such that the GFP biosensor has two sets of excitation/ emission spectra, one that varies with pH and another that does not. For these GFP variants, a calibration curve can be generated for the ratio of the spectra versus the pH. Nonratiometric GFP variants, such as EGFP [63] or ecliptic GFP [64], have pH dependent emission from the anionic chromophore (deprotonated) but almost no fluorescence of the neutral chromophore (protonated). These variants are used for reporting pH changes within cells when used as single molecule pH sensor or used in tandem with pH insensitive fluorescent partner (described below).

Ratiometric GFP pH biosensors have been generated by modification of a few key amino acids in the vicinity of the chromophore. Ratiometric pHluorin (RaGFP), the first ratiometric GFP described, contains a key S202H mutation and shows pH dependent change in excitation ratio between pH 5.5 and pH 7.5 [64]. The S202H mutation was shown to be important for the ratiometric property; pHluorins lacking the S202H were non ratiometric. Another class of GFP ratiometric pH sensors, deGFPs were generated from mutagenesis of the S65T GFP variant [65] resulting in substitutions H148G (deGFP1) or H148C (deGFP4) and T203C. The deGFPs are dual emission ratiometric GFPs emitting both blue and green light; blue light emission decreases with increase pH while green light emission increases with increased pH.

Variants pH GFP (H148D) [66] and E<sup>2</sup>GFP (F64L/S65T/T203Y/L231H) [67] function as dual excitation ratiometric pH indicators with pH-dependent excitation at 488 nm and relatively pH-independent excitation at 458 nm. In addition to its pH sensing properties, fluorescence emission from E<sup>2</sup>GFP is affected by the concentration of certain ions, including Cl<sup>-</sup>. The chloride ion sensitivity of E<sup>2</sup>GFP is a key component of the GFP-based chloride ion and pH sensor ClopHensor [68] (discussed in section **Fluorescent proteins as intrinsic ion sensors**).

In addition to single molecule based pH biosensors, ratiometric pH biosensors using tandem fluorescent protein variants have been constructed in which a pH sensitive GFP variant is linked to a less sensitive or pH insensitive GFP. GFP<sub>H</sub> and YFP<sub>H</sub> are tandem FRET pairs for

the detection of pH changes in the cytosol and nucleus of living cells. GfpH combines GFPuv, which has low pH sensitivity, with pH sensitive EGFP and YfpH combines GFPuv and EYFP [58, 69]. Not all tandem GFP biosensors are FRET pairs, however. pHusion is a ratiometric tandem GFP biosensor in which mRFP (pH insensitive) is tethered to EGFP (pH sensitive) via a linker. pH measurements are determined from the ratio of EGFP to mRFP fluorescence. pHusion biosensor was developed for analysis of intracellular and extracellular pH in developing plants [60].

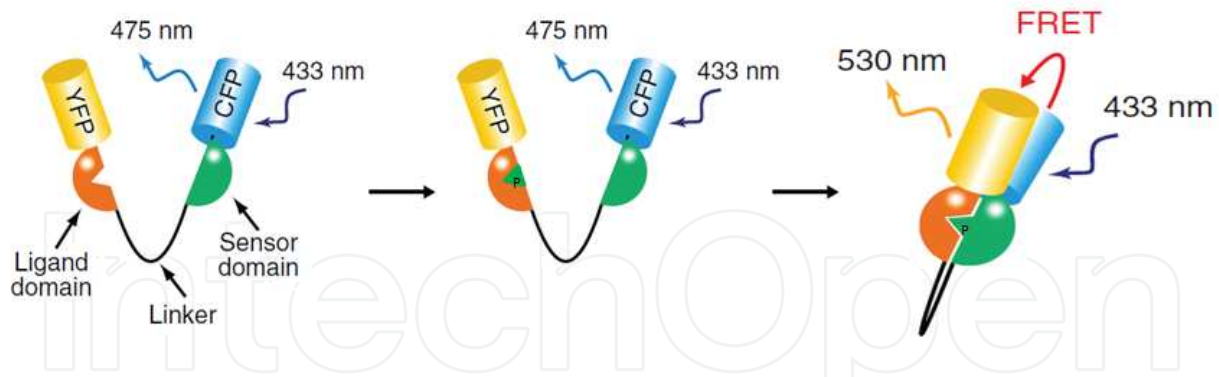
### 3.2. *In vivo* FRET-based biosensors

Genetically-encoded FRET-based biosensors can be applied in a variety of capacities to visualize intracellular spatiotemporal changes in real time. The evolution of these applications has progressed from cell culture systems that transiently express FRET biosensors to transgenic mouse models that express them in a heritable manner [57]. Production of transgenic mice with FRET biosensors arose in an effort to enhance our understanding of the differences that exist between tissue culture and living systems. Transgenic FRET GFP biosensor systems are very efficient and their fluorescence signals are easily distinguished from autofluorescence, which is analyte-independent fluorescence. The sensors themselves can be used to probe a variety of pathways for the activity of signaling enzymes as well as a number of post translational modifications.

#### 3.2.1. Detection of enzyme activity

In transgenic animal models, FRET biosensors can be used to study PKA activation by cAMP, ERK activation by TPA and their association with various physiological changes [57]. PKA and ERK are enzymes that transfer the  $\gamma$ -phosphate of ATP to a number of protein substrates thereby affecting a conformational change. Kinase induced conformational changes are important because they are involved in the control a number of critical cellular processes that include glycogen synthesis, hormonal response, and ion transport [70]. A number of signaling cascades that involve kinases require a means of dynamic control and spatial compartmentalization of the kinase activity; a requirement highlights the need for a mechanism to continuously track kinase activity in different compartments and signaling microdomains *in vivo*.

Traditional methods of assaying kinase activity fail to capture its dynamicity; a void that is filled by genetically encoded FRET-based biosensors. These sensors are constructed so that the substrate protein of the kinase of interest is flanked with a fluorescent protein pair in such a way that the conformational change imparted by phosphorylation translates into a change in the FRET signal (Figure 7) [70]. These biosensors can be localized to particular sites of interest with the aid of appropriate targeting signal sequences, allowing the imaging of site-specific kinase activity. G-protein coupled receptors, when used in a biosensor, provide a mechanism for transducing drug mediated effects on PKA activity into a light signal. Transgenic mice expressing FRET based biosensors provide an ideal system for studying the pharmacodynamics of these drugs.



**Figure 7.** Representation of the mode of action of an intramolecular FRET biosensor containing a molecular switch. The sensor domain and ligand domain of the construct are connected by a flexible linker with CFP and YFP serving as the donor and acceptor for the FRET pair. This switch can perceive various molecular events, such as protein phosphorylation, through binding to the ligand domain. This in turn induces an interaction between the ligand and sensor domains that facilitates a global change in the conformation of the biosensor, which serves to increase the FRET efficiency from the donor to the acceptor (CFP to YFP in this case) [71].

When used to study the signaling events in wound healing, the strength and duration of the fluorescent signals that are generated by these biosensors are dependent on the location within the tissue (tissue depth has a negative impact on the intensity of the fluorescent signal), its vicinity in relation to the site of injury, as well as the contributions made by chemical mediators (drugs) in sustaining kinase activity [57]. These model systems provide a means of visualizing in real-time the agonist/antagonist pharmacodynamics associated with a plethora of signaling molecules that do not necessarily have to be limited to PKA and ERK activity. They also provide a tool for resolving the maze of upstream signaling pathways that contribute to chemotaxis in the animals.

Genetically encodable FRET GFP biosensors have proven to be useful in characterizing the dynamic phosphorylation dependent regulation of small GTPases [70]. Ras GTPases play essential roles in regulating cell growth, cell differentiation, cell migration, and lipid vesicle trafficking. Upon binding GTP, the G-protein Ras recruits the serine/threonine kinase Raf. FRET biosensors for GTPase activity such as Raichu-Ras (Ras and Interacting protein Chimeric Unit for RAS) use this Ras-Raf interaction as the basis for the molecular switch. Raichu-Ras functions by using H-Ras as the sensor domain and the Ras Binding Domain (RBD) of Raf as the ligand domain in constructing a molecular switch that in turn is sandwiched by the FRET pair CFP/YFP (Figure 7). Such a design allows for the monitoring of Ras activation in living cells on the basis of fluctuations in the FRET signals generated.

### 3.2.2. Detection of antioxidant activity and reactive oxygen species

FRET-based GFP biosensors can also be employed in *in vitro* applications as an alternative tool for high throughput screening assays. These assays are simple, inexpensive, reproducible and highly specific. A good example can be observed in the use of bacterial cell-based assays for screening antioxidant activity of various substances for biological activity [72]. To achieve this objective *E.coli* biosensor strains that carry the plasmid that fuses *sodA* (manga-

nese superoxide dismutase) and *fumC* (fumarase C) promoters with GFP genes, called *sodA::gfp* and *fumC::gfp* respectively, were produced and used to evaluate antioxidant activity of a number of phenolic and flavonoid compounds in comparison with two DPPH radical scavenging and SOD activity assays (two more conventional assays). After paraquat treatment of *E. coli* cultures to induce oxidative stress, the putative antioxidant compounds were added and both the GFP fluorescence and cell culture density readings were taken to determine the role played by the respective compounds in reducing the free radical accumulation and intracellular oxidative stress. Genes *sodA* and *fumC* are turned on by SoxR and OxyR, respectively, which are the two main regulatory proteins involved in oxidative stress sensing. GFP fluorescence is therefore diminished by successful antioxidants. These constructs are important because they function as alternative screening tools that can be utilized to assess the activity of compounds with therapeutic potential against oxidative stress. Antioxidants have been shown to play a role in disease prevention.

### 3.2.3. Detection of calcium ions

FRET-based and single domain  $\text{Ca}^{2+}$  sensors have been constructed using the allosteric effect of calcium binding to receptors calmodulin or troponin [73]. In one construct, the CFP/YFP pairing is separated by a linker containing a calmodulin domain and a calmodulin ligand peptide called M13. When  $\text{Ca}^{2+}$  is present, it binds to the calmodulin domain, inducing a conformational change and binding of the proximal M13 peptide sequence. The M13 binding results in shortening of the linker, bringing CFP within FRET distance of YFP and changing the emission wavelength from cyan to yellow. The  $\text{Ca}^{2+}$  binding affinity was found to be highly variable, around 0.3  $\mu\text{M}$  with a Hill coefficient of  $n=4$ , depending on conditions. When used *in vivo*, the calmodulin-based biosensors suffered from endogenous interference by host proteins and did not always work [73]. To remedy this, the calmodulin/M13 linker was replaced with troponin C, whose N-to-C distance is shortened by  $\text{Ca}^{2+}$  binding, resulting in FRET. Using another strategy, calmodulin and M13 peptide sequences were separated by a circularly-permuted EGFP, which was quenched in the absence of  $\text{Ca}^{2+}$  but recovered fluorescence upon  $\text{Ca}^{2+}$ -induced binding of the calmodulin to M13. Improved genetically encoded  $\text{Ca}^{2+}$  indicators have been used *in vivo* to trace action potentials in neurons, with response times in the millisecond range [73, 74], becoming competitive with synthetic indicators and recording electrodes.

## 4. *In vitro* applications

GFP has great potential to work as an *in vitro* biosensor. Because of its remarkable stability, it can be used and manipulated in multiple ways to impart sensor functionality to the protein. Several approaches are described here, including creating a chimeric protein with antibody fragments, linking fluorescent proteins to quantum dots, manipulating the amino acid sequence to create analyte pores, as well as sequence manipulation that provides increased halide ion and/or pH sensitivity.



#### 4.1. GFP-antibody chimeric proteins

The goal of GFP-antibody chimeric proteins (GFPAbs) is to convert a multi-step experimental process for locating molecules via antibodies and enzyme-linked secondary antibodies, into a one-step process using a GFPAbs. This molecule could then work as a detection reagent in flow cytometry, for intracellular targeting, or fluorescence-based ELISAs [38]. However, in order to replace antibodies in these techniques, it is important to achieve the same nanomolar sensitivity that is found in the natural antibodies. To do this, [38] inserted two antigen-binding loops into the GFP structure, counting on cooperativity in binding to enhance affinity.

It became clear that adding loops impinges on the integrity of the native GFP structure. The binding loops must be placed such that their presence in the fluorescent protein does not jeopardize its structural fidelity, or that of the chromophore. There are only a few locations in the molecule that are amenable to such insertions: turn regions  $\beta 4/\beta 5$  (residue 102),  $\beta 7/\beta 8$  (residue 172) and  $\beta 8/\beta 9$  (residue 157). The latter two are too far apart in three-dimensional space to provide for cooperative binding (see Figure 5). The  $\beta 4/\beta 5$  and  $\beta 8/\beta 9$  loop regions are in close proximity, but these do not easily accommodate random loop insertions.

[38] used directed evolution with yeast surface display [75] to find sequences that stabilized the folded conformation in the context of loop insertions. The yeast secretory pathway does not allow unfolded protein to reach the surface of the cell, thus only mutants that yield fully folded GFP were displayed by yeast cells. Directed evolution revealed several mutations that conferred additional stability and increased fluorescence in the context of inserted loops: D19N, F64L, A87T, Y39H, V163A, L221V, and N105T. The F64L mutation has been shown to increase fluorescence of GFP and also to shift the excitation maximum to 488 nm. Y39H and N105T have been shown to improve refolding kinetics and refolding stability, respectively. V163A is linked to improved folding as a result of its increased expression in yeast surface display [38]. These mutations accommodated the insertions of antigen-binding loops from antibodies raised against streptavidin-phycoerythrin, biotin-phycoerythrin, TrkB, or GADPH, all while maintaining 40% of the fluorescence and 60% of the expression of wild type GFP. With dual loop insertion, dissociation constants as low as 3.2 nM have been achieved [38]. The success of this construct means that molecules such as GADPH can be located within cells without having to engineer a second round of antibodies, saving both time and resources.

#### 4.2. A chimeric fluorescent biosensor based on allostery

A general method for developing a biosensor for a specific receptor-ligand interaction has been described [76] in which a receptor protein is inserted into the GFP sequence between strand 8 and strand 9. The insertion puts enough of a strain on GFP that its fluorescence is reduced. Binding of the ligand to the GFP-receptor chimera may then impart enough of a change in its conformation that it causes a change in fluorescence, since the  $\beta 8/\beta 9$  loop is fairly close in space to the chromophore. This change may be found by plate screening for fluorescence. In [76], the receptor Bla1 was cloned into the loop, and random mutations were made to this construct. Mutant constructs that detected the Bla1 ligand BLIP were

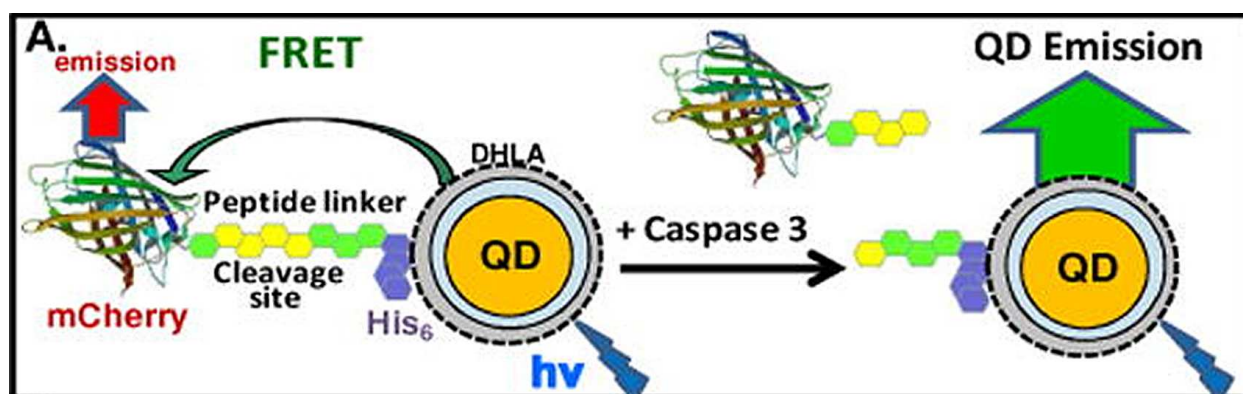


identified by a visual screen of colonies before and after the induced expression of BLIP. Using this method, a double mutant was found that was shown to detect BLIP *in vitro* with micromolar affinity. In principle, this method could be used to generate a sensor for any ligand that can be expressed in bacteria or added exogenously, as long as a receptor protein exists that can be inserted into the GFP loop.

#### 4.3. FRET-based biosensors using quantum dots

FRET-based *in vitro* biosensors may be constructed by linking fluorescent proteins to quantum dots (QDs). QDs are inorganic molecular nano-crystals whose absorption and emission spectra are dictated by the size of the QD. For example, a QD may be engineered to absorb ultraviolet light and emit light at 550 nm, which overlaps well with the excitation spectrum of mCherry, a variant of GFP [77], and produces FRET when the two fluorophores are in close proximity.

In order to make the FRET emission analyte-dependent, the QD was linked to the mCherry via an N-terminal linker peptide that contained a protease cleavage site and a 6 histidine tag. The imidazole side chains of the histidines electronically coordinate with the zinc atoms of the CdSe–ZnS core-shell semiconductor of the QD [77]. Multiple mCherry molecules can be coordinated with each QD. Splitting of mCherry from the QD by a protease may be detected by the loss of FRET. By placing the caspase-3 cleavage sequence into the linker between GFP and the QD, the FRET complex becomes a biosensor for the presence of caspase-3, glowing red at 610 nm in the absence of the protease, and reverting to the yellow fluorescence of the QD at 550 nm when the protease is present (Figure 8).



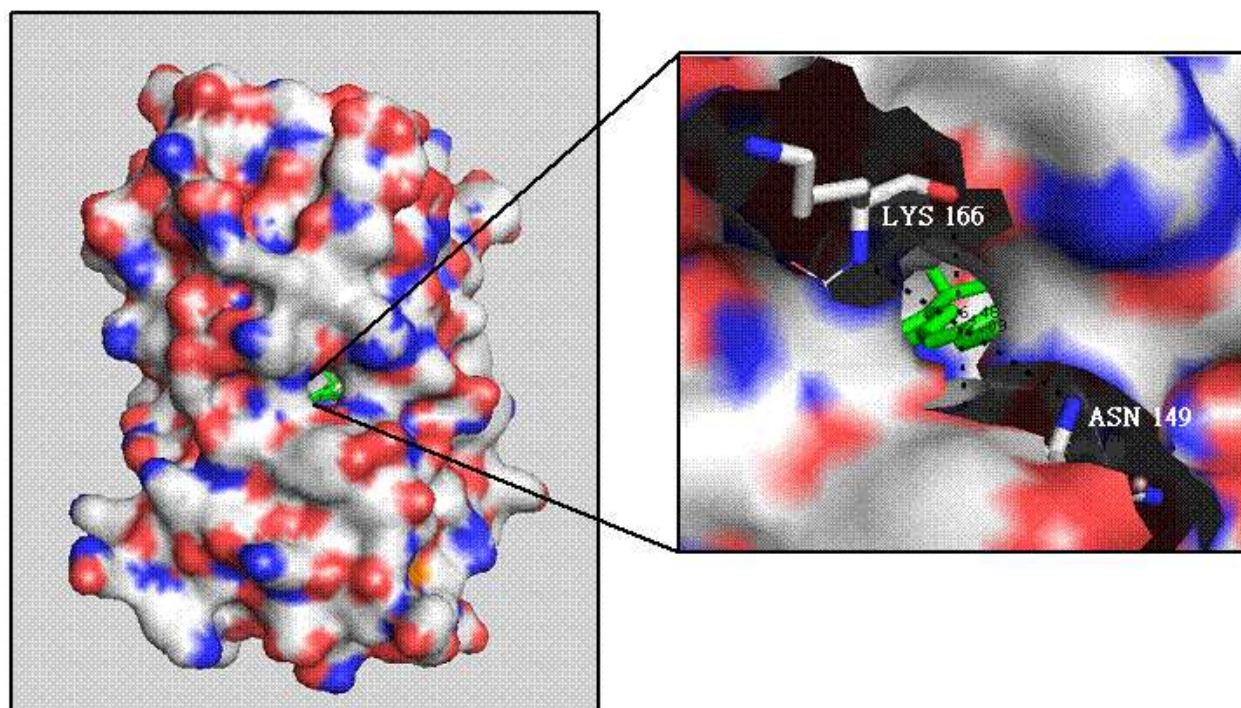
**Figure 8.** QD-FRET, showing emission of the chromophore only when in close proximity to the QD. When the two are split by caspase activity, FRET is lost. Figure used with permission from [78].

GFP/QD FRET emission may be also be manipulated by pH-induced changes in the spectral overlap, without having to spatially separate the QD from the fluorescent protein. It has been shown that fluorescent proteins such as GFP and mOrange experience a shift in excitation and emission spectra with changes in pH [78]. At a slightly acidic pH, there is very little spectral overlap between the QD emission and the mOrange excitation, which means that the QD emission is seen, in this case around 520 nm. However, as the pH increases, the exci-

tation spectrum of mOrange shifts such that there is more overlap with the QD emission, which subsequently causes an increase in FRET. The result is an upward shift in the emission wavelength with increasing pH. It is important to note that since there is a fluctuating hydrogen ion concentration, the histidine-QD coordination complex becomes unstable. In order to remedy this problem, a covalently linked quantum dot must be used.

#### 4.4. Fluorescent proteins as intrinsic ion sensors

Fluorescent proteins, especially E<sup>2</sup>GFP, have been shown to be sensitive not only to pH changes but also to the concentration of certain ions, particularly chloride ions. E<sup>2</sup>GFP provides an avenue for single domain ratiometric analysis of pH because it contains two excitation and emission peaks. Only the longer wavelength emission peak is pH dependent [68]. Therefore analysis of pH based on the ratio of green fluorescence to cyan. By coupling E<sup>2</sup>GFP to another fluorescent protein in a fusion construct, it is also possible to measure other intracellular chloride ion concentration. For example, DsRed is neither pH nor chloride ion sensitive, so it can be used to measure chloride ion concentration based on the ratio of its fluorescence to the cyan emission of E<sup>2</sup>GFP.



**Figure 9.** The analyte channel through which copper ions can pass through to the interior of the barrel structure and quench the fluorescence of the chromophore. Used with permission from [80].

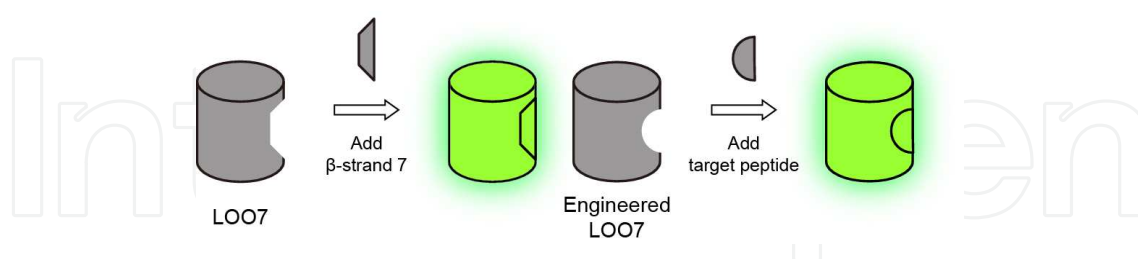
Making a few modifications can make GFP sensitive to the concentration of other ions. For example, superfolder GFP can be made sensitive to copper ions by mutating the arginine at position 146 to a histidine, which, as previously mentioned, coordinates well with metal ions [79]. GFP can also become sensitive to ions by creating channels in the structure through

which small molecules can pass through and access the chromophore (Figure 9). By mutating position 165 from a phenylalanine to a glycine, a channel is opened that is about 4 Å wide. This allows small molecules such as copper ions to enter the hydrophobic core of the protein and quench fluorescence [80]. GFP, thanks to its stability, has shown a remarkable ability to be modified, and thus shows great promise in visualizing a large variety of intracellular and extracellular substances.

## 5. Computationally designed LOO-GFPs

Recent work in the Bystroff lab has focused on programming GFP to accept any desired protein as a binding partner, like an antibody, and to switch on fluorescence only when the targeted protein is bound. The strategy combines Leave-One-Out split reconstitution with computational design and high throughput screening.

Leave-One-Out (LOO) was described earlier (**“Leave-One-Out” GFP**) as a technique for developing split proteins that spontaneously reconstitute function. Fluorescence is recovered in LOO-GFP when the left-out piece is encountered in the analyte. A promising application of LOO-GFP, knowing that it binds to the left-out segment and fluoresces [39, 40], is to engineer novel LOO-GFP molecules that recognize and sense desired peptides derived from other sources such as virus, bacteria and parasites. By modifying the sites of one of the eleven  $\beta$ -strands to complement shapes of given target peptides, the engineered LOO-GFP molecules will report the presence of specific target proteins, and therefore their host organism, through simple fluorescence readout (Figure 10). LOO-GFP biosensors can be engineered by generating mutations that accommodate the shape and charge of a desired target peptide. The target peptide may be made available for binding by denaturing the target protein.



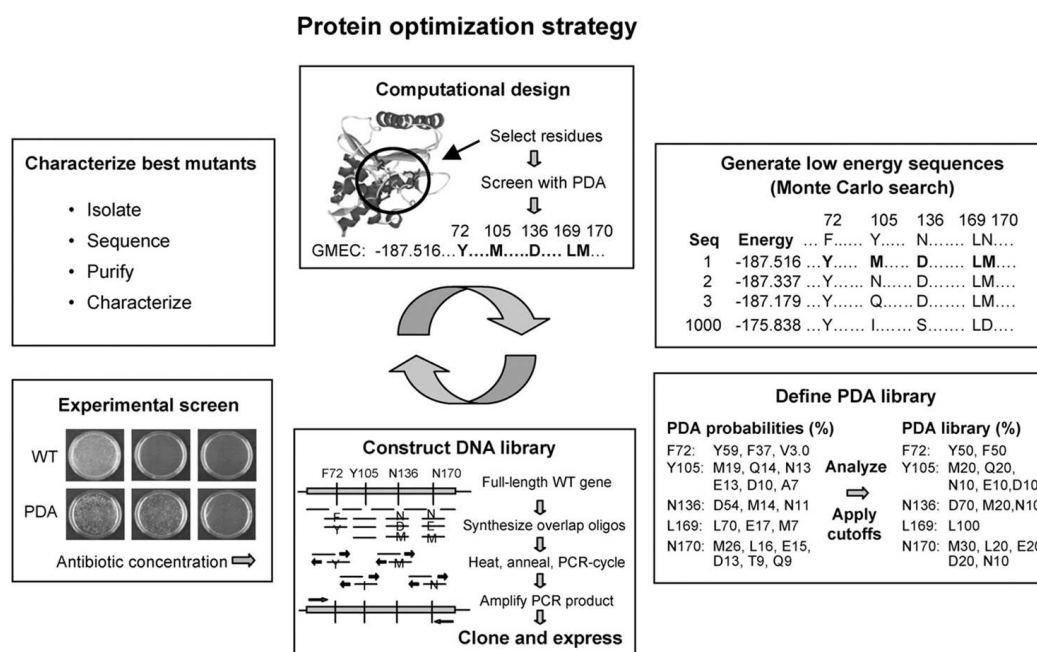
**Figure 10.** LOO-GFP peptide biosensors. Engineering LOO-GFP molecules to accommodate desired target peptides create specific sensing tools where fluorescence can be reconstituted upon adding back the left-out peptides and signals the detection.

Theoretically, this goal could be achieved by random mutation followed by high throughput screening to find mutants that glow in the presence of a peptide. However, random mutation would be extremely inefficient. Computational protein design methods offer a much better alternative for rationally generating sequence diversity before the labor-intensive experimental screen.



## 5.1. Computer-aided protein design

Computational protein design predicts protein sequences that fold into predefined protein structures. Proteins are described as a set of atoms with 3D spatial coordinates and physical/chemical properties [81-84]. Instead of mutating residues experimentally, mutations are explored *in silico* and selected using a computed goodness of fit (Figure 11). Mutations predicted to cause collisions between atoms, leave unsatisfied hydrogen bonding partners, cause charge-charge repulsion, or employ rare amino acid side chain conformations are down-weighted by assigning them a higher energy value. To facilitate the search for the best mutations, amino acid side chains are discretized into rotational isomers (called rotamers) [85-87]. Protein sequences that preserve the desired functionalities, such as the binding of a ligand, are obtained by searching the space of all side chain rotamers for the minimum free energy. There are few reviews of the methods used [88].



**Figure 11.** Computational protein design coupled with design library generation [89]. The entire designed sequence space of selected residues is computationally screened to determine the global minimum energy configuration (GMEC) for the given structure. Starting from the GMEC, sequence space is explored to obtain sub-optimal sequences that are also potentially predicted to be functional. A DNA library is constructed to cover all predicted sequences, and candidates are screened experimentally to select clones with desired functions. Information from analyzing obtained mutants is utilized to validate and improve the computational protein design strategy, and provides a better starting model for iterative optimization.

## 5.2. Protein biosensors versus other methods for detecting pathogens

Biosensors for specific proteins and pathogens offer potential advantages over the current state of the art, notably speed and simplicity. Laboratory diagnostics of infections commonly includes pathogen isolation using culture, direct antigen detection, or detection of pathogen specific DNA and/or RNA by polymerase chain reaction (PCR). The isolation method re-

quires a culture system to inoculate a specimen, followed by the examination of specific characteristics produced by pathogens, such as the cytopathogenic effect of virus and the distinct metabolism of bacteria. Although culture-based methods have higher detection sensitivity, they generally take three to ten days for diagnosis. Alternatively, immunoassays utilize pathogen specific antibodies and secondary anti-antibodies to detect and report a pathogen. Most of the rapid diagnostic tests only take 15 to 30 min for diagnosis, but raising specific antibodies against pathogens is time-consuming and expensive. Thirdly, molecular diagnosis using PCR takes the advantage of the gene amplification and provides a highly sensitive detection in diagnosis from minute amounts of pathogen genome within a short time. However, the need for real-time PCR and gel electrophoresis apparatus and reagents means it will not be possible in all settings, where a simple biosensor test would be possible. PCR assumes that DNA is present, but some pathogens such as anthrax toxin, snake venom and bovine spongiform encephalopathy contain no genetic material. All these point to a need for developing a diagnostic tool for proteins that is fast and easy to use, and suitable for rural, point-of-care facilities in developing nations.

The following describes how the computer-aided design of LOO-GFP was done, and the encouraging but preliminary results. The process has three steps: (1) the selection of a target peptide sequence from the genome of the pathogen, (2) the computational design of the LOO-GFP• target complex, and (3) the experimental screening of a library of potential biosensor sequences.

### 5.3. Target peptide selection

A target peptide for detection must be unique in order to avoid false positives, and must be conformable to the LOO-GFP binding site, which is the site of one of the eleven  $\beta$ -strands of GFP. From the examination of GFP and homolog fluorescent protein structures and sequences, we defined a set of signature patterns for each  $\beta$ -strand. These patterns define the limits of mutation. For example, no position within a target peptide may be a proline, since it must be hydrogen bonded on both sides to the neighboring  $\beta$ -strands. Cysteines are also disallowed, for experimental reasons. Target peptides are selected by searching the sequences of the target organism for a match to the signature pattern. Other considerations including the location of protease recognition sites, cellular location, and protein expression levels.

In the case study described here, a twelve-residue peptide (SSHEVSLGVSSA) was selected from hemagglutinin (HA) sequence of avian influenza virus H5N1, using the signature pattern of GFP  $\beta$ -strand 7. The target peptide retains the sequence pattern of the wild type  $\beta$ -strand 7, and it can be released by the chymotrypsin digestion of HA protein. A BLAST search of all known protein sequences confirmed that the HA target sequence occurs only in hemagglutinin from influenza virus type A.

### 5.4. Computational pre-screening of candidate biosensor sequences

To engineer customized LOO-GFP biosensors that sense a given peptide we developed a set of software called DEEdesign. DEEdesign uses a combination of physical properties and



statistical knowledge to energetically evaluate the fitness of rotamers in protein structures, along with sampling algorithms to search the space of all possible mutations. The parameters used in the fitness scoring system are trained by a machine learning technique to reproduce the true sidechain conformations in high-resolution crystal structures [90]. Sequence space is searched using one of two methods, either using Monte Carlo [91], with random mutations accepted or rejected based on the calculated energy, or using the dead-end elimination theorem (DEE), which holds that if energies can be decomposed into pairwise terms, then a solution to the problem of finding the lowest energy set of mutations can be found by a process of successive elimination [92].

However, inaccuracies in design due to the imperfect scoring system, the use of discretized side chains, and the lack of precise modeling of backbone flexibility, affect the reliability of the method. Therefore, instead of relying on the accuracy of the single lowest energy protein sequence, DEEdesign provides an ensemble of plausible mutants, all with reasonably low calculated energy scores. These are assembled into a single amino acid profile, from which a library of nucleotide sequences is derived, employing degenerate codons for those positions in the sequence that have more than one possible amino acid.

In our case study, residues 143-154 NSHNVYITADKQ of  $\beta$ -strand 7 were mutated *in silico* to the target peptide sequence SSHEVSLGVSSA from HA. All residues within 7Å of the target were mutated to all amino acids within the constraints of the evolutionary history of GFP, where the latter was derived from a multiple sequence/structure alignment of 34 fluorescent proteins, augmented by additional homologous sequences. If an amino acid was found at a given position in the evolutionary history of GFP, then that amino acid was allowed in the course of the sequence space search, otherwise it was disallowed. DEE and Monte Carlo were used to search this sequence space, identifying an ensemble of low-energy sequences such that the total complexity of the sequence space of the ensemble was only about ten thousand unique sequences, a number that can be efficiently screened on petri plates. The ensemble of sequences was back-translated to DNA and divided into overlapping degenerate-codon oligonucleotides of 60 bases each by the program DNAWorks [93]. The set of mixed oligos was assembled by PCR into a gene library for screening, using the protocols of gene assembly mutagenesis [94].

### 5.5. Experimental screening and diversity generation by *in vitro* evolution

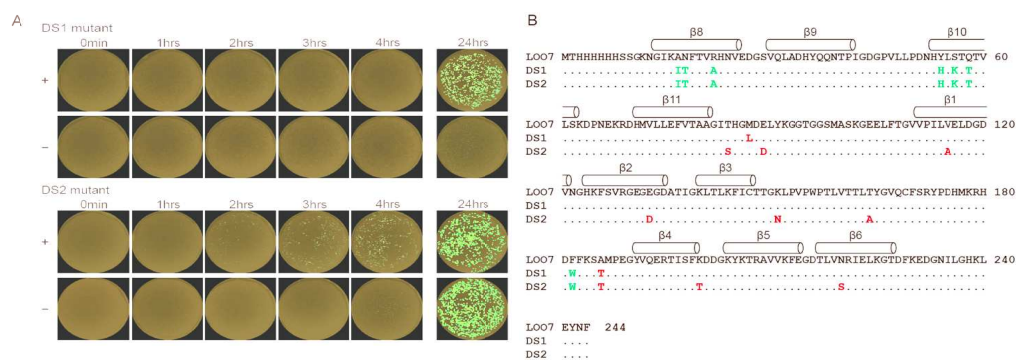
The computationally generated library for the H5N1 LOO-GFP biosensor had a complexity of around 10000 sequences and was relatively easy to screen in low to medium-throughput manner by looking for colonies that were fluorescent when co-expressed with its target peptide sequence. We fused the target peptide to intein [95] so that it would be cleaved immediately after expression and would exist as a free peptide.

However, potential mutations that are distant from the binding site of the target peptide (i.e. >10Å away from the binding site) may still have indirect effects on the binding of the target, or influence on LOO-GFP folding, are not easily captured in the computational design process because of time and memory limitations. To expand the screening, candidate mutant

genes can be subjected to rounds of *in vitro* evolution, such as error-prone PCR [96] and/or DNA shuffling [24].

We demonstrated the first proof-of-concept for designing LOO-GFP biosensors by combining computational protein design and *in vitro* evolution. DEEdesign was used to create a set of degenerate oligonucleotide primers for gene assembly. DNA shuffling was performed directly on this set of genes to further increase the diversity of the constructed library, since gene assembly mutagenesis does not ensure complete representation of all possible anticipated sequences [94]. DNA shuffling also introduces random mutagenesis beyond the predicted mutations on the gene variant.

Potential candidates for LOO-GFP biosensors were plate-screened in *E. coli* that co-expressed the biosensor gene library and the HA peptide fused to a carrier, intein. Expression of both peptide and biosensor library were induced simultaneously, and the intensity of fluorescence was monitored under excitation of 488 nm wavelength after the induction of 24 hours at room temperature. Two potential LOO-GFP biosensors, DS1 and DS2, that produced elevated fluorescence intensity in the presence of the HA peptide were found (Figure 12). There were nine and sixteen mutations found in DS1 and DS2 respectively, and seven of those mutations were from DEEdesign prediction and the remainder were from *in vitro* evolution.



**Figure 12.** Potential LOO-GFP biosensors against HA target peptides of influenza virus. (A) Time course study of fluorescence recovery upon expression of biosensor variants with [+] and without [-] HA peptides. Protein expression was induced with 0.5mM IPTG and under room temperature. Fluorescence was record every hour for 4 hours and after 24 hours. All pictures were taken with the same setting of digital camera. (B) Multiple sequence alignment of LOO7, DS1 and DS2 mutant. Mutations introduced by computational design (green) and *in vitro* evolution (red) in DS1 and DS2 mutants are shown.

When co-expressed with the HA peptide, the DS1 mutant exhibited target-dependent maturation of chromophore, while in the absense of the peptide it showed barely detectable fluorescence even after 24 hours, indicating a specific interaction between DS1 mutant and the HA peptide. DS2 mutant showed faster recovery of fluorescence within four hours in the presence of the HA peptide; however, a higher degree of nonspecific auto-fluorescence was also observed after 24 hours. The DS1 mutant chromophore formation showed a greater dependency on the left-out peptide (i.e. the HA peptide), implying better folding of designed LOO-GFP molecule, than DS2 mutant *in vivo*, showing DS1 mutant as a better HA-specific LOO-GFP biosensor.

## 6. Conclusions

The unique physical properties of GFP have made it a gold mine for the development of biosensors and biomarkers. GFP is kinetically super-stable. Its sequence may be readily permuted and mutated. Its engineered variants fluoresce at wavelengths across the visual spectrum, and some pairs of variants can interact via FRET. GFP is quenched by unfolding, by certain ions, and sometimes by light, and variants of GFP are pH sensitive. With many ways of generating a signal, it is no surprise that many types of biosensors have been developed that use GFP and its homolog fluorescent proteins. GFP and its variants can be immobilized and even dried while retaining structure and biosensor function, leading to the promise of future GFP-biosensor microarrays capable of detecting a wide variety of analytes in a single assay. In addition to being broadly useful, such material should be very cheap to produce, and would also be easily stored, used, and read. Arrays of GFP-based biosensors on paper or film may someday become available for household use, so that infections may be rapidly diagnosed without a trip to the hospital, or may become integral parts of devices that continuously monitor the water and air, making the world a healthier and safer place.

## Author details

Donna E. Crone, Yao-Ming Huang, Derek J. Pitman, Christian Schenkelberg, Keith Fraser, Stephen Macari and Christopher Bystroff

Department of Biology, Rensselaer Polytechnic Institute, Troy, New York, USA

## References

- [1] Ormö M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal Structure of the *Aequorea victoria* Green Fluorescent Protein. *Science* 1996;273(5280): 1392-5. <http://www.sciencemag.org/content/273/5280/1392.long> (accessed 15 July 2012)
- [2] Osamu Shimomura - Nobel Lecture: Discovery of Green Fluorescent Protein, GFP: Nobelprize.org [http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/2008/shimomura-lecture.html](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2008/shimomura-lecture.html) (accessed 15 July 2012)
- [3] Tsien RY. The Green Fluorescent Protein. *Annual Reviews of Biochemistry* 1998;67: 509-544. <http://www.annualreviews.org/doi/pdf/10.1146/annurev.biochem.67.1.509> (accessed 15 July 2012)
- [4] Cramer A, Whitehorn EA, Tate E, Stemmer WPC. Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling. *Nature Biotechnology* 1996;14(3): 315-319. <http://www.nature.com/nbt/journal/v14/n3/full/nbt0396-315.html> (accessed 15 July 2012)

- [5] Enoki S, Saeki K, Maki K, Kuwajima K. Acid Denaturation and Refolding of Green Fluorescent Protein. *Biochemistry* 2004;43(44): 14238-48. <http://pubs.acs.org/doi/abs/10.1021/bi048733%2B> (accessed 15 July 2012)
- [6] Fukuda H, Arai M, Kuwajima K. Folding of Green Fluorescent Protein and the Cycle3 Mutant. *Biochemistry* 2000;39(39): 12025-32. <http://pubs.acs.org/doi/abs/10.1021/bi000543l> (accessed 15 July 2012)
- [7] Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autooxidation of green fluorescent protein. *Proceedings of the National Academy of Science of the United States of America* 1994;91(26): 12501-4. <http://www.pnas.org/content/91/26/12501.long> (accessed 15 July 2012)
- [8] Heim R, Tsien RY. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology* 1996;6(2): 178-82. [http://dx.doi.org/10.1016/S0960-9822\(02\)00450-5](http://dx.doi.org/10.1016/S0960-9822(02)00450-5) (accessed 15 July 2012)
- [9] Rosenow MA, Huffman HA, Phail ME, Wachter RM. The Crystal Structure of the Y66L Variant of Green Fluorescent Protein Supports a Cyclization-Oxidation-Dehydration Mechanism for Chromophore Maturation. *Biochemistry* 2004;43(15): 4464-72. <http://pubs.acs.org/doi/abs/10.1021/bi0361315> (accessed 15 July 2012)
- [10] Henderson NJ, Ai HW, Campbell RE, Remington SJ. Structural basis for reversible photobleaching of a green fluorescent protein homologue. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104(16): 6672-7. <http://www.pnas.org/content/104/16/6672.long> (accessed 15 July 2012)
- [11] Ward WW, Prentice, HJ, Roth AF. Spectral Perturbations of the *Aequorea* Green-Fluorescent Protein. *Photochemistry and Photobiology* 1982;35(6): 803-808. DOI: 10.1111/j.1751-1097.1982.tb02651.x
- [12] Ward WW, Cody CW, Hart RC, Cormier MJ. Spectrophotometric Identity of the Energy Transfer Chromophores in *Renillia* and *Aequorea* Green-Fluorescent Proteins. *Photochemistry and Photobiology* 1980;31(6): 611-615. DOI: 10.1111/j.1751-1097.1980.tb03755.x
- [13] Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma sp.* red fluorescent protein. *Nature Biotechnology* 2004;22(12): 1567-72. doi: 10.1038/nbt1037 <http://www.nature.com/nbt/journal/v22/n12/full/nbt1037.html> (accessed 16 July 2012).
- [14] Petersen J, Wilmann PG, Beddoe T, Oakley AJ, Devenish RJ, Prescott M, Rossjohn J. The 2.0-Å Crystal Structure of eqFP611, a Far Red Fluorescent Protein from the Sea Anemone *Entacmaea quadricolor*. *The Journal of Biological Chemistry* 2003;278(45): 44626-31. <http://www.jbc.org/content/278/45/44626.full> (accessed 16 July 2012).



- [15] Gurskaya NG, Fradkov AF, Terskikh A, Matz MV, Labas YA, Martynov VI, Yanushevich YG, Lukyanov KA, Lukyanov SA. GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Letters* 2001;507(1): 16-20. <http://www.sciencedirect.com/science/article/pii/S0014579301029301> (accessed 16 July 2012).
- [16] Shkrob MA, Yanushevich YG, Chudakov DM, Gurskaya NG, Labas YA, Poponov SY, Mudrik NN, Lukyanov S, Lukyanov KA. Far-red fluorescent proteins evolved from a blue chromoprotein from *Actinia equina*. *Biochemical Journal* 2005;392(Pt 3): 649-54. <http://www.biochemj.org/bj/392/0649/bj3920649.htm> (accessed 16 July 2012).
- [17] Subach OM, Patterson GH, Ting L-M, Wang Y, Condeelis JS, Verkhusha VV. A photoswitchable orange-to-far-red fluorescent protein, PSmOrange. *Nature Methods* 2011;8(9): 771-7. <http://www.nature.com/nmeth/journal/v8/n9/full/nmeth.1664.html> (accessed 16 July 2012).
- [18] Chica RA, Moore MM, Allen BD, Mayo SL. Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proceedings of the National Academy of Sciences of the United States of America* 2010;107(47): 20257-62. <http://www.pnas.org/content/107/47/20257> (accessed 16 July 2012).
- [19] Pollok BA, Heim R. Using GFP in FRET-based applications. *Trends in Cell Biology* 1999;9(2): 57-60. <http://www.sciencedirect.com/science/article/pii/S0962892498014342> (accessed 16 July 2012).
- [20] Markwardt ML, Kremers G-J, Kraft CA, Ray K, Cranfill PJ, Wilson KA, Day RN, Wachter RM, Davidson MW, Rizzo MA. An Improved Cerulean Fluorescent Protein with Enhanced Brightness and Reduced Reversible Photoswitching. *PLoS One* 2011;6(3): e17896. <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0017896> (accessed 16 July 2012).
- [21] Albertazzi L, Arosio D, Marchetti L, Ricci F, Beltram F. Quantitative FRET Analysis With the EGFP-mCherry Fluorescent Protein Pair. *Photochemistry and Photobiology* 2009;85(1): 287-97. <http://onlinelibrary.wiley.com/doi/10.1111/j.1751-1097.2008.00435.x/full> (accessed 16 July 2012).
- [22] Day RN, Davidson MW. Fluorescent proteins for FRET microscopy: monitoring protein interactions in living cells. *Bioessays* 2012;34(5): 341-50. <http://onlinelibrary.wiley.com/doi/10.1002/bies.201100098/full> (accessed 16 July 2012).
- [23] Pritchard L, Corne D, Kell D, Rowland J, Winson M. A general model of error-prone PCR. *Journal of Theoretical Biology* 2005;234(4): 497-509. <http://www.sciencedirect.com/science/article/pii/S0022519304006071> (accessed 15 July 2012).
- [24] Stemmer WP. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* 1994;91(22): 10747-51 <http://www.pnas.org/content/91/22/10747.long> (accessed 15 July 2012).



- [25] Cormack BP, Valdivia RH, Falkow S. FACS-optimized mutants of the green fluorescent protein. *Gene* 1996;173: 33-8. <http://www.sciencedirect.com/science/article/pii/0378111995006850> (accessed 16 July 2012).
- [26] Waldo GS, Standish BM, Berendzen J, Terwilliger TC. Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology* 1999;17(7): 691-5. [http://www.nature.com/nbt/journal/v17/n7/full/nbt0799\\_691.html](http://www.nature.com/nbt/journal/v17/n7/full/nbt0799_691.html) (accessed 15 July 2012)
- [27] Pédelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* 2006;24(1): 79-88. <http://www.nature.com/nbt/journal/v24/n1/abs/nbt1172.html> (accessed 16 July 2012).
- [28] Lawrence MS, Phillips KJ, Liu DR. Supercharging Proteins Can Impart Unusual Resilience. *Journal of the American Chemical Society* 2007;129(33): 10110-2. <http://pubs.acs.org/doi/full/10.1021/ja071641y> (accessed 16 July 2012).
- [29] Melnik BS, Povarnitsyna TV, Glukhov AS, Melnik TN, Uversky VN. SS-Stabilizing Proteins Rationally: Intrinsic Disorder-Based Design of Stabilizing Disulphide Bridges in GFP. *Journal of biomolecular structure and dynamics* 2012;29(4): 815-24. DOI:10.1080/07391102.2012.10507414
- [30] Rosenman D, Huang YM, Xia K, Colon W, Van Roey P, Bystroff C. Green-lighting GFP folding by loop remodeling to eliminate a cis-trans peptide isomerization. (in preparation), 2012.
- [31] Evdokimov AG, Pokross ME, Egorov NS, Zaisky AG, Yampolsky IV, Merzlyak EM, Shkoporov AN, Sander I, Lukyanov KA, Chudakov DM. Structural basis for the fast maturation of Arthropoda green fluorescent protein. *EMBO Reports* 2006;7(10): 1006-12. <http://www.nature.com/embor/journal/v7/n10/full/7400787.html> (accessed 16 July 2012).
- [32] Dai M, Fisher HE, Temirov J, Kiss C, Phipps ME, Pavlik P, Werner JH, Bradbury AR. The creation of a novel fluorescent protein by guided consensus engineering. *Protein Engineering Design & Selection* 2007;20(2): 69-79. <http://peds.oxfordjournals.org/content/20/2/69.full> (accessed 16 July 2012).
- [33] Kiss C, Temirov J, Chasteen L, Waldo GS, Bradbury AR. Directed evolution of an extremely stable fluorescent protein. *Protein Engineering Design & Selection* 2009;22(5): 313-23. <http://peds.oxfordjournals.org/content/22/5/313.full> (accessed 16 July 2012).
- [34] Topell S, Hennecke J, Glockshuber R. Circularly permuted variants of the green fluorescent protein. *FEBS Letters* 1999;457(2): 283-9. <http://www.sciencedirect.com/science/article/pii/S0014579399010443> (accessed 16 July 2012).
- [35] Baird GS, Zacharias DA, Tsien RY. Circular permutation and receptor insertion within green fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America* 1999;96(20): 11241-6. <http://www.pnas.org/content/96/20/11241.long> (accessed 16 July 2012).

- [36] Reeder PJ, Huang Y-M, Dordick JS, Bystroff C. A Rewired Green Fluorescent Protein: Folding and Function in a Nonsequential, Noncircular GFP Permutant. *Biochemistry* 2010;49(51): 10773-9. <http://pubs.acs.org/doi/full/10.1021/bi100975z> (accessed 16 July 2012).
- [37] Gilbert D, Westhead D, Nagano N, Thornton J. Motif-based searching in TOPS protein topology databases. *Bioinformatics* 1999;15(4): 317-326 <http://bioinformatics.oxfordjournals.org/content/15/4/317.abstract%C2%B5%EF%A3%B9%CE%B0z> (accessed 16 July 2012)
- [38] Pavoov TV, Cho YK, Shusta EV. Development of GFP-based biosensors possessing the binding properties of antibodies. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106(29): 11895-900. <http://www.pnas.org/content/106/29/11895.long> (accessed 15 July 2012)
- [39] Huang YM, Bystroff C. Complementation and Reconstitution of Fluorescence from Circularly Permuted and Truncated Green Fluorescent Protein. *Biochemistry* 2009;48(5): 929-40. <http://pubs.acs.org/doi/full/10.1021/bi802027g> (accessed 16 July 2012).
- [40] Huang YM, Nayak S, Bystroff C. Quantitative *in vivo* solubility and reconstitution of truncated circular permutants of green fluorescent protein. *Protein Science* 2011;20(11): 1775-80. <http://onlinelibrary.wiley.com/doi/10.1002/pro.735/full> (accessed 16 July 2012).
- [41] Kent KP, Oltrogge LM, Boxer SG. Synthetic Control of Green Fluorescent Protein. *Journal of the American Chemical Society* 2009;131(44): 15988-9. <http://pubs.acs.org/doi/full/10.1021/ja906303f> (accessed 16 July 2012).
- [42] Nemethy G, Scheraga HA. A Possible Folding Pathway of Bovine Pancreatic RNase. *Proceedings of the National Academy of Sciences* 1979;76(12): 6050-6054. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC411798/?tool=pubmed> (accessed 16 July 2012)
- [43] Kent KP, Boxer SG. Light-Activated Reassembly of Split Green Fluorescent Protein. *Journal of the American Chemical Society* 2011;133(11): 4046-52. <http://pubs.acs.org/doi/full/10.1021/ja110256c> (accessed 12 June 2012).
- [44] Do K, Boxer SG. Thermodynamics, Kinetics, and Photochemistry of  $\beta$ -Strand Association and Dissociation in a Split-GFP System. *Journal of the American Chemical Society* 2011;133(45): 18078-81. <http://pubs.acs.org/doi/full/10.1021/ja207985w> (accessed 4 June 2012).
- [45] Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. *Science*. 1994;263(5148): 802-5. <http://www.sciencemag.org/content/263/5148/802.long> (accessed 15 July 2012)
- [46] Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene*. 1992;111(2): 229-33.

- [47] Chudakov DM, Matz MV, Lukyanov S, Lukyanov KA. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological Reviews* 2010;90(3): 1103-63. <http://physrev.physiology.org/content/90/3/1103.long> (accessed 15 July 2012)
- [48] Kremers GJ, Gilbert SG, Cranfill PJ, Davidson MW, Piston DW. Fluorescent proteins at a glance. *Journal of Cell Science* 2011;124(Pt 2): 157-60. doi: 10.1242/jcs.072744 <http://jcs.biologists.org/content/124/2/157.long> (accessed 15 July 2012)
- [49] Miyawaki A. Proteins on the move: insights gained from fluorescent protein technologies. *Nature reviews. Molecular cell biology* 2011;12(10): 656-68. doi: 10.1038/nrm3199. <http://www.nature.com/nrm/journal/v12/n10/full/nrm3199.html> (accessed 15 July 2012)
- [50] Cabantous S, Terwilliger TC, Waldo GS. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature Biotechnology*. 2005;23(1): 102-7. <http://www.nature.com/nbt/journal/v23/n1/full/nbt1044.html> (accessed 15 July 2012)
- [51] Goedert M. Tau protein and the neurofibrillary pathology of Alzheimer's disease. *Annals of the New York Academy of Sciences* 1996;777: 121-31. DOI:10.1111/j.1749-6632.1996.tb34410.x (accessed 15 July 2012).
- [52] Chun W, Waldo GS, Johnson GV. Split GFP complementation assay: a novel approach to quantitatively measure aggregation of tau in situ: effects of GSK3 $\beta$  activation and caspase 3 cleavage. *Journal of neurochemistry* 2007;103(6): 2529-39. DOI: 10.1111/j.1471-4159.2007.04941.x <http://onlinelibrary.wiley.com/doi/10.1111/j.1471-4159.2007.04941.x/abstract> (accessed 15 July 2012).
- [53] Chun W, Waldo GS, Johnson GV. Split GFP complementation assay for quantitative measurement of tau aggregation in situ. In: Roberson, ED. (ed.) *Alzheimer's Disease and Frontotemporal Dementia Methods and Protocols*. Clifton NJ: Springer. 2011. p109-23. Available from <http://www.springerlink.com/content/18444n2g5116g5t7/?MUD=MP> (accessed 15 July 2012)
- [54] Rothbauer U, Zolghadr K, Tillib S, Nowak D, Schermelleh L, Gahl A, Backmann N, Conrath K, Muyldermans S, Cardoso MC, Leonhardt H. Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nature methods* 2006;3(11): 887-889. <http://www.nature.com/nmeth/journal/v3/n11/full/nmeth953.html> (accessed 15 July 2012)
- [55] Betzig E; Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF. Imaging intracellular fluorescent proteins at nanomolar resolution. *Science* 2006;313(5793): 1642-1645. <http://www.sciencemag.org/content/313/5793/1642.long> (accessed 16 July 2012)
- [56] Ries J, Kaplan C, Platonova E, Eghlidi H, Ewers H. A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. *Nature methods* 2012; 9(6): 582-584. doi: 10.1038/nmeth.1991. <http://www.nature.com/nmeth/journal/v9/n6/full/nmeth.1991.html> (Accessed 15 July 2012).

- [57] Kamioka Y, Sumiyama K, Mizuno R, Sakai Y, Hirata E, Kiyokawa E, Matsuda M. Live imaging of protein kinase activities in transgenic mice expressing FRET biosensors. *Cell structure and function* 2012;37(1): 65-73. <http://dx.doi.org/10.1247/csf.11045> (accessed 15 July 2012).
- [58] Bizzarri R, Serresi M, Luin S, Beltram F. Green fluorescent protein based pH indicators for in vivo use: a review. *Analytical and bioanalytical chemistry* 2009;393(4): 1107-22. DOI: 10.1007/s00216-008-2515-9 <http://www.springerlink.com/content/y73v37381q82v3h6/> (accessed 15 July 2012).
- [59] Choi WG, Swanson SJ, Gilroy S. High-resolution imaging of Ca<sup>2+</sup>, redox status, ROS and pH using GFP biosensors. *The Plant journal* 2012;70(1): 118-28. doi: 10.1111/j.1365-313X.2012.04917.x. <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2012.04917.x/abstract> (accessed 15 July 2012).
- [60] Gjetting KS, Ytting CK, Schulz A, Fuglsang AT. Live imaging of intra- and extracellular pH in plants using pHusion, a novel genetically encoded biosensor. *Journal of experimental botany*. 2012;63(8): 3207-18. <http://jxb.oxfordjournals.org/content/63/8/3207.long> (accessed 15 July 2012).
- [61] Patterson GH, Lippincott-Schwartz J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* 2002;297(5588): 1873-7. <http://www.sciencemag.org/content/297/5588/1873.long> (accessed 15 July 2012).
- [62] Brakemann T, Stiel AC, Weber G, Andresen M, Testa I, Grotjohann T, Leutenegger M, Plessmann U, Urlaub H, Eggeling C, Wahl MC, Hell SW, Jakobs S. A reversibly photoswitchable GFP-like protein with fluorescence excitation decoupled from switching. *Nature Biotechnology* 2011;29(10): 942-7. doi: 10.1038/nbt.1952. <http://www.nature.com/nbt/journal/v29/n10/full/nbt.1952.html> (accessed 15 July 2012)
- [63] Kneen M, Farinas J, Li Y, Verkman AS. Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophysical journal* 1998;74(3): 1591-9. <http://www.sciencedirect.com/science/article/pii/S0006349598778701> (accessed 7 July 2012)
- [64] Miesenböck G, De Angelis DA, Rothman JE. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* 1998;394(6689): 192-5. <http://www.nature.com/nature/journal/v394/n6689/full/394192a0.html> (accessed 15 July 2012)
- [65] Hanson GT, McAnaney TB, Park ES, Rendell ME, Yarbrough DK, Chu S, Xi L, Boxer SG, Montrose MH, Remington SJ. Green fluorescent protein variants as ratiometric dual emission pH sensors. 1. Structural characterization and preliminary application. *Biochemistry* 2002;41(52): 15477-88. <http://pubs.acs.org/doi/abs/10.1021/bi026609p> (accessed 15 July 2012)
- [66] Elsliger MA, Wachter RM, Hanson GT, Kallio K, Remington SJ. Structural and spectral response of green fluorescent protein variants to changes in pH. *Biochemistry* 1999;27;38(17): 5296-301. <http://pubs.acs.org/doi/abs/10.1021/bi9902182> (accessed 15 July 2012)



- [67] Bizzarri R, Arcangeli C, Arosio D, Ricci F, Faraci P, Cardarelli F, Beltram F. Development of a novel GFPbased ratiometric excitation and emission pH indicator for intracellular studies. *Biophysical journal* 2006;90(9): 3300-14. <http://www.sciencedirect.com/science/article/pii/S0006349506725127> (accessed 15 July 2012)
- [68] Arosio D, Ricci F, Marchetti L, Gualdani R, Albertazzi L, Beltram F. Simultaneous intracellular chloride and pH measurements using a GFP-based sensor. *Nature methods* 2010;7(7): 516-8. <http://www.nature.com/nmeth/journal/v7/n7/full/nmeth.1471.html> (accessed 15 July 2012)
- [69] Awaji T, Hirasawa A, Shirakawa H, Tsujimoto G, Miyazaki S. Novel green fluorescent protein-based ratiometric indicators for monitoring pH in defined intracellular microdomains. *Biochemical and biophysical research communications* 2001;289(2): 457-62. <http://www.sciencedirect.com/science/article/pii/S0006291X01960048> (accessed 15 July 2012)
- [70] Zhou X, Herbst-Robinson K, Zhang J. Visualizing dynamic activities of signaling enzymes using genetically encodable FRET-based biosensors. From design to applications. In: Conn PM (ed.) *Methods in enzymology: Imaging and Spectroscopic Analysis of Living Cells Optical and Spectroscopic Techniques*. San Diego, Calif. : Elsevier Inc. 2012. 504: 317-340. Available from <http://www.sciencedirect.com/science/article/pii/B9780123918574000161> (accessed 16 July 2012)
- [71] Komatsu N, Aoki K, Yamada M, Yukinaga H, Fujita Y, Kamioka Y, Matsuda M. Development of an optimized backbone of FRET biosensors for kinases and GTPases. *Molecular biology of the cell* 2011;22(23): 4647-56. <http://www.molbiolcell.org/content/22/23/4647.long> (accessed 16 July 2012)
- [72] Eiamphungporn W, Prachayasittikul S, Isarankura-Na-Ayudha C, Prachayasittikul V. Development of bacterial cell-based system for intracellular antioxidant activity screening assay using green fluorescence protein (GFP) reporter. *African Journal of Biotechnology* 2012;11(27): 6934-45. DOI: 10.5897/AJB11.3790 <http://www.academicjournals.org/ajb/PDF/pdf2012/3Apr/Eiamphungporn%20et%20al.pdf> (accessed 15 July 2012)
- [73] Hires, S.A., L. Tian, and L.L. Looger, Reporting neural activity with genetically encoded calcium indicators. *Brain cell biology* 2008;36(1-4): 69-86. <http://www.springerlink.com/content/a145q526472454q4/> (accessed 15 July 2012).
- [74] Tian L, Hires SA, Mao T, Huber D, Chiappe ME, Chalasani SH, Petreanu L, Akerboom J, McKinney SA, Schreiter ER. Imaging neural activity in worms, flies and mice with improved GCaMPcalcium indicators. *Nature methods* 2009;6(12): 875-81. <http://www.nature.com/nmeth/journal/v6/n12/full/nmeth.1398.html> (accessed 15 July 2012)
- [75] Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology* 1997;15(6): 553-7. <http://www.nature.com/nbt/journal/v15/n6/full/nbt0697-553.html> (accessed 15 July 2012)

- [76] Doi N, Yanagawa H. Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. *FEBS letters* 1999;453(3): 305-7. [http://dx.doi.org/10.1016/S0014-5793\(99\)00732-2](http://dx.doi.org/10.1016/S0014-5793(99)00732-2) (accessed 15 July 2012)
- [77] Boeneman K, Mei BC, Dennis AM, Bao G, Deschamps JR, Mattoussi H, Medintz IL. Sensing caspase 3 activity with quantum dot-fluorescent protein assemblies. *Journal of the American Chemical Society* 2009;131(11): 3828-9. doi:10.1021/ja809721j <http://pubs.acs.org/doi/abs/10.1021/ja809721j> (accessed 15 July 2012)
- [78] Dennis, A. M., Rhee, W. J., Sotto, D., Dublin, S. N., & Bao, G. (2012). Quantum dot-fluorescent protein FRET probes for sensing intracellular pH. *ACS nano* 2012;6(4): 2917-24. doi:10.1021/nn2038077 <http://pubs.acs.org/doi/abs/10.1021/nn2038077> (accessed 15 July 2012)
- [79] Tansila N, Becker K, Na-ayudhya CI, Prachayasittikul V, Bu L. Metal ion accessibility of histidine-modified superfolder green fluorescent protein expressed in *Escherichia coli*. *Biotechnology Letters* 2008;30(8): 1391-1396. doi:10.1007/s10529-008-9692-7 <http://www.springerlink.com/content/3813241647413002/> (accessed 15 July 2012)
- [80] Tansila N, Tantimongcolwat T, Isarankura-Na-Ayudhya C, Nantasenamat C, Prachayasittikul V. Rational design of analyte channels of the green fluorescent protein for biosensor applications. *International journal of biological sciences* 2007;3(7): 463-70. <http://www.biolsci.org/v03p0463.htm> (accessed 15 July 2012)
- [81] Schueler-Furman O, Wang C, Bradley P, Misura K, Baker D. Progress in Modeling of Protein Structures and Interactions. *Science* 2005;310(5748): 638-42. <http://www.sciencemag.org/content/310/5748/638.long> (accessed 15 July 2012)
- [82] Street AG, Mayo SL. Computational Protein Design. *Structure* 1999;7(5): R105-R109.
- [83] Park S, Yang X, Saven JG. Advances in Computational Protein Design. *Current Opinion in Structural Biology* 2004;14(4): 487-94. <http://www.sciencedirect.com/science/article/pii/S0959440X04001009> (accessed 15 July 2012)
- [84] Lippow SM, Tidor B. Progress in Computational Protein Design. *Current Opinion in Biotechnology* 2007;18(4): 305-11. <http://www.sciencedirect.com/science/article/pii/S0958166907000778> (accessed 15 July 2012)
- [85] Dunbrack RL. Rotamer Libraries in the 21st Century. *Current Opinion in Structural Biology* 2002;12(4): 431-40. <http://www.sciencedirect.com/science/article/pii/S0959440X02003445> (accessed 16 July 2012)
- [86] Lovell SC, Word JM, Richardson JS, Richardson DC. The Penultimate Rotamer Library. *Proteins* 2000;40(3): 389-408. doi: 10.1002/1097-0134(20000815)40:3<389::AID-PROT50>3.0.CO;2-2 (accessed 16 July 2012)
- [87] Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure Validation by  $\alpha$  Geometry:  $\phi, \psi$  and  $C\beta$  Deviation. *Proteins* 2003;50(3) 437-450. doi: 10.1002/prot.10286 <http://onlinelibrary.wiley.com/doi/10.1002/prot.10286/abstract> (accessed 16 July 2012)

- [88] Pokala N, Handel TM, Review: protein design--where we were, where we are, where we're going. *Journal of structural biology* 2001;134(2-3): 269-81. <http://www.science-direct.com/science/article/pii/S1047847701943497> (accessed 16 July 2012)
- [89] Hayes RJ, Bentzien J, Ary ML, Hwang MY, Jacinto JM, Vielmetter J, Kundu A, Dahiyat BI. Combining Computational and Experimental Screening for Rapid Optimization of Protein Properties. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99(25): 15926-31 <http://www.pnas.org/content/99/25/15926.long> (accessed 15 July 2012)
- [90] Huang YM, Bystroff C. Exploring Objective Functions and Cross-terms in the Optimization of an Energy Function for Protein Design. (submitted) ACMBCB2012, 7-10 October 2012, Orlando, Florida, USA 2012
- [91] Shakhnovich EI, Gutin AM, A new approach to the design of stable proteins. *Protein Engineering* 1993;6(8): 793-800.
- [92] Desmet J, De Maeyer M, Hazes B, Lasters I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* 1992;356(6369): 539-42. doi: 10.1038/356539a0 <http://www.nature.com/nature/journal/v356/n6369/pdf/356539a0.pdf> (accessed 16 July 2012)
- [93] Hoover D. M. and Lubkowski J., DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Research* 2002;30(10): e43. <http://nar.oxfordjournals.org/content/30/10/e43.long> (accessed 16 July 2012)
- [94] Bessette PH, Mena MA, Nguyen AW, Daugherty PS. Construction of Designed Protein Libraries Using Gene Assembly Mutagenesis. *Methods in Molecular Biology* 2003;231(1) 29-37. DOI: 10.1385/1-59259-395-X:29
- [95] Intein – Wikipedia: <http://en.wikipedia.org/wiki/Intein>. (20 July 2012)
- [96] Cirino P. C., Mayer K. M., and Umeno D., Generating mutant libraries using error-prone PCR. In: Arnold FH, Georgiou G. (eds.) *Methods In Molecular Biology: Directed Evolution Library Creation: Methods and Protocols*. Totowa NJ: Humana Press; 2003. 231: p. 3-10. Available from <http://www.springerlink.com/content/978-1-58829-285-8/> (accessed 16 July 2012)