Novel fluorescent protein from *Discosoma* coral and its mutants possesses a unique far-red fluorescence

Arkady F. Fradkov[a], Ying Chen[b], Li Ding[b], Ekaterina V. Barsova[a], Mikhail V. Matz[a], Sergey A. Lukyanov[a,*]

[a] *Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117871 Moscow, Russia*
[b] *Clontech Laboratories, Inc., Palo Alto, CA, USA*

Received 29 June 2000; accepted 25 July 2000

Edited by Matti Saraste

Abstract A novel gene for advanced red-shifted protein with an emission maximum at 593 nm was cloned from *Discosoma* coral. The protein, named dsFP593, is highly homologous to the recently described GFP-like protein drFP583 with an emission maximum at 583 nm. Using the remarkable similarity of the drFP583 and dsFP593 genes, we performed a 'shuffling' procedure to generate a pool of mutants consisting of various combinations of parts of both genes. One 'hybrid gene' was chosen for subsequent random mutagenesis, which resulted in a mutant variant with a uniquely red-shifted emission maximum at 616 nm.

Key words: Green fluorescent protein; Red fluorescent protein; Random mutagenesis; drFP593; drFP583

1. Introduction

Since the introduction of *Aequorea victoria* green fluorescent protein (GFP) into cell biology and biotechnology practices [1] many attempts to expand the fluorescent color palette have been made [2]. In particular, red-shifted fluorescence would offer the convenience of minimizing background autofluorescence and would allow tracking two markers simultaneously. Until recently, the only way to achieve a red-shifted variant was via random and site-directed mutagenesis of GFP [3–5]. The lack of understanding GFP's structure–function relationship has limited the rational design of fluorescent proteins [4,5] to only a 20 nm red shift [5].

With the discovery that reef corals provide a rich new source of GFP-like fluorescent proteins [6], it became possible to simply find a natural protein with the desired properties. Taking into account the abundance of coral fluorescence and the impressive variety of colors [7,8], it can be assumed that coral reef contain an almost limitless inventory of fluorescent proteins. The fluorescent properties of some coral proteins demonstrate a dramatic red shift. In particular, a protein designated drFP583 has an emission maximum of 583 nm, which exceeds the previous red-shift record by more than 56 nm [6]. This fact makes the question of how the fluorophore is formed and tuned even more acute – it became evident that even basic factors determining the theoretical limits of fluorescence are still unclear.

In this work, we attempted to shift the fluorescence of a GFP-like protein into an even longer wave region. After identifying a good candidate from coral reef, we applied an 'old-fashioned' strategy – we shuffled coding region fragments between the new gene and drFP583 and then subjected the combinatorial gene to random mutagenesis using error-prone PCR. In addition to providing an advanced marker protein for various applications, we believe that our results will contribute to understanding the mechanisms of fluorescence in GFP-like proteins.

2. Materials and methods

2.1. cDNA preparation

Total RNA was isolated from the *Discosoma* sp. as described [9]. cDNA preparation and isolation of the full-length cDNA were performed as described [6,10].

2.2. Shuffling procedure

cDNA fragments encoding the drFP583 and dsFP593 proteins were amplified using the following PCR conditions: 20 cycles for 95°C 15 s; 68°C 1 min, 20 s. Plasmids containing drFP583 and dsFP593 ORFs were used as templates, and the PCR primers were as follows: ds5′-*Bam*: 5′-ACATGGATCCAGTTGTTCCAAGAATGTGAT and ds3′-*Xho*: 5′-TAGTACTCGAGGCCATTACCGCTAATC for dsFP593; dr5′-*Bam*: 5′-ACATGGATCCAGGTCTTCCAAGAATGTTATC and dr3′-*Xho*: 5′-TAGTACTCGAGCCAAGTTCAGCCTTA for drFP583 (0.2 μM each). Amplified fragments were simultaneously digested with *Eco*RI, *Hin*dIII and *Dra*I (New England Biolabs). Agarose gel-purified products of both digests were combined, ligated and amplified by PCR (20 cycles for 95°C 15 s; 68°C 1 min, 20 s) with all four primers. The PCR product (~700 bp) was digested with *Bam*HI and *Xho*I (New England Biolabs) and subcloned into the pQE-30 vector (Qiagen), which had been predigested with *Bam*HI and *Sal*I.

2.3. Random mutagenesis

Random mutagenesis of the ds/dr-hybrid gene was performed using the Diversify™ PCR Random Mutagenesis kit (Clontech) according to the manufacturer's protocol under conditions that provided a mutational frequency of three or four mutations per 1000 bp. Products of mutagenesis were cloned into the pQE-30/*Bam*HI–*Sal*I vector.

2.4. E. coli transformation and colony screening

In 10% glycerol, *Escherichia coli* DH5α (Clontech) were transformed with the ligation mixture by electroporation (0.1 cm cuvette, 12.5 kV/cm, 200 Ω, 25 μF). Transformants were plated onto LB/agar plates containing 100 μg/ml ampicillin and 0.1 mM isopropyl-β-D-thiogalactoside (IPTG) and were incubated overnight at 37°C. The resultant colonies (2–5000 per plate) were screened visually using a fluorescent microscope (Karl Zeiss) equipped with a 31001 filter (Chroma). The brightest red colonies, and colonies with different hues of red were selected for further analysis.

*Corresponding author. Fax: (7)-095-330 7056.
E-mail: luk@ibch.ru

0014-5793/00/$20.00
PII: S0014-5793(00)01895-0

2.5. Bacterial expression and protein purification

For protein expression, cells carrying the recombinant plasmid were grown to an OD_{600} of 0.6 in LB containing 100 μg/ml ampicillin and were induced with 1 mM IPTG. Bacteria were then incubated overnight at 37°C. The resultant biomass was pelleted by centrifugation, resuspended in 1×PBS buffer, and sonicated. The bacterial lysate was cleared by centrifugation, and protein was purified from the supernatant with Talon Metal Affinity Resin (Clontech) according to the manufacturer's protocol.

3. Results

To clone a novel gene encoding a putative red-shifted fluorescent protein, we chose a species from the *Discosoma* genus. Our species choice was based on the almost perfect correlation between fluorescence of animals in vivo and the emission spectrum of proteins isolated from them. Knowing that upon UV irradiation this *Discosoma* species exhibits an intensely bright red fluorescence, we began our search for a red-shifted fluorescent protein.

3.1. Structure and spectral properties of the new dsFP593 protein.

We isolated the full-length cDNA encoding a novel red-shifted fluorescent protein (GenBank accession number AF272711) using a cDNA cloning strategy previously described [6]. An amino acid sequence alignment of the dsFP593 protein with recently isolated, fluorescent proteins showed a high extent of identity between them (Fig. 1). Based on the alignment, the overall β-can structure of dsFP593 is similar to the β-can structure of other fluorescent proteins. The crucial amino acid residues involved directly in fluorophore formation (Tyr-66, Gly-67) and the amino acid residues

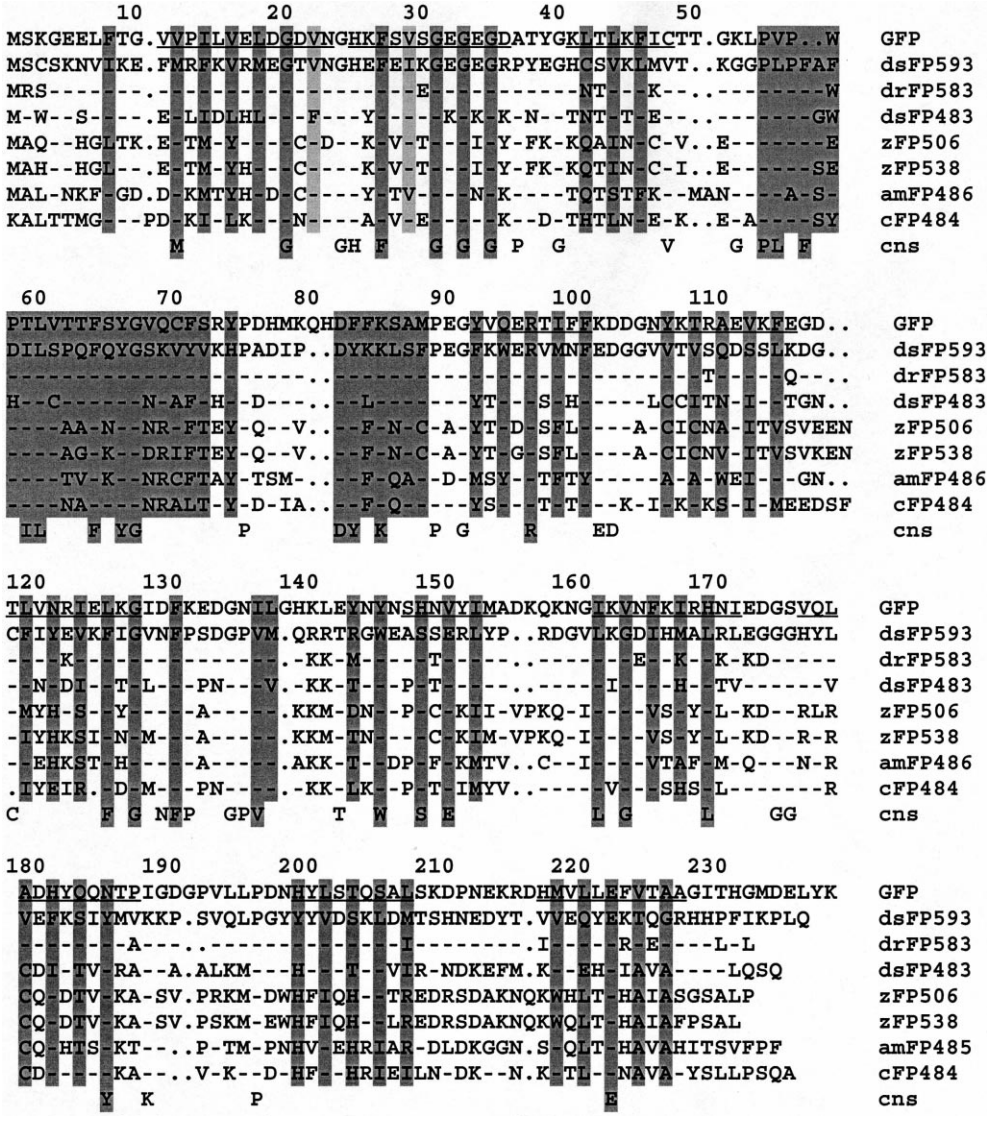

Fig. 1. Multiple alignment of fluorescent proteins. The numbering is based on *A. victoria* GFP. All proteins are compared to the protein from *Discosoma* sp. described here (second line, dsFP593). Residues identical to the corresponding ones in dsFP593 are represented by dashes. Introduced gaps are represented by dots. In the sequence of *A. victoria* GFP, the stretches forming β-sheets (according to [11]) are underlined. The residues whose side chains form the interior of the β-can are shaded. Amino acids residues whose positions are absolutely conserved in all six new proteins are marked as 'cns' (consensus) in an additional line. The N-terminal portion of cFP484, which shows no homology with the other proteins, is not shown [6].

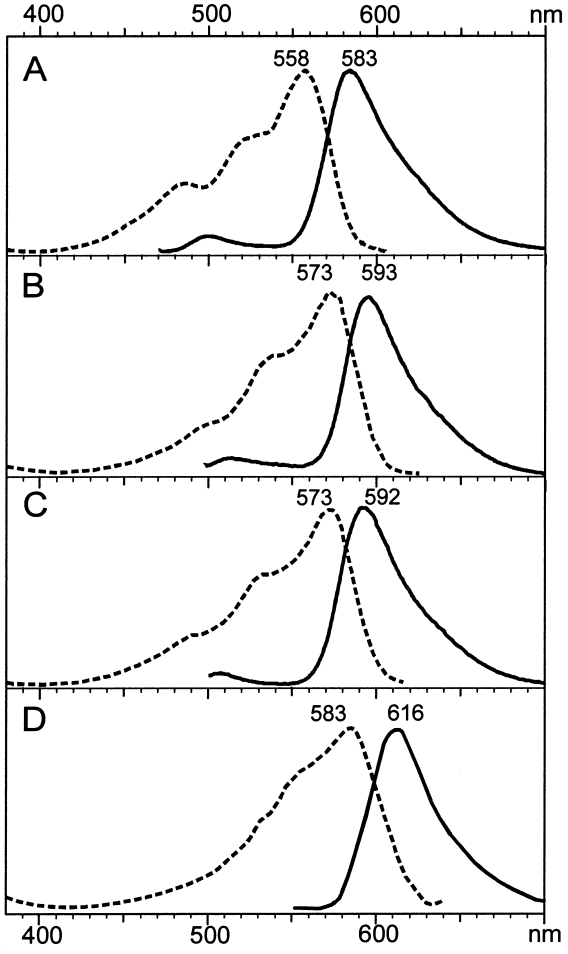

Fig. 2. The excitation (dotted lines) and emission (solid lines) spectra for the red-shifted fluorescent proteins as determined using recombinant *E. coli* proteins. A: drFP583. B: dsFP593. C: ds/dr-hybrid. D: ds/drFP616.

having a strong influence on the process (Arg-96, Glu-222) are conserved in dsFP593. Interestingly, the nearest homologue to dsFP593 is drFP583, the only known red-shifted fluorescent protein. The dsFP593 protein has 26 amino acid substitutions (11%) in comparison to drFP583, and only five of the substitutions are 'internal' (i.e. their side chains are directed to the inside of the β-can). Hence, only these five amino acids residues can be directly responsible for the spectral properties of dsFP593.

We expressed the full-length dsFP593 protein in bacteria and then purified it by His-6 metal affinity chromatography. The purified protein solution was intensely pink colored. When cells expressing either dsFP593 or drFP583 were plated next to each other, the difference in color and fluorescence of the bacterial colonies was quite evident. The excitation spectrum for dsFP593 has a rather wide major peak with a maximum at 573 nm and a shoulder at 525 nm (Fig. 2A). The emission spectrum is essentially a mirror of the excitation spectrum with a maximum at 593 nm (Fig. 2A), which in comparison to drFP583 (Fig. 2B) is red-shifted approximately 10 nm. In terms of quantum yield and overall brightness, the fluorescent characteristics of dsFP593 are considerably less than those of drFP583 (Table 1).

3.2. Generation of shuffled proteins

The nucleotide sequences of dsFP593 and drFP583 are extremely similar (~85% identity); in fact, some restriction sites are retained. We utilized this fact to generate a pool of combinatorial mutants in order to create new protein variants with unusual or improved spectral properties. Each gene was divided into four fragments using a mixture of restrictases. The fragments were then ligated in an effort to create 16 combinatorial mutants. Screening several hundred clones revealed that the majority of colonies fluoresced less than the initial drFP583 clone (the brighter 'parent'); thus, only visually brighter mutants were further investigated. Sequence analysis showed that all but one of the combinatorial mutants, designated 'ds/dr hybrid', corresponded to the original drFP583 gene.

The ds/dr-hybrid clone contains the larger N-terminal part (residues 1–180) of the dsFP593 gene and smaller C-terminal part (from residue 181 to the -IV (3′) end) of the drFP583 gene (Fig. 3). As expected, its spectral properties represent an obvious compilation of the 'parental' proteins' properties. The general form of the excitation and emission spectra, including the emission maximum position, resembles the dsFP593 spectra (Fig. 2C), whereas the quantum yield resembles that of drFP583. Moreover, the ds/dr-hybrid is brighter than drFP583, primarily due to the substantially increased extinction (Table 1).

3.3. Generation of a 'purple' mutant and its spectral properties

We performed error-prone PCR to introduce site-independent amino acid substitutions into the ds/dr-hybrid protein structure. An optimal PCR mutational frequency was determined by visually screening numerous colonies. The generation of a wide diversity of phenotypes was the criteria for PCR optimization. The optimal mutational frequency mutagenesis rate generated an average of 2 or 3 amino acid substitutions per protein. Several dozen recombinants were sequenced, 80% contained two non-synonymous mutations, while 20% contained three mutations.

Although the first round of mutagenesis resulted in several different mutants that exhibited some brightness variation,

Table 1
Comparative spectral properties of drFP583, dsFP593, and their mutants

Protein name	Absorbance maximum (nm)	Emission maximum (nm)	Maximum extinction coefficient	Relative quantum yield[a]	Relative brightness[a]
drFP583	558	583	22 500	0.29	0.24
dsFP593	573	593	21 800	0.11	0.09
ds/dr hybrid	573	592	35 000	0.24	0.30
ds/drFP616	583	616	N/D[b]	N/D[b]	N/D[b]

[a]As compared to the quantum yield and brightness of *A. victoria* GFP [12].
[b]Not determined due to strong precipitation of the purified protein in solution.

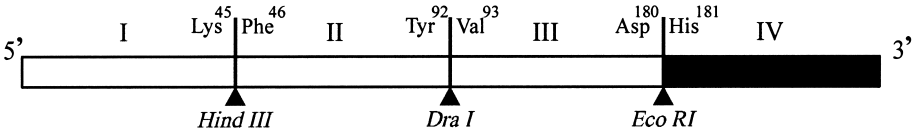

Fig. 3. Schematic diagram of the ds/dr-hybrid gene. The open bar corresponds to the dsFP593 gene fragment, the shaded bar corresponds to the drFP583 gene fragment.

these variations were not significantly different from the ds/dr-hybrid parent. Therefore, we pooled the mutant genes together and used this mixture in a second round of mutagenesis.

Screening a secondary mutagenesis library showed a notable increase in the number of non-fluorescent and weakly fluorescent colonies, while a few colonies with a different fluorescence hue appeared. Sequence analysis of these clones showed they all contained the same combination of amino acid substitutions: K15Q, K84M, and T224S (amino acid numeration corresponds to GFP) (Fig. 1). The spectral properties of the mutant form, designated ds/drFP616, were investigated in detail. The solution of purified ds/drFP616 protein was intensely purple-colored. The protein precipitated significantly from solution without, nevertheless, any loss of color or fluorescence in the precipitate. Precipitation could not be prevented or diminished by adding detergents or salts. The fraction that remained soluble was stable. The precipitation of ds/drFP616 protein was observed in vitro only, whereas in vivo the protein expression was quite similar to drFP583 and dsFP593, both in *E. coli* and in mammalian cells (data not shown).

The excitation spectrum for ds/drFP616 has a more or less structured major peak with a maximum at 583 nm, while the emission maximum is at 616 nm (Fig. 2D). In comparison to the parental ds/dr-hybrid (Fig. 2C), this is a 23 nm red shift. Notable features of the ds/drFP616 spectra are the remarkably increased Stock's shift (35 vs. initial 20 nm) and the very weak excitation at 280 nm. Quantitative spectral characteristics of the ds/drFP616 protein were not determined due to protein instability in solution; however, the fluorescence intensity in vivo (both in bacteria and mammalian cells) was not significantly different from drFP583- or dsFP593-expressing constructs.

4. Discussion

In this work we isolated a new gene, dsFP593, from a non-bioluminescent *Discosoma* species, which encodes a red-shifted fluorescent protein with an emission maximum at 593 nm. By means of shuffling different fragments between the dsFP593 and drFP583 [6] genes, we generated a library of combinatorial mutants. The brightest fluorescent variant was used as a template for subsequent random mutagenesis. Finally, a mutant protein ds/drFP616 was obtained having a unique emission maximum at 616 nm, which exceeds the emission maximum of all known fluorescent proteins. The expression of ds/drFP616 was demonstrated in a mammalian cell culture (Fig. 2).

The initial ds/dr-hybrid and its derivative ds/drFP616 only differ by three amino acid substitutions. Surprisingly, such a relatively small difference shifted the emission maximum by more than 20 nm. Of the three amino acid substitutions – K15Q, K84M, and T224S – the first (K15Q) corresponds to an external residue (i.e. its side chain is located outside the globule) and should not significantly affect fluorophore properties. Therefore, the spectral shift should be due to one of the latter two replacements (both of which occurred inside the protein globule), or be a cumulative effect of both.

Unlike drFR583 and dsFR593, ds/drFP616 precipitates from concentrated solutions, which may indicate that the whole scaffold of ds/drFP616 is altered in comparison to ds/dr-hybrid. Alternatively, some fraction of ds/drFP616 may have folded incorrectly in *E. coli* and thus become unstable in solution.

ds/drFP616 is the most red-shifted fluorescent protein to date. Its unique spectral properties are promising for different biological application. For instance, ds/drFP616 could become the most appropriate partner for GFPs in double and triple labeling systems. The longer wave length emission and virtual absence of excitation in the green spectral band should preclude any spectral overlap and background fluorescence, which allows the simultaneous detection of red and green fluorescent proteins in vivo.

Acknowledgements: We are especially grateful to Dr. Aleksandr P. Savitsky for spectral data generation and Dr. Nancianne Knipfer for her help in manuscript preparation. We also thank Dr. K.A. Lukyanov and Dr. Y.A. Labas for fruitful discussion.

References

[1] Kendall, J.M. and Badminton, M.N. (1998) Trends Biotechnol. 16, 216–224.
[2] Tsien, R. and Prasher, D. (1998) in: Green Fluorescent Protein Properties, Applications, and Protocols (Chalfie, M. and Kain, S., Eds.), pp. 97–118, Wiley-Liss, New York.
[3] Heim, R. and Tsien, R.Y. (1996) Curr. Biol. 6, 178–182.
[4] Heim, R., Cubitt, A.B. and Tsien, R.Y. (1995) Nature 373, 663–664.
[5] Ormö, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y. and Remington, S.J. (1996) Science 273, 1392–1395.
[6] Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L. and Lukyanov, S.A. (1999) Nature Biotechnol. 17, 969–973.
[7] Mazel, C.H. (1995) Mar. Ecol. Prog. Ser. 120, 185–191.
[8] Catala, R. (1959) Nature 183, 949.
[9] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[10] Matz, M., Shagin, D., Bogdanova, E., Britanova, O., Lukyanov, S., Diatchenko, L. and Chenchik, A. (1999) Nucleic Acids Res. 27, 1558–1560.
[11] Yang, F., Moss, L.G. and Phillips Jr., G.N. (1996) Nature Biotechnol. 14, 1246–1251.
[12] Ward, W.W. (1998) in: Green Fluorescent Protein: Properties, Applications, and Protocols (Chalfie, M. and Kain, S., Eds.), pp. 45–75, Wiley-Liss, New York.