A strategy for the generation of non-aggregating mutants of *Anthozoa* fluorescent proteins

Yurii G. Yanushevich[a], Dmitry B. Staroverov[a], Alexander P. Savitsky[b], Arkady F. Fradkov[a], Nadya G. Gurskaya[a], Maria E. Bulina[a], Konstantin A. Lukyanov[a], Sergey A. Lukyanov[a,*]

[a] *Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry RAS, Miklukho-Maklaya 16/10, 117997 Moscow, Russia*
[b] *Institute of Biochemistry RAS, Leninsky pr. 33, 117071 Moscow, Russia*

Received 4 October 2001; revised 21 November 2001; accepted 21 November 2001

First published online 19 December 2001

Edited by Giulio Superti-Furga

Abstract Recently, we cloned several fluorescent proteins of different colors homologous to *Aequorea victoria* green fluorescent protein, which have great biotechnological potential as in vivo markers of gene expression. However, later investigations revealed severe drawbacks in the use of novel fluorescent proteins (FPs), in particular, the formation of tetramers (tetramerization) and high molecular weight aggregates (aggregation). In this report, we employ a mutagenic approach to resolve the problem of aggregation. The elimination of basic residues located near the N-termini of FPs results in the generation of non-aggregating versions of several FPs, specifically, drFP583 (DsRed), DsRed-Timer, ds/drFP616, zFP506, zFP538, amFP486, and asFP595. © 2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: Green fluorescent protein; DsRed; Mutagenesis

1. Introduction

Recently, fluorescent proteins (FP) homologous to the well-known green fluorescent protein (GFP) were identified in *Anthozoa* species [1–3]. Due to the extensive popularity of GFP and mutants as in vivo markers of gene expression, protein interactions and trafficking, as well as molecular sensors, the discovery of novel FPs of different colors encouraged the further development of FP-based techniques. However, later studies revealed a number of disadvantages of novel FPs, specifically, the formation of tetramers [4,5], and subsequently high molecular weight aggregates.

The first disadvantage, tetramerization, complicates and consequently limits applications based on fusion of FP to a target protein. However, oligomerization does not limit the use of FPs as markers of gene expression. Details of the mechanism of tetramerization have been described using the red FP drFP583 (DsRed) as an example [6,7].

The second disadvantage, aggregation, may impede all possible applications due to considerable cellular toxicity. Aggregation renders proper targeting to cell compartments, investigation of protein interactions, and FRET-based applications impossible. To date, practically nothing is known about the possible causes and mechanisms of FP aggregation.

In this study, we employ site-directed mutagenesis of basic residues located near the N-termini of FPs to resolve the problem of aggregation. The effectiveness of this strategy is demonstrated in several *Anthozoa* FPs of different colors.

2. Materials and methods

2.1. Generation and characterization of mutant proteins

Site-directed mutagenesis was performed by PCR with primers containing appropriate target substitutions. All mutants were cloned between *Bam*HI and *Hin*dIII restriction sites of the pQE30 vector (Qiagen). Recombinant proteins were 6×Histidine-tagged to contain the sequence 'MRHHHHHHGS' instead of the first Met. After overnight expression in *Escherichia coli*, FPs were purified using Talon metal affinity resin (Clontech). SDS–PAGE analyses revealed that proteins were at least 95% pure.

To visualize aggregation, standard discontinuous 10% SDS–PAGE was employed. Protein aliquots were diluted with standard sample buffer and loaded on the gel without heating. Following electrophoresis, fluorescence was examined under UV illumination.

2.2. Expression in mammalian cell lines

For mammalian cell expression, all parental and non-aggregating mutants were cloned into pEGFP-C1 vector (Clontech) in lieu of the EGFP-coding region, between *Age*I and *Bgl*II restriction sites. The 293T cell-derived highly transfectable packaging line, Phoenix Eco, was transfected using the standard calcium–phosphate method (with the Life Technologies calcium phosphate transfection system).

3. Results and discussion

3.1. FP aggregation

Aggregation generally occurs during heterologous over-expression of FPs in both bacterial and eukaryotic cells. Although examples of aggregating FPs have been cited in the literature [5,8,9], the majority of failed experiments prob-

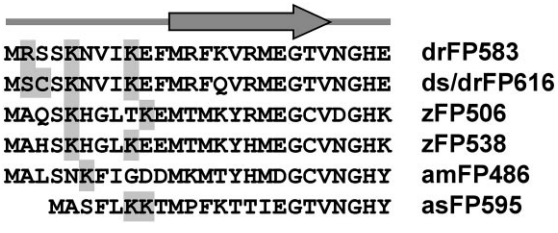

Fig. 1. Multiple alignment of N-terminal regions of the FPs. The arrow above the sequences represents the first β-sheet in the proteins, based on GFP and drFP583 structures. Amino acid residues substituted in the non-aggregating mutants are shaded.

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

0014-5793/02/$22.00 © 2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
PII: S0014-5793(01)03263-X

Table 1
Amino acid substitutions in parental and non-aggregating mutants

Wild type protein	Parental mutant [reference]	Differences between wild-type protein and parental mutant	Target substitutions in non-aggregating mutant
drFP583	'E57' (V105A, I161T, S197A) [unpublished]	faster and more complete folding in *E. coli*	R2A, K5E, K9T
drFP583	'Timer' (V105A, S197T) [13]	change color with time	R2A, K5E, K9T
dsFP593/drFP583	ds/drFP616 [2]	red-shifted fluorescence	S2del, C3del, K5E, K9T
zFP506	N66M [unpublished]	1.8-fold increase in brightness in *E. coli*	K5E, K10E
zFP538	M129V [unpublished]	more complete folding in *E. coli*	K5E, K9T
amFP486	K68M [unpublished]	1.5-fold increase in brightness in *E. coli*	K6E
asFP595	'M35-5' (F4L, K12R, F35L, T68A, F84L, A143S, K163E, M202L) [unpublished]	five-fold increase in brightness in *E. coli*, in comparison with mutant T68A, A143S [3]	K6T, K7E

ably remain unpublished. From our experience, FPs sometimes form large fluorescent granules inside transfected cells. More commonly, FP aggregation in cells results in 'smearing' of the fluorescent picture (see Fig. 3, left photos), so that nuclei and nucleoli are usually not visible. In these cases, even fluorescence that is concentrated in the central part of the cells is observed.

Exact kinetics of aggregation in vivo is unclear because it is difficult to measure this process within living cells. Probably, aggregation depends on FP concentration since brighter cells usually display more pronounced FP aggregates. Nevertheless, the aggregation picture can be observed even in low fluorescent cells as soon as the signal becomes visible. This indicates that the threshold value of FP concentration sufficient for aggregation is rather low.

Aggregation of purified FPs is additionally observed in vitro. For example, almost all *Anthozoa* FPs partially precipitate from solution, phosphate-buffered saline (PBS) without any loss of color or fluorescence. To visualize aggregation of purified FPs, we used 'pseudo-native' protein electrophoresis, based on the discontinuous SDS–PAGE of non-heated protein samples [4]. Under these conditions, FPs retain not only fluorescent properties, but also super-molecular structure: high molecular weight-aggregated proteins remain at the top of the gel, while oligomers migrate as bands of the high molecular weights.

3.2. Possible causes of aggregation

The recently solved X-ray structure of drFP583 (DsRed) [6,7] may provide clues on the possible reasons for aggregation of FP proteins. Since FP aggregates retain bright fluorescence, it is expected that they contain properly folded native protein molecules. Examples of native protein aggregation are very rare and the mechanisms of this phenomenon are poorly understood to date [10]. Aggregation may be caused by 'sticky' hydrophobic patches on the molecular surface [11]. However, the outer surface of tetrameric drFP583 does not contain pronounced hydrophobic areas that may result in strong interactions between tetramer, and therefore this explanation seems unlikely for the protein. Alternatively, protein aggregation may be explained by electrostatic interactions between positively and negatively charged surfaces [12,13]. A computer calculation of the electrostatic potential of tetrameric drFP583 showed that the protein surface is mostly negatively charged, except for a short N-terminal region of each monomer that contains a group of positively charged amino acid residues. It is speculated that each DsRed tetramer can form up to four salt bridges with adjacent tetramers. Once created, this net-like 'polymeric' structure should be very stable, due to four valencies for electrostatic interactions.

3.3. Mutagenesis

The above hypothesis highlights basic residues nearby N-termini as a possible cause of aggregation. Accordingly, in an attempt to generate non-aggregating variants of *Anthozoa* FPs, we substituted Lys and Arg residues close to the N-terminus with negatively charged or neutral residues (Fig. 1).

Red fluorescent drFP583 was the first protein subjected to mutagenesis. An improved mutant of DsRed (the commercially available version of drFP583 with altered codon usage optimized for expression in mammalian cells), designated E57, was used as a parental gene (Table 1). Mutants of E57 con-

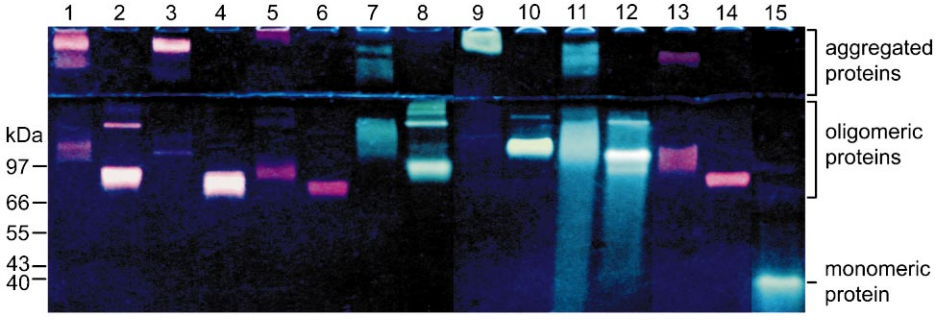

Fig. 2. Pseudo-native gel-electrophoresis of parental FPs (odd lanes) and their non-aggregating mutants (even lanes). The photograph was taken under UV illumination. Aggregated proteins are observed in the stacking gel. Oligomeric proteins migrate through the separating gel as bands of high molecular weights (expected MW of FP monomers and tetramers are about 27 and 108 kDa). Molecular weight standards are shown on the left of the gel. Lanes: 1 – DsRed mutant E57; 2 – E57-NA; 3 – DsRed mutant Timer; 4 – Timer-NA; 5 – ds/drFP616; 6 – ds/drFP616-NA; 7 – zFP506 mutant N66M; 8 – zFP506-N66M-NA; 9 – zFP538 mutant M129V; 10 – zFP538-M129V-NA; 11 – amFP486 mutant K68M; 12 – amFP486-K68M-NA; 13 – asFP595 mutant M35-5; 14 – M35-5-NA; 15 – EGFP.

taining the substitutions R2A, K5E, K9T (in different combinations) were generated. After expression in *E. coli* and purification, these proteins were analyzed by pseudo-native PAGE (see above). All mutants displayed lower levels of aggregation in comparison with parental E57, and the R2A substitution appeared to have the strongest impact on this outcome (not shown). A mutant, denoted E57-NA, containing all three substitutions (R2A, K5E, and K9T) that showed no aggregation (Fig. 2) and was very similar to E57 in terms of excitation–emission maxima, fluorescence brightness and maturation speed, was selected as the optimal protein.

E57-NA displayed an excellent fluorescence image in mammalian cells (Fig. 3). In most cells, nuclei and nucleoli were clearly visible, and cell borders and processes were well-defined. In contrast, the fluorescence of E57-expressing cells was smeared, with no visible intracellular structures. The borders of the fluorescent signal often did not coincide with cell borders and processes, so that cells looked rounded. Therefore, both in vivo and in vitro tests confirmed that the E57-NA protein was low-aggregating. Importantly, pilot trials of E57-NA in other laboratories showed greatly decreased toxicity of this protein in comparison with both DsRed and E57, during expression in cell lines (B. Angres, Clontech, Palo-Alto, CA, USA, personal communication), *Xenopus* embryos (A. Zaraisky, IBCh, Moscow, Russia, personal communication), and plants (A. Touraev, IMG, Vienna Biocenter, Vienna, Austria, personal communication). Now E57-NA is commercially available from Clontech as DsRed2.

Recently, an interesting mutant of DsRed entitled 'Timer', because it changes color with time, was described [14]. To generate a non-aggregating version of this protein, we employed the three substitutions mentioned above. The novel mutant, Timer-NA, was generated (Table 1), which possessed practically the same maturation properties with color change as parental Timer, but did not form aggregates on pseudo-native PAGE (Fig. 2). In mammalian cells, the differences between Timer and Timer-NA were analogous to those between E57 and E57-NA (Fig. 3).

The next FP targeted was ds/drFP616, which displays red-shifted fluorescence with a peak at 616 nm [2]. The protein was generated by the shuffling of two red FPs (dsFP593 and drFP583), followed by random mutagenesis. However, ds/drFP616 showed extremely high aggregation [2]. The mutation of two Lys residues at positions 5 and 9 at the N-terminal region of ds/drFP616 as for E57 and Timer (Fig. 1, Table 1), resulted in a significant decrease in the amount of aggregated protein on pseudo-native PAGE, although residual aggregation was still detected in the Lys mutant, ds/drFP616-K5E/K9T (not shown). After the screening of *E. coli* clones producing protein, one was selected that displayed complete absence of aggregation on pseudo-native PAGE (Fig. 2). Sequence analyses revealed that this clone contains two additional mutations in the N-terminal region of the protein (Ser-2 and Cys-3, deleted accidentally during cloning procedures). When expressed in eukaryotic cells, the mutant, designated ds/drFP616-NA, showed significant improvement in fluorescence image, similar to E57-NA and Timer-NA. In contrast to the bright but completely unstructured blot-like image of parental ds/drFP616, ds/drFP616-NA was more evenly distributed in nuclei and cytoplasm (Fig. 3).

The mutagenesis strategy described above was subsequently applied to four FPs of different colors: green zFP506, yellow

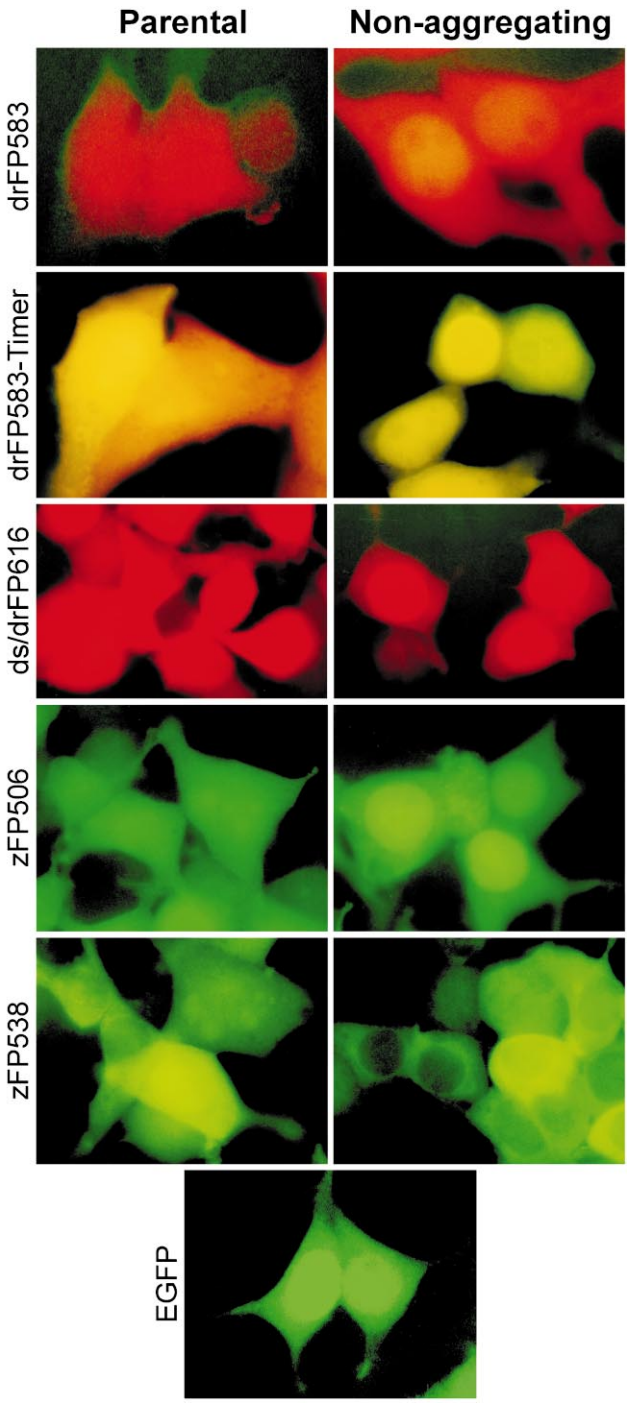

Fig. 3. Fluorescence images of cells expressing parental (left column) and corresponding non-aggregating (right column) FPs (see Table 1 for details). Protein names are shown on the left. EGFP-expressing cells are shown for comparison (below) as a well-known non-aggregating FP.

zFP538, red asFP595, and cyan amFP486 proteins, based on the improved mutants generated earlier by random mutagenesis (Table 1). When expressed in *E. coli*, mutant proteins displayed greater brightness and faster and more complete protein folding, in comparison with corresponding wild-type proteins (unpublished data). However, the introduced substitutions had no influence on the aggregation properties of these FPs. In an attempt to decrease aggregation tendency,

all lysines near the N-termini of the proteins were mutated (Fig. 1, Table 1). In vitro analyses of resulting secondary protein mutants confirmed no aggregation (Fig. 2). Additionally, all four non-aggregating mutants (zFP506-N66M-NA, zFP538-M129V-NA, amFP486-K68M-NA, and asFP595-M35-5-NA) displayed clear improvement in fluorescence images in mammalian cells, analogous to E57-NA, Timer-NA, and ds/drFP616-NA (Fig. 3, images for amFP486 and asFP595 are not shown).

In summary, we conclude that basic residues near the N-termini of *Anthozoa* FPs play a prominent role in the formation of protein aggregates. A number of examples of the significant effect of single amino acid substitutions on protein aggregation are documented [10–13]. Similarly, substitution of one to three residues in FPs led to a considerable increase in protein solubility. Possibly, these positively charged residues form ionic bonds between tetramers that result in FP conglomerates.

Acknowledgements: We are grateful to Dr. S.J. Remington for communication of crystallographic data prior to publication. This work was supported by Clontech Laboratories Inc. and the Russian Foundation for Fundamental Research (Grant 01-04-49037).

References

[1] Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L. and Lukyanov, S.A. (1999) Nat. Biotechnol. 17, 969–973.
[2] Fradkov, A.F., Chen, Y., Ding, L., Barsova, E.V., Matz, M.V. and Lukyanov, S.A. (2000) FEBS Lett. 479, 127–130.
[3] Lukyanov, K.A., Fradkov, A.F., Gurskaya, N.G., Matz, M.V., Labas, Y.A., Savitsky, A.P., Markelov, M.L., Zaraisky, A.G., Zhao, X., Fang, Y., Tan, W. and Lukyanov, S.A. (2000) J. Biol. Chem. 275, 25879–25882.
[4] Baird, G.S., Zacharias, D.A. and Tsien, R.Y. (2000) Proc. Natl. Acad. Sci. USA 97, 11984–11989.
[5] Mizuno, H., Sawano, A., Eli, P., Hama, H. and Miyawaki, A. (2001) Biochemistry 40, 2502–2510.
[6] Wall, M.A., Socolich, M. and Ranganathan, R. (2000) Nat. Struct. Biol. 7, 1133–1138.
[7] Yarbrough, D., Wachter, R.M., Kallio, K., Matz, M.V. and Remington, S.J. (2001) Proc. Natl. Acad. Sci. USA 98, 462–467.
[8] Jakobs, S., Subramaniam, V., Schonle, A., Jovin, T.M. and Hell, S.W. (2000) FEBS Lett. 479, 131–135.
[9] Lauf, U., Lopez, P. and Falk, M.M. (2001) FEBS Lett. 498, 11–15.
[10] Wetzel, R. (1994) Trends Biotechnol. 12, 193–198.
[11] Eaton, W.A. and Hofrichter, J. (1990) Adv. Protein Chem. 40, 63–279.
[12] Serina, L., Bucurenci, N., Gilles, A.M., Surewicz, W.K., Fabian, H., Mantsch, H.H., Takahashi, M., Petrescu, I., Batelier, G. and Barzu, O. (1996) Biochemistry 35, 7003–7011.
[13] Himanen, J.P., Popowicz, A.M. and Manning, J.M. (1997) Blood 89, 4196–4203.
[14] Terskikh, A., Fradkov, A., Ermakova, G., Zaraisky, A., Tan, P., Kajava, A.V., Zhao, X., Lukyanov, S., Matz, M., Kim, S., Weissman, I. and Siebert, P. (2000) Science 290, 1585–1588.