**Protein Engineering** vol. 16 no. 12 pp. 1099–1105, 2003
DOI: 10.1093/protein/gzg146

# Functional tuning of a salvaged green fluorescent protein variant with a new sequence space by directed evolution

**Sung-Hun Nam[1], Ki-Hoon Oh[2], Geun-Joong Kim[3,4] and Hak-Sung Kim[1,4]**

[1]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, 373-1, Kusung-dong, Yusung-gu, Taejon, 305-701, [2]R&D Center of Bioproducts, Institute of Science and Technology, CJ Corp., 522-1, Dokpyong-ri, Majang-myon, Ichon, 467-810 and [3]Institute of Biotechnological Industry, College of Engineering, Inha University, 253, Yonghyun-dong, Nam-gu, Incheon, 402-751, Korea

[4]To whom correspondence should be addressed.
E-mail: hskim@mail.kaist.ac.kr or geunkim@inha.ac.kr

**We previously reported a method, designated functional salvage screen (FSS), to generate protein lineages with new sequence spaces through the functional or structural salvage of a defective protein by employing green fluorescent protein (GFP) as a model protein. Here, in an attempt to mimic a step in the natural evolution process of proteins, the functionally salvaged mutant GFP-I5 with new sequence space, but showing low fluorescence intensity and stability, was selected and fine-tuned by directed evolution. During a course of functional tuning, GFP-I5 was found to evolve rapidly, recovering the spectral traits to those of the parent GFPuv. The mutant 3E4 from the third round of directed evolution possessed four substitutions; three (F64L, E111V and K166Q) were at the original GFP gene and the other (K8N) at the inserted segment. The fluorescence intensity of 3E4 was ~28-fold stronger than GFP-I5, and other spectral properties were retained. Biochemical and biophysical investigations suggested that the fine-tuning by directed evolution led the salvaged variant GFP-I5 to a functionally favorable structure, resulting in recovery of stability and fluorescence. Site-directed mutagenesis of the mutated amino acid residues in both GFPuv and GFP-I5 revealed that each amino acid residue has a different effect on the fluorescence intensity, which implies that 3E4 adopted a new evolutionary path with respect to fluorescence characteristics compared with the parent GFPuv. Directed evolution in conjunction with FSS is expected to be used for generating protein lineages with new fitness landscapes.**

*Keywords*: directed evolution/functional salvage screen/GFP/sequence space/tuning

## Introduction

In recent years, a number of methods have been proposed and implemented to design and create proteins/enzymes, accelerating the practical applications of proteins/enzymes as well as the understanding of many intrinsic questions regarding structure–function relationships, folding processes and structural organization of proteins (Stemmer, 1994a,b; Dahiyat and Mayo, 1997; Miyazaki and Arnold, 1999; Bittker *et al*., 2002). Of them, the directed evolution approach has been widely used for protein design, and a lot of promising results have been reported to date (Crameri *et al*., 1996; Ness *et al*., 1999; Oh *et al*., 2002). This approach basically resembles the evolution process of proteins in nature, and offers some advantages over other conventional approaches (Arnold, 2001; Burton *et al*., 2002).

Directed molecular evolution has proven to be a powerful tool for evolutionary protein design, but a challenge still remains. The reported methods generally give rise to the mutations in the identical sequence space to their parent, and a library pool of variants having the same sequence space as that of the parent gene is subjected to screening (Voigt *et al*., 2000). Thus, most of the variants lie within the pre-existing and structurally fated sequence space, thereby excluding the possibility of creating protein lineages with new fitness landscapes. The fated sequence space might be a limitation in the evolutionary protein design, when considering that the diverse protein lineages have been generated through insertion, deletion, duplication and recombination in the natural evolution process of proteins (Yogev *et al*., 1995; Wang *et al*., 1997; Todd *et al*., 2001). In this regard, various strategies have been attempted to generate the protein lineages with new sequence spaces by mimicking the natural evolution process. Matsuura *et al.* described the random elongation mutagenesis that leads to a change in the sequence space by the addition of random peptides to the C-terminus of proteins (Matsuura *et al*., 1999). Homology-independent recombination methods, designated ITCHY (Ostermeier *et al*., 1999) and SHIPREC (Sieber *et al*., 2001), were reported. Sub-domain shuffling (Hopfner *et al*., 1998), domain or module grafting (Aphasizheva *et al*., 1998; Nixon *et al*., 1998) were also proposed.

As a relevant approach, we developed a method, designated functional salvage screen (FSS), to generate a pool of protein variants with different sequence spaces from the parent gene through a rescuing process of a defective protein by employing green fluorescent protein (GFP) as a model protein (Kim *et al*., 2001). This process was found to effectively generate a diverse pool of protein variants with new sequence space by the incorporation of randomly fragmented genomic DNA segments into the defined region(s) of the defective template. The resulting GFP variants indeed revealed a variable nature in structural stability and fluorescence as expected compared with the parent GFP. It is generally accepted that gene rearrangement through various mechanisms such as insertion, deletion, recombination and duplication, followed by the adaptation/selection process in an environment, is a root in the generation of a protein pool for natural evolution (Arnold *et al*., 2001; Todd *et al*., 2001; Kinch and Grishin, 2002). Thus, the functional salvage process might correspond to a process generating a pool of protein variants in nature.

In this work, in an effort to delineate a process of adaptation/selection in the natural evolution of proteins, the salvaged GFP-I5 was tuned by using directed evolution. GFP-I5 was generated from the defective template GFP∇172-3/176(+2) in

  

our previous work (Kim *et al*., 2001). This variant contained the rescuing segment of 12 amino acid residues (PNGAAPLKGRSV) between 175 and 177 residues at the fifth upper loop of GFPuv. It showed ~4.0% of the fluorescence intensity of the parent GFPuv and low functional stability. A third round of directed evolution resulted in the mutant 3E4 with a comparable fluorescence intensity and stability to that of GFPuv. To get some insights into the structural property and evolutionary path, site-directed mutagenesis and a biochemical study were conducted for both the evolved 3E4 and the parent GFPuv. Details are reported herein.

## Materials and methods

*Strain, proteins and plasmids*

*Escherichia coli* strain JM109 was used for the cloning and expression of GFP variants. Functionally salvaged GFP-I5 in pTrc-99A was chosen as a starting template for directed evolution. The wild-type GFPuv was originated from the vector pGFPuv (Clontech) and used as a control. The pMAL-c2 (NEB) and pQE-30 (Qiagen) vectors were employed as expression vectors for evolved GFP variants. *Escherichia coli* cells harboring either the wild-type or GFP variants were grown in Luria–Bertani (LB) medium at 37°C. Ampicillin was added to a concentration of 50 μg/ml when needed.

*Error-prone PCR and DNA shuffling for library construction*

To introduce mutations randomly into the GFP-I5 variant, error-prone (ep) PCR was carried out using *Taq* DNA polymerase in the presence of $MnCl_2$. The mutation frequency of ep-PCR was controlled by adding the two divalent metal ions, $Mn^{2+}$ and $Mg^{2+}$, at an appropriate concentration of 0.3 and 3 mM, respectively (Spee *et al*., 1993). The PCR mixture (100 μl) contained 30 ng GFP-I5 gene, 0.2–1.0 mM dNTP, 2 units *Taq* DNA polymerase, 3 mM $MgCl_2$, 0.3 mM $MnCl_2$ and 30 pmol of each primer: ngfp, 5′-ATATATAGAATTCATG-AGTAAAGGAGAAG-3′; and cgfph, 5′-ATATATAT-AAGCTTTTATTTGTAGAGCTC-3′ flanked by *Eco*RI and *Hin*dIII sites, respectively. PCR was conducted under the following conditions: one cycle of 94°C, 5 min; 40 cycles of 94°C, 1 min; 54°C, 1 min; 72°C, 40 s; one cycle of 72°C, 5 min. The amplified DNA fragments (~0.7 kb) were eluted from 0.8% low melting point agarose gel using a Bio101 kit and restricted with *Eco*RI and *Hin*dIII. The resulting fragments were ligated with pMAL-c2 vector and transformed into electrocompetent *E.coli* JM109 cells using Gene Pulser II (Bio-Rad). The transformed cells were spread onto LB agar plates with 50 μg/ml ampicillin, and then cultured at 37°C.

A potential pool of positive clones showing an improved fluorescence was selected from the mutant library of ep-PCR and further subjected to DNA shuffling. The *in vitro* recombination procedure was performed according to the method described elsewhere with slight modifications (Stemmer, 1994a,b). PCR products were eluted, mixed and randomly fragmented by DNase I. Digestion was carried out depending on the concentration of mixed DNA at 15°C and followed by inactivation at 60°C for 20 min. Fragmented DNA (<~150 bp) was eluted from 2.0% agarose gel using DE81 paper. The purified DNA fragments were resuspended in a PCR mixture containing 0.2 mM of each dNTP, 2 units *Taq* polymerase, 3 mM $MgCl_2$, 10× buffer per 100 μl. Reassembly of fragments was carried without primers using a PCR program of: 94°C, 5 min; 40 cycles of 94, 48 and 72°C, 1 min; 72°C, 7 min. Fragmented genes were assembled into diverse sizes and the product of this reaction was diluted 40× into a new PCR mixture. The full-length GFP genes were amplified with the same primers in a PCR of: 94°C, 5 min; 20 cycles of 94, 52 and 72°C, 1 min; 72°C, 7 min. *In vitro* recombined GFP variants were sub-cloned, transformed and then screened on the basis of the fluorescence intensity as described in our previous work (Kim *et al*., 2001). The evolved genes after a round of ep-PCR and DNA shuffling were again selected and subsequently used as the templates for the next round of DNA shuffling.

*Selection and preliminary characterization of evolved variants*

Of the mutant library from GFP-I5, the positive clones emitting relatively high fluorescence were selected by direct observation under UV excitation (365 nm) using a hand type UV lamp (Vilber Lourmat). In order to select more florescent clones, primarily screened pools from independent experiments were combined and further compared in the same plate. As a control, *E.coli* cells harboring either the wild-type GFPuv or GFP-I5 were grown in the same plate under identical conditions.

The selected clones were further analyzed in terms of protein expression and fluorescence property by using crude extracts. To examine the expression patterns and levels of GFP variants, the mutant clones were cultured with or without 0.1 mM IPTG induction. After harvesting the cells by centrifugation, the resulting pellets were resuspended in a lysis buffer, 20 mM Tris–HCl (pH 7.4) containing 1 mM EDTA, 200 mM NaCl, 1% Triton X-100 and 1 mM PMSF, and then disrupted by using sonicator (Sonics, Inc.). After clarification, equal amounts of cell lysate (~30 μg) were loaded into SDS–PAGE. For a simple purification and comparison, evolved GFP mutants were expressed as the N-terminal MBP-fused and His-tagged fusion proteins, and then analyzed under the identical conditions described above. The selected mutants from each round were sequenced by using an automated DNA sequencer (ABI, Model 377), and the protein concentration was estimated according to the procedure of the Bradford assay.

*Protein purification*

MBP-fused GFP variants were purified to apparent homogeneity using affinity column chromatography. The cells expressing GFP variants were cultivated in 10 ml of LB medium at 37°C and induced with 0.1 mM IPTG when the OD 600 nm reached ~0.4. After 3 h of cultivation, cells were harvested, then resuspended in a lysis buffer. After cell lysis, the crude extracts were clarified by centrifugation and subsequently loaded onto amylose resin equilibrated with a column buffer containing 20 mM Tris–HCl buffer (pH 7.4), 1 mM EDTA and 200 mM NaCl. The bound proteins were completely washed with 10 volumes of the column buffer and eluted with an elution buffer containing 10 mM maltose. As for the expressed proteins as His-tag in an expression vector pQE-30, the tagged proteins were purified by using Ni-NTA affinity column chromatography, and further purified to apparent homogeneity by gel filtration chromatography using the supplier recommended buffer.

*Factor Xa digestion*

The site-specific cleavage of MBP-fused GFP variants was conducted using Factor Xa. The purified MBP-fused GFP variants were concentrated to be ~1 mg/ml by using a Centricon (Amicon). Factor Xa (1 μg) was added and then

incubated for ~12 h at 20°C. After reaction, the cleavage pattern and homogeneity were analyzed by SDS–PAGE (10%).

*Relative mobility on native PAGE*

Purified GFP variants (~2 μg) were loaded into 10% native polyacrylamide gel in the absence of denaturing agent such as SDS. Electrophoresis was performed at a constant voltage of 150 V for 1.5 h. The gel was stained with Coomassie brilliant blue G250.

*Site-directed mutagenesis*

Site-directed mutagenesis was carried out for both GFPuv and GFP-I5 variant by using the overlap extension method as described elsewhere (Sambrook and Russell, 2001). Each set of four primers was needed to introduce a site-specific mutation into a template gene. One pair of primers, referred in the library construction, flanking the N- and C-terminal regions of GFP was used. The second pair of primers (R and F) was designed to amplify the DNA fragment that contained the target site for mutation. The resulting constructs were sub-cloned into the *Eco*RI–*Hin*dIII sites of pMAL-c2 vector, and screened from transformed cells by DNA sequencing. The site-directed mutants were cultured, harvested and then used for preparation of proteins according to the procedures described above.

*Biochemical and biophysical characterization*

The fluorescent characteristics of quantified GFP variants were investigated at 20°C using a spectrofluorometer (SLM-AMINCO 8100; Spectronic). Bandwidth and integration time were kept at 5 nm and 1.0 s, respectively. Excitation and emission spectra of mutated variants were scanned in the range of 320–440 and 470–570 nm, respectively. For the analysis of spectral properties, GFP variants were excited at the wavelength of 395 nm and emitted at 508 nm. All fluorescence intensities were represented as a relative value compared with that of wild-type GFPuv and measured in triplicate. In order to avoid either the quenching or hindrance by fused MBP, the spectral properties of the MBP-removed and His-tagged variants were also determined under the same conditions.

To acquire more information on the structural properties, the purified GFP variants were treated with a random proteinase K (100 μU) for 0–20 min at 20°C. The digestion patterns and resulting fragments were analyzed by SDS–PAGE. Tryptophan fluorescence of each His-tagged GFP variant was measured with excitation at 283 nm and an emission scan at 300–400 nm, respectively, by using a spectrofluorometer. The pre-packed gel filtration column, Superdex 200 HR 10/30 (Amersham Pharmacia), was used to determine the oligomeric structure or non-specific association of the evolved GFP variants. His-tagged GFP variants (~5 mg/ml) were loaded into the gel filtration column at a flow rate of 0.25 ml/min. An eluted fraction of each variant was sampled and further confirmed by their fluorescence.

## Results

*Directed evolution for the functional tuning of a salvaged mutant GFP-I5*

In an approach to mimic a process of protein evolution in nature, fine-tuning of a salvaged GFP variant was conducted by using directed evolution. Of various GFP variants generated by FSS in our previous work (Kim *et al*., 2001), GFP-I5 having an additional segment of 12 amino acid residues was selected. GFP-I5 was structurally unstable and thus showed a relatively low fluorescence intensity compared with the parent GFPuv. First, ep-PCR was carried out to obtain a diverse pool of GFP variants in the course of directed evolution. Approximately 70 000 clones were screened, and eight clones (1E1–8) emitting distinct fluorescence intensities compared with GFP-I5 were selected. The DNA sequencing of these mutants revealed a mutation rate in the range of 0.4–0.7% (3–5/741 bp). The most fluorescent mutant, 1E7, had the three mutations (K52I, F64L and E111V) in the GFP gene without any substitution in the inserted segment. The primarily screened pool of variants (1E1–8) was used as templates for the next round of DNA shuffling.

The first round of DNA shuffling resulted in a huge variant pool of approximately 100 000 clones. Of them, approximately 80 strains were primarily selected based on the cell growth and fluorescence intensity, and the best mutant 2E1 emitting highest fluorescence was isolated. 2E1 was found to possess the two mutations F64L and E111V by DNA sequencing. The second round of DNA shuffling was carried out with the primarily screened 80 strains from the first round, and the resulting pool of variants (approximately 100 000 clones) was subjected to the screening for fluorescence emission. From the first step of screening, 10 more fluorescent clones were chosen and then further analyzed for their fluorescence intensity with crude cell extracts by using a spectrofluorometer. As a result, the brightest mutant 3E4 was finally selected, and its DNA sequence was determined. 3E4 possessed four substitutions; three (F64L, E111V and K166Q) were at the original GFP gene and the other (K8N) at the inserted segment. The F64L mutation was previously reported (Palm *et al*., 1997; Patterson *et al*., 1997) and located in the central fluorophore region of GFP (Ormo *et al*., 1996; Yang *et al*., 1996). E111V and K166Q were found to be located on the middle of the fifth and eighth strands, respectively (Figure 1).

The selected 3E4 from the three rounds of directed evolution was expressed in a solid or solution culture for the preliminary comparison of whole cell fluorescence with GFPuv. As expected, *E.coli* cells expressing the 3E4 in a solid plate were distinctly fluorescent, and their intensity was indeed comparable to that of the parent GFPuv. The fluorescence level of cells in solution was also similar to that in a solid plate. In a further finding, no significant difference in the fluorescence intensity was observed between MBP-fused and non-fused 3E4 (data not shown).

*Expression and spectral properties of the evolved GFPs*

The evolved GFP variants (1E7, 2E1 and 3E4) selected from each round of directed evolution were expressed as the MBP-fused proteins. GFPuv was also expressed as the control. SDS–PAGE (10%) analyses of whole cell extracts showed that the expression levels of the evolved GFPs were quite comparable to that of GFPuv. A similar result was also observed even under conditions of IPTG induction (data not shown). These indicate that the increase in the fluorescence intensities is mainly due to their improved fluorescence characteristics rather than the enhanced levels of expression. Soluble fraction of the evolved variants was also examined for various fusion partners. The MBP-fused GFP variants exhibited a similar trend to the parent GFPuv under various induction conditions, whereas the His-tagged GFPs showed remarkably different expression patterns. Most of the His-tagged GFPuv, 2E1 and 3E4 were well expressed as a soluble form under various IPTG inductions. On the other hand, a major portion of the His-tagged GFP-I5 was detected in the insoluble fraction. When the soluble fraction

was analyzed, its fluorescence intensity was very low, accounting for ~4% of GFPuv. The inserted segment of 12 amino acid residues for functional salvage appears to cause a distortion of the structural conformation, consequently resulting in an inappropriate folding and insolubility of GFP-I5, as observed in the expression without folding assistance like MBP. When it is considered that protein solubility is closely linked with structural stability and folding of protein, this result strongly implies that fine-tuning by directed evolution led the salvaged GFP variant to a favorable conformation and ultimately increased the soluble fraction and fluorescence intensity.

The MBP-fused GFP variants were purified and further analyzed for their fluorescence properties at the protein level. Excitation spectra of each mutant were found to be almost similar to those of the parent GFPuv (Crameri *et al*., 1996), showing a maximum at around 395 nm (Figure 2A). Emission spectra were also obtained at the excitation of 395 nm, and their emission maximum was observed at 508–509 nm as shown in Figure 2B. The mean fluorescence intensities of the evolved variants were measured at their maximum excitation and emission wavelengths (Table I). The fluorescence intensity of GFP-I5 was ~4% of GFPuv. The three evolved mutants, 1E7, 2E1 and 3E4, gave fluorescence intensities of ~30.7, 36.7 and 111.5% of GFPuv, respectively. The fluorescence intensity of GFP-I5, therefore, increased ~28-fold in a new sequence space through a fine-tuning process using directed evolution.

To check the effect of MBP fusion at the N-terminus of GFP variants on the spectral property, the fluorescence intensities of GFP variants were determined in the presence and absence of MBP. Digestion of MBP-fused variants (1E7, 2E1 and 3E4) with Factor Xa resulted in respective fragments corresponding to the MBP and GFP variants (Figure 3, lanes 3–5). On the other hand, GFP-I5 gave rise to a different cleavage pattern, showing a smaller fragment than expected (Figure 3, lane 2). When the cleaved GFP variants, except for GFP-I5, were analyzed for their fluorescence, the fluorescence intensity and spectrum were similar to those with the fused MBP. His-tagged GFP variants were also tested, and a negligible effect on the spectral properties was observed.

*Mobility and stability of the evolved variants*

To further confirm the change in the conformation of the salvaged GFP-I5 through a fine-tuning process, the relative mobility of the evolved variants on PAGE was determined under reducing or non-reducing conditions. The evolved GFPs were detected as an apparent single band on SDS–PAGE, but the resolution on native PAGE resulted in either a single or two bands as shown in Figure 4A. In addition, a fluorescence-emitting band was distinct, depending on the evolved variants

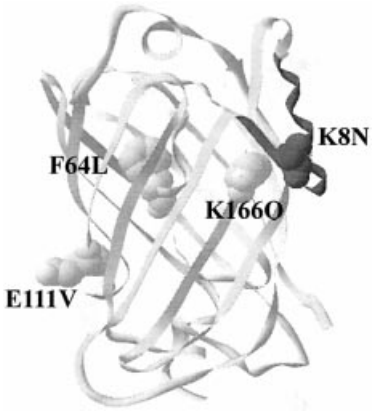


**Fig. 1.** Modified tertiary structure of GFP (PDB code 1EMA). The incorporated residues (PNGAAPLKGRSV) by FSS are represented as a dark ribbon. Each mutated residue by directed evolution is represented as a space filling mode. The figure was generated with the PDB viewer.

**Table I.** Spectral properties of GFP variants

| Variants | $\lambda_{max}$ (excitation, nm) | $\lambda_{max}$ (emission, nm) | Fluorescence intensity[a] |
|---|---|---|---|
| GFPuv | 397 | 508 | 100.0 |
| GFP-I5 | 396 | 509 | 4.0 ± 1.8 |
| 1E7 | 396 | 508 | 30.7 ± 13.3 |
| 2E1 | 397 | 509 | 36.7 ± 12.3 |
| 3E4 | 396 | 508 | 111.5 ± 23.5 |

[a]Each is a relative value compared with that of GFPuv and represents the average and deviation of triplicate measurements.

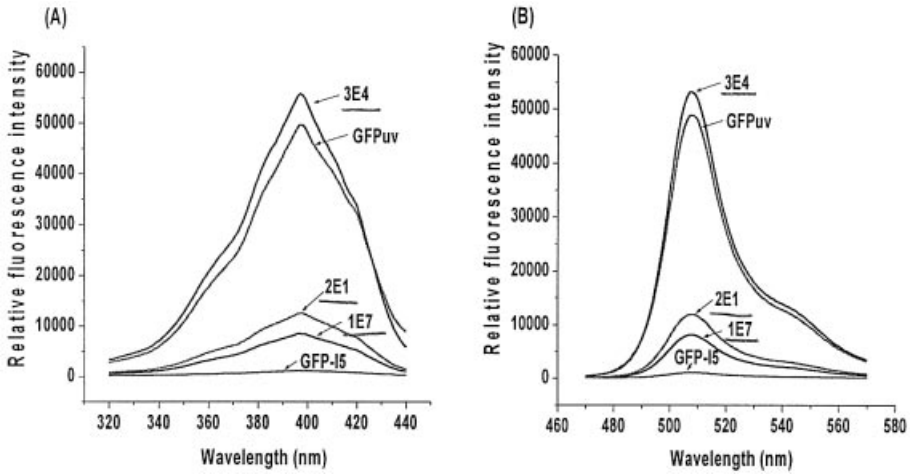


**Fig. 2.** The excitation and emission spectra of GFP variants. Excitation spectra (**A**) were scanned from 320 to 440 nm at an emission of 508 nm. Emission scans (**B**) were recorded from 470 to 570 nm at an excitation of 395 nm. The fluorescence intensity of each variant was represented as a relative value.

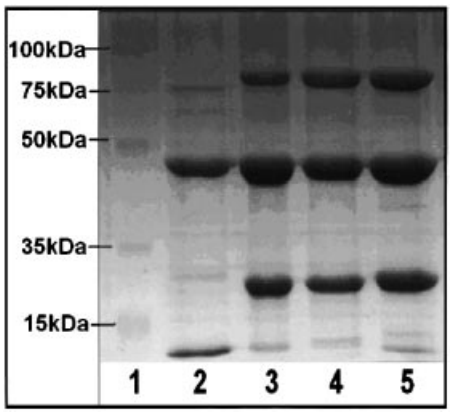


**Fig. 3.** SDS–PAGE analysis of the Factor Xa cleaved MBP fusion proteins. Lane 1, size marker; lane 2, GFP-I5; lane 3, 1E7; lane 4, 2E1; lane 5, 3E4.

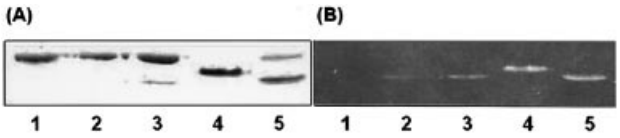


**Fig. 4.** Relative mobility of GFP variants on native PAGE (10%). (**A**) Coomassie brilliant blue stained native gel. (**B**) The photograph under UV excitation at 365 nm. Lane 1, GFP-I5; lane 2, 1E7; lane 3, 2E1; lane 4, 3E4; lane 5, GFPuv. The fluorescence of lanes 2–5 is observed in the lower band.

(Figure 4B). Unexpectedly, GFPuv was separated into two bands on native PAGE, and of them, only a protein band with higher mobility was observed to emit fluorescence. On the other hand, the evolved 3E4 migrated as a single fluorescent band which was located between the separated two bands of GFPuv. In the case of GFP-I5, the portion of the fluorescence-emitting protein was almost negligible, even though the fluorescence intensity of GFP-I5 is ~4% of GFPuv. It is interesting that the evolved 3E4 with only an inserted segment of 12 amino acids and mutations in four amino acids (F64L, E111V, K166Q and K8N) has a different migration pattern from other GFP variants including GFPuv. It might be due to the overall conformational changes by introduced mutations rather than the charge difference between GFP variants. Based on these observations, it seems that directed evolution leads the salvaged GFP-I5 to a more favorable conformation for fluorescence, resulting in an increase in the fluorescence-emitting portion of protein.

The stability of the evolved GFPs against proteolysis was investigated. As a general protease for random proteolysis, a serine protease K was employed due to its broad substrate spectrum towards aliphatic, aromatic and other hydrophobic amino acids (Parsell and Sauer, 1989; Chiang *et al*., 2001). Each GFP was incubated with Protease K for 20 min, and the digestion was analyzed by SDS–PAGE using a densitometer. The remaining portion of GFPuv was ~73%, whereas >62% of GFP-I5 was fragmented. In the case of the 3E4, ~69% of the protein remained intact, showing a similar digestion pattern to GFPuv (data not shown). This result suggests that the evolved 3E4 is more resistant to proteolysis than the salvaged variant GFP-I5.

**Table II.** Effect of site-directed mutagenesis on fluorescence intensity

| Substitution[b] | Fluorescence intensity[a] | |
|---|---|---|
| | GFPuv | GFP-I5 |
| No mutation | 100.0 | 4.0 |
| F64L | 121.3 | 7.2 |
| E111V | 116.4 | 0.8 |
| K166Q | 124.8 | 3.5 |
| K(Ins8)N | – | 0.7 |
| F64L, E111V, K166Q | 71.2 | – |
| F64L, E111V, K166Q, K8N | – | 111.5 |

[a]Relative value compared with that of GFPuv.
[b]Mutagenic primers: 64R, GAACACCATAAGAGAGAGTAGTGACAAGTG; 64F, CACTTGTCACTACTCTCTCTTATGGTGTTC; 111R, CCTTCAAACTTGACTACAGCACGCGTCTTG; 111F, CAAGACGCGTGCTGTAGTCAAGTTTGAAGG; 166R, GTTGTGGCGAATTTGGAAGTTAGCTTTG; 166F, CAAAGCTAACTTCCAAATTCGCCACAAC; Ins8R, GAACGGATCTGCCATTCAAAGGGGCGGC; Ins8F, GCCGCCCCTTTGAATGGCAGATCCGTTC.

### *Site-directed mutagenesis*

In order to analyze the effect of each mutated amino acid residues on the fluorescence intensity, mutation was incorporated individually into both GFP-I5 and GFPuv. As shown in Table II, the F64L mutation resulted in a slight and ~2-fold increase in the fluorescence intensity of GFPuv and GFP-I5, respectively. The mutation E111V caused a drastic decrease in the fluorescence intensity of GFP-I5, whereas a marginal increase was observed in GFPuv. When the lysine residue K166 was replaced with glutamine, there were no considerable effects on both proteins. The specific mutation, K8N, within the inserted segment also had a negative influence on the fluorescence intensity of GFP-I5. The substitution of an individual amino acid residue caused a negligible effect on the fluorescence intensity of GFP-I5, but a slight increase was observed in GFPuv. Interestingly, however, the simultaneous incorporation of three (F64L, E111V and K166Q) and four mutations (three residues plus K8N) into GFPuv and GFP-I5, respectively, resulted in a different feature: the fluorescence intensity of GFP-I5 increased ~28-fold. On the other hand, the replacement of three residues in GFPuv (GFPuv-3M) caused a reduction of fluorescence intensity. This result implies that the mutated amino acid residue has a different effect on the fluorescence intensity, depending on the sequence space of proteins. In other words, even the same amino acid residue of proteins has a different role on the trait as their sequence space changes.

### *Tryptophan fluorescence and oligomerization*

To get some insights into the conformation of the evolved GFPs, we investigated the tryptophan fluorescence and oligomeric structure of each GFP variant. The mutant GFPuv-3M, which was generated by the substitution of three residues in GFPuv as described in Table II, was also tested. One tryptophan residue exists in GFP as the 57th amino acid. It is located in the upper loop region of β-can, connecting the third strand and second helices. The fluorescence emission spectra of the tryptophan residue of His-tagged GFPs were recorded with excitation at 283 nm, as described elsewhere (Maeda *et al*., 2001; Vivian and Callis, 2001). As shown in Figure 5, both 3E4 and GFPuv-3M showed a maximum

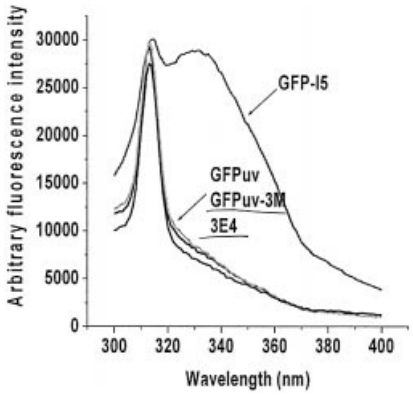


**Fig. 5.** Tryptophan emission spectra of GFP variants (GFP-I5, GFPuv, GFPuv-3M and 3E4). The proteins were excited at 283 nm and the fluorescence was scanned at 300–400 nm. The fluorescence intensity is in arbitrary units.

emission at around 313 nm, which is consistent with GFPuv. The emission spectrum of GFP-I5, however, was observed at 313 nm, with a shoulder peak at 330 nm. It has been reported that the tryptophan fluorescence is sensitive to the conformational change of the protein, especially the local environment around the residue. This result partly suggests that the environmental conditions (such as the exposure to the surface) around the tryptophan residue of GFP-I5 might be different from those of GFPuv and other variants. It is likely that the substitution of amino acid residues generated through a tuning process of GFP-I5 into 3E4 led to the restoration of tryptophan fluorescence to the level of the native state of GFPuv.

It was reported that the wild-type GFP dimerizes at high protein concentration (Tsien, 1998). Oligomeric structures of 3E4 and GFPuv-3M were analyzed and then compared with those of GFP-I5 and GFPuv by using gel filtration chromatography. As a result, GFP variants were separated into two distinct peaks corresponding to ~60 and ~30 kDa, and these appear to be dimeric and monomeric forms of protein, respectively (data not shown). Typically, the dimerized portion of GFPuv was estimated to be 30% under our experimental conditions. Both GFPuv-3M and 3E4 revealed a higher tendency to dimerization than GFPuv, resulting in ~60% of the dimerized portion. In contrast, GFP-I5 showed a marginal portion of dimerization (<10%).

## Discussion

We have demonstrated that the fluorescence characteristics of the functionally salvaged GFP variant can be recovered by functional tuning using directed evolution. This result suggests that directed evolution in conjunction with FSS is effective for the generation of protein lineages with new fitness landscapes. It is generally accepted that the natural evolution of proteins involves a process of generating a pool of diverse variants, followed by a process of adaptation/selection (Todd *et al*., 1999; Arnold *et al*., 2001; Kinch and Grishin, 2002). In this regard, FSS and the following directed evolution seem to mimic the natural evolution process of proteins; generation of lineages with new sequence spaces and functional/structural adaptation. Site-directed mutagenesis of both GFPuv and GFP-I5 revealed that even the same amino acid residue of proteins has a different effect on their traits as the sequence space of proteins changes, which strongly implies that proteins have a different evolutionary or maturation pathway, depending on the sequence space. This result also suggests a possibility of broadening the function of pre-existing amino acid residues through the change in the sequence space of the protein (Abedi *et al*., 1998; Baird *et al*., 1999).

GFP-I5 was generated from the defective GFP template by FSS in our previous work (Kim *et al*., 2001). This variant was found to contain the inserted segment of 12 amino acid residues, and showed a low fluorescence and structural stability compared with the parent GFPuv. It was likely that the change in the conformation of GFP occurred due to the inserted fragment and fluorescence intensity declined significantly to ~4% of GFPuv. Proteolysis of GFP-I5 using protease K revealed that >62% of GFP-I5 was digested, whereas the intact portion of GFPuv was ~73%. This result clearly supports the notion that GFP-I5 has a susceptible conformation to protease attack.

A third round of directed evolution of GFP-I5 generated the mutant 3E4 emitting a higher fluorescence than GFPuv. The evolved 3E4 was found to possess four mutations; three were at the original GFP gene and the other at the inserted segment. As expected, the substitution of four amino acid residues enabled GFP-I5 to recover the fluorescence intensity to higher than that of GFPuv. In other words, GFP-I5 has evolved to have a favorable conformation emitting higher fluorescence intensity. The digestion pattern of 3E4 with protease K showed that 3E4 has a similar pattern to GFPuv, which supports the above presumption. The fluorescence spectrum of the tryptophan residue also indicates that 3E4 exerts an adequate environment around the fluorophore of the protein for fluorescence emission. This implies that the perturbed conformation of GFP-I5 can be easily tuned to a correct conformation by directed evolution. In this regard, a step of directed evolution seems to correspond to a process of adaptation/selection in the natural evolution process.

The effect of the mutated amino acid residues on the fluorescence intensity was investigated for both GFP-I5 and GFPuv by site-directed mutagenesis, and an interesting result was observed. When four mutations (F64L, E111V, K166Q and K8N) were incorporated simultaneously into GFP-I5, the fluorescence intensity increased ~28-fold, as shown in 3E4. On the other hand, the counterpart mutant GFPuv-3M, which was generated from GFPuv, showed a decrease in the fluorescence intensity compared with GFPuv. This result suggests a possibility that the strategy attempted here can be applied to the generation of novel proteins/enzymes, since the same amino acid residue of a protein might have a different effect on its trait as its sequence space changes (Bergdoll *et al*., 1998; Bogarad and Deem, 1999; Riechmann and Winter, 2000). The finding that 3E4 revealed a unique migration on native PAGE compared with GFPuv might be an example of a protein with a novel feature. It is as yet unclear why 3E4 exhibits a different migration from other evolved variants and GFPuv, and this remains to be further explored. Further directed evolution of a salvaged GFP variant using a different screening method might be a way of creating a protein with novel properties. This suggestion is strictly based on the observation that proteins with different sequence spaces follow different evolutionary paths in nature. Determination of a three-dimensional structure might detail the effect of an inserted segment on the conformation and its interaction with the original backbone of GFP.

Directed evolution approaches have been widely used for tailoring proteins/enzymes with desired properties. Most of them, however, generally introduce the mutations on the target gene, and a pool of variants lie within the pre-existing and structurally fated sequence space (Voigt *et al*., 2000). Therefore, new strategies might be desirable to explore a new functional space available for beneficial mutations. A number of approaches to generate protein lineages with new sequence spaces have been reported (Matsuura *et al*., 1999; Ostermeier *et al*., 1999; Sieber *et al*., 2001), and we also developed a strategy, designated FSS. FSS was observed to generate a diverse pool of protein variants with a new sequence space when GFP was employed as a model protein. Based on this work, it is likely that the generation of protein lineages by FSS and then structural tuning using directed evolution might be a reasonable route to create a protein with novel properties. The approach attempted here can be an alternative for evolutionary protein engineering, and it should also be invaluable in elucidating the natural evolution of proteins.

## Acknowledgements

This work was supported by grants from the Brain Korea 21 Project of the Korean Ministry of Education, and the National Research Laboratory program of the Ministry of Science and Technology, South Korea.

## References

Abedi,M.R., Caponigro,G. and Kamb,A. (1998) *Nucleic Acids Res.*, **26**, 623–630.

Aphasizheva,I.Y., Dolgikh,D.A., Abdullaev,Z.K., Uversky,V.N., Kirpichnikov,M.P. and Ptitsyn,O.B. (1998) *FEBS Lett.*, **425**, 101–104.

Arnold,F.H. (2001) *Nature*, **409**, 253–257.

Arnold,F.H., Wintrode,P.L., Miyazaki,K. and Gershenson,A. (2001) *Trends Biochem. Sci.*, **26**, 100–106.

Baird,G.S., Zacharias,D.A. and Tsien,R.Y. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 11241–11246.

Bergdoll,M., Eltis,L.D., Cameron,A.D., Dumas,P. and Bolin,J.T. (1998) *Protein Sci.*, **7**, 1661–1670.

Bittker,J.A., Le,B.V. and Liu,D.R. (2002) *Nat. Biotechnol.*, **20**, 1024–1029.

Bogarad,L.D. and Deem,M.W. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 2591–2595.

Burton,S.G., Cowan,D.A. and Woodley,J.M. (2002) *Nat. Biotechnol.*, **20**, 37–45.

Chiang,C.F., Okou,D.T., Griffin,T.B., Verret,C.R. and Williams,M.N. (2001) *Arch. Biochem. Biophys.*, **394**, 229–235.

Crameri,A., Whitehorn,E.A., Tate,E. and Stemmer,W.P. (1996) *Nat. Biotechnol.*, **14**, 315–319.

Dahiyat,B.I. and Mayo,S.L. (1997) *Science*, **278**, 82–87.

Hopfner,K.P., Kopetzki,E., Kresse,G.B., Bode,W., Huber,R. and Engh,R.A. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 9813–9818.

Kim,G.J., Cheon,Y.H., Park,M.S., Park,H.S. and Kim,H.S. (2001) *Protein Eng.*, **14**, 647–654.

Kinch,L.N. and Grishin,N.V. (2002) *Curr. Opin. Struct. Biol.*, **12**, 400–408.

Maeda,H., Mizutani,O., Yamagata,Y., Ichishima,E. and Nakajima,T. (2001) *J. Biochem. (Tokyo)*, **129**, 675–682.

Matsuura,T., Miyai,K., Trakulnaleamsai,S., Yomo,T., Shima,Y., Miki,S., Yamamoto,K. and Urabe,I. (1999) *Nat. Biotechnol.*, **17**, 58–61.

Miyazaki,K. and Arnold,F.H. (1999) *J. Mol. Evol.*, **49**, 716–720.

Ness,J.E., Welch,M., Giver,L., Bueno,M., Cherry,J.R., Borchert,T.V., Stemmer,W.P. and Minshull,J. (1999) *Nat. Biotechnol.*, **17**, 893–896.

Nixon,A.E., Ostermeier,M. and Benkovic,S.J. (1998) *Trends Biotechnol.*, **16**, 258–264.

Oh,K.H., Nam,S.H. and Kim,H.S. (2002) *Protein Eng.*, **15**, 689–695.

Ormo,M., Cubitt,A.B., Kallio,K., Gross,L.A., Tsien,R.Y. and Remington,S.J. (1996) *Science*, **273**, 1392–1395.

Ostermeier,M., Shim,J.H. and Benkovic,S.J. (1999) *Nat. Biotechnol.*, **17**, 1205–1209.

Palm,G.J., Zdanov,A., Gaitanaris,G.A., Stauber,R., Pavlakis,G.N. and Wlodawer,A. (1997) *Nat. Struct. Biol.*, **4**, 361–365.

Parsell,D.A. and Sauer,R.T. (1989) *J. Biol. Chem.*, **264**, 7590–7595.

Patterson,G.H., Knobel,S.M., Sharif,W.D., Kain,S.R. and Piston,D.W. (1997) *Biophys. J.*, **73**, 2782–2790.

Riechmann,L. and Winter,G. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 10068–10073.

Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY, pp. 13.36–13.39.

Sieber,V., Martinez,C.A. and Arnold,F.H. (2001) *Nat. Biotechnol.*, **19**, 456–460.

Spee,J.H., de Vos,W.M. and Kuipers,O.P. (1993) *Nucleic Acids Res.*, **21**, 777–778.

Stemmer,W.P. (1994a) *Nature*, **370**, 389–391.

Stemmer,W.P. (1994b) *Proc. Natl Acad. Sci. USA*, **91**, 10747–10751.

Todd,A.E., Orengo,C.A. and Thornton,J.M (1999) *Curr. Opin. Chem. Biol.*, **3**, 548–556.

Todd,A.E., Orengo,C.A. and Thornton,J.M (2001) *J. Mol. Biol.*, **307**, 1113–1143.

Tsien,R.Y. (1998) *Annu. Rev. Biochem.*, **67**, 509–544.

Vivian,J.T. and Callis,P.R. (2001) *Biophys. J.*, **80**, 2093–2109.

Voigt,C.A., Kauffman,S. and Wang,Z.G. (2000) *Adv. Protein Chem.*, **55**, 79–160.

Wang,Y., Goligorsky,M.S., Lin,M., Wilcox,J.N. and Marsden,P.A. (1997) *J. Biol. Chem.*, **272**, 11392–11401.

Yang,F., Moss,L.G. and Phillips,G.N.,Jr (1996) *Nat. Biotechnol.*, **14**, 1246–1251.

Yogev,D., Watson-Mckown,R., Rosengarten,R., Im,J. and Wise,K.S. (1995) *J. Bacteriol.*, **177**, 5636–5643.

*Received March 31, 2003; revised October 30, 2003; accepted October 31, 2003*