

Generation of protein lineages with new sequence spaces by functional salvage screen

Geun-Joong Kim, Young-Hoon Cheon¹, Min-Soon Park¹, Hee-Sung Park¹ and Hak-Sung Kim^{1,2}

Department of Molecular Science and Technology, Ajou University, San5, Woncheon-dong, Paldal-gu, Suwon, 442-749 and ¹Department of Biological Sciences, Korea Advanced Institute of Science and Technology, 373-1, Kusung-dong, Yusung-gu, Taejeon, 305-701, Korea

²To whom correspondence should be addressed.
E-mail: hskim@mail.kaist.ac.kr

A variety of different methods to generate diverse proteins, including random mutagenesis and recombination, are currently available and most of them accumulate the mutations on the target gene of a protein, whose sequence space remains unchanged. On the other hand, a pool of diverse genes, which is generated by random insertions, deletions and exchange of the homologous domains with different lengths in the target gene, would present the protein lineages resulting in new fitness landscapes. Here we report a method to generate a pool of protein variants with different sequence spaces by employing green fluorescent protein (GFP) as a model protein. This process, designated functional salvage screen (FSS), comprises the following procedures: a defective GFP template expressing no fluorescence is first constructed by genetically disrupting a predetermined region(s) of the protein and a library of GFP variants is generated from the defective template by incorporating the randomly fragmented genomic DNA from *Escherichia coli* into the defined region(s) of the target gene, followed by screening of the functionally salvaged, fluorescence-emitting GFPs. Two approaches, sequence-directed and PCR-coupled methods, were attempted to generate the library of GFP variants with new sequences derived from the genomic segments of *E.coli*. The functionally salvaged GFPs were selected and analyzed in terms of the sequence space and functional properties. The results demonstrate that the functional salvage process not only can be a simple and effective method to create protein lineages with new sequence spaces, but also can be useful in elucidating the involvement of a specific region(s) or domain(s) in the structure and function of protein.

Keywords: functional salvage screen/green fluorescent protein/new lineage/protein engineering/sequence space

Introduction

Advances in protein engineering have accelerated the understanding of a number of intrinsic questions regarding proteins, such as structure–function relationships, folding processes and structural organization. From a practical standpoint, it has contributed to the improvement of protein properties for a number of applications (Nixon *et al.*, 1998). Recently, directed or *in vitro* evolution techniques based on sequential mutation and random recombination have been proved to be a very effective tool to generate the proteins with greater potential (Christians *et al.*, 1999; Joo *et al.*, 1999; Xirodimas and Lane,

1999) or novel function (Zhang *et al.*, 1997; Matsumura and Ellington, 2001). The directed evolution technique has also been used to address some issues regarding the regulation of protein molecules by endogenous or exogenous modulators (Doi and Yanagawa, 1999) and functional expression of protein in a host where intrinsic expression is restricted by genetic, translational and folding systems (Cramer *et al.*, 1996; Kim *et al.*, 2000).

Conventional mutagenesis and directed evolution techniques generally give rise to mutations on the target gene in a random fashion and a library of variants having the same sequence space as that of the parent gene is subjected to screening (Stemmer, 1994). Thus, most of the variants lie within the pre-existing and structurally fated sequence space, excluding the chances of creating protein lineages with new fitness landscapes. Recently, random elongation mutagenesis was attempted to generate the variants through the addition of random peptide tails to the C-terminus of the enzyme, providing a clue that changing the sequence space of a protein by incorporation of a random sequence could generate diverse protein lineages (Matsuura *et al.*, 1999). In line with this, a number of approaches including sub-domain swapping (Hopfner *et al.*, 1998; Kumar and Rao, 2000), domain or module grafting (Greenfeder *et al.*, 1995; Aphasizheva *et al.*, 1998; Nixon *et al.*, 1998), DNA homology-independent recombination (Ostermeier *et al.*, 1999) and scaffold design based on combinatorial methods (Altamirano *et al.*, 2000) have been carried out for the generation of new protein lineages.

Here we present a method, designated functional salvage screen (FSS), to generate protein lineages with new sequence spaces through functional or structural salvage of a defective protein by employing green fluorescence protein (GFP) as a model protein. The functional salvage process started with a construction of the defective GFP expressing no fluorescence by genetically disrupting a predetermined region(s) of the protein. The defective template was designed to be unable to recover the functional trait (i.e. fluorescence emission) *in vivo* through simple insertion of base(s). Thus, only a recombination between a defective template in a predetermined region(s) and DNA segments derived from *Escherichia coli* chromosome could rescue the protein function. For the generation of a library of GFP variants from the defective template, two independent approaches, sequence-directed and PCR-coupled recombination, were attempted. The functionally salvaged, fluorescence-emitting variants with considerable stability were selected and analyzed with respect to sequence space and functional properties.

Materials and methods

Strains and plasmids

E.coli JM109 was used for the cloning and expression of the GFP variants. The pGFPuv vector (CLONTECH) was used as a source for the wild-type GFPuv gene. Plasmid pTrc-99A, used for library construction, was obtained from Amersham

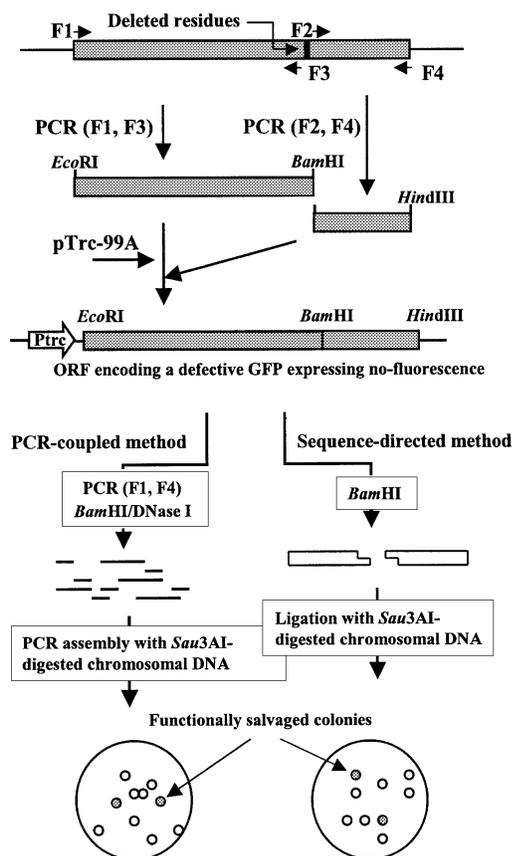


Fig. 1. Typical procedure for the functional salvage screen. The defective GFP genes expressing no fluorescence are constructed by genetically disrupting a predetermined region(s) of the protein as starting templates. For the functional salvage process, two independent approaches, PCR-coupled and sequence-directed methods, are attempted as described in the Materials and methods section. From the constructed library, the functionally salvaged, fluorescence-emitting clones are screened.

Pharmacia Biotech. The pMAL-c2 vector (New England Biolabs) was used to express the GFP variants as a fusion protein with the *E. coli* maltose binding protein (MBP). *E. coli* cells were grown at 37°C in Luria-Bertani (LB) broth supplemented with ampicillin (50 µg/ml) when needed.

Construction of the defective GFP templates

The defective GFP templates were constructed as depicted schematically in Fig. 1. In order to remove the residue encoding 176V with additional one or two bases from the wild-type GFPuv, PCR was carried out using the primers, F1(5'-GCGAATTCAGTAAAGGAGAAGAAGACTTTTCATCGGA-3')/F4(5'-AGTAAGCTTATTTGTAGAGCTCATCCATGCCATG-3') and F3(5'-GCGGATCCATCTTCAATGTTGTGGCG-3') flanked by *EcoRI* and *BamHI* sites, respectively. The amplified DNA fragment was cloned into the pTrc-99A vector, yielding pGFPN. The wild-type GFPuv gene was again amplified by PCR with the following two sets of primers: F2(-1)(5'-GAGGATCCAAGTACTAGCAGACCATTATCAACAAA-3')/F4(5'-AGTAAGCTTATTTGTAGAGCTCATCCATGCCATG-3') and F2(-2)(5'-GAGGATCCAAGTACTAGCAGACCATTATCAACAAA-3')/F4(5'-AGTAAGCTTATTTGTAGAGCTCATCCATGCCATG-3'). Underlined sequences indicate the *BamHI* and *HindIII* sites, respectively. Each of the amplified fragments was inserted into the *BamHI* and *HindIII* sites of the pGFPN, resulting in pGFP ∇ 176(+1) and pGFP ∇ 176(+2), respectively. Two more templates, GFP ∇ 172-3/176(+1) and

GFP ∇ 172-3/176(+2), were constructed from GFP ∇ 176(+1) and GFP ∇ 176(+2), respectively, by additional deletion of two residues 172E and 173D using PCR, according to a similar procedure to that described above. The resulting two constructs were designated pGFP ∇ 172-3/176(+1) and pGFP ∇ 172-3/176(+2), respectively.

To construct a defective template for the dual-point salvage process, we further deleted a region from 129D to 138G of the GFP ∇ 176(+2) by PCR using two sets of primers: F1/P1 (5'-GTGGATCCAATACCTTTTAACTC-3') and P2 (5'-ATGGATCCACAACTCGAGTC-3')/F4. The resulting template was designated GFP ∇ 129-138/176(+2).

All the constructed templates were tested to confirm whether they are able to express the chromoprotein under various induction conditions by fluorescence microscopy (Abedi *et al.*, 1998; Matz *et al.*, 1999).

Library construction for the functional salvage screen

A library for the functional salvage screen was constructed by two procedures, a sequence-directed (Jappelli and Brenner, 1999) and a PCR-coupled method (Kikuchi *et al.*, 1999). The sequence-directed process was basically similar to the shotgun cloning procedure. Four constructs containing each of the defective templates, pGFP ∇ 176(+1), pGFP ∇ 176(+2), pGFP ∇ 172-3/176(+1) and pGFP ∇ 172-3/176(+2), were linearized by digestion with *BamHI* and then eluted from agarose gel (0.8%). For oligonucleotide pools to be incorporated into the defective template genes, chromosomal DNA isolated from the *E. coli* MG1655 was digested with *Sau3AI* and the fragments ranging from 25 to 500 bp were eluted by using a DNA clean-up purification system (Promega). The resulting fragments were ligated with each of the previously linearized templates and then transformed into *E. coli* JM109 by electroporation.

For the PCR-coupled process, four defective GFPuv templates were amplified from each of the four constructs, pGFP ∇ 176(+1), pGFP ∇ 176(+2), pGFP ∇ 172-3/176(+1) and pGFP ∇ 172-3/176(+2), by PCR using two primers, F1 and F4. The amplified fragments were cleaved by *BamHI* and then further digested with DNase I. The DNA fragments ranging from 50 to 150 bp were excised and eluted from agarose gel (2.5%) and then reassembled with the *Sau3AI*-digested chromosomal DNA (25-500 bp) by PCR (94°C, 1 min; 45 ± 0.2°C/cycle, 1 min; 72°C, 40 s; total 40 cycles). The defective template (10-20 ng/µl) was mixed with chromosomal DNA in ratios of about 1:1, 1:0.5, 1:0.1 and 1:0.01. The reassembled DNA fragments were amplified by PCR (94°C, 1 min; 50.5°C, 1 min; 72°C, 40 s; total 25 cycles) with two primers, F1 and F4. The resulting DNA fragments (0.5-2 kb) were purified, digested and cloned into the plasmid pTrc-99A and the constructs were transformed into *E. coli* JM109 by electroporation. To expand the mutation space, the stringency at the reassembling PCR step was modulated by either increasing or decreasing the annealing temperature.

For construction of a library from the template GFP ∇ 129-138/176(+2) through the dual-point salvage process, we incorporated the DNA fragments simultaneously into the two points (129D-138G and 176V(+2) regions) of the template by using the PCR-coupled procedure as described above.

Screening of the functionally salvaged variants

Transformants were grown on agar plates in the presence and/or absence of IPTG and positive clones emitting fluorescence were first screened by direct observation under UV excitation

(365 nm) using a hand-type UV lamp (Vilber Lourmat). As a control, *E. coli* cells harboring each of the defective templates were grown under the same conditions.

With the primarily isolated clones, we further screened the salvaged GFPs for structural stability *in vivo* as described in our previous work (Doi and Yanagawa, 1999). *E. coli* cells harboring the salvaged GFPs were cultivated in Luria–Bertani medium at 37°C and induced with 0.2 mM of IPTG when the OD_{600 nm} reached about 0.5. After 2 h of cultivation, chloramphenicol (100 µg/ml) was added to the medium to block further protein synthesis and an aliquot of 0.5 ml was removed at the indicated times and analyzed by SDS–PAGE. A protein band corresponding to each of the functionally salvaged GFPs was scanned with a gel scanner. The clones showing a distinct protein band were isolated and the incorporated DNA segments were identified by DNA sequencing.

Protein purification and characterization

For the rapid purification and clear comparison with wild-type GFPuv, the genes encoding the selected GFP variants were subcloned into the *EcoRI* and *HindIII* sites of pMAL-c2 vector, expressed as MBP fusion proteins and purified (Kapust and Waugh 1999) according to the general procedure of the manufacturer.

The cleavage of fusion proteins was carried out to test the appropriate folding and thus accessibility of a site-specific protease factor Xa. During the reaction for 40 h at 8–10°C, aliquots were removed and analyzed by SDS–PAGE, along with measurements of fluorescence intensity. The cleaved GFP variants were further purified by reapplying them on to the amylose resin and concentrated by dialysis against 20 mM Tris–HCl (pH 8.0) buffer containing 5% glycerol.

Spectral properties of the salvaged GFPs were investigated at 25°C using a spectrofluorimeter (Amico). Bandwidths and integration times were kept at 5 nm and 0.5 s, respectively. Maximum emission and excitation wavelengths were first determined by automatic dual scanning and then confirmed by manual scanning at a predetermined wavelength. All scans were conducted in duplicate for various protein concentrations (5–500 µg/ml).

Results

Principle of functional salvage screen (FSS)

The FSS process consists of the following steps as depicted in Figure 1. (i) A predetermined region(s) of a target protein, such as residue(s), domain(s) or module(s), is genetically disrupted by deletion, insertion, inversion or duplication, which results in a functionally defective protein. (ii) A library is generated from the defective gene by incorporation of the diverse nucleotide pool, such as fragmented *E. coli* chromosomal DNA, into the defined region(s) of the defective gene through either sequence-directed insertion or PCR-coupled recombination. (iii) The variants that recover the functional or structural trait are screened by the functional or genetic screening system. In step (i), any region(s) of the parental protein can be selected randomly or rationally. Step (ii) involves the incorporation of the segments, which can be either synthetic oligonucleotides or digested genomic DNA. The incorporation event is carried out through a sequence specific insertion resembling a shotgun cloning strategy (Jappelli and Brenner, 1999) or homologous recombination using a PCR-like process (Kikuchi *et al.*, 1999). The critical point in the present strategy lies in designing a defective gene whose functional trait

should not be recovered by simple insertion of base(s), which guarantees the generation of protein lineages with new sequence spaces.

Construction of the defective GFP templates for functional salvage screen

We employed a well-characterized green fluorescent protein (GFPuv) with enhanced fluorescence intensity as a model protein (Kim *et al.*, 2000), which allows the rapid and simple detection of the functionally salvaged proteins from the defective one. A schematic procedure for construction of the defective GFP templates is described in Figure 1. Two GFP variants, expressing no fluorescence, were first constructed by introducing functional defect into the parent GFPuv. Considering the fact that the deletion in the β-strand of GFP (amino acid residues 175–188) results in the loss of fluorescence (Li *et al.*, 1997), we carried out internal deletion in this region. The nucleotides of GFPuv gene (gttc and gtcca) that encode 176V with successive one and two bases were deleted from the parent gene and the resulting templates were designated GFP∇176(+1) and GFP∇176(+2). In an effort to exclude the generation of variants salvaged by simple insertion of one or two base(s), we attempted further deletion of amino acid residues around the defective region and found that deletion of 172E and 173D leads to complete loss of fluorescence emission. Based on this finding, two more templates, GFP∇172–3/176(+1) and GFP∇172–3/176(+2), were constructed from GFP∇176(+1) and GFP∇176(+2), respectively, by additional deletion of the two residues 172E and 173D. This deletion is expected not only to expand the sequence space of the functionally salvaged GFPs, but also to eliminate the possibility that the defective GFPs might be salvaged by simple insertion of one or two base(s) countervailing the frame-shift mutations of GFP∇176(+1) and GFP∇176(+2).

All the constructed templates were tested as to whether they are able to emit the fluorescence. As a result, no fluorescence emission with excitation at 365–405 nm was observed. As the deletions in the templates gave rise to the frame-shift mutations, the constructs were defective in both structural and functional aspects. The resulting templates, therefore, were considered to be suitable as the starting target genes for the functional salvage process.

Library construction and screening of the salvaged GFPs

In an effort to explore the feasibility of the functional salvage process and to obtain the diverse pool of salvaged GFPs, we constructed the libraries by applying two procedures, sequence-directed and PCR-coupled recombination, to each of the four defective templates, GFP∇176(+1), GFP∇176(+2), GFP∇172–3/176(+1) and GFP∇172–3/176(+2). In the sequence-directed method, the oligonucleotide pool of *Sau3A1*-digested genomic DNA was incorporated into the target site. For construction of a more diverse library, we attempted a PCR strategy where a defective template, which had been cleaved with *Bam*HI, was further digested with DNase I and reassembled with *Sau3A1*-digested chromosomal DNA. For both cases, the pool of the randomly digested genomic DNA ranging from 25 to 500 bp was used for the functional salvage process.

Total colonies ranging in number from 3500 to 6000 were obtained from each library and about 50 colonies were randomly picked and analyzed in terms of the percentage of oligonucleotide-inserted genes and size variation to investigate the genetic diversity of each library. As a result, about 70%

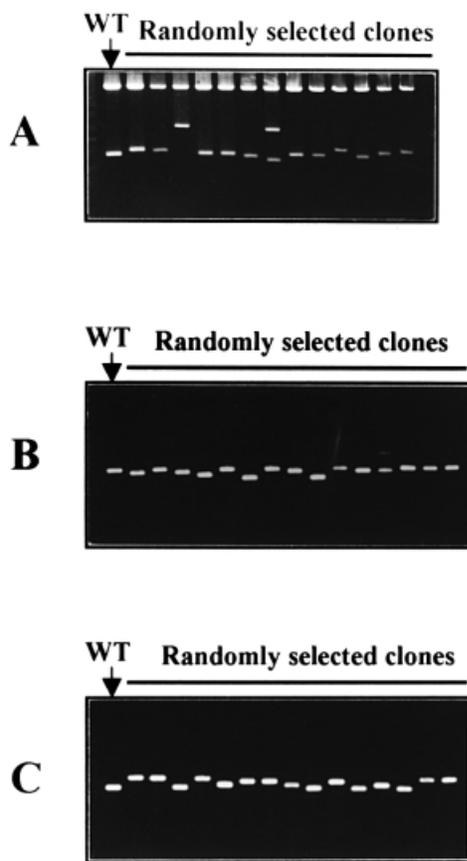


Fig. 2. Diversity and size variation of the GFP variants in the libraries constructed from the defective templates by the two different methods. Clones from each library were randomly selected and their plasmids were purified and analyzed on agarose gel (1.0%). (A) Clones from the library constructed from GFP Δ 176(+2) by the sequence-directed method. (B) Clones from the library generated from GFP Δ 172-3/176(+2) by the PCR-coupled method (annealing temperature started at 45°C). (C) Clones from the library by using the same process as in (B), except that the annealing temperature started at 30°C.

of the colonies from the library generated by the sequence-directed method were found to possess the insert, whereas the library obtained by the PCR-coupled process revealed an insertion frequency ranging from 12 to 41%, depending on the template and stringency of PCR conditions used (Abecassis *et al.*, 2000). With randomly selected colonies from the library constructed from GFP Δ 176(+2) by the sequence-directed method, we analyzed the size variation of the genes. As shown in Figure 2A, the size of the variants was found to be considerably diverse. The pool of genes generated from GFP Δ 172-3/176(+2) by the PCR-coupled method also revealed a random set of genes, as can be seen in agarose gel electrophoresis (Figure 2B). When the library was constructed from GFP Δ 172-3/176(+2) under less stringent PCR conditions (lower temperature and rate of annealing), a pool with more diverse size variation was obtained, as shown in Figure 2C. Similar trends were observed for the libraries constructed from the other templates (data not shown).

In order to screen the functionally salvaged GFPs, the transformants were grown on a plate containing 0.2 mM of IPTG as an inducer and subjected to two successive screenings. First, the variants that emitted the fluorescence persistently were selected. From the primarily selected pool, we further isolated the variants that were resistant to proteolysis *in vivo*

(Frank *et al.*, 1996; Doi and Yanagawa, 1999) as the wild-type GFPuv, which resulted in 375 clones showing both distinct fluorescence and structural stability. We classified them according to their origin of template and found that the clones from the library of GFP Δ 176(+2) and GFP Δ 172-3/176(+2) showed a more diverse nature in incorporated segment size and fluorescence intensity than those of GFP Δ 176(+1) and GFP Δ 172-3/176(+1). Based on the results, two sets of the salvaged library, generated by using the sequence-directed process for the GFP Δ 172-3/176(+2) and by using the PCR-coupled process for the GFP Δ 176(+2), were selected for further analyses.

Sequence diversity of the salvaged variants

From the library of GFP Δ 176(+2), 11 clones exhibiting different fluorescence intensities were randomly selected and analyzed. From them, we finally isolated nine clones having a distinct diversity based on restriction site analysis and SDS-PAGE. With the identical procedure, we also chose 12 clones from the library of GFP Δ 172-3/176(+2). The genes from the selected clones were retransformed into a freshly prepared host (*E. coli* JM109) and grown in LB medium to confirm the formation of chromoprotein.

The amino acid sequences of the incorporated segments were deduced from DNA sequencing (Figure 3). Of nine clones from the library of GFP Δ 176(+2), two variants had the same fragment with the GFP Δ 7. In addition, two variants from the GFP Δ 172-3/176(+2) library were found to contain the identical sequence with the GFP Δ 9 and GFP Δ 13. It is likely that the sequence of the segments rescuing the defective GFPs was not biased, but randomly distributed. The site where the random fragment of genomic DNA is incorporated for functional salvage of the defective GFP was tolerable against the insertion of various segments, ranging from nine (GFP Δ 7) to 52 amino acid residues (GFP Δ 13). Apparently, comparison of the library from GFP Δ 176(+2) with that from GFP Δ 172-3/176(+2) revealed that the sequence-directed salvage process results in more diverse lineages than the PCR-coupled method. It is worth noting that the incorporated segments of two libraries have different sequences, which strongly implies that further deletion in the template GFP Δ 172-3/176(+2) might affect the sequence space of the segment to be incorporated.

In the segments rescuing the defective GFPs, the proportion of hydrophobic amino acid residues to the total residues decreased from 45–60 to 30–45% as the segment size increased. As a further finding, hydrophobic amino acid residues with bulky side chains were relatively rare in the segments. This trend was almost the same in the clones derived from the two libraries, with the exception of GFP Δ 12. The incorporated segment of the GFP Δ 12 contained more hydrophobic amino acid residues compared with other salvaging sequences. The proportion of hydrophobic residues in the segments was observed to be closely linked with the stability of the resulting GFP variants, affecting their fluorescence intensity. Further analyses of the 10–12 clones with variable segment sizes from each library yielded a similar distribution of amino acid composition, which thus led us to believe that the sequence space of the rescuing segments is essentially random and has a distinct trend.

Characterization of the functionally salvaged GFPs

For a rapid and clear comparison, the seven genes from the GFP Δ 176(+2) library were subcloned into pMAL-c2 vector to express them as MBP fusion proteins. Most of clones were

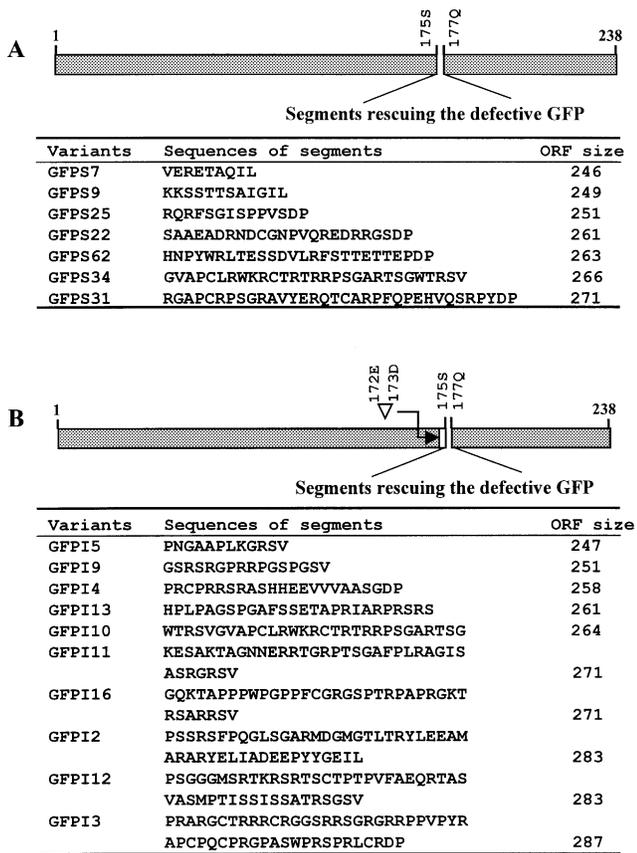


Fig. 3. Amino acid sequences of the rescuing segments. The clones expressing the fluorescent GFP with considerable stability were selected from the constructed library and the amino acid sequences of the incorporated segments were deduced. (A) Sequences of the rescuing segments from the library constructed from GFP ∇ 176(+2) by the PCR-coupled process. (B) Sequences of the rescuing segments from the library generated from GFP ∇ 172–3/176(+2) by the sequence-directed method.

also found to be fluorescent in the fusion state, except for the GFPS34, and similar results were obtained when induced at 37°C. From the SDS-PAGE analysis, all of the MBP-fused proteins were mainly expressed in the soluble fraction (>60%) with calculated molecular masses and accounted for about 10–15% of the total cell protein.

To investigate the functional properties of the salvaged GFPs, the fusion proteins were further purified to apparent homogeneity by using amylose resin. The two variants (Cont-1 and Cont-2) which emitted fluorescence but were excluded in the screening step because of their low structural stability were also expressed as fusion proteins and purified according to the same procedure for comparison. The purified fusion proteins showed different migration rates in SDS-PAGE, depending on the size of the incorporated segments (Figure 4A). To get some insight into the functional expression of the fusion proteins, the purified proteins were treated with factor Xa and analyzed by SDS-PAGE. As shown in Figure 4B, seven fusion proteins of the salvaged GFPs were efficiently cleaved and separated into their respective domains, which indicates that both domains fold independently and are accessible to site-specific protease factor Xa. In the case of the two variants (Cont-1 and Cont-2), however, the respective domains were not detected when cleaved with factor Xa, as shown in lanes 3 and 4 of Figure 4B. It is generally known that a misfolded

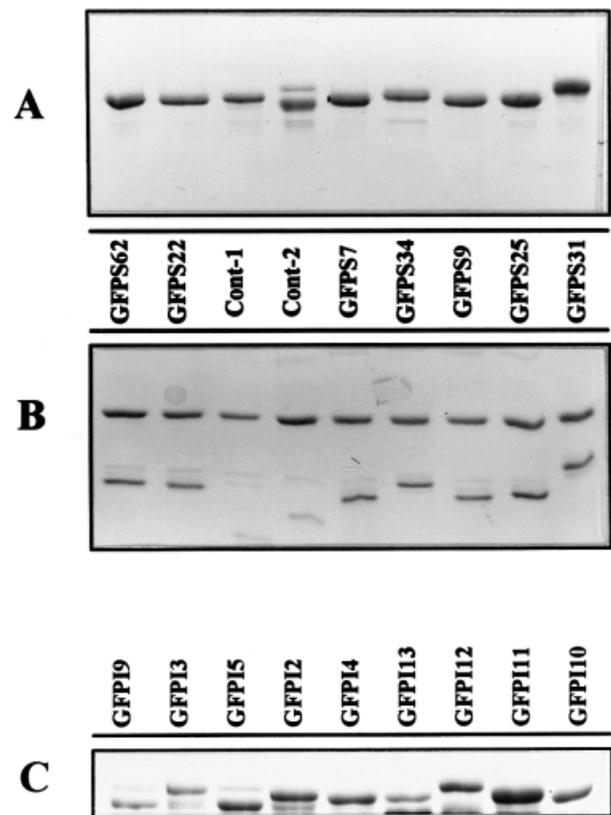


Fig. 4. SDS-PAGE analysis and factor Xa cleavage of the functionally salvaged GFP variants. (A) Aliquots (2 μ g) of the purified GFP variants fused with MBP, which were screened from the library generated from GFP ∇ 176(+2) by the PCR-coupled method, were analyzed on a 10% polyacrylamide gel under denaturing conditions. (B) Purified variants were treated with Factor Xa and analyzed by SDS-PAGE. In two lanes (3 and 4), a protein band corresponding to GFP variant was not observed. (C) Analysis of the salvaged GFP variants fused with MBP from the library constructed from GFP ∇ 172–3/176(+2) by the sequence-directed method. The purified fusion proteins (3 μ g) were analyzed on a 10% polyacrylamide gel under denaturing conditions.

or partly unfolded protein in the fusion state is structurally unstable and found to be tightly associated with bacterial chaperonin or protease. The result described here was the case that the two salvaged variants were partly misfolded and thus susceptible to protease, as reported by Keresztessy *et al.* (Keresztessy *et al.*, 1996). A similar result was observed in the control experiment using a general protease (Protease K). It is interesting that most of selected variants, when expressed in a single protein without MBP, emitted the fluorescence more persistently and retained their integrity than in the fusion state, and this might be attributed to a change in the fusion ability to proper folding in the salvaged variants with fusion partner MBP.

The functionally salvaged GFPs from the GFP ∇ 172–3/176(+2) library were also expressed as fusion proteins with MBP and purified according to the procedure described above. Most of the fusion proteins also exhibited the fluorescence and were expressed in a soluble fraction as shown in Figure 4C. However, the GFP ∇ 16 lost its fluorescence when fused with MBP. As a further observation, the fusion proteins were cleaved with factor Xa, yielding the same two domains as those from the GFP ∇ 176(+2) library (data not shown).

For the functional characterization of the salvaged GFPs, the fluorescence properties of the selected variants were

Table I. Spectral characteristics of the functionally salvaged GFP variants

Variant ^a	λ_{max} (excitation) (nm)	λ_{max} (emission) (nm)	Fluorescence intensity (%)
GFPuv(wt)	397	508	100
GFPS7	396	508	15.8
GFPS9	395	509	8.9
GFPS22	396	509	3.2
GFPS25	398	509	2.0
GFPS31	397	509	2.9
GFPS34 (NE)	—	—	—
GFPS62	395	510	13.6
GFPI2	397	509	0.6
GFPI3	395	506	5.3
GFPI4	398	506	0.9
GFPI5	396	509	4.9
GFPI9	397	508	15.3
GFP ∇ 10	399	509	1.4
GFP ∇ 11	396	507	12.5
GFP ∇ 12	397	508	1.7
GFP ∇ 13	398	507	16.2
GFP ∇ 16 (NE)	—	—	—

^aNE: not expressed when the mutants were fused with MBP under experimental conditions.

investigated (Table I). The wild-type GFPuv protein was first analyzed and excitation and emission maxima were observed at 397 and 508 nm, respectively. The functionally salvaged GFPs exhibited similar fluorescence properties, showing an excitation maximum at 395–399 nm and an emission maximum at 506–510 nm. The spectral properties were similar to each other, but the selected variants revealed a broad distribution of mean fluorescence intensity.

Multi-point salvage for more diverse protein lineages

In the natural evolutionary process of proteins, mutations occur randomly over the entire gene and the functional salvage process at multiple sites is expected to create more diverse protein lineages. In the multi-point salvage, the variants are supposed to recover their defective trait by complement at multiple sites. As a preliminary experiment to test the multi-point salvage, functional salvage at the dual-point was conducted. We designed a defective template in which an additional region (129D–138G) of the GFP ∇ 176(+2) was deleted and this template was designated GFP ∇ 129–138/176(+2). The deleted region (129D–138G) is part of a super-loop consisting of the residues from 129D to 142E (Ormo *et al.*, 1996; Yang *et al.*, 1996) and also functionally indispensable (Li *et al.*, 1997). For the multiple site salvage, the PCR-coupled process might be more effective in generating the functionally salvaged GFPs, because it enables the incorporation of segments into multiple sites simultaneously. Thus, we obtained a library from GFP ∇ 129–138/176(+2) by using the PCR-coupled process and picked clones randomly for analysis of their genetic diversity. As shown in Figure 5A, more diversity in genotype of the GFP variants was observed compared with that produced by a single point salvage. To confirm the diversity at the protein level, the genes were expressed in *E. coli* and expressed proteins were analyzed by SDS-PAGE (Figure 5B). The GFP variants bearing the incorporated segments were predominantly detected and their size was found to be diverse, as anticipated from the genetic diversity.

Discussion

We have attempted sequence-directed and PCR-coupled recombination approaches to create protein lineages with new

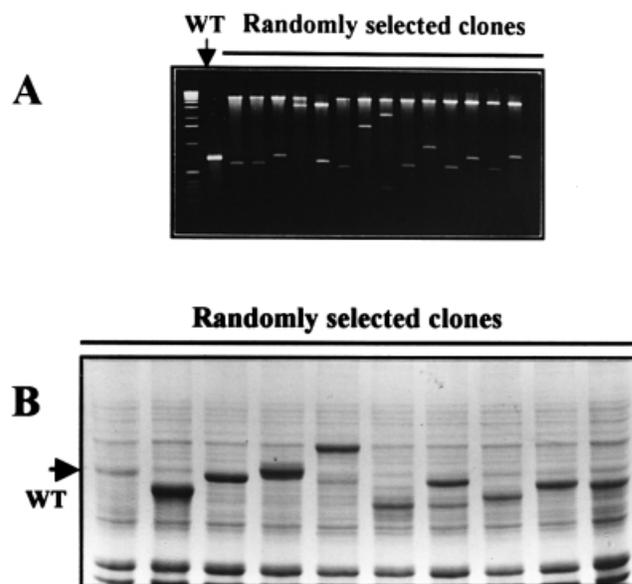


Fig. 5. Size variation and distribution of the library constructed from GFP ∇ 129–138/176(+2) by the PCR-coupled method at the dual-point. (A) Clones were randomly picked and their plasmids were analyzed on agarose gel. (B) SDS-PAGE analysis of the randomly picked clones. As a control, the wild-type gene or protein is shown in the left lane of the gel. The GFP variants from selected clones were expressed as major bands.

sequence spaces by employing a well-studied GFP as a model protein. The present study was primarily focused on whether the functional salvage process is able to generate the diverse protein lineages with new sequence spaces from the parent protein. As a conclusion, although the salvaged variants remain to be studied structurally in more detail, we suggest that the functional salvage process can be an effective tool for the creation of new protein lineages. The feasibility of the present method is partly supported by a recent report that the generation of rigid enzyme with a similar function to the parent enzyme is possible by incorporation of oligonucleotides from randomly fragmented DNA into an enzyme scaffold (Jappelli and Brenner, 1999).

It has been shown that GFPuv, with a nearly perfect shell around a chromophore by an 11-stranded β -barrel, is generated through a complex process, leading to a highly compact and rigid structure for fluorescence emission (Ormo *et al.*, 1996; Yang *et al.*, 1996). Thus, it has been considered that GFP is marginally tolerable against artificial transposition or insertion, displaying a vulnerable nature to structural change (Li *et al.*, 1997). It was also reported that deletions in either the internal or terminal region of GFP result in perturbation of the structural integrity and loss of fluorescence (Li *et al.*, 1997). Unexpectedly, nonetheless, a large number of functionally salvaged variants were generated by the functional salvage process, which indicates that this process provides an efficient route to create a pool of new protein lineages with different sequence spaces. In line with this, generation of the functionally salvaged GFPs supports the contention that the structural integrity of GFP can be further extended to the sequence space where further diversification is acceptable. Consistent with this view, experimental results have recently shown the expansion of sequence spaces in terms of structural and functional trait either in recombinant (Abedi *et al.*, 1998; Baird *et al.*, 1999) or natural GFP homologues (Matz *et al.*, 1999; Gross *et al.*,

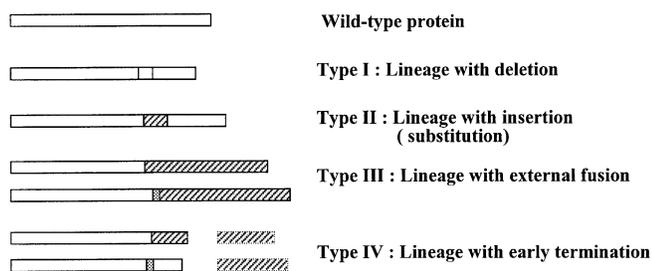


Fig. 6. Schematic representation of the lineage types that can be produced by the functional salvage process. The white boxes represent the conserved sequence of the wild-type protein. The hatched boxes indicate the rescuing sequences that are either expressed (solid lines) or not expressed (dotted lines) with functional domains. The gray boxes in types III and IV stand for the homologous sequences rescuing the defective templates.

2000; Lukyanov *et al.*, 2000; Wall *et al.*, 2000; Wiedenmann *et al.*, 2000).

The functionally salvaged GFP variants exhibited similar fluorescence properties in terms of maximum excitation and emission wavelengths when compared with the wild-type GFPuv, but their fluorescence intensities varied considerably. This result implies that the salvaged GFPs have different conformations or stabilities from the wild-type GFPuv, probably owing to the structural perturbation or subtle disorganization of the innate fluorophore by incorporation of segments. From a more detailed comparison, there appeared to be an apparent correlation between the portion of hydrophobic residues of the incorporated segments and fluorescence intensity. As observed in Figure 3 and Table I, the GFPs salvaged with more hydrophobic residues produced less fluorescence than those harboring a short peptide or hydrophilic residues, resulting in an ~25–100-fold lower intensity compared with the parent GFPuv. The salvaged variants, GFPS25, GFPS31, GFPI2 and GFP ∇ 10, contained a relatively high portion of hydrophobic amino acid residues in the segments compared with other variants and their fluorescence intensity was much lower than those of other variants. On the other hand, no apparent correlation was observed between the size of the incorporated segments and mean fluorescence value. The exact reason for these results remains to be elucidated, but it is plausible to suggest that the salvage point of the defective templates is tolerable to incorporation of diverse segments in a sequence-specific rather than a size-specific manner. Resolution of the structure of the functionally salvaged GFPs would demonstrate the effect of the sequence space of the rescuing segments on the properties of the resulting variants. Prior to structural resolution, it is also feasible that the salvaged variants might be further processed to improve their properties including stability and fluorescence intensity through *in vitro* evolution using random mutagenesis or DNA shuffling (Zhang *et al.*, 1997).

Considering the principle of the functional salvage process, four types of lineages are expected to be produced when the constructed templates were used, as depicted schematically in Figure 6. From the libraries generated from the four defective templates, 375 clones recovering the fluorescence through the salvage process were screened. Of these, 17 clones were randomly picked and typically characterized. As a preliminary result, random insertion of the segments was found to be a predominant event in the salvaged GFPs, mainly producing the type II lineage. In this work, screening of the salvaged GFPs relied on whether the green fluorescence is emitted at a

single wavelength excitation (365 nm) and this might partly contribute to a narrow range of the salvaged variants (Matz *et al.*, 1999; Lukyanov *et al.*, 2000; Wiedenmann *et al.*, 2000). In fact, we also detected a significant number of type III and IV lineages based on the fluorescence emission and protein expression. Of the primarily screened clones, about 30% were analyzed to be salvaged by fusion at the external region (type III) and by early termination (type IV). Type III variants are likely to be mainly generated through a fusion of a large fragment, resulting in a fusion protein-like feature by structural rearrangement. On the other hand, type IV might be created by either an early termination due to the appearance of a stop codon or proteolysis *in vivo*. Production of type I lineage is expected when the defective template is constructed by deletion of either multiple sites or a larger fragment or domain, but this type could not be produced in the single point salvage process attempted mainly in this work.

It is generally accepted that many proteins have divergently evolved from a common ancestor and that gene rearrangement including insertion, deletion, duplication and inversion has played a central role in the functional evolution of proteins (Heidmann and Rougeon, 1983; Yogev *et al.*, 1995; Wang *et al.*, 1997; Todd *et al.*, 1999). In a protein family, the natural evolution process has provided a huge diversity with respect to structures and mechanistic features (Holm and Sander, 1997), although several protein families were reported to display a relatively narrow scope (or fitness landscape). From the analysis of the libraries constructed by the functional salvage process, we believe that this process can generate a variety of protein lineages having diverse sequence spaces under the condition where the fate of the protein is not determined. Thus, the functional salvage process seems to be similar to the natural evolution process in that the diversity of proteins is generated from the pre-existing gene through gene rearrangement, such as insertion, deletion and inversion, resulting in new protein lineages with different dimensions from the parent protein (Pujadas *et al.*, 1996; Bergdoll *et al.*, 1998; Bogarad and Deem, 1999; Riechmann and Winter, 2000). In other words, given that a strategy is more rationally designed, diverse protein lineages generated by the functional salvage process may find corresponding counterparts in the naturally occurring protein pools. In the multi-point salvage, the variants are considered to recover their trait by complementing the defective gene at multiple sites. Consequently, more diverse protein lineages with lower homology compared with the parent protein are expected as the number of salvage points increases. As a preliminary experiment for multi-point salvage, our attempt at the dual-point provides a clue that versatile protein lineages with new sequence spaces could be generated by the multi-point salvage process, even though further characterization of the salvaged pools remains to be conducted. A recent report demonstrated that a new protein fold can be generated by shuffling of segments derived from the randomly fragmented genomic DNA from *E. coli* (Riechmann and Winter, 2000), and this also supports our suggestion.

The functional salvage process relies on the screening of the protein variants with appropriately reorganized structures complementing a defective trait. Either non-essential or essential regions of a protein should be the target sites for the functional salvage screen and these sites can be deduced from various methods such as multiple alignments of the related proteins or comparison of their 3-D structures, which would

lead to the development of a more rational strategy for the functional salvage screen. In this context, we believe that the functional salvage screen can be a valuable tool for the generation of protein lineages with new sequence spaces and the resulting variants can be further subjected to directed evolution to acquire the desired properties. The present approach may also be useful in studying the involvement of a specific region(s) or domain(s) in the structure and function of proteins.

References

- Abecassis,V., Pompon,D. and Truan,G. (2000) *Nucleic Acids Res.*, **28**, E88.
- Abedi,M.R., Caponigro,G. and Kamb,A. (1998) *Nucleic Acids Res.*, **26**, 623–630.
- Altamirano,M.M., Blackburn,J.M., Aguayo,C. and Fersht,A.R. (2000) *Nature*, **403**, 617–622.
- 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.
- 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.
- Bogard,L.D. and Deem,M.W. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 2591–2595.
- Christians,F.C., Scapozza,L., Crameri,A., Folkers,G. and Stemmer,W.P. (1999) *Nature Biotechnol.*, **17**, 259–264.
- Crameri,A., Whitehorn,E.A., Tate,E. and Stemmer,W.P. (1996) *Nature Biotechnol.*, **14**, 315–319.
- Doi,N. and Yanagawa,H. (1999) *FEBS Lett.*, **453**, 305–307.
- Frank,E.G., Gonzalez,M., Ennis,D.G., Levine,A.S. and Woodgate,R. (1996) *J. Bacteriol.*, **178**, 3550–3556.
- Greenfeder,S.A., Varnell,T., Powers,G., Lombard-Gillooly,K., Shuster,D., McIntyre,K.W., Ryan,D.E., Levin,W., Madison,V. and Ju,G. (1995) *J. Biol. Chem.*, **270**, 22460–22466.
- Gross,L.A., Baird,G.S., Hoffman,R.C., Baldrige,K.K. and Tsien,R.Y. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 11990–11995.
- Heidmann,O. and Rougeon,F. (1983) *Cell*, **34**, 767–777.
- Holm,L. and Sander,C. (1997) *Proteins*, **28**, 72–82.
- 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.
- Jappelli,R. and Brenner,S. (1999) *Biochem. Biophys. Res. Commun.*, **266**, 243–247.
- Joo,H., Lin,Z. and Arnold,F.H. (1999) *Nature*, **399**, 670–673.
- Kapust,R.B. and Waugh,D.S. (1999) *Protein Sci.*, **8**, 1668–1674.
- Keresztesy,Z., Hughes,J., Kiss,L. and Hughes,M.A. (1996) *Biochem. J.*, **314**, 41–47.
- Kikuchi,M., Ohnishi,K. and Harayama,S. (1999) *Gene*, **236**, 159–167.
- Kim,G.J., Cheon,Y.H. and Kim,H.S. (2000) *Biotechnol. Bioeng.*, **68**, 211–217.
- Kumar,L.V. and Rao,C.M. (2000) *J. Biol. Chem.*, **275**, 22009–22013.
- Li,X., Zhang,G., Ngo,N., Zhao,X., Kain,S.R. and Huang,C.C. (1997) *J. Biol. Chem.*, **272**, 28545–28549.
- Lukyanov,K.A. *et al.* (2000) *J. Biol. Chem.*, **275**, 25879–25882.
- Matsumura,I. and Ellington,A.D. (2001) *J. Mol. Biol.*, **305**, 331–339.
- Matsuura,T., Miyai,K., Trakulnaleamsai,S., Yomo,T., Shima,Y., Miki,S., Yamamoto,K. and Urabe,I. (1999) *Nature Biotechnol.*, **17**, 58–61.
- 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.
- Nixon,A.E., Ostermeier,M. and Benkovic,S.J. (1998) *Trends Biotechnol.*, **16**, 258–264.
- 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) *Nature Biotechnol.*, **17**, 1205–1209.
- Pujadas,G., Ramirez,F.M., Valero,R. and Palau,J. (1996) *Proteins*, **25**, 456–472.
- Riechmann,L. and Winter,G. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 10068–10073.
- Stemmer,W.P. (1994) *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.
- Wall,M.A., Socolich,M. and Ranganathan,R. (2000) *Nature Struct. Biol.*, **7**, 1133–1138.
- Wang,Y., Goligorsky,M.S., Lin,M., Wilcox,J.N. and Marsden,P.A. (1997) *J. Biol. Chem.*, **272**, 11392–11401.
- Wiedenmann,J., Elke,C., Spindler,K.D. and Funke,W. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 14091–14096.
- Xirodimas,D.P. and Lane,D.P. (1999) *J. Biol. Chem.*, **274**, 28042–28049.
- Yang,F., Moss,L.G. and Phillips,G.N.,Jr. (1996) *Nature Biotechnol.*, **14**, 1246–1251.
- Yogev,D., Watson-Mckown,R., Rosengarten,R., Im,J. and Wise,K.S. (1995) *J. Bacteriol.*, **177**, 5636–5643.
- Zhang,J.H., Dawes,G. and Stemmer,W.P. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 4504–4509.

Received February 14, 2001; revised June 8, 2001; accepted June 18, 2001