Proc. KSAM International Symposium
June 20-22, 2001, Chonan, Korea

Generation of Protein Lineages with New Sequence Spaces by Functional Salvage Screen

Geum-Joong Kim[1], Young-Hoon Cheon, Min-Soon Park,
Hee-Sung Park, and Hak-Sung Kim
Department of Biological Sciences, Korea Advanced Institute of Science and Technology
373-1, Kusung-dong, Yusung-gu, Taejon, 305-701, Korea
[1]Department of Molecular Science and Technology, Ajou University, San5, Woncheon-dong,
Paldal-gu, Suwon 442-749, Korea

Summary

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 firstly 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 *E. 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 property. 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.

Introduction

Advances in protein engineering have accelerated the understanding of a number of intrinsic questions of protein, such as the structure and function relationships, folding process 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 technique based on sequential mutation and random recombination has been proven to be a very effective tool to generate the proteins with greater potential (Christians *et al.*,1999; Xirodimas and Lane, 1999; Joo *et al.*,1999) or novel function (Matsumura and Ellington, 2001; Zhang *et al.*, 1997). Directed evolution technique has also been used to address some issues regarding the regulation of protein molecules by endoge-

nous 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 (Kim *et al.*, 2000; Crameri *et al.*, 1996).

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 with 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-terminal 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 (Kumar and Rao, 2000; Hopfner *et al.*, 1998), domain or module grafting (Nixon *et al.*, 1998; Aphasizheva *et al.*, 1998; Greenfeder *et al.*, 1995), DNA homology-independent recombination (Ostermeier *et al.*, 1999), and scaffold design based on combinatorial methods (Altamirano *et al.*, 2000) have been carried out for generation of new protein lineages.

Herein 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 *E. coli* chromosome could rescue the protein function. For 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 a considerable stability were selected, and analyzed with respect to sequence space and functional properties.

Experimental

Construction of the defective GFP templates

The defective GFP templates were constructed as schematically depicted in Figure1. 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-GCGAATTCAGTAAAGGAGAAGAACTTTTCACTGGA-3) and F3 (5-GCGGATCCATCTTCAATGTTGTGGCG-3) flanked by *Eco*RI and *Bam*HI 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-GAGGATCCAACTAGCAGACCATTATCAACAAA-3)/F4(5-AGTAAGCTTATTTGTAGAGCTCATCCATGCCATG-3) and F2(-2)(5-GAGGATCCACTAGCAGACCATTATCAACAAA-3)/F4(5-AGTAAGCTT ATTTGTAGAGCTCATCCATGC

CATG-3). Underlined sequences indicate the *Bam*HI and *Hind* III sites, respectively. Each of the amplified fragments was inserted into the *Bam*HI and *Hind*III 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 the similar procedure as 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-GTGGATCCAATAC-CTTTTAACTC-3) and P2 (5-ATGGATCCCACAAACTCGAGTC-3)/F4. The resulting template was designated GFP∇129-138/176(+2).

Library construction for the functional salvage screen

Four constructs containing each of defective templates, pGFP∇176(+1), pGFP∇176(+2), pGFP∇172-3/176(+1) and pGFP∇172-3/176(+2), were linearized by digestion with *Bam*HI, and then eluted from agarose gel (0.8 %). For oligonucleotides pools to be incorporated into the defective template genes, chromosomal DNA isolated from the *E. coli* MG1655 was digested with *Sau*3AI, and the fragments ranging from 25-500 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 *Bam*HI and then further digested with DNase I. The DNA fragments ranging from 50-150 bp were excised and eluted from agarose gel (2.5%), and then reassembled with the *Sau*3AI-digested chromosomal DNA (25-500 bp) by PCR (94 °C, 1 min; 45+0.2°C/cycle, 1 min; 72°C, 40 sec; total 40 cycles). The defective template (10-20 ng/ml) was mixed with chromosomal DNA at the 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 sec; 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.

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 plate 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. *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 hr cultivation, chlora With the primarily isolated clones, we further screened the salvaged mphenicol (100 mg/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.

References

1. Altamirano, M.M., Blackburn, J.M., Aguayo, C. and Fersht, A.R. (2000) *Nature,* **403,** 617-622.
2. 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.
3. Christians, F.C., Scapozza, L., Crameri, A., Folkers, G. and Stemmer, W.P. (1999) *Nat. Biotechnol.,* **17,** 259-264.
4. Crameri, A., Whitehorn, E.A., Tate, E. and Stemmer, W.P. (1996) *Nat. Biotechnol.,* **14,** 315-319.
5. Frank, E.G., Gonzalez, M., Ennis, D.G., Levine, A.S. and Woodgate, R. (1996) *J. Bacteriol.,* **178,** 3550-3556.
6. Greenfeder, S.A., Varnell, T., Powers, G., Lombard-Gillooly, K., Shuster, D., Mclntyre, K.W., Ryan, D.E., Levin, W., Madison, V. and Ju, G. (1995) *J. Biol. Chem.,* **270,** 22460-22466.
7. 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.
8. Jappelli, R. and Brenner, S. (1999) *Biochem. Biophys. Res. Commun.,* **266,** 243-247.
9. Joo, H., Lin, Z. and Arnold, F.H. (1999) *Nature,* **399,** 670-673.
10. Kapust, R.B. and Waugh, D.S. (1999) *Protein Sci.* 8, 1668-1674.
11. Keresztessy, Z., Hughes, J., Kiss, L. and Hughes, M. A. (1996) *Biochem. J.,* 314, 41-47.
12. Kikuchi, M., Ohnishi, K. and Harayama, S. (1999) *Gene,* **236,** 159-167.
13. Kim, G.J., Cheon, Y.H. and Kim, H.S. (2000) *Biotechnol. Bioeng.,* 68, 211-217.
14. Kumar, L.V. and Rao, C.M. (2000) *J. Biol. Chem.,* **275,** 22009-22013.
15. Matsumura, I. and Ellington, A.D. (2001) *J. Mol. Biol.,* **305,** 331-339.
16. Matsuura, T., Miyai, K., Trakulnaleamsai, S., Yomo, T., Shima, Y., Miki, S., Yamamoto, K. and Urabe, I. (1999) *Nat. Biotechnol.,* **17,** 58-61.
17. Nixon, A.E., Ostermeier, M. and Benkovic, S.J. (1998) *Trends Biotechnol.,* **16,** 258-264.
18. Ostermeier, M., Shim, J.H. and Benkovic, S.J. (1999) *Nat. Biotechnol.,* **17,** 1205-1209.
19. Xirodimas, D.P. and Lane, D.P. (1999) *J. Biol. Chem.,* **274,** 28042-28049.
20. Zhang, J.H., Dawes, G. and Stemmer, W.P. (1997) *Proc. Natl. Acad. Sci. USA,* **94,** 4504-4509.