

Full Length Research Paper

Enhancing the productivity of soluble green fluorescent protein through methionine-residue specific consensus approach

Nagasundarapandian Soundrarajan¹, Selvakumar Edwardraja¹, Sun-Gu Lee¹, Hyungdon Yun²
and Niraikulam Ayyadurai^{2*}

¹Department of Chemical Engineering, Pusan National University, Busan, South Korea.

²School of Biotechnology, Yeungnam University, Gyeongsan, South Korea.

Accepted 3 January, 2012

Protein sequences might have been evolved against different environmental pressures, which results in non-optimum properties in their stability, activity and folding efficiency. Directed evolution and consensus-based engineering of proteins are the protein engineering principles for the re-evolution of such natural proteins exhibiting non-optimal properties. Here, we propose an approach to improve the physical properties of target protein by engineering protein with new methionine residues. Our aim of this study was to investigate whether the physical property of protein can be improved by altering the negative effect caused by the introduction of additional methionine residue in the protein. First, we attempted to perform combinatorial mutagenesis of methionine residues of green fluorescent protein (GFP) using the consensus amino acids of conserved sequences. Each methionine residue in the internal sequence of GFP was combinatorially mutated by methionine or amino acid showing the highest frequencies in conserved sequences (I for M78, F for M88, Y for M153, V for M218, K for M233), and the mutants showing fluorescence were selected. Among the variants, the mutant of M218K showed an enhanced soluble expression in *Escherichia coli*. Our results indicate that it is possible to engineer protein by mutating methionine residues, specifically. We expect that the proposed approach can be exploited to enhance the expression of target protein in soluble form with avoiding the intensive labor of random mutagenesis and screening.

Key words: Protein sequence, GFP, directed evolution, consensus engineering, mutagenesis.

INTRODUCTION

One important goal of protein engineering is to improve the biotechnological or medical efficacy of proteins by adapting their biophysical or functional properties. However, improving these properties is not straight forward and this could be achieved through enhancing protein expression and folding properties, as well as protein solubility and stability (Forrer et al., 2004). Currently, a number of methods are available to improve such protein properties; among these, three approaches play vital role: (i) approaches based on structure and computation, (ii) those exploiting directed evolution and

et al., 2000; Arnold et al., 2001; Lehmann et al. (iii) those based on consensus design (Pluckthun et al., al., 2001; van den Burg 2002). Among these methods, consensus approach is recognized as a promising method (Kajander et al., 2006; Travaglini-Allocatelli et al., 2009). This method is time-consuming and less labor-intensive than traditional mutagenesis methods. It helps in redesigning a target protein using conserved amino acids of homologous sequences based on the assumption that the crucial amino acids for protein structure and stability may be retained during evolution (Van den Burg and Eijsink, 2002; McDonald, 2006).

Indeed the approach has been usefully employed to improve the physical properties of proteins such as thermal stability and folding efficiency. A major drawback of the consensus approach is that the success rate is not

*Corresponding author. E-mail: ayyadurai@ynu.ac.kr. Tel: 82-53-810-3918.

very high as it induces a high mutation rate of target sequence, which often disrupts the protein structure and function (Kajander et al., 2006).

In particular, when the activity of the protein is susceptible to mutations, the method cannot be easily utilized and it is often difficult for the identification of critical amino acids required for stabilization of protein. To overcome this problem, various derivatives of this method such as guided-consensus approach and combinatorial consensus approach were developed and tested (Lehmann et al., 2002). Here, we propose an alternative consensus protein engineering method to overcome such problems. Our basic idea is to narrow down the range of protein sequences for consensus approach from whole sequence level to specific amino acid level. Specifically, we suggest that methionine (Met) residues on a protein sequence would be a good target to apply the consensus concept due to the following reasons. Firstly, it is one of the “gainer” amino acids (Cys, His, Met, Ser and Phe) whose frequency in protein sequences are being increased during evolution (Jordan et al., 2005). According to the evolution theory, proteins are evolved by random mutagenesis. Therefore, the probability that the gainers include the amino acids residues which have negative influence on protein properties may be relatively high. Thus, we may have some chances to improve the protein properties by restoring the gainer amino acids that affect negatively. Secondly, Met is a very unique amino acid that contributes to protein structures with both hydrophobic interactions and hydrogen bonding, and becomes hydrophilic if it is oxidized (Wolschner et al., 2009). This indicates that the addition or removal of Met can alter the protein properties with high chances. Thirdly, being a weak hydrophobic residue, it is usually located in the solvent inaccessible regions of the protein, with only 15% solvent exposed (Budisa, 2005). This further allows us to look for an evolutionary compatible replacement approach to enhance the physicochemical property of protein especially its solubility. Therefore, in this study, we analysed whether we can improve protein properties through consensus approach at Met residue level by using green fluorescent protein (GFP) as a model protein. The use of traditional consensus concept is difficult to apply in GFP due to its high sensitivity against mutations (Nagasundarapandian et al., 2010). We analyzed GFP homologous proteins and selected the conserved amino acids for the Met residues of GFP. Then, we carried out Met residue specific combinatorial mutations using the conserved amino acids and Met. Among the generated mutants, we isolated a mutant GFP showing improved expression of protein in soluble form in *E. coli*. This residue specific combinatorial approach also allowed us to identify the important Met residues responsible for the proper folding of GFP in efficient manner.

MATERIALS AND METHODS

PCR reagents, T4 DNA ligase and restriction endonucleases were

purchased from Promega (Madison, WI, USA). The isopropyl-D-thiogalactopyranoside (IPTG) was purchased from sigma chemicals (St. Louis, MO, USA). The host bacterium *E. coli* strain XL1-blue (Stratagene, CA, USA) was used for plasmid DNA preparation in this study. *E. coli* cells with plasmids were grown aerobically in Luria-Bertani (LB) broth (Difco Laboratories, Detroit, Michigan, USA) or on LB agar plate, supplemented with appropriate antibiotics for the selection of transformants. The nickel-nitrilotriacetic acid (Ni-NTA) affinity column was purchased from Qiagen (Valencia, CA, USA).

Construction of plasmids and strains

The DNA manipulations were performed according to the procedures described by Sambrook and Russel (1989). The PCR reaction was conducted 10 pmole of primer, template DNA, 1x Taq DNA polymerase buffer, 2.5 units of Taq DNA polymerase, 2.5 mM deoxyribonucleotide triphosphates, and 1.5 mM MgCl₂ was used in this study. Amplification was performed in a DNA thermal cycler (Master Gradient thermal cycler, Eppendorf, Hamburg, Germany) programmed for an initial denaturation (94°C for 1 min) followed by 30 cycles of 1 min at 94°C, 1 min at 60°C and 0.5 min at 72°C with an extension at 72°C for 10 min. The construction of pET28a-GFP was previously described [15, 16]. pQE80-GFP was constructed by cloning the gene for GFP, amplified by two primers (5'-AAGGAT-CCCGTAAAGGAGAAGAACTT-3' and 5'-CGTAAGCTTTTATTAAT-GGTGATGATG GTGGTGGTATAGTTCATCCAT-3') using pET28a-GFP as template, into the pQE80-L using *Bam*H1 and *Hind*III restriction enzymes. The constructed pET plasmids and pQE plasmids were transformed into *E. coli* B834 for the expression and screening assay. All constructs were sequenced and confirmed for their target protein sequences.

Combinatorial mutagenesis and screening of library

Similarity searches for full length sequences and conserved domains were performed using a combination of standard bioinformatics programs and customized Python scripts.

Each assembled transcript was searched against UniProtKB database (Release 15.5, TrEMBL and Swiss-Prot at <http://www.uniprot.org>) resources using the BLASTX algorithm. Redundant sequences were then removed, and an alignment between GFPmut3.1 and the 31 remaining sequences was then generated using CLUSTAL W (Thompson et al. 1994). Further, consensus amino acid was identified for the each individual position (Table 1).

Based on the results, the following primers were synthesized for each individual gene mentioned in the Table 1.

The multi-site mutagenesis experiments were carried out with the PCR procedure as described earlier using Stratagene's QuickChange kit according to manufacturer's protocol (La Jolla, CA, USA). Out of 50 µl, 4 µl of purified PCR product was directly transformed into 200 µl TOP10 competent cells. The transformant was incubated in 5 ml LB medium containing 1 mM IPTG and 100 µg/ml ampicillin overnight. Finally, only fluorescent colonies were picked, propagated in 5 ml of LB liquid medium. Cells were collected and submitted for sequence analysis in CosmoTech, Korea (Seoul, South Korea).

Expression of GFP mutants, SDS PAGE analysis and purification

E. coli B834 (DE3) with the pQE80 carrying all the selected mutant gene sequences were grown to optical density at 600 (OD₆₀₀) of

Table 1. Position, sequence profile analysis and frequencies of amino acids of the five Met residues in GFP.

Met position	Primers used in the study	Frequency of amino acid (single letter code for the specific amino acid)																
		A	C	E	F	H	I	K	L	M	N	P	R	T	V	W	Y	gap
78	5'-GAGATACCCAGATCATATCAAACAGCATGAC-3'					1	19*	1		8				1			4	
88	5'-CAAGAGTGCCTTCCCCGAAGGTTACGTACAG-3'		3		17*				5	4	1		2					
153	5'-CACAAATGTATACATCTACGCAGACAAAACAAAAG-3'	1			3	2	5		1	4			6	1			11*	
218	5'-GAGACCACGTTGTCCTTCTTGAGTTTGTAAACAGC-3'				1		7			5				20*	1			
233	5'-GCTGGGATTACACATGGCAAAGATGAACACTATAC-3'			2	1			8*		4		1	2	3				10

*Highly conserved residue selected for mutation analysis.

0.6 at 37°C in 5mL LB medium containing 100 µg/mL ampicillin, induced with different concentration IPTG to optimize the protein production.

After 5 h, the cells were harvested by centrifugation at 4000 g at 4°C.

Total cell protein fractions were analyzed by SDS-PAGE (12% acrylamide gel), as described previously (Nagasundarapandian et al. 2010).

The selected GFP mutants were purified by affinity chromatography by using Ni-NTA resin according to the manufacturer's protocols. The purified GFP proteins were stored at 4°C until further use.

Protein quantification and fluorescence analysis

Absorption spectrum and quantification of proteins were measured as described in other studies (Ayyadurai et al., 2011).

Fluorescence spectra (2 µM protein samples were excited at 470 nm by using excitation/emission slits of 5.0 nm) were recorded on a JASCO FP-777 spectrofluorimeter equipped with digital software.

Measurement of refolding kinetics

Purified GFP and its variants (30 µM) were denatured under harsh condition in 1X PBS containing 8M urea and 5 mM DTT unfolded at 95°C for 5 min. Urea denatured samples were renatured at room temperature by 100 fold dilution into 1X PBS containing 5 mM DTT without urea. The concentrations of denatured proteins were adjusted to 0.3 µM and recovered fluorescence was measured using

Perkin-Elmer spectrometer (LS55) spectrophotometer (490 nm excitation, 511 nm emission, 5 nm excitation/emission slit) for 30 min with an interval of 3 s. The recovered fluorescence was normalized by dividing final fluorescence after 24 h value. The normalized values were fitted with Sigma Plot (Systat Software Inc., CA) using equations as described by previous report (Steiner et al., 2008).

RESULTS

Consensus analysis of Met residues on GFP

Among the various GFP mutants, we employed GFPmut3.1b as a target that contains 5 Met residues (M78, M88, M153, M218 and M233) excluding the N-terminal methionine. This variant can fold rapidly in *E. coli* compared to wild type we investigated whether the soluble expression level of the GFPmut3.1b could be further improved by our proposed approach.

To achieve the conserved amino acids of the Met residues, we analyzed the homologous fluorescent proteins of GFP obtained from the protein database (Figure 1) and multiple sequence alignment was carried out.

To create a valid consensus analysis, totally 28 homologous translated sequences were extracted from the Genbank with high homology. We used only the monomeric derivative of a fluorescent

protein derived from the different species.

We analysed the consensus residues at the methionine position as described in Table 1. This analysis showed the position, expossibility to solvent and frequencies of amino acids of the five Met residues. Among the five residues, 3 residues (M78, M88 and M218) reside on the secondary structure of protein and buried inside. The other 2 residues (M153, and M233) were on the loop region and exposed to solvent. The consensus amino acids for the Met residues were determined as the amino acids that showed the highest frequency in the particular Met position.

Combinatorial mutagenesis of the Met residues in GFPmut3.1

Further, to perform the combinatorial mutagenesis of the selected amino acids at each Met residues, we utilized Stratagene mutagenesis kit as described in Material and Methods Section. Each Met residue was mutated through combinatorial approach either with two amino acids, that is, Met or the conserved amino acids determined above by adjusting mutation efficiency at each site by varying primer concentration as described by manufacture's, which can generate 32 variants theoretically. From the library in LB Agar plate, we

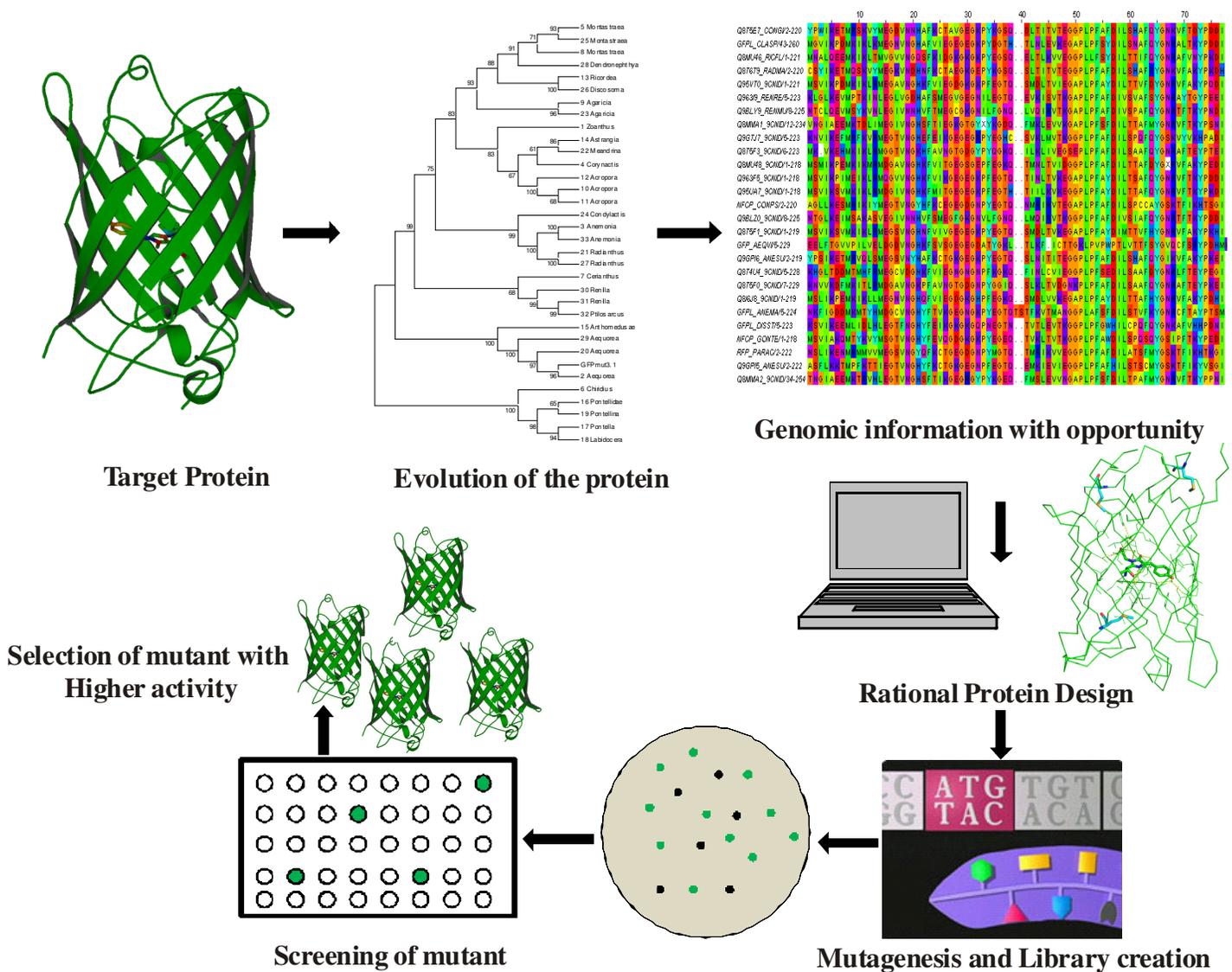


Figure 1. Schematic representation of the proposed approach to improve the solubility of protein through consensus engineering approach.

selected negative clones showing no fluorescence as well as positive clones showing green fluorescence, and analyzed the mutated positions.

The Table 2 shows the identified mutated positions and the approximate relative fluorescence intensities. Among the mutants, the GFP variants (M2 and M4) showed increased and moderate whole cell fluorescence, respectively. We confirmed that the fluorescence of other variants were almost inactivated or reduced by the mutations (data not shown).

Expression and characterization of the positive GFP variants

The mutant M2 expressed in *E. coli* and its soluble

productivities was compared with GFPmut3.1b by measuring their whole cell fluorescence (Figure 2A). Among the two variants, M2 and M4 showed comparable fluorescence with the GFPmut3.1b. These results indicate that the M78I (Table 2) and M233K mutations had positive influence on the expression level. The SDS-PAGE analysis allowed us to confirm that the M233K mutation led to the marginal enhancement of soluble expression level of GFPmut3.1b (Figure 2A).

On the other hand, the loss of fluorescence in other mutant suggests that the buried Met in hydrophobic core are sensitive against consensus mutation. Further, to analyze mutational sensitivity, we introduced combinatorial point mutations at the buried Met residues of GFPmut3.1b using amino acids of similar physico-chemical properties with Met, Leu, Ile, Val, Phe, and Ala.

Table 2. Identified mutated positions and relative fluorescence intensities.

Obtained mutant	Mutation	Amino acid position	RF	Fluorescence
GFP-M1	M78I	Buried	0.18	very weak
GFP-M7	M78I, M153Y, M218V, M233K	Buried, surface exposed, buried, surface exposed	0.11	very weak
GFP-M8	M78I, M218V, M233K	Buried, buried, surface exposed	0.10	very weak
GFP-M6	M78I,	Buried	0.18	very weak
GFP-M5	M78I, M88F	Buried, buried	0.12	very weak
GFP-M3	M153Y	Surface exposed	0.20	weak
GFP-M9	M78I, M88F	Buried, buried	0.20	weak
GFP-M4	M78I, M233K	Buried, surface exposed	0.56	moderate
GFP-M2	M233K	C-terminal surface exposed	1.30	strong

RF= Relative fluorescence intensity was measured with respect to the wild type GFP fluorescence. The resulting fluorescence intensity of GFP was normalized by dividing the fluorescence intensity in the sample by that of the mutant sample fluorescence. The experiment was measured in triplicates.

We found that all of the mutants showed negative effect of folding efficiency (data not shown). Surprisingly, M218A mutant showed more than fivefold reduction in the whole cell fluorescence when compare to the parent GFP (Data not shown). These results suggest that the buried Met are important to form the proper structure of GFP especially the M218 position, the most sensitive against mutation (Nagasundarapandian et al., 2010). Finally, in order to investigate the effect of M233K mutation resulted in increase soluble protein yield, the fluorescence level was compared with GFPmut3.1b. We analyzed their folding efficiency against urea and temperature. The purified proteins were denatured by boiling in 8M urea, and refolded upon dilution, and the recovery of the fluorescence was monitored fluorometrically. As shown in Figure 2 B, both the mutant GFP-M2 and GFPmut3.1b exhibited similar faster initial folding rates and as well as slower folding rates, respectively.

These results clearly indicate that the folding rates of the both protein were almost similar, hence the effect of M233K consensus mutation might slightly increase the soluble expression level.

DISCUSSION

Consensus engineering has been used to increase the stability of a number of different proteins, either by creating consensus proteins from scratch or by modifying existing proteins, so that their sequences closely match a consensus sequence. In the present study, we have shown that the expression level of GFP can be maintained through the methionine residue specific consensus approach. The work suggested here explains that consensus engineering does not always lead to more stable proteins, especially if applied in buried or more conserved amino acid. However, this approach could be used to maintain the protein stability. In general, to maintain the protein function during modification is an

important step and most of the time directed evolution will be in the choice. It is time consuming, cost effective and labor intensive. The consensus approach will help to modify the protein which is simple, and knowledge based approach. Nevertheless, our demonstration shows that consensus approach can be employed in residue-specific level instead of protein sequence level. Especially, when the protein is sensitive against high mutation rate, this approach may be considered one of the options. In our study, the mutation of Met residing at 233 positions to Lys will maintain the soluble expression level of GFP. As far as we know, the mutant position M233K in M2 is the first identified position for positive effect on soluble expression level of GFP. Since, methionine is an important amino acid that contributes to protein structures with specific hydrophobic interactions and hydrogen bonding. Usually, it is located in the core part of the proteins and only a few methionine residues are exposed to the surface of proteins (Jordan et al., 2005). So mutating the methionine residues is difficult and most of the time will led to loss of the protein function. It again prove that the proteins with both positive and negative effects are important to maintain the protein stability due to the non-optimized process in protein sequence evolution (Taverna and Goldstein, 2002; Hurst and Rocha, 2006). We also performed the comparison of refolding kinetics between M2 and parent in order to identify the factors that affect the soluble expression level of GFP, but their re-folding rates were almost same. This result indicates that the soluble expression level of M2 was maintained through protein folding nature. Although, we were disappointed that consensus engineering approach had a lower expression level for most of the mutants. It is interesting that the obtained mutant showed the different behavior of the two proteins during the fluorescence analysis. Our consensus engineering approach confirms the validity of this approach to maintain the protein function without losing their stability. We expect this to have significant biotechnological implications to modify the protein in

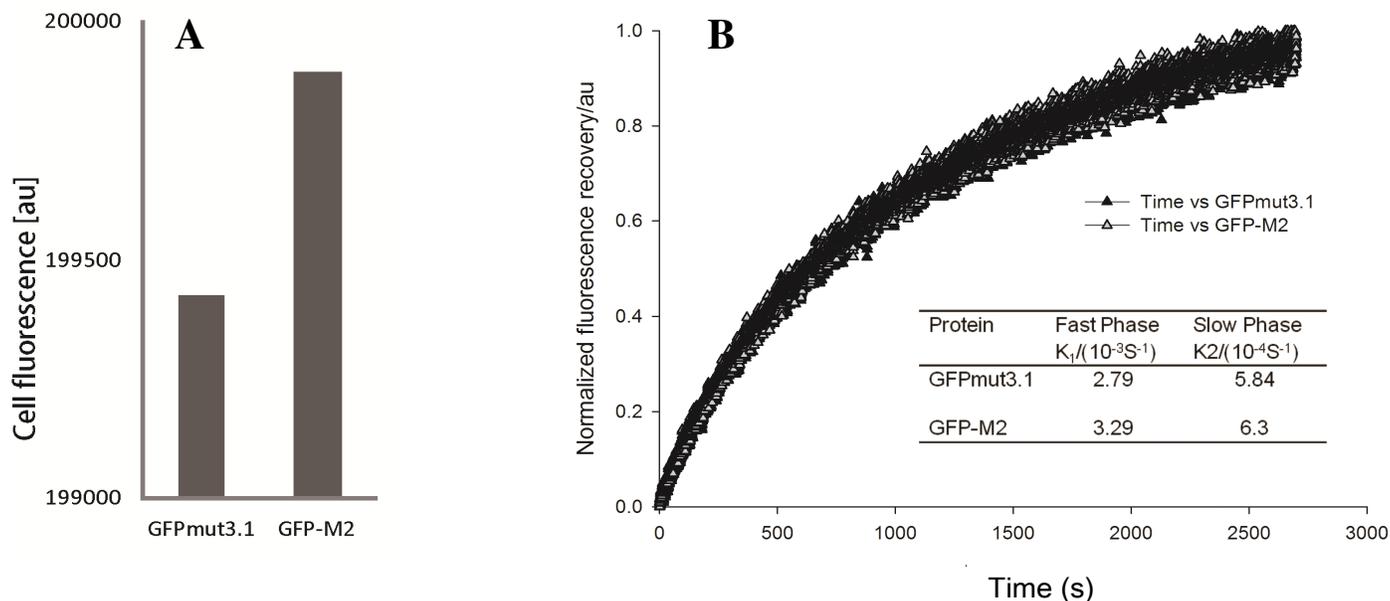


Figure 2. Expression level and functional productivity of GFPmut3.1 and GFP-M2 mutant. (A) Whole cell fluorescence of GFPmut3.1 and GFP-M2. (B) Refolding kinetics of GFPmut3.1 and GFP-M2.

limited time period.

ACKNOWLEDGMENTS

This research was supported by basic science research programme through the National Research foundation of Korea funded by Ministry of education, science and technology (2011-0011901) and this study also supported by 2011 Post doc development programme offered to first author in Pusan national university, Busan.

REFERENCES

- Arnold FH, Wintrodde PL, Miyazaki K, Gershenson A (2001). How enzymes adapt: lessons from directed evolution, *Trends Biochem. Sci.* 26: 100-106.
- Ayyadurai N, Saravanan Prabhu N, Deepankumar K, Lee SG, Jeong HH, Lee CS, Yun H (2011). Development of a selective, sensitive, and reversible biosensor by the genetic incorporation of a metal-binding site into green fluorescent protein. *Angew Chem. Int. Ed. Engl.* 50: 6534-6537.
- Budisa N (2005). *Engineering the Genetic Code - Expanding the amino acid repertoire for the design of novel proteins*, WILEY-VHC Weinheim, New York, Brisbane, Singapore, Toronto.
- Hurst LD, Feil EJ, Rocha EP (2006). Protein evolution: causes of trends in amino-acid gain and loss. *Nature*, 442: 11-12.
- Jordan IK, Kondrashov FA, Adzhubei IA, Wolf YI, Koonin EV, Kondrashov AS, Sunyaev S (2005). A universal trend of amino acid gain and loss in protein evolution. *Nature*, 433: 633-638.
- Kajander T, Cortajarena AL, Regan L (2006). Consensus design as a tool for engineering repeat proteins. *Methods Mol. Biol.* 340: 151-170.
- Lehmann M, Wyss M (2001). Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr. Opin. Biotechnol.* 12: 371-375.
- Lehmann M, Loch C, Middendorf A, Studer D, Lassen SF, Pasamontes L, van Loon AP, Wyss M (2002). The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* 15: 403-411.
- McDonald JH (2006). Apparent trends of amino Acid gain and loss in protein evolution due to nearly neutral variation. *Mol. Biol. Evol.* 23: 240-244.
- Nagasundarapandian S, Merkel L, Budisa N, Govindan R, Ayyadurai N, Sriram S, Yun H, Lee SG (2010). Engineering protein sequence composition for folding robustness renders efficient noncanonical amino acid incorporations. *Chem Bio. Chem.* 11: 2521-2524.
- Pluckthun A, Schaffitzel C, Hanes J, Jermutus L (2000). In vitro selection and evolution of proteins. *Adv. Protein Chem.* 55: 367-403.
- Sambrook J, Fritsch E, Maniatis T (1989). *Molecular cloning: A laboratory manual*. 2nd ed., Cold spring harbor laboratory press, Cold spring harbor, NY, USA.
- Steiner T, Hess P, Bae JH, Wiltschi B, Moroder L, Budisa N (2008). Synthetic biology of proteins: tuning GFPs folding and stability with fluoroproline. *PLoS One.* 3: e1680.
- Taverna DM, Goldstein RA (2002). Why are proteins marginally stable?. *Proteins.* 46: 105-109.
- Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.
- Tobin MB, Gustafsson C, Huisman GW (2000). Directed evolution: the 'rational' basis for 'irrational' design. *Curr. Opin. Struct. Biol.* 10: 421-427.
- Travaglini-Allocatelli C, Ivarsson Y, Jemth P, Gianni S (2009). Folding and stability of globular proteins and implications for function. *Curr. Opin. Struct. Biol.* 19: 3-7
- Van den Burg B, Eijssink VGH (2002). Selection of mutations for increased protein stability. *Curr. Opin. Biotechnol.* 13: 333-337.
- Wolschner C, Giese A, Kretschmar H, Huber R, Moroder L, Budisa N (2009). Design of anti and pro-aggregation variants to assess the effects of methionine oxidation in human prion protein. *Proc. Natl. Acad. Sci. USA* 106: 7756-7761.