Biotechnology and Bioprocess Engineering 19: 433-441 (2014)
DOI 10.1007/s12257-014-0162-x

RESEARCH PAPER

Modulation of Intracellular Protein Activity at Level of Protein Folding by Beta-turn Engineering

Bharat Madan and Sun-Gu Lee

Received: 9 March 2014 / Revised: 21 March 2014 / Accepted: 10 April 2014

Abstract Control of the intracellular protein activity is very important in various biological studies and biotechnology. This has generally been achieved at the transcription and translation levels. Although control of the intracellular activity at the protein folding level is conceptually possible, but there have been few studies. The present study examined this possibility by modulating the *in vivo* protein folding rate of green fluorescence protein (GFP) through beta-turn engineering. A type II' two residue beta-turn in GFP was targeted to generate two sets of mutants. First, a switch-off mutant was designed to stop the protein activity completely. The modulation mutants were then constructed to change the rates of GFP folding. The design of mutants was based on the rationale that residues i+1 and i+2 of a beta-turn have defined residue preferences, and their perturbation affects the rate of protein folding. The *in vivo* fluorescence activity of the designed GFP variants was switched off and modulated as expected. The change in the *in vivo* folding patterns of the mutants was confirmed by SDS-PAGE and found to be similar to the intracellular fluorescence activities of the mutants. The *in vitro* refolding kinetics performed with purified variants showed correlations with the *in vivo* folding patterns. These results showed that the beta-turns in a protein can be a target for modulating the *in vivo* protein folding pattern and activity.

Keywords: protein folding, protein activity, beta-turns, GFP

Bharat Madan, Sun-Gu Lee*
Department of Chemical Engineering, Pusan National University, Busan 609-735, Korea
Tel: +82-51-510-2786; Fax: +82-51-512-8563
E-mail: sungulee@pusan.ac.kr

1. Introduction

Controlling the intracellular protein activity is an important tool in biological research and biotechnology [1]. In particular, the emergence of synthetic biology has led to the development of more divergent and precise tuning systems for intracellular protein activities [2]. Most tuning systems rely on control of the transcription rate. For example, promoter engineering allows control of the intracellular protein activity at the transcription level [3]. Recent studies, such as untranslated region (UTR) engineering and anti-sense mRNA technology, have shown that it is also possible to control the intracellular activity at the translation stage [4]. Indeed, these artificial tuning systems facilitate the engineering of biological systems.

For proteins to function, the translated proteins need to fold into the desired three dimensional structures that are encoded in their amino acid sequences [5]. This suggests that the activity of a protein in a cell can also be controlled by modulating the protein folding process, but this hypothesis has not been tested seriously due to the lack of a proper understanding of the relation between the protein sequence and folding. On the other hand, extensive protein engineering studies on the protein folding dynamics have revealed the relationship between the protein sequence and folding [6,7]. For example, many studies have shown that a protein folding process can be modulated by engineering the beta-turn sequences rationally [8,9].

The abovementioned studies suggest that the protein folding process can be controlled by protein sequence design. On the other hand, most studies focused on modulating the *in vitro* protein folding property rather than controlling the *in vivo* protein activity. This study tested the hypothesis that the *in vivo* protein activity can be controlled at the protein folding level based on the following rationale.

In general, a delay in the protein folding rate increases the likelihood of misfolding a polypeptide chain in a cell, leading to the formation of inactive protein deposits, such as inclusion bodies [10]. Therefore, modulation of the protein folding rate can in turn help regulate the formation of the soluble active and insoluble inactive fractions of a protein. This suggests that the intracellular activity of a protein can be controlled by modulating the rate of folding. Here, the beta-turn, which is one of the most abundant non-repetitive secondary structures in the entire protein architecture [11], was selected as a target to modulate the protein folding rate. Beta-turns are found at the regions of chain reversal in proteins, and can be a control point in protein folding [12,13]. Their sequences, structures and critical roles in protein folding have been studied extensively over the past two decades [11,14,15].

Green fluorescent protein (GFP) was used as a model protein to test the possibility of intracellular protein activity control at the protein folding level by beta-turn engineering. GFP is a beta-sheet rich protein that contains a large number of beta-turns in its structure [16]. In addition, it is easier to monitor the folding and activity of GFP simultaneously by measuring its fluorescence. A two-residue beta-turn in GFP was selected as the target motif for controlling protein folding, and a series of mutants with various folding properties were designed based on the turn-type dependent statistical data and sequence alignment data for the target beta-turn. The designed mutants were expressed in *E. coli*, and their *in vivo* and *in vitro* folding properties were examined, which showed how the intracellular protein activity can be switched off completely or regulated by engineering the beta-turn sequences. This study can facilitate a consideration of protein sequence engineering and folding in controlling the intracellular protein activities.

2. Materials and Methods

2.1. Mutagenesis and cloning

The mutations were performed on an already cloned gene of GFPcon in pET30b expression vector [17]. The GFPcon gene contains a 6x His-tag at the N-terminal for purification purpose. The genes for Mut1, Mut2 and Mut3 were prepared by site directed mutagenesis method and the mutations were confirmed by sequence analysis.

2.2. Protein expression and purification

The plasmids carrying the genes were transformed in BL21 (DE3) *E. coli* cells for protein expression. Protein expression was induced by addition of 1 mM IPTG to a mid-log phase growth culture, at 37°C for 4 h. The cells were harvested by centrifugation at 4,000 rpm for 15 min at 4°C. The cell pellets were then resuspended in PBS for disruption using a French Press (Thermo IEC). The lysate was centrifuged at high speed (20,000 rpm) for 30 min at 4°C. The supernatant was subsequently filtered for column purification by Ni-NTA metal affinity chromatography using the standard procedures. The elution fractions were dialyzed against PBS pH 7.4 for further studies.

The samples for SDS-PAGE analysis were prepared from a cell pellet corresponding to 1 mL of 3 O.D. cells [18,19]. The cells were first disrupted using Bug Buster (Novagen), and centrifuged at 12,000 rpm for 30 min to separate the soluble and insoluble fractions. The SDS-PAGE analysis was performed in 12% polyacrylamide gels in Tris-glycine buffer. Gels were stained with Coomassie-blue staining solution.

2.3. Spectroscopic measurements

Protein concentration was determined as described previously [20,21]. Briefly, the concentration was calculated from the absorption at 280 nm. For the whole cell fluorescence (WCF) measurement, 200 μL of the expressed protein was centrifuged at 4,000 rpm for 10 min and the measurements were performed on a Perkin Elmer/Wallac Victor 2 Multilabel Counter (1420-011). The pellet was resuspended in 200 μL of PBS and dispensed in 96-well plates for measurement. The fluorescence was recorded at excitation 490 nm and emission 511 nm (slits = 5 nm). The fluorescence was normalized by dividing by OD_{600} nm of the 200 μL sample. The specific fluorescence activities of protein samples were measured by mixing equal concentration of each protein sample in PBS to a final volume of 200 μL and the fluorescence was read as mentioned above.

2.4. Refolding kinetics measurements

The protein refolding kinetics was measured as reported elsewhere [22]. Briefly, prior to refolding the protein was first fully unfolded in PBS containing 8 M urea and 5 mM DTT, at 95°C for 5 min. The denatured sample was then immediately diluted 100 fold to PBS with 5mM DTT. The refolding was monitored by measuring fluorescence using a Perkin Elmer LS-55 fluorescence spectrometer. The recovered fluorescence was normalized to the maximum fluorescence. The normalized data points were fitted to a single exponential model:

$$I(t) = y_o + A_1 \exp(-k_1 t) + C$$

where, I(t) is the intensity of the fluorescence signal at time t, y_o is the signal at time 0 sec. A_1 is the corresponding amplitude, and k_1 is the rate constants for refolding.

2.5. Molecular modeling and energy minimization studies

The modeling was performed on the type II' beta-turn

containing beta-hairpin extracted from GFPcon structure (PDB ID 1GFL). The mutations were made using Rosetta's fixed backbone design method [23,24]. In the fixed backbone design, the substituted amino acid side chain rotamers were applied to the fixed backbone of amino acid to be mutated and energy is calculated for the new structure. The structure is accepted or rejected based on the Metropolis criteria. For each mutant about 1,000 structures were generated and the structure with the lowest energy score was selected for further studies.

The energy minimization study was carried out using Gromacs (v 4.5.6) [25]. The modeled structures were centered in a cubic box with 1.0 nm edge and solvated with explicit solvent model SPC/E. The charges of the models were neutralized by adding counter ions Na^+/Cl^-. Finally, the energy minimization was formed for 500 ps until the potential energy, $E_{pot} > 1{,}000$ KJ/mol.

3. Results

3.1. Design of beta-turn variants

The native structure of GFP consists of 11 beta-strands joined by a range of loops and turns, which form 7 anti-parallel beta-hairpins. Among the turns and loops, a two residue tight β-turn at positions 115 ~ 118 was selected as the target turn for design because the sequences, structures and folding effects of these two residue beta-turns are well known. In general, a beta-turn consists of 4 residues; i to i+3, and are classified according to the phi/psi angles of the i+1 and i+2 residues [11,26]. In the case of a beta-turn in beta-hairpin, the residues, i+1 and i+2, form the turn region, whereas the i and i+3 residues occur at the beta-strands [27] (Fig. 1A). The residues comprising the target turn were $Glu115_i$- $Gly116_{i+1}$- $Asp117_{i+2}$- $Thr118_{i+3}$, and were identified to be a type II' turn. In a two residue beta-turn, all four residues are important for proper protein folding, but the central two residues are most important for driving the folding of a polypeptide chain. This was determined from previous studies on the role of beta-turns in protein folding [28,29]. The amino acid preferences in natural two residue beta-turns also support the importance of central two residues [30]. For example, the amino acid composition for the type II' beta-turns shows that positions i+1 and i+2 have very limited and specific preferences compared to positions i and i+3 [31] (Fig. 1B). Therefore, the variants with different folding rates were constructed by designing the central residues at the i+1 and i+2 positions of the target beta-turn in GFP based on the statistical data for type II' turns and the sequence alignment of the GFP homologues. Two types of variants were designed, a switch-off mutant that is expected to be misfolded completely, and a modulation mutant that is expected to exhibit a modulated folding rate.

3.2. Design of switch-off mutant

The design of a switch-off mutant suggests that the engineered mutation would lead to complete misfolding of the protein. The rationale used in designing this variant was the statistical preference of the residues at the i+1 and i+2 positions. As shown in Fig. 1B, Gly has the highest positional potential at position i+1, and position i+2 has very high preferences for Asp and Asn compared to the other amino acids [31]. This suggests that these amino acids may be important in the proper folding of a beta-hairpin with type II' turns. The native target turn in GFP already has the preferred amino acids at positions i+1 and i+2, *i.e.* Gly and Asp, respectively (GFP with the native target turn 'EGDT' is hereafter referred to as GFPcon). Therefore, it was expected that the substitution of both residues at positions i+1 and i+2 of the target turn with amino acids other than Gly and Asn/Asp would result in the complete misfolding of the polypeptide chains. The residues with the lowest positional preference at these

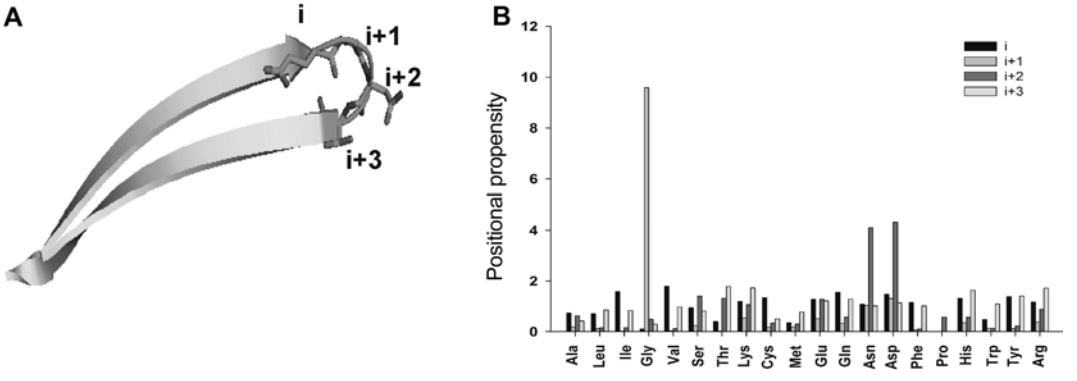
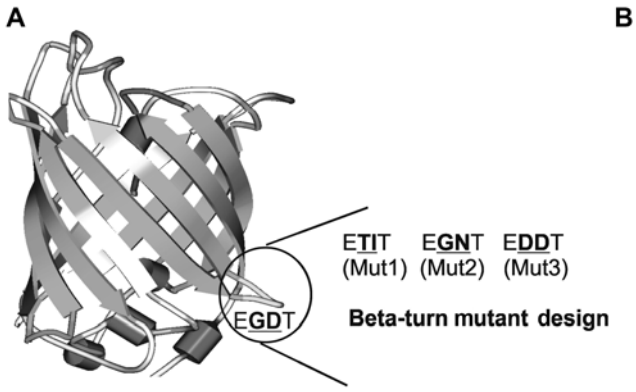


Fig. 1. (A) Schematic representation of a beta-hairpin with defined positions of turn residues. (B) Positional propensities of amino acids at positions i to i+3 of type II' beta-turns. The positional propensities were obtained from our previous study on beta-hairpin beta-turns [31].



B

Uniprot ID	Sequence alignment of i+1 and i+2 residues of β-turn EGDT	Turn type
Q8MU46	GD	II'
Q8T5F1	GD	II'
Q8I6J8	GD	II'
Q8MU48	GD	II'
Q963F5	GD	II'
Q9U6Y3	ED	II'
Q95UA7	GD	II'
Q9U6Y7	GN	II'
Q9GTJ7	DG	I'
Q95P04	GN	II'
Q8T5F0	GD	II'
Q9GPI6	GD	II'
Q8T5E7	GD	II'
Q8T6T9	GD	II'
Q8ISF8	DG	I'
Q9GPI5	GD	II'
Q95W11	GN	II'
Q963I9	DG	I'
Q9BLY9	ED	II'
P42212*	GD	II'
Consensus	.gd	

Fig. 2. (A) Beta-turn mutant design for a two residue turn (in yellow) in GFP. (B) Multiple sequence alignment of the designed turn in GFP homologues, obtained from pfam database. The MSA was performed using Multalin tool against GFPcon.

positions, *i.e.* Thr and Ile at positions i+1 and i+2, respectively, were selected to engineer the switch-off mutation, and the designed variant 'ETIT' was called Mut1 (Fig. 2A).

3.3. Design of modulation variant

As stated above, the positions, i+1 and i+2, of the type II' turn prefer Gly and Asp/Asn, respectively. In particular, i+1 showed a more exclusive preference than the i+2 position, suggesting that position i+1 would be more sensitive to mutations in terms of protein folding than i+2. This suggests that mutating the i+1 or i+2 residues to some other amino acid would have a modulating effect on the folding of GFP, but the effect would be different. Two variants of the target turn in GFP were designed to test this hypothesis. To reduce the possibility that the introduced mutation leads to the complete misfolding of GFP, the amino acids for the mutations were selected based on multiple sequence alignment (MSA) analyses of the target turn in the GFP homologues (Fig. 2B). For position i+2, Asp in the native GFP was the dominant one, and Asn was found to occur with the second highest preference. Therefore, position i+2 was mutated to Asn, and this variant with 'EGNT' was called Mut2. For position i+1, Gly was found to be a highly conserved residue, but Asp was also observed in some GFP homologues. This suggests that though Gly in general could be important for folding, but the presence of Asp in GFP homologues could also be accommodated in the turn structure of GFP, leading to a foldable protein. Therefore, position i+1 was substituted with Asp and this mutant, 'EDDT', was called Mut3.

3.4. Intracellular activities and folding patterns of the designed variants

$GFP_{mut3.1b}$, which was optimized for faster folding kinetics [32], was used as GFPcon with a native 'EGDT' turn sequence, and the designed Mut1, Mut2 and Mut3 sequences were introduced to GFPcon. To determine if the designed beta-turn sequences can alter the intracellular GFP activities at the folding level, the genes encoding GFP with the designed beta-turn sequences were expressed under the T7 promoter in *E. coli*. Their intracellular activities and folding patterns were then approximated by change in whole cell fluorescence (WCF) and *in vivo* soluble/insoluble expression levels.

3.5. *In vivo* expression profile of switch-off mutant

To understand the effects of the switch-off beta-turn sequence, 'ETIT', on the *in vivo* folding and activity of the protein, GFPcon and Mut1 were induced with IPTG, and their WCF and soluble and insoluble levels were monitored. As shown in Fig. 3A, the WCF of Mut1was found to be negligible compared to GFPcon. SDS-PAGE analysis confirmed that the entire protein of Mut1 was expressed as the insoluble form, whereas GFPcon showed higher soluble expression levels (Fig. 3B). This suggests that the designed switch-off mutant had been misfolded completely under these *in vivo* conditions, which led to switch-off of the intracellular activity, as expected. The rate of Mut1 growth after induction was much lower than that of GFPcon (Fig. 3C). The accumulation of the toxic inclusion bodies might reduce cell growth.

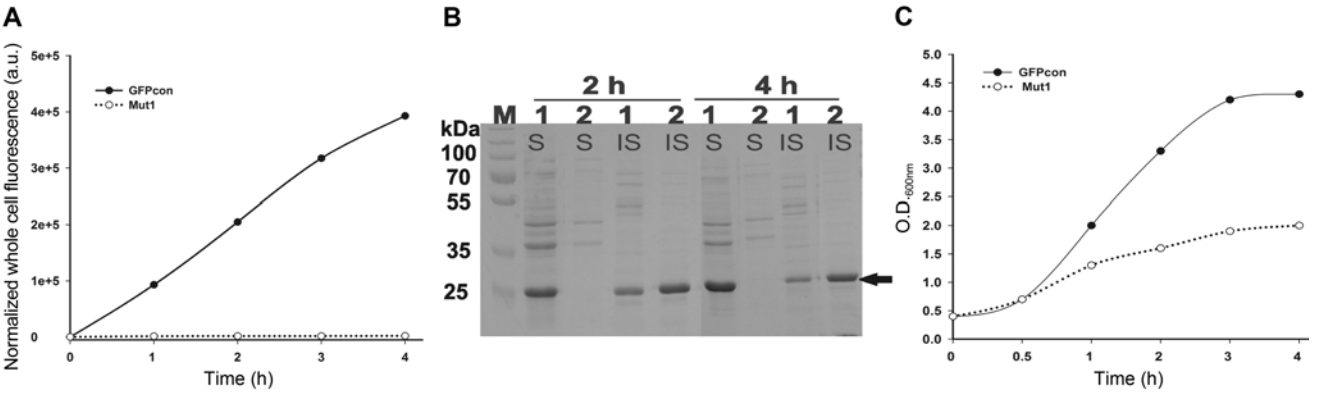


Fig. 3. Whole cell fluorescence, expression profile and cell growth curves of GFPcon and its variant Mut1. (A) Whole cell fluorescence (WCF), fluorescence values were normalized by dividing by the OD at 600 nm. (B) SDS PAGE of expressed protein samples at 2 and 4 h sampling time. Lanes 1 and 2 refer to soluble (S) and insoluble fractions (IS) of GFPcon and Mut1, respectively. The soluble and insoluble fractions are denoted by S and IS, respectively. (C) Growth curves for 4 h duration.

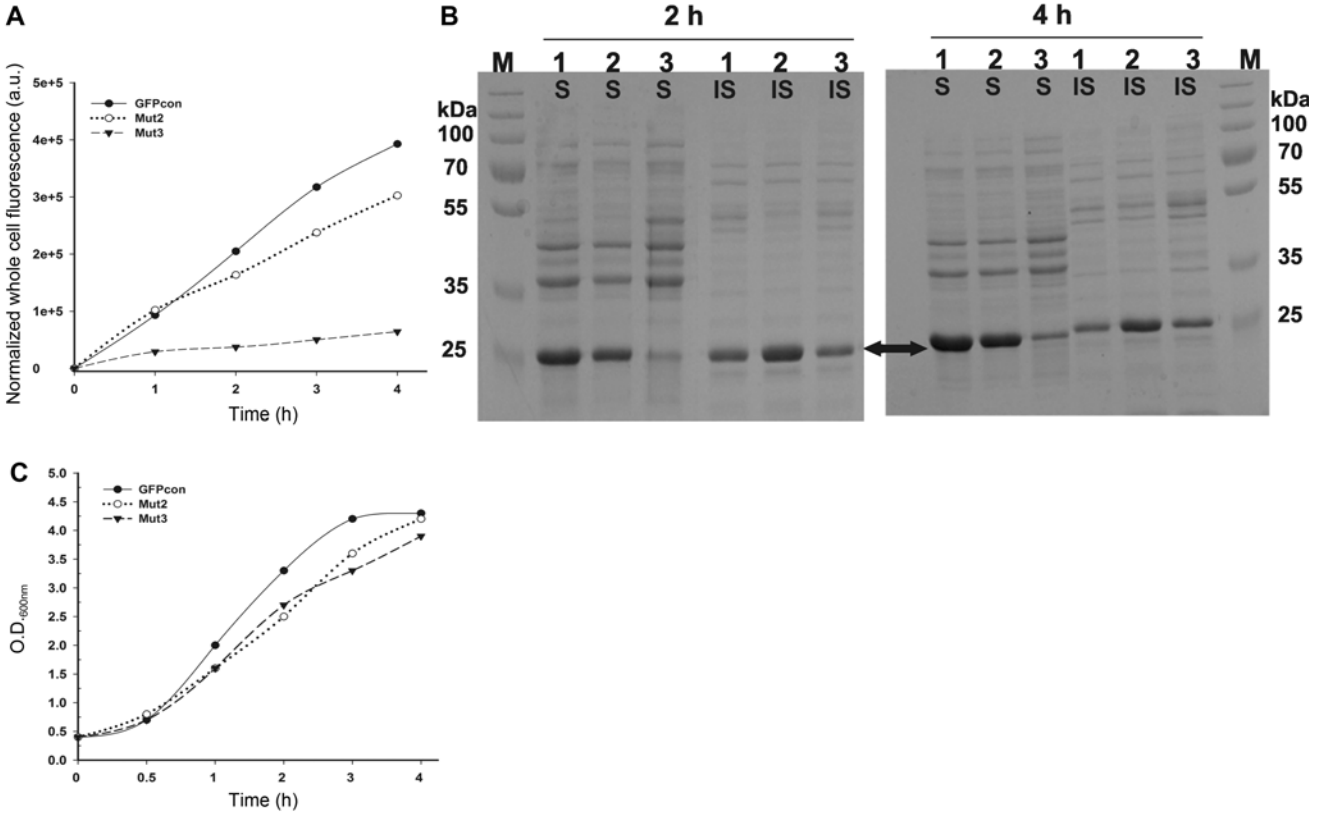


Fig. 4. Whole cell fluorescence, expression profile and cell growth plots of GFPcon and its mutants, Mut2 and Mut3. (A) Whole cell fluorescence; fluorescence values were normalized by dividing by the OD at 600 nm. Expression profile at (B) 2 h and 4 h from the time of induction. Lanes 1, 2, and 3 denote soluble (S) and insoluble fractions (IS) of GFPcon, Mut2 and Mut3, respectively. (C) Growth curves for 4 h duration.

3.6. *In vivo* expression properties of modulation variants

To determine if the designed modulation variants, *i.e.* Mut2 and Mut3, show a difference in the *in vivo* folding patterns and activities, as expected, they were expressed and monitored under the same conditions used for Mut1. As shown in Fig. 4A, the WCF of Mut2 was reduced slightly, and Mut3 showed a much lower WCF than GFPcon. When the proteins were expressed for 4 h, Mut2 and Mut3 showed approximately 77 and 16% WCF activities compared to GFPcon. The soluble and insoluble expression profiles of

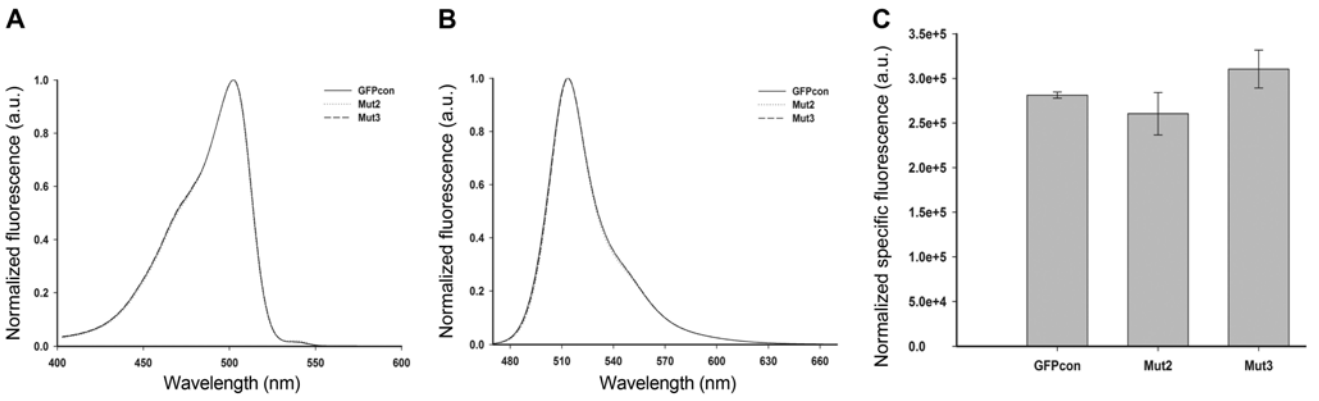


Fig. 5. Excitation and emission profiles and specific fluorescence activities of GFPcon and its variants Mut2 and Mut3. (A) Excitation maximum, (B) emission maximum, and (C) specific fluorescence activity.

GFPcon, Mut2 and Mut3 showed a correlation with the WCF (Fig. 4B). Mut2 was expressed as a slightly higher insoluble fraction than GFPcon. In the case of Mut3, only a limited fraction of the protein was expressed as the soluble form, which is responsible for the observed WCF, and most proteins were deposited as the insoluble fraction. These results suggest that, as designed, the mutation for the i+1 position of the target beta-turn affected the intracellular GFP folding patterns and activities more sensitively than the mutation for the i+2 position. These also confirmed that the intracellular activities of GFP can be modulated differently at the folding stage by beta-turn engineering. The cell growth of Mut2 and Mut3 was slower than GFPcon but faster than Mut1 (Fig. 4C).

3.7. *In vitro* characterization of the designed variants

The proposed strategy for modulating the intracellular activity of GFP was to adjust the intracellular protein folding pattern by changing the GFP folding rate through beta-turn design. The delay in the folding of polypeptide chains can lead to intracellular misfolding of the protein through the formation of non-native contacts or exposure of the hydrophobic core of the proteins. These results showed that the intracellular GFP activity under the induced promoter could be modulated at the folding level by a rational design of beta-turn, but it did not show directly that the folding rate of GFP was actually affected. To verify this directly, the *in vitro* folding rate of the designed variants should be compared *in vivo*, but this is a very difficult task. Therefore, the *in vitro* refolding kinetics of the variants was measured to determine if a correlation could be obtained between the *in vitro* refolding rates and *in vivo* folding properties of the designed mutants. For this study, the active variants, GFPcon, Mut2 and Mut3, were purified and characterized. The refolding rates for mutant Mut1 were not examined because the protein was expressed completely as insoluble inactive form.

First, the specific activities and excitation/emission profiles of the variants were compared to determine if the introduced mutation affects the chromophore environment (Fig. 5). The excitation and emission maximum were observed at 502 and 513 nm, respectively, for all proteins. The specific activities of all mutants were also almost similar to GFPcon. This suggests that the mutations for beta-turn engineering did not affect the intrinsic fluorescence of GFP.

To analyze the effects of mutations on the refolding rates of GFP, the refolding kinetics of the variants, Mut2 and Mut3, were examined and compared with those of GFPcon. The refolding rates of Mut2 and Mut3 at the fast phase were found to be 35 and 65% lower, respectively, than that of GFPcon (Fig. 6). These results suggest that there is an

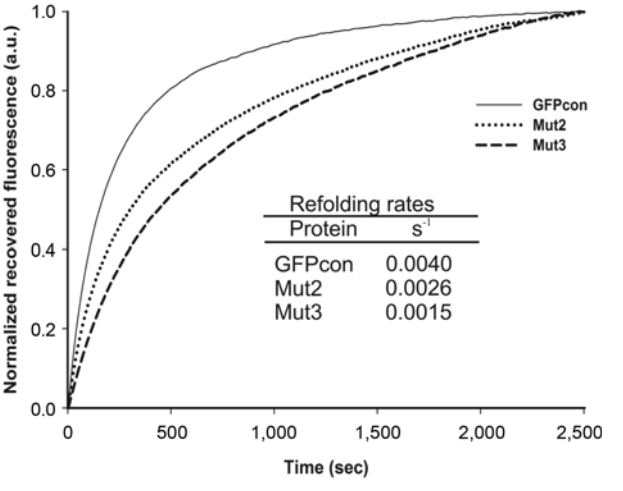


Fig. 6. *In vitro* refolding kinetics plots for GFPcon and its mutants Mut2 and Mut3. The inset shows the refolding rates obtained after fitting the normalized recovered fluorescence values using a single exponential model. The recovered fluorescence was normalized by dividing by the maximum fluorescence value.

approximate correlation between the *in vivo* expression pattern and *in vitro* refolding kinetics of the proteins. This also suggests that the rationale for controlling the *in vivo* protein folding properties and activities through the modulation of the folding rate is sound.

4. Discussion

This study reported the modulation of the *in vivo* protein activity at the protein folding level through the rational design of the beta-turns in GFP. Mutants were designed based on the assumptions that the positions i+1 and i+2 of the target beta-turn are important in protein folding, and their sensitivities to mutation are different. The intracellular GFP activity could be turned off by designing the beta-turn sequence that induces complete misfolding of the protein. In addition, it was also possible to control the intracellular activity of GFP by introducing mutations that can differentiate the protein folding rate. These results are meaningful in the following aspects. First, control of the intracellular activity, which has been limited to the transcription and translation levels, was attempted at the folding level. Second, the *in vivo* protein folding pattern could be modulated artificially by rationally designing the sequence of certain regions in the proteins, such as beta-turns.

In the present study, the mutants, Mut2 and Mut3, properly folded to the active form, but showed lower folding rates with differences compared to GFPcon. A direct understanding of the folding properties of these mutants at the molecular levels is quite difficult because the simulation of the folding dynamics for a large protein, such as GFP, is not feasible [33,34]. Therefore, this study examined the effect of the introduced mutations on the folding dynamics of GFP by analyzing indirectly the modeled static structures of the variants using energy minimization. This study analyzed the changes in the local environment of the designed turns in the variants, such as the turn residue interactions with neighboring residues or water because these local interactions affect protein folding, stability and aggregation properties. Fig. 7 shows

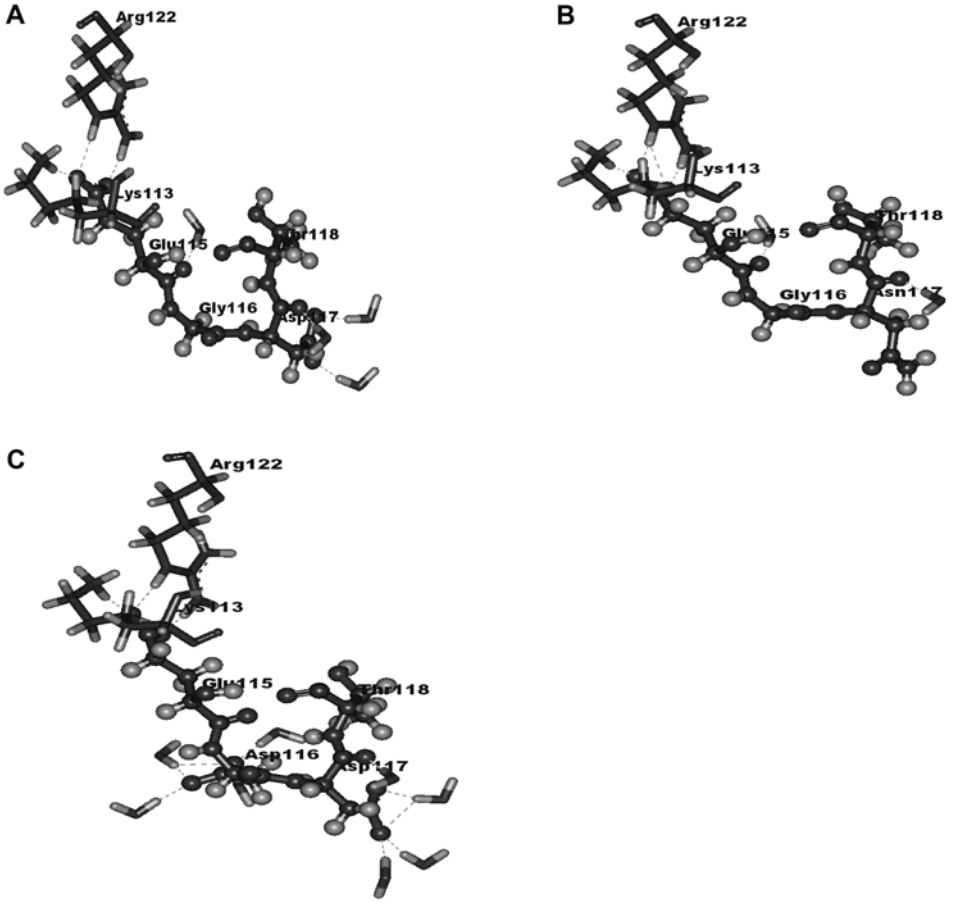


Fig. 7. Turn residue interactions with neighboring residues and surrounding water molecules in the modeled structures. The dashed lines in green color indicate hydrogen bonds. (A) GFPcon, (B) Mut2, and (C) Mut3.

Table 1. Turn residue interactions with neighboring residues and water molecules for the modeled structures in the energy minimization study

Protein	Hydrogen bonds with neighboring residues		Hydrogen bonds with water			
	115-113	115-122	115	116	117	118
GFPcon (E**GD**T)	1	2	1	0	3	0
Mut2 (E**GN**T)	1	3	1	0	1	0
Mut3 (E**DD**T)	1	2	1	5	5	0

the modeled turn structures and Table 1 lists the number of hydrogen bonds between the turn residues and water, as well as between the residues neighboring the turn obtained after the energy minimization study. The minimization study showed some changes in the turn residue interactions for the variants Mut2 and Mut3. In case of Mut2, two water mediated hydrogen bonds between the turn residue 117 were lost and an additional hydrogen bond interaction with the side chains of Glu115 and neighboring Arg122 residue was found. The variant Mut3 formed 7 additional hydrogen bonds with water compared to GFPcon. Hence, one possible explanation behind the difference in the folding patterns of GFPcon and the variant Mut2 and Mut3 could be the perturbations in the structural environment surrounding the turns in the variants. On the other hand, the examination of the modeled turn structures revealed no significant difference in the overall turn structures of the variants, which might support the results that the folding of the variants was not completely abolished. Further studies with more sophisticated simulation techniques, such as replica exchange dynamics [35] or meta-dynamics [36,37], will be needed to better understand the effect of the mutations on the protein folding rate.

Although it was found that control of the intracellular protein activity was possible by designing the protein folding rate, there may be many problems in the practical use of the idea. For example, the *in vivo* protein folding process is affected by many factors, such as temperature, protein expression rate, and size/structure of the target proteins [38,39]. Therefore, a change in these factors can alter the intracellular activity patterns for the same design concept. In addition, the tuning efficiency of the intracellular activity at a folding level might be much less efficient compared to the transcription level and translation level controls. The protein sequence design for the intracellular activity control is expected to become useful if the approach is combined with other methods. For example, modulation of the transcription rate using the designed variants with different folding rates might result in intracellular protein activities that cannot be obtained by transcription rate control alone.

Acknowledgements

This work was supported by a 2-Year Research Grant for Pusan National University.

References

1. Arpino, J. A., E. J. Hancock, J. Anderson, M. Barahona, G. B. Stan, A. Papachristodoulou, and K. Polizzi (2013) Tuning the dials of synthetic biology. *Microbiol*. 159: 1236-1253.
2. Grünberg, R. and L. Serrano (2010) Strategies for protein synthetic biology. *Nucleic Acids Res.* 38: 2663-2675.
3. Blount, B. A., T. Weenink, S. Vasylechko, and T. Ellis (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS One*. 7: e33279.
4. Olson, E. J. and J. J. Tabor (2012) Post-translational tools expand the scope of synthetic biology. *Curr. Opin. Chem. Biol*. 16: 300-306.
5. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Sci*. 181: 223-230.
6. Matouschek, A., J. T. Kellis, Jr., L. Serrano, and A. R. Fersht (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340: 122-126.
7. Huang, J., D. Xing, and W. Huang (2012) Relationship between protein folding kinetics and amino acid properties. *Amino Acids* 43: 567-572.
8. McCallister, E. L., E. Alm, and D. Baker (2000) Critical role of beta-hairpin formation in protein G folding. *Nat. Struct. Biol*. 7: 669-673.
9. Martinez, J. C., M. T. Pisabarro, and L. Serrano (1998) Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat. Struct. Biol*. 5: 721-729.
10. Ecroyd, H. and J. A. Carver (2008) Unraveling the mysteries of protein folding and misfolding. *IUBMB Life*. 60: 769-774.
11. Hutchinson, E. G. and J. M. Thornton (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Sci*. 3: 2207-2216.
12. Lewis, P. N., F. A. Momany, and H. A. Scheraga (1973) Chain reversals in proteins. *Biochim. Biophys. Acta* 303: 211-229.
13. Chou, P. Y. and G. D. Fasman (1979) Conservation of chain reversal regions in proteins. *Biophys. J.* 26: 385-399.
14. Gunasekaran, K., C. Ramakrishnan, and P. Balaram (1997) Beta-hairpins in proteins revisited: Lessons for de novo design. *Protein Eng*. 10: 1131-1141.
15. Marcelino, A. M. and L. M. Gierasch (2008) Roles of beta-turns in protein folding: From peptide models to protein engineering. *Biopolymers* 89: 380-391.
16. Yang, F., L. G. Moss, and G. N. Phillips (1996) The molecular

structure of green fluorescent protein. *Nat. Biotechnol*. 14: 1246-1251.
17. Sokalingam, S., G. Raghunathan, N. Soundrarajan, and S. G. Lee (2012) A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS One*. 7: e40410.
18. Muthukutty, P., T. Kato, and E. Park (2013) Production of human papillomavirus 6b L1 virus-like particles incorporated with enhanced green fluorescent whole protein in silkworm larvae. *Biotechnol. Bioproc. Eng.* 18: 514-519.
19. Tøndervik, A., S. Balzer, T. Haugen, H. Sletta, M. Rode, K. Lindmo, T. Ellingsen, and T. Brautaset (2013) High production of recombinant Norwegian salmonid alphavirus E1 and E2 proteins in *Escherichia coli* by fusion to secretion signal sequences and removal of hydrophobic domains. *Biotechnol. Bioproc. Eng*. 18: 742-750.
20. Raghunathan, G., S. Sokalingam, N. Soundrarajan, B. Madan, G. Munussami, and S. G. Lee (2013) Modulation of protein stability and aggregation properties by surface charge engineering. *Mol. Biosyst*. 9: 2379-2389.
21. Raghunathan, G., S. Sokalingam, N. Soundrarajan, G. Munussami, B. Madan, and S. -G. Lee (2013) A comparative study on the stability and structure of two different green fluorescent proteins in organic co-solvent systems. *Biotechnol. Bioproc. Eng.* 18: 342-349.
22. Merkel, J. S. and L. Regan (2000) Modulating protein folding rates *in vivo* and *in vitro* by side-chain interactions between the parallel β strands of green fluorescent protein. *J. Biol. Chem.* 275: 29200-29206.
23. Kaufmann, K. W., G. H. Lemmon, S. L. DeLuca, J. H. Sheehan, and J. Meiler (2010) Practically useful: What the rosetta protein modeling suite can do for you. *Biochem*. 49: 2987-2998.
24. Sokalingam, S., B. Madan, G. Raghunathan, and S. -G. Lee (2013) *In silico* study on the effect of surface lysines and arginines on the electrostatic interactions and protein stability. *Biotechnol. Bioproc. Eng.* 18: 18-26.
25. Pronk, S., S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 29: 845-854.
26. Venkatachalam, C. M. (1968) Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* 6: 1425-1436.
27. Sibanda, B. L., T. L. Blundell, and J. M. Thornton (1989) Conformation of β-hairpins in protein structures: A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol*. 206: 759-777.
28. Ramirez-Alvarado, M., F. J. Blanco, H. Niemann, and L. Serrano (1997) Role of beta-turn residues in beta-hairpin formation and stability in designed peptides. *J. Mol. Biol*. 273: 898-912.
29. de Alba, E., M. Rico, and M. A. Jimenez (1999) The turn sequence directs beta-strand alignment in designed beta-hairpins. *Protein Sci*. 8: 2234-2244.
30. Simpson, E. R., J. K. Meldrum, and M. S. Searle (2006) Engineering diverse changes in beta-turn propensities in the N-terminal beta-hairpin of ubiquitin reveals significant effects on stability and kinetics but a robust folding transition state. *Biochem*. 45: 4220-4230.
31. Madan, B., S. Seo, and S. -G. Lee (2014) Structural and sequence features of two residue turns in beta-hairpins. *Proteins: Struct. Funct. Bioinform.* In press: doi: 10.1002/prot.24526.
32. Cormack, B. P., R. H. Valdivia, and S. Falkow (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*. 173: 33-38.
33. Snow, C. D., E. J. Sorin, Y. M. Rhee, and V. S. Pande (2005) How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct*. 34: 43-69.
34. Lane, T. J., D. Shukla, K. A. Beauchamp, and V. S. Pande (2013) To milliseconds and beyond: Challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* 23: 58-65.
35. Sugita, Y. and Y. Okamoto (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett*. 314: 141-151.
36. Bussi, G., F. L. Gervasio, A. Laio, and M. Parrinello (2006) Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc*. 128: 13435-13441.
37. Spiwok, V., B. Kralova, and I. Tvaroska (2008) Continuous metadynamics in essential coordinates as a tool for free energy modelling of conformational changes. *J. Mol. Mod*. 14: 995-1002.
38. Hartl, F. U. and M. Hayer-Hartl (2009) Converging concepts of protein folding *in vitro* and *in vivo*. *Nat. Struct. Mol. Biol*. 16: 574-581.
39. Cabrita, L. D., C. M. Dobson, and J. Christodoulou (2010) Protein folding on the ribosome. *Curr. Opin. Struct. Biol*. 20: 33-45.