N-terminal domains of native multidomain proteins have the potential to assist de novo folding of their downstream domains in vivo by acting as solubility enhancers

CHUL WOO KIM,[1] KYOUNG SIM HAN,[2] KI-SUN RYU,[1] BYUNG HEE KIM,[1] KYUN-HWAN KIM,[3] SEONG IL CHOI,[1,4] AND BAIK L. SEONG[1,2,4]

[1]Department of Biotechnology, College of Engineering, Yonsei University, Seodaemun-Gu, Seoul 120-749, Korea
[2]Protheon Incorporated, Yonsei Engineering Research Center B120E, Seodaemun-Gu, Seoul 120-749, Korea
[3]Department of Pharmacology, School of Medicine, Konkuk University, 380-701, Korea
[4]Institute of Life Science and Biotechnology, Yonsei University, Seodaemun-Gu, Seoul 120-749, Korea

(RECEIVED May 7, 2006; FINAL REVISION November 13, 2006; ACCEPTED December 19, 2006)

Abstract

The fusion of soluble partner to the N terminus of aggregation-prone polypeptide has been popularly used to overcome the formation of inclusion bodies in the *E. coli* cytosol.

The chaperone-like functions of the upstream fusion partner in the artificial multidomain proteins could occur in de novo folding of native multidomain proteins. Here, we show that the N-terminal domains of three *E. coli* multidomain proteins such as lysyl-tRNA synthetase, threonyl-tRNA synthetase, and aconitase are potent solubility enhancers for various C-terminal heterologous proteins. The results suggest that the N-terminal domains could act as solubility enhancers for the folding of their authentic C-terminal domains in vivo. Tandem repeat of N-terminal domain or insertion of aspartic residues at the C terminus of the N-terminal domain also increased the solubility of fusion proteins, suggesting that the solubilizing ability correlates with the size and charge of N-terminal domains. The solubilizing ability of N-terminal domains would contribute to the autonomous folding of multidomain proteins in vivo, and based on these results, we propose a model of how N-terminal domains solubilize their downstream domains.

Keywords: fusion; multidomain proteins; de novo folding; N-terminal domains; solubility enhancers; charge; size

How proteins efficiently fold in the crowded cytosol remains one of the fundamental questions in biology. Three major molecular chaperones such as DnaK, GroEL/GroES, and trigger factor (TF) assist the folding of only a small fraction of newly synthesized proteins in the *Escherichia coli* cytoplasm (Hartl and Hayer-Hartl 2002; Kerner et al. 2005). The *E. coli* strains lacking DnaK or TF exhibit no apparent folding defects, whereas the combined deletion of two genes causes protein aggregation and synthetic lethality (Deuerling et al. 1999). However, the above problems are circumvented by the growth below 30°C or by overexpression of either GroEL/GroES or SecB (Ullers et al. 2004; Vorderwülbecke et al. 2004). The *E. coli* GroEL is essential for viability under all conditions tested (Fayet et al. 1989). Strikingly, however, GroEL/GroES are absent or not essential in the mycoplasmas, proposed to include minimal sets of genes for viability due to their small genome size (Wong and Houry 2004).

Small single-domain proteins are thought to fold spontaneously upon termination of translation and release

Reprint requests to: Seong Il Choi, Institute of Life Science and Biotechnology, Yonsei University, 134 Shinchon-Dong, Seodaemun-Gu, Seoul 120-749, Korea; e-mail: seongilchoi@daum.net; fax: 82-2-392-3582.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.062330907.

from ribosomes. In contrast, the folding of multidomain proteins is facilitated by a cotranslational folding process in vivo (Kramer et al. 2001). So far, the role of cotranslationally or independently folded domains in the folding of other domains remains largely unknown. On the other hand, the fusion technology, the fusion of soluble protein to the N terminus of aggregation-prone polypeptide, has been the most efficient and popular tool to overcome the formation of inclusion bodies in the *E. coli* cytosol (Braun and LaBaer 2003; Esposito and Chatterjee 2006). In contrast, the coexpression of molecular chaperones has been effective only for a limited number of heterologous proteins (Wall and Plückthun 1995). The *E. coli* maltose-binding protein (MBP) was proposed and referred to function as a general molecular chaperone in the context of a fusion protein (Kapust and Waugh 1999; Bach et al. 2001). Interestingly, the fusion proteins are kinds of multidomain proteins in which the N-terminal domains act as solubility enhancers for their downstream domains. This phenomenon could occur in the de novo folding of native multidomain proteins in vivo.

Here, we report that the N-terminal domains of native multidomain proteins promote the solubility of their C-terminal insoluble heterologous proteins. Based on the results, we suggest that like the commonly used solubility-enhancing fusion partners, the N-terminal domains as solubility enhancers could assist the folding of their authentic C-terminal domains. It was investigated what factors of N-terminal domains are important for their solubilizing ability. By combining our results with the well-known charge effect on protein solubility, we suggest a model of how the N-terminal domains solubilize their downstream domains.

Results

The N-terminal domains promote the solubility of their C-terminal heterologous proteins

To show the occurrence of the chaperone-like type in the artificial multidomain proteins during de novo folding of native multidomain proteins and to assess the potential role of the N-terminal domains in vivo, we used an indirect approach depicted in Figure 1. For the purpose, the native N-terminal domains were fused to the N terminus of the aggregation-prone heterologous proteins.

Three *E. coli* multidomain proteins such as lysyl-tRNA synthetase (LysRS, 57 kDa) encoded by *lysS* gene, threonyl-tRNA synthetase (ThrRS, 75 kDa), and aconitase (AcnB, 94 kDa) were selected. The corresponding N-terminal domains, LysN (residues 1–154 of LysRS), ThrN (residues 1–225 of ThrRS), and AcnN (residues 1–160 of AcnB), were selected on the basis of their known three-dimensional structures (Sankaranarayanan et al. 1999; Onesti et al. 2000; Williams et al. 2002). All N-terminal domains were expressed as soluble form at 37°C (Fig. 2A). As a control, thioredoxin (Trx) with a domain size, one of the well-known solubility-enhancing fusion partners (LaVallie et al. 1993), was compared. Aggregation-prone proteins such as *Aquorea victoria* green fluorescence protein (GFP), human granulocyte colony-stimulating factor (GCSF), and tobacco etch virus protease (TEVP) with the C-terminal hexahistidine tag were expressed predominantly or significantly as inclusion bodies at 30°C (Fig. 2B). The N-terminal domains promoted the solubility of their C-terminal proteins at 30°C and were more effective than Trx (Fig. 2C). In addition, the Western blot of proteins of interest was approximately in accord with the expression patterns on the SDS-PAGE. The solubility of proteins are summarized in Figure 2D. The solubilizing ability of N-terminal domains in the fusion proteins suggests that in the same context, the N-terminal domains could act as solubility enhancers for their authentic C-terminal domains in vivo.

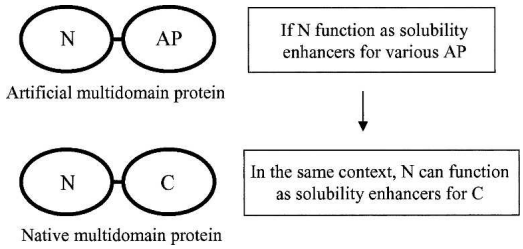

Figure 1. A schematic diagram for assessing the potential role of N-terminal domains as solubility enhancers for their authentic C-terminal domains in vivo. N, C, and AP represent N- and C-terminal domains of native multidomain proteins and aggregation-prone proteins, respectively.

To examine the role of the N-terminal domains in more detail, LysRS, LysN, and LysC (residues 155–505 of LysRS) were expressed at 37°C and 42°C. The expressed LysRS and LysN were almost soluble at both temperatures, whereas the significant fraction of LysC (~10% at 37°C and 20% at 42°C) was expressed as insoluble aggregates (Fig. 2E). The results suggest that the folded LysN with the solubilizing ability might be helpful for the de novo folding of LysC in vivo. Distinct from the LysN-fused proteins in Figure 2C, LysN exhibit the native interdomain interactions with LysC in the folded LysRS (Onesti et al. 2000). Probably LysN might be a more effective solubility enhancer for LysC than the heterologous proteins.

The N-terminal domain-fused proteins exhibit the functional activities of C-terminal proteins

The solubility of fusion proteins does not necessarily indicate the proper folding of the C-terminal reporter

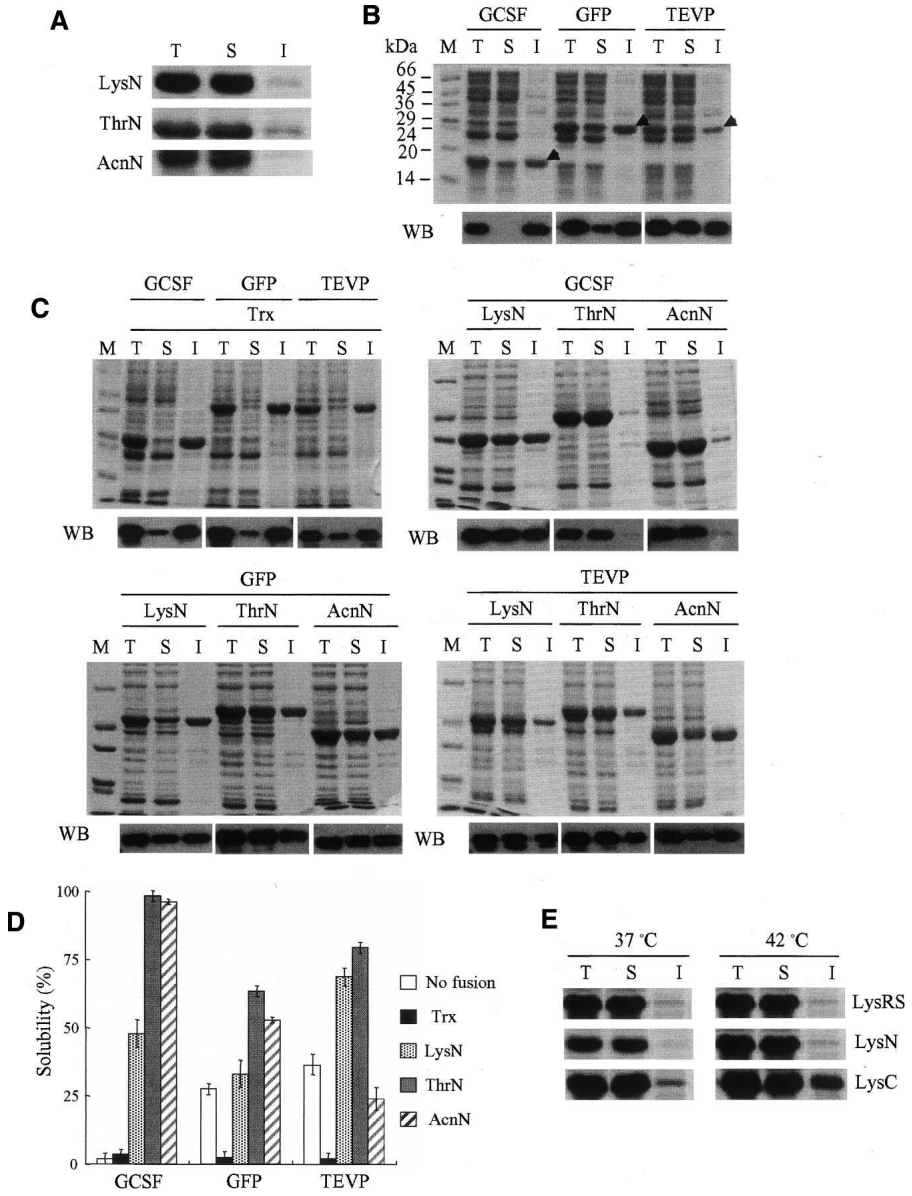

Figure 2. The solubilizing ability of the N-terminal domains of native multidomain proteins for aggregation-prone heterologous proteins. The expressed proteins were analyzed by SDS-PAGE. T, S, and I represent induced total lysates, soluble fraction, and insoluble fraction, respectively. The Western blot (WB) data were shown under the corresponding SDS-PAGE results. The same molecular weight marker was used in all SDS-PAGE. (*A*) The direct expression of N-terminal domains of native multidomain proteins at 37°C. (*B*) The direct expression of aggregation-prone heterologous proteins at 30°C. (*C*) The expression of fusion proteins at 30°C. (*D*) The solubility of tested proteins in *B* and *C* is summarized. (*E*) The expression of LysRS, LysN, and LysC at 37°C and 42°C.

proteins. To monitor the folding of the C-terminal proteins in the soluble fusion proteins, the functional assays of fusion proteins were performed. The purified TEVP fusion proteins (LysN-, ThrN-, and AcnN-TEVP) with C-terminal histidine tags and commercially available TEVP as a positive control were shown to efficiently cleave the substrate protein, MBP-GCSF containing TEVP recognition site between MBP and GCSF (Fig. 3A). Additionally, the relative fluorescence intensity of GFP and GFP fusion proteins were compared at an equimolar level. The relative fluorescence intensity obtained from each soluble extract containing GFP or GFP fusion protein showed that the soluble GFP fusion proteins exhibit the specific fluorescence intensity comparable with that of intact GFP (Fig. 3B).

Folding yield can be determined by multiplying relative specific activity by solubility. The relative specific fluorescence of GFP, LysN-GFP, ThrN-GFP, and AcnN-GFP

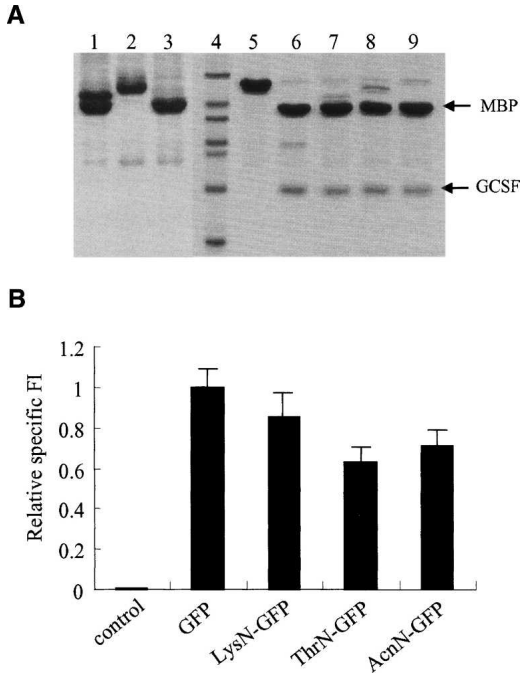

Figure 3. The functional assays of the N-terminal domain-fused proteins. (*A*) The TEVP fusion proteins, substrate protein, and cleaved mixtures were analyzed by SDS-PAGE. (Lanes *1–3*) Purified LysN-, ThrN- and AcnN-TEVP; (lane *4*) molecular weight marker; (lane *5*) purified substrate MBP-GCSF; (lanes *6–9*) MBP-GCSF was incubated with rTEVP (Invitrogen, lane *6*), LysN-TEVP (lane *7*), ThrN-TEVP (lane *8*), and AcnN-TEVP (lane *9*), respectively. (*B*) The relative specific fluorescence intensity (FI) of GFP and each GFP fusion protein in the soluble extract was compared. The value for GFP was set to one for comparison.

are 1, 0.85, 0.63, and 0.71, respectively, and their solubilities in Figure 2D are 27%, 33%, 63%, and 52%, respectively. Therefore, their folding yields are ~27%, 28%, 40%, and 37%, respectively. Despite the lower specific activity of GFP fusion proteins, the folding yields of GFP by the N-terminal domain fusion increase. It is possible that the N-terminal domains can interfere with the functions of their linked proteins in the fusion context, although the C-terminal proteins achieve native conformations, resulting in lower specific activity.

The solubilizing ability of N-terminal domains correlates with their charge

The solubility of fusion partners that closely correlates with their average net charge has served as a general indicator of their ability to solubilize their linked proteins (Davis et al. 1999). To investigate the contribution of the charge effect of N-terminal domains on their solubilizing ability, linker peptides harboring different charge contents were inserted between LysN (or Trx) and C-terminal GCSF. The insertion of the small linker peptide is expected not to alter the thermodynamic stability and folding rate of N-terminal domains significantly. The linker peptides tested here include six consecutive arginines(R_6), alanines (A_6), aspartic acids (D_6), $D_2(ST)_2$, and D_4ST, and were compared with serine-threonine repeated residue, $(ST)_3$, used in the original construct in Figure 2C.

As shown in Figure 4A, the negatively charged tags increased the solubility of the LysN-GCSF fusion protein, approximately proportional to the number of inserted aspartic acids (49%, 82%, 94%, and 95% for $[ST]_3$, $D_2[ST]_2$, D_4ST, and D_6, respectively). In contrast, R_6 and A_6 tags had negative or little effect on the solubility of fusion proteins when compared to the $(ST)_3$ tag (33% and 49% for R_6 and A_6, respectively). A similar or even more dramatic effect of inserted tags was observed in the Trx-GCSF fusion protein. As shown in Figure 4B, the insertion of negatively charged tags between Trx and GCSF resulted in great improvement of the solubility of the fusion proteins (7%, 17%, 58%, and 86% for $[ST]_3$, $D_2[ST]_2$, D_4ST, and D_6, respectively). Again, the

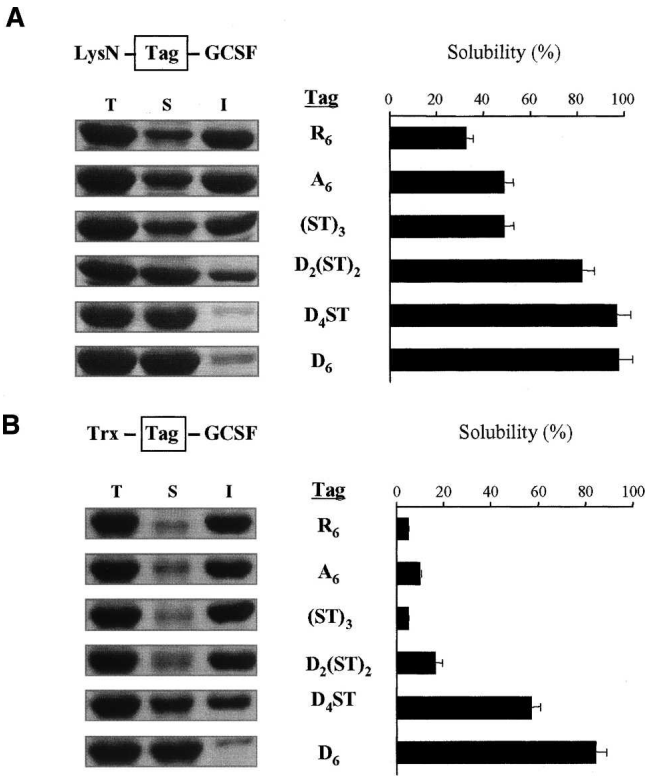

Figure 4. The charge effect of small tags between N-terminal domain and target protein on the solubility of fusion proteins. The various linker tags harboring different charge content were inserted between LysN and GCSF (*A*) and between Trx and GCSF (*B*), and their effect on the solubility of fusion proteins was analyzed by SDS-PAGE. All fusion proteins were expressed at 30°C.

insertion of R_6 and A_6 tags little affected the solubility of fusion proteins (10% and 7% for R_6 and A_6, respectively). The results seem to be in accord with the obvious charge effects on protein solubility (Otzen et al. 2000; Uversky et al. 2000; Chiti et al. 2002) and suggest that the solubilizing ability of N-terminal domains might correlate with their charge. However, the mechanism of the negatively charged linker peptide on the solubility is still unclear.

The solubilizing ability of N-terminal domains correlates with their size

Both solubility and average net charge are intensive properties, not representing quantitatively total net charge and protein size. Assuming that the total electrostatic repulsions by the surfaced-exposed charged residues of folded proteins are one of the important forces for the solubilization of their linked polypeptides, the protein size or number of domains is expected to correlate with their solubilizing potential.

In order to investigate the potential effect of the overall size of N-terminal domains on their solubilizing ability, we compared the single LysN domain with the tandem repeat of LysN ($LysN_2$), where the size and total net charge are doubled while keeping the average net charge unchanged. The solubility of LysN-GFP and $LysN_2$-GFP was ~30% and 50% at 30°C, respectively (Fig. 5A), suggesting that the size of N-terminal domains might correlate with their solubilizing ability. Importantly, the results suggest that as the number of folded domains increases during de novo folding with their solubility little changed, their solubilizing ability could be further increased.

LysRS and LysN, when expressed alone without fusion, exhibit a similar level of solubility (Fig. 2E). However, the size effect predicts that LysRS could be more effective than LysN in terms of solubilizing ability. LysN-GFP, LysC-GFP, and LysRS-GFP, were expressed at 30°C. As shown in Figure 5B, LysRS-GFP was almost soluble (95%), whereas the majority of the expressed LysN-GFP and LysC-GFP were insoluble. The results further support the correlation between the size of N-terminal domains and their solubilizing ability. Moreover, they suggest that the synergistic solubility-enhancing ability by folded LysN and LysC for GFP exist, probably due to the native interdomain packing between two-folded domains. The additive and/or synergistic solubilizing ability of folded domains might contribute to the autonomous folding of larger native multidomain proteins in vivo. Contrary to our results in vivo, the in vitro refolding experiments have revealed that domains interfere with the folding of other domains (Creighton 1992; Netzer and Hartl 1997; Inaba et al. 2000).

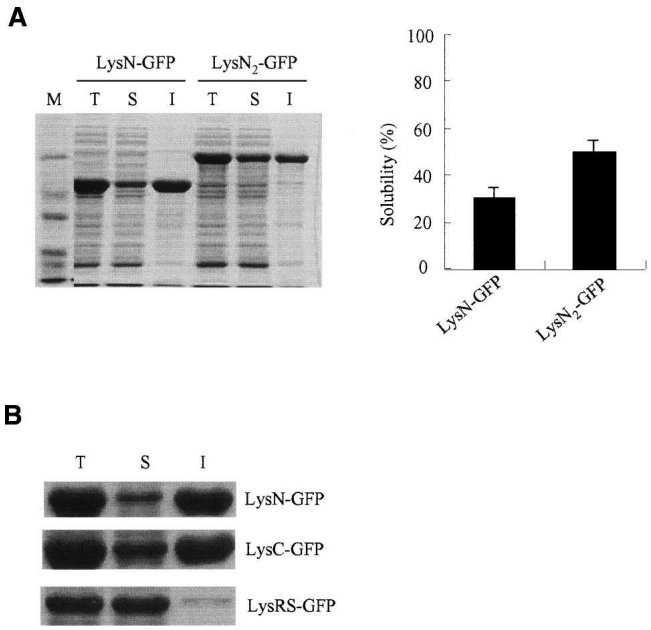

Figure 5. The size effect of N-terminal domains on their solubilizing ability. (*A*) LysN and tandem repeat ($LysN_2$) were fused to GFP, and fusion proteins were expressed at 30°C. (*B*) LysN, LysC, and LysRS, were fused to GFP, respectively, and the fusion proteins were expressed at 30°C.

The solubilizing ability of MBP is not inhibited by the insertion of a large tag between MBP and target protein

Mechanistically, MBP was proposed to act as a general molecular chaperone in the fusion context through transient and direct hydrophobic interactions between the exposed hydrophobic sites of MBP and aggregation-prone folding intermediates of downstream proteins (Kapust and Waugh 1999), similar to the mechanism of the current molecular chaperones. In contrast, the solubilizing ability based on the charge effect on protein solubility does not necessarily require the direct intramolecular interactions between fusion partner and downstream protein.

To distinguish between the two possibilities, we inserted large polypeptides such as Trx and LysN between MBP and target proteins (GFP and TEVP) to prevent the local intramolecular interactions between MBP and target proteins during the folding process. The GFP fusion proteins were expressed at 37°C. The insertion of Trx or LysN between MBP and GFP had little or no inhibitory effect on the solubilizing ability of MBP for GFP (Fig. 6A). Similar patterns were also observed when GCSF was used as target protein (data not shown). As shown in Figure 6B, in the case of TEVP, the insertion of Trx slightly decreased the solubility of fusion proteins, whereas the insertion of LysN increased the solubility of fusion proteins (2%, 34%, 58%, 54%, and 78% for Trx, LysN, MBP, MBP-Trx, and MBP-LysN, respectively). The results indicate that the solubilizing ability of MBP

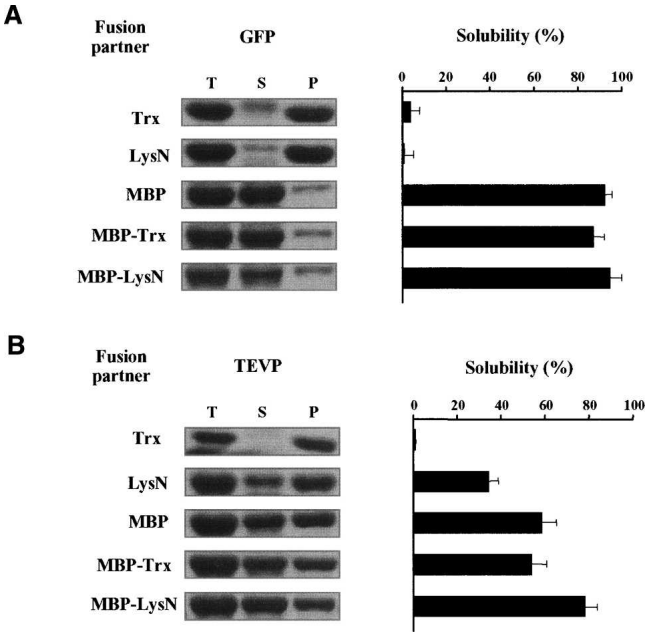

Figure 6. The effect of large inserted tags between MBP and target proteins on the solubilizing ability of MBP. LysN or Trx were inserted between MBP and GFP (*A*), between MBP and TEVP (*B*). The GFP fusion proteins and TEVP fusion proteins were expressed at 37°C and 34°C, respectively, and their solubility was analyzed by SDS-PAGE. The TEVP in B have no C-terminal histidine tag.

could be transferred to the target proteins via covalent linkage without intramolecular interactions. In accord with this reasoning, the mutagenesis of exposed hydrophobic sites of MBP, initially thought to be important for the solubilizing ability, showed no apparent effect on the solubility of MBP fusion proteins (Fox et al. 2001).

Discussion

The results suggest a strong correlation between the popular fusion technology and de novo folding of native multidomain proteins in vivo, using an indirect approach in Figure 1. In addition to the numerous whole proteins, the N- and C-terminal domains of *E. coli* proteins, including LysN as well as LysRS used here, have been known to promote the solubility and proper folding of their linked heterologous proteins (Anderluh et al. 2003; Sørensen et al. 2003; S.I. Choi, K.S. Han, C.W. Kim, B.H. Kim, K.H. Kim, S.I. Kim, K.S. Ryu, T.H. Kang, H.C. Shin, and B.L. Seong, unpubl.). However, there has been no report of linking this phenomenon to de novo folding of the native multidomain proteins. The indirect approach would be useful for assessing the intrinsic solubilizing ability of the folded N-terminal domains in vivo, although it does not provide any information for the effect of the native interdomain interactions on the folding of the interacting domains.

Proper folding of N-terminal domains and high solubility appear to be a prerequisite for their solubilizing functions. Cotranslational folding likely allows the solubilizing ability of folded domains to be more effective. Cotranslational folding of *E. coli* proteins and some eukaryotic proteins on *E. coli* ribosomes has been demonstrated (Kramer et al. 2001; Svetlov et al. 2006). In general, fusion partners, to be effective, have to be located upstream of target proteins; reversion of fusion order abolishes the solubilizing ability of MBP (Sachdev and Chirgwin 1998). Likewise, the fusion of aggregation-prone target protein upstream to the soluble reporter protein led to the insoluble aggregates of fusion proteins (Waldo 2003). These findings indicate that the proper folding of N-terminal domains is important for the efficient folding of multidomain proteins in vivo, consistent with Frydman et al. (1999).

The addition of negatively charged small peptides to the N or C terminus of aggregation-prone proteins can increase the protein solubility (Chen et al. 1998; Zhang et al. 2004). In particular, the electrostatic repulsions between protein molecules were suggested to prevent protein aggregation (Chiti et al. 2002). Similarly, the intermolecular electrostatic repulsions generated by the surface-exposed charged residues of folded domains might be responsible for their solubilizing ability. In accord with this assumption, a positive correlation between the solubility enhancement and the number of aspartic acids in the linker peptides was observed (Fig. 4). The theoretical isoelectric points of LysN, ThrN, and AcnN are 6.11, 5.58, and 4.81, respectively, indicating that they are anionic polypeptides. Moreover, the electrostatic potential surfaces of the N-terminal domains on the basis of the known three-dimensional structures show the exposed anionic surfaces (Fig. 7).

Besides the charge effect, there might be the steric hindrance of folded domains against intermolecular aggregation. Protein aggregation is a three-dimensional growing and specific process (London et al. 1974; Speed et al. 1996; Wright et al. 2005). Consequently, the structurally bulky folded domains are expected to greatly disturb the intermolecular aggregation driven by their linked aggregation-prone domains. Here, substantial fraction of the surfaces of C-terminal aggregation-prone domains around the linker region is expected to be inaccessible to the other folding intermediates due to the steric masking of folded domains. Importantly, the steric and electrostatic repulsions are likely proportional to the number of folded domains. The correlation of protein size with the solubilizing ability in Figure 5 supports this idea. In addition, both factors of folded domains seem to solubilize the downstream domains without the transient, intramolecular interactions between them. Consistently, little or no inhibitory effect of

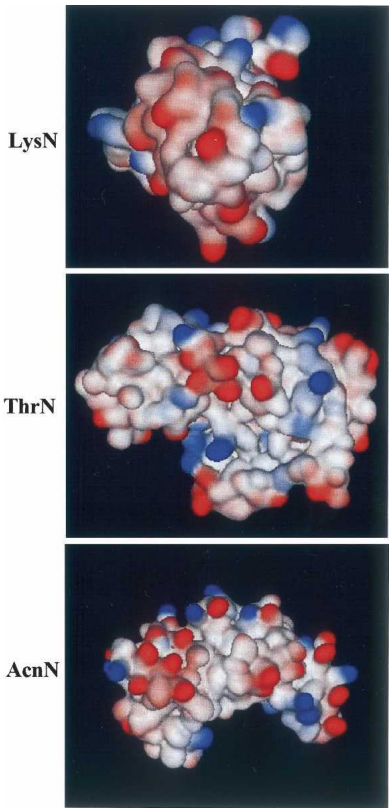

Figure 7. Electrostatic potential surfaces of the N-terminal domains. Red and blue represent negatively and positively charged surfaces, respectively. The electrostatic surfaces of LysN, ThrN, and AcnN were obtained from their known three-dimensional structures using the DS Visualizer (Accelrys).

inserted LysN or Trx between MBP and target proteins on the solubilizing ability of MBP was observed in Figure 6.

By combining the size effect and steric hindrance with the well-known charge effect on protein solubility, we propose a solubilizing mechanism as illustrated in Figure 8. Both electrostatic repulsions and steric hindrance of folded and soluble domains prevent the intermolecular associations driven by aggregation-prone domains. This would lead to the shift from oligomeric state to monomeric state in which the spontaneous folding of downstream domains is favored. This model is dependent on the reversibility of protein aggregation at the early phase. There is accumulated evidence supporting that protein aggregation is reversible (Silow et al. 1999; Ganesh et al. 2001; Carrió and Villaverde 2001). Additional factors, e.g., in vivo folding rates and stability of N-terminal domains in the whole proteins, transient interdomain interactions, and the presence of native interdomain interactions, which are not considered in this model, could greatly affect the solubilizing ability of N-terminal domains.

Nucleic acids and other polyanions can enhance in vitro refolding of proteins (Dabora et al. 1991; Rentzeperis et al. 1999). The negatively charged clusters of the cavity wall of GroEL are important for accelerating the folding of some proteins in vitro (Tang et al. 2006). Like the polyanions, the anionic surfaces of the folded N-terminal domains and anion linkers in Figure 4, especially in the close vicinity of their downstream domains, could enhance the folding of downstream domains. The N-terminal propeptides, usually polyanionic in nature (Jones et al. 2004), have been shown to directly assist the correct folding of their downstream proteins as intramolecular chaperones (Shinde and Inouye 2000). The intramolecular chaperone activity of propeptides would be mechanistically similar to the solubilizing ability of the N-terminal domains of multidomain proteins. Besides the cleavage of propeptides, however, there are some obvious differences between them. The N-terminal domains appear to assist the folding of their downstream heterologous proteins in a passive manner by simply solubilizing them.

Here we suggest that the chaperone-like type in the fusion technology could occur in de novo folding of native multidomain proteins in vivo. Further studies are required to understand the role of folded N-terminal domains in de novo folding. The potential role of N-terminal domains as solubility enhancers would contribute to the autonomous folding of native proteins in vivo.

Materials and Methods

Construction of expression vectors

The plasmid, derived from pGEMEX-1 (Promega), was used for the construction of protein expression vectors. The genes encoding proteins of interest are under the control of a T7 promoter. All *E. coli* genes were obtained from *E. coli* genomic

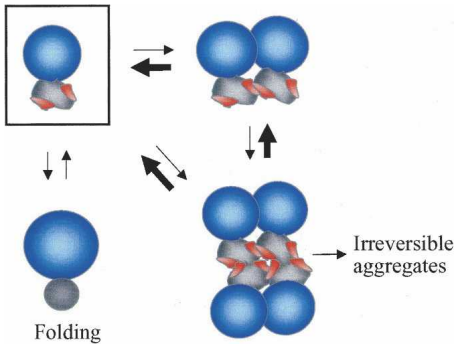

Figure 8. A model for how N-terminal domains solubilize their linked domains. The blue, gray, and wrinkled spheres represent the folded N- and C-terminal domains and incompletely folded C-terminal domains, respectively. The red spots on wrinkled spheres indicate the exposed regions involved in the intermolecular interactions. Thick arrows represent the shift from the oligomeric state to the monomeric state (boxed) of proteins driven by the electrostatic repulsions and steric hindrance of folded N-terminal domains.

DNA using PCR amplification. The gene encoding mature GCSF was used as described (Lee et al. 1999). TEVP is the catalytic domain of the NIa protease of tobacco etch virus (Kapust et al. 2001). For immunodetection and purification, the hexahistidine tag was inserted at the C terminus of TEVP. The linker sequence, GTGSSTSTST, was used between fusion partners (Trx, LysN, $LysN_2$, LysRS, LysC, ThrN, and MBP) and target proteins (GCSF, GFP, and TEVP) except GSSTSTST for AcnN as fusion partner. The GTGSSTSTST linker sequence was also used between fusion partners for the construction of MBP-LysN and MBP-Trx. One histidine residue was incorporated between LysN and LysN due to the restriction site of NdeI to make tandem repeat, $LysN_2$. MBP-GCSF has the hexahistidine tag at its C terminus and TEVP recognition site between MBP and GCSF.

Protein expression, solubility test, and Western blot analysis

The *E. coli* strain HMS174 (DE3) plysE (Novagen) was used as expression host. Each transformant was inoculated into 2 mL of LB containing 50 μg/mL ampicillin and 30 μg/mL chloramphenicol and then cultured overnight at 37°C. One milliliter of culture broth was diluted into 20 mL with fresh LB with the antibiotics. At the cell density of 0.5–0.8 at A_{600}, proteins were expressed for 5 h after addition of 1 mM IPTG at indicated temperature. The harvested cells from 10 mL of culture broth were suspended in 0.3 mL of PBS. After sonication and centrifugation, total lysates, soluble fraction, and insoluble fraction were obtained. The samples were analyzed by SDS-PAGE.

The solubility of proteins was estimated on SDS-PAGE with a densitometer using the following formula. Solubility (%) $= (S - S_E)/(T - T_E) \times 100$, where S and T are the band intensity of target proteins in soluble fraction and total lysates, and S_E and T_E are the band intensity of overlapped endogenous proteins that can be obtained from the parallel lines on the same SDS-PAGE. Three independent experiments were performed for the measurement of protein solubility.

The amount of proteins equivalent to one-sixteenth of SDS-PAGE samples in Figure 2, B and C, was transferred to PVDF membranes. The membranes were blocked with 10% skim milk in PBST (1× PBS supplemented with 0.1% Tween 20) for 1 h before the incubation of the primary antibody (anti-GCSF antibody [Sigma], anti-GFP antibody [Clontech], and Penta-His antibody [Qiagen], respectively), diluted 1:2000 in PBST, for 1 h. After washing three times for 15 min with PBST, the secondary antibody (anti-mouse IgG-horseradish peroxidase [Sigma] or anti-rabbit IgG-horseradish peroxidase [Sigma] for the detection of GFP and GFP fusion proteins), diluted 1:20,000, was treated for 1 h. After washing, blots were developed with ECL developing reagent (Intron biotechnology).

Protein purification and activity assays

Each prepared supernatant in buffer A (20 mM Tris-HCl at pH 7.5, 300 mM NaCl, 10% glycerol, 2 mM beta-mercaptoethanol, and 5 mM imidazole) was applied to a nickel-chelated prepacked HP column (Amersham Bioscience) pre-equilibrated with buffer A. After washing with buffer A, elution was performed with a linear gradient of imidazole by mixing buffer A and buffer B (buffer A supplemented with 300 mM imidazole). The eluted fractions were analyzed by SDS-PAGE and then pooled and dialyzed against buffer C (100 mM Tris-HCl at pH 8.0, 100 mM NaCl, 2 mM EDTA, and 2 mM DTT). The protein concentration was determined according to the BCA method (Pierce) using bovine serum albumin as standard. The dialyzed samples were mixed with the equal volume of 100% glycerol and then stored at −20°C.

TEVP cleavage reactions were carried out in 20 μL volumes containing 50 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, 1 mM DTT, 20 μg of MBP-GCSF as substrate, and 3 U of recombinant TEVP (Invitrogen) or 2 μM of each TEVP fusion protein for 2 h at 30°C. The reaction mixtures were analyzed by SDS-PAGE.

For the measurements of fluorescence intensity of the GFP and GFP fusion proteins in the soluble fractions of cell lysates, each fraction was diluted into 20-fold using 400 μL of PBS. The fluorescence emission was monitored at 509 nm with the excitation at 395 nm using Cary Eclipse fluorescence spectrophotometer (Varian). Each fluorescence intensity was divided by BI/MW where BI and MW represent the band intensity ($S - S_E$) on SDS-PAGE measured by densitometer and molecular weights of tested proteins, respectively. The resulting value of GFP was set to one and compared with those of GFP fusion proteins.

Acknowledgments

This work was supported in part by the Human Genomics Frontier Research and the Microbial Genomics R&D Grants from the Ministry of Science and Technology of the Korean Government, and the Korea Research Foundation Grant funded by the Korean Government (grant no. KRF-2004-005-C00148).

References

Anderluh, G., Gökçe, I., and Lakey, J.H. 2003. Expression of proteins using the third domain of the *Escherichia coli* periplasmic-protein TolA as a fusion partner. *Protein Expr. Purif.* **28:** 173–181.

Bach, H., Mazor, Y., Shaky, S., Shoham-Lev, A., Berdichevsky, Y., Gutnick, D.L., and Benhar, I. 2001. *Escherichia coli* maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *J. Mol. Biol.* **312:** 79–93.

Braun, P. and LaBaer, J. 2003. High throughput protein production for functional proteomics. *Trends Biotechnol.* **21:** 383–388.

Carrió, M.M. and Villaverde, A. 2001. Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Lett.* **489:** 29–33.

Chen, J., Skehel, J.J., and Wiley, D.C. 1998. A polar octapeptide fused to the N-terminal fusion peptide solubilizes the influenza virus HA_2 subunit ectodomain. *Biochemistry* **37:** 13643–13649.

Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G., and Dobson, C.M. 2002. Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc. Natl. Acad. Sci.* **99:** 16419–16426.

Creighton, T.E. ed. 1992. *Protein folding* pp. 405–454. Freeman, New York.

Dabora, J.M., Sanyal, G., and Middaugh, C.R. 1991. Effect of polyanions on the refolding of human acidic fibroblast growth factor. *J. Biol. Chem.* **266:** 23637–23640.

Davis, G.D., Elisee, C., Newham, D.M., and Harrison, R.G. 1999. New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol. Bioeng.* **65:** 382–388.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. 1999. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **400:** 693–696.

Esposito, D. and Chatterjee, D.K. 2006. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17:** 353–358.

Fayet, O., Ziegelhoffer, T., and Georgopoulos, C. 1989. The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J. Bacteriol.* **171:** 1379–1385.

Fox, J.D., Kapust, R.B., and Waugh, D.S. 2001. Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. *Protein Sci.* **10:** 622–630.

Frydman, J., Erdjument-Bromage, H., Tempst, P., and Hartl, F.U. 1999. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat. Struct. Biol.* **6:** 697–705.

Ganesh, C., Zaidi, F.N., Udgaonkar, J.B., and Varadarajan, R. 2001. Reversible formation of on-pathway macroscopic aggregates during the folding of maltose binding protein. *Protein Sci.* **10:** 1635–1644.

Hartl, F.U. and Hayer-Hartl, M. 2002. Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science* **295:** 1852–1858.

Inaba, K., Kobayashi, N., and Fersht, A.R. 2000. Conversion of two-state to multi-state folding kinetics on fusion of two protein foldons. *J. Mol. Biol.* **302:** 219–233.

Jones, L.S., Yazzie, B., and Middaugh, C.R. 2004. Polyanions and the proteome. *Mol. Cell. Proteomics* **3:** 746–769.

Kapust, R.B. and Waugh, D.S. 1999. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8:** 1668–1674.

Kapust, R.B., Tözsér, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., and Waugh, D.S. 2001. Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14:** 993–1000.

Kerner, M.J., Naylor, D.J., Ishihama, Y., Maier, T., Chang, H.C., Stines, A.P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., et al. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122:** 209–220.

Kramer, G., Ramachandiran, V., and Hardesty, B. 2001. Cotranslational folding–omnia mea mecum porto? *Int. J. Biochem. Cell Biol.* **33:** 541–553.

LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., and McCoy, J.M. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology (N. Y.)* **11:** 187–193.

Lee, J., Choi, S.I., Jang, J.S., Jang, K., Moon, J.W., Bae, C.S., Yang, D.S., and Seong, B.L. 1999. Novel secretion system of recombinant *Saccharomyces cerevisiae* using an N-terminus residue of human IL-1β as secretion enhancer. *Biotechnol. Prog.* **15:** 884–890.

London, J., Skrzynia, C., and Goldberg, M.E. 1974. Renaturation of *Escherichia coli* tryptophanase after exposure to 8 M urea. Evidence for the existence of nucleation centers. *Eur. J. Biochem.* **47:** 409–415.

Netzer, W.J. and Hartl, F.U. 1997. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388:** 343–349.

Onesti, S., Desogus, G., Brevet, A., Chen, J., Plateau, P., Blanquet, S., and Brick, P. 2000. Structural studies of lysyl-tRNA synthetase: Conformational changes induced by substrate binding. *Biochemistry* **39:** 12853–12861.

Otzen, D.E., Kristensen, O., and Oliveberg, M. 2000. Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: A structural clue to amyloid assembly. *Proc. Natl. Acad. Sci.* **97:** 9907–9912.

Rentzeperis, D., Jonsson, T., and Sauer, R.T. 1999. Acceleration of the refolding of Arcrepressor by nucleic acids and other polyanions. *Nat. Struct. Biol.* **6:** 569–573.

Sachdev, D. and Chirgwin, J.M. 1998. Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **244:** 933–937.

Sankaranarayanan, R., Dock-Bregeon, A.C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B., and Moras, D. 1999. The structure of threonyl-tRNA synthetase-tRNAThr complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* **97:** 371–381.

Shinde, U. and Inouye, M. 2000. Intramolecular chaperones: polypeptide extensions that modulate protein folding. *Semin. Cell Dev. Biol.* **11:** 35–44.

Silow, M., Tan, Y.J., Fersht, A.R., and Oliveberg, M. 1999. Formation of short-lived protein aggregates directly from the coil in two-state folding. *Biochemistry* **38:** 13006–13012.

Sørensen, H.P., Sperling-Petersen, H.U., and Mortensen, K.K. 2003. A favorable solubility partner for the recombinant expression of streptavidin. *Protein Expr. Purif.* **32:** 252–259.

Speed, M.A., Wang, D.I., and King, J. 1996. Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition. *Nat. Biotechnol.* **14:** 1283–1287.

Svetlov, M.S., Kommer, A., Kolb, V.A., and Spirin, A.S. 2006. Effective cotranslational folding of firefly luciferase without chaperones of the Hsp70 family. *Protein Sci.* **15:** 242–247.

Tang, Y.C., Chang, H.C., Roeben, A., Wischnewski, D., Kerner, M.J., Hartl, F.U., and Hayer-Hartl, M. 2006. Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125:** 903–914.

Ullers, R.S., Luirink, J., Harms, N., Schwager, F., Georgopoulos, C., and Genevaux, P. 2004. SecB is a bona fide generalized chaperone in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **101:** 7583–7588.

Uversky, V.N., Gillespie, J.R., and Fink, A.L. 2000. Why are ''natively unfolded'' proteins unstructured under physiologic conditions? *Proteins* **41:** 415–427.

Vorderwülbecke, S., Kramer, G., Merz, F., Kurz, T.A., Rauch, T., Zachmann-Brand, B., Bukau, B., and Deuerling, E. 2004. Low temperature of GroEL/ES overproduction permits growth of *Escherichia coli* cells lacking trigger factor and DnaK. *FEBS Lett.* **559:** 181–187.

Waldo, G.S. 2003. Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.* **7:** 33–38.

Wall, J.G. and Plückthun, A. 1995. Effects of overexpressing folding modulators on the in vivo folding of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6:** 507–516.

Williams, C.H., Stillman, T.J., Barynin, V.V., Sedelnikova, S.E., Tang, Y., Green, J., Guest, J.R., and Artymiuk, P.J. 2002. *E. coli* aconitase B structure reveals a HEAT-like domain with implications for protein-protein recognition. *Nat. Struct. Biol.* **9:** 447–452.

Wong, P. and Houry, W.A. 2004. Chaperone networks in bacteria: Analysis of protein homeostasis in minimal cells. *J. Struct. Biol.* **146:** 79–89.

Wright, C.F., Teichmann, S.A., Clarke, J., and Dobson, C.M. 2005. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* **438:** 878–881.

Zhang, Y.B., Howitt, J., McCorkle, S., Lawrence, P., Springer, K., and Freimuth, P. 2004. Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expr. Purif.* **36:** 207–216.