Foldon-guided self-assembly of ultra-stable protein fibers

ANSHUL BHARDWAJ, NANCY WALKER-KOPP, STEPHAN WILKENS,
AND GINO CINGOLANI

Department of Biochemistry and Molecular Biology, SUNY Upstate Medical University, Syracuse,
New York 13210, USA

(RECEIVED April 29, 2008; FINAL REVISION June 4, 2008; ACCEPTED June 4, 2008)

Abstract

A common objective in protein engineering is the enhancement of the thermodynamic properties of recombinant proteins for possible applications in nanobiotechnology. The performance of proteins can be improved by the rational design of chimeras that contain structural elements with the desired properties, thus resulting in a more effective exploitation of protein folds designed by nature. In this paper, we report the design and characterization of an ultra-stable self-refolding protein fiber, which rapidly reassembles in solution after denaturation induced by harsh chemical treatment or high temperature. This engineered protein fiber was constructed on the molecular framework of bacteriophage P22 tail needle gp26, by fusing its helical core to the foldon domain of phage T4 fibrin. Using protein engineering, we rationally permuted the foldon upstream and downstream from the gp26 helical core and characterized gp26-foldon chimeras by biophysical analysis. Our data demonstrate that one specific protein chimera containing the foldon immediately downstream from the gp26 helical core, gp26(1-140)-F, displays the highest thermodynamic and structural stability and refolds spontaneously in solution following denaturation. The gp26-foldon chimeric fiber remains stable in 6.0 M guanidine hydrochloride, or at 80°C, rapidly refolds after denaturation, and has both N and C termini accessible for chemical/biological modification, thereby representing an ideal platform for the design of self-assembling nanoblocks.

Keywords: bacteriophage P22; tail needle gp26; phage T4 foldon; spontaneous refolding; nanoscale building block for the design of protein nanodevices

Bacteriophages represent one of the simplest forms of life, which evolutionary pressure and adaptation have optimized to carry out just a few, yet greatly efficient, biological functions. Through this adaptation and optimization, the phages' structural proteins often provide unique structural and thermodynamic properties, which are of great interest in nanobiotechnology. In this study, we

have used rational design to generate a chimeric protein fiber consisting of the helical core of phage P22 tail needle gp26 (Bhardwaj et al. 2007; Olia et al. 2007) and the foldon domain of phage T4 fibrin (Tao et al. 1997). The tail needle gp26 is a 240 Å-long fiber formed by three identical polypeptide chains of 233 amino acids (Andrews et al. 2005; Olia et al. 2007). The trimer is stable even in the presence of >10% SDS and unfolds irreversibly in ~6.4 M guanidine hydrochloride (GdnHCl) or at 85°C (Bhardwaj et al. 2007). The domain structure of gp26 is well known from crystallography data (Olia et al. 2007) as well as mutational and truncation analyses (Bhardwaj et al. 2007). The needle consists of four structural subdomains (Fig. 1A), of which domain II (residues

Reprint requests to: Gino Cingolani, Department of Biochemistry and Molecular Biology, SUNY Upstate Medical University, 750 East Adams Street, Syracuse, NY 13210, USA; e-mail: cingolag@upstate.edu; fax: (315) 464-8750.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.036111.108>.

Figure 1. Design principles of gp26-foldon chimeric fibers. (A) Ribbon diagram of tail needle gp26 (RCSB Protein Data Bank ID: 2poh) (Olia et al. 2007). The seven constructs used in this study are enumerated 1–7. Constructs 1 and 2 include full-length gp26 and its helical core gp26(1-140), respectively. Five octad repeats (OR) mediating trimerization of gp26 are marked as OR1–OR5 (Bhardwaj et al. 2007). Constructs 3–5 were designed to have gp26 helical core fused to a C-terminal foldon domain (F). Constructs 6 and 7 have the foldon inserted at both N and C termini of the helical core, with the N-terminal foldon being either in a direct (6) or reverse (7) orientation. (B) SDS-PAGE analysis of purified gp26 fibers used in this study. All gp26 constructs were prepared in Laemmli sample buffer (Laemmli 1970) containing 0.1% SDS. Prior to loading, each sample was either boiled for 5 min at 95°C or incubated 5 min at 22°C (RT).

27–140) forms a triple-stranded coiled-coil helical core ~165 Å in length. Domain I (residues 1–27) folds like a hairpin on the helical core, where it adopts an extended conformation. Downstream from the helical core, the C terminus of the tail needle is built by domains III and IV, which fold into a triple-stranded β -helix, and an inverted

helical coiled coil named the *lazo*-domain, thought to be involved in lipid puncturing (Bhardwaj et al. 2007; Olia et al. 2007). Whereas deletion of domain I does not reduce the overall stability of the needle (Bhardwaj et al. 2007), domains III and IV (93 residues) are important for the stability of the gp26 helical core. Removal of domains III

and IV reduces the thermal and denaturant stability of gp26 by 10°C and ~2 M GdnHCl, respectively (Bhardwaj et al. 2007). Interestingly, isolated domains III–IV are not trimeric when expressed separately, suggesting that the stabilization of domain II is not dependent on the knotting of the helices, as previously reported for other trimeric coiled-coil fibrous proteins such as fibritin (Letarov et al. 1999) or collagen (Frank et al. 2003).

The foldon domain is a small 27-residue β -propeller-like trimeric structure found at the C terminus of bacteriophage T4 fibritin (Tao et al. 1997). Fibritin is a triple-stranded coiled-coil fibrous protein ~530 Å in length that forms the “whiskers” of bacteriophage T4. It is thought to function as a rudimentary environment-sensing device (Conley and Wood 1975). In the context of fibritin, the foldon serves as a registration motif that is necessary and sufficient, both in vitro and in vivo, to promote the correct folding of fibritin trimeric coiled coil (Letarov et al. 1999; Boudko et al. 2002). In vitro, foldon is a fast folder (Guthe et al. 2004), which has been evolutionarily optimized for rapid and specific initiation of trimer formation during fibritin assembly. The foldon has been shown to promote trimerization for a number of other engineered systems such as short collagen fibers (Frank et al. 2001; Stetefeld et al. 2003), HIV1 envelope glycoprotein (Yang et al. 2002), adenovirus fiber shaft (Papanikolopoulou et al. 2004a,b), and rabies virus glycoprotein (Sissoeff et al. 2005).

In this report, we have used the foldon domain to enhance the structural and thermodynamic stability of tail needle gp26 helical core fibers. The foldon domain was rationally permuted at the N and C termini of the gp26 helical core, in single or double copy. Biophysical analysis of gp26-foldon permutants indicates that the thermodynamic stabilization of the gp26 helical core by the foldon is position dependent but dose independent, as addition of multiple copies of the foldon fails to further enhance the fiber stability. The most stable gp26-foldon chimera identified in this study has the foldon immediately downstream from the gp26 helical core. This chimera represents an ideal nanoscale building block for the design of future protein nanodevices.

Results

Design principles of our study

The goal of this study was to generate a protein fiber of enhanced chemical and thermodynamic stability that could be used as a potential nanoscale building block in bionanotechnology. The desired protein fiber was designed to have four fundamental chemical and structural properties: (1) high structural stability, (2) low susceptibility to degradation, (3) reversible unfolding followed by harsh temper-

ature and chemical treatment, and (4) exposed N and C termini for engineering rational self-assembly. To achieve this goal we used the molecular framework of tail needle gp26, for which both the structure (Olia et al. 2007) and folding/unfolding pathways (Bhardwaj et al. 2007) have been characterized. Gp26 (construct 1 in Fig. 1A) is highly stable ($T_m \sim 85^\circ\text{C}$), greatly overexpressed in bacteria, and easy to purify, but its immediate application as a biomaterial is hindered by at least three factors. First, gp26 unfolds irreversibly when treated with denaturants or temperature $>85^\circ\text{C}$ (Bhardwaj et al. 2007); second, the C terminus of gp26 (domains III–IV) is flexibly connected to the helical core (domain II) and is susceptible to proteolysis; third, the C terminus of wild-type gp26 is not exposed, due to the peculiar folding of the *lazo*-domain (domain IV) (Olia et al. 2007). Likewise, the minimal helical core of gp26, gp26(1–140) (construct 2 in Fig. 1A), has exposed C termini but is significantly less stable than the full-length fiber ($T_m 75^\circ\text{C}$ versus 85°C) and also unfolds irreversibly following heat and chemical-induced denaturation (Bhardwaj et al. 2007).

To enhance the thermodynamic stability of the gp26 helical fiber and improve its versatility as a biomaterial, we fused the gp26 helical core to the foldon domain of phage T4 fibritin. The foldon domain, a 27-amino-acid-long trimer, was permuted upstream and downstream from the gp26 helical core using a two-step PCR approach, without incorporation of any linker sequences to avoid unintended flexibility in the chimeric fiber structure. Using this approach, five novel constructs were designed, as illustrated in Figure 1A. Constructs 3 and 4 had a single copy of the foldon domain fused downstream from the gp26 helical core. In construct 3, gp26(1–140)-F, the foldon was introduced immediately downstream from the gp26 helical core; in construct 4, gp26(1–170)-F, the foldon was fused downstream from the short β -helix domain III, which resembles the foldon domain itself (Olia et al. 2007). To determine if the foldon can stabilize gp26 fibers in a dose-dependent manner, we also designed constructs 5, 6, and 7 (Fig. 1A), which have two copies of the foldon fused to the gp26 helical core. In construct 5, gp26(1–140)-F-F, two copies of the foldon domain were added downstream from the helical core. In constructs 6 and 7, the foldon was introduced at the N terminus of the pre-built gp26(1–140)-F chimera, both in direct (F_D -gp26[27–140]-F) and reverse (F_{Rev} -gp26[27–140]-F) orientation. All gp26-foldon chimeric fibers, wild-type gp26, and the helical core gp26(1–140) were expressed in *Escherichia coli* and purified under identical expression conditions, as described in Materials and Methods. Interestingly, the two constructs bearing an N-terminal foldon were expressed with significantly lower yield and resulted in a mixture of degraded protein. Due to lack of homogeneity, both constructs were left out from further studies. In contrast, the C-terminal foldon chimeras displayed high solubility and stability and were purified to homogeneity.

Similar to wild-type gp26 and gp26(1-140), all gp26-foldon chimera migrated in their trimeric native form on a SDS-PAGE at room temperature (Fig. 1B), which suggests that the gp26 trimeric helical interface is not weakened by the addition of the foldon.

Structural stability of the gp26-foldon chimera

The thermal and chemical denaturant stability of wild-type gp26 and gp26(1-140) were previously characterized in great detail (Bhardwaj et al. 2007). Both proteins denature irreversibly with an apparent midpoint of thermal melting (T_m) and guanidine half-concentration (C_m) of 85°C/6.4 M and 75°C/4.7 M, respectively. In this study, we used far-ultraviolet circular dichroism (CD) spectroscopy to assess the structural stability of gp26-foldon chimeric fibers. GdnHCl- and heat-induced denaturation curves were measured for all foldon fusion constructs (Fig. 2A,B). Notably, insertion of the foldon at residue 170 (construct gp26[1-170]-F) resulted in a marginal increase in the stability of the gp26 helical core, with an apparent T_m and C_m of 80°C and 5.1 M, respectively. A more dramatic stabilization of the gp26 helical core was observed in the construct gp26(1-140)-F, where the foldon is inserted at position 141, immediately downstream from the gp26 coiled coil. This fiber melted reversibly with a T_m and C_m of 85°C and 6.4 M, respectively, which matches the stability profile observed for wild-type gp26 (Bhardwaj et al. 2007). Thus, the 27-residue foldon can rescue the stabilizing effect exerted by domains III–IV of gp26 (93 amino acids) in a position-dependent manner. Interestingly, however, the addition of a second foldon (construct gp26[1-140]-F-F) failed to further improve the stability of the gp26 helical core. As shown in Figure 2, A and B, GdnHCl- and heat-induced denaturation curves measured in the presence of two foldons (construct gp26[1-140]-F-F) were identical to those seen for the construct gp26(1-140)-F, with a T_m and C_m of 85°C and 6.4 M, respectively. Thus, the foldon-induced stabilization of the gp26 helical core is dose independent.

Synergistic stabilization of the foldon and gp26 helical core

After determining that the foldon domain stabilizes gp26(1-140) by ~10°C, we sought to determine if the helical core of gp26 can stabilize the foldon domain, and thus whether the two trimeric structural domains synergistically stabilize each other. To test this hypothesis, we took advantage of the tryptophan at position 20 of the foldon (Trp²⁰), which is the only Trp both in the free foldon and the gp26(1-140)-F chimera. Trp²⁰ falls in the first β -sheet of the foldon that is fully buried at the

trimeric interface (Tao et al. 1997). Consistently, the fluorescence emission spectrum of Trp²⁰ has a maximum at 328 nm, as expected from a tryptophan residue that is largely buried. This looks identical in the gp26(1-140)-F chimeric fiber as compared with the free foldon (data not

Figure 2. (Legend on next page)

shown). By monitoring the variations in intrinsic fluorescence emission of Trp²⁰ that occur during thermal unfolding, we were able to measure the structural stability of the foldon domain fused to the gp26 helical core as compared with the isolated foldon (Fig. 2C). For both the free foldon and gp26(1-140)-F, a stepwise increase in temperature resulted in a significant redshift in the fluorescence emission maximum wavelength toward 356 nm, suggesting the local environment of Trp²⁰ becomes solvent exposed. In both cases, heat-induced unfolding transitions resulted in a cooperative two-state transition, which was fit to the all-or-none unfolding model of a native trimer (N) to unfolded monomer (U). The global fit analysis gave a free energy of unfolding for free foldon of $\Delta G_u = 20.8 (\pm 0.5)$ kcal/mol and a T_m of $\sim 74^\circ\text{C}$ at a concentration of 30 μM , which is consistent with previous measurements (Frank et al. 2001). For the gp26(1-140)-F fiber, the free energy of unfolding is $\Delta G_u = 17.1 (\pm 0.7)$ kcal/mol and the T_m is $\sim 89^\circ\text{C}$, which is close to that measured by CD (Fig. 2B). In addition, while foldon trimerization ΔG_u and T_m depend strongly on the concentration of monomers (Frank et al. 2001), both association and dissociation of the gp26(1-140)-F fiber were virtually independent of the monomer concentration in a range of protein concentrations between high micromolar to low picomolar (data not shown). Taken together, these data strongly support the idea that in the context of the gp26(1-140)-F fiber, the stabilization of the foldon and gp26 helical core is mutual and synergistic.

Foldon-induced reversibility of assembly

Wild-type gp26 and gp26(1-140) have a strong tendency to aggregate upon thermal unfolding, resulting in irreversible precipitation (Bhardwaj et al. 2007). In contrast, the foldon domain has been reported to unfold in a completely reversible manner (Guthe et al. 2004). To deter-

mine if the chimeric gp26-foldon fibers designed in this study behaved like the gp26 template or followed the refolding pattern reported for the isolated foldon, we carried out refolding experiments in which gp26-foldon fibers were first thermally unfolded for 5 min at 95°C and then quickly cooled to 25°C . The degree of fiber reassembly, and thus reversible unfolding, was determined by monitoring the variations in the ellipticity at 220 nm during the thermal transition. Under these experimental conditions, refolding of thermally unfolded full-length gp26 resulted in less than $\sim 30\%$ of the initial helical signal (Fig. 3A), suggesting only partial refolding of the trimer. Likewise, cooling thermally unfolded gp26(1-140) resulted in a completely irreversible unfolding transition (Fig. 3B), which led to significant precipitation in the cuvette. Notably, refolding scans of gp26(1-140)-F and gp26(1-170)-F chimeras revealed surprising differences. While thermal refolding of denatured gp26(1-170)-F fibers displayed only a small gain in ellipticity (Fig. 3D), comparable to that seen for wild-type gp26, refolding of gp26(1-140)-F restored up to 95% of the initial helical signal (Fig. 3C). This suggests that in the context of the foldon-gp26 chimera, the β -sheet region of gp26 between residues 140 and 170 hinders foldon-induced reassembly. Even more surprisingly, the construct gp26(1-140)-F-F, which presented an identical stability profile and melting temperature as gp26(1-140)-F (Fig. 2), displayed limited thermal reversibility (Fig. 3E) as compared with gp26(1-140)-F. This may be explained by the nonspecific nucleation of two foldon domains that gives rise to aberrant structures upon refolding.

Refolded gp26(1-140)-F fibers adopt a native conformation

The gain in ellipticity at 220 nm upon thermal refolding is a good indicator of the formation of α -helices in the gp26 helical core. However, such a spectroscopic measurement cannot directly determine if the refolded gp26(1-140)-F fiber adopts a conformation of identical stability, oligomeric state, and structure as in the native state. To address this important point, we used three independent assays. First, we measured the thermal and chemical denaturant stability of gp26(1-140)-F fibers that had been denatured for 5 min at 95°C and then quickly cooled to 25°C to promote refolding. This yielded a T_m and C_m of 85°C and 6.4 M GdnHCl, respectively (data not shown), which matches exactly the stability profile of the native gp26(1-140)-F chimera (Fig. 2A). Thus, refolded gp26(1-140)-F fibers have identical thermodynamic stability as native fibers.

Second, to determine the oligomeric state, we analyzed the electrophoretic migration on SDS-PAGE of refolded gp26-F fibers, before and after boiling. As shown in Figure 1B, all gp26-foldon chimeric fibers run as trimers

Figure 2. Structural stability of gp26 helical fibers. Stability of gp26 fibers against GdnHCl (A) and thermal (B) denaturation. Unfolding curves were recorded by measuring changes in the mean ellipticity at 220 nm both at increasing GdnHCl concentrations and as a function of temperature measured at 1°C increments at constant intervals of 120 s. The protein concentration used in these experiments was 6 μM . The apparent T_m and half-denaturant concentration (C_m) values for gp26(1-140) and gp26(1-170)-F are $74^\circ\text{C}/4.7$ M and $80^\circ\text{C}/5.1$ M, respectively. Wild-type gp26, gp26(1-140)-F, and gp26(1-140)-F-F present an identical T_m/C_m value of $\sim 85^\circ\text{C}/6.4$ M. (C) Mutual stabilization of the gp26 helical core and the foldon. Unfolding curves for the free foldon (circle) and gp26(1-140)-F (triangle) were recorded by measuring the fluorescence emission maxima of Trp²⁰ at excitation wavelength of 295 nm as a function of increasing temperature. Both unfolding curves are fully reversible and gave a T_m $\sim 274^\circ\text{C}$ and $\sim 89^\circ\text{C}$ for free foldon and gp26(1-140)-F, respectively. Experiments were conducted with 30 μM sample and in 20 mM sodium phosphate buffer pH 8 and 170 mM NaCl.

Figure 3. Reversibility of gp26-foldon fibers unfolding assessed by circular dichroism. The far-UV CD spectra of wild-type gp26 (A), gp26(1-140) (B), and gp26-foldon chimeras, gp26(1-140)-F (C), gp26(1-170)-F (D), and gp26(1-140)-F-F (E), were measured between 197 and 260 nm with 6 μ M sample in 20 mM sodium phosphate buffer pH 8 and 170 mM NaCl. For each sample, CD spectra were measured at 25 °C ($\circ$), 95 °C ($\triangle$), and after 30 min of cooling from 95 °C to 25 °C (∇).

on SDS-PAGE when unboiled, and break down into monomers if boiled prior to electrophoretic separation. Similarly, after refolding, only correctly refolded fibers in native conformation migrated as SDS-resistant trimers at room temperature (Fig. 4A). Aggregated and partially

misfolded fibers either failed to enter the separating gel and were retarded at the top of the SDS-PAGE, or were dissociated by SDS and migrated as monomers relative to their molecular masses (Fig. 4A). For all constructs, prolonging the time course of renaturation up to 180 min

Figure 4. SDS-PAGE analysis of refolded gp26-foldon fibers. (A) Wild-type gp26, gp26(1-140), and gp26-foldon chimeras, gp26(1-140)-F, gp26(1-170)-F, and gp26(1-140)-F-F, were either boiled for 5 min at 95°C (indicated as 95°C) or incubated 5 min at 22°C (indicated as RT). The samples were then dissolved in Laemmli sample buffer containing 0.1% SDS and analyzed on SDS-PAGE. (B) The intensity of the trimer and monomer band was quantified and plotted in the form of histogram. The percentage of trimer or monomer obtained on gel is normalized by the final protein concentration.

neither resulted in a statistical significant gain in SDS-resistant quaternary trimer nor increased the yield of renaturation as accessed by CD spectroscopy (data not shown). The relative abundance of monomeric and trimeric fibers present on SDS-PAGE after refolding was quantified densitometrically using the NIH/Scion Image software and plotted in the form of a histogram (Fig. 4B). Consistent with the CD data, only the construct gp26(1-140)-F yielded >95% refolded, trimeric, SDS-resistant fibers, which supports the idea the refolded gp26(1-140)-F fiber adopts a native trimeric conformation. In contrast, neither the construct gp26(1-170)-F nor the double foldon-containing construct gp26(1-140)-F-F displayed a greater percentage of refolded trimeric fiber as compared with the wild-type gp26.

Finally, as a third independent proof that refolded gp26(1-140)-F fibers adopt a native trimeric conformation, we employed negative stain transmission electron

microscopy. This technique is very powerful, as it allows direct visualization of gp26 fibers at a single-molecule level. Despite being only ~77 kDa and ~55 kDa in molecular weight, respectively, both wild-type gp26 and gp26(1-140)-F fibers present a characteristic rod-like morphology. Wild-type gp26 forms a stick ~240 Å in length and ~20–30 Å in diameter (Olia et al. 2007). The gp26(1-140)-F fiber is expected to have the same diameter, and a length of ~185 Å, which corresponds to ~165 Å of gp26 helical core (Olia et al. 2007) plus the foldon domain that resembles a sphere ~20 Å in diameter (Guthe et al. 2004). Since the foldon and gp26 have approximately the same diameter (20–30 Å), at low resolution the foldon appears as a short extension of gp26 helical core. To visualize refolded gp26 fibers, native gp26 and gp26(1-140)-F were first thermally denatured and then cooled to 25°C to promote refolding. The two specimens were spotted on a grid, stained with 1% uranyl formate, and visualized in the electron microscope (Fig. 5A,B). In line with other experiments, wild-type gp26 was found to form only visible amorphous aggregates (data not shown), while refolded gp26(1-140)-F fibers were indistinguishable from the native fibers (Fig. 5B). Projection averages of negatively stained native wild-type gp26 and folded/refolded gp26(1-140)-F fibers were computed by averaging together ~500 particles from each specimen manually boxed from the micrographs. The average projection of wild-type gp26 shows a rod ~240 Å in length, which matches well the size expected from the crystal structure and supports the validity of this structural analysis (Fig. 5A). Interestingly, projection averages of both folded and refolded gp26(1-140)-F fibers showed a thin rod ~180 Å in length and ~30 Å in diameter, which matches the length of the gp26 (1-140)-F structural model in Figure 5B. Thus, based on this analysis, we conclude that refolded gp26(1-140)-F fibers have identical quaternary structure as the native fibers.

Discussion

Proteins represent a fertile and fast-growing territory for nanobiotechnology in light of their sophisticated architecture yet simple and engineerable compositions (Astier et al. 2005). As of 2006, the total number of structure-based protein superfamilies and folds in the COG (Clusters of Orthologous Groups) database is estimated to be ~4000 and ~1700, respectively (Sadreyev and Grishin 2006). This provides an excellent repertoire for the rational design of new chimeras that combine known structural motifs to generate novel materials with specific desired properties. In this paper, we have fused together the helical core of tail needle gp26 with the foldon domain of phage T4 fibrin to generate a ~180 Å

Figure 5. Visualization of negatively stained gp26-foldon fibers. (A) Electron micrograph and projection average of negatively stained wild-type native gp26 fibers. The approximate length of the wild-type gp26 rod estimated from the projection average is ~ 240 Å, which fits well with the crystal structure of gp26 determined at pH 4.6 (Olia et al. 2007). (B) Electron micrograph and projection averages of negatively stained native fibers (upper panel) and refolded gp26(1-140)-F fibers (lower panel). Refolded fibers are indistinguishable from the native gp26(1-140)-F fibers and form an ~ 180 Å long stick similar to a model of gp26 helical core fused to the foldon, shown in ribbon diagram (right).

biofiber of high structural stability (T_m 85°C), great chemical versatility, and unique self-assembly properties.

The results presented in this work emphasize three important properties of the foldon-induced stabilization of the gp26 helical core. First, the foldon exerts a stabilizing effect only when introduced at the C terminus of gp26 helical core; constructs bearing an N-terminal foldon failed to correctly fold in *E. coli*. Second, within the C terminus of the gp26-foldon fiber, foldon-induced stabilization of the gp26 helical core is strictly position-dependent. Only the gp26(1-140)-F fiber that has the foldon fused at position 141, immediately downstream from the gp26 helical core, has native stability and refolds spontaneously in solution after heat- and/or chemical-induced denaturation. In contrast, fusion of the foldon only 30 residues downstream from the gp26 helix, at position 170, results in a marginal increase in the stability of gp26(1-140) and yields irreversible unfolding. Third, foldon-induced stabilization of gp26 chimeric fibers is dose independent. The fiber gp26(1-140)-F-F chimera not only lacks higher stability than its counterpart with a

single foldon gp26(1-140)-F, but it refolds with significantly lower efficiency.

How does the foldon mediate reversible unfolding to the gp26 helical core? The answer to this question likely resides in the kinetics of foldon trimerization, which is one of the fastest bimolecular protein associations reported in the literature (Guthe et al. 2004). We hypothesize (Fig. 6) that during refolding of gp26(1-140)-F fibers, the foldon-triggered self-association of the C termini is kinetically faster than the assembly of gp26 strands, which is dependent on the chemical complementarity of gp26 trimerization octads (Bhardwaj et al. 2007; Olia et al. 2007). By acting as a kinetic seed, the foldon drives the correct pairing of gp26 helices and tethers the C-terminal ends of the nascent fiber together (Fig. 6B). This process can be envisioned as the ordered zippering of the gp26 helical core from the C to N termini. In the absence of the foldon domain, N and C termini of gp26 have equal probability to initiate the reassembly of the trimer, which

Figure 6. Model for foldon-induced stabilization of gp26 helical core. (A) Ribbon diagram of tail needle gp26 and a schematic hypothetical model for the irreversible unfolding of gp26 helical core. In the absence of the foldon, gp26 strands fail to find the correct register resulting in aggregated protein. (B) Ribbon diagram of gp26(1-140)-F fiber. Hypothetical function of the foldon, which is thought to kinetically drive nucleation of gp26(1-140)-F C termini followed by correct assembly of the helical core based on chemical complementarity of the helices.

leads in vitro to the formation of aberrant, nonnative structures (Fig. 6A). This problem is more pronounced in the construct gp26(1-140), which lacks C-terminal domains III–IV, than in wild-type gp26, which shows ~30% spontaneous refolding after denaturation. We hypothesize that the inefficient refolding of gp26(1-140) is due to the lack of domains III–IV, which are known to stabilize the gp26 helical core (Bhardwaj et al. 2007).

The lack of reversibility of folding seen for constructs gp26(1-170)-F and gp26(1-140)-F-F supports the idea that the foldon functions as a kinetic seed to promote gp26 helical pairing. In the former construct, the β -stranded region 141–170 of gp26 also folds into a β -like structure that likely interferes with the formation of the foldon seed and leads to aggregation upon refolding. Likewise, two copies of the foldon in the construct gp26(1-140)-F-F not only fail to increase the overall stability of the fiber (the T_m for this construct is identical to that with a single foldon), but they likely lead to mispairing of foldon units upon refolding, resulting in only 30% refolded fibers. Complete refolding is seen only when the foldon domain is placed immediately downstream from the helical core, with no linker residues between the two folding units. This confirms that foldon-induced stabilization of the gp26 helical core is both position dependent and dose independent.

Another interesting property of gp26(140)-F fibers resides in the synergistic stabilization of the foldon and the gp26 helical core in the context of the chimeric fiber. By monitoring the fluorescence emission of Trp²⁰ in the isolated foldon and in the foldon fused to the gp26 helical core, we were able to measure an increased stabilization in the latter by 15°C (T_m of 89°C versus 74°C). This clearly confirms one of the main strengths of protein engineering: By rationally combining two structural domains, novel and enhanced structural and chemical properties of an *ensemble* can be obtained. Although it remains difficult to predict what such properties could be, the rational permutation of the foldon domain led us to identify the gp26(1-140)-F fiber in a small library of five permutants (Fig. 1A).

In conclusion, the work presented in this paper emphasizes how simple modifications to the protein primary sequence, such as fusion of a 27-amino-acid domain, can be used to dramatically enhance the thermodynamic and chemical stability of a protein fiber. The high stability of the gp26(1-140)-F fiber and the accessibility of N and C termini makes it possible to carry out chemistry on the fiber in order to gain novel functions. The next challenge in this direction involves the rational functionalization of gp26(1-140)-F N and C termini with chemical groups that promote ordered self-assembly of gp26 fibers, thus generating nanowires of controllable length, high structural stability, and novel nanoscale functions.

Materials and Methods

Molecular biology techniques

The gene encoding gp26 (Andrews et al. 2005) was amplified and ligated between the XbaI and HindIII restriction sites of the vector pMal-c2e (New England Biolabs) (plasmid pMal-gp26), which expresses gp26 fused to an N-terminal maltose binding protein (MBP). A PreScission Protease cleavage site was engineered between the MBP and gp26 (plasmid pMal-PP-gp26). Constructs gp26(1-140) and gp26(1-170) were generated by introducing an amber stop codon at position 141 or 171, respectively, of gp26 using site-directed mutagenesis (plasmids pMal-PP-gp26[1-140] and pMal-PP-gp26[1-170]). The DNA coding sequence for the phage T4 foldon was inserted into the parental gp26 plasmids using a two-step approach. To generate the two constructs gp26(1-140)-F and gp26(1-170)-F, the parental templates, pMal-PP-gp26(1-140) and pMal-PP-gp26(1-170), were linearized by PCR using two antiparallel, adjacent oligonucleotides, which anneal just before and after the stop codon. These two oligonucleotides also encode the N and C termini of the foldon domain, yielding a first-step PCR product that began and terminated in the last and first five amino acids of the foldon, as well as the stop codon. A second set of 5'-phosphorylated oligonucleotides encoding the middle 15 amino acids of the foldon were annealed to the first-step PCR template and used for a second round of extension by PCR. The second amplification product was circularized by blunt ligation, yielding plasmids pMAL-PP-gp26(1-140)-F and pMAL-PP-gp26(1-170)-F. The construct F_{Rev}-gp26(1-140)-F, which contains a reversed foldon domain upstream of gp26 helical core, was also generated in a two-step PCR approach. First, the parental plasmid pMAL-PP-gp26(1-140)-F was linearized by PCR prior to the first amino acid of gp26. Second, phosphorylated oligonucleotides encoding amino acids 6–13 (reverse) and 14–20 (forward) of the foldon were used to amplify by PCR the linearized template plasmid and complete the reversed N-terminal foldon. This product was circularized by blunt ligation, yielding the plasmid pMAL-PP-F_{Rev}-gp26(1-140)-F. Finally, constructs containing two copies of the foldon domain in direct orientation, gp26(1-140)-F-F and F_D-gp26(1-140)-F, were generated using the plasmid pMAL-PP-gp26(1-140)-F as a template. The foldon-coding region was amplified by PCR and blunt ligated into the parental template previously linearized by long PCR either at the N terminus of gp26 (plasmid pMAL-PP-F-gp26[1-140]-F) or at the C terminus of the first foldon domain (pMAL-PP-gp26 [1-140]-F-F). Finally, the foldon-coding region was also ligated into the pMal-PP vector, generating the plasmid pMal-PP-F. The presence and fidelity of the DNA sequence for all constructs generated in this study were confirmed by DNA sequencing.

Protein expression and purification

All proteins used in this study were fused to an N-terminal MBP and overexpressed in the *E. coli* strain BL21 (pLysE). For expression, cells derived from a single colony were grown at 37°C to a density of $A_{600} \sim 0.6$, and induced for 16 h at 22°C with 0.5 mM isopropyl 1-thio- β -D-galactopyranoside. Cells were lysed by sonication in lysis buffer (250 mM NaCl, 20 mM TrisHCl pH 8.0, 3 mM β -mercaptoethanol, plus protease inhibitors), and MBP fusion proteins were purified by sequential passages over amylose-agarose (New England Biolabs). Fusion

proteins were digested with PreScission Protease in 170 mM NaCl, 20 mM Tris-HCl, pH 8.0, and 5 mM β -mercaptoethanol. Typically, 1 L of *E. coli* yielded ~ 5 mg of pure protein, which was concentrated using an Amicon ultracentrifugal filter device (Millipore) with a 10,000 Da molecular weight exclusion limit. The concentrated proteins were applied to a Superdex 200 gel filtration column (GE Healthcare) pre-equilibrated in gel filtration buffer (170 mM NaCl, 20 mM sodium phosphate buffer pH 8.0). Elution fractions were analyzed with SDS-PAGE, confirming a purity of $>95\%$. The concentration of all recombinant proteins used in this study was determined spectrophotometrically at 280 nm using theoretical absorption coefficients calculated from the protein primary sequence. In the SDS-resistance assay, 20 μ L of gp26-foldon chimeras at ~ 1 mg/mL were incubated for 5 min either at room temperature (22°C) or at 95°C. Thereafter, samples were left for 10 min at room temperature (22°C) to allow refolding. Laemmli buffer was added to the proteins, followed by electrophoretic separation on 12.5% SDS-PAGE (Laemmli 1970). The gels were scanned and band intensities integrated using the NIH/Scion image software. The relative intensities of trimer versus monomer were plotted using the Microsoft Excel package.

Spectroscopic techniques

Equilibrium unfolding studies for gp26, gp26(1-140), and gp26-foldon chimeras were carried out by monitoring variations in ellipticity at 220 nm as a function of guanidine hydrochloride (GdnHCl) or temperature as previously described (Bhardwaj et al. 2007). All spectroscopic measurements were performed with a final protein concentration of 6 μ M in 20 mM sodium phosphate (pH 8.0) and 170 mM NaCl. Circular dichroism (CD) spectra in the far-UV region (197–260 nm) were recorded using an AVIV 62A DS spectropolarimeter equipped with a Neslab CFT-33 refrigerated recirculator. A rectangular quartz cuvette with a path length of 0.1 cm was used to perform the CD measurements. Low-noise CD spectra were obtained by averaging three scans. Reversibility of unfolding by CD was monitored by recording variations in ellipticity at 220 nm as a function of temperature in 1°C increments. After each increment, samples were equilibrated for 60 s and CD spectra recorded with an integration time of 15 s. Slow cooling to 25°C followed by a second run was carried out to check the reversibility of unfolding.

Equilibrium unfolding denaturation curves for the isolated foldon domain and gp26(1-140)-F were measured by monitoring the intrinsic fluorescence emission of Trp²⁰ of the foldon. Fluorescence spectra were collected on a Jobin Yvon Fluoromax-3 spectrofluorometer equipped with a thermally controlled cell holder and a 0.5-cm cuvette as previously described (Bhardwaj et al. 2007). The fluorescence data for both free foldon and gp26(1-140)-F fit well to a two-state transition, assuming a monomer unfolding model (Bhardwaj et al. 2007), where the two states correspond to folded native (*N*) trimer and unfolded (*D*) monomeric foldon or gp26(1-140)-F. Free energy of unfolding (ΔG_u) was obtained by fitting the fluorescence unfolding data with a nonlinear curve-fitting function using the software ORIGIN 6.0.

Negative stain microscopy

For negative stain electron microscopy, gp26 fibers at a concentration of ~ 1 μ g/mL were applied to glow-discharged carbon-coated copper grids and stained with 1% uranyl formate.

Grids were examined in a Joel JEM-2100 transmission electron microscope operated at 200 kV. Images were recorded on a charge-coupled device (CCD, TVIPS F415MP) at an electron optical magnification of 100,000 $\times$ and a defocus of -1.5 μ m, placing the first zero of the contrast transfer function (CTF) at ~ 0.05 $\text{\AA}^{-1}$. The pixel size on the specimen scale was 1.14 $\text{\AA}$ and was calibrated with catalase two-dimensional crystals. Data sets of ~ 500 images for each specimen were manually selected using the EMAN program *boxer* (Ludtke et al. 1999). Images were CTF corrected using *ctfit*, band pass filtered to remove low and high spatial frequencies, and a circular mask was applied. Reference-free image alignment and classification were performed using IMAGIC-5 software (Image Science GmbH) (van Heel et al. 1996).

Acknowledgment

We thank Aethalie Alexandra Chabriol for technical assistance in the purification of gp26 chimeras.

References

- Andrews, D., Butler, J.S., Al-Bassam, J., Joss, L., Winn-Stapley, D.A., Casjens, S., and Cingolani, G. 2005. Bacteriophage P22 tail accessory factor GP26 is a long triple-stranded coiled coil. *J. Biol. Chem.* **280**: 5929–5933.
- Astier, Y., Bayley, H., and Howorka, S. 2005. Protein components for nano-devices. *Curr. Opin. Chem. Biol.* **9**: 576–584.
- Bhardwaj, A., Olia, A.S., Walker-Kopp, N., and Cingolani, G. 2007. Domain organization and polarity of tail needle GP26 in the portal vertex structure of bacteriophage P22. *J. Mol. Biol.* **371**: 374–387.
- Boudko, S.P., Londer, Y.Y., Letarov, A.V., Sernova, N.V., Engel, J., and Mesyanzhinov, V.V. 2002. Domain organization, folding and stability of bacteriophage T4 fibrin, a segmented coiled-coil protein. *Eur. J. Biochem.* **269**: 833–841.
- Conley, M.P. and Wood, W.B. 1975. Bacteriophage T4 whiskers: A rudimentary environment-sensing device. *Proc. Natl. Acad. Sci.* **72**: 3701–3705.
- Frank, S., Kammerer, R.A., Mechling, D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lustig, A., Bachinger, H.P., and Engel, J. 2001. Stabilization of short collagen-like triple helices by protein engineering. *J. Mol. Biol.* **308**: 1081–1089.
- Frank, S., Boudko, S., Mizuno, K., Schulthess, T., Engel, J., and Bachinger, H.P. 2003. Collagen triple helix formation can be nucleated at either end. *J. Biol. Chem.* **278**: 7747–7750.
- Guthe, S., Kapinos, L., Moglich, A., Meier, S., Grzesiek, S., and Kiefhaber, T. 2004. Very fast folding and association of a trimerization domain from bacteriophage T4 fibrin. *J. Mol. Biol.* **337**: 905–915.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.
- Letarov, A.V., Londer, Y.Y., Boudko, S.P., and Mesyanzhinov, V.V. 1999. The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibrin. *Biochemistry (Mosc.)* **64**: 817–823.
- Ludtke, S.J., Baldwin, P.R., and Chiu, W. 1999. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**: 82–97.
- Olia, A.S., Casjens, S., and Cingolani, G. 2007. Structure of phage P22 cell envelope-penetrating needle. *Nat. Struct. Mol. Biol.* **14**: 1221–1226.
- Papanikolopoulou, K., Forge, V., Goeltz, P., and Mitraki, A. 2004a. Formation of highly stable chimeric trimers by fusion of an adenovirus fiber shaft fragment with the foldon domain of bacteriophage t4 fibrin. *J. Biol. Chem.* **279**: 8991–8998.
- Papanikolopoulou, K., Teixeira, S., Belrhali, H., Forsyth, V.T., Mitraki, A., and van Raaij, M.J. 2004b. Adenovirus fibre shaft sequences fold into the native triple β -spiral fold when N-terminally fused to the bacteriophage T4 fibrin foldon trimerisation motif. *J. Mol. Biol.* **342**: 219–227.
- Sadreyev, R.I. and Grishin, N.V. 2006. Exploring dynamics of protein structure determination and homology-based prediction to estimate the number of

- superfamilies and folds. *BMC Struct. Biol.* **6**: 6. doi: 10.1186/1472-6807-6-6.
- Sissoeff, L., Mousli, M., England, P., and Tuffereau, C. 2005. Stable trimerization of recombinant rabies virus glycoprotein ectodomain is required for interaction with the p75NTR receptor. *J. Gen. Virol.* **86**: 2543–2552.
- Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R.A., Boudko, S., Landwehr, R., Okuyama, K., and Engel, J. 2003. Collagen stabilization at atomic level: Crystal structure of designed (GlyProPro)₁₀foldon. *Structure* **11**: 339–346.
- Tao, Y., Strelkov, S.V., Mesyanzhinov, V.V., and Rossmann, M.G. 1997. Structure of bacteriophage T4 fibrin: A segmented coiled coil and the role of the C-terminal domain. *Structure* **5**: 789–798.
- van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. 1996. A new generation of the IMAGiC image processing system. *J. Struct. Biol.* **116**: 17–24.
- Yang, X., Lee, J., Mahony, E.M., Kwong, P.D., Wyatt, R., and Sodroski, J. 2002. Highly stable trimers formed by human immunodeficiency virus type 1 envelope glycoproteins fused with the trimeric motif of T4 bacteriophage fibrin. *J. Virol.* **76**: 4634–4642.