Very Fast Folding and Association of a Trimerization Domain from Bacteriophage T4 Fibritin

Sarah Güthe1†, Larisa Kapinos1†, Andreas Möglich1†, Sebastian Meier2, Stephan Grzesiek2 and Thomas Kiefhaber1*

1Division of Biophysical Chemistry, Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland
2Division of Structural Biology, Biozentrum der Universität Basel, Klingelbergstrasse 70 CH-4056 Basel, Switzerland

The foldon domain constitutes the C-terminal 30 amino acid residues of the trimeric protein fibritin from bacteriophage T4. Its function is to promote folding and trimerization of fibritin. We investigated structure, stability and folding mechanism of the isolated foldon domain. The domain folds into the same trimeric β-propeller structure as in fibritin and undergoes a two-state unfolding transition from folded trimer to unfolded monomers. The folding kinetics involve several consecutive reactions. Structure formation in the region of the single β-hairpin of each monomer occurs on the submillisecond timescale. This reaction is followed by two consecutive association steps with rate constants of $1.9(\pm 0.5) \times 10^6$ M$^{-2}$ s$^{-1}$ and $5.4(\pm 0.3) \times 10^6$ M$^{-2}$ s$^{-1}$ at 0.58 M GdmCl, respectively. This is similar to the fastest reported bimolecular association reactions for folding of dimeric proteins. At low concentrations of protein, folding shows apparent third-order kinetics. At high concentrations of protein, the reaction becomes almost independent of protein concentrations with a half-time of about 3 ms, indicating that a first-order folding step from a partially folded trimer to the native protein ($k = 210(\pm 20) s^{-1}$) becomes rate-limiting. Our results suggest that all steps on the folding/trimerization pathway of the foldon domain are evolutionarily optimized for rapid and specific initiation of trimer formation during fibritin assembly. The results further show that β-hairpins allow efficient and rapid protein–protein interactions during folding.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: protein folding; protein association; trimeric proteins; prolyl isomerization; fast folding

Introduction

Fibritin is a rod-like structural protein of bacteriophage T4, which is attached to the neck of the virion via its N-terminal domain to form the collar structures ("whiskers"). Fibritin consists of an N-terminal anchor domain (residues 1–46), a large central coiled-coil part (residues 47–456) and a small C-terminal globular domain (residues 457–486).1 The 30 amino acid residue C-terminal domain was termed foldon, since it was shown to be essential for fibritin trimerization and folding in vivo and in vitro.1–3 Each subunit of the foldon domain consists of a single β-hairpin, which assemble into a β-propeller-like structure in the trimer.1 The trimer is stabilized by hydrophobic interactions involving Trp476 of each subunit, intermolecular salt-bridges between Glu461 and Arg471, and intermolecular backbone hydrogen bonds between Tyr469 and Arg471 (Figure 1). Expression of the isolated foldon domain (residues 457–483) yields a stable trimer, which shows a cooperative two-state thermal unfolding transition.4 Residues 484–486 were omitted from this study, since this region is unordered in the X-ray structure of fibritin.1

†S.Gü., L.K. and A.M. contributed equally to this work.
Abbreviations used: F-moc, N-(9-fluorenyl)-methoxycarbonyl; GdmCl, guanidinium chloride; RDC, residual dipolar coupling.
E-mail address of the corresponding author: t.kiefhaber@unibas.ch
motif, as its only known function is to promote folding of fibrin. The small size of its structured part (27 amino acid residues) and its simple fold make the foldon domain a perfect system for a detailed study on the mechanism of a folding reaction linked to intermolecular association steps. All previous folding studies on trimeric proteins investigated large filamentous proteins, which show extremely slow and complex folding kinetics, usually accompanied by irreversible aggregation reactions.

We expressed the foldon domain in *Escherichia coli* and synthesized it by solid-phase N-(9-fluorenyl)methoxycarbonyl (F-moc) chemistry to investigate its structure, stability and folding mechanism. For clarity, we are numbering the foldon sequence from residues 1 to 27:

\[
\begin{align*}
1 & \quad 11 & \quad 21 \\
GYIPEAPRDG & QAYVRKDEGW & VLLSTFL
\end{align*}
\]

corresponding to residues 457–483 in fibrin. All kinetic and stability data presented here were obtained using the chemically synthesized foldon domain, whereas the recombinant *E. coli* product was used for structural analysis. The *E. coli* product and the synthetic foldon domain showed identical stability and folding behavior. Further, the additional C-terminal amino acid residues Ser-Pro-Ala, which are present in the wild-type fibrin sequence, do not affect any thermodynamic or kinetic properties of the foldon domain.

**Figure 1.** A, Stereo view of a bundle of the 20 lowest-energy structures of the trimeric foldon domain determined by NMR spectroscopy. Each subunit is displayed in a different color. B, Side view of the foldon structure with the single Trp residues at position 20 of each chain highlighted in green and the two prolyl residues at positions 4 and 7 highlighted in red. C, Topology of the interactions of the three β-hairpins in the native foldon domain. The figures in A and B were prepared using the program MOLMOL and rendered with PovRay.
Results and Discussion
Structure and stability of the foldon domain

To test whether the 27 amino acid residue foldon domain adopts the same fold as in fibrin, we solved its solution structure to a backbone rmsd of 0.31 Å with 28 experimental restraints per residue (Figure 1 and Table 1). The structure largely resembles the crystal structure in constructs carrying the 75 and 120 C-terminal amino acid residues of fibrin, in which the foldon domain constitutes only a minor part of the total construct. 1, 6 Only at the immediate N terminus (residues 1–3) the isolated foldon domain assumes a slightly different and presumably more relaxed structure compared to fibrin. The trimer consists of an N-terminal hydrophobic stretch in left-handed polyproline II helix conformation between Pro4 and Pro7, which is connected to a β-hairpin (residues 12–23) and forms a hydrophobic cap of the hairpin on the N-terminal side (Figure 1). The hairpin terminates in a 3_10 helix at the C terminus with homophilic interactions of hydrophobic residues (Tyr2, Ile3, Val14, Leu23, Leu27) between the monomers along the symmetry axis. Large-scale nanosecond dynamics as evidenced by ^{15}N relaxation occur only at the most N-terminal residue, Tyr2. All other residues in the highly rigid foldon domain exhibit order parameters ^2 > 0.77 at 25 °C as determined by the program TENSOR. 7

The equilibrium unfolding properties of the foldon domain were measured by guanidinium chloride (GdmCl)-induced unfolding transitions at various concentrations of protein. Figure 2 shows unfolding curves at monomer concentrations of 5 μM and 30 μM. The coincidence of fluorescence and CD-monitored transition curves demonstrates that the trimer unfolds in a cooperative two-state transition at both concentrations of protein. Two-state unfolding is observed for all concentrations of protein between 2 μM and 100 μM. The sensitivity of the unfolding transitions to changes in protein concentration is expected for unfolding of a native trimer (N) to unfolded monomers (U):

\[ N = 3U \] (1)

The transitions at 5 μM and 30 μM can be fit globally to equation (1) (continuous line in Figure 2) by using:

\[ K_{eq} = \frac{3M_0^2}{1 - f_u} \] (2)

where \([M_0]\) indicates the total monomer concentration ([M]₀ = [U] + 3[N]), \(f_u\) is the fraction of

---

Table 1. Statistics of the foldon NMR structure

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rmsd from experimental distance constraints (Å)</td>
<td>0.046 ± 0.002</td>
</tr>
<tr>
<td>rmsd from NMR data</td>
<td>1.91 ± 0.07</td>
</tr>
<tr>
<td>NMR quality factor Q&lt;sup&gt;8&lt;/sup&gt;</td>
<td>0.199 ± 0.0062</td>
</tr>
<tr>
<td>rmsd (Hz) between measured and calculated dipolar couplings (81)&lt;sup&gt;9&lt;/sup&gt;</td>
<td>1.73 ± 0.29</td>
</tr>
<tr>
<td>Experimental dihedral constraints (deg.)&lt;sup&gt;9&lt;/sup&gt; (43)</td>
<td>0.89 ± 0.015</td>
</tr>
<tr>
<td>Total number of restraints per monomer</td>
<td>753</td>
</tr>
<tr>
<td>Deviation from the idealized covalent geometry</td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.0079 ± 0.0004</td>
</tr>
<tr>
<td>Angles (deg.)</td>
<td>0.92 ± 0.03</td>
</tr>
<tr>
<td>Improper* (deg.)</td>
<td>0.75 ± 0.06</td>
</tr>
<tr>
<td>Coordinate precision (Å)</td>
<td></td>
</tr>
<tr>
<td>Backbone non-hydrogen atoms</td>
<td>0.273</td>
</tr>
<tr>
<td>All non-hydrogen atoms</td>
<td>0.636</td>
</tr>
<tr>
<td>Non-Gly, non-Pro residues in Ramachandran regions&lt;sup&gt;8&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Most favored (%)</td>
<td>91.7</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>8.3</td>
</tr>
<tr>
<td>Generously allowed (%)</td>
<td>0.0</td>
</tr>
<tr>
<td>Disallowed (%)</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The statistics were obtained from a subset of the 40 best energy annealing structures out of 100 following a standard simulated annealing protocol with dipolar restraints incorporated. Individual simulated annealing structures are fitted to each other using residues 2–27 of all subunits. The number of the various constraints per monomer is given in parentheses.

<sup>a</sup> Distance restraints comprise: 111 intraresidual NOEs; 139 sequential NOEs (i–j = 1); 73 short range NOEs (1 < i – j ≤ 5); 128 long-range NOEs (i–j ≤ 5); 156 intermolecular NOEs; 11 H-bonds (eight intramolecular, three intermolecular). For each backbone hydrogen bond constraint, there are two distance restraints: _DQN<sub>1</sub>, 1.7–2.5 Å, _DQN<sub>1</sub>, 2.3–3.5 Å.

<sup>b</sup> The NMR quality factor Q is defined as the ratio of the rmsd between observed and calculated couplings and the rmsd of the observed couplings.

<sup>c</sup> The 81 RDCs comprise 22 _DQ<sub>21</sub>, 20 _DQ<sub>21</sub>, 12 _DQ<sub>21</sub>, 13 _DQ<sub>21</sub> (0.231), 14 _DQ<sub>21</sub>. Ramping the force constant for RDCs in the structure calculation from 0.001 kcal mol<sup>–1</sup> Hz<sup>–2</sup> to 0.5 kcal mol<sup>–1</sup> Hz<sup>–2</sup> was determined as optimal.

<sup>d</sup> The dihedral angle constraints comprise 69 φ and 60 ψ angles.

<sup>e</sup> The improper torsion restraints serve to maintain planarity and chirality.

<sup>f</sup> The coordinate precision is defined as the average rms difference between the individual simulated annealing structures and the mean coordinates. Values are reported for residues 2–27.

<sup>g</sup> These values are calculated with the program PROCHECK-NMR. Values are reported for all residues.

---

Figure 2. GdmCl-induced unfolding transition of the foldon domain at pH 7.1, 20 °C. Transitions at 5 μM (○, ○) and 30 μM (△, △) total monomer concentration ([M]₀) were measured by changes in Trp fluorescence (○, △) and in far-UV CD at 228 nm (○, △). The data were normalized to fraction of native molecules using the result of a global fit of all data according to equations (1a) and (1b) (continuous lines).
unfolded monomer \((f_u = [U]/[M_0])\) and \(K_{eq}\) is the equilibrium constant (for details, see Materials and Methods). The global fit yields a free energy of unfolding of \(\Delta G^0(H_2O) = 89.2(\pm0.6)\text{kJ/mol}\), which is unusually high compared to stabilities of small, single-domain proteins of similar size. However, this value applies to standard conditions of 1 M total monomer concentration. At typical physiological protein concentrations around 5 \(\mu\)M this corresponds to \(\Delta G = 29.7\text{kJ}\), which is similar to the stabilities found for small monomeric proteins. The change in free energy with GdmCl \((m_{eq} = \partial \Delta G^0/\partial (\text{GdmCl}))\) is \(-10.4(\pm0.2)\text{kJ/mol}/\text{M}\), which is the value expected for a monomeric globular protein of the size of the folded trimer. This shows that native foldon has properties comparable to those of small monomeric proteins with a compact hydrophobic core and a cooperative two-state unfolding transition.

Burst phase fluorescence changes

To investigate the folding kinetics of the foldon domain we performed stopped-flow refolding experiments starting from GdmCl-unfolded protein. Figure 3A shows a refolding trace at a residual denaturant concentration of 0.58 M and \([M_0]\) of 5 \(\mu\)M. The kinetics were monitored by the change in intrinsic tryptophan fluorescence above 320 nm. Within the first millisecond of refolding, a major burst phase reaction occurs, which leads to a significant increase in fluorescence intensity above the signals of both the unfolded and the native protein. This indicates very rapid structural changes in the dead-time of stopped-flow mixing (about 1 ms). The fluorescence intensity decreases slowly and reaches the value of the native protein after about 300 seconds. The burst phase increase in fluorescence is observed for all measured concentrations of protein (0.5 \(\mu\)M to 200 \(\mu\)M), indicating that the reaction occurs within the monomer. Even the fastest, diffusion-controlled association reaction to a partially folded dimer could not be complete within 1 ms at a monomer concentration of 1 \(\mu\)M and below, if we assume a maximum second-order rate constant of about \(1 \times 10^6 \text{M}^{-1} \text{s}^{-1}\). To further investigate the structural changes occurring in the burst phase, we monitored the folding kinetics at single wavelengths between 290 nm and 430 nm. Extrapolating the kinetic traces at the individual wavelengths to time zero allows the determination of the fluorescence spectrum of the burst phase intermediate (Figure 3B). Comparison of the fluorescence spectra of native and unfolded protein with the zero timepoint spectrum shows that the burst phase intermediate has a fluorescence emission maximum around 330 nm, which is between the emission maximum of the native protein \((\lambda_{max} = 317 \text{nm})\) and the unfolded state \((\lambda_{max} = 345 \text{nm})\). The significantly blue-shifted fluorescence maximum and the largely increased fluorescence intensity in the intermediate relative to the unfolded state suggest that the burst phase intermediate has a significantly more hydrophobic environment around the single tryptophan residue at position 20 in each β-hairpin (see Figure 1). The absence of a tyrosine fluorescence band at 303 nm in the burst phase intermediate further indicates significant chain compaction, which allows efficient energy transfer from the two tyrosine residues at positions 2 and 13 to Trp20. Similar fluorescence properties are observed for an acid-induced monomeric state (A-state) of the foldon domain, which shows virtually the same fluorescence emission spectrum as the burst phase intermediate but with reduced fluorescence intensity (Figure 3).

Fast and slow steps during association of the foldon domain

To determine the nature of the rate-limiting steps
during folding and association of the foldon domain, we analyzed the concentration-dependence of the refolding kinetics at final concentrations of protein between 0.5 μM and 200 μM. Figure 4A shows that the kinetics are strongly concentration-dependent. A very slow reaction on the hundreds of seconds timescale is observed at low concentrations of protein. At higher concentrations of protein ([M₀] > 10 μM) a concentration-dependent faster reaction with a half-time of about 4 ms at [M₀] = 102 μM and a concentration-independent slow reaction (τ = 50 s) are observed. The foldon domain contains two prolyl residues (Pro4 and Pro7) per subunit, which are in the trans conformation in the native state. This makes prolyl isomerization reactions a possible source for the slow, concentration-independent kinetics. Figure 5 shows that the slow reaction is catalyzed efficiently by human cyclophilin 18, a peptidyl-prolyl cis-trans isomerase, which identifies this reaction as a cis-trans isomerization at one or both of the two Xaa-Pro peptide bonds per monomer. The faster reaction is not affected by the presence of cyclophilin (data not shown).

We tested whether the faster, concentration-dependent reaction produces native protein or a folding intermediate by performing interrupted refolding experiments. In these experiments, the protein is allowed to refold for a certain time (tᵢ). The protein is then transferred to unfolding conditions and the resulting unfolding kinetics are monitored. The native state has a characteristic stability and barrier for unfolding, which results in a characteristic rate constant for its unfolding reaction. This distinguishes it from partially folded intermediates. Interrupted refolding experiments measure the increase in amplitude of the unfolding reaction of the native protein as a function of the refolding time, tᵢ. This corresponds to the time-course of formation of the native state. Thus, interrupted refolding experiments can distinguish whether a folding reaction produces native protein or a folding intermediate. Figure 4B shows that native molecules are formed in a fast and a slow reaction, which occur on the same timescale as the two fluorescence-detected folding reactions (Figure 4B). Obviously, both the faster, concentration-dependent process and the prolyl-isomerization limited process produce native protein. This suggests that the fast reaction reflects formation of the native trimer for molecules with all prolyl peptide bonds in the native trans conformation (fast-folding molecules, Uᵢ). The slow process is due to folding of molecules with at least one non-native cis isomer (slow-folding molecules, Uₛ), as shown by its catalysis by cyclophilin (Figure 5).

**Mechanism of folding and association**

The assignment of the different kinetic phases to direct folding and prolyl isomerization steps enables us to characterize the folding/association
process of the foldon domain in more detail. The analysis of the concentration-dependence of the half-lives ($t_{1/2}$) of the folding reaction starting from $U_b$ allows us to determine the apparent reaction order for the fast-folding pathway according to:

$$\log t_{1/2} = c - (n - 1) \log [M]_0$$

where $n$ is the reaction order and $c$ is a reaction-specific constant. Analysis of the concentration-dependence of folding thus yields information on contributions from concentration-dependent association reactions and from concentration-independent folding steps. We determined the half-time of the folding reaction from the concentration-dependence of the fluorescence-detected kinetics shown in Figure 4A. At high concentrations of protein, where folding and prolyl isomerization are well separated, we evaluated the half-lives of the fast reactions in order to obtain information on the direct folding reaction. Figure 6 shows that the slope of $\log t_{1/2}$ versus $\log [M]_0$ changes with protein concentration. At low concentrations of protein the slope is $-2.06 \pm 0.05$, which corresponds to apparent third-order kinetics. At monomer concentrations above 5 $\mu$M the slope decreases significantly, indicating a change in the apparent reaction order.

Between 100 $\mu$M and 200 $\mu$M initial monomer concentration there is only little effect of protein concentration on the folding kinetics. This shows that the fast-folding reaction approaches the first-order limit at high concentrations of protein.

The change from apparent third-order kinetics to first-order kinetics shows that folding is limited by association steps at low concentrations of protein and by a unimolecular folding reaction at high concentrations of protein. It is very unlikely that apparent third-order kinetics for a reaction in solution arise from a true trimolecular reaction. Third-order reactions are usually caused by a rapid monomer–dimer pre-equilibrium followed by a second bimolecular association step to form the trimer. This mechanism gives rise to apparent third-order kinetics if the dimer is populated to only very low levels. The observation of apparent third-order kinetics at low concentrations of protein confirms the finding that the rapid collapse observed within the first millisecond of refolding is due to a conformational change in the monomer.

The weak concentration-dependence of the fluorescence-detected kinetics at high concentrations of protein implies that a unimolecular folding step becomes rate-limiting when all association reactions are fast. The half-time for formation of the native state of about 3–4 ms at high concentrations of protein indicates that the rate constant of the unimolecular step is around 200–300 s$^{-1}$. It is reasonable to assume that this unimolecular folding process is due to a structural rearrangement in the trimer and represents a late step in the formation of the native structure, and might be similar to the final steps during folding of a single-domain protein. However, we cannot exclude completely the possibility that the first-order reaction, which becomes rate-limiting at high concentrations of protein, occurs at the level of the burst phase intermediate or a partially folded dimer. These considerations lead to the minimal folding model for the foldon domain:

$$\begin{align*}
81\% & \quad 3 \, U_t \quad \xrightarrow{k_m} \quad 3 \, I_t & \quad 1.9 \times 10^6 \text{M}^{-1}\text{s}^{-1} \\
& \quad \xleftarrow{k_m} \quad 3 \, I_c & \quad 59 \text{ s}^{-1} \\
19\% & \quad 3 \, U_c \quad \xrightarrow{k_m} \quad 3 \, I_c & \quad 5.7 \times 10^{-3} \text{s}^{-1}
\end{align*}$$

In this mechanism, $U$ denotes the completely unfolded monomer, $I$ is the monomeric burst phase intermediate, $D$ is a partially folded dimer, $T$ is a partially folded trimer and $N$ the native trimer. The subscripts $c$ and $t$ indicate monomers with a least...
one cis Xaa-Pro peptide bond and monomers with all trans peptide bonds, respectively. $k_a$ and $k_m$ represent rate constants for formation and unfolding of the burst phase intermediate, respectively, which occurs on the submillisecond timescale. To test whether this folding mechanism can describe all experimental data quantitatively, we fitted the folding kinetics at different concentrations of protein (Figure 4A) together with the time-course of formation of native foldon at 10 μM (Figure 4B) globally to the mechanism shown in equation (4). The results show that both the concentration-dependence of fluorescence kinetics and the time-course of formation of native molecules are well described by the model (continuous lines in Figure 4A and B).

The global fit allows the determination of the rate constants for all reactions that occur after the submillisecond burst phase, since the apparent reaction order changes from third-order to almost first-order. The results of the global fit are given in equation (4). In agreement with the lower limit of the estimate from the limiting value of the half-lives for folding at high concentrations of protein (Figure 6) the fit gives a rate constant ($k_i$) of structural rearrangement in the trimer of 210(±20) s$^{-1}$. This is in the same range as the rate constants observed for the fastest folding monomeric proteins of similar size. The bimolecular rate constants for formation of the dimer and of the trimer are 1.9(±0.5)×10$^9$ M$^{-1}$ s$^{-1}$ and 5.4(±0.3)×10$^8$ M$^{-1}$ s$^{-1}$, respectively, which is about 100 to 200-fold slower than the expected diffusion limit for bimolecular reactions of chains of this size. However, both association reactions on the foldon trimerization pathway are significantly faster than most bimolecular steps during folding of dimeric proteins like the well-characterized GCN4 leucine zipper and of many globular dimeric proteins. Rate constants similar to the two bimolecular steps during trimerization of foldon have been reported for wild-type arc repressor, the fast-folding naturally occurring dimeric protein known, and for some designed leucine zippers. However, an engineered arc repressor variant and a designed fragment of trp repressor show bimolecular rate constants of about 3×10$^8$ M$^{-1}$ s$^{-1}$, which are the fastest association reactions reported for folding of small dimeric proteins to date. In the case of the engineered arc repressor, the rate-enhancement relative to the wild-type protein was achieved by replacing the intermolecular salt-bridge/hydrogen bonding network by hydrophobic residues. The fast-folding trp repressor fragment is stabilized mainly by intermolecular hydrophobic interactions. Similar to wild-type arc repressor, the foldon domain is stabilized by intermolecular hydrogen bonds and by an intermolecular salt-bridge, which contributes substantially (17 kJ/mol) to trimer stability (S.M. et al., unpublished results). In this respect, it is interesting to note that the association steps in the foldon domain have virtually the same rate constants as the dimerization step of wild-type arc repressor.

The unfolding rate constants of foldon under native conditions (0.58 M GdmCl) obtained from the fit reveal that the major barrier for unfolding is represented by the reaction from the native trimer (N) to the partially folded trimer ($k = 4.2(±0.5)×10^{-4}$ s$^{-1}$). This value corresponds well to the rate constant for unfolding of native foldon measured at high concentrations of denaturant and extrapolated to 0.58 M GdmCl (S.Gü. & T.K., unpublished results). Unfolding of the partially folded trimer (T) and the dimer (D) are significantly faster with rate constants of 110(±40) s$^{-1}$ and 59(±4) s$^{-1}$, respectively, obtained from the fit.

In the folding mechanism shown in equation (4), we assumed that only the monomers with native trans prolyl isomers can form productive dimers and trimers. According to studies on model peptides the Ile3-Pro4 (12% cis) and Ala6-Pro7 (8% cis) peptide bonds in foldon should lead to 19% unfolded monomers with at least one cis prolyl peptide bond. This agrees well with 20(±1)% of slow-folding molecules observed in the interrupted refolding experiments (equation (4) and Figure 4B), if we assume that only monomers with both prolyl peptide bonds in trans can enter the productive folding pathway. The presence of 20(±1)% slow-folding molecules would, however, be observed if only the Ala6–Pro7 peptide bond, which is in a highly structured region of foldon (Figure 1), was essential for folding and if a cis bond at this position could be incorporated into partially folded dimers and trimers. For this mechanism, the presence of 22% slow-folding molecules would be expected. Our data do not allow us to discriminate between these mechanisms. However, the rate constants for the fast-folding pathway are not influenced significantly by the folding mechanism of the slow-folding molecules.

The determination of all rate constants on the folding/association pathway of the foldon domain enables us to calculate the population of each species during folding at various concentrations of protein (Figure 7). At an initial concentration of 1 μM monomer, only the burst phase intermediate (I) and the native state become populated significantly (>10%) during folding. At concentrations of protein above 5 μM, the dimer becomes populated transiently, in agreement with a change in reaction order around this concentration of protein (Figure 6). As a consequence, the fast-folding pathway changes from apparent three-state to apparent four-state with the burst phase intermediate and the dimer populated to significant amounts. Above a concentration of 50 μM monomer, also the trimeric intermediate becomes populated significantly. The significant population of dimeric and trimeric intermediates explains the low apparent half-time of the reaction at high concentrations of protein, which would suggest a unimolecular folding reaction faster than the 210(±20) s$^{-1}$ obtained from the global fit. Since the apparent half-time was determined from the fluorescence measurements (Figure 6), it will be influenced by
the kinetics of formation of dimers and trimers when these become populated significantly. These reactions are fast at high concentrations of protein, which results in apparently smaller and slightly concentration-dependent half-times for the folding reaction at high concentrations of protein. However, these half-times do not represent the true half-times for the first-order reactions.

During phage assembly in the cell, the estimated concentration of monomeric fibrin molecules is between 1 μM and 5 μM. Figure 7 shows that under these conditions neither the dimeric (D) nor the trimeric (T) intermediate becomes populated to significant amounts. At in vivo concentrations, the half-time for folding is between one and ten seconds, which is very fast compared to the generation time of the phage of around 30 minutes.

**Fast folding of multimeric proteins compared to monomeric proteins**

The folding mechanism of the foldon domain has several interesting differences compared to fast-folding reactions of small monomeric proteins. Small monomeric proteins usually do not populate partially folded states to significant amounts during folding, although high-energy intermediates were shown to be obligatory for folding of many apparent two-state folders.\(^{23,24}\) Obviously, small monomeric proteins are able to avoid transient high concentrations of intermediates, which minimizes the probability of aggregation side-reactions and optimizes the shape of the free energy barriers for rapid folding.\(^{25}\) In the case of the foldon domain, however, rapid formation of intermolecular interactions is essential for efficient folding. Formation of the burst phase intermediate (I) leads (ii) to pronounced fluorescence changes of the single Trp residue, which is part of the β-hairpin, (iii) to chain compaction as indicated by efficient energy transfer from the two tyrosine residues to the tryptophan and (iii) to significant changes in the far-UV CD signal (data not shown). This suggests that the monomeric intermediate involves structure formation of the β-hairpin, which is in agreement with the results of NMR studies on a monomeric state of the foldon domain (A-state) observed in equilibrium at low pH. In this state, the β-hairpin forms essentially the same structure as in the folded trimer (S.M. et al., unpublished results). This shows that the monomeric foldon domain has a high propensity to form a β-hairpin even in the absence of intermolecular interactions. The similarity between the burst phase intermediate and the A-state is supported by the identical fluorescence emission maxima of the two states (Figure 3B). Rapid formation of the β-hairpins in the monomers probably facilitates the subsequent association reactions, since the hairpins provide highly specific surfaces that allow fast formation of intermolecular interactions. This model is in agreement with the folding mechanisms of the dimeric GCN4 fragment, for which rapid formation of α-helical structure was shown to accelerate the subsequent bimolecular association step.\(^{26}\) Obviously, optimized energy landscapes for folding of monomeric and oligomeric proteins are different. Monomeric proteins fold fastest when the intermediates are marginally less stable than the unfolded state,\(^{25}\) whereas populated intermediates promote the subsequent concentration-dependent association reactions in oligomeric proteins. The rapid formation of a partially folded state in the 27 residue foldon monomer shows that small, single-domain proteins should be able to rapidly form partially folded intermediates, provided this was beneficial for efficient folding.
Conclusions

Our folding studies on the foldon domain provide the first detailed information on the folding mechanism of a fast-folding trimeric protein. Previous studies on trimeric proteins focused mainly on large proteins, which fold very slowly and are accompanied by irreversible aggregation reactions or on collagens, which are limited in folding by prolyl cis–trans isomerization reactions. Our results suggest that in the small 27 amino acid residue foldon domain, all folding and association steps are optimized for rapid formation of a stable trimer. The two consecutive association reactions show bimolecular rate-constants similar to the values found in the fastest-folding small dimeric proteins. Further, the first-order folding reaction, which becomes rate limiting at high concentrations of protein, has a rate constant comparable to those of fast-folding, small, single-domain proteins. In contrast to previously studied small multimeric proteins or protein fragments, where association involved mainly helix–helix interactions, the foldon domain trimerizes via β-hairpins. The high bimolecular rate constants for both association steps in the foldon domain demonstrate that β-hairpins allow very rapid and specific formation of intermolecular interactions during protein assembly.

Materials and Methods

Protein synthesis and purification

The foldon domain was synthesized chemically with an ABIMED economy peptide synthesizer EPS 221 (Abimed, Germany) using standard F-moc chemistry. Pre-coupled resin was purchased from Novabiochem, Switzerland. Amino acids were from Alexis Biochemicals, USA or from Iris Biotech, Marktredwitz, Germany. Solvents and other chemicals were from Fluka (Buchs, Switzerland). The protein was purified using HPLC with a C-8 reverse-phase preparative column (Hitachi, LiChrosorb®100 RP-8 from Merck, Darmstadt, Germany). 13C,15N-labeled foldon for NMR spectroscopy were prepared as 0.3 mM protein solutions at pH 7.1 in 100% 2H2O) of 300 mM sodium phosphate, 0.02% (w/v) NaN3 without 13C,15N-enriched M9 minimal medium. Protein purity was confirmed by nanospray mass spectrometry and analytical HPLC. Expression and purification of the E. coli was performed as described. 4

NMR spectroscopy

NMR samples ([U-15N]foldon 95% H2O/5% 2H2O; [U-13C,15N]foldon 95% H2O/5% 2H2O; [U-13C,15N]foldon 100% 2H2O) of 300 µl volume (Shigemi NMR microtubes) were prepared as 0.3 mM protein solutions at pH 7.1 in 5 mM sodium phosphate, 0.02% (w/v) NaN3 without any further addition of salt. Residual alignment of [U-13C,15N]foldon for the measurement of residual dipolar couplings was introduced by lamellar ether/n-hexanol phases. A set of standard triple and double-resonance assignment, quantitative J-coupling, nuclear Overhauser effect (NOE) spectroscopy (NOESY) and 15N relaxation experiments similar to those described were performed on a Bruker DRX 600 spectrometer at 25 °C. Standard data processing and analysis was carried out as described.

Structure calculation

Experimental NOE distance, torsion angle and residual dipolar restraints derived from the NMR data are listed in Table 1. Structure calculations were carried out with a simulated annealing protocol using the program CNS. The structural statistics for the best 40 structures are given in Table 1.

Protein Data Bank accession code

The structural data have been deposited in the Brookhaven Protein Data Bank with PDB accession code 1RFO.

Denaturant-induced equilibrium transitions

Denaturant-induced equilibrium transitions were recorded in an AMINCO Bowman series 2 spectrofluorimeter (SLM Aminco, USA) and with an Aviv 62ADS spectropolarimeter (Aviv, USA). All transitions were measured at 10 mM sodium phosphate (pH 7.1 at 20 °C). For fluorescence measurements at concentrations of total monomer ([M]0) of 5 µM and 30 µM, the excitation wavelengths were 278 nm and 298 nm, respectively, at 2 nm bandwidth. Emission was recorded at 320 nm (2 nm bandwidth). CD measurements were performed at 228 nm with 0.5 cm (5 µM) or 0.1 cm (30 µM) path-lengths.

Equilibrium transition curves were analyzed assuming a two-state transition from native trimer (N) to unfolded monomer (U):

\[
N = 3U
\]

The resulting equilibrium constant is given by:

\[
K_{eq} = \frac{[U]^3}{[N]^3}
\]

and the total monomer concentration ([M]0) can be expressed as:

\[
[M]_0 = [U] + 3[N]
\]

The fractions of monomers in the unfolded state (fU) and in the native state (fN) are given by:

\[
f_U = \frac{[U]}{[U] + 3[N]} = \frac{[U]}{[M]_0}
\]

\[
f_N = \frac{3[N]}{[U] + 3[N]} = \frac{3[N]}{[M]_0} = 1 - f_U
\]

Thus, equation (6) becomes:

\[
K_{eq} = \frac{[U]^3}{[N]^3} = \frac{3[N]^2[M]_0^2}{1 - f_U}
\]

To fit the equilibrium transitions to equation (9) we expressed the fraction of unfolded protein at a given denaturant concentration (x) as:

\[
f_U(x) = \frac{S_U(x) - S(x)}{(S_N(x) - S(x)) + (S(x) - S_U(x))}
\]

where S(x) corresponds to the measured signal at the given denaturant concentration, x. S_N and S_U represent the spectroscopic signals of the native and unfolded state, respectively, obtained from linear extrapolation of the baselines according to:

\[
S_N(x) = S_N(H_2O) + m_Nx
\]

\[
S_U(x) = S_U(H_2O) + m_Ux
\]
Using:

\[ \Delta G_0 = -RT \ln K_{eq} \]  

and a linear denaturant-dependence of \( \Delta G_0 \) we can obtain the free energy for unfolding in water from fitting the transition curve to equations (9)–(13) in a single step, similar to the procedure described by Santoro & Bolen for a two-state transition (9)–(13). This yields the following equation for a two-state unfolding transition from a native trimer to unfolded monomers:

\[
S(x) = S_U(x) - \frac{S_N(x) - S_U(x)}{3[M]_0} 
\times \left( e^{-\frac{-\Delta G_0}{R T}} \left( \frac{9}{2} [M]_0 + \sqrt{D} \right) 
+ e^{-\frac{-\Delta G_0}{R T}} \left( \frac{9}{2} [M]_0 - \sqrt{D} \right) \right)
\]

with:

\[ D = \left( \frac{9}{2} [M]_0^2 e^{-\frac{-\Delta G_0}{R T}} + e^{-\frac{-\Delta G_0}{R T}} \right) \]

Equations (14a) and (14b) were used to fit the individual transition curves monitored by fluorescence or CD and to globally fit all data obtained by measuring various spectroscopic probes at different concentrations of total monomer ([M]_0).

**Kinetic experiments**

All stopped-flow experiments were performed with a SX18.1MV stopped-flow instrument from Applied Photophysics (Leatherhead, UK) equipped with a Hamamatsu R1104 photomultiplier tube for single-wavelength kinetics and a Hamamatsu R6095 photomultiplier tube for all other measurements. For experiments using a cut-off filter (\( \sim 320 \) nm), the excitation wavelength was 278 nm (2 nm bandwidth). For single-wavelength refolding experiments, the excitation wavelength was 280 nm (4.5 nm bandwidth) for final concentrations of protein up to 10 \( \mu M \) and 295 nm (4.5 nm bandwidth) for higher concentrations. The emission bandwidth was 12 nm. For refolding experiments, the protein was allowed to unfold for at least ten hours in 6.38 M GdmCl, 10 mM sodium phosphate (pH 7.1) before the measurements. At final concentrations of protein above 10 \( \mu M \) monomer, the protein was unfolded at the same concentration of denaturant at pH 2. Unless stated otherwise, all measurements were performed in 10 mM sodium phosphate (pH 7.1 at 20°C). Refolding was initiated by 11-fold dilution to a final concentration of GdmCl of 0.58 M. The fluorescence intensity of the unfolded state under refolding conditions (Figure 3) was determined by extrapolating the fluorescence signal of the unfolded foldon domain (measured in the stopped-flow instrument) to 0.58 M GdmCl. The effect of human cyclophilin 18 on the slow-refolding reaction was measured by manual mixing fluorescence measurements in 10 mM sodium cacodylate (pH 7.1) at [M]_0 = 5 \( \mu M \) foldon domain in the absence and in the presence of 3.3 \( \mu M \) cyclophilin. The use of sodium cacodylate instead of sodium phosphate had no effect on the folding kinetics.

Stopped-flow interrupted refolding experiments were used to monitor the formation of native molecules during refolding in 0.58 M GdmCl, 20 mM sodium phosphate (pH 7.0) at 20.0°C. Completely unfolded foldon (in 3.4 M GdmCl, 20 mM sodium phosphate (pH 1.7), [M]_0 = 60 \( \mu M \) was diluted sixfold into final conditions of 10 \( \mu M \) protein, 0.58 M GdmCl, 20 mM sodium phosphate (pH 7.0) to initiate refolding. After various times (\( t_i \)), refolding was interrupted by transferring the solution into final conditions of 6.7 M GdmCl, 20 mM sodium phosphate (pH 7.0), final protein concentration 1.7 \( \mu M \). Native foldon unfolds with double-exponential kinetics under these conditions (S. Gü. & T.K., unpublished results) with \( t_1 \approx 7.8 \pm 0.5 \) seconds (80% amplitude) and \( t_2 \approx 0.24 \pm 0.02 \) seconds (20% amplitude). The relative amplitudes of the two reactions are independent of the refolding time. The amplitude of the major, slow-unfolding reaction (\( t_1 = 7.8 \pm 0.5 \) seconds) was used as a measure for the amount of native protein that was present after the time \( t_i \), when refolding was interrupted. The observed unfolding amplitudes after various times of refolding were normalized against the amplitude of completely refolded foldon to yield the fraction of native molecules that were present after \( t_i \).

**Data fitting and simulations**

Data evaluation was carried out using the programs ProFit (Quantumsoft, Zurich, Switzerland) and Matlab (The MathWorks, Natick, MA, USA). Interrupted refolding experiments at [M]_0 = 10 \( \mu M \) and ten direct fluorescence-detected refolding traces with [M]_0 ranging from 0.5 \( \mu M \) to 102 \( \mu M \) were analyzed globally by nonlinear, least-squares curve fitting. The experimental data were fit to the numerical solution of the kinetic scheme depicted in equation (4). Rate constants and relative signal amplitudes of the different kinetic species were fitted as global parameters. The equilibrium constant between native and unfolded protein determined by a global fit of the equilibrium unfolding transitions measured by fluorescence and CD at various concentrations of protein (Figure 2) was used as an additional constraint for the fit. To ensure that the fit converged to the global minimum, it was repeated 60 times with randomly chosen starting values for the fitting parameters.

**References**


5. Seckler, R. (2000). Assembly of multi-subunit...

*Edited by F. Schmid*