Mechanisms of Evolutionary Optimized Complex Folding Reactions

INAUGURALDISSERTATION

zur

Erlangung der Würde eines Doktors der Philosophie.

Vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Sarah Güthe

aus Deutschland

Basel, 2005

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Thomas Kiefhaber

Prof. Dr. Joachim Seelig

Basel, den 05. Juli 2005

Prof. Dr. Hans-Jakob Wirz (Dekan)

Abbreviations

aa	amino acid
A-state	acidic state
ASA	accessible surface area
CD	circular dichroism
δ	chemical shift
ER	endoplasmatic reticulum
F	fluorescence intensity
FRET	fluorescence resonance energy transfer
λ	observable rate constant; *also:* wavelength
GdmCl	guanidinium chloride
I	intermediate state of a protein folding reaction
N	native state of a protein
NMR	nuclear magnetic resonance
ppm	parts per million
RDC	residual dipolar coupling
RT	room temperature (25ºC, 298.15 K)
SFVP	Semliki Forest Virus Protease
SVD	singular value decomposition
U	unfolded state of a protein
wl	wavelength
wt	wild-type

The common one- and three- letter abbreviations for amino acids are used.
Concentrations are denoted in square brackets, and kinetic species by capital letters.
The foldon concentrations are always given in concentration of monomers.

Contents

1 Introduction

1.1 Proteins

Proteins are essential for all processes in living systems. Due to their structural diversity they are able to perform numerous different tasks: they are involved in metabolic processes or accomplish structural functions, catalyze chemical reactions or store substances. Various proteins transport different materials within and between cells, participate in signal transduction or have a function in the immune system.

Proteins are mainly created out of 20 different naturally occurring α-L-amino acids, which are joined by amide bonds to a one-dimensional chain whose sequence is specific for each protein (primary structure). The primary structure is encoded by the sequence of the four nucleotides in the genome. Genomic DNA is transcribed into mRNA, which is then translated into proteins at ribosomes. The term “secondary structure” is used for arrangements with a defined conformation of the peptide backbone such as α-helices, β-pleated sheets and hairpins. The three dimensional arrangement of the secondary structure is called tertiary structure, where also amino acids that are further apart in sequence come close together in space. It can be stabilised by salt bridges, hydrophobic interactions, hydrogen bonds and disulfide bonds. The expression quaternary structure is used when separate polypeptide chains form a multimeric protein stabilised by hydrogen bonds, hydrophobic interactions and ionic binding.

In this thesis, the formation of secondary structure elements and the acquisition of tertiary as well as quaternary structure is studied.

1.2 Equilibrium unfolding of proteins with denaturants

The stability of a native protein towards its unfolded state is usually very low, around 40 kJ/mol. This low stability has physiological reasons: it facilitates the flexibility of the protein, transport through membranes via partial unfolding, it avoids kinetic traps and us essential for easy digestion.[1] Several factors contribute to the stability of a native protein. Intramolecular interactions such as salt bridges and hydrogen bonds stabilise the native state, whereas it is destabilised by loss of chain entropy, as the conformations are restricted. These two opposing effects have roughly the same size. The reason why the native state is nevertheless more stable than the unfolded one can be explained by the hydrophobic effect: major contributions to protein stability arise from the interactions of water molecules with the chain. By adding up all the negative and positive contributions, a small netto stabilising effect results.

The measured value for stability is the free energy (G). In The Gibbs-Helmholtz equation, G is connected with entropy (S) and enthalpy (H):

$$\Delta G^0 = \Delta H^0 - T \cdot \Delta S^0 \qquad \text{(eq 1)}$$

For most proteins, a cooperative two-state transition can be observed with a pretransitional and a posttransitional baseline and a more or less steep transition region.

In the transition region, the logarithm of the equilibrium constant *K* between N and U varies linearly with the denaturant concentration, so that with

$$\Delta G^0 = -RT \ln K \qquad \text{(eq 2)}$$

the conformational free energy ΔG(H2O) can be determined by extrapolation to 0 M denaturant. This so-called linear extrapolation model was introduced by Greene and Pace[2].

Santoro and Bolen[3] introduced a single non-linear least square fit including the pre- and the posttransitional baseline, so that the complete transition could be fitted in one step.

The slope of the linear extrapolation, i.e. the change of ΔG with denaturant concentration, is called m-value (m_{eq}).[4] This linear dependence on denaturant concentration reflects the increase in solvent accessible surface area.

Sometimes, more than two states are populated in equilibrium. These thermodynamic intermediates can usually be induced by solvent conditions, such as low denaturant concentrations,[5] low pH[6; 7] or by alcohols such as trifluoroethanol (TFE).[8] Often intermediates are compact, and are thus called molten globules.[9] They contain a lot of secondary structure but almost no tertiary structure, and their radius of gyration is only 10-30% larger than that of the native protein. Their hydrophobic core is loosely packed, and the structure fluctuates on a timescale slower than nanoseconds. For equilibrium unfolding, bovine α-lactalbumin, cytochrome c and apomyoglobin belong to the best described proteins with molten globule intermediates.

In the native structure, the dihedral angles of the peptide backbone and the side chains are defined, but not rigid. In the unfolded protein on the other hand, many dihedral angles can be adopted, and conformations interchange rapidly, as they all have similar free energies. This unfolded structure is therefore not defined, and also depends on the chosen denaturation conditions. For heat and acid denaturation, residual structure might still be present.[10] High concentrations of chaotropic substances (such as urea and guanidinium chloride), on the contrary, destroy the water structure, which usually leads to completely unfolded proteins in random coil states.[11] There are exceptions: Neri et al.[12] were the first to show that in some proteins local structures still exist at high denaturant concentrations, and on the other hand, completely unfolded proteins are not necessarily in a random coil state, as Shortle[13] has shown for staphylococcal nuclease.

To investigate the stability of a protein, it is convenient to shift the equilibrium between the native and the unfolded protein so that the two states can be measured simultaneously. In addition to adding increasing amounts of denaturant, such as urea or guanidinium chloride

(GdmCl), proteins can also be unfolded by heating, cooling down, or by a shift in pH: at basic or acidic pH, side chains are protonated or deprotonated, leading to a disruption of salt bridges and electrostatic repulsion of alike charged groups. The experimentally determined stabilities for a protein usually coincide very good if different denaturation methods are used.

Unfolding with chemical denaturants is the most commonly used method and has long been known.[14-16] In 1964, Tanford was the first to investigate quantitative unfolding in urea.[17] Urea can form many H-bonds, so it is a good solvent for the peptide group. In unfolded proteins, however, also unpolar groups are exposed to the solvent, and the protein usually adopts a more extended state with a larger solvent accessible surface area (ASA).

Timasheff [18] and Record [19] suggested that urea prefers binding to proteins over binding to water. The denatured state with a larger surface and thus more binding sites for urea is thus preferred by urea compared to the compact native state. The excluded volume effect of co-solvents, on the other hand, favours the native state as the protein is bigger than the solvent water. The co-solvent acts as a denaturant if contact interactions are larger than the excluded volume effect, else it is an osmolyte and stabilises the native protein.[20] As direct interaction of urea with the protein is very weak, solvent exchange mechanism in addition to denaturant binding must also play a role.[21] Still, the dominant contribution of chemical denaturants to protein unfolding is preferential binding of newly exposed groups so that the protein unfolds with increasing urea or GdmCl concentration.

1.3 Protein folding

Inside the cell, the newly synthesised polypeptide chain adopts a specific three-dimensional structure, which is essential for the performance of the protein. This process is referred to as protein folding. As shown by Anfinsen and coworkers,[22; 23] the amino acid sequence contains the whole information necessary for the three-dimensional structure. Chemically unfolded ribonuclease A with reduced disulfide bonds regains functionality after removal of denaturants. This demonstrated that the protein must have reverted to its native structure (function follows form). This experiment proved that the information for structure and function is encoded within the protein itself.

1.3.1 Models in protein folding

The way how a protein finds its proper structure, and how the structure of a protein is encoded in the sequence of amino acids, have been major topics for scientists. It is not yet possible to predict protein structures; even proteins with the same fold can have only little sequence similarity. So far, the folding code has not been deciphered. Despite of advances in computational structure prediction, the atomic structure of a protein must still be determined experimentally, mainly by NMR-spectroscopy and X-ray crystallography.

As Levinthal argued[24; 25], a protein could never find its native structure just by random search of all the possible conformations, as this would take about 10^{27} years for a protein of 100 aa length. This became known as "Levinthal's paradox". Therefore it was assumed that intermediates which guide the protein folding process have to exist. Several models have been put forward to describe the folding process. They try to answer the general questions in protein folding: how does a protein find its native conformation on a biological timescale (ms to s),

how are losses in conformational entropy compensated, and how are meta-stable traps on the way to the native state avoided?

The framework model assumes that local elements of local native secondary structure forms independently of tertiary structure.[26; 27] These elements would diffuse, collide and adhere until the native structure is completed (also known as diffusion collision model[28; 29]).

In the nucleation model, neighbouring residues form native secondary structure to act as a nucleus from which native structure can propagate. [30; 31] A structural nucleus might also serve as a kinetic nucleus for the folding process, so that in successive folding events no intermediates are populated.

The hydrophobic collapse model[32] describes a protein collapsing rapidly around its hydrophobic side chains. In this intermediate, the side chains would then rearrange where the secondary structure would be directed by native-like tertiary interactions. Protein folding via a (molten globule) intermediate is also referred to as two-step hierarchical model: U → I → N.[33] Whether the intermediate can be observed, i.e. whether folding is two-state, depends on energy of the intermediate relative to U and N, and on the rate-limiting step.

An alternative explanation for rapid folding is illustrated by a funnel-like[34; 35] energy landscape with a small energy bias towards the native state. It describes the progress of a population of unfolded protein, advantageously high in entropy, but at the same time disadvantageously high in energy, that traverses its energy landscape towards the native state, which has unfavourable low entropy but favourable low energy. The sloping funnel walls guide the protein folding process, and parallel folding processes can occur. Intermediates are not necessary for the folding process, but they can be included in the funnel model. The loss of entropy upon folding is compensated by favourable pairwise interactions. [36]

1.3.2 Experimental data on protein folding

To investigate protein folding experimentally, often small monomeric single domain proteins are used as model systems. In many cases, only two states are observed in folding experiments,[37] and the accumulation of a folding intermediate is not a prerequisite for successful folding.[38-40] Chymotrypsin inhibitor 2 (CI2), e.g., folds rapidly via two-state kinetics (apart from 20-30% slow folding proteins with one or several prolyl residues in the *cis* conformation). CI2 has a single hydrophobic core formed by residues all over the protein, and secondary structure elements that are unstable in solution, so that folding proceeds in a cooperative and concerted step. Intermediates can be hidden if the they are higher in energy than both N and U at all denaturant concentrations, as shown for tendamistat,[41] a small all-ß protein. A variant exists which exhibits a nonlinear unfolding limb in the chevron plot. A chevron plot describes the dependence of the observed rate constants on denaturant concentration (see Figure 1.1 B, p. 19 for an example). At high and low denaturant concentrations, the chevron plot for this tendamistat variant is linear. This indicates a denaturant-induced switch between two distinct transition states. As a consequence, a model with a sequential folding mechanism can be used to describe the data. Tendamistat folds via consecutive transition states and a metastable high energy intermediate. In this case, the intermediate is on-pathway, if it is higher in energy than both the unfolded and the native state.

An indication for intermediates in the folding process are the observation of additional phases, deviation from linearity of the logarithm of the rate constant versus denaturant concentration, discrepancy in ΔG^0 (H_2O) obtained from kinetic and from equilibrium measurements and dissimilarities when using different spectroscopic probes (Trp fluorescence monitors tertiary and / or quaternary structure: surroundings of the Trp residue, far UV CD monitors secondary structure). In intermediates, some secondary structure is already formed. Kinetic

intermediates can only be detected if they are formed before the rate-limiting step and if their free energy is similar to or lower than that of the denatured state. There are two kinds of intermediates, a) ones that are more stable than U, but after rate-limiting transition state located between U and I, and b) ones that are less stable than U. Kinetically observed molten globule states in refolding can be detected by rapid mixing techniques monitoring CD, NMR, or small angle x-ray scattering signals. Cytochrome c, RNase A, lysozyme and apomyoglobin, for example, fold via well characterised molten globule intermediates. Very often burst-phase intermediate can be observed, that have accumulated during stopped-flow ms time resolution. Faster relaxation methods such as t-jump or continuous flow can be used to increase the time resolution.

The very first steps in folding include the hydrophobic collapse,[42] where hydrophobic residues are buried in the interior of the protein, and also the formation of native contacts. Early folding intermediates can for example be detected by protection of amide protons against exchange.[43] [44] For staphylococcal nuclease,[45] pulsed H-exchange experiments have shown that even after 10 ms refolding time a part of the amide protons were already protected, indicating an early folding intermediate. Isolated secondary structure elements have also been shown to exist in water: stable β-hairpins can form in water,[46] as well as reverse turns in short peptides.[47] α-Helices based on Ala peptides are stable in water,[48-52] too, hence the helix backbone itself must also be stable in water. Therefore, all classes of secondary structure can be present at very early stages of the folding process and can thus guide folding.

A recent investigation on 23 reported two-state proteins by Sánchez and Kiefhaber has shown that nonlinear activation free-energy relationships are caused by sequential folding pathways with consecutive distinct barriers and a few obligatory intermediates that are hidden from direct observation by the high free energies of the intermediates.

Complex folding kinetics have long been known for proteins such as cytochrome c, hen egg white lysozyme,[53] and RNase A.[54] Refolding and unfolding kinetics could be explained by the existence of intermediates. In RNase A, e.g., two phases lead to native protein,[55; 56] a fast phase (20% of the molecules) and a fifty times slower phase (80% of the molecules). The slow folding step is due to proline isomerisation, and was analysed in detail for RNase A and RNase T1.[57; 58] If the protein contains prolines in a non-native isomer, the isomerisation to form the correct one can take minutes to hours, as the activation energy is quite high (80 kJ/mol). Recent work has shown that also the *cis/trans* isomerisation of non-prolyl peptide bonds can be a rate-limiting step.[59]

A further slow step in protein folding is the correct formation of disulfide bonds, which takes place on the minute to hour time-scale. Wrong disulfide bonds may form as the native pattern is often one of many combinatorial possibilities. These intermediates can be trapped, as was analysed in detail by Creighton's work[60] on BPTI (bovine pancreatic trypsin inhibitor). Other slow steps include the addition of prosthetic groups such as the heme group in cytochrome c. These slow steps take place when the structure is already very close to that of the native state N.[61-65]

However, enzymes have evolved to catalyse these slow steps in protein folding: PPIase (peptidyl-prolyl-*cis/trans* isomerase)[66] is able to catalyse the peptidyl-prolyl bond *cis/trans* isomerisation, PDI, protein disulfide isomerase,[67-69] helps to isomerise disulfide bridges. These enzymes were shown to speed up folding *in vitro* when the substrate proteins contain accessible disulfide bonds or prolyl residues, respectively. An additional proteins that assist folding are the so-called molecular chaperones, or heat-shock proteins,[70] which prevent aggregation and assist the proper folding of other proteins inside the cell.

In summary, one can say that for large proteins, such as lysozyme, the folding process becomes more effective if intermediates are populated, whereas small proteins approximate

two-state folding with high energy intermediates to avoid aggregation of populated intermediates.

1.3.3 Folding of oligomers and multi-domain proteins

To date, most studies have dealt with small, monomeric single domain proteins. Less is known about folding and assembly pathways of oligomeric proteins[71] or folding of multi-domain proteins.[72]. In the 1970s, Jaenicke[37; 73] started to investigate folding coupled to subunit association in the concentration dependent folding reactions of oligomeric proteins. Studies on trimers or higher oligomers investigated large filamentous proteins, which show extremely slow complex folding kinetics competing with irreversible aggregation reactions.[74] They often assemble via native-like monomeric intermediates.

The upper limit for a bimolecular reaction is the diffusion limit, which was estimated 10^9 $M^{-1}s^{-1}$ for monomer subunits the size of an average protein domain[75]. For most oligomeric proteins investigated, the fastest bimolecular steps have rates in the range of 10^3–10^6 $M^{-1}s^{-1}$. Among the fastest folding dimeric proteins are the wild-type Arc repressor [75] and some designed leucine zippers,[76] with association rates constant around of $6 \cdot 10^6$ $M^{-1}s^{-1}$. The record is held by an engineered fragment of trp repressor,[77] it is $3 \cdot 10^8$ $M^{-1}s^{-1}$. In this case, rate enhancement was achieved by replacing the intermolecular salt bridge and hydrogen bonding network in the wild-type by hydrophobic residues.[78] In these cases, however, folding is tightly coupled to association, so that partially folded dimeric intermediates are involved in assembly, whereas large slow-folding oligomeric proteins usually form folded monomeric intermediates first before assembly takes place.

Another complexity in folding is introduced by the presence of more than one domain in a protein. Domains are still connected so that the protein is made up from one chain.

Most of the multi-domain proteins investigated fold rather slow. Generally, domain docking is the final, slow step. Even two-domain proteins such as the α-subunit of trp synthase[79] and phosphoglycerate kinase[80], which are among the fastest folding two-domain proteins, fold only with a time constant of about 50 s. Gene-3 product from phage fd[81] folds even slower. The rate-limiting reaction is a *cis/trans* isomerisation at a prolyl residue (τ = 6200 s) which regulates the final domain docking step.

These large proteins, consisting of several domains or subunits, usually fold very slow, and the native state can often only be reached with the assistance of molecular chaperones. In this study, two new model proteins for the folding of oligomeric and multi-domain proteins were investigated.

1.4 Protein folding kinetics

In the simplest case, protein folding can be described as a two-state process, without any observable intermediates:

$$U \underset{k_u}{\overset{k_f}{\rightleftharpoons}} N \qquad \text{(eq 3)}$$

There is a single observable rate constant, λ, that is the sum of the microscopic rate constants for folding (k_f) and unfolding (k_u), respectively:

$$\lambda = k_f + k_u \qquad \text{(eq 4)}$$

In equilibrium, $k_f \cdot [U]_{eq}$ equals $k_u \cdot [N]_{eq}$, and the equilibrium constant (K) is as follows:

$$K = \frac{[N_{eq}]}{[U_{eq}]} = \frac{k_f}{k_u} \qquad \text{(eq 5)}$$

The free energy of folding ΔG^0 can thus be derived from the folding kinetics, as well from equilibrium measurements. This can be used to check whether the right model is used to describe the data.

$$\Delta G^0 = -RT\ln K = -RT\ln\left(\frac{[N]_{eq}}{[U]_{eq}}\right) = -RT\ln\left(\frac{k_f}{k_u}\right) \qquad \text{(eq 6)}$$

To analyse folding, the transition state theory[82] is widely used. The transition-state theory focuses on the entire protein population. A broad ensemble of transitions state (‡) species is located on top of energy barrier between U (or I) and N (or I).

The rate constant k is connected to the free energy of activation $\Delta G^{0‡}$ for forming the transition state (‡).

$$k = \kappa\frac{k_B T}{h}e^{-\frac{\Delta G^{0‡}}{RT}} \qquad \text{(eq 7)}$$

k_B is the Boltzmann constant, h the Planck constant, k is a transmission factor with an upper limit of 1. The pre-exponential factor corresponds to the maximum rate of the reaction in the absence of free energy barriers. In protein folding, κ is usually assumed to be 1, so that

$$k = \frac{k_B T}{h}e^{-\frac{\Delta G^{0‡}}{RT}} \qquad \text{(eq 8)}$$

The transition state can also be assigned an activation free energy for folding and for unfolding ($\Delta G_f^{0‡}$ and $\Delta G_u^{0‡}$). The change of $\Delta G_f^{0‡}$ with denaturant concentration, the kinetic m-values (m_f, m_u), is thought to reflect the change in solvent accessible surface area of the transition state compared to the unfolded state (m_u) and the native state, respectively (m_f), like the equilibrium m-value reflects the change in ASA upon unfolding. Thus, the transition state can be characterised.

The activation free energy depends linearly on the concentration of denaturant:

$$\Delta G_u^{0‡} = \Delta G_u^{0‡}(H_2O) - m_u^{‡}\cdot[denaturant] \qquad \text{(eq 9a)}$$

$$\Delta G_f^{0‡} = \Delta G_f^{0‡}(H_2O) - m_f^{‡}\cdot[denaturant] \qquad \text{(eq9b)}$$

$$\Delta G^0 = \Delta G^0(H_2O) - m_{eq} \cdot [denaturant] \quad \text{(eq 9c)}$$

A plot of the logarithm of the observable rate constant versus denaturant concentration yields a V-shaped curve, which is also referred to as "chevron plot". With this analysis, k_f and k_u can be determined.

The total change in free energy is connected to $\Delta G_f^{0\ddagger}$ and $\Delta G_u^{0\ddagger}$. In the same way, $m_f^{\ddagger}$ and $m_u^{\ddagger}$ are related to m_{eq} as well:

$$\Delta G^0(H_2O) = \Delta G_f^{0\ddagger}(H_2O) + \Delta G_u^{0\ddagger}(H_2O) \quad \text{(eq 10)}$$

$$m_{eq} = m_f^{\ddagger} - m_u^{\ddagger} \quad \text{(eq 11)}$$

m_{eq} correlates with the change in solvent-accessible surface area upon unfolding,[56] and $m_f^{\ddagger}$ and $m_u^{\ddagger}$ as well. $a = m_f^{\ddagger} / m_{eq}$ describes how close the transition state is to the native state.

Nonlinearities in the chevron plot at low and/or high concentrations of denaturant occur when the rate-limiting step changes, and hint at the population of intermediates. The transient population of an intermediate leads to a change in the reaction mechanism and therefore to a deviation from linearity in the chevron plot, but a change in reaction order can also occur without the population of intermediates, e.g. when kinetic coupling takes place.[83]

If intermediates accumulate during folding or unfolding, multi-exponential kinetics are observed. The number of observable rate constants is determined by the number of kinetic species, including N and U. If *n* species exist, *n-1* rate constants are observed. and the number of intermediates is given by *n-2*.

To investigate the protein folding mechanism in detail, folding and unfolding kinetics at many different denaturant concentrations are measured. The number of exponentials needed to describe the kinetic traces accurately defines the number of kinetic species. It is also important to check for burst-phase reactions, which occur in the experimental deadtime. If the kinetic species and the rate constants are determined, this is not always enough to describe the

folding mechanism correctly. Proline *cis/trans* isomerisation, for example, also results in an observable rate constant, but is often a reason for parallel folding pathways. To test the mechanism, the formation of native molecules and refolding intermediates can be followed directly by sequential mixing experiments, as e.g. described by Schmid for RNase A.[84] For an N-test, (also called interrupted refolding,) completely unfolded protein is allowed to refold for a certain amount of time (t_i), and then diluted into unfolding conditions again. In this unfolding step, all intermediates and native molecules that have formed so far unfold with their own characteristic rate constant. The corresponding amplitudes correlate directly with the amount of corresponding species present when refolding was interrupted. By varying t_i, a time course of the population of N and I is obtained. With these interrupted refolding experiments it is possible to distinguish between sequential and parallel pathways.

To determine the time course of formation of U, to monitor the population of unfolding intermediates, and to detect slow equilibration reactions in the unfolded state, like non-prolyl *cis/trans* isomerisation, a similar experiment can be performed starting from native protein ("double jumps", introduced by Brandts[85]): In the first mixing step, the native protein is allowed to unfold for a certain amount of time t_i, where unfolded proteins and unfolding intermediates are formed. Refolding is initiated by dilution to native conditions in a second mixing step. The different species refold with their characteristic rate constant, and the amplitudes correspond to the amount of these species present when unfolding is interrupted, so that by varying t_i, the population of U and unfolding intermediates can be monitored. Slow phases originating from *trans* isomers, for example, would thus increase in amplitude with increasing t_i.

Sometimes folding reactions are faster than the time needed to mix the samples and start recording. These so-called deadtime reactions can be detected when complete unfolding and

refolding kinetics are measured (chevron plot in Figure 1-1 B) and the initial and final values of the kinetics are plotted versus the denaturant concentration (see Figure 1-1 C).

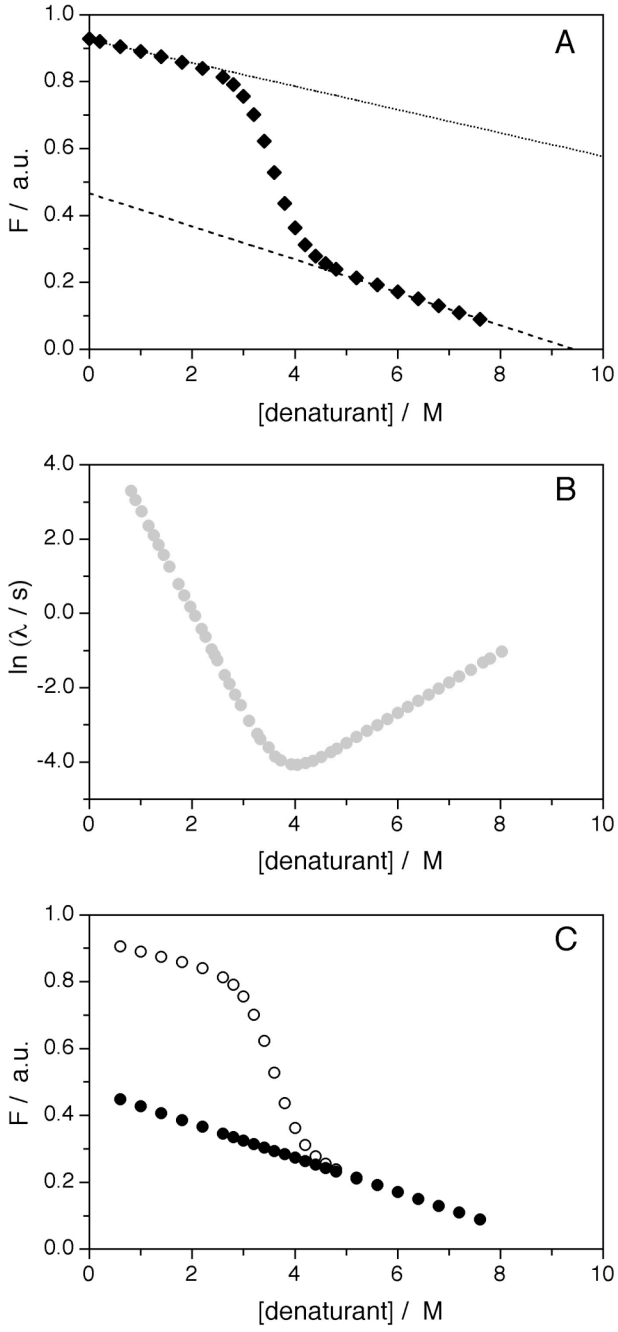


Figure 1-1

Hypothetical curve of the denaturant dependence of refolding kinetics and equilibrium stability. A: equilibrium transition with the native baseline (dotted line) and the unfolded baseline (dashed line). B: a chevron plot: unfolding and refolding kinetics. C: Initial (●) and end (○) points of refolding.

In the upper panel one can observe that the initial points (●) form a straight line representing the unfolded baseline of the corresponding equilibrium transition shown in Figure 1-1 A . The endpoints (○) mark the end of the kinetics, where equilibrium conditions prevail, thus they represent the equilibrium transition. If the reaction is not faster than the deadtime of stopped-flow mixing, which is about 1 to 3 ms, the whole amplitude of the kinetics can be resolved, and a plot of starting and end points looks like Figure 1-1 C. On the other hand, if refolding already starts during the mixing deadtime of about 1 ms, one cannot obtain the complete amplitude for the kinetic trace, and the initial points do not represent the native baseline. The larger the missing signal of the refolding kinetics is, the closer the initial points are to the end points. If an intermediate is formed in the deadtime, the initial points could show the equilibrium transition of the intermediate.

1.5 Foldon and SFVP as model proteins to study complex folding reactions

The proteins investigated in this thesis are both viral proteins optimised for folding. Foldon is a small homotrimer whose function is to promote association and folding of the 486 aa trimeric Fibritin. SFVP, the Semliki Forest Virus Protease, is a two-domain protein that has to fold co-translationally: it cleaves itself off the nascent chain and thus the rest of the polyprotein can be exported to the endoplasmatic reticulum (ER).

1.5.1 "Foldon" , a model for folding of a trimeric globular protein

“Foldon” is the C-terminal domain of Fibritin, a large trimeric phage protein. Fibritin is a 486 aa long coiled-coil surface protein of bacteriophage T4. It forms a fibre that is attached to

the collar of the phage, as can be seen in Figure 1-2. The parts of Fibritin whose X-ray structure have been solved by Tao et al.[86] and Boudko et al.[87] are pointed out.

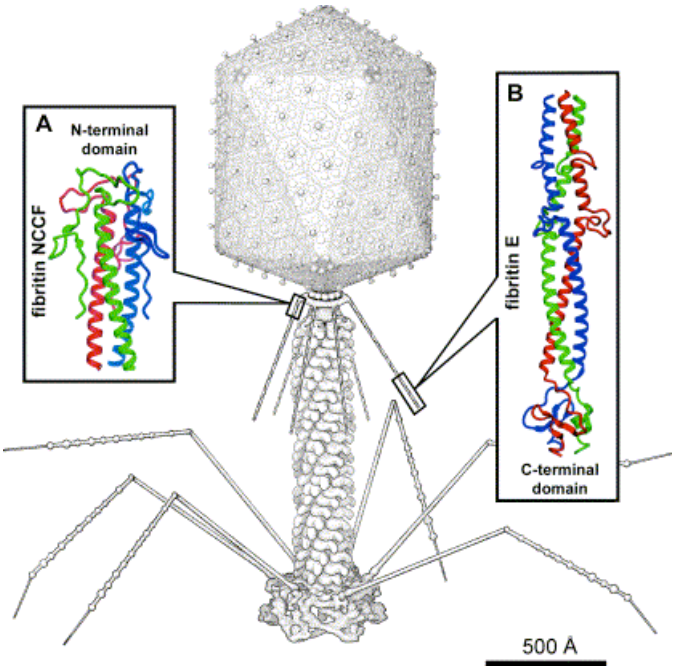


Figure 1-2

T4 phage with Fibritin pointed out. A: Enlarged ribbon diagram of the N-terminal part (residues 2–80) of fibritin NCCF. **B**: Ribbon diagram of fibritin E including residues 367–486 of the wild-type fibritin. The picture is taken from Boudko et al.[87]

The C-terminal domain comprises the last 30 residues (457–486) with the following sequence:

GYIPEAPRDG QAYVRKDGEW VLLSTFLSPA

The last three amino acids 484-486 were omitted in the foldon investigated, as they are not structured and therefore not visible in the X-ray structure. The residues were renumbered from 1-27. Twelve of these residues (12-23) form a β-hairpin. The three hairpins of the three subunits form a propeller-like structure, as shown in Figure 1-3. This symmetric trimer is stabilised by inter- as well as intersubunit salt bridges and hydrogen bonds.

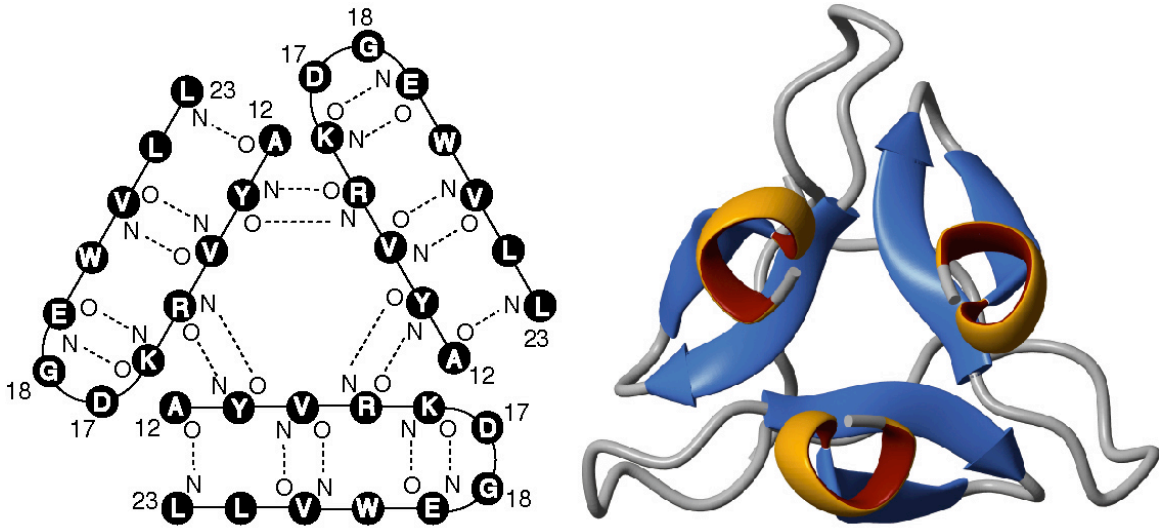


Figure 1-3

Structure of the foldon domain, left: scheme of hydrogen bonds within the hairpin, right: ribbon representation based on the solution structure. This figure is taken from section 4.1 and was created with MOLMOL [88])

In Fibritin, the C-terminal domain has the function of a folding domain; it is supposed to assist the assembly of Fibritin by trimerising very fast, thus aligning the three subunits for proper coiled-coil formation.[89] Hence the name it was given is foldon. The name "foldon" is also used for independent folding units, a term coined by Wolynes et al.[90; 91] It has been shown that the isolated trimerization domain is extremely stable, and that it promotes the assembly of engineered collagen-like fusion proteins,[92] as well as other trimeric proteins such as the envelope glycoprotein gp 140 from HIV-1[93]. To sum up, the function of foldon is to fold and associate fast and efficiently.

1.5.2 SFVP, a model for a two-domain protein

The second protein investigated in this study is the capsid protein of the Semliki Forest Virus. Like foldon, it has evolved to fold fast and efficiently. It is one of the few proteins for which co-translational folding has been shown.[94; 95]

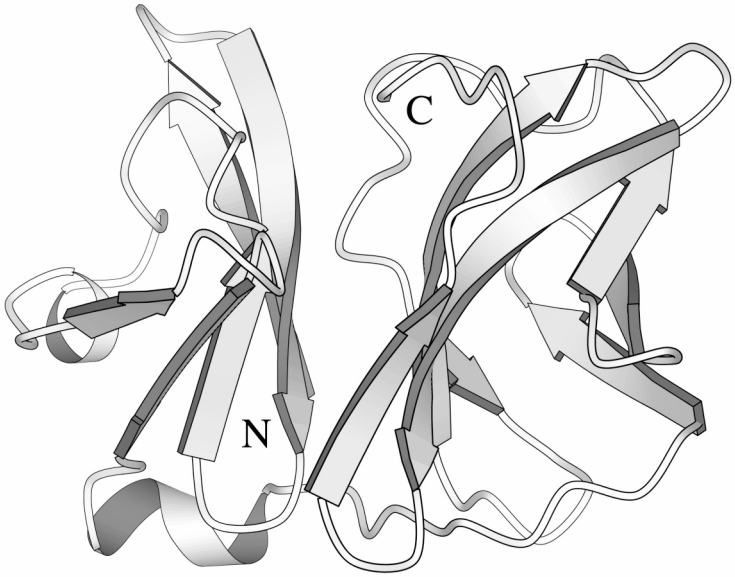


Figure 1-4

Ribbon representation of the X-ray structure of SFVP[96] (residues 119–267 of the viral polyprotein). Image created with MolScript.[97]

The capsid protein forms the N-terminal part (267 aa) of a large viral polyprotein containing structural proteins.[98; 99] The first 118 residues are positively charged and unstructured. They are important for binding to viral DNA and thus for capsid assembly. The second module (residues 119-267) is a serine protease (SFVP). The X-Ray structure[96] shows that the chymotrypsin-like fold making up the capsid shell is divided into two β-barrel domains (residues 119-182 and 183-267) containing a catalytic serine protease triad.[96] SFVP undergoes a single reaction turnover before it assembles into the capsid shell: it cleaves itself

off the nascent chain after Trp 267,[94; 100] which remains bound in the binding pocket and thus inhibits further proteolysis. By this cleavage a signal sequence at the new N-terminus is released, directing the polyprotein containing the envelope proteins to the endoplasmatic reticulum.[94] This self-cleavage has to occur during synthesis of the large polyprotein to ensure proper folding of the envelope proteins in the ER lumen. As a consequence, SFVP has to fold faster than the polyprotein is synthesised at the ribosome (3-5 residues per second in eukaryotic cells[101]). Sánchez *et al.*[102] investigated whether fast folding of SFVP is an intrinsic property of this two-domain protein, or whether spontaneous folding is slower than translation and requires additional catalysts. They also used SFVP as a model system for folding of two-domain proteins, as all two-domain proteins studied so far were reported to fold slowly *in vitro*, on the minutes to hours timescale.[79; 80; 103-105] The α-subunit of trp synthase[79] and phosphoglycerate kinase[80], which are among the fastest folding two-domain proteins, fold about 1000 times slower than the Semliki forest virus capsid protein SFVP (τ = 50 ms).

Equilibrium unfolding is two-state and completely reversible. They could show that ~ 60% of the proteins fold via a fast pathway with a time constant of 50 ms at 0 M denaturant. This corresponds to molecules with all seven prolyl residues in the native *trans* orientation. Three slower reactions are observed that are *cis/trans* isomerisation reactions of Xaa-Pro and Xaa-non-Pro peptide bonds. The two domains are probably formed sequentially, as shown by a lag phase in the formation of native molecules, and a rollover in the chevron plot. The lag phase in the N-test can be explained as the first formed N-terminal domain does not contain a Trp residue and thus does not contribute to the fluorescence signal.

In this thesis, a mutant of SFVP was investigated, where Phe160 was replaced by a Trp, in order to obtain an additional fluorescence signal in the N-terminal domain.

2 Aims of Research

The aim of this study was to examine complex folding reactions. The main questions addressed are: how fast can complex folding occur, and what are the mechanisms of evolutionary optimised complex folding reactions. Foldon and SFVP are very well suited model proteins since they have evolved under selective pressure for fast and efficient folding and association.

All previous studies on trimers or higher oligomers characterised large proteins that exhibit a complex and slow folding mechanism, often accompanied by aggregation. The foldon domain of T4 phage fibritin is a perfect model system for the investigation of the folding kinetics and the thermodynamics of a trimeric globular protein, since it is small, its 3D structure is known, as determined by x-ray crystallography, and folding and unfolding are completely reversible. SFVP, the protease from Semliki Forest virus, is a two-domain protein, and one of the few proteins for which co-translational folding was demonstrated. In order to fulfil its biological function, SFVP has to fold faster than synthesis of the rest of the polyprotein . Previous studies have shown that it folds fast in the absence of other proteins such as molecular chaperones. The two domains are probably formed sequentially. As there is no fluorescence probe in the N-terminal domain, the observed lag phase in the formation of native molecules indicates that the N-terminal domain, which is synthesised first at the ribosomes, also folds first.

In this thesis, the influence of guanidinium chloride on the wild-type protein, and the folding and stability of the mutant SFVP F160W is characterised, which has an additional fluorescence probe in the N-terminal domain.

3 Summary of Published Results

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

Sarah Güthe, Larisa Kapinos, Andreas Möglich, Sebastian Meier, Stephan Grzesiek and Thomas Kiefhaber.

J. Mol. Biol. (2004) 337, 905–915

The isolated foldon domain forms a homotrimer under native conditions, whose stability is concentration dependent. It acquires the same trimeric β-propeller structure as in fibritin, and unfolds in a single step from folded trimer to unfolded monomer.

GdmCl- induced equilibrium transitions at several protein concentrations, monitored by CD and fluorescence spectroscopy, showed two-state behaviour. A global fit yielded a stability of $\Delta G^0(H_2O)$ = 89.2±0.6 kJ/mol corresponding a protein concentration of 1 M. At a physiological protein concentration of 5 μM, however, ΔG is 29.7 kJ/mol. The change in free energy with GdmCl is m_{eq} = -10.4 (kJ/mol)/M, which is expected for globular monomeric proteins the size of the foldon trimer.[13]

Folding occurs in several consecutive steps. On the submillisecond time scale, a burst phase intermediate is observed, where structure is formed in the hairpin region of the monomer, followed by two subsequent association steps. This refolding intermediate must be monomeric, as the burst phase reaction occurs at all concentrations between 0.5 μM and 100 μM, whereas a bimolecular step would become observable at low protein concentrations.

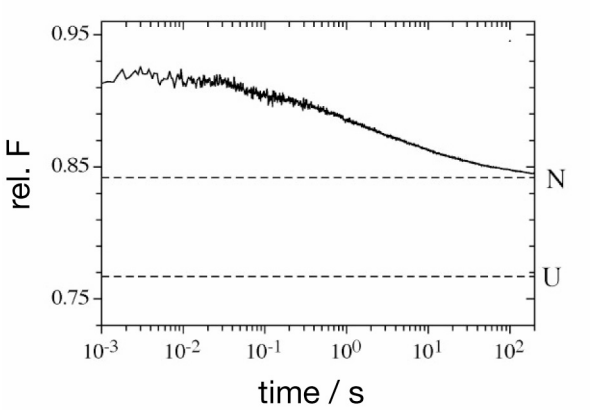


Figure 3-1

Refolding of the foldon domain in 0.58 M GdmCl, pH 7.1, protein concentration: 5 µM. Measured by the change in Trp fluorescence using a 320 nm emission cut-off filter. The broken lines represent the signals of the native and of the unfolded state at 0.58 M GdmCl, as indicated. The signal of the unfolded state is extrapolated from the unfolded baseline at high concentrations of GdmCl to 0.58 M GdmCl.

With the help of wavelength-dependent refolding experiments the fluorescence emission spectrum of this intermediate could be resolved. Refolding kinetics at 0.58 M GdmCl and 5 µM protein concentration were resolved by stopped-flow mixing at different fluorescence emission wavelengths. With the same method, the spectrum of the native protein at 0.58 M GdmCl and the unfolded spectrum at 6.4 M GdmCl as well as the corresponding buffer spectra were resolved. The native and the unfolded foldon spectra were then measured with the same settings on a fluorescence spectrometer, so that a correction factor for the stopped-flow fluorescence detector could be determined. From the starting points of each refolding trace, the fluorescence emission spectrum of the burst phase intermediate could be determined, as shown in Figure 3-2 A, which compares the intermediate spectrum to the fluorescence spectra of the native foldon at 0.58 M GdmCl, the GdmCl unfolded foldon, and to the A-state (at pH 2). Figure 3-2 B shows an overlay of the A-state foldon and the

intermediate, normalised to the fluorescence maximum. It has the same shape as foldon at pH 2 (A-state), indicating that the intermediate is probably as compact as the A-state.

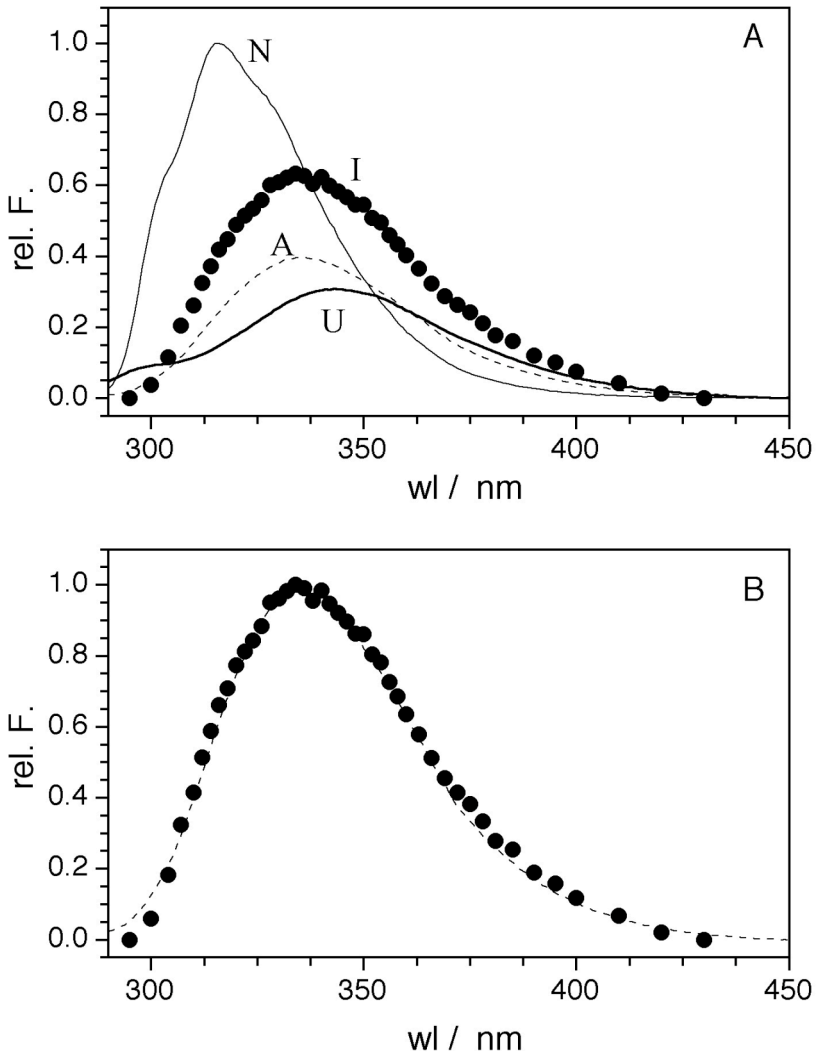


Figure 3-2

Fluorescence emission spectrum of foldon at different conditions. A: native foldon at 0.58 M GdmCl (thin solid line, N), GdmCl-unfolded foldon (8.2 M GdmCl, thick solid line, U), burst phase intermediate (closed circles, I), and the acid unfolded A-state (dashed line, A), all normalised relative to the native state. B: foldon at pH 2 (dashed line) and of the refolding intermediate (closed circles), Both spectra are normalised to the emission maximum.

Refolding kinetics are concentration dependent, and were investigated at several protein concentrations with stopped-flow and manual mixing techniques. The reaction order at low protein concentration (0.5-4 μM) approaches 3, and at high protein concentrations (≥ 200 μM), refolding becomes virtually concentration independent. To determine the formation of native molecules directly, interrupted refolding experiments were performed. The refolding kinetics, the time-course of formation of N and the equilibrium transitions at different protein concentrations were fitted globally. Thus, the rate constants for a minimal model for the mechanism of foldon folding could be obtained.

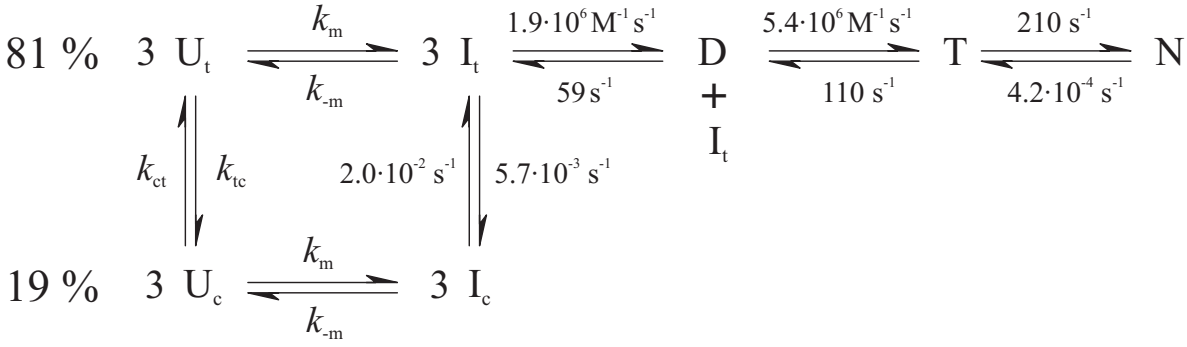


Figure 3-3

Minimal model for foldon folding, depicting the association and folding steps. The rate constants at 0.58 M GdmCl were obtained by a global fit using Matlab. This scheme is taken from Güthe et al.[106]

The rate constants for the two bimolecular association steps are $1.9(\pm0.5)\cdot10^6\ M^{-1}s^{-1}$ and $5.4(\pm0.3)\cdot10^6\ M^{-1}s^{-1}$ at 0.58 M GdmCl. This is significantly faster than most association reactions during the folding of dimeric proteins such as the GCN14 leucine zipper,[107] but in the same order of magnitude as the wild-type arc repressor[75], and some designed Leu zippers.[76] Only an engineered Arc repressor variant,[78] where the intermolecular salt-bridge in the hydrophobic core was replaced by hydrophobic residues, and a designed fragment of trp repressor,[77] with bimolecular rate constants of about $3\cdot10^8\ M^{-1}s^{-1}$, associate considerably faster.

3.2 Foldon, the natural trimerization domain of T4 Fibritin, dissociates into a monomeric A-state form containing a stable beta-hairpin: atomic details of trimer dissociation and local beta-hairpin stability from residual dipolar couplings

Sebastian Meier, Sarah Güthe, Thomas Kiefhaber and Stephan Grzesiek

J. Mol. Biol. (2004) 344, 1051–1069

The high stability of the foldon domain can be explained by the large number of interactions at the trimer interface, as shown in Figure 3-4.

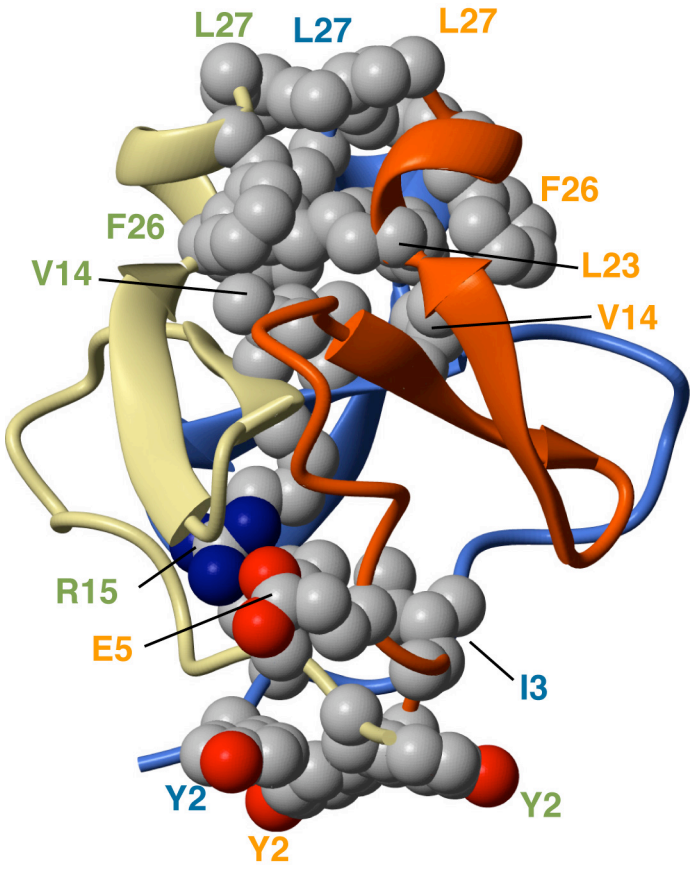


Figure 3-4

Inter-subunit contacts within the foldon trimer. The three monomers are shown in backbone worm representation in orange, blue and beige. Residues making inter-subunit contacts are indicated in space fill with CPK colours. For clarity the R15–E5 salt bridge (PDB code) is only shown between two monomers. This picture is taken from Meier *et al.*[108]

Of particular importance is a salt bridge between residues R15 and E5 connecting two adjacent subunits. For clarity, only one intermolecular salt bridge is depicted in Figure 3-4. At low pH, the salt bridge breaks up, as shown by a pH transition described in a previous study.[109] This as well as FRET measurements confirmed the existence of a low pH intermediate at pH 4. Additional measurements to investigate Tyr → Trp energy transfer at pH 2 have shown that also at pH 2 foldon is in a compact structure. (See chapter 5).

To investigate the low pH intermediate in more detail, particularly the formation of the A-state monomer from the trimer, its structure, thermodynamic stability, equilibrium association and folding dynamics, modern high-resolution NMR measurements were performed.

The NMR measurements validated the existence of the monomeric foldon A-state below pH 4.3. The solution structure could be determined(Figure 3-5), showing that the A-state forms a β-hairpin with intact and stable H-bonds similar to the monomer in the native foldon trimer, but lacking a defined structure in its N and C-terminal parts.

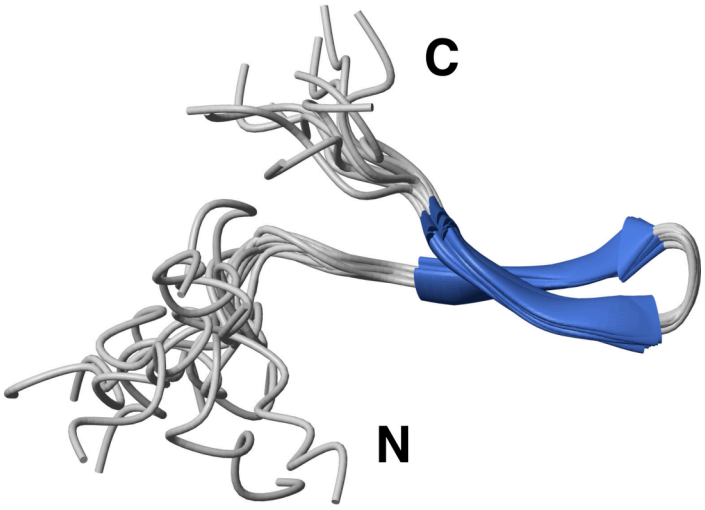


Figure 3-5

Solution structure of the A-state foldon monomer. Backbone worm representation of the ten lowest energy structures of the foldon monomer (PDB code: 1U0P). This picture is taken from Meier *et al.*[108]

By lowering the pH from pH 4 to pH 2, a second species besides the native trimer becomes increasingly populated. This A-state is monomeric, as the E5–R15 salt bridge disrupts due to

protonation of E5 at low pH. The pK_A of E5 is shifted to a value of ~1.75. From the difference in the protonation equilibrium constants for the E5 carboxylate in the monomer and trimer, the stabilization energy of $RT(\ln K_A^T - 3 \cdot \ln K_A^M) = 48$ kJ/(mol trimer) (16 kJ/(mol monomer)) at physiological pH and 25ºC was determined. This value corresponds to almost 20% of the total free energy between the native trimer and the unfolded protein of $\Delta G^0(H_2O)$ = 89.2±0.6 kJ/(mol monomer) as resolved by GdmCl-induced equilibrium transition.

With the help of residual dipolar couplings (RDCs) during thermal unfolding in mechanically strained polyacrylamide gels the structural order of $^{15}N–^1H^N$ and $^{13}C^a–^1H^a$ moieties could be observed directly and the different contributions from individual H-bonds and side-chain contacts became observable. The folding of the β-hairpin could thus be dissected into residue-specific phenomena. It was shown that local structures in the turn are exceptionally stable during thermal unfolding, and probably guide the collapsed hairpin to the native structure.

The thermodynamics of hairpin folding and ^{15}N relaxation data implies ms hairpin folding. This also confirms the folding model in Figure 3-3, where the compact monomeric burst-phase intermediate is formed on the submillisecond timescale.

doi:10.1016/j.jmb.2004.02.020

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

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

Sarah Güthe[1]†, Larisa Kapinos[1]†, Andreas Möglich[1]† Sebastian Meier[2], Stephan Grzesiek[2] and Thomas Kiefhaber[1]*

[1]*Division of Biophysical Chemistry, Biozentrum der Universität Basel Klingelbergstrasse 70, CH-4056 Basel, Switzerland*

[2]*Division of Structural Biology Biozentrum der Universität Basel, Klingelbergstrasse 70 CH-4056 Basel, Switzerland*

**Corresponding author*

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 $1.9(\pm 0.5) \times 10^6\,M^{-1}\,s^{-1}$ and $5.4(\pm 0.3) \times 10^6\,M^{-1}\,s^{-1}$ 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.

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]

The foldon domain was proposed to be an evolutionarily optimized trimerization/folding

† 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

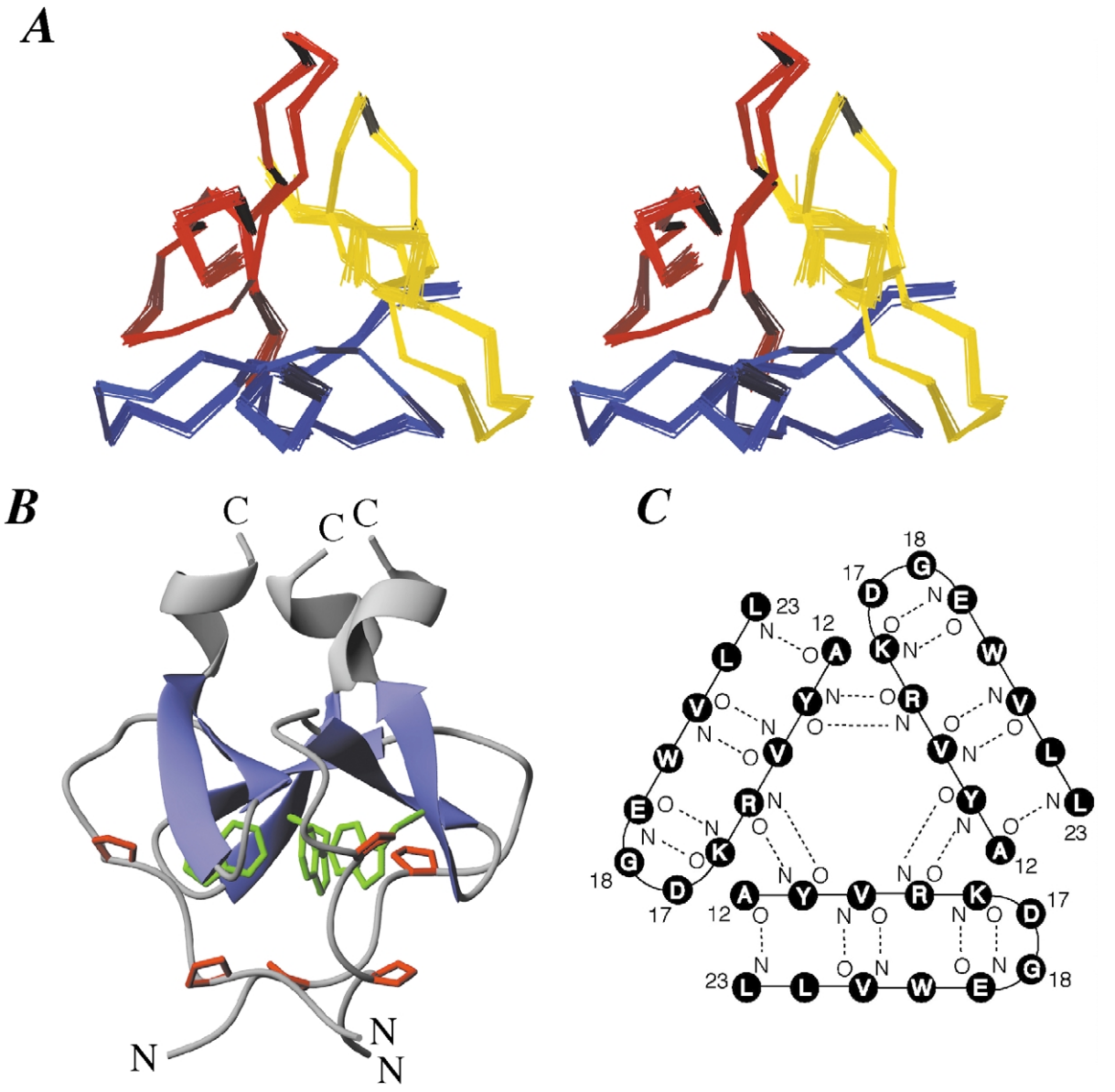


Figure 1. A, Stereo view of a bundle of the 20 lowest-energy structures of the trimeric foldon domain determined by NMR spectroscopy. Each subunit is displayed in a different color. B, Side view of the foldon structure with the single Trp residues at position 20 of each chain highlighted in green and 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[37] and rendered with PovRay.

motif, as its only known function is to promote folding of fibritin.[3] 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.[5]

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

```
1                  11                  21
GYIPEAPRDG QAYVRKDGEW VLLSTFL
```

corresponding to residues 457–483 in fibritin. 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 fibritin sequence, do not affect any thermodynamic or kinetic properties of the foldon domain.

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 fibritin, 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 fibritin, 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 fibritin. 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 $S^2 > 0.77$ at 25 °C as determined by the program TENSOR.[7]

Table 1. Statistics of the foldon NMR structure

rmsd from experimental distance constraints (Å)	
All (607)[a]	0.046 ± 0.002
rmsd from NMR data	
NMR quality factor Q[b]	0.199 ± 0.0062
rmsd (Hz) between measured and calculated dipolar couplings (81)[c]	1.91 ± 0.07
Experimental dihedral constraints (deg.)[d] (43)	1.73 ± 0.29
$^3J_{HNHA}$ coupling constants (Hz) (22)	0.89 ± 0.05
Total number of restraints per monomer	753
Deviation from the idealized covalent geometry	
Bonds (Å)	0.0079 ± 0.0004
Angles (deg.)	0.92 ± 0.03
Impropers[e] (deg.)	0.75 ± 0.06
Coordinate precision[f] (Å)	
Backbone non-hydrogen atoms	0.273
All non-hydrogen atoms	0.636
Non-Gly, non-Pro residues in Ramachandran regions[g]	
Most favored (%)	91.7
Allowed (%)	8.3
Generously allowed (%)	0.0
Disallowed (%)	0.0

The statistics were obtained from a subset of the 40 best energy 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.

[a] Distance restraints comprise: 111 intraresidual NOEs; 139 sequential NOEs ($|i-j|=1$); 73 short range NOEs ($1<|i-j|\leq 5$); 128 long-range NOEs ($|i-j|\leq 5$); 156 intermolecular NOEs; 11 H-bonds (eight intramolecular, three intermolecular). For each backbone hydrogen bond constraint, there are two distance restraints: r_{NH-O}, 1.7–2.5 Å, r_{N-O}, 2.3–3.5 Å.

[b] 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.[35]

[c] The 81 RDCs comprise 22 $^1D_{HN}$, 20 $^1D_{H\alpha C\alpha}$, 12 $^1D_{C\alpha C\beta}$, 13 $^1D_{NC'}$ (0.231), 14 $^1D_{CH3}$. Ramping the force constant for RDCs in the structure calculation from 0.001 kcal mol^{-1} Hz^{-2} to 0.5 kcal mol^{-1} Hz^{-2} was determined as optimal.

[d] The dihedral angle constraints comprise 69 ϕ and 60 ψ angles.

[e] The improper torsion restraints serve to maintain planarity and chirality.

[f] 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.

[g] These values are calculated with the program PROCHECK-NMR.[36] Values are reported for all residues.

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):

$$N \rightleftharpoons 3U \qquad (1)$$

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

$$K_{eq} = \frac{3f_U^3[M]_0^2}{1-f_U} \qquad (2)$$

where $[M]_0$ indicates the total monomer concentration ($[M]_0 = [U] + 3[N]$), f_u is the fraction of

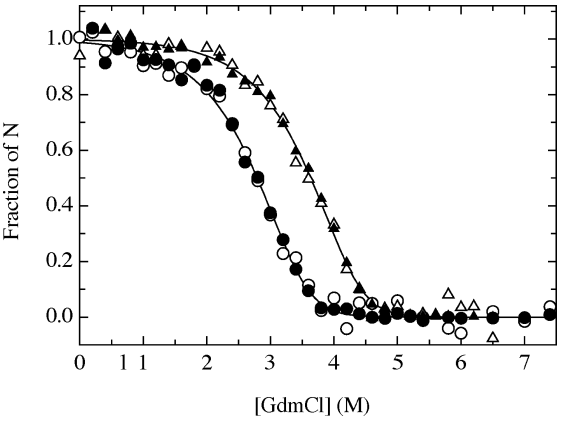


Figure 2. GdmCl-induced unfolding transition of the foldon domain at pH 7.1, 20 °C. Transitions at 5 μM (○,●) and 30 μM (▲,△) total monomer concentration ($[M]_0$) were measured by changes in Trp fluorescence (●,▲) and in far-UV CD at 228 nm (○,△). The data were normalized to fraction of native molecules using the result of a global fit of all data according to equations (14a) and (14b) (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 $\Delta G^0(H_2O) = 89.2(\pm 0.6)$ 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 μM this corresponds to $\Delta G = 29.7$ 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 $-10.4(\pm 0.2)$ (kJ/mol)/M, which is the value expected for a monomeric globular protein of the size of the folded trimer.[8] 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 μM. 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 μM to 200 μM), 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 μM and below, if we assume a maximum second-order rate constant[9] of about 1×10^9 M^{-1} 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 ($\lambda_{max} = 317$ nm) and the unfolded state ($\lambda_{max} = 345$ nm). The significantly blue-shifted fluorescence maximum and the largely increased fluorescence intensity in the intermediate relative

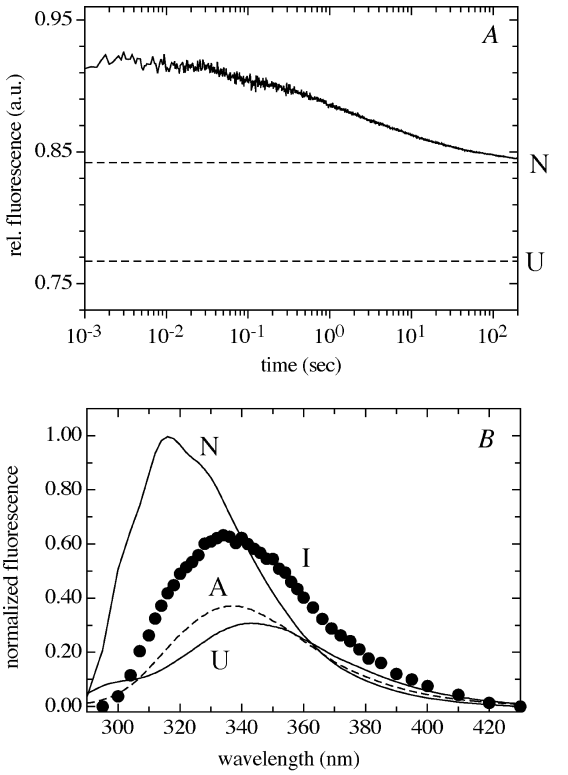


Figure 3. A, Refolding of the foldon domain in 0.58 M GdmCl, pH 7.1 ($[M]_0$ 5 μM) measured by the change in Trp fluorescence using a 320 nm emission cut-off filter. The broken lines represent the signals of the native and of the unfolded state at 0.58 M GdmCl, as indicated. The signal of the unfolded state is extrapolated from the unfolded baseline at high concentrations of GdmCl to 0.58 M GdmCl (see Materials and Methods). B, Comparison of the fluorescence spectrum of the kinetic burst phase intermediate (I) with the spectra of native (N) protein in 0.58 M GdmCl, the unfolded protein (U) in 8.2 M GdmCl and the monomeric A-state formed at pH 2 (A). The spectrum of I was determined in single-wavelength detection stopped-flow experiments. The fluorescence intensity extrapolated to $t = 0$ is shown. $[M]_0$ was 5 μM for all spectra.

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 ($[M]_0 > 10$ μM) a concentration-dependent faster reaction with a half-time of about 4 ms at $[M]_0 = 102$ μM and a concentration-independent slow reaction ($\tau = 50$ 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,[10] 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).

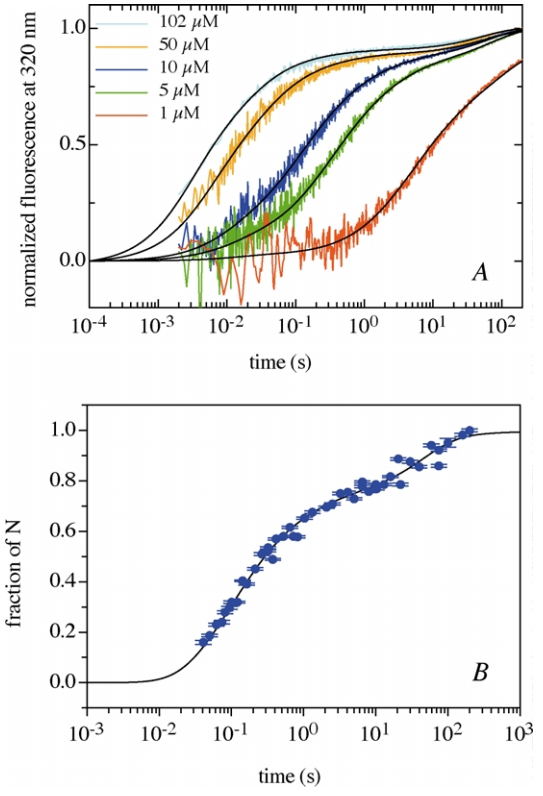


Figure 4. A, Fluorescence detected folding kinetics in the presence of 0.58 M GdmCl (pH 7.1), 20 °C at the indicated values of $[M]_0$ measured after stopped-flow mixing at an emission wavelength of 320 nm. B, Time-course of formation of native molecules measured in interrupted refolding experiments at pH 7.1, 20 °C, $[M]_0 = 10$ μM. The continuous lines in both panels represent the results from a global fit of the data. For global fitting, ten fluorescence-detected refolding traces at $[M]_0$ between 0.5 μM and 102 μM and the time-course of formation of native molecules shown in B, were fitted simultaneously to the kinetic model shown in equation (4).

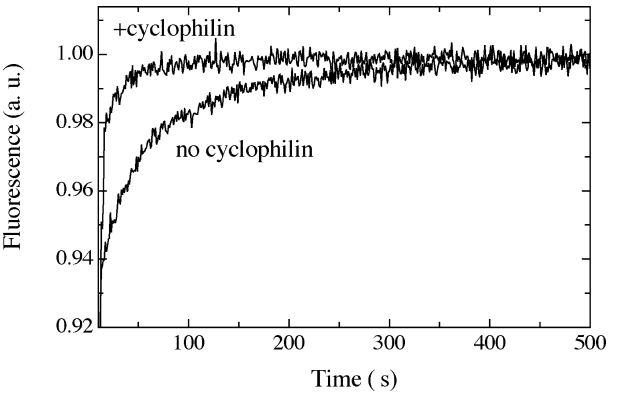


Figure 5. Effect of human cyclophilin 18 on the slow refolding reaction of the foldon domain. The plot compares refolding at pH 7.1, 0.58 M GdmCl, 20 °C in the absence and in the presence of 3.3 μM cyclophilin. The presence of cyclophilin increases the rate constant for the slow reaction from 7×10^{-3} s^{-1} to 4×10^{-2} s^{-1}.

We tested whether the faster, concentration-dependent reaction produces native protein or a folding intermediate by performing interrupted refolding experiments.[11–13] In these experiments, the protein is allowed to refold for a certain time (t_i). 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, t_i. 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.[12,13] 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, U_F). The slow process is due to folding of molecules with at least one non-native *cis* isomer (slow-folding molecules, U_S), 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

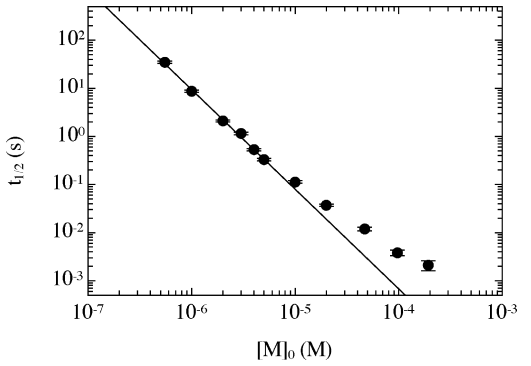


Figure 6. Effect of total monomer concentration ($[M]_0$) on the half-time of the fast-refolding reaction of the foldon domain at 0.58 M GdmCl (pH 7.1), 20 °C. The continuous line represents a fit of the data between 0.5 μM and 4 μM to equation (3). The fit gives a slope of -2.06 ± 0.05 indicating an apparent reaction order of 3 at low concentrations of protein.

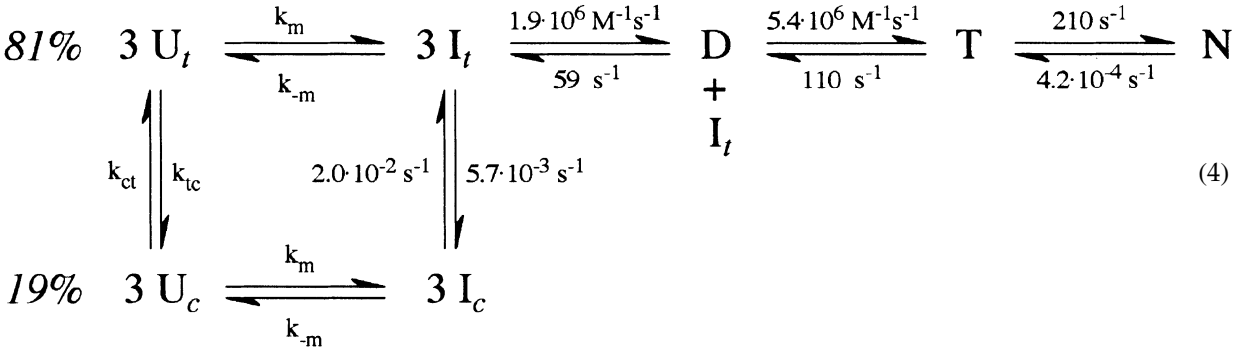
process of the foldon domain in more detail. The analysis of the concentration-dependence of the half-lives ($t_{1/2}$) of the folding reaction starting from U_F allows us to determine the apparent reaction order for the fast-folding pathway according to:

$$\log t_{1/2} = c - (n-1)\log [M]_0 \qquad (3)$$

where n is the reaction order and c is a reaction-specific constant.[14] 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 ± 0.05, which corresponds to apparent third-order kinetics. At monomer concentrations above 5 μM the slope decreases significantly, indicating a change in the apparent reaction order. Between 100 μM and 200 μM 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.[15] 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:



(4)

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_m 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_f) of structural rearrangement in the trimer of $210(\pm 20)\ s^{-1}$. This is in the same range as the rate constants observed for the fastest folding monomeric proteins of similar size.[16] The bimolecular rate constants for formation of the dimer and of the trimer are $1.9(\pm 0.5) \times 10^6\ M^{-1}\ s^{-1}$ and $5.4(\pm 0.3) \times 10^6\ M^{-1}\ s^{-1}$, respectively, which is about 100 to 200-fold slower than the expected diffusion limit for bimolecular reactions of chains of this size.[9] 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[17] leucine zipper and of many globular dimeric proteins.[18] Rate constants similar to the two bimolecular steps during trimerization of foldon have been reported for wild-type arc repressor,[9] the fastest folding naturally occurring dimeric protein known, and for some designed leucine zippers.[19] However, an engineered arc repressor variant[20] and a designed fragment of trp repressor[21] show bimolecular rate constants of about $3 \times 10^8\ M^{-1}\ s^{-1}$, 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.[20] 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 ($k = 4.2(\pm 0.5) \times 10^{-4}\ s^{-1}$). 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 $110(\pm 40)\ s^{-1}$ and $59(\pm 4)\ s^{-1}$, 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.[22] This agrees well with $20(\pm 1)$% 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 $20(\pm 1)$% 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 $210(\pm 20)\ s^{-1}$ obtained from the global fit. Since the apparent half-time was determined from the fluorescence measurements (Figure 6), it will be influenced by

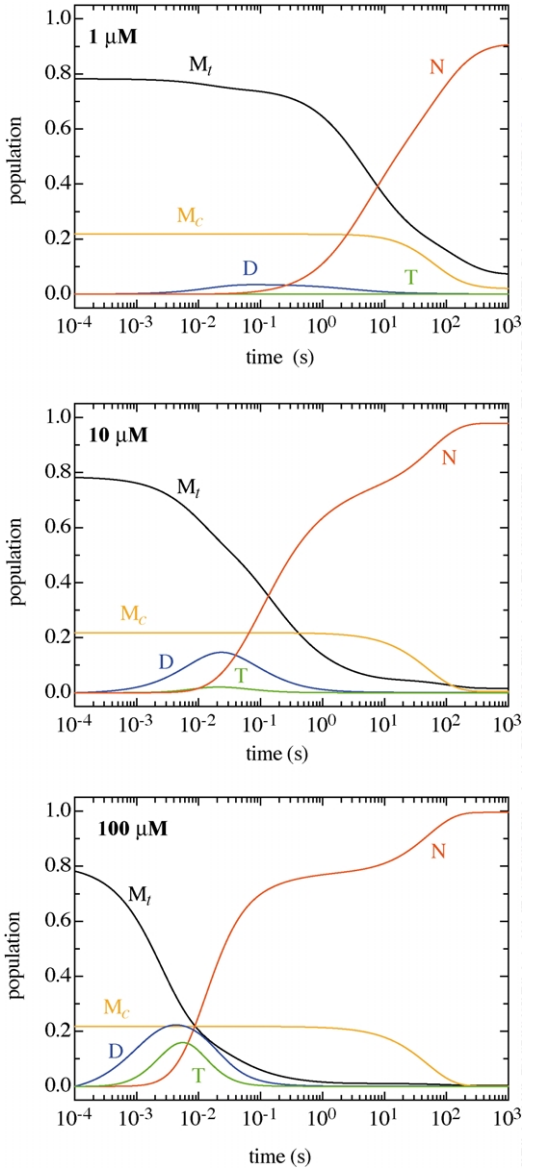


Figure 7. Simulation of the time-course of all kinetic species during folding of the foldon domain at 0.58 M GdmCl (pH 7.1), 20 °C. The results from the global fit of all kinetic data (see Figure 4) to the kinetic model shown in equation (4) were used to calculate the concentrations of the different species at the indicated values of $[M]_0$.

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 fibritin 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 (i) to pronounced fluorescence changes of the single Trp residue, which is part of the β-hairpin, (ii) 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[5] or on collagens, which are limited in folding by prolyl *cis–trans* isomerization reactions.[27] 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 (Hibar, LiChrosorb®100 RP-8 from Merck, Darmstadt, Germany). ^{13}C,^{15}N-labeled foldon for NMR spectroscopy was expressed in *E. coli* BL21(DE3) cells grown in ^{13}C,^{15}N-enriched M9 minimal medium. Protein purity was confirmed by nanospray mass spectrometry and analytical HPLC. Expression and purification of the foldon domain in *E. coli* was performed as described.[4]

NMR spectroscopy

NMR samples ([U-^{15}N]foldon 95% H_2O/5% ^2H_2O; [U-^{13}C,^{15}N]foldon 95% H_2O/5% ^2H_2O; [U-^{13}C,^{15}N]foldon 100% ^2H_2O) 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) NaN_3 without any further addition of salt. Residual alignment of [U-^{13}C,^{15}N]foldon for the measurement of residual dipolar couplings was introduced by lamellar ether/*n*-hexanol phases.[28] A set of standard triple and double-resonance assignment, quantitative *J*-coupling, nuclear Overhauser effect (NOE) spectroscopy (NOESY) and ^{15}N relaxation experiments similar to those described[29] were performed on a Bruker DRX 600 spectrometer at 25 °C. Standard data processing and analysis was carried out as described.[29]

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.[30] The structural statistics for the best 40 structures are given in Table 1.

Protein Data Bank accession code

The structural statistics 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):

$$N \rightleftharpoons 3U \tag{5}$$

The resulting equilibrium constant is given by:

$$K_{eq} = \frac{[U]^3}{[N]} \tag{6}$$

and the total monomer concentration ($[M]_0$) can be expressed as:

$$[M]_0 = [U] + 3[N] \tag{7}$$

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

$$f_U = \frac{[U]}{[U] + 3[N]} = \frac{[U]}{[M]_0} \tag{8a}$$

$$f_N = \frac{3[N]}{[U] + 3[N]} = \frac{3[N]}{[M]_0} = 1 - f_U \tag{8b}$$

Thus, equation (6) becomes:

$$K_{eq} = \frac{[U]^3}{[N]} = \frac{3f_u^3[M]_0^2}{1 - f_u} \tag{9}$$

To fit the equilibrium transitions to equation (9) we expressed the fraction of unfolded protein at a given denaturant concentration (x) as:[31]

$$f_U(x) = \frac{S_N(x) - S(x)}{(S_N(x) - S(x)) + (S(x) - S_U(x))} \tag{10}$$

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:[32]

$$S_N(x) = S_N(H_2O) + m_N x \tag{11a}$$

$$S_U(x) = S_U(H_2O) + m_U x \quad (11b)$$

Using:

$$\Delta G^0 = -RT \ln K_{eq} \quad (12)$$

and a linear denaturant-dependence of ΔG^0 :[32,33]

$$\Delta G^0(x) = \Delta G^0(H_2O) + mx \quad (13)$$

we can obtain the free energy for unfolding in water $\Delta G^0(H_2O)$ 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 of monomeric proteins.[34] This yields the following equation for a two-state unfolding transition from a native trimer to unfolded monomers:

$$S(x) = S_N(x) - \frac{S_N(x) - S_U(x)}{3[M]_0} \times \left(\sqrt[3]{e^{\frac{-\Delta G^0(H_2O)+mx}{RT}} \left(\frac{9}{2}[M]_0 + \sqrt{D} \right)} + \sqrt[3]{e^{\frac{-\Delta G^0(H_2O)+mx}{RT}} \left(\frac{9}{2}[M]_0 - \sqrt{D} \right)} \right) \quad (14a)$$

with:

$$D = \left(\frac{9}{2} \right)^2 [M]_0^2 + e^{\frac{-\Delta G^0(H_2O)+mx}{RT}} \quad (14b)$$

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.MV 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 (≥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 μM 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 μM 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 $[M]_0 = 5$ μM foldon domain in the absence and in the presence of 3.3 μM 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), $[M]_0 = 60$ μM) was diluted sixfold into final conditions of 10 μM 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 μM. Native foldon unfolds with double-exponential kinetics under these conditions (S.Gü. & T.K., unpublished results) with $\tau_1 = 7.8(\pm 0.5)$ seconds (80% amplitude) and $\tau_2 = 0.24(\pm 0.02)$ seconds (20% amplitude). The relative amplitudes of the two reactions is independent of the refolding time. The amplitude of the major, slow-unfolding reaction ($\tau_1 = 7.8(\pm 0.5)$ 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 $[M]_0 = 10$ μM and ten direct fluorescence-detected refolding traces with $[M]_0$ ranging from 0.5 μM to 102 μM 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

1. Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997). Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure*, **5**, 789–798.
2. Letarov, A. V., Londer, Y. Y., Boudko, S. P. & Mesyanzhinov, V. V. (1999). The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibritin. *Biochemistry (Moscow)*, **64**, 817–823.
3. Boudko, S. P., Londer, Y. Y., Letarov, A. V., Sernova, N. V., Engel, J. & Mesyanzhinov, V. V. (2002). Domain organization, folding and stability of bacteriophage T4 fibritin, a segmented coiled-coil protein. *Eur. J. Biochem.* **269**, 833–841.
4. Frank, S., Kammerer, R. A., Mechling, D., Schulthess, T., Landwehr, R. J. B., Guo, Y. *et al.* (2001). Stabilization of short collagen-like triple helices by protein engineering. *J. Mol. Biol.* **308**, 1081–1089.
5. Seckler, R. (2000). Assembly of multi-subunit

structures. In *Mechanisms of Protein Folding* (Pain, R., ed.), pp. 279–308, Oxford University Press, Oxford.
6. Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R. A., Boudko, S. P. *et al.* (2003). Collagen stabilization at atomic level: crystal structure of designed $(GlyProPro)_{10}$-foldon. *Structure*, **11**, 339–346.
7. Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR*, **16**, 23–28.
8. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
9. Milla, M. E. & Sauer, R. T. (1994). P22 arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry*, **33**, 1125–1133.
10. Fischer, G., Wittmann, L. B., Lang, K., Kiefhaber, T. & Schmid, F. X. (1989). Cyclophilin and peptidyl-prolyl *cis–trans* isomerase are probably identical proteins. *Nature*, **337**, 476–478.
11. Schmid, F. X. (1983). Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction *via* structural intermediates. *Biochemistry*, **22**, 4690–4696.
12. Kiefhaber, T. (1995). Protein folding kinetics. In *Methods in Molecular Biology* (Shirley, B. A., ed.), vol. 40, pp. 313–341, Humana Press, Totowa, NJ.
13. Bieri, O. & Kiefhaber, T. (2000). Kinetic models in protein folding. In *Protein Folding: Frontiers in Molecular Biology* (Pain, R., ed.), 2nd edit., pp. 34–64, Oxford University Press, Oxford.
14. Moore, J. W. & Pearson, R. G. (1981). *Kinetics and Mechanisms*, Wiley, New York.
15. Gutfreund, H. (1995). *Kinetics for the Life Sciences*, Cambridge University Press, Cambridge.
16. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91.
17. Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. E. & Matthews, C. R. (1995). Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry*, **34**, 12812–12819.
18. Jaenicke, R. (1987). Folding and association of proteins. *Prog. Biophys. Mol. Biol.* **49**, 117–237.
19. Wendt, H., Berger, C., Baici, A., Thomas, R. M. & Bosshard, H. R. (1995). Kinetics of folding of leucine zipper domains. *Biochemistry*, **34**, 4097–4107.
20. Waldburger, C. D., Jonsson, T. & Sauer, R. T. (1996). Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl Acad. Sci. USA*, **93**, 2629–2634.
21. Gloss, L. M. & Matthews, C. R. (1998). Mechanism of folding of the dimeric core domain of *Escherichia coli* Trp repressor: a nearly diffusion-limited reaction leads to the formation of an on-pathway intermediate. *Biochemistry*, **37**, 15990–15999.
22. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M. & Fischer, G. (1998). Side-chain effects on peptidyl-prolyl *cis/trans* isomerization. *J. Mol. Biol.* **279**, 449–460.
23. Bachmann, A. & Kiefhaber, T. (2001). Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* **306**, 375–386.
24. Sánchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376.
25. Wagner, C. & Kiefhaber, T. (1999). Intermediates can accelerate protein folding. *Proc. Natl Acad. Sci. USA*, **96**, 6716–6721.
26. Zitzewitz, J. A., Ibarra-Molero, B., Fishel, D. R., Terry, K. L. & Matthews, C. R. (2000). Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J. Mol. Biol.* **296**, 1105–1116.
27. Boudko, S. P., Frank, S., Kammerer, R. A., Stetefeld, J., Schulthess, T., Landwehr, R. *et al.* (2002). Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: transition from third to first order kinetics. *J. Mol. Biol.* **317**, 459–470.
28. Ruckert, M. & Otting, G. (2000). Alignment of biological macromolecules in novel non-ionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793–7797.
29. Kahmann, J. D., Sass, H. J., Allan, M. G., Seto, H., Thompson, C. J. & Grzesiek, S. (2003). Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *EMBO J.* **22**, 1824–1834.
30. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
31. Tanford, C. (1968). Protein denaturation. Part B. The transition from native to denatured state. *Advan. Protein Chem.* **23**, 218–282.
32. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.
33. Greene, R. F. J. & Pace, C. N. (1974). Urea and guanidine–hydrochloride denaturation of ribonuclease, lysozyme, alpha-chyomtrypsinn and beta-lactoglobulin. *J. Biol. Chem.* **249**, 5388–5393.
34. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.
35. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.*, 6836–6837.
36. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
37. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.

Edited by F. Schmid

(Received 24 November 2003; received in revised form 30 January 2004; accepted 5 February 2004)

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Foldon, The Natural Trimerization Domain of T4 Fibritin, Dissociates into a Monomeric A-state Form containing a Stable β-Hairpin: Atomic Details of Trimer Dissociation and Local β-Hairpin Stability from Residual Dipolar Couplings

Sebastian Meier[1], Sarah Güthe[2], Thomas Kiefhaber[2] and Stephan Grzesiek[1]*

[1]*Division of Structural Biology Biozentrum der Universität Basel, Klingelbergstrasse 70 CH-4056 Basel, Switzerland*

[2]*Division of Biophysical Chemistry, Biozentrum der Universität Basel Klingelbergstrasse 70, CH-4056 Basel, Switzerland*

**Corresponding author*

The C-terminal domain of T4 fibritin (foldon) is obligatory for the formation of the fibritin trimer structure and can be used as an artificial trimerization domain. Its native structure consists of a trimeric β-hairpin propeller. At low pH, the foldon trimer disintegrates into a monomeric (A-state) form that has similar properties as that of an early intermediate of the trimer folding pathway. The formation of this A-state monomer from the trimer, its structure, thermodynamic stability, equilibrium association and folding dynamics have been characterized to atomic detail by modern high-resolution NMR techniques. The foldon A-state monomer forms a β-hairpin with intact and stable H-bonds that is similar to the monomer in the foldon trimer, but lacks a defined structure in its N and C-terminal parts. Its thermodynamic stability in pure water is comparable to designed hairpins stabilized in alcohol/water mixtures.

Details of the thermal unfolding of the foldon A-state have been characterized by chemical shifts and residual dipolar couplings (RDCs) detected in inert, mechanically stretched polyacrylamide gels. At the onset of the thermal transition, uniform relative changes in RDC values indicate a uniform decrease of local $N–H^N$ and $C^\alpha–H^\alpha$ order parameters for the hairpin strand residues. In contrast, near-turn residues show particular thermal stability in RDC values and hence in local order parameters. This coincides with increased transition temperatures of the β-turn residues observed by chemical shifts. At high temperatures, the RDCs converge to non-zero average values consistent with predictions from random chain polymer models. Residue-specific deviations above the unfolding transition reveal the persistence of residual order around proline residues, large hydrophobic residues and at the β-turn.

Keywords: NMR spectroscopy; protein folding; thermodynamics; structure; order parameter

Introduction

Fibritin is an elongated, trimeric, 486 amino acid protein of the bacteriophage T4 head with structural, chaperone as well as sensory functions.[1–3] 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). The structures of C-terminal constructs comprising 120 and 75 amino acid residues of T4 fibritin have been solved by crystallography.[4,5] The C-terminal part of the fibritin trimer is formed from a trimeric coiled-coil domain, which is terminated by a C-terminal trimeric β-sheet propeller consisting of monomeric β-hairpin segments (Figure 1). This β-propeller is

Abbreviations used: RDC, residual dipolar coupling; RMSD, root-mean-square deviation; NOE, nuclear Overhauser enhancement.

E-mail address of the corresponding author: stephan.grzesiek@unibas.ch

necessary for the correct folding of the holoprotein[6] and hence has been termed foldon. Residues G457 to L483 (renumbered G1 to L27 for convenience in the present work) are sufficient to form an isolated trimeric foldon structure.[7] Trimerization occurs at temperatures of up to 75 °C and SDS concentrations of up to 2% (w/v) at room temperature.[6] Due to the extreme stability of the trimer, the foldon can be used as an artificial trimerization inducer or enhancer. Thus, collagen[8] and HIV gp140[9] trimers have been stabilized *via* attachment of the foldon domain.

Kinetic data indicate that folding of foldon is a multi-step process,[7] where trimerization starts from a collapsed monomeric intermediate formed in the sub-millisecond time scale. The monomeric intermediate has fluorescence emission properties (λ_{max} = 330 nm) that differ significantly both from the denatured state and the blue-shifted native trimer state (λ_{max} = 317 nm). However, the fluorescence maximum is virtually identical to a monomeric acidic state (A-state) that becomes strongly populated at pH values below pH 2 and concentrations of lower than ~200 μM.[7] Thus, the A-state may represent an early folding intermediate. Trimerization proceeds from this monomeric

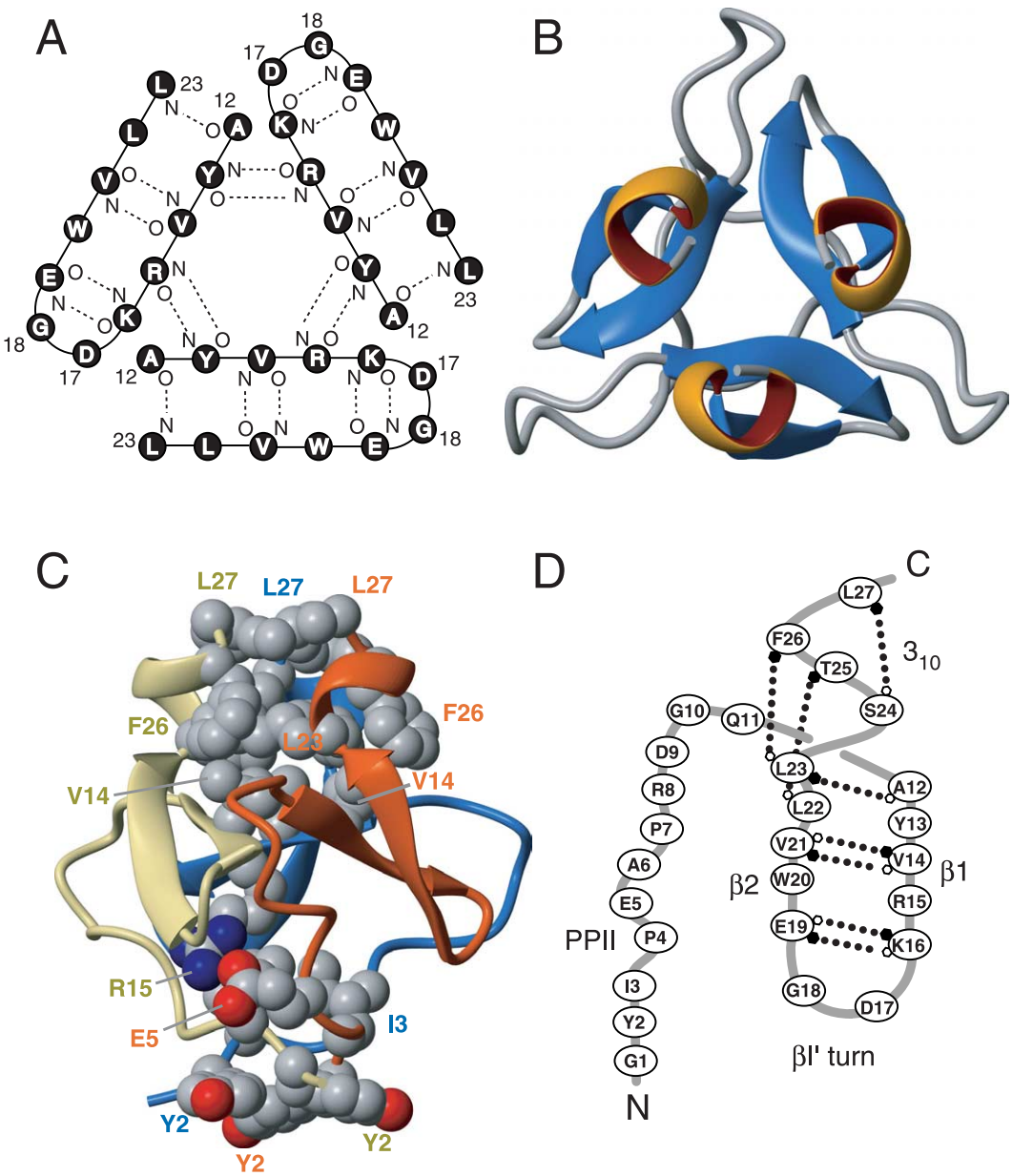


Figure 1. The structure of the T4 fibritin foldon trimer. A, H-bond topology of the trimeric foldon β-hairpin propeller. B, Backbone worm representation of the mean solution structure. C, Inter-subunit contacts within the foldon trimer. The three monomers are shown in backbone worm representation as orange, blue and beige. Residues that make inter-subunit contacts are indicated in space fill with CPK colors. For clarity the R15–E5 salt bridge is only shown between two monomers. D, Secondary structure topology of a foldon monomer within the trimer structure. The H-bonds are depicted by a dotted line between the amide proton donor (filled hexagons) and the oxygen acceptor (open hexagons).

intermediate by two sequential bimolecular association steps with rate constants that are close to the fastest association rates observed for proteins in solution. The extremely fast association reflects the biological role of foldon as a very effective trimerization domain during phage assembly, where fibritin concentrations in the host are on the order of a few micromolar.[7]

Clearly, the detailed understanding of protein folding is of fundamental importance.[10] The interest has been sparked further by the discovery of folding diseases where misfolded proteins reach a seemingly global energy minimum comprised of fibrillar β-strand amyloid structures.[11,12] While model peptides for the folding of α-helices have been described extensively for several decades,[13–15] the folding of isolated β-hairpins, representing the minimal model for β-sheet formation, has been characterized only more recently.[16,17] The understanding is complicated by the various non-sequential interactions involved in hairpin and β-sheet formation. The folding and unfolding of isolated β-hairpins has been shown to occur on a time scale of few microseconds, and the experimental evidence indicates that folding usually occurs in broad transitions due to the relatively small folding enthalpies.[18]

Solution NMR can monitor folding transitions and equilibria at atomic resolution. Traditionally, chemical shift and nuclear Overhauser enhancement (NOE) information are used,[19] but more recently also the weak alignment of proteins in anisotropic liquid phases[20] has become a tool to directly monitor average net orientations and order parameters of individual bonds by residual dipolar couplings (RDCs). The introduction of inert orienting media, such as strained polyacrylamide gels,[21,22] has made it possible to obtain weak alignment and RDC information even under relatively harsh unfolding conditions.[23–26] RDC techniques allow a quick characterization of the topology of folded and also of only partially folded states. Thus, the local propensities for α-helical conformations could be detected in unfolded forms of the S-peptide,[27] the acyl-coenzyme A binding protein,[28] and myoglobin.[29]

Here, we have used modern solution state NMR methods to characterize the details of the transition from the foldon trimer to the A-state monomer, to determine the monomer structure, and to characterize its thermal unfolding. The protonation of the E5 side-chain below pH 2, and the consequent opening of the E5–R15 salt bridge are responsible for the trimer–monomer transition. The A-state monomer forms a β-hairpin with native backbone structure, including native H-bonds and thermally stable side-chain contacts. Relaxation data show that the hairpin amides undergo microsecond motions, presumably representing the overall folding and unfolding transition. The determination of RDCs during thermal unfolding in mechanically strained polyacrylamide gels permits the direct observation of local order of NH^N and $C^\alpha H^\alpha$ moieties and a breakdown into different contributions from H-bonds and side-chain contacts. The apparent two-state folding of the hairpin can thus be dissected into residue-specific phenomena, which can be compared to predictions from simulations and statistical mechanical models. Local structures in the turn prove to be particularly stable during thermal unfolding and are therefore likely to steer the hairpin during the hydrophobic collapse towards the native H-bond register.

Results and Discussion

The native trimer structure

The isolated foldon trimer structure solved by NMR in solution at pH 7 (Figure 1) is very similar to the foldon part of the larger protein constructs in crystalline form.[7] Within the trimer, the foldon monomers (residues G1–L27) consist of an extended N-terminal region (G1-Q11), a β-hairpin (A12-L23), and a C-terminal 3_{10} helix (L23–L27). The extended N-terminal region contains a polyproline II helix between residues P4 and P7 and packs against one side of the β-hairpin by hydrophobic contacts. The central β-hairpin is short, with five residues per strand (A12–K16 and E19–L23) and strongly twisted. According to a definition used by Hermans and co-workers,[30] the right hand twist amounts to an angle of more than 60° between two cross-strand $C^\alpha C^\alpha$ vectors of sequential H-bonded residue pairs (e.g. $C^\alpha C^\alpha$ V14–V21 and $C^\alpha C^\alpha$ K16–E19; see Figure 1A). This twist leads to a very compact β-propeller topology and presumably disfavors aggregation into extended sheets. The hairpin is stabilized by hydrophobic residues at its N and C-terminal ends (A12–V14, V21–L23), and a diagonal interaction between Y13 and W20 seems to promote the hairpin twist. The hairpin strands are connected by a type 1 β-turn consisting of residues K16 to E19. Together with R15, residues K16, D17, and E19 form a charged cluster in the turn region of the hairpin.

The foldon trimer interface consists of many hydrophobic interactions along the trimer symmetry axis, involving residues Y2 and I3 at the N terminus, V14 in the hairpin, as well as L23, F26, and L27 in the C-terminal 3_{10} helix (Figure 1C). The crucial inter-subunit salt bridges (see below) between residues E5 and R15 and the inter-subunit H-bonds between Y13 and R15 (Figure 1A) further contribute to the extreme stability of the trimer.

The monomeric A-state

When changing the pH from neutral to acidic, a second species appears in the NMR spectra of foldon, which is in slow chemical exchange with the foldon trimer state (Figure 2A). This second species becomes increasingly populated between pH 4.3 and pH 1.6 (Figure 2C) and apparently corresponds to the acidic state (A-state) of foldon. Lowering the

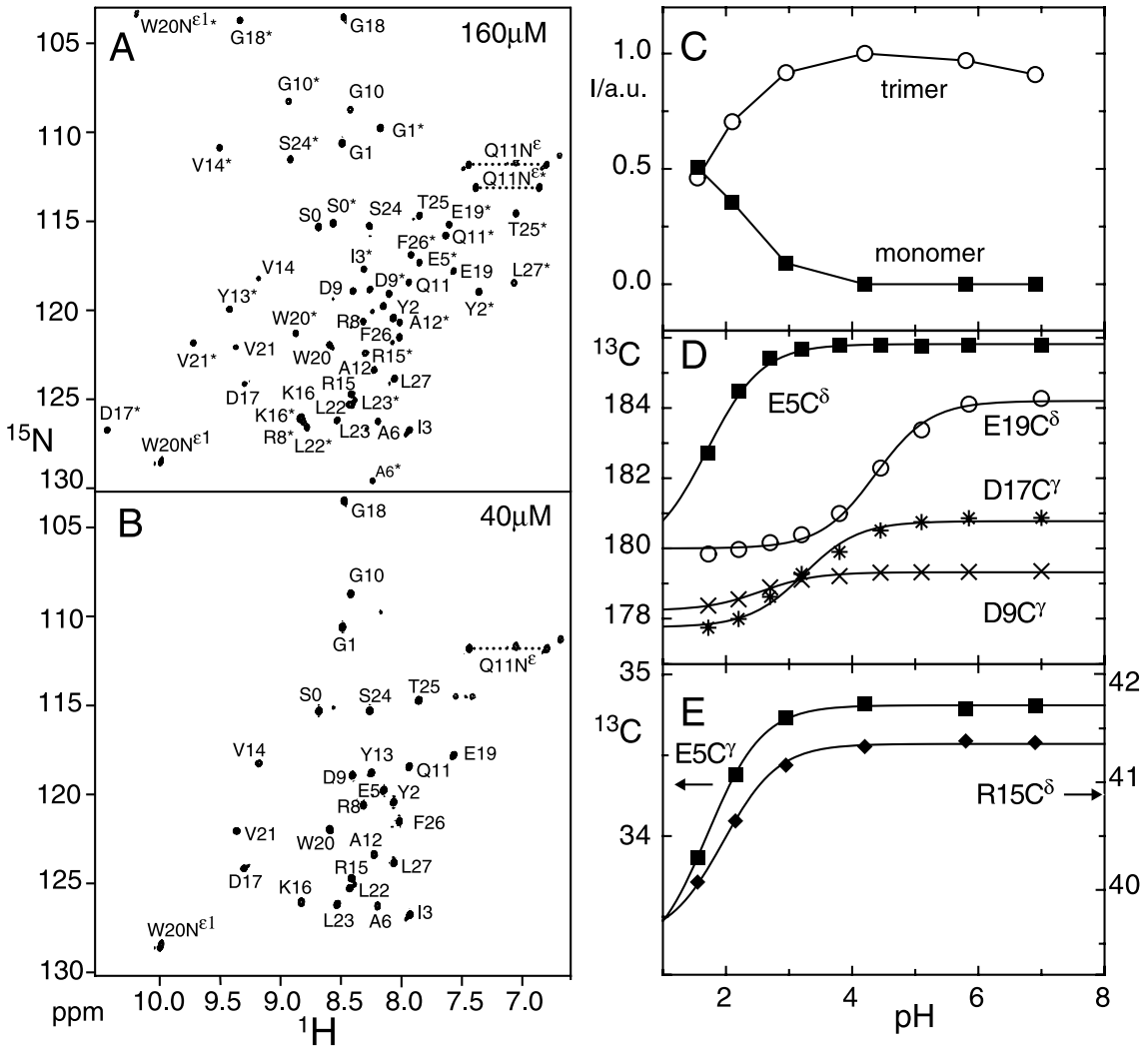


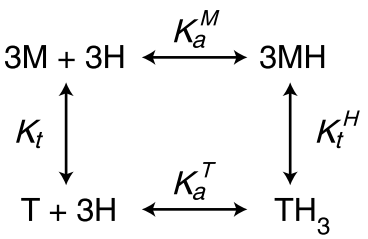
Figure 2. A and B, Concentration-dependent dissociation of the foldon trimer at pH 2 and 25 °C in solutions with 5 mM phosphate and 20 mM chloride at 40 μM (A) and 160 μM (B) peptide concentration. Trimer and monomer coexist in equilibrium in slow exchange. Amide resonance assignments for both species are given in A and B with trimer resonances indicated by asterisks(*). C, Average trimer and monomer peak intensities as a function of pH at 25 °C and 0.25 μM total peptide concentration. D, Titration of side-chain carboxylate ^{13}C resonances in trimeric foldon at 25 °C. Continuous lines are fits of the Henderson–Hasselbalch equation to chemical shifts. pK_a values are: E5: 1.70, D9: 2.55, D17: 3.23, E19: 4.37. E, Co-titration of E5-$^{13}C^{\gamma}$ and R15-$^{13}C^{\delta}$ aliphatic side-chain resonances.

protein concentration while keeping the pH at pH 2 leads to a relative intensity increase of this A-state species with respect to the trimer state (Figure 2B). The intensity dependence corresponds to a 3:1 stoichiometry of the two species. This fact, the ^{15}N relaxation data of the A-state (see below), its fast exchange with an unfolded species over a broad temperature range, and the high similarity of its fluorescence spectra with the monomeric folding intermediate indicate that the A-state of foldon is monomeric. The intensity of spectral lines corresponding to trimer and monomer becomes approximately equal at pH 1.7 for a 250 μM total protein concentration (Figure 2C). Thus, the trimer concentration is about one-third that of the monomer concentration under these conditions. As trimer and monomer coexist in slow chemical exchange, their separate sets of resonances could be assigned individually by standard heteronuclear 2D NMR experiments. In this way, complete backbone and side-chain chemical shift assignments were obtained. No further spectral species are detectable for a threshold of about 5% of the monomer A-state population. Thus, we conclude that other intermediates, such as dimers, are not significantly populated.

The reason for the disintegration of the trimer at low pH values can be traced back to the breaking of the inter-subunit E5–R15 salt bridge (Figure 1D). A titration of the side-chain carboxylate ^{13}C resonances for all glutamic and aspartic acid residues (Figure 2D) and of the aliphatic E5-$^{13}C^{\gamma}$ and R15-$^{13}C^{\delta}$ resonances (Figure 2E) shows that the pK_a value for E5 is strongly shifted to 1.7, and that

the aliphatic E5 and R15 side-chain resonances co-titrate according to this pK_a value. Thus, the protonation state of the E5–R15 salt bridge closely follows the monomer–trimer equilibrium, as judged from the monomer and trimer resonance intensities shown in Figure 2C.

The following minimal reaction scheme expresses this dependence of the monomer–trimer equilibrium on the monomer protonation state:



(1)

$$\text{where } K_a^M = \frac{[M][H]}{[MH]}, \quad K_a^T = \frac{[T][H]^3}{[TH_3]},$$

$$K_t^H = \frac{[MH]^3}{[TH_3]} \text{ and } K_t = \frac{[M]^3}{[T]}.$$

An analysis of the spectral data yields values for all four equilibrium constants: thus, the midpoint of the protonation equilibrium of E5 in the trimer is at pH 1.7, yielding a value for K_a^T of $10^{-5.1}$ M^3, whereas for the free glutamic acid side-chain in the monomer a K_a^M value of $10^{-4.4}$ M can be assumed. Neglecting an appreciable population of deprotonated monomer M at pH 1.7, the monomer–trimer equilibrium constant in the protonated state can be calculated as $K_t^H = [MH]^3/[TH_3] = [0.125\text{ mM}]^3/[0.125\text{ mM}/6] = 9.4\times10^{-8}$ M^2, because the observed monomer (MH) and total trimer (T+TH$_3$) intensities are approximately equal for a total foldon concentration of 250 μM, and [T]=[TH$_3$] at this pH value. Finally, the circular reaction scheme in equation (1) leads to the condition that $K_t = K_t^H(K_a^M)^3/K_a^T = 7.4\times10^{-16}$ M^2.

The difference in the protonation equilibrium constants for the E5 carboxylate in the monomer and trimer correspond to a stabilization energy of $RT(\ln K_a^T - 3\times\ln K_a^M) = 48$ kJ/mole trimer (16 kJ/mole monomer) at physiological pH and 25 °C. A total free energy of 89 kJ/mole monomer was determined for the transition from the folded trimer to the guanidinium chloride unfolded monomer.[7] The considerable strength of the inter-subunit salt-bridge can be explained by the highly hydrophobic interior of the foldon trimer, which also leads to pronounced blue shifts in the fluorescence emission spectrum of the interior residue W20.[7] A stabilization energy of very similar size due to the presence of a single salt-bridge has been observed in the pH-induced denaturation of T4 lysozyme.[31]

When both monomer and trimer species are present at low pH, no line-broadening from intermediate exchange can be observed in the ^{1}H–^{15}N correlation spectra, even for small chemical shift differences ($|\Delta\nu^1H|$) < 50 Hz between the two forms. This indicates that the exchange time between monomer and trimer forms must be significantly longer than 20 ms. Indeed, results of chemical exchange spectroscopy show that the transition from the monomer to the trimer occurs on a time scale > 800 ms at 25 °C. At 40 °C exchange peaks between monomer and trimer species become observable within 800 ms. Their intensities correspond to a pseudo-third-order association rate of $\sim1\times10^6$ M^{-2} s^{-1} for the monomers and a trimer dissociation rate of ~0.5 s^{-1}.

The A-state monomer structure

The resonances of the A-state monomer show strong chemical shift dispersion for hairpin residues A12–L23 (e.g. see Figure 2B). For this region, the secondary chemical shifts indicate an intact β-sheet structure. In particular, the pronounced downfield shifts of the $^1H^N$ amide resonances (V14, K16, D17, V21) are a sign of strong and intact β-sheet H-bonds.[32,33] In contrast, ^{1}H, ^{13}C, and ^{15}N chemical shifts for residues outside of the hairpin region, i.e. S0–A12 and S24–L27, are indicative of random coil conformations.

An analysis of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs (Figure 3A and B) obtained for the A-state monomer in horizontally compressed polyacrylamide gels

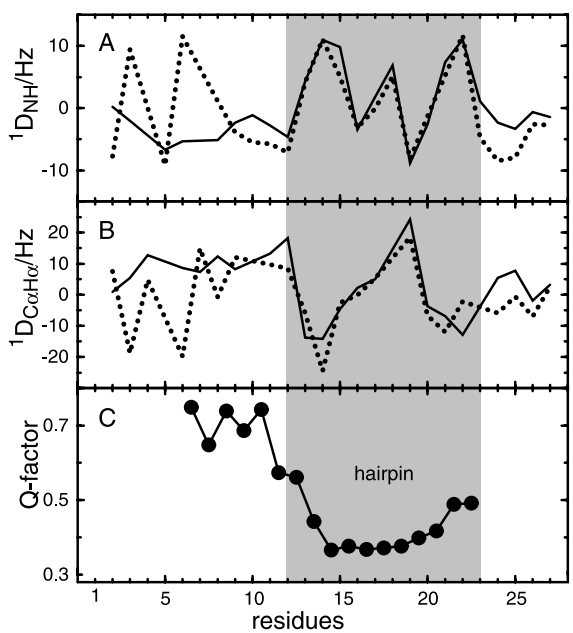


Figure 3. Analysis of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs in the foldon monomer. A, $^1D_{NH}$, B, $^1D_{C\alpha H\alpha}$ RDCs. Continuous lines represent measured couplings, dotted lines RDCs calculated from the average trimer structure. Residual orientation was obtained within mechanically strained acrylamide gels for 0.4 mM foldon, 5 mM phosphate, 20 mM chloride at 45 °C and pH 2. C, *Q*-factors of the combined $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs of segments of ten consecutive amino acid residues moved through the foldon sequence as a function of the average residue number (see the text).

corroborates this picture. RDCs in the hairpin are large and consistent with a rigid structure, whereas outside of this region RDCs adopt smaller negative ($^1D_{NH}$) or positive values ($^1D_{C\alpha H\alpha}$), consistent with a largely mobile polypeptide chain that orients relative to an external director[26] (see below). In order to test the structural similarity of foldon A-state monomer and trimer, orientation tensors were derived by linear fits[34] from the trimer structure and from all $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs observed within segments of ten consecutive amino acid residues. The position of these decapeptide segments was moved through the entire foldon sequence, i.e. spanning Y2–Q11, I3–A12, …, G18–L27, and orientation tensors were derived for each. Theoretical RDCs were then back-calculated from these 17 orientation tensors and compared to all observed RDCs. Figure 3C shows the NMR Q-factors ($=\mathrm{rms}(D^{obs}-D^{calc})/\mathrm{rms}(D^{obs})$[35]) for the different decapeptides as a function of their average position. Clearly, the Q-factors have distinct minimal values ($\sim$0.36) when the decapeptides are taken from the region between G10 and S24. Outside of this region, the Q-factors increase strongly, e.g. to values around 0.7 at the N terminus. This simple comparison proves that the A-state monomer structure is very similar to the trimer structure within the hairpin, but that it is different and unstructured elsewhere. This finding is illustrated in Figure 3A and B, which show calculated $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs for the entire foldon based on an orientation tensor that was derived from RDCs of the hairpin region only (residues A12–L23). Within the hairpin, the calculated and observed RDCs are almost identical. Elsewhere the calculated RDCs vary strongly and deviate significantly from the observed RDCs, which are more uniform and very moderate in size.

Despite of the inherent dynamics of the acid collapsed monomeric state and the low protein concentration ($<$0.2 mM) that is necessary to populate the monomer at pH 2, it was possible to determine its three-dimensional structure (Figure 4) from a limited set of distance, dihedral angle, and RDC restraints (Table 1). The structure is well defined with a heavy atom backbone RMSD of 0.29 Å for residues G10–L23 representing the hairpin and a small, kinked N-terminal extension (G10, Q11). Within the hairpin, the β-sheet H-bonds show regular in-line geometry. This is also directly evident from the NH^N RDCs for V14/L22, V21/R15, and K16/W20 (Figure 3A), which have pairwise very similar values due to the identical orientation of donor and acceptor amide groups within the H-bonds. As no NOEs are observed that would contradict this structural model, we conclude that other conformations are not highly populated.

Moreover, no NOE interactions were detectable that would support stable tertiary contacts between the hairpin and the unstructured N and C-terminal parts of the foldon monomer. The absence of such contacts is also evident from details of the chemical shifts. Thus, the ^1H chemical shift of the A6 methyl group is strongly shifted to -0.6 ppm in the trimer because of the packing against the side-chains of Y13 and W20, but adopts a random coil value of 1.2 ppm in the monomer. The disintegration of the trimer also induces some rearrangements within the hairpin side-chains. In particular, the variations in the NOE patterns show that the hydrophobic packing of the side-chains of Y13, R15, W20, and L23 is less compact in the monomer than in the trimer (Figure 4B and C).

The structure of the foldon monomer shows that the β-sheet hairpin conformation is intact even without further tertiary monomer or quartery

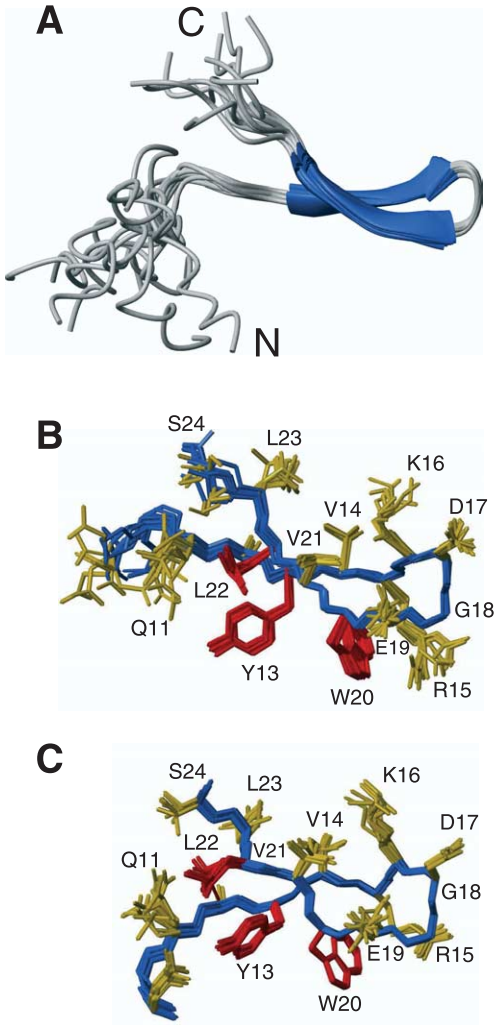


Figure 4. Solution structure of the A-state foldon monomer. A, Backbone worm representation of the ten lowest energy structures of the foldon monomer (PDB-code, 1U0P). B, The hairpin part of the ten lowest energy structures of the foldon monomer in the A-state in stick representation. Side-chains of residues Y13, W20 and L22, which form a hydrophobic cluster, are shown in red. For comparison, the hairpin part of the ten lowest energy structures of foldon within the native trimer (1RFO) is shown in C.

Table 1. Statistics of the foldon A-state NMR structure

RMSDs from experimental distance constraints[a] (Å) all (175)	0.068 ± 0.002
RMSDs from dihedral constraints (°) (28)[b]	1.0 ± 0.2
RMSDs from $^3J_{HNHA}$ coupling constraints (Hz) (20)	1.3 ± 0.1
NMR quality factor Q[c]	0.22 ± 0.01
Deviation from the idealized covalent geometry:	
bonds (Å)	0.0060 ± 0.0001
angles (°)	0.69 ± 0.02
impropers[d] (°)	0.50 ± 0.03
Coordinate precision (Å)[e]	
backbone non-hydrogen atoms	0.29
all non-hydrogen atoms	0.75
Percentage of non-gly, non-pro residues in Ramachandran regions:[f]	
core	92.2
allowed	7.8
generous	0.0
disallowed	0.0

The statistics were obtained from a subset of the ten best energy structures out of 100 calculated following the standard CNS[70] simulated annealing protocol with dipolar restraints incorporated. The number of the various constraints is given in parentheses.

[a] Distance constraints comprise 76 intra-residual NOEs, 61 sequential NOEs ($|i-j|=1$), 15 short-range NOEs ($1<|i-j|\leq 5$) and 23 long-range NOEs ($|i-j|>5$).

[b] The dihedral angle constraints comprise 16 ϕ and 12 ψ angles obtained from TALOS.[47]

[c] The NMR quality factor Q is defined as the ratio of the RMSD between observed and calculated couplings and the RMS of the observed couplings.[35] RDCs consist of 15 ^{1}H–^{15}N and 14 ^{1}H–^{13}C dipolar one-bond couplings from residues 8–22.

[d] The improper torsion restraints serve to maintain planarity and chirality.

[e] 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 10–23, i.e. for core residues, which show maximal {^{1}H}-^{15}N heteronuclear NOE values in the foldon A-state.

[f] These values are calculated with the program PROCHECK-NMR[72] for residues 10–23.

trimer interactions. Thus, foldon has an autonomously folding β-hairpin motif. Most likely, the foldon A-state structure resembles the monomeric on-path folding intermediate, which is stabilized at low pH by weakening the association reaction. This close resemblance is also evident from the similarity of the A-state and folding intermediate fluorescence spectra. Thus, the isolated foldon monomer with the intact hairpin appears as a stable, low energy intermediate in the folding pathway of the trimer. Early formation of hairpins in protein folding has also been reported, e.g. for staphylococcal nuclease[36] as well as for proteins L and G.[37]

Quantitative analysis of thermal β-hairpin unfolding from chemical shifts

Due to their pronounced structure dependence, chemical shifts are a sensitive and convenient probe for the stability of protein secondary structure elements. The temperature dependence of chemical shifts is usually highly reproducible for small reversibly unfolding peptides and has been used for the quantitative thermodynamic analysis of β-hairpins.[38–40] Here, the thermal equilibrium unfolding of the foldon monomer was followed simultaneously by δ^1H^α, $\delta^{13}C^\alpha$, $\delta^{15}N$ and δ^1H^N chemical shifts obtained from two-dimensional heteronuclear correlation spectra. Whereas δ^1H^α and $\delta^{13}C^\alpha$ primarily report on backbone and to a lesser extent on side-chain conformations, δ^1H^N and $\delta^{15}N$ are also very strongly influenced by the formation of H-bonds.[32,41]

All of the monitored chemical shifts within the hairpin show an apparent two-state transition behavior upon thermal unfolding (Figure 5). As the two states are in fast chemical exchange, observed shifts are a weighted average of both, folded and unfolded states. Thus, the temperature dependence of the measured chemical shifts could be fitted (Figure 5, continuous lines) to the functional form:

$$\begin{aligned}\delta X &= p_u\delta X_u + p_f\delta X_f \\ &= \delta X_u + (\delta X_f - \delta X_u)[\exp(-\Delta G^0/RT)]/ \\ &\quad [1+\exp(-\Delta G^0/RT)]\end{aligned} \tag{2}$$

where δX, δX_u, and δX_f are the observed, unfolded state, and folded state chemical shifts of nucleus X, p_u and p_f are the populations of the unfolded and folded state with $p_u+p_f=1$, and ΔG^0 is the free energy of folding obtained from the Gibbs–Helmholtz formula[42] as:

$$\begin{aligned}\Delta G^0 &= (\Delta H^0 + \Delta C_p^0(T-298\text{ K})) \\ &\quad - T(\Delta S^0 + \Delta C_p^0 \ln(T/298\text{ K}))\end{aligned} \tag{3}$$

The fits according to equations (2) and (3) did not prove very robust with respect to the value of ΔC_p. Since ΔC_p is usually very small for thermal denaturation of β-hairpins,[40] this parameter was approximated as $\Delta C_p=0$ to minimize errors in fitting of the other thermodynamic parameters.

The behavior of the chemical shifts in Figure 5 shows a high population of the folded A-state

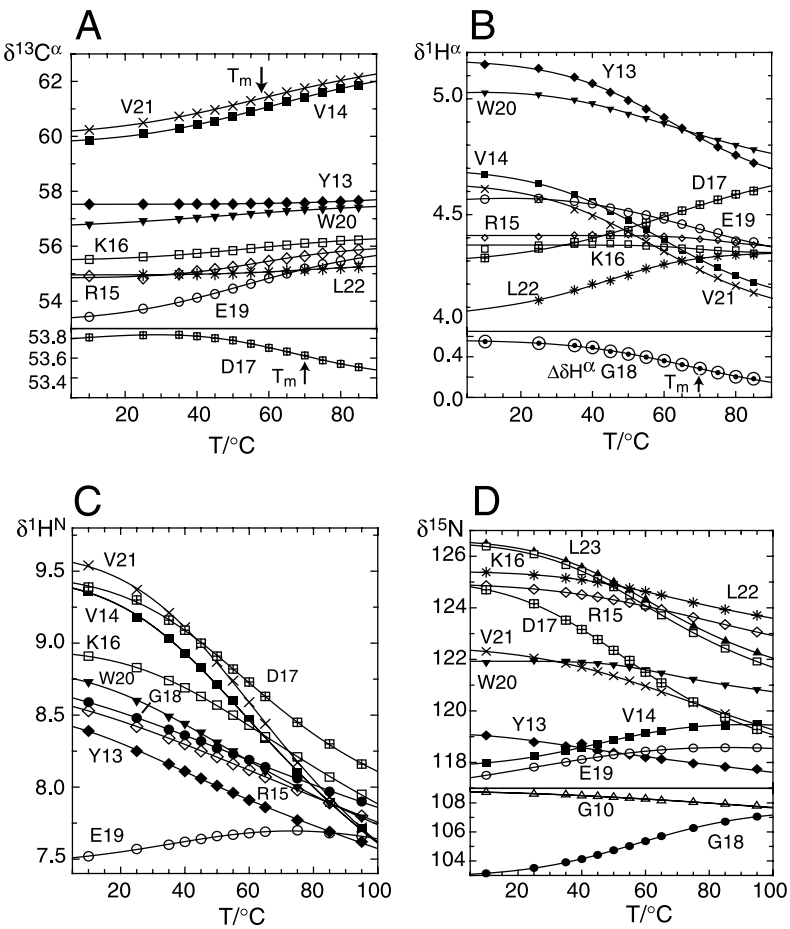


Figure 5. Residue-specific thermal transitions in the hairpin as monitored by backbone chemical shifts of $^{13}C^{\alpha}$ (A), $^{1}H^{\alpha}$ (B), $^{1}H^{N}$ (C), and ^{15}N (D) at (pH 2). Continuous lines are fits to the Gibbs–Helmholtz equation assuming two-state exchange. The derived thermodynamic parameters are given in Table 3.

β-hairpin at room temperature. Thus, the chemical shifts of this state can be determined to high accuracy. Fitted δX_f values correspond to a well-folded structure and are very similar to the shifts of the trimer structure. In contrast, the fitted δX_u values are close to random coil chemical shifts. Table 2 lists as an example the fitted δX_f, δX_u, trimer, and random coil values for the $^{13}C^{\alpha}$ shifts.

The fit of the thermal $^{1}H^{\alpha}$ and $^{13}C^{\alpha}$ chemical shift transitions yields apparent average folding enthalpies $\Delta H^0 = -50.7$ kJ/mol and folding entropies $\Delta S^0 = -152$ J/(mol×K) in H_2O (Table 3) for the hairpin strand residues Y13–K16 and E19–L22. This indicates that the foldon hairpin is populated to almost 90% at room temperature in the A-state. Thus, the hairpin is a highly stable structure with thermodynamic properties in aqueous solution comparable to designed hairpins stabilized by the addition of 50% methanol[39] or 30% trifluoroethanol.[40] Melting temperatures calculated as

Table 2. Comparison of foldon $^{13}C^{\alpha}$ chemical shifts in various states

	Trimer[a]	Monomer folded[b]	Monomer unfolded[b]	Random coil[c]
Y13	58.9	57.4	57.5	57.8
V14	59.4	59.7	62.2	61.9
R15	53.5	54.5	56.1	56.0
K16	55.7	55.6	56.3	56.1
D17	54.7	53.8	53.1	53.0
G18	45.6	45.5	45.5	45.2
E19	53.3	53.3	55.7	55.7
W20	55.9	56.6	57.6	57.5
V21	59.7	59.9	62.5	61.8
L22	55.7	55.0	54.9	55.1
L23	57.4	55.7	55.6	55.1

Values are given in ppm.
[a] Trimer chemical shifts at 25 °C, pH 2.
[b] Values $\delta^{13}C_f^{\alpha}$ and $\delta^{13}C_u^{\alpha}$ derived from the fit of the monomer thermal transition according to the Gibbs–Helmholtz equation (2).
[c] Random coil $\delta^{13}C^{\alpha}$ values derived by taking into account sequence-specific next neighbor effects.[73]

Table 3. Transition temperatures T_m and apparent folding enthalpies ΔH^0 and entropies ΔS^0 of the thermal unfolding of the foldon hairpin

	T_m (K)	ΔH^0 (kJ/mol)	ΔS^0 (J(mol×K))
Turn ($^{13}C^\alpha/^1H^\alpha$)[a]	342.6±0.8	−56.5±7.6	−165.4±22.0
Strands ($^{13}C^\alpha/^1H^\alpha$)[b]	332.6±2.7	−50.7±5.2	−152.4±16.4
Strands ($^1H^N$)[c]	330.5±3.7	−51.0±13.9	−154.4±43.3
Strands (^{15}N)[d]	334.6±5.2	−47.8±9.8	−142.7±29.0

Values are obtained by fits of the Gibbs–Helmholtz equation (2) to the $^{13}C^\alpha$, $^1H^\alpha$, $^1H^N$, and ^{15}N chemical shift transitions. Average values and standard deviations are given.

[a] Values indicate averages and standard deviations for $^{13}C^\alpha$(D17), $^1H^\alpha$(D17) and the glycine methylene proton dispersion Δ^1H^α(G18). The transition amplitude of $^{13}C^\alpha$(G18) is too small for a reliable fit.

[b] Values indicate averages and standard deviations for all $^{13}C^\alpha$ and $^1H^\alpha$ resonances of residues Y13–K16 and E19–L22 with the exception of $^{13}C^\alpha$(Y13) and $^1H^\alpha$(R15, K16), which have too small transition amplitudes for reliable fits.

[c] Values indicate averages and standard deviations for the $^1H^N$ resonances of hydrogen-bonded residues V14, K16, E19 and V21, which undergo significant cooperative transitions upon H-bond melting.

[d] Values indicate averages and standard deviations for all ^{15}N resonances of residues V14–K16 and E19–L22. The transition amplitude of the ^{15}N resonance of Y13 is too small for a reliable fit.

$T_m = \Delta H^0/\Delta S^0$ have values of $T_m = 332.6$ K for strand residues Y13–K16 and E19–L22. In contrast, the melting temperatures for the β-turn resonances of D17 and G18 are increased by 10 K (Table 3), as evidenced from the behavior of the $^{13}C^\alpha$ and $^1H^\alpha$ shift of D17 (Figure 5A) or the frequently used glycine methylene proton dispersion[43] $\Delta\delta^1H^\alpha$ (Figure 5B). An increase in thermal stability of up to 15 K in various turns has been reported from $^1H^\alpha$ chemical shifts in engineered hairpins[40] and interpreted as evidence for local stability and hairpin nucleation from the turn, which is also predicted by kinetic[44] and thermodynamic[45] zipper models. The apparent thermal stability of the turn residues is also consistent with the finding that the sequences DG and NG have the highest propensities for type 1 β-turns in the RCSB Protein Data Bank.[46]

$^1H^N$ chemical shifts are primarily reporters of H-bond distances.[32,33] The $^1H^N$ shift dispersion of the foldon A-state hairpin at room temperature with mostly downfield-shifted δ^1H^N shows the presence of well-formed H-bonds (Figure 5C). This differs from an incidental observation[19] that certain β-hairpins form collapsed states similar to well-folded structures but with ill-defined H-bonds. Monomer $^1H^N$ shifts of residues K16 and E19 near the turn are almost identical to the trimer and indicate that these H-bonds are very similar in geometry and population as in the native structure. This agrees with predictions from off-lattice models of β-hairpin formation[45] that the near-turn H-bonds are particularly highly populated and conformationally more stable than the neighboring in-line H-bonds. The temperature dependence of the $^1H^N$ shifts shows that the H-bonds melt in a cooperative manner during thermal unfolding with thermodynamic parameters (Table 3) close to those for $^1H^\alpha$ and $^{13}C^\alpha$, but slightly reduced transition temperatures (see also below). A more extended interpretation of these parameters is, however, difficult due to the simultaneous dependence of $^1H^N$ chemical shifts on H-bonding both within the polypeptide chain and with the solvent.

The thermal transitions of the hairpin ^{15}N chemical shifts (Figure 5D) are qualitatively and quantitatively very similar to the $^1H^\alpha$, $^{13}C^\alpha$, and $^1H^N$ chemical shifts (Table 3). However, a structural interpretation is even more difficult than for $^1H^N$ chemical shifts, because the ^{15}N chemical shifts depend not only on the direct H-bonds but also on the H-bonds to the preceding carbonyl group, as well as on backbone and side-chain conformations.[41] Noteworthy observations are the strong down- and upfield shifts of D17 and G18 at room temperature. Such strong down- and upfield shifts are characteristic for type 1 β-turns.[47] Accordingly, in all low-energy structures calculated of the monomer, residues D17 and G18 are in canonical type 1 β-turn conformation, although no dihedral angle restraints had been imposed on the turn during structure calculations. During the melting transition, the ^{15}N shifts of D17 and G18 move towards random coil values (Figure 5D). This clearly indicates an opening of the turn structure above the transition temperature.

Quantitative analysis of thermal β-hairpin unfolding from residual dipolar couplings

Chemical shifts and NOEs are widely used to characterize changes in local structures during folding transitions of polypeptides. However, due to the complicated structural dependencies of chemical shifts and the influence of local dynamics and spin diffusion on NOEs, the variations in these NMR parameters are in general difficult to relate to precise changes of geometric parameters. A characterization of folding phenomena based on RDCs offers various advantages: first, the measured dipolar coupling D_{IS} between two nuclei I and S represents a straightforward ensemble and time average over the relative nuclear conformation:

$$D_{IS} = -\frac{\gamma_I\gamma_S\hbar\mu_0}{4\pi^2 r^3}\left\langle\frac{3\cos^2\theta - 1}{2}\right\rangle = D_{IS}^{\max}\langle P_2(\cos\theta)\rangle \quad (4)$$

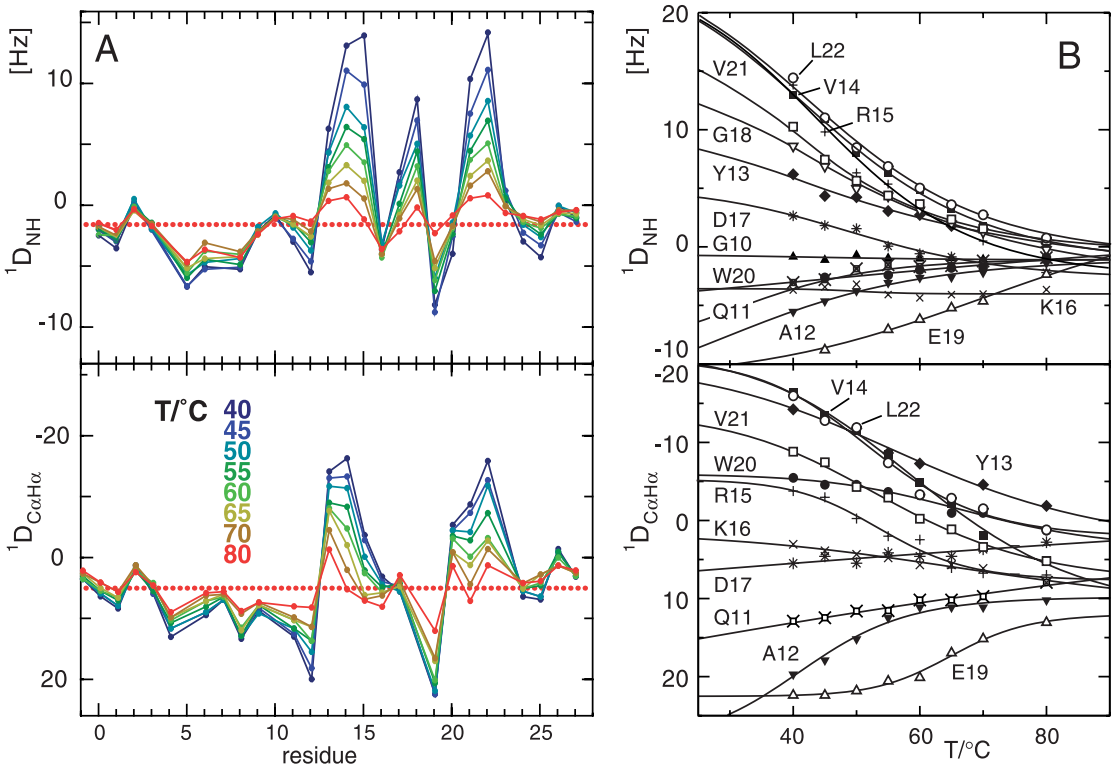


Figure 6. A, $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs in the foldon A-state between 40 °C and 80 °C. Vertical scales are adjusted to $^1D_{C\alpha H\alpha}^{max,solid\text{-}state} = -1.98\ ^1D_{NH}^{max,solid\text{-}state}$ such that the identical order parameters for $^{13}C^{\alpha}$–$^1H^{\alpha}$ and ^{15}N–^1H vectors result in identical vertical positions. The averages of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs at 80 °C are indicated as dotted red lines. B, Temperature dependence of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ fitted to two-state models.

where θ presents the angle between the inter-nuclear distance vector and the external magnetic field, r is the inter-nuclear distance (assumed as constant, e.g. within a stable chemical bond), and P_2 is the second-order Legendre polynomial. Thus, the RDCs correspond to well-defined order parameters. Second, long-range information on the relative orientation of different inter-nuclear vectors is contained in the RDC data, because the angle θ is determined relative to a common external direction, i.e. the magnetic field.[20] Thus, also non-local order can be monitored during the folding transition. Third, RDCs can be measured with high sensitivity and accuracy for a large number of sites. Therefore, a detailed description of the folding transition should be possible.

In order to characterize the thermal unfolding of the foldon A-state monomer, the temperature dependence of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs was determined in inert polyacrylamide gels. Figure 6 shows these data for temperatures between 40 °C and 80 °C (below 40 °C, the monomer population is too low for a precise determination of RDCs). It is evident that the large RDCs, which are observed at low temperatures within the hairpin region, converge to small, but non-zero values above the melting transition. Furthermore, the unfolded N and C-terminal stretches outside of the hairpin region also adopt small, non-zero RDC values across the entire temperature range. The all-residue average at 80 °C (shown as red dotted lines in Figure 6A) amounts to −1.4 Hz for $^1D_{NH}$ and 5.0 Hz for $^1D_{C\alpha H\alpha}$ RDCs, respectively. Apparently,

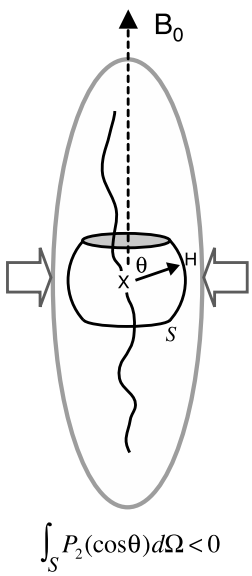


Figure 7. Schematic representation of the residual order for an elongated random chain that is confined within an elongated cavity. This situation is expected in horizontally compressed acrylamide gels (see the text).

the inter-nuclear N–H^N or C^α–H^α vectors have non-isotropic distributions even above the melting transition, such that $\langle P_2(\cos\theta)\rangle$ does not vanish. This behavior has been explained recently for weakly aligned, non-spherical random polymer chains.[26,48] The restrictions of bond movements due to the covalent geometry lead to non-vanishing RDCs even when the torsion angles along the polymer chain adopt random conformations. Figure 7 depicts this situation for an extended random peptide chain that is aligned parallel to the magnetic field in an elongated cavity. Similar conditions are expected within the horizontally compressed polyacrylamide gels of the present study. Under these conditions, the N–H^N and C^α–H^α (X–H) vector movements are restricted to an equatorial section of the unit sphere, and the average $\langle P_2(\cos\theta)\rangle$ becomes negative. Thus, due to the opposite signs of γ_N and γ_C that enter into equation (4), $^1D_{NH}$ RDCs should be negative, whereas $^1D_{C\alpha H\alpha}$ should be positive. The signs of the average $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDC values at 80 °C (Figure 6A) agree with these predictions. An inversion of the signs is observed when the foldon A-state monomers are aligned by longitudinally compressed polyacrylamide gels, which yield opposite sign orientation tensors[22] (data not shown).

The question arises to what extent the alignment properties of the gel itself change with temperature. This was investigated on a sample of gel-aligned ^{15}N-labelled ubiquitin (pH 4.7, data not shown). Over the temperature range of 20, 35, 50, 65, 80 °C, the norm of the alignment tensor[34] varied by less than 5% around its mean value, whereas the correlation coefficient between all tensors was larger than 99%. Thus, the temperature-induced changes of the gel alignment properties are very small and can be neglected relative to the large structural changes observed for the hairpin.

Thermal transitions of the RDCs in the hairpin strands fitted to two-state behavior (Figure 6B) yield average melting temperatures of 324.8 (±6.2) K for $^1D_{NH}$ and 330.7 (±3.9) K for $^1D_{C\alpha H\alpha}$ that are close to, but slightly lower than the transition temperatures obtained from the $^1H^\alpha$ and $^{13}C^\alpha$ chemical shifts (Table 3). $^1D_{NH}$ RDCs for residues V14–L22 and V21–R15 show almost parallel unfolding profiles with even lower average transition temperatures of 320.4 (±2.8) K. This proves that the strong H-bonds between V14 and V21, which connect the peptide groups of these residues, open in a cooperative way and that on average, the in-line geometries of the respective H-bond donor and acceptor groups remain virtually unchanged during the entire unfolding process. The lower transition temperatures for these RDCs as compared to the $^1H^\alpha/^{13}C^\alpha$ chemical shifts could be an indication that the V14↔V21 H-bonds melt before the backbone looses its specific conformation. A number of other RDCs also show considerable deviations from average behavior (Figure 6B). Thus, the transition temperatures of $^1D_{C\alpha H\alpha}$ and $^1D_{NH}$ RDCs for A12 are noticeably lower than average, possibly because of the location of this residue at the beginning of the hairpin and the rotational freedom of the small alanine side-chain. In contrast, the transition temperatures for $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs of near-turn residue E19 are significantly increased, whereas the RDCs of its H-bond partner K16 show almost no temperature dependence. The latter behavior is further evidence for a particular stabilization of the hairpin near the turn.

At 80 °C individual RDCs still show deviations from the respective $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ average values (Figure 6A), which are considerably larger than the reproducibility error of the experiments (<0.3 Hz for $^1D_{NH}$ and <0.8 Hz for $^1D_{C\alpha H\alpha}$). These deviations reveal residue-specific orientational order above the melting transition. Part of this order may be explained by the residual population of folded hairpin expected from the two-state folding model. At 80 °C, the folded population amounts to ~27% (~77% at 40 °C) as calculated from the average thermodynamic parameters of the $^1H^\alpha/^{13}C^\alpha$ chemical shift transitions. However, the non-uniform RDC values at 80 °C do not correspond to a simple population scaling of the 40 °C values. Thus for example, the reduction in RDCs at 80 °C is larger than expected from the $^1H^\alpha/^{13}C^\alpha$ chemical shift transitions for hairpin residues V14, R15, and V21. This is consistent with the low transition temperatures observed for these RDCs. In contrast, the reduction for residues close to the N-terminal proline residues P4 and P7 is considerably smaller. For these residues, the persistent and strong deviations from the average $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDC values at 80 °C indicate a restriction of conformational freedom, which likely arises from the presence of the two proline residues.

A qualitative approach to analyze such deviations from uniform RDC values was proposed by Louhivuori and co-workers[48] in a recent theoretical study on the alignment of random peptides: as the polypeptide chain exhibits a certain stiffness, consecutive fragmental alignment frames should not change too abruptly in a random chain. Therefore, sudden changes in the values of measured RDCs along the polypeptide chain indicate true residual order of the inter-nuclear vector orientations. According to this criterion, the abrupt sequential changes in $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ around proline residues P4 and P7 (Figure 6A) indicate non-random conformations. Indeed, strong effects of proline residues have been observed in an all-atom simulation of heteropolymers containing proline residues.[48] More interestingly, a certain fine structure becomes apparent also within the hairpin sequence. Thus, strong sequential changes for both $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ are detected at the beginning (A12/Y13) and end of the hairpin (L22/L23) as well as at the turn (K16–W20). A noteworthy observation is the strong sequential changes in $^1D_{C\alpha H\alpha}$ for residues Y13/V14, W20/V21, and V21/L22. These strong changes are not present at 40 °C. The deviations involve residues of the hydrophobic

cluster Y13/W20/L22 and present evidence for structural order of these residues above the melting transition. Thus, conformations favoring the hydrophobic collapse of the cluster may be particularly stabilized. Presumably, the structural order of these large, hydrophobic residues above the melting transition arises from local side-chain–backbone interactions, which force the side-chains into extended conformations and restrict the conformational space accessible to random polypeptide chains.[49] Accordingly, also the RDCs of residue Y2 within the unstructured N terminus exhibit such large sequential deviations over the entire temperature range.

A more quantitative description of conformation and local order during the thermal transition is clearly desirable. In practice, such an analysis is complicated due to the convolution of local and global order of the molecule during the unfolding transition. The global order resulting from the weak alignment of an entire molecule is usually described in terms of the Saupe order matrix, which can be written in irreducible form as S_m ($m = -2, -1, \ldots, 2$).[50] For rigid molecules in the absence of local disorder, expected RDC values are derived from S_m as:

$$D_{IS} = D_{IS}^{\max}\sqrt{\frac{4\pi}{5}}\sum_{m=-2}^{2} S_m^* Y_{2m}(\theta,\phi) \qquad (5)$$

where Y_{2m} are second-order spherical harmonics and θ and ϕ are the polar angles of the inter-nuclear distance vector in a molecule-fixed coordinate system. Local disorder can be incorporated into equation (5) as an average over a local angular probability $p(\theta,\phi)$:

$$\begin{aligned} D_{IS} &= D_{IS}^{\max}\sqrt{\frac{4\pi}{5}}\sum_{m=-2}^{2} S_m^*\langle Y_{2m}(\theta,\phi)\rangle \\ &= D_{IS}^{\max}\sqrt{\frac{4\pi}{5}} \\ &\quad\times\sum_{m=-2}^{2} S_m^*\int p(\theta,\phi)Y_{2m}(\theta,\phi)\sin\theta\,d\theta\,d\phi \end{aligned} \qquad (6)$$

It is easy to show that local diffusion with cylindrical symmetry around an average direction (θ_0,ϕ_0) simply reduces D_{IS} by the local order parameter S_l:

$$D_{IS} = S_l D_{IS}^{\max}\sqrt{\frac{4\pi}{5}}\sum_{m=-2}^{2} S_m^* Y_{2m}(\theta_0,\phi_0) \qquad (7)$$

For diffusion in a cone of half-angle $\Delta\theta$, S_l is given by:[51]

$$\begin{aligned} &S_l = \frac{\cos\Delta\theta}{2}(1+\cos\Delta\theta)\Leftrightarrow \\ &\Delta\theta = \cos^{-1}\left[\frac{1}{2}(\sqrt{1+8S_l}-1)\right], \\ &\text{for } 0 \le \Delta\theta \le 90^\circ \end{aligned} \qquad (8)$$

Equation (7) shows the principal difficulty that local order parameters can only be derived from the measured RDCs provided that the global order tensor S_m and the local inter-nuclear directions are known. As the structure of the molecule, and hence global order tensor and local geometry, change during unfolding, the problem of a complete description of local order and geometry is underdetermined from the currently measured number of ${}^1D_{\mathrm{NH}}$ and ${}^1D_{\mathrm{C}\alpha\mathrm{H}\alpha}$ RDCs.

An approximate change in the local order parameter may be obtained from the assumption that the global structural change is small and that local changes result only from an increase in disorder, but not from a change of the average inter-nuclear direction. This situation may be realized for small temperature changes at the onset of thermal denaturation, where the small increases in local fluctuations should lead to only minor changes of the overall structure and molecular alignment. As a result, the ratio of residual dipolar couplings D_{IS} measured at two temperatures T_1 and T_2 should be an approximate measure for the ratio of local order parameters:

$$\frac{D_{IS}(T_2)}{D_{IS}(T_1)} \approx \frac{S_l(T_2)}{S_l(T_1)} \qquad (9)$$

The assumption of only limited structural changes at the onset of the thermal denaturation of the foldon hairpin can be verified to a certain extent from the Pearson correlation coefficient of the measured ${}^1D_{\mathrm{NH}}$ and ${}^1D_{\mathrm{C}\alpha\mathrm{H}\alpha}$ RDCs. Using the values

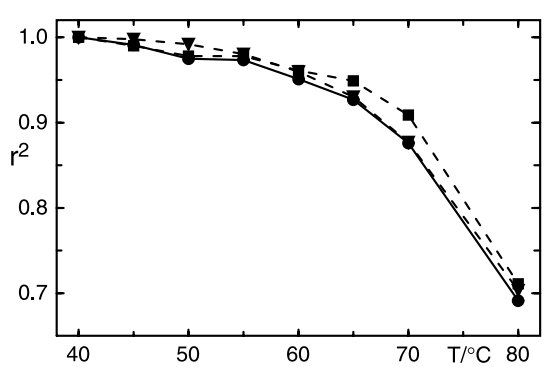


Figure 8. Pearson correlation coefficient r^2 between the measured hairpin RDCs at different temperatures. The correlation coefficient is calculated relative to the data measured at 40 °C. Filled squares and triangles indicate the correlation coefficients for the ${}^1D_{\mathrm{NH}}$ and ${}^1D_{\mathrm{C}\alpha\mathrm{H}\alpha}$ RDCs, respectively. Filled circles indicate the correlation coefficients of the combined RDCs with ${}^1D_{\mathrm{C}\alpha\mathrm{H}\alpha}$ values divided by −1.98.

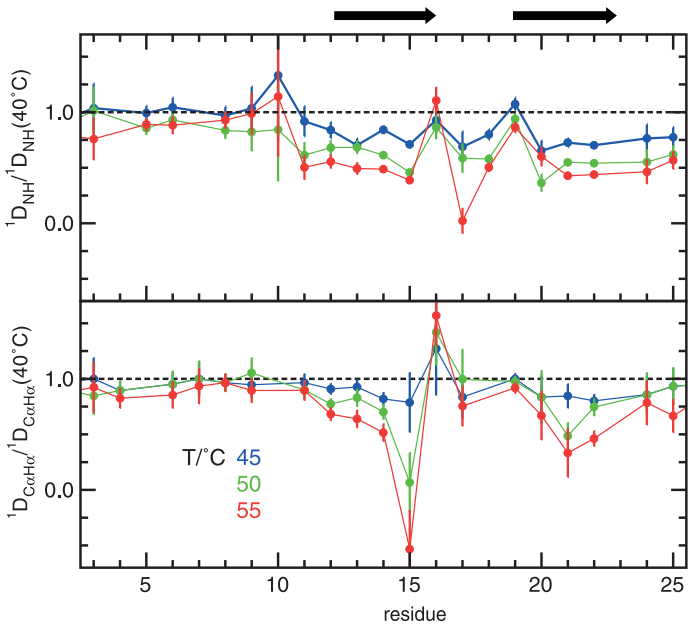


Figure 9. Measured $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs in the foldon A-state for temperatures between 45 °C and 55 °C divided by the respective values at 40 °C. These RDC ratios represent approximate values of the ratios of local order parameters at the respective temperatures (see the text). The positions of the two β-strands within the A-state structure are shown at the top.

at 40 °C as a reference, the correlation coefficient decreases only very moderately to ~0.93 at 65 °C (Figure 8), whereas a more pronounced drop to ~0.7 is observed over the temperature range from 65 °C to 80 °C. Correlation coefficients determined separately for either $^1D_{NH}$ or $^1D_{C\alpha H\alpha}$ RDCs show an identical behavior (Figure 8). Thus, the RDC correlation coefficients provide evidence for a certain robustness of the overall structure in the temperature range from 40 °C to 65 °C and hence the applicability of equation (9).

Figure 9 shows the relative changes of local order parameters $S_l(T_2)/S_l(T_1)$ from 40 °C to 55 °C as approximated according to equation (9) by the ratios of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs using the values at 40 °C as a reference. The changes are fairly uniform for residues A12 to V14 and V21 to T25. They correspond to a relative decrease in $N–H^N$ ($C^\alpha–H^\alpha$) order parameters of about 50% (40%) over the temperature range from 40 °C to 55 °C. As the RDCs for these residues cover a wide range of values at 40 °C and hence a wide range of inter-nuclear directions, the uniformity of the relative changes further supports the assumption that they result from decreased local order parameters and not from changes of the global order tensor or changes of the average inter-nuclear direction. Relative changes in RDC values for the unstructured N-terminal region from residue I3 to D9 are also rather uniform, but amount to only about 10–20% reduction in order over the same temperature range. In contrast to this uniform behavior, near-turn residues K16 and E19 show a considerably weaker relative reduction or even slight increases of the order parameters (Figure 9) consistent with a higher stability of the turn conformation. It is also evident from the behavior of $^1D_{C\alpha H\alpha}$ for R15 that the approximation of a strictly local order decrease fails already at 50 °C for this RDC, since its value changes sign at this temperature.

Relaxation data and folding kinetics

In order to characterize the dynamics of the foldon A-state monomer by NMR, ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates as well as {^1H}-^{15}N steady-state NOEs (Figure 10A) were determined below (25 °C) and near the hairpin folding transition (55 °C). The analysis of the ^{15}N relaxation data by the standard model-free formalism[52–54] (not shown) yields the sub-nanosecond order parameter S^2, the global molecular rotational correlation time τ_c, and contributions from chemical exchange R_{2ex} to the transverse relaxation rate R_2. At 25 °C, S^2 values are in a range of 0.6–0.85 for all hairpin residues, and $\tau_c = 2.1$ ns, consistent with isotropic rotational correlation times expected for monomeric foldon. At 55 °C, S^2 decreases to values of about 0.4 ($\tau_c = 1.7$ ns). The size of the S^2 parameters indicates that the hairpin $N–H^N$ vectors have rather low mobility on the pico- to nanosecond time scale at 25 °C, but the amplitude of such fast motions increases considerably at 55 °C.

The model-free analysis also reveals considerable exchange contributions R_{2ex} (Figure 10B). These are attributed to the exchange between folded and unfolded hairpin, which is fast on the chemical shift time scale (Figure 5). Under these conditions, the equilibrium exchange rate ($k_{ex} = 1/\tau_{ex} = k_f + k_u$) of a two-state folding reaction:

$$U \underset{k_u}{\overset{k_f}{\rightleftharpoons}} F \qquad (10)$$

is large compared to the chemical shift difference of the folded and unfolded states ($k_{ex} \gg |\Delta\omega| = |\omega_f - \omega_u|$). The chemical exchange contribution R_{2ex} to the

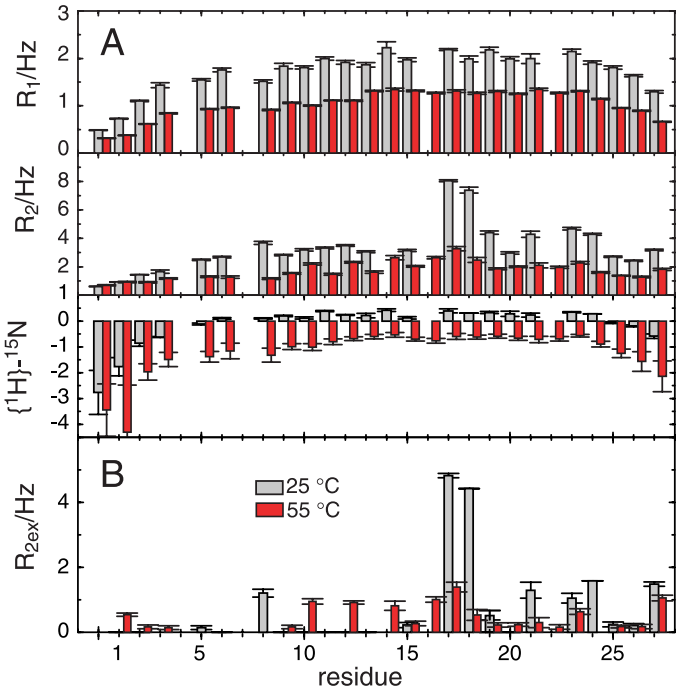


Figure 10. ^{15}N relaxation data of the foldon A state at 25 °C (grey) and 55 °C (red). A, Measured R_1, R_2 (CPMG) rates and {^{1}H}-^{15}N NOEs. B, Chemical exchange contributions R_{2ex} obtained by the model-free analysis.

transverse relaxation rate R_2 is then given by:[53,55,56]

$$R_{2ex} = p_f p_u (\omega_f - \omega_u)^2 \tau_{ex}. \quad (11)$$

where p_f and p_u are the equilibrium populations of states F and U, respectively, and $p_f + p_u = 1$.

It is evident from equation (11), that the populations of the two states, the frequency difference, and the exchange time cannot be derived independently from a single determination of R_{2ex}. However, in the case of the A-state foldon, populations and chemical shift differences are known from the thermal transitions of the chemical shifts (Figure 5). Therefore, the exchange times τ_{ex} can be calculated directly from the measured exchange contributions R_{2ex}. Values for p_f and p_u were taken from the thermal transitions as $p_f = 0.85$, $p_u = 0.15$ at 25 °C, and $p_f \approx p_u \approx 0.5$ at 55 °C, respectively.

Table 4 lists the determined exchange terms R_{2ex} together with the chemical shift differences $\Delta\omega(^{15}N)$ derived from the thermal transitions of the ^{15}N resonances (Figure 5D). Sizeable exchange terms are observed for residues that also show significant chemical shift differences. This applies to six A-state

Table 4. Exchange times between folded and unfolded states obtained for foldon hairpin residues

Res	25 °C				55 °C			
	$\lvert\nu_f-\nu_u\rvert$ (Hz)[a]	R_{2ex} (Hz)[b]	ΔR_{2ex} (Hz)[c]	$\tau_{ex}/\mu s$[d]	$\lvert\nu_f-\nu_u\rvert$ (Hz)[a]	R_{2ex} (Hz)[b]	ΔR_{2ex} (Hz)[c]	$\tau_{ex}/\mu s$[d]
13	21	–[e]	–	–	36	–[e]	–	–
14	186	Not fit[f]	–	–	171	0.72	0.10	2.5
15	104	0.33	0.10	6.1	119	0.24	0.04	1.7
16	293	Overlap[g]	–	–	308	0.98	0.04	1.1
17	361	4.79	0.04	7.3	376	1.41	0.02	1.0
18	365	4.40	0.03	6.6	350	0.65	0.12	0.5
19	153	0.64	0.15	5.4	138	0.19	0.06	1.0
20	88	–[e]	–	–	103	0.23	0.03	2.5
21	214	1.56	0.26	6.7	230	0.29	0.02	0.6
22	88	Overlap[g]	–	–	103	0.16	0.03	1.5
23	270	1.99	0.10	5.4	285	0.60	0.10	0.8

[a] Values are derived from two-state transitions of nitrogen shifts. $\lvert\nu_f - \nu_u\rvert = \lvert\omega_f - \omega_u\rvert/2\pi$ is the ^{15}N shift difference between the folded and the unfolded state. Values are temperature-corrected assuming a temperature-independent shift of the folded state and a linear temperature factor −0.5 Hz/°C for the unfolded state. These assumptions derive from nearly unchanged native trimer chemical shifts between 25 ° C and 55 °C and from the behavior of unfolded residues near the N and C terminus of the monomer. The corrections have only a very limited influence on τ_{ex}.

[b] Exchange rates R_{2ex} obtained from model-free spectral density fitting of T_1, T_2 and {^{1}H}-^{15}N NOE data (Figure 10) at pH 2.

[c] Errors in R_{2ex} estimated from the spectral density fit.

[d] The exchange time is obtained as: $\tau_{ex} = R_{2ex}/[4\pi^2 p_f p_u (\nu_f - \nu_u)^2]$.

[e] Analysis of relaxation data did not result in sizeable R_{2ex} terms for nuclei with small $\lvert\nu_f - \nu_u\rvert$ values.

[f] The model-free spectral density fitting of V14 relaxation data at 25 °C did not converge.

[g] Reliable ^{15}N relaxation data could not be determined due to spectral overlap.

hairpin residues, which do not overlap with trimer resonances at 25 °C, and to almost all hairpin residues (V14–L23) at 55 °C. For these cases, Table 4 lists the exchange times τ_{ex} calculated by equation (11) from the known equilibrium populations and frequency differences. At 25 °C average τ_{ex} values are 6.4 (±1.0) μs, corresponding to rate constants $k_f = 1.3(\pm 0.2) \times 10^5\ s^{-1}$ and $k_u = 3.0(\pm 0.4) \times 10^4\ s^{-1}$ for hairpin folding and unfolding, respectively. At 55 °C average τ_{ex} values are 1.5(±0.7) μs corresponding to $k_f = k_u = 4.7(\pm 2.4) \times 10^5\ s^{-1}$. Turn residue G18 exhibits particularly fast exchange with $\tau_{ex} = 0.5$ μs at 55 °C, possibly due to its conformational freedom in dihedral angle space. Studies of an isolated hairpin construct consisting only of residues A12–L23 revealed comparable hairpin formation and very similar exchange times of 6.6 (±0.5) μs at 25 °C as determined by $^1H^N$ relaxation rates (data not shown).

For the foldon A-state peptide, the chemical exchange contributions are mainly observed for contiguous stretches of hairpin residues. It is therefore very likely that the chemical exchange corresponds to the global folding–unfolding reaction of the hairpin. The folding times of a few microseconds are consistent with reaction times obtained by temperature-jump optical spectroscopy for hydrophobic cluster formation in GB1[18] and H-bond formation in engineered hairpins like trpzip4 and peptide I,[57] as well as with microsecond chemical exchange times in model hairpins derived from an analysis of ^{13}C NMR relaxation data.[58,59]

Conclusion

It has been proposed that foldon is an evolutionary optimized folding and trimerization domain,[60] as its only known biological function is to promote the correct folding of the fibritin trimer. Foldon's very efficient folding pathway was characterized in a previous kinetic investigation.[7] The association starts from a monomeric intermediate with partially buried hydrophobic side-chains and leads to the fully formed trimer with rates that are close to the fastest known protein dimerization reactions.[7] At pH 2, the trimer disintegrates into a monomer, which has similar fluorescence properties as the monomeric folding intermediate.

Here, the formation of this A-state monomer from the trimer, its structure, equilibrium association and folding dynamics, and its stability have been characterized to atomic detail by solution NMR. The destabilization of the foldon trimer at low pH could be traced to charge neutralization of residue E5 in the strong inter-subunit R15–E5 salt bridge. Equilibrium constants and association energies for this pH-dependent association reaction were quantified. The structural analysis by RDC and other NMR parameters shows that the foldon A-state monomer consists of a β-hairpin with intact and stable H-bonds that is similar to the monomer in the foldon trimer, but lacks a defined structure in its N and C-terminal parts. The hairpin is strongly twisted, presumably in order to provide suitable surfaces for the fast and specific association of the foldon trimer and to avoid the formation of extended β-sheets by competing reactions.

The thermal unfolding of the A-state β-hairpin was monitored by $^1H^\alpha$, $^{13}C^\alpha$, $^1H^N$ and ^{15}N chemical shifts, which yield average transition temperatures around 333 K. The derived folding enthalpies and entropies indicate that the foldon hairpin is highly populated in aqueous solution at room temperature. Its thermodynamic stability in pure water is comparable to designed hairpins that are stabilized in alcohol/water mixtures. The chemical shifts indicate increased local structural stability at the turn, as has been proposed for various hairpins from thermal $^1H^\alpha$ chemical shift transitions[40] and theoretical models.[45] The exchange between folded and unfolded forms occurs within a few microseconds as derived from the chemical exchange contribution to the ^{15}N transverse relaxation times. These folding times are comparable to observations for other hairpins.[18,57–59]

New qualitative and quantitative information on N–H^N and C^α–H^α bond order parameters during thermal unfolding was obtained from $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs obtained in inert, mechanically stretched polyacrylamide gels. The RDC data in the hairpin show a similar temperature dependence as the $^1H^\alpha$ and $^{13}C^\alpha$ chemical shifts, albeit with somewhat lower transition temperatures for the RDCs of the strand residues. This is particularly evident for $^1D_{NH}$ values from the donor and acceptor peptide planes of the two strong H-bonds between residues V14 and V21. This could be an indication that these H-bonds are destabilized before larger conformational changes of the backbone become detectable in the $^1H^\alpha$ and $^{13}C^\alpha$ chemical shifts. Above the melting transition, the RDCs converge to small, but non-zero average values consistent with expectations from oriented random chain models. The sequence-specific variations of these RDCs above the melting transition reveal residual order for many of the N–H^N and C^α–H^α bond vectors. It is particularly evident that residues close to the two N-terminal proline residues P4 and P7, within the hairpin turn, and residues with large hydrophobic side-chain deviate from random chain behavior. The structural preferences within the turn and the hydrophobic cluster of residues Y13, L22, and W20 are likely to direct the hydrophobic collapsed state towards the correct H-bond register of the monomer. It is interesting to note that the structural preferences within the "unstructured" N terminus of the monomer involve residues that form extended inter-monomer contacts within the trimer, i.e. Y2, I3, and E5 (Figure 1). It is thus possible that the trimer interface is already preformed to a certain extent in the monomer.

An attempt was made to translate the temperature-induced changes of the RDCs into changes of local order parameters. This seems possible for

small temperature variations at the onset of the thermal transition, where the high correlation coefficients between the hairpin RDCs measured at different temperatures indicate the preservation of the overall structure. In this situation, the ratio of RDCs measured at two temperatures should be approximately equal to the ratio of the local order parameters. These relative RDC values over the temperature range from 40 °C to 55 °C reveal a rather uniform decrease (40%–50%) of order for many of the β-strand residues and a smaller uniform decrease (~20%) for residues within the unstructured N terminus. The changes can be interpreted as a uniform temperature-induced decrease of local $N–H^N$ and $C^\alpha–H^\alpha$ order parameters. In contrast, near-turn residues K16 and E19 show only very small changes in RDC values and hence in local order parameters at the onset of the unfolding transition.

It has been suggested that the stability of β-turn conformations arises from an interplay of transient hydrophobic cross-strand interactions and local dihedral angle propensities.[61,62] Temporary hydrophobic side-chain contacts may stabilize the bent conformation by the stronger hydrophobic interactions at higher temperatures even after the melting of the H-bonds. This is in agreement with Monte-Carlo simulations, which show that transition temperatures of the hydrophobic collapse in molecular dynamics simulations are higher than the transition temperatures for the formation of backbone secondary structures in hairpins.[45] The present experimental results on the behavior of chemical shifts and RDCs corroborate these predictions for the foldon hairpin with respect to the particular stability of the hydrophobic cluster and the type 1 β-turn.

In summary, a large number of physico-chemical parameters of the foldon monomer–trimer equilibrium and of the thermal denaturation of its low pH monomeric form have been determined from a rigorous combination of modern high-resolution NMR techniques. In particular, $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs provide new information on structural preferences within the unfolded and partially unfolded monomer states. A further increase in the number of determined RDCs is possible, e.g. by the detection of side-chain or long-range RDCs or by using different alignment conditions. This may ultimately lead to a complete quantification of local order in folding–unfolding transitions.

Material and Methods

Purification

U–^{13}C, U–^{15}N-labeled foldon was expressed as a thioredoxin fusion protein in *Escherichia coli* BL21(DE3) from a plasmid based on pET-32a (Novagen).[8] The fusion protein carries an N-terminal 6xHis-tag and a thrombin cleavage site between thioredoxin and the C-terminal foldon. The non-native thrombin cleavage site residues $G^{-1}S^0$ are highly flexible and disordered at both pH 7 and pH 2. The complete sequence of the foldon construct is GS^0 GYIPEAPRDG[10] QAYVRKDGEW[20] VLLSTFL.[27]

Foldon was purified *via* affinity chromatography on a nickel-NTA column (Novagen) before and after thrombin cleavage and was dissolved in 5 mM potassium phosphate buffer (pH 7). Protein concentrations were determined using the method of Gill & von Hippel[63] with an extinction coefficient of 2.56 cm mg^{-1} ml^{-1}.

Heteronuclear NMR

NMR samples were prepared in $^1H_2O/^2H_2O$ 19:1 (v/v) and adjusted to the indicated pH values with a Russel electrode by addition of HCl. All spectra were recorded on a BRUKER DRX 600 spectrometer. Triple resonance assignment and NOE experiments were carried out similar as described.[64] Weak alignment of 0.3 mM [U–^{13}C, U–^{15}N]foldon in acrylamide gels of 7% (w/v) was achieved by horizontal compression in NEW-ERA sample tubes to an aspect ratio of 2.9 as described[65] or by vertical compression with the plunger of a Shigemi NMR tube.[22] $^1D_{NH}$ RDCs were measured with the IPAP experiment,[66] $^1D_{C\alpha H\alpha}$ RDCs were obtained *via* a gradient-selected constant time heteronuclear single quantum coherence (HSQC) without ^{13}C decoupling during acquisition. Errors are estimated as <0.3 Hz for $^1D_{NH}$ and <0.8 Hz for $^1D_{C\alpha H\alpha}$ dipolar couplings from reproducing the measurements. NMR spectra of the foldon A-state in polyacrylamide gels are identical to those in solution, indicating that the gels cause no major structural changes. Temperature calibration was performed as described on a standard polyethyleneglycol sample.[67] ^{1}H, ^{15}N, and ^{13}C chemical shifts were referenced relative to the ^{2}H lock resonance of water for all temperatures. Shifts of side-chain carboxylate resonances were measured at 25 °C with a gradient selected H(CA)CO experiment optimized for side-chain methlyene groups. ^{15}N relaxation data were recorded as described[64] on a sample of 0.3 mM [U–^{13}C, U–^{15}N]foldon. R_2 data were derived from Carr–Purcell–Meiboom–Gill (CPMG) spin-echo experiments with delays of 1 ms between the 180° pulses. Relaxation data were fitted to model-free spectral density functions with the program TENSOR.[54]

Data analysis and structure calculation

All data were processed with the NMRPipe[68] suite of programs and analyzed by the program PIPP.[69] Fits of thermal transitions according to the Gibbs–Helmholtz formula[42] were performed by Monte-Carlo fitting followed by a robust minimization of χ^2 with ProFit (QuantumSoft, Switzerland). pK_a values were obtained from ^{13}C chemical shift measurements at 25 °C by fitting the Henderson–Hasselbalch equation in ProFit. Structure calculations were performed using the program CNS[70] using a standard simulated annealing protocol. Structure representations were generated with MOLMOL.[71]

Data deposition

The atomic coordinates of the ten lowest energy CNS conformers of the foldon monomer have been deposited in the RCSB Protein Data Bank (www.rcsb.org) under PDB accession number 1U0P.

Acknowledgements

We thank Klara Rathgeb-Szabo for protein purification and the referees for very valuable suggestions. This work was supported by SNF grant 31-43′091.95 (to S.G.).

References

1. Coombs, D. H. & Eiserling, F. A. (1977). Studies on structure, protein-composition and assembly of neck of bacteriophage-T4. *J. Mol. Biol.* **116**, 375–405.
2. Terzaghi, B. E., Terzaghi, E. & Coombs, D. (1979). Role of the Collar–Whisker complex in bacteriophage T4d tail fiber attachment. *J. Mol. Biol.* **127**, 1–14.
3. Conley, M. P. & Wood, W. B. (1975). Bacteriophage-T4 Whiskers—Rudimentary environment-sensing device. *Proc. Natl Acad. Sci. USA*, **72**, 3701–3705.
4. Tao, Y. Z., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997). Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure*, **5**, 789–798.
5. Strelkov, S. V., Tao, Y., Shneider, M. M., Mesyanzhinov, V. V. & Rossmann, M. G. (1998). Structure of bacteriophage T4 fibritin M: a troublesome packing arrangement. *Acta Crystallog. sect. D*, **54**, 805–816.
6. Letarov, A. V., Londer, Y. Y., Boudko, S. P. & Mesyanzhinov, V. V. (1999). The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibritin. *Biochemistry (Moscow)*, **64**, 817–823.
7. Güthe, S., Kapinos, L., Möglich, A., Meier, S., Grzesiek, S. & Kiefhaber, T. (2004). Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. *J. Mol. Biol.* **337**, 905–915.
8. Frank, S., Kammerer, R. A., Mechling, D., Schulthess, T., Landwehr, R., Bann, J. *et al*. (2001). Stabilization of short collagen-like triple helices by protein engineering. *J. Mol. Biol.* **308**, 1081–1089.
9. Yang, X. Z., Lee, J., Mahony, E. M., Kwong, P. D., Wyatt, R. & Sodroski, J. (2002). Highly stable trimers formed by human immunodeficiency virus type 1 envelope glycoproteins fused with the trimeric motif of T4 bacteriophage fibritin. *J. Virol.* **76**, 4634–4642.
10. Onuchic, J. N. & Wolynes, P. G. (2004). Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75.
11. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E. *et al*. (1997). Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature*, **385**, 787–793.
12. Pan, K. M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D. *et al*. (1993). Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc. Natl Acad. Sci. USA*, **90**, 10962–10966.
13. Zimm, B. H., Doty, P. & Iso, K. (1959). Determination of the parameters for helix formation in poly-gamma-benzyl-L-glutamate. *Proc. Natl Acad. Sci. USA*, **45**, 1601–1604.
14. Zimm, B. H. & Bragg, J. K. (1959). Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* **31**, 526–535.
15. Lifson, S. (1961). Theory of Helix-coil transition in polypeptides. *J. Chem. Phys.* **34**, 1963.
16. Blanco, F. J., Rivas, G. & Serrano, L. (1994). A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nature Struct. Biol.* **1**, 584–590.
17. Gellman, S. H. (1998). Minimal model systems for beta sheet secondary structure in proteins. *Curr. Opin. Chem. Biol.* **2**, 717–725.
18. Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997). Folding dynamics and mechanism of beta-hairpin formation. *Nature*, **390**, 196–199.
19. Lacroix, E., Kortemme, T., Lopez de la Paz, M. & Serrano, L. (1999). The design of linear peptides that fold as monomeric beta-sheet structures. *Curr. Opin. Struct. Biol.* **9**, 487–493.
20. Tjandra, N. & Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, **278**, 1111–1114.
21. Tycko, R., Blanco, F. J. & Ishii, Y. (2000). Alignment of biopolymers in strained gels: a new way to create detectable dipole–dipole couplings in high-resolution biomolecular NMR. *J. Am. Chem. Soc.* **122**, 9340–9341.
22. Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T. & Grzesiek, S. (2000). Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR*, **18**, 303–309.
23. Shortle, D. & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science*, **293**, 487–489.
24. Ackerman, M. S. & Shortle, D. (2002). Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence. *Biochemistry*, **41**, 13791–13797.
25. Ohnishi, S. & Shortle, D. (2003). Observation of residual dipolar couplings in short peptides. *Proteins: Struct. Funct. Genet.* **50**, 546–551.
26. Louhivuori, M., Paakkonen, K., Fredriksson, K., Permi, P., Lounila, J. & Annila, A. (2003). On the origin of residual dipolar couplings from denatured proteins. *J. Am. Chem. Soc.* **125**, 15647–15650.
27. Alexandrescu, A. T. & Kammerer, R. A. (2003). Structure and disorder in the ribonuclease S-peptide probed by NMR residual dipolar couplings. *Protein Sci.* **12**, 2132–2140.
28. Fieber, W., Kristjansdottir, S. & Poulsen, F. M. (2004). Short-range, long-range and transition state interactions in the denatured state of ACBP from residual dipolar couplings. *J. Mol. Biol.* **339**, 1191–1199.
29. Mohana-Borges, R., Goto, N. K., Kroon, G. J., Dyson, H. J. & Wright, P. E. (2004). Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J. Mol. Biol.* **340**, 1131–1142.
30. Wang, L., O'Connell, T., Tropsha, A. & Hermans, J. (1996). Molecular simulations of beta-sheet twisting. *J. Mol. Biol.* **262**, 283–293.
31. Anderson, D. E., Becktel, W. J. & Dahlquist, F. W. (1990). pH-induced denaturation of proteins: a single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry*, **29**, 2403–2408.
32. Wagner, G., Pardi, A. & Wuthrich, K. (1983). Hydrogen-bond length and H-1-NMR chemical-shifts in proteins. *J. Am. Chem. Soc.* **105**, 5948–5949.
33. Cordier, F. & Grzesiek, S. (1999). Direct observation of hydrogen bonds in proteins by interresidue (3h)J(NC′) scalar couplings. *J. Am. Chem. Soc.* **121**, 1601–1602.
34. Sass, J., Cordier, F., Hoffmann, A., Cousin, A., Omichinski, J. G., Lowen, H. & Grzesiek, S. (1999).

Purple membrane induced alignment of biological macromolecules in the magnetic field. *J. Am. Chem. Soc.* **121**, 2047–2055.
35. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120**, 6836–6837.
36. Walkenhorst, W. F., Edwards, J. A., Markley, J. L. & Roder, H. (2002). Early formation of a beta hairpin during folding of staphylococcal nuclease H124L as detected by pulsed hydrogen exchange. *Protein Sci.* **11**, 82–91.
37. McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nature Struct. Biol.* **7**, 669–673.
38. Ramirez-Alvarado, M., Blanco, F. J. & Serrano, L. (1996). *De novo* design and structural analysis of a model beta-hairpin peptide system. *Nature Struct. Biol.* **3**, 604–612.
39. Maynard, A. J., Sharman, G. J. & Searle, M. S. (1998). Origin of beta-hairpin stability in solution: structural and thermodynamic analysis of the folding of model peptide supports hydrophobic stabilization in water. *J. Am. Chem. Soc.* **120**, 1996–2007.
40. Santiveri, C. M., Santoro, J., Rico, M. & Jimenez, M. A. (2002). Thermodynamic analysis of beta-hairpin-forming peptides from the thermal dependence of H-1 NMR chemical shifts. *J. Am. Chem. Soc.* **124**, 14903–14909.
41. Xu, X. P. & Case, D. A. (2002). Probing multiple effects on ^{15}N, ^{13}C alpha, ^{13}C beta, and ^{13}C′ chemical shifts in peptides using density functional theory. *Biopolymers*, **65**, 408–423.
42. Atkins, P. W. (1998). *Physical Chemistry* (6th edit.). Oxford University Press, Oxford, England.
43. Searle, M. S., Griffiths-Jones, S. R. & Skinner-Smith, H. (1999). Energetics of weak interactions in a beta-hairpin peptide: electrostatic and hydrophobic contributions to stability from lysine salt bridges. *J. Am. Chem. Soc.* **121**, 11615–11620.
44. Dill, K. A., Fiebig, K. M. & Chan, H. S. (1993). Cooperativity in protein-folding kinetics. *Proc. Natl Acad. Sci. USA*, **90**, 1942–1946.
45. Klimov, D. K. & Thirumalai, D. (2000). Mechanisms and kinetics of beta-hairpin formation. *Proc. Natl Acad. Sci. USA*, **97**, 2544–2549.
46. Ramirez-Alvarado, M., Blanco, F. J., Niemann, H. & Serrano, L. (1997). Role of beta-turn residues in beta-hairpin formation and stability in designed peptides. *J. Mol. Biol.* **273**, 898–912.
47. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
48. Louhivuori, M., Fredriksson, K., Paakkonen, K., Permi, P. & Annila, A. (2004). Alignment of chain-like molecules. *J. Biomol. NMR*, **29**, 517–524.
49. Street, A. G. & Mayo, S. L. (1999). Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc. Natl Acad. Sci. USA*, **96**, 9074–9076.
50. Moltke, S. & Grzesiek, S. (1999). Structural constraints from residual tensorial couplings in high resolution NMR without an explicit term for the alignment tensor. *J. Biomol. NMR*, **15**, 77–82.
51. Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* **104**, 4559–4570.
52. Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559.
53. Nicholson, L. K., Yamazaki, T., Torchia, D. A., Grzesiek, S., Bax, A., Stahl, S. J. *et al*. (1995). Flexibility and function in HIV-1 protease. *Nature Struct. Biol.* **2**, 274–280.
54. Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR*, **16**, 23–28.
55. Wennerström, H. (1972). Nuclear magnetic relaxation induced by chemical exchange. *Mol. Phys.*, **24**, 69–80.
56. Millet, O., Loria, J. P., Kroenke, C. D., Pons, M. & Palmer, A. G. (2000). The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. *J. Am. Chem. Soc.* **122**, 2867–2877.
57. Xu, Y., Oyola, R. & Gai, F. (2003). Infrared study of the stability and folding kinetics of a 15-residue beta-hairpin. *J. Am. Chem. Soc.* **125**, 15388–15394.
58. Nesmelova, I., Krushelnitsky, A., Idiyatullin, D., Blanco, F., Ramirez-Alvarado, M., Daragan, V. A. *et al*. (2001). Conformational exchange on the microsecond time scale in alpha-helix and beta-hairpin peptides measured by ^{13}C NMR transverse relaxation. *Biochemistry*, **40**, 2844–2853.
59. Friedrichs, M. S., Stouch, T. R., Bruccoleri, R. E., Mueller, L. & Constantine, K. L. (1995). Structural and dynamic properties of a beta-hairpin-forming linear peptide. 2. C-13 NMR relaxation analysis. *J. Am. Chem. Soc.* **117**, 10855–10864.
60. Boudko, S. P., Londer, Y. Y., Letarov, A. V., Sernova, N. V., Engel, J. & Mesyanzhinov, V. V. (2002). Domain organization, folding and stability of bacteriophage T4 fibritin, a segmented coiled-coil protein. *Eur. J. Biochem.* **269**, 833–841.
61. Munoz, V., Henry, E. R., Hofrichter, J. & Eaton, W. A. (1998). A statistical mechanical model for beta-hairpin kinetics. *Proc. Natl Acad. Sci. USA*, **95**, 5872–5879.
62. Espinosa, J. F., Munoz, V. & Gellman, S. H. (2001). Interplay between hydrophobic cluster and loop propensity in beta-hairpin formation. *J. Mol. Biol.* **306**, 397–402.
63. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.
64. Grzesiek, S., Bax, A., Hu, J. S., Kaufman, J., Palmer, I., Stahl, S. J. *et al*. (1997). Refined solution structure and backbone dynamics of HIV-1 Nef. *Protein Sci.* **6**, 1248–1263.
65. Chou, J. J., Gaemers, S., Howder, B., Louis, J. M. & Bax, A. (2001). A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *J. Biomol. NMR*, **21**, 377–382.
66. Ottiger, M., Delaglio, F. & Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378.
67. Raiford, D. S., Fisk, C. L. & Becker, E. D. (1979). Calibration of methanol and ethylene-glycol nuclear magnetic-resonance thermometers. *Anal. Chem.* **51**, 2050–2051.
68. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRpipe—a multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR*, **6**, 277–293.

69. Garrett, D. S., Gronenborn, A. M. & Clore, G. M. (1995). Automated and interactive tools for assigning 3d and 4d NMR-spectra of proteins—capp, stapp and pipp. *J. Cell. Biochem.* 71.
70. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
71. Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 29–32.
72. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
73. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995). ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR*, **5**, 67–81.

Edited by M. F. Summers

(Received 13 August 2004; received in revised form 27 September 2004; accepted 27 September 2004)

5 Unpublished Results on Foldon

To characterise folding of the foldon domain in more detail, the concentration dependence of the unfolding kinetics was investigated. In addition, equilibrium studies were performed, using urea as a denaturant. The question was addressed whether equilibrium intermediates can be detected, and additional data on the refolding intermediate and the A-state were obtained.

5.1 Urea-induced equilibrium transitions

In addition to the dependence of the foldon domain on GdmCl, as described in section 4, the stability of the foldon domain was also investigated by urea-induced equilibrium transitions. Fluorescence and circular dichroism signals were detected at 12, 5, and 1 µM foldon monomer concentration. The data, as shown in Figure 5-1, can be fitted to a two-state monomer- trimer equilibrium.

Single fits of the four data sets yielded comparable stabilities, so that the data was fit globally as well. The global fit yields a free energy of unfolding of $\Delta G^0 (H_2O) = 82.1 \pm 1.7$ kJ / (mol monomer), which is comparable to the $\Delta G^0 (H_2O)$ of 89.2±0.6 kJ/(mol monomer) found for GdmCl-induced equilibrium transition of the foldon domain (section 4.1.). This value applies to standard conditions of 1 M total monomer concentration. At typical physiological protein concentrations around 5 µM the free energy corresponds to $\Delta G^0 (H_2O) = 27.3 \pm 0.6$ kJ. The change in free energy with GdmCl ($m_{eq} = \partial \Delta G^0 / \partial [GdmCl]$) is 4.5 ±0.3 (kJ/mol) / M, which is the value expected for a monomeric globular protein of the size of the folded trimer.

The m_{eq} - value found for GdmCl is 10.4 (±0.2) (kJ/mol) / M (see section 4). The lower m_{eq} - value for urea reflects the weaker denaturing strength of urea compared to GdmCl.

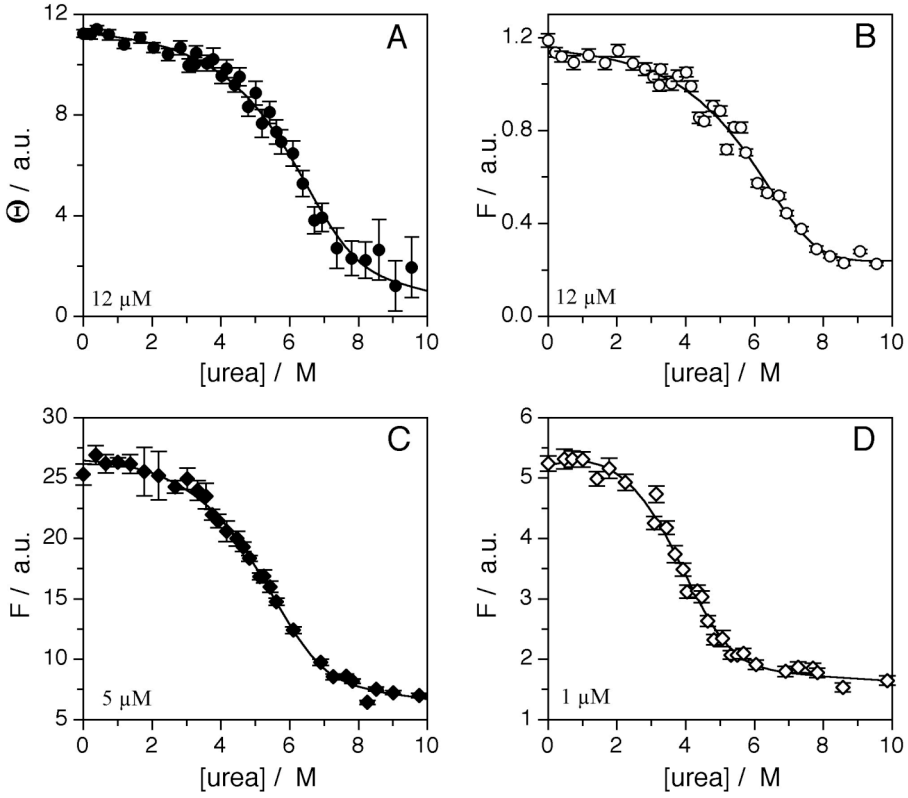


Figure 5-1

Urea- induced equilibrium transitions of the foldon domain, pH 7.1, 10 mM NaPi, 20ºC. A:12 µM foldon monomer, CD at 228 nm, 0.5 cm cuvette; B-D: fluorescence λ_{exc}: 278 nm, λ_{em}: 318 nm. B: 12 µM foldon monomer, C: 5µM foldon monomer, D: 1 µM foldon monomer; A-D: The straight line indicates the global fit to

5.2 Test for equilibrium unfolding intermediate

The stability of the foldon is concentration dependent. At low protein concentrations, the native state becomes less stable, and eventually an intermediate might become more stable than N. The stability of a dimer depends to a lesser extent on the protein concentration, so that an observed intermediate might well be dimeric or even a monomer.

To check whether foldon equilibrium unfolding is really two-state and if an intermediate is populated in equilibrium, GdmCl-induced equilibrium transitions were measured at several protein concentrations at several emission wavelengths. 318 nm and 340 nm were chosen as the wavelength where the signal for N and U, respectively, can be best observed. The isosbestic point of the fluorescence spectrum of N and U is around 356 nm. A change in signal with GdmCl concentration at this wavelength should not occur if only N and U are populated during equilibrium. A signal change at the isosbestic point thus indicates the population of an equilibrium intermediate. As can be seen in Figure 5-2, an intermediate seems to be populated at higher GdmCl concentrations at a protein concentration of 0.5 μM, at 1 μM the intermediate is populated transiently in the transition region, and at protein concentrations of 5 μM and higher, the intermediate seems to be stable also under native conditions, but it disappears with increasing GdmCl concentrations.

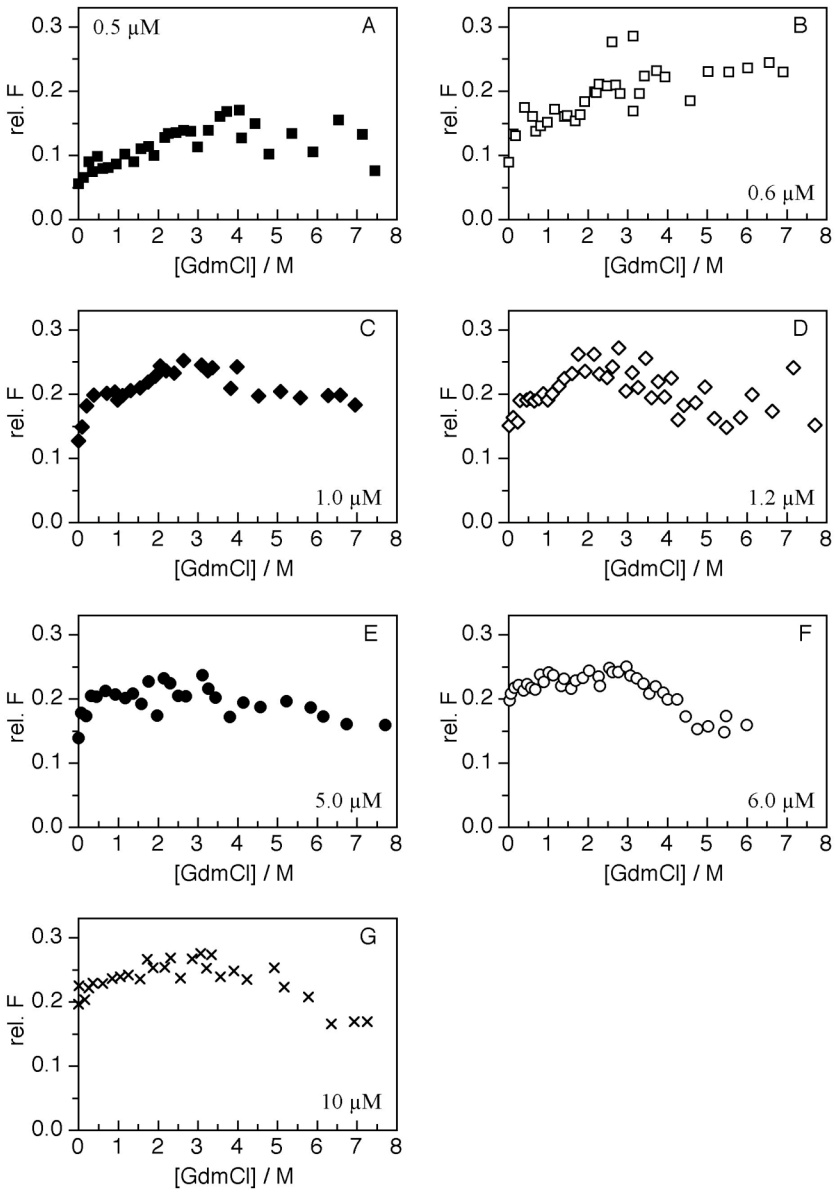


Figure 5-2

GdmCl- induced equilibrium transitions at different protein concentrations, monitored by fluorescence emission at the isosbestic point. A-D: λ_{exc}: 280 nm, pH 7.1, 20.0ºC; left side (A, C, E, G): λ_{em}: 357 nm, right side (B, D, F): λ_{em}: 355 nm.

As a consequence, an equilibrium transition measuring the whole fluorescence emission spectrum at each GdmCl concentration was performed, as shown in Figure 5-3, at a protein concentration of 1 µM.

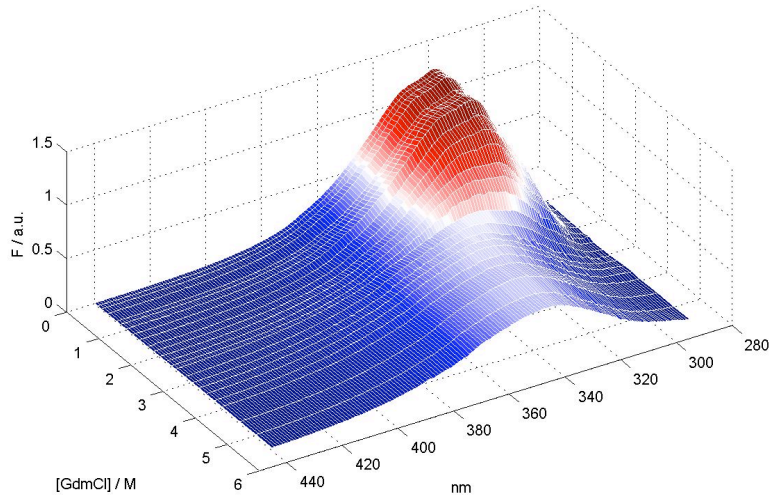


Figure 5-3

GdmCl- induced equilibrium transition of foldon at 1 µM protein concentration. Monitored by fluorescence emission, λ_{exc}: 280 nm, pH 7.1, 20.0ºC

The dataset was analysed using SVD analysis.[110; 111] A set of fluorescence emission spectra at different GdmCl concentrations was measured and can be represented as an $M \times N$ matrix A. Each column represents the fluorescence spectrum at k M GdmCl. According to SVD theory, any matrix may be written as the product of three matrices:

$A(S, k) = U(S, k)SVT$

where $U(S, k)$ is an $M \times N$ matrix of orthogonal columns, which form a complete set of basis fluorescence curves. The fluorescence spectrum at any GdmCl concentration can be represented as a linear superposition of U. S is a $N \times N$ diagonal matrix with non-negative elements called the singular values. VT is the transpose of an $N \times N$ orthogonal matrix V. Each column of SVT contains the set of GdmCl - dependent linear superposition coefficients corresponding to each basis function. Normally, there are some singular values that are so close to zero so they can be neglected. The number of the singular values that are necessary equals the number of species. The population of an intermediate becomes observable without assuming a specific model to analyse the dataset.

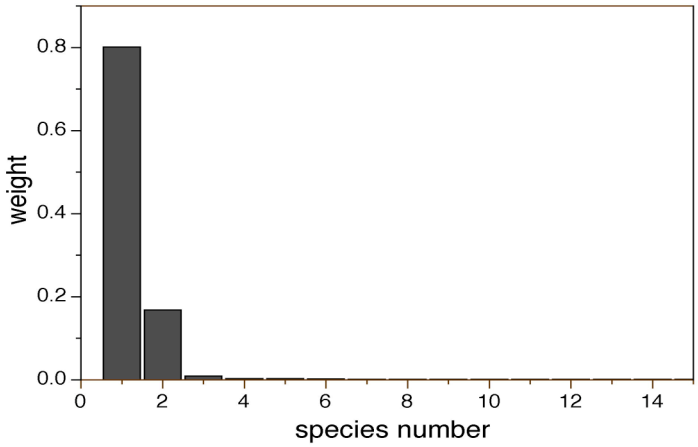


Figure 5-4

Weight of the different species obtained from SVD analysis.

In this case, only the first two values are significantly higher than the others, which represent noise, as shown in Figure 5-4. The third column is only marginally larger than the other values. This indicates that in equilibrium, at 1 µM protein concentration, only very small amounts of intermediate are populated. A look at the basis functions SVD analysis attributes to every species can also help to decide on the number of species needed to describe the dataset. On Figure 5-5 one can see the elementary spectra obtained from SVD analysis. The orange spectrum changes signs, but apart from that, the black and the orange spectrum, corresponding to species 1 and 2, resemble the "real" fluorescence spectra or U and N, especially when weighted (lower panel). From this dataset, it cannot be decided whether an intermediate is really needed to describe equilibrium behaviour of 1 µM foldon. At such a low protein concentration, the native protein is not very stable compared to the unfolded protein, so that an intermediate might become populated to some extent.

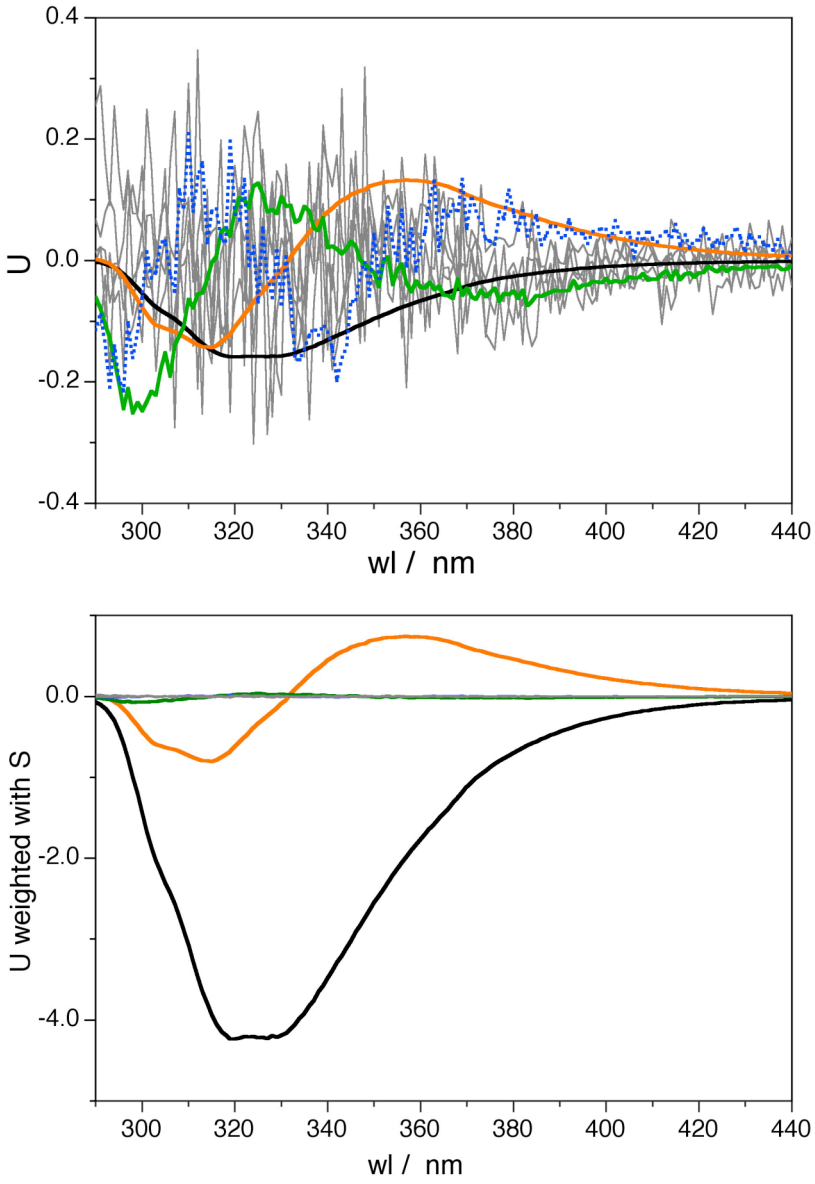


Figure 5-5

U vector from SVD analysis versus wavelength. Top: U directly from SVD analysis, bottom: U weighted with S. These basis functions do not correspond to the true fluorescence spectra. Black line: species 1, orange: species 2, green: species 3,blue:species 4. Species 5-10 in grey.

5.3 Concentration dependence of the unfolding kinetics

As folding of the trimeric foldon domain is concentration dependent, unfolding of foldon was also analysed at several protein concentrations. Unfolding was initiated by rapid dilution (stopped-flow mixing) of native protein into unfolding buffer to a final denaturant concentration of 7.4 M GdmCl. In Figure 5-6 one can observe that unfolding seems to be almost independent of the protein concentration. For clarity, only every second concentration is plotted.

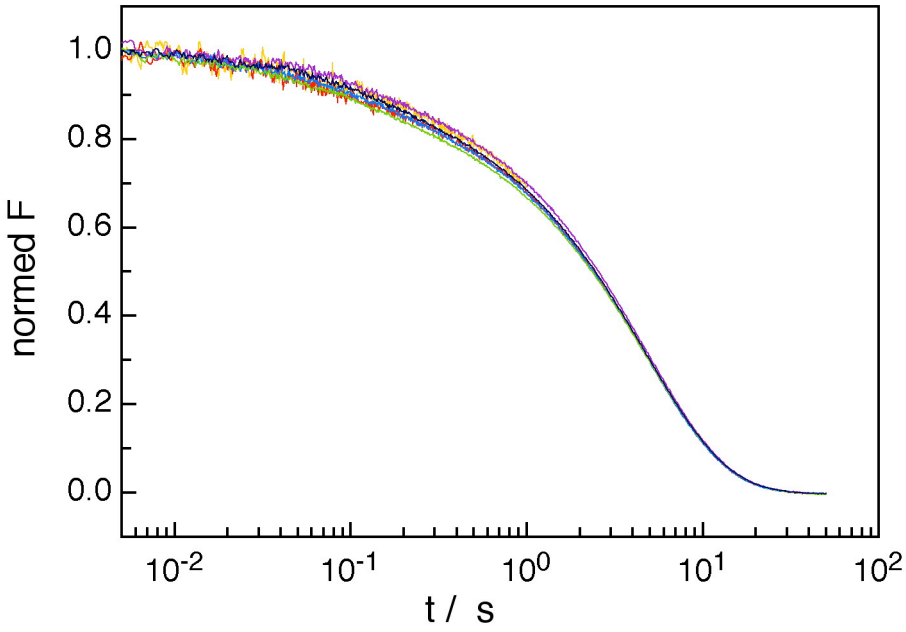


Figure 5-6

Concentration dependence of unfolding of the foldon domain. Stopped-flow 1:11 mixing from 0 M to 7.4 M GdmCl; λ_{exc}: 280 nm, λ_{em}: > 320 nm, pH 7.1, 10 mM NaPi, 20ºC. Protein concentration: yellow: 1 µM; red: 2 µM; purple: 4 µM; blue: 6 µM; green: 8 µM; black: 10 µM. Traces are normalised between 0 and 1.

As there is some variety in the region from 30 ms to 10 s, the traces were investigated more closely.

The traces could be fitted to the sum of three first order reactions. An example is shown in Figure 5-7 for unfolding of 5µM foldon.

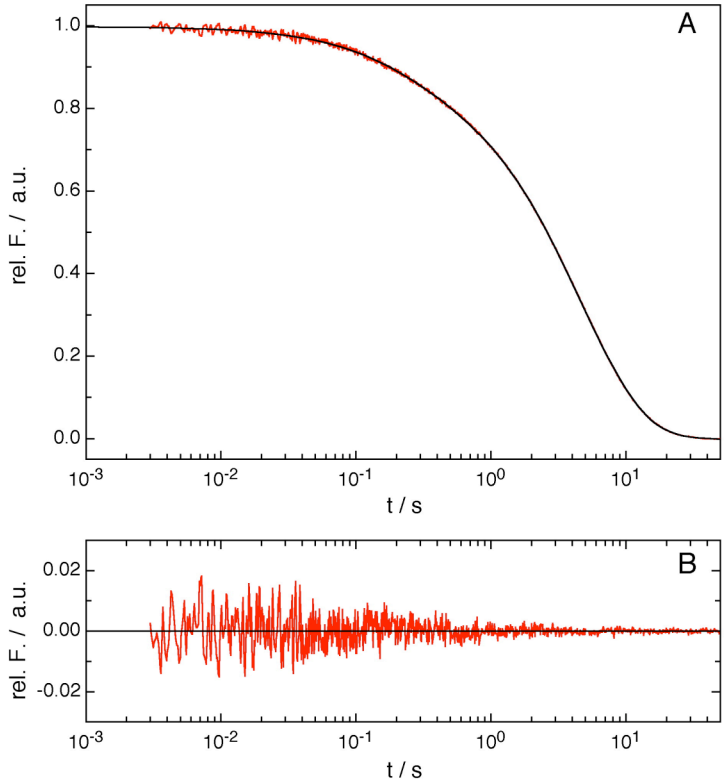


Figure 5-7

Unfolding of the foldon domain at 5 µM protein concentration. A: refolding trace and fit to the sum of three exponentials. A_1 = 0.116±0.027, λ_1= 10.5±0.618 s^{-1}; A_2 = 0.780±0.618, λ_2 = 0.197±2.55·10^{-3} s^{-1}; A_3 = 0.104 ±0.031, λ_3= 1.1325±0.159 s^{-1}; B: residuals. Stopped-flow 1:11 mixing from 0 M to 7.4 M GdmCl; λ_{exc}: 280 nm, λ_{em}: > 320 nm, pH 7.1, 10 mM NaPi, 20ºC.

In Figure 5.8, the dependence of the rate constants on protein concentration is shown, and also the corresponding amplitudes.

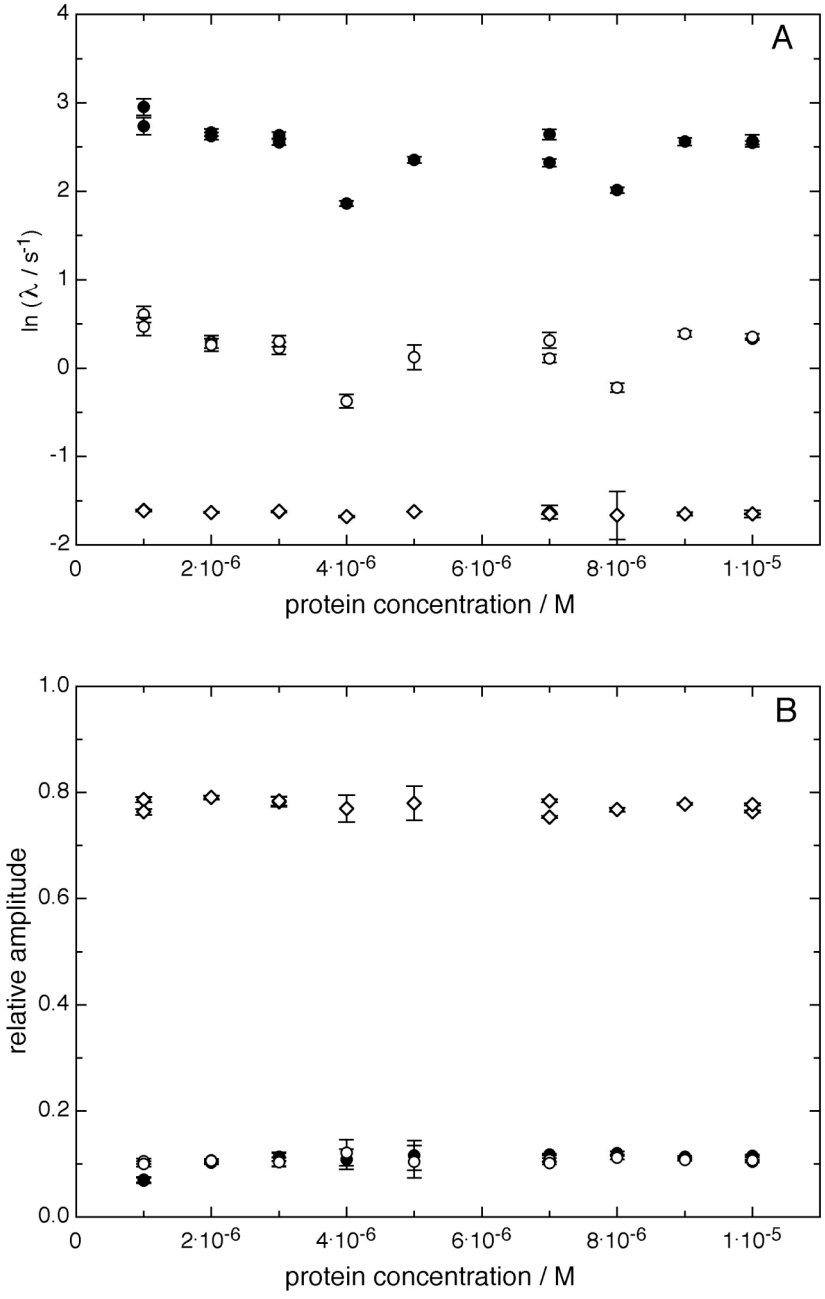


Figure 5-8

Concentration dependence of foldon unfolding. Stopped-flow 1:11 mixing from 0 M to 7.4 M GdmCl; λ_{exc}: 280 nm, λ_{em}: > 320 nm, pH 7.1, 10 mM NaPi, 20ºC. The traces could be fitted to a sum of three exponentials: phase 1: filled circles; phase 2: open diamonds; phase 3: open circles. Top: the rate constants (top panel) as well as the relative amplitudes (bottom panel) are independent of the protein concentration. The relative amplitudes were obtained by dividing the fitted amplitudes by the total amplitude (sum of the fitted amplitudes).

As can be seen in Figure 5-8, none of the three unfolding phases unfolding depend on the protein concentration. This is expected for a decay from trimer to monomer. The two fast phases tend to decrease with increasing protein concentration below 3 μM, but the scatter of the data points is larger than this tendency. An analysis of the reaction order below 4 μM yields values very close to unity for all three phases.

The three observed phases might originate from unfolding in several steps, via oligomeric intermediates, such as:

$$N \rightarrow I_3$$

$$I_3 \rightarrow I_2 + M$$

$$I_2 \rightarrow M$$

Additionally, there might be some contributions from folding reactions, even though at 7.4 M the equilibrium is strong on the side of the unfolded protein.

5.4 Starting values of refolding kinetics

Major fluorescence changes take place in the dead-time of stopped-flow unfolding (see Fig. 3A in section Güthe *et al.*[106]). As explained in section 1.4, in order to understand the processes that occur in the first milliseconds of foldon folding, the starting values of the refolding kinetics depending on denaturant concentration can be analysed to determine the stability of the intermediate.

For the foldon domain, this experiment was performed at 1.1 µM monomer concentration, Refolding was initiated by rapid dilution of 11 µM unfolded foldon in 5.5 M GdmCl to final conditions of 0.5 M to 7.7 M GdmCl.

As can be seen Figure 5-9 A, the starting values of the refolding kinetics (●) do not form a straight line in the low denaturant region, they rather resemble a transition. The amplitude of a whole refolding trace is also small compared to the change of the fluorescence values of the initial points at high and low denaturant concentration. The complete refolding trace could only be resolved below 3.5 M GdmCl, as kinetics became to slow at higher GdmCl concentrations to be determined by stopped-flow measurements. Above 3 M GdmCl, where no refolding kinetics can be observed, initial and end points are identical and represent the unfolded baseline of the corresponding equilibrium transition, which is shown in panel B. The equilibrium transition is compared to the initial values (C), and a difference can hardly be noticed, especially if the large scattering, that is best visible in the unfolded baseline above 3.5 M GdmCl, is taken into account.

Thus one can assume that the folding intermediate formed in the deadtime of stopped-flow mixing is almost as stable as the native protein at 1 µM protein concentration, and that most of the folding reaction takes place in the mixing deadtime.

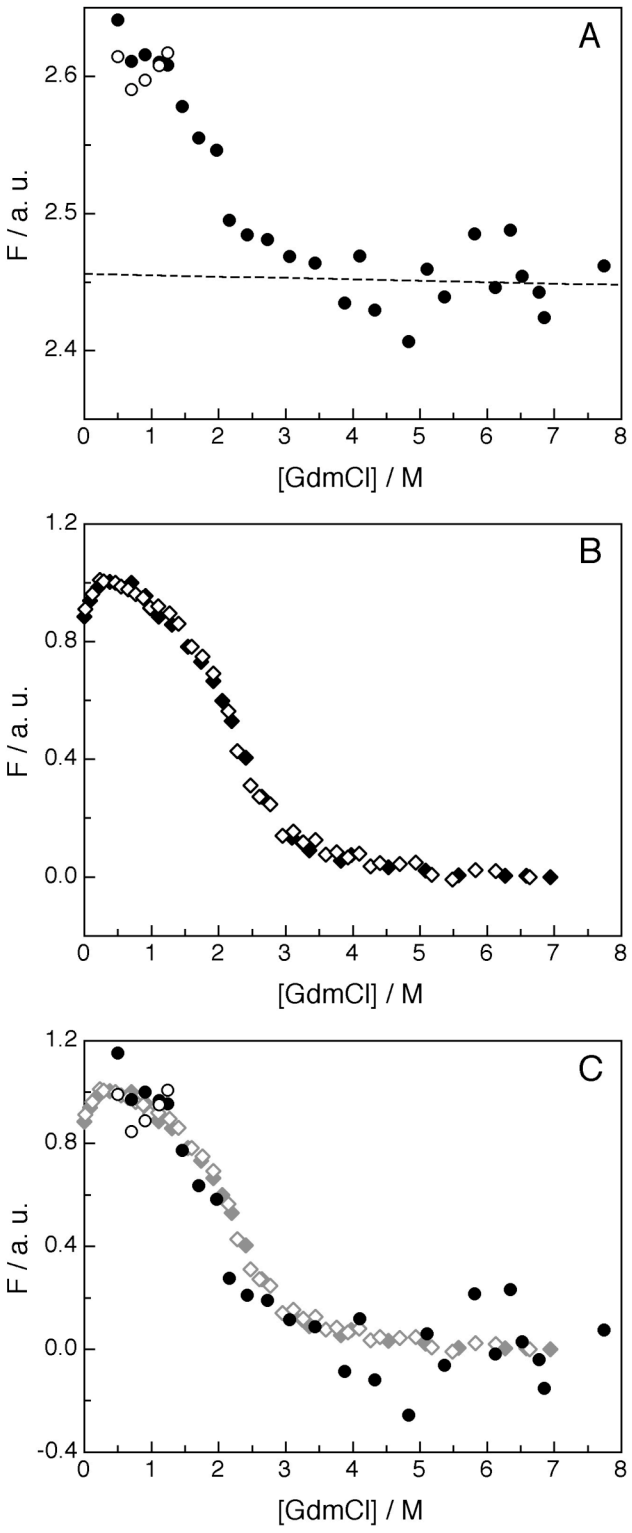


Figure 5-9

Initial values of refolding of the foldon domain compared to equilibrium. A: Initial values (○) for refolding kinetics at 1.1 µM foldon monomer concentration. Stopped flow, 320 nm filter. For refolding faster than 200 s the endpoints (●) could also be resolved. The dashed line indicates the unfolded baseline, where the refolding signal should start nothing would happen in the burst phase. B: The corresponding equilibrium transitions at 1.0 (◆) and 1.2 µM (◇) protein concentration (320 nm fluorescence emission). C: An overlay of the equilibrium transition (grey diamonds) with the initial and end points (black circles) of refolding kinetics. The initial and endpoints were normalised so that the endpoints at low urea concentrations correspond to the native signal in equilibrium, and that the initial points at high urea concentrations superimpose with the unfolded baseline in equilibrium. A C: Conditions: pH 7.1, 20ºC, 10 mM NaPi, l_{exc}: 280 nm

5.5 Energy transfer in the low pH intermediate

The A-state was further characterised by fluorescence spectroscopy. An intersubunit salt bridge between residues R15 and E5 connects two adjacent foldon subunits. At low pH, this salt bridge is titrated. From the pH transition and tyrosine to tryptophan resonance energy transfer measurements described section 7.5 it became evident that at pH 4 an intermediate is populated. Additional measurements to investigate Tyr → Trp energy transfer shown in Figure 5-10 demonstrated that also at pH 2 the foldon domain is in a compact structure, even though the maximum of fluorescence emission is redshifted compared to the native state.

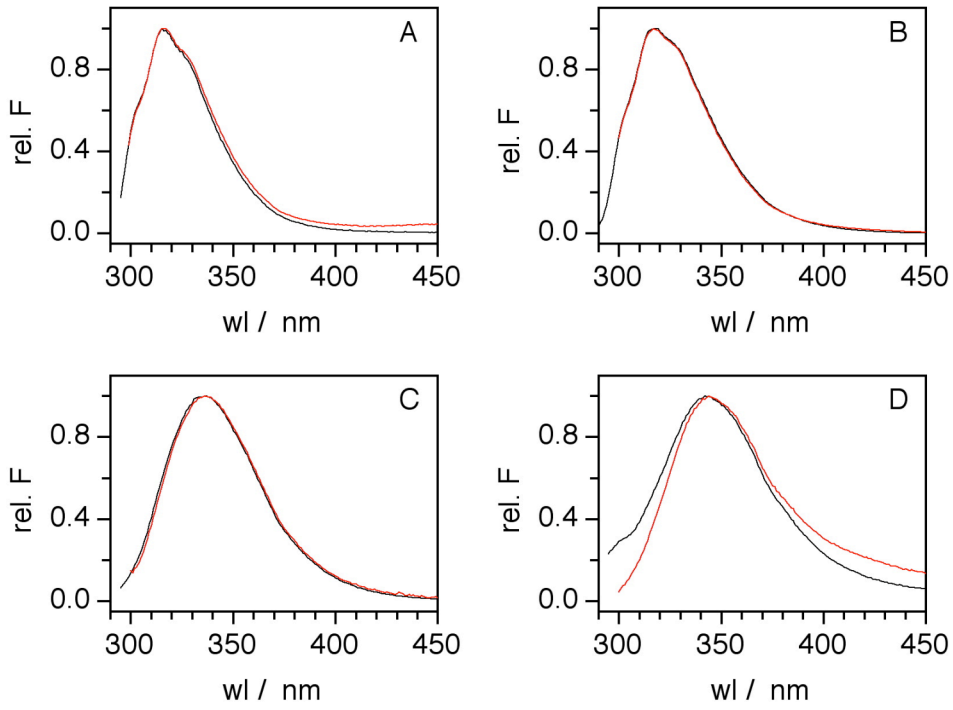


Figure 5-10

Tyr → Trp energy transfer in the foldon domain. protein concentration: 3 µM, 20ºC, 10 mM NaPi, λ_{exc}: 280 nm (black) 295 nm (red). A: native foldon at pH 7.1, λ_{max} = 316 nm; B: at pH 4, λ_{max} = 317 nm; C: at pH 2, λ_{max} = 337 nm; D: denatured in 7.8 M GdmCl at pH 7.1, λ_{max} = 344 nm.

5.6 Conclusion

Equilibrium unfolding at protein concentrations between 1 µM and 12 µM can be fit to a two-state model, and yields the same free energy irrespective of the protein concentration, so the data can be fit globally. From the global fit, similar stabilities for the foldon domain in urea are obtained compared to GdmCl, as shown in Table 5.1. The change in solvent accessible surface area upon unfolding, as reflected by the m_{eq} value, is larger for GdmCl than for urea, as urea is a weaker denaturant. The values are similar to the ones expected for small monomeric globular proteins the size of the foldon domain.

denaturant	$\Delta G^0(H_20)$ (kJ/mol monomer)	$\Delta G(H_2O)$ (kJ) at 5µM protein conc.	m_{eq} ((kJ/mol)/M)
urea	-82.1 ± 1.7	27.3	4.5 ± 0.3
GdmCl	-89.2 ± 0.6	29.7	10.4 ± 0.2

Table 5-1

Equilibrium data of foldon. The results are obtained from a two- -state fit according to equation 19, Materials and Methods.

In equilibrium, an intermediate seems to be populated at very low protein concentrations around 1 µM. The data, however, can still be fitted to a two-state model. At 1 µM, the kinetic intermediate in refolding is almost as stable as the native trimer, as the analysis of starting values from refolding traces have shown. The pH 2 intermediate (A-state) is compact, and the Tyr and Trp residues are close enough for Förster energy transfer to occur.

6 Results on the Semliki Forest Virus Protease

Previously, Sánchez *et al.*[102] investigated whether fast folding of SFVP is an intrinsic property of this two-domain protein, or whether spontaneous folding is slower than translation and requires additional catalysts. Around 60% of the proteins fold via a fast pathway with a time constant of 50 ms at 0 M denaturant (λ = 6.3 s^{-1} at 0.9 M urea, amplitude = 71%). SFVP contains seven prolyl residues, which are all *trans* in native state. Molecules with all prolyl residues in *trans* in the unfolded state fold via the fast pathway. Three slower parallel folding reactions correspond to *cis/trans* isomerisation reactions of Xaa-Pro and Xaa-non-Pro peptide bonds. The rate constants for these at 0.9 M urea are 1.4 s^{-1} (12%), which is compatible with non-prolyl *cis/trans* isomerisation, and 0.17 s^{-1} (4%), and 0.003 s^{-1} (13%), for prolyl *cis/trans* isomerisation. Unfolding is single-exponential, both domains seem to unfold in a concerted manner. Interrupted refolding experiments showed a lag phase in the formation of native molecules under these conditions. This indicates the population of an obligatory intermediate, as well as the downward curvature in the chevron plot at low urea concentrations and the presence of a fast phase with a very small amplitude at some low urea concentrations. The major unfolding/refolding phase was fitted to a three-state model with a populated on-pathway intermediate. Furthermore, the lag phase in the formation of native molecules, and a rollover in the chevron plot implies that the two domains are probably formed sequentially. The lag phase in the N-test and the lack of major signal in the fast phase can be explained as the N-terminal domain does not contain a Trp residue and thus does not contribute to tryptophan fluorescence.

In the following part, further experiments with the SFVP wild-type are described. To further elucidate the folding mechanism of SFVP, especially the fast refolding phase probably

originating from the on-pathway intermediate, a mutant of SFVP was investigated, where phenylalanine 160 is replaced by a tryptophan residue to obtain fluorescence signal in the N-terminal domain.

6.1 SFVP wild-type

In addition to the study of Sánchez et al.[102], who used urea as a denaturant, the effect of GdmCl on the stability of SFVP was investigated.

6.1.1 Determining the ideal buffer

The buffer used by Sánchez et al.[102] is 150 mM KCl, 0.1 mM EDTA, 20 mM HEPES at pH 7.5, in order to be consistent with the data from A. Helenius and coworkers.[94] EDTA is essential, but KCl did not influence stability at all, regardless of the denaturant used, GdmCl (see Figure 6-1 and Figure 6-2) or urea (Figure 6-3). HEPES has a rather high UV absorption, but buffers very well, especially when mixing two GdmCl solution of different concentrations, but same pH. 10 mM NaPi, on the other hand, does not buffer that well, so that the buffer conditions for SFVP were not changed, but extra pure fluorescent grade HEPES from FLUKA was used.

6.1.2 Equilibrium stability of SFVP

Previous experiments performed on SFVP[102] used urea as a denaturant. Here, SFVP was additionally characterised by GdmCl unfolding at 25ºC, using Trp fluorescence as a probe. As the native baseline was unexpectedly very steep, the population of an intermediate was suspected, and the equilibrium unfolding experiment was repeated at 10ºC to stabilise the putative intermediate. At 10ºC, the population of an intermediate can be observed. The data can be fit to a three-state model, ((eq16) in Materials and methods section), albeit with a high error as the baseline of the intermediate is not determined very well. At 25ºC, a three-state fit is only possible when the native baseline is fixed.

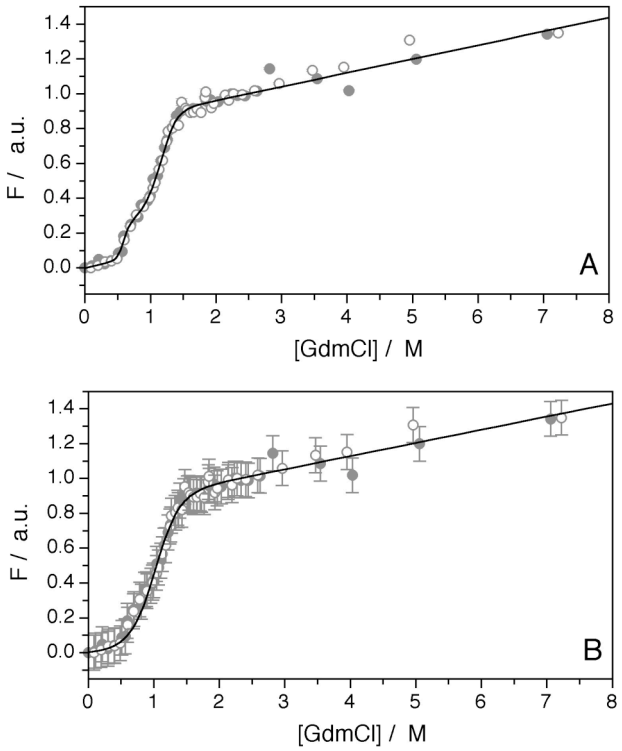


Figure 6-1

GdmCl-induced equilibrium transition of SFVP wt at 10ºC. A: data fit to a three-state model, ((eq 16), Materials and Methods) B: data fit to a two-state model (eq 15, Materials and Methods). 1 µM SFVP at pH 7.5, 20 mM HEPES, 0.1 mM EDTA, with 150 mM KCl (●) w/o KCl (○); λ_{exc}: 280±2 nm, λ_{em}: 355±4 nm.

Figure 6-1 shows the equilibrium transition at 10ºC. As can be seen on panel A, the three-state fit describes the dataset well, whereas the two-state fit (B) fails. The equilibrium stability of the intermediate relative to the native sate at 10ºC is $\Delta G_{NI}^{10ºC}(H_20)$ = 30±40 kJ/mol, with a change in free energy with urea of $m_{G_{NI}}$ = 50±770 (kJ/mol) / M, the stability of the unfolded protein relative to the intermediate is $\Delta G_{IU}^{10ºC}(H_20)$ = 24±91 kJ/mol with $m_{G_{IU}}$ = 20±60 (kJ/mol) / M. At 25ºC, a two-state fit of the data yields an $\Delta G^0(H_2O)$ of 37.3±4.5 kJ/mol with m_{eq} = 32.9±3.9 (kJ/mol) / M, as shown in Figure 6-2. The free energy is comparable to the $\Delta G^0(H_2O)$ = 32.3±0.9 kJ/mol obtained in urea by Sánchez et al.,[102] but the m_{eq} is higher as GdmCl is a stronger denaturant than urea, m_{eq}(urea) = 9.2±0.9 (kJ/mol) / M.

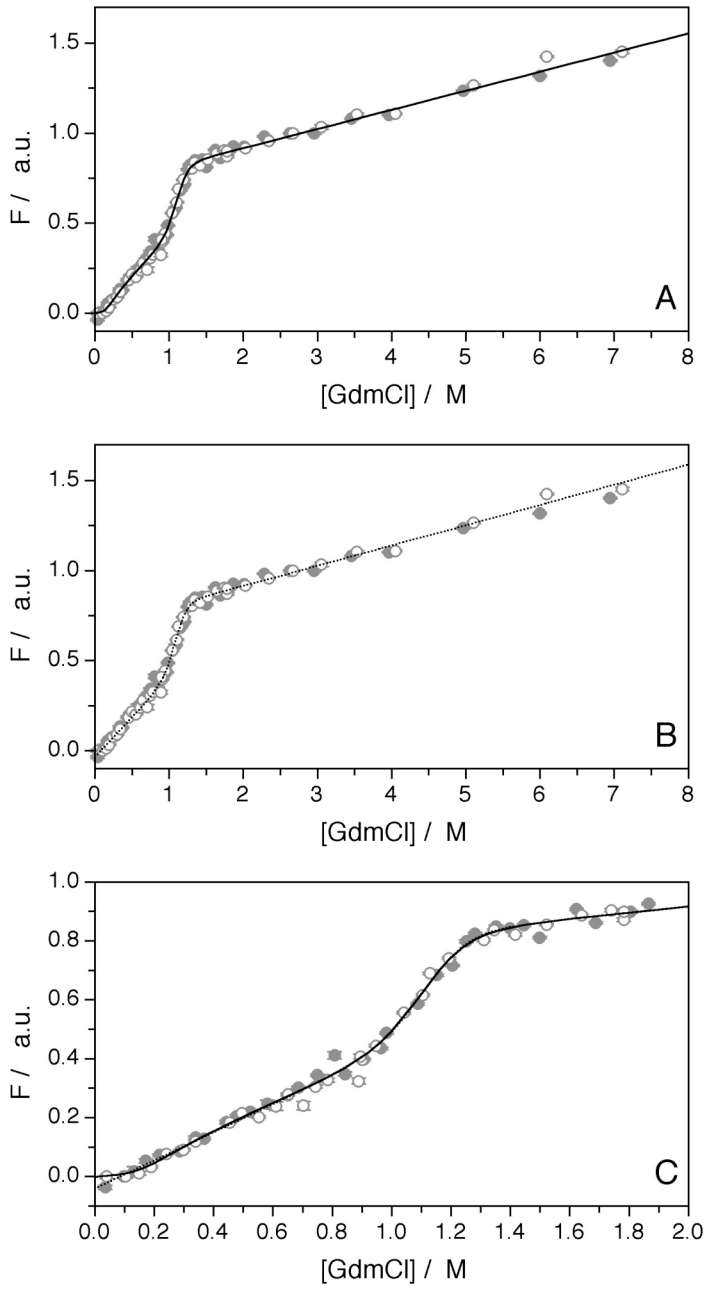


Figure 6-2

GdmCl-induced equilibrium transition of SFVP wt at 25ºC. A: data fit to a three-state model with fixed native baseline (eq 16). B: data fit to a two-state model (eq 15). C: overlay of A and B, low denaturant region. 1 µM SFVP at pH 7.5, 20 mM HEPES, 0.1 mM EDTA, with 150 mM KCl (●) w/o KCl (○); λ_{exc}: 280±2 nm, λ_{em}: 355±4 nm. The results are given in the text.

To test whether the SFVP preparation used behaved as described previously[102], an urea-induced equilibrium transition was performed, as shown in Figure 6-3.

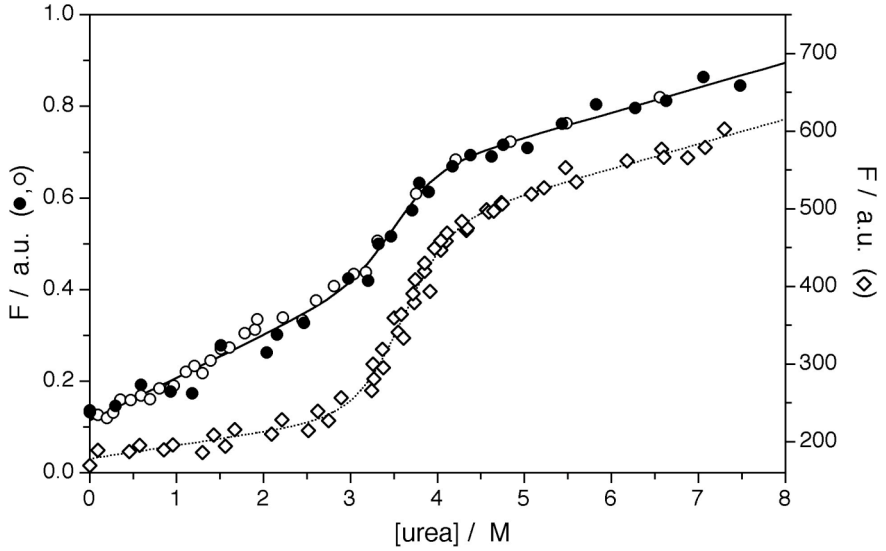


Figure 6-3

Urea- induced equilibrium transition. Open diamonds: data from Sánchez et al.[102], for comparison. 1 µM SFVP at pH 7.5, 25ºC, 20 mM HEPES, 0.1 mM EDTA, with 150 mM KCl (●, ◇) w/o KCl (○);λ_{exc}: 280±2 nm, λ_{em}: 355±4 nm. The solid lines represent individual two-state fits to (eq 15).

Even though the native baseline of ● and ○ is very steep, a global two-state fit yielded a free energy of $\Delta G^0(H_2O)$ = 34.0±4.21 kJ/mol, and a change in free energy with urea ($m_{eq} = \partial\Delta G^0/\partial$[urea]) of 9.6 ±1.7 (kJ/mol) / M, which is, within error, identical to the result obtained with a two-state fit of the data from Sánchez et al.[102] $\Delta G^0(H_2O)$ = 32.4±3.3 kJ/mol, m_{eq} = 9.0 ±0.9 (kJ/mol) / M.

6.2 SFVP F160W

The folding mechanism of SFVP is shown in Figure 6-4. As can be seen, the N-terminal domain folds first, followed by subsequent folding of the C-terminal domain.

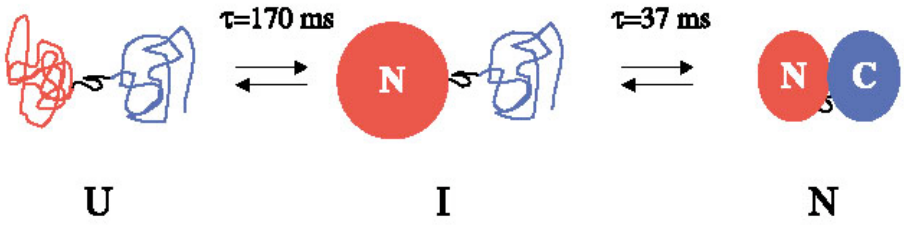


Figure 6-4

Schematic representation of the folding mechanism of SFVP, taken from Sánchez et al.[102]

As the N-terminal domain does not contain a Trp residue as a fluorescent probe, a mutant was designed [112] which contains an additional Trp residue, substituting Phe160. Equilibrium and kinetic studies were performed with SFVP F160W, to gather information on the first folding step of SFVP, the formation of the N-terminal domain.

The structure of the mutant protein has not been determined, but based on the crystal structure of the wild-type protein, a homology model of the SFVP F160W was created with SWISSMODEL[113].

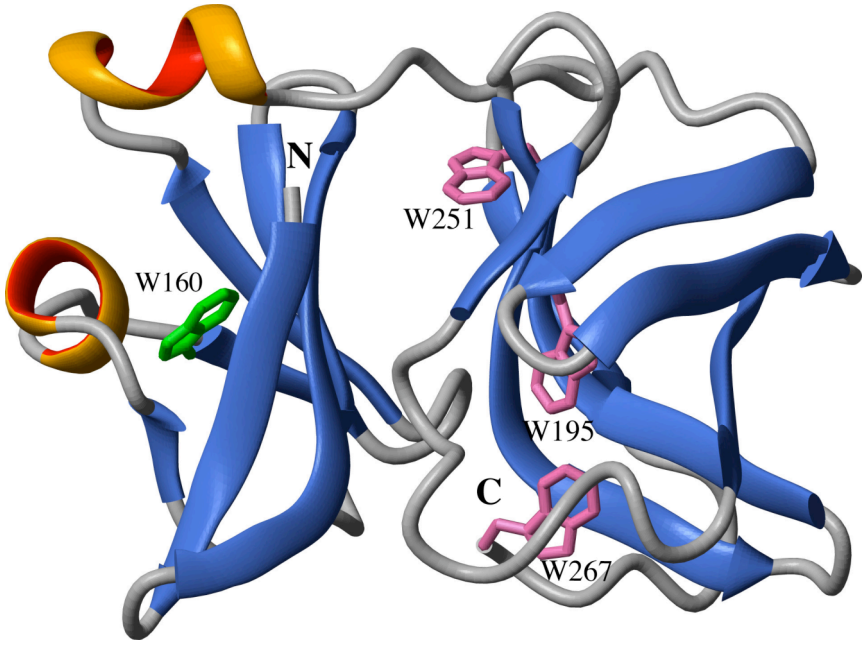


Figure 6-5

Putative structure of SFVP F160W based on the wild-type crystal structure. Ribbon representation of the homology model[113] created with MOLMOL[88]. The three Trp residues in the C-terminal domain are highlighted in pink, Trp 160 in the N-terminal domain is highlighted in green.

6.2.1 Spectroscopic characterization

The mutant was characterised by fluorescence emission spectroscopy and by far UV circular dichroism (CD). Fluorescence emission in the native state has a broad maximum at 335 nm. In the unfolded state, the maximum is shifted to 350 nm. In the native state, emission is quenched compared to the unfolded state, to about 60 % relative to the urea-unfolded state and to about 40 % relative to the GdmCl-denatured state.

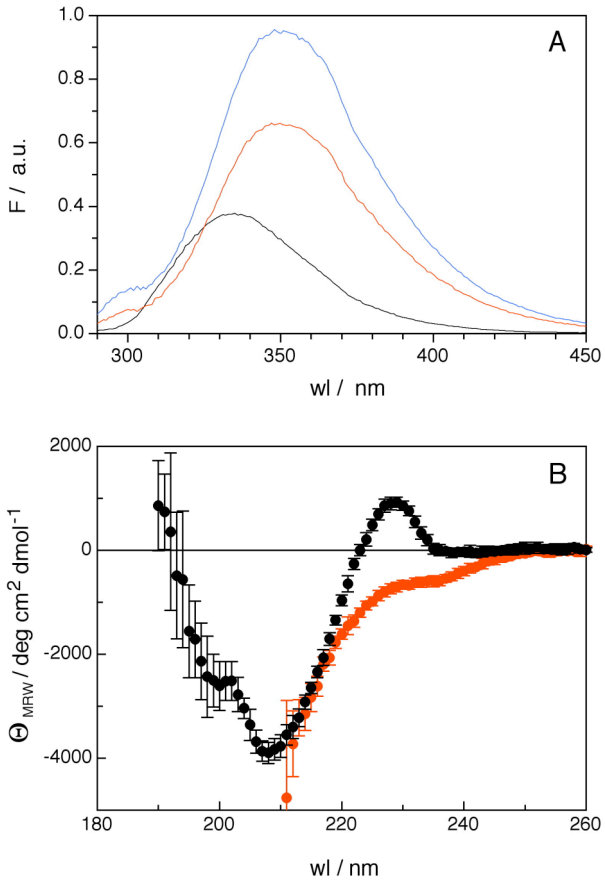


Figure 6-6

Fluorescence and far UV CD spectra of SFVP F160W. A: Fluorescence emission for λ_{exc} = 280 nm.1 µM protein, native (black), in 8.5 M GdmCl (blue) and in 6.6 M urea (red). B: far UV CD signal of the native protein (138 µM, 0.1 mm cuvette) and the denatured protein (20 µM, 1mm cuvette, in 8.1 M urea). pH 7.5, 25ºC, 0.1 mM EDTA, 20 mM HEPES, 150 mM KCl.

Far UV CD signal for the native SFVP F160W shows a small maximum at 229 nm and a pronounced minimum at 208 nm. The denatured spectrum is typical of a random coil state.

6.2.2 Equilibrium studies of SFVP F160W

Urea and GdmCl dependence of SFVP F160W stability were investigated. Denaturant-induced equilibrium transitions monitored by Trp fluorescence showed no sign of the population of an intermediate. Both urea- and GdmCl-induced denaturation can be described by a two state approximation.

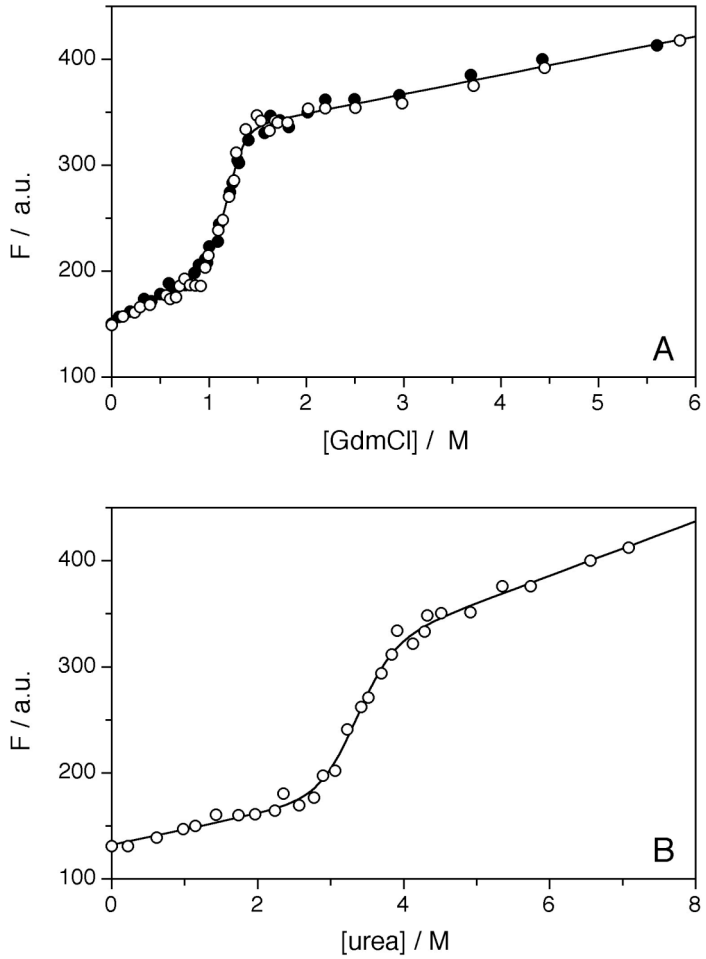


Figure 6-7

Denaturant-induced equilibrium studies of SFVP F160W. 20 mM HEPES, 0.1 mM EDTA, 150 mM KCl, pH 7.5, 25ºC, 1µM SFVP F160W., λ_{exc}: 280 nm, λ_{em}: 355 nm. A: GdmCl as denaturant, filled and open circles are two different experiments, B: urea as a denaturant. The solid line represents a two-state fit. The results are given in the text.

The free energy extrapolated to 0 M GdmCl is ΔG^0 (H_2O) = 32.9±0.4 kJ, and the change in free energy with GdmCl ($m_{eq} = \partial\Delta G^0/\partial$[GdmCl]) is 27.5±0.3 (kJ/mol) / M. Within error, the same free energy is obtained when using urea as a denaturant, ΔG^0 (H_2O) = 32.2±0.3 kJ/mol with m_{eq} = 9.65±0.01 (kJ/mol) / M. The mutation F160W thus does not seem to have influenced equilibrium stability of SFVP. For wild-type SFVP, a $\Delta G^0(H_2O)$ = 32.4±3.3 kJ/mol, m_{eq} = 9.0 ±0.9 (kJ/mol) / M in urea was obtained.[102] In GdmCl, the m_{eq} at 25ºC is 32.9±3.9 (kJ/mol) / M, (see 6.1.2), which is close to the value obtained for the mutant. This also validates the assumption that mutant and wt SFVP share the same structure. A comparison of the homology model and the wild-type structure (Figure 6-8) shows only minor rearrangements in the helix due to the F160W mutation.

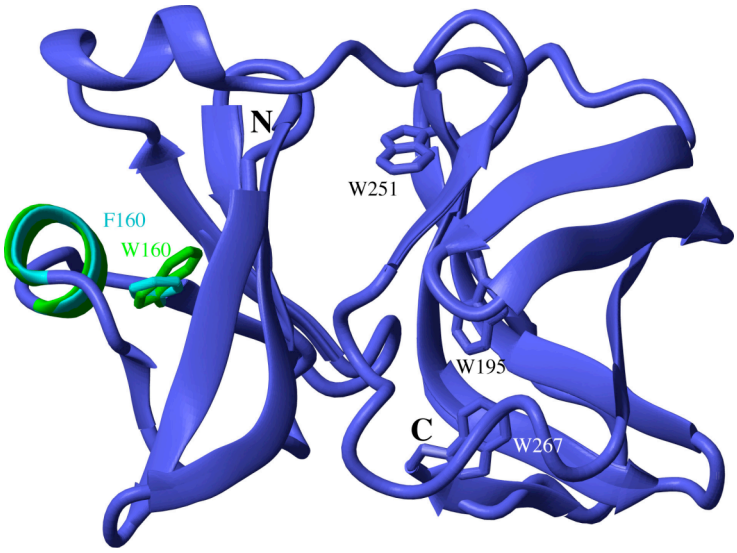


Figure 6-8

Comparison of the wild-type crystal structure and the homology model of the mutant. Slight differences between the wt (cyan) and SFVP F160W (green) can be observed at res. 160 and the neighbouring α-helix. Image created with MOLMOL.[88]

The mutant should therefore be well suited to study the N-terminal domain, as it does not disrupt the structure upon introducing a label.

6.2.3 Kinetic studies of SFVP F160W

In order to further characterise the mutant SFVP, kinetic experiments were performed. The urea dependence of unfolding and refolding was investigated using stopped-flow and manual mixing techniques, monitored by Trp fluorescence.

6.2.3.1 Urea- dependence of folding and unfolding kinetics of SFVP F160W

Unfolding and refolding kinetics were measured between 0.5 and 7.5 M urea with stopped-flow mixing at 25ºC, pH 7 and 1 µM protein concentration. An example trace is shown in Figure 6-9: refolding was initiated by rapid dilution to 0.9 M GdmCl. The trace could be fitted to the sum of five exponentials. The rate constants and amplitudes are summarised in Figure 6-9 and Table 6-1.

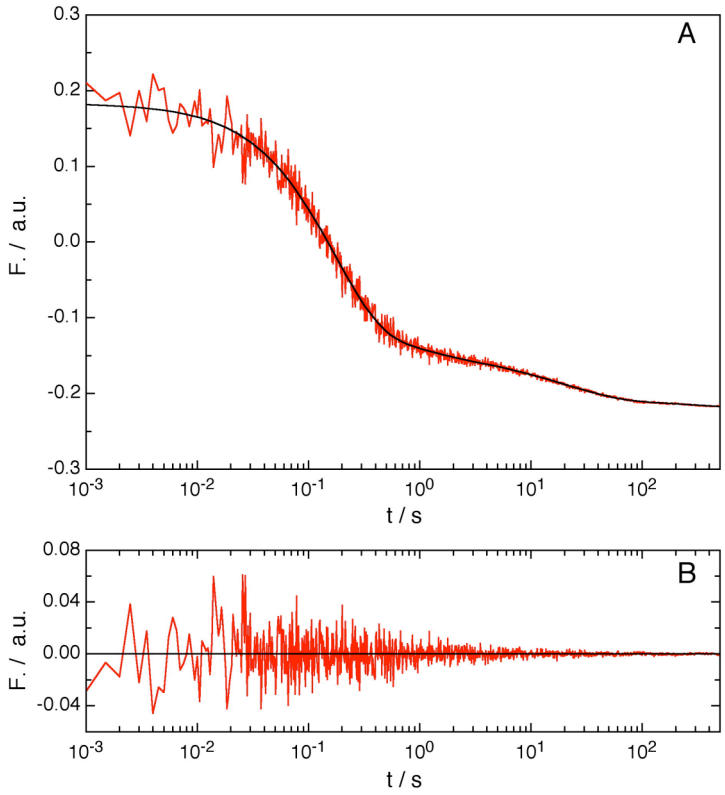


Figure 6-9

Refolding of SFVP F160W at 0.9 M urea. A: red: stopped-flow refolding trace; The solid black line indicates a fit to the sum of five first-order reactions. The residuals are shown in B. A_2 = 0.0303±9.86·10^{-3}; λ_2 = 1.02±0.09 s^{-1}; A_3 = 0.302±0.0114; λ_3 = 6.04±0.25 s^{-1}; A_4 = 0.0428±0.0382; λ_4 = 0.0362±0.0362 s^{-1}; A_5 = 0.0122±0.0857; λ_5 = 2.68e-3±0.0158 s^{-1}; A_7 = 0.01835±0.0423; λ_7 = 0.131±0.297 s^{-1}; Conditions: 20 mM HEPES, 0.1 mM EDTA, 150 mM KCl, pH 7.5, 25ºC, 1 µM SFVP F160W., λ_{exc}: 280 nm, λ_{exc}: ≥ 320 nm.

To describe unfolding and refolding traces, up to five exponentials were needed. The dependence of the kinetics on the urea concentration, and a comparison to equilibrium data is shown in Figure 6-10.

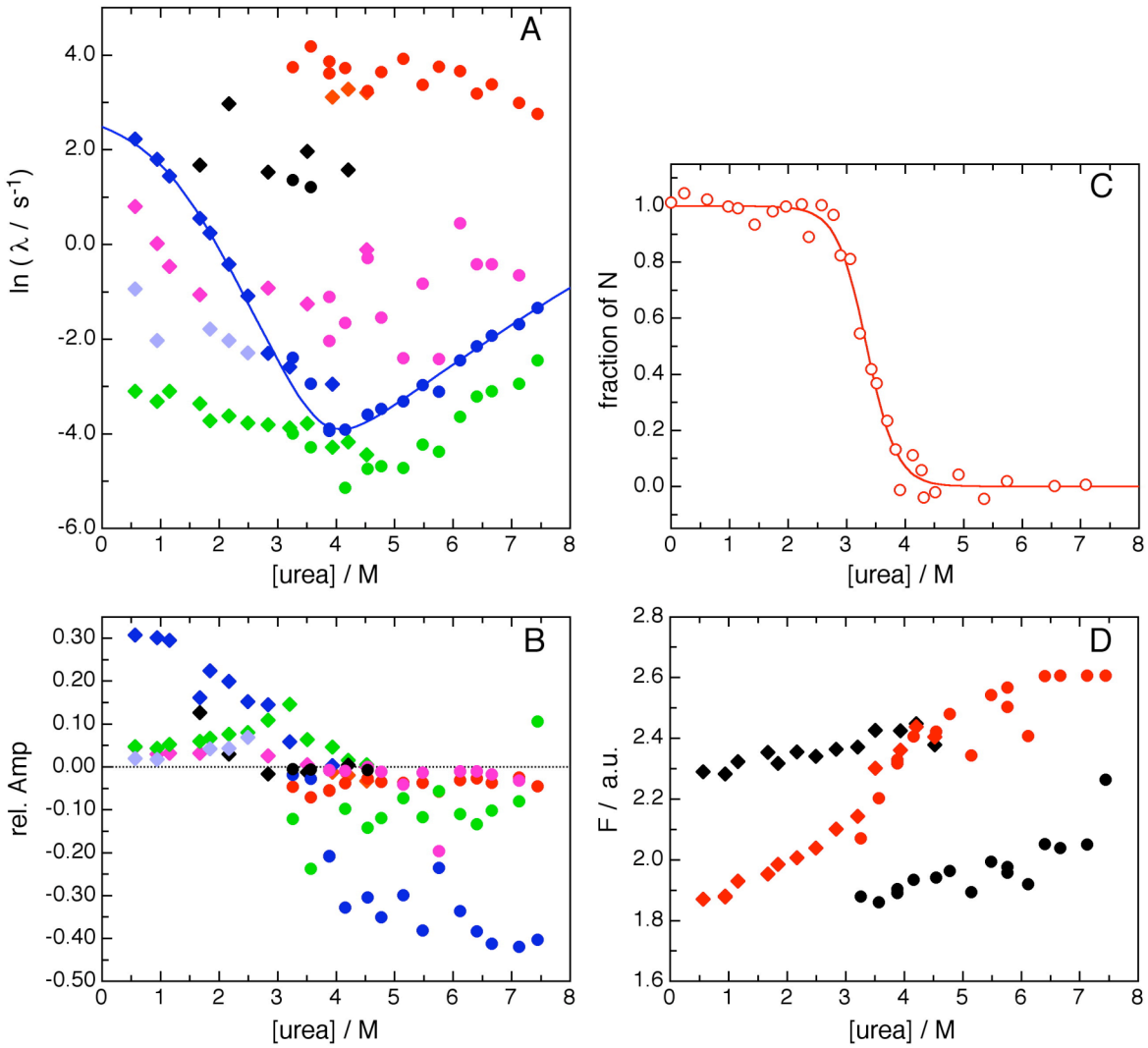


Figure 6-10

Urea dependence of SFVP F160W folding. A: GdmCl dependence of unfolding (circles) and refolding (diamonds). The solid line represents a fit of the major unfolding/refolding phase to a three-state model with a populated on-pathway intermediate. k_1 =13 s^{-1}, m_1 =-0.38, k_{-1} =990 s^{-1}, m_{-1} =0.62, k_2 =4.0·10^5 s^{-1}, m_2 =0.026, k_{-2} =0.16 s^{-1}, m_{-2} =0.30, K_{UI} =5.7·10^5 kJ/mol, m_{eq} =-2.2 (kJ/mol) / M. B: the corresponding amplitudes to A are given in the same colour code: phase 1: red, phase 2: pink, phase 3: blue, phase 4: green, phase 5: black, phase 6: pale blue. C: The Equilibrium transition correlates with the kinetic data. the solid line represents a two-state global fit with $\Delta G^0(H_2O)$ = -32.2 ± 0.31 kJ/mol and m_{eq} = 9.65 ± 0.09 (kJ/mol) / M D: Starting points (black) and endpoints (red) from unfolding (circles) and refolding (diamonds) traces. The complete amplitude seems to be resolved. Conditions: 20 mM HEPES, 0.1 mM EDTA, 150 mM KCl, pH 7.5, 25ºC, 1µM SFVP F160W., λ_{exc}: 280 nm, λ_{exc}: ≥ 320 nm.

The major refolding/unfolding phase contributes to about 70% of the complete fluorescence change. In refolding, four to five phases can be observed. Apart from the major refolding phase (blue diamonds), a fast phase (black diamonds) is detected. Three slower phases are also present in refolding (pink, light blue and green diamonds). At low urea concentrations, the main phase and the fast refolding phase (black) probably become very similar, which results in a roll-over of the main phase. The very fast phase (red) present in unfolding and refolding does not change with urea concentration, and could therefore be an artefact. Compared to the wild-type protein, unfolding has also become more complex, three to four unfolding phase are observed. Apart from the main refolding phase, there is an additional slower and an additional faster phase. The slower phase is not an artefact, it was also observed in manual mixing experiments, as shown in Figure 6-10 (blue circles).

The rate constants and amplitudes at 0.9 M urea are summarised in Table 6-1.

6.2.3.2 The fast phase of refolding

The fast phase at low urea concentrations was also investigated separately. To resolve the fast phase better, native protein was unfolded only for a short time, before refolding to different urea concentrations was initiated (interrupted unfolding setup[84], "double jumps"). The aim was to unfold for such a short time that Xaa-prolyl bonds are still in their native conformation, so that no slow *cis/trans* isomerisation reactions would take place.

As the mutant SFVP contains a Trp residue in the N-terminal domain that folds first, it was expected to get a better signal for the fast phase in refolding compared to the wild-type protein.

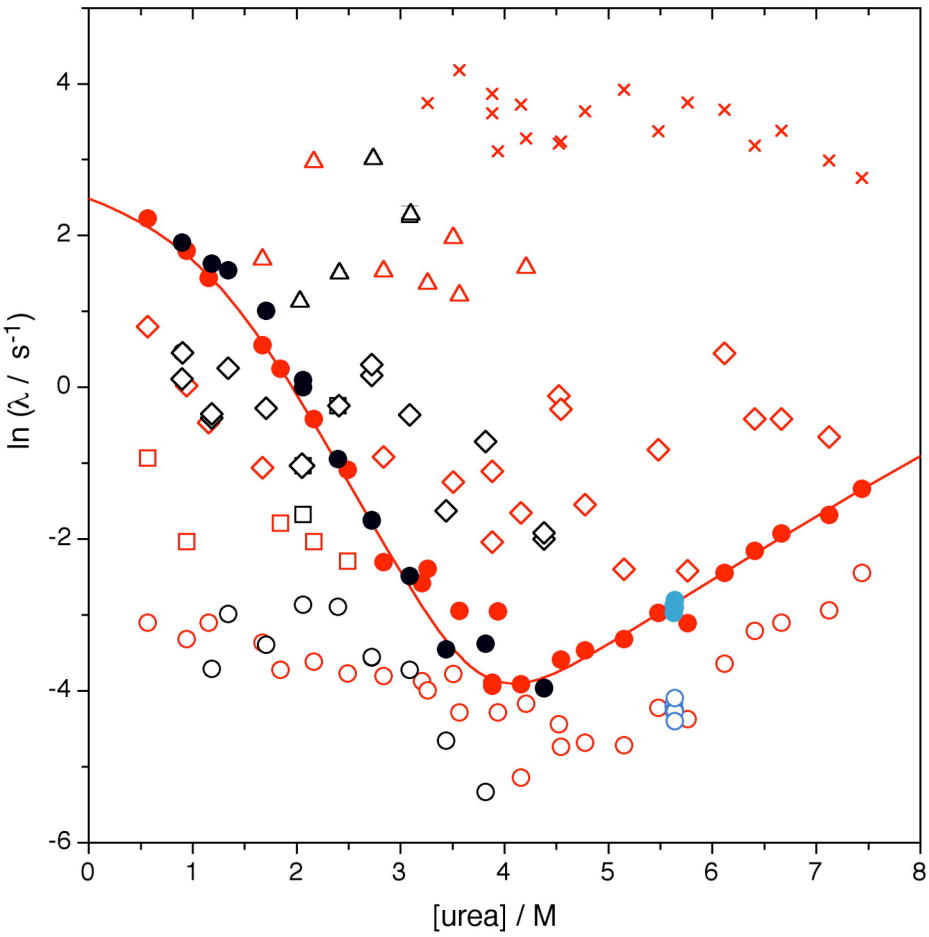


Figure 6-11

Overlay of the unfolding and refolding kinetics with the data from interrupted unfolding experiments. Data from single mixing experiments is represented by red symbols, data obtained with interrupted unfolding is shown in black. The open blue and green circles at 5.7 M urea are obtained from a manual mixing unfolding experiment. Conditions: 20 mM HEPES, 0.1 mM EDTA, 150 mM KCl, pH 7.5, 25ºC, 1µM SFVP F160W., λ_{exc}: 280 nm, λ_{exc}: ≥ 320 nm.

Due to the many refolding phases, the fast phase, however, could not be resolved better with the interrupted unfolding setup.

6.2.3.3 pH dependence of unfolding

The pH dependence of unfolding was investigated to find conditions where unfolding is fast enough for the stopped-flow double jump experiment described in section 6.2.3.2. Unfolding was initiated by sixfold dilution to different pH and 5.4 M urea. The buffer conditions were matched with the first step in the double jump experiment (unfolding). (See Materials and Methods, 7.2.2).

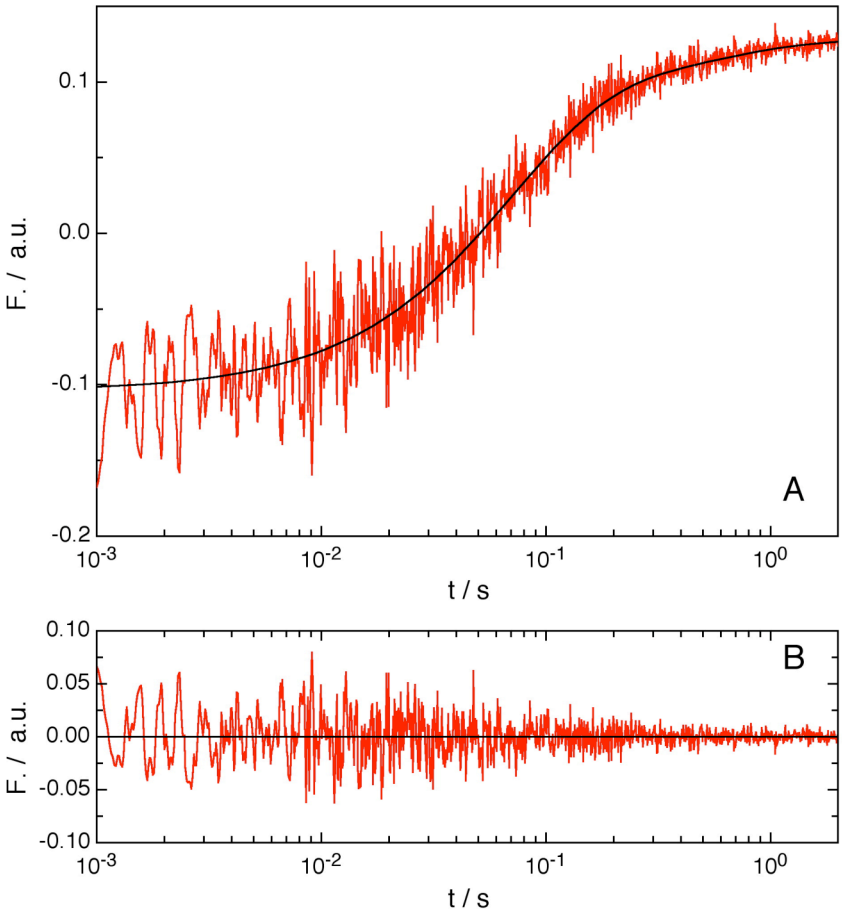


Figure 6-12

Unfolding of SFVP at pH 4.1, 5.4 M urea. The unfolding trace could be fitted to the sum of two first order reactions. A_1 = -0.194±9.74·10^{-3}, λ_1=14.4 ± 0.861 s^{-1}, A_2=-0.0377±7.78·10^{-3}, λ_2=1.900±0.921 s^{-1}. Conditions: pH 4.1, 5.4 M urea, mM glycine, 0.1 mM EDTA, 150 mM KCl, 2.2 mM HEPES, 1 µM protein. λ_{exc}: 280 nm, λ_{exc}: ≥ 320 nm.

The traces could be fitted to the sum of two exponentials. An example trace is shown in Figure 6-12. The observed rate constants and amplitudes obtained are: A_1 = -0.194±9.74·10^{-3}, λ_1=14.4 ± 0.861 s^{-1}, A_2=-0.0377±7.78·10^{-3}, λ_2=1.900±0.921 s^{-1}

A plot of the logarithm of all unfolding kinetics versus the different pH is shown in Figure 6-13. The two rate constants increase with decreasing pH. Below pH 2, the reaction is faster than the mixing dead time of the stopped-flow experiment.

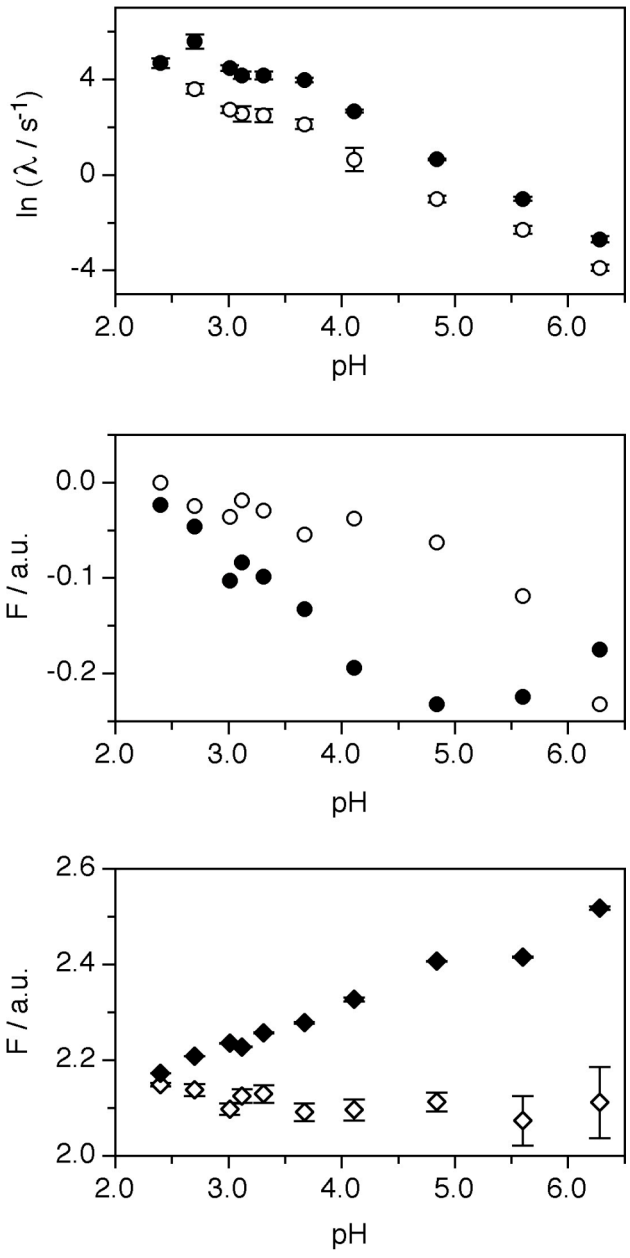


Figure 6-13

pH dependence of SFVP F160W unfolding. A: logarithm of the observed rate constant (l) versus pH; B: corresponding amplitudes; C: initial (open diamonds) and final values (closed diamonds) of unfolding traces. Conditions: 6 µM SFVP F160W in 150 mM KCl, 13 mM HEPES, at pH 7.5 was mixed in the first mixing step with 6.47 M urea, 12 mM glycine, 150 mM KCl, 0.1 mM EDTA of different pH, to achieve unfolding conditions of different pH, 5.39 M urea, 10 mM glycine, 0.1 mM EDTA, 150 mM KCl, 2.2 mM HEPES, 1 µM protein. l_{exc}: 280 nm, l_{exc}: ≥ 320 nm.

Unfolding at pH 2 and 5.4 M urea is sufficiently fast for complete unfolding of SFVP F160W within 50 ms, so that it provides ideal unfolding conditions for the first mixing step in the double jump experiment. As unfolding is complete in such a short time, slow equilibration processes in the unfolded state, such as *cis/trans* isomerisation, should not take place before refolding is initiated. Thus only fast refolding processes should be observed with peptide bonds still in the native conformation.

6.3 Conclusion

The kinetics of the SFVP wild-type and the F160W mutant look very similar. To illustrate this, in Figure 6-14 both chevrons are plotted together.

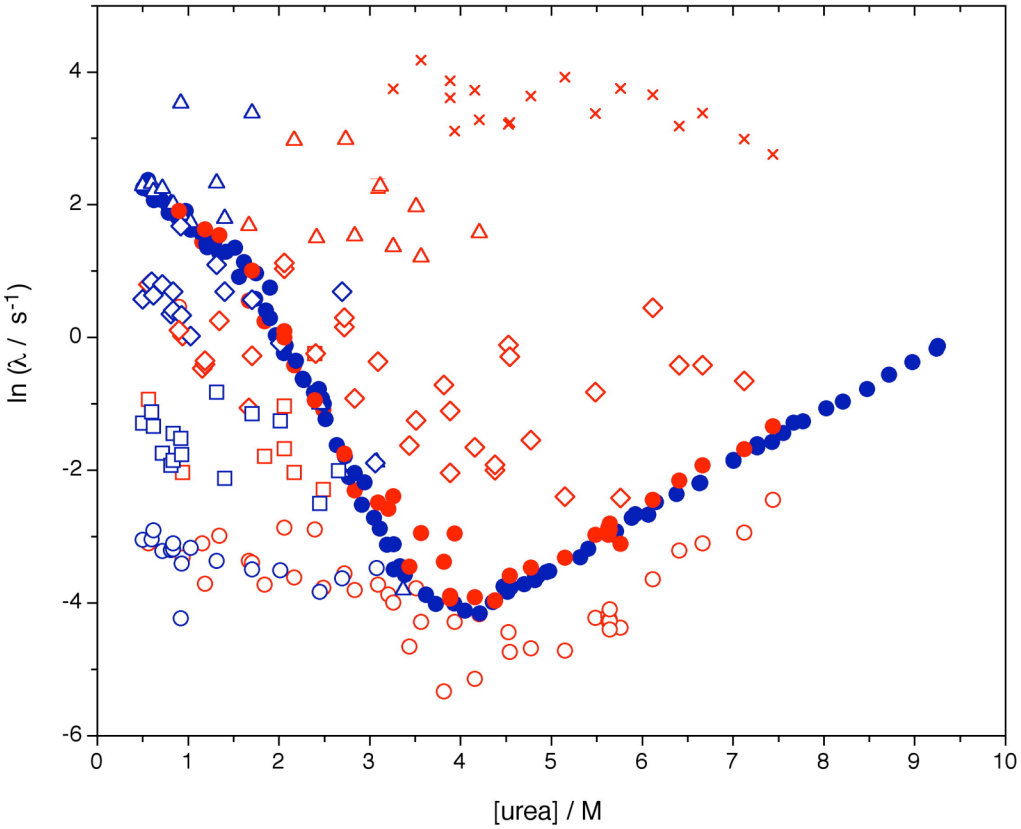


Figure 6-14

Comparison of the denaturant dependence of folding kinetics of SFVP and SFVP F160W. blue: SFVP wt, red: SFVP F160W chevron. Corresponding rate constants share the same symbols. Conditions: 20 mM HEPES, 0.1 mM EDTA, 150 mM KCl, pH 7.5, 25ºC, 1µM SFVP F160W., λ_{exc}: 280 nm, λ_{exc}: ≥ 320 nm.

The wild-type protein unfolds in a single step, whereas three additional unfolding phases could be detected for the mutant SFVP. By manual mixing unfolding measurements it was

shown that the very slow additional unfolding phase of the mutant is not a stopped-flow artefact. The other two additional unfolding phases are both faster than the main unfolding phase also present in the wild-type protein.

A comparison of the refolding rate constants at 0.9 M urea is shown in Table 6.1.

	SFVP F160W	wt SFVP	In wt, this phase is:
A_2 (%)	7.69 ± 32.3	11.9 ± 1.3	compatible with non-prolyl *cis/trans* isomerisation
L_2 (s^{-1})	1.02 ± 0.09	1.4 ± 0.3	
A_3 (%)	76.7± 3.74	71.3 ± 2.8	Main folding phase, all Pro residues in *trans*
L_3 (s^{-1})	6.04 ± 0.25	6.3 ± 0.2	
A_4 (%)	11.0 ± 84.4	12.7 ± 1.5	compatible with prolyl *cis/trans* isomerisation
L_4 (s^{-1})	0.0362 ± 0.036	0.033 ± 0.004	
A_6 (%)	4.57 ± 208	4.1 ±1.8	compatible with prolyl *cis/trans* isomerisation
L_6 (s^{-1})	0.131 ± 0.297	0.17 ± 0.10	

Table 6-1

Rate constants and amplitudes of SFVP and SFVP F160W folding at 0.9 M urea. Wild-type data is taken from literature.[102] pH 7.0, 25ºC, 20 mM Hepes, 150 mM KCl, 0.1 mM EDTA

A two-state fit of the equilibrium data yields an $\Delta G^0(H_2O)$ of 37.3±4.5 kJ/mol with m_{eq} = 32.9±3.9 (kJ/mol) / M,

As can be seen in Table 6.2, the equilibrium stability does not change much, either.

protein	Denaturant	$\Delta G^0(H_20)$ (kJ/mol)	m_{eq} ((kJ/mol)/M)
SFVP F160W	GdmCl	-32.9 ± 0.40	27.5 ± 0.33
SFVP F160W	urea	-32.2 ± 0.31	9.65 ± 0.09
SFVP wild-type	GdmCl	-37.3±4.5	32.9±3.90
SFVP wild-type	Urea	-32.4±3.29	9.03±0.91

Table 6-2

Equilibrium stability and m_{eq} for SFVP wt and SFVP F160W. pH 7.0, 25ºC, 20 mM Hepes, 150 mM KCl, 0.1 mM EDTA. Data fitted to a two-state model. (eq 15 in Materials and Methods). wild-type data with urea is taken from literature. [102]

This study has shown that SFVP and SFVP F160W have very similar thermodynamic and kinetic properties. The fast refolding phase, speculated to originate from folding of the N-terminal domain, could not be resolved in more detail with the new mutant.

7 Materials and Methods

7.1 Protein synthesis and purification

7.1.1 Foldon

^{13}C,^{15}N-labelled foldon for NMR studies was expressed in BL21 in ^{13}C,^{15}N enriched M9 medium as described in section 4.2

Foldon for other experiments was synthesised using the standard protocol for Fmoc solid phase peptide synthesis on a ABIMED economy peptide synthesiser EPS221 from Abimed, Germany. The C-terminal amino acid Leu was precoupled to the resin (Tentagel S Leu Fmoc from Rapp Polymere, Tübingen, Germany), one reaction circle was used per residue (7 min deprotection time, 25 min coupling time). Protected amino acids were purchased from IRIS Biotech (Marktredwitz, Germany) and from Alexis Biochemicals, (USA), solvents and other chemicals from Fluka (Switzerland). After HPLC purification on a C8 reverse phase preparative column (Hibar, LiChrosorb®100 RP-8 from Merck, Darmstadt, Germany), peptide fractions were lyophilised and stored at -20ºC. Mass and purity of the product was determined by nanospray mass spectroscopy and analytical HPLC.

7.1.2 SFVP

SFVP was expressed and purified by Manuel Morillas and Eva Zobeley in the group of R. Glockshuber (ETH Zürich), as described previously[102]. Protein fractions solved in buffer were stored at -20ºC.

7.2 Measurements

For all spectroscopic measurements, buffers were filtered and degased.

Standard conditions for foldon were 10 mM phosphate buffer at pH 7.1 and 20ºC, unless indicated otherwise. Protein concentration was determined by absorption measurement in 6 M GdmCl.[114] Foldon samples of different denaturant concentrations were prepared 10 h before the first measurement to make sure equilibrium conditions prevail.

Standard conditions for SFVP were 150 mM KCl, 0.1 mM EDTA, 20 mM HEPES at pH 7.5 and 25ºC, if not specified differently. For fluorescence experiments, a protein concentration of 1 µM protein was used, determined by absorption at 280 nm in the native state. SFVP sand SFVP F160W samples were prepared 30 min before measurement.

Salts used for buffers, as well as urea and ultra-pure fluorescence grade HEPES were from Fluka (Switzerland), and Guanidinium chloride (grade a.a.) from NIGU Chemie, Germany.

7.2.1 Equilibrium measurements

Fluorescence spectra and equilibrium transitions were recorded on an Aminco Bowman 2 Spectrofluorimeter (SLM Aminco, USA). Emission spectra were recorded at 280 nm and 295 nm excitation. The scan rate for spectra was 1 nm/sec, excitation and emission slots as well as the photo multiplier voltage were adjusted according to the strength of the fluorescence signal.

CD spectra were accumulated eight times with a step size of 1 nm and a time constant of 1 sec on a Aviv circular dichroism spectrometer (model 62ADS, Aviv, USA). Far UV spectra were recorded in a 1 mm cell and in a 0.1 mm cell.

7.2.2 Kinetic experiments

Kinetic traces were recorded with a stopped flow spectrometer (model SX.18MV, from Applied Photophysics, Leatherhead, UK, equipped with a Hamamatsu R1104 photomultiplier tube), as well as with a AB 2 fluorimeter after manual mixing.

Single mixing experiments with 1:10 dilution were performed with a stopped flow spectrometer and by manual mixing. For manual mixing experiments, after 10 minutes, shutters were closed repeatedly for several minutes to prevent photo bleaching.

For sequential mixing experiments, 36 µM SFVP F160W in 150 mM KCl, 13 mM HEPES, at pH 7.5 was mixed in the first mixing step with 6.47 M urea, 12 mM glycine, 150 mM KCl, 0.1 mM EDTA, pH 2.98 to achieve unfolding conditions of pH 2.95, 5.39 M urea, 10 mM glycine, 0.1 mM EDTA, 150 mM KCl, 2.2 mM HEPES, 6 µM protein. After unfolding for 500 ms, refolding to different urea concentrations between 0.9 and 4.4 M at pH 7.5 was achieved by mixing with refolding buffer (pH 7.5, 23.56 mM HEPES, 0.1 mM EDTA, 150 mM KCl, 0.1 mM EDTA and different amounts of urea. This refolding reaction at final conditions of pH 7.5, 20 mM HEPES, 150 mM KCl, 0.1 mM EDTA, 0.9 -4.4 M urea, 1.67 mM glycine was monitored by fluorescence emission above 320 nm, (excitation at 280 nm).

7.3 Data analysis

The data was analysed using proFit (Quantum Soft, Zürich, Switzerland) and Matlab (The MathWorks, Natick, MA, USA).

NMR data was processed with NMRPipe[115] and analyzed with PIPP,[116] as described in section 4.2.

Protein structure images were generated with MOLMOL[88] and MolScript.[97]

7.4 Analysis of equilibrium unfolding curves

7.4.1 Two-state behaviour

For a chemical reaction

$$A \rightleftharpoons B$$

ΔG^0 is related to the equilibrium constant K_{eq}

by $$K_{eq} = e^{-\Delta G^0/RT} = e^{-\Delta H^0/RT} \cdot e^{-\Delta S^0/R} \quad \text{(eq 12)}$$

$$\Delta G^0 = -RT \ln K_{eq} \quad \text{(eq 2)}$$

with $$K_{eq} = \frac{[B^\infty]}{[A^\infty]} \quad \text{(eq 13)}$$

The stability can be determined in denaturant-induced equilibrium transitions, where the amount of native protein depends on the amount of a denaturant (Den). It has been found experimentally that the stability depends linearly on the denaturant concentration:[3; 117]

$$\Delta G^0 = \Delta G^0(H_2O) + m_{eq} \cdot [denaturant] \quad \text{(eq 14)}$$

The m_{eq} value reflects the dependence of the free energy of unfolding on the denaturant concentration and was shown to correlate with the change in solvent-accessible surface area upon unfolding.[13] ΔG^0 and the m_{eq} value can be obtained by plotting the different $\ln K_{eq}$ within the transition region versus the denaturant concentration: From a linear extrapolation one obtains $\Delta G^0(H_2O)$ at the intercept of the ordinate (where [Den] = 0), and the slope equals the m_{eq} value. The denaturant dependence of protein stability can also be described in one step, as described in (eq 15).

$$f([Den]) = \frac{(N_0 + m_n \cdot [Den]) + (U_0 + m_u \cdot [Den]) \cdot e^{\frac{-(\Delta G - m_{eq} \cdot [Den])}{RT}}}{1 + e^{\frac{-(\Delta G - m_{eq} \cdot [Den])}{RT}}} \qquad \text{(eq 15)}$$

N_0 and m_n are the intercept and the slope of the native baseline, U_0 and m_U of the unfolded baseline. [Den] is the denaturant concentration.

7.4.2 Three state model for equilibrium unfolding

$$f([Den]) = \frac{F_N([Den]) + F_I([Den]) \cdot K_{NI} + F_U([Den]) \cdot K_{NI} \cdot K_{IU}}{1 + K_{NI} + K_{NI} \cdot K_{IU}} \qquad \text{(eq 16)}$$

with the native baseline:

$$F_N([Den]) = N_0 + m_N \cdot [Den] \qquad \text{(eq 16a)}$$

the intermediate baseline:

$$F_I([Den]) = I_0 + m_I \cdot [Den] \qquad \text{(eq 16b)}$$

and the unfolded baseline:

$$F_U([Den]) = U_0 + m_U \cdot [Den] \quad \text{(eq 16c)}$$

as well as with the equilibrium between N and I, and I and U:

$$K_{NI} = e^{-\frac{\Delta G_{NI} - m_{eq,NI} \cdot [Den]}{RT}} \quad \text{(eq 17a)}$$

$$K_{IU} = e^{-\frac{\Delta G_{IU} - m_{eq,IU} \cdot [Den]}{RT}} \quad \text{(eq 17b)}$$

N_0 and m_n are the intercept and the slope of the native baseline, I_0 and m_i of the intermediate baseline and U_0 and m_U of the unfolded baseline. [Den] is the denaturant concentration. $m_{eq,NI}$ and $m_{eq,IU}$ are the m-values for the N-to-I transition and the I-to-U transition, respectively. ΔG_{NI} and ΔG_{UI} are the free energies differences between the native and the intermediate state, and between the unfolded and the intermediate state.

7.4.3 Analysis of two-state monomer-trimer equilibrium transitions

Section 9.4 describes the equilibrium behaviour of monomeric proteins, with only two states populated in equilibrium. One of the protein investigated, the foldon domain, is trimeric. Therefore, a monomer-trimer transition has to be considered:

$$3\,U \rightleftharpoons N_3$$

with $$K = \frac{[N_3]}{[U]^3} \quad \text{(eq 18)}$$

From combination of the monomer two-state approximation with the monomer - trimer equilibrium the following function is obtained. The derivation is explained in detail by Krause S (2001)[109] and in Güthe *et al.*[106]

$$f([Den]) =$$

$$F_N([Den]) - \frac{F_N([Den]) - F_U([Den])}{3 \cdot [M_0]} \cdot \left(\sqrt[3]{e^{\frac{-\Delta G_{H_2O} + m_{eq}[Den]}{RT}} \cdot \frac{9}{2}[M_0] + \sqrt{D}} + \sqrt[3]{e^{\frac{-\Delta G_{H_2O} + m_{eq}[Den]}{RT}} \cdot \frac{9}{2}[M_0] - \sqrt{D}} \right) \quad \text{(eq 19)}$$

with the native baseline

$$F_N([Den]) = N_0 + m_N \cdot [Den] \quad \text{(eq 19a)}$$

and the unfolded baseline

$$F_U([Den]) = U_0 + m_U \cdot [Den] \quad \text{(eq 19b)}$$

and

$$D = \left(\frac{9}{2}\right)^2 \cdot [M_0]^2 + e^{\frac{-\Delta G_{H_2O} + m_{eq}[Den]}{RT}} \quad \text{(eq 19c)}$$

The midpoint $[Den]_{1/2}$ of the transition, which is defined here as the point where the amount of native proteins equals the amount of denatured protein, is given by

$$[Den]_{1/2} = \frac{\Delta G^0{}_{H_2O} + RT \cdot \ln\left(0.75 \cdot [M_0]^2\right)}{m_{eq}} \quad \text{(eq 20)}$$

where $\Delta G^0(H_2O)$ and m_{eq} are obtained from the fit. m_{eq} reflects the sensitivity to denaturants of $\Delta G^0(H_2O)$

7.5 Tyrosine → Tryptophan Energy Transfer

Förster energy transfer or fluorescence resonance energy transfer (FRET) describes the photon free transfer of an initially excited donor to an acceptor, whose absorption spectrum must overlap with the emission spectrum of the donor. As the transfer results from long-range dipole-dipole interactions it depends on the orientation of donor and acceptor and on their distance. The quantum yield and the overlap of the donor's emission spectrum with the acceptor's absorption spectrum also contributes to the extent of energy transfer.

The rate of energy transfer kT is given by the following expression:

$$k_T(r) = \frac{Q_D \kappa^2}{\tau_D r 6} \cdot \left(\frac{9000 \cdot \ln 10}{128 \pi^5 N_A n^4} \right) \cdot \int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda \qquad \text{(eq 21)}$$

or simpler $$k_T(r) = \frac{1}{\tau_D} \left(\frac{R_0}{r} \right)^6 \qquad \text{(eq 22)}$$

where r is the distance between donor and acceptor, Q_D the quantum yield of the donor, κ^2 describes the relative orientation, τ_D the life time of the donor in absence of the acceptor, N_A Avogadros number, n is the refractive index of the medium, $F_D(\lambda)$ is the corrected fluorescence intensity of the donor in a wavelength range with the total intensity normalised to 1, $\varepsilon_A(\lambda)$ is the extinction coefficient of the acceptor at the wavelength λ. R_0 is the Förster distance, the distance where energy transfer is 50% efficient. R_0 depends on the donor-acceptor pair and is usually in the range of 10 Å to 60 Å.[118] As can be seen from equation 22, energy transfer depends on the overlap integral, the orientation of donor and acceptor and on the wavelength. The efficiency of energy transfer, E, can be expressed as follows:

$$E = \frac{k_T}{(\tau D)^{-1} + k_T} \qquad \text{(eq 23)}$$

which together with (eq 47) yields:

$$E = \frac{R_0^6}{R_0^6 + r^6} \qquad \text{(eq 24)}$$

In proteins, efficient energy transfer is also possible. In most of the cases, energy is transferred from tyrosine to tryptophan, but tryptophan→ tryptophan and phenyl-alanine→tryptophan energy transfer has also been observed. R_0 for tyrosine→ tryptophan energy transfer, is between 9 Å and 18 Å. Usually the energy transfer takes place only in the native protein, as the distance between donor and acceptor usually is larger than R_0 in the unfolded protein. The occurence of energy transfer is an indication that the protein is in a compact state.

Tyrosine→tryptophan energy transfer can be tested quite easily by measuring emission spectra at two different excitation wavelengths: at 280 nm, where Tyr as well as Trp absorb, and at 295 nm, where only Trp absorbs light significantly. If energy is transferred from tyrosine to tryptophan, then the contribution of tyrosine to the emission spectrum is not visible and only the emission spectrum of tryptophan at is observed both excitation wavelengths. If the spectra at the two wavelengths differ, then no energy transfer occurs.

8 Summary

The mechanisms and speed limits of evolutionary optimised complex folding reactions were investigated in this study. The two viral model proteins, foldon and SFVP, have evolved under selective pressure for fast and efficient folding and association. The function of foldon, the C-terminal domain of the trimeric phage T4 protein fibritin, is to support rapid folding and association of fibritin. SFVP, a protease from the Semliki forest virus, has to cleave itself out of a large polyprotein chain co-translationally during protein synthesis. This promotes the biogenesis of the more C-terminal peptide segments in the viral polyprotein.

This work has revealed that even complex folding reactions like quaternary structure formation in foldon and folding of the two-domain SFVP can occur very fast and efficiently via on-path folding intermediates.

The kinetic and thermodynamic studies on the foldon domain provide the first detailed information on a fast-folding trimeric protein. The foldon domain forms a β-propeller-like structure. All folding and association steps seem to be optimized for rapid formation of a stable trimer. At low protein concentration, the folding reaction shows apparent third-order kinetics. From the compact monomeric intermediate formed within the first milliseconds, the native trimer is formed in two consecutive association reactions with bimolecular rate constants up to $6 \cdot 10^6$ $M^{-1}s^{-1}$. This is close to the fastest association reactions reported for dimeric proteins. At low protein concentrations, an intermediate species is populated to a small amount at equilibrium. This influences the unfolding kinetics only to a small extent, and does not affect the equilibrium stability at all. With increasing protein concentrations folding becomes independent of the protein concentration, indicating that a first-order folding step from a partially folded trimer to the native protein becomes rate-limiting. The half-time of

about 3 ms is comparable to fast-folding small single domain proteins. In contrast to small monomeric proteins, however, intermediates become populated very fast, probably providing suitable interfaces to assist trimerisation. Foldon is the first protein where these interfaces are shown to be formed by β-hairpins. At pH 2, the trimer disintegrates into a monomeric A-state with almost identical fluorescence properties as the monomeric folding intermediate. Thus an equilibrium species very similar to the kinetic intermediate is accessible to NMR studies. The A-state consists of a β-hairpin with the same hydrogen bond pattern as the native trimer, but lacks structure in the N- and C- terminus. The hairpin forms within a few microseconds, and its stability is comparable to designed hairpins in alcohol/water mixtures. With state-of-the-art solution NMR techniques, in particular residual dipolar coupling measurements, new insights on structural preferences in the unfolded and partially folded monomer states were gained.

SFVP folds 1000 times faster than any other two-domain protein investigated so far. Studies with wild-type SFVP demonstrated that an intermediate is populated in GdmCl-induced equilibrium transitions at 10ºC. At 25ºC, only two states are populated in GdmCl and in urea. A mutant of SFVP containing an additional fluorescence probe in the N-terminal domain (SFVP F160W) shows identical equilibrium behaviour and very similar folding and unfolding kinetics. Additional unfolding phases besides the main phase also present in the wild-type protein are observed: a slower one and a faster one. The fast folding phase probably originates from the formation of the N-terminal domain. It is difficult to separate the fast and the main folding phase at low urea concentrations, as the two rate constants become to similar. These results provide further evidence for the folding model of SFVP, where an intermediate is slowly formed from the unfolded state (τ = 170 ms), and then rapidly converts to the native protein (τ = 37 ms), so that the intermediate is only populated to a small extent.

9 Acknowledgements

This work was carried out from December, 2001 to May, 2005 in the laboratory of Prof. Dr. Thomas Kiefhaber in the Department of Biophysical Chemistry at the Biozentrum of the University of Basel, Switzerland.

First of all, I would like to thank Thomas Kiefhaber for guiding me through the exciting field of protein folding. I have learned a lot from his superb scientific approach.

I also thank all present and former members of the Kiefhaber lab, for technical help on protein folding and peptide synthesis, and for social interactions.

Furthermore I would like to express my thanks to Jürgen Engel and his group, (Biozentrum, Dept. Biophysical Chemistry), for igniting the foldon project, as well as to Stephan Grzesiek for being interested in working with foldon, and Sebastian Meier in his group for the excellent collaboration.

I would also like to thank Rudi Glockshuber at the ETH Zürich and Manuel Morillas in his group to start the SFVP project as well as the collaboration with Ignacio Sánchez from the our lab; my special gratitude goes to Manuel for expression and purification of SFVP, and design of the Trp variant.

Leo, Gernot and Hans from the workshop did excellent work on instrument maintenance and repair. Further thanks go to Paul Jenö und Thierry Mini for mass spectroscopy.

I am also grateful to Felix, Andi and Sebastian for valuable comments on the manuscript.

And, last but not least, special hug to Linus and Felix for their support and patience, and to my parents and Holger for “Linus-sitting” in times of need.

10 Bibliography

General Reading

- Protein Folding Handbook (2005), edited by Johannes Buchner and Thomas Kiefhaber, Wiley-VCH, Weinheim, Germany, ISBN: 3-527-30784-2.
- Mechanisms in Protein Folding (2000), edited by Roger Pain, Oxford University Press, UK, ISBN: 0-199-63788-1.

1. Stryer, L. (1995). *Biochemistry*. 4th edit, W. H. Freeman and Company, New York.
2. Greene, R. F. J. & Pace, C. N. (1974). Urea and guanidine-hydrochloride denaturation of ribonuclease, lysozyme, alpha-chyomtrypsinn and beta-lactoglobulin. *J. Biol. Chem.* **249**, 5388-5393.
3. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063-8068.
4. Shortle, D. (1995). Staphylococcal nuclease, a showcase of m-value effects. *Adv. Protein Chem.* **46**, 217-247.
5. Fink, A. L., Calciano, L. J., Goto, Y., Kurotsu, T. & Palleros, D. R. (1994). Classification of acid denaturation of proteins: intermediates and unfolded states. *Biochemistry* **33**, 12504-12511.
6. Liu, Z.-P., Rizo, J. & Gierasch, L. M. (1994). Equilibrium folding studies of cellular retinoic acid binding protein, a redominantly β-sheet protein. *Biochemistry* **33**, 134-142.

7. Shiraki, K., Nishikawa, K. & Goto, Y. (1995). Trifluoroethanol-induced stabilization of the α-helical structure of β-lactoglobulin: implications for non-hierarchical protein folding. *J. Mol. Biol.* **245**, 180-194.

8. Ohgushi, M. & Wada, A. (1983). 'Molten glogule state": A compact form of globular proteins with mobile side chains. *FEBS Lett.* **164**, 21-24.

9. Sánchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367-376.

10. Neri, D., Billeter, M., Wider, G. & Wüthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **257**, 1559-1563.

11. Goldenberg, D. P. (2003). Computational simulation of the statistical properties of unfolded proteins. *J Mol Biol* **326**, 1615-1633.

12. Gillespie, J. R. & Shortle, D. (1997). Characterization of long-range structure in the denatured state of Staphylococcal Nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268**, 158-169.

13. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138-2148.

14. Spiro, K. (1900). Über die Beeinflussung der Eiweisscoagulation durch stickstoffhaltige Substanzen. *Z Physiol Chem* **30**, 182 - 199.

15. Ramsden, W. (1902). Some new properties of urea. *J Physiol* **28**, 23 - 27.

16. Greenstein, J. P. (1938). Sulfhydryl groups in proteins. *J Biol Chem* **125**, 501-513.

17. Tanford, C. (1964). Isothermal unfolding of globular proteins in aqueous urea solutions. *J Am Chem Soc* **86**, 2050-2059.

18. Timasheff, S. & Xie, G. (2003). Preferential interactions of urea with lysozyme and their linkage to protein denaturation. *Biophys Chem* **105**.

19. Courtenay, E. S., Capp, M. W. & Record, M. T. J. (2001). Thermodynamics of interactions of urea and guanidinium salts with protein surface: relationship between solute effects on protein processes and changes in water-accessible surface area. *Protein Sci.* **10**, 2485-2497.

20. Schellman, J. A. (2003). Protein stability in mixed solvents: a balance of contact interactions and excluded volume effects. *Biophys. J.* **85**, 108-125.

21. Schellman, J. A. (1987). Selective binding and colvent? cosolvent? solvent? denaturation. *Biopolymers* **26**, 549-559.

22. Anfinsen, C. B. & Haber, E. (1961). Studies on the reduction and reformation of protein disulfide bonds. *J. Mol. Biol.* **236**, 1361-1363.

23. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

24. Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44.

25. Levinthal, C. (1969). *Mössbauer Spectroscopy in Biological Systems*, *Allerton House, Monticello, Ill.*

26. Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631-660.

27. Ptitsyn, O. B. (1973). [Stages in the mechanism of self-organization of protein molecules]. *Dokl Akad Nauk SSSR* **210**, 1213-1215.

28. Bashford, D., Cohen, F. E., Karplus, M., Kuntz, I. D. & Weaver, D. L. (1988). Diffusion-collision model for the folding kinetics of myoglobin. *Proteins* **4**, 211-27.

29. Karplus, M. & Weaver, D. L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci* **3**, 650-68.

30. Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA* **70**, 697-701.

31. Wetlaufer, D. B. (1990). Nucleation in protein folding--confusion of structure and process. *Trends Biochem Sci* **15**, 414-5.

32. Ptitsyn, O. B. (1995). How the molten globule became. *Trends Biochem Sci* **20**, 376-379.

33. Arai, M. & Kuwajima, K. (2000). Role of the molten globule state in protein folding. *Adv. Protein Chem.* **53**, 209-282.

34. Ansari, A., Berendzen, J., Browne, S., Frauenfelder, H., Iben, I., Sauke, T., Shyamsunder, E. & Young, R. (1985). Protein states and proteinquakes. *Proc. Natl. Acad. Sci. USA* **82**, 5000-5004.

35. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Struct. Funct. Genetics* **21**, 167-195.

36. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10-19.

37. Jackson, S. E. (1998). How do small single-domain proteins fold? *Folding & Design* **3**, R81-R91.

38. Perl, D., Welker, C., Schindler, T., Schröder, K., Marahiel, M. A., Jaenicke, R. & Schmid, F. X. (1998). Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic proteins. *Nat. Struct. Biol.* **5**, 229-235.

39. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428-10435.

40. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry* **30**, 10436-10443.

41. Bachmann, A. & Kiefhaber, T. (2001). Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* **306**, 375-386.

42. Schmid, F. X. & Baldwin, R. L. (1979). Detection of an early intermediate in the folding of ribonuclease A by protection of amide protons againts exchange. *J. Mol. Biol.* **135**, 199-215.

43. Wagner, G. A. & Wüthrich, K. (1982). Amide proton exchange and surface conformation of BPTI in solution. *J. Mol. Biol.* **160**, 343-361.

44. Walkenhorst, W. F., Edwards, J. A., Markley, J. L. & Roder, H. (2002). Early formation of a beta hairpin during folding of staphylococcal nuclease H124L as detected by pulsed hydrogen exchange. *Prot. Sci.* **11**, 82-91.

45. Blanco, F. J., Rivas, G. & Serrano, L. (1994). A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nat. Struct. Biol.* **1**, 584-590.

46. Dyson, H. J., Cross, K. J., Houghten, R. A., Wilson, I. A., Wright, P. E. & Lerner, R. A. (1985). The immunodominant site of a synthetic immunogen has a conformational preference in water for a type-II reverse turn. *Nature* **318**, 480-483.

47. Marqusee, S., Robbins, V. H. & Baldwin, R. L. (1989). Unusually stable helix formation in short alanine-based peptides. *Proc. Natl. Acad. Sci. USA* **86**, 5286-5290.

48. Goto, Y., Calciano, L. J. & Fink, A. L. (1990). Acid-induced folding of proteins. *Proc. Natl. Acad. Sci. USA* **87**, 573-577.

49. Goto, Y., Takahashi, N. & Fink, A. L. (1990). Mechanism of acid-induced folding of proteins. *Biochemistry* **29**, 3480-3488.

50. Kuwajima, K., Hiraoka, Y., Ikeguchui, M. & Sugai, S. (1985). Comparison of the transient folding intermediates in lysozyme and α-lactalbumin. *Biochemistry* **24**, 874-881.

51. Matthews, C. R. & Westmoreland, D. G. (1975). Nuclear Magnetic Resonance studies of residual structure in thermally unfolded ribonuclease A. *Biochemistry* **14**, 4532-4538.

52. Jeng, M. F. & Englander, S. W. (1991). Stable submolecular folding units in a non-compact form of cytochrome c. *J. Mol. Biol.* **221**, 1045-1061.

53. Tsong, T. Y., Baldwin, R. L. & McPhee, P. (1972). A sequential model of nucleation-dependent protein folding: kinetic studies of ribonuclease A. *J. Mol. Biol.* **63**, 453-457.

54. Garel, J. R. & Baldwin, R. L. (1973). Both the fast and slow folding reactions of ribonuclease A yield native enzyme. *Proc. Natl. Acad. Sci. USA* **70**, 3347-3351.

55. Kiefhaber, T. & Schmid, F. X. (1992). Kinetic coupling between protein folding and prolyl isomerization. II. Folding of ribonuclease A and ribonuclease T1. *J. Mol. Biol.* **224**, 231-240.

56. Kiefhaber, T., Kohler, H. H. & Schmid, F. X. (1992). Kinetic coupling between protein folding and prolyl isomerization. I. Theoretical models. *J. Mol. Biol.* **224**, 217-229.

57. Odefey, C., Mayr, L. & Schmid, F. X. (1995). Non-prolyl cis/trans peptide bond isomerization as a rate-determinig step in protein unfolding and refolding. *J. Mol. Biol.* **245**, 69-78.

58. Pappenberger, G., Aygün, H., Engels, J. W., Reimer, U., Fischer, G. & Kiefhaber, T. (2001). Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics. *Nat. Struct. Biol.* **8**, 452-458.

59. Creighton, T. E. (1986). Disulfide bonds as probes of protein folding pathways. *Meth. Enzymol.* **131**, 83-86.

60. Matthews, C. R. (1993). Pathways of protein folding. *Curr. Opin. Struct. Biol.* **1**, 17-21.

61. Lang, K., Schmid, F. X. & Fischer, G. (1987). Catalysis of protein folding by prolyl isomerase. *Nature* **329**, 268-270.

62. Fischer, G., Wittmann, L. B., Lang, K., Kiefhaber, T. & Schmid, F. X. (1989). Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins. *Nature* **337**, 476-478.

63. Fischer, G. & Schmid, F. X. (1990). The mechanism of protein folding. Implications of in vitro refolding models for de novo protein folding and translocation in the cell. *Biochemistry* **29**, 2205-2212.

64. Fischer, G. & Bang, H. (1985). The refolding of urea-denatured ribonuclease A is catalyzed by peptidyl-prolyl cis-trans isomerase. *Biochim. Biophys. Acta* **828**, 39-42.

65. Gething, M. J. & Sambrook, J. (1992). Protein folding in the cell. *Nature* **335**, 33-45.

66. Freedman, R. B., Brockway, B. E. & Lambert, N. (1984). Protein disulphide-isomerase and the formation of native disulphide bonds. *Biochem Soc Trans* **12**, 929-32.

67. Horwich, A. L., Neupert, W. & Hartl, F. U. (1990). Protein-catalyzed protein folding. *Trends Biotechnol.* **8**, 126-131.

68. Frydman, J. & Hartl, F. U. (1996). Principle of chaperone-assisted protein folding: differences between in vitro and in vivo mechanisms. *Science* **272**, 1497-1502.

69. Ellis, R. J. (1990). The molecular chaperone concept. *Semin. cell. biol.* **1**, 1-9.

70. Dill, K. A. (1985). Theory for folding and stability of globular proteins. *Biochemistry* **24**, 1501-1509.

71. Jaenicke, R. (1987). Folding and association of proteins. *Progr. Biophys. Mol. Biol.* **49**, 117-237.

72. Neet, K. E. & Timm, D. E. (1994). Conformational stability of dimeric proteins: quantitative studies by equilibrium denaturation. *Prot. Sci.* **3**, 2167.

73. Fersht, A. R. (1997). Nucleation mechanisms in protein folding. *Curr Opin Struct Biol* **7**, 3-9.

74. Seckler, R. (2000). Assembly of multi-subunit structures. In *Mechanisms of protein folding* (Pain, R., ed.), pp. 279-308. Oxford University Press, Oxford.

75. Milla, M. E. & Sauer, R. T. (1994). P22 arc repressor: Folding kinetics of a single-domain, dimeric protein. *Biochemistry* **33**, 1125-1133.

76. Wendt, H., Berger, C., Baici, A., Thomas, R. M. & Bosshard, H. R. (1995). Kinetics of folding of leucine zipper domains. *Biochemistry* **34**, 4097-4107.

77. Gloss, L. M. & Matthews, C. R. (1998). Mechanism of folding of the dimeric core domain of Escherichia coli Trp repressor: a nearly diffusion-limited reaction leads to the formation of an on-pathway intermediate. *Biochemistry* **37**, 15990-15999.

78. Waldburger, C. D., Jonsson, T. & Sauer, R. T. (1996). Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl. Acad. Sci. USA* **93**, 2629-2634.

79. Chrunyk, B. A. & Matthews, C. R. (1990). Role of diffusion in the folding of the alpha-subunit of tryptophan synthase from Escherichia coli. *Biochemistry* **29**, 2149-2154.

80. Parker, M. J., Spencer, J., Jackson, G. S., Burston, S. G., Hosszu, L. L., Craven, C. J., Waltho, J. P. & Clarke, A. R. (1996). Domain behavior during the folding of a thermostable phosphoglycerate kinase. *Biochemistry* **35**, 15740-15752.

81. Martin, A. & Schmid, F. X. (2003). The folding mechanism of a two-domain protein: folding kinetics and domain docking of the gene-3 protein of phage fd. *J. Mol. Biol.* **329**, 599-610.

82. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering methods: evidence for a nucleation-condesation mechanism for protein folding. *J. Mol. Biol.* **254**, 260-288.

83. Schmid, F. X. (1983). Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction via structural intermediates. *Biochemistry* **22**, 4690-4696.

84. Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. *Biochemistry* **14**, 4953-4963.

85. Jaenicke, R. (1995). Folding and association versus misfolding and aggregation of proteins. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **348**, 97-105.

86. Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997). Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure* **5**, 789-798.

87. Boudko, S. P., Strelkov, S. V., Engel, J. & Stetefeld, J. (2004). Design and crystal structure of Bacteriophage T4 mini-fibritin NCCF. *J. Mol. Biol.* **339**, 927-935.

88. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51-55.

89. Letarov, A. V., Londer, Y., Y., Boudko, S. P. & Mesyanzhinov, V. V. (1999). The Carboxy-terminal domain initiates trimerization of bacteriophage T4 fibritin. *Biochemistry (Moscow)* **64**, 817-823.

90. Panchenko, A. R., Luthey-Schulten, Z., Cole, R. & Wolynes, P. G. (1997). The foldon universe: a survey similarity and self-recognition of independently folding units. *J. Mol. Biol.* **227**.

91. Panchenko, A. R., Luthey-Schulten, Z. & Wolynes, P. G. (1997). Foldons as independently folding units of proteins. *Physica D* **107**, 312-315.

92. Frank, S., Kammerer, R. A., Mechling, D., Schulthess, T., Landwehr, R., J., B., Guo, Y., Lustig, A., Bächinger, H. P. & J., E. (2001). Stabilization of short collagen-like triple helices by protein engineering. *J. Mol. Biol.* **308**, 1081-1089.

93. Yang, X. Z., Lee, J., Mahony, E. M., Kwong, P. D., Wyatt, R. & Sodroski, J. (2002). Highly stable trimers formed by human immunodeficiency virus type 1 envelope glycoproteins fused with the trimeric motif of T4 bacteriophage fibritin. *Journal of Virology* **76**, 4634-4642.

94. Nicola, A. V., Chen, W. & Helenius, A. (1999). Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat. Cell. Biol.* **1**, 341-345.

95. Kolb, V. A., Makeyev, E. V. & Spirin, A. S. (2000). Co-translational folding of an eukaryotic multidomain protein in a prokaryotic translation system. *J. Biol. Chem.* **275**, 16597-16601.

96. Choi, H. K., Lu, G., Lee, S., Wengler, G. & Rossmann, M. G. (1997). Structure of Semliki Forest virus core protein. *Proteins: Struct. Funct. Genetics* **27**, 345-359.

97. Kraulis, P. (1991). MolScript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.

98. Boege, U., Wengler, G. & Wittmann-Liebold, B. (1981). Primary structures of the core proteins of the alphaviruses Semliki Forest virus and Sindbis virus. *Virology* **113**, 293-303.

99. Strauss, J. H. & Strauss, E. G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* **58**, 491-562.

100. Kowarik, M., Küng, S., Martoglio, B. & Helenius, A. (2002). Protein folding during cotranslational translocation in the endoplasmic reticulum. *Mol. Cell* **10**, 769-778.

101. Braakmann, I., Hoover-Litty, H., Wagner, K. R. & Helenius, A. (1991). Folding of influenza hemagglutinin in the endoplasmic reticulum. *J. Cell. Biol.* **114**, 401-411.

102. Sánchez, I. E., Morillas, M., Zobeley, E., Kiefhaber, T. & Glockshuber, R. (2004). Fast folding of the two-domain Semliki Forest virus capsin protein explains co-translational proteolytic activity. *J. Mol. Biol.* **338**, 159-167.

103. Rudolph, R., Siebendritt, R., Nesslauer, G., Sharma, A. K. & Jaenicke, R. (1990). Folding of an all-beta protein: independent domain folding in gamma II-crystallin from calf eye lens. *Proc. Natl. Acad. Sci. USA* **87**, 4625-4629.

104. Jaenicke, R. (1999). Stability and folding of domain proteins. *Progr. Biophys. Mol. Biol.* **71**, 155-241.

105. Martin, A. & Schmid, F. X. (2003). The Folding Mechanism of a Two-domain Protein: Folding Kinetics and Domain Docking of the Gene-3 Protein of Phage fd. *J. Mol. Biol.* **329**, 599-610.

106. Güthe, S., Kapinos, L., Möglich, A., Meier, S., Grzesiek, S. & Kiefhaber, T. (2004). Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. *J Mol Biol* **337**, 905-15.

107. Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. E. & Matthews, C. R. (1995). Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry* **34**, 12812-12819.

108. Meier, S., Güthe, S., Kiefhaber, T. & Grzesiek, S. (2004). Foldon, the natural trimerization domain of T4 fibritin, dissociates into a monomeric A-state form containing a stable beta-hairpin: atomic details of trimer dissociation and local beta-hairpin stability from residual dipolar couplings. *J Mol Biol* **344**, 1051-69.

109. Krause, S. (2001). Folding and Stability of Foldon, the trimeric C-terminal Domain from Bacteriophage T4 Fibritin. Diploma Thesis, Biozentrum der Universität Basel.

110. Henry, E. R. & Hofrichter, J. (1992). Singular value decomposition: application to analysis of experimental data. *Meth. Enzymol.* **210**, 129-192.

111. Chen, L., Hodgson, K. O. & Doniach, S. (1996). A lysozyme folding intermediate revealed by solution X-ray scattering. *J. Mol. Biol.* **261**, 658-671.

112. Manuel Morillas, Eva Zobeley & Rudi Glockshuber. ETH Zürich.

113. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* **31**, 3381-5.

114. Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948-1954.

115. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of Biomolecular Nmr* **6**, 277-293.

116. Garrett, D. S., Gronenborn, A. M. & Clore, G. M. (1995). Automated and Interactive Tools for Assigning 3d and 4d Nmr-Spectra of Proteins - Capp, Stapp and Pipp. *Journal of Cellular Biochemistry*, 71-71.

117. Aune, K. C. & Tanford, C. (1969). Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. II Dependence on denaturant concentration at 25 degrees. *Biochemistry* **11**, 4586-4590.

118. Lakowicz, J. R. (1999). *Principles of Fluorescence Spectroscopy*. 2 edit, Plenum Press, New York.