doi:10.1016/j.jmb.2006.11.014 J. Mol. Biol. (2007) 366, 650–665

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

A Comparison of the Folding of Two Knotted Proteins: YbeA and YibK

Anna L. Mallam and Sophie E. Jackson*

University Chemical Laboratory Lensfield Road, Cambridge CB2 1EW, UK

**Corresponding author*

The extraordinary topology of proteins belonging to the α/β-knot superfamily of proteins is unexpected, due to the apparent complexities involved in the formation of a deep trefoil knot in a polypeptide backbone. Despite this, an increasing number of knotted structures are being identified; how such proteins fold remains a mystery. Studies on the dimeric protein YibK from *Haemophilus influenzae* have led to the characterisation of its folding pathway in some detail. To complement research into the folding of YibK, and to address whether folding pathways are conserved for members of the α/β-knot superfamily, the structurally similar knotted protein YbeA from *Escherichia coli* has been studied. A comprehensive thermodynamic and kinetic analysis of the folding of YbeA is presented here, and compared to that of YibK. Both fold *via* an intermediate state populated under equilibrium conditions that is monomeric and considerably structured. The unfolding/refolding kinetics of YbeA are simpler than those found for YibK and involve two phases attributed to the formation of a monomeric intermediate state and a dimerisation step. In contrast to YibK, a change in the rate-determining step on the unfolding pathway for YbeA is observed with a changing concentration of urea. Despite this difference, both proteins fold by a mechanism involving at least one sequential monomeric intermediate that has properties similar to that observed during the equilibrium unfolding. The rate of dimerisation observed for YbeA and YibK is very similar, as is the rate constant for formation of the kinetic monomeric intermediate that precedes dimerisation. The findings suggest that relatively slow folding and dimerisation may be common attributes of knotted proteins.

Keywords: topological knot; protein folding; chevron plot; global analysis; dimer kinetics

Introduction

Current theories on folding mechanisms suggest that proteins can undergo a variety of conformational changes during the folding process.[1,2] However, that a polypeptide chain could "knot" itself to form a functional protein was thought highly improbable, if not impossible. Nevertheless, a recently identified group of proteins have revealed a somewhat unexpected topological twist; contrary to all previous protein-folding models, they do indeed possess a deep knot in their structure formed by the polypeptide backbone.[3–6] Elucidation of the folding mechanism of these proteins represents an important new challenge in the protein-folding field.

A growing number of proteins containing topological knots have been identified over the last five years. Most display a deep trefoil knot in their structure and, to date, over 15 protein structures with deep trefoil knots have been deposited in the RCSB Protein Data Bank.[3,7–15] All are thought to function as methyltransferases (MTases), a type of enzyme involved in the transfer of the methyl group of *S*-adenosyl methionine (AdoMet) to carbon, nitrogen or oxygen atoms of DNA, RNA, proteins and other small molecules,[16] and all form dimers in the crystalline form, with the knotted region forming a large part of the dimer interface. Another

Abbreviations used: FL, fluorescence; MTase, methyltransferase; AdoMet, *S*-adenosyl methionine; SEC, size-exclusion chromatography; *SASA*, solvent-accessible surface area.

E-mail address of the corresponding author: sej13@cam.ac.uk

part of the knotted structure is thought to be the location of the AdoMet binding site.[3,7–15] In recognition of the above similarities, MTases with knotted structures have been combined into one family, known as the α/β-knot superfamily.[7,17] While the majority of recently discovered knotted proteins belong to this superfamily, the occurrence of knots is not restricted to members the α/β-knot clan. A deep trefoil knot with different features has been observed in the chromophore-binding domain of *Deinococcus radiodurans* phytochrome,[6] and an impressive deep figure-of-eight knot has been identified in the plant protein acetohydroxy acid isomeroreductase.[4] The most complicated knot discovered to date in human ubiquitin hydrolase displays five projected crossings.[18]

The first folding studies on the α/β-knot family of knotted proteins were carried out on the protein YibK from *Haemophilus influenzae*.[19,20] Despite its complicated knotted topology, the unfolding of YibK was found to be fully reversible *in vitro*, molecular chaperones not being required for efficient folding. YibK was shown to undergo equilibrium denaturation in a manner similar to that of many unknotted dimers, and *via* a monomeric equilibrium intermediate with considerable structure and stability.[19] Furthermore, a complex kinetic mechanism with four reversible kinetic folding phases was observed. The behaviour of these phases at different pH values allowed a folding mechanism to be proposed. Two different intermediates from parallel pathways were seen to fold *via* a third sequential, monomeric intermediate that formed native dimer in a slow, rate-limiting dimerisation step. All the intermediates were structurally distinct and on-pathway, and the parallel channels were shown to arise from heterogeneity in the denatured state, most likely caused by proline isomerisation.[20]

Whether protein folding pathways are conserved for a given fold is an increasingly important topic.[21,22] Studies on a number of small, two-state, monomeric proteins found a correlation between folding rates and topology that led to the hypothesis that topology, not sequence, is the major contributing factor in how a protein folds.[23,24] If protein folding rates and mechanisms are determined largely by the topology of the native state, then the complexity of the protein-folding problem would be greatly simplified and it would no longer be necessary to determine the folding pathway of every single protein of interest.[25] However, studies performed on proteins that possess similar tertiary structures but divergent sequences give conflicting evidence for the conservation of folding pathways of

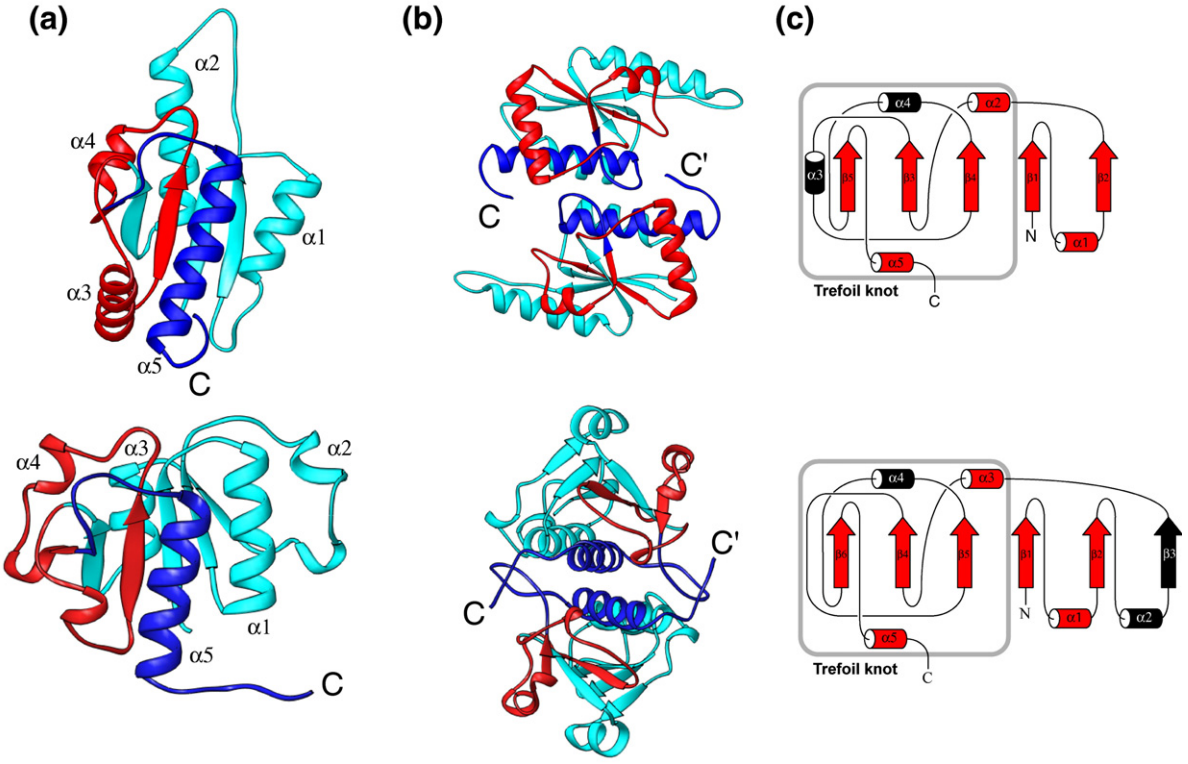

Figure 1. Structure of YbeA from *Escherichia coli* (top) and YibK from *Haemophilus influenzae* (bottom). Both proteins contain a topological trefoil knot formed by the polypeptide backbone; a substantial length of polypeptide chain (approximately 40 residues) has threaded through a loop during folding. (a) Ribbon diagram of a monomer subunit, showing the deep trefoil knot at the C terminus. Structures are coloured according to definitions given by Nureki *et al.*,[12] with the knotting loop highlighted in red and the knotted chain highlighted in dark blue. (b) Dimeric structures coloured as in (a). YibK is a parallel homodimer, while YbeA dimerises in an antiparallel fashion. Ribbon diagrams were generated using Ribbons.[48] (c) Topological diagrams of YbeA (top) and YibK (bottom). Structural elements common to members of the α/β-knot superfamily are shown in red.

proteins in the same structural class.[21,22] This, along with the widespread observation of highly variable folding rates for proteins of the same structure,[26] suggests that topology alone is not enough to predict folding rates and mechanism.

In the context of the α/β-knot superfamily, it is important to consider whether proteins that contain a deep topological knot share a common folding mechanism. Here, we focus on YbeA from *Escherichia coli*, a 155-residue protein similar in structure to YibK (Figure 1). A number of YbeA-like structures all possessing deep knots have been deposited in the Brookhaven Protein Data Bank, and all belong to the α/β-knot superfamily of MTases. Many are the result of structural genomics studies, including sr145 from *Bacillus subtilis* (PDB code 1TO0), and the hypothetical proteins Tm0844 (1OD6) and Sav0024 (1VH0). YbeA has a deep trefoil knot in its backbone structure formed by the threading of the last 35 residues (residues 120–155) through a 45-residue knotting loop (residues 74–119) (Figure 1(a)). YbeA crystallises as an antiparallel dimer, and the protein interface involves close-packing of α1 and α5 from each monomer (Figure 1(b)). Its topological features are very similar to those of YibK, and YbeA displays the structural elements characteristic of all α/β-knot MTases (Figure 1(c)). However, YibK and YbeA share only 19% sequence identity.

In this study, the thermodynamic and kinetic folding properties of YbeA are characterised, and a folding mechanism is proposed. Comparisons are made to the folding of YibK, and the similarities and differences between the two proteins are discussed.

Results

To enable a direct comparison of the folding parameters of the knotted proteins YbeA and YibK, experiments on YbeA were carried out under the same conditions as those used for YibK: specifically, at pH 7.5 in a buffer containing 200 mM KCl and 10% (v/v) glycerol. These stabilising agents were needed to prevent aggregation during studies of YibK at pH 7.5.[19] YbeA, however, remains soluble in buffer alone at all experimental concentrations of protein studied, allowing additional studies to be performed under these conditions. Unless stated otherwise, all experiments were done in buffer with stabilising agents (50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% (v/v) glycerol, 1 mM DTT), and in buffer without stabilising agents (50 mM Tris–HCl (pH 7.5), 1 mM DTT).

The oligomeric state of YbeA

Although YbeA crystallises as a homodimer (Figure 1(b)), its oligomeric state under solution conditions may be different.[27] Size-exclusion chromatography (SEC) was used to investigate the oligomeric state of YbeA for concentrations of protein between 5 μM and 40 μM, and the results are shown in Figure 2 for buffer with stabilising agents. The protein interacted with the gel-filtration column if salt was not present; therefore, experiments using buffer without salt were not possible. YbeA eluted at a volume of 10.6 ml, which was independent of protein concentration, corresponding to a molecular mass of 36.8 kDa (see the insets in Figure 2 for a calibration plot). This agrees well with the expected molecular mass of 34.6 kDa for homodimeric YbeA.

Degree and reversibility of unfolding of YbeA

The chemical denaturant urea was used to induce unfolding of YbeA, and changes in both intrinsic protein fluorescence and far-UV CD signal were monitored. YbeA contains five tryptophan and two tyrosine residues distributed throughout its structure, and addition of urea to a final concentration of 8 M resulted in an overall fluorescence increase along with a simultaneous red-shift in λ_{max} from 328 nm to 350 nm, indicative of an unfolding event (Figure 3, inset). The maximum fluorescence change was observed at 350 nm, and unfolding was monitored at this wavelength during subsequent experiments. Far-UV CD spectra under the same unfolding conditions showed a complete loss of secondary structure (Figure 3, inset), suggesting that addition of urea causes a global, and not just a local, unfolding event. Native fluorescence and far-UV CD fingerprints were independent of buffer conditions (Figure 3).

The reversibility of the YbeA unfolding reaction was investigated using probes of secondary and tertiary structure. Refolded YbeA retains approximately 100% of native fluorescence and far-UV CD signal (Figure 3, inset), suggesting that the folding of

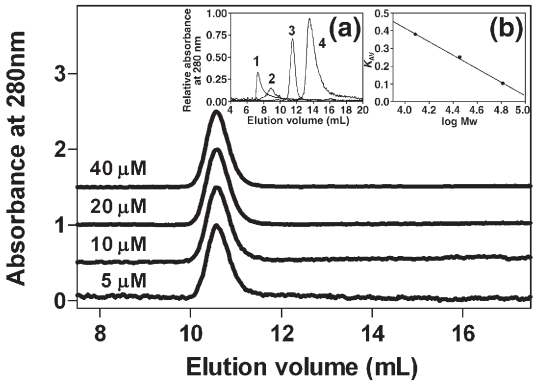

Figure 2. Determination of the oligomeric state of YbeA by size-exclusion chromatography. Main: Elution profiles for 40 μM, 20 μM, 10 μM and 5 μM protein. Absorbance has been normalised against protein concentration. An upper limit of 15 nM for the dissociation constant of the YbeA dimer can be estimated from these data. Insets: (a) Elution profile of (1) blue dextran 200 (2000 kDa), (2) bovine serum albumin (66.3 kDa), (3) carbonic anhydrase (28.8 kDa) and (4) cytochrome *c* (12.4 kDa). (b) Calibration curve. Conditions: room temperature in 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% (v/v) glycerol, 1 mM DTT.

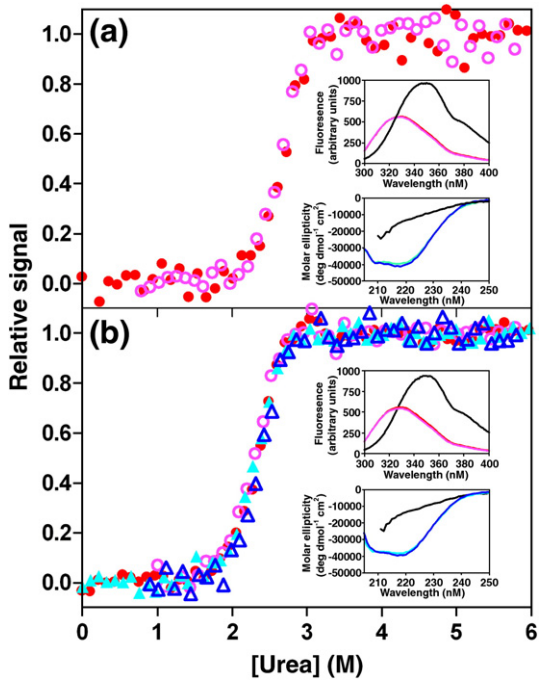

Figure 3. Reversibility of the YbeA folding reaction in (a) buffer with salt and glycerol and (b) buffer only. Main: YbeA fluorescence (circles) and far-UV CD (triangles) denaturation (filled symbols) and renaturation (open symbols) profiles at 1 μM protein. Insets: fluorescence and far-UV CD spectra of native (red and dark blue continuous lines, respectively), refolded from 8 M urea (pink and light blue continuous lines, respectively) and denatured (continuous black lines) YbeA at 2 μM protein. Conditions for (a): 25 °C, 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% (v/v) glycerol, 1 mM DTT. Conditions for (b): 25 °C, 50 mM Tris–HCl (pH 7.5), 1 mM DTT.

YbeA secondary and tertiary structure is fully reversible under both buffer conditions used. Additionally, unfolding and refolding equilibrium titrations at the same concentration of YbeA measured using fluorescence and far-UV CD superimpose, confirming the reversibility of the unfolding reaction (Figure 3).

Equilibrium unfolding experiments on YbeA

Equilibrium denaturation experiments were performed on YbeA under buffer conditions with and without stabilising agents at pH 7.5 over an 80-fold and 20-fold change in protein concentration, respectively. Profiles measured using intrinsic protein fluorescence and far-UV CD are shown in Figure 4 (far-UV CD was used to measure denaturation data only for buffer without stabilising agents). Both probes show a single unfolding transition that is protein concentration-dependent, with a midpoint that increases with increasing concentration of YbeA. This dependence on protein concentration- was used to assign a dimer equilibrium-unfolding model.

Data were first fit to the simplest dimer denaturation model involving only native dimer and unfolded monomers (equation (1)).[19] Datasets for each concentration of protein were treated separately, and the results of the fit are shown in Figure 4(a) and summarised in Table 1. Figure 4(a) shows that the data appear to be described well by a two-state dimer denaturation model, however, Table 1 illustrates that $\Delta G^{N_2\leftrightarrow 2D}_{H_2O}$ and $m_{N_2\leftrightarrow 2D}$ values from this fit for both fluorescence and far-UV CD denaturation data show a general increase with increasing protein concentration. This suggests that the two-state dimer denaturation model is not adequately describing the YbeA equilibrium unfolding data, as $\Delta G^{N_2\leftrightarrow 2D}_{H_2O}$ and $m_{N_2\leftrightarrow 2D}$ values should remain constant with protein concentration.[19,28] Furthermore, the variation in *m*-value with protein concentration indicates that an intermediate state is populated under equilibrium conditions.[29]

An increase in apparent $\Delta G^{N_2\leftrightarrow 2D}_{H_2O}$ and $m_{N_2\leftrightarrow 2D}$ value of the denaturation profiles with protein concentration is consistent with a dimer unfolding *via* a three-state denaturation model involving a monomeric intermediate; in the simplest case, the increase in *m*-value with protein concentration can be explained by dissociation of the dimer in the transition region at low concentrations of protein.[19,28] Accordingly, YbeA equilibrium data were globally fit to this model (equation (2)); fluorescence and far-UV CD datasets were treated separately, as were data for each buffer condition. The results of these fits are shown in Figure 4(b) and summarised in Table 2. Considering the parameters from fluorescence denaturation data, the *m*-values calculated under different buffer conditions are in excellent agreement (Table 2). However, a noticeable difference in the stability of the dimer can be seen. The free energy difference between the dimer and the monomeric intermediate, $\Delta G^{N_2\leftrightarrow 2I}_{H_2O}$, is calculated to be 15.2 kcal mol^{-1}and 13.3 kcal mol^{-1}, for experiments performed in buffer with and without stabilising agents, respectively. In comparison, the stability of the monomeric intermediate is similar for both buffers, and is in the range of 2.5–2.8 kcal mol^{-1}. Parameters calculated from fluorescence and far-UV CD under the same buffer conditions are in good agreement (Figure 4(b), right, and (c); Table 2).

Estimation of *m*-value from changes in solvent-accessible surface area

The *m*-value of a protein is related to the change in solvent-accessible surface area ($\Delta SASA$) that occurs upon unfolding.[30] The $\Delta SASA$ expected for dissociation and unfolding of a YbeA dimer were calculated in order to estimate the *m*-values corresponding to these changes. The estimated *m*-value for the dissociation of dimer to fully folded monomer is 0.38 kcal mol^{-1} M^{-1}, that for the complete unfolding of a fully folded monomer 1.8–2.3 kcal mol^{-1} M^{-1}, and for the full unfolding of the dimer is 3.8–4.9 kcal mol^{-1} M^{-1} (Table 3). The latter is in good agreement with the $m_{N_2\leftrightarrow 2D}$ value of 4.4–4.6 kcal mol^{-1} M^{-1} calculated for YbeA from the equilibrium denaturation data (Table 2). The $m_{I\leftrightarrow D}$

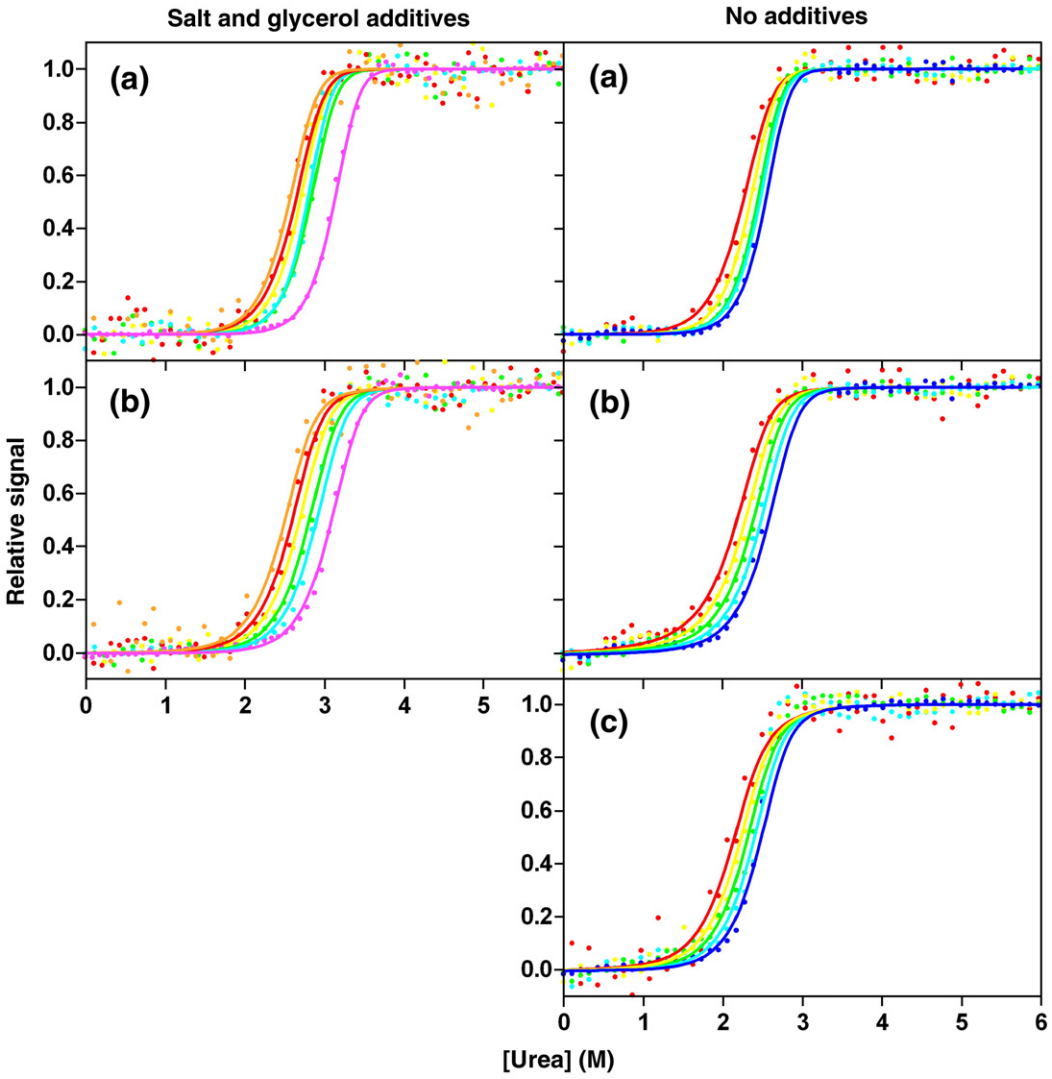

Figure 4. YbeA equilibrium denaturation profiles for 0.25 μM (orange), 0.5 μM (red), 1 μM (yellow), 2.5 μM (green), 5 μM (light-blue), 10 μM (dark-blue) and 20 μM (pink) protein, as measured by (a) and (b) fluorescence emission at 350 nm, and (c) far-UV CD signal at 220 nm. The continuous lines in (a) represent the best fit to a two-state dimer denaturation model, while those in (b) and (c) represent the global fit to a three-state dimer denaturation model with a monomeric intermediate. Conditions for buffer with and without salt and glycerol were as described for Figure 3.

value of 1.38–1.52 kcal mol^{-1} M^{-1} calculated from the equilibrium unfolding data is slightly smaller than the estimated *m*-value for a monomeric subunit, suggesting that the monomeric intermediate has lost some structure relative to the fully folded monomer in the dimer.

Table 1. Thermodynamic parameters for the fit of YbeA equilibrium unfolding data to a two-state dimer denaturation model

	Buffer only			Buffer salt and glycerol		
P_t (μM)	$[D]_{50\%}$ (M)	$m_{N_2 \leftrightarrow 2D}$ (kcal mol^{-1} M^{-1})	$\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ (kcal mol^{-1})	$[D]_{50\%}$ (M)	$m_{N_2 \leftrightarrow 2D}$ (kcal mol^{-1} M^{-1})	$\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ (kcal mol^{-1})
0.25	-	-	-	2.56±0.09	4.67±0.09	21.0±0.8
0.5	2.25±0.02	4.34±0.05	18.4±0.4	2.64±0.02	4.55±0.4	20.6±0.5
1	2.34±0.01	5.03±0.01	20.0±0.1	2.69±0.01	4.94±0.01	21.5±0.1
2	2.43±0.01	5.35±0.01	20.8±0.1	2.83±0.01	5.03±0.04	22.0±0.3
5	2.46±0.01	5.28±0.01	20.2±0.1	2.79±0.01	5.51±0.01	22.6±0.1
10	2.53±0.01	5.40±0.1	20.5±0.4	-	-	-
20	-	-	-	3.12±0.01	5.28±0.07	22.9±0.5

Fluorescence was monitored at 350 nm. Parameters are quoted with their standard errors. Urea denaturation profiles were fit singularly to a two-state dimer denaturation model[19] using Prism, version 4, and $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ was calculated from $\Delta G_{H_2O} = -RT \ln(P_t) + m[D]_{50\%}$. Profiles measured using far-UV CD at 220 nm showed similar parameters (data not shown).

Table 2. Thermodynamic parameters for the fit of YbeA and YibK equilibrium unfolding data to a three-state dimer denaturation model with a monomeric intermediate

Protein	Buffer additives	Probe	Y_I[a]	$\Delta G^{N_2\text{--}2I}_{H_2O}$ (kcal mol^{-1})	$m_{N_2\text{--}2I}$ (kcal mol^{-1} M^{-1})	$\Delta G^{I\text{--}D}_{H_2O}$ (kcal mol^{-1})	$m_{I\text{--}D}$ (kcal mol^{-1} M^{-1})	$\Delta G^{N_2\text{--}2D}_{H_2O}$[b] (kcal mol^{-1})	$m_{N_2\text{--}2D}$[c] (kcal mol^{-1} M^{-1})
YbeA	Salt, glycerol	FL	0.32±0.1	15.2±0.1	1.65±0.01	2.5±0.01	1.38±0.01	20.2±0.1	4.4±0.02
YbeA	None	FL	0.75±0.1	13.3±0.01	1.56±0.01	2.8±0.01	1.50±0.02	18.9±0.02	4.6±0.03
YbeA	None	Far-UV CD	0.21±0.1	12.7±0.01	1.57±0.01	2.8±0.01	1.52±0.01	18.3±0.02	4.6±0.02
YibK[d]	Salt, glycerol	FL	0.61±0.1	18.9±0.4	1.80±0.09	6.5±0.2	1.53±0.05	31.9±1.2	4.9±0.3

Fluorescence (FL) and far-UV CD datasets were analysed separately using equation (2). Global analysis was performed with the non-linear least-squares fitting program Prism, version 4. Errors quoted are the standard errors calculated by the fitting program and the small errors quoted for some parameters are a reflection of the global analysis, rather than a true experimental error.

[a] Y_I is the experimentally determined spectral signal for the intermediate, relative to a signal of 0 for a native monomeric subunit in a dimer and 1 for a denatured monomer. Despite the low fitting errors associated with Y_I, this value is the most changeable during fitting; for example, when a different number of datasets are included in the global analysis, and so some variation is not unexpected. Fluorescence can be extremely sensitive to experimental conditions, such as pH, [urea], buffer additives and temperature.

[b] $\Delta G^{N_2\text{--}2D}_{H_2O} = \Delta G^{N_2\text{--}2I}_{H_2O} + 2\Delta G^{I\text{--}D}_{H_2O}$.

[c] $m_{N_2\text{--}2D} = m_{N_2\text{--}2I} + 2m_{I\text{--}D}$.

[d] Data for YibK are taken from Mallam & Jackson.[19]

Kinetic folding experiments on YbeA

Kinetic folding studies were performed on YbeA under the buffer conditions used for the thermodynamic studies. Stopped-flow mixing techniques were employed to measure folding/unfolding rate constants, and folding/unfolding was monitored using intrinsic protein fluorescence. Typical traces for YbeA single-jump unfolding and refolding at various final concentrations of urea and a final protein concentration of 1 μM are shown in Figure 5. Results were similar for experiments performed in buffer with and without stabilising agents. Unfolding traces measured under strongly unfolding conditions were best described by a first-order reaction with two exponentials (Figure 5(a)), while unfolding profiles at lower concentrations of urea were best fit to a first-order reaction with a single exponential (Figure 5(b)). Refolding traces were best described by a first-order reaction with two exponentials (Figure 5(c)). There was no observable burst phase, as all amplitude change was accounted for by the kinetic traces (Figure 6(c)), and all rate constants appeared to be independent of protein concentration (Figure 5(c), left, inset).

The urea concentration-dependence of the unfolding and refolding of YbeA was examined. A global analysis of all folding and unfolding kinetic traces was undertaken using equation (3) to obtain unfolding and refolding rate constants in the absence of denaturant and unfolding and refolding *m*-values for any observed phases. Traces within each set of buffer conditions were analysed together, and the results of the fit to equation (3) are shown in Figure 7 and summarised in Table 4. A chevron plot calculated from the parameters in Table 4 is shown in Figure 6(a). Analysis of the kinetic transients in a global fashion allowed data measured at urea concentrations where the two observed rate constants are different enough to be dynamically uncoupled to be used to define the chevron plot at concentrations of urea where the two phases are too close together for accurate rate constants to be extracted from analysis of separate folding transients. The good global fit of the kinetic data to equation (3) indicates that YbeA has two reversible folding phases, denoted 1 and 2 in Table 4, and coloured red and blue, respectively. Figure 6(a) shows that the unfolding arms of the two phases cross at 5.4 M urea and 3.4 M urea for buffer with and without stabilising agents, respectively. The amplitudes corresponding to the rate constants obtained from the global fits are shown in Figure 6(b). The magnitude of the amplitude of the red unfolding phase decreases with decreasing concentrations of urea until it reaches zero at approximately 5.4 M urea and 3.4 M urea for buffer with and without stabilising agents, respectively. This reflects the changing nature of the YbeA unfolding traces from double to single-exponential (Figure 5(b)). These observations are consistent with a change in rate determining step in the YbeA unfolding reaction at concentrations of urea where

Table 3. Changes in SASA for YbeA upon dimer dissociation and unfolding, along with estimated *m*-values

A. *Dissociation*

Native dimer (N_2) *SASA*[a] (Å^2)	Fully folded monomer subunit (N) *SASA*[a] (Å^2)	Δ*SASA* for dissociation $N_2 \leftrightarrow 2N$[b] (Å^2)	*m*-value estimate for dissociation $N_2 \leftrightarrow 2N$[c] (kcal mol^{-1} M^{-1})
15215	8962	2709	0.38±0.03

B. *Unfolding*

	SASA for folded protein[a] (Å^2)	*SASA* estimate for unfolded protein (Å^2)		Δ*SASA* for unfolding[d] (Å^2)		*m*-value estimate for unfolding[c] (kcal mol^{-1} M^{-1})	
		Tripeptide method[e]	Upper boundary method[f]	Tripeptide method	Upper boundary method	Tripeptide method	Upper boundary method
Native dimer (N_2)	15215	50082	42606	34867	27391	4.9±0.3	3.8±0.3
Fully folded monomer (N)	8962	25041	21303	16079	12341	2.3±0.2	1.8±0.1

[a] Calculated using the web-based program GETAREA version 1.1.[44]
[b] Δ*SASA* upon dissociation = 2 [*SASA* of monomer subunit]–[*SASA* of dimer]. The value calculated for the Δ*SASA* upon dimer dissociation assumes no unfolding of the monomer subunits.
[c] Estimated using equation (5).[3]
[d] Δ*SASA* for unfolding = [*SASA* unfolded protein]–[*SASA* folded protein].
[e] Calculated using values from tripeptide studies.[45]
[f] Calculated using data given by Creamer *et al*.[47]

the unfolding arms of the phases cross on the chevron plots.

The parameters calculated from the global analysis of the kinetic data for buffers with and without stabilising agents can be compared, and those corresponding to phase 1 (red) are in good agreement (Table 4). Parameters calculated for phase 2 (blue) also compared well, except for the unfolding rate constant in the absence of denaturant, which is notably larger under conditions without salt and glycerol (Table 4).

Interrupted-refolding experiments on YbeA

Interrupted-refolding experiments involve refolding a protein for various amounts of time before unfolding is initiated. The amplitudes of the unfolding reactions are directly proportional to the population of refolding species present after the delay, allowing the time-course of refolding intermediates to be monitored.[31] The method assumes that any intermediates formed during refolding will unfold faster than the native protein.[32,33] Interrupted refolding was performed on YbeA at pH 7.5 in buffer containing salt and glycerol, and the following analysis refers only to this buffer. Two sets of experiments were undertaken involving refolding to 1 M urea, and subsequent unfolding to either 4.3 M urea or 7.7 M urea. These final concentrations of urea were chosen to probe the change in rate-determining step observed on the unfolding pathway during single-jump experiments, and allowed investigation of how the same population of refolding intermediates behaved under different unfolding conditions. YbeA single-jump unfolding experiments at 4.3 M urea and 7.7 M urea were best described by a first-order reaction with one and two exponentials, respectively, and the chevron plots show that the blue and red phases are rate limiting, respectively, at these concentrations of urea (Figure 6(a)). Traces from interrupted-refolding experiments to final concentrations of 4.3 M urea and 7.7 M urea were best fit to a first-order reaction with two exponentials (data not shown), and unfolding amplitudes are shown in Figure 8(a) and (b), respectively. The unfolding rate constants obtained from these fits agreed well with those obtained from global analysis of all the kinetic data (Figure 6(a)).

The unfolding amplitudes in Figure 8(a) measured at 4.3 M urea show that the species corresponding to phase 1, the red refolding phase, is formed immediately with no discernible lag. Its population reaches a maximum after a refolding time of 12 s before it decreases to zero over the next 300 s. In contrast, there is a lag in the formation of the species corresponding to phase 2, the blue refolding phase, before its population increases to dominate the refolding ensemble. This is consistent with an intermediate preceding its formation.[32] These data allow the assignment of a folding mechanism for YbeA, provided that several assumptions are made. The first is that during refolding, the species formed in the red refolding phase is an obligatory intermediate preceding the formation of the species corresponding to the blue refolding phase. The lag observed for the blue species during refolding while the species corresponding to the red refolding phase accumulates is consistent with this (Figure 8(a)). The second assumption is that the final refolding step involves formation of native dimer. These assumptions leave two possible folding mechanisms that could describe the YbeA experimental data; folding could occur by a three-state sequential mechanism

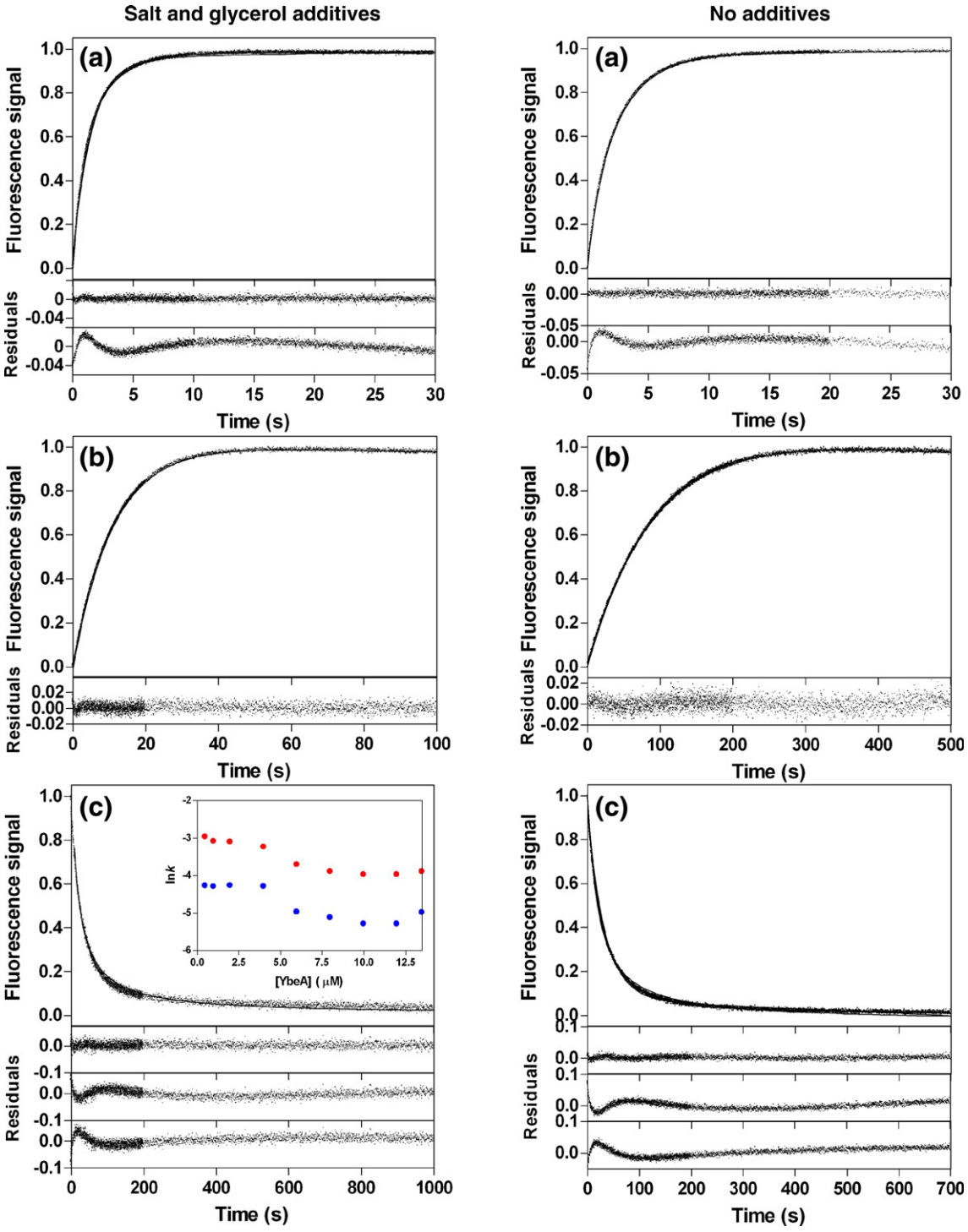

Figure 5. Typical YbeA kinetic traces for experiments performed in buffer with (left) and without (right) stabilising agents. Traces are normalised relative to a native dimer signal of zero and a denatured monomer signal of 1. (a) Unfolding in 8 M urea; residuals are for the fit of the trace to a first-order reaction with two exponentials (top) and a first-order reaction with one exponential (bottom). (b) Unfolding at 5.3 M urea (left) and 3.3 M urea (right). Residuals are for the fit of the trace to a first-order reaction with one exponential. (c) Refolding at 1 M urea; residuals are for the fit of the trace to a first-order reaction with two exponentials (top), a first-order reaction with one exponential (middle) and a second-order reaction with one-exponential (bottom). Inset: Protein concentration-dependence of the refolding rate constants at 1.75 M urea, coloured red and blue for fast and slow, respectively, calculated from the fit of the traces to a first-order reaction with two exponentials. Symbols are larger than the error in the rate constants. Conditions were as described for Figure 3.

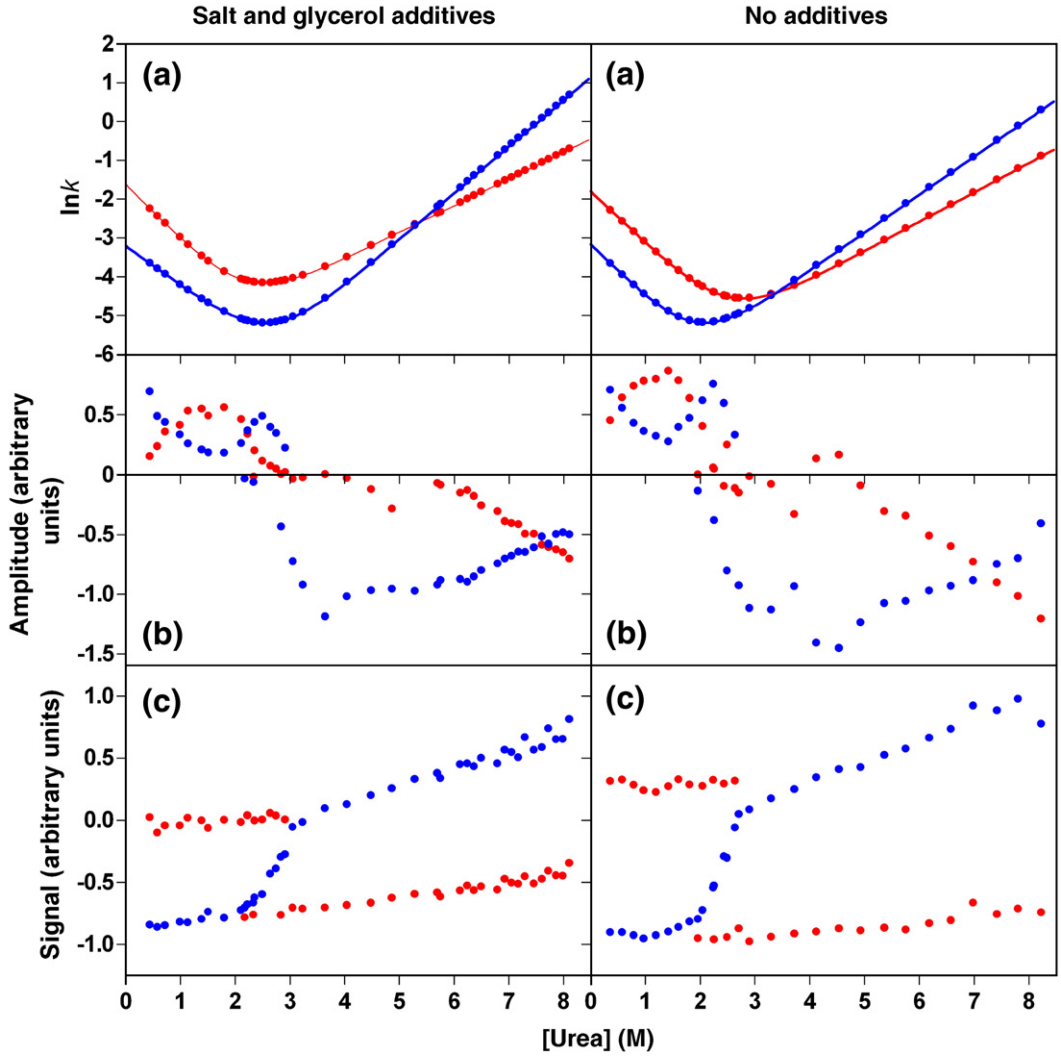

Figure 6. (a) Chevron plots for the folding and unfolding kinetics of YbeA in buffer conditions with (left) and without (right) stabilising agents, calculated from the global kinetic parameters shown in Table 4. Phases 1 and 2 are coloured red and blue, respectively. (b) Fluorescence amplitudes for the kinetic phases shown in (a) calculated from the global fit of all kinetic traces to equation (3). Amplitudes of refolding and unfolding reactions are positive and negative, respectively, and are coloured according to their corresponding phase in (a). (c) Initial (red) and final (blue) fluorescence signals for YbeA kinetic traces. All amplitude is accounted for in the kinetic traces, and there is no apparent fluorescence burst phase. Conditions for buffers with and without stabilising agents were as described for Figure 3(a) and (b), respectively.

involving either a dimeric (Scheme 1) or a monomeric (Scheme 2) kinetic intermediate:

$$2\mathrm{D} \overset{\text{Red}}{\longleftrightarrow} \mathrm{I}_2 \overset{\text{Blue}}{\longleftrightarrow} \mathrm{N}_2 \qquad \text{Scheme 1}$$

$$2\mathrm{D} \overset{\text{Red}}{\longleftrightarrow} 2\mathrm{I} \overset{\text{Blue}}{\longleftrightarrow} \mathrm{N}_2 \qquad \text{Scheme 2}$$

Since both red and blue refolding phases appear to be independent of protein concentration, assignment of either phase to a dimerisation step is difficult. The expected kinetic m-values associated with each scheme were calculated using the parameters shown in Table 4, and the following relationships:

$$m_{\text{Scheme1}} = m_{\text{phase1}} + m_{\text{phase2}}$$

$$m_{\text{Scheme2}} = 2m_{\text{phase1}} + m_{\text{phase2}}$$

The total m-value for Scheme 1 is 2.6 kcal mol^{-1} M^{-1}. This is much smaller than the $m_{\mathrm{N_2 \leftrightarrow 2D}}$ value of 4.4 kcal mol^{-1} M^{-1} calculated for YbeA from equilibrium studies (Table 2), and the m-value of 3.8–4.9 kcal mol^{-1} M^{-1} estimated from the associated *SASA* changes during the unfolding of a YbeA dimer (Table 3). The total m-value for Scheme 2 is 3.9 kcal mol^{-1} M^{-1} and, in contrast to that for Scheme 1, is in very good agreement with the m-values calculated from the equilibrium experiments and $\Delta SASA$ estimates. A pathway by which YbeA folds *via* a kinetic monomeric intermediate is therefore most consistent with all the equilibrium and kinetic data (Figure 8(c)). A simulation of the time-course of monomeric intermediate and native dimer present during refolding *via* this mechanism at 1 M urea was performed using the program KINSIM[34] and the appropriate rate constants from the chevron plot, and is shown in Figure 8(a). The good agreement of the amplitudes from the interrupted-

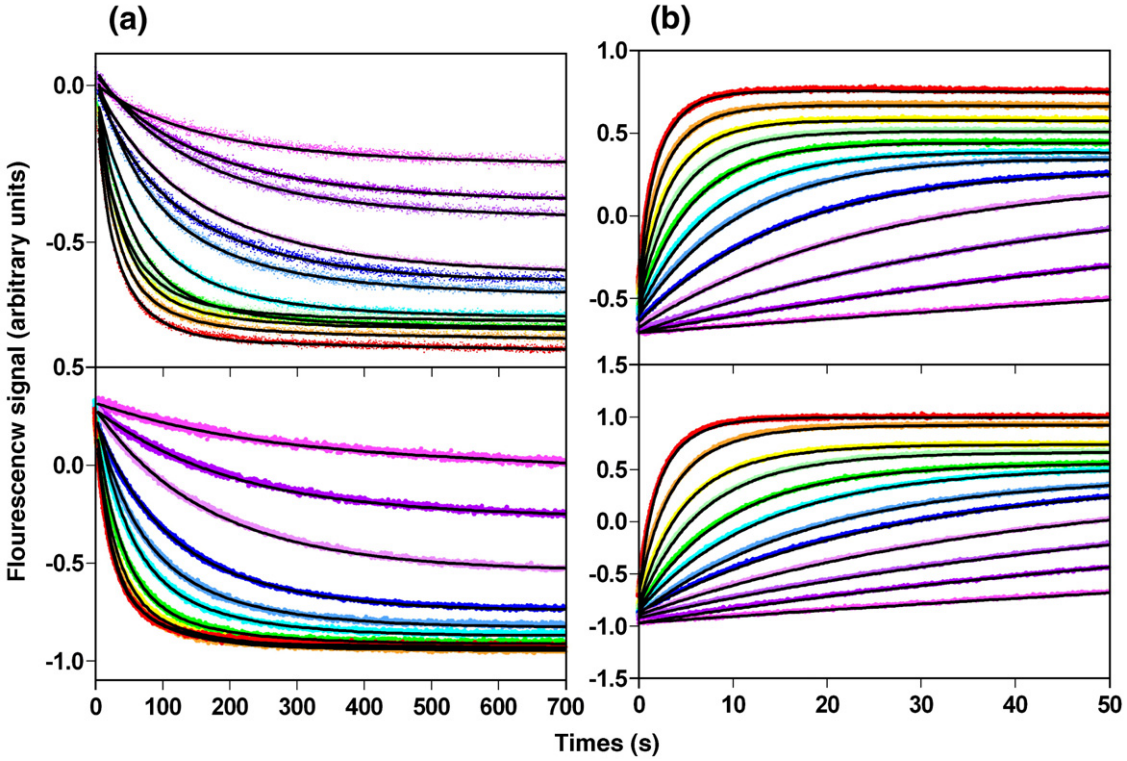

Figure 7. Global analysis of the folding and unfolding kinetics of YbeA for conditions with (top) and without (bottom) stabilising agents. (a) Refolding transients (0.8–2.6 M urea in 0.2 M increments, red to pink) and (b) unfolding transients (2.8–7.2 M urea in 0.4 M increments, pink to red). Continuous black lines represent the global fit of the data to equation (3). Conditions were as described for Figure 3.

refolding experiments with the results of the numerical simulation suggests that, despite being quite similar in value, the apparent relaxation rate constants calculated from the interrupted refolding experiments at 4.3 M urea do approximate the underlying microscopic rate constants at this concentration of urea.

Interrupted-refolding amplitudes when unfolding YbeA at 7.7 M urea differ from those at 4.3 M urea, and are shown in Figure 8(b). At 7.7 M urea, the population of the red refolding species increases steadily with delay time, with no lag, and reaches a stable maximum. The amplitude corresponding to the blue refolding species also increases with delay time, but a lag during the first 5 s is observed. If the YbeA folding mechanism shown in Figure 8(c) is correct, the observation of two unfolding phases after all refolding delay times at 7.7 M urea suggests that the intermediate I is populated during unfolding at this concentration of urea, and accumulates after the $N_2 \leftrightarrow 2I$ step (blue) has occurred, before the rate-determining $I \leftrightarrow D$ (red) step takes place. The unfolding amplitude of the red refolding phase after various delays is therefore not a direct measure of the population of I present after a given refolding delay, but instead is proportional to the amount of $[I+N_2]$ present. As mentioned earlier, the use of interrupted refolding to monitor population of an intermediate species assumes that any intermediate will unfold faster than the native protein.[31,32] In this case, however, the intermediate I unfolds slower than N_2 above concentrations of urea of 5.4 M, and

Table 4. Kinetic parameters for the global fit of YbeA unfolding and refolding kinetics at pH 7.5 in buffers with and without stabilising agents at 1 μM final protein concentration

Phase	Colour	Buffer additives	$k_f^{H_2O}$ (s^{-1})	$k_u^{H_2O}$ (s^{-1})	m_{k_f} (kcal mol^{-1} M^{-1})	m_{k_u} (kcal mol^{-1} M^{-1})	m_{kin}[a] (kcal mol^{-1} M^{-1})	$\Delta G^{kinb}_{H_2O}$ (kcal mol^{-1})
1	Red	Salt, glycerol	0.20±0.003	1.8 (±0.02)×10^{-3}	0.86±0.01	0.41±0.002	1.3±0.01	2.8±0.02
		None	0.16±0.001	8.1 (±0.2)×10^{-4}	0.80±0.01	0.45±0.003	1.3±0.01	3.1±0.03
2	Blue	Salt, glycerol	4.1 (±0.02)×10^{-2}	1.2 (±0.02)×10^{-4}	0.61±0.006	0.71±0.002	1.3±0.01	11.2±0.02
		None	4.2 (±0.03)×10^{-2}	4.3 (±0.06)×10^{-4}	0.82±0.01	0.58±0.003	1.4±0.01	10.5±0.02

Global analyses were performed with Prism, version 4 (GraphPad Software) using equation (3), and rate constants are first order. Errors quoted are the standard errors calculated by the fitting program; the small errors quoted for some parameters are a reflection of the global analysis, rather than a true experimental error.

[a] $m_{kin} = m_{k_f} + m_{k_u}$.

[b] $\Delta G^{kin}_{H_2O} = -RT \ln(k_u^{H_2O}/k_f^{H_2O})$ except for phase 2, where $\Delta G^{kin}_{H_2O} = -RT \ln(2k_u^{H_2O}/k_{2nd}^{H_2O})$, according to folding *via* the mechanism shown in Figure 8(c).

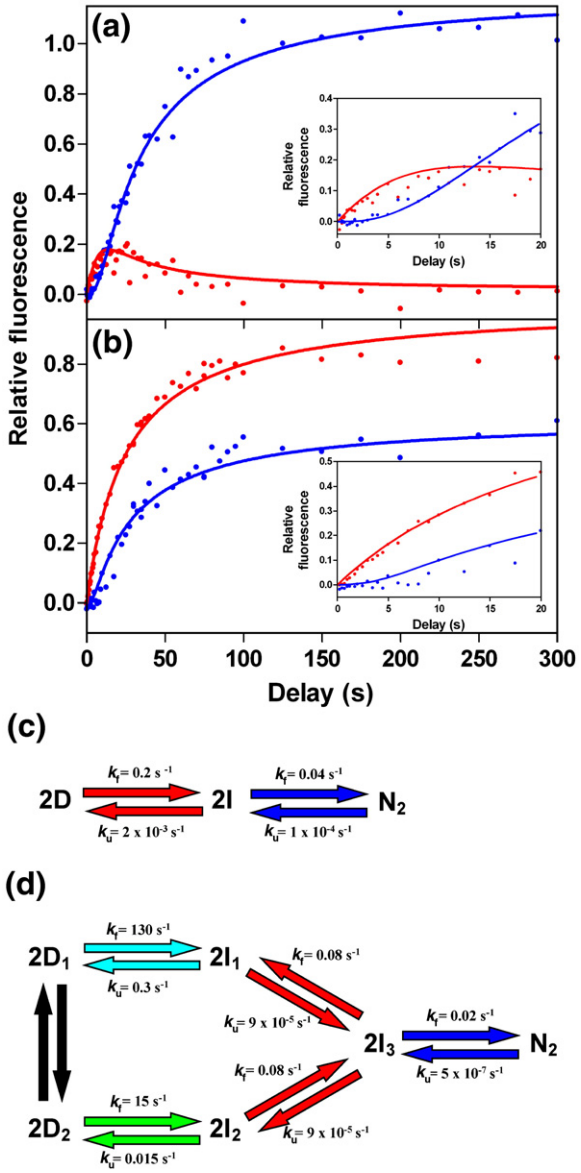

Figure 8. Relative amplitudes for the two YbeA unfolding phases seen during interrupted-refolding experiments after refolding to 1 M urea and unfolding to a final concentration of (a) 4.3 M urea and (b) 7.7 M urea. Insets show an expanded view for delay times up to 20 s. Amplitudes are coloured according to their corresponding phase in Figure 6(a). (c) The folding pathway of YbeA most consistent with all data. Rate constants are shown for buffer at 25 °C, pH 7.5 with salt and glycerol, and arrows are coloured to match their corresponding phase in Figure 6(a). Continuous lines in (a) and (b) represent the KINSIM simulation of the time-course of intermediate monomeric species and native YbeA dimer during refolding *via* the mechanism shown in (c). The continuous red line represents the population of I in (a) and the population of [I+N_2] in (b), where unfolding of N_2 is not rate limiting. (d) The folding pathway of YibK is taken from Mallam & Jackson.[20] The rate constants are for buffer at 25 °C, pH 7.5 with salt and glycerol. Conditions: 25 °C, 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% (v/v) glycerol, 1 mM DTT.

the presence of both I and N_2 molecules will result in the observation of a I↔D unfolding event. The continuous lines in Figure 8(b) represent the time-course of [I+N_2] (red) and N_2 (blue) molecules during refolding at 1 M urea, simulated using KINSIM.[34]

A free energy diagram illustrating how the rate-determining step in unfolding changes between 7.7 M urea and 4.3 M urea is shown in Figure 9. The free energy of unfolded YbeA has been set arbitrarily to zero, and the relative energies of the intermediate, I, and native dimer, N_2, were calculated using the parameters given in Table 4. Activation barriers were estimated using a derivation of the Eyring equation, and an empirical estimate of the pre-exponential factor.[35] Figure 9 illustrates the change in unfolding rate-determining step; at 7.7 M urea, the barrier for unfolding of N_2↔2I is lower than that for I↔D, hence the latter is rate limiting. The reverse is true at 4.3 M urea, where N_2↔2I is rate limiting and therefore the only reaction observed.

Discussion

The complicated backbone topology of YbeA involving the formation of a deep trefoil knot makes it an interesting and challenging candidate for a protein-folding study. It is classified on the basis of its structure as a member of the α/β-knot superfamily of MTases (Figure 1). Extensive studies

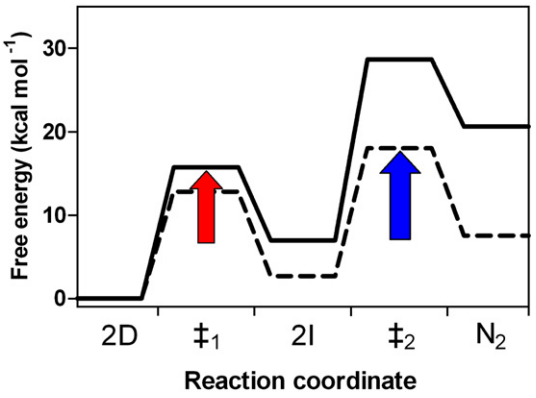

Figure 9. Profile of the reaction coordinate for the unfolding of YbeA at 7.7 M urea (continuous line) and 4.3 M urea (broken line) to illustrate the change in unfolding rate-determining step with concentration of urea. At 7.7 M urea, $\Delta G^{\ddagger}_{\ddagger_2\text{-}N2}$ (8.1 kcal mol^{-1}) < $\Delta G^{\ddagger}_{\ddagger_1\text{-}I}$ (8.8 kcal mol^{-1}), and the I↔D unfolding transition is rate limiting (red arrow). At 4.3 M urea, $\Delta G^{\ddagger}_{\ddagger_2\text{-}N2}$ (10.5 kcal mol^{-1}) > $\Delta G^{\ddagger}_{\ddagger_1\text{-}I}$ (10.2 kcal mol^{-1}) and unfolding of N_2↔2I is the rate-determining step (blue arrow). Stabilities are for 1 μM YbeA in 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% (v/v) glycerol, 1 mM DTT. Ground state stabilities were calculated from the kinetic parameters in Table 4, and the free energy of the denatured state was arbitrarily set to zero. Activation energies were estimated using the relationship $k_{obs} = k_a \exp(-G^{\ddagger}/RT)$. An empirical estimate of 10^6 s^{-1} was used for the pre-exponential factor k_a.[35]

on the folding of another member of this group of proteins, YibK from *Haemophilus influenzae*, have been undertaken.[19,20] The aim of this study is to characterise the thermodynamic and kinetic folding mechanism of YbeA, and to compare it to that of YibK.

Experiments have been performed on YbeA using two buffer conditions at pH 7.5: one with and one without salt and glycerol as stabilising agents. The former was to enable a direct comparison of stabilities and folding rates to YibK, where aggregation issues made it necessary to carry out experiments in a buffer with additives.[19] The latter was to allow the effect of stabilising agents on the folding properties of YbeA to be ascertained.

α/β-Knotted proteins are dimeric in solution

SEC studies were undertaken on YbeA and results are consistent with YbeA existing as a dimer at all experimental concentrations of protein studied, down to 5 μM (Figure 2). An upper limit for the dissociation constant between dimer and monomer, K_d, of 15 nM can be estimated from these data. These observations are similar to those made for YibK, which was shown to be dimeric in solution with a dissociation constant of less than 1 nM.[19] The strong association of monomers near pH 7 appears to be a common characteristic of proteins belonging to the α/β-knot superfamily and has been observed for other family members.[8,13]

YbeA unfolds *via* a monomeric equilibrium intermediate

Equilibrium denaturation studies have been performed on YbeA and, under all conditions examined, unfolding profiles appeared monophasic and protein concentration-dependent (Figure 4). A protein concentration-dependence in the equilibrium unfolding in dimeric systems is expected, and can be used to assign a denaturation model. YbeA denaturation profiles display an increase in apparent $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ and $m_{N_2 \leftrightarrow 2D}$ values with increasing protein concentration when analysed using a simple two-state dimer denaturation model (Table 1; Figure 4(a)), suggesting that YbeA unfolds *via* a three-state dimer denaturation involving a monomeric intermediate.[19,28] The thermodynamic parameters from the fit to this model show that YbeA has a total free energy of unfolding of native dimer to two unfolded monomers, $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$, of 20.2 kcal mol^{-1} and 18.9 kcal mol^{-1}, and an $m_{N_2 \leftrightarrow 2D}$ value of 4.4 kcal mol^{-1} M^{-1} and 4.6 kcal mol^{-1} M^{-1} for buffer with and without stabilising agents, respectively (Figure 4 (b); Table 2). These *m*-values agree well with those obtained from Δ*SASA* estimates (Table 3). $\Delta G^{I \leftrightarrow D}_{H_2O}$ and $m_{I \leftrightarrow D}$ values are similar for both buffer conditions, suggesting that stabilising agents have little effect on the stability and structure of the equilibrium monomeric intermediate (Table 2). In comparison, $\Delta G^{N_2 \leftrightarrow 2I}_{H_2O}$ values do show a dependence upon buffer conditions, and are 15.2 kcal mol^{-1} and 13.3 kcal mol^{-1} for conditions with and without stabilising agents, respectively. This suggests that the glycerol and salt have stabilised the native dimer by some 2 kcal mol^{-1}. Dissociation constants for dimer dissociating to a monomeric intermediate, $K_d^{N_2 \leftrightarrow 2I}$, can be calculated from the denaturation data, and are 7×10^{-12} M and 2×10^{-10} M for buffer with and without stabilising agents, respectively. These low values confirm that YbeA is dimeric at all experimental concentrations of protein studied under both buffer conditions, and the increased $K_d^{N_2 \rightarrow 2I}$ seen for buffer without stabilising agents reflects the slight reduction in stability of the dimer in the absence of salt and glycerol.

The equilibrium unfolding mechanism for YbeA described here can be compared to that for YibK. Comparisons are drawn from experiments performed under the same buffer conditions with salt and glycerol stabilising agents. Unfolding of both knotted proteins is fully reversible in urea (Figure 3), suggesting that their complicated topology has not hindered their folding efficiency.[19] Reversible folding in urea was observed also during the purification of the knotted protein TrmH from *Aquifex aeolicus*,[3] and therefore appears to be a common trait of all members of the α/β-knot clan. Equilibrium denaturation studies show that YbeA and YibK are stable homodimers that denature by the same equilibrium mechanism. They unfold *via* monomeric equilibrium intermediates that have comparable structure (Table 2). Intermediates for both proteins have undergone some partial loss of secondary and tertiary structure relative to a fully folded monomeric subunit in a dimer; this is reflected in the Y_I values calculated from the analysis of fluorescence and far-UV CD data, and the observation that the *m*-value predicted for the unfolding of a monomer subunit from analysis of Δ*SASA* is larger than that observed for unfolding of the monomeric intermediate measured experimentally (Tables 2 and 3).[19] YibK is the most stable dimer of the two by some 11 kcal mol^{-1}, and its monomeric intermediate is considerably more stable than that of YbeA (Table 2).

The folding mechanism of YbeA

The folding kinetics of YbeA were studied using single-jump and double-jump experiments. For each buffer condition, all kinetic traces were considered together and analysed globally to give the chevron plots and kinetic parameters shown in Figure 6 and Table 4, respectively. Global analysis allowed the characterisation of the two reversible YbeA folding phases, shown in red and blue (Figure 6(a)). The unfolding arms of the phases crossed at 5.4 M urea and 3.4 M urea for buffer with and without stabilising agents, respectively, indicating a change in rate-determining step on the YbeA unfolding pathway at these concentrations of urea.

Interrupted-refolding experiments were performed to both moderate unfolding conditions (4.3 M urea), where the blue unfolding phase is

rate limiting, and to strongly unfolding conditions (7.7 M urea), where both red and blue unfolding phases are observed, and these were used to assign a folding mechanism to YbeA (Figure 8). The former allowed the population of refolding intermediates to be mapped out; the species corresponding to the red refolding phase increased immediately to a maximum value before decaying, while the population of the species corresponding to the blue refolding phase displayed a lag, consistent with an obligatory intermediate preceding its formation. These data were most consistent with a sequential folding mechanism involving three species and including a kinetic monomeric intermediate. This scheme is shown in Figure 8(c) and the continuous lines in Figure 8(a) represent a simulation of species present during refolding *via* this mechanism.

Under strongly unfolding conditions at 7.7 M urea, unfolding of native dimer is no longer rate limiting. The change in rate-determining step with the concentration of urea on the YbeA unfolding pathway can be illustrated by considering the free energies along an unfolding reaction coordinate under strong and moderate unfolding conditions (Figure 9). Estimates of the transition-state energies involved for each reaction show that at 7.7 M urea the barrier for I↔D unfolding is highest, and therefore rate limiting. At 4.3 M urea, the N_2↔2I transition has the higher activation energy, and becomes the rate-limiting step (Figure 9).

The kinetic parameters from the global analysis of YbeA kinetic traces agree well with those from equilibrium denaturation experiments for both buffer conditions. The kinetic and equilibrium monomeric intermediates have comparable stabilities and *m*-values, suggesting that they are similar species (Tables 2 and 4). As with the YbeA dimerisation phase in equilibrium studies, the blue kinetic dimerisation phase (2I↔N_2) is destabilised during unfolding in buffer without stabilising agents compared to that with salt and glycerol. The agreement of kinetic and thermodynamic parameters gives confidence in the YbeA folding model assigned in each case.

Comparison of the folding pathways of YbeA and YibK: elements common to knotted proteins

The kinetic folding pathway of YbeA can be compared to that of YibK, shown in Figure 8(d). The pathway for YbeA is much simpler, and involves only three species compared to the six species involved in the folding of YibK (Figure 8). The complex folding of YibK is thought to be, in part, a consequence of proline isomerisation in the denatured state.[20] Unlike YibK, YbeA does not possess a *cis*-proline residue, nor any other proline residues in its native structure that appear to affect folding. Furthermore, it is possible that intermediate states similar to I_1 and I_2 seen in the folding of YibK exist for YbeA, but are simply too unstable to be populated significantly. Both YbeA and YibK fold *via* sequential mechanisms that involve monomeric kinetic intermediates that are the precursors to native dimer formation (Figure 8). Equally, they display a slow dimerisation phase that has an apparent rate constant in the region of 2×10^{-2} s^{-1}–4×10^{-2} s^{-1} at 1 μM protein (Figure 8). The corresponding second-order rate constants are some five orders of magnitude below the diffusion limit of 10^8–10^9 M^{-1} s^{-1}, implying that association is not diffusion limited in the folding of YbeA or YibK. It is interesting to note that their dimerisation is much slower than that observed for other dimeric proteins,[36–40] suggesting that a slow dimerisation step may be a characteristic of knotted protein folding. The dimerisation of YbeA displayed no obvious dependence on protein concentration at pH 7.5, suggesting that it is limited by a conformational change rather than a collision event at this pH, and so becomes a first-order reaction (Figure 5(c)). An identical situation was seen for the dimerisation reaction observed at pH 7.5 for YibK, which was also independent of protein concentration.[20] Both YibK and YbeA populate kinetic monomeric intermediates that are similar to those observed during equilibrium unfolding, shown by the good agreement of *m*-values and stability (Tables 2 and 4).[20] These monomeric intermediates are the precursors to native dimer formation, and are formed in a relatively slow folding step with a rate constant in the region of 0.1–0.2 s^{-1} (Figure 8). The increased stability of the YibK dimer compared to that of YbeA means that, under the conditions studied, dissociation is always rate-limiting during unfolding of wild-type dimeric YibK, and no change in the unfolding rate-determining step with urea concentration is observed.[20]

Conclusions

The folding of the knotted homodimer YbeA has been studied under a variety of buffer conditions using equilibrium denaturation and kinetic single-jump and double-jump experiments. YbeA unfolds under equilibrium conditions by a three-state dimer denaturation model involving a monomeric intermediate of appreciable structure and moderate stability. Kinetics experiments show that YbeA folds *via* a simple three-state sequential mechanism where monomeric precursors form native dimer in a slow dimerisation step. The kinetic and equilibrium monomeric intermediates have similar properties.

The folding of YbeA has been compared to that of the related knotted dimer YibK. Reversible folding in urea is a shared trait, and both have considerable stabilities and a common equilibrium unfolding mechanism. Strong dimerisation appears to be a characteristic of knotted proteins, and no evidence of dissociation of either protein is seen in buffer near neutral pH. Additionally, both fold *via* sequential mechanisms that involve the slow formation of a kinetic monomeric intermediate followed by an even slower dimerisation step. These similarities suggest that the mechanism of folding and knot formation in both proteins may be alike.

Materials and Methods

Materials

The gene encoding the hypothetical protein YbeA was amplified from *Escherichia coli* genomic DNA and subcloned into the pET-17b vector (Novagen). Chromatography columns and media were obtained from GE Health Sciences, and molecular biology grade urea was purchased from BDH Laboratory Supplies. All other chemicals and reagents were of analytical grade and were purchased from Sigma or Melford Laboratories. Millipore-filtered, double-deionised water was used throughout.

Protein expression and purification

YbeA was purified by the protocol used for YibK.[19] Protein yield was approximately 40 mg l^{-1}.

Buffers

Unless stated otherwise, all experiments were carried out in both 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% (v/v) glycerol, 1 mM DTT and 50 mM Tris–HCl (pH 7.5), 1 mM DTT. Aggregation assays were used to confirm that YbeA remained soluble under all conditions used (data not shown).[41]

Size-exclusion chromatography

Size-exclusion chromatography (SEC) was performed on YbeA by the methods used for YibK, and these are described elsewhere.[19] YbeA samples at various concentrations of protein between 5 μM and 40 μM, pre-equilibrated in 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% glycerol, 1 mM DTT, were injected (100 μl) onto an analytical gel-filtration column equilibrated in the same buffer. The relative elution volume was compared to that of molecular mass standards. SEC was not performed in 50 mM Tris–HCl (pH 7.5), 1 mM DTT, as YbeA interacted with the column without salt present.

Spectroscopic measurements

All measurements were made at 25 °C using a thermostatically controlled cuvette or cell. For fluorescence studies, an excitation wavelength of 280 nm (4 nm band-pass) was used in all experiments. An SLM-Amico Bowman series 2 luminescence spectrometer with a 1 cm path-length cuvette was used for fluorescence equilibrium denaturation experiments, and scans were recorded from 310–350 nm. Far-UV CD spectra were measured using an Applied Photophysics Chirascan. Scans were taken between 190 nm and 260 nm at a scan rate of 1 nm s^{-1} using a 0.1 cm path-length cuvette. Rapid-mixing fluorescence data were collected using an Applied Photophysics SX.18MV stopped-flow fluorimeter with a 335 nm cut-off filter.

Equilibrium denaturation experiments

Equilibrium denaturation experiments on YbeA were performed using the chemical denaturant urea and the methods described for YibK.[19] Samples were left for at least 1 h to equilibrate, after which no change in spectroscopic signal was seen.

Kinetic unfolding and refolding experiments

Kinetic folding experiments using rapid-mixing techniques were performed on YbeA by the methods used for YibK as described.[20]

Data analysis

All data analysis was performed using the non-linear, least-squares fitting program Prism, version 4 (GraphPad Software). A detailed description and derivation of the equations used to analyse YbeA equilibrium denaturation curves and kinetic traces can be found elsewhere.[19,20] Briefly, equilibrium data were first fit to a two-state dimer denaturation model:

$$N_2 \overset{K_U}{\longleftrightarrow} 2D \qquad \text{Scheme 3}$$

where the fraction of unfolded monomers, F_D, can be defined as:

$$F_D = \frac{\sqrt{K_U^2 + 8K_U P_t} - K_U}{4P_t}$$

where P_t is the total protein concentration in terms of monomer, and the equilibrium constant K_U, is defined as:

$$K_U = \exp\left(\frac{RT\ln(P_t - m([D]_{50\%} - [\text{denaturant}])}{RT}\right)$$

Fluorescence and far-UV CD datasets for each protein concentration were fit individually to:

$$Y_0 = Y_N(1 - F_D) + Y_D F_D \qquad (1)$$

where Y_0 is the spectroscopic signal at a given concentration of urea, and Y_N and Y_D are the spectroscopic signals for native and denatured monomeric subunits, respectively.

Additionally, data for each buffer were globally fit over all concentrations of protein to a three-state dimer denaturation model involving a monomeric intermediate:

$$Y_{re1} = Y_N\left(\frac{2P_t F_I^2}{K_1}\right) + Y_I(F_I) + Y_D(K_2 F_I) \qquad (2)$$

where Y_{re1} is the normalised spectral signal, Y_N, Y_I and Y_D are the spectroscopic signals of the native, intermediate and denatured state, respectively, F_I represents the fraction of monomeric subunits involved in the intermediate state, and K_1 and K_2 are the equilibrium constants for the first and second transitions, respectively.

A global analysis of YbeA kinetic traces was undertaken where all refolding and unfolding kinetic transients at different concentrations of urea were considered together and fit to a single equation.[42,43] This analysis makes several assumptions: first, that there are two, reversible folding phases; and second, that the folding limbs of both phases are linear. The second assumption is valid, as refolding rate constants obtained from the analysis of separate refolding traces show a linear dependence on the concentration of urea; no rollover is observed (data not

shown). All YbeA kinetic traces measured under the same buffer conditions were fit globally to:

$$Y(t) = Y_{\text{Native}} + Y_1\exp(-k_{\text{obs1}}t) + Y_2\exp(-k_{\text{obs2}}t) \quad (3)$$

where $k_{\text{obs}} = k_{\text{f}}^{\text{H}_2\text{O}}\exp(-m_{k_f}[\text{urea}]) + k_{\text{u}}^{\text{H}_2\text{O}}\exp(m_{k_u}[\text{urea}])$ for phases 1 and 2, and Y_1 and Y_2 are the corresponding fluorescence amplitude changes. Parameters $k_{\text{f}}^{\text{H}_2\text{O}}$, $k_{\text{u}}^{\text{H}_2\text{O}}$, m_{k_f} and m_{k_u} for each phase were shared throughout all datasets, and were utilised to construct the chevron plots shown in Figure 6(a) using:

$$\ln k_{\text{obs}} = \ln(k_f^{\text{H}_2\text{O}}\exp(-m_{k_f}[\text{urea}]) + k_u^{\text{H}_2\text{O}}\exp(m_{k_u}[\text{urea}])) \quad (4)$$

Traces from interrupted-refolding experiments for different delay times to the same final conditions were fit globally to equation (3), with values for the first-order unfolding rate constants shared throughout all datasets.

Kinetic simulations to model the time-course of species present during refolding *via* the folding mechanism shown in Figure 8(c) were performed using the numerical modelling program KINSIM,[34] and the appropriate rate constants from the chevron plots.

Solvent-accessible surface area and *m*-value calculations

The *SASA* of native dimeric YbeA was calculated from the coordinates of its X-ray crystal structure, using the web-based program GETAREA version 1.1.[44] The *SASA* was calculated also for a fully folded monomer subunit. The *SASA* of an unfolded monomer was estimated using values for individual residues obtained from tripeptide studies.[45] These studies used Gly-X-Gly tripeptides as model compounds for the *SASA* of side-chains in the unfolded state. However, tripeptide models are thought to often overestimate the *SASA* of the unfolded state; therefore, the *SASA* of an unfolded monomer was also estimated using values obtained from hard-sphere simulations, termed the upper bound model.[46,47]

It has been shown that the *m*-value of a protein is highly correlated to the Δ*SASA* between native and denatured states, and the following relationship has been observed for proteins without crosslinks:[30]

$$\text{Urea } m\text{-value} = \Delta SASA(0.14 \pm 0.01) \quad (5)$$

This relationship, along with the Δ*SASA* calculated for the unfolding transition, was used to estimate the *m*-value associated with complete unfolding from native dimer to two unfolded monomers. The theoretical *m*-value associated with dimer dissociation to fully folded monomer subunits was also estimated using this method; the Δ*SASA* upon dissociation is the difference in *SASA* between native dimer and two fully folded monomers.

Acknowledgements

Financial support was gratefully received from the Welton Foundation. A.L.M holds an MRC studentship. We thank Edward Coulstock for helpful discussions regarding molecular biology techniques.

References

1. Daggett, V. & Fersht, A. (2003). The present view of the mechanism of protein folding. *Nature Rev. Mol. Cell Biol.* **4**, 497–502.
2. Rousseau, F., Schymkowitz, J. W., Wilkinson, H. R. & Itzhaki, L. S. (2002). The structure of the transition state for folding of domain-swapped dimeric p13suc1. *Structure*, **10**, 649–657.
3. Pleshe, E., Truesdell, J. & Batey, R. T. (2005). Structure of a class II TrmH tRNA-modifying enzyme from *Aquifex aeolicus*. *Acta Crystallog. sect. F*, **61**, 722–728.
4. Taylor, W. R. (2000). A deeply knotted protein structure and how it might fold. *Nature*, **406**, 916–919.
5. Taylor, W. R. & Lin, K. (2003). Protein knots: a tangled problem. *Nature*, **421**, 25.
6. Wagner, J. R., Brunzelle, J. S., Forest, K. T. & Vierstra, R. D. (2005). A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature*, **438**, 325–331.
7. Ahn, H. J., Kim, H. W., Yoon, H. J., Lee, B. I., Suh, S. W. & Yang, J. K. (2003). Crystal structure of tRNA (m1G37)methyltransferase: insights into tRNA recognition. *EMBO J.* **22**, 2593–2603.
8. Elkins, P. A., Watts, J. M., Zalacain, M., van Thiel, A., Vitazka, P. R., Redlak, M. *et al.* (2003). Insights into catalysis by a knotted TrmD tRNA methyltransferase. *J. Mol. Biol.* **333**, 931–949.
9. Forouhar, F., Shen, J., Xiao, R., Acton, T. B., Montelione, G. T. & Tong, L. (2003). Functional assignment based on structural analysis: crystal structure of the yggJ protein (HI0303) of *Haemophilus influenzae* reveals an RNA methyltransferase with a deep trefoil knot. *Proteins: Struct. Funct. Genet.* **53**, 329–332.
10. Lim, K., Zhang, H., Tempczyk, A., Krajewski, W., Bonander, N., Toedt, J. *et al.* (2003). Structure of the YibK methyltransferase from *Haemophilus influenzae* (HI0766): a cofactor bound at a site formed by a knot. *Proteins: Struct. Funct. Genet.* **51**, 56–67.
11. Michel, G., Sauve, V., Larocque, R., Li, Y., Matte, A. & Cygler, M. (2002). The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot. *Structure*, **10**, 1303–1315.
12. Nureki, O., Shirouzu, M., Hashimoto, K., Ishitani, R., Terada, T., Tamakoshi, M. *et al.* (2002). An enzyme with a deep trefoil knot for the active-site architecture. *Acta Crystallog. sect. D*, **58**, 1129–1137.
13. Nureki, O., Watanabe, K., Fukai, S., Ishii, R., Endo, Y., Hori, H. & Yokoyama, S. (2004). Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme. *Structure*, **12**, 593–602.
14. Zarembinski, T. I., Kim, Y., Peterson, K., Christendat, D., Dharamsi, A., Arrowsmith, C. H. *et al.* (2003). Deep trefoil knot implicated in RNA binding found in an archaebacterial protein. *Proteins: Struct. Funct. Genet.* **50**, 177–183.
15. Mosbacher, T. G., Bechthold, A. & Schulz, G. E. (2005). Structure and function of the antibiotic resistance-mediating methyltransferase AviRb from *Streptomyces viridochromogenes*. *J. Mol. Biol.* **345**, 535–545.
16. Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K. & McCann, P. P. (1996). S-Adenosylmethionine and methylation. *FASEB J.* **10**, 471–480.
17. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S. *et al.* (2004). The Pfam protein families database. *Nucl. Acids Res.* **32**, D138–D141.
18. Virnau, P., A.M., L. & Kardar, M. (2006). Intricate

knots in proteins: function and evolution. *PLoS Comput. Biol.* **2**, 1074–1079.
19. Mallam, A. L. & Jackson, S. E. (2005). Folding studies on a knotted protein. *J. Mol. Biol.* **346**, 1409–1421.
20. Mallam, A. L. & Jackson, S. E. (2006). Probing nature's knots: the folding pathway of a knotted homodimeric protein. *J. Mol. Biol.* **359**, 1420–1436.
21. Gunasekaran, K., Eyles, S. J., Hagler, A. T. & Gierasch, L. M. (2001). Keeping it in the family: folding studies of related proteins. *Curr. Opin. Struct. Biol.* **11**, 83–93.
22. Zarrine-Afsar, A., Larson, S. M. & Davidson, A. R. (2005). The family feud: do proteins with similar structures fold via the same pathway? *Curr. Opin. Struct. Biol.* **15**, 42–49.
23. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
24. Plaxco, K. W., Simons, K. T., Ruczinski, I. & Baker, D. (2000). Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry*, **39**, 11177–11183.
25. Baker, D. (2000). A surprising simplicity to protein folding. *Nature*, **405**, 39–42.
26. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, 81–91.
27. Ponstingl, H., Henrick, K. & Thornton, J. M. (2000). Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins: Struct. Funct. Genet.* **41**, 47–57.
28. Hobart, S. A., Meinhold, D. W., Osuna, R. & Colon, W. (2002). From two-state to three-state: the effect of the P61A mutation on the dynamics and stability of the factor for inversion stimulation results in an altered equilibrium denaturation mechanism. *Biochemistry*, **41**, 13744–13754.
29. Soulages, J. L. (1998). Chemical denaturation: potential impact of undetected intermediates in the free energy of unfolding and *m*-values obtained from a two-state assumption. *Biophys. J.* **75**, 484–492.
30. 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.
31. Schmid, F. X. (1983). Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction via structural intermediates. *Biochemistry*, **22**, 4690–4696.
32. Heidary, D. K., O'Neill, J. C., Jr, Roy, M. & Jennings, P. A. (2000). An essential intermediate in the folding of dihydrofolate reductase. *Proc. Natl Acad. Sci. USA*, **97**, 5866–5870.
33. Wallace, L. A. & Matthews, C. R. (2002). Sequential vs. parallel protein-folding mechanisms: experimental tests for complex folding reactions. *Biophys. Chem.* **101-102**, 113–131.
34. Dang, Q. & Frieden, C. (1997). New PC versions of the kinetic-simulation and fitting programs, KINSIM and FITSIM. *Trends Biochem. Sci.* **22**, 317.
35. Hagen, S. J., Hofrichter, J., Szabo, A. & Eaton, W. A. (1996). Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proc. Natl Acad. Sci. USA*, **93**, 11615–11617.
36. Milla, M. E. & Sauer, R. T. (1994). P22 Arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry*, **33**, 1125–1133.
37. Zeeb, M., Lipps, G., Lilie, H. & Balbach, J. (2004). Folding and association of an extremely stable dimeric protein from *Sulfolobus islandicus*. *J. Mol. Biol.* **336**, 227–240.
38. 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 dimeric intermediate. *Biochemistry*, **37**, 15990–15999.
39. Doyle, S. M., Bilsel, O. & Teschke, C. M. (2004). SecA folding kinetics: a large dimeric protein rapidly forms multiple native states. *J. Mol. Biol.* **341**, 199–214.
40. de Prat-Gay, G., Nadra, A. D., Corrales-Izquierdo, F. J., Alonso, L. G., Ferreiro, D. U. & Mok, Y. K. (2005). The folding mechanism of a dimeric beta-barrel domain. *J. Mol. Biol.* **351**, 672–682.
41. Bondos, S. E. & Bicknell, A. (2002). Detection and prevention of protein aggregation before, during, and after purification. *Anal. Biochem.* **316**, 223–231.
42. Gloss, L. M. & Matthews, C. R. (1998). The barriers in the bimolecular and unimolecular folding reactions of the dimeric core domain of *Escherichia coli trp* repressor are dominated by enthalpic contributions. *Biochemistry*, **37**, 16000–16010.
43. Placek, B. J. & Gloss, L. M. (2005). Three-state kinetic folding mechanism of the H2A/H2B histone heterodimer: the N-terminal tails affect the transition state between a dimeric intermediate and the native dimer. *J. Mol. Biol.* **345**, 827–836.
44. Fraczkiewicz, R. & Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19**, 319–333.
45. Miller, S., Janin, J., Lesk, A. M. & Chothia, C. (1987). Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641–656.
46. Creamer, T. P., Srinivasan, R. & Rose, G. D. (1995). Modeling unfolded states of peptides and proteins. *Biochemistry*, **34**, 16245–16250.
47. Creamer, T. P., Srinivasan, R. & Rose, G. D. (1997). Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry*, **36**, 2832–2835.
48. Carson, M. (1997). Ribbons. *Methods Enzymol.* **277**, 493–505.

Edited by F. Schmid

(Received 10 August 2006; received in revised form 19 October 2006; accepted 3 November 2006)
Available online 10 November 2006