doi:10.1016/j.jmb.2004.12.055 J. Mol. Biol. (2005) 346, 1409–1421

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Folding Studies on a Knotted Protein

Anna L. Mallam and Sophie E. Jackson*

Chemistry Department
Lensfield Road, Cambridge
CB2 1EW, UK

YibK is a 160 residue homodimeric protein belonging to the SPOUT class of methyltransferases. Proteins in this group all display a unique topological feature; the backbone polypeptide chain folds to form a deep trefoil knot. Such knotted structures were completely unpredicted, it being thought impossible for a protein to fold efficiently in this way. However, they are becoming more common and there are now a growing number of examples in the Protein Data Bank. These intriguing knotted structures represent a new and significant challenge in the field of protein folding. Here, we present an initial characterisation of the folding of YibK, one of the smallest knotted proteins to be identified. This is the first detailed folding study on a knotted protein to be reported. We have established conditions under which the protein can be denatured reversibly *in vitro* using urea, thereby showing that molecular chaperones are not required for the efficient folding of this protein. A series of equilibrium unfolding experiments were performed over a 400-fold range of protein concentration. Both secondary and tertiary structural probes show a single, protein concentration-dependent unfolding transition, and data are most consistent with a three-state equilibrium denaturation model involving a monomeric intermediate. Thermodynamic parameters obtained from the fit of the data to this model indicate that the intermediate is a stable species with appreciable secondary and tertiary structure; whether the topological knot remains in the intermediate state is still to be shown. Together, these results demonstrate that, despite its complex knotted structure, YibK is able to fold efficiently and behaves remarkably similarly to other dimeric proteins under equilibrium conditions.

Keywords: protein folding; fluorescence; circular dichroism; dimer thermodynamics; topological knot

*Corresponding author

Introduction

The past two decades have seen the folding pathways of many proteins characterised in great detail, using both experimental and computational approaches. These studies have been biased towards small, monomeric proteins, lacking in disulphide bridges, cofactors and *cis* proline residues because they represent simple folding systems. As a result, many models for the different mechanisms by which small proteins fold have been proposed and tested.[1,2] A newly identified class of proteins, however, challenges current folding theories. These proteins contain trefoil knots formed by the backbone polypeptide chain, a rather unexpected structural feature.[3–10]

Knots in proteins are fairly common if the entire covalent network is considered; disulphide cross-links or metal-atom bridges often create "covalent knots", which can form either during or after folding.[11,12] The path of the backbone polypeptide chain alone defines "topological knots"; until relatively recently, the only examples of these involved the tucking of a few (ten at most) amino acid residues through a wide loop in the polypeptide chain.[13,14] These knots are termed shallow trefoil knots, but remain unconvincing, as they can disappear if the structure is viewed from a different angle.[15,16] A few years ago, a mathematical algorithm was developed that allowed the easy identification of knots in proteins.[16] The algorithm smoothes the polypeptide chain repeatedly while keeping the termini fixed; if a straight line is obtained the protein is not knotted. If a knot is

Abbreviations used: MTase, methyltransferase; AdoMet, *S*-adenosylmethionine; SASA, solvent-accessible surface area; CD, circular dichroism.

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

found, the location of the knotted core can be pinpointed; "deep" knots have more than 20 amino acid residues on either side of the core.[17] The algorithm led to the identification of an impressive deep figure-of-eight knot in acetohydroxy acid isomeroreductase (AIR) with over 200 amino acid residues on one side and 70 on the other. Thus, speculation began on how such a deep and complicated knot can form during folding.[16,17]

A deep topological trefoil knot was first identified in the catalytic domain of the hypothetical RNA 2′-*O*-ribose methyltransferase from *Thermus thermophilus* (RrmA), a member of the SpoU family.[8] To date, 12 further protein structures containing deep trefoil knots have been deposited in the Protein Data Bank, all belonging to the SpoU, TrmD or YbeA families.[3–7,9,10] These knotted proteins share some common features: structural similarities to other family members as well as functional assignment make it likely that all are methyltransferases (MTases), a type of enzyme involved in the transfer of the methyl group of *S*-adenosylmethionine (AdoMet) to carbon, nitrogen or oxygen atoms of DNA, RNA, proteins and other small molecules.[18] All form dimers in solution, with the knot forming a large part of the dimer interface it is also thought to be the location of the AdoMet binding site.[3–10] In recognition of the above similarities between MTases with knotted structures, a new class of fold has been defined known as the SPOUT (SpoU-TrmD) class MTase fold.[3] Features of MTases belonging to the SPOUT class include the presence of a deep trefoil knot that provides the AdoMet cofactor binding site, and dimer formation. Interestingly, SpoU, TrmD and YbeA families had been unified into the SPOUT class on the basis of their evolutionary origin, and it was predicted that they might share a similar fold.[19]

Proteins adopting the SPOUT class fold represent a novel folding problem. It is not obvious how a substantial length of polypeptide chain manages to thread itself through a loop during the process of protein folding. One proposal is that such proteins fold from a knotted unfolded state,[15] although theoretical studies suggest that the chance of a disordered polypeptide chain becoming entangled is small for proteins the size of those considered here.[20] To understand further the mechanisms involved in protein knot formation, we have undertaken a folding study on YibK, a member of the SPOUT class of MTases.[6]

YibK of *Haemophilus influenzae* is a 160 amino acid residue protein sharing significant sequence homology with the SpoU family of MTases. Crystallography studies indicate that it adopts a structure consistent with the SPOUT class MTase fold (Figure 1).[6] Specifically, a deep trefoil knot is formed at the C terminus by the threading of the last 40 residues (residues 121–160) through a knotting loop of approximately 39 residues (residues 81–120) (Figure 1(a)). Like other SPOUT class members, YibK exists as a dimer in solution. The dimer interface is located at the C-terminal α helix (α5), and consists of two closely packed monomers arranged in a parallel fashion (Figure 1(b)). YibK is one of the smallest SPOUT MTases identified so far, and therefore an ideal candidate for a folding study.

Here, we present the first characterisation of the folding of a protein containing a deep trefoil knot. We have examined the folding of YibK under equilibrium conditions, using probes of both secondary and tertiary structure. Conditions are found under which chemical denaturation by urea is fully reversible, and the equilibrium unfolding transition is monitored over a wide range of protein concentrations. The data obtained are used to assign an equilibrium-denaturation model in which the YibK dimer unfolds *via* a monomeric intermediate retaining significant structure.

Results

The homodimeric protein YibK was chosen for study as it contains a C-terminal deep trefoil knot in its backbone architecture. Understanding the mechanisms involved in the formation of such a knot represents a significant challenge, as in the native structure a region of polypeptide chain of about 40 residues has been threaded through a loop (Figure 1). Current general theories of protein folding offer little explanation as to how this type of structure may fold. Here, we present an initial characterisation of the folding of YibK; in particular, a detailed analysis of its unfolding equilibrium.

Determination of the oligomeric state of YibK

The oligomeric state of YibK was examined using analytical size-exclusion chromatography (SEC) at various concentrations of protein between 0.25 μM and 100 μM. Data throughout this concentration range are consistent with a dimeric form of YibK (Figure 2), and no monomer peak is seen. YibK elutes at a volume of 10.6 ml, which corresponds to a K_{AV} of 0.20 and a molecular mass of 36.7 kDa (see the calibration curve shown as inset (b) in Figure 2). This is close to the calculated molecular mass of 36.803 kDa for a YibK dimer.

Unfolding of YibK monitored by intrinsic fluorescence and far-UV circular dichroism

Intrinsic protein fluorescence provides a convenient way to monitor loss of tertiary structure during unfolding, as each YibK monomer contains two tryptophan residues and five tyrosine residues. Addition of urea to a final concentration of 8 M was found to result in an unfolding event, indicated by a large decrease in fluorescence with a concomitant red shift in the λ_{max} from 328 nm to 347 nm (see inset (a), Figure 3). The maximum fluorescence change occurred at 319 nm, and this wavelength was used to monitor unfolding in the equilibrium experiments. To ensure intrinsic fluorescence

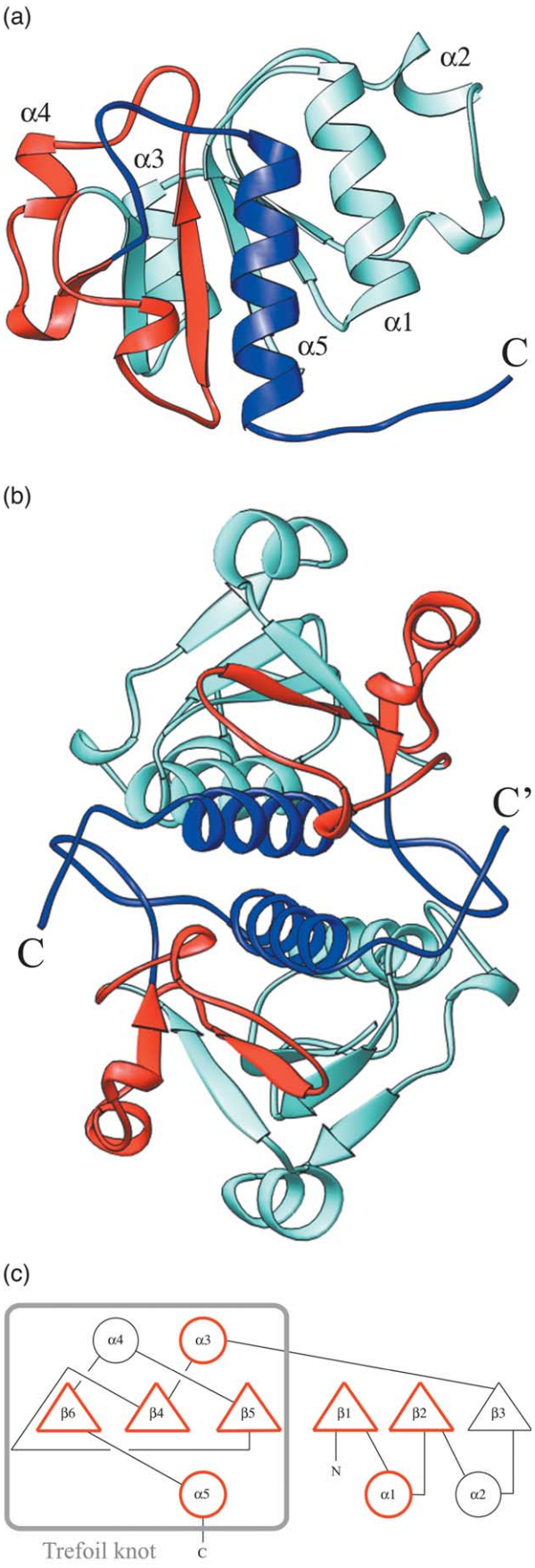

Figure 1. Structure of YibK from *Haemophilus influenzae*. (a) A ribbon diagram of a monomer subunit (PDB code 1MXI), showing the deep trefoil knot at the C terminus. The structure is coloured according to definitions given by Nureki *et al.*:[8] the knotting loop is highlighted in red (residues 81–120), while the knotted chain appears dark blue (residues 121–160). (b) Structure of dimeric YibK, coloured as in (a). YibK dimerises in a parallel fashion, with the knot forming a substantial part of the dimerisation interface. The ribbon diagrams were generated using Ribbons.[47] (c) A topological diagram of YibK. Structural elements that are common to the SPOUT class of MTases are outlined in red.

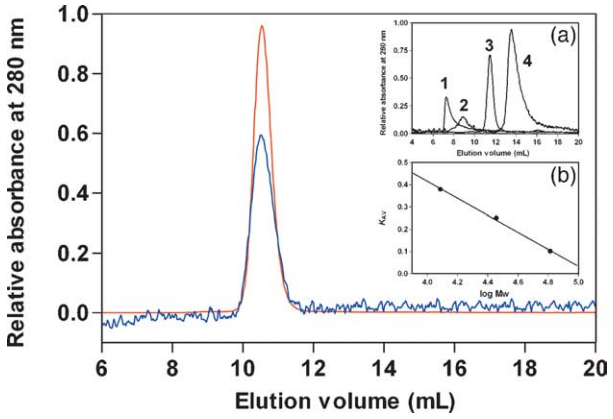

Figure 2. Determination of the oligomeric state of YibK by size-exclusion chromatography. Main: Elution profile of YibK at 100 μM (red) and 0.25 μM (blue) protein. Insets: (a) Elution profile of (1) blue dextran 200 (2000 kDa), (2) bovine serum albumin (66.3 kDa), (3) carbonic anhydrase (28.8 kDa) and (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.

monitors a global not a local unfolding event, far-UV circular dichroism (far-UV CD) was used as a probe of secondary structure. The helical signal observed for native YibK was completely lost on unfolding in 8 M urea, and there is no evidence for significant residual structure under the conditions used (see Figure 3, inset (b)).

Test of reversibility

Both intrinsic protein fluorescence and far-UV CD were used to investigate the reversibility of the YibK unfolding reaction. Unfolded and subsequently refolded YibK retains approximately 100% of its native fluorescence and far-UV CD signal (see Figure 3, insets (a) and (b)). Furthermore, refolding fluorescence equilibrium titrations (performed by diluting unfolded protein into buffer) superimpose onto unfolding equilibrium titrations at the same concentration of YibK (Figure 3), confirming that the unfolding of YibK is fully reversible under the buffer conditions used.

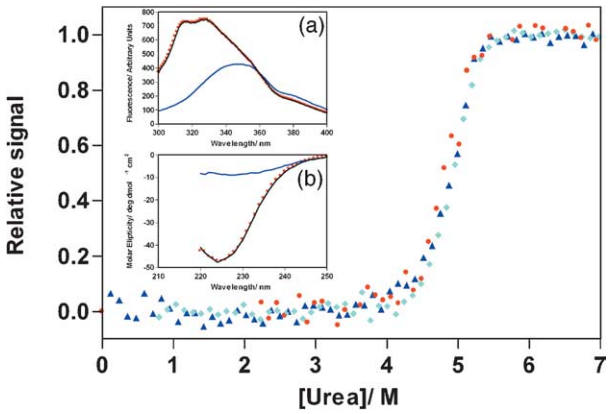

Figure 3. Reversibility of the YibK unfolding reaction. Main: YibK fluorescence denaturation (dark blue), fluorescence renaturation (light blue), and far-UV CD denaturation (red) profiles at 1 μM protein. Far-UV CD data points were not recorded at concentrations of urea below 2 M, as there was no change in baseline. Insets: (a) fluorescence and (b) far-UV CD emission spectra of native (black line), denatured (blue line) and renatured from 8 M urea (red circles) YibK at 5 μM protein. Conditions: 25 °C in 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% (v/v) glycerol, 1 mM DTT.

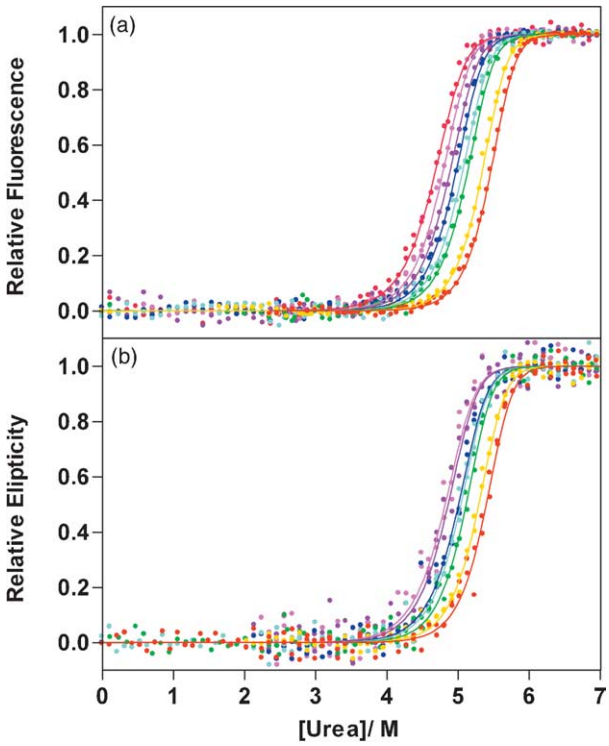

Figure 4. YibK equilibrium denaturation profiles for 100 μM (red), 50 μM (yellow), 10 μM (green), 5 μM (light blue), 2.5 μM (dark blue), 1 μM (purple), 0.5 μM (lilac) and 0.25 μM (dark pink) protein, as measured by (a) fluorescence emission at 319 nm and (b) far-UV CD signal at 225 nm. The data have been normalised for ease of comparison, and continuous lines represent the fit to a two-state dimer denaturation model (equation (7)). The conditions are as described for Figure 3.

Equilibrium unfolding experiments: protein concentration dependence

YibK equilibrium denaturation curves were measured using intrinsic fluorescence and far-UV CD over a 400-fold and a 200-fold range in protein concentration, respectively. The results of these experiments are shown in Figure 4 (the data have been normalised for ease of comparison). Both secondary and tertiary structural probes show a single unfolding transition over the range of protein concentrations studied. As expected for a dimer, the denaturation curves are protein concentration-dependent; an increase in the midpoint of the unfolding transition is seen with increasing protein concentration. This is a unique characteristic of oligomeric protein systems, and is due to the coupled denaturation and dissociation reactions that occur during unfolding.[21] Loss of secondary and tertiary structure appears concomitant, as fluorescence and far-UV CD curves superimpose (Figure 3). The concentration-dependence of the YibK unfolding curves can be used to assign the correct dimer denaturation model.

Data fitting

Two-state dimer model

Equilibrium unfolding data were first fit to a two-state dimer denaturation model in which native dimer, N_2, is in equilibrium with denatured monomer, D (Scheme 1). Fluorescence and far-UV CD datasets were treated separately, and curves for each concentration of protein were fit individually to equation (7). Figure 4 shows the results of these fits, with the associated thermodynamic parameters calculated shown in Table 1. While the fits appear good, there are several indicators that the two-state dimer model is not describing the YibK equilibrium unfolding data adequately. First, Table 1 shows that for the fluorescence and far-UV CD data, $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ increases with concentration of protein (P_t, measured in terms of monomer). $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ is the free energy difference between 1 mol of dimer and 2 mol of unfolded monomer in the absence of denaturant. If the correct dimer denaturation model is being used, $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ should remain constant at all concentrations of protein examined.[22,23] Second, the slope of the unfolding transition ($m_{N_2 \leftrightarrow 2D}$) shows an overall increase with protein concentration, suggesting that at higher concentrations of YibK unfolding becomes more co-operative. When examined using Student's *t*-test, the difference in $m_{N_2 \leftrightarrow 2D}$ of two datasets closest in protein concentration is not statistically significant. However, the $m_{N_2 \leftrightarrow 2D}$ values of the lowest and highest protein concentration datasets are considered to be very statistically different, and Student's *t*-test gives the probability of this difference having occurred by

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

	Fluorescence			Far-UV CD		
P_t (μM)	$[D]_{50\%}$ (M urea)	$m_{N_2\leftrightarrow 2D}$ (kcal mol^{-1} M^{-1})	$\Delta G^{N_2\leftrightarrow 2D}_{H_2O}$ (kcal mol^{-1})	$[D]_{50\%}$ (M urea)	$m_{N_2\leftrightarrow 2D}$ (kcal mol^{-1} M^{-1})	$\Delta G^{N_2\leftrightarrow 2D}_{H_2O}$ (kcal mol^{-1})
0.25	4.68±0.01	3.81±0.09	26.8±0.6	–	–	–
0.5	4.79±0.01	4.19±0.07	28.7±0.5	4.82±0.02	3.90±0.25	27.4±1.7
1	4.88±0.01	4.12±0.11	28.3±0.8	4.86±0.01	4.04±0.25	27.8±1.7
2.5	4.97±0.01	4.28±0.07	28.9±0.5	5.01±0.01	4.10±0.20	28.2±1.4
5	5.06±0.01	4.26±0.10	28.8±0.7	5.04±0.01	4.24±0.15	28.6±1.0
10	5.13±0.01	4.39±0.08	29.3±0.5	5.11±0.01	4.46±0.18	29.6±1.2
50	5.34±0.01	4.61±0.06	30.5±0.4	5.29±0.01	4.40±0.12	29.1±0.8
100	5.46±0.01	4.66±0.05	30.9±0.3	5.40±0.01	4.52±0.17	29.9±1.1

Fluorescence was monitored at 319 nm and far-UV CD at 225 nm. The parameters are quoted with their standard errors. Urea-denaturation profiles were fitted singularly to equation (7) using Prism, version 4, and $\Delta G^{N_2\leftrightarrow 2D}_{H_2O}$ was calculated using equation (4).

chance to be less than 1%. This observation is not expected for a simple two-state denaturation mechanism, and indicates that a more complex scheme is necessary to describe the system fully.[24,25] In particular, the variation in *m*-value with protein concentration implies that an equilibrium intermediate state is populated under equilibrium conditions.[26,27]

Three-state dimer models

Three-state dimer models involving either a dimeric (Scheme 2) or a monomeric (Scheme 3) intermediate were applied to the YibK equilibrium unfolding data. For both, the complexity of the system results in a large number of variable parameters (see equations (12) and (16)), and it is not possible to fit an individual dataset with any degree of accuracy. However, the models can be applied globally with thermodynamic parameters shared throughout all datasets. Denaturation profiles for all concentrations of protein were fit globally to both models in this manner; fluorescence and far-UV CD datasets were considered separately. The best fit to the data was achieved using a three-state dimer denaturation model with a monomeric intermediate, and the results of this fit along with the thermodynamic parameters obtained are shown in Figure 5 and Table 2, respectively. $\Delta G^{N_2\leftrightarrow 2D}_{H_2O}$, the free energy change corresponding to complete unfolding of dimer to two unfolded monomers in the absence of denaturant for a standard state of 1 mol of dimer, is calculated to be 31.9 kcal mol^{-1} and 33.7 kcal mol^{-1}, using fluorescence and far-UV CD data sets, respectively. The free energy difference between dimer and monomeric intermediate, $\Delta G^{N_2\leftrightarrow 2I}_{H_2O}$, is 18.9 kcal mol^{-1} and 19.1 kcal mol^{-1} for fluorescence and far-UV CD datasets, respectively. The monomeric intermediate has a stability of 6.5 kcal mol^{-1} and 7.3 kcal mol^{-1} relative to denatured monomer, for fluorescence and far-UV CD datasets, respectively. The equilibrium unfolding data do not agree well with a three-state dimer model involving a dimeric intermediate, and the thermodynamic parameters could not be determined with any precision (data not shown).

Relating *m*-values and the change in solvent-accessible surface area

The change in solvent-accessible surface area (ΔSASA) upon unfolding is strongly correlated with the *m*-value of a protein.[28] The crystal structure of YibK was used to calculate the SASA of both a native dimer and a fully folded monomeric subunit, allowing the ΔSASA upon dissociation to be determined (Table 3). The *m*-value associated with dissociation from native dimer to fully folded monomers ($N_2\leftrightarrow 2N$) is 0.4 kcal mol^{-1} M^{-1},

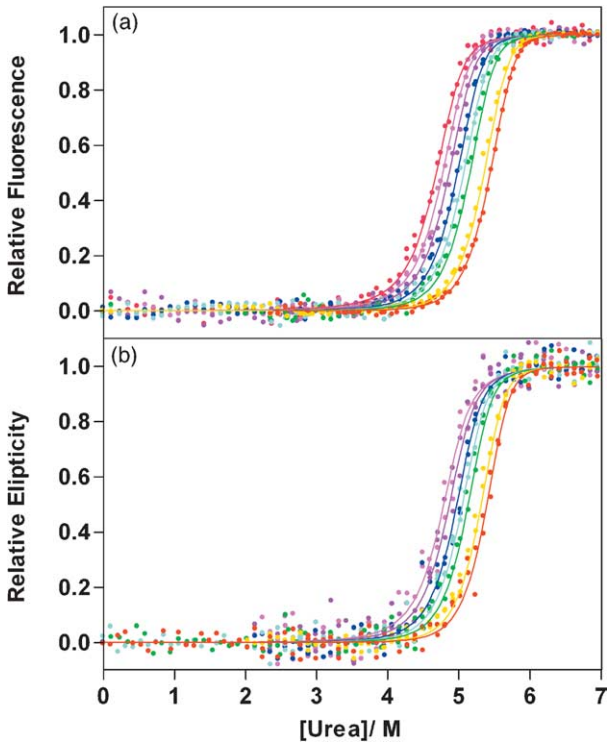

Figure 5. Global fit of (a) fluorescence and (b) far-UV CD YibK equilibrium denaturation data to a three-state dimer model with a monomeric intermediate (equation (16)). Protein concentrations are indicated by colour as in Figure 4, and the continuous lines represent the fit to the model.

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

	Y_I	$\Delta G^{N_2\leftrightarrow 2I}_{H_2O}$ (kcal mol^{-1})	$m_{N_2\leftrightarrow 2I}$ (kcal mol^{-1} M^{-1})	$\Delta G^{I\leftrightarrow D}_{H_2O}$ (kcal mol^{-1})	$m_{I\leftrightarrow D}$ (kcal mol^{-1} M^{-1})	$\Delta G^{N_2\leftrightarrow 2D}_{H_2O}$ (kcal mol^{-1})[a]	$m_{N_2\leftrightarrow 2D}$ (kcal mol^{-1} M^{-1})[b]
Fluorescence	0.61 ± 0.04	18.9 ± 0.4	1.80 ± 0.09	6.5 ± 0.2	1.53 ± 0.05	31.9 ± 1.2	4.86 ± 0.29
Far-UV CD	0.39 ± 0.05	19.1 ± 0.1	2.03 ± 0.01	7.3 ± 0.1	1.60 ± 0.02	33.7 ± 0.5	5.23 ± 0.07

Fluorescence and far-UV CD datasets were analysed separately. 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.

[a] $\Delta G^{N_2\leftrightarrow 2D}_{H_2O} = \Delta G^{N_2\leftrightarrow 2I}_{H_2O} + 2\Delta G^{I\leftrightarrow D}_{H_2O}$.

[b] $m_{N_2\leftrightarrow 2D} = m_{N_2\leftrightarrow 2I} + 2m_{I\leftrightarrow D}$.

estimated using equation (17). Evaluating ΔSASA for complete unfolding involves an estimate of SASA of each residue in the denatured state. Two alternative methods are used here: one based on tripeptide model compound studies,[29] the other on a model of the unfolded polypeptide chain generated from hard-sphere simulations.[30,31] The values calculated for YibK are shown in Table 3, along with the corresponding ΔSASA for unfolding and the associated *m*-value estimated using equation (17). The *m*-value corresponding to complete unfolding of dimeric YibK into two denatured monomers is estimated to be 4.4–5.5 kcal mol^{-1} M^{-1}, depending on the method used to model the denatured state SASA. This is in good agreement with the $m_{N_2\leftrightarrow 2D}$ values shown in Table 2 of 4.86 kcal mol^{-1} M^{-1}and 5.23 kcal mol^{-1} M^{-1} using fluorescence and far-UV CD data, respectively.

Discussion

YibK is a member of the SPOUT class of methyltransferases, a group of homodimeric proteins that contain deep trefoil knots in their backbone topology. The knot structure forms an integral part of the dimer interface, and consists of a segment of polypeptide chain, approximately 40 residues in length, threaded through a loop (Figure 1). YibK is one of the smallest SPOUT MTases identified so far, making it an ideal candidate for studies aimed at understanding how this class of knotted protein folds. Here, we report the results of experiments on the equilibrium denaturation of YibK, and propose a model for the equilibrium unfolding reaction.

Dimeric YibK

YibK and other SPOUT MTases are known to be dimers in the crystalline form.[3–10] Size-exclusion chromatography (SEC) was used to confirm the oligomeric state of YibK over the range of protein concentrations used in these experiments. All data were consistent with a dimeric structure, confirming that YibK is a stable dimer down to sub-micromolar concentrations (Figure 2). From these

Table 3. Changes in SASA for YibK 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)	ΔSASA for dissociation $N_2\leftrightarrow 2N$[b] (Å^2)	*m*-value estimate for dissociation $N_2\leftrightarrow 2N$[c] (kcal mol^{-1} M^{-1})
14,461	8806	3151	0.4 ± 0.1

B. *Unfolding*

	SASA for folded protein[a] (Å^2)	SASA estimate for unfolded protein (Å^2)		ΔSASA 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)	14,461	53,814	45,740	39,353	31,279	5.5 ± 0.4	4.4 ± 0.3
Fully folded monomer (N)	8806	26,907	22,870	18,101	14,064	2.5 ± 0.2	2.0 ± 0.1

[a] Calculated using the web-based program GETAREA, version 1.1.[46]

[b] ΔSASA upon dissociation = 2(SASA of monomer subunit)–(SASA of dimer). The value calculated for the ΔSASA upon dimer dissociation assumes no unfolding of the monomer subunits.

[c] Estimated using equation (17), a correlation equation given by Myers *et al.*[28] The errors were estimated also using this equation.

[d] ΔSASA for unfolding = (SASA unfolded protein)–(SASA folded protein).

[e] Calculated using values from tripeptide studies.[29]

[f] Calculated using data given by Creamer *et al.*[31]

data, an upper limit of 1 nM can be estimated for K_d, the equilibrium dissociation constant for monomer–dimer interconversion ($N_2 \leftrightarrow 2N$), and the absence of monomers even at low concentrations of protein indicates that dimer association is strong under the conditions used.

Equilibrium unfolding of YibK occurs *via* a monomeric intermediate

The equilibrium denaturation of YibK was studied using both intrinsic protein fluorescence and far-UV CD over a large range of protein concentrations. Within this range, the unfolding transition was shown to be fully reversible, monophasic and protein concentration-dependent. The protein concentration-dependence of equilibrium unfolding curves in dimer systems is expected, and can be rationalised from the way the equilibrium constant K_U is defined: $K_U = [D]^2/[N_2]$. At a given concentration of denaturant, K_U and ΔG remain the same for all concentrations of protein; only the fraction of each equilibrium species present changes with P_t. In a dimeric two-state system, $[D]_{50\%}$ is defined as the concentration of denaturant where the fraction of unfolded monomers is equal to the fraction of monomers present as dimers. This is also where $K_U = P_t$ (see the section on data analysis for more details); therefore, the concentration of denaturant where $[D]_{50\%}$ occurs will depend on P_t, and a change in $[D]_{50\%}$ with total protein concentration is expected. $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ remains constant with changing P_t and $[D]_{50\%}$ (equation (4)). The *m*-value of a protein is related to ΔSASA.[26,32] It is, therefore, a constant for each protein, and should not be affected by protein concentration. Fluorescence and far-UV CD YibK unfolding data were treated separately and fit to both a two-state model and three-state models with either a dimeric or a monomeric intermediate. If the denaturation mechanism were two-state, the $m_{N_2 \leftrightarrow 2D}$ and $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ values determined should remain the same for all concentrations of protein examined.[22,23] However, when YibK unfolding data were analysed according to a two-state denaturation model, an increase in $m_{N_2 \leftrightarrow 2D}$ and $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ values with increasing protein concentration was observed (Table 1). Thus, a two-state model does not describe the experimental results adequately. Global fitting of the YibK denaturation curves to both three-state models showed that a denaturation mechanism involving a monomeric intermediate best described the data (Figure 5 and Table 2). A three-state model with a dimeric intermediate was ruled out, as there were very large errors associated with the thermodynamic parameters obtained from the global fit of the data to this model (data not shown).

The feasibility of a particular three-state model can be thought about intuitively by considering the slope of the unfolding curves (i.e. the apparent *m*-value calculated in the two-state analysis, shown in Table 1). In a dimer system, the presence of an intermediate is obvious if the equilibrium unfolding transition is biphasic[33–37] and each transition can be considered separately. If the transition appears monophasic, however, an intermediate cannot be observed directly, but can manifest itself in a protein concentration-dependent slope change.[24] With many monomeric proteins a decrease in apparent *m*-value upon mutation is due to the population of an intermediate state.[27,38] This decrease is dependent on the spectral properties of the intermediate and its concentration; a significant population of the intermediate will result in an underestimation of the *m*-value for an $N \leftrightarrow D$ process.[26,27] With dimeric proteins, the concentration of any intermediate depends on the values of K_1 and K_2, and the total concentration of protein. At a particular concentration of denaturant, K_1 and K_2 remain constant and do not change with P_t. They are defined, respectively, in terms of the concentrations of species present as:

$$K_1 = \frac{[I_2]}{[N_2]} \text{ and } K_2 = \frac{[D_2]}{[I_2]}$$

for a three-state model with a dimeric intermediate, and as:

$$K_1 = \frac{[I]^2}{[N_2]} \text{ and } K_2 = \frac{[D]}{[I]}$$

for a three-state model with a monomeric intermediate. If a dimeric intermediate is involved, the ratio of $[I_2]$ to $[N_2]$ remains the same for all protein concentrations. However, the definition of K_2 means that as $[I_2]$ increases with the total protein concentration, [D] will increase only in proportion to the square-root of $[I_2]$. Overall, this will lead to an increase in the fraction of I_2 present. Therefore, in a dimer three-state system with a dimeric intermediate, the slope of the transition would be expected to decrease as the total protein concentration increases, due to an increased population of the intermediate state. Table 1 shows that this is not the case for YibK, making a dimeric intermediate unlikely. The same argument can be applied to a three-state model involving a monomeric intermediate. In this case, for a particular value of K_2, the ratio of [I] to [D] will remain constant at all P_t. Conversely, the definition of K_1 means that as P_t increases, [I] increases with only the square-root of $[N_2]$. This will lead to an overall decrease in the fraction of I present with increasing protein concentration. Reflecting this, the slope of the equilibrium unfolding transition in a dimeric system with a monomeric intermediate is expected to increase as P_t increases. Table 1 shows that this is indeed what happens with YibK, and, as expected, the best fit of the data is to a three-state dimer denaturation model involving a monomeric intermediate (Figure 5 and Table 2).

A recent study by Gunasekaran and co-workers revealed a distinct correlation between dimers composed of stable monomeric subunits and the ratio of the dimer interface area to the total surface

area.[39] This ratio for YibK is comparable to that of complexes consisting of stable monomers.

The thermodynamic parameters obtained from the fit to a three-state dimer denaturation model with a monomeric intermediate gave the free energy of unfolding from native dimer to denatured monomers, $\Delta G_{H_2O}^{N_2 \leftrightarrow 2D}$, as 32–34 kcal mol^{-1} (Table 2). This is comparable to values reported for other dimer systems.[21] Here, however, it is important to bear in mind that experiments were done in a buffer containing glycerol, an additive known to stabilise the native states of proteins.[40] This was necessary to maintain 100% protein solubility during experiments at higher concentrations of YibK, which were required for an accurate global analysis. Experiments carried out in the absence of glycerol suggest that its addition stabilises YibK by some 5 kcal mol^{-1} (data not shown). The associated m-value for complete unfolding, $m_{N_2 \leftrightarrow 2D}$, was calculated to be 4.9–5.2 kcal mol^{-1} M^{-1}, in good agreement with the m-value of 4.4–5.5 kcal mol^{-1} M^{-1} estimated from the ΔSASA of unfolding (Table 3), therefore giving confidence in the parameters obtained from this model.

The monomeric intermediate

The thermodynamic parameters shown in Table 2 provide information about the monomeric intermediate state. The free energy of the intermediate relative to the denatured monomer, $\Delta G_{H_2O}^{I \leftrightarrow D}$, is 6.5–7.3 kcal mol^{-1}, and within the range found for many small globular proteins.[41] This stability suggests that the intermediate state has significant structure. The experimentally determined m-values, $m_{N_2 \leftrightarrow 2I}$ and $m_{I \leftrightarrow D}$ (Table 2), can be compared to m-values estimated from the SASA changes involved in dissociation and unfolding (Table 3). The m-value predicted for dissociation of a YibK dimer into two native-like monomers is 0.4 kcal mol^{-1} M^{-1}. This is substantially less than the experimental $m_{N_2 \leftrightarrow 2I}$ value measured of 1.8–2.0 kcal mol^{-1} M^{-1}, and indicates that each monomer has partially unfolded on dissociation to form the intermediate state. Likewise, the m-value predicted for the unfolding of a native-like monomer is 2.0–2.5 kcal mol^{-1} M^{-1} (Table 3), which is higher than the experimental m-value for the I↔D transition of 1.5–1.6 kcal mol^{-1} M^{-1}. Again, this indicates a loss of structure in the intermediate state compared to the monomeric unit in the crystal structure. The experimentally determined spectral signal for the intermediate (Y_I) is given in Table 3, relative to a signal of 0 for a native monomeric subunit in a dimer and 1 for a denatured monomer. The global fit to far-UV CD data yields a value of Y_I of 0.39, suggesting that the monomeric intermediate retains significant helical structure. Fluorescence data give a Y_I value of 0.61. This indicates that changes in quaternary and tertiary structure on formation of the intermediate state lead to an increased exposure of tyrosine and tryptophan residues to solvent. The thermodynamic parameters, therefore, are consistent with a monomeric intermediate that has undergone partial unfolding, involving some loss of both secondary and tertiary structure.

Modelling the equilibrium unfolding of YibK

The thermodynamic parameters in Table 2 can be used to model the equilibrium unfolding of YibK at different concentrations of protein. Figure 6(a) and (b) shows the fraction of native dimer, monomeric intermediate and denatured protein present as a function of urea concentration for YibK concentrations of 100 μM and 0.25 μM, respectively. Figure 6 shows that at 100 μM YibK the intermediate state is hardly populated, while at 0.25 μM YibK it is present at fractions up to 0.14 at concentrations of urea close to the midpoint of the unfolding transition. The YibK denaturation profiles are monophasic within the protein concentration range used in this study, where the two unfolding transitions $N_2 \leftrightarrow 2I$ and $I \leftrightarrow D$ remain closely coupled. Figure 7 illustrates how the equilibrium unfolding curves are predicted to vary over a larger range of YibK concentrations. A biphasic profile becomes apparent at sub-nanomolar concentrations of protein; here, the protein concentration-

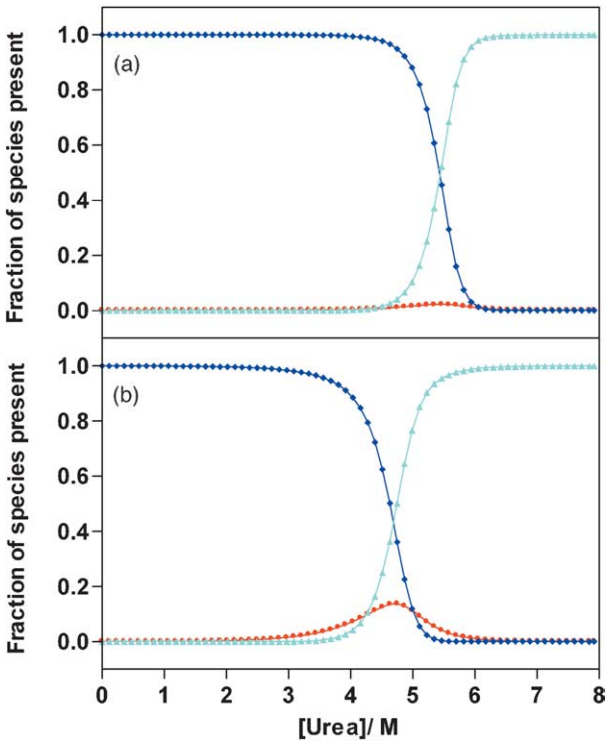

Figure 6. Modelling the fraction of native (N_2, dark blue diamonds), monomeric intermediate (I, red circles) and denatured (D, light blue triangles) species present as a function of the concentration of urea for (a) 100 μM YibK and (b) 0.25 μM YibK. Fractions are calculated using the thermodynamic parameters obtained from the global fit of YibK fluorescence unfolding data to a three-state denaturation model involving a monomeric intermediate (equation (16)).

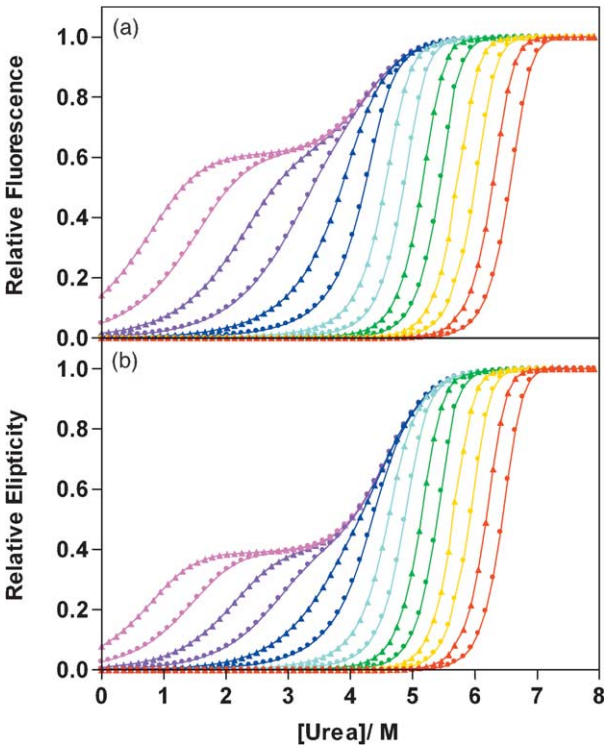

Figure 7. The theoretical YibK denaturation profiles for 0.1 pM (lilac triangles), 1 pM (pink circles), 10 pM (purple triangles), 0.1 nM (purple circles), 1 nM (dark blue triangles), 10 nM (dark blue circles), 0.1 μM (light blue triangles), 1 μM (light blue circles), 10 μM (green triangles), 0.1 mM (green circles), 1 mM (yellow triangles), 10 mM (yellow circles), 0.1 M (red triangles) and 1 M (red circles) protein modelled using the thermodynamic parameters obtained from the best fit of (a) fluorescence and (b) far-UV CD unfolding data to a three-state model with a monomeric intermediate (equation (16)).

dependent $N_2 \leftrightarrow 2I$ transition occurs at lower concentrations of denaturant than the protein concentration-independent $I \leftrightarrow D$ transition. The value of Y_I dictates the relative height of each transition, and so fluorescence and far-UV CD unfolding profiles no longer superimpose where a biphasic transition is seen (Figure 7).

Conclusions

Proteins belonging to the SPOUT class of knotted MTases represent an intriguing protein-folding problem; during the course of folding, a considerable segment of polypeptide chain has to thread through a loop. Here, we have established conditions under which YibK, a homodimeric SPOUT class member, can be unfolded reversibly using the chemical denaturant urea, thus demonstrating that molecular chaperones are not required for the efficient folding of this protein. A series of equilibrium unfolding experiments using secondary and tertiary structural probes demonstrate that the protein unfolds through a monomeric intermediate state. Using global fitting analysis, we have shown that this intermediate state is quite stable with respect to the denatured state, with a free energy of unfolding of 6.5–7.3 kcal mol^{-1}, comparable to many small monomeric proteins. The thermodynamic parameters obtained suggest that, in addition to the dissociation of the dimer, there has been a partial loss of secondary and tertiary structure of the monomer subunit on formation of the intermediate state. Whether the topological knot continues to exist in the intermediate state remains to be shown. The growing number of knotted structures deposited in the Protein Data Bank implies that knots in proteins may be more common then previously thought. Moreover, the results of this study show that, despite the presence of a deep trefoil knot, YibK undergoes equilibrium denaturation in the same manner as many unknotted dimers. Therefore, it appears that Nature has not only evolved mechanisms to successfully fold polypeptide chains, but in some cases, to knot them as well.

Materials and Methods

Materials

A plasmid containing the YibK gene inserted into a pET-17b vector was a kind gift from Dr Osnat Herzberg (University of Maryland Biotechnology Institute, MD). Chromatography columns and media were obtained from Amersham Biosciences, 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.

YibK expression and purification

C41 (DE3) cells containing the YibK plasmid were grown at 37 °C in 2×YT medium containing 100 μg ml^{-1} ampicillin until they reached mid-log phase. Expression was induced with IPTG to a final concentration of 0.1 mM, followed by continued incubation at 37 °C for four hours. Cells were harvested by centrifugation, resuspended in 20 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% (v/v) glycerol, 1 mM DTT, and lysed by sonication. The cell debris was removed by centrifugation, and the supernatant applied to a Q-Sepharose anion-exchange column. This was used to bind contaminants, while YibK appeared in the flow-through. The fractions enriched with YibK (as assessed by SDS-PAGE) were pooled, dialysed into 50 mM sodium acetate (pH 5.8), 125 mM KCl, 5% glycerol, 1 mM DTT and applied to an SP-Sepharose cation-exchange column. The column was eluted with a linear gradient of 0–0.5 M NaCl over 20 column volumes; YibK eluted at approximately 0.1 M NaCl. Fractions containing purified protein (as assessed by SDS-PAGE) were pooled and dialysed against 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% glycerol, 1 mM DTT. Protein yield was approximately 70 mg l^{-1}. Purity was assessed by SDS-PAGE and analytical size-exclusion chromatography. Protein was flash-frozen and stored at −80 °C until required. YibK concentration, in monomer

units, was verified spectrophotometrically using an extinction coefficient at 280 nm of 18,700 M^{-1} cm^{-1}, determined using the method of Gill & von Hippel.[42] Since oxidised DTT absorbs at 280 nm,[43] care was taken to ensure reference and sample buffers were identical.

Size-exclusion chromatography

SEC was performed on an ÄKTA FPLC system using a Superdex 75 10/300 GL analytical gel-filtration column equilibrated at 25 °C in 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% glycerol, 1 mM DTT. Pre-equilibrated samples of YibK (three hours) at concentrations between 0.25 μM and 100 μM were injected (100 μl), and the relative elution volume was compared to that of molecular mass (M) standards. The relative elution volume was calculated as:

$$K_{AV} = \frac{V_e - V_o}{V_g - V_o} \quad (1)$$

where V_e is the elution volume, V_o is the void volume determined by the elution of blue dextran 2000 (2000 kDa), and V_g is the geometric column volume. A standard curve was plotted of K_{AV} *versus* log (M). Molecular mass standards were bovine serum albumin (66.3 kDa), carbonic anhydrase (28.8 kDa) and cytochrome c (12.4 kDa).

Equilibrium unfolding and refolding experiments

All experiments were performed in a thermostatically controlled cuvette at 25 °C in 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% glycerol, 1 mM DTT. Aggregation of the native state prevents fully reversible unfolding at this pH without salt and glycerol as stabilising agents. For urea equilibrium denaturation curves, a stock solution of urea (approximately 9 M) in buffer was prepared in a volumetric flask and stored at −20 °C. The concentration of this stock was determined from its refractive index,[44] measured using an Atago 1T refractometer (Abbe). The stock urea solution was diluted with buffer using a Hamilton Microlab apparatus to give 800 μl aliquots of urea at various concentrations between 0 M and 9 M. For denaturation profiles, YibK stock solution (100 μl) at various concentrations of protein was added to each aliquot to give final concentrations of urea from 0 M to 8 M, and final concentrations of YibK from 0.25 μM to 100 μM (measured in monomer units). For renaturation profiles, YibK at various concentrations was left to unfold in 7.2 M urea. At this concentration of urea, unfolding is complete after 90 minutes. Unfolded protein (100 μl) was added to each aliquot to give final concentrations of urea from 0.8 M to 8.7 M. Samples were left for at least nine hours to equilibrate, after which no further change in spectroscopic signal was seen.

Spectroscopic measurements

Fluorescence measurements were taken with a SLM-Amico Bowman series 2 luminescence spectrometer using a 1 cm path-length cuvette. The excitation wavelength was 280 nm with a band pass of 4 nm for both excitation and emission. Emission spectra were taken between 300 nm and 355 nm at a scan rate of 1 nm s^{-1}. The largest difference in fluorescence between native and denatured YibK was observed at 319 nm, and emission at this wavelength was used in subsequent analysis. Denaturation profiles were measured for various YibK concentrations between 0.25 μM and 100 μM, with at least two curves recorded in all cases.

Far-UV CD spectra were acquired using a Jasco J-720 spectropolarimeter with an emission band pass of 2 nm. Scans were taken between 210 nm and 250 nm at a scan rate of 1 nm s^{-1}. The largest difference in signal between native and denatured YibK was observed at 225 nm, and this wavelength was used to monitor unfolding. Denaturation profiles were measured for YibK concentrations of 0.5–2.5 μM using a 1 cm cell, 5–10 μM using a 0.3 cm cell and 50–100 μM using a 0.1 cm cell, with at least two curves recorded in all cases.

Data analysis

Two-state dimer denaturation model

Equilibrium unfolding data were fit to a two-state dimer model, with native dimer (N_2) in equilibrium with unfolded monomer (D):

$$N_2 \overset{K_U}{\leftrightarrow} 2D \qquad \text{Scheme 1}$$

The equilibrium constant for concerted unfolding, K_U, is defined as $K_U = [D]^2/[N_2]$, and the total protein concentration in terms of monomer, P_t, is $P_t = 2[N_2] + [D]$. The fraction of monomers involved in native dimers, F_N, is given as $F_N = 1 - F_D$, where F_D is the fraction of unfolded monomers, and can be written as $F_D = [D]/P_t$. These equations can be combined to give F_D in terms of K_U and P_t:

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

K_U can be defined according to the linear free energy model, which states that the free energy of unfolding varies linearly with concentration of denaturant:[26,32]

$$\Delta G_U = -RT \ln(K_U) = \Delta G_{H_2O} - m[\text{denaturant}] \quad (3)$$

ΔG_{H_2O} is the free energy of unfolding in the absence of denaturant, and m is a constant of proportionality relating to the solvent exposure difference between native and denatured states.

In a two-state dimer system, $[D]_{50\%}$ is defined as the concentration of denaturant where the fraction of unfolded monomers equals the fraction of monomers involved in native dimers ($F_N = F_D = 0.5$). At this concentration of denaturant:

$$[D] = 0.5P_t \text{ and } 2[N_2] = 0.5P_t$$

Thus, at $[D]_{50\%}$:

$$K_U = \frac{[D]^2}{[N_2]} = \frac{0.25P_t^2}{0.25P_t} = P_t$$

This can be compared to monomeric two-state systems where at $[D]_{50\%}$, $K_U = [D]/[N] = 1$. Substituting this into equation (3) gives:

$$\Delta G_{H_2O} = -RT \ln(P_t) + m[D]_{50\%} \quad (4)$$

K_U now becomes:

$$K_U = \exp\left(\frac{RT \ln(P_t) - m([D]_{50\%} - [\text{denaturant}])}{RT}\right) \quad (5)$$

For ease of comparison, denaturation data are normalised to give the relative spectral signal, Y_{rel}:

$$Y_{rel} = \frac{(Y_0 - Y_N)}{(Y_D - Y_N)} \quad (6)$$

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 YibK monomeric subunits at a concentration of P_t, respectively. The terms Y_N and Y_D vary linearly with urea concentration;[45] therefore, $Y_N = \alpha_N + \beta_N$[denaturant] and $Y_D = \alpha_D + \beta_D$ [denaturant] where α_N and α_D are the intercepts, and β_N and β_D are the gradients of the native and denatured baselines, respectively. In the case of a two-state dimer system, Y_{rel} is equal to F_D, therefore:

$$Y_0 = Y_N(1 - F_D) + Y_D F_D \quad (7)$$

Fluorescence and far-UV CD datasets for each concentration of YibK were fit individually to equation (7), with F_D and K_U defined as in equations (2) and (5), respectively. Fits were performed using Prism, version 4 (GraphPad Software) to give thermodynamic parameters $m_{N_2 \leftrightarrow 2D}$, $[D]_{50\%}$; $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ was calculated from equation (4). The values obtained for Y_N and Y_D were then used to normalise the data according to equation (6).

Three-state dimer denaturation models

There are two possible three-state denaturation models, in which native dimer (N_2) is in equilibrium with either a dimeric (I_2) or monomeric (I) intermediate and unfolded monomer (2D):

$$N_2 \overset{K_1}{\leftrightarrow} I_2 \overset{K_2}{\leftrightarrow} 2D \qquad \text{Scheme 2}$$

$$N_2 \overset{K_1}{\leftrightarrow} 2I \overset{K_2}{\leftrightarrow} 2D \qquad \text{Scheme 3}$$

For a three-state model involving a dimeric intermediate (Scheme 2), the equilibrium constants for the first (K_1) and second (K_2) transitions can be defined, respectively, as $K_1 = [I_2]/[N_2]$ and $K_2 = [D]^2/[I_2]$. The total protein concentration, in terms of monomer, is $P_t = 2[N_2] + 2[I_2] + [D]$, and the sum of the fractions of individual species is equal to 1: $F_N + F_I + F_D = 1$, where F_I represents the fraction of monomeric subunits involved in the intermediate state. Combining these relationships gives:

$$K_1 = \frac{F_I}{F_N} \quad (8)$$

and

$$K_2 = \frac{2P_t F_D^2}{F_I} \quad (9)$$

By defining F_I and F_N solely in terms of F_D, K_1, K_2 and P_t, F_D can be expressed as a function of K_1, K_2 and P_t:

$$F_D = \frac{-K_1K_2 + \sqrt{(K_1K_2)^2 + 8(1 + K_1)(K_1K_2)P_t}}{4P_t(1 + K_1)} \quad (10)$$

In a three-state model the relative spectroscopic signal, obtained by normalising the data using equation (6) becomes:

$$Y_{rel} = Y_N F_N + Y_I F_I + Y_D F_D \quad (11)$$

Y_I is the spectroscopic signal of the intermediate. Substituting equations (8) and (9) into equation (11) gives the final fitting equation:

$$Y_{rel} = Y_N\left(\frac{2P_t F_D^2}{K_1K_2}\right) + Y_I\left(\frac{2P_t F_D^2}{K_2}\right) + Y_D(F_D) \quad (12)$$

For a three-state model involving a monomeric intermediate (Scheme 3), the equilibrium constants K_1 and K_2 are defined, respectively, as $K_1 = [I]^2/[N_2]$ and $K_2 = [D]/[I]$. The total protein concentration in terms of monomer is $P_t = 2[N_2] + [I] + [D]$, and again $F_N + F_I + F_D = 1$. Combining these relationships results in:

$$K_1 = \frac{2P_t F_I^2}{F_N} \quad (13)$$

$$K_2 = \frac{F_D}{F_I} \quad (14)$$

F_N and F_D can now be defined solely in terms of F_I, K_1, K_2 and P_t, and therefore F_I in terms of K_1, K_2 and P_t:

$$F_I = \frac{-K_1(1 + K_2) + \sqrt{K_1^2(1 + K_2)^2 + 8P_tK_1}}{4P_t} \quad (15)$$

The fitting equation is obtained by substituting equations (13) and (14) into equation (11):

$$Y_{rel} = Y_N\left(\frac{2P_t F_I^2}{K_1}\right) + Y_I(F_I) + Y_D(K_2F_I) \quad (16)$$

Global fitting to three-state dimer denaturation models

For both three-state models, the number of variables involved means that global analysis involving fitting the unfolding data for all YibK concentrations to the model simultaneously is necessary to achieve accurate thermodynamic parameters. Analysis was carried out using the non-linear, least-squares fitting programme Prism, version 4, and fluorescence and far-UV CD datasets were considered separately. For both three-state models, K_1 and K_2 are defined from equation (3) as:

$$K_1 = \exp\left(\frac{-\Delta G^1_{H_2O} + m_1[\text{denaturant}]}{RT}\right)$$

$$K_2 = \exp\left(\frac{-\Delta G^2_{H_2O} + m_2[\text{denaturant}]}{RT}\right)$$

For a three-state model involving a dimeric intermediate, normalised data were globally fit to equation (12), with F_D defined as in equation (10), and K_1 and K_2 defined as above. For a three-state model involving a monomeric intermediate, normalised data were globally fit to equation (16), with F_I defined as in equation (15), and again K_1 and K_2 defined as above. For both models, the thermodynamic parameters $\Delta G^1_{H_2O}$, $\Delta G^2_{H_2O}$ (stabilities of the first and second unfolding transitions, respectively), m_1, m_2 (the m-values for the first and second unfolding transitions, respectively) and Y_I (the spectral signal of the intermediate) were obtained; Y_I was assumed not to vary with denaturant concentration to minimise the number of parameters needed. To reduce any baseline artefacts, the baselines were constrained to zero slope, with Y_N and Y_D defined as zero and 1, respectively. To ensure the parameters calculated by the fitting programme represented a global minimum in the fitting procedure, initial estimates were varied individually by approximately $\pm 15\%$. A global minimum in the non-linear, least-squares fit was assumed if the parameters calculated remained the same, despite the change of initial estimates.

Data modelling

The three-state denaturation mechanism of YibK involving a monomeric intermediate was modelled

using the equations described above and $\Delta G^{N_2\leftrightarrow 2I}_{H_2O}$, $\Delta G^{I\leftrightarrow D}_{H_2O}$, $m_{N_2\leftrightarrow 2I}$, $m_{I\leftrightarrow D}$ and Y_I values determined from global analysis (Table 2). Denaturation profiles were modelled for various values of P_t using Microsoft Excel 2000, as was the change in the fraction of species present with denaturant concentration.

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

The SASA of native dimeric YibK was calculated from the coordinates of its X-ray crystal structure,[6] using the web-based program GETAREA version 1.1.[46] The SASA was also calculated for a fully folded YibK monomer subunit. The SASA of an unfolded monomer was estimated using values for individual residues obtained from tripeptide studies.[29] 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 YibK monomer was estimated using values obtained from hard-sphere simulations, termed the Upper Bound Model.[30,31]

It has been shown that the *m*-value of a protein is highly correlated to the ΔSASA between native and denatured states,[28] and the following relationship has been observed for proteins without crosslinks:

$$\text{Urea } m\text{-value} = \Delta\text{SASA}(0.14 \pm 0.01) \qquad (17)$$

This relationship, along with the ΔSASA calculated for the YibK unfolding transition, was used to estimate the *m*-value associated with complete YibK 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 ΔSASA upon dissociation is the difference in SASA between native dimer and two fully folded monomers.

Acknowledgements

A.L.M. holds an MRC PhD studentship. We thank Alexei Murzin for introducing us to knotted proteins, Dr Osnat Herzberg for donation of the YibK expression vector, and Andrew Brown and other members of the Jackson group for helpful discussions. We thank Professor Sir Alan Fersht, and Professor Chris Dobson for access to biophysical equipment. The work was funded, in part, by The Welton Foundation.

References

1. Daggett, V. & Fersht, A. R. (2003). The present view of the mechanism of protein folding. *Nature Rev. Mol. Cell. Biol.* **4**, 497–502.
2. Daggett, V. & Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28**, 18–25.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. Liang, C. Z. & Mislow, K. (1995). Topological features of protein structures - knots and links. *J. Am. Chem. Soc.* **117**, 4201–4213.
12. Liang, C. Z. & Mislow, K. (1994). Knots in proteins. *J. Am. Chem. Soc.* **116**, 11189–11190.
13. Mansfield, M. L. (1994). Are there knots in proteins? *Nature Struct. Biol.* **1**, 213–214.
14. Takusagawa, F. & Kamitori, S. (1996). A real knot in protein. *J. Am. Chem. Soc.* **118**, 8945–8946.
15. Mansfield, M. L. (1997). Fit to be tied. *Nature Struct. Biol.* **4**, 166–167.
16. Taylor, W. R. (2000). A deeply knotted protein structure and how it might fold. *Nature*, **406**, 916–919.
17. Taylor, W. R. & Lin, K. (2003). Protein knots: a tangled problem. *Nature*, **421**, 25.
18. 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.
19. Anantharaman, V., Koonin, E. V. & Aravind, L. (2002). SPOUT: a class of methyltransferases that includes SpoU and TrmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J. Mol. Microbiol. Biotechnol.* **4**, 71–75.
20. Finkelstein, A. V. & Badretdinov, A. (1997). Influence of chain knotting on the rate of folding. *Fold. Des.* **3**, 67–68.
21. Neet, K. E. & Timm, D. E. (1994). Conformational stability of dimeric proteins: quantitative studies by equilibrium denaturation. *Protein Sci.* **3**, 2167–2174.
22. Gloss, L. M. & Matthews, C. R. (1997). Urea and thermal equilibrium denaturation studies on the dimerization domain of *Escherichia coli* Trp repressor. *Biochemistry*, **36**, 5612–5623.
23. Hobart, S. A., Ilin, S., Moriarty, D. F., Osuna, R. & Colon,

W. (2002). Equilibrium denaturation studies of the *Escherichia coli* factor for inversion stimulation: implications for *in vivo* function. *Protein Sci.* **11**, 1671–1680.
24. 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.
25. Park, Y. C. & Bedouelle, H. (1998). Dimeric tyrosyl-tRNA synthetase from *Bacillus stearothermophilus* unfolds through a monomeric intermediate. A quantitative analysis under equilibrium conditions. *J. Biol. Chem.* **273**, 18052–18059.
26. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.
27. 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.
28. 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.
29. Miller, S., Janin, J., Lesk, A. M. & Chothia, C. (1987). Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641–656.
30. Creamer, T. P., Srinivasan, R. & Rose, G. D. (1995). Modeling unfolded states of peptides and proteins. *Biochemistry*, **34**, 16245–16250.
31. Creamer, T. P., Srinivasan, R. & Rose, G. D. (1997). Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry*, **36**, 2832–2835.
32. Tanford, C. (1970). Protein denaturation. Part C. Theoretical models for the mechanism of denaturation. *Advan. Protein Chem.* **24**, 1–95.
33. Clark, A. C., Sinclair, J. F. & Baldwin, T. O. (1993). Folding of bacterial luciferase involves a non-native heterodimeric intermediate in equilibrium with the native enzyme and the unfolded subunits. *J. Biol. Chem.* **268**, 10773–10779.
34. Doyle, S. M., Braswell, E. H. & Teschke, C. M. (2000). SecA folds *via* a dimeric intermediate. *Biochemistry*, **39**, 11667–11676.
35. Grimsley, J. K., Scholtz, J. M., Pace, C. N. & Wild, J. R. (1997). Organophosphorus hydrolase is a remarkably stable enzyme that unfolds through a homodimeric intermediate. *Biochemistry*, **36**, 14366–14374.
36. Hornby, J. A., Luo, J. K., Stevens, J. M., Wallace, L. A., Kaplan, W., Armstrong, R. N. & Dirr, H. W. (2000). Equilibrium folding of dimeric class mu glutathione transferases involves a stable monomeric intermediate. *Biochemistry*, **39**, 12336–12344.
37. McLaughlin, S. H. & Jackson, S. E. (2002). Folding and stability of the ligand-binding domain of the glucocorticoid receptor. *Protein Sci.* **11**, 1926–1936.
38. Spudich, G. & Marqusee, S. (2000). A change in the apparent *m* value reveals a populated intermediate under equilibrium conditions in *Escherichia coli* ribonuclease HI. *Biochemistry*, **39**, 11677–11683.
39. Gunasekaran, K., Tsai, C. J. & Nussinov, R. (2004). Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *J. Mol. Biol.* **341**, 1327–1341.
40. Gekko, K. & Timasheff, S. N. (1981). Mechanism of protein stabilization by glycerol: preferential hydration in glycerol–water mixtures. *Biochemistry*, **20**, 4667–4676.
41. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, 81–91.
42. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.
43. Cleland, W. W. (1964). Dithiothreitol, a new protective reagent for SH groups. *Biochemistry*, **35**, 480–482.
44. Pace, C. N., Shirley, B. A. & Thomson, J. A. (1990). Measuring the conformational stability of a protein. In *Protein Structure. A Practical Approach* (Creighton, T. E., ed.), pp. 311–330, Oxford University Press, Oxford.
45. Schmid, F. X. (1990). Spectral methods of characterizing protein conformation and conformational changes. In *Protein Structure. A Practical Approach* (Creighton, T. E., ed.), pp. 251–285, Oxford University Press, Oxford.
46. 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.
47. Carson, M. (1997). Ribbons. *Methods Enzymol.* **277**, 493–505.

Edited by F. Schmid

(Received 4 November 2004; received in revised form 20 December 2004; accepted 23 December 2004)