Exploring knotting mechanisms in protein folding

Anna L. Mallam, Elizabeth R. Morris, and Sophie E. Jackson[1]

University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, England

Edited by S. Walter Englander, University of Pennsylvania School of Medicine, Philadelphia, PA, and approved September 23, 2008 (received for review July 10, 2008)

One of the most striking topological features to be found in a protein is that of a distinct knot formed by the path of the polypeptide backbone. Such knotted structures represent some of the smallest ''self-tying'' knots observed in Nature. Proteins containing a knot deep within their structure add an extra complication to the already challenging protein-folding problem; it is not obvious how, during the process of folding, a substantial length of polypeptide chain manages to spontaneously thread itself through a loop. Here, we probe the folding mechanism of YibK, a homodimeric α/β-knot protein containing a deep trefoil knot at its carboxy terminus. By analyzing the effect of mutations made in the knotted region of the protein we show that the native structure in this area remains undeveloped until very late in the folding reaction. Single-site destabilizing mutations made in the knot structure significantly affect only the folding kinetics of a late-forming intermediate and the slow dimerization step. Furthermore, we find evidence to suggest that the heterogeneity observed in the denatured state is not caused by isomerization of the single *cis* proline bond as previously thought, but instead could be a result of the knotting mechanism. These results allow us to propose a folding model for YibK where the threading of the polypeptide chain and the formation of native structure in the knotted region of the protein occur independently as successive events.

intermediate states | methyltransferases | parallel pathways | knotted proteins | trefoil knot

The α/β-knot methyltransferases (MTases) are a family of homodimeric proteins that exhibit an unusual trefoil knot deep within in their native structure (1–9). Such knots are particularly impressive because they are defined by the path of the polypeptide backbone alone and therefore require that a considerable segment of protein chain (at least 40 residues) has threaded through a loop. The question of how such complex topologies arise during protein folding is an intriguing one, and is of growing importance with the increasing number and complexity of knotted structures observed (10–14). In addition to trefoil knots, a highly intricate figure-of-eight knot (10) and a knotted structure with 5 projected crossings (15) have been observed. Consequently, when contemplating how a given polypeptide chain might fold, the possibility that it might knot must also be considered; if a global solution to the protein-folding problem is to be found, the puzzle of how such knotted structures form must be solved.

We have addressed this question by using the 160-residue homodimer YibK, one of the smallest α/β-knot proteins observed to date. YibK contains a deep trefoil knot at the C terminus of its structure where at least 40 amino acid residues have passed through a similarly sized loop (Fig. 1*A*) (7). Wild-type YibK folds with a complex mechanism involving 4 reversible kinetic phases (Fig. 2). Two different intermediates formed in parallel (I_1 and I_2) fold via a third, sequential monomeric intermediate (I_3) to produce native dimer (N_2) in a slow, rate-limiting step (16). The exact mechanism of knot formation, however, is still not clear. The early intermediates in the folding reaction, I_1 and I_2, have considerable structure as determined by their *m*-values, a measure of the solvent-accessible surface area change on folding (17, 18). Knotting would presumably be more straightforward before or during the folding of these species, because after their formation there is less unstructured protein to allow for easy threading of the polypeptide chain through a loop. Strong evidence for this comes from studies involving the fusion of a ''molecular plug'' in the form of another small protein to either the amino terminus, carboxy terminus, or both termini of YibK (19); this modification does not appear to hinder the threading motions of the polypeptide chain. It has therefore been proposed that knotting is not rate limiting during folding, but instead a loose knot is formed in a denatured-like state. However, studies on monomeric variants of YibK indicate that the knotted region of the protein is not completely folded in these mutants, suggesting that this part of the native structure only fully forms on dimerization (20).

Here, we have probed the folding pathway of YibK by using single-site mutants. Protein-engineering techniques have been used extensively on small, model protein systems to obtain comprehensive, near-atomic resolution information on their folding pathways (21, 22). We use an equivalent method to examine the folding of YibK, a significantly more complex system, to obtain residue-specific detail about the folding mechanism of a knotted protein. Substitution of the single *cis* proline residue in the protein has a negligible effect on the folding reaction, suggesting that the complicated folding kinetics are caused by some other phenomenon. Additionally we have analyzed the effect of mutations made in the knotted region of the structure to inform on its formation along the folding pathway. Measurement of thermodynamic and kinetic parameters has allowed us to establish the role of the mutated residues during folding and we find that only the slowest folding phases are affected. Consequently, whereas our previous work indicated that the polypeptide chain threads early in the folding of a knotted protein, here we propose a mechanism where the native knotted region of YibK does not become structured until late in the folding reaction. We conclude that knotting and folding are therefore effectively independent events.

Results and Discussion

Destabilizing Mutations in the Knotted Core of YibK Do not Disrupt the Knot Structure. The knotted region of YibK is defined as the ''knotted chain'' (residues 121–160) and the knotting loop (residues 81–120) (Fig. 1*A*) (4, 7). The core of the knot structure consists of 3 β-strands: β4, β5, and β6 (Fig. 1*C*). Single-site mutations to alanine were made at 4 positions within this region to give the mutant proteins L75A, F76A, L78A, and Y96A (Fig. 1*B*); these mutations delete side chains that make considerable interactions within the knotted core [supporting information (SI) Table S1]. Urea denaturation profiles for mutant proteins, as monitored by the change in fluorescence signal at 319 nm, all

Author contributions: A.L.M. and S.E.J. designed research; A.L.M. and E.R.M. performed research; A.L.M. analyzed data; and A.L.M. and S.E.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: sej13@cam.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0806697105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

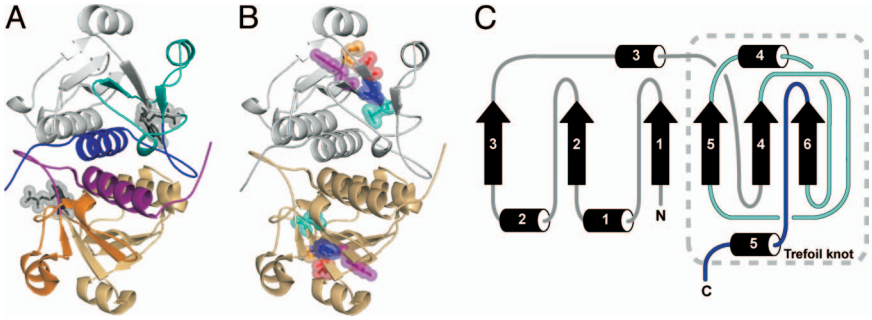

Fig. 1. The structure of dimeric YibK from *Haemophilus influenzae*. (*A*) Ribbon diagram colored to highlight the deep trefoil knot at the C terminus (PDB ID code 1MXI). The knotting loop is cyan or orange (residues 81–120), and the knotted chain appears blue or purple (residues 121–160). The cofactor binding site is also shown with AdoHcy molecules displayed as ball-and-stick models. The crystal structure contains 1 AdoHcy binding site per monomer, located in the knotted region of the protein. (*B*) Residues targeted by mutagenesis in this study. Leu-75 (red), Phe-76 (blue), Leu-78 (cyan), and Tyr-96 (purple) are located in the knotted region of the protein, and Pro-34 (orange) is in a loop. (*C*) Topological diagram of YibK. Numbers refer to secondary structure elements and the structure is colored as in *A*.

display a midpoint of unfolding at a notably lower concentration of denaturant to wild-type YibK for the same concentration of protein, indicating that the mutation has been significantly destabilizing (Fig. 3). As expected for a dimer system, because of the coupled denaturation and dissociation reactions that occur during unfolding, mutant denaturation curves are protein concentration dependent (23). An increase in the concentration of urea for the unfolding midpoint is seen with increasing concentration of protein. We analyzed mutant equilibrium unfolding data for each concentration of protein by using a 2-state dimer denaturation model (Fig. 3 and Table S2) (23). In general, both

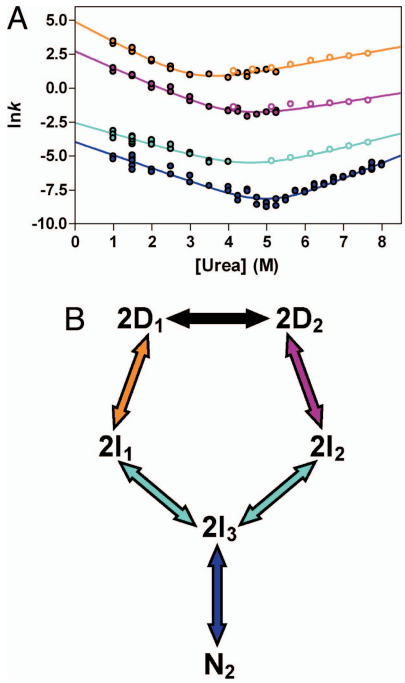

Fig. 2. The folding mechanism of wild-type YibK. (*A*) A V-shaped plot of the natural logarithm of rate constants observed during folding and unfolding at various concentrations of urea (16). Rate constants from single-jump and double-jump experiments are represented by filled and open circles, respectively. Phases are orange, purple, and cyan in order from fastest to slowest, respectively, and the phase that corresponds to dimerization is shown in blue. Continuous lines represent the fit of each phase to a 2-state model (Eq. **2**). (*B*) The folding mechanism proposed for wild-type YibK dimer (16). Arrows are colored according to their corresponding phase in *A*.

$\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$, the free-energy difference between 1 mol of dimer and 2 mol of unfolded monomer in the absence of denaturant, and the slope of the unfolding transition ($m_{N_2 \leftrightarrow 2D}$) for the mutants increase with increasing concentration of protein. Similar to what has been observed for wild-type YibK, this indicates that the 2-state model is not adequately describing the data and is consistent with mutants unfolding with a 3-state denaturation mechanism involving a monomeric intermediate (23). The complexity of this model results in a large number of variable parameters, and accurate $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$, values for YibK mutants are more readily obtained from analysis of kinetic folding and unfolding data.

α/β-knotted proteins such as YibK bind the MTase cofactor *S*-adenosylmethionine (AdoMet) in a binding crevice formed by the knotted region of the protein (Fig. 1*A*) (3, 7, 9). It has been shown that the binding affinity for *S*-adenosylhomocysteine (AdoHcy), the product of AdoMet after methyl-group transfer to the substrate has taken place, can be used to confirm the integrity of the YibK cofactor binding site and therefore the presence of the native knotted structure in the protein (19, 20). Binding of AdoHcy was measured by using isothermal titration calorimetry (ITC) for all of the mutants to verify that the knotted core remains intact; all mutants displayed a similar affinity for AdoHcy as wild-type YibK (Fig. S1 and Table S3). Additionally, analytical size-exclusion chromatography (SEC) was used to confirm the oligomeric state of the mutant proteins. Like wild-type YibK, all mutants elute as a single peak consistent with

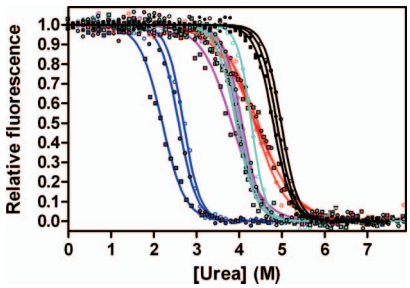

Fig. 3. Mutations in the knotted core affect the midpoint of urea-induced unfolding. Normalized fluorescence denaturation profiles for L75A (red), F76A (blue), L78A (cyan), and Y96A (purple) compared with wild-type protein (black) at 2.5 μM (open circles), 1 μM (filled circles), and 0.5 μM (squares) protein to demonstrate the relative midpoints of the unfolding transitions. The fits to a 2-state dimer denaturation model are shown as a guide for the eye.

a homogeneous dimer population at the concentration of protein studied (Fig. S2). This provides further evidence that the knot structure is correctly folded in the mutants as it forms part of the dimer interface (Fig. 1*A*). Thus, despite being notably destabilized, all of the mutants studied remain homodimeric and knotted.

Mutations in the Knotted Core Predominantly Affect Late Folding Events and Dimerization. The folding of wild-type YibK has been studied previously and the protein displays complex urea concentration-dependent kinetics (Fig. 2*A*) (16). In single-jump experiments, refolding is best described by a first-order reaction with 4 exponentials, whereas the unfolding reaction shows slow, single first-order exponential behavior. Double-jump unfolding experiments, where refolding is allowed for short amounts of time before unfolding is initiated to various final concentrations of urea, have been used to detect faster unfolding phases corresponding to the nonnative species populated on the refolding pathway (Fig. 2*A*) (16). The urea concentration dependence of the natural logarithm of the unfolding and refolding rate constants has been determined to give a V-shaped plot for each of the 4 reversible phases (Fig. 2*A*). Furthermore, the time course of the population of different intermediates along the wild-type folding pathway has been mapped and a mechanism most consistent with all of the kinetic data proposed (Fig. 2*B*) (16, 20). Two fast-forming intermediates, I_1 and I_2, are thought to fold via parallel folding channels and kinetic analysis over a wide range of conditions indicates that they are structurally distinct (16). Additionally, the population of molecules folding to either I_1 and I_2 was found to depend on the period that the protein was left unfolded, suggesting that the parallel folding routes are caused by heterogeneity developing in the denatured state. Here, we used similar single- and double-jump experiments to characterize the folding kinetics of the mutant proteins. All mutants showed 4 refolding and unfolding phases similar to the wild-type protein, indicating that they are likely folding with the same mechanism (Fig. S3 and Fig. S4). This is comparable to monomeric variants of YibK that contain significantly more disruptive mutations than those used here and display kinetic behavior similar to wild-type protein (20). Because folding transients for YibK can contain up to 4 exponential reactions occurring over different timescales, kinetic rate constants extracted from fluorescence folding data can be inadvertently affected by the analysis method; for example, factors such as the timescale over which the reaction is monitored, the number of exponentials used to fit the data, and the relative rates of each phase at a particular concentration of urea can all affect the precision with which the final rate constants are obtained. To circumvent this problem, and to minimize any disparity in the analysis that might affect the accuracy of the parameters calculated, we globally evaluated the kinetic data. Kinetic transients at different concentrations of urea for each refolding or unfolding phase were considered together in an identical manner for each mutant by using Eq. **1** (Fig. S5). Only traces under conditions where either the refolding or unfolding reactions were dominant were used so that a linear response of the natural logarithm of the rate constant to urea concentration could be assumed (24, 25). Additionally, to minimize the number of variable parameters involved in the analysis, mutant *m*-values for each phase were fixed to those of wild-type YibK. This parameter relates to the solvent-accessible surface area change associated with each phase and should not alter significantly on mutation (18, 24, 25). Unfolding and refolding rate constants in the absence of denaturant, $k_f^{H_2O}$ and $k_u^{H_2O}$, respectively, for each phase were obtained by using this approach (see *Materials and Methods*) and V-shaped plots of the urea concentration dependence of the natural logarithm of the rate constants were obtained by using Eq. **2** (Fig. 4 and Table 1). The validity of the global analysis was confirmed by the good agreement of the V-shaped plots obtained by using this method for wild-type YibK with those measured previously by considering all kinetic transients separately (Fig. 4). In this way, we can confidently attribute any differences in mutant $k_f^{H_2O}$ or $k_u^{H_2O}$ values to the effect of the mutation rather than to a variation in the analysis method.

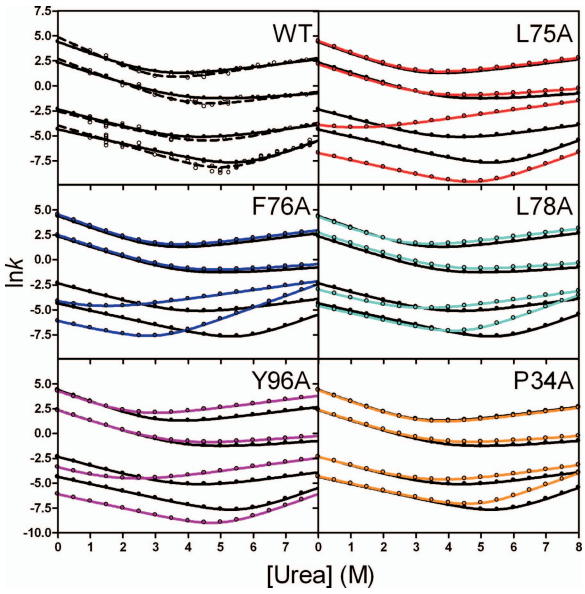

Fig. 4. Global analysis of urea kinetics of mutant proteins at 1 μM protein. V-shaped plots calculated from Eq. **2** of the natural logarithm of rate constants observed during folding and unfolding at various concentrations of urea for L75A (red), F76A (blue), L78A (cyan), Y96A (purple), and P34A (orange) compared with wild-type protein (black filled circles). Folding and unfolding data were measured by using fluorescence and analyzed as described in *Materials and Methods*. V-shaped plots calculated previously for wild-type YibK by considering all kinetic data separately are included for comparison (black open circles) (16).

The total destabilization energy, $\Delta\Delta G_{H_2O}^{N_2 \leftrightarrow 2D}$, for the mutants in the knotted region of YibK varies between 2.5 and 5.4 kcal mol^{-1} (Table 1). To assess the relative destabilization effect of a particular mutation on each kinetic intermediate, we calculated the parameter Φ_n from the ratio of the destabilization energy observed for each phase to the total destabilization energy (Table 1). Large values of Φ_n on mutation suggest that the wild-type side chain at that position makes significant interactions in the corresponding intermediate. It is clear from mutant V-shaped plots and Φ_n values that single-site mutations in the knotted region of YibK primarily affect the slower folding phases that correspond to the formation of the late obligatory monomeric intermediate I_3 (Φ_3 varies between 0.33 and 0.69) and its subsequent dimerization to fully folded YibK homodimer (Φ_4 varies between 0.19 and 0.53); the sum of Φ_n values for both these phases ($\Phi_3 + \Phi_4$) is between 0.67 and 0.94 for all knotted core mutants (Fig. 4 and Table 1). It follows that these residues develop the majority of their side-chain interactions only late on in the folding reaction. Consequently, we conclude that the knotted region of YibK remains largely unstructured until the last stages of folding.

Intermediate Species Formed in Parallel During the Folding of YibK Are Unlikely to be Related to Proline Isomerization. YibK contains 10 proline residues with one, Pro-34, adopting a *cis* conformation in the native structure. The apparent fast parallel folding channels observed during the folding of wild-type YibK (to the

Table 1. Summary of the parameters obtained from the global analysis of mutant kinetic folding data

Phase	Color in Fig. 2	Protein	$k_f^{H_2O}$, s^{-1}	$k_u^{H_2O}$, s^{-1}	m_{kf}, kcal mol^{-1} M^{-1}	m_{ku}, kcal mol^{-1} M^{-1}	m_{kin}, kcal mol^{-1} M^{-1}	$\Delta G^{kin}_{H_2O}$, kcal mol^{-1}	$\Delta\Delta G^{kin}_{H_2O}$, kcal mol^{-1}	$\Delta\Delta G^{kin,total}_{H_2O}$, kcal mol^{-1}	Φ_n
1	Orange	Wild-type	81 ± 3	0.60 ± 0.01	0.69	0.24	0.93	2.9 ± 0.04	—	—	—
		L75A	90 ± 1	0.69 ± 0.03				2.9 ± 0.04	0.03 ± 0.06	—	0.01 ± 0.02
		F76A	94 ± 1	0.80 ± 0.03				2.8 ± 0.04	0.08 ± 0.06	—	0.02 ± 0.01
		L78A	73 ± 1	0.93 ± 0.01				2.6 ± 0.02	0.32 ± 0.04	—	0.13 ± 0.02
		Y96A	71 ± 1	1.83 ± 0.01				2.2 ± 0.02	0.74 ± 0.04	—	0.23 ± 0.02
		P34A	78 ± 1	0.54 ± 0.01				2.9 ± 0.02	−0.03 ± 0.04	—	−0.02 ± 0.04
2	Purple	Wild-type	10 ± 0.9	0.11 ± 0.01	0.62	0.11	0.73	2.7 ± 0.13	—	—	—
		L75A	7 ± 0.2	0.19 ± 0.01				2.3 ± 0.06	0.44 ± 0.14	—	0.12 ± 0.04
		F76A	12 ± 0.1	0.18 ± 0.01				2.5 ± 0.06	0.23 ± 0.14	—	0.04 ± 0.03
		L78A	15 ± 0.2	0.17 ± 0.01				2.7 ± 0.06	0.06 ± 0.14	—	0.02 ± 0.06
		Y96A	10 ± 0.1	0.19 ± 0.01				2.4 ± 0.06	0.34 ± 0.14	—	0.10 ± 0.04
		P34A	10 ± 0.1	0.12 ± 0.01				2.7 ± 0.09	0.03 ± 0.15	—	0.02 ± 0.1
3	Cyan	Wild-type	(9.5 ± 1.4) × 10^{-2}	(5.6 ± 0.1) × 10^{-4}	0.52	0.27	0.79	3.0 ± 0.15	—	—	—
		L75A	(1.4 ± 0.02) × 10^{-2}	(6.3 ± 0.4) × 10^{-3}				0.49 ± 0.07	2.54 ± 0.16	—	0.69 ± 0.06
		F76A	(1.3 ± 0.01) × 10^{-2}	(3.3 ± 0.1) × 10^{-3}				0.79 ± 0.03	2.25 ± 0.15	—	0.42 ± 0.03
		L78A	(5.0 ± 0.01) × 10^{-2}	(1.2 ± 0.3) × 10^{-3}				2.2 ± 0.24	0.84 ± 0.29	—	0.33 ± 0.12
		Y96A	(3.1 ± 0.02) × 10^{-2}	(2.4 ± 0.1) × 10^{-3}				1.5 ± 0.04	1.52 ± 0.15	—	0.47 ± 0.06
		P34A	(9.6 ± 0.01) × 10^{-2}	(1.2 ± 0.05) × 10^{-3}				2.6 ± 0.04	0.42 ± 0.15	—	0.33 ± 0.14
4	Blue	Wild-type	(1.3 ± 0.04) × 10^{-2}	(3.4 ± 0.03) × 10^{-7}	0.42	0.70	1.12	14.0 ± 0.03	—	—	—
		L75A	(1.2 ± 0.03) × 10^{-3}	(1.0 ± 0.01) × 10^{-7}				13.3 ± 0.03	0.70 ± 0.04	3.7 ± 0.2	0.19 ± 0.02
		F76A	(2.2 ± 0.01) × 10^{-3}	(7.2 ± 0.01) × 10^{-6}				11.2 ± 0.01	2.85 ± 0.03	5.4 ± 0.2	0.53 ± 0.02
		L78A	(9.7 ± 0.1) × 10^{-3}	(2.4 ± 0.01) × 10^{-6}				12.7 ± 0.01	1.31 ± 0.03	2.5 ± 0.3	0.52 ± 0.07
		Y96A	(2.2 ± 0.07) × 10^{-3}	(1.8 ± 0.01) × 10^{-7}				13.4 ± 0.03	0.67 ± 0.05	3.3 ± 0.2	0.20 ± 0.02
		P34A	(1.3 ± 0.08) × 10^{-2}	(1.5 ± 0.01) × 10^{-6}				13.2 ± 0.06	0.84 ± 0.07	1.3 ± 0.2	0.67 ± 0.13

Analyses were performed with Prism, version 4 (GraphPad Software). Errors quoted are the standard errors calculated by the fitting program. $k_f^{H_2O}$ and $k_u^{H_2O}$ are the rate constants for refolding and unfolding, respectively, in the absence of denaturant. m_{kf} and m_{ku} are the kinetic refolding and unfolding *m*-values, respectively. *m*-values for mutants were fixed to those calculated for wild-type protein, which had errors of approximately ± 5%; $m_{kin} = m_{kf} + m_{ku}$. $\Delta G^{kin}_{H_2O} = -RT \ln(k_u^{H_2O}/k_f^{H_2O})$, except for phase 4 where $\Delta G^{kin}_{H_2O} = -RT \ln(2k_u^{H_2O}/k_{2nd}^{H_2O})$; $k_{f,app}^{H_2O} = P_t k_{2nd}^{H_2O}$ for a dimerization reaction, where P_t is the concentration of protein.

$\Delta\Delta G^{kin}_{H_2O} = \Delta G^{kin}_{H_2O}$ (wild-type) − $\Delta G^{kin}_{H_2O}$ (mutant)

$\Delta\Delta G^{kin,total}_{H_2O} = \Delta\Delta G^{kin}_{H_2O}$ (phase 1) + $\Delta\Delta G^{kin}_{H_2O}$ (phase 2) + $\Delta\Delta G^{kin}_{H_2O}$ (phase 3) + $\Delta\Delta G^{kin}_{H_2O}$ (phase 4)

$\Phi_n = \Delta\Delta G^{kin}_{H_2O}/\Delta\Delta G^{kin,total}_{H_2O}$.

intermediates I_1 and I_2 in Fig. 2*B*) were thought to arise from heterogeneity in the denatured state from the isomerization of this *cis* proline residue (16). On unfolding, the *cis* peptidyl-prolyl bond will isomerize to the energetically more favorable *trans* conformation in the majority of molecules, resulting in a predominantly denatured state ensemble with nonnative-like proline isomers (26, 27). Here, we have investigated the role of Pro-34 in the folding of YibK by examining the kinetics of the mutant P34A. Pro-34 is located in a loop region of the protein and is relatively solvent exposed (Fig. 1*B* and Table S1). P34A displays similar kinetics to wild-type YibK and refolding traces are best described by a first-order reaction with 4 exponentials (Fig. S6). Kinetic transients for P34A at different concentrations of urea were measured and globally analyzed in the same way as for the knotted core mutants (Fig. 4 and Table 1). In contrast to what is observed for many proline-related folding phases on substitution of the *cis* proline residue (28–30), no refolding phase was eliminated in the kinetics of P34A. This suggests that the heterogeneity in the denatured state causing the fast parallel reactions remains. Furthermore, the mutation has no effect on the refolding and unfolding rate constants (Table 1) or on the relative amplitudes of the folding phases corresponding to the species forming in parallel, which vary between 0.35 and 0.65 as a fraction of the total fluorescence amplitude change associated with their folding (Fig. S6). This makes it unlikely that P34A has simply retained a peptide bond in the *cis* conformation that still isomerizes once unfolded, as a change in folding rate and/or relative amplitudes of the parallel channels would be expected in this case (31). Instead, as with the knotted core mutants, only the 2 slowest folding phases are affected by the Pro-Ala mutation (Fig. 4 and Table 1). These results therefore suggest that the *cis* proline at this position is not the cause of the apparent parallel folding pathways seen in wild-type YibK. Although less likely to have such a significant effect, it is possible that the heterogeneity observed in the denatured state of YibK is the result of isomerization of 1 or more of the other proline residues that adopt a *trans* conformation in the native protein. However, the folding characteristics of I_1 and I_2 are not consistent with the proline-limited folding reactions observed in other proteins. First, both I_1 and I_2 fold in the absence of denaturant with a rate constant that is much larger than that observed for proline isomerization reactions (32). Furthermore, pH double-jump experiments performed on wild-type YibK indicated that the population of molecules folding via I_1 increases from zero with increasing time spent in the denatured state with a rate constant of 0.05 s^{-1}, whereas the population of molecules folding via I_2 decreases (16). Because the folding rate constant of I_1 is much higher than that of I_2, and larger than the rate constant of 0.05 s^{-1} observed for the heterogeneity to develop in the acid unfolded state of YibK, this implies that nonnative-like isomers in the denatured state block folding to I_2 and cause faster folding to a structurally different intermediate I_1. This is very different from what has been observed for a number of proteins where nonnative-like proline isomers in the denatured state cause a slower parallel reaction to the same species, which is limited by a proline isomerization event (32). There is evidence to suggest, therefore, that the apparent parallel channels seen during the folding of YibK are the result of heterogeneity in the denatured state caused by a phenomenon other than proline isomerization.

Mechanism for the Knotting and Folding of YibK. The mutational studies presented here allow us to propose a preliminary model

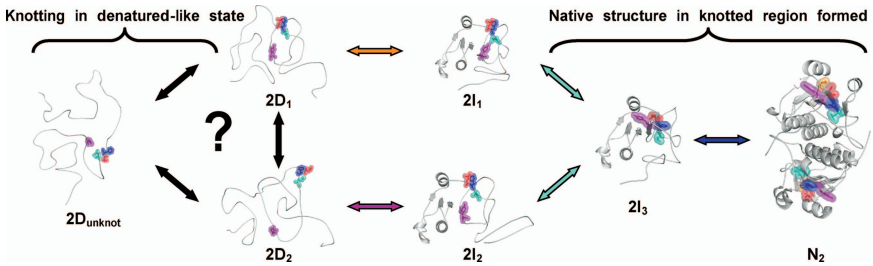

Fig. 5. A model for the folding mechanism of YibK to demonstrate the independent knotting and folding events. Heterogeneous loosely knotted conformations in a denatured-like state (D_1 and D_2) fold via parallel channels to form the intermediates I_1 and I_2. The knotted region remains relatively unstructured until it forms during the folding of I_3 and the subsequent dimerization of the protein. The exact nature of the heterogeneity in the denatured state leading to the parallel folding channels and the structure of the intermediate species remains unknown, and so the representations shown here are for illustrative purposes only. Arrows are colored to match those in Fig. 2.

for the knotting and folding of YibK that incorporates residue-specific detail (Fig. 5). Kinetic analysis of mutations made in the knotted region indicates that the native structure in this area does not develop until late on in the folding reaction and is not fully formed until the protein dimerizes in a slow association step. Earlier studies on an engineered monomeric version of YibK, which we demonstrated to be an excellent model for the folding intermediate I_3, are consistent with this; the YibK monomer displayed no binding affinity for AdoHcy, implying that the cofactor binding site, and therefore the knotted region of the protein, becomes only completely structured in the dimeric state of the protein (20).

We have shown that destabilizing mutations in the knotted region of the protein do not significantly affect the kinetics of fast-forming intermediates I_1 and I_2. This indicates that the structure in this part of the protein remains relatively unfolded in these intermediate states (Fig. 5). However, as noted in previous studies on YibK, these species have significant structure (at least 60% of a fully folded monomer, as estimated from kinetic *m*-values). We have therefore proposed that I_1 and I_2 are unlikely to contain a knotting loop large enough for the polypeptide chain to easily thread through and that it is plausible that the protein chain is already knotted in these intermediates (19). Consequently, our results suggest that threading of the polypeptide chain (at some stage before or during the folding of I_1 and I_2) and formation of the native structure in the knotted region of the protein (primarily during folding of I_3 and N_2) are not concurrent events; knotting appears to precede folding. This conclusion can be taken together with previous studies on engineered knotted fusion proteins where we observed that YibK is still able to knot and fold despite the presence of another small folded protein attached to both termini (19). The fusion of these extra domains did not affect the folding rate of the knotted domain. To allow for the threading of the extra domains on the termini of YibK, we concluded that loose knotting of the polypeptide chain most likely occurs in a denatured-like state (19). The biophysical data presented here are consistent with this, and taken together the results suggest a folding model for YibK where threading of the polypeptide chain into a knot occurs early in folding, but formation of the native structure in the knotted region of the protein happens late and is relatively slow (Fig. 5). This provides a basis for further research to evaluate such a mechanism.

Our data show that proline isomerization is unlikely to be the cause of the apparent fast parallel folding reactions that result from heterogeneity in the denatured state of YibK. Instead, we propose that the heterogeneity causing these parallel folding channels is a consequence of the polypeptide chain adopting several, distinct, loosely knotted conformations in the denatured state (Fig. 5); these fold at different rates to form the intermediates I_1 and I_2. Consistent with this idea are simulated and physical experiments that demonstrate that long flexible strings similar to unfolded protein chains are almost certain to become knotted (33, 34). It is interesting to note that α/β-knot MTases such as YibK adopt a right-handed trefoil knot and so threading in only 1 direction would lead to productive knotting. It remains to be determined whether the knot ever becomes completely untied in the chemically denatured state of a knotted protein and therefore, the relevance of in vitro folding studies to the formation of knotted protein structures from a nascent polypeptide chain. However, studies to establish the existence of such denatured knotted conformers and the mechanisms involved in their interconversion, which could be via an unknotted state as suggested previously (19) (Fig. 5), remain an exciting prospect for future research.

The atypical folding pathway for YibK suggested here represents an interesting addition to the variety of mechanisms by which small proteins have been observed to fold to their native structures (35, 36). Perhaps unexpectedly, although threading of the polypeptide chain appears fast and not rate limiting in the folding reaction (19), development of the native structure in the knotted region of YibK is slow. It will be interesting to see whether such a mechanism is common to knotted proteins in general, because initial studies suggest that other α/β-knot MTases fold in a comparable manner (19, 37). An early threading event may be the defining feature of a polypeptide-knotting mechanism, with the ensuing folding occurring in a similar fashion to unknotted proteins; the folding of a knotted protein differs only with an initial knotting event in a denatured-like state. Consequently, to predict whether a given polypeptide chain will fold to a knotted structure it may be necessary to focus on early folding events rather than native structure formation.

Materials and Methods

Protein Purification. Point mutations were introduced into the gene encoding for YibK wild-type protein by using the Quick Change site-directed mutagenesis kit (Stratagene). Mutant proteins were expressed and purified as described for wild-type YibK (23). The identity of mutants was confirmed by DNA sequencing and mass spectrometry. All experiments were performed in a buffer of 50 mM Tris·HCl (pH 7.5), 200 mM KCl, 10% glycerol (vol/vol), 1 mM DTT, except for ITC experiments where β-mercaptoethanol replaced DTT as the reducing agent. All protein concentrations are in monomer units.

Protein Characterization. Analytical gel-filtration chromatography, ITC, and thermodynamic and kinetic folding experiments were carried out as described in refs. 16, 20, and 23. Analysis of urea denaturation profiles was performed as in ref. 23. A global analysis of the kinetic traces for each mutant was undertaken with refolding or unfolding kinetic transients at different concentrations of urea for each observable phase considered together and fit to a single equation. This assumes a linear dependence of the natural logarithm of the folding and unfolding rate constants on urea concentration, an assumption shown to be valid by analysis of separate refolding traces that display no rollover (data not shown) (24, 25). The exponential reaction for

each observable phase at different concentrations of urea outside the transition region were globally fit to:

$$Y(t) = Y_{\text{native}} + Y_1 \exp(-k_{\text{obs}} t), \quad [1]$$

where k_{obs} is equal to $k_f^{H_2O}$ exp($-m_{kf}$ [urea]) or $k_u^{H_2O}$ exp(m_{ku} [urea]) for refolding and unfolding reactions, respectively, and Y_1 is the corresponding fluorescence amplitude change. The parameters $k_f^{H_2O}$ and $k_u^{H_2O}$, the refolding and unfolding rate constants in the absence of denaturant, respectively, for each phase were shared throughout all datasets, and m_{kf} and m_{ku} values for mutants, a measure of the response of the folding or unfolding rate constants, respectively, to changing urea concentration, were fixed to wild-type values. V-shaped plots were constructed by using the equation (38):

$$\ln k_{\text{obs}} = \ln(k_f^{H_2O} \exp(-m_{k_f}[\text{urea}]) + k_u^{H_2O} \exp(m_{k_u}[\text{urea}])) \quad [2]$$

Structural figures were made by using PyMOL (www.pymol.org).

ACKNOWLEDGMENTS. This work was supported in part by the Leverhulme Trust. A.L.M. is a Research Fellow at St John's College, Cambridge, UK.

1. Bateman A, *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res* 32:D138–D141.
2. Pleshe E, Truesdell J, Batey RT (2005) Structure of a class II TrmH tRNA-modifying enzyme from *Aquifex aeolicus*. *Acta Crystallogr F* 61:722–728.
3. Nureki O, *et al.* (2004) Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme. *Structure* 12:593–602.
4. Nureki O, *et al.* (2002) An enzyme with a deep trefoil knot for the active-site architecture. *Acta Crystallogr D* 58:1129–1137.
5. Mosbacher TG, Bechthold A, Schulz GE (2005) Structure and function of the antibiotic resistance-mediating methyltransferase AviRb from *Streptomyces viridochromogenes*. *J Mol Biol* 345:535–545.
6. Michel G, *et al.* (2002) The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot. *Structure* 10:1303–1315.
7. Lim K, *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.
8. Elkins PA, *et al.* (2003) Insights into catalysis by a knotted TrmD tRNA methyltransferase. *J Mol Biol* 333:931–949.
9. Ahn HJ, *et al.* (2003) Crystal structure of tRNA(m1G37)methyltransferase: Insights into tRNA recognition. *EMBO J* 22:2593–2603.
10. Taylor WR (2000) A deeply knotted protein structure and how it might fold. *Nature* 406:916–919.
11. Taylor WR (2007) Protein knots and fold complexity: Some new twists. *Comput Biol Chem* 31:151–162.
12. Taylor WR, Lin K (2003) Protein knots: A tangled problem. *Nature* 421:25.
13. Kolesov G, Virnau P, Kardar M, Mirny LA (2007) Protein knot server: Detection of knots in protein structures. *Nucleic Acids Res* 35:W425–W428.
14. Yeates TO, Norcross TS, King NP (2007) Knotted and topologically complex proteins as models for studying folding and stability. *Curr Opin Chem Biol* 11:595–603.
15. Virnau P, Mirny LA, Kardar M (2006) Intricate knots in proteins: Function and evolution. *PLoS Comput Biol* 2:1074–1079.
16. Mallam AL, Jackson SE (2006) Probing Nature's knots: The folding pathway of a knotted homodimeric protein. *J Mol Biol* 359:1420–1436.
17. Pace CN (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131:266–280.
18. Myers JK, Pace CN, Scholtz JM (1995) Denaturant *m* values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 4:2138–2148.
19. Mallam AL, Onuoha SC, Grossmann JG, Jackson SE (2008) Knotted fusion proteins reveal unexpected possibilities in protein folding. *Mol Cell* 30:642–648.
20. Mallam AL, Jackson SE (2007) The dimerization of an alpha/beta-knotted protein is essential for structure and function. *Structure* 15:111–122.
21. Matouschek A, Kellis JT, Jr, Serrano L, Fersht AR (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340:122–126.
22. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224:771–782.
23. Mallam AL, Jackson SE (2005) Folding studies on a knotted protein. *J Mol Biol* 346:1409–1421.
24. Tanford C (1968) Protein denaturation. Part A. Characterisation of the denatured state; Part B. The transition from native to denatured state. *Adv Protein Chem* 23:121–282.
25. Tanford C (1970) Protein denaturation. Part C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24:1–95.
26. Schmid FX, Mayr LM, Mucke M, Schonbrunner ER (1993) Prolyl isomerases: Role in protein folding. *Adv Protein Chem* 44:25–66.
27. Brandts JF, Halvorson HR, Brennan M (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. *Biochemistry* 14:4953–4963.
28. Schultz DA, Schmid FX, Baldwin RL (1992) *Cis* proline mutants of ribonuclease A. II. Elimination of the slow-folding forms by mutation. *Protein Sci* 1:917–924.
29. Munoz V, Lopez EM, Jager M, Serrano L (1994) Kinetic characterization of the chemotactic protein from Escherichia coli, CheY. Kinetic analysis of the inverse hydrophobic effect. *Biochemistry* 33:5858–5866.
30. Kiefhaber T, Grunert HP, Hahn U, Schmid FX (1990) Replacement of a *cis* proline simplifies the mechanism of ribonuclease T1 folding. *Biochemistry* 29:6475–6480.
31. Odefey C, Mayr LM, Schmid FX (1995) Non-prolyl *cis-trans* peptide bond isomerization as a rate-determining step in protein unfolding and refolding. *J Mol Biol* 245:69–78.
32. Wedemeyer WJ, Welker E, Scheraga HA (2002) Proline *cis-trans* isomerization and protein folding. *Biochemistry* 41:14637–14644.
33. Virnau P, Kantor Y, Kardar M (2005) Knots in globule and coil phases of a model polyethylene. *J Am Chem Soc* 127:15102–15106.
34. Raymer DM, Smith DE (2007) Spontaneous knotting of an agitated string. *Proc Natl Acad Sci USA* 104:16432–16437.
35. Daggett V, Fersht AR (2003) Is there a unifying mechanism for protein folding? *Trends Biochem Sci* 28:18–25.
36. Daggett V, Fersht A (2003) The present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol* 4:497–502.
37. Mallam AL, Jackson SE (2007) A comparison of the folding of two knotted proteins: YbeA and YibK. *J Mol Biol* 366:650–665.
38. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30:10428–10435.