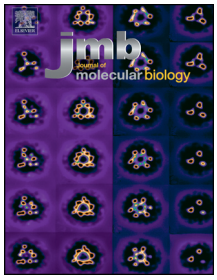

Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*

Nicole C.H. Lim[1,2] **and Sophie E. Jackson**[1]

*1 - **Department of Chemistry,** University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*
*2 - **Faculty of Sciences,** Universiti Brunei Darussalam, Gadong BE 1410, Brunei Darussalam*

Correspondence to Sophie E. Jackson: *sej13@cam.ac.uk*
http://dx.doi.org/10.1016/j.jmb.2014.09.007
Edited by J. Buchner

Abstract

The importance of knots and entanglements in biological systems is increasingly being realized and the number of proteins with topologically complex knotted structures has risen. However, the mechanism as to how these proteins knot and fold efficiently remains unclear. Using a cell-free expression system and pulse-proteolysis experiments, we have investigated the mechanism of knotting and folding for two bacterial trefoil-knotted methyltransferases. This study provides the first experimental evidence for a knotting mechanism. Results on fusions of stable protein domains to N-terminus, C-terminus or both termini of the knotted proteins clearly demonstrate that threading of the nascent chain through a knotting loop occurs *via* the C-terminus. Our results strongly suggest that this mechanism occurs even when the C-terminus is severely hindered by the addition of a large stable structure, in contrast to some simulations indicating that even the folding pathways of knotted proteins have some plasticity. The same strategy was employed to probe the effects of GroEL-GroES. In this case, results suggest active mechanisms for the molecular chaperonin. We demonstrate that a simple model in which GroEL-GroES sterically confines the unknotted polypeptide chain thereby promoting knotting is unlikely, and we propose two alternatives: (a) the chaperonin facilitates unfolding of kinetically and topologically trapped intermediates or (b) the chaperonin stabilizes interactions that promote knotting. These findings provide mechanistic insights into the folding of knotted proteins both *in vitro* and *in vivo*, thus elucidating how they have withstood evolutionary pressures regardless of their complex topologies and intrinsically slow folding rates.

Introduction

Over the past decade, the emergence of classes of topologically complex knotted proteins has intrigued researchers in the protein structure and folding fields [1–3]. To date, there are nearly 400 knotted proteins in the PDB database, each containing either a trefoil (3_1), figure-of-eight (4_1), Gordian (5_2) or Stevedore (6_1) knot in their structures, which is formed by the path of the polypeptide backbone chain [3–5]. These knotted topologies have been found to be conserved across different families [5], suggesting that the knot itself may be advantageous and important to the function of the protein. As yet, little is known about the functional advantages, if any, of these complex knotted structures over their unknotted counterparts. Although experimentally, the elucidation of the pathways by which knotted proteins fold remains challenging; in recent years, some progress has been made.

Most of the experimental investigations on knotted proteins have been focussed on the trefoil-knotted α/β methyltransferases, YibK from *Haemophilus influenzae* and YbeA from *Escherichia coli* [6–13]. Both proteins are homodimers and contain a trefoil knot at the C-terminus in which at least 40 residues pass through a similarly sized loop (Fig. 1a–c). Biophysical techniques have been used to probe the knotting and folding mechanisms of purified recombinantly expressed YibK and YbeA [6,7,9]. However, through *in vitro* folding experiments on circularized variants of YibK and YbeA, it was discovered that the denatured ensembles, even in high concentrations of chemical denaturant, contained kinetically trapped knotted polypeptide chains [12]. Consequently, *in vitro* folding

Please cite this article as: Lim Nicole C.H., Jackson Sophie E., Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.09.007

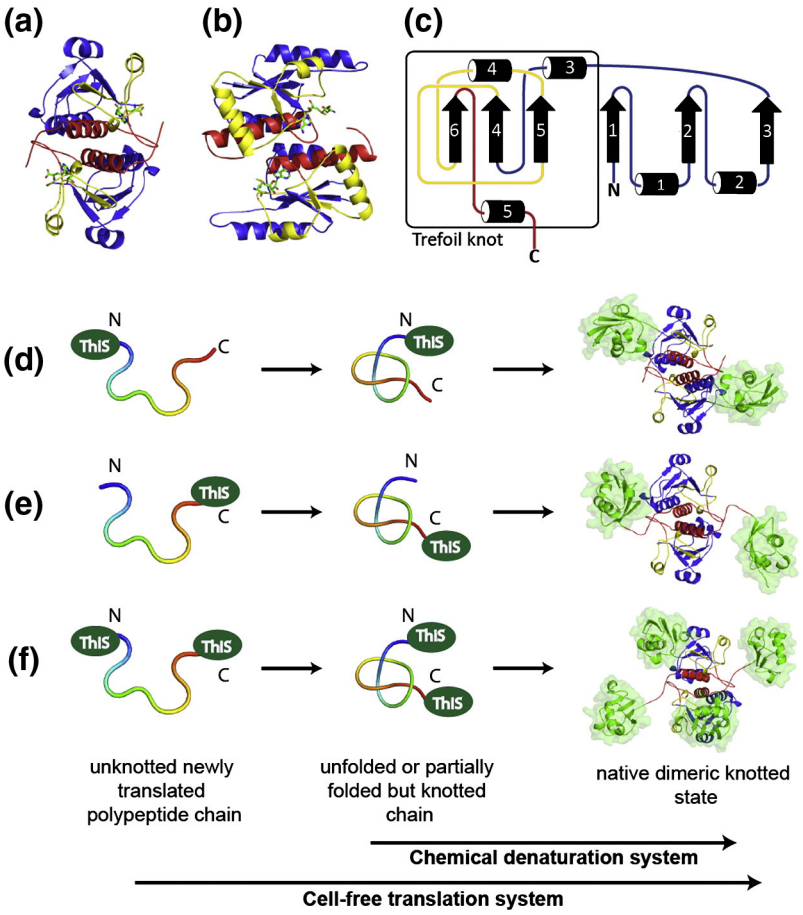

Fig. 1. Structures of the α/β trefoil-knotted methyltransferases (top panel) and a schematic diagram for the knotting and folding of the different constructs of the newly translated ThiS fusions (bottom panel). Top panel: Ribbon diagrams of homodimeric knotted proteins YibK (PDB code 1MXI) (a) and YbeA (PDB code 1NS5) (b). Structures are colored to show the knotting loop highlighted in yellow and the knotted chain at the C-terminus is red. The methyltransferase cofactor, AdoHcy, is shown as a stick model to indicate the cofactor-binding site in the knotted region of the proteins. (c) Topological diagram of YibK. Numbers refer to secondary structure elements and the structure is colored as in (a). Bottom panel: The ThiS domain is fused at the amino-terminus (d), carboxy-terminus (e) or both termini (f) for the different constructs. Folding rates of a newly translated unknotted polypeptide chain can be determined using novel pulse-proteolysis methods in the cell-free translation system while the rates of folding from the chemically denatured states have been determined previously using chemical denaturants in *in vitro* folding experiments of the purified recombinant knotted fusion proteins. Knot representations and ribbon diagrams were generated using KnotPlot (http://www.knotplot.com/) and PyMOL (http://www.pymol.org).

experiments from a chemically denatured state probe the refolding from a knotted denatured state into a native, knotted structure. Recently, with the use of a coupled *in vitro* transcription–translation system and novel methodology developed based on existing pulse-proteolysis experiments, Mallam and Jackson were able to investigate the folding rates of nascent chains of knotted proteins after they were first synthesized by the ribosome [13]. The results showed that the nascent chains could fold correctly to their trefoil-knotted structure, albeit very slowly and that the rate-determining step in folding was associated (either directly or indirectly) with threading. Additionally, the GroEL-GroES chaperonin was found to have a dramatic effect on the folding rate of the newly translated polypeptide chain, thus establishing that chaperonins are likely to be important in the post-translational folding of these bacterial knotted proteins *in vivo*.

In the study presented here, the same coupled *in vitro* transcription–translation system [13] was used to synthesize nascent chains of three different constructs of ThiS fusions of YibK and YbeA (Fig. 1d–f). These had been previously made by fusing the stable 91-residue thermophilic protein *Archaeoglobus fulgidus* ThiS to the amino-terminus (ThiS-YibK and ThiS-YbeA), carboxy-terminus (YibK-ThiS and YbeA-ThiS) or both termini (ThiS-YibK-ThiS and ThiS-YbeA-ThiS) of the knotted methyltransferases such that it acts as a "molecular plug" hindering threading movements of the polypeptide chain [11]. The structure, stability and ligand binding of the different fusion proteins of YibK and YbeA have been previously reported and no significant differences between the behavior were observed [11]. Ligand binding is a good reporter of native structure and dimer formation. Thus, N- and C-terminal ThiS fusions minimally perturb the system. The cell-free expression PURExpress system used here contains only the components needed for transcription and translation in *E. coli* [14–16]. Like in all typical prokaryotic systems, *in vitro* transcription and translation in the PURExpress system are coupled and occur simultaneously; translation begins while the mRNA is still being synthesized. This system is advantageous as it not only provides a defined environment in which the process of protein synthesis can be studied but also lacks molecular chaperones and thus the role that they play in the folding of nascent polypeptide chains can be investigated [14,17].

Please cite this article as: Lim Nicole C.H., Jackson Sophie E., Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.09.007

By using the cell-free expression system, we directly monitored the knotting and folding behavior of the nascent chains of the different ThiS fusions of YibK and YbeA. Here, we show that even with the addition of a ThiS domain to either the amino-terminus or the carboxy-terminus or to both termini, these multi-domain proteins with extremely deep knots can be synthesized *in vitro* and spontaneously knot without the help of any molecular chaperones. Established pulse-proteolysis kinetic experiments were then used to investigate the effects of the fusion on the folding rate, thus providing the first experimental evidence of the mechanism of knotting. In addition, the effect of the GroEL-GroES chaperonin system on the folding of nascent chains of the fusion constructs was used to gain information on the possible mechanisms for its mode of action on the folding of this class of protein. This provides additional insights as to how knotted proteins have withstood evolutionary pressures despite their complex topologies and intrinsically slow rates of folding.

Results

N- and C-terminal fusions of ThiS to YibK and YbeA can be synthesized *in vitro*

The *in vitro* synthesis of the different constructs was initiated by addition of the appropriate template DNA to the PURExpress reaction mix and it was incubated at 37 °C for 4 h. Products of the translation reaction were analyzed by SDS-PAGE and visualized with SYPRO Ruby stain. Bands corresponding to the respective ThiS fusion proteins were prominent compared with a control reaction with no template DNA (Supplementary Fig. 1). This indicated that the different ThiS fusions can be successfully synthesized using the PURExpress system. Visualization of the gel after the addition of modified lysine tRNA BODIPY-Lys-tRNA$_{Lys}$ to the translation reactions also confirmed that the translated products resulted in ThiS fusion proteins of the correct molecular weight in which BODIPY-FL had been successfully incorporated (Supplementary Fig. 1).

The yield and solubility of the various ThiS fusions of YibK and YbeA synthesized *in vitro* were also determined by SDS-PAGE of the translation reactions (Supplementary Fig. 2). All of the fusion proteins were found to be ~90–98% soluble and the yields were comparable to those of the wild-type proteins [13].

N- and C-terminal ThiS fusions of YibK and YbeA synthesized *in vitro* fold correctly to their native, knotted structures

Mallam and Jackson have previously validated the use of the cell-free expression system in synthesizing YibK and YbeA by probing their structures, stabilities and cofactor binding [13]. Here, we simply measured the stability of the fusion proteins synthesized *in vitro* and compared the results to the stability of purified proteins expressed recombinantly in *E. coli*.

To determine the free energy of unfolding, $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ (where N_2 is native dimer and D is denatured monomer) of the fusion proteins synthesized *in vitro* and those bacterially expressed and purified to homogeneity, we employed pulse-proteolysis experiments [13,18]. As this technique only requires small amounts of sample and can be performed in lysate mixtures [19], the measurements on the samples synthesized *in vitro* were carried out without purification from the other components in the PURExpress system. Thermolysin was incubated with the pre-equilibrated mixtures of protein and urea for 1 min, during which it selectively digests unfolded or partially folded proteins. Fully folded protein in the samples was quantified using SDS-PAGE to determine the fraction of folded molecules present at different concentrations of urea for the different constructs (Supplementary Fig. 3). The chemical denaturation profiles of the fusion proteins were then obtained and protein stabilities were determined (Fig. 2 and Table 1). The $\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ values calculated for the fusion proteins synthesized *in vitro* were in agreement with the recombinantly expressed proteins measured under the same conditions. Although some differences in stability between the different ThiS fusion constructs are seen, the variations between the free energies of unfolding are within 15% for the YibK constructs and within 10% for the YbeA constructs and not significant (Table 1). The *m*-values, a measure of the cooperativity of the system, are all within error (Table 1). Thus, both N- and C-terminal fusions have not significantly perturbed the native basin or cooperativity of the system, as one would expect given they are fused to the knotted methyltransferases through flexible linkers and structural modeling based on small-angle X-ray scattering

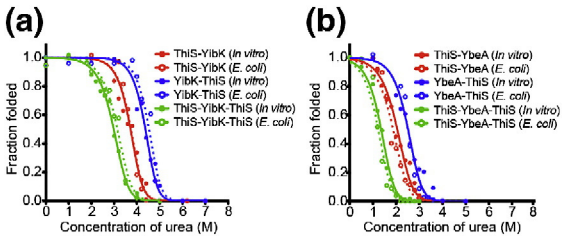

Fig. 2. Analysis of the stability of the different ThiS fusion knotted proteins. Pulse-proteolysis experiments were performed on the different ThiS fusions of YibK (a) and YbeA (b) expressed in *E. coli* and purified to homogeneity or produced by cell-free translation. Fractions of folded protein, determined by pulse-proteolysis experiments with thermolysin, are shown. Continuous lines represent the best fit of the data to a two-state equilibrium dimer unfolding model (see Supplementary Methods).

Please cite this article as: Lim Nicole C.H., Jackson Sophie E., Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.09.007

data suggests that the additional ThiS domains make few interactions with the methyltransferase domains [11]. This suggests that that majority of the protein molecules produced by the reconstituted, cell-free expression system fold correctly to their native, knotted structures without the presence of significant amounts of any misfolded species even in the absence of molecular chaperones.

Spontaneous knotting after translation occurs *via* threading of the C-terminus of the nascent chain

Here, we applied the same modified pulse-proteolysis methodology used by Mallam and Jackson [13] to monitor the time course of *in vitro* translation and subsequent folding of the fusion proteins. After initiating the translation reaction with the addition of plasmid DNA, we removed several aliquots at various time points and stopped protein synthesis. Pulse-proteolysis experiments were then carried out on half of the quenched translation reaction mixture so that any unfolded or partially folded protein molecules were digested. SDS-PAGE analysis of the undigested and digested samples was then used to obtain the time courses for the appearance of full-length translated protein and full-length folded protein (Supplementary Fig. 4). Kinetic data acquired were then simultaneously modeled and fitted to a simplified model to obtain estimated rate constants for the translation reaction (k_{trans}) and subsequent protein folding reaction (k_{fold}). As a control, the translation and subsequent folding reaction of ThiS on its own was monitored using the cell-free expression system under the same conditions. Notably, in this case, no delay between the formation of translated protein and the formation of the folded ThiS protein was observed (Supplementary Fig. 5), indicating that the ThiS domain folds very rapidly to a proteolytically resistant folded state. This important feature of ThiS allows us to examine the translation and subsequent folding of the knotted domains while attached to one or more folded ThiS domains. Additionally, the translation and folding behavior of YibK and YbeA alone were also monitored so that a direct comparison of the folding rates with the fusion proteins could be made. As previously observed, there was a delay of 10–20 min between the formation of translated product and the formation of the stable folded protein (Fig. 3) [13].

For the N-terminal ThiS fusion proteins (ThiS-YibK and ThiS-YbeA), folding rates observed were very similar to that of YibK or YbeA alone, which were estimated to be 0.07 min^{-1} and 0.09 min^{-1}, respectively (Fig. 3a and d and Table 2). This suggests that making an N-terminal fusion does not hinder knotting or folding of the nascent polypeptide chain. In contrast, a dramatic deceleration in the folding rates of the C-terminal ThiS fusion proteins (k_{fold} values of 0.02 min^{-1} and 0.03 min^{-1} for YibK-ThiS and YbeA-ThiS, respectively) was observed (Fig. 3b and e and Table 2). This could be due to either the hindrance of the threading of the C-terminus by the folded ThiS domain or the nascent chain threading through a knotting loop *via* a less effective pathway potentially involving the N-terminus. However, with the use of the double ThiS fusion proteins (ThiS-YibK-ThiS and ThiS-YbeA-ThiS), we observed similar folding rates to that of the C-terminal fusion proteins (Fig. 3c and f and Table 2). Consequently, this suggests that threading must still occur by the C-terminus of the newly translated polypeptide chains passing through a knotting loop as the greatly retarded folding rates signify that the ThiS

Table 1. Thermodynamic parameters for the ThiS fusions of YibK and YbeA synthesized *in vitro* and recombinantly expressed and purified.

Fusion protein	Expression conditions	C_m (M)	m_{app} (kcal mol^{-1} M^{-1})	$\Delta G_{H_2O}^{N_2 \leftrightarrow 2D}$ (kcal mol^{-1})
ThiS-YibK	Recombinant expression in *E. coli*	3.7 ± 0.1	3.1 ± 0.3	19.7 ± 1.2
	Cell-free *in vitro* translation	3.7 ± 0.1		19.7 ± 1.2
YibK-ThiS	Recombinant expression in *E. coli*	4.5 ± 0.1	3.2 ± 0.4	22.6 ± 1.8
	Cell-free *in vitro* translation	4.4 ± 0.1		22.3 ± 1.8
ThiS-YibK-ThiS	Recombinant expression in *E. coli*	3.1 ± 0.1	2.6 ± 0.2	16.3 ± 0.7
	Cell-free *in vitro* translation	3.0 ± 0.1		16.0 ± 0.7
ThiS-YbeA	Recombinant expression in *E. coli*	1.9 ± 0.1	2.5 ± 0.2	13.0 ± 0.5
	Cell-free *in vitro* translation	2.1 ± 0.1		13.5 ± 0.5
YbeA-ThiS	Recombinant expression in *E. coli*	2.5 ± 0.1	2.8 ± 0.3	15.2 ± 0.8
	Cell-free *in vitro* translation	2.5 ± 0.1		15.2 ± 0.8
ThiS-YbeA-ThiS	Recombinant expression in *E. coli*	1.2 ± 0.1	2.9 ± 0.3	11.7 ± 0.5
	Cell-free *in vitro* translation	1.4 ± 0.1		12.3 ± 0.5

C_m is the concentration of urea at the midpoint of the unfolding transition. m_{app} is the apparent *m*-value, is a measure of the slope of the unfolding transition and is shared during the analysis of the denaturation curves for each protein. $\Delta G_{H_2O}^{N_2 \leftrightarrow 2D}$ is the free energy of unfolding in the absence of denaturant and is a measure of the conformational stability of a protein. Fitting errors are shown. The midpoints of denaturation are all similar to those reported using fluorescence-based unfolding experiments [11].

Please cite this article as: Lim Nicole C.H., Jackson Sophie E., Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.09.007

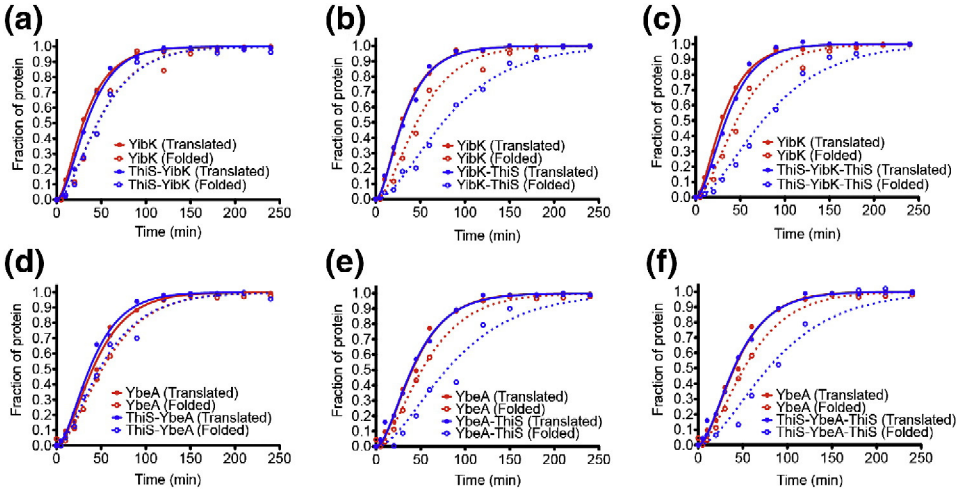

Fig. 3. Kinetics of translation and folding for the various ThiS fusion YibK and YbeA constructs synthesized *in vitro*. The time courses for the appearance of translated and folded ThiS-YibK (a), YibK-ThiS (b), ThiS-YibK-ThiS (c), ThiS-YbeA (d), YbeA-ThiS (e) and ThiS-YbeA-ThiS (f) in a cell-free translation reaction are shown. The kinetic data were fit to the simplified reaction Scheme 2 (see Materials and Methods). Each ThiS fusion YibK and YbeA kinetic dataset (blue) was compared to the kinetic data of wild-type YibK and YbeA synthesized *in vitro*, respectively (red).

domain attached to the C-terminus in both cases hinders the threading motion of the polypeptide chain.

Action of the bacterial GroEL-GroES chaperonin on the ThiS fusion knotted proteins

The results described above confirm that the newly translated fusion proteins can spontaneously knot and fold in the absence of any cellular chaperones. Given the effects of GroEL-GroES on the folding of nascent chains of wild-type YibK and YbeA previously reported [13], here, we examined the effect of the bacterial chaperonin on the folding of the fusion proteins as they are being translated *in vitro* in the cell-free expression system.

Time courses for the appearance of full-length translated fusion proteins and the appearance of full-length folded fusion proteins were measured as described above with the addition of the GroEL-GroES complex (0.15 μM) (Fig. 4 and Supplementary Fig. 6). In the case of the N-terminal ThiS fusion proteins, no delay between the formation of the translated products and the folded knotted proteins was observed and a lower limit for k_{fold} of 2 min^{-1} for ThiS-YibK and ThiS-YbeA in the presence of GroEL-GroES was estimated (Fig. 4a and d and Table 2). This is remarkably similar to the rate constants obtained for folding of wild-type YibK and YbeA under the same conditions [13], indicating that GroEL-GroES also enhanced the rate of knotting and/or folding of the newly translated N-terminal ThiS fusions. In contrast, the chaperonin complex had surprisingly little or no effect on the rate of folding of the C-terminal (Fig. 4b and e) and the double ThiS fusion proteins (Fig. 4c and f) as no significant increase in the estimated rate constants for protein folding was observed (at least 20-fold increase in k_{fold} for the N-terminal ThiS fusion proteins compared to a maximum of 2-fold increase in

Table 2. Kinetic parameters describing the appearance of translated (k_{trans}) and folded (k_{fold}) proteins synthesized *in vitro*.

Fusion protein ± GroEL-GroES	k_{trans} (min^{-1})	k_{fold} (min^{-1})
YibK	0.11 ± 0.02	0.07 ± 0.04
YibK + GroEL-GroES	0.08 ± 0.02[a]	>2[a]
ThiS-YibK	0.11 ± 0.02	0.07 ± 0.04
ThiS-YibK + GroEL-GroES	0.10 ± 0.02	>2
YibK-ThiS	0.11 ± 0.02	0.02 ± 0.01
YibK-ThiS + GroEL-GroES	0.10 ± 0.02	0.04 ± 0.01
ThiS-YibK-ThiS	0.11 ± 0.02	0.02 ± 0.01
ThiS-YibK-ThiS + GroEL-GroES	0.10 ± 0.02	0.04 ± 0.02
YbeA	0.08 ± 0.03	0.09 ± 0.06
YbeA + GroEL-GroES	0.08 ± 0.02[a]	>2[a]
ThiS-YbeA	0.09 ± 0.02	0.09 ± 0.06
ThiS-YbeA + GroEL-GroES	0.09 ± 0.01	>2
YbeA-ThiS	0.09 ± 0.02	0.03 ± 0.02
YbeA-ThiS + GroEL-GroES	0.09 ± 0.02	0.07 ± 0.06
ThiS-YbeA-ThiS	0.09 ± 0.01	0.03 ± 0.02
ThiS-YbeA-ThiS + GroEL-GroES	0.09 ± 0.02	0.06 ± 0.05

Rate constants were obtained from the fit of the data to the simplified Scheme 2 (see Materials and Methods).

[a] Values were those obtained by Mallam and Jackson [13]. A Student's *t*-test indicates that the lower limit for the value of k_{fold} observed for either ThiS-YibK or ThiS-YbeA in the presence of GroEL-GroES is significantly different from the value of k_{fold} observed in the absence of GroEL-GroES ($P < 0.0001$, unpaired *t*-test, $n = 4$). In contrast, there is no significant difference between the values of k_{fold} observed in the absence of GroEL-GroES compared to those observed in the presence of GroEL-GroES for either YibK-ThiS or YbeA-ThiS and for either ThiS-YibK-ThiS or ThiS-YbeA-ThiS ($P > 0.1$, unpaired *t*-test, $n = 4$).

Please cite this article as: Lim Nicole C.H., Jackson Sophie E., Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.09.007

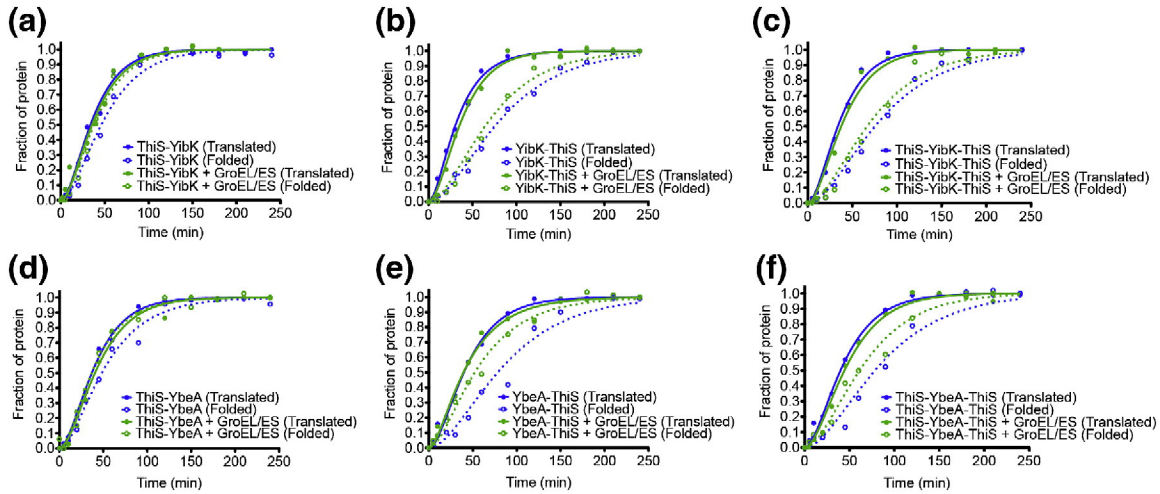

Fig. 4. Kinetics of translation and folding for the various ThiS fusion YibK and YbeA constructs synthesized *in vitro* in the presence of the GroEL-GroES bacterial chaperonin. The time courses for the appearance of translated and folded ThiS-YibK (a), YibK-ThiS (b), ThiS-YibK-ThiS (c), ThiS-YbeA (d), YbeA-ThiS (e) and ThiS-YbeA-ThiS (f) with GroEL-GroES in a cell-free translation reaction are shown. The kinetic data were fit to the simplified reaction Scheme 2 (see Materials and Methods). Each kinetic dataset (green) was compared to that of the kinetic data obtained without GroEL-GroES (blue).

k_{fold} for the C-terminal or double ThiS fusion proteins) (Table 2). Consequently, these results show that the GroEL-GroES chaperonin was only able to facilitate the knotting and/or folding of the newly translated polypeptide chain post-translationally when the ThiS domain is fused to the N-terminus but not the C-terminus of the knotted protein.

Discussion

In recent years, our understanding of how knotted proteins undergo efficient folding both *in vitro* and *in vivo* despite their complex topology has increased significantly through various *in silico* and experimental efforts [2–5,20,21]. In the study presented here, we have shown that newly translated polypeptide chains of various ThiS fusions of YibK and YbeA can spontaneously knot and fold into their native, knotted structures without the help of any molecular chaperones, albeit slowly. This is remarkably intriguing as the double ThiS fusion proteins synthesized in this study are the most deeply embedded protein knots observed to date, as at least 125 residues have to be theoretically removed from each termini before the structure becomes unknotted [11].

By using the cell-free expression system and modified pulse-proteolysis methodology employed by Mallam and Jackson [13], we were able to monitor the knotting and folding behavior of the fusion proteins after they are first synthesized by the ribosomal machinery and measure folding rates. We use the relative folding rate constants between the different fusion constructs to inform on folding pathways. Of course, the folding rate in itself does not provide information on the mechanism of folding; however, in the case of the different fusions, the relative rates provide a strong indication of the pathway. We cannot completely rule out the fact that there may be significant plasticity in the folding route and that fusions folding with the same rate constant do so with completely different pathways, but we think this unlikely.

To understand how the knotted fusion proteins can knot and fold under these conditions, we first need to consider the rate of folding of ThiS on its own. The data show that ThiS folds rapidly with an estimated k_{fold} with a lower limit of 2.3 min^{-1} (Supplementary Fig. 5). This is much faster than the folding of wild-type YibK or YbeA (k_{fold} of 0.07 min^{-1} or 0.09 min^{-1}, respectively) (Table 2). As the rate of translation in *E. coli* is approximately 40 amino acids per second [22], the various knotted fusion proteins would be synthesized within 10 s. This means that the ThiS domains in the fusion proteins will fold first, whether they are at the N-terminus or at the C-terminus, be it in a co-translational or post-translational manner, and before any knotting or folding of the YibK or YbeA domains. Based on this, the ThiS domains should act as molecular plugs and hinder the threading of the polypeptide chain through a knotting loop, which should slow down folding. It should be noted that this experimental approach enables us to probe the dominant folding pathway of nascent chains of YibK and YbeA; however, it is not possible with this method to assess the importance of less populated folding routes that may exist and play a role in the folding of these types of proteins. Nevertheless, the experimental strategy employed is extremely suitable for probing the dominant folding pathways of nascent chains and also establishing the effects of molecular chaperones.

Please cite this article as: Lim Nicole C.H., Jackson Sophie E., Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.09.007

Our results demonstrate that the folding rates of the N-terminal ThiS fusion constructs are similar to that of YibK or YbeA alone (Fig. 3a and d), whereas the folding rates of the C-terminal ThiS fusions are dramatically reduced (Fig. 3b and e). These results clearly show that threading occurs *via* a mechanism in which the C-terminal end of the chain passes through a knotting loop to form the knot. For the C-terminal ThiS fusions, the reduction in folding rates may be due to the steric hindrance of the ThiS domain passing through the knotting loop in a mechanism where threading still occurs *via* the C-terminus. On the other hand, it may also result from an alternative pathway involving the N-terminus, as has been observed in simulations [21,23]. However, the experimental data obtained for the double ThiS fusions showed that the folding rates were similar to those of the C-terminal fusions (Fig. 3c and f). This strongly suggests that the C-terminal fusions do not switch to a threading mechanism that involves the N-terminal region of the chain. Consequently, we conclude that the C-terminus is crucial for the threading of the newly translated polypeptide chain through a knotting loop and efficient folding in all the constructs studied. These results provide the first experimental evidence for the knotting mechanism of trefoil-knotted proteins and it is consistent with various folding simulations [21,24,25]. Recent Monte Carlo simulations, which examined the effect of surface tethering on the folding of a trefoil-knotted protein, also demonstrated that the mobility of the terminus closest to the knotted core is vital for successful folding and hindrance leads to a decrease in folding rate [23] consistent with our results. As the C-terminus of the nascent chain is required for the threading motion, it is highly likely that knotting is a post-translational rather than a co-translational event as we have suggested before [13].

The results on the N- and C-terminal single fusion constructs are very clear cut and strongly suggest C-terminal threading events as expected. However, the results on the double ThiS fusions are equally revealing. Simulations, which are based mainly on Gō-type, native centric models, have shown that, for trefoil-knotted chains, hindering the chain at the terminus closest to the knot, in this case by tethering, can affect the knotting pathway, such that knotting occurs *via* an alternative mechanism involving the other terminus and a so-called spindle mechanism [23]. In addition, in other *in silico* studies, varying the temperature of the simulation was also found to result in mechanistic changes [21]. Therefore, in these simulations, which frequently use coarse-grained models, the folding pathway has a certain plasticity. In contrast, our experimental results suggest that even on extending the chain by some 91 residues that knotting and folding still occur along the same, or a very similar pathway, which always involves threading of the C-terminus of the chain through a loop. Thus, the polypeptide chain cannot access multiple pathways as in the simulations. This suggests that the protein has evolved a distinct pathway along which it folds. It is interesting to speculate on why the experimental and computational studies differ in this regard. One possibility is that non-native interactions play a much more important role in defining the folding pathway for these proteins with complex topologies in contrast to the folding of smaller proteins with relatively simple folds. Another is that the coarse-grained models, which are frequently used, are too simplistic and do not fully represent the conformational dynamics of a polypeptide chain. In this regard, it is very interesting to note that Noel *et al.* found that a simple C^{α} model of the chain with no side chains showed a more diverse set of routes leading to the rate-limiting transition state than a model in which side chains were explicitly incorporated [21].

The GroEL-GroES chaperonin system from *E. coli* has been studied extensively [26–33] and several possible mechanisms have been proposed [26,34,35]. It has been suggested that the GroEL-GroES chaperonin does not have any direct specific effect on the substrate protein but acts passively as an isolated folding environment for the polypeptide chain such that it avoids aggregation in the overcrowded cellular environment [36–38]. In addition, GroEL-GroES has also been found to enhance the rate of folding of a substrate protein by actively unfolding kinetically trapped misfolded intermediates repeatedly until the native state is reached (iterative annealing) [39–41]. Furthermore, it has been proposed that steric confinement of folding intermediates by GroEL-GroES can actively accelerate protein folding, as it reduces conformational entropy thereby enhancing their conversion to the native state [33,37,42].

How does the GroEL-GroES chaperonin facilitate folding of knotted proteins? Previously, it was found that the GroEL-GroES complex accelerated knot formation after *in vitro* synthesis of YibK and YbeA [13]. It was suggested that this might be due to the steric confinement within the chaperonin central cavity whereby the entropic barrier to knotting is reduced and the probability of the threading event to populate the knotted intermediate state increases (Fig. 5a). In this study, the effect of GroEL-GroES on the various fusion proteins was investigated. Our results show that GroEL-GroES increases the rate of folding of the newly translated N-terminal ThiS fusions by at least 20-fold, similar to the results obtained for wild-type YibK and YbeA (Fig. 4a and d). This suggests that the N-terminal fusion proteins have the same folding pathway as the wild-type proteins and the folding reaction in the presence of the chaperonin is limited by translation. This is in line with our model in which the N-terminal ThiS domain folds rapidly followed by the independent slow folding of the YibK or YbeA domains with a mechanism in which the C-terminal region of the chain threads through a knotting loop. Previous

Please cite this article as: Lim Nicole C.H., Jackson Sophie E., Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.09.007

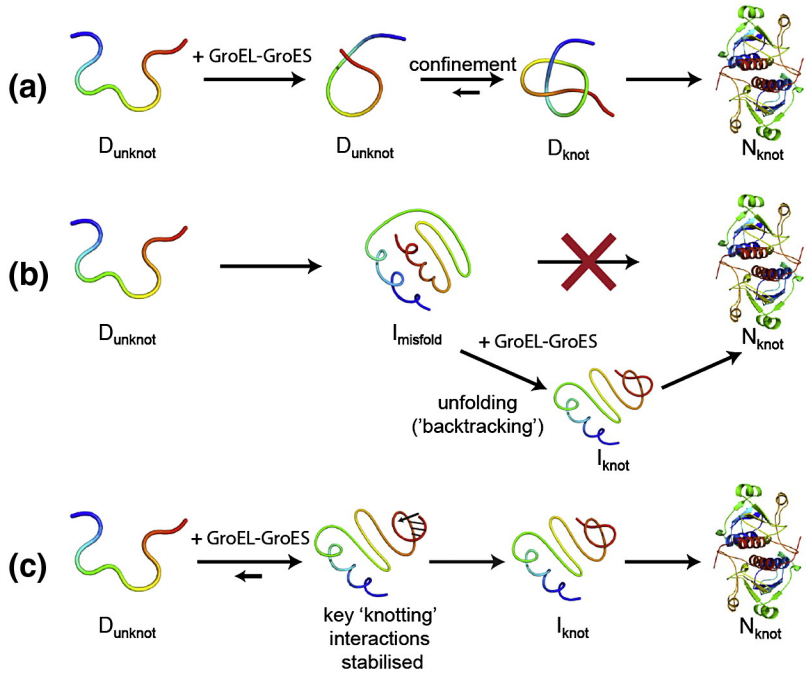

Fig. 5. Schematic diagram illustrating the possible mechanisms for the bacterial GroEL-GroES chaperonin action on the folding of bacterial trefoil-knotted methyltransferases *in vitro* and *in vivo*. The substrate protein is encapsulated within the GroEL-GroES central cavity and may lead to (a) the steric confinement of the unknotted polypeptide chain, (b) the active facilitation of the unfolding of any misfolded intermediates (kinetic topological traps) that slow down folding or (c) the further stabilization of key interactions that promote knotting. D, denatured; I, intermediate; N, native. The exact secondary structure and relative position of the termini during folding is not known, and it is shown here for illustrative purposes only.

simulations have revealed that the folding pathway of knotted proteins may involve repeated unfolding ("backtracking") events of misfolded intermediates so that key native contacts needed for the successful folding can be formed and topological barriers can be overcome [21,43]. This is consistent with a mechanism of GroEL-GroES action in which there is active unfolding of misfolded intermediates. As both wild-type and N-terminal ThiS fusion proteins fold with similar pathways, the effect of GroEL-GroES on the misfolded intermediates is, in these cases, the same.

In the case of the C-terminal and the double ThiS fusion proteins, very little effect of the GroEL-GroES complex on the folding rates is observed (an approximate 2-fold increase in k_{fold}) (Fig. 4b and c and e and f). This suggests either that these fusion proteins have different folding pathways and rate-determining steps to that of the wild-type YibK and YbeA or that the misfolded intermediates are sufficiently different when the ThiS plug is fused to the C-terminus such that GroEL-GroES can no longer act on them. These results also provide strong evidence that the chaperonin does not simply act by confining the chain thereby promoting knotting; otherwise, we would expect to see similar effects on all constructs.

Our experimental data provide new insights into the possible mode of actions of the bacterial GroEL-GroES chaperonin on the folding of newly translated ThiS knotted fusion proteins. We propose that GroEL-GroES encapsulates the substrate protein and actively facilitates the unfolding of any misfolded intermediates (kinetic topological traps) that slow down folding or may further stabilize key interactions that promote knotting (Fig. 5b and c). As wild-type ThiS is found to be a very stable protein that folds remarkably fast [11], it is highly unlikely that the ThiS domains fused to the N- or C-termini are unfolded by GroEL-GroES. Thus, in the case of the C-terminal and double ThiS fusion proteins, the ThiS plug is still present and continues to hinder the threading motion of the polypeptide chain. This corroborates with our findings that demonstrate no significant enhancement in the folding rates of the proteins in the presence of GroEL-GroES when a ThiS domain is fused to the C-terminus.

By replicating the bacterial cellular environment with the use of the cell-free expression system, not only we are able to determine the knotting mechanism of newly translated knotted proteins but also we further investigate the mode of action of the GroEL-GroES chaperonin on the folding of knotted proteins. From our study, we can unequivocally conclude that the C-terminus of trefoil-knotted proteins is critical to the threading of the polypeptide chain to form the knot. We have also demonstrated that GroEL-GroES actively assists the folding of knotted proteins by a mechanism that may involve the unfolding of kinetically trapped misfolded intermediates until interactions that stabilize the knotting event are achieved. These key

Please cite this article as: Lim Nicole C.H., Jackson Sophie E., Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.09.007

observations provide us with important information into the complex folding pathway of trefoil-knotted proteins and provide us of a stepping stone toward further insights into how topologically knotted proteins have withstood evolutionary pressures and achieve efficient folding *in vivo*.

Materials and Methods

Recombinant protein expression and purification

YibK and YbeA were expressed and purified as described elsewhere [6,7]. The different fusion constructs were expressed and purified as previously described [11]. Protein concentrations were determined spectrophotometrically and are reported in monomer units.

Cell-free protein expression

The coupled *in vitro* transcription and translation of YibK, YbeA and the different ThiS fusions was carried out using the PURExpress system (New England Biolabs) following the manufacturer's protocol. This system is reconstituted from only the purified components needed for *E. coli* translation and transcription; thus, it does not contain any molecular chaperones. Template plasmid DNA (250 ng) containing the appropriate gene was added to the protein synthesis reaction (25 μL) and incubated at 37 °C for 4 h. Protein synthesis was also carried out in the presence of BODIPY-Lys-tRNA$_{Lys}$ (FluoroTect Green$_{Lys}$ *in vitro* Translation Labeling System; Promega) to incorporate fluorescently labeled lysine residues into the nascent chains. Reaction products were analyzed by SDS-PAGE followed by staining with SYPRO Ruby Protein Gel Stain (Invitrogen). Samples containing BODIPY-Lys-tRNA$_{Lys}$ were treated with RNase A (0.1 mg mL^{-1}; Fermentas) and heated at 37 °C for 10 min before running on a SDS-PAGE gel. Protein gels were visualized with a Typhoon Imager (GE Healthcare). BODIPY-FL fluorescence was measured by direct-blue excitation and emission through a 520-nm bandpass filter while SYPRO Ruby was measured by direct-green excitation and emission through a 610-nm bandpass filter.

Stability measurements by pulse proteolysis

The stabilities of the fusion proteins expressed *in vitro* and the purified recombinant fusion proteins were measured using pulse-proteolysis experiments [18,19]. This technique allows the fraction of folded proteins in equilibrium mixtures of folded and unfolded proteins to be determined as it exploits the fact that folded structures are more proteolytically resistant than unfolded conformations. The concentration of translated product in each PURExpress reaction was calculated by a comparison of the band intensity of the translated product to that of a control, which was purified recombinant proteins of known concentration. All measurements were made in a buffer of 50 mM Tris–HCl (pH 7.5), 200 mM KCl, 10% (v/v) glycerol, 1 mM DTT and 10 mM $CaCl_2$. Samples of the completed *in vitro* translation reactions or purified recombinant proteins were added to aliquots of urea to give a range of final denaturant concentrations between 0 and 7 M and a final protein concentration of 1 μM. Samples were incubated at 25 °C overnight. A stock solution of 1 mg mL^{-1} of thermolysin from *Bacillus thermoproteolyticus rokko* (Sigma-Aldrich) was prepared in 2.5 M NaCl containing 10 mM $CaCl_2$. Thermolysin was added to each aliquot to a final concentration of 0.2 mg mL^{-1} and left for 1 min at 25 °C to digest any unfolded or partially unfolded protein. The proteolysis reaction was then quenched by the addition of phosphoramidon (Sigma-Aldrich) to a final concentration of 13.5 μM to inhibit thermolysin. Each quenched reaction was then mixed with ethylenediaminetetraacetic acid (Invitrogen) to a final concentration of 12.5 mM. Each reaction mixture was analyzed by SDS-PAGE and the band intensity of the protein of interest was quantified with Image Quant to determine the fraction of folded protein present at each urea concentration.

As all the ThiS fusion proteins are homodimers [11], a simplified, two-state dimer model, with native dimer in equilibrium with unfolded monomer, was used to fit the equilibrium unfolding datasets measured by pulse proteolysis:

$$N_2 \overset{K_U}{\longleftrightarrow} 2D \qquad \textit{Scheme } 1$$

Further details are described in Supplementary Methods.

Translation and folding kinetics

To measure the time course for the appearance of newly translated, folded knotted proteins, we modified pulse-proteolysis experiments described above. *In vitro* transcription and translation were initiated by the addition of the appropriate DNA template and the reaction was placed at 37 °C. Aliquots were removed at various time points after the DNA template was added and translation was halted by the addition of chloramphenicol (2 mM). Half of this quenched translation reaction was immediately subjected to pulse proteolysis to digest any full-length translated protein that was not yet folded. Both undigested and digested samples were analyzed by SDS-PAGE to compare the time course for the appearance of full-length translated protein to that of full-length folded protein, respectively. These measurements were repeated under identical conditions in the presence of the GroEL-GroES chaperonin complex (0.15 μM; Enzo Life Sciences). ATP and an ATP regeneration system are present in the cell-free expression system [15,17]. For the ThiS-YibK-ThiS construct, BODIPY-Lys-tRNA$_{Lys}$ was added into the mixture and the samples analyzed by SDS-PAGE were measured based on the BODIPY-FL fluorescence, which is a direct measure of the incorporation of fluorescently labeled lysine residues into the nascent chains.

A simplified sequential reaction scheme consisting of three consecutive, irreversible steps was used to model the kinetic data and estimate the rate constants to describe the appearance of full-length translated protein and full-length folded protein resistant to proteolysis:

$$\text{DNA template} \xrightarrow{k_1} \text{Transcript} \xrightarrow{k_2} \text{Translated protein}$$
$$\xrightarrow{k_3} \text{Folded protein}$$

Scheme 2

Please cite this article as: Lim Nicole C.H., Jackson Sophie E., Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.09.007

All reactions were considered separately to account for any differences in the PURExpress kit components that may alter the rate of translation or folding with or without the assistance of GroEL-GroES. The rate constant describing the steps involved in transcription (k_1) was assumed to be a constant under a given set of conditions and was a shared parameter when calculating k_{trans} and k_{fold} from a single reaction. Rate constants for translation (k_{trans}) and protein folding (k_{fold}) were estimated from the fit of this model to the data using KinTek Explorer [44], which allowed both the simulation and fitting of kinetic data to reaction Scheme 2. The errors quoted are the standard errors from the non-linear regression. For the ThiS-YibK and ThiS-YbeA translation reactions measured in the presence of GroEL-GroES, the rate of protein folding appeared limited by the rate of translation, and thus, a lower limit for k_{fold} was estimated from these data by simulating the folding reactions in KinTek Explorer.

Acknowledgements

We thank A. Mallam (University of Texas, Austin) for helpful discussions. N.C.H.L. is supported by a UBD Chancellor scholarship from the Brunei Government.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.09.007.

Received 3 July 2014;
Received in revised form 5 September 2014;
Accepted 6 September 2014

Keywords:
protein folding;
knotted;
chaperonins;
GroEL-GroES;
nascent chain

References

[1] Taylor WR. Protein knots and fold complexity: some new twists. Comput Biol Chem 2007;31:151–62.
[2] Mallam AL. How does a knotted protein fold? FEBS J 2009; 276:365–75.
[3] Boelinger D, Sulkowska JI, Hsu HP, Mirny LA, Kardar M, Onuchic JN, et al. A Stevedore's protein knot. PLoS Comput Biol 2010;6:e1000731.
[4] Virnau P, Mallam A, Jackson S. Structures and folding pathways of topologically knotted proteins. J Phys Condens Matter 2011;23:033101.
[5] Sulkowska JI, Rawdon EJ, Millett KC, Onuchic JN, Stasiak A. Conservation of complex knotting and slipknotting patterns in proteins. Proc Natl Acad Sci USA 2012;109:1715–23.
[6] Mallam AL, Jackson SE. Folding studies on a knotted protein. J Mol Biol 2005;346:1409–21.
[7] Mallam AL, Jackson SE. Probing nature's knots: the folding pathway of a knotted homodimeric protein. J Mol Biol 2006; 359:1420–36.
[8] Mallam AL, Jackson SE. The dimerization of an alpha/beta-knotted protein is essential for structure and function. Structure 2007;15:111–22.
[9] Mallam AL, Jackson SE. A comparison of the folding of two knotted proteins: YbeA and YibK. J Mol Biol 2007;366:650–65.
[10] Mallam AL, Morris ER, Jackson SE. Exploring knotting mechanisms in protein folding. Proc Natl Acad Sci USA 2008; 105:18740–5.
[11] Mallam AL, Onuoha SC, Grossmann JG, Jackson SE. Knotted fusion proteins reveal unexpected possibilities in protein folding. Mol Cell 2008;30:642–8.
[12] Mallam A, Rogers JM, Jackson S. Experimental detection of knotted conformations in denatured proteins. Proc Natl Acad Sci USA 2010;107:8189–94.
[13] Mallam AL, Jackson SE. Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. Nat Chem Biol 2012;8:147–53.
[14] Ohashi H, Kanamori T, Shimizu Y, Ueda T. A highly controllable reconstituted cell-free system—a breakthrough in protein synthesis research. Curr Pharm Biotechnol 2010; 11:267–71.
[15] Shimizu Y, Ueda T. In: Endo Y, Takai K, Ueda T, editors. PURE Technology. Cell-Free Protein Production: Methods and Protocols. Methods in Molecular Biology Humana Press; 2010. p. 11–21.
[16] Boelinger D, Sulkowska JI, Hsu HP, Mirny LA, Kardar M, Onuchic JN, et al. Cell-free translation reconstituted with purified components. Nat Biotechnol 2001;19:751–5.
[17] Shimizu Y, Kanamori T, Ueda T. Protein synthesis by pure translation systems. Methods 2005;36:299–304.
[18] Park CW, Marqusee S. Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. Nat Methods 2005;2:207–12.
[19] Park C, Marqusee S. Quantitative determination of protein stability and ligand binding by pulse proteolysis. Curr Protoc Protein Sci 2006;46:20.11.1–20.11.14.
[20] Faisca PFN, Travasso RDM, Charters T, Nunes A, Cieplak M. The folding of knotted proteins: insights from lattice simulations. Phys Biol 2010;7:16009.
[21] Noel JK, Sulkowska JI, Onuchic JN. Slipknotting upon native-like loop formation in a trefoil knot protein. Proc Natl Acad Sci USA 2010;107:15403–8.
[22] Berg JM, Tymoczko JL, Stryer L. Section 29.1: protein synthesis requires the translation of nucleotide sequence into amino acid sequences. Biochem 5th edit. New York: W H Freeman; 2002.
[23] Soler MA, Faisca PFN. How difficult is it to fold a knotted protein? *In silico* insights from surface-tethered folding experiments. PLoS One 2012;7:e52343.
[24] Wallin S, Zeldovich KB, Shakhnovich EI. The folding mechanics of a knotted protein. J Mol Biol 2007;368:884–93.
[25] Prentiss MC, Wales DJ, Wolynes PG. The energy landscape, folding pathways and the kinetics of a knotted protein. PLoS Comput Biol 2010;6:e1000835.
[26] Hartl FU, Hayer-Hartl M. Converging concepts of protein folding *in vitro* and *in vivo*. Nat Struct Mol Biol 2009;16:574–81.
[27] Horwich AL, Fenton WA. Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. Q Rev Biophys 2009;42:83–116.

[28] Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature 2011;475:324–32.
[29] Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. Molecular chaperone functions in protein folding and proteostasis. Annu Rev Biochem 2013;82:323–55.
[30] Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H. A systematic survey of *in vivo* obligate chaperonin-dependent substrates. EMBO J 2010;29:1552–64.
[31] Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, et al. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. Cell 2005;122:209–20.
[32] Dahiya V, Chaudhuri TK. Chaperones GroEL/GroES accelerate the refolding of a multidomain protein through modulating on-pathway intermediates. J Biol Chem 2014;289:286–98.
[33] Hartl FU, Hayer-Hartl M. Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 2002;295:1852–8.
[34] Lin Z, Rye HS. GroEL-mediated protein folding: making the impossible, possible. Crit Rev Biochem Mol Biol 2006;41:211–39.
[35] Weissman JS, Hohl CM, Kovalenko O, Kashi Y, Chen S, Braig K, et al. Mechanism of GroEL action—productive release of polypeptide from a sequestered position under GroES. Cell 1995;83:577–87.
[36] Weber F, Keppel F, Georgopoulos C, Hayer-Hartl MK, Hartl FU. The oligomeric structure of GroEL/GroES is required for biologically significant chaperonin function in protein folding. Nat Struct Biol 1998;5:977–85.
[37] Brinker A, Pfeifer G, Kerner MJ, Naylor DJ, Hartl FU, Hayer-Hartl M. Dual function of protein confinement in chaperonin-assisted protein folding. Cell 2001;107:223–33.
[38] Takagi F, Koga N, Takada S. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. Proc Natl Acad Sci USA 2003;100:11367–72.
[39] Thirumalai D, Lorimer GH. Chaperonin-mediated protein folding. Annu Rev Biophys Biomol Struct 2001;30:245–69.
[40] Lin Z, Madan D, Rye HS. GroEL stimulates protein folding through forced unfolding. Nat Struct Mol Biol 2008;15:303–11.
[41] Sharma S, Chakraborty K, Müller BK, Astola N, Tang YC, Lamb DC, et al. Monitoring protein conformation along the pathway of chaperonin-assisted folding. Cell 2008;133:142–53.
[42] Chakraborty K, Chatila M, Sinha J, Shi Q, Poschner BC, Sikor M, et al. Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. Cell 2010;142:112–22.
[43] Sulkowska JI, Sulkowski P, Onuchic J. Dodging the crisis of folding proteins with knots. Proc Natl Acad Sci USA 2009;106:3119–24.
[44] Johnson KA, Simpson ZB, Blom T. Global Kinetic Explorer: a new computer program for dynamic simulation and fitting of kinetic data. Anal Biochem 2009;387:20–9.

Please cite this article as: Lim Nicole C.H., Jackson Sophie E., Mechanistic Insights into the Folding of Knotted Proteins *In Vitro* and *In Vivo*, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.09.007