nature chemical biology

ARTICLE

PUBLISHED ONLINE: 18 DECEMBER 2011 | DOI: 10.1038/NCHEMBIO.742

Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins

Anna L Mallam[1,2] & Sophie E Jackson[1]*

Topological knots are found in a considerable number of protein structures, but it is not clear how they knot and fold within the cellular environment. We investigated the behavior of knotted protein molecules as they are first synthesized by the ribosome using a cell-free translation system. We found that newly translated knotted proteins can spontaneously self-tie and do not require the assistance of molecular chaperones to fold correctly to their trefoil-knotted structures. This process is slow but efficient, and we found no evidence of misfolded species. A kinetic analysis indicates that the knotting process is rate limiting, occurs post-translationally, and is specifically and significantly ($P < 0.001$) accelerated by the GroEL–GroES chaperonin complex. This demonstrates a new active mechanism for this molecular chaperone and suggests that chaperonin-catalyzed knotting probably dominates *in vivo*. These results explain how knotted protein structures have withstood evolutionary pressures despite their topological complexity.

Most newly translated proteins must fold to their native states to function *in vivo*. To facilitate this process, evolutionary pressures give rise to protein sequences that optimize folding efficiency as far as possible while maintaining function[1–3]. Although evidence suggests that the cooperative folding and smooth free-energy landscapes observed for small proteins are probably the product of natural selection[4,5], it is not obvious how this concept is applicable to topologically knotted structures. Proteins that contain trefoil, figure-of-eight, penta and stevedore knots with three, four, five and six projected crossings of the polypeptide backbone, respectively, have been observed in all three domains of life[6–9]. So far, no apparent functional benefit has been found for these complex structures over their unknotted counterparts, and it is unclear how they have survived the evolutionary pressure for efficient folding given the requirement for their polypeptide chains to thread to form their native, active states.

Much of our current understanding of how knotted proteins fold is based on studies on the bacterial α/β-knotted methyltransferases YibK from *Haemophilus influenzae* and YbeA from *Escherichia coli*[10–21]. YibK and YbeA are single-domain, homodimeric proteins that contain a right-handed trefoil knot in which at least 40 residues pass through a similarly sized loop (**Fig. 1**). Biophysical experiments to characterize their knotting and folding mechanisms have been done with purified, bacterially expressed recombinant proteins. These studies show that knotted proteins can be reversibly unfolded *in vitro* using chemical denaturants to a state with no detectable structure and can even tolerate the fusion of additional domains to both their N and C termini[12,13,16]. Recent experiments on circularized variants of YibK and YbeA, however, reveal that their polypeptide chains adopt trefoil-knotted conformations even in their chemically unfolded states[19]. This means that *in vitro* folding experiments on these purified recombinant proteins report on the refolding of a knotted denatured polypeptide chain into a native knotted structure. They give no insights into how the knot is first formed, which must occur during the initial expression of the proteins in *E. coli*. It is therefore necessary to investigate the folding of newly translated knotted proteins to provide information on the knotting mechanism. However, such studies *in vivo* are particularly challenging because of the complex environment of the cell[22].

In this study, we directly examined the behavior of knotted proteins as they were first synthesized by the ribosomal machinery. We used a coupled *in vitro* transcription-translation system, containing only the components necessary for protein translation, to synthesize YibK and YbeA. This approach allowed us to investigate the knotting and folding of these proteins in an environment similar to that found in the cell. During translation, the nascent polypeptide is initially produced as an unknotted linear chain that at some point must undergo a knotting event to fold and form the native, active structure. We also developed a strategy based on existing techniques used to measure the stability and ligand binding of proteins in complex solutions[23–25] to simultaneously monitor the translation reaction and the appearance of folded protein. We used this new methodology to measure the folding rates of polypeptide chains of YibK and YbeA after they were synthesized.

We found that newly translated knotted protein molecules can self-tie and do not require the assistance of molecular chaperones to correctly fold into their trefoil-knotted structures. This spontaneous folding is efficient, and there is no evidence for misfolded species. We demonstrate that knotting of the polypeptide chain is rate limiting, occurs post-translationally, and is significantly enhanced by the bacterial chaperonin GroEL–GroES. We found evidence suggesting that encapsulation in the chaperonin cage leads to a greater rate of knot formation, representing a new mode of action by which the GroEL–GroES system actively promotes folding. Furthermore, our data indicate that the intermediate species in chaperonin-catalyzed knot formation may mimic that of a knotted chemically denatured state. By selectively probing the folding of YibK and YbeA as they are first synthesized, we experimentally characterized important aspects of the polypeptide knotting mechanism that are relevant to the cellular environment. We also show how both the efficient spontaneous and chaperonin-catalyzed knotting of these proteins explains how they withstand evolutionary pressure despite their complicated topology.

RESULTS

Knotted proteins can be synthesized *in vitro*

To examine the folding of newly translated chains of YibK and YbeA in a controlled environment, we used a reconstituted cell-free

[1]Department of Chemistry, The University of Cambridge, Cambridge, UK. [2]Present address: Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, Texas, USA. *e-mail: sej13@cam.ac.uk

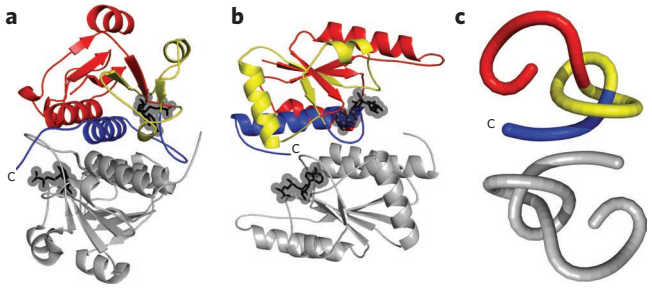

Figure 1 | Structures of the trefoil-knotted methyltransferases YibK and YbeA. (**a**,**b**) Ribbon diagrams of homodimeric knotted proteins YibK (160 residues, PDB 1MXI, **a**) and YbeA (155 residues, PDB 1NS5, **b**). One subunit of each protein is highlighted to show the deep trefoil knot in the structure; the knotting loop is yellow, and the chain knotted at the C terminus is blue. Both the knotting loop and the knotted chain are ~40 residues long. The methyltransferase cofactor, AdoHcy, is shown as a ball-and-stick model to indicate the cofactor-binding site in the knotted region of the proteins[10,29]. (**c**) Simplified representation of path of polypeptide chain in **a** to show the position of the right-handed trefoil knot within the structure.

translation system consisting of only those purified components required for transcription and translation in *E. coli*[26,27]. These include translation factors, aminoacyl-tRNA synthetases, T7 RNA polymerase and enzymes for energy recycling along with 70S ribosomes, amino acids, rNTPs and tRNAs. This system lacks the intrinsic molecular chaperones found in a bacterial cell, and so the effect of these on newly translated polypeptide chains can be investigated[27,28]. *In vitro* synthesis of YibK and YbeA was initiated by addition of the appropriate template DNA and was allowed to proceed at 37 °C for 2 h. We analyzed the products of translation by SDS-PAGE and visualized them with SYPRO Ruby stain (**Supplementary Results**, **Supplementary Fig. 1**). Bands corresponding to YibK and YbeA were prominent compared with a control reaction with no template DNA. In cell-free translation systems, components can be manipulated to generate proteins with useful properties[27]. We added the modified lysine tRNA BODIPY-Lys-$tRNA_{Lys}$ to the translation reaction to incorporate the fluorophore BODIPY-FL into YibK and YbeA. We confirmed these translation products by visualization after SDS-PAGE (**Supplementary Fig. 1**). We found that *in vitro*–translated YibK and YbeA were highly soluble, and ~0.14 mg of protein was produced per milliliter of translation reaction (**Supplementary Fig. 2**). This yield is comparable to that of other proteins that have been synthesized using an equivalent cell-free system[26].

Newly translated knotted proteins can fold spontaneously

We investigated the structure, stability and cofactor binding of *in vitro*–translated YibK and YbeA to assess whether the newly synthesized proteins can successfully fold to their native knotted state. We compared these results to the behavior of the same proteins that were recombinantly produced in *E. coli* and that are known to correctly fold[12–14]. We used analytical size-exclusion chromatography (SEC) to determine the oligomeric state of YibK and YbeA translated in the cell-free system. These proteins form highly stable native dimers (K_d < 1 nM) when they are expressed in bacteria[11–14]. We monitored SEC elution profiles of the *in vitro*–synthesized proteins by optical density at 280 nm (OD_{280}) and did SDS-PAGE analysis on acetone-precipitated fractions of eluted protein (**Fig. 2** and **Supplementary Fig. 3**). We selectively monitored elution profiles of BODIPY-FL-labeled proteins during SEC by OD_{502}, and in SDS-PAGE by fluorescence emission through a 520-nm–bandpass filter. We found that proteins synthesized *in vivo* in *E. coli* and *in vitro* using the cell-free translation system eluted from the column at the same volume. Elution volumes were 10.3 ml and 10.5 ml for YibK and YbeA, respectively, which are consistent with those expected for dimeric species (**Supplementary Fig. 4**). We observed no monomeric protein, and the peaks that eluted at volumes greater than 10.3 ml or 10.5 ml did not correspond to YibK or YbeA in SDS-PAGE analysis (**Fig. 2b–i**). These experiments establish that *in vitro*–synthesized knotted proteins fold to native-like dimers in solution and that the dimer interface, which is located near the knotted core in YibK and YbeA, is correctly formed.

We used intrinsic protein fluorescence to assess the internal packing arrangements and the tertiary and quaternary structure of the knotted proteins produced using the cell-free system compared with those produced in *E. coli*. Fluorescence spectra measured for *in vitro*–synthesized YibK and YbeA that were further purified by SEC are very similar to those recorded for the bacterially expressed proteins and are consistent with the presence of the correct tertiary and quaternary contacts (**Supplementary Fig. 5**). This suggests that *in vitro*– and *in vivo*–synthesized knotted proteins adopt equivalent global structures.

A change in the stability of knotted proteins synthesized *in vitro* compared with their bacterially expressed equivalents could indicate a difference in structure or the presence of a globular but misfolded species. We used pulse-proteolysis experiments to determine the free energy of unfolding, $\Delta G_{H_2O}^{N_2 \leftrightarrow 2D}$ (where N_2 is a native dimer and D is a denatured monomer), of YibK and YbeA produced using both the cell-free system and *E. coli* expression[24]. This technique requires only small amounts of sample and can be carried out in lysate mixtures[23]. We estimated $\Delta G_{H_2O}^{N_2 \leftrightarrow 2D}$ values for the *in vitro*–synthesized proteins without purification from the other components in the cell-free translation system. Pre-equilibrated mixtures of protein and urea were incubated for 1 min with the protease thermolysin, which selectively digests any unfolded protein (**Supplementary Methods**). We quantified the remaining folded protein in the sample by SDS-PAGE to determine the fraction of folded molecules present at different concentrations of urea (**Supplementary Fig. 6**), and estimated protein stability (**Fig. 3** and **Table 1**). We also obtained denaturation profiles and $\Delta G_{H_2O}^{N_2 \leftrightarrow 2D}$ values for bacterially expressed and purified YibK and YbeA in the same manner using pulse proteolysis; these compared well with previous results[11,13] using intrinsic protein fluorescence to probe the folded state of the protein (**Supplementary Fig. 7**). This suggests that the pulse-proteolysis experiments are sensitive to the presence of native structure and can be used to estimate the global stability of these proteins. Notably, the denaturation profiles and the values for $\Delta G_{H_2O}^{N_2 \leftrightarrow 2D}$ determined from pulse proteolysis for *in vitro*–synthesized YibK and YbeA are in agreement with those measured under the same conditions for the knotted proteins expressed in *E. coli* (**Fig. 3** and **Table 1**). This is consistent with the results above and confirms that the native knotted structure is present in the majority of protein molecules produced in the cell-free system.

YibK and YbeA bind the cofactor *S*-adenosylmethionine (AdoMet) in a binding pocket formed by the knotted region of the protein (**Fig. 1a**,**b**)[10,29]. The binding affinity for *S*-adenosyl homocysteine (AdoHcy), the product of AdoMet after methyl-group transfer to the substrate has taken place, can be used to confirm the integrity of both the cofactor-binding site and the native, knotted structure[14,16,19]. AdoHcy selectively binds the native, dimeric state of knotted methyltransferases, and, therefore, association with this ligand should lead to an apparent increase in the stability of YibK and YbeA (**Supplementary Methods**)[14,24]. We used pulse-proteolysis experiments to measure the stability of the knotted proteins in the presence of AdoHcy and to assess cofactor-binding affinity. We observed an apparent increase in the stability of both *in vitro*– and *in vivo*–synthesized YibK and YbeA upon addition of AdoHcy (**Fig. 3**, **Table 1** and **Supplementary Fig. 6**). We used this increase to estimate cofactor dissociation constants of 21–50 μM and 2.7–18 μM for YibK and YbeA, respectively. These values are in reasonable

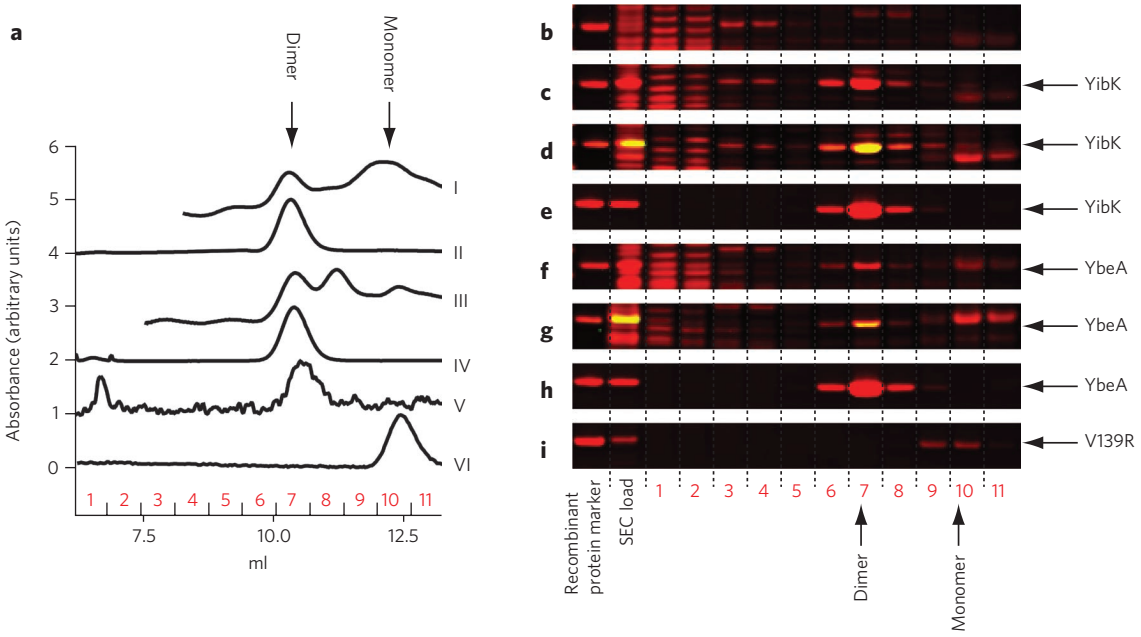

Figure 2 | Characterization of the oligomeric state of *in vitro*-translated knotted proteins. (**a**–**i**) SEC elution profiles were measured by absorbance (**a**), and SDS-PAGE analysis of acetone-precipitated fractions was conducted (**b**–**i**; also see **Supplementary Fig. 3**). Expected elution volumes for dimeric and monomeric YibK, arrows; see **Supplementary Figure 4** for calibration curve. (**a**) Elution profile of the *in vitro* synthesis reactions of YibK (I) and YbeA (III). Profiles of purified recombinant dimeric YibK (II) and YbeA (IV) are shown for comparison, as well as a monomeric mutant of YibK, V139R[14] (VI). A profile for YibK translated *in vitro* in the presence of BODIPY-Lys-tRNA$_{Lys}$ (V) is also shown. OD_{280} was measured, except for V, for which OD_{502} was measured. (**b**–**d**,**f**,**g**) SDS-PAGE of SEC profiles for control cell-free expression reaction with no DNA template (**b**), *in vitro* synthesis of YibK (**c**), YibK synthesized *in vitro* in the presence of BODIPY-Lys-tRNA$_{Lys}$ (**d**), *in vitro*–synthesized YbeA (**f**) and YbeA synthesized *in vitro* in the presence of BODIPY-Lys-tRNA$_{Lys}$ (**g**). (**e**,**h**,**i**) SDS-PAGE analysis of SEC profiles of purified, bacterially expressed YibK (**e**), YbeA (**h**) and the monomeric mutant YibK V139R (**i**) are included for comparison. BODIPY-FL fluorescence is yellow. Each gel shows a purified recombinant protein marker of either YibK or YbeA (lane 1), total SEC load (lane 2) and fractions 1–11 of the SEC elution profile.

agreement with those measured previously using isothermal titration calorimetry for the purified knotted proteins expressed in *E. coli* (K_d values of 20 μM and 2.5 μM for YibK and YbeA, respectively)[19]. The stability measurements therefore suggest that both YibK and YbeA expressed in the cell-free system adopt a structure that binds the cofactor AdoHcy.

Taken together, our probes of structure, stability and cofactor binding indicate that the newly translated knotted proteins can spontaneously form the correct knotted, homodimeric, native structure without molecular chaperones. Folding efficiency is a measure of the fraction of unfolded protein molecules that fold correctly to form the native, functional state. Misfolded species can manifest in several ways. First, SDS-PAGE analysis of the fractions from the size-exclusion column would indicate any molecules of YibK or YbeA that are either monomeric or higher oligomers and therefore do not have the dimeric structure of the native state; we observed no such species. The stability and ligand-binding measurements would show whether a substantial fraction of molecules had misfolded to a proteolytically resistant state; these would not bind ligand and would have a different chemical denaturation profile from the correctly folded native state. We found no evidence of misfolded species, and thus the folding reaction of *in vitro*–synthesized knotted proteins occurs with high efficiency.

Spontaneous knotting after translation is slow

The above results indicate that the folding and knotting of newly translated polypeptide chains of YibK and YbeA can be probed with a reconstituted cell-free expression system. We used a modified pulse-proteolysis method to monitor the time course of *in vitro* translation and subsequent folding of YibK and YbeA. We removed aliquots of the translation reaction at various time points after addition of plasmid DNA and halted protein synthesis by adding chloramphenicol. At this instant, fully synthesized knotted protein molecules were present in both unfolded and folded states. We immediately subjected half of the reaction mixture to pulse proteolysis to digest any unfolded protein. We analyzed both undigested and digested samples by SDS-PAGE to obtain a time course for the appearance of full-length translated protein and full-length folded, knotted protein. We fit these kinetic data to a simple model to estimate rate constants for the translation (k_{trans}) and protein folding (k_{fold}; **Supplementary Methods**). For both

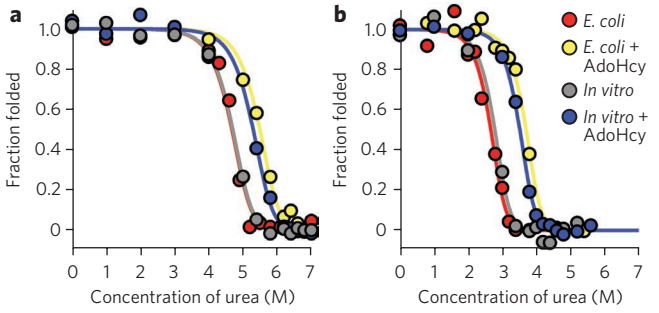

Figure 3 | Analysis of stability of knotted proteins produced by cell-free translation. (**a**,**b**) Pulse-proteolysis experiments for YibK (**a**) and YbeA (**b**) expressed in *E. coli* and purified to homogeneity or translated *in vitro*. Measurements were made without and with methyltransferase cofactor AdoHcy and at 1 μM protein. Fraction of unfolded protein, determined by pulse proteolysis with thermolysin, is shown. Solid lines represent the fit of data to a two-state equilibrium dimer unfolding model (**Supplementary Methods**).

Table 1 | Thermodynamic parameters for the unfolding of *in vitro*– or *in vivo*–translated knotted proteins

Expression conditions	Concentration of AdoHcy (mM)	C_m (M)	m_{app} (kcal mol^{-1} M^{-1})	$\Delta G^{N_2 \leftrightarrow 2D}_{H_2O}$ (kcal mol^{-1})	K_d^{AdoHcy} (μM)
YibK					
In vivo recombinant expression in *E. coli*	0	4.7 ± 0.1	2.9 ± 0.1	21.8 ± 0.5	-
	1.0	5.5 ± 0.1		24.1 ± 0.6	21 ± 9
In vitro expression in a cell-free system	0	4.7 ± 0.1		21.8 ± 0.5	-
	1.0	5.3 ± 0.1		23.6 ± 0.6	50 ± 20
YbeA					
In vivo recombinant expression in *E. coli*	0	2.7 ± 0.1	3.7 ± 0.3	18.2 ± 0.9	-
	1.4	3.7 ± 0.1		21.9 ± 1.2	2.7 ± 0.9
In vitro expression in a cell-free system	0	2.8 ± 0.1		18.5 ± 0.9	-
	1.4	3.5 ± 0.1		21.1 ± 1.1	18 ± 6

C_m is the concentration of urea at the midpoint of the unfolding transition. m_{app} is the apparent m value, a measure of the slope of the unfolding transition, and it was determined separately for YibK and YbeA from pulse-proteolysis experiments. This parameter relates to the solvent-accessible surface area change that occurs upon unfolding and should not change significantly between data sets for the same protein. The value for m_{app} was therefore shared during the analysis of denaturation curves for each protein; this was necessary to reduce the number of variable parameters and prevent overfitting of the data. For measurements made with the cofactor AdoHcy, the apparent stability in the presence of ligand is quoted.

proteins, we observed a delay of about 10–20 min between the formation of translated product and the formation of folded protein (**Fig. 4a**,**b**). Rate constants for the folding of polypeptide chains of YibK and YbeA to native-like structure, as assessed by resistance to proteolysis, were estimated to be 0.05 min^{-1} and 0.09 min^{-1}, respectively (**Supplementary Table 1**). We also monitored the refolding of purified, bacterially expressed YibK and YbeA from a urea-denatured state under conditions identical to those used in the translation reaction (**Supplementary Fig. 8**). Refolding was considerably faster in these cases (k_{fold} of 1.8 min^{-1} and 0.3 min^{-1} for YibK and YbeA, respectively). This result can be considered in the context of our recent work showing that chemically denatured knotted proteins retain a knotted conformation[19]. In comparison, a newly translated chain must at first be knot free. The ~3- to 35-fold greater rate constant for refolding for a knotted denatured state compared with an unknotted denatured state suggests that knotting of the polypeptide chain is rate limiting in the folding of newly translated knotted proteins.

Knotting is catalyzed by bacterial chaperonins

These data establish that newly synthesized YibK and YbeA can spontaneously knot and fold independently of cellular protein folding factors. However, the presence of molecular chaperones could facilitate or hinder the knotting and/or folding of these proteins. We investigated the effect of the bacterial chaperonin GroEL in complex with its cofactor GroES on the folding reaction of knotted proteins as they are translated *in vitro*. GroEL is a large, double-ringed cylindrical complex, with each ring comprising seven monomeric subunits, each ~57 kDa (ref. 30). GroES is a heptameric ring of subunits of ~10 kDa that caps the GroEL cylinder to form a chaperonin cage that can encapsulate a protein substrate[30]. Time courses for the appearance of full-length translated protein and the appearance of full-length folded protein were measured as described above with GroEL–GroES (0.15 μM complex) included in the reaction mixture (**Fig. 4c**,**d**). In these experiments, we observed no delay between the formation of translated protein and the formation of folded knotted protein, suggesting that folding is now limited by translation. The estimated rate constants for protein folding were at least 20-fold higher under these conditions (a lower limit for k_{fold} of 2 min^{-1} was estimated from the data for both YibK and YbeA), indicating that the folding rate of newly synthesized knotted proteins increased significantly in the presence of the chaperonin (**Supplementary Table 1**). This can be compared to the rates of refolding measured by pulse proteolysis in the presence of GroEL–GroES from the urea-denatured states of purified, bacterially expressed YibK and YbeA, which are knotted (**Supplementary Fig. 8**). In these cases, the rate constants for refolding remained unaltered by GroEL–GroES. This means that the chaperonin specifically increases the rate of structure formation only in newly translated knotted proteins. Because chemically denatured YibK and YbeA retain a knotted conformation[19], these data indicate that GroEL–GroES specifically enhances the rate of knotting in the unknotted newly synthesized

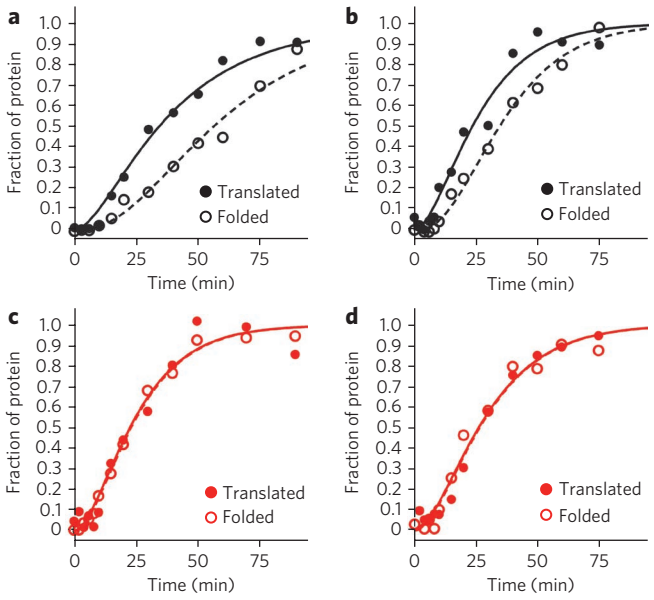

Figure 4 | Kinetics of translation and folding for *in vitro*–synthesized knotted proteins. (**a**–**d**) Representative time courses for the appearance of translated and folded YibK (**a**,**c**) and YbeA (**b**,**d**) in a cell-free translation reaction. These were characterized without (**a**,**b**; black) and with (**c**,**d**; red) the GroEL–GroES bacterial chaperonin (0.15 μM complex). Fit of the kinetic data to a simplified reaction scheme (**Methods**) is shown to describe appearance of translated protein formation of translated-folded protein; this was used to estimate rate constants for translation and protein folding reactions (**Supplementary Table 1**). In **c** and **d**, time courses for the appearance of translated protein and folded protein in the presence of GroEL–GroES are very similar (solid and dashed lines overlay).

polypeptide chains of YibK and YbeA. Chaperonins such as the GroEL–GroES complex are thought to facilitate the folding of protein substrates in a post-translational manner[30,31]. Consequently, the observed effect of the chaperonin on the knotting of newly translated YibK and YbeA suggests that knot formation is a post-translational event.

DISCUSSION

It is unclear how topologically knotted proteins have withstood evolutionary pressures for efficient folding within the cell, especially as they have no obvious functional advantage over their unknotted equivalents[8]. Here we have shown that newly translated polypeptide chains of the bacterial trefoil-knotted methyltransferases YibK and YbeA fold into their native, proteolytically resistant states capable of cofactor binding and, thus, establish unequivocally that molecular chaperones are not essential for the knotting and folding of these proteins. Despite their complex topology, YibK and YbeA fold spontaneously with a high yield, and we found no evidence for misfolded or aggregated species during or after translation. These experimental data confirm that the knotting and folding pathways of these structures are robust, even when folding initiates from an unknotted, newly translated state. This is consistent with recent simulations of all-atom representations of YibK and the smaller (82-residue) trefoil-knotted protein MJ0366 from *Methanocaldococcus jannaschi*, which suggest that a native-biased free-energy landscape is sufficient for the successful folding of these knotted proteins[17,32].

The half-lives for the spontaneous knotting and folding of unknotted, newly translated polypeptide chains of YibK and YbeA are ~15 min and ~8 min, respectively. This is considerably slower than many folding rates observed *in vitro* for single-domain proteins with simple topologies[33] and is in agreement with the slow folding kinetics reported in earlier theoretical folding simulations of knotted proteins from an unknotted state[17,18,34]. Notably, folding is ~3–35 times slower for an unknotted chain than for a knotted one, indicating that the knotting step is responsible for the slow, spontaneous folding of YibK and YbeA after translation. This corroborates theoretical folding simulations of the trefoil-knotted protein MJ0366 (ref. 32) and a truncated model of the tRNA methyltransferase TrmD from *Haemophilus influenzae*[34], as well as *in vitro* experimental studies on a designed knotted protein[35], all of which suggest a rate-limiting knotting event during folding. Many larger proteins with complex (but unknotted) topologies have slow folding rates because they are often stabilized by long-range contacts, which give rise to a large entropic barrier to folding[36,37]. The slow folding rates observed here for YibK and YbeA that are limited by knotting of the polypeptide chain are consistent with a large entropic contribution to the folding barrier during the knotting step.

Our results demonstrate that GroEL–GroES increases the rate of spontaneous folding in newly translated polypeptide chains of YibK and YbeA by at least 20-fold, and under these conditions folding seems limited by the rate of translation (**Fig. 4** and **Supplementary Table 1**). Notably, the size and complex α/β topology of YibK and YbeA are typical of many proteins that have been observed to interact with GroEL[30,38]. The chaperonin has an effect only on the folding of newly translated, unknotted polypeptide chains and does not further enhance the folding of chemically denatured, knotted chains. This suggests a specific post-translational mechanism in which the chaperonin has a direct role in catalyzing protein knot formation (rather than folding) and suggests a new active chaperonin mechanism. Our findings corroborate proteomic studies that have identified another similar bacterial knotted protein, the *E. coli* methyltransferase TrmD, as a GroEL–GroES substrate[38]. Consistent with the results described here, this protein also remains soluble even in GroEL–GroES–depleted cells[39], suggesting that, like YibK and YbeA, it can fold without the chaperonin.

As the *in vitro* translation experiments somewhat mimic the bacterial cellular environment, we propose that knotting and folding of newly translated topologically knotted proteins occurs *in vivo* by a similarly efficient mechanism that can be catalyzed by a chaperonin (**Fig. 5**). This is substantiated by the high yields of soluble, native protein obtained from the bacterial expression of both YibK and YbeA, which are about 70 mg and 40 mg per liter of culture, respectively[12,13]. The markedly faster knotting speed of YibK and YbeA in the presence of GroEL–GroES suggests that chaperonin-catalyzed knotting pathways are probably fast relative to translation

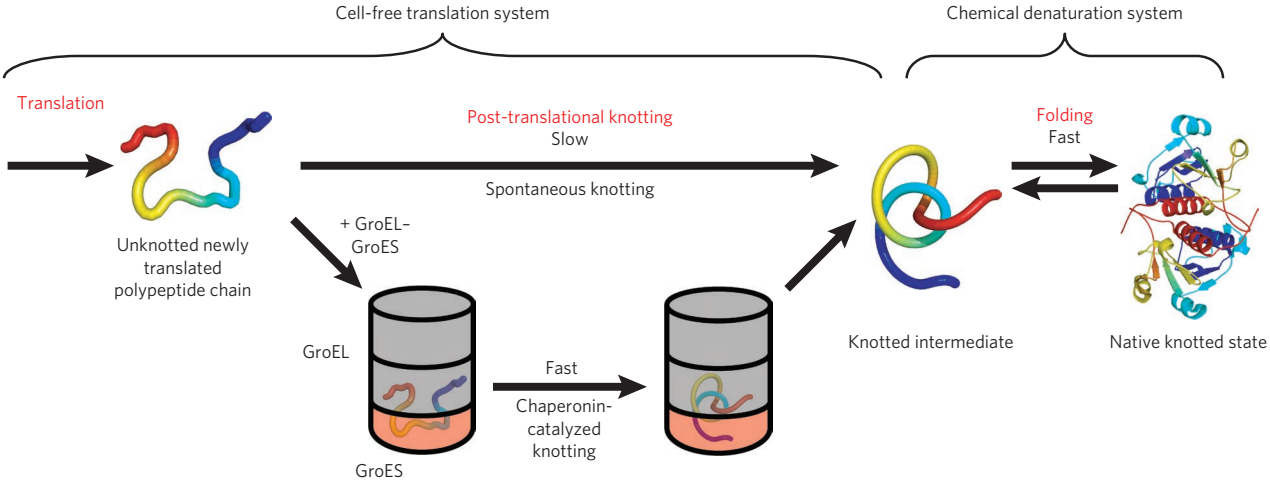

Figure 5 | A model for the knotting and folding of newly translated topologically knotted proteins in bacteria. The rate of spontaneous folding from a newly translated unknotted polypeptide chain is slow and depends on GroEL–GroES, indicating that knotting is rate limiting and occurs post-translationally. Therefore, after synthesis of the polypeptide chain, post-translational knotting can be either spontaneous or catalyzed by the chaperonin; however, knotting is >20-fold faster in the presence of GroEL-GroES, suggesting that this folding pathway probably dominates *in vivo*. In comparison, folding experiments *in vitro* show that knotted proteins reversibly unfold to a chemically denatured state in which the chain remains in a trefoil-knotted conformation. Folding from a knotted chemically denatured state is fast and chaperonin independent, suggesting that this state could mimic the GroEL–GroES–induced knotted polypeptide chain. The unknotting of a knotted protein has yet to be experimentally characterized, and the conditions under which knotted polypeptide chains unknot remain to be determined; they could remain in a kinetically trapped knotted state.

and dominate *in vivo*. These observations explain the conservation of knotted structures throughout evolution despite their complexity. In addition, earlier studies have proposed that chaperonins such as GroEL–GroES have a 'buffering' function to allow for the folding of proteins with deleterious mutations[30,38,40], which is thought to be important in the evolution of new proteins. The ability of GroEL–GroES to promote knot formation suggests a possible role of the chaperonin in the evolution of topologically knotted structures.

GroEL–GroES is thought to actively accelerate protein folding by encapsulating folding intermediates to limit their conformational entropy[30,41]. Steric confinement within the chaperonin compartment probably decreases the number of accessible conformations available to the unfolded and unknotted chains of YibK and YbeA; this would reduce the entropic barrier to knotting and increase the probability of an initial threading event to populate a knotted intermediate state. This idea is consistent with existing models of chaperone-mediated protein folding[30,41–43] and numerical studies on model homopolymers that have shown an increase in the probability of knotting upon confinement[44,45]. Because we did not observe a rate enhancement when knotted denatured chains folded in the presence of chaperonin, our data suggest that the chaperonin cage induces a knotted intermediate state that may be similar to that of the knotted chemically denatured state (**Fig. 5**). No further rate enhancement by GroEL–GroES is observed when folding takes place from a knotted denatured state, as the data here and from previous studies show that structure formation during this process is relatively fast[19].

Notably, the unknotting of a knotted polypeptide chain has yet to be observed during experimental studies, even under highly denaturing conditions[19]. This raises the possibility that knotted structures thread only once after the polypeptide chain has been synthesized, and, after this, during subsequent unfolding and refolding events, the protein chain remains in a knotted state that may reflect kinetic trapping. This model can account for all of the experimental observations so far on the folding mechanism of YibK and YbeA and, in particular, the successful refolding reactions seen from both a knotted chemically denatured state and an unknotted newly translated polypeptide chain[11,13,19]. By considering together the results from refolding studies on chemically denatured proteins and on newly synthesized polypeptide chains, we have gained important information about the knotting and folding mechanisms of YibK and YbeA *in vivo* (**Fig. 5**) and have demonstrated a new activity for the chaperonin GroEL–GroES. In addition, the methodologies developed here can be applied to probe the folding reactions of many other proteins directly after translation. They, therefore, could be useful in future studies aimed at understanding the folding of nascent polypeptide chains and the role of molecular chaperones.

METHODS

Cell-free protein expression. The coupled *in vitro* transcription and translation of YibK and YbeA was carried out using the PURExpress system (New England Biolabs). This system is reconstituted from only the purified components required for *E. coli* expression and, therefore, does not contain any molecular chaperones[26–28]. Template plasmid DNA (250 ng) containing the gene encoding either YibK or YbeA inserted into a pET-17b vector (Novagen) was added to the protein synthesis reaction (25 µl) and incubated at 37 °C for 2 h. Protein synthesis was also carried out in the presence of BODIPY-Lys-tRNA$_{Lys}$ (FluoroTec GreenLys *in vitro* Translation Labeling System; Promega) to incorporate fluorescently labeled lysine residues into the nascent chains during translation. 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 5 min before SDS-PAGE. Protein gels were visualized with a Typhoon Imager (GE Healthcare). Yields from *in vitro* translation reactions were quantified using Image Quant (GE Healthcare) and a protein standard of known concentration of either purified YibK or YbeA produced from expression in *E. coli*. After synthesis, samples of reaction mixtures were centrifuged at 14,000*g* for 30 min. The total protein, precipitate protein and supernatant protein fractions of each sample were analyzed by SDS-PAGE to verify the solubility of YibK and YbeA expressed *in vitro* (**Supplementary Fig. 2**).

Protein characterization. *In vitro*– and *in vivo*–expressed YibK and YbeA were characterized under identical conditions by SEC and spectroscopic measurements (see **Supplementary Methods**). The stability of YibK and YbeA expressed *in vitro* and *in vivo* was measured using pulse-proteolysis experiments[24]. This method takes advantage of the fact that folded structures are more resistant to proteolysis than unfolded conformations. 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, 10 mM $CaCl_2$. Samples of either completed *in vitro* translation reactions or purified recombinant proteins were added to aliquots of urea at final denaturant concentrations ranging between 0 M and 7 M and a final protein concentration of 1 µM and were incubated at 25 °C overnight. We used the protease thermolysin for pulse-proteolysis measurements, as it selectively digests only unfolded YibK and YbeA under the conditions required to measure their stability. This was confirmed by preliminary proteolysis experiments, which indicated that for short exposure times all the native protein remained intact compared with an untreated sample, whereas unfolded protein was completely digested. 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 protein. The proteolysis reaction was then quenched by addition of EDTA to a final concentration of 12.5 mM. Each quenched reaction 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 concentration of urea.

Translation and folding kinetics. The pulse-proteolysis methods described above were modified to measure the time course for the appearance of newly translated, folded knotted proteins. *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 to a concentration of 2 mM. Half of this quenched translation reaction was immediately subjected to pulse proteolysis to digest any full-length translated protein that was still unfolded. 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. These measurements were repeated under identical translation conditions in the presence of GroEL–GroES chaperonin complex (0.15 µM; Stressgen Biotechnologies). A greatly simplified sequential reaction scheme comprising three consecutive, irreversible steps was used to model the kinetic data and to estimate the rate constants to describe the appearance of full-length translated protein and full-length folded protein resistant to proteolysis:

$$\text{DNA template} \xrightarrow{k_1} \text{Transcript} \xrightarrow{k_{trans}} \text{Translated protein} \xrightarrow{k_{fold}} \text{Folded protein}$$

The kinetic data were modeled using this scheme in KinTek Explorer[46]. The kinetic rate constants for translation (k_{trans}) and protein folding (k_{fold}) were extracted by obtaining the best fit of the model to the pulse-proteolysis data measured for the appearance of translated protein and the appearance of folded protein (**Fig. 4** and **Supplementary Table 1**). Each reaction was considered separately to account for possible small differences in conditions, for example the amount of energy source present. During analysis, data sets for the time course for the appearance of translated and translated-folded protein after addition of DNA template for a given reaction were modeled together, and k_1 was a shared parameter. k_1 describes the lag between the addition of the DNA template and the appearance of translated protein and probably reflects the many rate constants involved in the transcription process.

Figures. Structural figures were made using PyMOL (http://www.pymol.org/). Knot representations were generated using KnotPlot (http://www.knotplot.com/).

Other methods. Any remaining experimental procedures are described in **Supplementary Methods**.

Received 21 July 2011; accepted 27 September 2011; published online 18 December 2011

References

1. Onuchic, J.N. & Wolynes, P.G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).
2. Dill, K.A. & Chan, H.S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
3. Friel, C.T., Smith, D.A., Vendruscolo, M., Gsponer, J. & Radford, S.E. The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary contraints. *Nat. Struct. Mol. Biol.* **16**, 318–324 (2009).
4. Watters, A.L. *et al.* The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* **128**, 613–624 (2007).
5. Baker, D. A surprising simplicity to protein folding. *Nature* **405**, 39–42 (2000).
6. Bölinger, D. *et al.* A Stevedore's protein knot. *PLOS Comput. Biol.* **6**, e1000731 (2010).

7. Taylor, W.R. Protein knots and fold complexity: some new twists. *Comput. Biol. Chem.* **31**, 151–162 (2007).
8. Virnau, P., Mirny, L.A. & Kardar, M. Intricate knots in proteins: function and evolution. *PLoS Comput. Biol.* **2**, e122 (2006).
9. Virnau, P., Mallam, A.L. & Jackson, S.E. Structures and folding pathways of topologically knotted proteins. *J. Phys. Condens. Matter* **23**, 033101 (2011).
10. Lim, K. *et al.* Structure of the YibK methyltransferase from *Haemophilus influenzae* (HI0766): a cofactor bound at a site formed by a knot. *Proteins* **51**, 56–67 (2003).
11. Mallam, A.L. & Jackson, S.E. Folding studies on a knotted protein. *J. Mol. Biol.* **346**, 1409–1421 (2005).
12. Mallam, A.L. & Jackson, S.E. Probing Nature's knots: the folding pathway of a knotted homodimeric protein. *J. Mol. Biol.* **359**, 1420–1436 (2006).
13. Mallam, A.L. & Jackson, S.E. A comparison of the folding of two knotted proteins: YbeA and YibK. *J. Mol. Biol.* **366**, 650–665 (2007).
14. Mallam, A.L. & Jackson, S.E. The dimerization of an α/β-knotted protein is essential for structure and function. *Structure* **15**, 111–122 (2007).
15. Mallam, A.L., Morris, E.R. & Jackson, S.E. Exploring knotting mechanisms in protein folding. *Proc. Natl. Acad. Sci. USA* **105**, 18740–18745 (2008).
16. Mallam, A.L., Onuoha, S.C., Grossmann, J.G. & Jackson, S.E. Knotted fusion proteins reveal unexpected possibilities in protein folding. *Mol. Cell* **30**, 642–648 (2008).
17. Sułkowska, J.I., Sulkowski, P. & Onuchic, J. Dodging the crisis of folding proteins with knots. *Proc. Natl. Acad. Sci. USA* **106**, 3119–3124 (2009).
18. Wallin, S., Zeldovich, K.B. & Shakhnovich, E.I. The folding mechanics of a knotted protein. *J. Mol. Biol.* **368**, 884–893 (2007).
19. Mallam, A.L., Rogers, J.M. & Jackson, S.E. Experimental detection of knotted conformations in denatured proteins. *Proc. Natl. Acad. Sci. USA* **107**, 8189–8194 (2010).
20. Sułkowska, J.I., Sulkowska, P., Szymczak, P. & Cieplak, M. Untying Knots in Proteins. *J. Am. Chem. Soc.* **132**, 13954–13956 (2010).
21. Mallam, A.L. How does a knotted protein fold? *FEBS J.* **276**, 365–375 (2009).
22. Cabrita, L.D., Dobson, C.M. & Christodoulou, J. Protein folding on the ribosome. *Curr. Opin. Struct. Biol.* **20**, 33–45 (2010).
23. Kim, M.S., Song, J. & Park, C. Determining protein stability in cell lysates by pulse proteolysis and western blotting. *Protein Sci.* **18**, 1051–1059 (2009).
24. Park, C. & Marqusee, S. Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. *Nat. Methods* **2**, 207–212 (2005).
25. Schlebach, J.P., Kim, M.S., Joh, N.H., Bowie, J.U. & Park, C. Probing membrane protein unfolding with pulse proteolysis. *J. Mol. Biol.* **406**, 545–551 (2011).
26. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
27. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* **36**, 299–304 (2005).
28. Ying, B.W., Taguchi, H., Ueda, H. & Ueda, T. Chaperone-assisted folding of a single-chain antibody in a reconstituted translation system. *Biochem. Biophys. Res. Commun.* **320**, 1359–1364 (2004).
29. Purta, E., Kaminska, K.H., Kasprzak, J.M., Bujnicki, J.M. & Douthwaite, S. YbeA is the m3Psi methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA. *RNA* **14**, 2234–2244 (2008).
30. Hartl, F.U. & Hayer-Hartl, M. Converging concepts of protein folding *in vitro* and *in vivo*. *Nat. Struct. Mol. Biol.* **16**, 574–581 (2009).
31. Hartl, F.U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).
32. Noel, J.K., Sulkowska, J.I. & Onuchic, J.N. Slipknotting upon native-like loop formation in a trefoil knot protein. *Proc. Natl. Acad. Sci. USA* **107**, 15403–15408 (2010).
33. Jackson, S.E. How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91 (1998).
34. Prentiss, M.C., Wales, D.J. & Wolynes, P.G. The energy landscape, folding pathways and the kinetics of a knotted protein. *PLOS Comput. Biol.* **6**, e1000835 (2010).
35. King, N.P., Jacobitz, A.W., Sawaya, M.R., Goldschmidt, L. & Yeates, T.O. Structure and folding of a designed knotted protein. *Proc. Natl. Acad. Sci. USA* **107**, 20732–20737 (2010).
36. Hsu, S.T., Blaser, G. & Jackson, S.E. The folding, stability and conformational dynamics of β-barrel fluorescent proteins. *Chem. Soc. Rev.* **38**, 2951–2965 (2009).
37. Plaxco, K.W., Simons, K.T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994 (1998).
38. Kerner, M.J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220 (2005).
39. Fujiwara, K., Ishihama, Y., Nakahigashi, K., Soga, T. & Taguchi, H. A systematic survey of *in vivo* obligate chaperonon-dependent substrates. *EMBO J.* **29**, 1552–1564 (2010).
40. Maisnier-Patin, S. *et al.* Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nat. Genet.* **37**, 1376–1379 (2005).
41. Chakraborty, K. *et al.* Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell* **142**, 112–122 (2010).
42. Cheung, M.S. & Thirumalai, D. Nanopore-protein interactions dramatically alter stability and yield of the native state in restricted spaces. *J. Mol. Biol.* **357**, 632–643 (2006).
43. Baumketner, A., Jewett, A. & Shea, J.E. Effects of confinement in chaperonin assisted protein folding: Rate enhancement by decreasing the roughness of the folding energy landscape. *J. Mol. Biol.* **332**, 701–713 (2003).
44. Virnau, P., Kantor, Y. & Kardar, M. Knots in globule and coil phases of a model polyethylene. *J. Am. Chem. Soc.* **127**, 15102–15106 (2005).
45. Micheletti, C., Marenduzzo, D., Orlandini, E. & Sumners, D.W. Knotting of random ring polymers in confined spaces. *J. Chem. Phys.* **124**, 64903 (2006); erratum **7**, 219903 (2006).
46. Johnson, K.A., Simpson, Z.B. & Blom, T. Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem.* **387**, 20–29 (2009).

Acknowledgments

We thank E. O'Brien, D. Hsu, F. Andersson, G. Blaser and E. Werrell for helpful discussions. This research was supported by a fellowship and a grant from St. John's College, University of Cambridge, UK (to A.L.M.).

Author contributions

A.L.M. and S.E.J. designed research; A.L.M. carried out research; A.L.M. and S.E.J. analyzed data, and A.L.M. and S.E.J. wrote the paper.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available online at http://www.nature.com/naturechemicalbiology/. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to S.E.J.