REVIEW ARTICLE

How does a knotted protein fold?

Anna L. Mallam

St John's College and University Chemical Laboratory, Cambridge, UK

Keywords
acetohydroxy-acid isomeroreductase; dimeric proteins; knotted proteins; methyltransferases; multistate kinetics; protein folding; topological complexity; topological knots; trefoil knot; ubiquitin hydrolase

Correspondence
A. L. Mallam, St John's College, Cambridge CB2 1TP, UK
Fax: +44 1223 336362
Tel: +44 1223 767042
E-mail: alm38@cam.ac.uk

(Received 30 September 2008, revised 5 November 2008, accepted 14 November 2008)

doi:10.1111/j.1742-4658.2008.06801.x

The issue of how a newly synthesized polypeptide chain folds to form a protein with a unique three-dimensional structure, otherwise known as the 'protein-folding problem', remains a fundamental question in the life sciences. Over the last few decades, much information has been gathered about the mechanisms by which proteins fold. However, despite the vast topological diversity observed in biological structures, it was thought improbable, if not impossible, that a polypeptide chain could 'knot' itself to form a functional protein. Nevertheless, such knotted structures have since been identified, raising questions about how such complex topologies can arise during folding. Their formation does not fit any current folding models or mechanisms, and therefore represents an important piece of the protein-folding puzzle. This article reviews the progress made towards discovering how nature codes for, and contends with, knots during protein folding, and examines the insights gained from both experimental and computational studies. Mechanisms to account for the formation of knotted structures that were previously thought unfeasible, and their implications for protein folding, are also discussed.

The protein-folding problem continues to be a major challenge for structural, molecular and computational biologists. The past two decades have seen the folding pathways of many proteins characterized in detail using experimental and computational approaches. Current theories suggest that proteins can collapse, rearrange, form intermediates and even swap parts of their structure in order to reach their native conformation [1,2]. Yet, it was once thought impossible that a polypeptide chain could fold to form a knot in a protein. It was somewhat surprising, therefore, when a group of proteins possessing this entirely unexpected structural property was identified [3–7]. Such knotted structures were completely unpredicted as, because of the apparent complexities involved, it was thought unfeasible for a protein to fold efficiently in this way. To determine how these proteins knot represents a fundamental and exciting new challenge in the protein-folding field. This review highlights some of the most complex knotted structures identified to date and summarizes the recent developments made towards understanding the mechanisms involved in their formation.

Why are protein knots so unexpected?

In its simplest form, the protein-folding problem can be broken down into two parts: first, how a given amino acid sequence specifies the final functional structure of a protein and, second, how a protein reaches this native state from an initially unfolded (or denatured) chain. Answers to these questions will have practical consequences in medicine, drug development and bio- and nanotechnology [8–10]. A wealth of data on protein-folding mechanisms has been acquired since the first reported high-resolution, three-dimensional protein structures prompted research into the field nearly five decades ago [11]. Presently, the majority of protein-folding studies have focused on easily manipu-

Abbreviation
MTase, methyltransferase.

lated, single-domain monomeric proteins, as they represent simple folding systems [12]. These have led to many models being proposed and tested for the different mechanisms by which small proteins fold [1,13]. Currently, the combination of experimental data and all-atom molecular dynamics simulations means that it is possible to monitor the folding of a small protein at atomic resolution [1]. It is hoped that information gained from such studies will be applicable to larger proteins with more complex topologies.

The final structure of a protein is of particular importance compared to other biological polymers since it is this specific three-dimensional shape that allows it to perform its function. As the number of solved protein structures continues to grow, an increasing variety of unique protein topologies have been observed [14,15]. The likelihood of a protein developing a knotted structure was first reflected on over 30 years ago [16], but it was thought improbable that folding could occur efficiently in this way [17]. Assuming that a polypeptide chain cannot pass through itself, a knot in a protein would have to nucleate at one terminus and a threading event would be required at some stage during folding. Current protein-folding theories do not anticipate such an event and therefore imply that proteins should generally be knot free. For example, evolved proteins tend to fold co-operatively in an all-or-none fashion; in the simplest case, molecules fold with a two-state mechanism and exist only in native or denatured forms [18]. As a result, folding is assumed to occur spontaneously and in a single step under conditions in which the native state of the protein is favoured. It is difficult to imagine how a precise knot could form during folding in this manner; when considering human-scale examples, specific knots are unable to self-assemble spontaneously and any threading must be performed with intent. Further to this, many recent protein-folding models involve the concept of folding energy landscapes (Fig. 1) [19,20]. It is thought that, for folding to take place efficiently and on a biological time scale, a protein must have a funnel-shaped energy landscape under folding conditions. The width of the funnel at a particular energy represents the chain entropy, resulting in a broad top that indicates the large number of conformations available to the denatured state. Natural proteins have evolved to have relatively smooth funnels so that their low-energy native configuration can be approached efficiently from a wide ensemble of denatured states (Fig. 1). Folding is assumed to occur with an increasing degree of 'nativeness' as the protein progresses down the funnel; the native topology of a protein determines its folding mechanism [20]. This notion would effectively preclude knotting of the polypeptide chain if non-native interactions are required to initiate a threading event. Furthermore, the necessity for knot formation during folding would significantly reduce the number of denatured conformations that could successfully reach the native state and, consequently, restrict the folding landscape. Interestingly, a protein with knotted topology was thought to be so unlikely that protein structure prediction studies sometimes make use of algorithms that rapidly detect and discard any protein models containing knotted conformations, as they are deemed 'impossible structures' [21,22]. It was quite unexpected when, contrary to all existing protein-folding models, a group of proteins possessing a knot in their structure was identified [3,4].

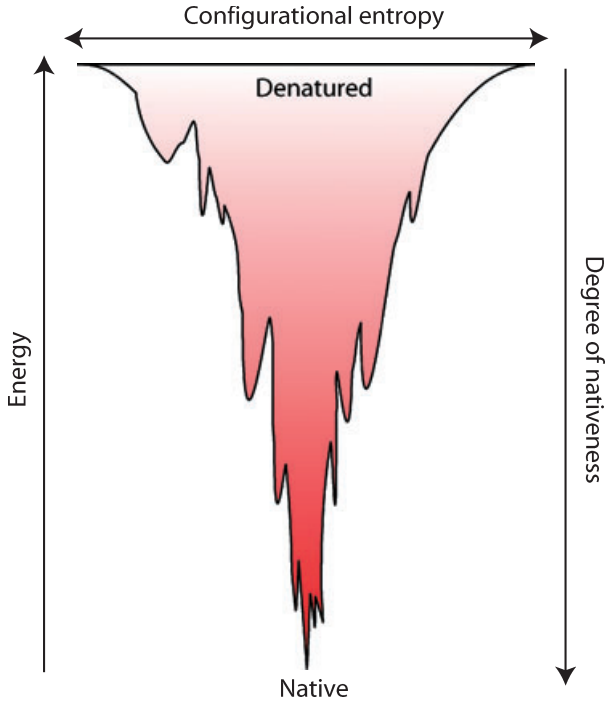

Fig. 1. Cross-section of a protein-folding energy landscape that describes folding from the denatured to the native state of a protein. Such folding landscapes are thought to be robust and funnel like [20]. In simple terms, the system can be described by a configurational entropy term on the *x*-axis, whereas the *y*-axis represents the energy of the conformation and also the fraction of native contacts, or the 'degree of nativeness'.

Protein knots – a surprising case of topological complexity

Knots and other entanglements occur frequently in biological polymers. Knotted DNA molecules were observed as early as 1976 [23], and have since been studied extensively [24–28]. Long strands of DNA can

form loose random knots of varying complexity. Similarly, RNA can adopt knotted conformations [29]. In the case of proteins, structures with a range of intricate assemblies have been reported [14,15]. Interlocked topologies can occur when two protein chains interconnect and subsequently become inseparable. Examples include natural and engineered catenanes that consist of two interlocking rings [30–33] and pseudorotaxanes that comprise a chain threaded through a ring [34]. Knots in proteins are fairly common if the entire covalent network is considered; disulfide cross-links or metal-atom bridges often create 'covalent knots' that can form either during or after folding [35,36]. A cysteine knot occurs when two disulfide bonds and their connecting backbone segments form a ring that is threaded by a third disulfide bond. Examples of this include the cyclotide family of plant-derived miniproteins that are approximately 30 amino acids in size. These contain a cyclized backbone, presumed to arise from a post-translational modification, and a knotted arrangement of three disulfide bonds [15,35,37]. 'Protoknots' have been observed in small peptides when a linear segment loops back and threads through a cyclic component formed by a backbone side-chain linkage [38,39]. Such structures do not present an obvious folding problem as a covalent knot can be introduced after the backbone folds; a specific threading event is not necessarily required. Finally, protein 'slip-knots' can exist if the protein chain forms a knot, but then folds back to effectively untie itself and render the structure unknotted when considered in its entirety [40].

The path of the backbone polypeptide chain exclusively defines protein 'topological knots'. The first of these to be identified nearly 15 years ago were only 'shallow' knots, with one end of the chain extending through a wide loop by just a few residues. Examples include carbonic anhydrase B from *Neisseria gonorrhoeae* [17] and *Escherichia coli S*-adenosylmethionine synthetase [41,42]. It is easy to see how such knots might form from a wandering chain during folding, and they only exist because a few residues at a terminus pass on one side of a neighbouring strand rather than another. Often these structures become unknotted if viewed from a different angle [3,6]. This brings about the issue of what defines a knot in a protein, and how they can be identified [6]. Detecting protein knots is often not straightforward, and sometimes impossible simply by examination of the structure by eye [3]. From a mathematical viewpoint, formal knot theory defines knots as closed paths; no unspliced ends are allowed by which the knot can untie [43]. In this strict sense, an amino acid chain can never form a true knot. However, the ends of the protein chain can be theoretically joined by a long loop. This can often be done unambiguously, as protein termini, because of their charged nature, tend to lie on the surface of the structure. The polypeptide backbone then becomes a closed path and the knot state of the resulting structure can be determined by its Alexander polynomial [43–46]. Problems still occur if the termini of a protein do not lie on the surface, and an algorithm developed by William Taylor offers an alternative approach to the detection of knots [3]. The algorithm 'shrinks' the protein in on itself, whilst the termini are left fixed by repeatedly replacing the coordinates of the α-carbon of each residue with the average of itself and its two neighbours. If this is continued indefinitely, unknotted strings are reduced to a straight line connecting the termini, whilst those containing knots become blocked. Using this algorithm, Taylor detected the first deeply embedded knot in a protein approximately 8 years ago – an intricate figure-of-eight knot in the plant protein acetohydroxy acid isomeroreductase – that had not been identified visually beforehand because of its structural complexity (Fig. 2A). The method allows the location of the knotted core to be pinpointed and the 'depth' of the knot to be calculated from the smallest number of residues that can be removed from each side before the structure becomes unknotted [3,44]. 'Deep' knots have more than 20 amino acid residues on either side of the knot core [4]. Improvements in detection methods, together with the growing number of solved protein structures, has led to the identification of an increasing number of knotted proteins [44–47]. To date, over 250 knotted structures have been discovered in the Protein Data Base, equivalent to approximately 0.5% of all entries [46]. As well as the figure-of-eight knot in acetohydroxy acid isomeroreductase, notable examples include a knot with five projected crossings in human ubiquitin hydrolase (Fig. 2B) and a significant number of α/β proteins containing deep trefoil knots (Fig. 2C). The advantages of such knotted topologies remain unknown, although it has been suggested that they may confer stability and/or functional benefits [7,14,46].

A deep topological trefoil knot was first seen in the catalytic domain of the hypothetical RNA 2′-*O*-ribose methyltransferase (MTase) from *Thermus thermophilus* (RrmA), an α/β protein and a member of the SpoU family [48]. Since then, over 15 α/β proteins containing similarly structured trefoil knots have been reported [5,44,49–56]. These knotted proteins share some common features. All are likely to function as MTases, a type of enzyme involved in the transfer of the methyl group of *S*-adenosylmethionine to DNA, RNA,

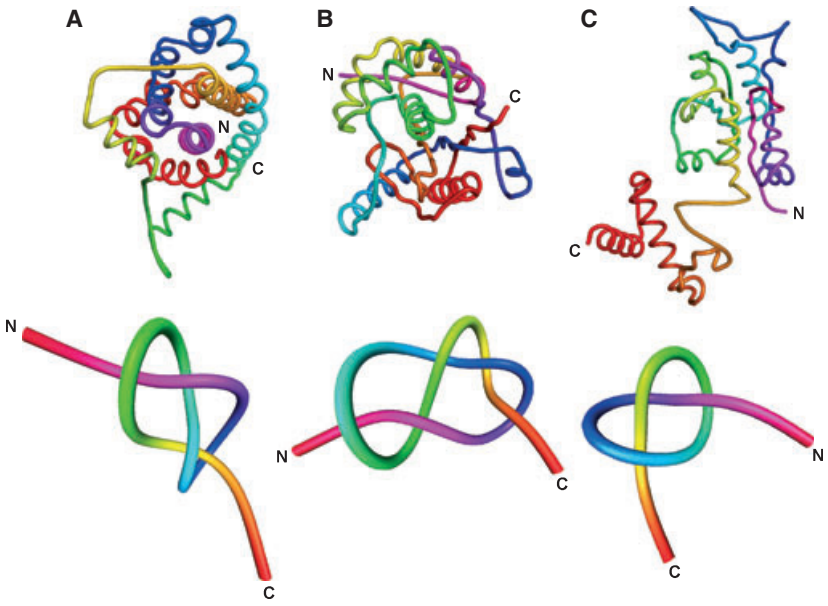

Fig. 2. Example structures of complex knotted proteins. (A) Plant protein acetohydroxy acid isomeroreductase [Protein Data Bank (PDB) code 1YVE] boasts a very complicated figure-of-eight knot and the most deeply embedded natural knot observed to date, with over 200 residues on one side and 70 on the other [4]. (B) Human ubiquitin hydrolase (PDB code 1XD3) contains the most complex protein knot discovered, with five projected crossings. (C) *Haemophilus influenzae* tRNA(m^1G37) methyltransferase (TrmD) (PDB code 1UAJ) has one of the deepest natural trefoil knots identified, with 92 residues on its shortest side. Crystal structures are coloured pink to red from amino to carboxy terminus, respectively. Reduced representations of the various knots generated using KNOTPLOT (http://www.knotplot.com) are shown below each structure. Protein structures were produced using PYMOL (http://www.pymol.org).

proteins and other small molecules [57]. The knot region comprises part of the *S*-adenosylmethionine-binding site [5,48–53,55]. In addition, all form dimers in solution, with the knot structure also involved in the dimer interface. In recognition of the above similarities between MTases with knotted topologies, a new superfamily of proteins was defined, known as the α/β-knot superfamily of MTases [49,58]. The characteristics of members of this family include dimer formation and the presence of a deep trefoil knot that provides the *S*-adenosylmethionine cofactor-binding site. Trefoil knots have been identified in proteins other than α/β-knot MTases, the most recently reported example being a zinc finger protein containing a new trefoil-knot fold [59].

Since these structures are unaccounted for by present-day folding models, proteins that contain a knot represent a unique protein-folding conundrum. Unlike the random unstructured knots observed in DNA molecules that can be compared with accidental tangles, proteins adopt specific topologies and have defined folding mechanisms. It is not obvious how, during folding, a substantial length of polypeptide chain manages to spontaneously and reproducibly thread itself through a loop. Several important advances have been made in the last few years to address the question of how a knotted protein folds.

How does a knotted protein fold? Experimental and computational insights

The mechanisms involved in protein knotting have been probed using a combination of *in vitro* and *in silico* techniques. Experimental studies have primarily examined the folding of two of the smallest homodimeric α/β MTases identified, YibK from *Haemophilus influenzae* and YbeA from *Escherichia coli* (Fig. 3). Both YibK and YbeA can be unfolded reversibly *in vitro* using the chemical denaturant urea, which demonstrates that their complicated knotted structure has not hindered their folding efficiency [60,61]. Their folding pathways have been characterized using kinetic single-jump and double-jump mixing experiments. The folding of YibK is complex because of its dimeric nature and the existence of heterogeneous species in the unfolded state that cause multiple folding pathways [62]. The kinetic mechanism most consistent with the experimental data involves two different intermediates from apparent parallel folding channels that fold via a

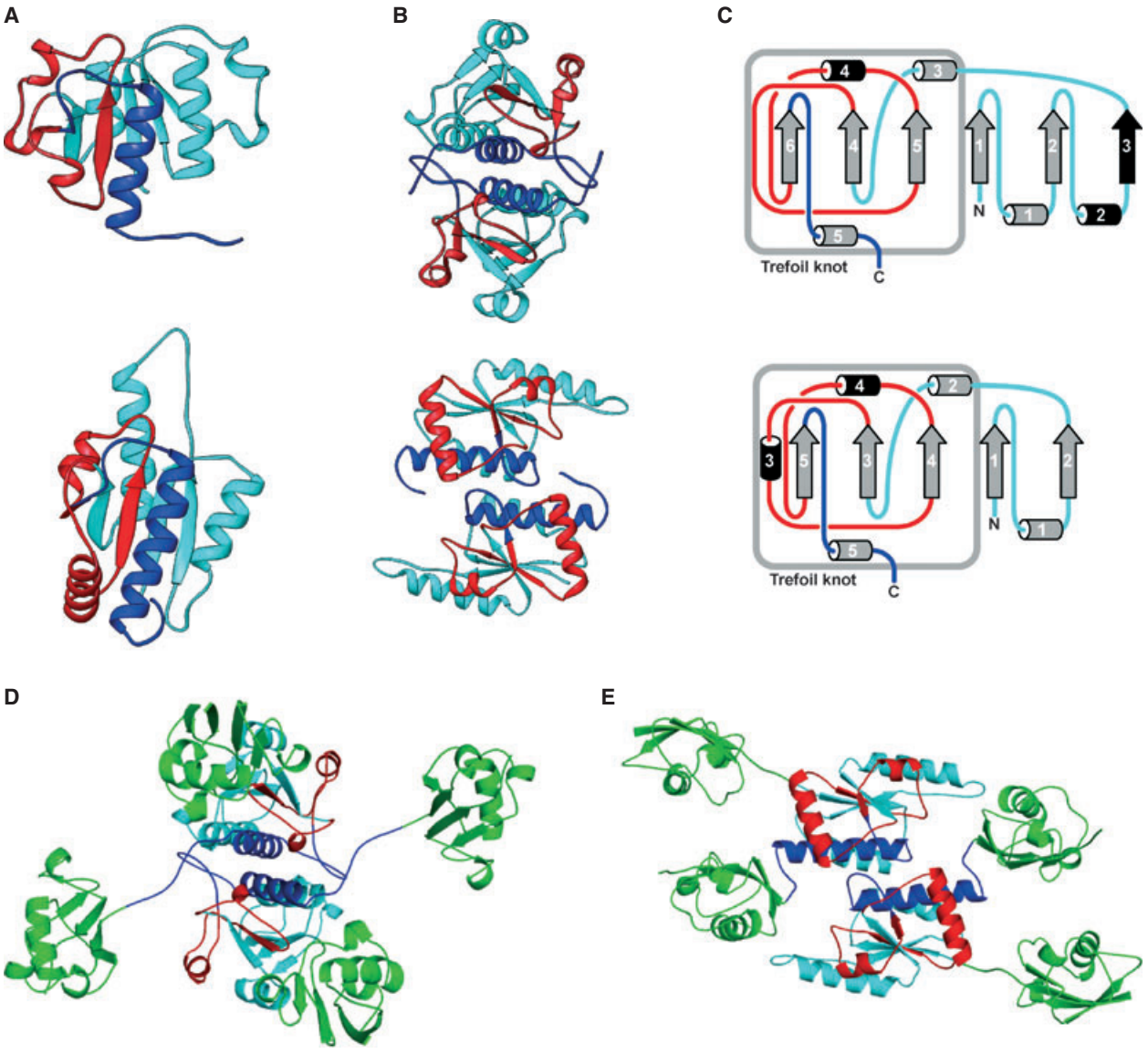

Fig. 3. The α/β-knot MTases YibK and YbeA. (A–C) The X-ray crystal structures of YibK from *Haemophilus influenzae* (top, PDB code 1MXI) and YbeA from *Escherichia coli* (bottom, PDB code 1NS5). Both proteins contain a topological trefoil knot formed by the polypeptide backbone; a substantial length of polypeptide chain (approximately 40 residues) has threaded through a loop during folding. (A) Ribbon diagram of a monomer subunit, showing the deep trefoil knot at the C-terminus. Structures are coloured to show the knotting loop highlighted in red and the knotted chain in dark blue. (B) Dimeric structures coloured as in (A). YibK is a parallel homodimer, whereas YbeA dimerizes in an antiparallel fashion. (C) Topological diagrams indicating the knot region and structural elements common to members of the α/β-knot superfamily, which are shown in grey. (D, E) Structures of the knotted fusion proteins ThiS-YibK-ThiS and ThiS-YbeA-ThiS, respectively, obtained from small-angle X-ray scattering experiments. These artificial constructs contain the deepest knots known, with over 125 residues on either side of the knot core. Knotted domains are coloured as in (A), and ThiS domains are highlighted in green. Ribbon diagrams were generated using RIBBONS [78] and PYMOL (http://www.pymol.org).

third sequential monomeric intermediate to form native dimer in a slow rate-limiting dimerization reaction (Fig. 4A). Although YbeA appears to fold via a simpler pathway with only one observable monomeric intermediate, similarities between the folding of YibK and YbeA imply that the mechanisms involved in knot formation in both proteins may be related (Fig. 4B) [61]. Both show considerable resistance to denaturation and share a common equilibrium unfolding mechanism involving a populated monomeric intermediate. Strong dimerization appears to be a characteristic of α/β-knotted proteins, and there is no evidence for

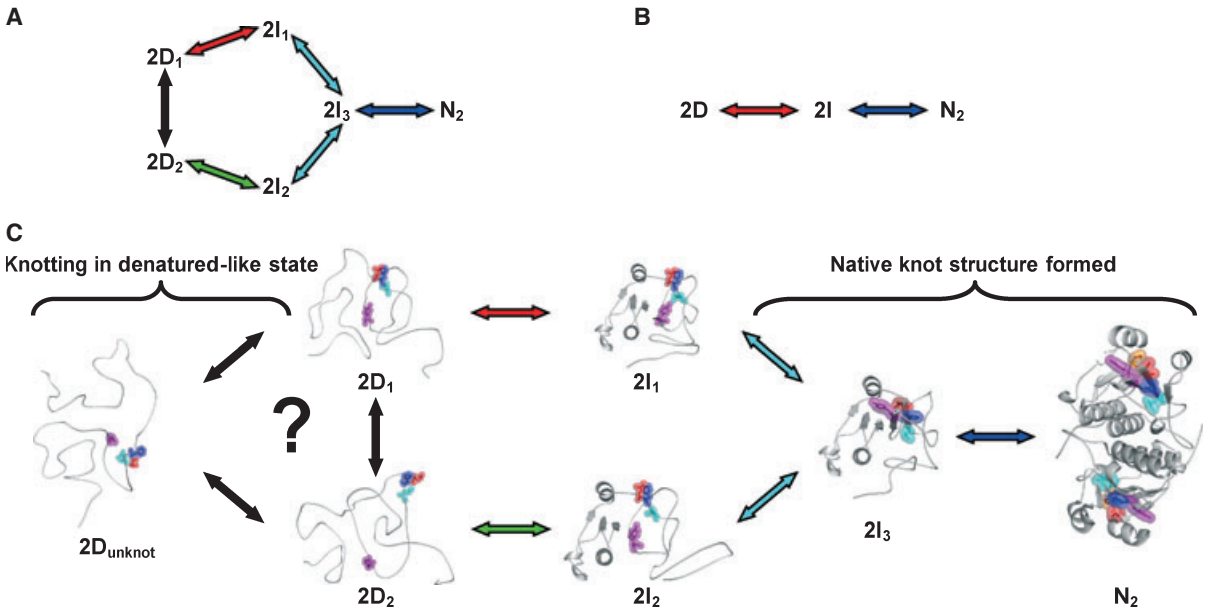

Fig. 4. The proposed folding pathways of the knotted proteins YibK (A) and YbeA (B) most consistent with kinetic experimental data [61,62]. (C) A mechanism for the knotting and folding of YibK based on data from mutational studies [66]. The mutations made in the knot region to probe the folding mechanism are highlighted. Mutant kinetic data were consistent with independent knotting and folding events. It has been suggested that heterogeneous loosely knotted conformations in a denatured-like state (D_1 and D_2) fold via parallel channels to form intermediates I_1 and I_2. The knotted region of the protein remains relatively unstructured until it forms during the folding of I_3 and subsequent dimerization. The exact nature of the heterogeneity in the denatured state, leading to the apparent parallel folding channels, and the structure of the intermediate species remain unknown, and so the representations shown here are for illustrative purposes only. Arrows are coloured to match those in (A). Ribbon diagrams were generated using PYMOL (http://www.pymol.org). Figure adapted from [66].

dissociation of either protein in buffer at near-neutral conditions. Furthermore, both fold via sequential mechanisms that involve the slow formation of a kinetic monomeric intermediate, followed by an even slower dimerization step [61]. Perhaps somewhat surprisingly, investigations have revealed no folding attributes that can be directly linked to knot formation; the apparent intricacies involved do not seem to cause a folding problem. In addition, it has been shown that the dimerization of YibK is essential for maintaining its native structure and function, as cofactor-binding experiments indicate that the knotted region is not fully structured in a monomeric version of the protein [63]. The folding of YibK has also been studied using molecular dynamics simulations [64]. These suggest that formation of the knotted structure is only possible when specific, non-native, attractive interactions are introduced during folding simulations. The results indicate that knot formation occurs before dimerization of the protein and, in agreement with experimental data, that parallel folding pathways exist to the native structure [64].

The events that occur during the folding and knotting of YibK and YbeA have been probed by the construction of a set of novel multidomain proteins that involve the fusion of another small protein, *Archeoglobus fulgidus* ThiS, to either YibK or YbeA at their amino terminus, carboxy terminus or both termini (Fig. 3D,E) [65]. ThiS is a 91-residue monomeric protein that was used as a 'molecular plug' in an attempt to hinder the threading motions of the polypeptide chain or to prevent it from knotting altogether. Interestingly, cofactor-binding and small-angle X-ray scattering experiments indicated that the artificial multidomain constructs were all able to knot and fold [65]. Furthermore, their folding kinetics were comparable with those of the equivalent wild-type protein, despite the fact that a considerably longer segment of chain must be threaded through a loop. These results suggested that a threading event was not the rate-limiting step during the *in vitro* folding of these proteins. The fusion proteins with ThiS attached to both termini contained the most deeply embedded protein knots observed to date, as over 125 residues can theoretically be removed from each side before the structure becomes unknotted (Fig. 3D,E). In order to account for the ability of an additional protein domain to thread during folding, it was concluded that the

formation of α/β-knotted proteins probably propagates from a loosely knotted, denatured-like state [65]. Whether the knot becomes completely untied in the urea-denatured state of a knotted fusion protein remains to be determined; however, the existence of a rapid pre-equilibrium between unknotted and loosely knotted conformations was proposed [65].

Most recently, the folding pathway of YibK has been examined using single-site mutants [66]. The effect of mutations made in the knot region was investigated to provide information on its formation along the folding pathway. Data were consistent with a folding mechanism for YibK in which loose knotting of the polypeptide backbone occurs very early on in folding, but formation of the native structure in the knotted region of the protein happens late and is relatively slow. These results suggest that the threading and folding of the protein chain are therefore successive events, and a preliminary folding model for YibK was proposed (Fig. 4C). The idea that the heterogeneity observed in the denatured state of YibK is a result of the knotting mechanism, and caused by multiple unfolded knotted conformers, was also suggested [66]. This unusual folding model raises questions about the relative importance of early folding events in predicting how a given polypeptide chain will fold.

Other strategies have been used to probe knot formation in protein molecules. For example, experiments on engineered interlocking protein complexes have provided information on the possible mechanisms involved in the knotting of a polypeptide chain. To investigate the kinetics of protein threading directly, a designed protein catenane based on the small p53 tetramerization domain was engineered into a protein pseudo-rotaxane [32,34]. In order to fold, a linear portion of protein chain was required to thread through a cyclic segment; this was found to be a slow but highly efficient process [34]. Studies have also been undertaken to investigate the effect of a knot structure on the mechanical response of a protein. A protein can be mechanically unfolded at the single-molecule level using atomic force microscopy [67]. In these experiments, the protein of interest is attached between two surfaces and a force is applied by increasing the distance between the tethered ends. It is interesting to consider what would happen to a knotted protein as it is 'pulled'. Presumably, the presence of a knot would cause the molecule to become tightened rather than loosened if it is pulled from both ends; after the structure is completely destabilized, the protein would remain as a straight chain with a single knot. The mechanical properties of bovine carbonic anhydrase B, a protein that contains a shallow trefoil knot at its carboxy terminus [68], have been examined [69,70]. On unfolding, the protein extends to a distance much shorter than its theoretical stretching length, which indicates that the knot structure has indeed become taut on mechanical unfolding. The effect of pulling knotted structures has also been examined theoretically, and stretching simulations suggest that, on tightening, knots in proteins will behave differently from those in homopolymers [71]. Atomic force microscopy studies on proteins with deeper knots, to discover more about their folding and unfolding pathways, remain an exciting prospect for future research, and a preliminary description of the first successful unfolding of a protein with a deep figure-of-eight knot has been reported [72].

Although the current understanding of protein folding mechanisms implies that productive knot formation during folding should be a rare, or impossible, event, a significant number of simulation studies suggest that protein chains are likely to frequently become entangled [45,73–76]. The number of knots expected in a protein-like homopolymer has been investigated using polyethylene models [73]. These simulations indicate that both the frequency and complexity of knotted structures ought to increase with chain length; a polymer that is equivalent in length to a few thousand amino acids should almost certainly be knotted. In addition, the presence, size and type of knot were found to depend on the solvent conditions. These ideas suggest that large protein chains have a high chance of becoming knotted of their own accord, that more complex knotted topologies should arise from longer polypeptide chains and that the solvent could have a notable effect on the folding mechanism of a knotted protein. Interestingly, a study comparing the knotting probabilities in proteins with those in random polymers has shown that native protein conformations have statistically fewer knots than random compact loops, which suggests that proteins have evolved to specifically avoid knotted topologies [45]. The development of knotted structures during the collapse of polymer chains has been investigated using simulation techniques [74,75]. These experiments indicated that knotting occurs by the tunnelling of the ends of the polymer chain in and out of the polymer globule. This knotting mechanism may not be applicable to proteins since, because of their charged nature, protein termini favour the solvent-exposed surface; it has been suggested that this could account for the apparent lack of knots in protein structures [75]. Furthermore, it was noted that the model is not applicable to the folded state of a protein where the chain is immobile. The knotting properties of a simple piece of string as it is

agitated inside a cubic box have also been examined [76]. In agreement with polymer simulations, these experiments suggest that long flexible strings are almost certain to become knotted after being rotated for only a few seconds. The results allow a simplified model for knot formation to be proposed based on the stiff string forming a coil in the box; when multiple parallel strands are situated near the termini of the string, knots can form as the end segment weaves under and over adjacent segments [76].

Towards solving the folding puzzle of knotted proteins

The insights gained from investigations into the knotting properties of homopolymers are more likely to be relevant to the flexible state of a denatured protein, or perhaps a partially folded 'molten-globule' intermediate, rather than the final rigid native structure. Such studies may therefore suggest that knotting of a sufficiently long polypeptide chain in the denatured state is a frequent event, consistent with the model of folding for the α/β-knot MTases proposed from experimental data that involves knotting early on in the reaction (Fig. 4A) [65,66]. Experiments to establish the existence of such denatured knotted conformers and the mechanisms involved in their interconversion are an area for future research.

The prospect that productive knotting to form a functional protein originates in the denatured state raises some interesting issues: given that long flexible strings similar to polypeptide chains appear to have a high chance of becoming entangled, it may be that knotted denatured conformers are a folding feature of all sufficiently large proteins [65]; where the native state structure is unknotted, productive folding would only be possible from an untangled chain. Such knotting could preclude successful folding and result in kinetic traps, parallel folding channels and complex protein-folding kinetics – large proteins are often overlooked as candidates for folding experiments because of their complicated or irreversible folding. The relevance of the knot state of a chemically denatured protein to folding events in the cellular environment remains to be addressed; it is possible that one role of molecular chaperones is to prevent or reverse unproductive knotting events in the cell [65].

The mechanisms discussed in this review suggest that nature may have overcome the problem of knotting a protein by separating the processes of threading and folding, so that they occur as successive events. Experiments are consistent with early threading in a loose, denatured-like state, and simulations suggest that knotting is not a problem for polymers similar to the denatured state of a protein. If threading and folding occur sequentially, the inconsistencies between knotted protein structures and current protein-folding models cease to exist. Knotted and unknotted proteins could be considered to fold in a similar fashion, the former differing only with an initial knotting event in a denatured-like state. Should this theory prove to be correct, the challenging task to determine the interactions responsible for knotting in a denatured-like state presents itself. Denatured states are inherently difficult to characterize because of their flexible heterogeneous nature [77]. However, it may be necessary to focus on events that occur early on in folding rather than native structure formation to predict whether a given polypeptide chain will fold to a knotted structure. Further studies to establish the early folding interactions that are required to productively knot a polypeptide chain will likely prove to be essential for protein structure prediction, simulation and design.

Acknowledgements

The author thanks Sophie Jackson for helpful discussions and is grateful to Simon Humphrey and Jane Clarke for critical reading of the manuscript. A. L. M. is a Research Fellow at St John's College, Cambridge, UK.

References

1 Daggett V & Fersht A (2003) The present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol* **4**, 497–502.
2 Rousseau F, Schymkowitz JW, Wilkinson HR & Itzhaki LS (2002) The structure of the transition state for folding of domain-swapped dimeric p13suc1. *Structure* **10**, 649–657.
3 Taylor WR (2000) A deeply knotted protein structure and how it might fold. *Nature* **406**, 916–919.
4 Taylor WR & Lin K (2003) Protein knots: a tangled problem. *Nature* **421**, 25.
5 Michel G, Sauve V, Larocque R, Li Y, Matte A & Cygler M (2002) The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot. *Structure* **10**, 1303–1315.
6 Mansfield ML (1997) Fit to be tied. *Nature Struct Biol* **4**, 166–167.
7 Taylor WR (2007) Protein knots and fold complexity: some new twists. *Comput Biol Chem* **31**, 151–162.
8 Yue P, Li Z & Moult J (2005) Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol* **353**, 459–473.

9 Bullock AN & Fersht AR (2001) Rescuing the function of mutant p53. *Nature Rev Cancer* **1**, 68–76.
10 Dobson CM (2003) Protein folding and misfolding. *Nature* **426**, 884–890.
11 Fersht AR (2008) From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat Rev Mol Cell Biol* **9**, 650–654.
12 Jackson SE (1998) How do small single-domain proteins fold? *Fold Des* **3**, 81–91.
13 Daggett V & Fersht AR (2003) Is there a unifying mechanism for protein folding? *Trends Biochem Sci* **28**, 18–25.
14 Yeates TO, Norcross TS & King NP (2007) Knotted and topologically complex proteins as models for studying folding and stability. *Curr Opin Chem Biol* **11**, 595–603.
15 Kent S (2004) Novel forms of chemical protein diversity – in nature and in the laboratory. *Curr Opin Biotechnol* **15**, 607–614.
16 Crippen GM (1974) Topology of globular proteins. *J Theor Biol* **45**, 327–338.
17 Mansfield ML (1994) Are there knots in proteins? *Nature Struct Biol* **1**, 213–214.
18 Jackson SE & Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428–10435.
19 Dill KA & Chan HS (1997) From Levinthal to pathways to funnels. *Nature Struct Biol* **4**, 10–19.
20 Onuchic JN & Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* **14**, 70–75.
21 Khatib F, Weirauch MT & Rohl CA (2006) Rapid knot detection and application to protein structure prediction. *Bioinformatics* **22**, e252–e259.
22 Tramontano A, Leplae R & Morea V (2001) Analysis and assessment of comparative modeling predictions in CASP4. *Proteins* **45** (Suppl. 5), 22–38.
23 Liu LF, Depew RE & Wang JC (1976) Knotted single-stranded DNA rings: a novel topological isomer of circular single-stranded DNA formed by treatment with *Escherichia coli* omega protein. *J Mol Biol* **106**, 439–452.
24 Bao XR, Lee HJ & Quake SR (2003) Behavior of complex knots in single DNA molecules. *Phys Rev Lett* **91**, 265506.
25 Arsuaga J, Vazquez M, Trigueros S, Sumners D & Roca J (2002) Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids. *Proc Natl Acad Sci USA* **99**, 5373–5377.
26 Dean FB, Stasiak A, Koller T & Cozzarelli NR (1985) Duplex DNA knots produced by *Escherichia coli* topoisomerase I. Structure and requirements for formation. *J Biol Chem* **260**, 4975–4983.
27 Shaw SY & Wang JC (1993) Knotting of a DNA chain during ring closure. *Science* **260**, 533–536.
28 Sogo JM, Stasiak A, Martinez-Robles ML, Krimer DB, Hernandez P & Schvartzman JB (1999) Formation of knots in partially replicated DNA molecules. *J Mol Biol* **286**, 637–643.
29 Wang H, Di Gate RJ & Seeman NC (1996) An RNA topoisomerase. *Proc Natl Acad Sci USA* **93**, 9477–9482.
30 Cao Z, Roszak AW, Gourlay LJ, Lindsay JG & Isaacs NW (2005) Bovine mitochondrial peroxiredoxin III forms a two-ring catenane. *Structure* **13**, 1661–1664.
31 Boutz DR, Cascio D, Whitelegge J, Perry LJ & Yeates TO (2007) Discovery of a thermophilic protein complex stabilized by topologically interlinked chains. *J Mol Biol* **368**, 1332–1344.
32 Blankenship JW & Dawson PE (2003) Thermodynamics of a designed protein catenane. *J Mol Biol* **327**, 537–548.
33 Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW & Johnson JE (2000) Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133.
34 Blankenship JW & Dawson PE (2007) Threading a peptide through a peptide: protein loops, rotaxanes, and knots. *Protein Sci* **16**, 1249–1256.
35 Cemazar M, Joshi A, Daly NL, Mark AE & Craik DJ (2008) The structure of a two-disulfide intermediate assists in elucidating the oxidative folding pathway of a cyclic cystine knot protein. *Structure* **16**, 842–851.
36 McDonald NQ & Hendrickson WA (1993) A structural superfamily of growth factors containing a cystine knot motif. *Cell* **73**, 421–424.
37 Craik DJ, Cemazar M, Wang CK & Daly NL (2006) The cyclotide family of circular miniproteins: nature's combinatorial peptide template. *Biopolymers* **84**, 250–266.
38 Bayro MJ, Mukhopadhyay J, Swapna GV, Huang JY, Ma LC, Sineva E, Dawson PE, Montelione GT & Ebright RH (2003) Structure of antibacterial peptide microcin J25: a 21-residue lariat protoknot. *J Am Chem Soc* **125**, 12382–12383.
39 Rosengren KJ, Clark RJ, Daly NL, Goransson U, Jones A & Craik DJ (2003) Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. *J Am Chem Soc* **125**, 12464–12474.
40 King NP, Yeates EO & Yeates TO (2007) Identification of rare slipknots in proteins and their implications for stability and folding. *J Mol Biol* **373**, 153–166.
41 Takusagawa F & Kamitori S (1996) A real knot in protein. *J Am Chem Soc* **118**, 8945–8946.
42 Takusagawa F, Kamitori S & Markham GD (1996) Structure and function of S-adenosylmethionine synthetase: crystal structures of S-adenosylmethionine synthetase with ADP, BrADP, and PPi at 28 angstroms resolution. *Biochemistry* **35**, 2586–2596.

43 Adams CC (1994) *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*. W. H. Freeman, New York, NY.
44 Kolesov G, Virnau P, Kardar M & Mirny LA (2007) Protein knot server: detection of knots in protein structures. *Nucleic Acids Res* **35**, W425–W428.
45 Lua RC & Grosberg AY (2006) Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS Comput Biol* **2**, e45.
46 Virnau P, Mirny LA & Kardar M (2006) Intricate knots in proteins: function and evolution. *PLoS Comput Biol* **2**, 1074–1079.
47 Lai YL, Yen SC, Yu SH & Hwang JK (2007) pKNOT: the protein KNOT web server. *Nucleic Acids Res* **35**, W420–W424.
48 Nureki O, Shirouzu M, Hashimoto K, Ishitani R, Terada T, Tamakoshi M, Oshima T, Chijimatsu M, Takio K, Vassylyev DG *et al.* (2002) An enzyme with a deep trefoil knot for the active-site architecture. *Acta Crystallogr Sect D* **58**, 1129–1137.
49 Ahn HJ, Kim HW, Yoon HJ, Lee BI, Suh SW & Yang JK (2003) Crystal structure of tRNA(m1G37)methyltransferase: insights into tRNA recognition. *EMBO J* **22**, 2593–2603.
50 Elkins PA, Watts JM, Zalacain M, van Thiel A, Vitazka PR, Redlak M, Andraos-Selim C, Rastinejad F & Holmes WM (2003) Insights into catalysis by a knotted TrmD tRNA methyltransferase. *J Mol Biol* **333**, 931–949.
51 Forouhar F, Shen J, Xiao R, Acton TB, Montelione GT & Tong L (2003) Functional assignment based on structural analysis: crystal structure of the yggJ protein (HI0303) of *Haemophilus influenzae* reveals an RNA methyltransferase with a deep trefoil knot. *Proteins: Struct Funct Genet* **53**, 329–332.
52 Lim K, Zhang H, Tempczyk A, Krajewski W, Bonander N, Toedt J, Howard A, Eisenstein E & Herzberg O (2003) Structure of the YibK methyltransferase from *Haemophilus influenzae* (HI0766): a cofactor bound at a site formed by a knot. *Proteins: Struct Funct Genet* **51**, 56–67.
53 Zarembinski TI, Kim Y, Peterson K, Christendat D, Dharamsi A, Arrowsmith CH, Edwards AM & Joachimiak A (2003) Deep trefoil knot implicated in RNA binding found in an archaebacterial protein. *Proteins: Struct Funct Genet* **50**, 177–183.
54 Mosbacher TG, Bechthold A & Schulz GE (2005) Structure and function of the antibiotic resistance-mediating methyltransferase AviRb from *Streptomyces viridochromogenes*. *J Mol Biol* **345**, 535–545.
55 Nureki O, Watanabe K, Fukai S, Ishii R, Endo Y, Hori H & Yokoyama S (2004) Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme. *Structure* **12**, 593–602.
56 Pleshe E, Truesdell J & Batey RT (2005) Structure of a class II TrmH tRNA-modifying enzyme from *Aquifex aeolicus*. *Acta Crystallogr Sect F* **61**, 722–728.
57 Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K & McCann PP (1996) S-Adenosylmethionine and methylation. *FASEB J* **10**, 471–480.
58 Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res* **32**, D138–D141.
59 van Roon AM, Loening NM, Obayashi E, Yang JC, Newman AJ, Hernandez H, Nagai K & Neuhaus D (2008) Solution structure of the U2 snRNP protein Rds3p reveals a knotted zinc-finger motif. *Proc Natl Acad Sci USA* **105**, 9621–9626.
60 Mallam AL & Jackson SE (2005) Folding studies on a knotted protein. *J Mol Biol* **346**, 1409–1421.
61 Mallam AL & Jackson SE (2007) A comparison of the folding of two knotted proteins: YbeA and YibK. *J Mol Biol* **366**, 650–665.
62 Mallam AL & Jackson SE (2006) Probing nature's knots: the folding pathway of a knotted homodimeric protein. *J Mol Biol* **359**, 1420–1436.
63 Mallam AL & Jackson SE (2007) The dimerization of an alpha/beta-knotted protein is essential for structure and function. *Structure* **15**, 111–122.
64 Wallin S, Zeldovich KB & Shakhnovich EI (2007) The folding mechanics of a knotted protein. *J Mol Biol* **368**, 884–893.
65 Mallam AL, Onuoha SC, Grossmann JG & Jackson SE (2008) Knotted fusion proteins reveal unexpected possibilities in protein folding. *Mol Cell* **30**, 642–648.
66 Mallam AL, Morris ER & Jackson SE (2008) Exploring knotting mechanisms in protein folding. *Proc Natl Acad Sci USA* **105**, 18740–18745.
67 Borgia A, Williams PM & Clarke J (2008) Single-molecule studies of protein folding. *Annu Rev Biochem* **77**, 101–125.
68 Saito R, Sato T, Ikai A & Tanaka N (2004) Structure of bovine carbonic anhydrase II at 1.95 Å resolution. *Acta Crystallogr Sect D Biol Crystallogr* **60**, 792–795.
69 Alam MT, Yamada T, Carlsson U & Ikai A (2002) The importance of being knotted: effects of the C-terminal knot structure on enzymatic and mechanical properties of bovine carbonic anhydrase II. *FEBS Lett* **519**, 35–40.
70 Wang T, Arakawa H & Ikai A (2001) Force measurement and inhibitor binding assay of monomer and engineered dimer of bovine carbonic anhydrase B. *Biochem Biophys Res Commun* **285**, 9–14.
71 Sulkowska JI, Sulkowski P, Szymczak P & Cieplak M (2008) Tightening of knots in proteins. *Phys Rev Lett* **100**, 058106.
72 Ball P (2008) Material witness: get knotted. *Nat Mater* **7**, 772.

73 Virnau P, Kantor Y & Kardar M (2005) Knots in globule and coil phases of a model polyethylene. *J Am Chem Soc* **127**, 15102–15106.
74 Mansfield ML (2007) Efficient knot group identification as a tool for studying entanglements of polymers. *J Chem Phys* **127**, 244901.
75 Mansfield ML (2007) Development of knotting during the collapse transition of polymers. *J Chem Phys* **127**, 244902.
76 Raymer DM & Smith DE (2007) Spontaneous knotting of an agitated string. *Proc Natl Acad Sci USA* **104**, 16432–16437.
77 Shortle D (1996) The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J* **10**, 27–34.
78 Carson M (1997) Ribbons. *Methods Enzymol* **277**, 493–505.