Structures and folding pathways of topologically knotted proteins

2011 J. Phys.: Condens. Matter 23 033101

(http://iopscience.iop.org/0953-8984/23/3/033101)

J. Phys.: Condens. Matter **23** (2011) 033101 (17pp) doi:10.1088/0953-8984/23/3/033101

TOPICAL REVIEW

Structures and folding pathways of topologically knotted proteins

Peter Virnau[1], **Anna Mallam**[2,3] **and Sophie Jackson**[2]

[1] Institut für Physik, Johannes Gutenberg-Universität Mainz, Staudinger Weg 7, 55128 Mainz, Germany
[2] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
[3] Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA

E-mail: virnau@uni-mainz.de, alm@mail.utexas.edu and sej13@cam.ac.uk

Received 22 September 2010, in final form 27 October 2010
Published 10 December 2010
Online at stacks.iop.org/JPhysCM/23/033101

Abstract
In the last decade, a new class of proteins has emerged that contain a topological knot in their backbone. Although these structures are rare, they nevertheless challenge our understanding of protein folding. In this review, we provide a short overview of topologically knotted proteins with an emphasis on newly discovered structures. We discuss the current knowledge in the field, including recent developments in both experimental and computational studies that have shed light on how these intricate structures fold.

(Some figures in this article are in colour only in the electronic version)

Contents

1. Introduction 1
2. Computer-aided knot detection and visualization 2
3. Knots in the protein data bank 3
4. Experimental studies to investigate the mechanisms of knot formation in proteins 5
   4.1. Experiments on trefoil-knotted proteins 8
   4.2. Folding studies on 5_2 knotted proteins: ubiquitin C-terminal hydrolases (UCHs) 10
   4.3. Experiments on slipknotted proteins 11
   4.4. Folding and stability studies on other topologically complex systems 11
5. Folding simulations 13
6. Conclusions and outlook 14
Acknowledgments 14
References 14

1. Introduction

A fine grasp of the underlying mechanisms by which proteins fold into their native states has been a long-standing goal of biophysics and chemistry. While some progress has been achieved in understanding the folding of comparatively simple structures, intriguingly complex folds have emerged in recent years that pose a considerable challenge. In this review, we discuss a particularly fascinating class of proteins which contain a knot in their backbone [1–9]. These can be thought of as proteins with structures that do not disentangle completely after being pulled from both ends.

For a long time it was believed that it was impossible for a protein to contain a real knot. The only known structure that contained something akin to a knot (carbonic anhydrase) disentangled completely when two or three amino acids were removed from the C-terminus. In fact, structural biologists have long used the notable absence of knots in protein structures to produce (unknotted) models from electron density maps before computers automated this task (see e.g. [10]). Nowadays, protein structure prediction programs scan (and typically dismiss) knotted conformations from candidate structures [11]. This is because the frequency with which knots are predicted by these programs is typically too high.

For many years, knots in protein structures went undetected. This is because finding them by visual inspection

Table 1. Examples of proteins with knotted backbones (data was taken 01/2010). For each fold an example pdb code is given. Chain start–stop refers to the first and the last amino acid that are resolved in the high resolution structure. The knotted core is the minimum configuration which stays knotted after a series of deletions from either terminus as given by our web server [32]. This 'knot size' is determined by an automated procedure, and results should only be regarded as a guideline. Knots marked by + and − are right and left-handed, respectively (see text). Examples of slipknots are not listed in this table.

Protein family	pdb	Chain start–stop	Knot type	Knotted core
RNA methyltransferase, (α/β knot)	1ns5	1–153	3_1+	69–121
Carbonic anhydrase	1lug	2–260	3_1+	31–257
SAM synthetase	1fug	1–383	3_1+	33–260
Transcarbamylase fold	1js1	1–324	3_1+	169–267
Zinc-finger fold	2k0a	−1–107	3_1-	18–78
Ribbon-helix-helix superfamily	2efv	6–87	3_1-	19–66
Artificially knotted protein	3mlg	13–181	3_1-	44–147
CII Ketol acid reductoisomerase	1yve	83–595	4_1	321–533
Chromophore binding domain	2o9c	4–322	4_1	37–296
Ubiquitin C-terminal hydrolase	2etl	1–223	5_2-	10–216
α-haloacid dehalogenase I	3bjx	−14–296	6_1+	71–268

is particularly hard. In recent years, however, the exponential rise of protein structures in the protein data bank (PDB) [12] allowed for systematic structural analyses and consequently new discoveries of knotted topologies. The first of such studies was performed by Mansfield in 1994 [1]. Even though the study was unsuccessful in finding novel protein knots (at that point only about 400 protein structures were available) it nevertheless paved the way for future investigations. In 2000, the first novel protein knots were detected by a search of the PDB by Taylor [3], and since then several other groups have performed similar investigations [4–7, 11, 13]. Notably, Virnau *et al* [4] included homologous structures in their study which allowed the first systematic investigation of the evolutionary origin of protein knots.

This review focuses on protein structures that contain a knot formed by the path of the polypeptide chain. However, other topologically challenging classes of protein, for example slipknots [6, 14, 15] (backbone knots which effectively untie themselves when the protein is considered in its entirety) [16–22] are also discussed in section 4.3. After providing a comprehensive overview of knotted backbone structures and explaining some of the pitfalls of computer-aided knot detection, we will address recent experimental work on the folding of knotted proteins [23–30]. We conclude with a brief review of recent folding simulations. Taken together, these studies are providing some insight into how these intricate structures are able to fold.

2. Computer-aided knot detection and visualization

From a mathematical point of view, although perhaps contrary to daily experience, knots are properties of closed curves [31]. Strictly speaking, knots cannot be defined in open chains because every entanglement can be unknotted by an appropriate closure. However, almost all biological or synthetic macromolecules of interest are open chains. The concept of 'knottedness' as a measure of entanglement can be expanded to account for these 'physical knots'. For a layman, a string is knotted if it does not disentangle after being pulled from both ends, and this idea is typically applied to detect 'knots' in proteins. This approach is equivalent to connecting the termini of a protein by a loop and analysing the resulting closed curve.

Fortunately, N- and C-termini are often located on the surface of proteins such that they can be connected unambiguously. In rare cases (for example when the termini are buried in the centre of the structure) closures may create knots in otherwise 'unknotted' proteins. Therefore, it is prudent to visually double-check each protein that is identified as knotted using this method to reject false positives. This last step cannot be performed computationally, and consequently a completely automated knot detection approach is unfeasible.

A list of knotted proteins is shown in table 1. This was generated by applying the following closure procedure to the structures from the PDB [4, 32]: first, a protein was reduced to its C_α-backbone and two lines were drawn outward starting at the termini in the direction of the connection line between the centre of mass of the backbone and the respective ends. These two lines are joined by a loop. As an alternative, one can apply a statistical method as proposed in [1]. A sphere (considerably larger than the protein) is put around the centre of mass. Then, two points on the sphere are chosen at random and the two termini are connected to these points and the points with each other. The closure is repeated several times to determine the majority knot type. This method works slightly better for proteins, albeit at increased computational cost [4].

Once the chain is closed, computers and the tools of mathematics can be employed to determine which type of knot is present in the protein. Knots are typically classified according to the minimum number of crossings in a projection onto a plane. A closed, unknotted loop (crossing number 0) is usually referred to as the unknot. The simplest non-trivial knot is the trefoil knot (3_1) with three crossings. There is one type of knot with four crossings—the so-called figure-of-eight knot (4_1), two with five crossings (5_1, 5_2) and three with six crossings (6_1, 6_2, 6_3). So far, protein structures that contain 3_1, 4_1, 5_2, and 6_1 knots have been identified (see figure 1). Unfortunately, there is still no algorithm available that can identify every type of knot unambiguously from an entangled polymer (to provide the reader with some impression

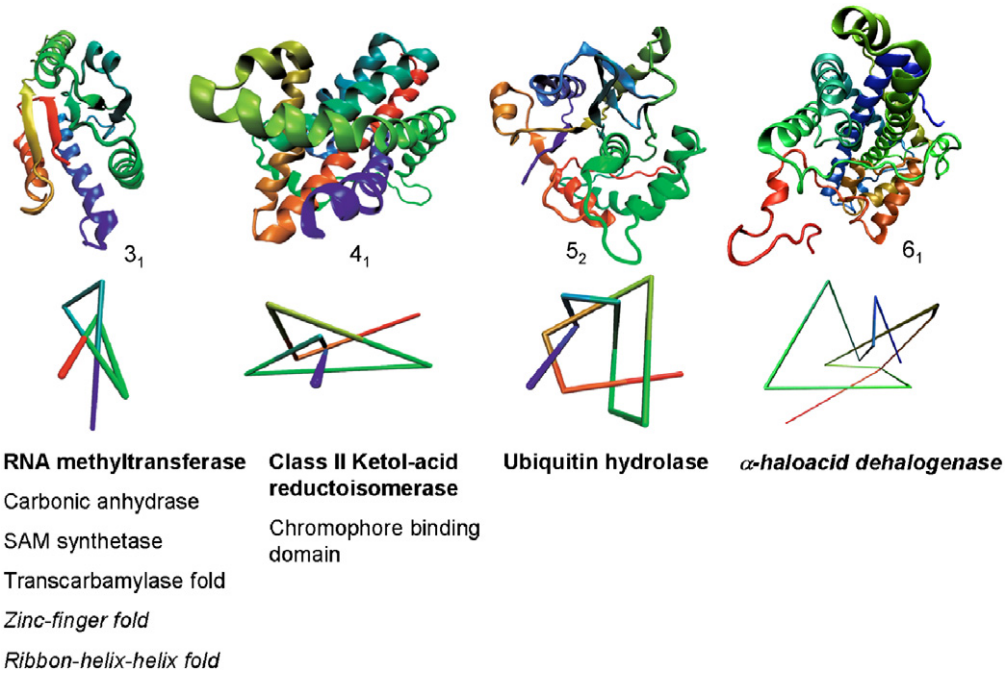

Figure 1. Examples of the four different types of knots (3_1, 4_1, 5_2, 6_1) which have been identified in the backbone of proteins. Colours change continuously from red (light grey) (N-terminus) to blue (dark grey) (C-terminus). The lower panel shows a reduced representation of the knot based on an algorithm described in [34]. Bold names describe the actual proteins depicted in the figure—italic names refer to recently discovered protein knots. Slipknotted proteins are not included. Reproduced from [13]. Copyright 2010 Bölinger *et al*.

of the tremendous complexity of this task, note that there are 1701 936 distinct knots with 16 crossings or less and more than 1 million knots with 16 crossings alone [33]). From a practical point of view, however, simple knots, like the ones which occur in proteins, can be distinguished. The Alexander polynomial was applied to classify the knotted structures listed in table 1. This is one of the first, simplest (and also weakest) knot detection algorithms, but is sufficient for this purpose. For a detailed discussion of the algorithmic implementation the reader is referred to [4, 34].

Due to the difficulties associated with visual identification of knots in protein structures, many of these complex topologies went undiscovered until recently (compare upper and lower row of figure 1). To ease the visualization of knots in proteins, one typically simplifies structures by omitting all the amino acids that are not essential for maintaining the knot. An algorithm to achieve this was first introduced by Koniaris and Muthukumar in 1991 [35] to depict knots in model polymers. A similar scheme was also developed independently in the context of protein knots by Taylor [8]. This reduction scheme is therefore sometimes referred to as the KMT reduction. In the method used to produce the reduced configurations in figure 1, we check if triangles made up of three successive beads are intersected by any part of the remaining chain. If this is not the case, the middle bead is removed. After several iterations of going back and forth between the termini, a highly reduced representation of the original chain remains. As a side benefit, the computation of the Alexander polynomial is considerably faster for reduced configurations. (We need to compute the determinant of a square matrix of size $(n - 1)$ with n being the number of all crossings obtained by projecting the three-dimensional structure onto a plane. Computing the determinant typically takes $O(n^3)$ time with a LU decomposition.) Reduction algorithms of this type have also been used to detect knots directly as an alternative to the calculation of knot polynomials [3, 11]. However, this mechanical approach is less powerful, as sometimes, entangled configurations with no knots remain, and if the termini are not connected before reducing the configuration, knots may also vanish [36].

Finally, we would like to distinguish between 'shallow' knots, which disappear when a few amino acids are cleaved from each terminus and 'deep' knots, which persist. This can be done by defining the size of the knotted core. The extent of the knotted cores of the structures listed in table 1 can be calculated by a series of deletions from each side as suggested by Taylor [3]. First, amino acid monomers are successively removed from the N-terminus until the knot that was originally identified in the structure disappears. The same deletion procedure is then repeated from the C-terminus using the last N-terminal deleted structure that contained the original knot. Note, however, that the size of the knotted core determined using this method should only serve as a guideline, and a visual inspection of the structure generally provides better results.

3. Knots in the protein data bank

The search for knotted proteins is a challenging problem and the detection of knots requires a combination of computational tools to screen promising candidates and extremely careful visual inspection. In many cases, crystallographers have overlooked the presence of a knotted topology in a structure [37–40]. The type of several knots has also been

misclassified [3, 4, 41]. Fortunately, in more recent times web servers and data banks alleviate considerably the task of knot identification. Apart from the protein data bank [12] in which all structures are stored, a classification data bank such as SCOP [42], a tool for protein structure comparison like SSM [43] and, of course, a web server to determine and visualize protein knots [32, 44] enable even non-specialists to check and compare structures and look for promising knotted candidates.

An excellent review of protein structures that contain knots in their backbone can be found in [15]. In this section, we also discuss more recent additions (up to January 2010) and a novel trefoil-knotted fold, which was created artificially by King *et al* [45]. The list of protein folds with knots in their backbone shown in table 1 only contains knots of which the authors are certain. Several candidate structures, like the bluetongue viral core protein [3] (pdb code: 2btv), were deliberately excluded because they are knotted in some reported structures and unknotted in others.

The first protein knot to be identified was a trefoil knot discovered in carbonic anhydrase (as mentioned e.g. in [46])—a well-studied family of proteins which catalyze the reaction of carbon dioxide to hydrogen carbonate and H^+. This trefoil, however, is rather shallow at the C-terminus and a deletion of two to three amino acids typically suffices to disentangle it. Therefore, it is often omitted in the literature even though it is by far the most numerous knotted protein found in the protein data bank. (Roughly two out of three knotted proteins in the current pdb are carbonic anhydrases [7].)

The first trefoil knot approved by all pundits was detected in S-adenosylmethionine synthetase by Takusagawa and co-workers in 1996 [9, 47, 48]. This enzyme catalyzes the reaction from methionine and ATP to S-adenosylmethionine, which is a primary donor of methyl groups. A knotted subunit consists of three unknotted domains which are related to each other by a pseudo-3-fold symmetry.

The most well-studied and numerous family of deeply knotted proteins in the protein data bank is the so-called α/β knot fold (see figure 1, left)—a class of RNA-methyltransferases which are currently subdivided into eight superfamilies in the SCOP classification scheme (v1.75) [42]. These structures all contain a deep trefoil knot, the first of which was discovered in 2002 [49, 50]. The folding pathways of two members, YibK from *Haemophilus influenzae* and YbeA from *Escherichia coli*, have been studied extensively and will be discussed in section 4 of the review, which will also provide additional details on the structure.

A trefoil knot in *N*-succinylornithine transcarbamylase remained initially undetected when the structure was solved in 2002 [39]. Although the protein contains a simple trefoil knot, it is difficult to identify simply by visual inspection of the structure. The knot was finally detected independently by three groups in 2006 [4, 5, 11] by computational methods. Two knotted homologues of this protein exist: AOTCase in *X. campestris* [51] (pdb code: 1yh0) which catalyzes a reaction from *N*-acetylornithine and carbamyl phosphate to acetylcitrulline, and SOTCase in *B. fragilis* (1js1) that promotes the carbamylation of *N*-succinylornithine [39]. From a functional point of view, the two are closely related and it has recently been noted that the substitution of a single amino acid interchanges the substrate specificity [52] (SOTCase was originally misidentified as an AOTCase). Intriguingly, another more frequently observed homologue (OTCase carbamylating *N*-ornithine) adopts an unknotted structure. As noticed by Virnau *et al* [4], the knotted variant contains a rather rigid proline-rich loop (located between residues 178–185), through which residues 252–256 are threaded. This loop is missing in the more common unknotted OTCase, and consequently, the active site and the catalytic reaction are modified (for more details see [4]). Intriguingly, the two knotted variants share only about 35% sequence identity [43] while e.g., knotted AOTCase in *X. campestris* has 41% sequence identity with unknotted OTCase in *P. furiosus* (1a1s) [53]. This conundrum was addressed recently by Potestio *et al* [7], who generated a phylogenetic tree of transcarbamylase-like folds. The two knotted homologues separated early on from the unknotted ones and split again, soon after. One of the authors of this review (PV) originally thought that transcarbamylase-like proteins are ideal candidates for creating artificially knotted proteins from unknotted ones. However, protein engineering experiments took a different route (see below).

In 2008, a rather short trefoil knot was detected in a novel zinc-finger structure in *S. cerevisiae* (Rds3p, pdb code: 2k0a) by van Roon *et al* [54]. Rds3p is well conserved among eukaryotic species and required for pre-mRNA splicing. Another trefoil-knotted nucleic acid binding protein was found in the ribbon-helix-helix (RHH) superfamily of DNA-binding proteins [13, 55]. The subunit is similar to the dimeric folds of typical RHH proteins and likely resulted from a gene duplication/fusion event of two unkotted RHH motifs, which are connected by a linker. Intriguingly, the backbone of the yet uncharacterized protein MJ0366 from *M. jannaschii* (pdb code 2efv) [13, 56] contains only 92 amino acids (of which 82 are resolved in the pdb structure). It is thus the smallest knotted protein discovered so far. In the same superfamily, VirC2 from *A. tumefaciens* (pdb code 2rh3) [13, 55, 57] also folds into a trefoil and contains a similar RHH motif, even though the two proteins show only 9% sequence identity. It is worth noting that other members of the superfamily are unknotted, and the generation of a phylogenetic tree would probably be informative in this case as well.

The most deeply embedded protein knot known so far can be observed in a plant ketol acid reductoisomerase in spinach (pdb code: 1yve) [37] and rice (pdb code: 3fr8) [58]. As first noted in [3], more than 60 amino acids can be cleaved from the C-terminus of this all-alpha protein and more than 300 amino acids from the N-terminus without destroying the embedded figure-of-eight (4_1) knot (figure 1, second from left). This enzyme is involved in the biosynthesis of branched amino acids that cannot be synthesized in animals, and it is therefore a target for herbicides.

Another intriguing figure-of-eight knot can be found in the chromophore binding domain of phytochrome in *D. radiodurans* [32, 41]—a bacterium which is extremely resistant against radiation, dehydration, vacuum, acid and cold. Phytochromes are red/far-red photosensors, which can

be found in bacterial, fungal and plant kingdoms. The first structure was missing several amino acids and originally misclassified as a trefoil knot [41]. However, a computational analysis of knotted protein structures [4] pointed out that the embedded knot was likely a figure-of-eight, which was confirmed once the complete structure (pdb code: 2o9b) became available [59].

A rather intricate knot with five crossings in a projection onto a plane (5_2, figure 1, second from right) can be found in ubiquitin carboxyl-terminal hydrolases UCHs [38, 60–63]. Its discovery is interesting on its own and once again showcases the difficulties associated with knot detection in proteins. Essential parts of the structure (pdb code: 1uch) were already determined in 1997 by Johnston *et al* [38] even though the knot was not noticed back then. Two years later, the same group solved the structure of the yeast homologue YUH1 [60] (pdb code: 1cmx). Again, segments were missing, but a superposition of the two structures, at this stage, would have identified the complex knot without ambiguity. Soon after, a trefoil knot (3_1) was detected in 1cmx [3], which occurs when missing segments in the first structure of the file are connected by a straight line (the second structure in the file as well as 1uch would have yielded a 5_2 knot). In the end, it took six more years until the knot was noticed [4, 15] and the (now complete) structure of UCH-L3 [61] was categorized independently as containing a 5_2-knotted topology by two groups [4, 5]. Humans have four UCHs (UCH-L1, UCH-L3, UCH-L5 and BAP1), whereas yeast only has one (1YUH). The UCH domain consists of roughly 230 amino acids, which form a pretzel shaped knot, and around 10 amino acids can be cleaved from the N-terminus before the knot disentangles completely. UCHs are deubiquitinating enzymes (DUBs) that catalyze the deconjugation of ubiquitin from lysine side chains of protein adducts and other molecules and are thus involved in the ubiquitin–proteasome pathway. Recently [63], the structure of UCH-L5 was solved, which, in addition to its UCH domain, contains an (unknotted) C-terminal tail domain. Intriguingly, the brain also contains a large number of knotted UCH-L1 molecules [62] (pdb code: 2etl) which make up 1–2% of the brain protein mass. The normal function of UCH-L1 is still unknown, however, abnormal over-expression of the protein can be found in certain types of cancer. Mutations of UCH-L1 have also been implicated in neurodegenerative diseases like Parkinson's [62]. The experimental section of this paper summarizes recent folding experiments on UCH-L1 and UCH-L3.

The most complex protein knot known to date was detected recently [13] in α-haloacid-dehalogenase DehI in *P. putida* (pdb code: 3bjx) [40] (figure 1 right). Haloacid dehalogenases are of interest for their potential use in industrial chemical processes and for bioremediation because they catalyze the cleavage of carbon–halogen bonds from organic haloacids. A reduced representation of the protein in figure 1, right, reveals six crossings which form a so-called Stevedore knot (6_1) (a stopper knot which in former times was used by stevedores to prevent large blocks from running through the line while raising or lowering cargo [64]). The knot in DehI is deep and does not vanish until approximately 20 amino acids are cut from the C-terminus and roughly 65 residues from the N-terminus. The DehI monomer consists of two (unknotted) regions (around 130 a.a. each) which have very similar structures and are linked by a proline-rich loop that forms an arc and is an essential part of the knot structure. In the computational section, we will discuss how such a complicated structure may fold.

Whilst writing this review, we became aware of an exciting new development [45, 65]. For the first time, a group has managed to create an artificially knotted protein (pdb code: 3mlg). This protein is based on two copies of the all-alpha helical protein HP0242 (pdb code: 2ouf) from *H. pylori*. These are covalently linked together via a nine amino acid long peptide that consists of an alternating serine–glycine sequence (SGSGSGSSG). The resulting trefoil is rather deep and around 30 amino acids can be cleaved from both sides before it disappears.

For completeness, we have also included the handedness of all protein knots in table 1, which is determined as follows and described in [3, 7]. In a projection onto a plane, an orientation is chosen and each minimal crossing either contributes $+1$ if the crossing is right-handed or -1 if the crossing is left-handed. The handedness of a crossing can be determined by pointing the thumb of the right hand in the direction of the arrow on the overcrossing segment and the index finger in the direction of the arrow on the undercrossing segment. If the remaining fingers point upwards, the crossing is right-handed. If they point downwards, the crossing is left-handed. The sign of the sum determines the handedness or chirality of the knot. As pointed out in [7] both left- and right-handed variants of trefoils exist (the artificial knot is left-handed), the figure-of-eight knots are achiral by definition, UCH-L3 is left-handed and the 6_1 knot in the haloacid dehalogenase right-handed. Hence, no apparent bias in chirality exists.

4. Experimental studies to investigate the mechanisms of knot formation in proteins

Present-day theories based upon both experimental and computational studies generally suggest that cooperativity and an increasing degree of nativeness are required for a rapid and efficient protein folding reaction [66, 67]. These ideas imply that proteins should be knot free, and therefore the existence of knotted protein structures is somewhat unexpected from a protein folding point of view. It is not obvious how, during folding, a substantial length of polypeptide chain manages to spontaneously and reproducibly thread itself through a loop. Furthermore, it is possible that non-native interactions are required to initiate a threading event. An understanding of the mechanisms involved in knot formation in protein structures is therefore important for the successful calculation of *de novo* folding pathways after protein synthesis on the ribosome. Experimental studies are likely to play a key role in this due to the difficulties in predicting protein knot formation *in silico* [13, 68, 69] (see the later section on folding simulations).

The folding pathways of proteins with a variety of complicated structures have been experimentally probed to

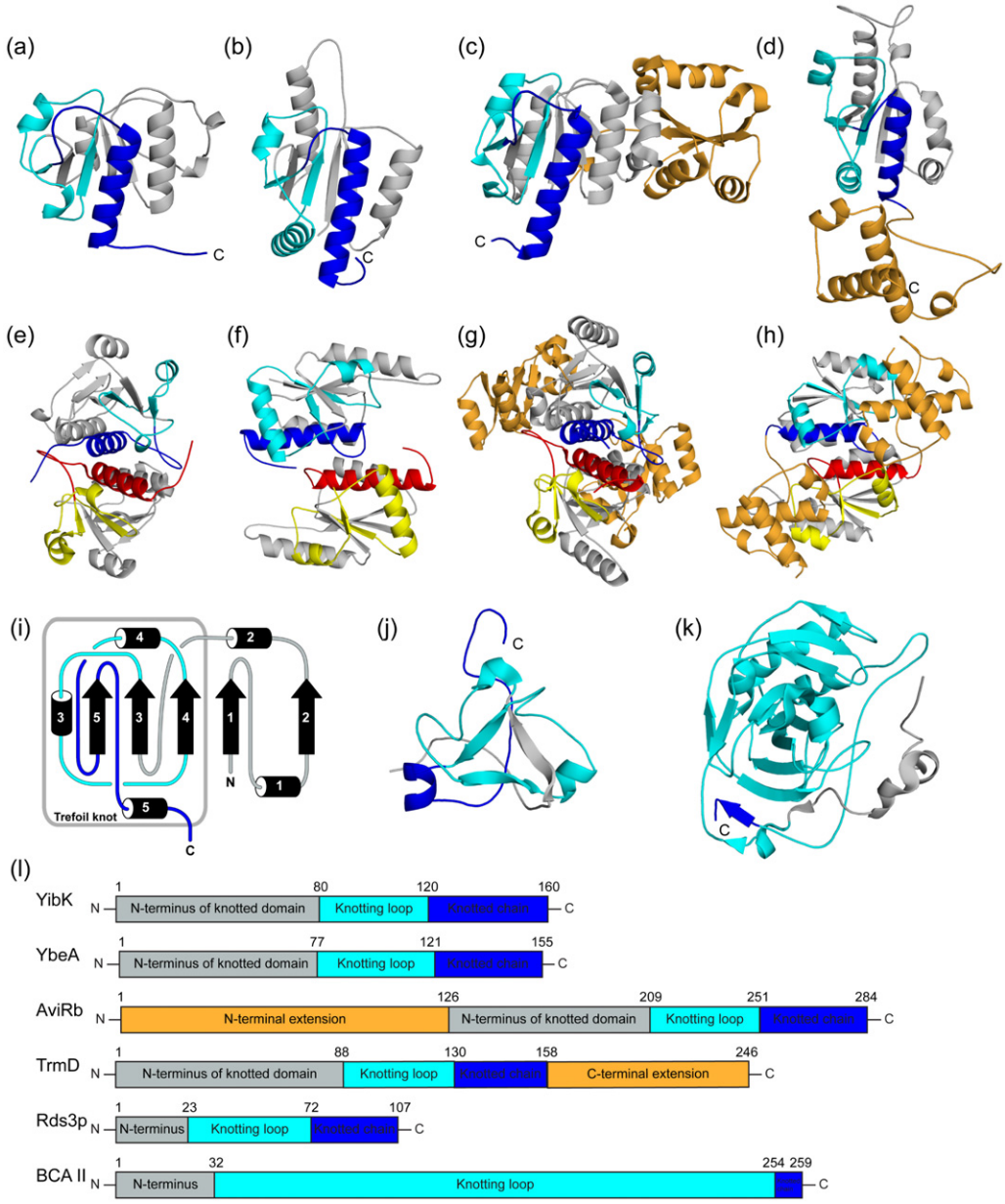

Figure 2. Trefoil knots in protein structures. (a)–(d) Ribbon diagrams of the monomeric subunits of representative α/β-knot MTases. (a) YibK from *Haemophilus influenzae* (PDB code 1mxi), (b) YbeA from *Escherichia coli* (PDB code 1ns5), (c) AviRb from *Streptomyces viridochromogenes* (PDB code 1x7o) and (d) TrmD from *Escherichia coli* (PDB code 1p9p). The knotted domains in the structures are coloured according to definitions given by Nureki *et al* [49], with the 'knotting loop' coloured in cyan (mid-grey) and the 'knotted chain' shown in blue (dark grey). Both AviRb and TrmD have additional appending domains at the N- and C-termini, respectively, and these are highlighted in orange and which in the black and white version can be identified as the domain on the right of the knotted domain (c) or below the knotted domain (d). (e)–(h) Dimeric structures of (e) YibK, (f) YbeA, (g) AviRb and (h) TrmD. All known α/β-knot MTases exist as dimers in solution; YibK and AviRb are parallel homodimers, while YbeA and TrmD dimerize in an antiparallel fashion. (i) Topological diagram to illustrate the conserved topology of the knotted domain in an α/β-knot MTase. Numbers refer to secondary structure elements. (j) and (k) Examples of other types of protein structures that contain a trefoil knot. Rds3p (j) is one of the smallest proteins (107 residues) with a trefoil knot topology to be identified. Bovine carbonic anhydrase II (k) contains a very shallow trefoil knot, where only a few residues are tucked through a wide loop in the polypeptide chain, and this has been probed in AFM studies. (l) Schematic representation to illustrate the architecture of the trefoil-knotted proteins.

gain insights into how such topological complexities arise during folding [6, 70]. These include trefoil, figure-of-eight and penta-knotted proteins, in addition to slipknotted proteins and protein catenane and rotaxanes (figure 2(a)). In these studies, various biochemical and biophysical techniques have been employed to examine folding and knot formation from simple bulk unfolding experiments using chemical denaturants, to single-molecule atomic force microscopy (AFM) measurements (figure 3). The use of protein engineering techniques to examine the effects of perturbing the knotted protein structure, from simple single-site mutations to the addition of substantial appending domains, has also

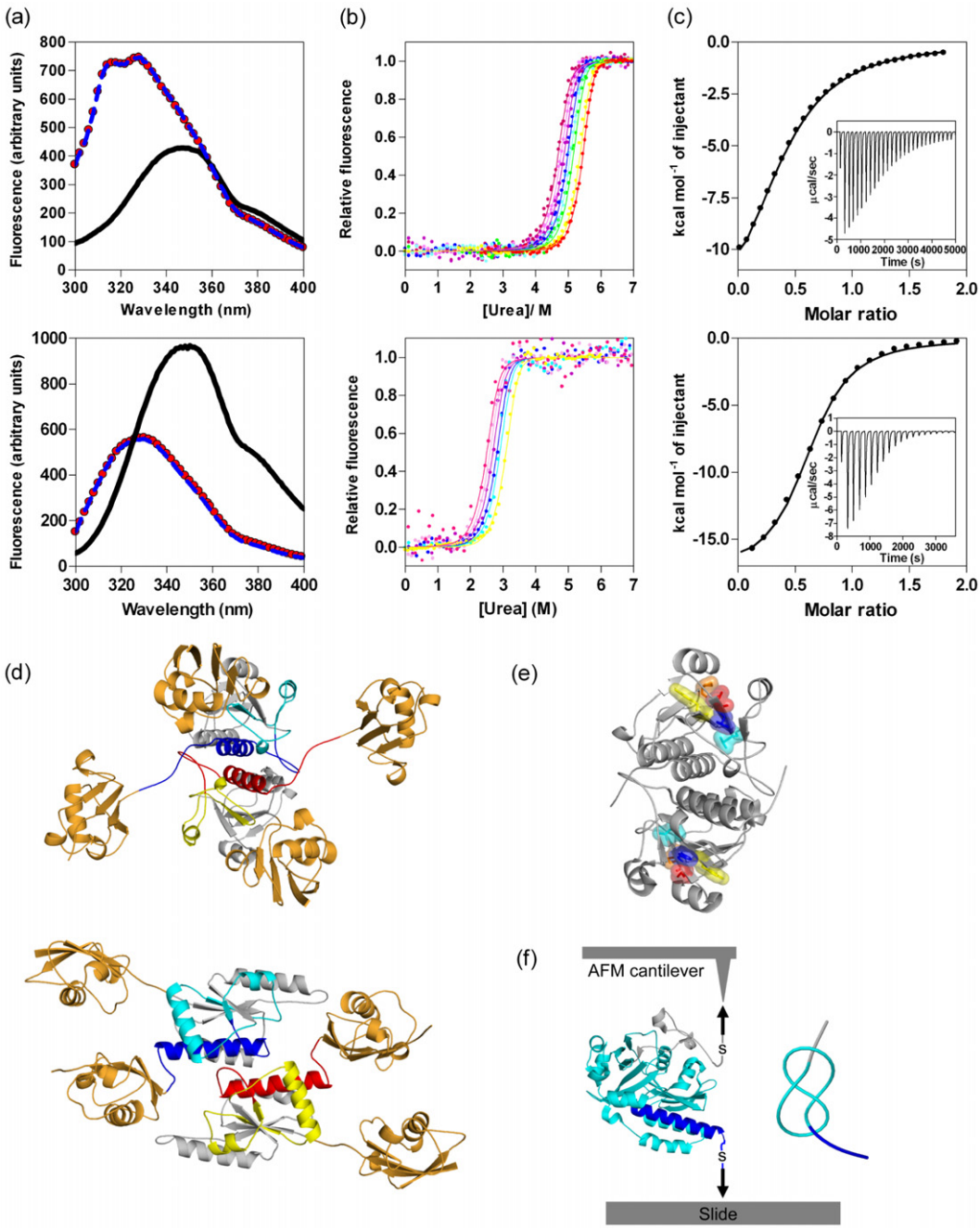

Figure 3. Examples of experimental techniques used to probe knot formation in proteins. (a) Bulk fluorescence experiments can report on the reversibility of folding of a knotted protein and are also used to measure protein stability and un/folding kinetics. Fluorescence spectra are shown for native (red circles/dark grey), unfolded (solid black line) and refolded (dashed blue/dark grey line) YibK (top) and YbeA (bottom) wild-type proteins. (b) Urea denaturation profiles for YibK (top) and YbeA (bottom) measured at various concentrations of protein (100–0.25 μM, red to purple from left to right (lower to higher concentrations of denaturant) give information about the stability of the dimeric knotted proteins. (c) Cofactor binding experiments have been used to confirm the native structure and the presence of the knotted structure in α/β-knot MTases. The binding pocket of the cofactor (S-adenosyl methionine) is located in the knotted region of the protein. Isothermal calorimetry (ITC) measurements to measure the cofactor binding affinity are shown for YibK (top) and YbeA (bottom). The insets show the original ITC traces. (d) and (e) Protein engineering has been employed to examine the effects of additional appending domains (d) or single-site point mutations (e) on the folding pathway a knotted protein. (d) Artificial multidomain constructs consisting of YibK (top) and YbeA (bottom) fused at both their N- and C-termini to another small protein, *Archaeoglobus fulgidus* ThiS (orange/mid-grey, the ThiS domains are at the far top and bottom and left and right of the dimeric knotted methyltransferase as shown) have been successfully constructed as determined by structures reconstructed from small angle x-ray scattering data. Proteins are coloured as in figure 2. (e) Single-site mutations made in the knotted region of YibK (highlighted in colour, the side chains are shown in the space filling mode) were used to probe the folding pathway of this part of the structure. (f) The effect of 'pulling' a knotted protein has been examined by single-molecule atomic force microscopy (AFM) experiments. A simplified schematic representation of an atomic force microscope is shown with the knotted chromophore binding domain (CBD) of phytochrome (PDB code 2o9c) attached to the AFM slide and cantilever by a disulfide bond. The CBD of phytochrome contains a figure-of-eight knot and the knotted region is coloured according to the definitions given in figure 2; a simplified representation of the knot is also included for clarity.

provided information on protein knotting mechanisms and folding pathways for these complex structures (figure 3).

4.1. Experiments on trefoil-knotted proteins

Trefoil knots are the simplest and the most prevalent type of knot to be identified in proteins [4, 15, 32] and their topology is such that the polypeptide chain has to thread once during folding (figure 2). Some proteins contain only 'shallow' trefoil knots, with one end of the chain extending through a wide loop by just a few residues, e.g., carbonic anhydrase B (CAB) [1]. 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. In contrast, 'deep' trefoil knots represent a more challenging protein folding problem [8]. A large number of these are found in proteins that belong to a group known as the α/β-knot methyltransferases (MTases), (figures 2(a)–(i)). Accordingly, knotted proteins within this family have been the focus of many of the experimental studies aimed at exploring the folding mechanisms involved in protein knot formation [70]. Other examples of deep trefoil knots that have been identified within protein structures include several transcarbamylase-like folds and a small zinc-finger protein (figure 2(j)) [15, 54].

The first deep topological trefoil knot to be identified was found in the catalytic domain of the hypothetical RNA 2′-*O*-ribose MTase from *Thermus thermophilus* (RrmA), an α/β protein and a member of the SpoU family [49]. Since then, many more protein structures containing similar deep trefoil knots have been deposited in the protein data bank, and have been classed together in a superfamily known as the α/β-knot MTases [71]. These knotted proteins share some common features: structural similarities to other family members as well as functional assignment make it likely that all are methyltransferases (MTases), a type of enzyme involved in the transfer of the methyl group of S-adenosyl methionine (AdoMet) to carbon, nitrogen or oxygen atoms of DNA, RNA, proteins and other small molecules [72]. All form dimers in solution, with the knot forming a large part of the dimer interface. The knot is also the location of the AdoMet binding site [49, 50, 73–78].

Experimental studies have examined the folding of two α/β MTases, YibK from *H. influenzae* and YbeA from *E. coli*. YibK is a 160-amino-acid residue protein sharing significant sequence homology with the SpoU family of MTases. Crystallographic studies indicate that it adopts a structure consistent with the α/β-knot MTase fold (figure 2(a)) [76]. Specifically, a deep trefoil knot is formed at the C-terminus by the threading of the last 40 residues (residues 121–160) through a knotting loop of approximately 39 residues (residues 81–120) (figure 3(a)). Like other α/β-knot superfamily members, YibK is homodimeric. The dimer interface involves the N- and C-terminal α helices (α1 and α5), and consists of two closely packed monomers arranged in a parallel fashion (figure 2(e)). YbeA is a 155-residue protein similar in structure to YibK. YbeA has a deep trefoil knot in its backbone structure formed by the threading of the last 35 residues (residues 120–155) through a 45-residue knotting loop (residues 74–119) (figure 2(b)). YbeA crystallizes as an antiparallel dimer, and the dimer interface involves close packing of α1 and α5 from each monomer (figure 2(f)). Its topological features are very similar to those of YibK, and YbeA displays the structural elements characteristic of all α/β-knot MTases (figure 2(i)). However, YibK and YbeA have only 19% sequence identity.

YibK and YbeA are single domain knotted proteins and are two of the smallest α/β-knot MTases identified so far, and therefore are suitable candidates for protein folding studies. It is interesting to note, however, that similar knotted domains have been observed with additional amino-or carboxy-terminal domains (figures 2(c), (d), (g), (h)) [79]. These knotted multidomain structures represent an even more challenging protein folding problem, and their existence suggests that either threading of the polypeptide chain can occur from either terminus or that long polypeptide chains are able to thread through a loop [79].

As mentioned above, α/β-knotted proteins are dimeric. Characterizing the folding pathway of a dimeric species is generally more complicated than that of a single domain monomeric protein, as at some stage during folding an intermolecular step must occur [80]. However, the presence of the dimeric structure can also be a useful probe to help to confirm that the correct knotted structure is retained after mutations made to probe the folding mechanism are introduced into the protein.

Both YibK and YbeA can be unfolded reversibly using the chemical denaturant urea to a state that lacks secondary or tertiary structure (figures 3(a) and (b)) [23, 25]. This demonstrates that their complicated knotted structure has not hindered their folding efficiency *in vitro*. Furthermore, it appears that molecular chaperones are not required for the successful folding of YibK and YbeA during simple refolding experiments using recombinantly expressed purified protein [23, 25]. The folding pathways of these α/β-MTases have been characterized using kinetic single-jump and double-jump mixing experiments (figure 4). The folding of YibK is complex due to its dimeric nature and the existence of heterogeneous species in the unfolded state that give rise to multiple folding pathways [24]. The kinetic mechanism that is most consistent with the experimental data involves the formation of two different intermediates from parallel fast-folding channels that fold via a third sequential monomeric intermediate to form the native dimer in a slow rate-limiting dimerization step (figure 4(c)). Although YbeA appears to fold via a simpler pathway with only one observable monomeric intermediate (figure 4(d)), similarities between the folding of YibK and YbeA suggest that the folding mechanisms in both proteins may be related [25]. Both show considerable resistance to chemical denaturation and share a common equilibrium unfolding mechanism involving a populated monomeric intermediate. There is no evidence for dissociation of either protein in buffer at near neutral conditions suggesting that strong dimerization appears to be a characteristic of these α/β-knotted proteins. Furthermore, both fold via sequential mechanisms that involve the slow formation of a kinetic monomeric intermediate followed by an even slower dimerization step [25]. In addition, it has been shown

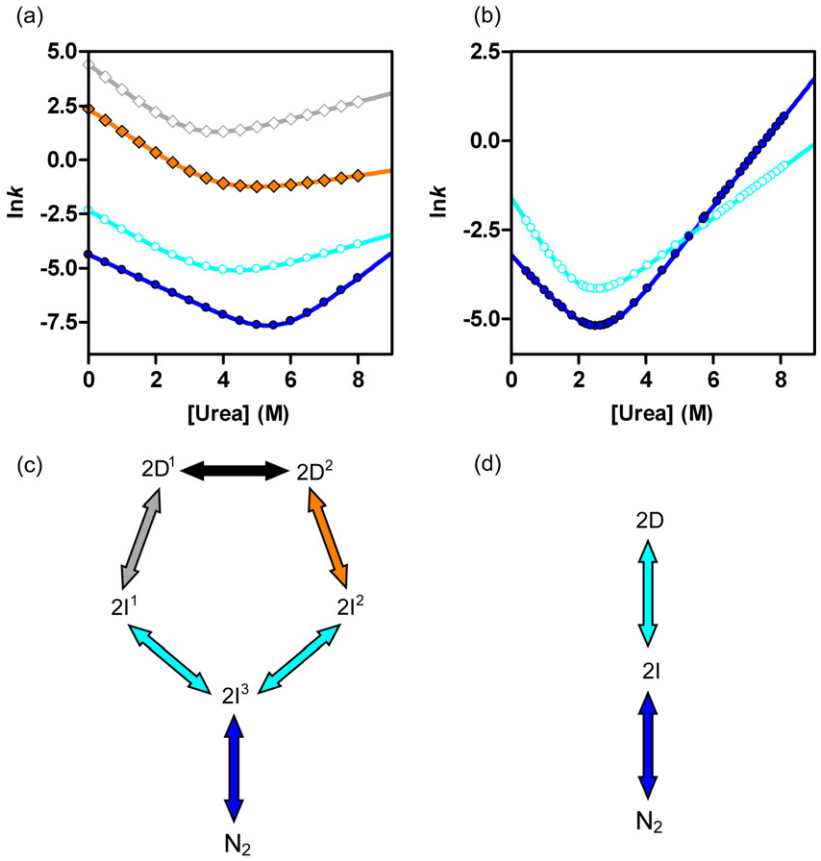

Figure 4. The folding mechanisms of wild-type YibK and YbeA. (a) and (b) Chevron plots of the natural logarithm of rate constants observed during folding and unfolding kinetic experiments at various concentrations of urea for YibK (a) and YbeA (b) [25, 27]. Phases are coloured grey, orange and cyan in order from fastest to slowest, respectively, and the phase that corresponds to dimerization is shown in blue. In monochrome the phases can be identified as follows: the fastest refolding/unfolding phase in (a) corresponds to $2D^1$ to $2I^1$ in (c), the next fastest refolding/unfolding phase in (a) corresponds to $2D^2$ to $2I^2$ in (c), the next fastest refolding/unfolding phase in (a) corresponds to both $2I^1$ to $2I^3$ and $2I^2$ to $2I^3$ in (c), whilst the slowest refolding/unfolding phase in (a) corresponds to $2I^3$ to N_2 in (c). Continuous lines represent the fit of each phase to a two-state model. (c) and (d) The folding mechanism proposed for wild-type dimeric YibK (c) and YbeA (d) [24, 25]. Arrows are coloured according to their corresponding phase in (a) and (b). In monochrome, the fastest refolding/unfolding phase at low concentrations of denaturant in (b) corresponds to 2D to 2I in (d), and the slowest refolding/unfolding phase at low concentrations of denaturant in (b) corresponds to 2I to N_2 in (d).

that the dimerization of YibK is essential for maintaining its native structure and function by cofactor binding experiments that indicate that the knotted region is not fully structured in a monomeric variant of the protein [26]. Taken together, these studies on YibK and YbeA revealed no folding features that could be directly linked to the formation of the knot.

The successful characterization of the *in vitro* folding pathways of YibK and YbeA provided the necessary groundwork for more ambitious use of protein engineering to experimentally probe the knotting and folding mechanisms of these proteins. A set of novel multidomain proteins that involved the fusion of another small protein, *A. fulgidus* ThiS, to either YibK or YbeA at their amino, carboxy or both termini were constructed to specifically examine the threading events that take place during folding (figure 3(d)) [28]. ThiS is a very stable, 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. However, cofactor binding and small angle x-ray scattering (SAXS) experiments indicated that all six of the artificial multidomain constructs were able to successfully knot and fold (figure 3(d)) [28]. Furthermore, their rates of folding were comparable to those of the equivalent wild-type protein, despite the fact that a considerably longer segment of chain must be threaded through a loop to form the native structure. Interestingly, the artificial fusion proteins with ThiS attached to both termini contain 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 (figure 3). 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 likely propagates from a loosely knotted, denatured-like state [28]. This is consistent with the existence of natural knotted α/β-MTase proteins that have additional large amino or carboxy domains [79].

The folding pathway of YibK has also been examined using single-site mutants [27]. The effect of mutations made in the knot region of YibK was investigated to provide information on its formation along the folding pathway (figure 3(e)). The kinetic data for these mutants were consistent with a folding mechanism for YibK where loose knotting of

the polypeptide backbone occurs very early on in folding, possibly even in the denatured state, but formation of the native structure in the knotted region of the protein happens late and is relatively slow [27]. These results implied that threading and folding of the protein chain *in vitro* are therefore independent, successive events. Additionally, the results of this study suggested that the heterogeneity observed in the denatured state of YibK could be a result of the knotting mechanism and be caused by multiple unfolded knotted conformers [27]. This folding model highlights the importance of early folding events in predicting how a given polypeptide chain will fold.

Most recent experimental studies have probed the chemically denatured state of knotted proteins. It has been demonstrated that it is possible to experimentally trap and detect knots in non-native protein chains by the construction of mutant knotted proteins that contain terminal cysteine residues [81]. These constructs were circularized by formation of a disulfide bond. The ability of the circularized protein to fold and form the native knotted structure was used to verify the presence of a knot under the conditions in which the disulfide bond was first introduced. Using this strategy, evidence was found to suggest that the denatured ensembles of both YibK and YbeA contain molecules that predominantly exist in the correct trefoil knot conformation necessary for productive folding to the native knotted structure. Denatured state topologies such as these may have a significant, and as yet unexplored, role in general protein folding and misfolding processes.

Another experimental technique that has been used to study the folding pathways of trefoil-knotted proteins is atomic force microscopy (AFM). Proteins can be mechanically unfolded at the single-molecule level using AFM where the protein of interest is attached, usually at its termini, between two surfaces and a force applied by increasing the distance between the tethered ends [82] (figure 3(f)). It is interesting to consider the effect of mechanical stress on a knotted protein structure, especially as several theoretical studies have suggested that a possible functional role of the knotted topology is to increase resistance to cellular translocation and degradation pathways [4, 83, 84]. Studies using this technique have been undertaken to investigate the effect of a shallow trefoil knot structure on the mechanical response of a protein. The mechanical properties of bovine carbonic anhydrase B (CAB), a protein that contains a shallow trefoil knot at its carboxy terminus (figure 2(k)) [85], have been examined [86, 87]. Upon unfolding, the protein extends to a distance much shorter than its theoretical stretching length, which indicates that the knot structure has indeed been retained and becomes taut upon mechanical unfolding. The effect of pulling knotted structures has been examined theoretically, and stretching simulations suggest that, upon tightening, knots in proteins will behave differently from those in homopolymers [88]. AFM studies on the figure-of-eight knot in the chromophore binding domain of phytochrome have also been undertaken [89] (figure 3(f)). These experiments allowed for the determination of the size of the knot under force, which appears to be a tightened chain of approximately 17 residues.

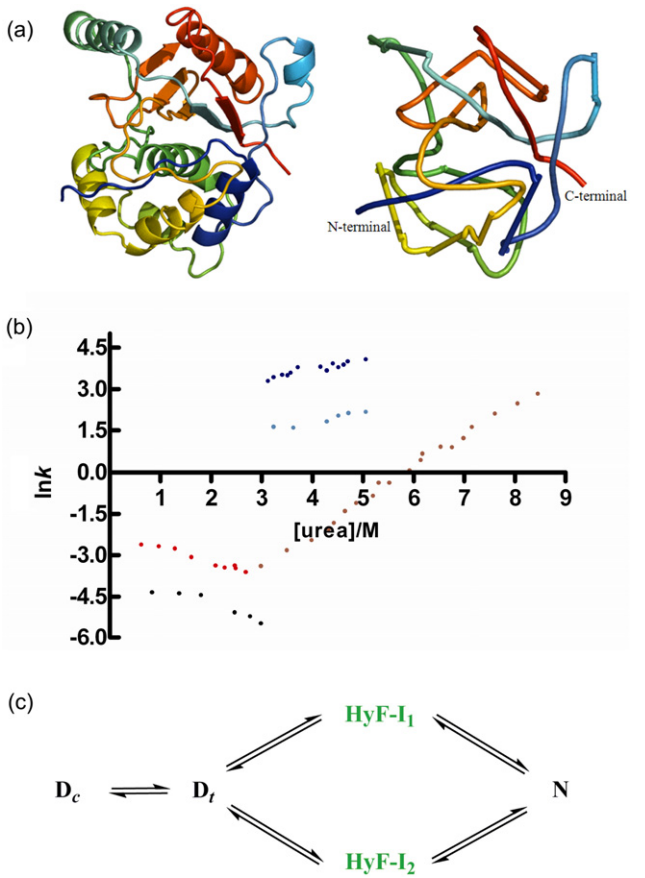

Figure 5. (a) Crystal structure of UCH-L3 and representation of the 5_2-crossings. (b) Chevron plot for the unfolding and folding kinetics of UCH-L3. Depicted are the rate constants for the only observed unfolding phase in the single-jump experiments (●, with ln k values between −3 and 2), the two slow refolding phases: k_1 (●, with ln k values between −2 and −3.5) and k_2 (●, with ln k values between −4.5 and −6), and the unfolding rate constants $k_{\mathrm{HyF\_unfold\_1}}$ (●, with ln k values between 3 and 4) and $k_{\mathrm{HyF\_unfold\_2}}$ (●, with ln k values between 1.5 and 2) corresponding to the unfolding of the hyperfluorescent intermediates. See [30] for further details. (c) Proposed kinetic scheme for the folding of UCH-L3. See [30] for further details.

4.2. Folding studies on 5_2 knotted proteins: ubiquitin C-terminal hydrolases (UCHs)

In addition to the trefoil-knotted methyltransferases described in detail above, the only other family of knotted proteins for which there has been any significant experimental characterization of their folding pathways are the 5_2-knotted ubiquitin C-terminal hydrolases (UCHs) [30]. These proteins are characterized by a structure in which there are five crossings of the polypeptide backbone chain to produce the 5_2 topology, also known as a Gordian knot (figure 5(a)). (It is interesting to note that the name Gordian knot comes from a Greek legend and this type of knot was supposedly impossible to untie (until Alexander the Great cut it in two with a sword!).) The term has therefore subsequently become associated with intractable problems. The structures of three human knotted ubiquitin C-terminal hydrolases have been solved: UCH-L1, a neuronal form associated with Parkinson's Disease [62], a more general form UCH-L3 [38, 61] and most recently UCH-

L5 [63]. It has been proposed that the knotted structure of this class of UCH helps prevent its degradation by the proteasome [4]. The stability of UCH-L1 and UCH-L3 has been determined [30, 90], and, in both cases, the *in vitro* unfolding/refolding with chemical denaturants was shown to be fully reversible under the conditions used [30]. Using both fluorescence (to probe tertiary structure) and far-UV CD (to probe secondary structure), the unfolding data for UCH-L3 could be fit to a simple two-state model showing that the unfolding/folding of this protein is highly cooperative. These measurements were used to calculate a free energy of unfolding in water of some 7 kcal mol^{-1}, typical of many other non-knotted globular proteins of similar size. In contrast, more recent studies on the unfolding of UCH-L1 under equilibrium conditions have established that an intermediate state is populated and the system is three state [90]. Hydrogen–deuterium exchange (HDX) experiments and NMR have been used to show that whilst many of the α-helices have unfolded, the central β-sheet core of the protein remains structured in the intermediate state, which is still compact and stable with respect to the denatured state [90].

The unfolding and refolding kinetics of UCH-L3 have also been studied [30], and, similar to the knotted methyltransferases discussed above, they are complex and multiple phases were observed in both unfolding and refolding experiments (see the Chevron plot in figure 5(b)). A careful analysis of both the unfolding and refolding kinetics by employing interrupted unfolding and refolding double-jump experiments was used to establish the nature of each of the kinetic phases observed and how they related to each other in a manner similar to that used to elucidate the folding pathway of YibK and YbeA. The results were used to produce a kinetic scheme for the folding of UCH-L3, which is shown in figure 5(c). The main features of the scheme are that there are parallel pathways in the unfolding/refolding reactions on which lie two intermediate states. These are hyperfluorescent and are rapidly populated during the folding reaction. It has been proposed that the hyperfluorescent intermediate states are formed by a rapid local collapse of the chain in the N-terminal region of the protein around residues Trp6 and Trp29, which become partially buried. This highly fluorescent state decays as the native structure is formed in which the fluorescence of these tryptophan residues is quenched possibly by the action of Cys50. There was also evidence for heterogeneity in the denatured state, most likely the result of proline isomerization events which give rise to a very slow phase observed in the refolding but not the unfolding kinetics, and common with many other proteins [91]. The unfolding kinetics of UCH-L1 have also been investigated and, unusually, two unfolding phases observed even in single-jump unfolding experiments (the multiple unfolding phases in UCH-L3 are only observed when using interrupted refolding experiments) [90]. Although folding kinetics and a complete analysis of the folding pathway have yet to be undertaken, there appear to be similarities with UCH-L3. The multiple unfolding phases point towards parallel unfolding pathways, and the population of metastable intermediate states.

4.3. Experiments on slipknotted proteins

Slipknotted structures have also been found in proteins [14]. These are structures in which part (but not all) of the polypeptide chain threads through a loop to form a knot, however, the chain doubles back on itself such that when pulled at the termini the knot becomes undone, figure 6(f). Because of this, slipknots are not identified using the normal methods for knot detection in proteins and have been overlooked until relatively recently. Slipknots become real knots at some point when the termini are shortened, i.e., a knot is present in some part of partial structure within the protein. Yeates and co-workers used this fact to develop a method to detect slipknots in proteins [14]. By screening 14 870 polypeptide chains in a database of unique protein structures, their approach revealed 37 protein chains which contained slipknots belonging to one of four types of protein fold: *E. coli* alkaline phosphatase (pdb code: 1ALK), equine herpesvirus thymidine kinase (pdb code: 1P6X), Glt_{Ph}, an aspartate transporter (pdb code: 2 NWL) and Leu_{Taa}, a bacterial homolog of a Na+/Cl− dependent neurotransmitter transporter (pdb code: 2A65). *E. coli* alkaline phosphatase was found to be representative of the largest family of proteins which contain deep slipknots, and the deletion of 30 residues from the C-terminus of the chain resulted in a knotted conformation. An analysis of the slipknot region showed that it is located in a five-strand region within the ten-stranded β-sheet. Two other water soluble proteins were demonstrated to contain slipknots—thymidine kinase from equine herpesvirus and protein B116 from the turreted icosahedral virus of *Sulfolobus*. Interestingly, two membrane proteins were also shown to have slipknots—the sodium-dependent aspartate transporter Glt_{Ph}, and Leu_{TAa}, a bacterial homologue of the sodium/chloride-dependent neurotransmitter transporter. The Yeates group used the homodimeric *E. coli* alkaline phosphatase to test the possible role of the slipknot within the structure. As the slipknot is formed by a protruding loop that is located near the dimer interface, mutants were created in which intermolecular disulfide bonding promoted slipknot formation. This mutant was more stable with respect to denaturation than either wild-type or control mutants, which were disulfide bonded elsewhere. This suggested that slipknots may contribute to the stability of a native state [14]. For an excellent review of topologically complex proteins and peptides see the review by Yeates [6].

While such slipknotted proteins do not technically contain a topological knot in their structure, it has been suggested that the mechanisms involved in their formation could offer insights into how deeply knotted proteins fold [6, 14]. Indeed, folding simulations on various trefoil-knotted proteins, including the experimentally characterized α/β MTases YibK and YbeA described in detail above, have suggested that such proteins may fold via an intermediate configuration which contains a slipknot [69].

4.4. Folding and stability studies on other topologically complex systems

Can we get any clues about how knotted proteins fold from studies on other topologically complex systems? Topological complexity in peptide and protein structures can take a

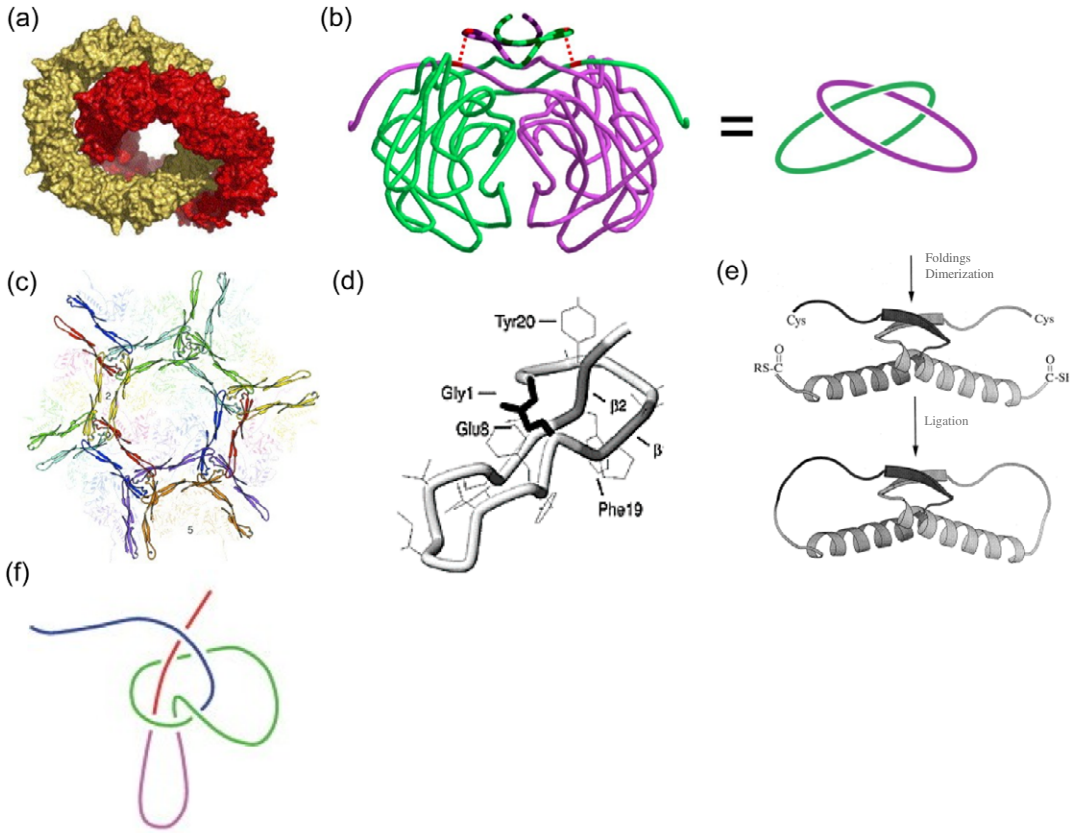

Figure 6. (a) Catenane structure formed by the dodecameric rings of bovine mitochondrial peroxiredoxin III. Figure adapted from [92]. (b) Catenane structure of the homodimeric citrate synthase from the hyperthermophilic archaeon *Pyrobaculum aerophilum.* The two subunits are shown in green and purple (light grey and mid-grey) and the disulfide bonds in red (dark grey dashed lines). Figure adapted from [93]. (c) The structure of the viral capsid protein HK97 which forms a catenated chain mail structure. Figure adapted from [95]. (d) The lariat protoknot structure of the bacterial polypeptide microcin J25. Figure adapted from [96]. (e) Catenane formation using thioester chemistry and a dimeric form of the tetramerization domain of p53. Figure adapted from reference [97, 98]. (f) Schematic representation of the slipknot formed by *E. coli* alkaline phosphatase. Figure adapted from reference [6].

number of forms. Interlinked, oligomeric rings have been observed such as those formed by bovine mitochondrial peroxiredoxin III. In this case, the protein chain folds to form subunits, twelve of which pack together to produce a ring structure. Another twelve subunits form a second ring which interlocks with the first ring to form a catenane structure [92] (figure 6(a)). However, these types of structures are very rare in proteins. Catenane structures have also been observed for the homodimeric citrate synthase from the hyperthermophilic archaeon *Pyrobaculum aerophilum* [93] (figure 6(b)) as well as the lysyl oxidase from *Pichia pastoris* [94]. The ring structures are formed by two intramolecular disulfide bonds. It has been suggested that the complex topology that arises increases the stability of the native state because a mutant linear, non-disulfide bonded homodimeric form has a much lower melting temperature [93]. The viral capsid protein HK97 also forms an interesting catenated structure [95]. In this case, an unusual isopeptide bond forms between a lysine residue in one subunit and an asparagine side chain in the neighbouring monomer. This results in interlocking rings which form a chain mail organization [95] (figure 6(c)). The interlocking rings cover the whole capsid and are thought to provide extreme stability, despite forming a rather thin protein shell. As these types of coliphages are found in severe conditions in nature, a high level of stability is presumably necessary [95]. The bacterial polypeptide microcin J25 is another example of an unusual structure with complex topology—a lariat protoknot structure made with a backbone side chain amide linkage [96]. This consists of an eight-residue cyclic structure followed by a 13-residue peptide chain, which loops back and threads through the cyclic segment (figure 6(d)). This structure is extremely stable and resistant to denaturation in 8 M urea at 95 °C over a 10 h period. It has been proposed that the formation of this type of structure is a means of obtaining a unique and stable three-dimensional fold even with short peptides [96].

Covalently linked and topologically connected protein chains have also been engineered using thioester chemistry and the tetramerization domain of p53 [97, 98] (figure 6(e)). The first attempt by the Dawson group was on a tetrameric form of the protein and used native chemical ligation [98]. Although formation of the catenane was fast and efficient, accurate thermodynamic measurements proved challenging with the tetrameric system [98]. A follow up study used a double mutant of the p53 tetramerization domain known to be dimeric [97]. In this case, catenane formation was also rapid and efficient and the structure unaffected by the ligation of the chain. With this system, rigorous stability studies were made and the topological links introduced shown

to massively stabilize the native structure by some 59 °C or 4.5 M GdmCl [97]. The formation of the links also made the protein more resistant to proteolysis [97]. Such studies are interesting because they demonstrate that a range of complex topologies can be attained with a polypeptide chain and simple chemistry, and several show that the topologically complex molecules formed have considerable stability. However, it has yet to be determined how these structures form and the chains fold, so it is unclear whether their folding pathways share any similarity with the knotted proteins discussed above. The rate of threading events in the formation of catenane structures has been addressed by the Dawson group. They used the p53 system and constructed both a protein pseudo-rotaxane and a hetero-catenane which required threading to fold [99]. They attached a variety of fluorophores to the chains in order to obtain independent reporters of both threading and folding [99]. Their results showed that whilst the threading events were slower than the corresponding folding kinetics, the threading itself was very efficient [99].

Another type of topologically complex structure is the cystine knot motif, where the complexity is a direct result of the disulfide bonds formed [17, 18, 22]. As the folding of cystine knots is an oxidative process [16] and the complexity a result of the covalent chemistry involved, it is unlikely that the folding of such species can cast much light on the folding of knotted proteins where only changes in non-covalent bonding occur.

5. Folding simulations

In the following section, we would like to discuss the computational approaches that have been used to address how knotted proteins may fold. Here, we do not intend to review protein folding simulations in general, nor simulations of coarse-grained protein-like polymer models [5, 100]. Instead, we focus on recent attempts to fold specific knotted proteins with computer simulations.

It is instructive to bring to mind some simple mechanical arguments laid out by Taylor in the context of proteins [15]. Even though it may be difficult to imagine how knotted proteins fold in the first place, it is not so difficult to create complicated knots with strings. Imagine twisting an unknotted loop over and over again before threading one end through the loop. The result will be a knot with a large number of essential crossings which can be tied or untied in a single movement. Indeed, there are quite a number of knots [101–103] which can be untied by single chain crossing (or corresponding movement), and all observed protein knots so far belong to this class with unknotting number 1.

Nevertheless, even unknotted proteins are notoriously difficult to fold in simulations. All-atom molecular dynamics folding simulations of small polypeptides in water nowadays extend to several microseconds (on supercomputers), but, it is still not possible to fold knotted proteins in this way. Unfolding at high temperatures or by mechanical stretching is easier, however, in such simulations the stretching or unfolding rates (and corresponding relaxation processes) are many orders of magnitude faster (due to current limitations of today's computers) than observed in experiments even for coarse-grained models. Atomistic simulations of this type [104] may be used to interpret AFM data. On a more coarse-grained level, unfolding simulations may reveal generic features of knotted [88] and slipknotted proteins [105], and allow for a comparison of mechanical and thermal stability of knotted and unknotted protein variants as shown for the case of transcarbamylase-like proteins [106]. Folding simulations of knotted proteins, on the other hand, are mostly restricted to the so-called Gō-models [107].

Nowadays, Gō-models typically reduce the protein to its C_α-backbone. A potential based on a generic polymer model is used, which includes terms for excluded volume, bond angles, bending and torsion. Additional attractive interactions are included for residues that are in contact with each other in the native state [108–113]. As the model is based on the final structure, which has to be known beforehand, an unfolded protein will typically fold back into the native state. From the observation of folding trajectories, conclusions on folding routes and dynamics are drawn. Knotted proteins, however, pose additional obstacles because when using Gō-models they have the tendency to fold into topologically frustrated, unknotted globules. The first study on knotted proteins following this approach was undertaken by Wallin *et al* in 2007 [68]. The authors studied the folding of the trefoil-knotted YibK (see section 3). Similar to experimental studies, they observed two parallel folding pathways, which were attributed to an early and a late stage of knot formation. However, the initial rate of successful folds was very low and necessitated additional, carefully selected non-native interactions in order to fold YibK successfully. In [69] Sulkowska *et al* folded YibK and YbeA without non-native interactions and succeeded in forming a knot in 1–2% of all cases. These trajectories were facilitated by the formation of a slipknot (not observed in the non-native simulations of [68]), which helped reduce the barriers imposed by topology.

The same approach was also pursued in the folding of the 6_1 knot in DehI [13], the most complex protein knot known to date. Even though the rate of successful folds was even lower (6 out 1000 trajectories folded into the native state), the successful trajectories (figure 7) nevertheless provide hints on how this complicated protein may fold. Two similar routes were observed. In the beginning, two aligned loops form (figure 7(b)). In the first route, the C-terminus is threaded through the loop coloured in red (dark grey) (which needs to twist once again) before the green (light grey) coloured loop flips over the red (dark grey) coloured loop (figure 7(c)). In the second route, the two steps are interchanged.

The discovery of the smallest knotted protein (MJ0366) [13] on the other hand allowed for simulations with a more detailed all-atom Gō-model in which all heavy atoms were included [114]. The underlying model [115] alleviates some of the problems of simpler models and appears less prone to topological traps. The study suggests a three-state folding mechanism for MJ0366. The first barrier is posed by the formation of a nucleation site that creates a correctly twisted loop. The rate-limiting barrier is overcome by two parallel mechanisms one of which includes the formation of a slipknot.

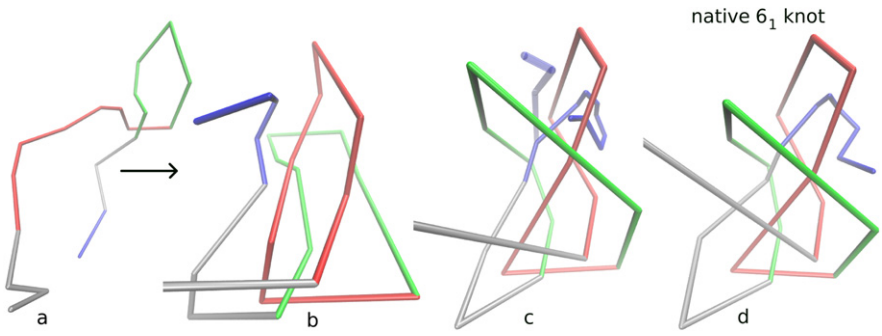

Figure 7. Snapshots taken from a simulated folding trajectory of DehI ($0 \rightarrow 6_1$). Adapted from [13].

Even though folding simulations of knotted proteins have only been undertaken for a few years and underlying models are still rather basic, simulations already provide useful information on potential folding pathways, which may help interpret and guide future experiments.

6. Conclusions and outlook

Proteins with knotted backbones, mostly disregarded or viewed as fascinating oddities in the past, are gradually being recognized as challenging and significant structural motifs. In this review, we have attempted to shed some light on how these exciting proteins form. We have provided an overview of knotted structures and placed an emphasis on recent developments in experiments and simulations that examine their folding properties.

Experimental folding studies on two different classes of knotted proteins, the α/β knotted methyltransferases and ubiquitin C-terminal hydrolases, have revealed some important features of the folding energy landscape for these topologically complex structures. Both these 3_1 and 5_2 knotted proteins can be reversibly unfolded *in vitro* using chemical denaturants to states which are largely unstructured [23, 25]. However, in the case of the methyltransferases there is now strong evidence that they retain their knotted topologies even under highly denaturing conditions [81]. This has yet to be established for the UCHs. In addition, it has been shown that all the knotted proteins studied to date have complex folding pathways where intermediate states are populated, and frequently there is heterogeneity and multiple folding paths [24, 30]. Protein engineering strategies have been employed to investigate the folding of the knotted methyltransferases and these studies have shown that the interactions in the native state that help stabilize the knotted structure are formed only late on the folding pathway [27], again highlighting the fact that non-native interactions may play a critical role in the efficient folding of these types of structure. Furthermore, protein engineering techniques have been used to create novel proteins with the deepest knotted structures yet observed [28], opening up many possibilities for the future manipulation of these structures.

Even though we have made progress into understanding how these structures knot and fold, many questions remain: it is unclear why so few proteins contain a knot. Are knots simply eliminated by evolution because they pose an obstacle to folding, or can the bias against knots be explained in terms of statistical mechanics? Globular homopolymers tend to be highly knotted in simulations [5, 116–119] like a string of Christmas beads which is rubbed in the palms of our hands [8]. Similarly, protein prediction models typically overestimate the number of knotted structures [11]. Swollen polymers, on the other hand, tend to be unknotted and remain in this state for some time after a collapse to a so-called crumpled globule [120] before 'topological equilibration' kicks in. If folded proteins are pinned in such an unknotted state, similar to a string of Christmas beads dipped in honey [8], can this idea be reconciled with the observation that some knotted proteins reliably fold back into their knotted state after unfolding [23, 24]? Is the absence of knots a consequence of non-ergodic protein dynamics or does secondary structure simply shift the length scales at which knots typically occur? The answer is probably a combination of all these factors and may be addressed with more sophisticated computational models in the future.

Other open questions revolve around the role of protein knots. Even though they tend to be highly preserved throughout evolution, their function is often unknown. They appear to modify active regions (as demonstrated for transcarbamylase-like proteins) and it has been speculated that certain mechanical quantities, like thermal stability [8] or resistance against degradation [4], are improved. These properties may play an important role in future protein engineering applications and are currently being tested with simulations [106] and experiments. It has also been suggested that a knot in the denatured state of a knotted protein could promote efficient folding by lowering the entropy of the denatured polypeptide chain [81].

At some point in the future, topology and knots in particular may become standard building blocks in the design of artificial proteins with defined function and mechanical properties. The first steps towards this have already been achieved with the design of a novel monomeric, knotted protein from an entangled but unknotted dimer [65]. Until then, however, knotty protein problems will persist!

Acknowledgments

PV would like to thank Cristian Micheletti for providing a preprint of [7] and Alexey Murzin for informing us about 3mlg.

References

[1] Mansfield M L 1994 Are there knots in proteins? *Nat. Struct. Biol.* **1** 213–4
[2] Mansfield M L 1997 Fit to be tied *Nat. Struct. Biol.* **4** 166–7
[3] Taylor W R 2000 A deeply knotted protein structure and how it might fold *Nature* **406** 916–9
[4] Virnau P, Mirny L A and Kardar M 2006 Intricate knots in proteins: function and evolution *PLoS Comput. Biol.* **2** 1074–9
[5] Lua R C and Grosberg A Y 2006 Statistics of knots, geometry of conformations, and evolution of proteins *PLoS Comput. Biol.* **2** e45
[6] Yeates T O, Norcross T S and King N P 2007 Knotted and topologically complex proteins as models for studying folding and stability *Curr. Opin. Chem. Biol.* **11** 595–603
[7] Potestio R, Micheletti C and Orland H 2010 Knotted versus unknotted proteins: evidence of knot-promoting loops *PLOS Comput. Biol.* **6** e1000864
[8] Taylor W R and Lin K 2003 Protein knots: a tangled problem *Nature* **421** 25
[9] Takusagawa F and Kamitori S 1996 A real knot in protein *J. Am. Chem. Soc.* **118** 8945–6
[10] Bryant T N, Watson H C and Wendell P L 1974 Structure of yeast phosphoglycerate kinase *Nature* **247** 14–7
[11] Khatib F, Weirauch M T and Rohl C A 2006 Rapid knot detection and application to protein structure prediction *Bioinformatics* **22** e252–9
[12] Berman H M *et al* 2000 The protein data bank *Nucleic Acids Res.* **28** 235–42
[13] Bölinger D *et al* 2010 A Stevedore's protein knot *PLOS Comput. Biol.* **6** e1000731
[14] King N P, Yeates E O and Yeates T O 2007 Identification of rare slipknots in proteins and their implications for stability and folding *J. Mol. Biol.* **373** 153–66
[15] Taylor W R 2007 Protein knots and fold complexity: some new twists *Comput. Biol. Chem.* **31** 151–62
[16] Cemazar M, Gruber C W and Craik D J 2008 Oxidative folding of cyclic cystine knot proteins *Antioxid. Redox. Signal.* **10** 103–11
[17] Craik D J 2009 Circling the enemy: cyclic proteins in plant defence *Trends Plant Sci.* **14** 328–35
[18] Kolmar H 2008 Alternative binding proteins: biological activity and therapeutic potential of cystine-knot miniproteins *FEBS J.* **275** 2684–90
[19] Mcdonald N Q and Hendrickson W A 1993 A structural superfamily of growth-factors containing a cystine knot motif *Cell* **73** 421–4
[20] Liang C Z and Mislow K 1994 Knots in proteins *J. Am. Chem. Soc.* **116** 11189–90
[21] Craik D J, Daly N L and Waine C 2001 The cystine knot motif in toxins and implications for drug design *Toxicon* **39** 43–60
[22] Craik D J, Cemazar M and Daly N L 2007 The chemistry and biology of cyclotides *Curr. Opin. Drug Discov. Devel.* **10** 176–84
[23] Mallam A L and Jackson S E 2005 Folding studies on a knotted protein *J. Mol. Biol.* **346** 1409–21
[24] Mallam A L and Jackson S E 2006 Probing nature's knots: the folding pathway of a knotted homodimeric protein *J. Mol. Biol.* **359** 1420–36
[25] Mallam A L and Jackson S E 2007 A comparison of the folding of two knotted proteins: YbeA and YibK *J. Mol. Biol.* **366** 650–65
[26] Mallam A L and Jackson S E 2007 The dimerization of an alpha/beta-knotted protein is essential for structure and function *Structure* **15** 111–22
[27] Mallam A L, Morris E R and Jackson S E 2008 Exploring knotting mechanisms in protein folding *Proc. Natl Acad. Sci. USA* **105** 18740–5
[28] Mallam A L *et al* 2008 Knotted fusion proteins reveal unexpected possibilities in protein folding *Mol. Cell.* **30** 642–8
[29] Mallam A L and Jackson S E 2008 Use of protein engineering techniques to elucidate protein folding pathways *Prog. Mol. Biol. Transl. Sci.* **84** 57–113
[30] Andersson F I *et al* 2009 Untangling the folding mechanism of the 5_2-knotted protein UCH-L3 *FEBS J.* **276** 2625–35
[31] Adams C C 2004 *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots* vol xiii (Providence, RI: American Mathematical Society) p 306
[32] Kolesov G *et al* 2007 Protein knot server: detection of knots in protein structures *Nucleic Acids Res.* **35** W425–8 Web Server issue
[33] Hoste J, Thistlethwaite M and Weeks J 1998 The first 1,701,936 knots *Math. Intell.* **20** 33–48
[34] Virnau P 2010 Detection and visualization of physical knots in macromolecules *Phys. Proced.* **6** 117–25
[35] Koniaris K and Muthukumar M 1991 Self-entanglement in ring polymers *J. Chem. Phys.* **95** 2873–81
[36] Millett K, Dobay A and Stasiak A 2005 Linear random knots and their scaling behavior *Macromolecules* **38** 601–6
[37] Biou V *et al* 1997 The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog determined at 1.65 angstrom resolution *EMBO J.* **16** 3405–15
[38] Johnston S C *et al* 1997 Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 A resolution *EMBO J.* **16** 3787–96
[39] Shi D S *et al* 2002 Crystal structure of a transcarbamylase-like protein from the anaerobic bacterium *Bacteroides fragilis* at 2.0 angstrom resolution *J. Mol. Biol.* **320** 899–908
[40] Schmidberger J W *et al* 2008 The crystal structure of Dehl reveals a new alpha-haloacid dehalogenase fold and active-site mechanism *J. Mol. Biol.* **378** 284–94
[41] Wagner J R *et al* 2005 A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome *Nature* **438** 325–31
[42] Murzin A G *et al* 1995 Scop—a structural classification of proteins database for the investigation of sequences and structures *J. Mol. Biol.* **247** 536–40
[43] Krissinel E and Henrick K 2004 Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions *Acta Crystallogr.* D **60** 2256–68
[44] Lai Y L *et al* 2007 pKNOT: the protein KNOT web server *Nucleic Acids Res.* **35** W420–4 (Web Server issue)
[45] Yeates T private communication
[46] Richardson J S 1977 Beta-sheet topology and relatedness of proteins *Nature* **268** 495–500
[47] Takusagawa F *et al* 1996 Crystal structure of S-adenosylmethionine synthetase *J. Biol. Chem.* **271** 136–47
[48] Takusagawa F, Kamitori S and Markham G D 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** 2586–96
[49] Nureki O *et al* 2002 An enzyme with a deep trefoil knot for the active-site architecture *Acta Crystallogr.* D **58** 1129–37
[50] Michel G *et al* 2002 The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot *Structure* **10** 1303–15
[51] Shi D S *et al* 2005 Crystal structure of N-acetylornithine transcarbamylase from xanthomonas campestris—a novel enzyme in a new arginine biosynthetic pathway found in several eubacteria *J. Biol. Chem.* **280** 14366–9
[52] Shi D S *et al* 2007 A single mutation in the active site swaps the substrate specificity of *N*-acetyl-L-ornithine transcarbamylase and *N*-succinyl-L-ornithine transcarbamylase *Protein Sci.* **16** 1689–99

[53] Villeret V *et al* 1998 The crystal structure of pyrococcus furiosus ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures *Proc. Natl Acad. Sci. USA* **95** 2801–6

[54] van Roon A M *et al* 2008 Solution structure of the U2 snRNP protein Rds3p reveals a knotted zinc-finger motif *Proc. Natl Acad. Sci. USA* **105** 9621–6

[55] Andreeva A and Murzin A G 2010 Structural classification of proteins and structural genomics: new insights into protein folding and evolution *Acta Crystallogr.* F **66** 1190–7

[56] Kumarevel T S *et al* 2010 *Crystal Structure of a Hypothetical Protein(MJ0366)* in preparation

[57] Lu J *et al* 2009 Agrobacterium tumefaciens VirC2 enhances T-DNA transfer and virulence through its C-terminal ribbon-helix-helix DNA-binding fold *Proc. Natl Acad. Sci. USA* **106** 9643–8

[58] Leung E W W and Guddat L W 2009 Conformational changes in a plant ketol-acid reductoisomerase upon Mg2+and NADPH binding as revealed by two crystal structures *J. Mol. Biol.* **389** 167–82

[59] Wagner J R *et al* 2007 High resolution structure of deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution *J. Biol. Chem.* **282** 12298–309

[60] Johnston S C *et al* 1999 Structural basis for the specificity of ubiquitin C-terminal hydrolases *EMBO J.* **18** 3877–87

[61] Misaghi S *et al* 2005 Structure of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate *J. Biol. Chem.* **280** 1512–20

[62] Das C *et al* 2006 Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase UCH-L1 *Proc. Natl Acad. Sci. USA* **103** 4675–80

[63] Nishio K *et al* 2009 Crystal structure of the deubiquitinating enzyme UCH37 (human UCH-L5) catalytic domain *Biochem. Biophys. Res. Commun.* **390** 855–60

[64] Ashley C W 1944 *The Ashley Book of Knots* (Garden City, NY: Doubleday) p 620

[65] King N P, Jacobitz A W, Sawaya M R, Goldschmidt L and Yeates T O 2010 Structure and folding of a designed knotted protein *Proc. Natl Acad. Sci. USA* at press

[66] Onuchic J N and Wolynes P G 2004 Theory of protein folding *Curr. Opin. Struct. Biol.* **14** 70–5

[67] Daggett V and Fersht A 2003 The present view of the mechanism of protein folding *Nat. Rev. Mol. Cell Biol.* **4** 497–502

[68] Wallin S, Zeldovich K B and Shakhnovich E I 2007 The folding mechanics of a knotted protein *J. Mol. Biol.* **368** 884–93

[69] Sulkowska J I, Sulkowski P and Onuchic J 2009 Dodging the crisis of folding proteins with knots *Proc. Natl Acad. Sci. USA* **106** 3119–24

[70] Mallam A L 2009 How does a knotted protein fold? *FEBS J.* **276** 365–75

[71] Bateman A *et al* 2004 The Pfam protein families database *Nucleic Acids Res.* **32** D138–41 Database issue

[72] Chiang P K *et al* 1996 S-Adenosylmethionine and methylation *FASEB J.* **10** 471–80

[73] Ahn H J *et al* 2003 Crystal structure of tRNA(m1G37)methyltransferase: insights into tRNA recognition *EMBO J.* **22** 2593–603

[74] Elkins P A *et al* 2003 Insights into catalysis by a knotted TrmD tRNA methyltransferase *J. Mol. Biol.* **333** 931–49

[75] Forouhar F *et al* 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** 329–32

[76] Lim K *et al* 2003 Structure of the YibK methyltransferase from Haemophilus influenzae (HI0766): a cofactor bound at a site formed by a knot *Proteins: Struct. Funct. Genet.* **51** 56–67

[77] Nureki O *et al* 2004 Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme *Structure* **12** 593–602

[78] Zarembinski T I *et al* 2003 Deep trefoil knot implicated in RNA binding found in an archaebacterial protein *Proteins: Struct. Funct. Genet.* **50** 177–83

[79] Tkaczuk K L *et al* 2007 Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases *BMC Bioinform.* **8** 73

[80] Jaenicke R and Lilie H 2000 Folding and association of oligomeric and multimeric proteins *Adv. Protein Chem.* **53** 329–401

[81] Mallam A L, Rogers J M and Jackson S E 2010 Experimental detection of knotted conformations in denatured proteins *Proc. Natl Acad. Sci. USA* **107** 8189–94

[82] Borgia A, Williams P M and Clarke J 2008 Single-molecule studies of protein folding *Annu. Rev. Biochem.* **77** 101–25

[83] Dzubiella J 2009 Sequence-specific size, structure, and stability of tight protein knots *Biophys. J.* **96** 831–9

[84] Huang L and Makarov D E 2008 Translocation of a knotted polypeptide through a pore *J. Chem. Phys.* **129** 121107

[85] Saito R *et al* 2004 Structure of bovine carbonic anhydrase II at 1.95 A resolution *Acta Crystallogr.* D **60** 792–5

[86] Alam M T *et al* 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

[87] Wang T, Arakawa H and 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

[88] Sulkowska J I *et al* 2008 Tightening of knots in proteins *Phys. Rev. Lett.* **100** 058106

[89] Bornschlogl T *et al* 2009 Tightening the knot in phytochrome by single-molecule atomic force microscopy *Biophys. J.* **96** 1508–14

[90] Andersson F I, Werrell E F, McMorran L, Crone W J K, Das C, Hsu S-T D and Jackson S E 2010 The effect of Parkinson's disease associated mutations on the deubiquitinating enzyme UCH-L1 *J. Mol. Biol.* submitted

[91] Schmid F X 1993 Prolyl isomerase—enzymatic catalysis of slow protein-folding reactions *Annu. Rev. Biophys. Biomol. Struct.* **22** 123–43

[92] Cao Z *et al* 2005 Bovine mitochondrial peroxiredoxin III forms a two-ring catenane *Structure* **13** 1661–4

[93] Boutz D R *et al* 2007 Discovery of a thermophilic protein complex stabilized by topologically interlinked chains *J. Mol. Biol.* **368** 1332–44

[94] Duff A P *et al* 2003 the crystal structure of pichia pastoris lysyl oxidase *Biochemistry* **42** 15148–57

[95] Helgstrand C *et al* 2003 The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution *J. Mol. Biol.* **334** 885–99

[96] Bayro M J *et al* 2003 Structure of antibacterial peptide microcin J25: a 21-residue lariat protoknot *J. Am. Chem. Soc.* **125** 12382–3

[97] Blankenship J W and Dawson P E 2003 Thermodynamics of a designed protein catenane *J. Mol. Biol.* **327** 537–48

[98] Yan L Z and Dawson P E 2001 Design and synthesis of a protein catenane *Angew. Chem. Int. Edn* **40** 3625–7

[99] Blankenship J W and Dawson P E 2007 Threading a peptide through a peptide: protein loops, rotaxanes, and knots *Protein Sci.* **16** 1249–56

[100] Faisca P F N *et al* 2010 The folding of knotted proteins: insights from lattice simulations *Phys. Biol.* **7** 016009

[101] Weisstein E W, Unknotting number *From Mathworld—A Wolfram Web Resource*

[102] Flammini A, Maritan A and Stasiak A 2004 Simulations of action of DNA topoisomerases to investigate boundaries and shapes of spaces of knots *Biophys. J.* **87** 2968–75

[103] Flammini A and Stasiak A 2007 Natural classification of knots *Proc. R. Soc.* A **463** 569–82

[104] Ohta S *et al* 2004 Origin of mechanical strength of bovine carbonic anhydrase studied by molecular dynamics simulation *Biophys. J.* **87** 4007–20
[105] Sulkowska J I, Sulkowski P and Onuchic J N 2009 Jamming proteins with slipknots and their free energy landscape *Phys. Rev. Lett.* **103** 268103
[106] Sulkowska J I *et al* 2008 Stabilizing effect of knots on proteins *Proc. Natl Acad. Sci. USA* **105** 19714–9
[107] Taketomi H, Ueda Y and Go N 1975 Studies on protein folding, unfolding and fluctuations by computer-simulation. 1. Effect of specific amino-acid sequence represented by specific inter-unit interactions *Int. J. Peptide Protein Res.* **7** 445–59
[108] Clementi C, Jennings P A and Onuchic J N 2000 How native-state topology affects the folding of dihydrofolate reductase and interleukin-1 beta *Proc. Natl Acad. Sci. USA* **97** 5871–6
[109] Clementi C, Nymeyer H and Onuchic J N 2000 Topological and energetic factors: what determines the structural details of the transition state ensemble and 'en-route' intermediates for protein folding? An investigation for small globular proteins *J. Mol. Biol.* **298** 937–53
[110] Kaya H and Chan H S 2003 Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *J. Mol. Biol.* **326** 911–31
[111] Knott M and Chan H S 2006 Criteria for downhill protein folding: calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity *Proteins-Struct. Funct. Bioinform.* **65** 373–91
[112] Levy Y, Wolynes P G and Onuchic J N 2004 Protein topology determines binding mechanism *Proc. Natl Acad. Sci. USA* **101** 511–6
[113] Wallin S and Chan H S 2006 Conformational entropic barriers in topology-dependent protein folding: perspectives from a simple native-centric polymer model *J. Phys.: Condens. Matter* **18** S307–28
[114] Noel J K, Sulkowska J I and Onuchic J N 2010 Slipknotting upon native-like loop formation in a trefoil knot protein *Proc. Natl Acad. Sci. USA* **107** 15403–8
[115] Whitford P C *et al* 2009 An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields *Proteins-Struct. Funct. Bioinform.* **75** 430–41
[116] Grosberg A and Nechaev S 1993 Polymer topology *Adv. Polym. Sci.* **106** 1–29
[117] Mansfield M L 1994 Knots in hamilton cycles *Macromolecules* **27** 5924–6
[118] Lua R, Borovinskiy A L and Grosberg A Y 2004 Fractal and statistical properties of large compact polymers: a computational study *Polymer* **45** 717–31
[119] Virnau P, Kantor Y and Kardar M 2005 Knots in globule and coil phases of a model polyethylene *J. Am. Chem. Soc.* **127** 15102–6
[120] Grosberg A Y, Nechaev S K and Shakhnovich E I 1988 The role of topological constraints in the kinetics of collapse of macromolecules *J. Physique* **49** 2095–100