Available online at www.sciencedirect.com

ScienceDirect

Computational Biology and Chemistry

Computational Biology and Chemistry 31 (2007) 151–162

www.elsevier.com/locate/compbiolchem

Review

Protein knots and fold complexity: Some new twists

William R. Taylor[*]

Division of Mathematical Biology, National Institute for Medical Research,
The Ridgeway, Mill Hill, London NW7 1AA, UK

Received 16 March 2007; accepted 17 March 2007

Abstract

The current knowledge on topological knots in protein structure is reviewed, considering in turn, knots with three, four and five strand crossings. The latter is the most recent to be identified and has two distinct topological forms. The knot observed in the protein structure is the form that requires the least number of strand crossings to become un-knotted. The position of the chain termini must also correspond to a position that allows (un) knotting in one move. This is postulated as a general property of protein knots and other more complex knots with this property are proposed as the next most likely knots that might be found in a protein. It is also noted that the "Jelly-roll" fold found in some all-β proteins would provide likely candidates. Alternative measures of knottedness and entanglement are reviewed, including the occurrence of slip-knots. These measures are related to the complexity of the protein fold and may provide useful filters for selecting predicted model structures.

Keywords: Protein knots; Fold complexity

Contents

[*] Tel.: +44 208 816 2298; fax: +44 208 816 2460.
E-mail address: wtaylor@nimr.mrc.ac.uk.

1476-9271/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2007.03.002

1. Introduction

That a given protein sequence folds into a unique 3D structure is remarkable, but even more remarkable is that some have been found to fold into knotted structures.[1] Partially knotted protein structures have been known for some time but these were not very securely tied (in human terms) (Mansfield, 1994, 1997). In the last few years, however, some well-tied knots have been found. These include the topologically simplest knot types with three and four crossings.[2] The former being the simple trefoil knot seen in a RNA methyltransferase (Michel et al., 2002) and the latter the figure-of-eight knot in a plant acetohydroxy acid isomeroreductase (Biou et al., 1997). These knots occur deeply embedded in their protein folds, requiring that a considerable length of the chain must be threaded through a loop for their formation. Roughly 30 residues for the methyltransferase and 70 for the isomeroreductase (Taylor, 2000). The protein knot story took a new twist recently with the identification of a protein containing a knot with five crossings (Virnau et al., 2006) and although this knot is less deeply embedded than the others, it opens some interesting topological issues. In this review, I will review the current state of knowledge of the two simplest knot types then examine the topological implications of this new knot. This leads to a consideration of slipknots and tangles followed by a general consideration of the complexity of protein folds and ways in which this can be measured. Finally, the implications of these topological concepts for the prediction of protein structure are examined.

2. Trefoil knots

2.1. RNA methyltransferases

Over the period since the identification of a trefoil knot in the RNA methyltransferase structure (PDB codes: **1ipa** (Nureki et al., 2002), **1gz0** (Michel et al., 2002)) several new variations have been reported (Zarembinski et al., 2003; Cuff et al., in press). While the nature of the knotted carboxy-terminal domain is common, the amino terminal domain can vary. The knotted domain is now classed in SCOP as a β/α-knot fold and is arranged in 5 sub-families with a total of 15 structures,[3] all of which are dimeric.

These proteins all bind *S*-adenosylmethionine (SAM) as a cofactor in the knotted binding domain that shares some characteristics with the Rossmann fold of the more familiar nucleotide binding domain.[4] A closer approximation of the Rossmann fold is found in a different family of methyltransferases which also bind the same co-factor. These are the catechol *O*-methyltransferases (e.g. **1vid** (Vidgren et al., 1994)), referred to below as ‘classic’. The differences in these folds lie in their carboxy-terminal halves (Fig. 1). For the classic methyltransferase fold, the simple deletion of the C-terminal β-strand restores it to the dinucleotide binding fold. For the RNA methyltransferases, a swap between the location of strands 4 and 5 is needed to regenerate the dinucleotide binding fold. The former is a change the might easily occur through chance insertion and deletion but the latter change requires the knot to be undone. Although this would seem to be a greater degree of change, it may be sufficient simply to reverse the relative hydrophobicity between β-strands 4 and 5 to restore the Rossmann fold. It has been noted by Shakhnovich and colleagues (Wallin et al., 2007) that strand 5 is very hydrophobic (as would be expected for its buried location), if this property were to be transfered to strand 4 then the folding of the chain might be switched into a different pathway.

If there is any evolutionary connection between these folds, it is maintained more strongly in the N-terminal half of the domain, possibly suggesting an independent common ancestor for the N-terminal half of the domain comprising an ancient adenine binding function. Such a history would be supported by the position of the adenine in the NAD binding dehydrogenases and the classic methyltransferases, however, in the RNA methyltransferases the adenine is bound in the knotted C-terminal part of the domain (Lim et al., 2003). Furthermore, in the most minimal occurrence of the knotted domain (**1to0**), the N-terminal part is further reduced to a single βα-unit, suggesting any evolutionary link through this part may not be significant.

The folding pathway of a simple member of the family (YibK) has been investigated (Mallam and Jackson, 2005, 2006).

[1] The term “knot” can only be applied to a closed path and protein knots should properly be called “open knots”. This distinction is only a problem if it is not clear how to join the free ends (for example if one is buried deep in the structure).

[2] The number of crossings in a knot is defined as the minimal number of times a piece of chain must lie over another in the projected planar image of the knot. These are more properly referred to as “over-crossings”. Each crossing has a hand so knots are often chiral with left and right-handed forms. Much more information on knots can be found in “The Book of Knots” (Adams, 1994) and many pictures of knots (along with the theory of their classification) can also be found on the internet. A few such sites are: http://knotplot.com/, http://mathworld.wolfram.com/SolomonsSealKnot.html, http://www.popmath.org.uk/exhib/pagesexhib/table.html.

[3] A recent search using the Dali program (http://www.ebi.ac.uk/dali/) found 14 proteins, 1 of which (1z85A) did not have knot because of a break in the chain.

[4] The term “Rossmann fold” strictly applies only to a sequence of three βα-units which by itself is not found as an isolated domain. The dinucleotide binding domains, such as the dehydrogenases, have a double Rossmann fold and the term is often applied to this larger intact functional domain which has a six-stranded β-sheet with helices packed on both sides.

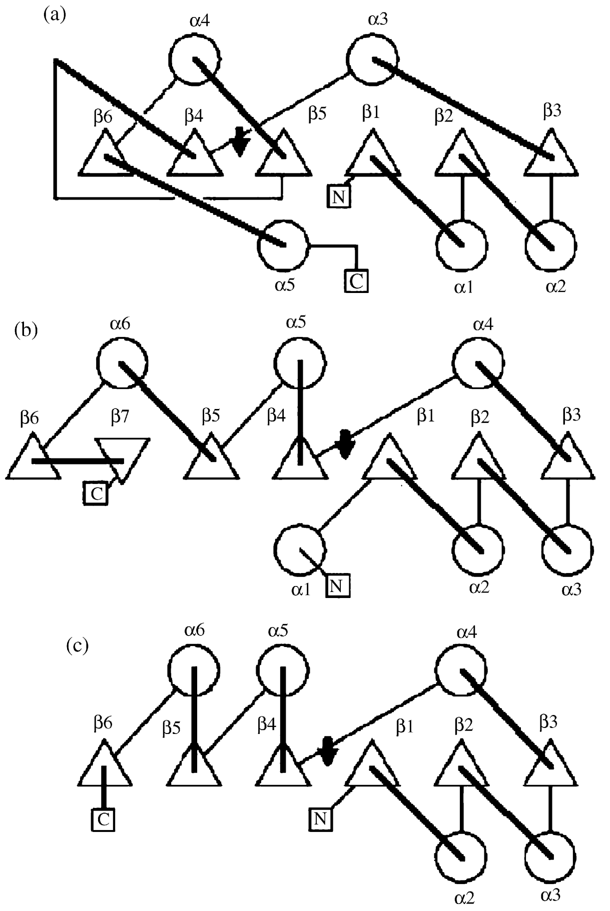


Fig. 1. Topological relationships in adenine cofactor binding folds. The folds of three βα class domains that all bind adenine based nucleotide co-factors are shown with β-strands as triangles and α-helices as circles. (a) The knotted C-terminal domain of an RNA methyltransferase (*e.g.* **1j85**). (b) A 'classic' methyltransferase fold (*e.g.* **1vid**) from which two additional helices have been removed from the amino terminus for clarity. (c) The double Rossmann fold found in the dinucleotide binding domain of NAD utilising enzymes (*e.g.* **31dh**). All these folds have a common start (β1–α4). The classic methyltransferase fold (b) differs from the dinucleotide binding fold (c) only through the insertion of an additional C-terminal strand (β7) whereas in the RNA methyltransferase fold (a) strands β4 and β5 have swapped positions. The smallest RNA methyltransferase fold also lacks an αβ-unit (α2, β3) on the edge of the domain. The figures were based on Ref. (Lim et al., 2003), Fig. 3. Arrows mark the cofactor binding centre.

Although this protein contains just the knotted domain, it is only functional as a dimer which adds some complications to the kinetics (Mallam and Jackson, 2007a). It appears that the folding kinetics of the knotted protein is not dissimilar to other unknotted proteins of a similar type. However, it may be noteworthy that the YibK protein contains 10 proline residues, one of which is required in the rare *cis* conformation. The time taken for *cis*/*trans* isomerisation may add a delay to facilitate loop penetration. (See Gloss, 2007 for further discussion.) There is also an element of ambiguous secondary structure propensity in the knotted region which may play some part (Wallin et al., 2007). Recent studies on a related protein (YbeA), however, cast doubt on the importance of the *cis* proline as it has similar folding properties to YibK but no *cis* proline (Mallam and Jackson, 2007b).

2.2. *Transcarbamylase fold*

A simple trefoil knot can be found in the structure of two transcarbamylase-like enzymes (**1js1**, **1yh1**) (Khatib et al., 2006). The knot occurs in the carboxy terminal domain of these double domain structures and is formed by the intertwining of two loops in adjacent βα-units. Except for this feature, the fold of this domain is a conventional double Rossmann fold and the knot is formed by linking the loops connecting β1–α2 and β4–α5 in the numbering of Fig. 1c. Both loops are relatively long and occur at the pseudo-symmetry point between the Rossmann folds. Crossing loops are rather rare in proteins and the few other examples (*e.g.* **2csmA** and **1b65A**) are not associated with a knotted topology. Long loops can sometimes be poorly defined in X-ray diffraction studies leading to the possibility that the path of the chain has been mistraced through the electron density. However with two independent structures this seems to be an unlikely explanation for these crossed loops.

An interesting observation on this family of proteins is that the knotted protein *N*-acetylornithine transcarbamylase (**1yh1**) has almost the same structure as the unknotted enzyme ornithine transcarbamylase (**1c9y**)) (Virnau et al., 2006), with 2.6 Å over 300 α-carbon positions. Except for an extended C-terminus in **1c9y**, the only differences are a minor insertion in the N-terminal domain and the swapping of the two adjacent loop positions in the C-terminal domain. However, this latter change makes the difference between having a trefoil knot in **1yh1** and none in **1c9y**. This is not a difference caused by a small change in the length of the termini but is a fundamental change of topology deep in the protein fold. It would appear that, in this family, the knot is not an integral part of the fold but is a minor variation being used to adjust the placement of catalytic residues in the formation of the active site.

2.3. *SAM synthetase*

The only other protein family to contain a trefoil knot also binds the SAM co-factor but as this protein has a completely unrelated fold to the RNA methyl-transferase (or any other), there can be no functional or evolutionary connection. The SAM synthetase is a three-domain protein with pseudo-three-fold symmetry, having derived from a domain triplication followed by strand swapping between the β-sheets in a cyclic manner. It is the path of the chain through the three domains that creates the knot and not the fold within any isolated domain. Even extracting the more complex C-terminal domain with its swapped strand does not create a knot.

3. Figure-of-eight knot

The figure-of-eight knot seen in the C-terminal α-helical domain of the plant ace-tohydroxy acid isomeroreductase (**1yve** and homologues **1yrl**, **1qmg**) remains the most deeply embedded knot known. This protein, which is a class-II ketol-acid reductioIsomerase (KARI), has a 250 residue αβ nucleotide binding domain to its amino terminus and 70 residues trailing on its carboxy terminus. As proposed previously, the knotted domain can best be explained by a duplication followed by a helix swap (Taylor, 2000) (with deletion of the second αβ domain). A dimeric precursor of the knotted domain can be found in the class-I KARI structure (Ahn et al., 2003) (PDB code: **1np3**) in which the terminal segments of the monomers in the dimer make a closest approach of just over 10 Å, suggesting that a simple duplication (which requires these ends to join) would be quite viable. The subsequent deletion of one of the nucleotide binding domains would then leave a larger gap of over 20 Å but this can be closed easily by remodelling a few of the residues on each terminus at either end of the gap. The resulting (knotted) dimeric fusion has an RMSD with the true knotted domain of 5 Å over 260 residues.

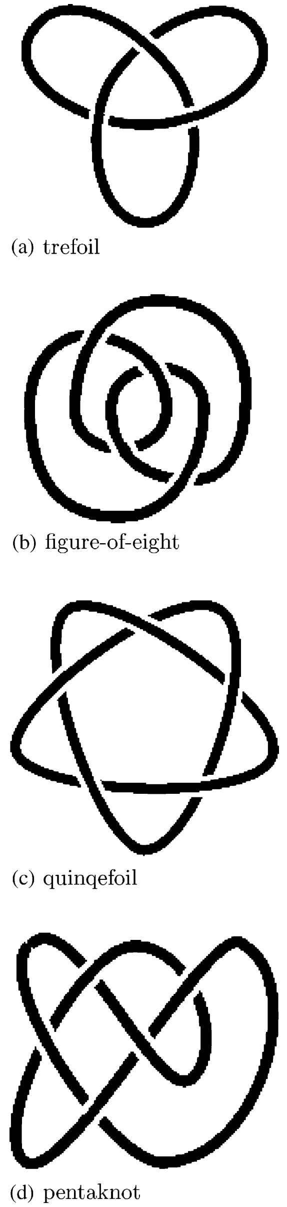
Other duplication, swapping and deletion events have led to a number of related folds including glycerol-3-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase (PGDH) and similar oxidoreductases (Andreeva and Murzin, 2006), none of which are knotted. These enzymes, collectively referred to as the PGDH-like oxidoreductases, all contain a conserved nucleotide binding domain, without which the relationships among the all-α catalytic domains would be very difficult to deconvolute. The all-α domain of PGDH and the corresponding knotted domain in the KARI class-II structure both contain a clear internal duplication but in the knotted domain, the core helices are swapped across the pseudo-two-fold.

The relationship between these domains is not just a simple exchange of two helix positions and there is no single rearrangement that would transform one fold to the other. However, an indirect link can be traced through the all-α-dimerisation domains of two further dehydrogenases, UDP-glucose dehydrogenase (**1dliA**, UDPGDh) and GDP-manose dehydrogenase (**1mv8A**, GPDMDh) which are related by a swapped pair of helices (Andreeva and Murzin, 2006). When a dimeric fusion is constructed from the GPDMDh structure, there is a plausible superposition over the core. This comprises the long central helices at the start of each duplication and two following helices but the link between these symmetric halves has no correspondence. Between the KARI class-I and the GDPMDH structures, the RMSD for these segments is 8.5 Å over 94 matched positions, with most of the error contributed by the third helix.

4. A new five-crossing knot

4.1. Pretzel and Donut knots

The identification of a new five-crossing knot in a ubiquitin hydrolase protein (**1xd3** (Misaghi et al., 2005)) (Virnau et al., 2006) has given a new twist to the protein knot story. For this new knot, the authors report that it is necessary to delete 12 residues from the carboxy terminus before the knot is undone. To determine this (and other knot sizes) they used a method of making multiple random external extensions to the exterior of the protein (following Mansfield, 1994). However, using an alternative approach that makes an internal connection through an iteratively smoothed model (Taylor, 2000) only seven deletions were necessary.


(a) trefoil

(b) figure-of-eight

(c) quinqefoil

(d) pentaknot

Fig. 2. The four simplest knots: (a) the trefoil with three crossings, (b) the figure-of-eight with four crossings, and two forms of the five crossing knot, (c) the Seal of Solomon, more commonly seen as the pentagram of Black Magic, denoted 5_1 in knot nomenclature and (d) a less symmetric knot can be made with five crossings denoted 5_2. Knots were drawn using the spline tool in GIMP.

Although this knot is not deeply embedded, five crossings is the first chance for a knot to adopt distinct topologies (apart from the trivial mirror image forms). For the five-crossing knot there are two distinct forms: one form is the seal of Solomon and is denoted as 5_1, The other is the twisted form, denoted 5_2 which can be drawn in the shape of a pretzel. The former can be inscribed on the surface of a torus (donut) while the latter is most simply made by linking the loops at the ends of a twisted circle (Fig. 2)

As proteins are not circular, the twist form of the knot can be made by threading a free end through the loop at the top of a twisted hairpin. This is quite a simple operation requiring only a single strand to be fed through a loop once. To make the donut form, however, there is no way of making it that does not involve two operations of feeding a terminus through a loop. This undoubtedly explains why the pentaknot[5] found in a protein has the pretzel-like form (5_2) and not the donut form (5_1).

As loop penetration is likely to be the rate limiting event, knot formation will depend firstly on the number of times a threading event must occur, secondly, on the number of residues that must be fed through the loop and only lastly on the number of crossings in the knot.

4.2. *An accidental strand crossing move*

In the earlier survey, Taylor (Taylor, 2000) also included a knotted ubiquitin hydrolase (**1cmx**) which is a clear homologue of the protein (**1xd3**) with the new pentaknot. The SAPit program calculates a good superposition of 2 Å RMSD over 200 equivalent α-carbon positions with 33.7% sequence identity, yet **1cmx** has a left-handed trefoil knot that differs from the pentaknot in **1xd3**. The reason for this is immediately apparent when the two structures are compared. The older **1cmx** structure has a missing section of chain between Asp.79 (res. 57) and Asn.62 (res. 58). If these broken ends are connected with a straight line, the resulting 'virtual' chain path passes through a β-sheet, just inside the edge strand by 1 Å and executes the move needed to convert the pentknot to a trefoil. If any physically reasonable connection had been made between these ends, the pentaknot would have been retained. It is difficult to understand why this was not considered in the original study.

In terms of knot theory, this is not an unexpected relationship as a change in a single crossing (changing an over-crossing to an under-crossing in the projected form of the knot) is sufficient to transform the quinfoil knot to the trefoil (Flammini et al., 2004). One such move can reduce the quinfoil to trefoil and one more will then unknot the resulting trefoil. Because of the symmetry of the toroidal form, such a move at any point where the strands cross will reduce the quinfoil to a trefoil. However, the knot in the ubiquitin hydrolase is the less symmetric twist form and if the move was made on the loops at the ends of the twisted circle that close the knot, it would become immediately unknotted. This 'loop' is where the chain termini lie in the open form of the protein knot and does not correspond to the chain break in **1cmx**, therefore any other change in crossing topology will down-grade the knot to a trefoil.

5. The handedness of knots

Each over-crossing in a knot can be assigned a hand that can be determined by pointing a thumb along the chain direction of the over-crossing strand. If the under-crossing strand passes right to left then it is right-handed (R), otherwise left (L). This corresponds to the direction of a magnetic field around a current carrying wire and not to the hand determined between strands in a β-sheet where the direction of the chain is not significant. As proper knots are circular there is no unique chain direction but the rule works in both directions. In knot terminology, 'R' is designated by '+1' and 'L' by '−1' and a trefoil made up from three '+1' crossings is, called right-handed. A simple knot nomenclature (similar to the Dowker notation) can be obtained by simply listing the hand of the crossings as a string of L/Rs in the sequential occurrence of their over-crossings (Taylor, 2000). In this nomenclature a right-handed trefoil is "RRR" and left-handed is "LLL", while the figure-of-eight is "RLRL" or "LRLR".[6] In a circular chain these would not be distinct (*i.e.* the figure-of-eight knot is achiral) but in a protein, the termini mark a start point in the circle and the natural chain direction (N → C) defines a unique direction, giving two enantiomers. The knot in **1yve** is the latter form.

Virnau et al. (2006) provided a survey of the known knots in the protein databank, of which there are now almost 40. These are almost all the simple trefoils with just the figure-of-eight knot in the isomeroreductase (**1yve** and homologues) and the ubiquitin hydrolase and a new homologue (**1xd3**, **2etl**). Unfortunately, the authors do not distinguish the chirality of the knots in their compilation but from a previous study (Taylor, 2000) both chiral forms of the trefoil were observed, however, the only left-handed trefoil was in the ubiquitin hydrolase **1cmx**. As this knot can now be reclassified as an incomplete pentaknot, it becomes even more unexpected that all other known trefoils are right-handed.

Given the observation that all trefoil containing proteins are of the βα-class, it is tempting to associate this bias with the distinct preference for the right-handed connection between βαβ-units in proteins. To investigate this, a knot can be constructed on a secondary structure lattice (Taylor, 2002) such that the knot derives entirely from the positions of the secondary structure elements and not from any unusual loop crossings. For a protein with five β-strands and four α-helices in strict alternation, a knotted fold containing a trefoil is one of the few topologies that preserve the right-handed connections and have no crossing loops (Taylor and Aszódi, 2005). While not a proof, it is suggestive that the knot obtained in this theoretical construct has the same hand as the exclusive hand observed in native proteins (Fig. 3).

Despite having almost 40 known structures with knots, they fall into only a few distinct families and for the trefoil, there is only four: the carbonic anhydrases, the RNA methyltransferases, the SAM synthetase and the transcarbamylase. Despite the possible link to the βα hand, with a sample of four, the bias towards one hand does not seem so striking and might easily be due to chance, especially as the SAM synthetase knot is an interdomain construct. In addition, the observation that the crossings in the pentaknot, which is also a βα-class protein, are of opposite hand

[5] Following a progression from the trefoil, the terms quinquefoil (contracted to quinfoil) or pentafoil (contracted to pentoil) are sometimes used for the Seal of Solomon. Retaining the same dead language, the tre-/quin-pairing should be preferred. These terms are specific to the torus forms of the knot (5_1) and there is no generally accepted trivial name given to refer to both knots with five over-crossings. To avoid repeated use of the phrase "five-crossing knot" the term "pentaknot" will be used here for both five-knots jointly.

[6] Taylor originally used the definition of the hand across the strands (as in a β-sheet) so switching left and right. The alternative knot convention is used throughout this review.

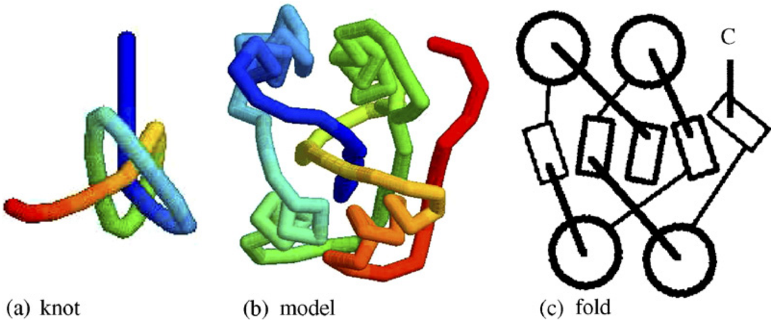


Fig. 3. A knot in a fake βα-fold. A unique trefoil knot can be constructed in a three-layer αβα architecture using only secondary structure combinations with right-handed connection and no crossing loops. This creates a right-handed trefoil knot (a) from the model (b) based on the topology (c). The C-terminal αβ-unit is not essential and the first four strands with positions 2413 in the sheet are sufficient to create the knot. This topology in not found in native proteins (Ruczinski et al., 2002).

to the crossings in the trefoils would seem to negate any simple molecular explanation of the bias. No additional support is provided either way by the crossings in the figure-of-eight knot which requires alternating handedness in successive crossovers and this knot is contained in an all-α domain.

6. Predictions for larger knots

6.1. Topological constraints on protein knots

If human experience with tying knots has any relevance to the approach used by proteins, then the distinction between the two forms of the pentaknot can be taken as predictive of the probability of observing more complex knots in the future. Any torus knot (except the trefoil) will involve repeated feeding of the chain through the centre (the hole in the donut) and should be very rare while any of the twist-knots require only a final tuck of the terminus through a single loop in the midpoint of the hairpin (Fig. 4). In theory, these should all be equally probable except that as more twists are put into the hairpin, the termini and the loop may become more separated. This means that a knot of the 5_2 class, despite having five crossings still needs only one loop penetration event.

Neglecting handedness, the two simplest knots are unique and the pentaknot has only two forms. However, these properties pertain to ideal closed circular knots and as noted previously (Taylor, 2000), the point where the protein termini mark a break in the circle introduces an additional element of asymmetry. All breaks in the toroidal forms (trefoil and quinfoil) are equivalent so there can be no distinction either in structure or in folding where the termini lie. In the figure-of-eight knot there are two distinct pairs of linked loops but these differ only in handedness which would not be expected to provide any preferred location for the termini. With the twist (5_2) form of the pentaknot, however, it was seen above that a loop crossing that unlinks the end loops leads to no knot (called the "unknot") while one in any other crossing leaves a trefoil knot. It can generally be stated therefore that it will be very advantageous for knot formation if the termini lie in a loop that can untie the knot with one unlinking move, or in terms of protein folding, it is necessary to pass only one segment of chain through a loop. Such preferences might be investigated experimentally by cyclic permutation.

As the number of crossings increases, so do the number of topological variations, with three for the six-crossing knot and seven for the seven-crossing knot (neglecting handedness in both). If any more complex knots are discovered, the above reasoning would suggest that any 'loop' crossing that leads directly to the unknot, and contains the termini, would be a likely candidate to occur in proteins. Such knots have an unknotting number of 1 and besides the twist series: 3_1, 4_1, 5_2, 6_1, 7_2, 8_1, 9_2, ... (see Fig. 4 for the first three topologies), all the six-crossing knots and roughly a third of all knots up to nine crossings have this property (Flammini et al., 2004). (See also Figs. 3–9 in the supplementary material to Flammini and Stasiak, 2006.)

6.2. The size of protein knots

Unlike DNA, proteins are not shoelaces and have considerable bulk for a their length. This imposes limits on the minimal length required to form these more complex knots which can

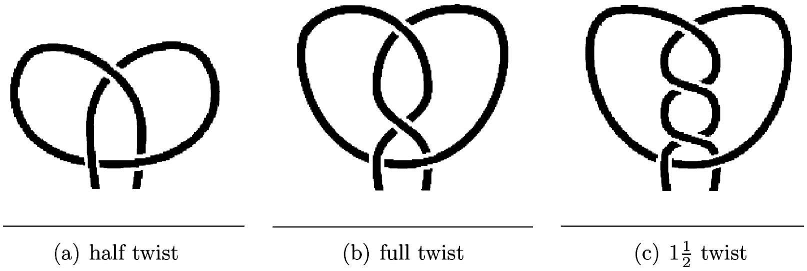


Fig. 4. A series of twist knots can all be constructed (or unknotted) with only one linking (unlinking) move. These differ only in the number of twists given to the ends before the link is made. For protein folding, this requires only one terminus to be threaded through one loop and all protein knots have this form.

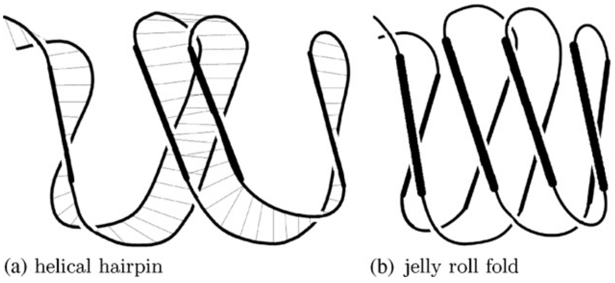


Fig. 5. Folding a double helix into a Jelly-roll. A long β-hairpin (a) can be rolled-up to create a Jelly-roll fold (b). The twist of the β-sheet can bring the loop at the mid-point of the hairpin close to the termini when these lie diagonally opposite on the same layer. (Top left and bottom right corners.)

be calculated using a homogeneous 'rubber' tube construction (Katritch et al., 1996). Most proteins fall far from these theoretical limits but a general trend of linear increasing length with knot complexity can be expected. Despite the limited data (one point for the pentknot), the theoretical line can be fitted to the minimal protein knotted core size giving the relationship of $60N - 140$ residues for the number of knot over-crossings (N). This is roughly 40 residues to make a trefoil, 140 to make a pentaknot and a projected 280 to make a protein knot with 7 crossings.

These estimate are based on the known knotted structures which all contain α-helices. The β-sheet is more extended than the α-helix and it would seem probable that larger knots could more easily be constructed in the all-β class of protein. The formation of twist-type knots requires the chain to be twisted as a double helix then locked by a loop penetration and the "Jelly-roll" topology in the all-β class would be particularly well suited to this type of knot formation. The double winding of the chain in this fold is topologically equivalent to a double helix and the twist of the β-sheet can bring the mid-point hairpin and either of the termini into close proximity (Fig. 5). With a typical Jelly-roll fold (**1lr5A**), the C-terminus can be threaded through the mid-point hairpin with remodelling of only five residues to create a pentaknot (Fig. 6). Taking a larger structure (**1tg7A**), the N-terminus of the domain lies near the midpoint hairpin and when threaded through the loop, creates a 6-crossing knot (6_1). Both these knots were created using far fewer residues than estimated above, with just over 100 for 5 crossings and 150 for 6 crossings.

7. The utility of protein knots

Do topological knots in proteins serve any biological function or are they just good for knotting? By analogy with the cysteine knots (formed by disulphide crosslinks) (Vitt et al., 2001), it might be suspected that topological knots will provide some added stability to the protein structure. Cysteine knots are always found in small extracellular domains where their lack of any significant hydrophobic core makes them vulnerable to variations in the more extreme extracellular environment. However, this combination of factors cannot be argued for the proteins with topological knots, which are all large intracellular enzymes. Some however, in particular the RNA methyltransferase, derive from thermophilic organisms in which added stability would clearly be useful.

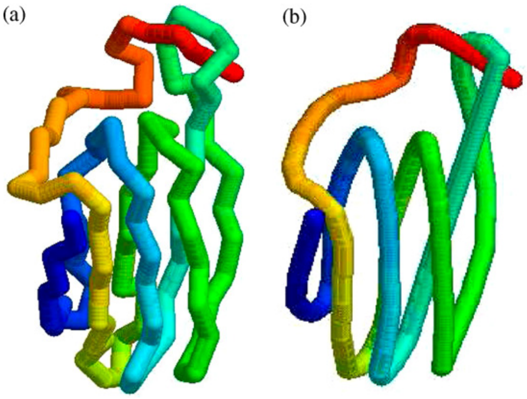
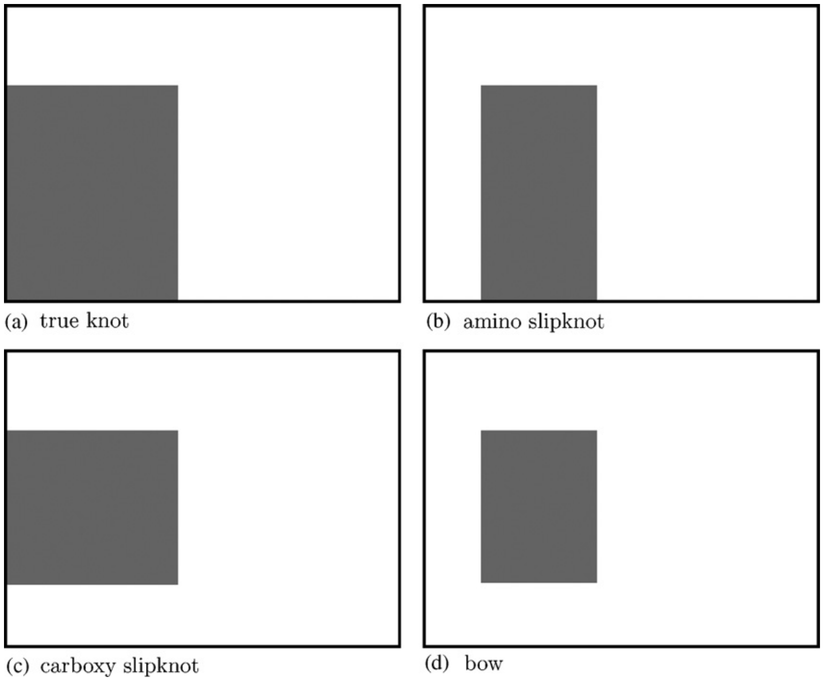


Fig. 6. An artificial knot in a Jelly-roll fold. The structure **1lr5A** (residues 40–148) has a Jelly-roll fold in which the C-terminus passes close to the midpoint loop (78–85). The five terminal residues (red) were redirected to pass through the loop (green) forming a pentaknot. (a) The α-carbon backbone (coloured N = blue to C = red) and (b) the smoothed chain. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

Given that the known proteins with topological knots are all enzymes, it may be important that the location of the knot is usually in the catalytic domain and in some cases, encompass the active site. However, it remains difficult to imagine what special geometric features could be provided by a knotted chain that would not be equally easy to support with an unknotted chain. Again the argument of additional stability can be used as the active site is a location where flexibility for substrate binding must be combined with stability for catalysis. However, some doubt may be cast on this proposal by the observation that the active site in the RNA methyltransferases lies at their dimer interface. The ubiquitin hydrolase discussed by Virnau and colleagues rescues proteins marked by ubiquitinilation for degradation in the proteosome by removing their ubiquitin label. The authors suggest that the knot prevents the protein from being drawn into the proteosome by the protein it is trying to save from destruction. This is an attractive but highly speculative proposal and not one that could explain the more widespread occurrence of knots in proteins.

8. Detecting knotted models

While knotted topologies are rare in real proteins, they can be much more common in models, occasionally created through loop entanglement when working from a native template (Tramantano et al., 2001) but more frequently they are found in models constructed using *de novo* or *ab initio* methods. This is especially true for methods using distance geometry (Aszódi and Taylor, 1996) where there are no kinetic constraints on the folding pathway. At the other extreme, chains constructed by accretion around a core contain no knots (Taylor, 2005). The *de novo* methods based on fragment assembly (*e.g.* the ROSETTA method; Bradley et al., 2005) lie somewhere between. The algorithm of Taylor (Taylor, 2000) was originally designed for fast



(a) true knot

(b) amino slipknot

(c) carboxy slipknot

(d) bow

Fig. 7. Knot plots of terminal deletions. In each plot, the effect of making terminal deletions on the knotted state are plotted. The *X*-axis measures the increasing numbers of deletions from the N-terminus and the *Y*-axis increasing number of residues removed from the C-terminus. (a) True knots always include the origin (the full protein), whereas slipknots require a portion to be removed from one terminus before they become true knots. (An equivalent construct is used by Yeates and King, 2007 to find slipknots.) This can occur either at the amino (b) or carboxy (c) termini. If a knot appears following deletions at both ends then it is a bow (c). No bows have yet been found in proteins.

knot detection in models (typically many thousands) and was also applied at runtime to test partially constructed models both in *de novo* (Taylor et al., 2003a) and in threading (Taylor et al., 2006). The published accounts of the algorithm, however, are more aimed at visualisation of knots in native structures.

Khatib et al. (2006) have described a 'new' method[7] specifically designed for the analysis of models that has been applied to collections of models and also to native structures. They find all the knotted proteins discussed above, including the ubiquitin hydrolase structure **1xd3** yet, strangely, they associate this protein (which contain the new pentaknot discussed above) with the degenerate left hand trefoil found in the homologue **1cmx** (also discussed above). Either the authors have missed the new knot on the assumption that it was the same as **1cmx** or their algorithm is sensitive to the order of triangle reduction and does not always retain the true knot topology. This might not be unexpected as it has been shown that the triangle removal algorithm is not robust (Millett et al., 2005) and this would be exacerbated where the triangles are large. Confusingly, the authors also report that they do not find four knots previously listed by Taylor, however, the structures to which they refer were knots formed only by hydrogen bond links (Taylor and Lin, 2003) and were never classed as topological knots.

As would be expected, the number of knots found in homology based models are low (less than 5%) but this rises to over 15% when the targets become more difficult and fold recognition methods are used.[8] Khatib et al. concentrated only on models generated on the basis of structural similarity and made no separation of template free (new fold) models. As discussed above, it is this distinction where the frequency of knots would give some insight into the *de novo* construction methods. By concentrating on homology models, the types of knots they found resulted from loop entanglement and could be classed according to whether the loop made a knot itself, with another loop or with the framework.

9. Fold complexity

9.1. Tangles slip-knots and bows

It had been observed that some unknotted protein chains take more steps to reach a straight line than others and that these

[7] In their paper, Khatib et al. contrast their algorithm to the algorithm of Taylor. However, they appear to have misread the previous work which contains identical steps to their triangle removal algorithm. The only difference is that the Taylor algorithm does not always remove a triangle on every cycle. (However, adding the condition that the smallest triangle must always be removed on each cycle makes little difference to speed.) Also, when determining the core of the knot by the Taylor method of the terminal deletions, all deletions are made before application of the knot testing algorithm so there can be no ambiguity in interpreting the result as suggested by Khatib et al.

[8] This statement is based on the CASP5 results for ROSETTA models in Khatib et al. (their Fig. 7a). The corresponding CASP6 graph shows much lower absolute numbers yet the authors state in the text that CASP6 still showed significant occurrence of knots. Due to poor annotation and description in the text, it is not clear if the CASP6 results were filtered somehow or what the circles are on these graphs.

often correspond to what would be called a tangle or a slip-knot in which a loop has been encircled by another portion of the chain (Taylor and Aszódi, 2005). Proper slip-knots and bows (of the type we use to tie-up shoes) are distinct from tangles and require that a true knot will be created if the ends of the knotted loops are pulled once through the encircling noose. In the algorithm of Taylor, this change in knotted state can be detected by making the series of all possible deletions from the amino and carboxy termini and a similar construct has been used by Yeates and King (2007) to find such knots. Such a series is usually made to determine the core location of a knot and when the number of crossings in the knot are plotted on a graph with axes corresponding to the size of the terminal deletions (a knot plot), then knotted regions appear clearly and any change in knot complexity can be visualised (Taylor, 2005). In such a knot plot, the knotted regions are rectangular which indicates that, to a first approximation, the two ends of the chain can be treated independently (Fig. 7).

True topological knots always occupy the corner region of the knot plot that includes the two undeleted termini but a slip knot would be detached from one axis and a bow would be a free floating region. On the assumption that the termini are independent, slip knots can be checked easily by making a series of amino then carboxy terminal deletions (checking just along the axes of the knot plot). A survey of a nonredundant PDB (Yeates and King, 2007) found several slip-knots and one of these, in an alkaline phosphatase (**1alkA**), was quite substantial. This forms part of an even more interesting story involving cysteine crosslinking (Todd Yates, personal communication). The larger slip-knots were right-handed trefoils and all were created from C-terminal deletions. Although this is a small sample, almost all the other slip-knots (held by just one or two residues) were formed by C-terminal deletions.

10. Unravelling folds

10.1. Simulated folding/unfolding

This idea of how easy it is to unravel a protein structure has been addressed most directly both experimentally and theoretically by taking the ends of the protein and pulling. In the experimental context this can be done using atomic force microscopy (AFM) while theoretically, an artificial extension force can be applied during a molecular dynamics (MD) simulation. Knotted proteins have been the target of these investigations for some time, concentrating on the original knot seen in carbonic anhydrase. Both approaches have been applied to the unfolding of carbonic anhydrase (Ohta et al., 2004) in which the stability of the fold is attributed, not to its knot (a barely threaded trefoil), but to a histidine chelated zinc ion.

The deeper trefoil in the methyltransferases has also been the subject of theoretical study (Wallin et al., 2007), not unfolding but folding under a Gō potential. This is a potential that derives from the contacts in the known structure and so is not unlike unfolding in reverse. Kinetically there is a distinction as it is easier to pull a thread out of the eye of a needle rather than thread it in the first place. This is effectively what is found in the simulations as additional artificial constraints had to be added to guide the path to the native structure. Their α-carbon representation cannot address the *cis*/*trans* proline isomerisation effect proposed as a potential factor that might influence folding, however, Wallin et al. do note that one of the regions involved in the knot is an α-helix in the structure but predicts strongly as a β-strand. This is the C-terminal α-helix which is the segment of chain that must thread through the loop to form the knot. Some delay in adopting the bulky helical native conformation may derive from this secondary structure ambiguity and the authors suggest that adopting a transient β-strand conformation may assist the threading process.

10.2. Topological frustration

The entanglement of a protein fold has been recently formalised into an algorithm to both measure what the authors call “topological frustration” and identify the ‘sticking’ points in the fold (Norcross and Yeates, 2006). This ingenious algorithm is based on representing the protein (α-carbon backbone) as a network or adjacent connections calculated by Delaunay tessellation (the dual construct of the more familiar Vorinii volumes). In this network, rings of connections can be found that encircle parts of the chain. For the chain to unravel, these rings must firstly be broken, imposing an order in which parts of the chain must be unravelled. The optimal order for this (causing minimal frustration) is calculated by a dynamic programming algorithm. The algorithm is also suitable for application to knotted topologies and to proteins containing crosslinks. For disulphide crosslinks the authors were able to compare their results with some experimental data and found good agreement in the order of disulphide bond formation. More generally, the algorithm might be extended to encompass ‘knots’ that are only defined by hydrogen-bonds, as discussed previously in relation to the SET domain (Taylor et al., 2003b).

By localising the points of frustration along the sequence, Norcross et al. were able to reexamine the old suggestion that proteins have a preference to fold more from their amino terminus (Phillips, 1966), with the assumption that this would occur co-translationally. If a segment of chain is encircled by parts all lying towards the carboxy terminus then this would suggest an N-terminal nucleation point. For segments lying close to the termini (within 10 residues) there appears to be a clear asymmetry for the N-terminus to be encircled more often than the C-terminus, up to a factor of two for the most terminal segments. Rather than having a simple explanation in co-translational folding, it may also be possible that the bias derives from the operation of mechanisms within the chaperone system.

10.3. Topological accessibility

A different approach has recently been used to shed some light on the N-terminal folding hypothesis. In an algorithm called “topological accessibility”, the ease with which protein structures can be created using only local contacts has been quantified

(Taylor, 2006). In this algorithm, the regeneration of the chain extends in both directions from a starting residue and the ease of success depends on where this point lies in the structure. For some proteins, the fold is accessible from almost all points along the chain while for others there is a restricted nucleation segment and a few proteins cannot be regenerated at all in this way. (These are called topologically inaccessible.)

For most proteins there is no overall bias for the nucleation region to lie towards either terminus, except for the βα class of protein where there is a very strong bias. Much, but not all, of this derives from the Rossman-like folds which typically bind nucleotides and while the bias might well be structural in nature, these proteins are also some of the oldest we know. Even if the folding of these proteins is now under the control of the chaperone system, it was suggested that the fold in these ancient proteins might be a relic from pre-chaperone days. It would be interesting to check if the Norcross et al. signal derived from the same class.

11. Fold complexity and structure prediction

The measures discussed above are related to the complexity of a protein fold. This is a difficult quantity to formalise but one that is important as it sheds light on the mechanism of protein folding (including knot formation) and would provide useful constraints for structure prediction. A simple measure proposed previously was the “contact order” (Plaxco et al., 1998) which measured the relative frequency of sequentially local and non-local contacts and showed good correspondence with observed folding rates. More recently, Rost has examined this relationship (Punta and Rost, 2005a).

The complexity of a fold is more easily investigated by considering just the strand order in β-sheets and has been a topic of analysis for many years (Richardson, 1977). Complex sheets pose a challenge to *de novo* and *ab initio* prediction methods and information from the analysis of β-sheet topology has been used as a filter in the ROSETTA method to improve selection of the correct topology (Ruczinski et al., 2002). More recently, this approach has been implemented in a more active way in the form of a guide tree to direct the construction of β-sheet by forcing the pairings of β-strands (Bradley and Baker, 2006).

The prediction of long range interactions in proteins is the key to tertiary structure prediction but remains a very difficult problem. Using a variety of information sources integrated through an artificial neural net, Punta and Rost (2005b) describe one of the best recent attempts to tackle this problem. Similarly, Kinjo et al. (2005) have attempted to predict the contact number directly from sequence data based on a one-dimensional encoding (Kinjo and Nishikawa, 2005). The remaining level of error in all these studies, however, remains large and by itself would not have sufficient accuracy to determine a tertiary structure but should prove useful in combination with other constraints. Concentrating on the more constrained environment of the β-sheet may also help in this problem and some useful contact preferences have been noted in an extensive recent analysis (Fooks et al., 2006).

The required constraints for structure prediction may simply be that we need more accurate secondary structure prediction and Rose and colleagues (Fleming et al., 2007) have demonstrated that given well defined secondary structure definitions, the generation of compact folds gives a high chance of including the native. Unfortunately, this is likely to remain an academic exercise as the level of accuracy required is unlikely to be attained and the method also makes use of conformational preferences in the loop regions that are probably even more difficult to predict than secondary structure. Indeed, it has also been shown recently that for non-local interactions, where the most accurate predictions for secondary structure are needed, is exactly where they are least accurate (Kihara, 2005). The implications for structure prediction are already well known: that the definition of secondary and tertiary structure must interact (Meiler and Baker, 2003) and when this is done, good results can be obtained at least for proteins under 100 residues in length (Bradley et al., 2005; Yang et al., 2007).

12. Conclusions

The occurrence of the three simplest types of knot was reviewed and it was seen that there is yet too few examples to say whether there is a preferred handed to these. More generally, it was inferred that protein knots should be more likely to adopt topologies that require only one loop threading event. This corresponds to the form seen in a new pentaknot and a prediction on this basis is that the knots adopted by proteins should have a loop that when unlinked will undo the knot completely and that this should be the location of the protein termini. There is a series of twist-knots that have this property and several others. Some projections were made on the expected size of more complex knots and it was suggested that these might be formed more easily in all-β type proteins, especially those with a Jelly-roll fold.

Some slip-knots were identified in proteins but only a few were deeply knotted (by more than a few residues). Nevertheless, there seemed to be some trend for these to be formed in the carboxy-terminal regions of the protein. This may be associated with two recent studies that provided some support for amino-terminal folding as it might be expected that a greater number of local interactions at the amino terminus would restrict topological complexity in this region.

Topological filters are becoming increasingly used in protein fold prediction and even modelling. While more common folds may be better predicted (or more frequently sampled) using this information, it may mean that rare complex or even knotted folds will never be predicted. Current methods that attempt to predict long-range contacts (or complexity) directly from sequence are not yet able to provide information of sufficient specificity but a combination of these restraints along with topological constraints may prove better than either approach alone.

This review has attempted to gather some loose ends in the area of protein topology and weave (or braid) from them some general principles. In this, the focus has remained at the level of the folded protein chain but there remain many other topological aspects at the higher level, involving crosslinking and catenation that remain to be explored.

Acknowledgements

Andrzej Stasiak, Todd Yeates and Eugene Shakhnovich are thanked for their comments on the manuscript.

References

Adams, C.C., 1994. The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots. W.H. Freeman, New York.

Ahn, H.J., Eom, S.J., Yoon, H.J., Lee, B.I., Cho, H., Suh, S.W., 2003. Crystal structure of class I acetohydroxy acid isomerase from pseudomonas aeruginosa. J. Mol. Biol. 328, 505–515.

Andreeva, A., Murzin, A.G., 2006. Evolution of protein fold in the presence of functional constraints. Curr. Opin. Struct. Biol. 16, 399–408.

Aszódi, A., Taylor, W.R., 1996. Homology modelling by distance geometry. Fold. Des. 1, 325–334.

Biou, V., Dumas, R., Cohen-Addad, C., Douce, R., Job, D., Pebay-Peyroula, E., 1997. The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transtition state analog at 1.65 Å resolution. EMBO J. 16, 3405–3415.

Bradley, P., Baker, D., 2006. Improved beta-protein structure prediction by multilevel optimisation of nonlocal strand pairings and local backbone conformation. Prot. Struct. Funct. Bioinf. 65, 922–929.

Bradley, P., Misura, K.M.S., Baker, D., 2005. Toward high-resolution de novo structure prediction for small proteins. Science 309, 1868–1871.

Cuff, M.E., Mussar, K.E., Li, H., Moy, S., Joachimiak, A., in press. The structure of a putative RNA methyltransferase of the TrmH family from porphyromonas gingivalis. Midwest Center for Structural Genomics (MCSG). PDB code:2I6D, http://www.rcsb.org/pdb/explore.do?structureId=2I6D.

Flammini, A., Maritan, A., Stasiak, A., 2004. Simulations of action of DNA topoisomerases to investigate boundaries and shapes of spaces of knots. Bio-phys. J. 87, 2968–2975.

Flammini, A., Stasiak, A., 2006. Natural classification of knots. Proc. Roy. Soc. A 463, 569–582.

Fleming, P.J., Gong, H., Rose, G.D., 2007. Secondary structure determines protein topology. Prot. Sci. 15, 1829–1834.

Fooks, H.M., Martin, A.C.R., Woolfson, D.N., Sessions, R.B., Hutchinson, E.G., 2006. Amino acid pairing preferences in parallel β-sheets in proteins. J. Mol. Biol. 356, 32–44.

Gloss, L.M., 2007. Tying the knot that binds. Structure 15, 2–4.

Katritch, V., Bednar, J., Michoud, D., Scharein, R.G., Dubochet, J., Stasiak, A., 1996. Geometry and physics of knots. Nature 384, 142–145.

Khatib, F., Weirauch, M.T., Rohl, C.A., 2006. Rapid knot detection and application to protein structure prediction. Bioinformatics 22, e252–e259.

Kihara, D., 2005. The effect of long-range interactions on the secondary structure formation of proteins. Prot. Sci. 14, 1955–1963.

Kinjo, A.R., Horimoto, K., Nishikawa, K., 2005. Predicting absolute contact numbers of native protein structure from amino acid sequence. Prot. Struct. Funct. Bioinf. 58, 158–165.

Kinjo, A.R., Nishikawa, K., 2005. recoverable one-dimensional encoding of three-dimensional protein structures. Bioinformatics 21, 2167–2170.

Lim, K., Zhang, H., Tempcyzk, 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. Prot.: Struct., Funct., Genet. 51, 56–67.

Mallam, A.L., Jackson, S.E., 2005. Folding studies on a knotted protein. J. Mol. Biol. 346, 1409–1421.

Mallam, A.L., Jackson, S.E., 2006. Probing natures's knots: the folding pathway of a knotted homodimeric protein. J. Mol. Biol. 359, 1420–1436.

Mallam, A.L., Jackson, S.E., 2007a. The dimerisation of an α/β-knotted protein is essential for structure and function. Structure 15, 111–122.

Mallam, A.L., Jackson, S.E., 2007b. A comparison of the folding of two knotted proteins: YbeA and YibK. J. Mol. Biol. 366, 650–665.

Mansfield, M.L., 1994. Are there knots in proteins. Nat. Struct. Biol. 1, 213–214.

Mansfield, M.L., 1997. Fit to be tied. Nat. Struct. Biol. 4, 116–117 (News and Views).

Meiler, J., Baker, D., 2003. Coupled prediction of protein secondary structure and tertiary structure. Proc. Natl. Acad. Sci. U.S.A. 100, 12105–12110.

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.

Millett, K., Dobay, A., Stasiak, A., 2005. Linear random knots and their scaling behaviour. Macromolecules 38, 601–606.

Misaghi, S., Galardy, P.J., Meester, W.J.N., Ovaa, H., Ploegh, H.L., Gaudet, R., 2005. Structure of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate. J. Biol. Chem. 280, 1512–1520.

Norcross, T.S., Yeates, T.O., 2006. A framework for describing topological frustration in models of protein folding. J. Mol. Biol. 362, 605–621.

Nureki, O., Shirouzu, M., Hashimoto, K., Ishitani, R., Terada, T., Tamakoshi, M., Oshima, T., Chijimatsu, M., Takio, K., Vassylyev, D.G., Shibata, T., Inoue, Y., Kuramitsu, S., Yokoyama, S., 2002. An enzyme with a deep trefoil knot for the active-site architecture. Acta Cryst., Sect. D 58, 1129–1139.

Ohta, S., Alam, M.T., Arakawa, H., Ikai, A., 2004. Origin of mechanical strength of bovine canbonic anhydrase studied by molecular simulation. Biophys. J. 87, 4007–4020.

Phillips, D.C., 1966. The three-dimensional structure of an enzyme. Sci. Am. 215, 78–90.

Plaxco, K.W., Simons, K.T., Baker, D., 1998. Contact order, transition state placement and the refolding rates of single domain proteins. J. Mol. Biol. 277, 985–994.

Punta, M., Rost, B., 2005a. Protein folding rates estimated from contact predictions. J. Mol. Biol. 348, 507–512.

Punta, M., Rost, B., 2005b. PROFcon: novel prediction of long-range contacts. Bioinformatics 21, 2960–2968.

Richardson, J.S., 1977. β-sheet topology and the relatedness of proteins. Nature 268, 495–500.

Ruczinski, I., Kooperberg, C., Bonneau, R., Baker, D., 2002. Distributions of beta sheets in proteins with application to structure prediction. Prot. Struct. Funct. Genet. 48, 85–97.

Taylor, W.R., 2000. A deeply knotted protein and how it might fold. Nature 406, 916–919.

Taylor, W.R., 2002. Protein structure comparison using bipartite graph matching. Mol. Cell. Proteom. 1, 334–339.

Taylor, W.R., 2005. Protein folds, knots and tangles. In: Calvo, J., Millett, K., Rawdon, E., Stasiak, A. (Eds.), Physical and Numerical Models in Knot Theory, Series on Knots and Everything, vol. 36. World Scientific, Singapore, ISBN 981-256-187-0, pp. 171–202 (chapter 10).

Taylor, W.R., 2006. Topological accessibility shows a distinct asymmetry in the folds of βα proteins. FEBS Lett. 580, 5263–5267.

Taylor, W.R., Aszódi, A., 2005. Protein Geometry, Classification, Topology and Symmetry. Institute of Physics. Currently published by CRC Press.

Taylor, W.R., Lin, K., 2003. A tangled problem. Nature 421, 25 (concept).

Taylor, W.R., Lin, K., Klose, D., Fraternali, F., Jonassen, I., 2006. Dynamic domain threading. Prot., Struct. Funct. Bioinfo. 64, 601–614.

Taylor, W.R., Munro, R.E.J., Petersen, K., Bywater, R.P., 2003a. ab initio modelling of the N-terminal domain of the secretin receptors. Comp. Biol. Chem. 27, 103–114.

Taylor, W.R., Xiao, B., Gamblin, S.J., Lin, K., 2003b. A knot or not a knot? SETting the record ‘straight’ on proteins. Comp. Biol. Chem. 27, 11–15.

Tramantano, A., Leplae, L., Morea, V., 2001. Analysis and assessment of comparative modelling predictions in CASP4. Prot.: Struct. Funct. Genet. 45 (Sup. 5), 22–38.

Vidgren, J., Svensson, L., Liljas, A., 1994. Crystal structure of catechol *O*-methyltransferase. Nature 368, 354–358.

Virnau, P., Mirny, L.A., Kardar, M., 2006. Intricate knots in proteins: function and evolution. PLoS, Comput. Biol. 2, 1074–1079.

Vitt, U.A., Hsu, Y., Hsueh, A.J.W., 2001. Evolution and classification of cystein knot-containing hormones and related extracellular signaling moleculas. Mol. Endocrinol. 15, 681–694.

Wallin, S., Zeldovich, K.B., Shakhnovich, E.I., 2007. The folding mechanics of a knotted protein. J. Mol. Biol. 368, 884–893.

Yang, J.S., Chen, W.W., Skolnick, J., Shakhnovich, E.I., 2007. All-atom ab initio folding of a diverse set of proteins. Structure 15, 53–63.

Zarembinski, T.I., Kim, Y., Peterson, K., Christendat, D., Dharamsi, A., Arrowsmith, C.H., Edwards, A.M., Joachimiak, A., 2003. Deep trefoil knot implicated in RNA binding found in an archaebacterial protein. Prot.: Struct., Funct., Genet. 50, 177–183, PDB code: 1K3R.