Published in final edited form as:

Curr Opin Chem Biol. 2007 December ; 11(6): 595–603.

Knotted and topologically complex proteins as models for studying folding and stability

Todd O. Yeates^{1,2}, Todd S. Norcross¹, and Neil P. King¹

¹ *UCLA Dept. of Chemistry and Biochemistry, Los Angeles, CA*

² *UCLA-DOE Institute of Genomics and Proteomics, Los Angeles, CA*

SUMMARY

Among proteins of known three dimensional structure, only a few possess complex topological features such as knotted or interlinked (catenated) protein backbones. Such unusual proteins offer potentially unique insights into folding pathways and stabilization mechanisms. They also present special challenges for both theorists and computational scientists interested in understanding and predicting protein folding behavior. Here we review complex topological features in proteins with a focus on recent progress on the identification and characterization of knotted and interlinked protein systems. Also, an approach is described for designing an expanded set of knotted proteins.

Keywords

Protein knots; protein links; protein folding; protein stability; protein topology

INTRODUCTION

A central goal in biochemistry is to understand the mechanisms by which proteins reliably fold into and maintain their native three-dimensional structures. A number of recent conceptual advances have focused on how the native three-dimensional structure or ‘fold’ of a protein affects its folding properties. For instance, the recently developed concept of ‘contact order’ explicitly defines a relationship between the geometries of the native structures of proteins and their rates of folding [1–3]. Energy landscape theories have also provided an important framework [4–10]. Depending in part on the geometric properties of its fold, a protein’s energy landscape may contain multiple local minima and dead end pathways, leading to frustration during folding [11–14]. In the last decade, theoretical, experimental, and computational investigations have focused mainly on proteins having relatively simple folds. Small proteins with simple folding kinetics have provided tractable systems for analysis and a valuable testing ground for theories of protein folding [15–19].

Various lines of research have begun to clarify the folding mechanisms of relatively simple proteins, while at the same time efforts in structural biology have continued to reveal novel protein structures with surprisingly complex folds. Rare proteins whose backbones adopt knotted configurations offer particularly interesting challenges for folding theories. For

Corresponding author contact information: Todd O. Yeates, UCLA Dept. Chem. and Biochem., 611 Charles Young Dr. East, Los Angeles, CA 90095-1569, (tel 310-206-4866) (yeates@mbi.ucla.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

example, according to current thinking the folding energy landscapes of natural proteins (at least those that fold easily) are funneled in such a way that the low-energy native configuration can be approached smoothly from a broad basin of attraction [4,5,10]. Deeply knotted proteins present an intriguing counter-scenario. To the extent that some degree of threading is required to generate a deep knot, the native configuration must be approached by traversal of more restricted valleys through the folding landscape. In turn, departure from the folded configuration would entail traversal of the same entropically constricted valleys, suggesting how knotting might provide kinetic stabilization. Similar issues arise in another kind of rare situation wherein separate protein chains are entwined to the point of being topologically linked together. Here again, the threading of protein chains through constricted spaces has important implications for folding and unfolding mechanisms.

Topologically complex proteins offer potentially valuable model systems for conducting theoretical, computational, and experimental studies of protein folding. Such proteins have only begun to fall under investigation [20–26]. In this review we summarize recent observations and experiments on knotted and interlinked proteins.

FORMS OF TOPOLOGICAL COMPLEXITY

The term ‘topology’ is sometimes used loosely in structural biology. Here we restrict our use to the stricter mathematical notion of whether or not curves in space are knotted or linked together. This still admits a wide range of topologically interesting features in proteins, particularly if non-bonded interactions are included as parts of the curves to be considered. Here we touch on a variety of topological features in proteins before focusing on a few special types (Figure 1). Situations where non-covalent interactions have led to interesting topological features include (i) interlocked, oligomeric rings of protein subunits (Figure 1a) [27], and (ii) so-called ‘topological folding barriers’ (Figure 1b) [14•], in which a group of non-covalently connected residues in a protein form a ring in the native structure through which another segment of the protein must be threaded. Because the threaded segment would have more difficulty coming into place after the ring, the situation has implications for which pathways to the folded state are most accessible.

Other cases of topological complexity arise from covalent bonding. Such cases are of interest in part because of the strength and effective irreversibility of the interactions involved. The ‘cystine knot’ superfamily of growth factors and toxins provides a well-reviewed example [28,29]. In these proteins, a disulfide bond between two beta strands passes through a ring formed by two other beta strands and the two disulfide bonds that connect them (Figure 1c). More recently, lariat-like pseudorotaxane topologies have been observed in the structures of a class of short antimicrobial polypeptides [30,31]. Formation of an isopeptide bond in those peptides results in cyclization of the N-terminal portion, through which the C-terminal portion is threaded to form the pseudorotaxane. A property these cases share is that the topological features are present mainly by virtue of additional bonds connecting different parts of the protein chain. Such structures present relatively simple folding puzzles; the topological complexity arising from the additional covalent bonds can be introduced as a final step, after the backbone folds.

Special situations arise when the protein backbone itself embodies some kind of topological complexity, such as knotting or interlinking. These cases are of particular interest in the present paper because, as noted above, questions arise immediately about how such proteins can fold efficiently. In analyzing both knotting and linking in proteins, some liberty is taken in including cases where the topological feature of interest relies to a degree on the presence of additional bonds or connections in the protein, as long as the key features are evident in the protein backbone considered in isolation.

IDENTIFYING KNOTTING AND INTERLINKING IN PROTEINS

The problem of identifying knots in proteins is a challenging one. Making conclusive identifications in large structures by visual inspection can be extremely difficult. In fact, the first deeply knotted protein was identified computationally by Taylor [32] some time after the structure was first reported [33]. There are also cases where a knot was reported [34] which does not exist [35,36], or where the type of knot differed from that reported [37,38•]. To identify the presence of knots in large structures, or in the large database of known structures, computational methods are called for. Two kinds of computational approaches have been developed, those that are effectively mechanical [32,39•], and those based on knot theory [26••,38•,40,41•]. In the mechanical approaches, the protein backbone is repeatedly simplified or smoothed under the constraint that the backbone is not permitted to cross through itself. The effect is to mechanically straighten the chain. If during the process the backbone converges to a straight line, then the original protein chain is determined to be unknotted. If a straight line cannot be obtained, the protein is judged to be knotted. Such mechanical methods are rapid, but suffer from two potential drawbacks. First, entanglement can occur even in an unknotted chain – a situation understandable from everyday experience – leading to potential false positives. Second, there is considerable interest in different kinds of knots that might be formed; mechanical approaches do not offer any information about what kind of knot is present. Methods based on knot theory address those problems, although at the expense of algorithmic complexity. In the language of knot theory, methods have been reported for classifying knots in proteins according to their Alexander polynomials [38•,41•], their Vassiliev invariants [41•], or their Jones' polynomials [26••].

With protein knots, we must also deal with the mathematical problem that the protein backbone is in general an open curve, while knots are technically defined only for closed curves. In practice, the ends of the protein chain are projected away from the protein and joined externally before deciding mathematically whether a knot is present [26••,38•,40,41•]. In general, this procedure does not create problems, particularly since the protein termini are usually at or near the surface of the protein. However, some important qualitative features of protein knots are clarified by looking more carefully at issues concerning the termini. It has long been recognized that many spurious or 'incipient' knots can be seen in proteins, e.g. due to the very end of a protein protruding slightly through a loop [40]. These are viewed as relatively insignificant because the knot vanishes when only a few residues are omitted from the end; the obstacle to folding here is judged to be minor. This leads to the concept of the depth of a knot, which is obtained by considering how many residues (e.g. the smaller of the values from the two termini) need to be omitted before the knot is eliminated [26••,32,38•]. A related idea is the knot tightness, which relates to the smallest substructure (allowing truncation at both termini) that retains the knot.

The key features of a knotted protein can be captured using a protein 'knot plot', which encodes the presence of knots, and their types, across the possible substructures within a complete three-dimensional protein structure [26••,42,43•] (Figure 2). As seen in Figure 2a, the RNA methyltransferases of the α/β -knot superfamily contain a particularly deep knot of the right-handed trefoil type (a 'three-crossing knot'). In the structure reported by Nureki *et al.*, 41 residues can be deleted from the C-terminus without eliminating the knot (Figure 1e) [44]. The structure of acetohydroxy acid isomeroreductase is an example of a knot (in this case a figure eight, or four-crossing knot) that is significant in depth, but looser than the knot in the α/β -knot superfamily (Figure 2b) [32]. The smallest substructure from the isomeroreductase that retains a knot is about 180 residues long, compared to 44 for the α/β -knot superfamily. The most complex knot so far (i.e. a knot of 5-crossings) has been identified in ubiquitin hydrolase UCH-L3 [38•], but the knot is very shallow (Figure 2d). Finally, an interesting variation arises when one considers the possibility of a protein chain that is not knotted when examined in its complete

form, but which becomes knotted when one (or both) termini are deleted. Such a structure is referred to as a 'slipknot'. Slipknots occur when the path of some part of the chain forms a knot, which is then effectively undone when the terminus doubles back on itself, like a tied shoelace. Because slipknots do not reveal themselves during computational examinations of intact protein structures, they had not been detected by routine applications of programs used to find knots in the protein database. By looking specifically for slipknots, the first deep slipknot was discovered by King *et al.* in alkaline phosphatase [26••] (Figures 1d,2c). Another complex slipknot was identified in a transmembrane protein, LeuT_{Aa} [26••]. A list of proteins containing knots or slipknots is given in Table 1, with a representative from each known family.

In addition to the knotted structures that have been found, a few structures have been observed where two or more protein backbones are interlinked topologically; to achieve true linkage, one additional bond within each protein is required to form closed chains, as shown in Figure 1f. The particular examples observed so far have been identified by visual inspection [23••, 45,46], following a prediction from biochemical studies in one case [47]. From a computational standpoint, determining whether or not two (closed) curves are interlinked is a relatively simple problem, as noted by Connolly *et al.* in the context of protein chains [48]. To our knowledge, no recent systematic search has been made for proteins that are interlinked (or could be interlinked) by the presence of an intramolecular disulfide bond. Such an investigation might identify new cases of interest.

IMPLICATIONS OF KNOTTED AND LINKED SYSTEMS FOR PROTEIN FOLDING AND STABILITY

Folding studies in knotted and linked proteins

The first studies on the folding of knotted proteins were conducted recently on the α/β -knot superfamily of methyltransferases [20–22]. Numerous structures from this large family of dimeric bacterial enzymes have revealed a conserved, deep trefoil knot comprised of residues that contribute to both the dimerization interface and the S-adenosyl methionine cofactor binding site [49]. Mallam and Jackson initiated investigations on this knotted system by first characterizing the equilibrium unfolding of one member of the family, the YibK protein from *Haemophilus influenzae* [20]. This study established that unfolding of the knotted protein was reversible *in vitro* without molecular chaperones, and suggested the existence of a partially unfolded monomeric intermediate. The equilibrium behavior of the protein was quite similar to that of other small, dimeric proteins. A subsequent characterization of the unfolding and refolding kinetics of the same protein led to the proposal of a folding pathway involving two distinct monomeric intermediates (arising from proline isomerization in the unfolded protein) converging upon a third monomeric intermediate, which then slowly converts to the native dimer in a rate-limiting dimerization step [21••]. Characterization of another member of the α/β -knot superfamily, the YbeA protein from *Escherichia coli*, revealed a similar equilibrium unfolding mechanism and a similar folding pathway involving a stable monomeric intermediate and a slow dimerization step [22••]. Given the low sequence identity between the two enzymes (19%), it seems likely that these shared behaviors are characteristic of other members of the α/β -knot superfamily. Despite those informative investigations, however, questions about how and when the knots form during protein folding remain unanswered by the experiments performed so far.

Those questions were recently addressed by Shakhnovich and coworkers using molecular dynamics simulations of the folding of YibK [24••]. The central observation of their work is that specific, nonnative interactions (an extension of [50]) are required for reliable folding to the native, knotted state, while an exclusively native-centric energy function [51] fails to result in successful folding. This intuitively appealing result offers a plausible explanation for the

mechanics of knot formation in this system, along with a hypothesis for experimental testing. The study also addresses the timing of knot formation by obtaining the distribution of the fraction of native contacts present when the knot is first formed, calculated over a large number of folding trajectories. The observed bimodal distribution shows that, at least according to simulation, there are two pathways by which the knot is formed, one occurring in the early and the other in the late stages of folding. Also of note is the observation that during some folding trajectories, the knot is actually formed by threading the C-terminal portion of the protein through the knotting loop in a hairpin-like conformation, transiently producing a slipknot before the final residue of the protein is threaded through to form the mature knot. In light of the recent discovery of slipknots in proteins [26••], this may hint at a common folding mechanism for both knots and slipknots.

The folding pathways of linked proteins have not been studied yet in any of the natural systems that have been identified, but one synthetic system has been explored. Blankenship and Dawson recently engineered the small p53 tetramerization (p53tet) domain in order to generate a topologically linked dimer [52]. To investigate the process of threading one polypeptide through another [25••], they mixed a population of p53tet that had been cyclized via native chemical ligation with a population of linear p53tet protein under denaturing conditions. When the denaturant was diluted out, the linear molecules threaded through the cyclized molecules to form the native-like structure. Fitting kinetic parameters to the data revealed that the threading rate, although slower than the folding of wild-type p53tet, was within a biologically relevant range, while the unthreading rate was unusually slow. This result, like the results of Mallam and Jackson discussed above, demonstrate that threading events during protein folding may be exceptional cases, but they are not forbidden. It also suggests that topological complexity may result in strong kinetic stabilization of the folded state.

Stability studies in knotted and linked proteins

The observation of knotted regions participating in the active or binding sites of various enzymes [37,44,49] has prompted speculation that such knots may confer stability or rigidity to those regions, thereby influencing the catalytic properties of the enzymes [37,53]. Similarly, a hypothesis predicting a functional role for the complex five-crossing knot in human ubiquitin hydrolase is attractive, yet remains to be tested [38•]. In a recent paper reporting the discovery of a deep slipknot in alkaline phosphatase, engineered disulfide bonds were used to probe the contribution that the slipknot makes to the unusual stability of that enzyme [26••]. Although not definitive, the results were consistent with the slipknot playing a stabilizing role. The nascent area of research on knotted proteins will require new experimental approaches in order to provide conclusive answers about the roles of knots in proteins.

Interlinked or catenated proteins, on the other hand, have already provided clear evidence for stabilization. The stability afforded by topological linking appears to derive mainly from a reduction in the entropy of the unfolded state, owing to the inability of the protein chains to fully unfold and dissociate. This effect seems to be more pronounced in topologically linked proteins than in proteins containing simple intermolecular disulfide bond cross-links [52,54]. Four topologically linked protein systems have been characterized to date. The mature capsid of the bacteriophage HK97 contains an isopeptide bond between subunits, which results in a topologically linked network reminiscent of chain mail [45]. The viral capsid is unusually thin, and the topological linking has been found to contribute to the maintenance of capsid stability [47] and infectivity [45]. The second characterization of a linked protein system was the engineered p53 tetramerization domain discussed above [52]. The stability studies were conducted on a system where both chains of the dimeric construct had been cyclized to form a linked structure whose chains were inseparable. The increase in the stability of the p53tet dimer due to catenation was dramatic, raising the melting temperature by 59 °C and the

midpoint of guanidine hydrochloride denaturation by 4.5 M. The final two linked systems are both interlinked dimers effected by natural intramolecular disulfide bonds that cyclize intertwined protein chains [23••,46]. In the more recently discovered case, citrate synthase from the hyperthermophilic archaeon *Pyrobaculum aerophilum*, an engineered mutant lacking the disulfide bond (and therefore lacking covalently linked topology) was shown to have reduced stability compared to the wild-type enzyme [23••].

PROSPECTS FOR DESIGN

Analyses of folding and stability in knotted proteins have thus far suffered from a lack of unknotted controls. In order to pinpoint the effects of knotting, it would be desirable to compare a knotted protein to a control protein having a similar core structure, but lacking the knot. In their analysis of the knotted RNA methyltransferase YibK, Lim *et al.* noted that the knot could be resolved (i.e. removed) by altering the connectivity of the protein backbone at two points [49]. As diagrammed in Figure 3, this approach could be generally applied in either direction to convert knotted proteins into unknotted versions, or vice-versa. The operation required at the sequence level can be likened to two DNA recombination events, with each occurring where two loops of the protein come into proximity. The result is the swapping of two segments of the protein sequence. Only certain choices for the recombination points lead to interconversion of knotted and unknotted topologies, and not all proteins may be suitable subjects for topological interconversion. Nonetheless, a wide variety of corresponding knotted and unknotted protein pairs could be generated. Such pairs of proteins could be valuable in both experimental and computational studies. In addition, if an increase in stability frequently accompanies knotting, then synthetic knotting could become a new method for engineering novel proteins with enhanced stabilities.

Acknowledgements

The authors thank Eugene Shakhnovich and Phil Dawson for critical readings of the manuscript. This work was supported by NIH Grant GM081652.

References

1. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 1998;277:985–994. [PubMed: 9545386]
2. Miller EJ, Fischer KF, Marqusee S. Experimental evaluation of topological parameters determining protein-folding rates. *Proc Natl Acad Sci U S A* 2002;99:10359–10363. [PubMed: 12149462]
3. Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, Finkelstein AV. Contact order revisited: influence of protein size on the folding rate. *Protein Sci* 2003;12:2057–2062. [PubMed: 12931003]
4. Leopold PE, Montal M, Onuchic JN. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci U S A* 1992;89:8721–8725. [PubMed: 1528885]
5. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 1995;21:167–195. [PubMed: 7784423]
6. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 1997;48:545–600. [PubMed: 9348663]
7. Chan HS, Dill KA. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins* 1998;30:2–33. [PubMed: 9443337]
8. Onuchic JN, Nymeyer H, Garcia AE, Chahine J, Socci ND. The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv Protein Chem* 2000;53:87–152. [PubMed: 10751944]
9. Plotkin SS, Onuchic JN. Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q Rev Biophys* 2002;35:111–167. [PubMed: 12197302]
10. Wolynes PG. Recent successes of the energy landscape theory of protein folding and function. *Q Rev Biophys* 2005;38:405–410. [PubMed: 16934172]

11. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* 1987;84:7524–7528. [PubMed: 3478708]
12. Shea JE, Onuchic JN, Brooks CL 3rd. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc Natl Acad Sci U S A* 1999;96:12512–12517. [PubMed: 10535953]
13. Thirumalai D, Klimov DK. Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr Opin Struct Biol* 1999;9:197–207. [PubMed: 10322218]
- 14. Norcross TS, Yeates TO. A framework for describing topological frustration in models of protein folding. *J Mol Biol* 2006;362:605–621. [PubMed: 16930616]The authors use computational geometry and dynamic programming to investigate how topology restricts protein folding. The paper provides evidence that proteins favor folding around the N terminus, consistent with the idea that proteins tend to fold co-translationally
15. Kim PS, Baldwin RL. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem* 1982;51:459–489. [PubMed: 6287919]
16. Jackson SE, Fersht AR. Folding of chymotrypsin inhibitor 2.1 Evidence for a two-state transition. *Biochemistry* 1991;30:10428–10435. [PubMed: 1931967]
17. Matouschek A, Kellis JT Jr, Serrano L, Fersht AR. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 1989;340:122–126. [PubMed: 2739734]
18. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 1998;282:740–744. [PubMed: 9784131]
19. Shimada J, Shakhnovich EI. The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc Natl Acad Sci U S A* 2002;99:11175–11180. [PubMed: 12165568]
20. Mallam AL, Jackson SE. Folding studies on a knotted protein. *J Mol Biol* 2005;346:1409–1421. [PubMed: 15713490]
- 21. Mallam AL, Jackson SE. Probing nature's knots: the folding pathway of a knotted homodimeric protein. *J Mol Biol* 2006;359:1420–1436. [PubMed: 16787779]In this study, a thorough investigation of the folding mechanism of the knotted methyltransferase YibK was carried out using a number of techniques. [Urea]-jump and pH-jump experiments at various protein concentrations were used to assign the slowest phase of unfolding/refolding to dissociation/association of the subunits of the dimer. Interrupted refolding and unfolding experiments probed the nature of the folding intermediates. A folding model consistent with all kinetic data was proposed
- 22. Mallam AL, Jackson SE. A comparison of the folding of two knotted proteins: YbeA and YibK. *J Mol Biol* 2007;366:650–665. [PubMed: 17169371]To investigate whether different knotted proteins exhibit similar folding behavior, the authors characterized the folding pathway of YbeA, a member of the α/β -knot superfamily from *E. coli*, and compared the results to their previous studies on YibK. Equilibrium denaturation and kinetic single- and double-jump experiments provided data consistent with a folding model similar in many respects to that proposed for YibK
- 23. Boutz DR, Cascio D, Whitelegge J, Perry LJ, Yeates TO. Discovery of a thermophilic protein complex stabilized by topologically interlinked chains. *J Mol Biol* 2007;368:1332–1344. [PubMed: 17395198]A proteomics approach was taken to identify disulfide-bonded proteins and protein complexes in the hyperthermophilic archaeon *Pyrobaculum aerophilum*. One of the disulfide-bonded complexes identified, the homodimeric citrate synthase, was crystallized and the structure revealed intramolecular disulfide bonds which topologically linked the two chains of the dimer. Mutation of the cysteine residues involved in the disulfide bonds to serine resulted in a significant decrease in the stability of the enzyme
- 24. Wallin S, Zeldovich KB, Shakhnovich EI. The folding mechanics of a knotted protein. *J Mol Biol* 2007;368:884–893. [PubMed: 17368671]Molecular dynamics simulations of the folding of the knotted protein YibK were carried out to specifically address the mechanics and timing of knot formation during folding. The authors found that specific, nonnative interactions were necessary for successful folding. A bioinformatics analysis of protein sequences related to YibK suggested possible candidates for the nonnative interactions necessary to drive folding
- 25. Blankenship JW, Dawson PE. Threading a peptide through a peptide: protein loops, rotaxanes, and knots. *Protein Sci* 2007;16:1249–1256. [PubMed: 17567748]The engineered dimeric p53tet system was used to investigate the process of threading during protein folding. Using fluorescence quenching as a specific probe for the threading process, the authors showed that threading is an

efficient process. Moreover, a very slow unthreading rate was observed, which has implications for the role of topological complexity in protein stabilization

- 26. King NP, Yeates EO, Yeates TO. Identification of rare slipknots in proteins and their implications for stability and folding. *J Mol Biol*. 2007A systematic survey for slipknotted topologies in proteins was performed by calculating the knottedness of partial protein structures. A few rare cases of significant slipknots in proteins were found, including two transmembrane proteins. Engineered disulfide bonds were used to probe the role of the complex topology in the stability of one of the slipknotted proteins, alkaline phosphatase
- 27. Cao Z, Roszak AW, Gourlay LJ, Lindsay JG, Isaacs NW. Bovine mitochondrial peroxiredoxin III forms a two-ring catenane. *Structure (Cambridge, Mass)* 2005;13:1661–1664.
- 28. McDonald NQ, Hendrickson WA. A structural superfamily of growth factors containing a cystine knot motif. *Cell* 1993;73:421–424. [PubMed: 8490958]
- 29. Craik DJ, Daly NL, Waine C. The cystine knot motif in toxins and implications for drug design. *Toxicon* 2001;39:43–60. [PubMed: 10936622]
- 30. Bayro MJ, Mukhopadhyay J, Swapna GV, Huang JY, Ma LC, Sineva E, Dawson PE, Montelione GT, Ebricht RH. Structure of antibacterial peptide microcin J25: a 21-residue lariat protoknot. *J Am Chem Soc* 2003;125:12382–12383. [PubMed: 14531661]
- 31. Iwatsuki M, Tomoda H, Uchida R, Gouda H, Hirono S, Omura S. Lariatins, antimycobacterial peptides produced by *Rhodococcus* sp K01-B0171, have a lasso structure. *J Am Chem Soc* 2006;128:7486–7491. [PubMed: 16756302]
- 32. Taylor WR. A deeply knotted protein structure and how it might fold. *Nature* 2000;406:916–919. [PubMed: 10972297]
- 33. Biou V, Dumas R, Cohen-Addad C, Douce R, Job D, Pebay-Peyroula E. The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog determined at 1.65 Å resolution. *Embo J* 1997;16:3405–3415. [PubMed: 9218783]
- 34. Jacobs SA, Harp JM, Devarakonda S, Kim Y, Rastinejad F, Khorasanizadeh S. The active site of the SET domain is constructed on a knot. *Nat Struct Biol* 2002;9:833–838. [PubMed: 12389038]
- 35. Yeates TO. Structures of SET domain proteins: protein lysine methyltransferases make their mark. *Cell* 2002;111:5–7. [PubMed: 12372294]
- 36. Taylor WR, Xiao B, Gamblin SJ, Lin K. A knot or not a knot? SETting the record ‘straight’ on proteins. *Comput Biol Chem* 2003;27:11–15. [PubMed: 12798035]
- 37. Wagner JR, Brunzelle JS, Forest KT, Vierstra RD. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* 2005;438:325–331. [PubMed: 16292304]
- 38. Virnau P, Mirny LA, Kardar M. Intricate knots in proteins: Function and evolution. *PLoS Comput Biol* 2006;2:e122. [PubMed: 16978047]The authors use the Alexander polynomial to look for new knots in the Protein Data Bank. The authors identify a shallow five crossing knot in ubiquitin hydrolase UCH-L3 from *Homo sapiens*. They hypothesize the knot makes the protein resistant to degradation by the proteasome
- 39. Khatib F, Weirauch MT, Rohl CA. Rapid knot detection and application to protein structure prediction. *Bioinformatics* 2006;22:e252–259. [PubMed: 16873480]The authors introduce a modified version of Taylor’s chain smoothing algorithm. The new algorithm is fast enough to be used in structure prediction and the authors apply it to model structures generated by the Rosetta homology-based structure prediction method
- 40. Mansfield ML. Are there knots in proteins? *Nat Struct Biol* 1994;1:213–214. [PubMed: 7656045]
- 41. Lua RC, Grosberg AY. Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS Comput Biol* 2006;2:e45. [PubMed: 16710448]The authors use knot invariants to compare the knotting probabilities in native proteins and random compact loops. From this analysis the authors conclude that the known “protein universe” has avoided knots over the course of evolution
- 42. Taylor, W. Protein folds, knots and tangles. In: C, JA.; M, KC.; R, EJ., editors. *Physical and numerical models in knot theory*. World Scientific; 2005. p. 171–202.
- 43. Taylor WR. Protein knots and fold complexity: some new twists. *Comput Biol Chem* 2007;31:151–162. [PubMed: 17500039]The different types of knots observed in proteins are reviewed from a

theoretical perspective. The implications for structure prediction are discussed, and predictions for the types of knots that may be expected from future structural studies are made

44. Nureki O, Shirouzu M, Hashimoto K, Ishitani R, Terada T, Tamakoshi M, Oshima T, Chijimatsu M, Takio K, Vassylyev DG, et al. An enzyme with a deep trefoil knot for the active-site architecture. *Acta Crystallogr D Biol Crystallogr* 2002;58:1129–1137. [PubMed: 12077432]
45. Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* 2000;289:2129–2133. [PubMed: 11000116]
46. Duff AP, Cohen AE, Ellis PJ, Kuchar JA, Langley DB, Shepard EM, Dooley DM, Freeman HC, Guss JM. The crystal structure of *Pichia pastoris* lysyl oxidase. *Biochemistry* 2003;42:15148–15157. [PubMed: 14690425]
47. Duda RL. Protein chainmail: catenated protein in viral capsids. *Cell* 1998;94:55–60. [PubMed: 9674427]
48. Connolly ML, Kuntz ID, Crippen GM. Linked and threaded loops in proteins. *Biopolymers* 1980;19:1167–1182. [PubMed: 7378549]
49. Lim K, Zhang H, Tempczyk A, Krajewski W, Bonander N, Toedt J, Howard A, Eisenstein E, Herzberg O. Structure of the YibK methyltransferase from *Haemophilus influenzae* (HI0766): a cofactor bound at a site formed by a knot. *Proteins* 2003;51:56–67. [PubMed: 12596263]
50. Clementi C, Plotkin SS. The effects of nonnative interactions on protein folding rates: theory and simulation. *Protein Sci* 2004;13:1750–1766. [PubMed: 15215519]
51. Clementi C, Jennings PA, Onuchic JN. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1 β . *Proc Natl Acad Sci USA* 2000;97:5871–5876. [PubMed: 10811910]
52. Blankenship JW, Dawson PE. Thermodynamics of a designed protein catenane. *Journal of molecular biology* 2003;327:537–548. [PubMed: 12628256]
53. Taylor WR, Lin K. Protein knots: A tangled problem. *Nature* 2003;421:25. [PubMed: 12511935]
54. Matsumura M, Becktel WJ, Levitt M, Matthews BW. Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proc Natl Acad Sci U S A* 1989;86:6562–6566. [PubMed: 2671995]
55. McDonald NQ, Lapatto R, Murray-Rust J, Gunning J, Wlodawer A, Blundell TL. New protein fold revealed by a 2.3-Å resolution crystal structure of nerve growth factor. *Nature* 1991;354:411–414. [PubMed: 1956407]

Figure 1.

Types of topological complexity observed in proteins. In each panel, a simplified view of the protein is shown on the left, with a stylized diagram of the topology of the system on the right. (a) A unique case of non-covalent catenation. The crystal structure of bovine mitochondrial peroxiredoxin III (PDB code 1zye) revealed two interlinked rings of twelve subunits each [27]. (b) A topological folding barrier [14•] in human superoxide dismutase (1hl4). The red segment of the protein backbone is threaded through a ring formed by the surrounding blue residues. (c) The crystal structure of nerve growth factor (1bet) revealed the first view of the cystine knot motif [55]. The three disulfide bonds which define the motif are shown as red bars. (d) The backbone of the RNA 2'-O-ribose methyltransferase RrmA (1ipa) contains a deep trefoil knot, colored to facilitate visualization [44]. (e) *E. coli* alkaline phosphatase (1alk) was recently identified as having a deeply slipknotted topology [26••]. The magenta segment of the chain is threaded through the 'knot core' (green), but the C-terminal portion of the chain (red) returns through the knot core to effectively unknot the protein as a whole. (f) The dimeric citrate synthase from *P. aerophilum* (2ibp) is topologically linked by two intramolecular disulfide bonds, shown as red bars [23].

Figure 2.

Protein knot plots of four representative knotted proteins. The key (top left) associates various knot types with colors in the plots: green = right-handed trefoil (knot designation 3_1), red = left-handed trefoil, blue = figure eight knot (4_1), yellow = 5_2 knot. Within a given plot, each point in the square matrix indicates a partial structure contained within the protein of interest. The point at the lower left corner of a matrix indicates the complete protein chain, while points closer to the diagonal indicate smaller partial structures. Truncating the N-terminus of a protein corresponds to moving from the lower left corner in a horizontal direction, while truncation of the C-terminus corresponds to moving vertically upwards. White regions are unknotted and colored regions are knotted. (a) RNA 2'-O-ribose methyltransferase (PDB code 1ipa) showing

a right-handed trefoil knot that is deep (about 41 residues can be truncated from the C-terminus before the knot is eliminated) and tight (the smallest knotted substructure is only about 44 residues long) [44]. (b) Acetohydroxy acid isomeroreductase (1qmg), showing a deep figure eight knot [32]. A tight trefoil, not previously noted, is also visible within the structure. (c) Alkaline phosphatase (1alk) showing a slipknot structure [26••]; a right-handed trefoil is found within the structure, but the complete protein chain is unknotted. (d) The enzyme ubiquitin hydrolase UCH-L3 (1xd3) showing a complex five-crossing knot [38•]. Note that the five-crossing knot is formed only by the last few C-terminal residues. Otherwise, the structure contains a shallow left-handed trefoil.

Figure 3.

Unknotting and knotting proteins by design. (a) Schematic of the fold of a hypothetical knotted protein. Altering the connectivity of the protein chain at the two indicated crossings (*) and (#) results in a protein with a nearly identical core structure, but an unknotted topology. (b) The fold of the hypothetical unknotted protein generated from the knotted protein in (a). Note how the reverse operation could be applied to the unknotted protein to regenerate the knotted version. (c) Schematic of the primary and secondary structures of the knotted (top) and unknotted (bottom) proteins. The operations necessary to unknot or knot the protein are indicated by pairs of dashed arrows (middle).

TABLE 1

Representative Protein Knots and Slipknots

Protein	Organism	PDB Code	Type**
RNA 2'-O Ribose Methyltransferase	<i>Thermus thermophilus</i>	1IPAA	3 ₁ knot
Hypothetical tRNA/rRNA Methyltransferase HI0766	<i>Haemophilus influenzae</i>	1MXIA	3 ₁ knot
Transcarbamylase	<i>Bacteroides fragilis</i>	1JS1X	3 ₁ knot
Hypothetical Protein HI0303	<i>Haemophilus influenzae</i>	1VHYA	3 ₁ knot
Acetohydroxy Acid Isomeroreductase	<i>Spinacia oleracea</i>	1YVEL	4 ₁ knot
Conserved Protein MT0001	<i>Methanobacterium thermoautotrophicum</i>	1K3RA	3 ₁ knot
Bacteriophytochrome	<i>Deinococcus radiodurans</i>	1ZTUA	4 ₁ knot
tRNA (Guanine-N(1)-)- Methyltransferase	<i>Escherichia coli</i>	1P9PA	3 ₁ knot
Hypothetical UPF0247 Protein TM0844	<i>Thermotoga maritima</i>	1O6DA	3 ₁ knot
* Ubiquitin hydrolase UCH-L3	<i>Homo sapiens</i>	1XD3A	5 ₂ knot
Alkaline Phosphatase	<i>Escherichia coli</i>	1ALKA	3 ₁ slipknot
Thymidine kinase	<i>Equine herpesvirus</i>	1P6XA	3 ₁ slipknot
Glutamate Symport Protein	<i>Pyrococcus horikoshii</i>	2NWL A	3 ₁ slipknot
Na(+):Neurotransmitter Symporter (SNF Family)	<i>Aquifex Aeolicus VF5</i>	2A65A	3 ₁ & 4 ₁ slipknots
STIV B116*	<i>Sulfolobus Turreted Icosahedral Virus</i>	2J85A	3 ₁ slipknot

* Indicates a knot shallower than 10 residues. All others listed are deeper than 20 residues.

** All of the 3₁ (trefoil) knots observed are right-handed, although ubiquitin hydrolase contains a left-handed trefoil as a substructure.