doi:10.1016/j.jmb.2007.07.042 J. Mol. Biol. (2007) 373, 153–166

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Identification of Rare Slipknots in Proteins and Their Implications for Stability and Folding

Neil P. King[1], Eric O. Yeates[1] and Todd O. Yeates[1,2]*

[1]*UCLA Department of Chemistry and Biochemistry, 611 Charles Young Dr. East, Los Angeles, CA 90095-1569, USA*

[2]*UCLA-DOE Institute of Genomics and Proteomics, Los Angeles, CA 90095-1570, USA*

Received 16 May 2007; received in revised form 17 July 2007; accepted 19 July 2007
Available online
2 August 2007

Edited by I. Wilson

Among the thousands of known three-dimensional protein folds, only a few have been found whose backbones are in knotted configurations. The rarity of knotted proteins has important implications for how natural proteins reach their natively folded states. Proteins with such unusual features offer unique opportunities for studying the relationships between structure, folding, and stability. Here we report the identification of a unique slipknot feature in the fold of a well-known thermostable protein, alkaline phosphatase. A slipknot is created when a knot is formed by part of a protein chain, after which the backbone doubles back so that the entire structure becomes unknotted in a mathematical sense. Slipknots are therefore not detected by computational tests that look for knots in complete protein structures. A computational survey looking specifically for slipknots in the Protein Data Bank reveals a few other instances in addition to alkaline phosphatase. Unexpected similarities are noted among some of the proteins identified. In addition, two transmembrane proteins are found to contain slipknots. Finally, mutagenesis experiments on alkaline phosphatase are used to probe the contribution the slipknot feature makes to thermal stability. The trends and conserved features observed in these proteins provide new insights into mechanisms of protein folding and stability.

Keywords: protein knots; protein folding; alkaline phosphatase; thermophilic proteins; protein topology

Introduction

Do protein structures offer hints about the paths their chains take as they fold into their final, native configurations? The idea that the folded configuration of a protein chain should be informative about the pathway by which the folded state is reached dates back at least to the 1970s.[1,2] Recent investigations into the relationship between native structure and folding pathways have led to the development of ideas such as "contact order" and related geometric measures as predictors of folding rates.[3–5] According to those ideas, a protein may not fold easily if pathways to the folded state tend to involve unlikely configurations. Situations where this is particularly likely to be the case include proteins that have complex topological features, such as knotted or linked chains. Although such proteins are rare, a number of cases have been described.[6–12] Computational, structural, and biophysical investigations of these unusual proteins should lead to valuable insight into how proteins fold and how they maintain their folded configurations, often under extreme conditions.

The question of how likely it would be for a linear protein chain to fold up into a knot was first considered by Crippen in 1974.[1] None of the known protein structures at that time had backbones that were knotted. Twenty years later, the issue of whether the database of known protein structures might contain knots was addressed computationally by Mansfield.[13] In a pure mathematical sense, only a closed curve can be truly knotted, but ideas were introduced under which open curves (i.e., typical protein chains) might be considered to be practically knotted; the chain termini could be extended away from the body of the protein and joined virtually.[13] When this first computational survey was conducted, approximately 400 known structures were

*Corresponding author. E-mail address: yeates@mbi.ucla.edu.

Abbreviations used: EHV, equine herpesvirus; PDB, Protein Data Bank; STIV, *Sulfolobus* turreted icosahedral virus.

examined, but still no striking cases of deep knotting could be found.[13,14] By then, the absence of knots in proteins was noteworthy. Simulations on random polymers of increasing length suggested that knots would have been seen in the PDB if they had not been systematically selected against in nature.[15] The importance of the apparent rarity of knots in proteins was discussed.[13,14] If a major fraction of the possible configurational space for proteins was never observed, it might be concluded that protein folding is not "ergodic" (i.e., does not sample all configurations).

The perspective on protein knots changed in 2000 when Taylor[10] discovered that the structure of a deeply knotted protein had been determined,[16] but had gone unnoticed. That finding reopened the question of how "forbidden" knots are in proteins. It also highlighted the importance of using computational tools to examine protein structures for such topological features. Recent computational investigations have identified a total of five or six distinct protein folds bearing knots of significant depth,[17–19] and experimental studies aimed at understanding the effects of knots on protein folding have been initiated in one family of knotted proteins.[20,21]

Knots in proteins bear not only on folding pathways but also on stability and function. Nureki *et al*. noted the participation of the knotted region of the protein chain in the active site of a thermophilic RNA methyltransferase.[9] The knot in the phytochrome domain from the radiation-resistant *Deinococcus radiodurans* also participates in the active site.[11] Similar stabilization issues may be relevant in topologically linked structures as well. The HK97 viral capsid is constructed from protein chains that are covalently linked and topologically connected.[22] The resulting stabilization may be what allows that virus to make do with an unusually thin shell. In the thermophilic microbe *Pyrobaculum aerophilum*, the protein chains of the dimeric enzyme citrate synthase have been found to be cyclized and linked together.[7] In that case, the topological connection was shown to contribute to thermal stability.[7] Stabilizing topological connections have also been noted in the bactericidal polypeptide microcin J25,[6] and engineered topological connectivity has been used to stabilize a polypeptide system based on the p53 tetramerization domain.[23]

Recent computational[17–19,24] and experimental[7,20,21] investigations highlight a renewed interest in knots and links in proteins. Here, we investigate known protein structures for the presence of slipknots. We use the term slipknot to refer to a situation where some part of the protein chain is tied in a knot, but the protein chain considered in its entirety would be judged to be unknotted in the mathematical sense. This occurs when the path of a protein chain forms a knot, but the chain then folds back upon itself in a way that effectively unties the knot. A tied shoelace provides a familiar example. As with knots, slipknots would seem to present important considerations for protein folding, but slipknots are not detected by typical searches for knots among known structures. Here, we identify a few rare cases of slipknots in proteins and provide preliminary experimental characterization of the effects of the slipknot in one protein, the thermostable enzyme alkaline phosphatase.

Results

Identification of a deep slipknot in alkaline phosphatase

During a visual survey of dimeric structures in the Protein Data Bank (PDB), we noted the presence of a deep slipknot in the structure of *Escherichia coli* alkaline phosphatase (PDB code 1ALK[25]). Alkaline phosphatase is a functionally and structurally well-characterized enzyme, yet its peculiar slipknotted topology had not, to our knowledge, been noted before. The smallest substructure of the protein that constitutes a knot is the segment between residues 51 and 371, which forms a right-handed trefoil knot (Fig. 1). The slipknot in alkaline phosphatase is deep in the sense that the protein chain extends significantly on either side of the knot core. Approximately 50 residues precede the knot core on the N-terminal side of the protein. On the C-terminal side, 49 residues (371–419) extend out from the knot core to form the encircled "protruding loop," but the chain then threads back through the knot core, with the final 29 residues of the protein forming a tail that effectively unknots the chain as a whole. Therefore, a partial structure lacking the 30 most C-terminal residues (the "slipknot tail") would be classified as knotted. Following ideas developed for knotted chains, we take the depth of a slipknot to be the lesser of the lengths of the two termini that extend out of the knot core in the knotted partial structure lacking the slipknot tail. In the case of alkaline phosphatase, this is 49 residues (the lesser of 49 and 50), which is quite deep for a protein knot. Among the known knotted proteins, only the knots in certain RNA methyltransferases (e.g. PDB code 1P9P), transcarbamylase (1JS1), and acetohydroxyacid isomeroreductase (1QMG) are deeper. Although alkaline phosphatase is a large protein, the topological core of the slipknot can be reduced to five β-strands within the enzyme's 10-stranded β-sheet core. Strands 1, 2, 7, 8, and 9 are contiguous in the sheet in the order 2, 9, 8, 1, 7, with strand 9 antiparallel with the others. The result of this motif is that the segment of the protein from the end of strand 1 (around residue 51) to the end of strand 8 (around residue 371) forms a trefoil knot, while the domain between strands 8 and 9 protrudes out of the knot core to form the slipknot.

Computational identification and visual representation of knots and slipknots in proteins

If a protein contains a slipknot, this is manifested by the presence of a knot in some partial structure

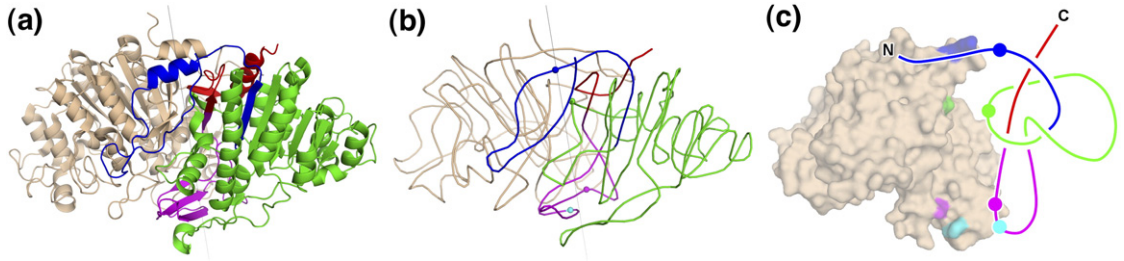

Fig. 1. Structure of the deeply slipknotted protein alkaline phosphatase. (a) Cartoon representation of the dimer. The N-terminal knotting extension is blue; the knot core, green; the protruding loop, magenta; and the C-terminal slipknot tail, red. The twofold axis relating the subunits is shown as a grey line. Note the difficulty of visualizing the slipknot. (b) A smoothed representation of the protein backbone, colored as in (a). Residues 34, 81, 384, and 403 (which were mutated to cysteine in this study) of chain A are shown as blue, green, cyan, and magenta spheres, respectively. (c) A highly simplified version of the slipknot in alkaline phosphatase, colored as in (a), pictured next to a surface representation of subunit B (residues 1–30 were omitted for clarity). Residues 34, 81, 384, and 403 are colored as in (b), and represented as circles in the simplified slipknot image. All figures of protein structures were made with PyMOL.[56]

(constituting a single continuous segment) within the entire protein chain. Finding slipknots therefore requires an examination of the possible partial structures within a protein chain that arise from different possible choices for the beginning and ending residues. The set of possible partial structures of a protein chain can be represented by a square matrix in which the coordinates of each element in the matrix specify the beginning and ending residues of a partial structure. The corner of the matrix represents the intact protein chain, whereas points closer to the diagonal represent smaller partial structures (Fig. 2). Similar representations have been used before to characterize how potential protein-folding pathways might be affected by various topological features of proteins, including knots.[24,26,27] Following Taylor, here we color each point in the knot matrix according to whether the corresponding partial structure is a knot.[27] This representation provides a useful device for visualizing knots and slipknots (Fig. 2). If a protein is knotted, then the corner region of the knot matrix—which represents the intact protein chain—will be colored. If a protein contains a slipknot, the corner will not be colored. Instead, a region shifted along one of the edges will be colored. Moving along an edge of the matrix, away from the corner, is equivalent to progressively truncating one of the termini of the protein chain; for a protein that contains a slipknot, removing the terminus generates a knotted structure. The knot matrix also provides a description of the depth of a knot, which is useful in judging whether the knot that has been detected is likely to be substantial enough to influence the folding or stability of the protein.

Algorithms for evaluating knots in protein chains come in two varieties, those that are effectively mechanical and those based on knot theory. The mechanical approach is to try to numerically pull straight or smooth out a chain without allowing it to

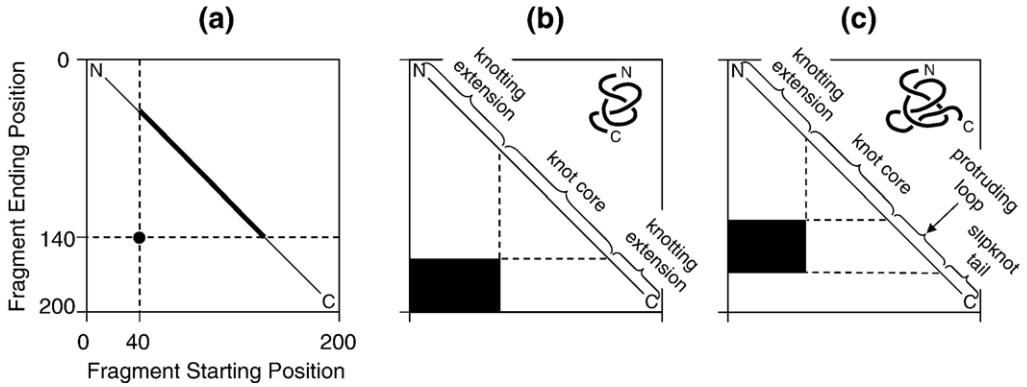

Fig. 2. Explanation of the knot matrix for visualizing knots and slipknots in proteins. (a) Each point in the lower left region of the square matrix corresponds to a fragment or partial structure (bold line segment) within the complete protein chain, which is 200 amino acids long in this hypothetical diagram. The bottom left corner of the square corresponds to the complete protein chain. (b) The knot matrix is shown for a hypothetical knotted protein. The colored region comprises the partial structures that would be judged to be knotted. The lengths of the sides of the rectangular colored region describe how deeply knotted the protein is at the N and C termini, with the length of the shorter side providing a potential measure of folding difficulty. (c) The knot matrix is shown for a hypothetical slipknotted protein. Note that the complete protein chain would be judged to be unknotted in the mathematical sense. Similar representations of knotting and topological complexity in partial structures have been described.[24,27]

pass through itself. If this is impossible, then one concludes that the chain is knotted. This is the approach that was taken first by Taylor.[10] Such approaches are intuitively appealing and rapid, but there are pitfalls. Entanglement sometimes occurs, so there is no guarantee that a structure that could not be straightened is truly knotted. An approach similar to Taylor's was reported recently by Khatib *et al.*, wherein the straightening is performed in two different ways in order to reduce the occurrence of false positives.[28]

The other approach to protein knot analysis is using knot theory. The question of whether a given curve is knotted or not has a long history, and is still not completely solved. In 1928, J. W. Alexander provided the first mathematical system—the Alexander polynomial—for categorizing knots by computing invariant properties of their two-dimensional representations.[29] This was the categorization employed by Mansfield[13] when he examined protein chains, and by Virnau in a recent report.[19] Some 50 years after Alexander's work, Vaughn Jones developed more powerful invariants for knots;[30] the Jones polynomial discriminates between more complex knots, including their handedness. It is possible with the Jones polynomial to calculate from a two-dimensional representation of a curve a reduced polynomial representation that is unique, at least out to a level of knot complexity sufficient to handle any kind of knot likely to be encountered in proteins. Following the development of the Jones polynomial, other knot invariants have been developed. One recent analysis of protein knots relies on the Alexander polynomial, combined with the Vassiliev invariants.[18] The knot theory approaches offer some trade-offs compared to the mechanical methods. They provide a firm conclusion about whether or not a curve is knotted. In addition, they provide a potentially richer description. While mechanical methods simply conclude that a curve could not be straightened, knot theory approaches can classify knots that are found as being trefoils (left- or right-handed), figure-eight knots, or even more complex knots.[18,19] Among the disadvantages of the knot theory approaches is a tendency to be computationally intensive for large proteins whose two-dimensional representations contain many crossings. Here, we evaluate the Jones polynomial for protein structures using a recursive algorithm (see Methods).

Discovery of slipknots in a survey of the PDB

The knot and slipknot calculations were performed on 14,870 protein chains obtained from the PISCES database of unique protein structures extracted on April 4, 2007, based on a 95% sequence identity cutoff between protein chains.[31] In the initial analysis over all proteins, in order to make the computational survey more tractable, only the edges of the knot matrix were sampled (Fig. 2). Especially for large proteins, it was not unusual to find among the many partial structures tested one or a few that were loosely wound and were determined computationally to be knotted because of the particular way the ends of the partial fragment were connected computationally. In order to eliminate spurious situations and to identify only proteins having significant knots or slipknots, the extent of the knotting along the edges of the knot matrix was required to exceed a given threshold, such as a region of knotting of at least 20 residues.

The calculations resulted in a list of knotted protein chains and a list of slipknotted protein chains. The knotted chains identified included all those that have been identified recently as being deeply knotted.[17–19] For the slipknot category, 37 protein chains were identified. These were then grouped by homology, as judged by structural similarity, in order to single out a representative from each fold type.[32] Four distinct protein folds, represented by PDB codes 1ALK, 1P6X, 2NWL, and 2A65, were judged to have deep slipknot features. Remarkably, two of these (2NWL and 2A65) are transmembrane proteins. In addition, a fifth protein fold (PDB code 2J85) that did not satisfy the strict criterion of having a knotting region of at least 20 residues was retained for further analysis, owing to its remarkably small size. These five protein chains were subjected to full knot calculations over the entire two-dimensional knot matrix (Fig. 3).

The three water-soluble proteins identified as having significant slipknots were alkaline phosphatase (PDB code 1ALK[33]), equine herpesvirus (EHV) thymidine kinase (1P6X[34]), and protein B116 from *Sulfolobus* turreted icosahedral virus (STIV; 2J85[35]). Of the three, the slipknot in alkaline phosphatase is the deepest, with a depth of 49 residues as described above. The knot matrix for alkaline phosphatase is shown in Fig. 3a. It is noteworthy that the knotted region of the alkaline phosphatase plot extends significantly along both the horizontal and the vertical directions, indicating that both the N-terminal extension and the protruding loop extending out of each side of the knot core are of considerable size. The knotted region occurs along the left side of the diagram, indicating that the protein is a C-terminal slipknot; truncating the C terminus generates a right-handed trefoil knot. Alkaline phosphatase is just one representative of a large family of homologous proteins, all of which contain similar slipknots, indicating that these proteins have retained their slipknotted topologies since diverging from a slipknotted common ancestral protein. This family includes acid phosphatase, arylsulfatase, phosphonoacetate hydrolase, and 2,3-bisphosphoglycerate-independent phosphoglycerate mutase.

The next most significant slipknot among soluble proteins was found in EHV thymidine kinase (PDB code 1P6X). The slipknot in this protein is also of the C-terminal variety (Fig. 3b), but there are notable differences compared to alkaline phosphatase in terms of the depth and tightness of the knot. As illustrated by the elongated nature of the rectangular knot region, a large portion of the protein (residues 139–328) makes up the protruding loop that is encircled by the rest of the chain in the

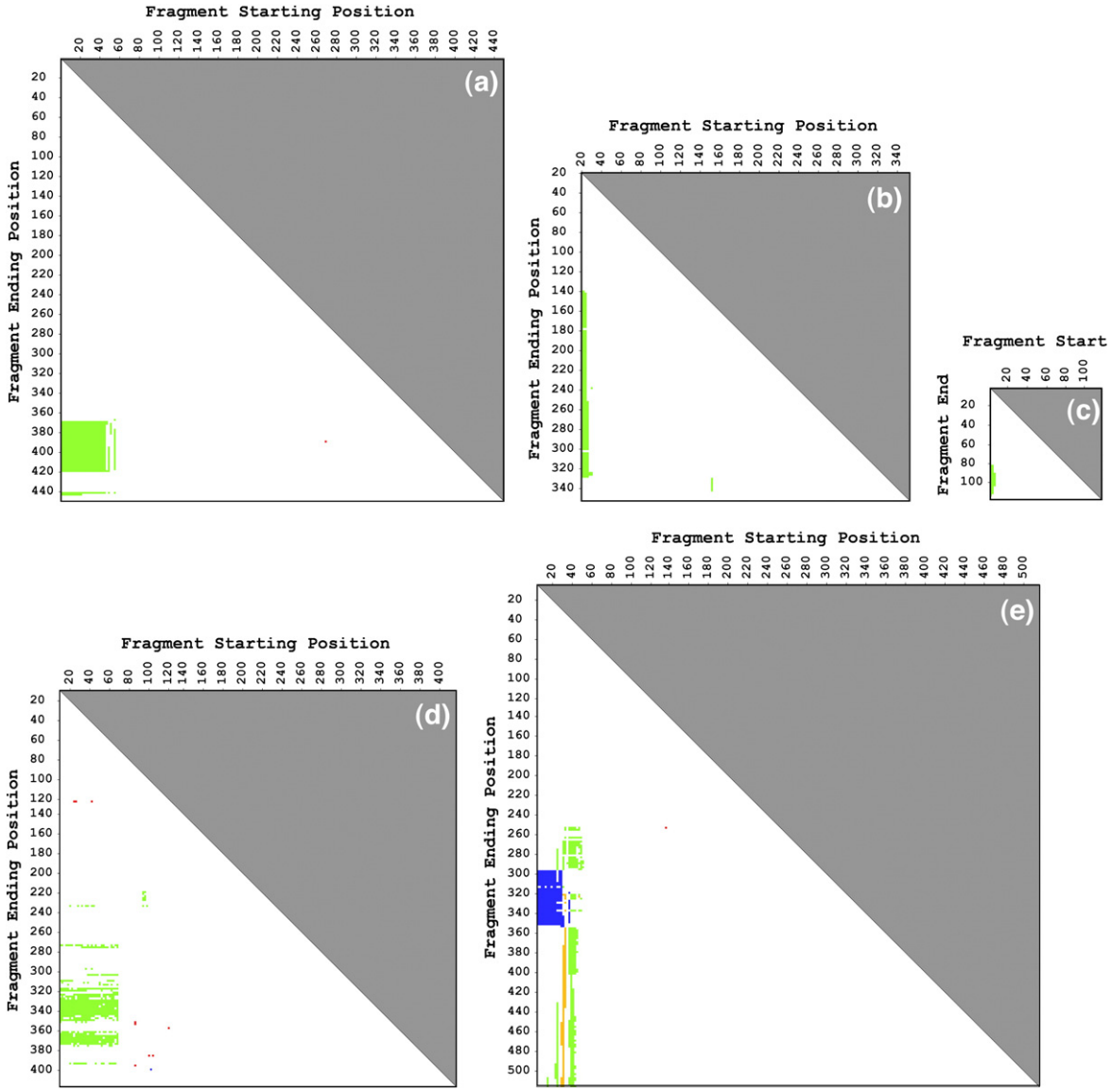

Fig. 3. Plots of the knot matrices for five protein folds in which significant slipknots were found. (a) *E. coli* alkaline phosphatase (PDB code 1Y6V[33]). (b) Equine herpesvirus thymidine kinase (1P6X[34]). (c) Protein B116 from STIV (2J85[35]). (d) Glt_{Ph}, a sodium-dependent aspartate transporter (2NWL[39]). (e) $LeuT_{Aa}$, a bacterial homolog of Na^+/Cl^--dependent neurotransmitter transporters (2A65[40]). Colored regions indicate partial structures determined to be knotted, colored according to the following scheme: green, right-handed trefoil; red, left-handed trefoil (of which there are only spurious occurrences); blue, figure-eight knot; orange, the five-crossing knot (5_2), which appears as a minor, transient feature in the $LeuT_{Aa}$ protein.

slipknot. Whereas the slipknot in thymidine kinase is therefore very deep on the C-terminal side, it is shallower on the N-terminal side, with only about 10 residues extending out of the knot core. However, the length of the N-terminal extension is most likely longer in the native protein than indicated by the plot shown here, as the native protein contains 19 additional residues at the N terminus that are not part of the reported crystal structure.[34] The difference in depth on the two sides of the thymidine kinase knot core suggests that the knot is most likely formed by the N-terminal segment of the protein passing through a preformed opening in the protein during the later stages of protein folding. Because the N-terminal extension is shorter, the slipknot in thymidine kinase is considerably shallower than the one in alkaline phosphatase. However, the knot is much tighter. The knot core—the smallest substructure that constitutes a knot—is only about 114 residues (26–140) in thymidine kinase, compared to 320 residues in alkaline phosphatase, as noted above.

The third slipknot described here (PDB code 2J85) is in protein B116 from STIV (Fig. 3c). This virus infects thermophilic *Sulfolobus* species found in hot springs in Yellowstone National Park. Disulfide bonding may be an important mechanism for stabilizing proteins from this virus (E. T. Larson and C. M. Lawrence, personal communication), and indeed protein B116 does contain a disulfide bond.[35] In addition to the disulfide bond, the slip-

knot feature may also contribute to thermal stability. The slipknot in this protein is rather shallow, particularly on the N-terminal side, but it is remarkable that a slipknot can be formed in such a small protein. The knot core is only about 76 residues long.

Despite the differences in overall size and extent between the slipknots in the three soluble proteins highlighted here, there are some striking similarities in the detailed architectures of the slipknots. As mentioned above, the topological core of the alkaline phosphatase slipknot can be reduced to a sheet of five β-strands. The slipknots in EHV thymidine kinase and STIV protein B116 are both built on quite similar β-sheet cores, although the ordering of one of the strands differs in the case of thymidine kinase (Fig. 4). The three slipknots are all of the C-terminal variety. In thymidine kinase and especially in STIV protein B116, the knot core is shallow and could probably be formed rather easily by threading the N terminus into the core of the structure. In the case of alkaline phosphatase, the N-terminal extension is longer, and threading the 50-residue N-terminal extension through the protein would seem to present a more serious topological barrier to folding. In fact, alkaline phosphatase fused C-terminally to a large number of proteins retains activity,[36,37] showing that its ability to fold to the native state is insensitive to the depth of the slipknot on the N-terminal side. This is consistent with the notion of the slipknot being formed by threading of the C-terminal protruding loop. The three proteins noted above have also been identified by Taylor in a search for slipknot proteins that was conducted concurrently with the present study.[38]

The other two proteins identified as containing significant slipknots are transmembrane proteins. To our knowledge, this is the first report of slipknots or knots in transmembrane proteins. The transmembrane protein Glt_{Ph}, a sodium-dependent aspartate transporter (PDB code 2NWL[39]), contains a slipknot that is similar in some respects to those described above in the soluble proteins. The knot core is a right-handed trefoil, and the slipknot is again of the C-terminal variety (Figs. 3d and 5a). The major difference is that the knot is formed by a bundle of helices and is therefore entirely unrelated to the β-sheet architectures seen in the soluble slipknot proteins. The knot core is formed by residues 55–319, a region that encompasses transmembrane helices 3 through 7a. This makes the slipknot loose, although it is rather deep, with 54 residues extending on the N-terminal side including transmembrane helices 1 and 2, and 59 residues constituting the protruding loop, which actually passes and returns through the knot core twice as the chain forms a helical hairpin in this region. Interestingly, this helical hairpin contributes residues to the aspartate binding site in the protein. We note that the veracity of the slipknot identified in the aspartate transporter depends on the accuracy of the reported structure. Because the diffraction resolution in this case is only 2.96 Å, higher resolution data would be useful in ruling out any possibility that the loops of the protein chain might cross each other differently.

The second transmembrane protein, $LeuT_{Aa}$, a bacterial homolog of Na^+/Cl^--dependent neurotransmitter transporters (PDB code 2A65[40]), contains a slipknot that is more complex than any of the others identified. In this case, a knot is formed when either one of the termini is truncated (Figs. 3e and 5b and c). If about 34 residues are truncated from the N terminus, a right-handed trefoil is formed. If 165 residues are truncated from the C terminus, a figure-eight knot is formed. In both cases, the knots are substantial. As in the other transmembrane slipknot, the knots are formed predominantly by bundles of alpha helices, along with the loops that connect them. It is remarkable that a protein backbone that is not knotted can be converted to two different kinds of knots by truncation of the individual termini. A simplified diagram illustrates how this is possible (Fig. 5c). The structure of $LeuT_{Aa}$ was determined at high resolution (1.65 Å), which provides a high level

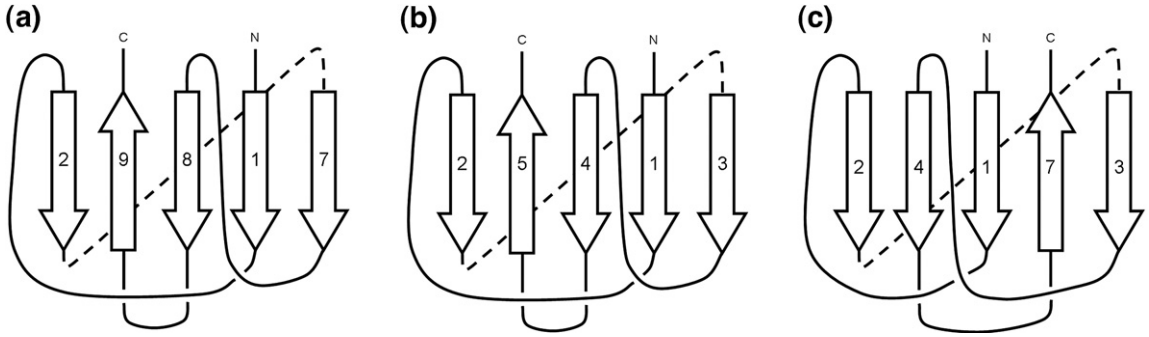

Fig. 4. Diagrams of the five-stranded slipknot cores observed in soluble slipknotted proteins. (a) A simplified diagram of the AP slipknot core, highlighting the five β-strands involved in the knot core. The strands are numbered according to the order in which they occur in the full-length protein. Strands 3–6 and 10 are not shown since they do not directly contribute to the topology of the slipknot. (b) A similar diagram of the slipknot core in protein B116 from STIV. Note the conservation of both strand order and direction when compared to the slipknot in AP. (c) A simplified diagram of the slipknot core of thymidine kinase from EHV. The final strand of the motif is in a different position in the sheet compared to the previous two slipknots, but the directions of the corresponding strands are conserved. Strands 5 and 6 are not shown since they do not directly contribute to the topology of the slipknot.

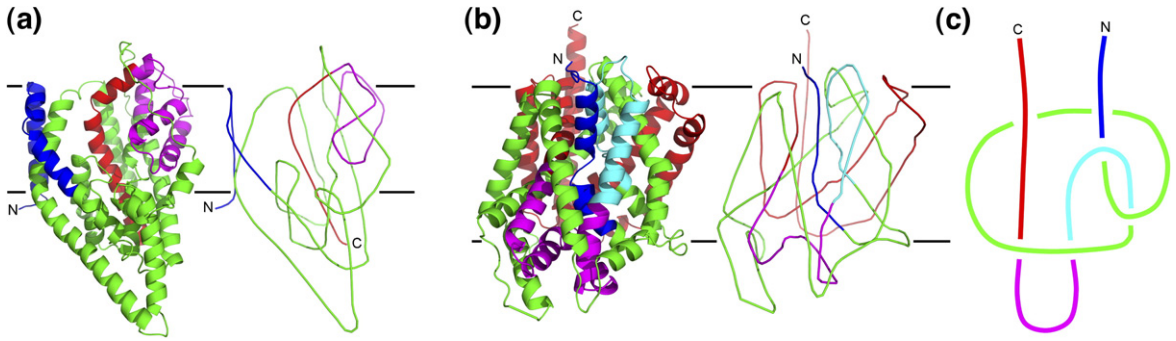

Fig. 5. Simplified representations of slipknotted transmembrane proteins. (a) Cartoon (left) and smoothed backbone (right) representations of Glt_{Ph} with the N-terminal knotting extension colored blue; the knot core, green; the protruding loop, magenta; and the C-terminal slipknot tail, red. (b) Similar representations of $LeuT_{Aa}$, except that an additional section of the protein, the "figure-eight" segment, is cyan. In (a) and (b) the horizontal black lines represent the approximate boundaries of the lipid bilayer. (c) A highly simplified schematic of the topology of the $LeuT_{Aa}$ backbone. In this diagram, it can be seen how deletion of the N-terminal knotting extension (blue) would result in a trefoil knot, but deletion of the C-terminal slipknot tail (red) would result in a figure-eight knot.

of confidence in the model and the slipknot features observed.

Experimental analysis of the stabilizing effects of the alkaline phosphatase slipknot

Preliminary protein engineering experiments were conducted to investigate a potential stabilizing role for the slipknot in alkaline phosphatase. Alkaline phosphatase is a homodimer, and the protruding loop that is essential for forming the slipknot passes near the twofold axis of symmetry. This made it possible to introduce single cysteine residues in the protruding loop at positions that would permit disulfide bond formation between the two subunits. Connecting the protruding loop from one subunit to the same loop from the other subunit strongly enhances the knotting characteristics of the protein by preventing those loops from retracting through the knot core, which would untie the protein chains (Fig. 6). The experiments described here are based on single cysteine substitutions. This restricted the sites of cysteine substitution to positions very near the twofold axis of symmetry in the dimer. Two single cysteine mutants were introduced in the protruding loop, one at position 384 (P384C; Fig. 6b) and the other at position 403 (G403C; Fig. 6c). Single cysteine mutations were also generated at two other positions, one at residue 34 (R34C; Fig. 6c), which was chosen to probe the effects of enhancing the knotting on the N-terminal side, and one at residue 81 (T81C; Fig. 6a), which was intended to act as a control by linking the subunits together in a way that does not affect the knotting. An AP(R34C/G403C) double mutant was also constructed in order to make it possible to create a cross-linked structure that contains two trefoil knots in a covalently closed chain (Fig. 6c).

In each of the purified mutants of alkaline phosphatase, the engineered cysteines were oxidized to generate an intermolecular disulfide bond using cupric sulfate, a strong oxidizing agent. No single set of conditions was found that resulted in complete disulfide formation for all mutants, but under the conditions established (see Methods), the disulfide bonded forms represented the major species for all of the mutant proteins as judged by SDS-PAGE under nonreducing conditions (Fig. 7a). The disulfides at positions 384 and 403 (the knotting disulfides) formed essentially quantitatively, while the disulfides at positions 81 and 34 were formed in most and about half of the protein molecules,

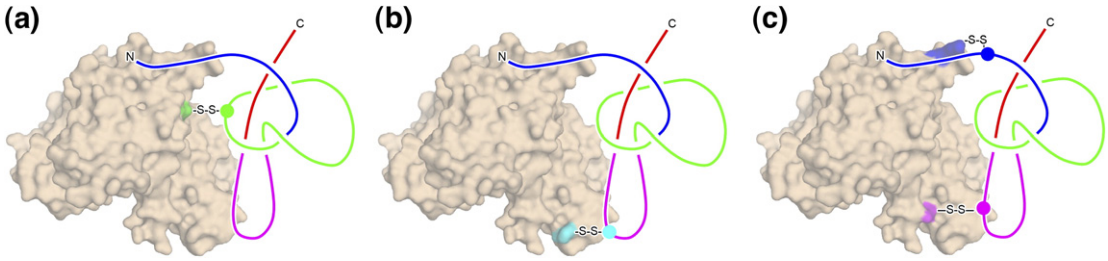

Fig. 6. Schematic diagrams of cross-linked mutant APs. (a) A representation of the AP dimer similar to that in Fig. 1c except that a schematic of the engineered disulfide at position 81 (shown in green) has been shown connecting the two subunits. (b) A similar representation of the designed disulfide bond connecting the two subunits at position 384 (shown in cyan). (c) A schematic of the AP(R34C/G403C) double mutant, with two intersubunit cross-links at positions 34 (blue) and 403 (magenta), which mathematically knot the dimer in the sense that a covalently closed loop containing a knot in each subunit is formed by the two protein molecules. Diagrams of AP(R34C) and AP(G403C) are not shown since both disulfides appear in (c).

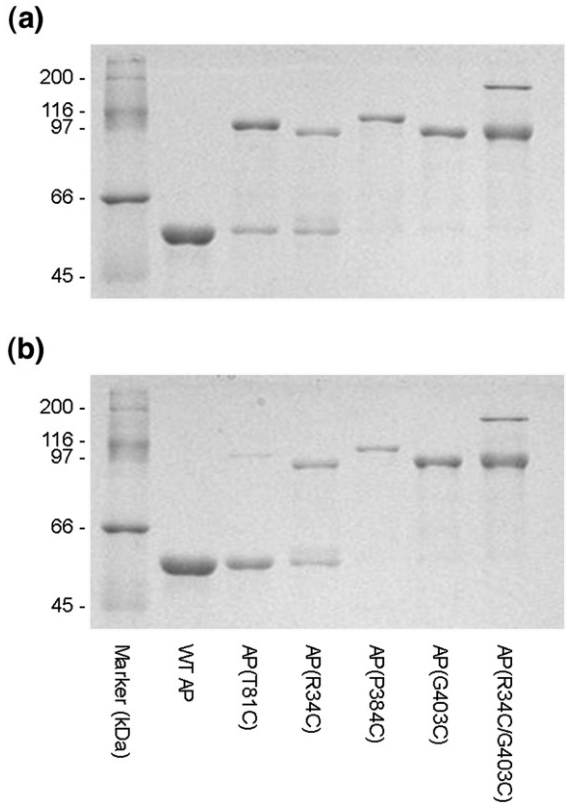

Fig. 7. Disulfide bond formation and retention in mutant APs. (a) Nonreducing SDS-PAGE of all APs after oxidation with $CuSO_4$. Mutant protein populations that do not contain the engineered intersubunit disulfide migrate as monomers with WT AP. Disulfide-bonded mutant proteins migrate as dimers (upper bands). The highest band in the AP(R34C/G403C) lanes is taken to be the doubly cross-linked species. (b) Nonreducing SDS-PAGE of all APs after demetallation. The only detectable difference from (a) is a decrease in the fraction of AP (T81C) molecules containing the intermolecular disulfide bond.

respectively (Fig. 7a). After the intermolecular disulfide bonds were formed in the mutant proteins, all of the protein samples were demetallated. Alkaline phosphatase binds two zinc ions and one magnesium ion per subunit, which contribute significantly to the stability of the enzyme. The proteins were therefore stripped of their bound metals prior to thermal denaturation experiments to eliminate the possibility of differential metal binding between the mutants, which would have complicated our analysis. After demetallation, the extent of disulfide bond retention in each of the samples was evaluated (Fig. 7b). This experiment revealed that while the disulfide linkages at positions 34, 384, and 403 were unaffected by demetallation and were stable over time, the disulfide at position 81 was labile even at pH 5.5, well below pK_a values typical of thiol groups. The extent of disulfide loss at position 81 was identical in protein samples that went through two independent demetallation protocols, suggesting an inherent instability of the disulfide at this position. The T81C mutation was designed to provide a simple covalent connection between the subunits of the dimer with no effect on knotting. The instability of the T81C disulfide therefore made it impossible to evaluate what effects a simple intermolecular covalent bond might have on alkaline phosphatase stability. The thermal denaturation studies on the other mutants, however, were informative, although the incomplete formation of the disulfide bond in the R34C mutants led to complex thermal melting curves representative of mixtures of proteins with or without the disulfide.

The thermal stabilities of the apo forms of the wild-type and mutant alkaline phosphatases were evaluated using CD spectroscopy. Wavelength scans in the far-UV region at 20 °C revealed very little difference in the secondary structure content of the proteins prior to thermal melts (data not shown). Our observed melting transition for apo wild-type alkaline phosphatase (apo-WT AP) at 62 °C (Fig. 8) matches well to those determined previously using differential scanning calorimetry (57.5 °C[41,42]) and CD (59 °C[43]). Apo-AP(T81C) exhibited an essentially identical transition at 62 °C, although, as discussed above, this represents the transition for a protein sample that is mutated but not cross-linked (Fig. 8a). Two transitions were observed for the N-terminally cross-linked apo-AP(R34C), one around 45 °C and one around 59 °C, presumably representing the two populations in which the disulfide is either intact or reduced (Fig. 8a). Because the two populations are approximately equivalent in size, it is not obvious whether the higher or lower transition represents the disulfide cross-linked population. In either case, however, the disulfide-bonded species is less stable than the wild-type protein. In contrast, the two mutants in which the knotting was enhanced by linking through the protruding loop were both significantly stabilized compared to apo-WT AP. The melting temperatures of the apo-AP(P384C) and apo-AP(G403C) mutants were increased to 75 °C and 77 °C, respectively (Fig. 8b). Finally, the double mutant apo-AP(R34C/G403C) exhibited a broad melting transition centered around 73 °C, between those of apo-WT AP and apo-AP(G403C) (Fig. 8b). This suggests stabilization by the disulfide cross-link at position 403 in the protruding loop, combined with a partially offsetting destabilization by the mutation (whether disulfide-bonded or not) at position 34. All six apo-AP proteins (the wild type and five mutants) recovered most of their original signal in the far-UV region after being cooled to 20 °C (data not shown), suggesting that all the mutants unfold reversibly, as has been shown to be the case for the wild-type protein.[43]

Discussion

Slipknots in known protein structures

Slipknots present a variation on knotted proteins. Like knotted proteins, they bring up questions about

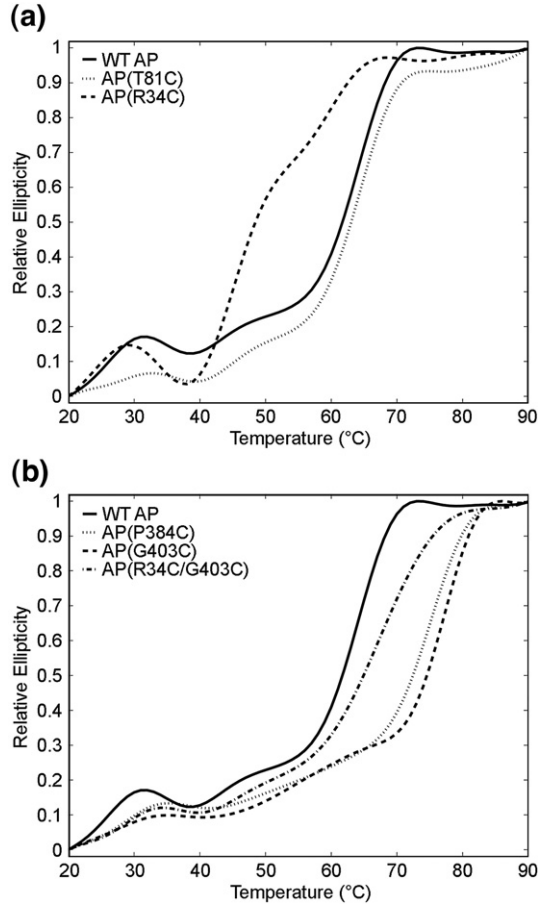

Fig. 8. Thermal stabilities of wild-type and mutant apo-APs. (a) The thermal melting curves of the two nonknotted mutants are plotted along with apo-WT AP (solid curve). The thermal stability of apo-AP(T81C) (dotted curve) is comparable to apo-WT AP, while apo-AP(R34C) (dashed curve) exhibits two transitions, both of which are lower than that of apo-WT AP. (b) The thermal melting curves of the three mutant proteins containing knotting disulfides are plotted along with apo-WT AP (solid curve). The thermal stabilities of the single knotted mutants apo-AP(P384C) (dotted curve) and apo-AP (G403C) (dashed curve) are both significantly higher than that of apo-WT AP, while apo-AP(R34C/G403C) (dash-dot curve) exhibits an intermediate melting transition. The raw CD data for each curve were normalized to the minimum and maximum CD signal on the temperature range for the purposes of comparison. All samples contained 0.25 mg mL^{-1} protein in 0.05 M metal-free Tris (pH 8.0).

folding pathways and how proteins find their native configurations when confronted by topological barriers. Slipknots are not identified by straightforward algorithms that check for knots in intact structures; partial structures within the protein chain must be examined. We find that, like knots, slipknots are very rare in natural proteins. In fact, among soluble proteins, only the fold adopted by alkaline phosphatase and others members of its broad family was judged to have a very deep slipknot. Somewhat shallower slipknots were found in a few other soluble proteins, including two viral proteins, while deep slipknots were found in two transmembrane proteins.

A common motif among soluble slipknotted proteins

The three most significant slipknots found in soluble proteins share a common motif of five β-strands with nearly conserved order and direction (Fig. 4). This observation raises the question of whether this slipknot motif is the result of divergent or convergent evolution. The three proteins bearing this motif do not show any statistically significant sequence similarity to each other, and their structures are quite different in the regions outside the knot motif. This suggests that the cases observed probably arose independently by convergent evolution. One explanation for the independent evolution of similar five-stranded slipknots would be architectural parsimony. A motif with five elements may be the simplest structure that can satisfy the topological requirements for slipknot formation. However, this hypothesis would not necessarily explain the observed conservation of β-sheet architecture; a five-helix slipknot could also be imagined. The observed preference for slipknots based on a β-sheet architecture may relate to preferred folding pathways in these proteins, which have not yet been elucidated.

Experimental and computational folding studies on the slipknotted proteins described here could be used to evaluate hypotheses about how topologically complex proteins fold. Recent simulations and experiments on the knotted protein YibK have suggested that specific nonnative interactions are necessary for the successful folding of that protein.[20,21,44] It would be interesting to see if simulations of slipknotted proteins behave similarly. The protein AFV3-109,[45] a homolog of protein B116 from STIV that lacks disulfide bonds, might be a good candidate for such a study due to its small size.

Implications for the folding pathways of membrane proteins

It is notable that among the few proteins with slipknots, two are transmembrane proteins. This is particularly remarkable in view of the relative paucity of transmembrane proteins in the PDB. The observations may suggest that, in comparison to soluble proteins, it is easier for transmembrane proteins to fold into slipknot structures. This might relate to the mechanisms and environments in which these two different classes of proteins fold.

For soluble proteins, whether one is considering knotted or slipknotted architectures, one envisions the protein chain having to thread its way through an opening in some potentially constricted structure formed by the rest of the protein. Whether or not

significant energy barriers might be encountered, such a pathway to the folded state would seem to be relatively confined and potentially costly in an entropic sense. The driving force for folding along such a threading pathway is harder to imagine than for a pathway that more nearly resembles a collapse from an unfolded configuration. But transmembrane proteins fold by mechanisms that are inherently different from those followed by soluble proteins. Those differences allows for speculative ideas regarding the formation of slipknots in transmembrane proteins. Transmembrane proteins composed of alpha helices may be inserted into the membrane in sequential fashion. For soluble proteins, insertion of part of the protein chain through the core would seem to be difficult in general, whereas for transmembrane proteins, insertion of the protein chain into the core may be a natural consequence of the membrane insertion mechanism. The role of insertion as a fundamental part of transmembrane protein folding mechanisms might explain why slipknots are apparently not so rare in those proteins. Following helix insertion, it is likely that subsequent movement of transmembrane helices around each other would be required in order to form a slipknot. If this is a reasonable representation of how slipknots can be formed in transmembrane proteins, it argues for a high degree of fluidity of helices in the membrane during the folding process. Finally, it is notable that only slipknots, rather than true knots, have been found in transmembrane proteins so far. The reasons for this are not yet clear, but it may be that having protein synthesis machinery that inserts proteins into the membrane while remaining confined to one side makes it permissible to form a slipknot, but difficult or impossible to form a true knot. These possibilities should be viewed as hypotheses to be tested.

Evidence for stabilizing effects of slipknots

Because deep knots and slipknots in proteins are rare, and because they seem to present challenges for folding, it is natural to suspect that in cases where they occur they might provide certain advantageous properties, such as additional stability or rigidity. Such arguments have been made before in the context of knotted proteins.[9,11,46] The present findings regarding slipknots are consistent with those ideas. The deepest slipknot found is in alkaline phosphatase, a protein noted for its extreme thermal stability. Much of this stability is known to derive from metal binding and disulfide bonding, but the preliminary protein engineering experiments reported here suggest that the slipknot feature also plays a role in stability. In particular, two different mutations that enhanced the depth of the knot on the C-terminal loop side both stabilized the protein against thermal denaturation. Two other mutations examined, which either enhanced the N-terminal extension or were intended to serve as a control, tended to destabilize the protein. Engineered disulfide bonds have been used extensively to stabilize proteins and to probe their folding pathways.[47,48] Interpreting the results of protein engineering experiments can be challenging, however, as highlighted by the present study. Additional experiments on alkaline phosphatase and other knotted and slipknotted proteins will be necessary to arrive at definitive conclusions and quantitative estimates of the effects these topological features have on protein stability. Folding and engineering experiments have not yet been conducted on the other two soluble slipknot proteins identified. It is notable, however, that both are viral proteins, for which extra stability might be an important advantage. The STIV protein B116 is particularly interesting in that the host species is thermophilic, requiring the virus to survive in extreme conditions.

Methods

Strains and plasmids

Escherichia coli strain SM547 (Δ(*phoA–phoC*), *phoR*, tsx: :Tn5, Δ*lac*, *galK*, *galU*, *leu*, str^r) was a gift from J. Beckwith. The deletion of the *phoA* gene ensured that plasmid-borne alkaline phosphatases were expressed and purified without contamination from endogenous wild-type protein. Plasmid pEK154 containing the wild-type *phoA* gene with its natural, phosphate-repressible promoter in pUC119 was a gift of E. Kantrowitz. Plasmids carrying the mutant alkaline phosphatases used in this study were constructed using plasmid pEK154 as a template and the QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's protocol. The mutated genes in all plasmids were sequenced to confirm that the desired mutations were present and no other mutations had been introduced.

Protein expression and purification

Wild-type and mutant alkaline phosphatases were expressed in *E. coli* strain SM547 and purified by osmotic shock, heat purification, ammonium sulfate precipitation, and anion exchange chromatography as previously described.[49] The final purification step was achieved using a Q-Sepharose HP column (Amersham Biosciences). A 100-mL gradient from 0 to 0.1 M NaCl in TMZP buffer (0.01 M Tris, pH 7.4, 0.001 M $MgCl_2$, 10^{-5} M $ZnSO_4$, 10^{-4} M NaH_2PO_4, 0.003 M NaN_3) was used to elute each protein. Purity of the proteins was assessed by SDS-PAGE.[50] Absorbance measurements at 280 nm using an extinction coefficient of 0.71 cm^2 mg^{-1} were used to determine the concentration of purified WT AP samples.[51] Concentrations of mutant APs were determined by the Bio-Rad version of Bradford's dye-binding assay[52] using WT AP as the standard.

Disulfide bond formation in mutant APs

Purified mutant APs were first dialyzed at 4 °C into TMZ buffer (0.01 M Tris, pH 8.0, 0.001 M $MgCl_2$, 10^{-5} M

$ZnSO_4$). The proteins were then dialyzed overnight at room temperature in TMZ buffer supplemented with 0.004 M $CuSO_4$ to facilitate disulfide bond formation. Because the engineered disulfides covalently cross-link the subunits of the AP dimer, the extent of disulfide bond formation could be assessed by SDS-PAGE under non-reducing conditions.[7]

Preparation of apoproteins

All purified APs were rendered metal free using two independent demetallation protocols in parallel, namely, extensive dialysis in either 0.05 M EDTA (pH 5.5)[53] or 2 M $(NH_4)_2SO_4$ (pH 9.0).[54] The apoproteins were then dialyzed extensively into metal-free 50 mM Tris (pH 8.0) in preparation for CD experiments. The efficacy of both demetallation protocols was checked by inductively coupled plasma mass spectrometry analysis, which revealed only ~2% occupancy of zinc binding sites and ~6% occupancy of the magnesium binding site in apoprotein samples prepared using the EDTA method (essentially no Cu^{2+} was detected). Roughly twice the amounts of each metal were found in the apoprotein samples prepared by ammonium sulfate dialysis. Standard steps to prevent metal ion contamination were taken when handling apoproteins, such as avoiding glassware and pretreating plasticware with acid followed by extensive rinsing with metal-free buffer.

Circular dichroism

CD measurements were recorded using a Jasco J-715 spectrometer with a Peltier temperature-controlled cell holder (Jasco). All samples contained 0.25 mg mL^{-1} apoprotein in 50 mM Tris (pH 8.0), which had been pretreated by filtration through Chelex100 resin to remove trace metals. Wavelength scans were conducted from 260 to 200 nm with a bandwidth of 1 nm and a scan rate of 50 nm min^{-1}. Thermal denaturations were performed by monitoring the CD signal at 222 nm with a bandwidth of 2 nm while increasing the temperature linearly from 20 °C to 90 °C at a rate of 1 °C min^{-1}. After the thermal denaturations were complete, the samples were subjected to a second far-UV CD scan at 90 °C to observe the spectra of the unfolded proteins before rapid cooling to 20 °C. Refolding was allowed for 10 min at 20 °C before another wavelength scan was conducted to judge whether unfolding was reversible for each mutant. CD experiments on apo-APs prepared using two independent demetallation protocols (see above) yielded identical results; plots generated using the proteins demetallated with EDTA are shown in this report.

Protein knot and slipknot calculations

The knottedness of a protein structure was determined by representing the three-dimensional path of the backbone as a planar diagram, which is a list of crossings and their orientations for a two-dimensional curve, after which the planar diagram was used to evaluate the Jones polynomial, which defines the type of knot (or unknot). First, a series of computational steps are required to convert the three-dimensional coordinates of the protein backbone to a planar diagram. The steps include backbone smoothing, end joining (to form a closed curve), and planar projection. Initial backbone smoothing reduces unnecessary crossings, which tend to occur in projections of an alpha helix, for example. The backbone smoothing is performed as described earlier.[24] End joining is required to generate a closed curve from an open curve. Both probabilistic and deterministic approaches to end joining have been employed. In probabilistic approaches, multiple possible end joinings are considered, such as would arise by extending the termini in many possible directions to an outer sphere.[13] In the deterministic approaches, the ends are joined by some algorithm based on the general notion that the extensions that join the ends should not penetrate through the protein interior. One deterministic approach that has been used is to project the ends directly away from the protein center of mass.[18,19] We have implemented a strategy that is a slight variation on that. To avoid incorrect knot assignments as much as possible, we extend the ends of the protein chain as if they are repelled from the protein surface, until they reach an outer sphere. To accomplish this, a series of small steps are taken, with each subsequent step chosen to be in a direction that maximizes the resulting distance from the protein. This is continued until an outer sphere larger than the protein is reached, after which the ends are joined essentially along a great circle on that sphere. Especially when partial structures are considered, the artificial termini are occasionally disposed so as to make extension to the outside problematic, such as when a terminus of a partial structure points into an interior space. In our implementation, we keep track of such cases and do not make a determination of the Jones polynomial. This helps to avoid problems with false-positive knots in partial structures. After a closed curve is defined, a planar diagram is obtained by projecting the closed curve into two dimensions.

From the planar diagram, the Jones polynomial is calculated according to a recursive procedure.[55] An example of how this proceeds is illustrated in Fig. 9 for a trefoil knot. Briefly, the Jones polynomial for a particular knot is related to the Jones polynomial for two other knots that would be obtained from the first one by application of the two Conway operations—flipping and smoothing—at a given crossing point (Fig. 9). In a two-dimensional drawing of a closed curve, the flipping operation takes the curve segment that passes underneath at some crossing point and causes it to cross in front instead. The smoothing operation is analogous to DNA recombination; it alters the connectivity of the curve at the crossing point in question. The Jones polynomials for the three knots—the original one and the two arising from flipping and smoothing—are related mathematically by a so-called skein relationship: $t^{-1} \times X(K_+) - t \times X(K_-) = (t^{1/2} - t^{-1/2}) \times X(K_0)$, where $X(K)$ is the Jones polynomial (written in terms of the variable t) for the knot K, and K_- and K_0 are the knots generated from the knot K_+ by flipping and smoothing, respectively, at some crossing point. A further relationship describes the joining of a circle to another knot: $X(K+O) = (-t^{1/2} - t^{-1/2}) \times X(K)$. Finally, the Jones polynomial for the unknot (the circle) is 1. By repeated application of these relationships, the Jones polynomial for a given curve can be determined by branching operations that lead eventually to unknots at the terminal nodes. Figure 9 shows how the Jones polynomial for the left-handed trefoil is determined to be $t^{-1} - t^{-3} - t^{-4}$.The program for calculating the Jones polynomial from a planar diagram was validated by applying it to knots whose planar diagrams and Jones polynomials have been tabulated†, including all 165

† http://katlas.math.toronto.edu

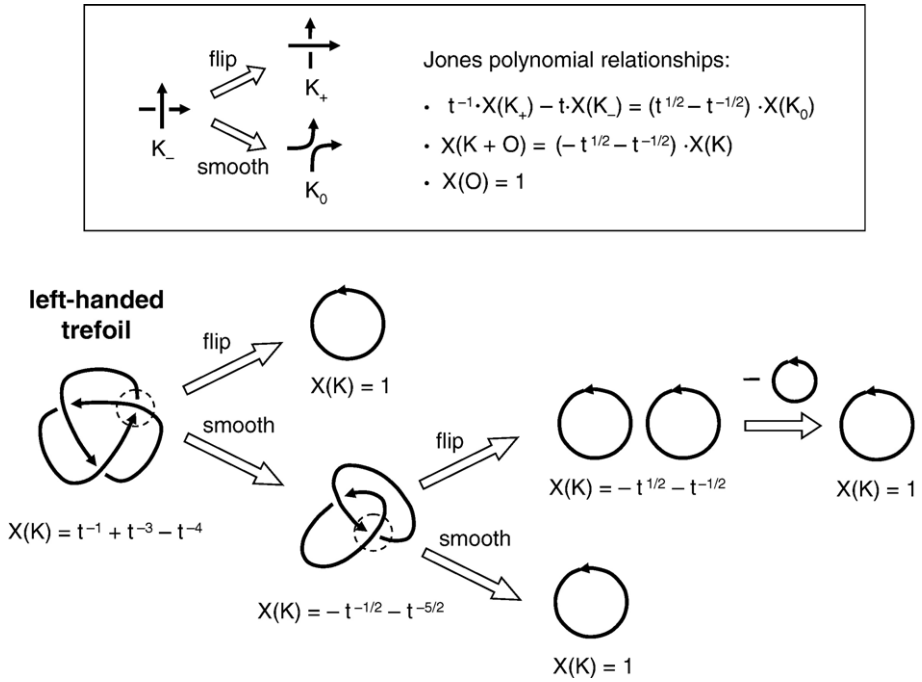

Fig. 9. An illustration of knot classification according to the Jones polynomial. $X(K)$ is the Jones polynomial for the knot K, written in terms of the symbol t. The diagram and equations in the box describe the relationships obeyed by the polynomials for three knots (or links), K_+, K_-, and K_0, which differ from each other only by their geometry at a single crossing. The trefoil example (bottom) shows how a knot can be simplified by flipping and smoothing operations to unknots. The crossings at which flipping and smoothing are performed are indicated by dotted circles. Working backwards from the terminal unknots, the equations shown (top) can be used to establish the Jones polynomial for the original knot. The Jones polynomial for the left-handed trefoil shown is $t^{-1}+t^{-3}-t^{-4}$.

distinct knots whose minimum crossing number is 10. The time required to determine whether a protein of approximately 300 residues was knotted was about 0.5 s on a single 3.8-GHz Intel Pentium IV processor.

In order to identify and characterize slipknots in proteins, knot calculations were performed on partial structures defined by choosing the beginning and ending residue positions for the protein chain. In the complete survey of the PDB, the N and C termini were truncated individually but not together. In other words, only the edges of the square knot matrix described in the text were sampled in the PDB survey. Calculation of the edges of the knot matrix required approximately 5 min per protein chain on average on a single Intel Pentium processor, but the time was strongly dependent on protein size. In cases where significant slipknots were identified, a full calculation of the knot matrix was performed, which involved a pairwise sampling of the possible N- and C-terminal residue positions for each fragment. In all cases, the positions of the fragment termini were sampled at two residue intervals. This sampling helped limit the intensity of the computations and was judged to be fine enough to capture the important topological properties of the chain.

Acknowledgements

The authors thank Martin Phillips for assistance with CD spectrometry, Tom Holton, Jason Nerothin, and Alex Lisker for computational assistance, and Todd Norcross, Douglas Rees, and Katelyn Connell for helpful discussions. This work was supported by the BER program of the Department of Energy Office of Science.

References

1. Crippen, G. M. (1974). Topology of globular proteins. *J. Theor. Biol.* **45**, 327–338.
2. Connolly, M. L., Kuntz, I. D. & Crippen, G. M. (1980). Linked and threaded loops in proteins. *Biopolymers*, **19**, 1167–1182.
3. Ivankov, D. N., Garbuzynskiy, S. O., Alm, E., Plaxco, K. W., Baker, D. & Finkelstein, A. V. (2003). Contact order revisited: influence of protein size on the folding rate. *Protein Sci.* **12**, 2057–2062.
4. 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.
5. Miller, E. J., Fischer, K. F. & Marqusee, S. (2002). Experimental evaluation of topological parameters determining protein-folding rates. *Proc. Natl. Acad. Sci. USA*, **99**, 10359–10363.
6. Bayro, M. J., Mukhopadhyay, J., Swapna, G. V., Huang, J. Y., Ma, L. C., Sineva, E. *et al.* (2003). Structure of antibacterial peptide microcin J25: a 21-residue lariat protoknot. *J. Am. Chem. Soc.* **125**, 12382–12383.
7. Boutz, D. R., Cascio, D., Whitelegge, J., Perry, L. J. & Yeates, T. O. (2007). Discovery of a thermophilic

protein complex stabilized by topologically interlinked chains. *J. Mol. Biol.* **368**, 1332–1344.
8. Duff, A. P., Cohen, A. E., Ellis, P. J., Kuchar, J. A., Langley, D. B., Shepard, E. M. *et al.* (2003). The crystal structure of *Pichia pastoris* lysyl oxidase. *Biochemistry*, **42**, 15148–15157.
9. Nureki, O., Shirouzu, M., Hashimoto, K., Ishitani, R., Terada, T., Tamakoshi, M. *et al.* (2002). An enzyme with a deep trefoil knot for the active-site architecture. *Acta Crystallogr. D*, **58**, 1129–1137.
10. Taylor, W. R. (2000). A deeply knotted protein structure and how it might fold. *Nature*, **406**, 916–919.
11. Wagner, J. R., Brunzelle, J. S., Forest, K. T. & Vierstra, R. D. (2005). A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature*, **438**, 325–331.
12. Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W. & Johnson, J. E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. *Science*, **289**, 2129–2133.
13. Mansfield, M. L. (1994). Are there knots in proteins? *Nat. Struct. Biol.* **1**, 213–214.
14. Mansfield, M. L. (1997). Fit to be tied. *Nat. Struct. Biol.* **4**, 166–167.
15. Grosberg, A. Y. & Nechaev, S. (1993). *Adv. Polym. Sci.* **106**, 1–29.
16. 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 transition state analog determined at 1.65 A resolution. *EMBO J.* **16**, 3405–3415.
17. Khatib, F., Weirauch, M. T. & Rohl, C. A. (2006). Rapid knot detection and application to protein structure prediction. *Bioinformatics*, **22**, e252–e259.
18. Lua, R. C. & Grosberg, A. Y. (2006). Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS Comput. Biol.* **2**, e45.
19. Virnau, P., Mirny, L. A. & Kardar, M. (2006). Intricate knots in proteins: function and evolution. *PLoS Comput. Biol.* **2**, e122.
20. Mallam, A. L. & Jackson, S. E. (2005). Folding studies on a knotted protein. *J. Mol. Biol.* **346**, 1409–1421.
21. Mallam, A. L. & Jackson, S. E. (2006). Probing nature's knots: the folding pathway of a knotted homodimeric protein. *J. Mol. Biol.* **359**, 1420–1436.
22. Helgstrand, C., Wikoff, W. R., Duda, R. L., Hendrix, R. W., Johnson, J. E. & Liljas, L. (2003). The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 A resolution. *J. Mol. Biol.* **334**, 885–899.
23. Blankenship, J. W. & Dawson, P. E. (2003). Thermodynamics of a designed protein catenane. *J. Mol. Biol.* **327**, 537–548.
24. Norcross, T. S. & Yeates, T. O. (2006). A framework for describing topological frustration in models of protein folding. *J. Mol. Biol.* **362**, 605–621.
25. Kim, E. E. & Wyckoff, H. W. (1991). Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. *J. Mol. Biol.* **218**, 449–464.
26. Alm, E. & Baker, D. (1999). Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. USA*, **96**, 11305–11310.
27. Taylor, W. (2005). Protein folds, knots and tangles. In *Physical and Numerical Models in Knot Theory* (JA, C., KC, M. & EJ, R., eds), pp. 171–202, World Scientific, Singapore.
28. Taylor, W. R., Xiao, B., Gamblin, S. J. & Lin, K. (2003). A knot or not a knot? SETting the record 'straight' on proteins. *Comput. Biol. Chem.* **27**, 11–15.
29. Alexander, J. W. (1928). Topological invariants of knots and links. *Trans. Am. Math. Soc.* **30**, 275–306.
30. Jones, V. (1985). A polynomial invariant for knots via von Neumann algebras. *Bull. Am. Math. Soc.* **12**, 103–111.
31. Wang, G. & Dunbrack, R. L., Jr (2003). PISCES: a protein sequence culling server. *Bioinformatics*, **19**, 1589–1591.
32. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D*, **60**, 2256–2268.
33. Wang, J., Stieglitz, K. A. & Kantrowitz, E. R. (2005). Metal specificity is correlated with two crucial active site residues in *Escherichia coli* alkaline phosphatase. *Biochemistry*, **44**, 8378–8386.
34. Gardberg, A., Shuvalova, L., Monnerjahn, C., Konrad, M. & Lavie, A. (2003). Structural basis for the dual thymidine and thymidylate kinase activity of herpes thymidine kinases. *Structure*, **11**, 1265–1277.
35. Larson, E. T., Eilers, B. J., Reiter, D., Ortmann, A. C., Young, M. J. & Lawrence, C. M. (2007). A new DNA binding protein highly conserved in diverse crenarchaeal viruses. *Virology*, **363**, 387–396.
36. Calamia, J. & Manoil, C. (1990). lac permease of *Escherichia coli*: topology and sequence elements promoting membrane insertion. *Proc. Natl. Acad. Sci. USA*, **87**, 4937–4941.
37. Sugiyama, J. E., Mahmoodian, S. & Jacobson, G. R. (1991). Membrane topology analysis of *Escherichia coli* mannitol permease by using a nested-deletion method to create mtlA–phoA fusions. *Proc. Natl. Acad. Sci. USA*, **88**, 9603–9607.
38. Taylor, W. R. (2007). Protein knots and fold complexity: some new twists. *Comput. Biol. Chem.* **31**, 151–162.
39. Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K. & Gouaux, E. (2007). Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature*, **445**, 387–393.
40. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. (2005). Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. *Nature*, **437**, 215–223.
41. Chlebowski, J. F. & Mabrey, S. (1977). Differential scanning calorimetry of apo-, apophosphoryl, and metalloalkaline phosphatases. *J. Biol. Chem.* **252**, 7042–7052.
42. Chlebowski, J. F., Mabrey, S. & Falk, M. C. (1979). Calorimetry of alkaline phosphatase. Stability of the monomer and effect of metal ion and phosphate binding on dimer stability. *J. Biol. Chem.* **254**, 5745–5753.
43. Dirnbach, E., Steel, D. G. & Gafni, A. (2001). Mg^{2+} binding to alkaline phosphatase correlates with slow changes in protein lability. *Biochemistry*, **40**, 11219–11226.
44. Wallin, S., Zeldovich, K. B. & Shakhnovich, E. I. (2007). The folding mechanics of a knotted protein. *J. Mol. Biol.* **368**, 884–893.
45. Keller, J., Leulliot, N., Cambillau, C., Campanacci, V., Porciero, S., Prangishvilli, D. *et al.* (2007). Crystal structure of AFV3-109, a highly conserved protein from crenarchaeal viruses. *Virol. J.* **4**, 12.
46. Taylor, W. R. & Lin, K. (2003). Protein knots: a tangled problem. *Nature*, **421**, 25.
47. Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987). Enhanced protein thermostability from site-directed

mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. USA*, **84**, 6663–6667.
48. Clarke, J. & Fersht, A. R. (1993). Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation. *Biochemistry*, **32**, 4322–4329.
49. Chaidaroglou, A., Brezinski, D. J., Middleton, S. A. & Kantrowitz, E. R. (1988). Function of arginine-166 in the active site of *Escherichia coli* alkaline phosphatase. *Biochemistry*, **27**, 8338–8343.
50. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
51. Stec, B., Holtz, K. M. & Kantrowitz, E. R. (2000). A revised mechanism for the alkaline phosphatase reaction involving three metal ions. *J. Mol. Biol.* **299**, 1303–1311.
52. Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* **72**, 248–254.
53. Applebury, M. L. & Coleman, J. E. (1969). *Escherichia coli* co (II) alkaline phosphatase. *J. Biol. Chem.* **244**, 709–718.
54. Gettins, P. & Coleman, J. E. (1983). 113Cd nuclear magnetic resonance of Cd(II) alkaline phosphatases. *J. Biol. Chem.* **258**, 396–407.
55. Sossinsky, A. (2002). *Knots.* Harvard Univ. Press, Cambridge, Massachusetts.
56. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, Palo Alto, California.