Knotting pathways in proteins

Joanna I. Sułkowska*†[1], Jeffrey K. Noel‡, César A. Ramírez-Sarmiento§, Eric J. Rawdon‖, Kenneth C. Millett¶ and José N. Onuchic‡

*Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland, †Center for Theoretical Biological Physics, University of California San Diego, 9500 Gilman Drive, San Diego, CA 92037, U.S.A., ‡Center for Theoretical Biological Physics, Rice University, 6100 Main Street, Houston, TX 77005, U.S.A., §Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Casilla 6553, Santiago, Chile, ‖Department of Mathematics, University of St. Thomas, 2115 Summit Avenue, St. Paul, MN 55105, U.S.A., and ¶Department of Mathematics, University of California Santa Barbara, 552 University Road, Santa Barbara, CA 93106, U.S.A.

Abstract

Most proteins, in order to perform their biological function, have to fold to a compact native state. The increasing number of knotted and slipknotted proteins identified suggests that proteins are able to manoeuvre around topological barriers during folding. In the present article, we review the current progress in elucidating the knotting process in proteins. Although we concentrate on theoretical approaches, where a knotted topology can be unambiguously detected, comparison with experiments is also reviewed. Numerical simulations suggest that the folding process for small knotted proteins is composed of twisted loop formation and then threading by either slipknot geometries or flipping. As the size of the knotted proteins increases, particularly for more deeply threaded termini, the prevalence of traps in the free energy landscape also increases. Thus, in the case of longer knotted and slipknotted proteins, the folding mechanism is probably supported by chaperones. Overall, results imply that knotted proteins can be folded efficiently and survive evolutionary pressure in order to perform their biological functions.

Introduction

For a long time, it was believed that knots were too complicated to exist in protein structures [1]. The situation changed in 2000 [2] when the first deeply embedded protein knot was discovered. Since then, knots and slipknots (slipknots arise from threading one loop through another, while the entire chain remains unknotted [3]) have been discovered in 2% of the proteins deposited in the PDB [4–7]. Although this number is significant, it is small in comparison with the ubiquitous knots in globular homopolymers [8,9]. This suggests that protein knots are limited either by the challenges of folding and misfolding or that knots have been systematically discriminated against by Nature for not providing an advantage to the organism [10]. Which case is true is a very intriguing and challenging question.

Recent results show that knotted motifs can be conserved across different families despite very low sequence similarity [11]. This suggests that these knots are conserved to preserve some functional advantage since their complex folding is likely to be disadvantageous for their host organisms. Currently, the functions of knots are not known. However, different types of stabilizing capacities of knots and slipknots have been suggested [7,11–15].

How complicated are protein knotting mechanisms? In the case of unknotted small or midsize proteins, it is known that they fold upon minimally frustrated funnel-like energy landscapes, which allow for fast and robust folding [16,17]. Native contacts play a dominant role in guiding these proteins to their native states in the range of microseconds to seconds. Relatively few complicated folding mechanisms have been proposed; examples are backtracking in interleukin 1β [18,19] or GFP (green fluorescent protein) [19a]. Thus, in principle, the low frequency of knotted proteins may be a consequence of the topological barrier to folding. Proteins that fold too slowly will be eliminated by evolution. Therefore it is important to understand how existing knotted proteins can find their native states. Are there multiple pathways possible for folding knots? Are native contacts [20] always sufficient to fold knotted proteins, and, if not, how are chaperones involved?

Currently, research is focused on three main aspects of knotted proteins: their evolutionary pathway, their tying mechanism and the function of the knotted topology [2,6,7,11,21,22]. In the present review, we summarize our present understanding of the tying process in proteins. Other reviews on the topic of knotted proteins can be found in [7,15,22,23], and, of particular relevance to the present review, the mechanism for untying is discussed in [23].

The folding mechanism of trefoil knotted proteins

The tying process has been most intensely investigated for the proteins YibK, from *Haemophilus influenzae*, and YbeA, from *Escherichia coli* [24], members of the SPOUT methyltransferase superfamily that conserve a deeply knotted trefoil at the C-terminus [11]. Recent *in vitro* results

Key words: artificial knot, chaperone, free energy landscape, knotted protein, protein folding, slipknotted protein.

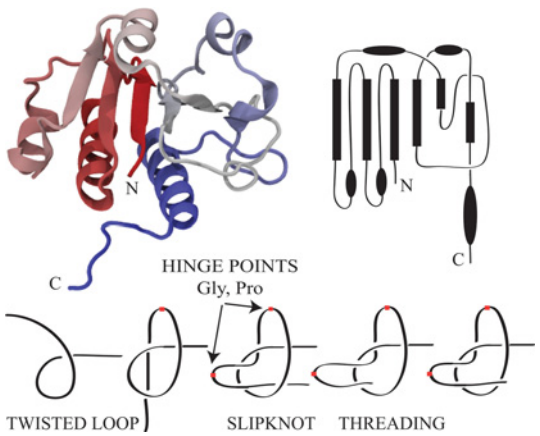
Abbreviations used: AOTCase, aspartate/ornithine carbamoyltransferase; SBM, structure-based model; UCH, ubiquitin C-terminal hydrolase.

[1]To whom correspondence should be addressed (email jsulkowska@chem.uw.edu.pl).

©The Authors Journal compilation ©2013 Biochemical Society

Figure 1 | Tying proteins through slipknots

Upper panel: cartoon representation of YibK (PDB code 1MXI), which forms a trefoil knot. Lower panel: knotting mechanism of YibK based on numerical simulations.



show that both YibK and YbeA can fold spontaneously in approximately 20 min [25]. However, chaperones are shown to significantly accelerate the knotting mechanism. The most important advance in this study was initiating folding from an unknotted and denatured protein since other studies have shown that the knotted topology persists in the denatured state [26] and during mechanical manipulations [12,27,28]. Comparison of folding times between unknotted [25] and knotted-unfolded [26] protein chains suggests that knotting is the rate-limiting step during folding.

Whereas experiments have shown unequivocally that isolated proteins can fold into knots, a structural explanation of the folding process is still beyond experimental resolution. The detailed structural information available in theoretical simulations is shedding considerable light on knotting mechanisms. Before experimental confirmation, it was shown in 2009 that native contacts are able to guide the folding process of YibK and YbeA [29] using SBMs (structure-based models) [20,30]. Their folding process comprises two main steps: a native twisted loop formation and the threading of the shorter terminus via a slipknot conformation (Figure 1). The low success rate (<5%) for reaching the native knotted configuration was explained as a consequence of deep energetic and topological traps. Knot threading was identified as a rate-limiting step for folding in these proteins since an unknotted construct achieved a much higher success rate (~73%). This slipknot folding mechanism was preferred over forming shallow knots that could be converted into native knots. This is surprising in the light of experimental results showing that flexible polymers (strings) can knot spontaneously [31].

Heterogeneous interactions in the knotted region were able to increase the folding rate of YibK [29,32]. It was shown that the slipknot intermediate and the twist (region closing the native knotted loop) is formed by the amino acids glycine (flexible) and proline (stiff). In a homogeneous SBM, these amino acids and all native contacts interact with the same strength. Adding flexibility to glycine, stiffening proline and introducing heterogeneity to the native contact interactions in the knotted region of the SBM increased the successful knotting rate [29,32]. In a different study, introduction of targeted non-native contacts between the C-terminus and the knotted loop significantly decreased the topological barrier and allowed YibK to fold with near 100% efficiency [29,32]. However, these contacts completely change the knotting mechanism from slipknotting to pulling (threading shallow knots). Appropriate interplay between non-native and native contacts can knot proteins even twice the size of YibK, as was observed for AOTCase (aspartate/ornithine carbamoyltransferase) [33].

The predicted tying mechanism (slipknot intermediate) has been seen in other models as well. The potential energy surface of truncated YibK [34] shows that the threading of the terminus through a twisted loop occurs late in the folding process, thereby creating a large energy barrier that serves as the rate-limiting step in the folding process. Interestingly, various unfolding simulations using Cα, all-atom SBM or explicit solvent models all showed that the slipknot conformation is commonly used to untie YibK [35] or AOTCase [14,33]. Folding a trefoil knot (which mimics YibK) on a lattice model [36] showed properties similar to those in a continuous model [29]. These results, along with those presented in the next section, imply that folding knots via slipknots is a robust feature of knotted protein energy landscapes.

The free energy landscape of the two smallest knotted proteins

In 2010, the smallest knotted protein (PDB code 2EFV) was found [37]. The protein 2EFV creates a trefoil knot located at least ten amino acids deep from both termini (Figure 2A). Using an SBM [38] based solely on the native contacts [29,33], 2EFV became the first knotted protein for which a free energy landscape, $F(Q)$ (Figure 2A), was determined [39]. The thermodynamics data revealed an intermediate with twisted loop formation and knot threading at the top of the free energy barrier. The transition state ensemble showed two folding routes: a plugging route and a slipknotting route. We present, as well, new data for VirC2, a protein that has the same fold as 2EFV, but with a deeper knot and only 33% sequence similarity. Figure 2 shows that the two proteins have analogous free energy profiles with clear pre-knotted intermediate states. The analysis shows the dominant effect of topology on the folding route.

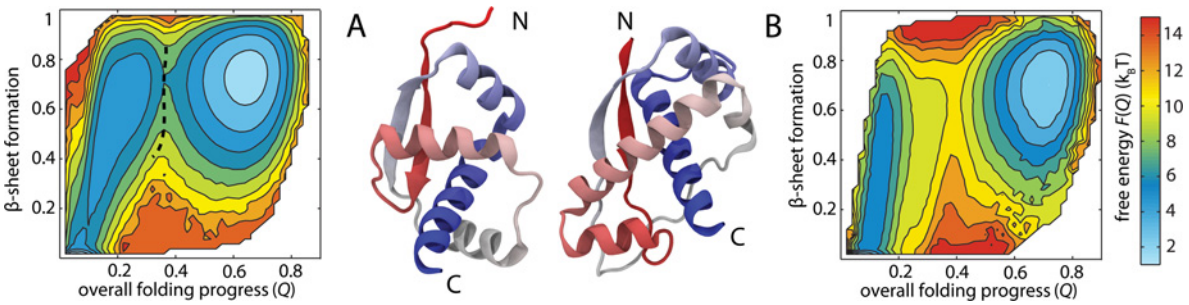
Artificial knots through protein design

The first detailed *in vitro* experimental characterization of equilibrium protein folding/knotting was achieved by designing a novel knotted protein based on genetic fusion of the tandem repeat of a gene of an unknotted dimeric protein [40]. This designed protein reversibly ties itself into

©The Authors Journal compilation ©2013 Biochemical Society

Figure 2 | Comparison of free energy landscapes of a structurally similar hypothetical protein MJ0366 (A) and VirC2 (B) (PDB codes 2EFV and 2RH3 respectively)

The proteins fold the β-sheet first, which defines the twisted-loop intermediate that precedes threading of the C-terminus. The broken line in (**A**) shows the contour where the probability of having a knot in the simulation is 50%. This indicates that the knot is threading at the top of the free energy barrier, the rate-limiting step to folding. The contours in (**A** and **B**) define steps of 1.5 k_BT.



the intended natively knotted configuration. Its folding time was shown to be 20-fold slower than that of a twin protein, which was designed to have an analogous tertiary structure, but to be unknotted. This reversibly foldable knotted protein is amenable to simulations, and showed close qualitative and quantitative agreement between experimental and theoretical results, suggesting further that SBMs (and simulations in general) are an important tool for explaining the tying process that is still hidden from experimental resolution. Simulation with SBMs was possible since the native topology was known from X-ray crystallography [40]. Key conclusions from the simulations were that (i) knotting occurs in the transition state ensemble, (ii) kinetics data reflect the experimentally observed rollover in the folding limbs of chevron plots, (iii) correct folding or tying is restricted to a narrow range of temperatures in comparison with the unknotted protein [11], and (iv) the presence of a dead-end intermediate that lacks the knot suggests the importance of backtracking events [41]. Moreover, the results [11] suggest that gene fusion coupled with evolutionary optimization of the stiffness and size of the knotted protein core is a feasible process for creating small knotted proteins that can fold reversibly.

Proteins with complex knotted fingerprints

Identified in 2010, the 6_1 knot in DehI [37] represents the most complex knot known [11]. Hints of how this protein can fold were obtained from SBM simulations. Even though the folding rate was very low (even lower than for YibK [29]), successful folding trajectories showed a simple tying process. DehI first folds two aligned unknotted loops then ties a knot by (i) threading the C-terminus through the smaller loop (through a slipknot conformation), and (ii) flipping the larger loop over the smaller loop, where the order of the steps (i) and (ii) can be reversed. Flexible regions with glycine and proline are involved in flipping the loop, the twisted arc and the temporary slipknotted configuration.

In 2006, a shallow 5_2 knot [7] was identified in the UCHs (ubiquitin C-terminal hydrolases) UCH-L1 and UCH-L3, whose function is the degradation of proteins. Available *in vitro* data on UCHs suggest reversible folding/knotting under chemical denaturation [42]. The kinetics suggest parallel pathways with two intermediate states. Such a landscape is plausible because of the very shallow knot that allows spontaneous knotting through thermal fluctuations after folding the core of the protein. However, verification of an unfolded/unknotted state is still required in order to confirm this interpretation of the kinetic data.

Natively slipknotted proteins [3,11], although untied when pulled from their termini [43,44], create a very interesting class of proteins with non-trivial folding processes. Currently, it is known that several slipknots in tandem can create complex knotted motifs that are conserved despite low sequence similarity [3,11]. Results suggest a reversible unfolding and folding process composed of multiple routes for alkaline phosphatase, a protein with a simple slipknot motif [3,11]. SBM simulations of a protein with the same slipknot motif, thymidine kinase, explained the folding process by a 'flipping mechanism', in which the slipknot formation is initiated by rotation of the slip-loop over the unknotted native core of the protein [29]. Glycine and proline are found in the hinge regions that probably support this rotation. The flipping mechanism represents the simplest way to fold a slipknot; however, the small number of successful folding events suggests participation of external help to overcome deep topological traps.

Analogous knotting mechanisms for other complex knotted proteins were suggested by considering sequence and knotted motif conservation [3,11]. It was shown that families with a conserved knotted fingerprint, despite having low sequence similarity, show a preference for the conservation of glycine in the hinge regions. These amino acids can facilitate folding and suggest a knotting mechanism composed of the steps observed in smaller proteins. Much shorter slipknotted proteins, such as those encoded by

©The Authors Journal compilation ©2013 Biochemical Society

crenarchaeal virus AFV3 (*Acidianus* filamentous virus 3) open reading frame 109, fold employing only the first three steps in Figure 1 [29]. This provides strong support for a common knotting process between knotted and slipknotted proteins.

Discussion

In summary, numerical simulations suggest a set of related folding mechanisms for knotted proteins. It consists of twisted loop formation, and then threading by either slipknot configurations or flipping motions. This general knotting mechanism is robust across all knotted or slipknotted protein topologies that can be untied in one step (which is the case for all known proteins) as suggested in [22]. Combination of experimental and theoretical results shows that small knotted proteins can be reversibly self-tied, whereas proteins with deep or complex knots may need some external guidance to fold, for example from chaperones [25]. The precise action of these chaperones remains unknown, but their general role would be to enable protein conformations to overcome topological traps and to facilitate the proposed folding mechanisms. In the future, detailed methods such as single-molecule fluorescence need to be incorporated into *in vitro* studies in order to strengthen the combined experimental and theoretical efforts to elucidate the driving force for tying knots in proteins.

Acknowledgements

J.I.S., E.J.R. and K.C.M. thank the Isaac Newton Institute for Mathematical Sciences for sponsoring the Topological Aspects of DNA Function and Protein Folding workshop and for hosting our stays at the Institute.

Funding

J.I.S., J.K.N. and J.N.O. were supported by the Center for Theoretical Biological Physics sponsored by the National Science Foundation (NSF) [grant numbers PHY-0822283 and MCB-1214457]. J.N.O. is a Cancer Prevention and Research Institute of Texas (CPRIT) Scholar in Cancer Research sponsored by CPRIT. J.I.S. was supported by the Fundacja na rzecz Nauki Polskiej (FNP) and the Foundation for Polish Science TEAM Project [grant number TEAM/2011-7/6]. E.J.R. was supported by the National Science Foundation [grant number 1115722]. C.A.R.-S. was supported by a Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) doctoral fellowship.

References

1 Mansfield, M.L. (1994) Are there knots in proteins. Nat. Struct. Biol. **1**, 213–214
2 Taylor, W.R. (2000) A deeply knotted protein structure and how it might fold. Nature **406**, 916–919
3 King, N.P., Yeates, E.O. and Yeates, T.O. (2007) Identification of rare slipknots in proteins and their implications for stability and folding. J. Mol. Biol. **373**, 153–166
4 Khatib, F., Weirauch, M.T. and Rohl, C.A. (2006) Rapid knot detection and application to protein structure prediction. Bioinformatics **22**, E252–E259
5 Lua, R.C. and Grosberg, A.Y. (2006) Statistics of knots, geometry of conformations, and evolution of proteins. PLoS Comput. Biol. **2**, e45
6 Taylor, W.R. and Lin, K. (2003) Protein knots: a tangled problem. Nature **421**, 25–25
7 Virnau, P., Mirny, L.A. and Kardar, M. (2006) Intricate knots in proteins: function and evolution. PLoS Comput. Biol. **2**, 1074–1079
8 Lua, R., Borovinskiy, A.L. and Grosberg, A.Y. (2004) Fractal and statistical properties of large compact polymers: a computational study. Polymer **45**, 717–731
9 Virnau, P., Kantor, Y. and Kardar, M. (2005) Knots in globule and coil phases of a model polyethylene. J. Am. Chem. Soc. **127**, 15102–15106
10 Grosberg, A. and Nechaev, S. (1993) Polymer topology. Adv. Polym. Sci. **106**, 1–29
11 Sułkowska, J.I., Rawdon, E.J., Millett, K.C., Onuchic, J.N. and Stasiak, A. (2012) Conservation of complex knotting and slipknotting patterns in proteins. Proc. Natl. Acad. Sci. U.S.A. **109**, E1715–E1723
12 Bornschlögl, T., Anstrom, D.M., Mey, E., Dzubiella, J., Rief, M. and Forest, K.T. (2009) Tightening the knot in phytochrome by single-molecule atomic force microscopy. Biophys. J. **96**, 1508–1514
13 Sayre, T.C., Lee, T.M., King, N.P. and Yeates, T.O. (2011) Protein stabilization in a highly knotted protein polymer. Protein Eng., Des. Sel. **24**, 627–630
14 Sułkowska, J.I., Sułkowski, P., Szymczak, P. and Cieplak, M. (2008) Stabilizing effect of knots on proteins. Proc. Natl. Acad. Sci. U.S.A. **105**, 19714–19719
15 Yeates, T., Norcross, T.S. and King, N.P. (2007) Knotted and topologically complex proteins as models for studying folding and stability. Curr. Opin. Chem. Biol. **11**, 595–603
16 Leopold, P.E., Montal, M. and Onuchic, J.N. (1992) Protein folding funnels: a kinetic approach to the sequence–structure relationship. Proc. Natl. Acad. Sci. U.S.A. **89**, 8721–8725
17 Onuchic, J.N., Luthey-Schulten, Z. and Wolynes, P.G. (1997) Theory of protein folding: the energy landscape perspective. Annu. Rev. Phys. Chem. **48**, 545–600
18 Chavez, L.L., Gosavi, S., Jennings, P.A. and Onuchic, J.N. (2006) Multiple routes lead to the native state in the energy landscape of the β-trefoil family. Proc. Natl. Acad. Sci. U.S.A. **103**, 10254–10258
19 Mallam, A.L. and Jackson, S.E. (2007) A comparison of the folding of two knotted proteins: YbeA and YibK. J. Mol. Biol. **366**, 650–665
19a Andrews, B.T., Capraro, D.T., Sułkowska, J.I., Onuchic, J.N. and Jennings, P.A. (2013) Hysteresis as a marker for complex, overlapping landscapes in proteins. J. Phys. Chem. Lett. **4**, 180–188
20 Clementi, C., Nymeyer, H. and Onuchic, J.N. (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J. Mol. Biol. **298**, 937–953
21 Potestio, R., Micheletti, C. and Orland, H. (2010) Knotted vs. unknotted proteins: evidence of knot-promoting loops. PLoS Comput. Biol. **6**, e1000864
22 Taylor, W.R. (2007) Protein knots and fold complexity: some new twists. Comput. Biol. Chem. **31**, 151–162
23 Sułkowska, J.I., Sułkowski, P., Szymczak, P. and Cieplak, M. (2010) Untying knots in proteins. J. Am. Chem. Soc. **132**, 13954–13956
24 Nureki, O., Shirouzu, M., Hashimoto, K., Ishitani, R., Terada, T., Tamakoshi, M., Oshima, T., Chijimatsu, M., Takio, K., Vassylyev, D.G. et al. (2002) An enzyme with a deep trefoil knot for the active-site architecture. Acta Crystallogr. Sect. D: Biol. Crystallogr. **58**, 1129–1137
25 Mallam, A.L. and Jackson, S.E. (2012) Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. Nat. Chem. Biol. **8**, 147–153
26 Mallam, A.L., Rogers, J.M. and Jackson, S.E. (2010) Experimental detection of knotted conformations in denatured proteins. Proc. Natl. Acad. Sci. U.S.A. **107**, 8189–8194
27 Dzubiella, J. (2009) Sequence-specific size, structure, and stability of tight protein knots. Biophys. J. **96**, 831–839
28 Sułkowska, J.I., Sułkowski, P., Szymczak, P. and Cieplak, M. (2008) Tightening of knots in proteins. Phys. Rev. Lett. **100**, 058106
29 Sułkowska, J.I., Sułkowski, P. and Onuchic, J. (2009) Dodging the crisis of folding proteins with knots. Proc. Natl. Acad. Sci. U.S.A. **106**, 3119–3124
30 Sułkowska, J.I. and Cieplak, M. (2008) Selection of optimal variants of Go-like models of proteins through studies of stretching. Biophys. J. **95**, 3174–3191

©The Authors Journal compilation ©2013 Biochemical Society

31 Raymer, D.M. and Smith, D.E. (2007) Spontaneous knotting of an agitated string. Proc. Natl. Acad. Sci. U.S.A. **104**, 16432–16437
32 Wallin, S., Zeldovich, K.B. and Shakhnovich, E.I. (2007) The folding mechanics of a knotted protein. J. Mol. Biol. **368**, 884–893
33 Skrbić, T., Micheletti, C. and Faccioli, P. (2012) The role of non-native interactions in the folding of knotted proteins. PLoS Comput. Biol. **8**, e1002504
34 Prentiss, M.C., Wales, D.J. and Wolynes, P.G. (2010) The energy landscape, folding pathways and the kinetics of a knotted protein. PLoS Comput. Biol. **6**, e1000835
35 Tuszynska, I. and Bujnicki, J.M. (2010) Predicting atomic details of the unfolding pathway for YibK, a knotted protein from the SPOUT superfamily. J. Biomol. Struct. Dyn. **27**, 511–520
36 Faisca, P.F., Travasso, R.D., Charters, T., Nunes, A. and Cieplak, M. (2010) The folding of knotted proteins: insights from lattice simulations. Phys. Biol. **7**, 16009
37 Bolinger, D., Sułkowska, J.I., Hsu, H.P., Mirny, L.A., Kardar, M., Onuchic, J.N. and Virnau, P. (2010) A Stevedore's protein knot. PLoS Comput. Biol. **6**, e1000731
38 Noel, J.K., Whitford, P.C., Sanbonmatsu, K.Y. and Onuchic, J.N. (2010) SMOG@ctbp: simplified deployment of structure-based models in GROMACS. Nucleic Acids Res. **38**, W657–W661
39 Noel, J.K., Sułkowska, J.I. and Onuchic, J.N. (2010) Slipknotting upon native-like loop formation in a trefoil knot protein. Proc. Natl. Acad. Sci. U.S.A. **107**, 15403–15408
40 King, N.P., Jacobitz, A.W., Sawaya, M.R., Goldschmidt, L. and Yeates, T.O. (2010) Structure and folding of a designed knotted protein. Proc. Natl. Acad. Sci. U.S.A. **107**, 20732–20737
41 Li, W., Terakawa, T., Wang, W. and Takada, S. (2012) Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. Proc. Natl. Acad. Sci. U.S.A. **109**, 17789–17794
42 Andersson, F.I., Pina, D.G., Mallam, A.L., Blaser, G. and Jackson, S.E. (2009) Untangling the folding mechanism of the 5_2-knotted protein UCH-L3. FEBS J. **276**, 2625–2635
43 Sułkowska, J.I., Sułkowski, P. and Onuchic, J.N. (2009) Jamming proteins with slipknots and their free energy landscape. Phys. Rev. Lett. **103**, 268103
44 He, C., Genchev, G.Z., Lu, H. and Li, H. (2012) Mechanically untying a protein slipknot: multiple pathways revealed by force spectroscopy and steered molecular dynamics simulations. J. Am. Chem. Soc. **134**, 10428–10435

Received 21 November 2012
doi:10.1042/BST20120342

©The Authors Journal compilation ©2013 Biochemical Society