Tightening of Knots in Proteins

Joanna I. Sułkowska,[1] Piotr Sułkowski,[2,3,4] P. Szymczak,[2] and Marek Cieplak[1]

[1]*Institute of Physics, Polish Academy of Sciences, Aleja Lotników 32/46, 02-668 Warsaw, Poland*
[2]*Institute of Theoretical Physics, University of Warsaw, Hoża 69, 00-681 Warsaw, Poland*
[3]*Sołtan Institute for Nuclear Studies, Hoża 69, 00-681 Warsaw, Poland*
[4]*Physikalisches Institut der Universität Bonn, Nussallee 12, 53115 Bonn, Germany*

(Received 13 June 2007; published 8 February 2008)

We perform theoretical studies of stretching of 20 proteins with knots within a coarse-grained model. The knot's ends are found to jump to well defined sequential locations that are associated with sharp turns, whereas in homopolymers they diffuse around and eventually slide off. The waiting times of the jumps are increasingly stochastic as the temperature is raised. Knots typically do not return to their native locations when a protein is released after stretching.

DOI: 10.1103/PhysRevLett.100.058106 PACS numbers: 87.15.A−, 87.10.−e, 87.14.E−, 87.15.La

Time and again, objects of nontrivial topology turn out to be relevant in physics. Polymers provide examples of such a relevance as they may acquire topologically nontrivial configurations known as knots [1–3]. In DNAs—polymers which are nearly homogeneous—knots arise spontaneously and abundantly under bad solvent conditions and for large system sizes [4–6]. In proteins, however, they are a rarity. Knots in the native states of proteins were first discovered by Mansfield in 1994 [7]. Further research [8–10], and especially a survey by Virnau *et al.* [11], has led to an identification of 273 examples of proteins with knots, which constitutes less than 1% of the structures deposited in the Protein Data Bank. The biological function of knots in proteins remains to be elucidated, but it is likely that such shapes are not accidental. It should also be noted that these 273 proteins correspond to only three different topologies denoted as 3_1 (the trefoil knot), 4_1, and 5_2, where the main integer indicates the number of crossings and the subscript a particular shape. (To identify a knot properly from a mathematical perspective, it is assumed implicitly that the protein terminals are connected by an outside segment that transforms an open chain into a closed loop.)

In this Letter, we explore the dynamical behavior of a knot when a protein is stretched, for example, by a tip of an atomic force microscope. Experiments [12] and all-atom simulations [13] on a knotted protein have been performed recently for the bovine carbonic anhydrase (coded 1v9e), and the manipulation involved *unwinding* of the knot. Our study considers *tightening* of knots and is based on molecular dynamics simulations in a coarse-grained model that represents a protein as a chain of the C^α atoms with effective attractive contact interactions [14,15]. In contrast to the all-atom simulations, a coarse-grained approach allows for a survey of many proteins, incorporation of much larger statistics, slower rates of pulling, and extensive variation of parameters.

We observe that the knot tightening process in a stretched protein is dominated by jumps, i.e., sudden displacements of positions of the knot's ends along the sequence toward each other. These jumps have definite lengths and together with the final location of a tightened knot they are specified by a local geometry of a protein chain. The larger the size of a knot or its topological complication, the larger the number of jumps is observed before its final tightening. However, such jumps are not observed in the dynamics of knot motion on stretched polymers. In this case, a motion is of a diffusive character [4,16,17].

In order to define the knotted core, i.e., a minimal segment of amino acids that can be identified as a knot, we use the Koniaris-Muthukumar-Taylor algorithm [1,8]. It involves removing the C^α atoms, one at a time, as long as the backbone does not intersect a triangle set by the atom under consideration and its two immediate sequential neighbors. As a result of this procedure, two end points of the knot are identified. The knot's ends depend on the conformation and, as the protein gets stretched, they may depin and come closer together. We have studied 18 proteins with the trefoil knot 3_1 (1j85, 1o6d, 1dmx, 1jd0, 1j86, 1ipa, 1js1, 1k3r, 1kop, 1nxz, 1v9e, 1x7p, 1v2x, 1fug, 1vh0, 1zrj, 1hcb, and 1keq) and two 5_2 proteins (2etl and 1xd3) [11]. We have found that once the knot shrinks from its native size, one end of a knot invariably lands in a sharp turn of a protein backbone. Then it moves again until a final position corresponding to the tightest knot is reached. In most cases, such turns contain proline which stiffens a backbone through a ring structure that forms a backbone angle $\sim$75 deg. The second frequent knot-stopping turn contains glicyne (in 1o6d, 1fug, 1vho, 1zrj, 1keq, and 1v9e, the latter also has a turn with proline) which, due to the lack of the side chain, leads to strongly sinuous local conformations of the backbone. In one case (1hcb), the knot-stopping turn involved alanine. In the absence of a sharp turn in a protein backbone, the knot is stopped at the beginning of a helix. It should be noted that proteins with knots have a shorter effective end-to-end length available for stretching, which is similar to the case of proteins with covalent disulfide bonds between cysteins (not present in the proteins considered here). However, there are also

important differences between the two: disulfide bonds stay in place whereas knots may move.

The details of our modeling of stretching are described in Refs. [18,19]. Native contacts are defined through heavy atom overlaps and are assigned the Lennard-Jones potentials with an amplitude ϵ and length parameters tuned in such a way as to guarantee that the native conformation of a protein corresponds to the global minimum of potential energy. The remaining non-native contacts are repulsive. We take $\epsilon/k_B = 900$ K, which correlates well with the experimental data on protein unfolding ($T = k_BT/\epsilon \sim 0.3$ corresponds to the room temperature). Unlike Wallin *et al.* [20] who consider *folding* of protein 1j85, we do not need to introduce additional non-native attractive contacts leading to a knot formation, since our configurations are already knotted.

The presence of a solvent is mimicked by velocity dependent friction and fluctuational forces corresponding to a temperature T. The stretching was accomplished by attaching the protein to a pulling spring which moves with the velocity v_p of 0.005 Å/ns. Our approach and its variants passed many benchmark tests for protein stretching and agrees favorably with the experimental results [19] and all-atom molecular dynamics simulations [21]. In particular, predictions of our model are consistent with studies of knot unwinding in 1v9e [12,13], as discussed in the supplementary material [24].

In order to represent motion of the ends of a knot, we use diagrams such as the one shown in the middle panel of Fig. 1. The panel corresponds to protein 1j85 which contains $N = 156$ amino acids that make a simple trefoil knot with the ends set at amino acids numbered $n_1 = 75$ and $n_2 = 119$ in the native state. The diagram shows what happens to the values of n_i when the protein gets stretched as the pulling tip moves by a distance d and the corresponding $F(d)$ curve for a protein (the right panel in Fig. 1). It is seen that despite the presence of several force peaks the ends of the knot stay put for most of the stretching trajectory. However, at the final force peak, i.e., around $d = 400$ Å, both ends jump toward each other along the sequence and then undergo another jump about 50 Å later. This jumpy behavior is not found when the protein is heated up or replaced by a homopolymer with purely repulsive contact interactions (the left panel in Fig. 1). In the homopolymeric case, we start with the native conformation of a protein, but remove attractive contacts. Another possibility of observing homopolymerlike behavior in a protein is to increase the temperature of the system above that of the specific-heat maximum. In a homopolymer, the positions of the knot ends diffuse around and, particularly in the initial stages, the distance between them may increase considerably which corresponds to swelling of the knot. Eventually, however, they come closer together but remain mobile and, in most cases, slide off the polymer chain. These results agree with earlier studies on the dynamics of knots in polymers and DNA, in which the diffusive character of knot motion was analyzed both experimentally [4] and theoretically [16,17].

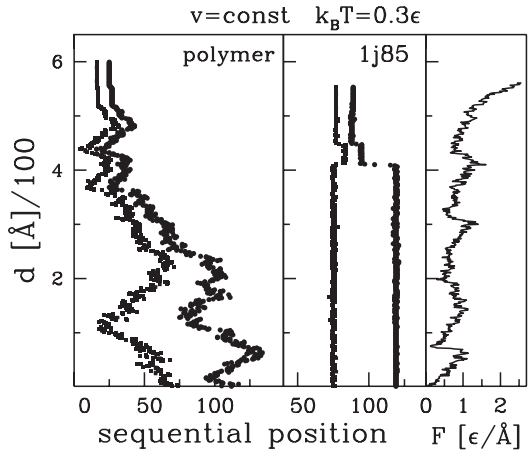

FIG. 1. Motion and tightening of a knot on a homopolymer (first panel) and on protein 1j85 (second panel) during stretching with constant velocity. Squares and circles indicate positions of the ends of the knot along the chain. Knots typically slide off homopolymeric chains. Here, however, we have chosen a less frequent example in which a knot tightens close to one end of the chain and may remain immobile in the absence of an adequate thermal activation [17,22,23]. In contrast, knots in proteins always tighten in a specific position inside its initial configuration, after a series of jumps. Each jump corresponds to a definite force peak in the force-displacement curve shown in the third panel. Reducing the pulling speed by an order of magnitude makes only small shifts in the curves shown in all panels. A video pertaining to knot tightening is available in the supplementary material [24].

Both for the homopolymer and the protein, the motion of the knot's ends depends on the particular trajectory even if the $F(d)$ curves look nearly the same. In particular, the ends may sometimes depin on an earlier force peak. The stretching process affects the knotted core of a protein much less than the outside region and thus leaves the geometry inside the knotted core and its secondary structures nearly nativelike. For instance, a well tightened knot in 1o6d contains an entire α helix in its nearly native conformation.

The description of a knot dynamics is reduced and involves only the movement of its end points n_i along the sequence. We have found, however, that the real space distances between the residues in the knotted core turn out to be mostly unchanged in between the knot jumps and undergo rapid changes as the knot ends jump. This indicates the existence of a coupling of the real space dynamics of a knot to its motion in the sequence space.

The final and metastable locations of the knot ends coincide with the sharp turns in the protein backbone (and/or the end points of a helix), as seen in Figs. 2 and 3. The stopping points correspond to the deep local minima of the angle θ between every second vector along the C_α backbone (i.e., between the vector $C_{\alpha,i}C_{\alpha,i+1}$ and

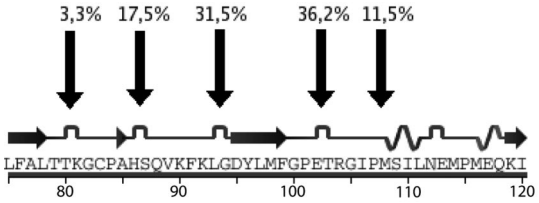

FIG. 2. The ends of a knot in the 1j85 protein in the native state are located at amino acids $n_1 = 75$ and $n_2 = 119$. In a tightened configuration, the ends of the knot are located between n_1 and n_2, with one end either in a sharp turn or at the end of a helix. The arrows indicate these characteristic places. The numbers show percentages of situations (based on 700 trajectories) in which a knot's end is pinned at the feature after moving from the native state. The innermost features correspond to the tightest knot.

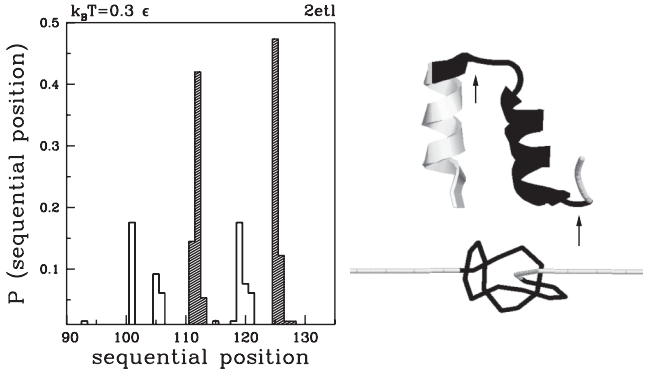

FIG. 3. The preferred final locations of a knot's ends in the 2etl protein found in 700 trajectories. The darker peaks indicate the most likely outcome: $n_1 \sim 111$ (the end of a helix) and $n_2 \sim 126$ (a turn). The corresponding tightened knot conformation is shown on the bottom right. The relevant sequential segment is shown on the top right in the native conformation where the arrows indicate the values of n_i. The less probable outcomes are shown by the lighter peaks. Here $n_1 \sim 101$ or 106, whereas $n_2 \sim 119$.

$C_{\alpha,i+2}C_{\alpha,i+3}$), which coincides with Kuntz's criterion [25] for detection of turns (and is also satisfied at the end points of a helix). Such turns are usually stabilized by hydrogen bonds and are thus harder to break. At high temperatures ($kT > 0.5\epsilon$), the motion of a knot gradually becomes less predictable, and the final position of the knot ends is no longer always connected to the turn in the native structure. Additionally, the knot may wander outside the initial knotted core. Finally, for $kT \gg \epsilon$, a homopolymeric behavior is observed, with the knot freely diffusing along the backbone.

A protein typically contains several sharp turns in the native state. Thus there are several pinning centers on which the knot's ends may settle during stretching. This is illustrated in Fig. 2 for the 1j85 protein. Another example is given in Fig. 3 for the 2etl protein which supports a 5_2 knot spanning 174 (out of all 223) sites in the native state. In this case, there are two characteristic pinning centers leading to the final knot tightening either between sites 110–126 or 101–119 for a range of temperatures. It should be noted that the preference for a knot to begin or end on a turn does not appear to apply to the native conformation. It arises only during stretching.

In addition to the simple stretching (whether at a constant speed or at a constant force), we have also studied processes in which one pulls a protein to a certain extension and then releases it abruptly. If the stretching stage lasts sufficiently long (so that several force peaks are observed and the knot gets tightened substantially) then the protein misfolds on releasing and the knot ends continue to reside at the metastable locations. We have observed such irreversibility effects in 2etl ($N = 223$), 1vho ($N = 157$), and 1v2x ($N = 191$) and in 80% of trajectories for 1o6d ($N = 147$). However, apart from a few trajectories (such as the one shown in Fig. 4), the knot in protein 1j85 ($N = 156$) is usually found to return to its native location, even though it is longer than 1o6d and of a similar length as 1vho. Thus this irreversibility appears not to be (strongly) related to N. The different behavior of 1j85 compared to the other four proteins may be due to the fact that 1j85 has low equilibrium stability [26] and it easily unfolds (and unties itself) through heating [20].

We now consider a distribution of waiting times, δt, between the jumps. In fact, it is convenient to measure these times in terms of a respective displacement of the pulling tip $\delta d = v_p \delta t$ (in addition, pulling distances corresponding to jumps are only weakly sensitive to the choice of v_p). At $T = 0$, the process is deterministic, lasts for a relatively long time, with δd reaching 400 Å before the first (and only) jump is made. At the time of the jump, the knotted core constitutes the only portion of the original protein structure that has not been unfolded yet. This unfolding route is denoted as pathway 1 and corresponds to the rightmost peak in the top and middle panels of Fig. 5. As the temperature is increased an alternative pathway 2

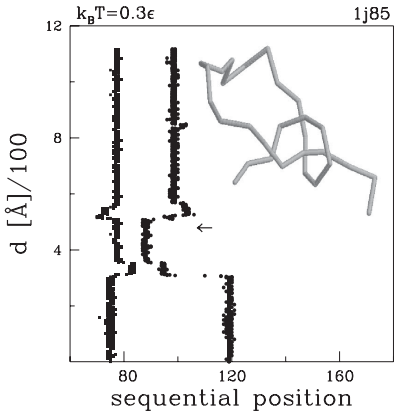

FIG. 4. After terminating the pulling process at $d = 500$ Å (indicated by the arrow) 1j85 returns to its native state in most cases. Sometimes, it ends up in a metastable state (such as shown on the right).

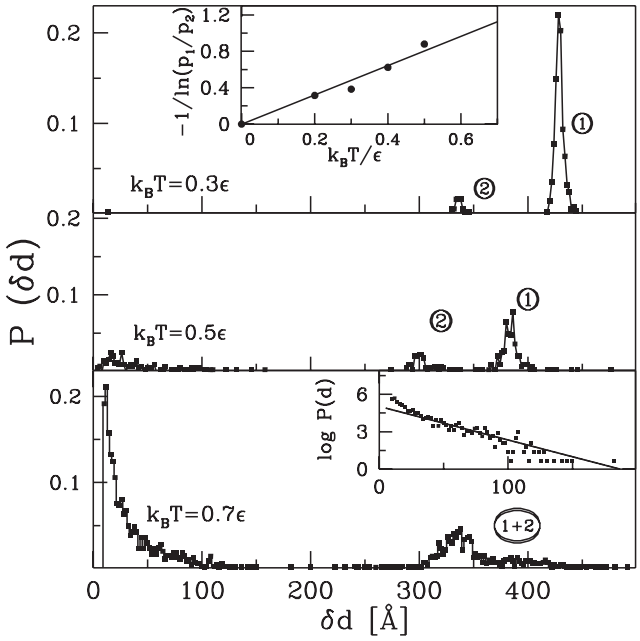

FIG. 5. Distribution of waiting distances δd of the left end of a knot in protein 1j85 at various temperatures T. Pathways 1 and 2 are indicated by symbols in circles. The panels are explained in the text.

becomes stochastically available. In this pathway the knot is tightened at $\delta d \sim 340$ Å, which is before the protein gets fully unfolded. The ratio of probabilities of choosing these pathways can be then described as

$$\frac{p_1}{p_2} = \exp\left(-\frac{\Delta F}{k_B T}\right), \tag{1}$$

where ΔF is the free energy barrier associated with the transition between pathway 1 and pathway 2. The data points shown in the inset of the top panel of Fig. 5 suggest $k_B T/\Delta F \approx 1.6$. As T increases, the jumps on each pathway get shorter and are usually followed by another jump with a much shorter jumping distance ($d < 100$ Å in the middle and bottom panels). Above $k_B T/\epsilon = 0.5$ the peaks corresponding to pathways 1 and 2 merge. At this stage, the short distance part of the distribution may be approximated by the exponential distribution $P(d) = \alpha^{-1}\exp(\alpha d)$, as shown in the bottom panel for $k_B T/\epsilon = 0.7$. In the inset in the bottom panel $\log P(d)$ is fitted to a line whose slope yields $\alpha \approx -0.027$.

In summary, we have found that the process of knot tightening in proteins is qualitatively distinct from that occurring in homopolymers. The proteinic knots shrink in size and one of their ends gets pinned on a sharp turn. The movement of knot ends in the protein along the sequence is characterized by sudden jumps, whereas in polymers knots perform a diffusive motion and, in most cases, slide off the chain. We predict that most proteins will not refold to the native conformation after a stretch and release. It would be interesting to devise stretching experiments that would monitor knot tightening and end jumping in proteins, analogous to those reported for nucleic acids [4].

We appreciate the useful comments of R. Kutner and R. Stolarski. This work was funded by MNiSW Grants No. N202 021 31/0739 and No. N N202 085233.

[1] K. Koniaris and M. Muthukumar, Phys. Rev. Lett. **66**, 2211 (1991).
[2] N. Moore, R. C. Lua, and A. Y. Grosberg, Proc. Natl. Acad. Sci. U.S.A. **101**, 13 431 (2004).
[3] V. Katritch, J. Bednar, D. Michoud, R. G. Scharein, J. Dubochet, and A. Stasiak, Nature (London) **384**, 142 (1996).
[4] X. R. Bao, H. J. Lee, and S. R. Quake, Phys. Rev. Lett. **91**, 265506 (2003); Phys. Rev. Lett. **95**, 199901 (2005).
[5] B. Marcone, E. Orlandini, A. L. Stella, and F. Zonta, J. Phys. A **38**, L15 (2005).
[6] P. Virnau, Y. Kantor, and M. Kardar, J. Am. Chem. Soc. **127**, 15 102 (2005).
[7] M. L. Mansfield, Nat. Struct. Biol. **1**, 213 (1994).
[8] W. Taylor, Nature (London) **406**, 916 (2000).
[9] R. C. Lua and A. Y. Grosberg, PLOS Comp. Biol. **2**, e45 (2006).
[10] W. R. Taylor, Comput. Biol. Chem. **31**, 151 (2007).
[11] P. Virnau, L. A. Mirny, and M. Kardar, PLOS Comp. Biol. **2**, e122 (2006).
[12] M. T. Alam, T. Yamada, U. Carlsson, and A. Ikai, FEBS Lett. **519**, 35 (2002).
[13] S. Ohta, M. T. Alam, H. Arakawa, and A. Ikai, Biophys. J. **87**, 4007 (2004).
[14] H. Abe and N. Go, Biopolymers **20**, 1013 (1981).
[15] M. Cieplak and T. X. Hoang, Biophys. J. **84**, 475 (2003).
[16] A. Arend, R. Rzehak, and W. Zimmermann (to be published).
[17] R. Metzler, W. Reisner, R. Riehn, R. Austin, J. O. Tegenfeldt, and I. M. Sokolov, Europhys. Lett. **76**, 696 (2006).
[18] M. Cieplak, T. X. Hoang, and M. O. Robbins, Proteins: Struct. Funct. Bio. **56**, 285 (2004).
[19] J. I. Sułkowska and M. Cieplak, J. Phys. Condens. Matter **19**, 283201 (2007); Biophys. J. **94**, 6 (2008).
[20] S. Wallin, K. B. Zeldovich, and E. I. Shakhnovich, J. Mol. Biol. **368**, 884 (2007).
[21] D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, and E. Paci, Biophys. J. **90**, 287 (2006).
[22] P.-G. de Gennes, Macromolecules **17**, 703 (1984).
[23] P. Pieranski, S. Przybyl, and A. Stasiak, Eur. Phys. J. E **6**, 123 (2001).
[24] See EPAPS Document No. E-PRLTAO-100-044806 for supplementary movies and material. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.
[25] I. D. Kuntz, J. Am. Chem. Soc. **94**, 4009 (1972).
[26] A. L. Mallam and S. E. Jackson, J. Mol. Biol. **359**, 1420 (2006).