J|A|C|S COMMUNICATIONS

Published on Web 09/21/2010

Untying Knots in Proteins

Joanna I. Sułkowska,*,† Piotr Sułkowski,‡,# Piotr Szymczak,¶ and Marek Cieplak§

Center for Theoretical Biological Physics, University of California−San Diego, 9500 Gilman Drive, La Jolla, California 92037, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, Institute of Theoretical Physics, University of Warsaw, ul. Hoża 69, 00-681 Warsaw, Poland, and Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

Received April 1, 2010; E-mail: jsulkow@physics.ucsd.edu

Abstract: A shoelace can be readily untied by pulling its ends rather than its loops. Attempting to untie a native knot in a protein can also succeed or fail depending on where one pulls. However, thermal fluctuations induced by the surrounding water affect conformations stochastically and may add to the uncertainty of the outcome. When the protein is pulled by the termini, the knot can only get tightened, and any attempt at untying results in failure. We show that, by pulling specific amino acids, one may easily retract a terminal segment of the backbone from the knotting loop and untangle the knot. At still other amino acids, the outcome of pulling can go either way. We study the dependence of the untying probability on the way the protein is grasped, the pulling speed, and the temperature. Elucidation of the mechanisms underlying this dependence is critical for a successful experimental realization of protein knot untying.

Several tens of protein strutures, which belong to different folds and classes, are currently known to contain knots.[1−3] The knotted proteins are fascinating to biologists: it is not clear why nature uses such molecules and how their folding proceeds.[4,5] Understanding how to accomplish untying these knots may suggest a proper way to design single-molecule experiments aiming at identifying folding pathways of knotted proteins: if untied conformations are used in refolding studies, one avoids situations in which unfolded states already contain knots. Recent experimental studies indicate that a knot may either get tightened[6] or untied,[7,8] depending on the way the protein is pulled. Here, we explore the pulling direction dependence of the untying probability in a systematic way through molecular dynamics (MD) simulations. We study two methyltransferases with the Protein Data Bank structure codes 1o6d and 1v2x. Their geometries are different, yet the results found are qualitatively similar (see Supporting Information). As an illustration, we focus our discussion on 1o6d, which contains a trefoil knot and comprises $N = 147$ residues. In the native state, the knotted core (i.e., the minimal segment of amino acids that can be identified as a knot) is located between $k_1 = 65$ and $k_2 = 119$ (see Figure 1). Note that the distance between k_1 and the N-terminus is about double the distance between k_2 and the C-terminus. Thus, placing one of the attachment points near the C-terminus makes untying easier than placing it near the other terminus. We have selected 12 attachment points: five in the knotted region (73, 76, 94, 101, 113), five on the N-terminal side (1, 10, 27, 40, 63), and two on the C-terminal side (122, 147). None of them is buried, and connecting them to a cantilever tip through molecular linkers should be implementable experimentally.[9]

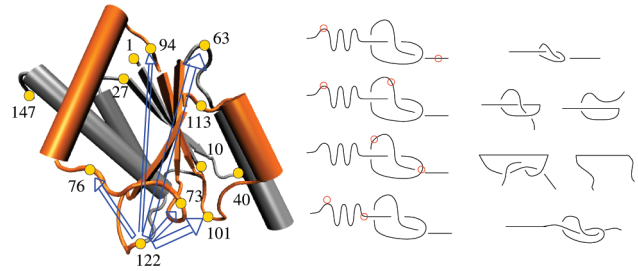

Figure 1. Left: Structure of protein 1o6d, in which the locations of 12 representative amino acids used in pulling are indicated (yellow circles). The knotted core (in orange) extends between sites 65 and 119. It consists of two nearly symmetric loops "attached" to the β-sheet. The arrows represent the energy landscape for untying in a schematic way; their width corresponds to the free energy barrier F_0 as listed in eq 4. The arrows are shown for $p_2 = 122$. Middle: Schematic representation of a trefoil knot in a protein, with the circles indicating the main geometrical directions of pulling. Right: The resulting final conformations. In the second and third cases they can be either knotted or unknotted.

The four possible pulling directions are shown schematically in Figure 1. We shall discuss the untying process first in the absence of thermal fluctuations, i.e., at temperature $T = 0$. In the *first* case, the attachment points (denoted by p_1 and p_2) are located on opposite sides of the knotted core, and the final stretched structure is always knotted (for theoretical analysis of this case, see refs 10 and 11 and the first experimental realization in ref 6). The *second* case corresponds to the experimental situation for carboanhydrase.[7] Here, p_1 is located on one side of the knot and p_2 inside the knotted core. Such a knot can be untied if the distance between p_2 and the C-terminus ($N-p_2$) is sufficiently short to allow for dragging of the C-terminus out of the knotted loop (i.e., is smaller than p_2-p_1). This yields the condition

$$p_1 < 2p_2 - N \tag{1}$$

which for $p_2 =$113, 101, and 94 in 1o6d gives $p_1 < 79$, 57, and 39, respectively. For $p_2 = 76$ we get $p_1 < 5$. In the reverse situation with $p_1 > p_2$, the above condition reads simply $p_1 > 2p_2$. In the *third* case, both p_1 and p_2 are located inside the knotted core. As discussed in the Supporting Information, for our choice of p_1 and p_2, untying at $T = 0$ is not possible in this case. In the *fourth* case, both p_1 and p_2 are between one of the termini and the knotted core. This way of pulling does not affect the knot, so it remains entangled.

These theoretical considerations are confirmed by our MD simulations. For example, at $T = 0$ and for $p_2 = 101$, each choice of $p_1 = 1$, 10, 28, and 43 leads to untying, in agreement with the constraint $p_1 < 57$ from eq 1, while the choice of $p_1 = 63$ leads to tightening of the knot.

† University of California−San Diego.
‡ California Institute of Technology.
¶ University of Warsaw.
§ Polish Academy of Sciences.
# On leave from Institute for Nuclear Studies, Warsaw, Poland.

10.1021/ja102441z

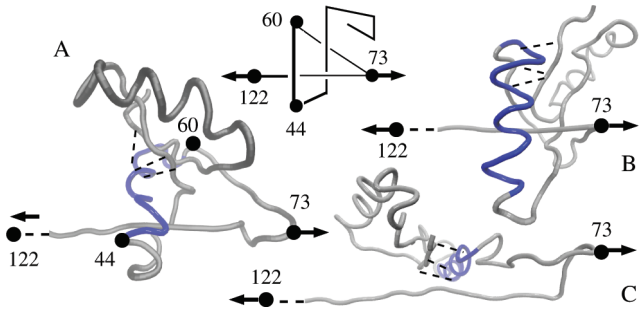

Figure 2. Topological transitions resulting from pulling at $p_1 = 73$ and $p_2 = 122$; the sense of pulling is shown by the arrows. At a certain stage, the system is in conformation A (also sketched schematically at the top). Its topology bifurcates at this point, depending on the behavior of the conformation between residues 44 and 60, shown in blue, during a critical and short stretching distance Δd. If it continues to be helical, the end result is the knotted conformation, B. However, if the helix unfolds, then the loop spanned by residues 44−60−73 gets enlarged, allowing for the N-terminal part (amino acids 1−40, which includes another helix) to exit through the loop and form the untied conformation, C. Eventually, the 44−60 regions re-forms the helix and re-establishes contacts (broken lines) with the other parts of the protein, C.

At nonzero temperatures, there appears a region in the pulling direction space (cf. upper left panel of Figure 3) in which the outcome of stretching is no longer unique: the knot survives with a certain probability P, which is a function of temperature T, velocity υ, and the pulling direction. Interestingly, in all the cases analyzed by us, the decision as to which pathway to follow is taken during a specific short time span (characterized by stretching distance Δd) and is associated with the stability of a certain secondary structure against thermal fluctuations. For instance, selecting $p_1 = 73$ and $p_2 = 122$ leads to either tightening or untying, which depends on the behavior of one of the helices 40−60. During the stretching process, this helix becomes exposed. If it unfolds temporarily, the loop gets wider, which may result in untying. Otherwise, the knot gets tightened, as illustrated in Figure 2. For other choices of p_1 and p_2, analogous mechanisms arise. Characteristics of the unfolding process, such as force−displacement curves, positions of the knot's ends k_1 and k_2, fraction of native contacts, and radius of gyration, are discussed further in the Supporting Information.

The effect of thermal fluctuations on the knot untying process is illustrated in Figure 3. Note that, at $T = 0$, untying takes place inside a wedge-like region colored red in the p_1−p_2 plane, which is determined by the geometry: the vertical edge of the wedge at $p_2 = 119$ corresponds to the end point of the knotted core k_2, while the tilted edge follows the geometrical constraint of eq 1. The region in the pulling direction space in which there is a nonzero probability of knot untying gets considerably larger as the temperature increases. However, only in the wedge-like region described above does the knot untie with 100% certainty.

For trajectories that do not lead to untying, relative frequencies of various final knot-containing conformations for $T = 0.3\varepsilon/k_B$ are shown in the top right panel of Figure 3. The regions in red or yellow correspond to the locations of knot ends (k_1,k_2) that appear most often. There are three typical final states: (1) a knot that is tightened maximally, with $|k_1 - k_2| \approx 10$ (these conformations form a tilted line at the upper left corner of the graph), (2) a knot with k_1 and k_2 at their native locations (this is represented by the red region in the vicinity of point (119,65)), and (3) no knot. Analogous maps for $T = 0$ and $T = 0.5\varepsilon/k_B$ are shown in the Supporting Information.

There is a complicated energy landscape that underlies protein stretching in various directions. This landscape also characterizes

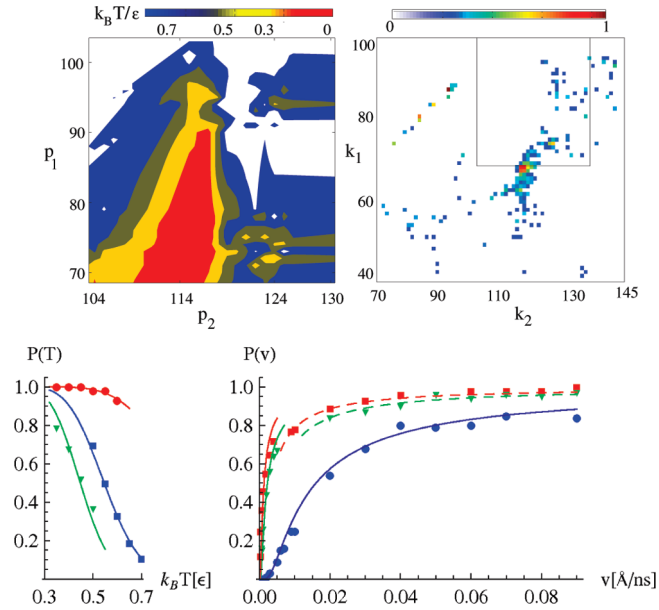

Figure 3. (Top) Left: Attachments points (p_2, p_1) that may lead to an unknotted configuration at various temperatures (denoted by colors, as indicated on the scale above). Right: Distribution of the knot end locations, (k_2,k_1) at $T = 0.3\varepsilon/k_B$. The color scale indicates the probability of attaining particular values of k_2 and k_1. The framed part in the right panel indicates the region covered by the left panel. (Bottom) Left: Dependence of the probability of knot survival on T upon stretching for $p_2 = 122$ and $p_1 = 63$ (circles), 76 (squares), and 94 (triangles) at a fixed pulling speed. Right: Dependence of the survival probability, for the 76−122 pulling direction, on υ for $k_BT/\varepsilon = 0.3$ (squares), 0.5 (triangles), and 0.7 (circles). For $k_BT/\varepsilon = 0.3$ and 0.5, the two unfolding pathways are represented by continuous and dashed lines. For $k_BT/\varepsilon = 0.7$, there is a unique unfolding pathway. All lines are fitted to the model $P(T,\upsilon)$ with the same $\Delta d/\tau_0 = 0.7$ Å/ns and slightly different E_0 reflecting a choice of the pathway.

topological properties of proteins by encoding heights of energy barriers for untying for various p_1 and p_2. We discuss a patch in such a landscape and consider stretching for a set of directions with $p_1 = 63, 73, 76, 94, 101$ and p_2 fixed at 122. For each choice of p_1, we determine the dependence of knot survival probability, $P(t|T, \upsilon)$, by measuring, in a large ensemble of simulations, a fraction of conformations that remain knotted when pulling at speed υ. As mentioned above, for each pulling direction, there is a well-defined short range of stretching distances, Δd (and a corresponding time range, $t_1 < t < t_2$), during which the knot may untie. This implies that it should be possible to describe the untying process simply in terms of crossing a single kinetic barrier of height E_0:

$$\frac{dP(t|T,\upsilon)}{dt} = -\frac{1}{\tau_0} e^{-E_0/k_BT} P(t|T,\upsilon), \quad t_1 < t < t_2 \qquad (2)$$

where τ_0 is an intrinsic period and the height E_0 is measured relative to the state of the system at time t_1, and where additionally we neglect the effect of the time-dependent external force onlowering the barrier during the time interval $t_2 - t_1$. In such case the final survival probability (at $t = t_2$) reads (see also Supporting Information)

$$P(T,\upsilon) = \exp\left(-\frac{t_2 - t_1}{\tau_0} e^{-E_0/k_BT}\right) = \exp\left(-\frac{\Delta d}{\upsilon\tau_0} e^{-E_0/k_BT}\right) \qquad (3)$$

It is striking that such a simple model indeed describes the complicated nonlocal change in topology of the protein, as seen, e.g., in the corresponding fit for $\upsilon = 0.005$ Å/ns and $p_1 = 63, 76, 94$ shown in Figure 3 (bottom left). There is one subtlety, though: in

some cases, e.g., $p_1 = 76, 94$, there may exist two configurations with the knot blocked. One configuration is chosen at lower values of T and v, and the other at their higher values. This model leads to the following barrier heights for pulling in directions involving $p_2 = 122$ (for $p_1 = 76, 94$, the data concern the high-T pathway):

p_1	63	73	76	94	101
E_0/ε	4.6	3.3	2.9	2.5	5.1

(4)

The corresponding patch of the energy landscape is indicated in Figure 1 by the arrows connecting relevant pairs of residues, with their widths proportional to the barrier heights. The dependence of the survival probability on v, for a fixed $p_1 = 76$ and various temperatures, is shown in the bottom right panel of Figure 3. Notably, for $k_BT/\varepsilon = 0.7$ and the direction 76−122, two independent fits to the temperature and velocity dependence of $P(v,T)$ give the same value of $E_0 = 2.9\varepsilon$, which supports the applicability of eq 3 to the description of the knot untying process.

Understanding the dynamics of knot untying is important for the proper interpretation of unfolding data. We provide criteria that can be used to untie knotted proteins. When, due to thermal fluctuations, the outcome of the untying process becomes random, it is possible to describe its kinetics in terms of a simple barrier crossing. Accomplishing knot untying still poses experimental challenges, but overcoming them may be essential for experimental studies of how knotted proteins fold. In chemical denaturation experiments,[12] the difficult issue is to tell whether the denatured state is knotted, since there is no direct way of detecting a knot (in experiments on knot tightening through stretching,[6] the knot is detected as an apparent shortening of the chain length). To ensure unknotting, such a molecule can be stretched in one of the directions for which the geometrical conditions (1) are fulfilled. Then, refolding in a force clamp[13] or applying related experimental methods should allow the dynamics of knot formation to be explored. Note that the model we use[14] accounts for nonuniform mechanical properties of proteins, which is another feature that distinguishes proteins from shoelaces. Knots formed on the more homogeneous DNA molecules, however, should be still affected by thermal fluctuations.

Acknowledgment. This work was supported by the EC FUNMOL FP7-NMP-2007-SMALL-1 project, grant agreement 213382, and the EU Innovative Economy grant POIG.01.02. J.S. was funded by NSF-PHY-0822283 and NSF-MCB-0543906 and P. Sułkowski by a Marie Curie IOF Fellowship and Foundation for Polish Science.

Supporting Information Available: Details of the simulation model and algorithm to determine the knot position during unfolding. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Lua, R. C.; Grosberg, A. Y. *PLoS Comp. Biol.* **2006**, *2*, 5.
(2) Virnau, P.; Mirny, L. A.; Kardar, M. *PLoS Comp. Biol.* **2006**, *2*, 1074.
(3) Taylor, W. R. *Comp. Biol. Chem.* **2007**, *31*, 151.
(4) Yeates, T. O.; Norcross, T. S.; King, N. P. *Curr. Opin. Chem. Biol.* **2007**, *11*, 595.
(5) Sułkowska, J. I.; Sułkowski, P.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3119.
(6) Bornschloegl, T.; Anstrom, D. M.; Mey, E.; Dzubiella, J.; Rief, M. T.; Forest, K. *Biophys. J.* **2009**, *96*, 1508.
(7) Alam, M. T.; Yamada, T.; Carlsson, U.; Ikai, A. *FEBS Lett.* **2002**, *519*, 35.
(8) Ohta, S.; Alam, M. T.; Arakawa, H.; Ikai, A. *Biophys. J.* **2004**, *87*, 4007.
(9) Dietz, H.; Berkemeier, F.; Bertz, M.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12724.
(10) Sułkowska, J. I.; Sułkowski, P.; Szymczak, P.; Cieplak, M. *Phys. Rev. Lett.* **2008**, *100*, 058106.
(11) Dzubiella, J. *Biophys. J.* **2009**, *96*, 831.
(12) Mallam, A. L.; Jackson, S. E. *J. Mol. Biol.* **2007**, *366*, 650.
(13) Fernandez, J. M.; Li, H. *Science* **2004**, *303*, 1674.
(14) Sułkowska, J. I.; Cieplak, M. *J. Phys.: Condens. Matter* **2007**, *19*, 283201.

JA102441Z