Sequence-Specific Size, Structure, and Stability of Tight Protein Knots

Joachim Dzubiella*
Physics Department, Technical University Munich, Garching, Germany

ABSTRACT Approximately 1% of known protein structures display knotted configurations in their native fold, but the function of these configurations is not understood. It has been speculated that the entanglement may inhibit mechanical protein unfolding or transport, e.g., as in cellular threading or translocation processes through narrow biological pores. Protein knot manipulation, e.g., knot tightening and localization, has become possible in single-molecule experiments. Here, we investigate tight peptide knot (TPK) characteristics in detail by pulling selected 3_1 and 4_1-knotted peptides using all-atom molecular dynamics computer simulations. We find that the 3_1- and 4_1-TPK lengths are typically $\Delta l \approx 47 \pm 4$ Å and 69 ± 4 Å, respectively, for a wide range of tensions ($0.1\ \text{nN} \lesssim F \lesssim 1.5\ \text{nN}$). The 4_1-knot length is in agreement with recent atomic force microscopy pulling experiments. Calculated TPK radii of gyration point to a pore diameter of ~20 Å, below which a translocated knotted protein might get stuck. TPK characteristics, however, may be sequence-specific: we find a different size and structural behavior in polyglycines, and, strikingly, a strong hydrogen bonding and water trapping capability of hydrophobic TPKs. Water capture and release is found to be controllable by the tightening force in a few cases. These mechanisms result in a sequence-specific "locking" and metastability of TPKs, which might lead to a blocking of knotted peptide transport at designated sequence positions. We observe that macroscopic tight 4_1-knot structures are reproduced microscopically ("figure of eight" versus the "pretzel") and can be tuned by sequence, in contrast to mathematical predictions. Our findings may explain a function of knots in native proteins, challenge previous studies on macromolecular knots, and prove useful in bio- and nanotechnology.

INTRODUCTION

After the discovery of the first knotted structure in the native fold of a protein in 1994 (1), additional studies (2,3) and, in particular, a recent survey, have identified almost 300 more knotted proteins, constituting ~1% of known structures in the protein database (4). Most of them have the simplest 3_1 (trefoil) topology; only a few have been found to possess the more complicated 4_1 (typically called a ''figure-of-eight'' knot) and 5_2 types of prime knots (6). (The nomenclature ''X_n knot'' refers to the number of strand crossings of the knot structure projected onto a two-dimensional plane, i.e, $X = 3$ for the trefoil and 4 for the figure-of-eight topology. The index n refers to the type of prime knot with the same number of crossings. For the three- and fourfold crossings, only one prime knot exists; thus, for them, always, $n = 1$.) Fig. 1 provides an illustration of tightened 3_1 and 4_1 open knot topologies. (Mathematically these knots are open, i.e., no closed loops, but throughout the work we just use the term ''knot'' for simplicity.) Whereas the question of the physiological relevance of protein knots is still a matter of debate (3,8,9), it has been proposed that the entangled structure might have a stabilizing effect against thermal or mechanical protein unfolding (4,9–11). An interesting possible consequence is the inhibition of knotted protein translocation and threading through the narrow pores of biological membranes or proteasomes (4). In this respect, it is tempting to speculate that the steric blocking of narrow pathways by a localized or tightly pulled protein knot may have bio(techno)logical significance. Very recently, another complex entanglement referred to as a ''slipknot'' has been discovered in proteins, the existence of which is also linked to a stabilizing function (12). Also relevant in this respect are cyclotides, a superstable family of proteins that feature a cyclic peptide backbone and a tightly entangled topology in their interior, showing strong biological activity and high pharmaceutical potential (13). In contrast to that of protein knots, however, the tight structure of cyclotides is generated by covalent connections between cysteine side chains. We note that the synthesis and design of artificially interlocked molecules has become possible in supramolecular chemistry with applications in bio- or nanotechnology, e.g., as molecular receptors, locks, or machines (14,15).

The study of tight-knot characteristics in (bio)polymers has been of interest for a long time, as knots easily self-tie and localize in any long chain (16–19). More than 20 years ago, de Gennes argued that knots may self-tie in crystallizing or sheared polymer melts, changing the macroscopic relaxation behavior of the latter (20). Possible self-tying mechanisms may be based on electrostatic repulsion (21), entropic tightening in wormlike chains (22), or localization of polymer knots either in confinement (23–25) or in bad solvent conditions (17). Externally controlled manipulation and characterization of microscopic knots has become accessible experimentally by employing optical tweezer methods (26,27) or atomic force microscopy (AFM) (9,28,29). Molecular knot behavior has been addressed from a theoretical perspective using scaling arguments (22), vacuum quantum calculations (30), or coarse-grained computer simulations (17,31–35). Previous studies focused almost

Submitted August 12, 2008, and accepted for publication October 20, 2008.
*Correspondence: jdzubiel@ph.tum.de
Editor: Nathan Andrew Baker.

0006-3495/09/02/0831/9 $2.00
doi: 10.1016/j.bpj.2008.10.019

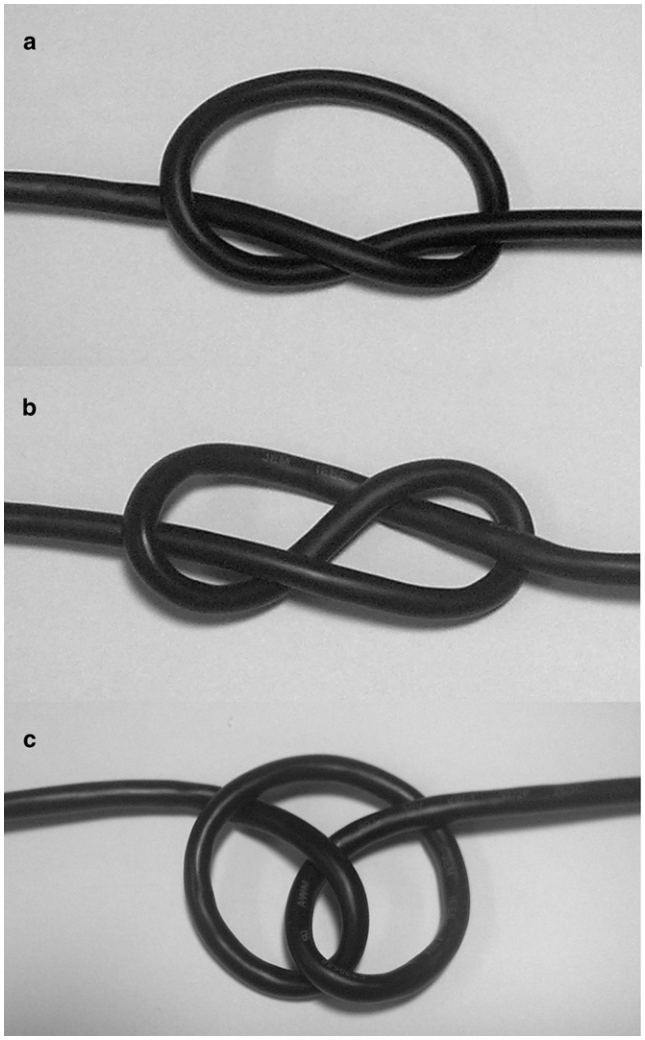

FIGURE 1 Self-made photographs of open tight-knot configurations in a tensioned computer cable. (*a*) A trefoil (3_1)-knot. (*b*) A ''figure-of-eight'' configuration in a 4_1 knot. (*c*) A ''pretzel'' configuration in a 4_1 knot.

exclusively on homogeneous systems such as polyethylene, DNA, or actin filaments; only recently have tight protein knots (TPKs) with specific, inhomogeneous sequences been investigated by Sułkowska et al. using an implicit-solvent Go model (36). Due to local geometry (side-chain size or kinks in the peptide), knot localization and diffusion was dominated by jumps of the knot's ends to specific peptide locations, suggesting qualitatively different, sequence-dependent fluctuations of TPKs when compared to homopolymers (21,27,34,35,37).

Although highly relevant for transport, translocation, and threading processes of knotted proteins, the size of a TPK has not been determined before in a systematic way to our knowledge. For a first rough guess, consider a rope of contour length l_c, tie a knot in it, and tighten it by pulling the rope's ends. The end-to-end distance of the rope, l, will be reduced by the presence of the knot by $\Delta l = l_c - l$. In the following, we refer to Δl as the ''tight knot length'', i.e., the length of the rope involved in the open knot. By dividing Δl by the rope thickness, D, we obtain the characteristic quantity (38)

$$\Lambda = \frac{\Delta l}{D}, \tag{1}$$

which has been shown to be minimized by the tight knot conformation, and is $\Lambda = 10.1$ and 13.7 for 3_1 and 4_1 knots, respectively, for idealized hard-core ropes (38). If we apply this simplified picture to molecular entities and assume a typical peptide thickness of the order of an atomic size $D \simeq 3.5 - 5$ Å, we would anticipate $\Delta l = \Lambda D \simeq 35 - 51$ Å and 48–69 Å for tight 3_1- and 4_1-peptide knots, respectively. This coarse size estimate for idealized, nonelastic hard-core peptides agrees well with recent AFM predictions of the 4_1-TPK length $\Delta l = 62 \pm 10$ Å in bacterial phytochrome (29). If the knot is further assumed, naively, to be a circle with circumference $\Delta l = 2\pi R$, we infer that a typical TPK radius would be $R \simeq 6 - 11$ Å.

In this work, we take a more detailed look at TPKs by performing explicit-water molecular dynamics (MD) computer simulations (39) of 3_1 and 4_1 knots in selected polypeptides. The simulated peptides involve up to 30 amino acids and we systematically study knot sizes and their structural behavior. We find TPK lengths of $\Delta l \simeq 47 \pm 4$ Å (involving 13 ± 1 amino acids) and 69 ± 4 Å (19 ± 1 amino acids) for the 3_1 and 4_1 knots, respectively. The knot sizes are found to be surprisingly constant for a wide range of stretching forces ($F \lesssim 1.5$ nN). Typical tight-knot radii of gyration are $R_g \simeq 7 - 8$ Å. All sizes are in the range of the macroscopic estimate (1). The 4_1-TPK length is in agreement with recent AFM pulling experiments on the natively knotted bacterial phytochrome (29). Detailed tight-knot characteristics, however, may be sequence-specific: we find smaller knots, and different structural and stability behavior, in the special case of polyglycines. We find that TPKs have a strong water-capturing and hydrogen-bonding capability within their closely packed interior that is sequence-specific and promoted by nonpolar side chains. Buried water and long-lived intraknot hydrogen bonds lead to surprisingly rigid and stable tight knots in free simulations on a ~100-ns timescale. An intriguing finding is that macroscopic tight 4_1-knot structures are reproduced microscopically (in a ''figure-of-eight'' versus a ''pretzel'' configuration) but depend on peptide sequence. This stands in contrast to mathematical predictions of the tight 4_1-knot structure, which is the ''figure-of-eight'' structure (38). We predict strongly localized tight knots after peptide stretching, and a preferential affinity toward regions with dominantly nonpolar side chains. We demonstrate that the accurate modeling of specific side chains and the aqueous environment is crucial for a full understanding of protein knot characteristics.

METHODS AND SYSTEMS

MD simulations

Our all-atom MD simulations are performed using the software package Amber9.0 with the ff03 force-field and TIP3P solvent (40). Systems are maintained at a fixed pressure, $P = 1$ bar, and temperature, $T = 300$ K, by coupling to a Berendsen barostat and Langevin thermostat, respectively. System sizes vary between $N \simeq 4000$ and $N \simeq 8000$ atoms. Electrostatic neutrality is assured by additional Na^+ counterions compensating the net peptide charge given at pH 7. The rectangular and periodically repeated simulation box has lateral edge lengths $L_x \simeq L_y \simeq 30 - 35$ Å while in the peptide stretching direction, $L_z \simeq 55 - 70$ Å, depending on amino acid and peptide size. Given the observed maximum extension in the x- or y-direction of TPKs of ~15–20 Å, this allows for at least a 15-Å distance between the peptide and its nearest image. Box sizes are based on thorough testing of finite-size effects before the production runs. Electrostatic interactions are calculated by particle mesh Ewald summation and real-space interactions have a cut-off of 9 Å. Polypeptides are generated using the Amber ''tleap'' tool. Knots are tied into them utilizing interactive MD (IMD) in visual MD (VMD) (41): while a Langevin simulation of the peptide is running and visualized, a force can be applied to selected fragments by using the computer mouse. Thus, the peptide can be dragged by hand into a finally knotted configuration. Thereafter, the system is equilibrated for $\simeq 5$ ns with Langevin dynamics, solvated with TIP3P water, and further equilibrated by an $\simeq 5$-ns MD simulation. For peptide stretching and loosening, we utilize the Amber steered MD (SMD) tool: a constant pulling velocity of 0.1 Å/ns (0.01 m/s) drives the first and last atom (in a distance l) of the peptide backbone in opposite directions, and force-extension curves $F(l)$ are calculated. Pulling is terminated after the mean force reaches ~1.5 nN, a value at which covalent bond breaking can occur experimentally (28). For every investigated system, at least one stretching-loosening (''reverse pulling'') loop is performed to check for reproducibility, a possible hysteresis, and nonequilibrium effects. This leads to a simulation time of $\simeq$ 200–300 ns per peptide and stretching-loosening loop. Simulation snapshots are generated using VMD (41). Hydrogen bonds, radii of gyration, and root mean-square deviations (RMSDs) are analyzed using the Amber ''ptraj'' tool.

Systems

Knots of types 3_1 and 4_1 are investigated. To study the influence of amino acid type on the tight knot structure we opt for three different homopeptides: the hydrophobic polyleucine (sequence $L_{N_{aa}}$), the partly hydrophilic and charged polyglutamic acid ($E_{N_{aa}}$), and the slim, amphiphilic polyglycine ($G_{N_{aa}}$). The peptides have a total number of amino acids of $N_{aa} = 21$ and 30 for the 3_1 and 4_1 knots, respectively. Furthermore, two randomly picked pieces from the knotted cores of the natively 3_1-knotted YibK methyltransferase (42) and the 4_1-knotted Class II ketol-acid reductoisomerase (43) are considered to directly connect to naturally occurring protein knots. In the following, we name the knotted peptides by knot type and sequence, e.g., ''3_1L'' for a polyleucine trefoil and ''4_1mix'' for the 4_1 knot in a mixed sequence. The different knotted-peptide systems, their amino acid (aa) sequence and number (N_{aa}) are summarized in Table 1. The nomenclature ''X_n''-knot refers to the number of strand crossings of the knot structure projected onto a two-dimensional plane, i.e, $X = 3$ for the trefoil and $X = 4$ for the figure-of-eight topology. The index n refers to the type of prime knot with the same number of crossings. For the three- and fourfold crossings only one prime knot exists, thus for them always $n = 1$.

A comment is in order here regarding the chirality of the knots. Each crossing in a protein knot can be assigned a ''handedness'' (6). If the undercrossing strand passes the direction of the overcrossing strand from right to left, then it is righthanded (R); if the reverse is true, it is left-handed (L). A knot nomenclature can be defined by listing the handedness of the crossings according to their sequential occurrence, so that a righthanded trefoil is ''RRR'' and a left-handed trefoil is ''LLL'', whereas the figure of eight is ''RLRL'' or ''LRLR'' (8). All types of handedness (or chirality) have been observed in native proteins (8,42,44,45). As our work connects to the natively lefthanded trefoil in the YibK protein (42) and the ''LRLR'' reductoisomerase (8,43,), all our 3_1 and 4_1 knots are lefthanded and of ''LRLR'' chirality, respectively. For the TPK properties investigated in this work, however, we do not expect any noticeable influence of chirality.

RESULTS AND DISCUSSION

Tight-knot size and structure

A typical initial configuration of a 3_1-knotted peptide in our simulation is shown in Fig. 2 *a*, where a snapshot of 3_1G is sketched before pulling it tight. The end-to-end extension here is $l \simeq 25$ Å. A tight-knot situation for the same peptide is shown in Fig. 2 *b* for a large stretching force of ~1.5 nN ($l \simeq 45$ Å). For an elastic peptide, as considered in this study, the final ''tightness'' of the knot will naturally depend on the external stretching force, F. The calculated force-extension curve, $F(l)$, for 3_1G is shown in Fig. 3 *a* together with the data for 3_1E and 4_1L. We observe an overall monotonic nonlinear increase of the force. Fluctuations are moderate on that scale and have local standard deviations ranging from ~20 pN to ~50 pN. We also plot $F(l)$ of knot loosening, i.e., ''reverse pulling'', showing complete reversibility of the process and no obvious hysteresis within the

TABLE 1 Simulated knotted peptide systems

Knot	Amino acid sequence	N_{aa}	$l_c(F_1)$ (Å)	$l(F_1)$ (Å)	$\Delta l(F_1)$ (Å)	$n_{aa}(F_1)$	$l_c(F_2)$ (Å)	$l(F_2)$ (Å)	$\Delta l(F_2)$ (Å)	$n_{aa}(F_2)$
3_1L	L_{21}	21	77.3	27.5	49.8	14	79.0	32.5	46.5	12
3_1E	E_{21}	21	77.3	31.0	46.3	12	79.0	32.5	46.5	12
3_1G	G_{21}	21	77.3	33.0	44.3	12	79.0	42.0	37.0	10
3_1mix	AHSQVKFKLG DYLMFGPETRG	21	77.3	27.8	49.5	13	79.0	32.9	46.1	12
4_1L	L_{30}	30	110.4	40.0	70.4	19	112.8	44.0	68.8	18
4_1E	E_{30}	30	110.4	42.0	68.4	19	112.8	45.1	67.3	18
4_1G	G_{30}	30	110.4	50.0	60.4	16	112.8	63.0	49.8	13
4_1mix	TKGMLALYNS LSEEGKKDFQ AAYSASYYPS	30	110.4	38.7	71.7	19	112.8	44.3	68.5	18

Systems are named by knot type and sequence, e.g., ''3_1L'' for a polyleucine trefoil, ''3_1E'' for a polyglutamic acid trefoil, ''3_1G'' for a polyglycine trefoil, and ''4_1mix'' for the 4_1-knot in a mixed sequence. The peptides have N_{aa} amino acids with the sequence shown. l_c is the estimated contour length of the unknotted peptide, l the measured end-to-end distance of the knotted peptide, and Δl the tight-knot length involving n_{aa} amino acids. The lengths are evaluated at a stretching force of $F_1 = 200$ pN and $F_2 = 1$ nN.

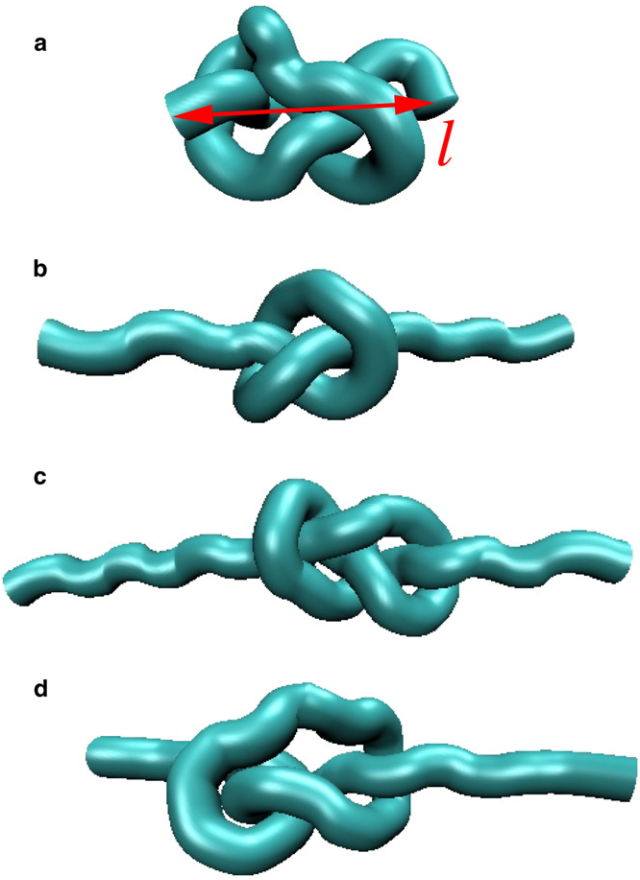

FIGURE 2 MD simulation snapshots of different protein knots in a "cartoon" representation. (*a*) Initial configuration of peptide 3_1G, where $l \simeq 25$ Å is the end-to-end distance. (*b*) Tight-knot configuration of peptide 3_1G. The end-to-end distance is $l \simeq 45.0$ Å at stretching force $F \simeq 1.5$ nN. (*c*) A tight "figure-eight" knot configuration of peptide 4_1G at $F \simeq 1$ nN. (*d*) A tight "pretzel" knot configuration of peptide 4_1L at $F = 1$ nN.

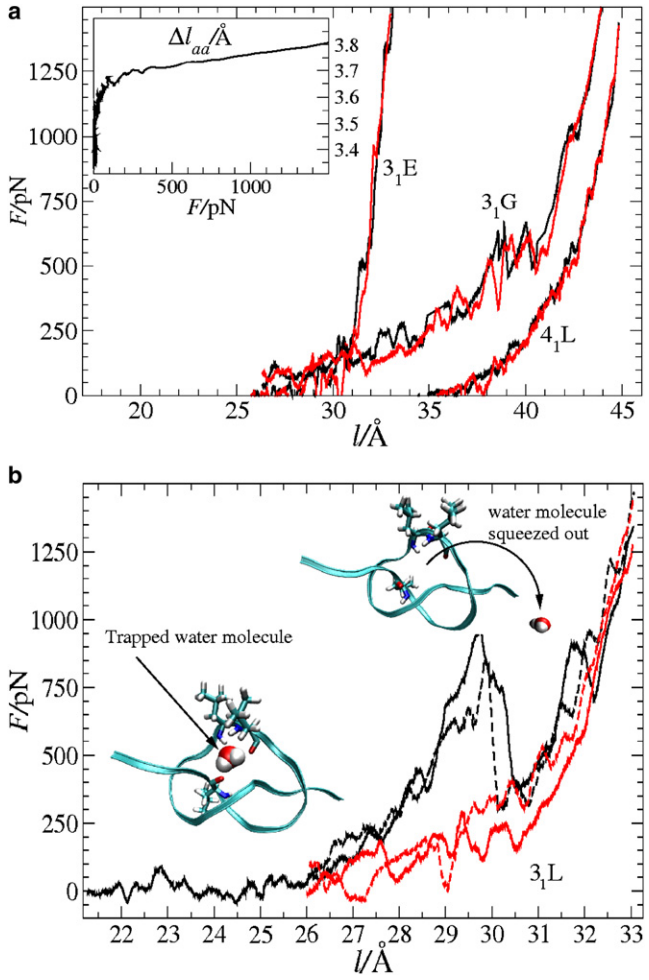

FIGURE 3 (*a*) Force (F)-extension (l) curves for the peptides 3_1E, 3_1G, and 4_1L. Stretching curves (*black lines*) and loosening curves (*red lines*) lie on top of each other, indicating a small hysteresis. The pulling rate is 0.1 Å/ns. The inset shows the mean distance Δl_{aa} between neighboring backbone nitrogen atoms versus stretching force, F, in an unknotted peptide. (*b*) Force-extension curves for the polyleucine 3_1L. Black lines correspond to stretching, whereas red lines correspond to loosening of the knot. While stretching, for extensions $l \lesssim 30$ Å, a single water molecule is permanently trapped by the polar backbone of the peptide knot (*left inset*). When $l \gtrsim 30$ Å, the water molecule is squeezed out (*right inset*), giving rise to a significant peak in the force-extension stretching curve. This transition leads to a considerable hysteresis when stretching and loosening curves are compared. The effect is reproducible when the stretching-loosening loop is repeated (*dashed lines*).

statistical fluctuations. This indicates that our systems are close to equilibrium at the chosen pulling rate of 0.1 Å/ns.

To determine the tight knot length, $\Delta l(F) = l_c(F) - l(F)$, an accurate estimate of the force-dependent contour length $l_c(F)$ of the unknotted peptide is needed. For this, we calculate the average amino acid length $\Delta l_{aa}(F)$ by measuring the mean distance between neighboring backbone nitrogen atoms in short ($N_{aa} = 8$), unknotted peptides. From the resulting force-extension curves $F(l)$ we obtain, by inversion and division by the number of amino acids, $\Delta l_{aa}(F) = l(F)/N_{aa}$. The result, which we find to be independent of the choice of amino acid sequence, is presented in the inset to Fig. 3 *a*: below a stretching force of ~10 pN the length thermally fluctuates around $\Delta l_{aa}(F) \simeq 3.5$ Å, then rises quickly with force in a nonlinear fashion in the low-stretching, thermal regime ($F \sim$ 10–150 pN) to eventually increase linearly in the high stretching regime, $F \gtrsim 150$ pN. At $F = 1.5$ nN, a value of $\Delta l_{aa}(F) \simeq 3.8$ Å is reached. From the slope, b, of the linear part, we estimate the linear elastic modulus $\Gamma = \Delta l_{aa}(F = 0)/b \simeq 42$ nN, which is in agreement with AFM pulling experiments, where $\Gamma \simeq 50 \pm 15$ nN (46). This agreement is remarkable, since MD force fields are typically not benchmarked to be accurate at the large tensions considered in this work.

In pulling experiments, rupture of some terminal bonds at the AFM tip can occur at forces of ~100–200 pN (28), supplying thereby the relevant range for comparing to experimental TPK lengths. At $F_1 = 200$ pN we find $\Delta l_{aa}(F_1) \simeq 3.68$ Å, leading to contour length estimates $l_c(F_1) = N_{aa}\,\Delta l_{aa} \simeq 77.3$ Å and $l_c(F_1) \simeq 110.4$ Å for the trefoil and 4_1 peptide, respectively. Consequently, subtracting the

calculated end-to-end distances at F_1, we find that the tight-knot lengths for the trefoil peptides are between $\Delta l(F_1) \simeq 44.3$ Å (3_1G) and 49.8 Å (3_1L). The number of amino acids involved in the knot is thus $n_{aa} = \Delta l/\Delta l_{aa} \simeq 13 \pm 1$. For the 4_1 knots, three of the four values lie between $\Delta l \simeq 68.4$ Å and 71.9 Å ($n_{aa} \simeq 19$), whereas for the polyglycine knot (4_1G), we find $\Delta l\,(F_1) \simeq 60.4$ Å ($n_{aa} \simeq 16$), ~14% smaller. The lengths are summarized in Table 1. A typical error of these values is given by the fluctuations of the $F(l)$ curve and is roughly of amino acid size (± 4 Å).

Let us now consider more intense stretching and study the knot lengths at a larger force, $F_2 = 1$ nN. $\Delta l_{aa}(F)$ increases to $\simeq 3.76$ Å, giving rise to a slightly larger contour length for the unknotted peptides. In evaluating the particular knot lengths (see Table 1), we observe that the knots shrink in size (while the whole peptide is more stretched), as could have been anticipated. Although the pulling force is substantially increased, typically only one amino acid less is involved in a single knot, so that, surprisingly, the knot sizes vary by only a few percent for a wide range of tensions. Both of the polyglycine peptides are exceptions, however: here, the tightening effect is considerable and the final knot lengths are 20–25% smaller than those of the other studied peptides. All lengths are summarized in Table 1.

The knot lengths of $\simeq 37 - 50$ Å and $\simeq 50 - 72$ Å for the 3_1 and 4_1 knots, respectively, fall within the predicted range (1). This agreement indicates that TPK lengths are primarily determined by generic packing effects, with an effective excluded volume thickness, D, that is similar for most of the peptides, including the mixed sequences. Furthermore, macroscopic arguments roughly hold on the molecular scale. In contrast, hydrophobicity and hydrophilicity seem to have no direct influence on tight knot size in the considered force regime. A closer examination of the nature of amino acid side chains supports this statement: while glycine has basically no side chain, and packing is performed by its backbone only, a typical-sized residue with a few carbon atoms gives rise to a more difficult molecular arrangement close to or inside the tight knot. This presumably leads to the 20–25% smaller knots in the special case of polyglycine. We thus find a smaller effective thickness, $D \simeq 3.7$ Å, for polyglycine than for the other peptides, where $D \simeq 4.6 - 5.0$ Å). It is important to note that, apart from the polyglycine, the 4_1-TPK lengths are in agreement with recent AFM pulling experiments on the natively knotted bacterial phytochrome, where $\Delta l = 62 \pm 10$ Å has been measured at an $\simeq 70$ pN pulling force (29).

Illustrative simulation snapshots are shown in Fig. 4, where we plot tight knot situations for peptides 4_1G and 4_1mix, including their side chains. Large side chains obviously impede tight peptide packing. We also calculate the radius of gyration, R_g, of the knots by averaging RMS atomic distances from the geometric center of the atoms involved in the knot. We measure $R_g \simeq 7.2 \pm 0.2$ Å and $R_g \simeq 7.8 \pm 0.2$ Å for the 3_1 and 4_1 knots, respectively, with only weak dependence on the stretching force for all considered peptides apart from the polyglycine. For the latter, radii of gyration are found to be close to the values above for moderate stretching ($F \simeq 200$ pN), but 20% smaller for strong stretching ($F \gtrsim 1$ nN). These TPK sizes are larger than the typical size of biological channels, e.g., those found in the protease enzyme (11,47), which is responsible for protein degradation. Thus, a translocation or threading of a protein would indeed be blocked in vivo by the presence of a tightened knot.

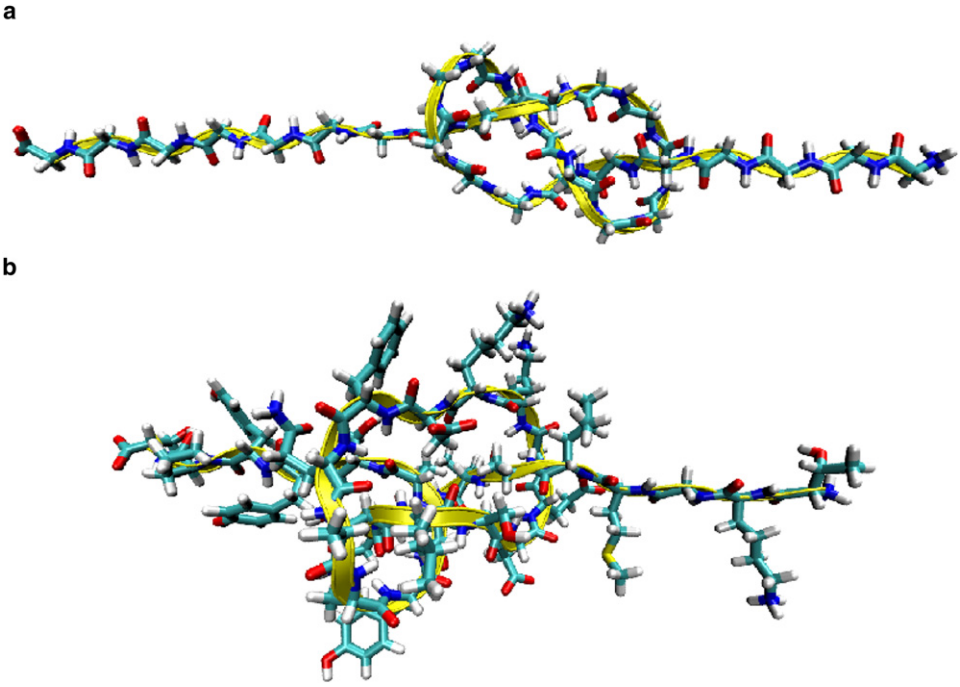

FIGURE 4 MD simulation snapshots of the tightly knotted 4_1G (*a*) and 4_1mix (*b*). The backbone is sketched in a yellow ribbon for better identification and all amino acids are resolved in a "licorice" representation ($F \simeq 1$ nN).

An intriguing topological feature appears when inspecting the overall 4_1-knot structure without the obscuring side chains, as in Fig. 2, *c* and *d*. Although the 4_1G knot is, figuratively indeed, in a ''figure-of-eight'' configuration, 4_1L displays a ''pretzel'' configuration, as illustrated in Fig. 1 *c*. Actually, we find that all of the 4_1 knots considered except 4_1G reproducibly prefer the pretzel shape when inspected by eye. This comes as a surprise, as the tight 4_1-knot configuration that minimizes (1) has been shown to be the ''figure of eight'', at least using simplifying mathematical assumptions (38). Presumably, the reasons are rather physics-based, i.e., possibly the amino acid packing and their interactions in a pretzel-like configuration minimizes the system free energy. It is interesting to note that the pretzel-like configuration can be a stable 4_1-configuration in macroscopic knots under tension, e.g., as can be easily self-demonstrated using a simple computer cable (see Fig. 1 *c*) or as taught in books on cowboy rope tricks (48). This is believed to be the first observation of a tight pretzel-like structure on microscopic scales. Recently, a somewhat looser pretzel configuration was found in collagen by transmission electron microscopy (49).

Water trapping, hysteresis, and hydrogen bonding

A striking structural feature observed in this study is the capability of some peptide knots to capture and strongly bind water molecules in their interior. The simulated peptides show this effect with varying magnitude: we find no bound water in polyglycine (3_1G and 4_1G) and the mixed peptide 3_1mix for any simulated peptide extension, whereas in 3_1E, a single trapped water molecule is reproducibly found only in the case of very close peptide packing at high forces, $F \gtrsim 1$ nN. We find stronger water-binding qualities for the other four peptides, 3_1L, 4_1E, 4_1L, and 4_1mix, for a wider range of simulated peptide extensions. Here, water was bound for simulation times of the order of ~10–100 ns per peptide, pointing to a mechanism that is quite stable. It is surprising that, on first glance, both homopeptides with the purely hydrophobic leucine side chains show the strongest water-trapping capability.

Simulation snapshots are shown in Fig. 5 for peptides 4_1E and 4_1mix: the water molecule bonds to the backbone amides in the knot interior, involving at least three hydrogen bonds per molecule, and is rotationally immobilized. Apparently, the water binding is made possible by the tight peptide packing in the highly bent knot, allowing for multiple bonds of a water molecule to the polar backbone. A particularly interesting case is the water binding in 3_1L. Here, the bound water molecule is ''squeezed out'' of the knot interior for large stretching forces ($F \simeq 1$ nN). This behavior leads to a high force peak in the force extension curve, as shown in Fig. 3 *b*: for extensions $l \lesssim 30$ Å, the water molecule is bound as shown in the left snapshot. At $l \simeq 30$ Å and $F \simeq 1$ nN, the bound water is ''wrung'' out and the force

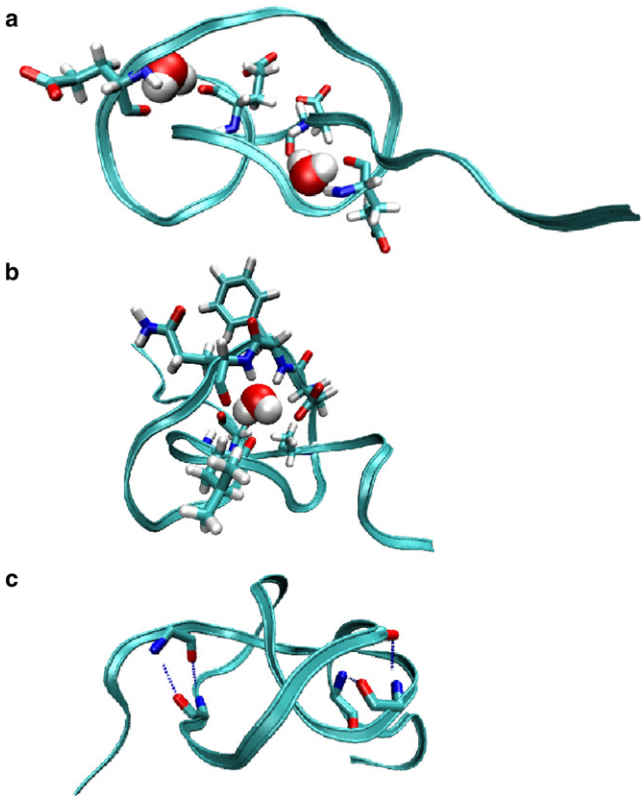

FIGURE 5 MD simulation snapshots of water trapped in peptide 4_1E (*a*) and 4_1mix (*b*) at force $F \simeq 1$ nN. The backbone is shown in ribbon structure and only those residues are sketched (in ''licorice'' representation) that are actively involved in water binding. Water (*red and white spheres*) is hydrogen-bonded to the backbone amides. (*c*) MD snapshot of 4_1mix in an unconstrained MD simulation. Four long-lived hydrogen bonds between backbone amides at the knot's ends are explicitly drawn (*dotted blue lines*).

drops significantly before further increasing. When the knot is loosened, $F(l)$ shows a considerable hysteresis. However, a water molecule is captured by the knot again during loosening at extensions $l \lesssim 27$ Å and $F \simeq 200$ pN. Repeating the stretching-loosening loop twice shows quantitative reproducibility of this effect (cf. Fig. 3 *b*). The occurrence of the hysteresis points to the fact that the water binding-unbinding events fluctuate on large timescales, and this simulation deviates therefore from equilibrium. The magnitude of the hysteresis can be estimated by integrating over the $F(l)$ stretching-loosening cycle, which gives rise to a large dissipation energy of about $\Delta G \simeq 30 - 35\ k_BT$. This value is indeed comparable to the energy of three to four hydrogen bonds between a water molecule and a peptide environment (8–10 k_BT per hydrogen bond) (50).

It is a well-known fact that buried water molecules constitute an integral part of many native protein structures, contributing to stability, flexibility, folding, and mechanical and enzymatic function (51–54). It is noteworthy that our measured ΔG is very close to the binding enthalpy of a buried water molecule in the polar pocket of bovine pancreatic trypsin inhibitor (BPTI) (55), where four hydrogen bonds

constitute $\Delta H \simeq 36\ k_BT$. We find a similar large dissipation energy in the peptide 4_1mix ($\Delta G \simeq 20\ k_BT$) and less pronounced dissipation energy in 4_1E ($\Delta G \simeq 12\ k_BT$) and 3_1E ($\Delta G \simeq 5\ k_BT$) due to partial water hydrogen binding events during knot tightening. No hysteresis is found in 4_1L, as water is bound here during the full stretching-loosening loop without any binding/unbinding transition.

It is interesting to note that in the sequence 3_1mix, we find no water trapped in the knot interior for the entire peptide extension, in contrast to 4_1mix, where we observe water trapping on a ~10-ns timescale, with three binding/unbinding events for tensions $F \lesssim 500$ pN. A closer inspection of the MD trajectory reveals that the immediate surrounding of the buried water molecule consists of six amino acids, ALD FQS, which create a mostly hydrophobic environment (see Fig. 5 *b*). This observation and the strong water binding capabilities of the polyleucines indicate that a nonpolar-side-chain environment promotes water-hydrogen bonding to the tightly packed polar backbone. We explain this by the textbook fact that hydrogen bonds are generally stronger in a nonpolar and/or desolvated protein environment (50,56), where electrostatic interactions are only weakly screened. We suspect, in addition, that the hydrophobic side chains impose a large energy barrier for possible escape of a water molecule. In the polyglutamic acids, water screening and the (probably lower) barrier is likely to be provided by the methylene groups of the side chains. A nonpolar environment is clearly absent in the polyglycines. However, a strong water trapping capability seems to result from a unique and delicate combination of local backbone structure and a specific, but rather nonpolar, amino acid side chain environment.

Related to this, another consequence of the tight peptide packing, as further revealed by our simulations, is the existence of long-lived hydrogen bonds between particular backbone amide groups. During the $\simeq$ 200-ns stretching and loosening loop of polyleucine 4_1L, for instance, we find that the backbone of amino acids 10 and 24 hydrogen-bonds for $\simeq$ 80% of the simulation time. Detailed analysis yields similar behavior for the other peptides, yielding stable intrapeptide hydrogen bonds on a long ~10- to 100-ns timescale. An exception is polyglycine, where the longest hydrogen bond life expectancy is found to be one or two orders of magnitude shorter.

Free simulations and tight knot stability

We also conduct free simulations of the knotted peptides without any constraints to check whether the knots dissolve on a typical simulation timescale. Initial configurations are taken from a stretched situation at $F \simeq 200$ pN. Dissolution of a knot is defined here by connecting the peptide ends with an imaginary line and observing whether or not we find a knot in the closed loop. Only the two polyglycine knots, 3_1G and 4_1G, show strong fluctuations and unknot quickly, on a timescale of ~10 ns. All other investigated knots do not dissolve in an $\simeq$ 100-ns simulation, pointing to a (meta)stable tight knot situation. To quantify this, we measure the root mean-square deviation (RMSD) from the initial structure of the knotted backbone part only. By our definition, this includes the amino acids 5–17 and 6–25 in the 3_1 and 4_1-knotted peptides, respectively. We find values of RMSD $\simeq$ 2 Å, increasing quickly within $\simeq$ 10 ns to 7–8 Å for the polyglycines. For the other peptides, however, the RMSD value stays small, at $\simeq$ 2 Å, for the total simulation time, supporting the observation that, apart from the polyglycines, tight knots stay stable and quite rigid after peptide stretching on relatively long timescales. The RMSD and illustrating snapshots for 4_1G and 4_1mix are presented in Fig. 6. The initial knot in 4_1G quickly dissolves into a random coil structure within 10–20 ns. In contrast, system 4_1mix displays almost the same knotted core structure after 110 ns when compared to the initial structure. An RMSD value of 1–2 Å is typical for thermal fluctuations of a stable protein structure (57). We note that dissolution of the polyglycine proceeds via a ''swelling'' rather than a ''slithering'' mechanism of the knot (where the knot stays tight and diffuses to the end), possibly to relax the highly bent backbone. This needs not to be in contrast to the study of Grosberg and Rabin (22), where an entropic tightening and slithering was predicted, as this might be the dominant mechanism for somewhat looser tight knots.

As in the constrained case, closer inspection of the knot structure reveals a few long-lived hydrogen bonds in all stable knots also in the free simulations. A representative illustration is shown in Fig. 5 *c*, where we plot an MD snapshot of 4_1mix after an $\simeq$ 100-ns free simulation. Four hydrogen bonds are found between amide backbone groups right at the knot's ends, clearly inhibiting the opening of the knot. Typically, we find that these hydrogen bonds persist on average for 80–90% of the total free simulation time, even noticeably longer for the polyleucines. In 3_1mix, the longest

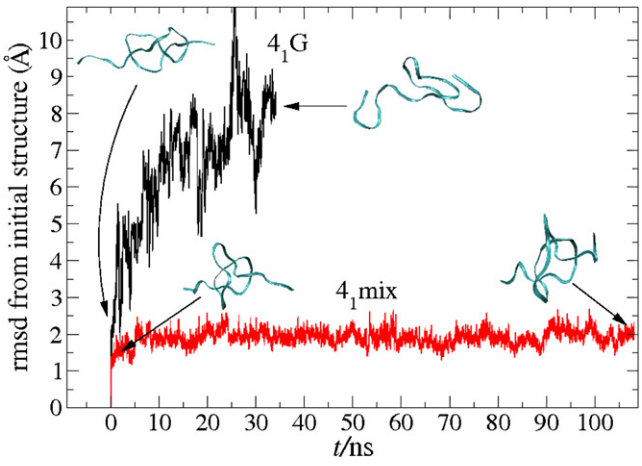

FIGURE 6 RMSD of the knotted backbone part (amino acids 6–25) from the initial configuration at $t = 0$ for 4_1G and 4_1mix in a free, unconstrained MD simulation run. The MD snapshots correspond to initial and final configurations. Note that the knotted structure of 4_1mix hardly changes in time.

hydrogen-bond lifetime is shorter and is 50–60 ns. (Recall that also in 3_1mix, no buried water could be detected.) It is worthy of note, in addition, that in 3_1L and 4_1L, one water molecule was trapped during the full, unconstrained simulation, constituting a total of seven to eight long-lived hydrogen bonds within the knots! Again the high quantity and persistence strength of hydrogen bonds in 3_1L and 4_1L must be attributed to the desolvated, strongly nonpolar side-chain environment of the tight knot.

CONCLUDING REMARKS

In summary, our MD study has revealed some generic and some specific structural behavior of tight peptide knots and provokes a few interesting conclusions and future prospects: TPKs exhibit an unexpectedly strong stability after stretching, and their radii of gyration are all indeed bigger (~7–8 Å) at moderate stretching ($F \lesssim 200$ pN) than the radius of the protease pore (~6.5 Å) (11,47). As a consequence, the steric blocking of narrow pathways by a localized or tightly pulled knot might be possible in vivo. We predict that a translocated knotted protein should get stuck in pores of diameter $\lesssim 20$ Å.

Of interest for future investigation, not only from a topological point of view (57), is the observation that most 4_1 knots are not in a figure of eight, but rather in a ''pretzel'' configuration. The pretzel might be a (meta)stable configuration in ''physical'' open tight knots, in contrast to those underlying simplifying mathematical assumptions (38). The pretzel may be preferred by the lowering of the system free energy due to favorable amino acid arrangements. This seems to be in contrast to the macroscopic pretzel, which is determined by the way the knot is tied and is then stabilized by friction. On microscopic scales, however, the configuration can be tuned by the molecular sequence.

A striking result is our finding that the TPK backbone has a strong water-binding and hydrogen-bonding capability, promoted by rather nonpolar side-chain environments. This mechanism results in ''locking'' of the tight knot structure and surprisingly stable and rigid TPKs after peptide stretching. The observed quantitative reproducibility of squeezing out and capturing a water molecule at well-defined tensions may allow for an external mechanical control of single (water)-molecule binding. This might be a useful feature in biotechnological applications. It is mportant, in this respect, that buried water is known to be an integral part of native protein structures and can be essential for protein flexibility, folding (52,53,55), and catalytic action (54).

Furthermore, TPKs might resemble structural elements of the cyclotide protein family—constituted of short ($\simeq 30$ amino acids), tightly entangled cyclic peptides—which has strong potential for drug design (13). In view of their structural complexity and molecule-binding potential, engineered protein knots thus may serve as an important model system for a deeper understanding of protein stability (58) and enzymatic activity, and may be useful for pharmaceutical purposes due to a possible catalytic function.

Finally, we would like to encourage further experimentation in this stimulating field, readily accessible by AFM (29) or optical tweezer methods (26,27). Studies on protein knot size, stability, and diffusion behavior along stretched peptides, as well as some on the refolding of knotted proteins after stretching, are desirable. Of particular interest would be a study focusing on knotted peptide translocation to verify/falsify our prediction about whether and where the protein gets stuck. Knotted protein threading may be possible in vitro using narrow biological or solid-state nanopores (59). Buried water molecules are detectable by nuclear magnetic relaxation dispersion methods (52), allowing for experimental exploration of the existence of bound water in protein knots, and consequently a means to further explore TPK fluctuations and energy landscapes.

J.D. is grateful to Thomas Bornschlögl, Katrina Forest, and Matthias Rief for pointing him toward this interesting project, Lyderic Bocquet, Ralf Metzler, Roland Netz, and Joachim Seel for useful comments, and Michael Hinczewski for a critical reading of the manuscript. Computing time on the HLRBII computer cluster of the Leibniz-Rechenzentrum München is acknowledged.

The author acknowledges the support of the Deutsche Forschungsgemeinschaft (DFG) through the Emmy-Noether-Program.

REFERENCES

1. Mansfield, M. L. 1994. Are there knots in proteins? *Nat. Struct. Biol.* 1:213–214.
2. Taylor, W. R., and K. Lin. 2003. A tangled problem. *Nature*. 421:25.
3. Lua, R. C., and A. Y. Grosberg. 2006. Statistics of knots, geometry of conformations, and evolution of proteins. *PLOS Computational Biology*. 2:e45.
4. Virnau, P., L. A. Mirny, and M. Kardar. 2006. Intricate knots in proteins: function and evolution. *PLOS Comput. Biol.* 2:e122.
5. Reference deleted in proof.
6. Adams, C. C. 1994. The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots. W.H. Freeman, New York.
7. Reference deleted in proof.
8. Taylor, W. R. 2007. Protein knots and fold complexity: some new twists. *Comput. Biol. Chem.* 31:151–162.
9. Alam, M. T., T. Yamada, U. Carlsson, and A. Ikai. 2002. The importance of being knotted: effects of the C-terminal knot structure on enzymatic and mechanical properties of bovine carbonic anhydrase II. *FEBS Lett.* 519:35–40.
10. Wallin, S., K. B. Zeldovich, and E. I. Shakhnovich. 2007. The folding mechanics of a knotted protein. *J. Mol. Biol.* 368:884–893.
11. Prakash, S., and S. Matouschek. 2004. Protein unfolding in the cell. *Trends Biochem. Sci.* 29:593–600.
12. King, N. P., E. O. Yeates, and T. O. Yeates. 2007. Identification of rare slipknots in proteins and their implications for stability and folding. *J. Mol. Biol.* 373:153–166.
13. Rosengren, K. J., N. L. Daly, M. R. Plan, C. Waine, and D. J. Craik. 2002. Twists, knots, and rings in proteins: structural definition of the cyclotide framework. *J. Biol. Chem.* 278:8606–8616.
14. Breault, G. A., C. A. Hunter, and P. C. Mayers. 1999. Supramolecular topology. *Tetrahedron.* 55:5265–5293.

15. Williams, A., B. H. Northrop, T. Chang, J. F. Stoddart, A. J. P. White, et al. 2006. Suitanes. *Angew. Chem. Int. Ed.* 40:6665–6669.

16. Katritch, V., W. K. Olson, A. Vologodskii, J. Dubochet, and A. Stasiak. 2000. Tightness of random knotting. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 61:5545–5549.

17. Virnau, P., V. Kantor, and M. Kardar. 2005. Knots in globule and coil phases of a model polyethylene. *J. Am. Chem. Soc.* 127:15102–15106.

18. Belmonte, A. 2007. The tangled web of self-tying knots. *Proc. Natl. Acad. Sci. USA.* 104:17243–17244.

19. Kardar, M. 2007. The elusiveness of polymer knots. *Eur. Phys.J. B.* 64:519–523.

20. de Gennes, P. -G. 1984. Tight knots. *Macromolecules.* 17:703–704.

21. Dommersnes, P. G., Y. Kantor, and M. Kardar. 2002. Knots in charged polymers. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 66, 031802.

22. Grosberg, A. Y., and Y. Rabin. 2007. Metastable tight knots in a wormlike polymer. *Phys. Rev. Lett.* 99:217801.

23. Metzler, R., A. Hanke, P. G. Dommersnes, Y. Kantor, and M. Kardar. 2002. Equilibrium shapes of flat knots. *Phys. Rev. Lett.* 88:188101.

24. Ercolini, E., F. Valle, J. Adamcik, G. Witz, R. Metzler, et al. 2007. Fractal dimension and localization of DNA knots. *Phys. Rev. Lett.* 98, 058102.

25. Marcone, B., E. Orlandini, and A. L. Stella. 2007. Knot localization in adsorbing polymer rings. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76, 051804.

26. Arai, Y., et al. 1999. Tying a molecular knot with optical tweezers. *Nature.* 399:446–448.

27. Bao, X. R., H. J. Lee, and S. R. Quake. 2003. Behavior of complex knots in single DNA molecules. *Phys. Rev. Lett.* 91:265506.

28. Hugel, T., and M. Seitz. 2001. The study of molecular interactions by AFM force spectroscopy. *Macromol. Rapid Commun.* 22: 989–1016.

29. Bornschlögl, T., D. Anstrom, J. Dzubiella, M. Rief, and K. T. Forest. 2009. Tightening the knot in phytochrome by single molecule atomic force microscopy. *Biophys. J.* In press.

30. Saitta, A. M., P. D. Soper, E. Wasserman, and M. L. Klein. 1999. Influence of a knot on the strength of a polymer strand. *Nature.* 399:46–48.

31. Mansfield, M. L. 1997. Tight knots in polymers. *Macromolecules.* 31:4030–4032.

32. Farago, O., Y. Kantor, and M. Kardar. 2002. Pulling knotted polymers. *Europhys. Lett.* 60:53–59.

33. Arteca, G. A. 2007. Externally steered relaxation of tight polyethylene tangles with different initial knot topologies. *Theor. Chem. Acc.* 118:549–556.

34. Vologodskii, A. 2006. Brownian dynamics simulation of knot diffusion along a stretched DNA molecule. *Biophys. J.* 90:1594–1597.

35. Huang, L., and D. E. Makarov. 2007. Langevin dynamics simulations of the diffusion of molecular knots in tensioned polymer chains. *J. Phys. Chem. A.* 111:10338–10344.

36. Sułkowska, J. I., P. Sułkowski, P. Szymczak, and M. Cieplak. 2008. Tightening of knots in proteins. *Phys. Rev. Lett.* 100:058106.

37. Metzler, R., W. Reisner, R. Riehn, R. Austin, J. O. Tegenfeldt, et al. 2006. Diffusion mechanisms of localised knots along a polymer. *Europhys. Lett.* 76:696–702.

38. Pierański, P., S. Przybył, and A. Stasiak. 2001. Tight open knots. *Eur. Phys. J. E.* 6:123–128.

39. Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of macromolecules: a perspective. *Nat. Struct. Mol. Biol.* 9:646–652.

40. Case, D. A., T. Darden, T. E. Cheatham, III, C. Simmerling, J. Wang, et al. 2006. AMBER9.0. University of California, San Francisco.

41. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38.

42. Lim, K., H. Zhang, A. Tempczyk, W. Krajewski, N. Bonander, et al. 2003. Structure of the YibK methyltransferase from *Haemophilus influenzae* (HI0766): a cofactor bound at a site formed by a knot. *Proteins Struct. Funct. Genet.* 51:56–67.

43. Biou, V., R. Dumas, C. Cohen-Addad, R. Douce, D. Job, et al. 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 Å resolution. *EMBO J.* 16:3405–3415.

44. Pleshe, E., J. Truesdell, and R. T. Batey. 2005. Structure of a class II TrmH tRNA-modifying enzyme from *Aquifex aeolicus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 61:722–728.

45. van Roon, A. -M. M., N. M. Loening, E. Obayashi, J. -C. Yang, A. J. Newman, et al. 2008. Solution structure of the U2 snRNP protein Rds3p reveals a knotted zinc-finger motif. *Proc. Natl. Acad. Sci. USA.* 105:9621–9626.

46. Ptak, A., S. Takeda, C. Nakamura, J. M. M. Kageshima, S. P. Jarvis, et al. 2001. Modified atomic force microscope applied to the measurement of elastic modulus for a single peptide molecule. *J. Appl. Phys.* 90:3095–3099.

47. Pickart, C. M., and A. P. VanDemark. 2000. Opening doors into the proteasome. *Nat. Struct. Mol. Biol.* 7:999–1001.

48. Mason, B.S. 1928. How to Spin a Rope: Lariat Throwing, Rope Spinning and Trick Cowboy Knots. Self-published, Columbus, OH. http://www.inquiry.net/outdoor/spinrope/trickknots.htm.

49. Myers, J. C., P. S. Amenta, A. S. Dion, J. P. Sciancalepore, C. Nagaswami, et al. 2007. The molecular structure of human tissue type XV presents a unique conformation among the collagens. *Biochem. J.* 404:535–544.

50. Jackson, M. B. 2006. Molecular and Cellular Biophysics. Cambridge University Press, Cambridge, UK.

51. Baker, E. N. 1995. Solvent interactions with proteins as revealed by X-ray crystallographic studies. *In* Protein-Solvent Interactions. R. B. Gregroy, editor. Marcel Dekker, New York, pp. 143–189.

52. Denisov, V. P., J. Peters, H. D. Hörlein, and B. Halle. 1996. Using buried water molecules to explore the energy landscape of proteins. *Nat. Struct. Biol.* 3:505–509.

53. Dougan, L., G. F. H. Lu, and J. M. Fernandez. 2008. Solvent molecules bridge the mechanical unfolding transition state of a protein. *Proc. Natl. Acad. Sci. USA.* 105:3185–3190.

54. Ball, P. 2008. Water as an active constituent in cell biology. *Chem. Rev.* 108:74–108.

55. Fischer, S., and C. S. Verma. 1999. Binding of buried structural water increases the flexibility of proteins. *Proc. Natl. Acad. Sci. USA.* 96:9613–9615.

56. Roseman, M. A. 1988. Hydrophobicity of the peptide C=OH—N hydrogen-bonded group. *J. Mol. Biol.* 201:621–623.

57. Katritch, V., J. Bednar, D. Michoud, R. G. Scharein, J. Dubochet, et al. 1996. Geometry and physics of knots. *Nature.* 384:142–145.

58. Yeates, T., S. Norcross, and N. P. King. 2007. Knotted and topologically complex proteins as models for studying folding and stability. *Curr. Opin. Chem. Biol.* 11:595–603.

59. Dekker, C. 2007. Solid-state nanopores. *Nat. Nanotechnol.* 2:209–215.