Supercoiling in a Protein Increases its Stability

Szymon Niewieczerzał
Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland

Joanna I. Sulkowska
Department of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland

(Received 11 March 2019; published 26 September 2019)

The supercoiling motif is the most complex type of nontrivial topology found in proteins with at least one disulfide bond and, to the best of our knowledge, it has not been studied before. We show that a protein from extremophilic species with such a motif can fold; however, the supercoiling changes a smooth landscape observed in reduced conditions into a two-state folding process in the oxidative conditions, with a deep intermediate state. The protein takes advantage of the hairpinlike motif to overcome the topological barrier and thus to supercoil. We find that the depth of the supercoiling motif, i.e., the length of the threaded terminus, has a crucial impact on the folding rates of the studied protein. We show that fluctuations of the minimal surface area can be used to measure local stability, and we find that supercoiling introduces stability into the protein. We suggest that the supercoiling motif enables the studied protein to live in physically extreme conditions, which are detrimental to most life on Earth.

DOI: 10.1103/PhysRevLett.123.138102

Introduction.—More than one thousand proteins deposited in the Protein Data Bank (PDB) contain knots or slipknots [1]. They come from all kingdoms of life, in organisms separated even by one billion years of evolution. These proteins create a challenge from many perspectives—protein folding, degradation, influence of topology on biological activity of these proteins [2,3]. Nonetheless, only the simplest types of knots have been identified in the PDB [4]. Recent study, however, has revealed that proteins also form another type of a nontrivial topology: the so-called lassos [5]. In this case the nontrivial topology of the protein chain arises in the presence of disulfide bonds in appropriate positions, so that a fragment of a protein forms a closed loop through which each of the termini can be threaded (see Supplemental Material [6], Fig. S1). These structures should not be confused with cystine knots [18]. Currently, 18% of nonredundant proteins deposited in the PDB form lassos [7]. The most intriguing lasso proteins are those with the most entangled motif identified so far, the so-called supercoiled lasso, where one of the tails crosses the loop more than once in the same direction as shown in Fig. 1(a).

Supercoiling topology is found, for example, in Pfu-542154 protein from an extremophilic species of Archaea that thrives in physically or geochemically extreme conditions that are detrimental to most life on Earth [19,20]. Such organisms have an optimum growth temperature of 100 °C—a temperature that would destroy most living organisms. It appears as if, in such cases, nature invented supercoiling to protect proteins against harsh conditions as well as to allow for the internal flexibility necessary, e.g., to bind ligands. Note that an ortholog to the Pfu-542154 protein from a human [8] has a trivial topology.

The presence of disulfide bonds is directly related to the protein's flexibility and its thermodynamical and mechanical stability [21]. More recently, it was also shown that a protein with a nontrivial topology is more stable than the same one with a trivial topology and the same number of disulfide bonds [22]. The analysis of the simplest lasso showed that this motif facilitates binding to a receptor, but disturbs the folding pathway of a protein [23,24]. Thermostable enzymes are commonly engineered, e.g., via disulphide bridges, for industrial applications [25]. Structures with different topology have been also synthesized using proper polymeric compounds and are intensively investigated [26–29].

To the best of our knowledge, the lasso supercoiling as a topological motif has not been studied in detail in proteins to date. On the other hand, the standard supercoiling has been extensively studied in polymeric systems, especially in the context of DNA [30–32]. The existence of supercoiled proteins leads to many intriguing questions never asked before. Why did nature design such complex structures? How do they fold? Does the topology indeed introduce stability? How to measure topological stability? In this Letter, based on the analysis of a supercoiled protein from extremophilic species, we answer these questions and introduce general methods to study the threading mechanism.

Materials and methods.—We analyzed the dynamics of a protein by means of the C_α coarse-grained structure based model [9], using Gromacs with SMOG software [10]. The progress of the folding process was studied via the minimal surface approach [5]. By Q we denote a fraction of native

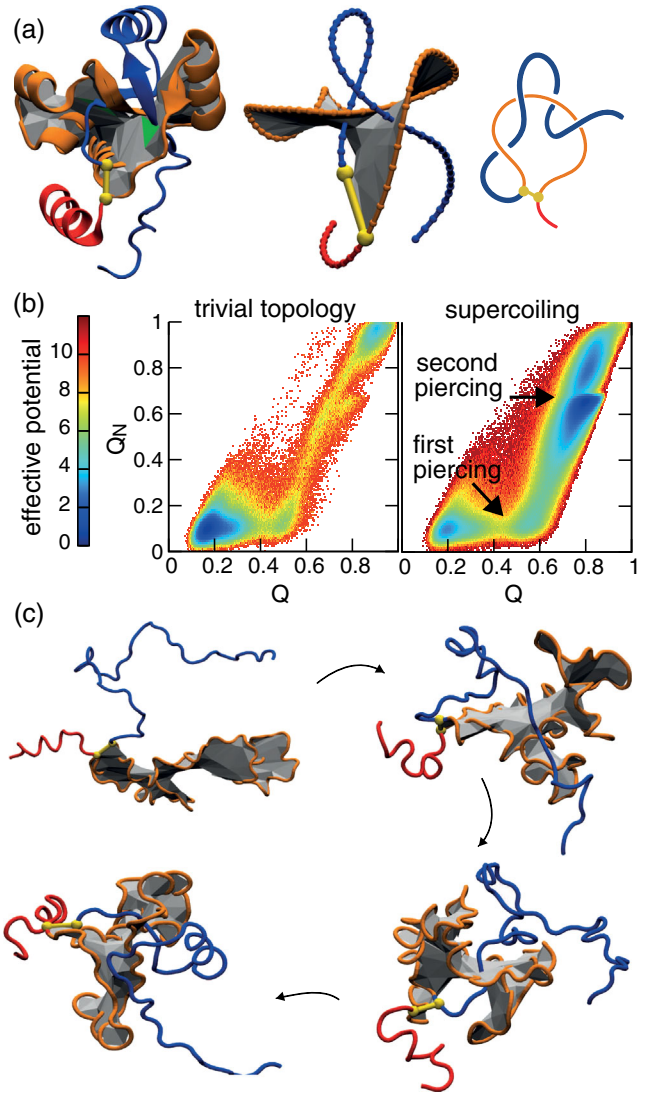

FIG. 1. (a) Protein 1zd0 in ribbon representation (left), with smoothed backbone (middle), and in schematic representation (right). The minimal surface (gray), spanned by the protein's backbone (orange) closed by a disulfide bond (yellow), is pierced twice in the same direction by the N terminal. (b) Effective energy landscape as a function of fraction of total native contacts and native contacts formed by amino acids at the N-terminal tail, obtained from the folding process data for 1zd0 without a disulfide bond, and with a ds bond with 3 amino acids removed. (c) First stage of the 1zd0 folding process. First fragments forming the loop gain secondary structure and adopt nativelike conformations within the loop. The intermediate stabilizes the structure and allows the N terminal to find itself in the vicinity of the accessible fragment of the loop, through which it can be threaded to adopt the supercoiled lasso topology and reach the native state. For a broader comparison see the Supplemental Material [6], Fig. S5.

contacts present in a given conformation [11]. The third reaction coordinate, denoted Q_N, is the fraction of contacts formed by first the 36 amino acids in the native structure of 1zd0; see the Supplemental Material [6].

In order to identify lassos and to determine whether and how a given terminal of the chain pierces a covalent loop, we further develop the method described in Ref. [5]. Our procedure is as follows: we determine a triangulated surface of minimal area spanned by the loop closed by a disulfide bond. Then, we check which segment of the tail crosses one of the triangles comprising the surface and, finally, we assign the direction of crossing to distinguish a slipknot from a supercoiling geometry.

Results and discussion.—Supercoiling in Pfu-542154 protein: We focus on Pfu-542154 (PDB code 1zd0), the protein with an unknown function from Pyrococcus furiosus, an extremophilic species of Archaea. In the native state, the protein has two cysteine residues, 48 and 131. Under oxidative conditions, they form a disulfide bond. In this way, a fragment of the protein backbone forms a closed loop. Under reducing conditions, the disulfide bond breaks and the loop disappears. This loop is pierced twice, each time in the same direction, by the N-terminus fragment. Thus, in oxidative conditions, Pfu-542154 forms a supercoiled lasso—one of the most complex entangled motifs identified in proteins to date [7].

We use the following abbreviations. The protein in reduced and the oxidative state is denoted, respectively, as 1zd0-noSS and 1zd0. The protein with a different depth of supercoiling motif (length of the N-terminal tail) is denoted according to the number of amino acids removed; e.g., 1zd0-2aa and 1zd0-3aa denote, respectively, a protein in oxidative conditions with 2 and 3 residues removed. Note that removing up to 8 amino acids from the N terminus does not affect the enthalpy contribution to the free energy landscape (see the Supplemental Material [6], Fig. S2).

Folding thermodynamics and kinetics: The folding mechanism of 1zd0-noSS turns out to be a two-state process with no stable intermediate (the free-energy landscape is provided in the Supplemental Material [6], Fig. S3). The presence of the disulfide bond significantly alters this process leading to a profound folding slowdown, due to which we were unable to collect data at equilibrium conditions (details in the Supplemental Material [6]). Therefore, we concentrated on the folding kinetics of 1zd0, which we studied at T_f characteristic for 1zd0-noSS.

The potential of mean force (PMF) calculated for 1zd0-noSS and 1zd0 reveals another minimum at $Q \approx 0.8$ [see Figs. 1(b), 1(c)]. The related stable intermediate state is structurally close to the native state; only the final fragment of the N-terminal tail does not form all its native contacts. In the beginning of the 1zd0 folding process, the loop together with the C-terminal tail adopt the nativelike arrangement, and the first crossing event takes place, in which the N-terminal tail close to the loop crosses the topological barrier (the closed loop) in a slipknot conformation (see the Supplemental Material [6], Fig. S4). In the second stage, the final fragment of the N-terminal tail must be threaded again through the loop.

The comparison of 1zd0-noSS and 1zd0 kinetics is not straightforward, because we observe no intermediate state in reduced conditions. Therefore, we arbitrarily chose the moment when the protein reaches $Q = 0.8$. In the case of 1zd0-noSS, this value of Q is reached slower then the intermediate state for 1zd0; see Fig. 2. However, since there is no conformational or topological barrier along the folding pathway for 1zd0-noSS, very shortly afterwards the protein reaches its native conformation. We stress that conformations with $Q = 0.8$ for 1zd0-noSS do not constitute the same ensemble as in the case of 1zd0. In the presence of the disulfide bond, the fragment of the polypeptide chain containing the loop and a fragment of the N-terminal tail, stabilized by the disulfide bond, arranges quite fast in the nativelike way in comparison with the folding time of the whole protein. These results suggest that the closed loop plays a stabilizing role as a nucleation site to accelerate the initial folding stage.

The kinetics of the second stage of the 1zd0 folding strongly depends on the length of the threading N-terminal fragment, and it is discussed in detail below.

The second piercing is a rate limiting step: The loop in the nativelike arrangement does not leave enough space for the N terminus to enable its piercing. Meanwhile, the N-terminal fragment tries to push itself through the loop, to form the native contacts with amino acids on the other side of the loop (for the schematic representation of the final folding stage, a more detailed description and qualitative characteristics such as the fluctuation of a minimal area and number of piercings, see the Supplemental Material [6], Figs. S6 and S7). Mostly, during these efforts, two additional piercings are observed as a result of a slipknot formed by the N terminus in an attempt to cross the loop (see the Supplemental Material [6], Fig. S7). The successful

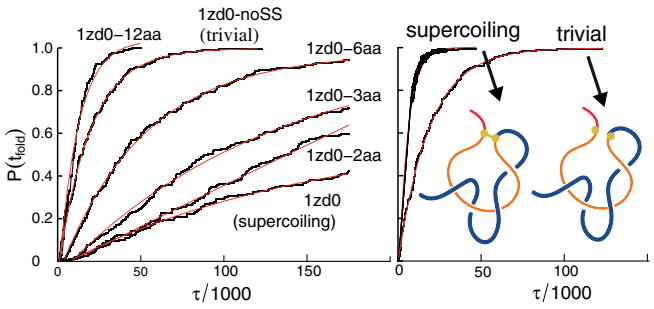

FIG. 2. A fraction of the trajectories in which the protein reached the intermediate state (right) and the native state (left) as a function of time (in the reduced units) for various depths of the supercoiling motif (lengths of the N-terminal tail). The curves were fitted to the function $\mathbf{f}(\mathbf{t}) = \mathbf{a}(\mathbf{1} - \mathbf{e}^{-\mathbf{b}(\mathbf{t}-\mathbf{t_0})})$ describing the kinetics first-order increase of the product. Fitted parameters are presented in the Supplemental Material [6], Table SI. Data from simulations for 1zd0, −2aa, −3aa, −6aa, and −12aa, converge for the intermediate state; therefore, we present the fitted curve for 1zd0 only. Removing up to 8 amino acids does not affect the enthalpic part of the effective energy landscape of the protein.

piercing occurs after a temporary disruption of the loop resulting from the action of pushing the N terminus. Then, the terminus attracted by the structure on the other side of the loop can go through it, or be threaded to an extent sufficient to get through gradually in spite of the following closure of the loop.

Local stabilization due to the second piercing: To understand the process of entanglement and its influence on the stability of the 1zd0, we introduce new coordinates that enable us to parametrize the topological and structural features of the system. To this end we employ the area of the minimal surface and the number of piercings through the loop by the polypeptide chain. In the first folding stage almost 80% of native contacts are formed, which is strongly correlated with the first piercing through the loop, and the surface of the loop decreases by about 20%, from 1330 ± 150 Å^2 in the unfolded state to 1080 ± 34 Å^2 in the intermediate state, while fluctuations drop by about 77% (see Fig. 3). When the protein stays in the intermediate state, we observe two distinct cases well described by the loop surface area and its fluctuations. When the N-terminal tail crosses the loop in the slipknotlike way, the observed average value of the loop surface area is equal to 1095 ± 34 Å^2 (for 1zd0 conformation with three piercings see the Supplemental Material [6], Fig. S8). In the case when the N-terminal tail does not pierce the loop, the loop surface area is smaller: 1078 ± 31 Å^2.

The final threading frequently occurs shortly after a local backtracking [12], see Fig. 3, and Fig. S7 and Fig. S9 in the Supplemental Material [6]. After the second piercing occurs, the average surface of the loop does not change much and on average it is equal to 1077 ± 22 Å^2, but we observe a decrease of its fluctuations by one-third—denoted by a black circle in Fig. 3. Just after the second piercing the remaining native contacts are formed, as the protein adopts the native conformation. The direct correlation of the threading event

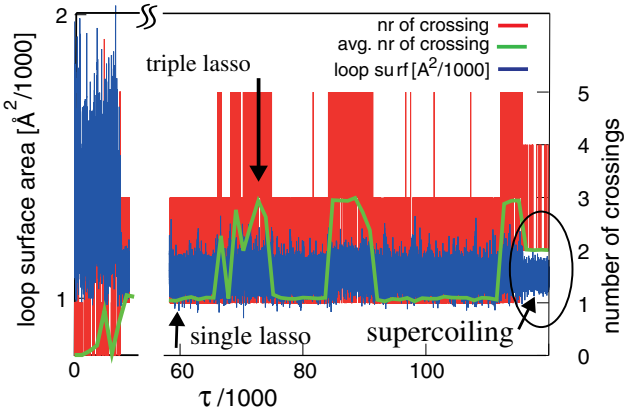

FIG. 3. Time evolution of the surface area of the loop (blue), discrete (red), and averaged (green) number of crossings of the polypeptide chain through the loop for a sample folding trajectory of 1zd0.

and the decrease of fluctuations implies that the change of topology introduces stability to the protein.

Simulation studies of knotted polymers (and also star polymers) show that fluctuations of variables such as hydrodynamic radius, radius of gyration and intrinsic viscosity, decrease with the topological complexity, and at the same time, the rigidity of such polymers increases [33,34]. Such features might be expected in a supercoiled protein behavior in the early stages of the folding process, when there is a relatively small number of native contacts present. For 1zd0, the first piercing occurs in the initial stage of folding, and it is accompanied by a significant decrease in fluctuations of the surface area. The protein then quickly becomes highly structured and stays in one of the two states: (i) with one piercing, or (ii) with three piercings (with a slipknot at the N-terminal end), before it reaches (iii) the native state (with 2 piercings through the loop).

Generally, introduction of the nontrivial topology is expected to decrease polymer's fluctuations. In the case of knotted polymers, reduction in fluctuations of R_g for increasing knot's complexity (and hence crossing number), is a consequence of a global increase in topological constraints. In the case of slipknot lassos, the constraints are localized inside the minimal surface of the loop. In the supercoiled lasso conformation, constraints are inside as well as outside of the spanning surface of the loop. Therefore, the expected fluctuations should be larger in the case of a slipknot lasso than for a supercoiled lasso. This is what we observe comparing the native state (iii) with the state in which the nativelike arranged loop is pierced by a slipknot (ii). Moreover, we would expect fluctuations in (ii) to be smaller than in (i) because of the constraints imposed on the slipknot loop. However, in this case, the presence of a slipknot in the state (ii) leads to unfavorable conformational changes which lead to the opposite relation, although the fluctuations are comparable.

Depth of supercoiling (N-terminal end extension length) and slowing down kinetics: The time needed for the N-terminal end to get through the available part of the loop very strongly depends on the length of the threaded fragment. We studied this dependence by considering 1zd0 in oxidized conditions with different N-terminal lengths (see Fig. 4). The median times needed to arrive at the intermediate state and in the folded state are presented in the Supplemental Material [6], Table SII. For all considered versions of 1zd0, times needed to reach the intermediate state were very similar, but it is worth mentioning that the shortest time was obtained for the shortest version of 1zd0 (1zd0-12aa, time: $3.4 \times 10^3\ \tau$), while 1zd0-noSS needed $10^3\ \tau$ longer time to reach the same Q.

The total folding time for 1zd0-12aa is the fastest of all considered versions. For 20% of folding trajectories the threading event was almost instantaneous and in the rest of the trajectories, we observed some delay in reaching the

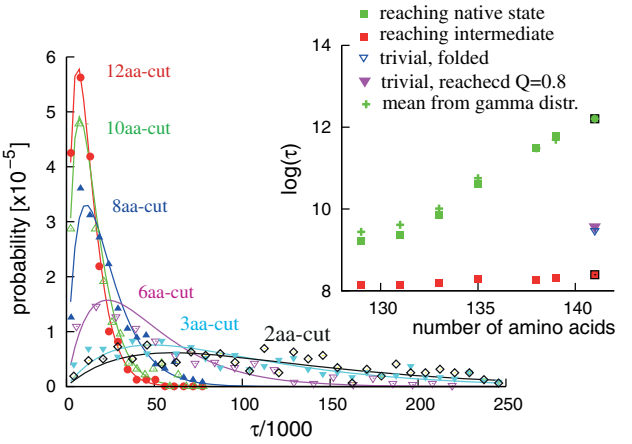

FIG. 4. Distribution of folding times of 1zd0 with the disulfide bond and varying length of the N-terminal end. Γ-distribution fits to data points from 500 trajectories for each protein length. Inset: Median folding time as a function of the number of cutted amino acids at the N-terminal end of 1zd0.

folded state. For longer N termini there were no fast folding events, and the mean folding times are significantly longer than the time needed for reaching the second stage. For 1zd0, the topological factor plays a crucial role in the folding process. The presence of the stable intermediate state slows folding significantly, even though about 80% of the structure adopts the nativelike conformation much quicker than in the case of 1zd0-noSS.

Conclusions.—In this work we presented the first detailed study of a protein with the supercoiling topology. Based on different reaction coordinates we showed directly how the nontrivial topology affects the free (effective) energy landscape. The native contacts are sufficient to fold (self-tie) a protein. The supercoiling topology introduces intermediate states in contrast to a smooth folding in reduced conditions. Based on kinetics data we showed that for the final threading event there is a topological barrier, which increases with the depth of supercoiling.

We described the final threading—the second stage of folding, after reaching the intermediate state—quantitatively and qualitatively based on the analysis of the surface area of the loop and the number of piercings along the folding process in oxidized conditions. The time evolution of the surface area is strictly correlated with the number of crossings through the loop. First, a significant reduction takes place when the first piercing and the intermediate is reached. Then, fluctuations of the surface area become much smaller, and finally decrease after the final piercing event. The rise of the topological complexity in two knottinglike events leads in both cases to a significant relative decrease of fluctuations of the investigated protein. This rise of the stability of the system can be a result of two factors: (i) imposed topological restraints, and (ii) increase of established native contacts. In that case our analysis shows that the nontrivial topology (in majority the second

piercing) is responsible for stabilization of the protein, as it was expected for a long time but hard to observe explicitly in the case of knotted proteins.

We found that due to the reduction of conformational entropy in the presence of the disulfide bond (supercoiling topology), the calculated time of reaching the value of Q characteristic for the intermediate state observed in reduced conditions, was 3 times longer than in the case of the absence of the disulfide bond. But the total folding time calculated for the protein without the disulfide bond is almost tenfold shorter than for the oxidized variant 1zd0-2aa.

Our bioinformatics approach reveals that the protein 1zd0 is an ortholog of the human and archaeal Cgi121 [8] (for details see also the Supplemental Material [6], Fig. S10). Even though all orthologs (domains) share the same fold, any of CYS—the amino acid which has very high probability to be conserved—is not conserved in the human and archaeal Cgi121 domain. Consequently, only the 1zd0 possesses a nontrivial topology. This suggests that even though supercoiling topology hinders folding, it may be crucial to secure internal flexibility while protecting the 1zd0 domain from geochemically extreme conditions that are detrimental to most life on Earth [19,20]. Since the genes coding this domain evolved from the same common gene ancestor, further bioinformatics study can reveal the origin of the nontrivial topology. Furthermore, our findings show that such topology can be used as a natural template to design superstable artificial biomolecules.

Scientific work was financed from Polish Ministry for Science and Higher Education No. 0003/ID3/2016/64 to J. I. S., National Science Centre No. 2012/07/E/NZ1/01900 to J. I. S., and EMBO YIP to J. I. S. The authors thank Vasilina Zayats and Paweł Dabrowski-Tumanski for constructive discussions.

[1] M. Jamroz, W. Niemyska, E. J. Rawdon, A. Stasiak, K. C. Millett, P. Sułkowski, and J. I. Sulkowska, Nucleic Acids Res. **43**, D306 (2014).
[2] P. Dabrowski-Tumanski and J. I. Sulkowska, Polymers **9**, 454 (2017).
[3] S. E. Jackson, A. Suma, and C. Micheletti, Curr. Opin. Struct. Biol. **42**, 6 (2017).
[4] J. I. Sułkowska, E. J. Rawdon, K. C. Millett, J. N. Onuchic, and A. Stasiak, Proc. Natl. Acad. Sci. U.S.A. **109**, E1715 (2012).
[5] W. Niemyska, P. Dabrowski-Tumanski, M. Kadlof, E. Haglund, P. Sułkowski, and J. I. Sulkowska, Sci. Rep. **6**, 36895 (2016).
[6] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.123.138102 for additional figures containing schemes of threading scenarios, free energy landscape for folding in reduced conditions, and additional sequential and structural analysis, which includes Refs. [7–17].
[7] P. Dabrowski-Tumanski, W. Niemyska, P. Pasznik, and J. I. Sulkowska, Nucleic Acids Res. **44**, W383 (2016).
[8] D. Y. Mao, D. Neculai, M. Downey, S. Orlicky, Y. Z. Haffani, D. F. Ceccarelli, J. S. Ho, R. K. Szilard, W. Zhang, and C. S. Ho, Mol. Cell **32**, 259 (2008).
[9] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[10] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic, Nucleic Acids Res. **38**, W657 (2010).
[11] P. C. Whitford, J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu, and J. N. Onuchic, Proteins: Struct., Funct., Bioinf. **75**, 430 (2009).
[12] D. T. Capraro, M. Roy, J. N. Onuchic, and P. A. Jennings, Proc. Natl. Acad. Sci. U.S.A. **105**, 14844 (2008).
[13] H. Lammert, A. Schug, and J. N. Onuchic, Proteins Struct. Funct. Bioinf. **77**, 881 (2009).
[14] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).
[15] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).
[16] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, J. Comput. Chem. **25**, 1605 (2004).
[17] J. K. Noel, J. I. Sułkowska, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **107**, 15403 (2010).
[18] D. J. Craik, N. L. Daly, T. Bond, and C. Waine, J. Mol. Biol. **294**, 1327 (1999).
[19] P. H. Rampelotto, Sustainability **2**, 1602 (2010).
[20] L. J. Rothschild and R. L. Mancinelli, Nature (London) **409**, 1092 (2001).
[21] A. A. Dombkowski, K. Z. Sultana, and D. B. Craig, FEBS Lett. **588**, 206 (2014).
[22] P. Dabrowski-Tumanski and J. I. Sulkowska, Proc. Natl. Acad. Sci. U.S.A. **114**, 3415 (2017).
[23] E. Haglund, J. I. Sułkowska, Z. He, G.-S. Feng, P. A. Jennings, and J. N. Onuchic, PLoS One **7**, e45654 (2012).
[24] E. Haglund, J. I. Sulkowska, J. K. Noel, H. Lammert, J. N. Onuchic, and P. A. Jennings, PLoS Comput. Biol. **10**, e1003613 (2014).
[25] F. Rigoldi, S. Donini, A. Redaelli, E. Parisini, and A. Gautieri, APL Bioeng. **2**, 011501 (2018).
[26] T. Yamamoto and Y. Tezuka, Polym. Prepr. **2**, 1930 (2011).
[27] Y. Doi, Y. Ohta, M. Nakamura, A. Takano, Y. Takahashi, and Y. Matsushita, Macromolecules **46**, 1075 (2013).
[28] C.-L. Fu, Z.-Y. Sun, H.-F. Li, L.-J. An, and Z. Tong, Polymer **49**, 3832 (2008).
[29] E. Uehara, R. Tanaka, M. Inoue, F. Hirose, and T. Deguchi, React. Funct. Polym. **80**, 48 (2014).
[30] L. Coronel, A. Suma, and C. Micheletti, Nucleic Acids Res. **46**, 7533 (2018).
[31] G. Witz, G. Dietler, and A. Stasiak, Proc. Natl. Acad. Sci. U.S.A. **108**, 3608 (2011).
[32] J. Huang, T. Schlick, and A. Vologodskii, Proc. Natl. Acad. Sci. U.S.A. **98**, 968 (2001).
[33] F. Vargas-Lara, A. M. Hassan, M. L. Mansfield, and J. F. Douglas, Sci. Rep. **7**, 13374 (2017).
[34] N. Kanaeda and T. Deguchi, Phys. Rev. E **79**, 021806 (2009).