

Molecular Dynamics Simulation of the α -Helix to β -Sheet Transition in Coiled Protein Filaments: Evidence for a Critical Filament Length Scale

Zhao Qin and Markus J. Buehler*

Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue Room 1-235A&B, Cambridge, Massachusetts, 02139, USA

(Received 27 September 2009; published 12 May 2010)

The alpha-helix to beta-sheet transition (α - β transition) is a universal deformation mechanism in alpha-helix rich protein materials such as wool, hair, hoof, and cellular proteins. Through a combination of molecular and theoretical modeling, we examine the behavior of alpha-helical coiled-coil proteins with varying lengths under stretch. We find that the occurrence of the α - β transition is controlled by the length of constituting alpha-helical proteins. In the asymptotic limit, short proteins with less than 26 amino acids or 3.8 nm length reveal interprotein sliding, whereas proteins with greater lengths feature an α - β transition, leading to a significant increase in the protein's stiffness, strength, and energy dissipation capacity at large deformation. Our study elucidates the fundamental physics of this mechanism and explains why the α - β transition typically occurs in protein filaments with long alpha-helical domains.

DOI: 10.1103/PhysRevLett.104.198304

PACS numbers: 82.37.Rs, 62.20.M-, 62.25.-g, 87.15.bd

Alpha helices and beta sheets are major secondary structural motifs that organize the three-dimensional structure of proteins [1]. Whereas the sequence of amino acids defines a protein's specific secondary and higher level structures [2], the transition between alpha helices and beta sheets (α - β transition) is observed universally for a broad range of alpha-helix based protein filaments. The α - β transition was observed under variations of pH [3], temperature changes [4], and solvent composition alterations [5], and under mechanical deformation as shown by experiment and simulation [6–10]. Many alpha-helix rich proteins are crucial for structural support, in mechano-transduction and cell motility, and in defining the cell's stretchiness. Experimental studies revealed that α keratin [6,7] (found in wool, hair, and hooves), hagfish slime threads [8], desmin [9], and vimentin [10] all display a α - β transition under deformation. The transition significantly thins filaments, but toughens and strengthens it at large deformation [10]. Furthermore, the α - β transition may also have important implications for diseases related to amyloid formation (s.a. Alzheimer's, Type II Diabetes, Parkinson's), where native proteins convert to beta-sheet rich proteins [11,12].

The experimental observation of this phenomenon for a variety of proteins suggests that the α - β transition under stretching is a universal mechanism of deformation, and as such is crucial for providing a link between a protein material's hierarchical structure and its mechanical response [13]. The understanding of the α - β transition in protein materials may be similarly important as studies of dislocation nucleation and propagation in metals or ceramics, which present the fundamental unit deformation mechanism in crystalline materials. However, compared to dislocations, the understanding of deformation mechanisms in protein materials remain poorly understood.

What is the molecular mechanism of the α - β transition, and is there a key structural parameter that controls it? Here, we report a systematic study to address this issue.

We consider 13 randomly picked alpha-helical coiled-coil proteins selected from the protein data bank (PDB) [14] with varying lengths, all subjected to mechanical stretching (for information on coiled-coiled proteins, see [15]). Each protein is first energy minimized [16] and then equilibrated at 300 K. All simulations are performed using an *NVT* ensemble (Nose-Hoover thermostat 1 fs time step, CHARMM19 all-atom energy function and an effective Gaussian model for the water solvent [17,18] to facilitate rapid sampling of configurations under loading [19]). After equilibration, each structure is stretched by steered molecular dynamics (SMD), following the shear loading [Fig. 1(a)]. Thereby, the C $^{\alpha}$ atom at the left end of one helical strand is fixed to resemble the attachment at a substrate, while the C $^{\alpha}$ atom at the right end of the other strand is linked to a harmonic spring, spring constant 10 kcal/mol/Å²). We pull at a constant velocity of 0.01 Å/ps along the axial direction [Fig. 1(a)], and the displacement ΔL and pulling force are recorded. The tensile strain is expressed as $\varepsilon = \Delta L/L_0$ where L_0 is the initial end-to-end length of the protein [20]. By analyzing both the H-bond patterns and the backbone geometry with the STRIDE algorithm [21], we assess the secondary structure and associated changes during deformation. For each protein, the stretching simulations are repeated 10 times with different random seeds for the initial velocity distribution for good statistics. Based on the results of an ensemble of 10 simulations for each protein, we then calculate the percentage of the occurrence of the α - β transition as a function of the coiled-coil length.

We begin by considering a very short coiled coil and a very long one (short coiled coil PDB ID 1ajy,

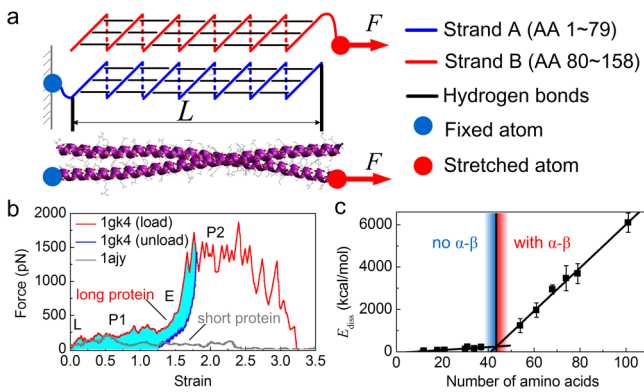


FIG. 1 (color online). Model for studying the strength of coiled-coil proteins under shear loading. (a) Upper: Schematic of the structure and loading condition of the coiled-coil (AA = amino acid residue). Lower: Example molecular structure of the 2B segment of the vimentin protein (PDB ID 1gk4). (b) Force-strain behavior of two coiled-coil proteins with different lengths (PDB ID 1gk4, where the modulus in the L region is 0.3 GPa average diameter 2.4 nm, and in the E region reaches 8.9 GPa average diameter 1.1 nm; PDB ID 1ajy, short). The shaded area reflects the energy dissipated during unfolding. (c) Dissipated energy during the entire stretching process, for coiled-coil proteins with different lengths.

$L_0 \approx 2.6$ nm, long coiled coil 1gk4, $L_0 \approx 11.5$ nm). In the short coiled coil, we see that the stretching force remains at a low value [Fig. 1(b)], where the maximum force approaches 200 pN at 55% strain as the two helical strands slide against each other [Fig. 2(a)]. We observe that only a few amino acids close to the ends of the protein domain are unfolded during the simulation [Fig. 2(c)]. In the long coiled coil, even though all simulation parameters are identical, the force-strain relation shows a dramatically different behavior compared with the short one, and now displays four phases [Fig. 1(b)]. First, we find a linear rise [phase (L)] towards a plateau at 200 pN [phase ($P1$)]. At the end of phase ($P1$), we observe a regime with significant stiffening [phase (E)], followed by another, bumpier plateau phase ($P2$) at relatively large force levels close to 1500 pN. A structural analysis in Fig. 2(d) suggests that the secondary structure remains intact within phase (L), indicating that the protein undergoes solely elastic deformation below 15% strain (no H-bond rupture). The snapshots shown in Fig. 2(b) reveal that within phase ($P1$), each strand twists as it unravels, and that groups of H bonds along the alpha-helix rupture, causing a sudden unfolding of helical turns. (Up until this point, the coiled coil behaves similarly as myosin [22,23].) Phase (E) begins at force levels of 360 pN, but the force increases to more than 1000 pN at strains beyond 150%. In phase ($P2$), a large number of beta sheets have formed, and an increase in the strain leads to the sliding of the two strands against one another, at a force level ≈ 1500 pN (1200–1800 pN; fluctuations due to the stretching, rupture, and rebinding

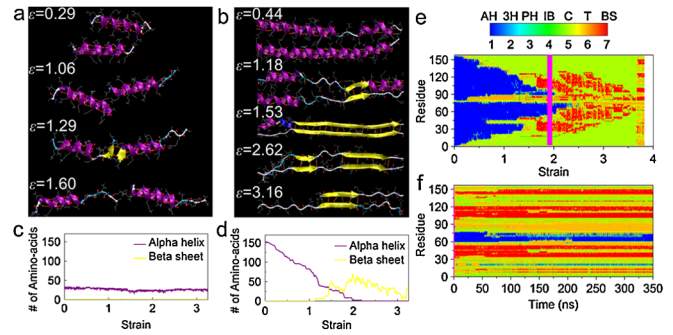


FIG. 2 (color online). Simulation snapshots of two coiled-coil proteins under shear (proteins with PDB ID 1ajy [(a), (c)] and 1gk4 [(b), (d)]). (a) and (b) Structure under applied strain during stretching of the proteins (H bonds are shown as dashed lines). (c) and (d) Number of amino acids associated with alpha-helix and beta-sheet secondary structures as a function of strain. (e) Structure spectrum for structure of the protein 1gk4 under stretching. (f) Structure spectrum of the protein 1gk4 during equilibration. The color bar distinguishes 7 structures: Alpha helix (AH), 3–10 helix (3H), Phi helix (PH), Isolated bridge (IB), Coil (C), Turn (T), beta sheet (BS).

of the H bonds in the sheared beta-sheet segments resembling a stick-slip motion). We note that this force level agrees with the strength of natural beta sheets reported earlier [24]. The much greater shear resistance is due to the fact that beta sheets in shear display a significantly larger strength [24–26] than alpha-helical proteins [22,27]. Phase ($P2$) lasts until 325%, where the structure is completely detached. Figure 2(e) shows an analysis of the secondary structure as a function of strain, providing additional details into structural changes. This transformation of the secondary structure from an alpha-helix to a beta-sheet rich structure resembles the α - β transition observed in experiments.

We record the unloading F - ϵ curve after the α - β transition has occurred. To achieve this, we begin from the configuration obtained at a strain of 180%, and subsequently relax the applied force to zero, then set the two ends free and equilibrate the structure. During 350 ns equilibration, no structural changes are observed as confirmed by the spectrum analysis [Fig. 2(f)]. This irreversible character of the transition is also found in experimentally observed α - β transitions [7,10]. The unloading curve is depicted in Fig. 1(b), showing that the protein filament retracts during relaxation. However, the protein does not approach the length at the beginning of the simulation, suggesting permanent deformation. The shaded area in Fig. 1(b) represents the dissipated energy during the unfolding-refolding process [dissipated energy is 846 kcal/mol, calculated as $E_{\text{diss}} = \int_0^{180\%} (F_{\text{load}} - F_{\text{unload}}) L_0 d\epsilon$]. Since the number of broken H bonds is $N_{\text{H}} = 119$ at 180% strain [compared with the initial structure, Fig. 1(a)], we estimate the energy associated with each H bond as $E_{\text{HB}} = E_{\text{tot}}/N_{\text{H}} = 7.11$ kcal/mol.

The initiation of the α - β transition process features three major phases. From phase (L) onwards, a single turn in the middle of the intact strand unravels locally inside one of the alpha helices (this explains why the ultimate force in this phase approaches the unfolding force of a single alpha helix). This structural defect, locally highly flexible, forms a nucleation point from which on nearby amino acids begin to unfold, turning them into random coils, as shown in Fig. 2(b) at $\varepsilon = 0.44$ (the increase in flexibility can be explained by the significant decrease in persistence length, from several tens of nanometers for an alpha helix to approximately 0.4 nm for a random coil). The increasing number of unfolded turns in both helices increases the probability for hydrophobic interactions to take effect, which squeezes two unfolded turns together to form a first parallel beta-sheet cluster [Fig. 2(a), $\varepsilon = 1.29$; Fig. 2(b), $\varepsilon = 1.18$], here referred to as a *beta-sheet seed*. In the shorter coiled coil, there is either no beta-sheet seed or the beta-sheet seed is not strong enough to induce further unfolding of alpha-helical turns, and rather ruptures, leading to interprotein sliding [Figs. 2(a) and 2(c)]. In the longer coiled coil, however, the beta-sheet seed acts as an effective clamp that prevents sliding and thereby enforces unfolding of all other alpha-helical turns [Fig. 2(b), $\varepsilon = 1.53$], resulting in the occurrence of the transition process throughout the rest of the chain [Fig. 2(d)]. At the end of the transition process, the structure is composed of an array of many small beta-sheet clusters [Fig. 2(b), $\varepsilon = 2.62$]. The key to understand this process is to consider the ratio of the strength of the beta-sheet seed (f_B) versus the strength of the alpha helix (f_A), where the ratio of f_B/f_A controls whether ($f_B/f_A \geq 1$) or not ($f_B/f_A < 1$) the α - β transition occurs.

The off rate of a reference structure (alpha helix or beta sheet) with a reference length N_0 is [27–29]: $\chi_0 = \omega_0 \exp[-(E_b - F_0 x_b)/(k_B T)]$, where $\omega_0 = 10^{13} \text{ s}^{-1}$ is the natural frequency of bond vibration [30], x_b the unfolding transition point, k_B the Boltzmann constant, E_b the energy barrier, F_0 the averaged unfolding force, and T the temperature. The off rate of alpha helices of general length N is $\chi = \omega_0 (N/N_0) \exp[-(E_b - f_A x_b)/(k_B T)]$, which considers the fact that alpha-helical turns are arranged in series (so that increasing the length increases the probability of unfolding). Noting the fact that pulling speed $v = \chi x_b$ is constant, we obtain

$$f_A(N) = F_0 - \frac{k_B T}{x_b} \ln\left(\frac{N}{N_0}\right) \sim -\ln(N). \quad (1)$$

We note that longer alpha helices result in a reduced mechanical strength with a negative logarithmic dependence on N , a finding that has been validated in experiments and simulations [31]. We now calculate the strength of the beta-sheet seed, f_B . The energy barrier in the beta-sheet shearing is $E_b = E_{\text{HB}} n_{\text{cr}}$, where $n_{\text{cr}} = 3-4$ is the number of H bonds that are involved in a unit rupture event

[32]. Combining with the expression of χ_0 , we arrive at

$$f_B(n_{\text{cr}}) = \frac{1}{x_B} \left[k_B T \ln\left(\frac{v}{\omega_0 x_B}\right) + E_{\text{HB}} n_{\text{cr}} \right] \quad (2)$$

where v is the pulling speed and x_B the rupture distance along the shearing direction. Since the size of the beta-sheet seed is constant at n_{cr} regardless of the length of the alpha-helical proteins in the coiled coil, $f_B(n_{\text{cr}}) = f_B = \text{const}$.

We now identify numerical values of the parameters in Eqs. (1) and (2) directly from the atomistic model. For Eq. (1), the strength of an alpha-helical protein with $N_0 = 79$ at 0.01 Å/ps pulling speed is $F_0 = 200$ pN, and $x_b = 0.2$ Å (parameters extracted from a pulling experiment of a single alpha-helical protein [27]). For Eq. (2), $E_{\text{HB}} = 7.11$ kcal/mol (close to the value predicted from density functional theory [33]), and $x_B = 4$ Å [25]. As shown in Fig. 3(a), f_B of an initial beta-sheet seed falls within the range of 346.5 ± 61.5 pN, whereas f_A varies greatly as a function of the length. The condition $f_B = f_A(N_{\text{cr}})$ yields the critical amino acid number $N_{\text{cr}} = 41 \pm 12$ from which on the α - β transition occurs.

We carry out simulations for 13 different coiled-coil proteins [corresponding PDB IDs shown in Fig. 3(b)], and measure the probability for the occurrence of the α - β transition). The results reveal that the points fall in two distinct regions and that the transition occurs at a critical number of $N_{\text{cr}} = 38.7$ amino acids (5.7 nm), which agrees with the theoretical analysis, thus corroborating the concept of a critical length for the α - β transition. It is noted that the outcome of the test for each protein is very robust. However, there are greater fluctuations for those proteins near the critical length (for these, the repeated simulations give appropriate statistics and convergence).

We find that the critical length decreases slightly under extreme variations of the pulling speed. For example, at $v = 0.005$ Å/ps, we obtain $F_0 = 167$ pN, $f_B = 340$ pN, and $N_{\text{cr}} = 34$; at $v = 0.002$ Å/ps, $F_0 = 146$ pN, $f_B = 330$ pN, and $N_{\text{cr}} = 32$; at $v = 0.001$ Å/ps, $F_0 = 116$ pN, $f_B = 323$ pN, and $N_{\text{cr}} = 29$. At vanishing pulling velocities corresponding to experimental and physiological conditions (at 1 $\mu\text{m/s}$ and less), we estimate a critical number of 26 amino acids, or 3.8 nm ($F_0 = 14$ pN, $N_0 = 288$, $x_b = 1.2$, $f_B = 96$ pN, and $N_{\text{cr}} = 26$ based on asymptotic strength estimates as reported in the early work [25–27,34]) [Fig. 3(c)].

Because of the distinct force-strain behavior, the critical length has major implications for the energy dissipation capacity of a protein filament [Fig. 1(c)]. Notably, energy dissipation increases manifold with the protein's length in the second regime where the α - β transition occurs. This finding provides a potential explanation for the universality of *long* alpha-helical protein filaments in mechanically relevant proteins in biology, as they provide an innate

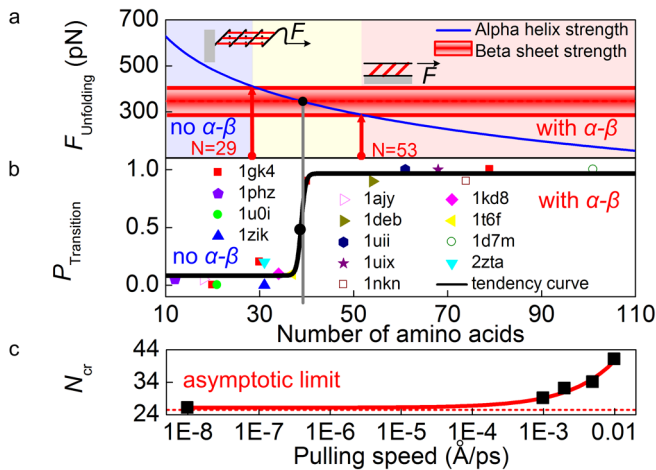


FIG. 3 (color online). (a) Theoretical analysis of the critical condition for the α - β transition, where the shaded area in the middle represents the critical condition. The left side of the critical point corresponds to sliding and unfolding of the alpha-helices without occurrence of the α - β transition, while the right side corresponds to the regime in which the α - β transition occurs. (b) Percentage of α - β transition occurrence for various coiled-coil proteins (simulation). Each point reflects the percentage of the transition within the simulation ensemble. The fitted curve is a sigmoid function $p = A_2 + (A_1 - A_2)\{1 + \exp[(n - n_0)/dx]\}^{-1}$, with $A_1 = 0.083$, $A_2 = 0.967$, $n_0 = 38.835$, and $dx = 0.465$. The fit enables us to identify the transition point from the simulation data, leading to $N_{cr} = 38.7$ (probability 50%). (c) Critical number of amino acids N_{cr} as a function of pulling speed (continuous curve is power law fit to the simulation data; asymptotic limit $N_{cr} = 26$).

capacity to heightened energy dissipation at large deformations.

Our results provide a critical geometrical condition of the α - β transition and associated molecular-level effects. The concept put forth in our simple model may be used to explain other structural transitions, stability, and flexibility in proteins or polymers, and as such, may find future applications in bottom-up materials design [35,36]. The insight into the critical length scales may have applications in the design of novel peptide-based materials with high stiffness, novel hierarchical fibers, and materials with high extreme absorption capacity.

Support from AFOSR (Grant No. FA9550081-0321) is acknowledged.

*Corresponding author: mbuehler@MIT.EDU

- [1] C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland Pub., New York, 1999).
- [2] D.L. Minor and P.S. Kim, *Nature (London)* **367**, 660 (1994).

- [3] T. Koga *et al.*, *Chem. Eur. J.* **9**, 1146 (2003).
- [4] F. Ding *et al.*, *Proteins: Struct. Funct. Genet.* **53**, 220 (2003).
- [5] M. Meier and J. Seelig, *J. Am. Chem. Soc.* **130**, 1017 (2008).
- [6] J.S. Church, G.L. Corino, and A.L. Woodhead, *J. Mol. Struct.* **440**, 15 (1998).
- [7] L. Kreplak *et al.*, *Biophys. J.* **87**, 640 (2004).
- [8] D.S. Fudge *et al.*, *Biophys. J.* **85**, 2015 (2003).
- [9] L. Kreplak *et al.*, *Biophys. J.* **94**, 2790 (2008).
- [10] Z. Qin *et al.*, *PLoS ONE* **4**, e7294 (2009).
- [11] M. Gross, *Current Protein & Peptide Science* **1**, 339 (2000).
- [12] I. Daidone *et al.*, *Proteins: Struct. Funct. Bioinf.* **57**, 198 (2004).
- [13] M.J. Buehler and Y.C. Yung, *Nature Mater.* **8**, 175 (2009).
- [14] H.M. Berman *et al.*, *Nucleic Acids Res.* **28**, 235 (2000).
- [15] Coiled coils are double-stranded protein motifs (e.g., found in myosin and kinesin), where each strand is an alpha-helix with heptad repeated substrings. This seven residue repeat generally has apolar residues at the first and fourth position, forming a left-handed hydrophobic stripe. Following the orientation of this stripe, the two alpha-helical strands wrap around each other.
- [16] We minimize the potential energy via steepest descent and adopted basis Newton-Raphson (ABNR) methods.
- [17] T. Lazaridis and M. Karplus, *Science* **278**, 1928 (1997).
- [18] T. Lazaridis and M. Karplus, *Proteins: Struct. Funct. Genet.* **35**, 133 (1999).
- [19] E. Paci and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6521 (2000).
- [20] We use visual molecular dynamics (VMD), W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* **14**, 33 (1996), for visualization. The rupture length of H bonds is defined to be 5 Å for visualization purposes.
- [21] D. Frishman and P. Argos, *Proteins: Struct. Funct. Genet.* **23**, 566 (1995).
- [22] I. Schwaiger *et al.*, *Nature Mater.* **1**, 232 (2002).
- [23] D.D. Root *et al.*, *Biophys. J.* **90**, 2852 (2006).
- [24] M. Sotomayor and K. Schulten, *Science* **316**, 1144 (2007).
- [25] S. Keten and M.J. Buehler, *Nano Lett.* **8**, 743 (2008).
- [26] S. Keten and M.J. Buehler, *Phys. Rev. Lett.* **100**, 198301 (2008).
- [27] T. Ackbarow *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16410 (2007).
- [28] M. Rief *et al.*, *Phys. Rev. Lett.* **81**, 4764 (1998).
- [29] E. Evans, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 105 (2001).
- [30] G.I. Bell, *Science* **200**, 618 (1978).
- [31] H. Dietz and M. Rief, *Phys. Rev. Lett.* **100**, 098101 (2008).
- [32] We find that the initial beta-sheet seed formed under shearing includes 3–4 H bonds, corresponding to the number of H bonds in an alpha-helical turn [27].
- [33] K. Tsemekhman *et al.*, *Protein Sci.* **16**, 761 (2007).
- [34] Z. Qin *et al.*, *Nanotechnology* **20**, 425101 (2009).
- [35] M. de Leeuw *et al.*, *PLoS ONE* **4**, e7296 (2009).
- [36] R. Burioni *et al.*, *Proteins: Struct. Funct. Genet.* **55**, 529 (2004).