Atomistically Informed Mesoscale Model of Alpha-Helical Protein Domains

*Jérémie Bertaud, Zhao Qin, & Markus J. Buehler**

Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ABSTRACT

Multiscale mechanical properties of biological protein materials have been the focal point of extensive investigations over the past decades. In this article, we present the development of a mesoscale model of alpha-helical (AH) protein domains, key constituents in a variety of biological materials, including cells, hair, hooves, and wool. Our model, derived solely from results of full atomistic simulations, is suitable to describe the deformation and fracture mechanics over multiple orders of magnitude in time- and length scales. After validation of the mesoscale model against atomistic simulation results, we present two case studies, in which we investigate, first, the effect of the length of an AH protein domain on its strength properties, and second, the effect of the length of two parallel AH protein domain arrangement on its shear strength properties and deformation mechanisms. We find that longer AHs feature a reduced tensile strength, whereas the tensile strength is maximized for ultrashort protein structures. Moreover, we find that the shearing of two parallel AHs engenders sliding, rather than AH unfolding, and that the shear strength does not significantly depend on the length of the two AHs.

KEYWORDS

hierarchical material, nanomechanics, materiomics, biological protein materials, fracture, deformation, experiment, simulation, multiscale modeling

*Address all correspondence to mbuehler@mit.edu.

1543-1649/08/$35.00 © 2009 by Begell House, Inc.

1. INTRODUCTION

Proteins constitute critical building blocks of life, forming biological protein materials (BPMs) such as hair, bone, skin, spider silk, and cells, which play an important role in providing essential mechanical functions to biological systems [1–5]. Flaws and failure of these materials can cause serious diseases and malfunctions in biological organisms, for example, due to misfolded protein structures. However, the fundamental deformation and failure mechanisms of biological protein materials remain largely unknown, partly due to a lack of understanding of how individual protein building blocks respond to mechanical load and how they participate in the function of the overall biological system. Significant advances in experimental, theoretical, and computational materials science have enabled a deeper understanding of BPMs through the linking of structure-process-property (SPP). The material properties of biological materials have been the focal point of extensive studies over past decades, leading to formation of a research field that connects biology and materials science, referred to as materiomics. Materiomics utilizes mechanistic insight, based on SPP relations in their biological context, to provide a basis for understanding disease processes; to develop new approaches to treating genetic and infectious diseases, injury, and trauma; and to enhance engineered materials via translating material concepts from biology. Computational bottom-up multiscale approaches provide methods for rapid screening of SPP relations.

The behavior of materials, in particular, their mechanical properties, are intimately linked to the atomic microstructure of the material. Whereas crystalline materials show mechanisms such as dislocation spreading or crack extension [6–8], biological materials feature molecular unfolding or sliding, with a particular significance of rupture of chemical bonds such as hydrogen bonds (H-bonds), covalent cross-links, or intermolecular entanglement. Additional mechanisms operate at larger length scales, where the interaction of extracellular materials with cells and of cells with one another, different tissue types, and the influence of tissue remodeling (at longer timescales) become more evident. The dominance of specific mechanisms is controlled by geometrical parameters, the chemical nature of the molecular interactions, and the structural arrangement of the protein elementary building blocks across many hierarchical scales, from nano to macro. The multiscale understanding of how molecular structures participate in macroscale deformation of biological tissues remains an outstanding challenge, and multiscale computational approaches are believed to play a crucial role in advancing this field.

What is the significance of mechanics of biological protein materials? The mechanical properties of biological materials have wide-ranging implications for biology. In cells, for instance, mechanical sensing is used to transmit signals from the environment to the cell nucleus or to control tissue formation and regeneration [1–5, 9]. The structural integrity and shape of cells is controlled by the cell's cytoskeleton, which resembles an interplay of complex protein structures and signaling cascades arranged in a hierarchical fashion. Bone and collagen, providing structure to our bodies, or spider silk, used for prey procurement, are examples of materials that have incredible elasticity, strength, and robustness unmatched by many synthetic materials, which can mainly be attributed to its structural formation with molecular precision. This transfer of concepts observed in biology into technological applications and new materials design remains a big challenge with a potentially big payoff. In particular, the combination of nanostructural and hierarchical features into materials developments could lead to significant breakthroughs in the development of new materials that mimic or exceed the properties found in biological analogs. The characterization of material properties for biological protein materials may also play a crucial role in developing a better understanding of diseases. Injuries and genetic diseases are often caused by structural changes in protein materials (e.g., defects, flaws, changes to the molecular structure), resulting in failure of the material's intended function. This approach enables one to probe how mutations in structure alter the properties of protein materials. In the case of osteogenesis imperfecta (brittle bone disease), for instance, molecular-scale models predict a softening of bone's basic collagen constituent [10]. These observations may eventually provide explanations of the molecular origin of certain diseases. Additionally, these findings provide evidence that material properties play an essential role in biological systems, and that the current paradigm of focusing on biochemistry

alone as the cause of diseases is insufficient. It is envisioned that the long-term potential impact of this work can be used to predict diseases in the context of diagnostic tools by measuring material properties, rather than focusing on symptomatic chemical readings alone. Such approaches have been explored for cancer and malaria, for instance [11, 12].

This article is focused on the analysis of AH protein domains, as they appear, for instance, in intermediate filaments in the cell's cytoskeleton. Figure 1 displays the multiscale hierarchical structure of this protein network. The work reported here is focused on the smallest of all elements in this protein material: the mechanics of a single AH (see Fig. 1, bottom).

1.1 Review: Multiscale Methods for Protein Materials

Multiscale simulation models for protein materials have become increasingly popular in recent years and have enabled the direct link between experiment and theoretical bottom-up descriptions of materials (Fig. 2). Here we provide a review of popular

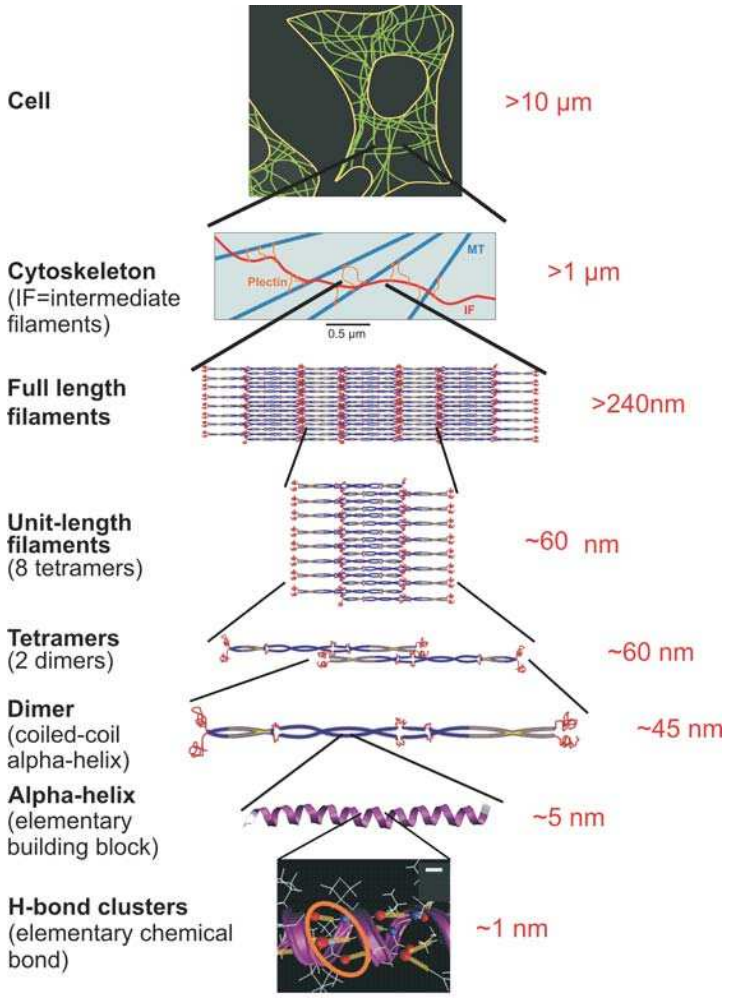

FIGURE 1. Overview of different material scales, from nano to macro, here exemplified for the cellular protein network vimentin intermediate filaments. To understand their deformation and fracture mechanisms, it is crucial to elucidate atomistic and molecular mechanisms at each scale. Computational multiscale approaches thereby play a crucial role in transcending through multiple scales in length and time

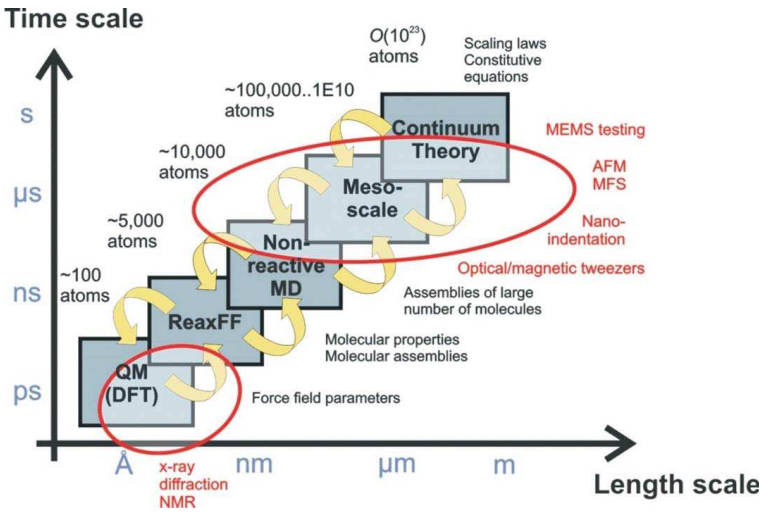

FIGURE 2. Overview of computational and experimental methods. Hierarchical coupling of different computational tools can be used to traverse throughout a wide range of length- and timescales. Such methods enable one to provide fundamental insight into deformation and fracture phenomena, across various time- and length scales. Handshaking between different methods enables one to transport information from one scale to another. Eventually, results of atomistic, molecular, or mesoscale simulation may feed into constitutive equations or continuum models. While continuum mechanical theories have been very successful for crystalline materials, biological materials require statistical theories (e.g., the Bell model discussed in Section 2.3 and related formulations) to describe elasticity and strength. Experimental techniques, such as the atomic force microscope, molecular force spectroscopy, nanoindentation, or magnetic/optical tweezers, now overlap into atomistic and molecular approaches, enabling direct comparison of experiment and simulation. Techniques such as X-ray diffraction, infrared spectroscopy, or nuclear magnetic resonance (NMR) provide atomic-scale resolution information about the 3-D structure of protein molecules and protein assemblies

multiscale approaches to describe the mechanics of proteins in a multiscale setting.

Single-bead models are the most direct approach taken for studying macromolecules. The term *single bead* derives from the idea of using single beads, that is, point masses, for describing each amino acid in a protein structure. The Elastic Network Model (ENM) [13], Gaussian Network Model [14], and Go-model [15] are well-known examples that are based on such bead model approximations. These models treat each amino acid as a single bead located at the C_α position, with mass equal to the mass of the amino acid. The beads are connected via harmonic bonding potentials, which represent the covalently bonded protein backbone. In Go-like models, an additional Lennard-Jones–based term is included in the potential to describe short-range nonbonded interactions between atoms within a finite cutoff separation. Despite their simplicity, these models have been extremely successful in explaining thermal fluctuations of proteins [16] and have also been implemented to model the unfolding problem to elucidate atomic-level details of deformation and rupture that complement experimental results [17–19]. A more recent direction is coupling of ENM models with a finite element–type framework for mechanistic studies of protein structures and assemblies [20].

Using more than one bead per amino acid provides a more sophisticated description of protein molecules. In the simplest case, the addition of another bead can be used to describe specific side-chain interactions in proteins [21]. Higher-level models, for instance, four- to six-bead descriptions, capture more details by explicit or united atom description for backbone carbon atoms, side chains, and carboxyl and amino groups of amino acids [22, 23]. Even coarser-level multiscale modeling methods have been reported more recently, applied to model biomolecular systems at larger time- and length scales. These models typically employ superatom descriptions that treat clusters of amino

acids as beads. In such models, the elasticity of the polypeptide chain is captured by simple harmonic or anharmonic (nonlinear) bond and angle terms. These methods are computationally quite efficient and capture shape-dependent mechanical phenomena in large biomolecular structures [24], and can also be applied to collagen fibrils in connective tissue [25] as well as mineralized composites such as nascent bone [26]. In this article, we develop such a coarse-level description of AH protein domains.

1.2 Outline of This Article

The plan of this article is as follows. In Section 2, we begin with a presentation of the computational setup. This includes a detailed description of the mesoscopic model formulation, the fitting procedure, and validation. We present a review of strength models for protein domains, with a focus on descriptions derived from the Bell model. Section 3 is dedicated to a discussion of computational results obtained using the mesoscale model, here applied to study the length dependence of strength properties of AH protein domains in single and parallel arrangements. We conclude in Section 4 with a discussion and an outlook for future research.

2. COMPUTATIONAL METHODS

In the following sections, we describe our atomistic-based multiscale simulation approach used to develop a mesoscale description of AH protein domains. The setup of the coarse-grained model for AH protein domains is based on the geometry of an AH, which features a linear array of turns or convolutions; during rupture, any one of these convolutions ruptures, as reported in earlier simulation studies [27]. In our model, every convolution is represented by one bead so that an entire AH is represented by a linear combination of multiple beads.

2.1 Coarse-Graining Approach

To achieve the coarse-grained description, the entire sequence of amino acids that makes up the AH structure is replaced by a collection of beads that are linked to their neighbors by a single-type bond, as shown in Fig. 3. An AH consists of a series of convolutions, where each features 3.6 residues. This protein secondary structure is stabilized through the

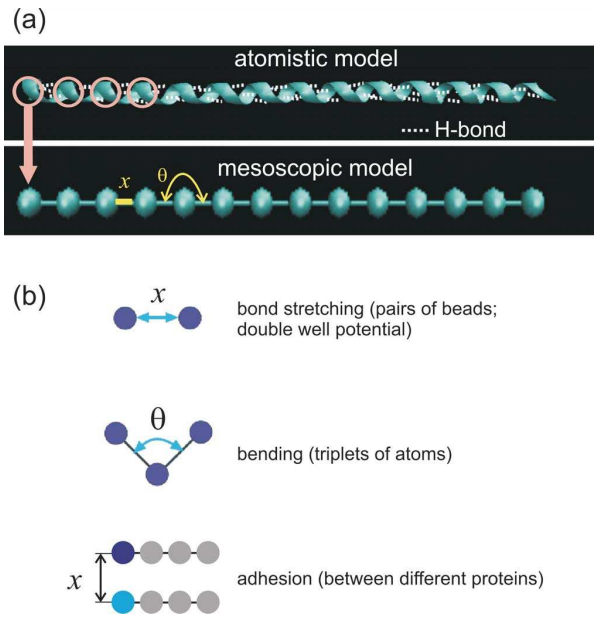

FIGURE 3. Schematic of the coarse-graining procedure, in which we replace the full atomistic representation by a mesoscopic bead model. A pair of beads represents one turn in the alpha-helix (AH; also called a convolution) and thus 3.6 residues (each bead also has the corresponding mass). In the atomistic representation, the folded states of the turns are stabilized by the presence of 3.6 H-bonds between turns. In the mesoscopic bead model, this is represented by using a double-well potential to describe the energy landscape under bond stretching. Different AH chains interact with a Lennard-Jones potential to describe intermolecular adhesion

presence of H-bonds between the O atom of residue n and the N atom of residue $n + 4$ (there are 3.6 H-bonds between turns, on average).

In our coarse-grained model, we aim at capturing the main structural and energetic features of an AH protein domain. Full atomistic molecular dynamics (MD) simulations have confirmed that H-bonds break in clusters of three to four (corresponding to one convolution) for pulling velocities lower than 0.3 m/s [27]. Thus, to capture the role of the convoluted structure on AH protein strength, the size of one bead (the smallest unit of our system) represents one convolution of the AH, with the corresponding mass that reflects all atoms that are represented by that bead. Our model is set up to study AH protein domain strength for pulling velocities lower than 0.3 m/s.

The beads interact according to a bond potential and an angle potential. We choose a double-well bond potential to capture the existence of two equilibrium states for a convolution, corresponding to the folded and unfolded configuration (see Fig. 4 for the energy landscape and snapshots of atomistic geometries of the folded and unfolded states). The model does not involve explicit solvent; rather, the effect of solvent on the breaking dynamics of AH convolutions is captured in the effective double-well potential. Through this formulation, the bond potential can describe the microscopic details of the rupture mechanism of the 3.6 H-bonds between each convolution under force, and the transition from the folded states to the unfolded states of convolutions through an energy barrier that separates the two states. Yet the description is sufficiently coarse so that it enables a significant computational speedup and efficiency compared with the full atomistic description. To characterize interactions between AH protein domains in a parallel arrangement, we use a Lennard-Jones (LJ) potential between pairs of beads of different strands. This describes the adhesion energy between AHs.

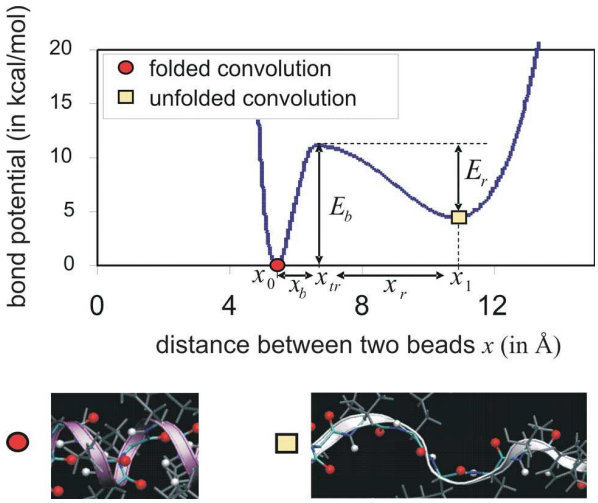

FIGURE 4. Double-well profile of the bond stretching potential in the bead model, representing the energy landscape associated with unfolding of one convolution. The numerical values of the equilibrium states (x_0 and x_1), energy barriers (E_b and E_r), and the transition state x_{tr} are obtained from geometric analysis of the AH geometry as well as full atomistic simulations. The transition state (local energy peak of the potential) corresponds to the breaking of the 3.6 H-bonds between two convolutions of the AH. After failure of these weak bonds, the convolution unfolds to a second equilibrium state with a larger inter-bead distance. Under further loading, its covalent bonds begin to be stretched, which leads to a second increase of the potential at large deformation

The mathematical expression for the total energy of the system is given by

$$E = E_T + E_B + E_S \tag{1}$$

where E_T is the total tensile energy, E_B is the total bending energy, and E_S is the total intermolecular interaction energy. The total bending energy is given by the sum over all triples of beads,

$$E_B = \sum_{\text{triplets}} \phi_B(x) \tag{2}$$

The angle potential is given by

$$\phi_B(\theta) = \frac{1}{2} K_B(\theta - \theta_0)^2 \tag{3}$$

where K_B relates to the bending stiffness of the molecule EI, with θ as the interbead angle (in triplets of atoms) and θ_0 as the equilibrium angle. The bending stiffness parameter K_B is given by

$$K_B = \frac{3EI}{x_0} \tag{4}$$

with x_0 denoting the equilibrium bead distance, which corresponds to the equilibrium distance of one folded convolution, and EI as the bending stiffness of the alpha helix. The total tensile energy is given by the sum over all pair-wise interactions,

$$E_T = \sum_{\text{pairs}} \phi_T(x) \tag{5}$$

The double-well bond potential $\phi_T(x)$ is given by (see also Fig. 4 for a schematic)

$$\phi_T(x) = \begin{cases} \dfrac{E_b}{x_b^4}\,(x-x_{tr})^2\Big(x-x_{tr}-\sqrt{2}\cdot x_b\Big) \\ \times\,\big(x-x_{tr}+\sqrt{2}\cdot x_b\big)+E_b, \quad x<x_{tr} \\ \dfrac{E_r}{x_r^4}\,(x-x_{tr})^2\Big(x-x_{tr}-\sqrt{2}\cdot x_r\Big) \\ \times\,\big(x-x_{tr}+\sqrt{2}\cdot x_r\big)+E_b, \quad x_{tr}\le x \end{cases} \tag{6}$$

The first equilibrium with reaction coordinate x_0 (first potential minimum) corresponds to the folded

state of one turn of an AH under no force. The transition state (energy barrier E_b), with position x_{tr} (peak of the potential between two wells), corresponds to the breaking of the 3.6 H-bonds between two turns of the AH. After failure of these weak bonds, the turn unfolds to a second equilibrium state. This corresponds to the second potential minimum with a larger interbead distance, x_1. Under further loading, its backbone bonds begin to be stretched, which leads to a second increase of the potential. The parameters x_b and E_b represent the distance and energy barrier required to unfold one convolution. Similarly, x_r and E_r correspond to the refolding process.

The total intermolecular interaction energy E_S is given by the sum over all pair-wise interactions between beads of different AH protein domains:

$$E_S = \sum_{\text{pairs}} \phi_S(x) \tag{7}$$

The adhesion potential ϕ_S is described by a LJ potential:

$$\phi_S(x) = 4 \cdot \varepsilon \cdot \left[\left(\frac{\sigma}{x}\right)^{12} - \left(\frac{\sigma}{x}\right)^{6} \right] \tag{8}$$

The energy minimum ε of the LJ potential corresponds to the adhesion energy per convolution between two AHs in equilibrium divided by 3.51 (factor that takes into account the fact that we do not use a cutoff distance for the LJ potential, and therefore the adhesion energy for each bead is the same as the adhesion energy estimated from full atomistic simulation); and x is the distance between mesoscale particles. The zero-crossing distance σ is linked to the equilibrium bead distance d_0 (equilibrium spacing between two AHs) by

$$d_0 = 2^{1/6} \cdot \sigma \tag{9}$$

2.2 Parameter Fitting: Linking Atomistic and Mesoscale

The parameters are determined through a fitting procedure against geometric properties of AHs as well as full atomistic MD simulation results in explicit solvent. We fit the energy barrier measured from MD simulation to the energy barrier in the mesoscale model formulation. The mass of each bead corresponds to the approximate average mass of one turn or 3.6 residues, leading to 400 amu.

The two parameters of the angle potential are introduced in Eq. (3). The value of the equilibrium angle θ_0 is 180°, based on the geometry of the AH structure. The bending stiffness parameter K_B is linked to molecular parameters as described by Eq. (4) and therefore can be determined from full atomistic simulations of bending studies of AH protein domains (we use the results reported in [28]). We find that $K_B = 21.589$ kcal/mol/rad^2. The parameters of the tensile double potential are introduced in Eq. (6) (see also Fig. 4). We find that $x_0 = 5.4$ Å for the equilibrium bead distance of the folded state, which corresponds to the length of one folded convolution. The distance x_b between the folded state equilibrium and the transition state corresponds to the distance to break 3.6 H-bonds, which leads to the unfolding of the convolution, and the parameter E_b is the corresponding energy barrier (full atomistic MD simulations [27] have confirmed that H-bonds in convolutions indeed break in clusters). These two parameters are determined from fitting against full atomistic simulations of tensile loading experiments of AH domains, for a range of pulling rates below 0.3 m/s [27]. We find that $x_b = 1.2$ Å and $E_b = 11.1$ kcal/mol by fitting the theoretical strength model described in the next section to these full atomistic studies. The equilibrium bead distance of the unfolded state, x_1, is determined by fitting the mesoscopic force-strain curve against the atomistic simulation results at large deformation. The parameter describes at what strain levels an AH convolution is completely unfolded and when further strain leads to significant stiffening due to stretching of the protein backbone. Therefore we adjust x_1 so that the angular point between the plateau regime and the backbone stretching regime occurs at the same strain as in the atomistic simulation curve. We find that $x_1 = 10.8$ Å, which corresponds to twice the length of the folded state. Last, the energy barrier E_1 to refold a so-called broken convolution must be smaller than E_b since the folded state is the most favorable state for a convolution in equilibrium. On the basis of a suggestion put forth in [29], we determine that $E_1 = 0.6E_b$ (it is noted, however, that the resulting mechanical properties of the AH are insensitive to variations of choices of E_1, as long as $E_1 < E_b$).

The two parameters of the intermolecular potential are introduced in Eqs. (8) and (9). The parameter ε is obtained from the measurement of the min-

imum of the adhesion energy per convolution between two AHs in equilibrium, and the parameter d_0 is obtained from the equilibrium spacing between two AHs. They are determined from full atomistic simulations, as shown in Fig. 5. We consider two identical AHs in parallel. Each AH contains 100 glutamine amino acids, with an approximate length of 150 Å. The two AHs are kept parallel to each other at a distance of d by fixing all the C_α atoms in each chain, while all other atoms (including side chains) are free to move. We use the CHARMM-19 force field with excluded volume implicit solvation model (Effective Energy Function 1, EEF1) [30, 31]. Full atomistic simulations are carried out with the Chemistry at HARvard Macromolecular Mechanics (CHARMM) package. We consider a series of molecular arrangements with systematically increasing intermolecular distance d; for each case, we record the potential energy after energy minimization. Figure 5 shows the results of the adhesive properties between two AHs, where we systematically change the intermolecular spacing between two parallel AHs. The MD studies reveal that

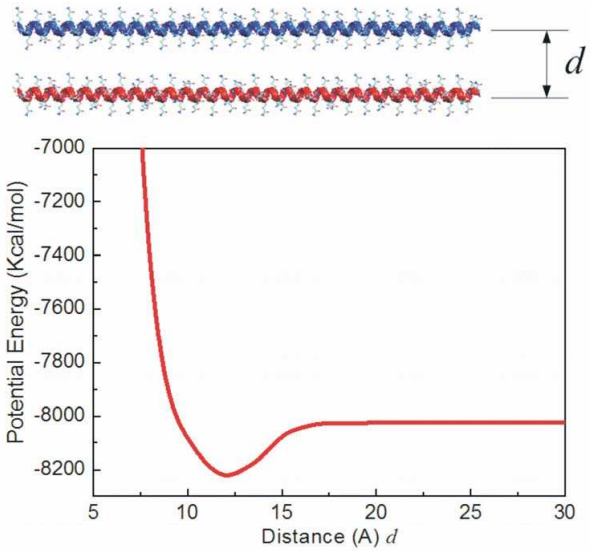

FIGURE 5. Full atomistic model with implicit solvent calculation with CHARMM, used here to identify the intermolecular adhesion between two parallel AHs. This simulation reveals that the adhesion energy is $\gamma = 1.262$ kcal/mol/Å, that is, 6.815 kcal/mol per convolution, with an equilibrium spacing of 12.1 Å. The full atomistic simulation is carried out for a AH sequence made of glutamine amino acids

the adhesion energy is $\gamma = 1.262$ kcal/mol/Å (corresponding to the minimum of the adhesive energy function), reached at $d_0 = 12.1$ Å equilibrium spacing, leading to a LJ distance parameter $\sigma = 10.8$ Å. The adhesion energy per convolution is linked to the LJ parameter ε, which is determined to be $\varepsilon = 6.815$ kcal/mol.

The complete set of parameters of the mesoscopic model and their physical meanings are summarized in Table 1.

2.3 Theoretical Modeling of Alpha-Helix Rupture: Chemomechanical Strength Model

Here we briefly review a theoretical strength model that can be applied to describe the mechanical behavior of AH protein domains: the Bell model. The Bell model is a simple phenomenological model that describes the frequency of failure of reversible bonds [32]. The concept of reversibility means that an individual bond can break under no force if one waits a sufficiently long time, and that it can reform spontaneously. Such bonds may be associated with electrostatic, van der Waals, or H-bond interactions (as they apply to AHs). The frequency of failure, also called the *dissociation rate* or *off rate*, is defined as the inverse of the bond lifetime and is used as a concept to describe the dynamical behavior of such bonds.

Bell's theory explains the force dependence of the off rate and thus shows the significant role of mechanical force in biological chemistry. For instance, this theory can be applied to describe the forced unbinding of biological adhesive contacts such as adhesion of cells to cells. Bell's theory is an extension of the transition state theory for reactions in gases developed by Eyring and others [33]. Inspired also by Zhurkov's work on the kinetic theory of the strength of solids [34], Bell predicted for the first time that the off rate of a reversible bond, which is the inverse of the bond lifetime, increases when subjected to an external force f. Indeed, the rupture of bonds occurs via thermally assisted crossing of an activation barrier E_b, which is reduced by $f \cdot x_b$ as the applied force f increases, x_b being the distance between the bound state and the transition state. Thus the Bell off-rate expression is

$$k = \omega_0 \exp\left(-\frac{E_b - f \cdot x_b}{k_B T}\right) \qquad (10)$$

TABLE 1. Summary of parameters of the mesoscale model, derived from geometrical analyses and atomistic simulations[a]

Parameters	Numerical values
Equilibrium bead distance of the folded state x_0 [Å]	5.4
Equilibrium bead distance of the unfolded state x_1 [Å]	10.8
Distance between folded state and transition state x_b [Å]	1.2
LJ distance parameter between beads of two AHs, σ [Å]	10.8
Energy barrier between folded state and transition state E_b [kcal/mol]	11.1
Energy barrier between unfolded state and transition state E_r [kcal/mol]	6.7
Energy minimum between beads of two AHs, ε [kcal/mol]	6.815
Equilibrium angle θ_0 [deg]	180
Bending stiffness parameter K_B [kcal/mol/rad^2]	21.589
Mass of each mesoscale bead [amu]	400

[a] Table corresponds to Eqs. (1)–(9) as well as to the discussion presented throughout Section 2.2. Abbreviations are as follows: AH, alpha-helix; LJ, Lennard-Jones.

where ω_0 is the natural vibration frequency and k_BT is the thermal energy. The force $f_0 = E_b/x_b$ represents the force needed for the energy barrier to vanish completely. This expression can be rearranged to yield an expression of strength as a function of pulling velocity v:

$$f(v, E_b, x_b) = \frac{k_BT}{x_b}\ln(v) - \frac{k_BT}{x_b}\ln(x_b\omega_0) + \frac{E_b}{x_b} \quad (11)$$

with $v = \Delta x/\Delta t$ as the pulling speed at which a protein structure is deformed. We will use Eq. (11) in the following sections in a comparison between the theoretical model and simulation results (for further details regarding this model and its derivation, we refer the reader to [27]).

3. COMPUTATIONAL RESULTS

The first step of the application studies is the validation of the mesoscale results against full atomistic simulation results. Figure 6 presents the validation of our mesoscale model by direct comparison of the strength of AH protein domains with full atomistic results of the rupture mechanics of an AH protein domain. The straight line in this plot corresponds to the predictions by the theoretical Bell model (Eq. (11)), discussed in Section 2.3.

Figure 7 depicts the entire force-strain curve for a stretching experiment on the 14-bead mesoscopic model of an AH with a length of 70.2 Å, at a temperature of 300 K and a pulling rate of 0.1 m/s. The curve shows the three typical regimes observed in full atomistic simulations: an elasticity regime at low strain, an energy dissipation regime that corresponds to the unfolding of the 13 bonds (13 peaks on the curve), and the subsequent regime of stretching of the backbone bonds. The peaks of force, which correspond to the AH rupture force, fit atomistic re-

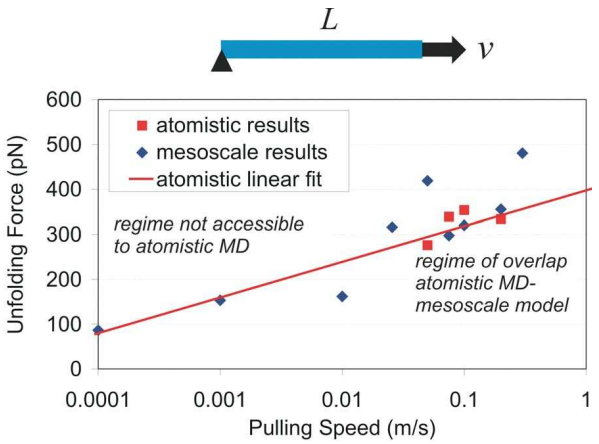

FIGURE 6. Validation of the mesoscopic model by comparison with full atomistic results (MD results taken from [27]). The plot shows the rate dependence of the unfolding force for both the mesoscopic and atomistic models. The mesoscale model is in very good agreement with the full atomistic simulations, validating the fitting of the mesoscopic bond potential. The figure further illustrates that the mesoscale model is capable of reaching much slower pulling rates than those accessible to full atomistic simulation studies, here shown for the slowest pulling speed of 0.0001 m/s

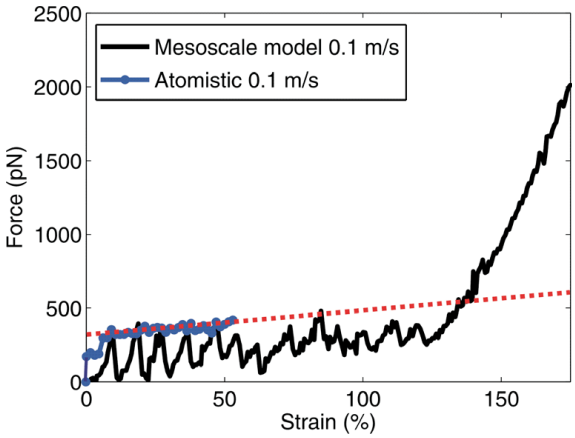

FIGURE 7. Entire force-strain curve for a stretching experiment on the 14-bead mesoscopic model of an AH (with a length of 70.2 Å at a temperature of 300 K and a rate of 0.1 m/s). The mesoscale curve shows the three typical regimes observed in full atomistic simulations: an elasticity regime (before the first peak occurs), an energy dissipation regime, which corresponds to the unfolding of the 13 bonds (corresponding to the 13 peaks on the curve), and the regime of stretching of the backbone bonds. The peaks in the mesoscale model, corresponding to the rupture forces, agree very well with the rupture force measured from full atomistic simulation. We note that the full atomistic simulation curve presents only small fluctuations since the atomistic system has about 3,000 times more particles (atoms) than the mesoscale model, and thus a larger number of interactions contribute to the damping of force relaxation during unfolding of convolutions. The dotted line approximates the increase of rupture force with increasing strain. In agreement with atomistic simulations reported earlier [28], the third regime sets in at approximately 135% strain

sults very closely. Both full atomistic and mesoscale models predict an initial rupture force, referred to as the angular point, of approximately 350 pN, with a slight increase as the strain is increased. The results shown in Fig. 6 illustrate a key advantage of the coarse-grained model in reaching much longer timescales than what could be achieved in full atomistic simulations (the current limit in MD simulations is 0.01 m/s, whereas we have easily reached a 100-fold increase in accessible timescales using our mesoscale model). Experimental results of stretching and breaking single AH domains [35, 36] (with a length of less than 100 Å) report forces between 140 and 240 pN during unfolding, corresponding to the force level predictions at ultraslow pulling speeds.

Next, we illustrate the application of this model to AH protein domains with three distinct lengths, L: 5, 14, and 81 beads (thus corresponding to $L = 21.6$ Å, 70.2 Å, and 432 Å lengths, respectively, or equivalently, 4, 13, and 80 convolutions). Figure 8 shows the strength obtained from our model for these three different lengths. We observe that the rupture strength decreases as the length increases. The first force peak of the five-convolution system is approximately 500 pN and is thus almost twice as high as the 81-convolution structure, where the strength approaches 250 pN. We note that the simulation of the system with 432 Å length would not have been possible with a full atomistic simulation (the atomistic simulation for a 70.2 Å protein structure took several weeks of computational time). We have used the simulation results to fit an empirical equation of the form

$$f(L) = a \ln(L/L_0) + b \tag{12}$$

where we find $a = -74.463$ pN, $b = 689.08$ pN, and $L_0 = 1$ Å. Extrapolating the strength values predicted from Eq. (12) to ultralong protein lengths yields a strength of 125 pN at a length of $2,000$ Å

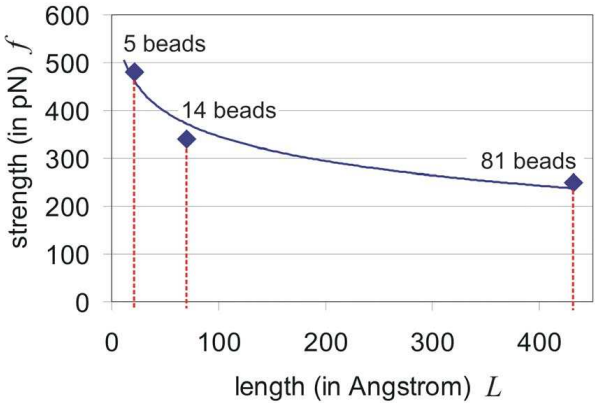

FIGURE 8. Strength properties of AHs with different lengths, ranging from $L = 21.6$ Å to $L = 432$ Å, at a 0.1 m/s pulling rate. The results illustrate that the strength decreases as the length of the AH increases. This behavior also explains the continuous strengthening effect observed in Fig. 7. As more turns are broken at increasing strain, the force required to break the increasingly short AH segments increases. The continuous line shows a fit to the data obtained from mesoscale simulations (see Eq. (12) for the equation and numerical fitting parameters)

(= 200 nm), thus only one-fourth of the strength at 21.6 Å. In summary, our model predicts that the strength of AHs decreases as the length increases. Notably, this weakening behavior with increasing molecular lengths also explains the continuous strengthening effect observed in Fig. 7 (indicated using the dotted line). This is because as more turns are broken at increasing strains, the force required to break the increasingly short AH segments increases. We leave further investigation of this length-scale effect to future studies.

We present a second application of our mesoscopic model to study the shear strength of two parallel AH proteins with the same three lengths: 21.6 Å, 70.2 Å, and 432 Å. Figure 9 shows a schematic of the coarse-grained model for the shearing of two AH protein domains in parallel (Fig. 9(a)). The extremity of one strand is fixed, and the second strand is pulled in the opposite direction at a constant velocity $v = 0.1$ m/s. Figure 9(b) depicts the entire force-strain curve for a shearing experiment on the mesoscopic model of two parallel AH protein domains (here we use a length of 14 beads for each AH strand, a pulling velocity of 0.1 m/s, and a temperature of 300 K). The shear strain is defined as the ratio between the extension and the equilibrium distance between two AH strands. This plot shows three distinct regimes: first, an elasticity regime, as the shearing begins; then a plateau regime, which corresponds to the sliding of the pulled strand along the fixed strand; and last, a zero-force regime, once the two strands separate. The average of the force value over the plateau regime provides an estimate for the mean shear strength. Figure 10 shows snapshots during shearing of two AHs: a short and a long AH structure. The predominant deformation mechanism is shearing, with only a few unfolded convolutions appearing in the long case (red segments). This behavior is expected since the shear resistance of 83 pN is lower than the rupture strength of AHs, which approaches values of more than 200 pN (see the results in Fig. 8). Figure 11 shows the shear strength prop-

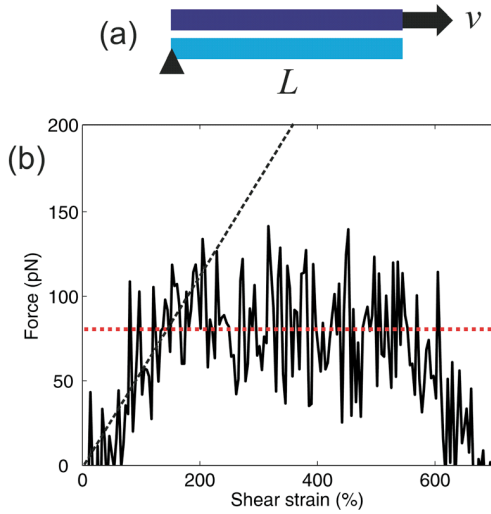

FIGURE 9. Shear strength of an assembly of two AHs. **(a)** Schematic of the coarse-grained model set up for the shearing of two aligned AH protein domains. One strand has its end fixed and the second strand is pulled in the opposite direction at a constant velocity $v = 0.1$ m/s. **(b)** Entire force strain curve for a shearing experiment on the mesoscopic model of two parallel AH protein domains (here we use 14 beads per AH strand, a pulling velocity of 0.1 m/s, and a temperature of 300 K). The shear strain is defined as the ratio between the extension and the equilibrium distance between two AH strands. This plot shows three different regimes: first, an elasticity regime, as we start shearing; then a plateau regime, which corresponds to the sliding of the pulled strand along the fixed strand; and last, a zero-force regime, once the two strands separate. The average of the force value over the plateau regime gives the mean shear strength (indicated with the horizontal red dashed line)

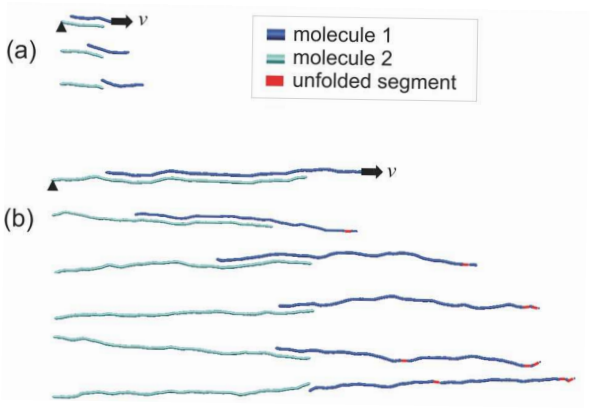

FIGURE 10. Dynamical mechanisms of shearing two AHs for **(a)** a short length of 70.2 Å (14 beads) and for **(b)** a long length of 432 Å (81 beads). The predominant deformation mechanism is sliding, with some unfolded convolutions appearing in the long case (visualized by red segments)

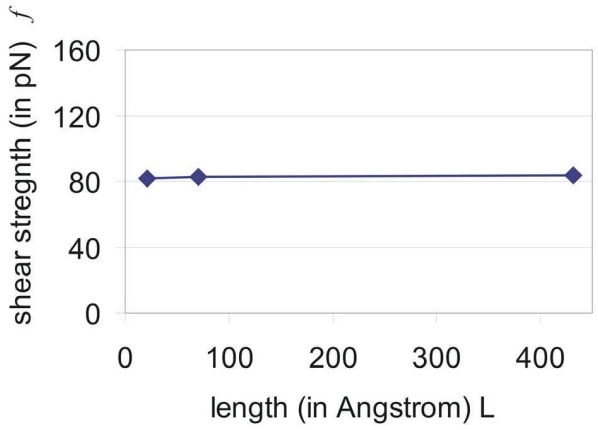

FIGURE 11. Shear strength properties of two parallel AHs with different lengths ranging from $L = 21.6$ Å to $L = 432$ Å, at a 0.1 m/s pulling rate. For different lengths, the plot shows the mean shear strength obtained from the force strain curve of the shearing experiment. The results illustrate that the shear strength does not significantly depend on the length, as opposed to the tensile strength, where the mean shear strength value is approximately 83 pN. The constant shear strength behavior may be explained by a simple theoretical model based on variation of surface energy (model described in Eq. (13)). We note that for this range of lengths, the shear strength is below the tensile strength (which is > 200 pN; see Fig. 8), and thus we do not observe unfolding of the whole AH strand as shear deformation occurs

erties of two parallel AHs obtained with three different lengths ranging from $L = 21.6$ Å to $L = 432$ Å, at a 0.1 m/s pulling rate. The results illustrate that the shear strength does not significantly depend on the length as opposed to the tensile strength. On average, the mean shear strength value is about 83 pN.

The shear strength may be estimated by a simple theoretical model based on the variation of the total surface energy of the system as the two AHs are sheared. During the shearing experiment, once failure due to intermolecular shear initiates, the pulled AH slides relative to the fixed AH, and thus the surface of contact between the two strands decreases. Therefore the total surface energy of the system increases as the strands are being separated. Thus, for a shear extension ΔL, the work done by the pulling force f should equal the increase of surface energy of the system:

$$f \cdot \Delta L = \gamma \cdot \Delta L \tag{13}$$

where γ is the adhesion energy per length as measured from full atomistic simulations ($\gamma = 1.262$ kcal/mol/Å, dimension of force). Equation (13) leads to

$$f = \gamma \tag{14}$$

According to this model, the theoretical shear strength equals 87.6 pN. The theoretical value of the shear strength is constant and thus agrees with the existence of a shear strength plateau regime, as observed in the mesoscopic simulations. Moreover, the theoretical value of the shear strength approximately equals the shear strength value measured from simulations (83 pN). Therefore this simple theoretical model seems to be relevant to describe the shear strength between parallel AHs. The small difference between the numerical values of theory and simulation may be explained by the fact that our simple theoretical model does not include the effect of the pulling rate, which might shift the measured shear strength to larger values.

4. DISCUSSION AND CONCLUSION

The coarse-grained model enables us to simulate the dynamics of large systems over a large range of length and timescales. The model is capable of reaching timescales of several microseconds and longer, with a quantitative accuracy comparable with full atomistic MD simulations. Such relatively long simulations can be carried out within several days of computational time (on a single Intel Xeon CPU). In comparison, MD simulations of the dynamical behavior at fractions of microseconds can take weeks and months of computational time (even on a large parallelized simulation setup). This reflects a considerable speedup due to the coarse-graining approach, while the model is still capable of describing the small- and large-deformation force-strain response characteristics (e.g., softening at $\approx 10\%$ strain and stiffening at $\approx 135\%$ strain) as well as strength values quite accurately. The application to structures of different lengths and arrangements (see the results shown in Figs. 8–11) illustrates the unique ability of our mesoscale model to describe how structure (that is, size, length, geometry) influences strength properties and mechanisms of deformation. Earlier attempts at describing the rupture mechanics of protein domains with coarse mesoscale models have not included such a level of

detail and have thus not been capable of describing how structure and strength properties are linked. Our results show that intermolecular sliding is the dominating mechanism of deformation, suggesting that AHs remain intact during deformation of AH assemblies. Further investigations to elucidate the effects of varying amino acid sequences on this behavior will be necessary to understand better how the chemical structure and overall mechanical deformation mechanisms and strength properties are related.

Potential applications of the coarse-grained AH protein domain model presented here could be further studies of length-scale effects on AH strength and elasticity, and the effects of hierarchical arrangements of AH-based protein domains. Further studies could focus on larger variations of timescales (e.g., to extend to even slower pulling speeds) and the development of similar formulations for coiled-coil proteins or larger-level hierarchical protein structures.

ACKNOWLEDGMENTS

This research was supported by the Air Force Office of Scientific Research (AFOSR). Additional support from NSF, the Office of Naval Research (ONR), and DARPA is greatly acknowledged. J.B. acknowledges support from the Schoettler Fellowship Program at the Massachusetts Institute of Technology.

REFERENCES

1. Alberts, B., et al., *Molecular Biology of the Cell*. Taylor & Francis, New York, 2002.
2. Astbury, W. T., and Street, A., X-ray studies of the structures of hair, wool and related fibres. I. General. *Trans. R. Soc. London, Ser. A*. **230**:75–101, 1931.
3. Weiner, S., and Wagner, H. D., The material bone: Structure mechanical function relations. *Annu. Rev. Mater. Sci.* **28**:271–298, 1998.
4. Vincent, J. F. V., *Structural Biomaterials*. Princeton University Press, Princeton, NJ, 1990.
5. Fratzl, P., and Weinkamer, R., Nature's hierarchical materials. *Prog. Mater. Sci.* **52**:1263–1334, 2007.
6. Courtney, T. H., *Mechanical Behavior of Materials*. McGraw-Hill, New York, 1990.
7. Broberg, K. B., *Cracks and Fracture*. Academic Press, San Diego, 1990.
8. Hirth, J. P., and Lothe, J., *Theory of Dislocations*. Wiley-Interscience, New York, 1982.
9. Engler, A. J., Sen, S., Sweeney, H. L., and Discher, D. E., Matrix elasticity directs stem cell lineage specification. *Cell.* **126**:677–689, 2006.
10. Gautieri, A., Vesentini, S., Redaelli, A., and Buehler, M. J., Single molecule effects of osteogenesis imperfecta mutations in tropocollagen protein domains. *Protein Sci.* **18**(1):161–168, 2009.
11. Suresh, S., Spatz, J., Mills, J. P., Micoulet, A., Dao, M., Lim, C. T., Beil, M., and Seufferlein, T., Connections between single-cell biomechanics and human disease states: Gastrointestinal cancer and malaria. *Acta Biomater.* **1**:15–30, 2005.
12. Jin, Y.-S., Rao, J., and Gimzewski, J. K., Nanomechanical analysis of cells from cancer patients. *Nature Nanotechnol.* **2**:780–783, 2007.
13. Tirion, M., Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett*. **77**:1905–1908, 1996.
14. Haliloglu, T., Bahar, I., and Erman, B., Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* **79**:3090–3093, 1997.
15. Hayward, S., and Go, N., Collective variable description of native protein dynamics. *Annu. Rev. Phys. Chem.* **46**:223–250, 1995.
16. Tozzini, V., Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **15**:144–150, 2005.
17. West, D. K., Brockwell, D. J., Olmsted, P. D., Radford, S. E., and Paci, E., Mechanical resistance of proteins explained using simple molecular models. *Biophys. J*. **90**:287–297, 2006.
18. Dietz, H., and Rief, M., Elastic bond network model for protein unfolding mechanics. *Phys. Rev. Lett.* **100**(9): paper # 098101, 2008.
19. Sulkowska, J. I., and Cieplak, M., Mechanical stretching of proteins—a theoretical survey of the Protein Data Bank. *J. Phys. Condens. Matter*. **19**:283201, 2007.
20. Bathe, M., A finite element framework for computation of protein normal modes and mechanical response. *Proteins Struct. Function Bioinformatics*. **70**:1595–1609, 2008.
21. Bahar, I., and Jernigan, R., Inter-residue poten-

tials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* **266**:195–214, 1997.

22. Nguyen, H., and Hall, C., Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U. S. A.* **101**:16180–16185, 2004.
23. Nguyen, H., and Hall, C., Spontaneous fibril formation by polyalanines: Discontinuous molecular dynamics simulations. *J. Am. Chem. Soc.* **128**:1890–1901, 2006.
24. Freddolino, P., Arkhipov, A., and Schulten, K., Coarse-grained molecular dynamics simulations of rotation-induced structural transitions in the bacterial flagellum. *Biophys. J.* **91**: 4589–4597, 2007.
25. Buehler, M. J., Nature designs tough collagen: Explaining the nanostructure of collagen fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **103**:295102, 2006.
26. Buehler, M., Molecular nanomechanics of nascent bone: Fibrillar toughening by mineralization. *Nanotechnology*. **18**:XX–XX, 2007.
27. Ackbarow, T., Chen, X., Keten, S., and Buehler, M. J., Hierarchies, multiple energy barriers and robustness govern the fracture mechanics of alpha-helical and beta-sheet protein domains. *Proc. Natl. Acad. Sci. U. S. A.* **104**:16410–16415, 2007.
28. Ackbarow, T., and Buehler, M. J., Superelasticity, energy dissipation and strain hardening of vimentin coiled-coil intermediate filaments: Atomistic and continuum studies. *J. Mater. Sci.* **42**:8771–8787, 2007.
29. Karcher, H., Lee, S. E., Kaazempur-Mofrad, M. R., and Kamm, R. D., A coarse-grained model for force-induced protein deformation and kinetics. *Biophys. J.* **90**:2686–2697, 2006.
30. Lazaridis, T., and Karplus, M., Effective energy function for proteins in solution. *Proteins Struct. Function Genet.* **35**:133–152, 1999.
31. Lazaridis, T., and Karplus, M., "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science*. **278**:1928–1931, 1997.
32. Bell, G. I., Models for the specific adhesion of cells to cells. *Science*. **200**:618–627, 1978.
33. Hanggi, P., Talkner, P., and Borkovec, M., Reaction-rate theory: Fifty years after Kramers. *Rev. Mod. Phys.* **62**:251–341, 1990.
34. Zhurkov, S. N., Kinetic concept of the strength of solids. *Int. J. Fract. Mech.* **1**:311–323, 1965.
35. Lantz, M. A., Jarvis, S. P., Tokumoto, H., Martynski, T., Kusumi, T., Nakamura, C., and Miyake, J., Stretching the alpha-helix: A direct measure of the hydrogen-bond energy of a single-peptide molecule. *Chem. Phys. Lett*. **315**:61–68, 1999.
36. Kageshima, M., Lantz, M. A., Jarvis, S. P., Tokumoto, H., Takeda, S., Ptak, A., Nakamura, C., and Miyake, J., Insight into conformational changes of a single alpha-helix peptide molecule through stiffness measurements. *Chem. Phys. Lett*. **343**:77–82, 2001.