HIERARCHICAL CHEMO-NANOMECHANICS OF PROTEINS: ENTROPIC ELASTICITY, PROTEIN UNFOLDING AND MOLECULAR FRACTURE

Markus J. Buehler

HIERARCHICAL CHEMO-NANOMECHANICS OF PROTEINS: ENTROPIC ELASTICITY, PROTEIN UNFOLDING AND MOLECULAR FRACTURE

Markus J. Buehler

Proteins are an integral part of nature's material design. Here we apply multiscale modeling capable of providing a bottom-up description of the nanomechanics of chemically complex protein materials under large deformation and fracture. To describe the formation and breaking of chemical bonds of different character, we use a new reactive force field approach that enables us to describe the unfolding dynamics while considering the breaking and formation of chemical bonds in systems that are comprised of several thousand atoms. We particularly focus on the relationship between secondary and tertiary protein structures and the mechanical properties of molecules under large deformation and fracture. Our research strategy is to systematically investigate the nanomechanics of three protein structures with increasing complexity, involving alpha helices, random coils and beta sheets. The model systems include an alpha helical protein from human vimentin, a small protein α-conotoxin PnIB from *conus pennaceus*, and lysozyme, an enzyme that catalyzes breaking of glycosidic bonds. We find that globular proteins can feature extremely long unfolding paths of several tens of nanometers, displaying a characteristic sawtooth shape of the force-displacement curve. Our results suggest that the presence of disulfide crosslinks can significantly influence the mechanics of unfolding. Fibrillar proteins show shorter unfolding paths and continuous increase of forces until molecular rupture occurs. In the last part of the article we outline how a mesoscale representation of the alpha helical protein structure can be developed within the framework of hierarchical multiscale modeling, utilizing the results of atomistic modeling, without relying on empirical parameters. We apply this model to describe the competition between entropic and energetic elasticity in the mechanics of a single alpha helical protein molecule, at long time scales reaching several microseconds. We conclude with a discussion of hybrid reactive-nonreactive modeling that could help to overcome some of the computational limitations of reactive force fields.

1. Introduction

The behavior of biological systems is controlled by a complex interplay of a large set of macromolecules, chemical solvents and external stimuli such as mechanical forces or strain. Cells, for example, represent exceptionally complicated systems that feature heterogeneous structures across many length- and time-scales, including ribosomes, protein networks, microtubules, DNA and the cell membrane. Many structural materials found in nature, such as bone and nacre, display a clever heterogeneous design that includes proteins, inorganic phases and solvents.

To date, the function and mechanical response of biological structures and materials is largely limited to phenomenological concepts that characterize the behavior of biological systems from a macroscopic perspective, by introducing a set of empirical parameters, using a top-down approach. Quantitative theories, in particular those that link the scale of chemistry and molecular properties to the scales of materials, or to the scale of complex biological systems that comprise of several thousands of molecules, are still missing.

This lack of understanding is partly due to difficulties in handling and measuring properties of such tiny structures with dimensions of several nanometers and below, and force levels that are often limited to several pN or nN. Carrying out highly specific experiments with high spatial and temporal resolution at these force levels represents a significant challenge.

Experimental methods developed recently now enable us to investigate the nanoscale behavior of materials using quantitative analysis techniques. For example, nanoindentation, AFM, optical and magnetic tweezers enables scientists to probe the origins of elastic and plastic deformation of materials, with forces in the range of pN to μN, and at scales approaching that of individual atoms, molecules or cells [Gouldstone et al. 2001; Sun et al. 2001; Dao et al. 2003; Sun et al. 2004]. At the same time availability of computational resources and new theoretical approaches have led to significant advances in addressing nanomechanics from a first principles viewpoint [Marko and Siggia 1995; MacKerell et al. 1998]. Combining experimental with computational or theoretical studies could lead to an alternative to the classical top-down engineering approach, by providing a bottom-up materials description linking small to large [Whitesides and Wong 2006]. The multiscale approach is visualized schematically in Figure 1.

Materials found in nature often feature hierarchical structures ranging from the atomistic and molecular to macroscopic scales [Hulmes et al. 1995; Sasaki and Odajima 1996; Jager and Fratzl 2000; Puxkandl et al. 2002; Aizenberg et al. 2005; Whitesides and Wong 2006]—a variation which renders this class of materials both fascinating and extremely challenging. Examples of such materials include bone, tendon, or nacre. Moreover, many biological materials found in living organisms primarily utilize proteins as fundamental building blocks, creating fascinating materials whose functions range from load bearing and serving as catalysts to intercellular signaling. Proteins are a particularly intriguing class of biopolymers representing a complex three-dimensional folded structure of one or more polypeptide chains. Proteins play a particularly important role in many biological tissues and functions, including tendon, bone, teeth, or cartilage and even in the cardiovascular system. Severe mechanical tensile and shear loading of proteins can occur under physiological conditions, as in joints and in bone. In other cases, extreme mechanical stimulation can lead to malfunction and disease. The properties of proteins represent a complicated and intertwined interplay of mechanics, chemistry and biological function, creating multifunctional, active or *smart* materials out of primarily only 20 distinct building blocks, the naturally occurring amino acids.

Our long-term objective is to contribute to develop a rigorous understanding of the mechanics of complex biological protein materials, while considering atomistic and molecular scales, bridging to larger time and length scales. To reach this goal we develop atomistic models of the nanomechanical properties of globular and fibrillar proteins. In this article we focus on the source of elasticity, deformation and fracture of single protein molecules. We apply a new modeling approach based on reactive force fields that enables us to treat complex chemistry in systems comprising several thousand atoms.
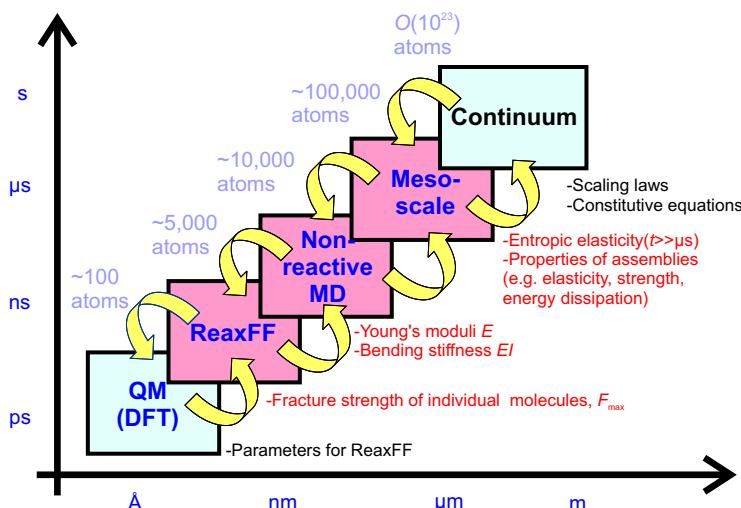
Figure 1. Summary of a hierarchical multiscale scheme used to gain an understanding of the behavior of biological materials, across scales in length and time. First principles quantum mechanics (QM) calculations (for example, Density Function Theory, DFT [Springborg 1997]) are carried out to train a reactive force field ReaxFF [van Duin et al. 2001]. The reactive force field is used together with nonreactive force fields [MacKerell et al. 1998] to obtain properties of individual protein molecules and assemblies of several molecules [Buehler 2006a; 2006b]. The results of atomistic calculations can then be coupled to continuum scale models, for example, by using scaling laws [Buehler 2006a; 2006d]. We note that reactive force field calculations are significantly more expensive, which typically limits us to sub-ns time scales. This can influence the observed trajectories, as strain rates may be unrealistically large. This limitation could be overcome by parallelization or development of faster computational resources.

1.1. *Nanomechanics of protein materials: laboratory experiments.* When materials are deformed, they display a small regime in which deformation is reversible or elastic. Once the forces on the material are increased, deformation becomes irreversible and can involve fracture. Deformation and fracture of materials is controlled by atom-by-atom processes that are eventually governed by quantum mechanics, or quantum chemistry. The deformation mechanics of brittle materials (for example, ceramics, silicon, glass, some polymers) and ductile materials (for example, copper, nickel) has been subject to extensive and very successful research over the past decades [Buehler and Gao 2006b]. Figures 2(a) and (b) depict a schematic of the fundamental deformation mechanisms in these materials that include crack propagation or dislocation nucleation and interaction [Buehler et al. 2003; 2004; 2005; Hartmaier et al. 2005; Buehler and Gao 2006a].

However, similar mechanisms are not yet well understood for biological materials, and rigorous deformation theories are still missing. Figure 2c depicts a schematic of a hierarchical biological material that consists of a heterogeneous assembly of building blocks. The response of the material depends on the mechanical and interface properties of its building blocks (for example, protein molecules, nanocrystals
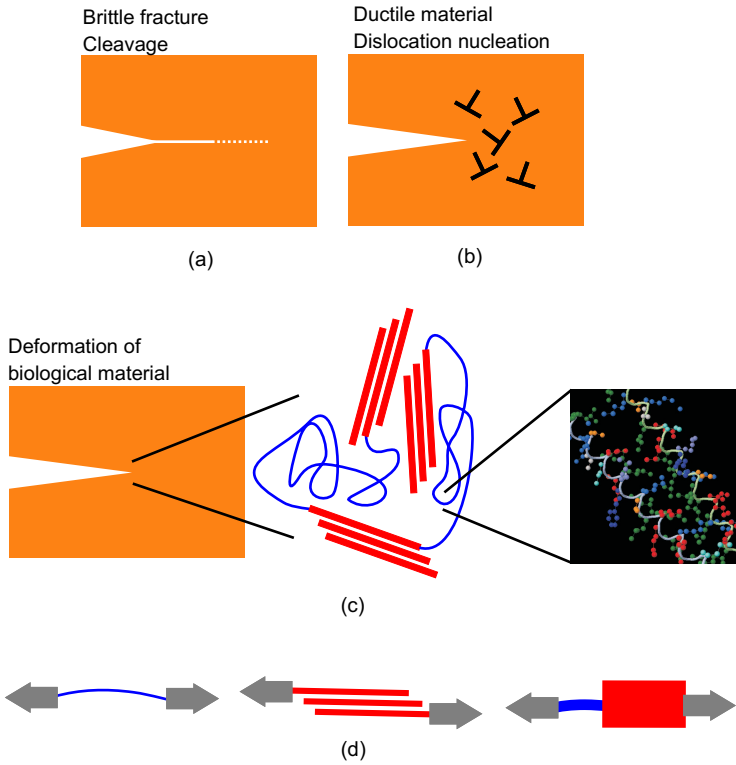
Figure 2. Response of different classes of materials to extreme mechanical stimulation. The response of brittle (subplot (a)) and ductile (subplot (b)) materials is relatively well understood, with theories describing crack extension and dislocation nucleation and propagation. However, the response of biological materials to mechanical loading, for example, the response of materials to the large stresses at the tip of a microcrack, remains an active area of research, since little understanding exists about how specific components of hierarchical materials participate in deformation and how they contribute to the macroscopic material properties (subplot (c)). Biological materials feature several layers of complexity, and the key to understanding their behavior is the ability to decipher the response of its building blocks (for example, random coils, beta sheets and assemblies of those, or globular folded structures). Subplot (d) shows a schematic decomposition of the complex three-dimensional structure into its building blocks. Each building block is studied under a variety of different types of mechanical loading conditions. Information from these studies is then used to build models that are capable of describing the entire, realistic material nano-structure of the material.

or other components), as illustrated in Figure 2d. The apparent need to understand the properties of nature's buildings better has motivated systematic investigations of the nanomechanical properties of individual protein molecules.

The nanomechanics of individual proteins has been subject to an intense debate over the last decades that has led to significant advances in understanding the behavior of their mechanical response, leading

to estimates for their Young's modulus, the persistence length or the bending stiffness [Rief et al. 1997; 2000]. For example, in a recent study carried out using optical tweezers researchers have measured the persistence length of individual proteins [Rief et al. 1997; 2000; Mehta et al. 1999; Sun et al. 2001; Schwaiger et al. 2002]. A widely used approach determine the persistence length is to employ optical tweezers to obtain force-stretching curves. The persistence length can be extracted from such experiments by assuming that the elastic response can be described using a worm-like chain (WLC) model. If the contour length and temperature is known, the persistence length is the only parameter in the WLC model [Bustamante et al. 1994; Marko and Siggia 1995] and can thus be determined to fit the experimental results. For example, this approach has also been used to study the persistence length of a type I tropocollagen molecule [Sun et al. 2004]. Similar techniques has also been successfully applied to titin [Rief et al. 1997; 2000; Gao et al. 2001; 2002], DNA and many other biomolecules.

However, forces in experiment are typically limited to a few nN, which approximately equals the strength of covalent bonds. This limits the applicability of such experiments to study the large-strain elastic properties and fracture properties of individual molecules or assemblies of molecules. Further, some experimental techniques are limited in terms of their temporal and spatial resolution. There is indirect experimental evidence of large-deformation and fracture of protein materials, as outlined in several experimental studies of bone, mineralized tendon and tendon [Landis et al. 2002; Nalla et al. 2003; 2005; Ritchie et al. 2004; Gupta et al. 2004; 2005; Peterlik et al. 2006].

1.2. *Hierarchical multiscale modeling of protein molecules: reactive versus nonreactive force fields.*
Atomistic and molecular modeling [Wang et al. 2001; Karplus and McCammon 2002; Buehler 2006c] can supplement experiment by providing a highly specific, controlled and fundamental method to describe the nanomechanical properties of biological matter in general, in particular those of proteins.

Atomistic models are capable of simulating the motion of all atoms in a material, with systems comprised of up to several billion particles [Kadau et al. 2004], reaching scales of several micrometers, that are getting close to macroscopic scales where material behavior that can be directly observed in experiment [Buehler 2006a, 2006d]. Atomistic models enable us to probe the macroscopic response of materials, for example, due to mechanical stimulation, based on their fundamental, atomistic ultrastructure, including the complexities of the chemical interactions in the material without introducing empirical parameters. The key input parameter in atomistic modeling are the interatomic forces. Classical molecular dynamics [Leach 2001] force fields suitable to study the properties of proteins include CHARMM, AMBER and DREIDING [Wang et al. 2001; Karplus and McCammon 2002; Li and Arteca 2005] that provide a reasonably accurate representation of the molecular structure and energetics.

MD simulation using an approach referred to as steered MD allows application of atom-specific forces in large molecules. This method has been applied to study the unfolding dynamics of several fibrillar and globular proteins [Gao et al. 2001; 2002; Arteca 2003; Cieplak and Marszalek 2005; Lorenzo and Caffarena 2005]. Some of these simulations have been used to resemble AFM experiments of protein unfolding [Rief et al. 1997; 2000].

Even though molecular simulation has been successfully applied to describe some properties of protein molecules, most classical molecular simulation techniques are limited to describing molecular states close to the equilibrium configuration, and cannot be applied to describe large strain elastic response and fracture of covalent bonds [Wang et al. 2001; Karplus and McCammon 2002; Li and Arteca 2005;

Buehler 2006b]. The reason for these limitations is that these methods are not capable of describing formation or breaking of chemical bonds, as the methods often do not allow a unified treatment of the various chemical interactions with different bonding strength, treating covalent interactions by using harmonic potentials that lead to an infinitely large energy barrier for bond breaking. So far this is only possible by using quantum mechanical methods (for example, [Papamokos and Demetropoulos 2004] describes a study of nanomechanical properties of a small helical protein).

Figure 3 depicts a schematic representation of this behavior, emphasizing on the limitation of classical force fields so that they are not able to describe the transition state A-B during bond breaking or formation, but can only provide a representation of the ground states A or B. An important consequence of this limitation in regards to the mechanics of proteins is that molecular fracture cannot be described, making biological molecules unrealistically strong [Buehler 2006b]. In the remainder of this article we refer to these models as nonreactive force fields.
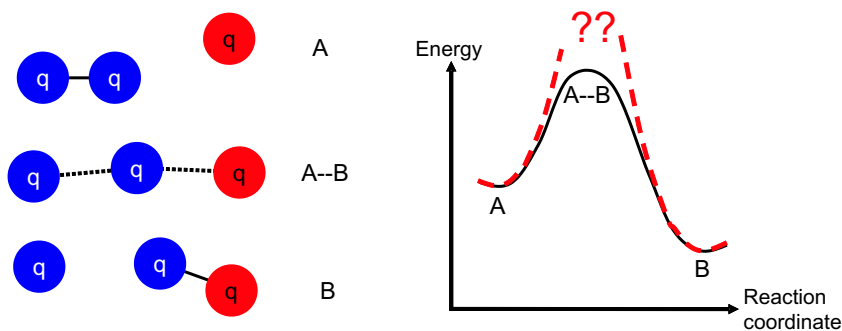

Figure 3. The concept of reactive versus nonreactive force fields. The reactive force fields enable describing the transition state energies between two ground states, as it is important for example during chemical reactions, including bond rupture. Nonreactive force fields (dashed line in the right subplot) are only capable of describing the ground states A and B, but not the transition state A-B.

To overcome the limitations of the nonreactive force fields we propose a new generation reactive force field ReaxFF [van Duin et al. 2001] that is capable of describing formation and breaking of chemical bonds, including the flow of charges during chemical reactions. This method enables us to describe the full reactivity of large systems that include several thousand atoms with quantum mechanical accuracy, providing a more realistic description of the large-deformation elastic behavior, including permanent deformation and fracture. The work described in this article is focused on demonstrating the applicability of this new method to describe the unfolding and stretching dynamics of proteins, enabling, for the first time, the description of molecular fracture of molecules comprised of several thousand atoms.

Both reactive and nonreactive models are limited to time scales of several nanoseconds. However, using a combination of hierarchies of simulation methods, vast time and length scales can be captured, a method referred to as multiscale modeling. Multiscale modeling combines a set of computational tools ranging from first principles quantum mechanics and molecular dynamics to reactive force fields and

	Force regime	Time scale	Entropic elasticity	Specificity, applicability
Experiment (optical tweezer, AFM, etc.)	Typically < 10 nN	$\gg$ fs (slow strain rates)	Yes	Difficult, analysis of atomistic processes often not possible
Atomistic simulation (reactive and nonreactive force fields)	$0, \ldots, 100$ pN (nonreactive force fields) $0, \ldots, 10$ nN (reactive force fields)	Typically < ns (high strain rates)	No (time scale typically too short, molecular structures rather small)	Highly specific, reveals atomistic processes
Mesoscale particle models (for example, bead models)	$0, \ldots, 100$ nN (wide range)	$< \mu$sec$, \ldots,$ sec (depending on level of discretization)	Yes	Highly specific, but no atomistic resolution

Table 1. Differences between various experimental and computational methods that allow probing single molecule mechanics.

mesoscale models. The concept of integrating various simulation methods by handshaking to bridge across the scales is schematically represented in Figure 1.

Table 1 includes an overview over various properties that can be probed using nanomechanical experiments and atomistic or molecular simulation, illustrating how the different methods can supplement each other.

1.3. *Outline.* The focus of this article is application of reactive atomistic models to describe the large-deformation elastic and fracture behavior of proteins. After this introduction we briefly review the physical foundations of elasticity in biological molecules and biological materials in Section 2. In Section 3 we describe our multiscale modeling approach, including a discussion of the reactive force fields. In Section 4 we review a detailed analysis of three specific applications. First, we describe atomistic and mesoscale studies of stretching a single alpha helix, a building block of other proteins and protein materials including keratin, elastin and intermediate filaments in the cell's cytoskeleton. We continue with a discussion of atomistic modeling of the unfolding dynamics of a crosslinked protein α-conotoxin PnIB from *conus pennaceus*. Finally, we discuss the unfolding of lysozyme, a well-studied enzyme that catalyzes breaking of glycosidic bonds. In Section 5 we present a broader discussion on chemical-mechanical interactions in biological materials and methods to bridge full atomistic to larger time and length scales, including coarse-graining techniques. We conclude with a brief outline of how such methods enable us to describe entropic elasticity using molecular dynamics and an example that demonstrates how reactive and nonreactive force fields can be combined in a concurrent multiscale scheme.

2. Physical foundations of elasticity: entropic and energetic contributions

Elasticity stems from the collective interactions of atoms, and, thus, it is intimately linked to chemistry. The elastic properties of materials can be expressed as the partial derivative of the free energy density with respect to the strain tensor that characterizes the resistance to deformation. Starting with free energy $A = U - TS$ composed of energetic U and entropic contributions TS, we can define the free energy density, that is, free energy per unit volume V, via $\Phi = A/V$. The (scalar) stress σ and the elastic modulus E are then given by

$$\sigma = \frac{\partial \Phi}{\partial \varepsilon}, \qquad E = \frac{\partial^2 \Phi}{\partial \varepsilon^2}. \tag{1}$$

Note that stress and strain are related by Hooke's law, $\sigma = E\varepsilon$ [Courtney 1990]. Whereas the elasticity of crystalline materials is primarily controlled by energetic changes of the internal energy, in natural and biological materials elasticity can be significantly influenced by entropic contributions. This is because in many crystalline materials the entropic term can be neglected, so that Φ in Equation (1) can be directly substituted by U/V, the internal energy. However, in biopolymers entropic contributions can dominate the elasticity, in particular for small deformations, and therefore Φ in Equation (1) can be substituted by $-TS/V$.

Dominance of entropic behavior is a well-known phenomenon in many polymers. The contributions to the entropic term to elasticity can be described in several ways, including classical descriptions such as the WLC or the freely jointed Gaussian chain model [Bustamante et al. 1994; Marko and Siggia 1995]. Such descriptions are similar to constitutive models in continuum elasticity, and require input parameters that are typically determined empirically. In contrast to these models, molecular dynamics modeling can provide a first principles based description of entropic elasticity without any additional fitting parameters beyond the atomic interactions. This can be achieved by calculating the bending stiffness using full atomistic simulations.

The persistence length is defined as the molecular length at which entropic contributions to elasticity become important, as the molecule shows significant bending purely due to its thermal energy. A molecule with length far beyond the persistence length will bend, even without application of forces, and assume a conglomerated, wiggly shape. With the bending stiffness of a molecule denoted as EI, the persistence length is defined as $\xi_p = EI/(k_B T)$. When the length of molecules, denoted by L, is greater than the persistence length, that is, $L \gg \xi_p$, thermal energy can bend the molecule, and entropic elasticity typically plays a role. On the other hand, when $L \ll \xi_p$, entropic effects play a minor role, and energetic elasticity governs. Entropic effects become important and appear in measurements, for example, when one stretches a convoluted molecule.

Assuming that the initial point-to-point distance is $x < L$ (expressing the fact that the molecule is convoluted), the force that resists stretching can be approximated by

$$F(x; L, \xi_p) = \frac{kT}{\xi_p} \left(\frac{1}{4} \frac{1}{(1-x/L)^2} - \frac{1}{4} + \frac{x}{L} \right). \tag{2}$$

This model is called the WLC or Marko–Siggia equation [Bustamante et al. 1994; Marko and Siggia 1995]. The molecular properties enter this equation in form of the persistence length, which is a function of the bending stiffness. If these properties are known from atomistic calculations, the WLC model

provides a quantitative estimate of entropic elasticity. The WLC model is only valid for deformations sufficiently far away from the contour length, as $F(x)$ diverges when $x \to L$.

Stretching single molecules typically involves a transition between two extreme regimes of sources of elasticity. Entropic elasticity controls the resistance to stretch the molecule when the end-to-end distances below the contour length. When the molecules is stretched to its contour length $x = L$, the entropic contributions decrease to zero and energetic elasticity starts to dominate. Thereafter the force increases proportionally to the stretching distance, according to a specific spring constant that relates distance of stretch and restoring force via

$$F(x; k) = k(x, L)(x - L). \tag{3}$$

Note that this equation is only valid when $x \geq L$, since the molecule can easily buckle under compressive load. When $x < L$, the restoring force assumes values close to zero. Also, note that $k(x, L)$ is a function of the deformation, in general, indicating that the elastic properties can change with deformation, a phenomenon referred to as nonlinear elasticity or hyperelasticity. At the atomistic scale energetic elasticity is characterized by the stretching of atomic, metallic, covalent or ionic bonds that leads to a change in potential energy in the material volume [Buehler 2006b].

3. Computational modeling of atomic interactions: from chemistry to mechanics

We provide a brief review of atomistic modeling techniques used for the studies reported in this article. The behavior of molecules is intimately linked to the interactions of atoms, which are governed by the laws of quantum chemistry. In metals, for example, bonding is primarily nondirectional, and can be characterized by positive ions embedded in a gas of electrons. Other materials show much greater chemical complexity, often featuring many different chemical bonds with varying strength.

In biological materials it is vital to consider the interplay of chemical interactions that include (ordered by their approximate strength): covalent bonds (due to overlap of electron orbitals); electrostatic interactions (Coulombic interactions); hydrogen bonds; and weak or dispersive van der Waals (vdW) interactions. We note that electrostatic interactions can be significantly weakened by screening due to electrolytes, which can lead to interactions that are weaker than vdW interactions [Feig and Brooks 2004].

In proteins covalent interactions are primarily responsible for the chemistry within the polypeptide amino acid chains. The three-dimensional folded structure is stabilized by a combination of hydrogen bonding, dispersive interactions and electrostatic interactions. In addition to the weak interactions, the structure of proteins is sometimes stabilized by covalent cross-links, such as disulfide bridges between different amino acids as they can be formed between two cysteine (CYS) residues. Even though some of these interactions are relatively weak (up to 1,000 times weaker than covalent bonding), they play an overarching in stabilizing many biological molecules.

3.1. *Atomistic model: classical nonreactive force fields.* Our basic modeling approach is based on the classical force field CHARMM [MacKerell et al. 1998], implemented in the MD program NAMD [Nelson et al. 1996]. The CHARMM force field is widely used in the protein and biophysics community, and provides a reasonably accurate description of proteins. Other popular examples are the AMBER force field [Wang et al. 2001] and the DREIDING force field [Mayo et al. 1990]. Force fields of this type are

typically based on harmonic and anharmonic terms describing covalent interactions, in addition to long-range contributions describing van der Waals (vdW) interactions, electrostatic (Coulomb) interactions, as well as hydrogen bonding. Water molecules are described using the TIP3 water model [Wang et al. 2001].

Since the bonds between atoms are modeled by harmonic springs or its variations, bonds between atoms cannot be broken, and new bonds cannot be formed. Moreover, the charges are fixed and cannot change, and the equilibrium angles do not change depending on stretch. For example, the energy penalty due to stretching of a bond r_{ij} between atoms i and j is given by a harmonic function

$$\varphi(r_{ij}) = \tfrac{1}{2}k_t(r_{ij} - r_0)^2. \tag{4}$$

The harmonic approximation leads to a linear dependence of the stretching force and distance, $F(r_{ij}) \sim k_t r_{ij}$. While such a description is certainly appropriate for small deformation from the equilibrium bond length r_0, it fails to describe the correct energy-distance relationship for large deformation (we emphasize that k_t in Equation (4) is not a function of deformation as in the general formulation provided in Equation (3)). Clearly, this is an unrealistic description for the behavior of biological molecules under large stretch.

Typical time steps in nonreactive force fields are chosen to be $\Delta t = 1$ fs. Thus, by carrying out 10^6-10^7 steps one can reach time scales of several nanoseconds.

3.2. *Reactive force fields.* So far, all attempts have failed to accurately describe the transition energies during chemical reactions using more empirical descriptions than relying on purely quantum mechanical (QM) methods [Papamokos and Demetropoulos 2004] (see also Figure 3).

Reactive force fields [Stuart et al. 2000; van Duin et al. 2001; Brenner et al. 2002] represent a new strategy to overcome some of the limitations classical force fields, in particular the fact that these descriptions are not able to describe chemical reactions. In fact, the behavior of chemical bonds at large stretch has major implications on the mechanical response, as it translates into the properties of molecules at large-strain, a phenomenon referred to nonlinear elasticity or hyperelasticity [Buehler et al. 2003; Buehler and Gao 2006a].

Reactive potentials are based on a more sophisticated formulation than most nonreactive potentials. A bond-length to bond-order relationship is used to obtain smooth transition from nonbonded to single, double, and triple bonded systems. All connectivity-dependent interactions (that is, valence and torsion angles) are formulated to be bond-order dependent. This ensures that their energy contributions disappear upon bond dissociation so that no energy discontinuities appear during reactions. The reactive potential also features nonbonded interactions (shielded van der Waals and shielded Coulomb).

Several flavors of reactive potentials have been proposed in recent years [Stuart et al. 2000; van Duin et al. 2001; Brenner et al. 2002; van Duin et al. 2003]. Reactive potentials can overcome the limitations of empirical force fields and enable large-scale simulations of thousands of atoms with quantum mechanics accuracy. The reactive potentials, originally only developed for hydrocarbons [Stuart et al. 2000; van Duin et al. 2001; Brenner et al. 2002], have been extended recently to cover a wide range of materials, including metals, semiconductors and organic chemistry in biological systems such as proteins [van Duin et al. 2003; Strachan et al. 2003; Nielson et al. 2005; Han et al. 2005; Chenoweth et al. 2005; Strachan et al. 2005; Cheung et al. 2005]. Here we use the ReaxFF formulation [van Duin et al. 2001]. We employ

a particular flavor of the ReaxFF potentials as suggested in [van Duin et al. 2001; Strachan et al. 2005], with slight modifications to include additional QM data suitable for protein modeling [Datta et al. 2005].

Key features of the ReaxFF reactive force fields are [van Duin et al. 2001; van Duin et al. 2003; Strachan et al. 2003; Nielson et al. 2005; Strachan et al. 2005; Chenoweth et al. 2005]:

- a bond-length/bond-order relationship is used to obtain smooth transition (Pauling) from nonbonded to single, double, and triple bonded systems;

- all connectivity-dependent interactions (that is, valence and torsion angles) are made bond-order dependent to ensure that their energy contributions disappear upon bond dissociation;

- they feature shielded nonbonded interactions that include van der Waals and Coulomb interactions;

- ReaxFF uses a geometry-dependent charge calculation scheme (similar to QEq [Rappe and Goddard 1991]) that accounts for polarization effects;

- most parameters in the formulation have physical meaning, such as corresponding distances for bond-order transitions, atomic charges and others.

The reactive formulation uses a geometry-dependent charge calculation scheme similar to QEq [Rappe and Goddard 1991] that accounts for polarization effects and modeling of charge flow. This is a critical breakthrough leading to a new bridge between QM and empirical force fields. All interactions feature a finite cutoff of $R_{\text{cut}} = 10 \text{ Å}$.

Since charge equilibration is carried out at every step, the reactive method can also describe flow of charges during deformation processes of the protein. Such information cannot be obtained from nonreactive force fields. It is apparent that such charge flows may be very significant in biological processes, such as enzymatic reactions, or solvent-protein interactions in general. Even though we will not discuss this aspect in detail, it is an important potential of the reactive force field that could be investigated in future studies.

In ReaxFF the total energy of a system is expressed as the sum of different contributions that account for specific chemical interactions. The total energy is given by

$$E_{\text{system}} = E_{\text{bond}} + E_{\text{vdWaals}} + E_{\text{Coulomb}} + E_{\text{val,angle}} + E_{\text{tors}} + E_{\text{over}} + E_{\text{under}} + E_{\text{res}}.$$

The terms $E_{\text{bond}} + E_{\text{vdWaals}} + E_{\text{Coulomb}}$ are two-body contributions, the terms $E_{\text{val,angle}} + E_{\text{tors}}$ are 3-body and 4-body terms, and $E_{\text{over}} + E_{\text{under}}$ correspond to multibody contributions due to the local chemical environment. The term E_{res} describes energetic contributions of resonance effects.

ReaxFF is based on the concept of bond-orders that dates back to early work by Abell, Tersoff, Brenner and others [Tersoff 1988; Brenner 1990; Stuart et al. 2000; Brenner et al. 2002]. The basic concept of bond-orders is simple to explain. The key idea is to modulate the bond strength based on the atomic environment, taking advantage of some theoretical chemistry principles.

Consider a pair potential in which the total energy of the system is given by the sum over all pairs of atoms (note the factor 1/2 to avoid double counting):

$$U_{\text{total}} = \frac{1}{2} \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} \varphi(r_{ij}). \tag{5}$$

Instead of writing $\varphi(r_{ij})$ as a harmonic function (see above), in the Abell–Tersoff approach the interaction between two atoms is expressed as $\varphi(r_{ij}) = \varphi_R(r_{ij}) - M_{ij}\varphi_A(r_{ij})$, where $\varphi_R(r_{ij})$ and $\varphi_A(r_{ij})$ are pair repulsive and attractive interactions, respectively. The parameter M_{ij} that multiplies the attractive interactions represents a many-body interaction parameter. This parameter describes how strong the attraction is for a particular bond, from atom i to atom j. Most importantly, the parameter M_{ij} can range from zero to one, and describes how strong this particular bond is, depending on the environment of atom i. It can thus be considered a normalized bond-order, following the concept of the Pauling relationship between bond-length and bond-order. Abell suggested that

$$M_{ij} \sim Z^{-\delta}, \tag{6}$$

where δ depends on the particular system, and Z is the coordination number of atom i that depends on the bond radius.

Equations (5) and (6) immediately lead to a relationship between bond-length, binding energy and coordination, through the parameter M_{ij}. The modulation of the bond strength effectively leads to a change in spring constant as a function of bond environment, $k(r) \sim k_0 M(Z, \delta)$. Note that k_0 is a reference spring constant, which is modulated by the atomic environment that is essentially dependent on the bond radius. This method has been very successful in describing the interatomic bonding in several materials, for example the C-C bonds in diamond, graphite and even hydrocarbon molecules. It is also the basis for the ReaxFF force field.

The concept of the bond-length/bond-order dependence used in the reactive force field ReaxFF is shown schematically in Figure 4. Table 2 shows a systematic comparison between reactive and nonreactive force fields.

We note that the coordination number is a concept widely used in lattice systems, for example crystals. In organic molecules, the coordination number can be thought of as the amount of covalent bonds that an atom has made.

We refer the reader to specific publications for the derivation and development of reactive force fields [van Duin et al. 2001; van Duin et al. 2003; Strachan et al. 2003; Chenoweth et al. 2005; Strachan et al. 2005; Cheung et al. 2005]. In the remainder of this paper, we focus on the application of this method to describe large-deformation behavior of proteins.

Due to the increased complexities of force field expressions and the charge equilibration step that is carried out at each force calculation, reactive force fields are between 50 and 100 times more expensive than nonreactive force fields, yet several orders of magnitude faster than DFT-level calculations that would be able to describe bond rupture as well.

Typical time steps in reactive force fields are chosen to be $\Delta t = 0.25$ fs. Thus, by carrying out 10^6-10^7 steps, one can reach time scales of approximately one to two nanoseconds. However, since each step of force calculation is significantly more expensive in reactive force fields than in nonreactive formulations, we can typically only reach time scales less than tens picoseconds, leading to large strain rates. This represents a significant limitation of this method that could be overcome, for instance, by parallelization or development of faster algorithms.

The large strain rates and large forces applied to the protein may induce deformation mechanisms that are different from those that would be seen at slower strain rates of μm/sec, as may, for example, be
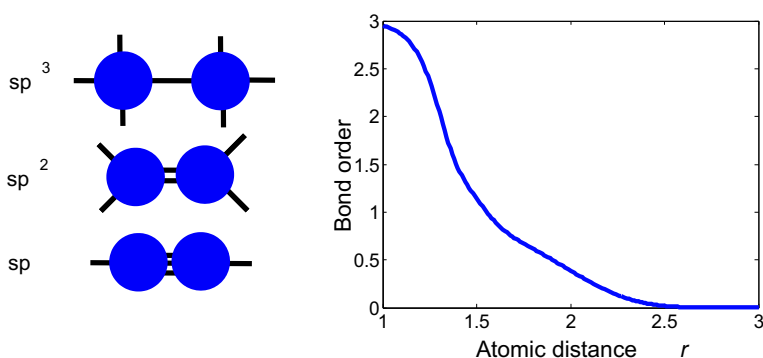
Figure 4. This plot illustrates the concept of bond-orders, here in an example for a C-C bond. Depending on the distance between atoms, different bond-orders are obtained through a bond-order mapping function. This enables us to distinguish different quantum chemical states such as sp^3 (single bond), sp^2 (double bond) and sp (triple bond). The continuous change of bond-orders as a function of distance ensures that reactive force fields are energy continuous, which is critical to carrying out constant energy simulations. At large distances the bond-order vanishes indicating breaking of the covalent bond. In ReaxFF the spring constant that characterizes the strength of atomic bonding is modulated by the bond order leading to vanishing bond strength or dissociation at large stretch.

	Nonreactive FFs	Reactive FFs
Ground state energies (for example, distinguish sp3-sp2-sp...)	Yes (few states)	Yes
Excited / transition states (go from one to another ground state; see also Figure 4)	No	Yes
Breaking of bonds and continuous energies during reactions	No (sometimes: Morse functions for bond breaking but energetics are often wrong)	Yes
Formation of bonds	No	Yes
Charge flow during reactions	No	Yes
Organo-Inorganic interfaces (or between other materials)	No (mostly)	Yes (bridging FFs)
Retyping necessary after reaction	Yes (have C_2,C_3 etc. for different hybridization [Mayo et al. 1990])	No (atom types are element types [van Duin et al. 2001])
Accessible time scales	Several ns on single CPU easily reached	Typically < ns for single CPU simulation

Table 2. Distinctions between reactive and nonreactive force fields.

used in experiment [Balsera et al. 1997; Rief et al. 1997; Lu et al. 1998; Arteca 2003]. We leave a more systematic investigation of these phenomena to future work.

Even though many biological processes are controlled by the interplay of weak interactions, there exist several examples of material deformation during which fracture of covalent bonds is critical. Specific examples include fracture of bone during which cross-links in the collagen phase or cross-links between hydroxyapatite and collagen may rupture, fracture of cross-links in biopolymers such as highly cross-linked or aged collagen, or large deformation of elastin networks that contain a large number of disulfide bonds [Landis et al. 2002; Nalla et al. 2003; 2005; Ritchie et al. 2004; Gupta et al. 2004; 2005; Peterlik et al. 2006]. Even though the applied stress may be moderate, stress concentrations at, for example, locations of cracks or flaws in the material [Anderson 1991] may induce forces large enough to lead to rupture of covalent bonds.

3.3. *Molecular simulation procedure.*

The atomic structure of many proteins has been determined according to X-ray diffraction data obtained by experiment. Atomistic structures can often be taken directly from the Protein Data Bank (PDB) and provide reasonably starting structures for atomistic simulation. Typically, charges of atoms and the locations of hydrogen atoms are not included. To add hydrogen atoms and water molecules, and to assign atomic partial charges, we use the *psfgen* program that is part of the NAMD molecular simulation package. The charges of each atom are assigned according to the CHARMM rules, while hydrogen atoms and the protonation states of individual amino acids are assigned according to pH 7. The CHARMM input files (structure and topology files) are then used to perform NAMD calculations. In addition to crystallographic water, we add a skin of water of a few Å surrounding each tropocollagen molecule.

We note that even though the entire molecule is embedded in water at the beginning of the simulation, under large deformation of the molecule, water begins to cluster around certain regions in the protein, leaving other regions exposed to vacuum. Since hydrophobic side chains are typically buried in aqueous environment, charged side chains tend to come together in vacuum, which can induce additional unfolding paths. The effect of these processes on the unfolding dynamics are hard to predict and represent a potential source of error, and need to be considered with caution. A possible remedy to this situation could be using much larger water skins, for example, periodic boxes whose space is completely filled with water molecules. Such models are, however, prohibited with today's computational resources, in particular when reactive force fields are employed.

For reactive calculation, no information about the chemical bonding is necessary, so that the only input parameters are chemical element atom types (C, N, S, H, . . .) and the coordinates of all atoms in the system (see also comparison between reactive and nonreactive force fields in Table 2).

Before finite temperature dynamical calculations are performed, we carry out an energy minimization for several thousand steps, making sure that convergence is achieved, thus relieving any potential overlap in vdW interactions after adding hydrogen atoms. In the second step, we anneal the molecule after heating it up to a temperature $T = 300$ K. The heat up rate is $\Delta T = 25$ K every 25 steps, and we keep the temperature fixed after the final temperature $T = 300$ K is achieved (then we apply a temperature control in an NVT ensemble [Leach 2001]). We also ensure that the energy remains constant after the annealing procedure.

For stretching calculations, we typically fix one end of the molecules and apply a force at the other end of the molecule using a method called Steered Molecular Dynamics (SMD). SMD is based on the concept of adding restraint force to groups of atoms by extending the Hamiltonian by an additional restraint potential of the form $k_{SMD}(r - r_\lambda)^2/2$. The SMD approach is implemented in both ReaxFF and NAMD [Nelson et al. 1996; Phillips et al. 2005]. Unless indicated otherwise, we use an SMD scheme with spring constant $k_{SMD} = 10\,\text{kcal/mol/Å}^2$.

While the force is increased, we investigate the response of the molecule due to the applied loading. Typically, we obtain force-versus-displacement data, which is then used to extract the bending stiffness, using continuum mechanical concepts by drawing analogies between the molecular level and continuum mechanical theories [Buehler 2006a; 2006d].

We note that in this paper we typically present simulations only at a single deformation rate, and – due to the higher computational expenses of reactive force fields – at rather short time scales at fractions of nanoseconds. Future studies should focus on a more careful investigation of the rate dependence of the results. This is particularly critical in order to compare the results with experiments. Results of different loading rates could be extrapolated to rates comparable to experimental efforts.

4. Computational results

Our focus is to investigate the relationship between protein structure and its mechanical properties. Our strategy is to focus on three protein structures with increasing complexity, involving alpha helices, random coils and beta sheets.

These model systems include alpha-helical (AH) structure, a small protein α-conotoxin PnIB from *conus pennaceus*, and lysozyme, a well-studied enzyme that catalyzes breaking of glycosidic bonds. Table 3 summarizes the most significant structural features of these three proteins. Figure 5 shows an overview of the three protein structures considered. This figure illustrates the increase in complexity and distinct structural features in the three proteins studied.

	1GK7 (alpha helix)	1AKG (conotoxin PnIB)	194L (lysozyme)
Alpha helices	Yes	Yes	Yes
Random coils	—	Yes	Yes
Beta-sheets	—	—	Yes
Disulfide cross-links	—	Yes	Yes
Function: physiological or general	Structural protein	Antagonist	Enzyme

Table 3. Summary of the most significant structural features of the three proteins in this article.

4.1. *Stretching and bending of a single alpha-helical protein.* First, we study the elastic, plastic and fracture behavior of a small AH motif. AH motifs constitute the molecular building blocks of coiled structural motifs. We consider a recently crystallized AH protein with PDB ID 1GK7 [Strelkov et al. 2002]. This protein is part of the human vimentin coil 1A fragment with 38 residues that belongs to
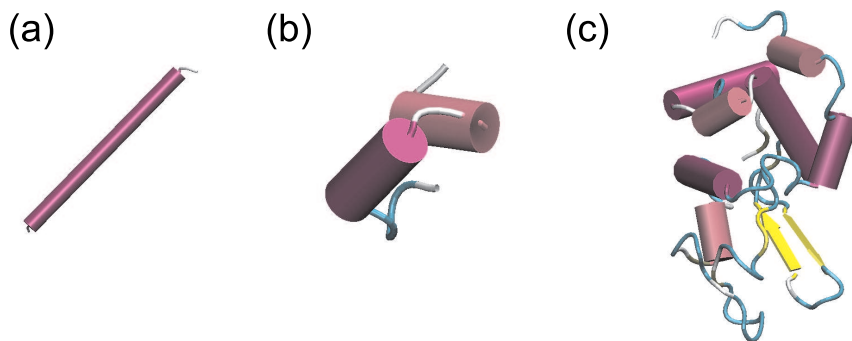
Figure 5. Overview over the three protein structures studied in this article. Subplot (a) shows a simple alpha-helical structure (PDB ID 1GK7) [Strelkov et al. 2002], subplot (b) depicts a small protein α-conotoxin PnIB from *conus pennaceus* (PDB ID 1AKG) [Fainzilber et al. 1994] and subplot (c) depicts the protein lysozyme, an enzyme that catalyzes cleaving glycosidic bonds (PDB ID 194L) [Vaney et al. 1996]. This sequence of structures is chosen since it represents a systematic increase in complexity of the folded arrangement and existence of structural motifs (alpha helices are drawn as cylinders, random coils as thin curved lines, and beta shears are drawn as yellow arrows). The pure alpha helical structure represents a class of coiled-coil proteins (subplot (a)). Additional polypeptide motifs and covalent cross-links are added in the structure of 1AKG (subplot (b)). Lysozyme (subplot (c)) represents the most complex structure, which features beta sheets, alpha helices, random coils and covalent cross-links.

the intermediate filament family of proteins. The intermediate filaments together with microtubules and actin microfilaments constitute the cytoskeleton, the structural backbone of eukaryotic cells. Vimentin filaments are an important structural feature of eukaryotic cell, serving as cross-links between microtubules and actin filaments, and thus these filaments are subject to severe mechanical loading under physiological conditions of cells, serving as its *safety belt* [Alberts et al. 2002]. The sequence of the protein considered here is GLY SER ASN GLU LYS VAL GLU LEU GLN GLU LEU ASN ASP ARG PHE ALA ASN TYR ILE ASP LYS VAL ARG PHE LEU GLU GLN GLN ASN LYS ILE LEU LEU ALA GLU LEU GLU GLN LEU.

Earlier studies [Arteca and Li 2004; Li and Arteca 2005; Ortiz et al. 2005; Contera et al. 2005] of the response of AH structures to mechanical loading were carried out using nonreactive force fields, with force levels below the critical load that leads to molecular fracture. Some studies were carried out using rather short sequences of AH structures composed of 10 and 20 residues [Li and Arteca 2005], using a soft steered MD approach (SSMD) that allows for a more gentle force application to molecules by redistributing internal energy during pulling [Arteca 2003]. Other molecular simulations were carried out using a coarse grained representation of the helical structure, using a Lennard–Jones model [Cieplak et al. 2002]. Here we focus particularly on the large-deformation regime of such AH structures, including molecular rupture, using a longer sequence of AH structure composed of 39 residues and having a total length of 58 Å.

Our computational experiments are designed to resemble an AFM or optical tweezers experiment, where a continuously increasing force is applied at the ends of the molecule, while the end-to-end distance is measured. This results in force-displacement data that provides information about the mechanical properties. A structural analysis of the protein during deformation enables insight into the atomistic and molecular deformation mechanisms. This computational experiment is realized by using a steered MD scheme, with a thin layer of water surrounding the molecule to represent realistic solvent conditions. The forces are applied at the C_α atom at the first (N-terminus) and last (C-terminus) residue of the AH structure.

Figure 6 depicts the entire force-stretching response of a single AH protein, for a loading rate of 0.0005 Å/iteration. The loading case is depicted in Figure 6a. The results shown in Figure 6b indicate that there are three different deformation regimes, resulting in a strongly nonlinear mechanical response. The first regime is characterized by a relatively strong increase of the force with respect to strain, up to approximately 19% strain or stretch or 10 Å (point A in Figure 6). This regime corresponds to the initial stretching of the intact helical structure.

This initial regime is followed by a plateau-like behavior (between points A and B), during which we observe uncoiling of the helical arrangements, leading to an almost straight polypeptide chain at approximately 129% strain or 75 Å (point B in Figure 6), while the applied force increases only slightly with increased end-to-end distance. This unfolding regime is primarily controlled by continuous *breaking* of H-bonds that provide structural integrity of the AH motif.

Once the helical structure is lost, covalent bonds in the protein backbone are being stretched, leading to a sharp increase of the force with respect to strain, at point C. The molecule fractures at approximately 163% strain or approximately 95 Å, at a maximum force close to 7.8 nN.

Snapshots of the entire deformation process are shown in Figure 7, using a cartoon representation that visualizes the AH structure. The visual representation clearly illustrates breakdown of the AH structure, leading to a flat polypeptide structure at large deformation. Eventually, molecular rupture occurs at large values of force.

The force-stretch information can be used to extract a local, in terms of strain or stretch, Young's modulus $E(L)$, which is given by

$$E(L) = \frac{L_0}{A_C} \frac{\partial F(L)}{\partial L}. \tag{7}$$

In Equation (7), the parameter L_0 is the initial, undeformed length of the AH molecule, where $L_0 \approx 58$ Å. Note that Young's modulus is independent of the length L_0 of the molecule. The definition in Equation (7) is a consequence of the fact that the stretching force is expressed as a function of stretch d rather than strain ($\sigma = E\varepsilon$ and note that $\varepsilon = (L - L_0)/L_0$).

With $A_C = \pi R_C^2$ and $R_C \approx 7.27$ Å, where the value of R_C is based on an estimate of the cross-sectional distance between the ends of AH side chains across the diameter of the molecule, Young's modulus for deformations up to approximately 18% is approximately 5 GPa. The secant modulus up to point B is approximately 0.8 GPa. The tangent modulus under large deformation (beyond point B up to fracture at point C) is approximately 9.2 GPa. If an additional layer of water molecules surrounding the AH molecule is considered, the effective cross-sectional area increases (the two method to approximate the cross-sectional area of the molecule are shown in the inlay of Figure 7). In this case, $R_C \approx 10.27$ Å, and Young's modulus for small deformation is given by 2.5 GPa, the secant modulus is approximately
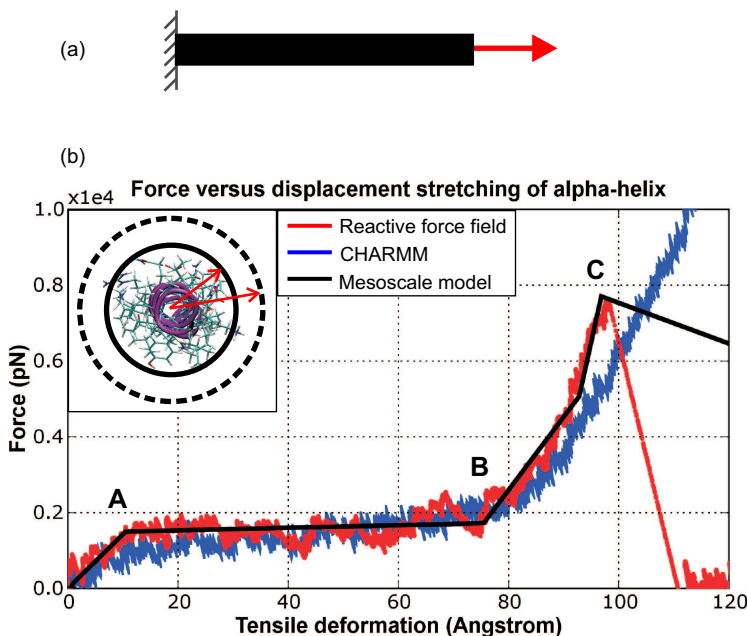
(a)

(b)

Figure 6. Stretching of a single alpha-helix with approximately 5.8 nm length, comparing reactive and nonreactive force fields. Both models yield similar force-displacement curves for small deformation. For large deformation, however, the two models disagree. The reactive model predicts a fracture strength of approximately 7,800 pN, at approximately 172% strain. The deformation behavior at strains below 75 Å is characterized by a homogeneous uncoiling of the helical structure. The initial uncoiling regime is accompanied by a slowly increasing force that approaches approximately 2,000 pN. Once the entire molecule has lost the helical structure, the forces increase significantly until the molecule fractures. The inlay depicts a view into the molecular axis of the AH protein, illustrating our method to estimate the cross-sectional area of the molecule. The dashed circle corresponds to the size estimate considering a thin layer of water molecules as part of the molecular cross-sectional area. The pulling simulation was carried out over a time scale of 6.8×10^{-11} seconds.

0.4 GPa, and the large-deformation tangent modulus is 4.6 GPa. The fracture tensile stress of the AH protein is approximately 4.8 GPa. If only the helical core is considered, $R_C \approx 3.64$ Å, and Young's modulus for small deformation is given by 20 GPa, the secant modulus is approximately 1.6 GPa, and the large-deformation tangent modulus is 18.4 GPa.

Figure 7 also includes the results obtained using a nonreactive (CHARMM) force field denoted by a red curve. The prediction by CHARMM is similar to the result obtained using the reactive force field, albeit forces are slightly smaller than those obtained in the reactive calculation. However, the two descriptions disagree for large deflections from the initial length, with the most significant difference being the continuous increase in force even for forces that approach 10 nN and more. This result suggests
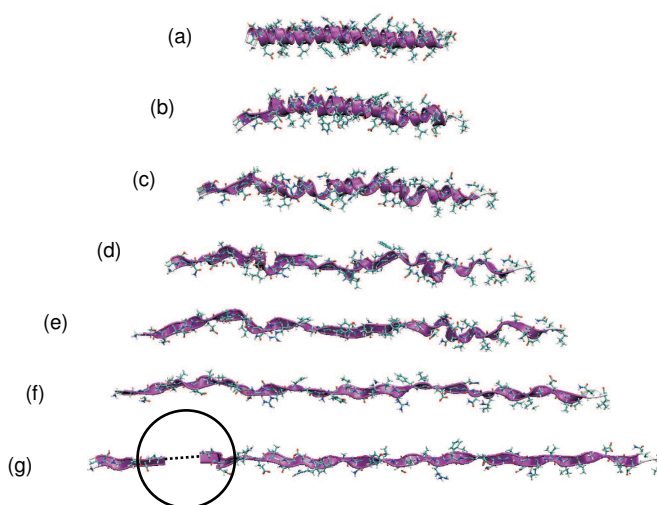
Figure 7. Unfolding of an alpha-helix, due to mechanical stimulation along the axial direction. The initial deformation behavior is characterized by homogeneous stretching of the alpha-helical structure (up to point A in Figure 4), followed by continuous uncoiling of the helical structure (between points A and B in Figure 4)). Once the entire molecule has lost the helical structure (see point B in Figure 4), the forces increase significantly until the molecule fractures (see point C in Figure 4). A snapshot right after fracture has occurred is shown in subplot (g). To make the alpha-helical structure (and its disappearance) better visible, we render it using VMD's *Ribbons* method.

that such models are incapable of correctly describing the molecular fracture mechanics of biological molecules.

The total simulation time for the stretching calculation is 6.8×10^{-11} seconds. The resulting high strain rates may introduce artificial effects, as the molecule may not be in equilibrium at all times during deformation. However, our results are nevertheless in qualitative agreement with earlier simulation work using nonreactive force fields, as well as experimental efforts [Schwaiger et al. 2002; Akkermans and Warren 2004; Fudge and Gosline 2004; Li and Arteca 2005; Root et al. 2006; Kiss et al. 2006].

Figure 8 shows the results of a bending calculation. Subplot (a) depicts the loading case of a double supported three-point bending test. Figure 8b shows the force versus bending displacement, obtained using the CHARMM force field. The data shown in Figure 8b is collected in a regime where the molecule undergoes pure bending.

The bending simulation is carried out using the identical SMD technique, but using a different spring constant and a much lower strain rate. The loading rate used for this simulation is 0.000002 Å/iteration, with $k_{SMD} = 0.01$ kcal/mol/Å^2. These choices reflect the fact that the molecule is much softer under bending than under tension.

This study leads to force information about F_{appl} versus bending displacement d. This information can be used to estimate the bending stiffness, for the case of a double supported beam with a bending
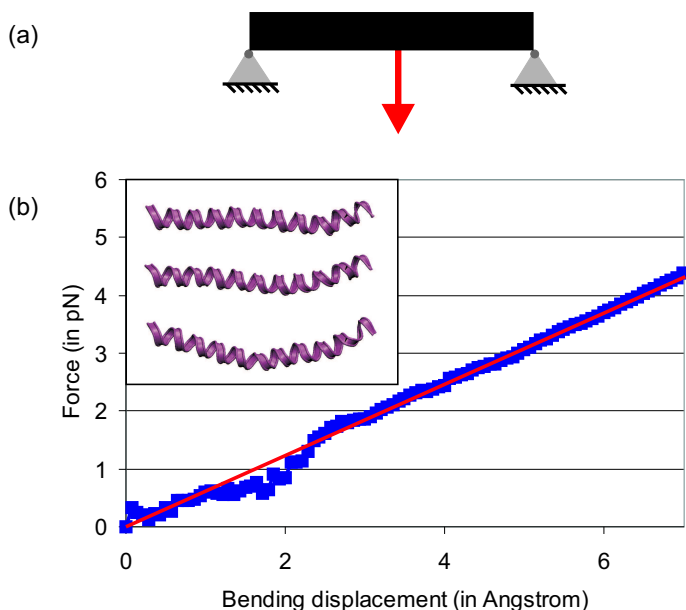
Figure 8. Bending of a single alpha-helix with approximately $L \approx 5.8$ nm length. Subplot (a) shows the loading case, and subplot (b) depicts the force increase versus the bending displacement increase. The red line is a linear fit to the data, whose slope is used to determine the bending stiffness according to Equation (8) (the ratio F/d corresponds to the slope measured from atomistic simulation results). The inlay shows a few snapshot as the AH protein undergoes bending deformation. The initial force-displacement regime is slightly noisier. We believe that this is due to thermal fluctuations that disappear once the molecule is under increased tensile load due to the bending.

force applied in the center of the molecule:

$$EI = \frac{F_{\text{appl}}L^3}{48d}. \tag{8}$$

Assuming that the AH protein can be represented by a continuum beam, the resulting bending stiffness of this molecule is 2.27×10^{-29} Nm^2. This atomistic result is a key input parameter for the development of the mesoscale representation (see Section 5.1). Using the bending stiffness, we estimate the persistence length of the AH protein (at $T = 300$ K) to approximately be 5.5 nm. This result agrees somewhat with experimental studies that found a persistence length of a AH structure on the order of 1-2 nm [Papadopoulos et al. 2006]. The reason for this disagreement could, for example, be rate effects.

We note that the observation that the AH structure vanishes under large deformation is not contracting Astbury's X-ray diffraction based observations in the 1930s that the structure of such proteins changes significantly under stretch [Astbury and Street 1932].

4.2. *Unfolding dynamics of globular, cross-linked proteins: 1AKG.* The protein studied in the previous example has a relatively simple structure, consisting of only one alpha helix. Now we consider a protein
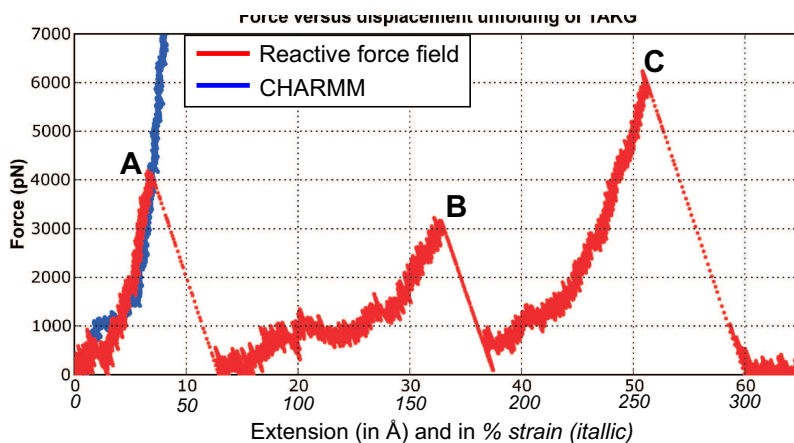
Figure 9. Mechanically stimulated unfolding of a small protein PnIB 1AKG, modeling results obtained using CHARMM (red curve) and the reactive force field ReaxFF (blue curve). We apply a slowly increasing force at the N- and C-terminus of the polypeptide. It is apparent that although the ReaxFF and CHARMM descriptions agree for small deformation (below $\approx 7\,\text{Å}$), they disagree strongly for larger deformation. The difference can be explained based on the fact that the CHARMM potential is incapable of describing breaking of the disulfide bonds. The points A–C characterize different stages of deformation; A and B correspond to breaking of the two disulfide bonds, with lower force for the second one. Point C characterizes breaking of the backbone chain. The strength of the backbone chain is slightly larger than the disulfide bonds. The characteristic saw tooth shape is reminiscent of experimental results. However, experiment is somewhat limited with respect to the force levels that can be reached, which makes it difficult to study breaking of strong covalent bonds. The pulling simulation was carried out over a time scale of 2×10^{-11} seconds.

that has a slightly more complex structure. We consider a small protein α-conotoxin PnIB from a cone-shell species *conus pennaceus* that appears primarily in the Indian Ocean. This protein is an extracellular protein that serves as acetylcholine receptor antagonist. It consists of 16 residues that resemble two distinct building blocks, alpha helices and random coils, which are connected by disulfide cross-links formed at the CYS residue. This protein has recently been crystallized and deposited in the PDB with ID 1AKG [Fainzilber et al. 1994]; the structure is shown in Figure 5b.

To simulate the mechanical stimulated unfolding of the protein, we apply a slowly increasing force at the N-terminus (beginning of the polypeptide chain, in this example the GLY1 residue) and C-terminus of the protein (end of the polypeptide chain, in this example the CYS16 residue).

Figure 9 shows the force-displacement curve during mechanical unfolding of this protein, comparing the predictions made by the CHARMM force field and ReaxFF. Individual rupture events of covalent bonds can be associated with peaks in the force-displacement curve, leading to a characteristic sawtooth shape. The first rupture occurs at approximately 4 nN, and is due to breaking of the disulfide bond between CYS3 and CYS16. The second rupture features a lower rupture force of 3 nN, and is due to
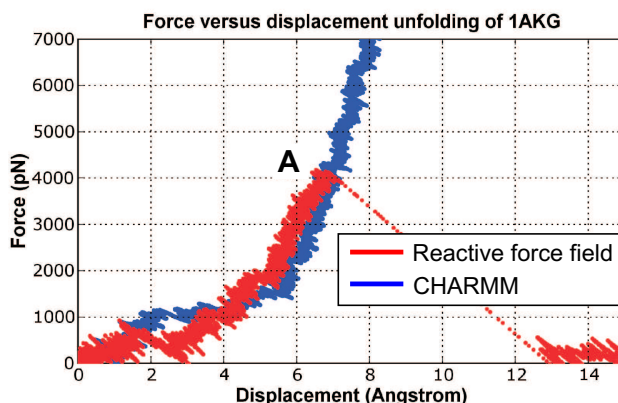
Figure 10. Mechanically stimulated unfolding of a small protein PnIB 1AKG. This plot shows a zoom into smaller displacements (same data as shown in the previous figure). It is apparent that although the ReaxFF and CHARMM descriptions agree for small deformation (for displacements below $\approx 17\,\text{Å}$), they disagree strongly for larger deformation, until the first covalent disulfide bond ruptures (*A*).

breaking of the disulfide bond between CYS2 and CYS8. The final rupture of the protein backbone peptide bond occurs at approximately 6 nN, at a larger force than any of the previous breaking events. These results suggest that the disulfide cross-link is weaker than the peptide bond. The total simulation time for the results shown in Figure 9 is 2×10^{-11} seconds, which is only a fraction of a nanosecond. As in the previous case, the resulting high strain rates may introduce artificial effects, as the molecule is not in equilibrium at all times during deformation.

Figure 10 shows a more detailed view of this system, focusing in on the small displacements and the first bond rupture event. It is apparent that the reactive force field model and the CHARMM description agree reasonably well for small deformation. The CHARMM model, however, predicts a very different unfolding sequence with forces raising to unrealistically large values. This is an immediate consequence of the nonreactive character of this force field as the forces become arbitrarily large for large bond stretch (see Equation (4)).

Figure 11 depicts several snapshots of the atomic structure as the protein undergoes deformation, for the studies using the reactive force field. Figure 12 depicts a similar sequence of snapshots for the nonreactive CHARMM force field. The sequence of snapshots shown in this figure shows that the covalent cross-links never break, even at large deformation, leading to incorrect prediction of the folding dynamics under large stretch.

From the results shown in Figures 9–12, it is apparent that classical CHARMM-like descriptions are not capable of addressing the various chemical events properly, and lead to incorrect behavior at large deformation. Our results clearly demonstrate the difference between the reactive potential and a classical CHARMM potential in a study of unfolding of a small protein PDB ID 1AKG.

4.3. *Unfolding dynamics of a globular, cross-linked protein: lysozyme.* Now we focus on the unfolding mechanics of a more complex protein structure, the enzyme lysozyme. Figure 13 depicts the three-dimensional structure of lysozyme, without (subplot (a)) and with (subplot (b)) the substrate attached.
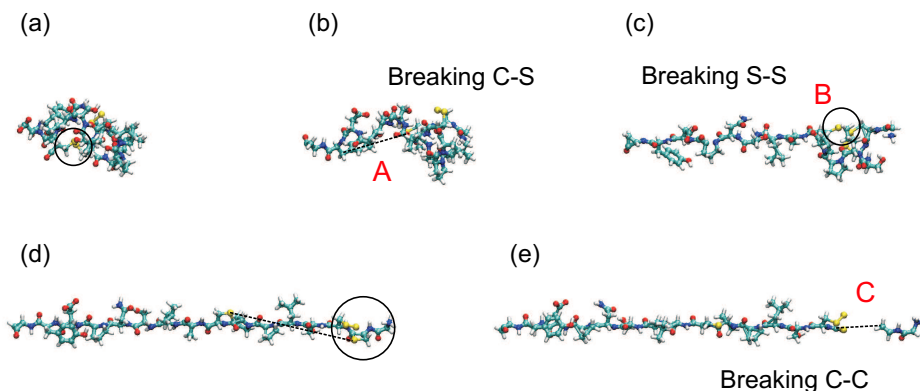
Figure 11. Snapshots of mechanically stimulated unfolding of a small protein, simulation carried out using the reactive potential. The plot depicts results obtained using the reactive force field ReaxFF The points A–C represent rupture events and correspond to those shown in Figure 9.
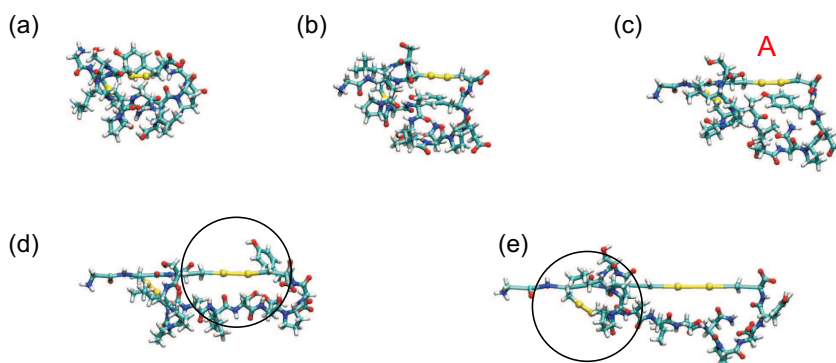

Figure 12. Snapshots of mechanically stimulated unfolding of a small protein. Simulation is carried out using the CHARMM potential. It is evident that the disulfide cross-links never rupture (see the disulfide bond indicated with *A* in subplot (c)).

As can be verified in this figure, the substrate binds to the active site of the enzyme. The structure shown in subplot (a) is based on PDB ID 194L [Vaney et al. 1996] (hen egg white lysozyme), and the structure on the right is based on PDB ID 1LJN [Harata and Kanai 2002] (turkey egg lysozyme with a N-acetyl-D-glucosamine substrate). The basis for all calculations discussed subsequently is the crystal structure of lysozyme 194L [Vaney et al. 1996].

Lysozyme represents a more complex structure than those studied in Sections 4.1 and 4.2. In addition to alpha helices and random coils, this protein also contains one beta sheet. Of particular interest to us are the four disulfide cross-links inside the protein, and its participation and influence on the unfolding dynamics of the protein. Lysozyme consists of 129 residues. Disulfide cross-links are present between CYS6 and CYS127, CYS30 and CYS115, CYS64 and CYS80, as well as between CYS76 and CYS94.
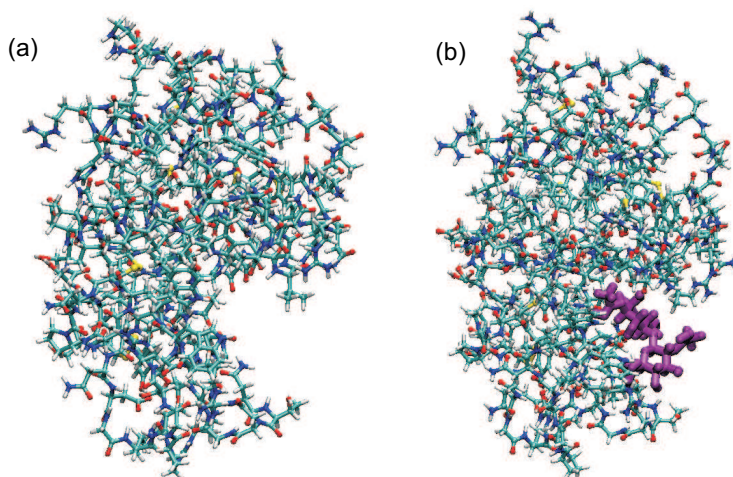
Figure 13. This figure depicts the three-dimensional structure of the enzyme lysozyme, without (subplot (a)) and with (subplot (b)) the substrate attached, without showing any water molecules. As can be verified in this figure, the substrate binds to the active site of the enzyme. The structure shown in subplot (a) is based on PDB ID 194L [Vaney et al. 1996] (hen egg white lysozyme), and the structure on the right is based on PDB ID 1LJN [Harata and Kanai 2002] (turkey egg lysozyme with a N-acetyl-D-glucosamine substrate).

Figure 14 depicts the locations of disulfide bonds in lysozyme, where subplot (a) shows a view of the entire protein, and subplot (b) depicts a detailed view on one of the disulfide bonds.

A particular interest of this study is to shed light on the unfolding dynamics such as the sequence of specific unfolding events during deformation [Cieplak et al. 2002].

Figure 15 depicts the force-displacement curve during mechanically stimulated unfolding of lysozyme. The unfolding is induced by applying a slowly increasing force between the N- and C-terminus of the protein, using a steered MD scheme. The force is applied between residues LYS1 and LEU239. The force-displacement scheme indicates two maxima (points A and D) at which the forces approach several nN, corresponding to two cross-link fracture events. The forces during uncoiling of the folded structure is characterized by force levels of approximately 1 nN and 2 nN. We observe a small maximum at approximately 120 Å separation, where the forces reach 2.2 nN.

The initial increase in stretching force is due to breaking of the first covalent disulfide bond (A) at approximately 4.5 nN, followed by its rupture and a regime in which the forces approach an approximately constant value of 1.2 nN (B). This regime is followed by a slight increase to 2.5 nN (C). The second disulfide cross-link breaks at point D, leading to an increase in force similar to that at point A. The first peak (A) is due to breaking of the cross-link between CYS6 and CYS127, and the second peak is due to breaking of the cross-link between CYS30 and CYS115. This result can be explained since these residues are in closest geometric vicinity to the ends of the polypeptide chain, at which the forces are being applied.
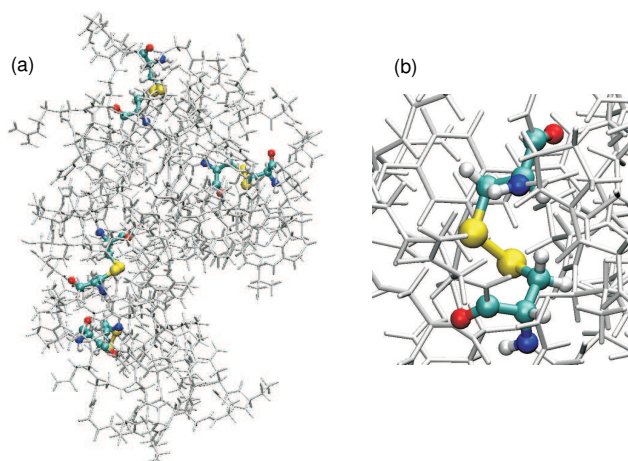
Figure 14. Locations of disulfide bonds in lysozyme. Subplot (a) shows a view of the entire protein, highlighting the four disulfide cross-links (in color, and using larger spheres). The substrate binding site can be seen in the lower right part of the protein (similar angular view as that used in Figure 13). Subplot (b) depicts a detailed view on one of the disulfide bonds, including all atoms in the neighboring CYS residues. The yellow atoms are the two sulfur atoms that connect two CYS residues.
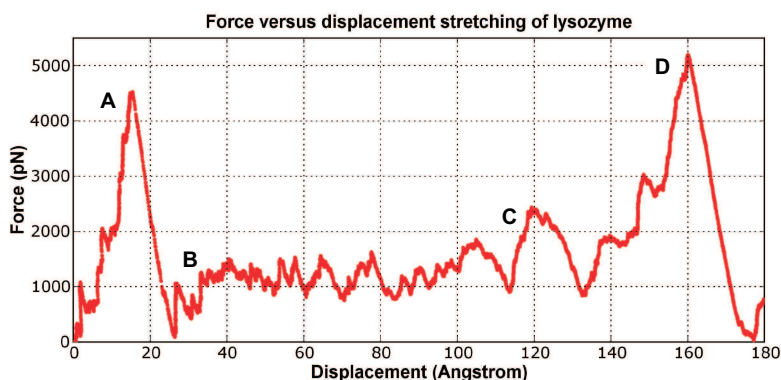

Figure 15. Mechanically stimulated unfolding of lysozyme, force versus displacement. The initial increase in stretching force is due to breaking of the first covalent disulfide bond (A) at approximately 4,500 pN, followed by its rupture and a regime in which the forces approach an approximately constant value of 1,200 pN (B). This regime is followed by a slight increase to 2,500 pN (C). The second disulfide cross-link breaks at point D, leading to an increase in force similar to that at point A, but with slightly higher maximum rupture force of approximately 5,200 pN. The first peak (A) is due to breaking of the cross-link between CYS6 and CYS127, and the second peak (D) is due to breaking of the cross-link between residues CYS30 and CYS115. The pulling simulation was carried out over a time scale of 4.5×10^{-11} seconds.

Figure 16 depicts snapshots of the different stages during unfolding dynamics of lysozyme. In this simulation, the distance between the ends of the protein (C-alpha atom of the terminal residues) is continuously increased by applying a continuously increasing force. The direction of the force is given by the instantaneous distance vector of the two ends.

The molecular simulation enables us to perform a detailed analysis of the deformation mechanics. We observe that the N-terminus of the protein unfolds first. Once the cross-link that is attached at residues CYS6 and CYS 127 is reached, the force increases significantly due to the resistance of the covalent cross-link (this corresponds to the first peak (A)).

After rupture of the first cross-link, the C-terminus of the protein starts to unfold, and additional unfolding at the N-terminus ceases. This is because the cross-link between residues CYS30 and CYS115 pins this end leading to shutdown of unfolding. Once the C-terminus is pulled so far that the cross-link between CYS30 and CYS115 is reached, the forces increase significantly, leading to uncoiling of the alpha-helical structure that is attached close to the C-terminus. This is confirmed in Figure 16(f-h) by considering the increase of the length of the cylinder that represents the AH structure.

Our simulations reveal that unfolding of lysozyme occurs sequentially, and that peaks that appear in the force-stretch curve are due to sequential breaking of cross-links. Unfolding of this protein is a sequential rather than homogeneous process, as observed in several earlier studies. Smaller force peaks appear when unfolding of domains is observed (see Figure 15, point C).

The key predictions of our reactive simulations are:

- the N-terminus unfolds first, while the C-terminus remains in its folded configuration until the fist disulfide cross-link ruptures;
- large peaks occur at specific end-to-end distances, whenever a covalent cross-link is reached.

5. Discussion, conclusion and outlook

The results reported in this article illustrate how reactive force fields can be successfully used to describe the stretching and unfolding dynamics of small proteins, including molecular fracture. Using this computational technique, we have studied the unfolding and stretching dynamics of three proteins with increasing complexity.

Our model provides a reactive treatment of the unfolding problem, considering not only rupture of vdW and H-bonds, but also rupture of covalent bonds. We have demonstrated that in proteins in which disulfide cross links are present, a reactive treatment is essential, and that CHARMM type potentials do not allow a description of the unfolding processes. Including the possibility of bond rupture is essential to describe the unfolding processes correctly, and neglecting such effects may lead to incorrect force-displacement curves as shown in Figures 9 and 10. A nonreactive description may lead to unrealistically large forces and an incomplete unfolding of the protein, even though extremely large forces are applied that exceed several nN.

Our studies show a distinctions in the force-displacement unfolding curves between fibrillar and globular proteins. Whereas stretching a single AH structure suggests a continuous increase of the stretching force with stretching distance followed by molecular rupture, globular proteins feature a sawtooth like force-displacement curve with extremely long unfolding paths, featuring several local maxima that correspond to domain unfolding or cross-link rupture.
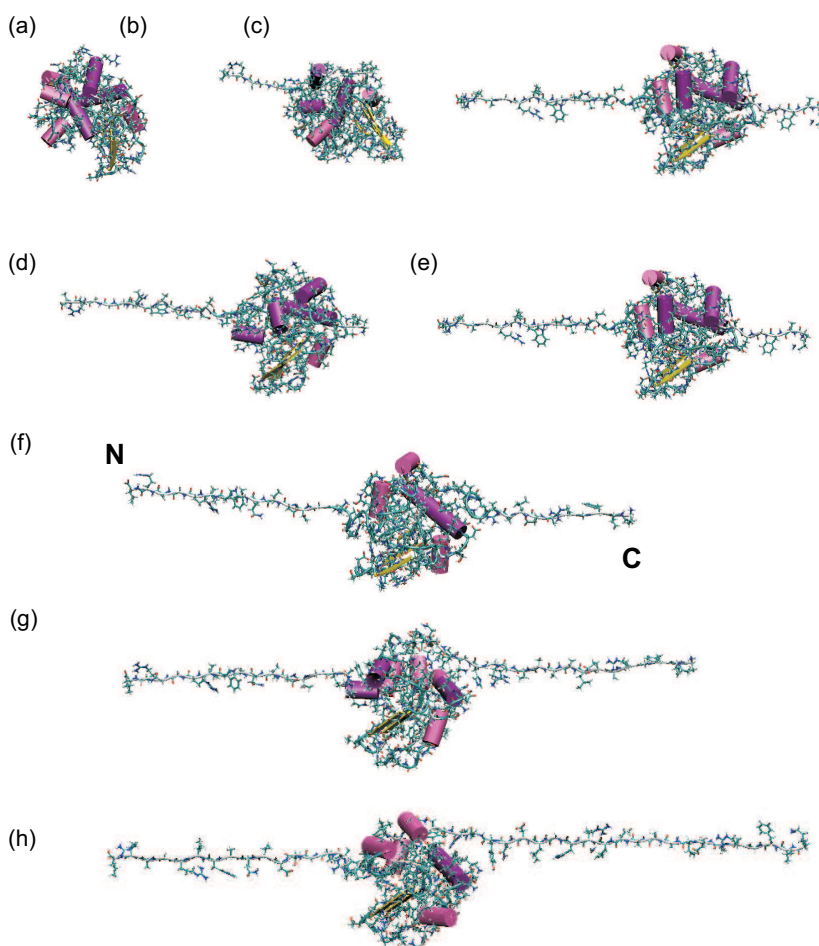
Figure 16. Different stages during unfolding dynamics of lysozyme. In this simulation, the distance between the ends of the protein (C_α atom of the terminal residues) is continuously increased by applying a continuously increasing force. The direction of the force is given by the instantaneous distance vector of the two ends. The molecular simulation enables us to perform a detailed analysis of the deformation mechanics. We observe that one end of the protein unfolds first. Once the cross-link that is attached at residue CYS6 is reached, the force increases significantly due to the resistance of the covalent cross-link. After this cross-link is broken and the displacement further increases, the second cross-link is reached (close to residue CYS30) the second end of the protein starts to unfold as well, and the further displacement of the first one ceases. Unfolding of the other end proceeds until another cross-link is reached, which is then followed by significant stretch of the alpha-helical structure that is attached close to the other end. Subplot (h) depicts the configuration slightly before the second covalent cross-link is broken.

Our analysis of unfolding of lysozyme suggests that the activity of the enzyme may be severely impeded if a stretching force is applied at the ends of the polypeptide strand. This is because the binding pocket of the enzyme undergoes a shape change that may make it difficult or impossible to bind the substrate and thus significantly hinder the enzyme's main function to stabilize the transition states of the catalyzed chemical reactions. Our MD modeling further provided an analysis of the temporal sequence of unfolding of different domains, of both lysozyme and the small protein with PDB ID 1AKG.

For both globular proteins studied here (1AKG and 194L), we observe that the force increases rapidly even for modest deformation (see Figures 9 and 15). This may be due a specific molecular design that prevents unfolding of the protein, thus providing additional stability. In both cases, this design objective is realized by placing a covalent cross-link close to the end and the beginning of the polypeptide chain, respectively. The AH protein resembles a different behavior, showing a continuous increase of stress with respect to increasing strain, leading to eventual rupture when the force approaches its maximum. Similar behavior has been reported about the deformation mechanics of individual tropocollagen molecules.

Our reactive studies provide estimates for the fracture strength of an AH protein, including an estimate for Young's modulus and the bending stiffness. Our molecular analyses suggest a persistence length at room temperature of approximately 5.5 nm.

Extreme deformation of structural and functional proteins can be relevant under physiological conditions. Cells, for example, can exert large forces that can exceed several nN. As our studies show, such large force levels can lead to breaking of covalent bonds or other severe deformation with nonlinear elasticity. Thus, sound understanding of the nanomechanical responses of materials may be critical to shed light on associated biological or bioengineering processes. We note that even though cells can exert such large forces, it does not necessarily mean that individual cells sense such large forces. Other critical applications are in fracture of tissue such as bone or skin. Once covalent cross-links are present, covalent bond breaking becomes important under fracture conditions.

Although full atomistic studies provide a fundamental view into the atomistic mechanisms during deformation of matter, it has intrinsic limitations, the most significant of which is related to the high deformation and strain rates, due to the limitations of time scales to several nanoseconds. This may lead to overestimation of the unfolding forces; it has been discussed extensively in earlier work [Balsera et al. 1997; Rief et al. 1997; Lu et al. 1998; Arteca 2003].

5.1. *Development of a mesoscale model: entropic elasticity.* To overcome the time- and length-scale limitations, the atomistic simulation results can be used to develop a mesoscale model of the protein structure that features less degrees of freedom, but is capable of capturing the essential physical properties of the molecules. Here we demonstrate how such a model can be applied to the AH protein.

To achieve this, the entire sequence of amino acids that makes up the AH structure is replaced by a collection of beads (see Figure 17a). The beads interact according to a intermolecular multibody potential. The parameters of this bead model are determined from the full atomistic simulation results. Appendix A describes the detailed mathematic formulation of this model. For example, the tensile stiffness parameters are obtained from the stretching calculations. The bending stiffness is obtained from a bending calculation of the single molecule, similar as described in future publications. We leave details about how the parameters are obtained to a future work, and focus on application of this model.
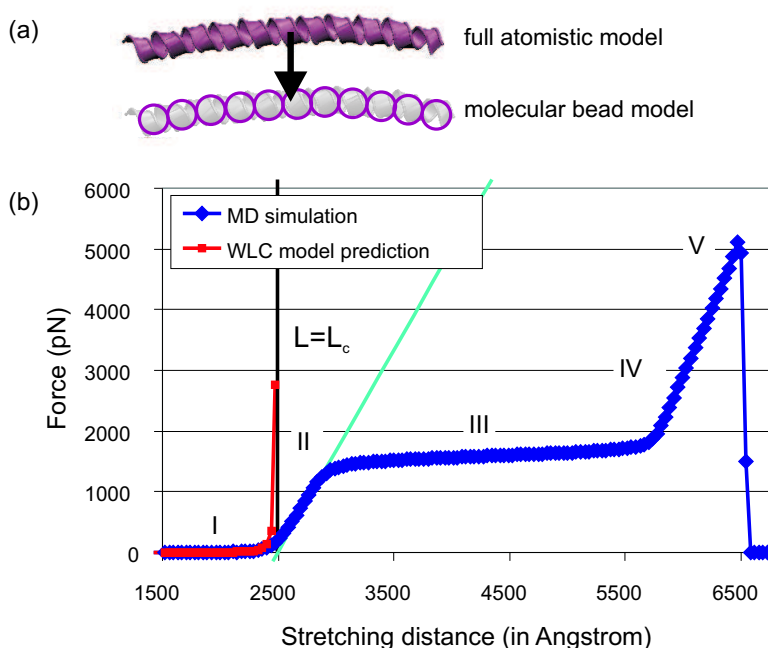
Figure 17. Stretching experiment of an alpha helix molecule, with length $L = 249.5$ nm, at 300 K, carried out using the mesoscale model. Subplot (a) shows a schematic of the coarse-graining procedure, changing the full atomistic representation arriving at the molecular bead model. Subplot (b) shows the entire force-displacement curve including regime I, entropic elasticity, regime II, energetic elasticity, regime III uncoiling of the AH structure, IV stretching of the backbone bonds, and regime V, molecular fracture. Whereas the WLC model predicts divergence of forces at the contour length, molecular modeling predicts a smooth transition from entropic to energetic elasticity. The plateau regime reached beyond approximately $3{,}000$ Å is due to unfolding of alpha helical structures. The long stretching distances are a consequence of the fact that the molecule is extremely long. Conformations at different stages are shown in Figure 20.

Figure 17b depicts the force-versus stretching curve for this case, showing five regimes: (I) entropic elasticity, (II) energetic elasticity due to homogeneous stretching of the AH structure, (III) uncoiling of the AH structure, (IV) stretching of backbone bonds, followed by (V) molecular rupture. Figure 18 shows a detailed view of the entropic regime, comparing the WLC model given in Equation (2) with the molecular modeling results. Figure 19 depicts several snapshots as the molecule undergoes tensile deformation.

These results show that the mesoscale representation of the molecule is capable of modeling entropic contributions to elasticity, at large time scales reaching several microseconds. Our results confirm the hypothesis that the molecular mechanics is controlled by a transition from entropic to energetic elasticity.

5.2. *Hybrid reactive-nonreactive models.* Reactive force fields can be computationally very expensive. In some cases, the region of atoms where a reactive treatment is actually needed is quite localized. A
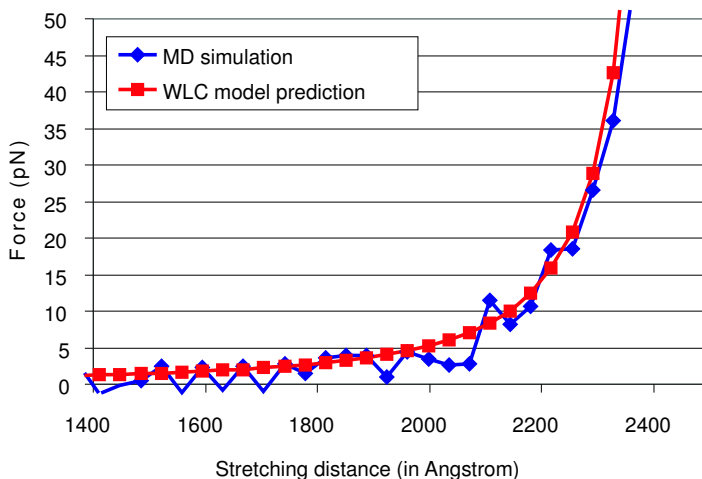
Figure 18. This plot shows a zoom into entropic elasticity before the molecule is completely stretched to its contour length, including a quantitative comparison with the WLC model. The contour length is indicated in the plot as a straight line.

prominent example for such a case are enzymes, where reactions are largely limited to the active site where the substrate is attached (see, for example, the substrate binding site shown in Figure 13b) [Wang et al. 2001].

To reduce the computational cost of reactive force fields, we can use a spatial decomposition scheme, similar to the scheme used in QM/MM approaches [Wang et al. 2001; Cui et al. 2002]. These methods, however, have limitations with respect to the number of atoms that can be treated and difficulties associated with terminating quantum mechanical regions and their handshaking to the empirical force field.

As an alternative approach to the QM/MM methods, we suggest a concurrent coupling of ReaxFF [van Duin et al. 2001] and DREIDING [Mayo et al. 1990]. DREIDING is a nonreactive force field similar to CHARMM. The central idea is to use a smooth transition of force contributions along a spatial decomposition, as shown schematically in Figure 20. We have implemented this algorithm in the CMDF framework [Buehler et al. 2006]. The CMDF framework is a simulation framework that is capable of integrating various simulation methods, facilitating multiscale and multiparadigm modeling [Buehler et al. 2005].

Different algorithms can be used to determine the reactive region. Here we have used a simple approach that is based on selection a set of specific atoms that are required to be embedded in a reactive domain. During the simulation, a spherical region surrounds these atoms whose quantum numbers are dynamically updated. Regions of reactive atoms are linked to a nonreactive force field using the concept of mixed Hamiltonians. The reactive region is updated every N_Ω steps with typical values $N_\Omega = 10\text{-}20$.

Figure 20a depicts the concept of handshaking reactive and nonreactive force fields using force mixing, slowing transitioning from one force field to another. The figure depicts the theoretical method of smoothly transitioning from one force field to another. The specific transition is characterized by two
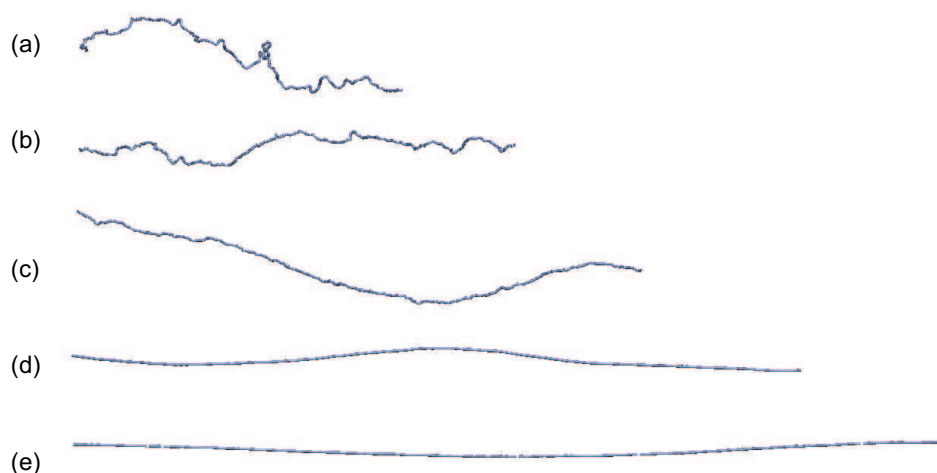
Figure 19. Stretching dynamics of a long alpha-helical molecule, with length of 249.5 nm, which is approximately 50 times larger than the persistence length. The plot depicts snapshots that illustrate the transition from entropic elasticity when the molecule is highly convoluted (subplots (a) and (b)) to a regime where the molecule is beginning to be stretched out entirely. The final subplots (d) and (e) correspond to the regime when energetic elasticity starts to govern the elastic behavior. The stretching force is applied using a steered MD algorithm.
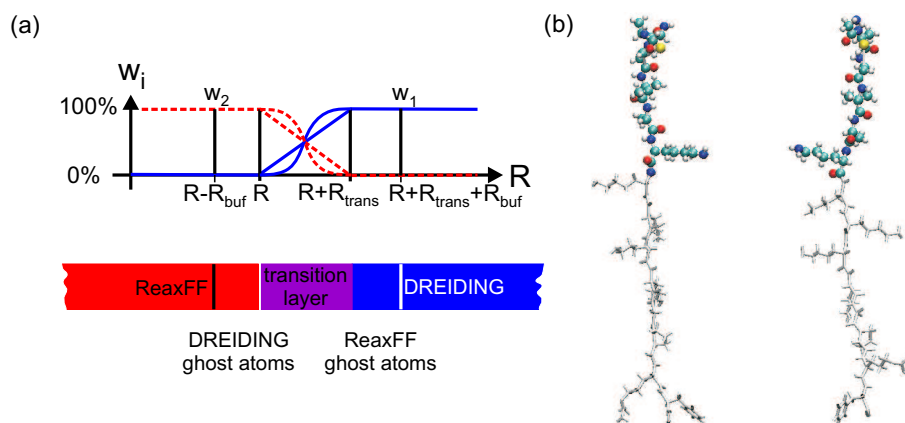

Figure 20. Concept of handshaking reactive and nonreactive force fields using force mixing. Subplot (a) depicts the theoretical method of smoothly transitioning from one force field to another, and subplot (b) shows two snapshots of an example calculations of a small polypeptide molecule. The upper end of the molecule is chosen to be reactive. The colored atoms are those that are included in the ReaxFF treatment, including ghost atoms and the atoms in the transition region. The size of the ghost atom region is chosen equal or larger than the cutoff radius the corresponding force field.

parameters, the width of the transition region R_{trans} and the width of the ghost atom region R_{buf}. While the size of the transition region can be chosen arbitrarily (although ideally, the transition is made very smooth so that energy is conserved), the size of the ghost atom region must be larger than the potential cutoff region, $R_{\text{buf}} > R_{\text{cut}}$. Using a decomposition scheme, each atom is assigned to one or more force engines. This flag determines which atoms are treated by a particular force engine that is available. In addition, each atom features an array weights w_i that determine how much force is included from each force engine. The sum of all weights sum up to one, $\sum w_i = 1$. After each force calculation step with all N force engines, the contributions from each engine (denoted by $\bar{F}_{j,i}$ for atom j) is summed up, yielding the resulting force on atom j, $\bar{F}_j = \sum \bar{F}_{j,i} w_i$. Here we demonstrate that this combination enables large reactive regions of the system to describe the formation and breakage of bonds as needed, while allowing other regions not participating in the reaction to be treated more efficiently. We illustrate our new hybrid method in a study of stretching of small organic molecules, wave propagation studies and modeling of proton transfer in a small protein. Our results suggest that our hybrid model is a practical tool for certain applications including modeling protein dynamics. The possibility of performing such fast screening of different reaction paths may have a major impact on because it may replace existing QM/MM schemes for some cases. However, our model does not conserve energy if the reactive and nonreactive regions are updated dynamically during the simulation.

Figure 20b shows two snapshots of an example calculations of a small polypeptide molecule, obtained using the approach described above, studying the dynamics of the molecule in equilibrium, at a temperature of 300 K. The upper end of the molecule is chosen to be reactive. The colored atoms are those that are included in the ReaxFF treatment, including ghost atoms and the atoms in the transition region. In our example, we chose $R_{\text{buf}} = 10\,\text{Å}$ and $R_{\text{trans}} = 5\,\text{Å}$. The hybrid model is computationally much more efficient than a pure reactive treatment. We leave further analyses to future publications.

5.3. *Summary and outlook.* In this article, we have reported several stretching and unfolding studies of proteins with increasing structural complexity. The results indicate that the secondary and tertiary protein structure has significant effect on the deformation mechanics.

The most important contributions of this paper are:

- to the best of our knowledge, we have reported the first fully reactive treatment of protein molecule mechanics, enabling a coherent description of the intimate links between molecular chemistry and molecular mechanics;

- we have reported large-deformation studies of unfolding and molecular fracture of a AH protein structure, including estimates for Young's modulus as a function of strain and a prediction of its fracture stress, including an atomistic analysis of the bending stiffness of a AH protein;

- we have reported the unfolding dynamics of two proteins, 1AKG and 194L (lysozyme), which illustrates the significance of a reactive treatment in particular when covalent cross-links are present in the molecular structure, showing the significance of such models to describe their nanomechanical properties;

- we have reported development of a mesoscale representation of AH proteins;

- we have described a hybrid reactive-nonreactive model that enables us to perform computationally efficient atomistic reactive calculations.

Restrictions in time and length scales accessible to full atomistic studies indicate that coarse graining techniques (for example, bead models, as discussed in Section 5.1, see Figure 17a) are critical to enable a unified treatment of elastic contributions (for example, the competition between entropic and energetic elasticity, as shown in Figure 17b). Our mesoscale studies predict a smooth transition between these two sources of elasticity as the molecule is stretched shown in Figure 18, where the two extreme cases agree with the predictions by Equations (2) and (3). We emphasize that the bead model can only be as accurate as the input parameters from full atomistic simulation. For example, the high strain rates in MD can induce inaccuracies in the unfolding behavior, thus rendering the corresponding mesoscale model incorrect. This issue can only be addressed by developing more accurate full atomistic methods that can traverse across different time scales. In addition to tensile and bending loading, other loading conditions such as torsional loading or shear of assemblies molecules [Buehler 2006a] could be investigated. Such additional information could also help to render the mesoscale molecular model more accurate.

We further have introduced a hybrid concurrent multiscale model (Section 5.2) that combines reactive and nonreactive force fields using mixed Hamiltonians, which can be an alternative to full reactive studies. An example application of this method is provided in Figure 20.

Our studies could eventually find useful applications in several scientific disciplines. For example, a better understanding of the mechanical response of proteins to mechanical stimulation could lead to advances in the biological sciences. Other applications may be related to provide design suggestions for new biopolymers that could be designed and synthesized based on recombinant DNA technologies [Petka et al. 1998; Tirrell 2002; Langer and Tirrell 2004]. Such new strategies to synthesize materials represent exciting opportunities at the intersections of materials science, biology and chemical engineering. The hybrid reactive-nonreactive modeling scheme could find useful applications in studies and design of enzymes, as it represents an efficient alternative to QM/MM methods [Wang et al. 2001].

Finally, we emphasize that theoretical modeling approaches as discussed in this article are not intended to replace experiment, but rather work hand in hand with experimental efforts to improve our understanding of the complex mechanics of protein materials. We believe that the combination of theory, computer simulation and experiment represents a particularly promising combination to advance the science of deformation and fracture of natural and biological materials.

Appendix A. Mathematical formulation of the reactive mesoscale model

We use a reactive mesoscopic model describing alpha helical (AH) molecules as a collection of beads interacting according to interparticle multibody potentials (see Figure 18a). The total energy of the system is given by

$$E = E_T + E_B + E_{\text{weak}}. \tag{A.1}$$

The total energy is given by the sum over all pair-wise and three-body interactions,

$$E_I = \sum_{\text{pairs}} \phi_I(r), \qquad E_B = \sum_{\text{angles}} \phi_B(\varphi). \tag{A.2}$$

The bending energy is

$$\phi_B(\varphi) = \tfrac{1}{2} k_B (\varphi - \varphi_0)^2, \tag{A.3}$$

with k_B relating to the bending stiffness of the molecule. The bending stiffness parameter is given by

$$k_B = \frac{3EI}{2r_0}.$$

The bending stiffness of a AH molecule is obtained from full atomistic simulation (see main text). The mass of each bead is determined by distributing the mass of the entire AH molecule homogeneously into all beads.

We approximate the nonlinear stress-strain behavior under tensile loading by a multilinear model. The multilinear model is a combination of four spring constants $k_T^{(i)}$, which are turned on at specific intervals of molecular stretch. A similar model has been used successfully in earlier studies of fracture [Buehler et al. 2003; Buehler and Gao 2006a], where the function E_T is given by integrating $F_T(r)$ over the radial distance. The force between two particles is

$$F_T(r) = -\frac{\partial \phi_T(r)}{\partial r} = H(r_{\text{break}} - r) \begin{cases} k_T^{(1)}(r - r_0), & r_1 > r, \\ R_1 + k_T^{(2)}(r - r_1), & r_1 \leq r < r_2, \\ R_2 + k_T^{(3)}(r - r_2), & r_2 \leq r < r_3, \\ R_3 + k_T^{(4)}(r - r_3), & r_3 \leq r, \end{cases} \quad \text{(A.4)}$$

where $H(r - r_{\text{break}})$ is the Heaviside function $H(a)$, which is defined to be zero for $a < 0$, and one for $a \geq 0$. The parameters $R_1 = k_T^{(1)}(r_1 - r_0)$, $R_2 = k_T^{(2)}(r_2 - r_1)$ and $R_3 = k_T^{(3)}(r_3 - r_2)$ come from force continuity conditions. They are fitted to reproduce the force-stretch behavior obtained using the full atomistic model with the molecular formulation.

All parameters are calculated directly from full atomistic results, without empirical fitting. The entire set of parameters of the mesoscopic model is summarized in Table 4.

Since these equations are derived from atomistic simulations at relatively large strain rates, Equations (A.1)–(A.4) are only literally valid for comparable deformation rates. Under much smaller deformation rates, the parameters may change significantly.

Equilibrium bead distance r_0 (in Å)	5.00
Critical distances r_1, r_2 and r_3 (in Å)	5.90, 11.50, 13.0
Tensile stiffness parameters $k_T^{(1)}$, $k_T^{(2)}$, $k_T^{(3)}$, $k_T^{(4)}$ (in kcal/mol/Å^2)	23.80, 0.56, 32.20, 54.60
Bond breaking distance r (in Å)	13.35
Equilibrium angle φ_0 (in degrees)	180.00
Bending stiffness parameter k_B (in kcal/mol/rad^2)	3.44
Mass of each mesoscale particle (in amu)	400.00

Table 4. Summary of the parameters used in the mesoscopic molecular model, based on full atomistic modeling of alpha-helical molecules (1 kcal/mol/Å = 69.479 pN).

Appendix A Molecular visualization scheme

We use the Visual Molecular Dynamics (VMD) program to visualize MD simulation results [Humphrey et al. 1996]. Proteins are depicted in various ways, depending on the type of information that is to be highlighted. In the standard format, atoms are plotted as spheres, and bonds between atoms are drawn as thin lines. Note that whether an atom is bonded or not is determined based on a cutoff scheme, which may not reflect the actual treatment of atoms in the reactive potential, which features a smooth bond order-distance relationship.

In the cartoon mode, helices are drawn as cylinders, beta sheets as solid ribbons with arrows, and all other structure including coils and turns as a tube or wire-like structure. In our plots we typically do not show water molecules and only render atoms that are part of the protein.

Acknowledgement

References

[Aizenberg et al. 2005] J. Aizenberg, J. C. Weaver, M. S. Thanawala, V. C. Sundar, D. E. Morse, and P. Fratzl, "Skeleton of Euplectella sp.: structural hierarchy from the nanoscale to the macroscale", *Science* **309**:5732 (2005), 275–278.

[Akkermans and Warren 2004] R. L. C. Akkermans and P. B. Warren, "Multiscale modelling of human hair", *Philos. T. Roy. Soc. A* **362**:1821 (2004), 1783–1793.

[Alberts et al. 2002] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of the cell*, 4th ed., Garland Science, 2002.

[Anderson 1991] T. L. Anderson, *Fracture mechanics: fundamentals and applications*, CRC Press, 1991.

[Arteca 2003] G. A. Arteca, "Stress-induced shape transitions in polymers using a new approach to steered molecular dynamics", *Phys. Chem. Chem. Phys.* **5**:2 (2003), 407–414.

[Arteca and Li 2004] G. A. Arteca and Z. Y. Li, "Effect of proline kinks on the mechanical unfolding of alpha-helices", *Chem. Phys. Lett.* **399**:4–6 (2004), 496–502.

[Astbury and Street 1932] W. T. Astbury and A. Street, "X-ray studies of the structure of hair, wool and related fibres, I: general", *T. Roy. Soc. Lond. A* **230** (1932), 75–101.

[Balsera et al. 1997] M. Balsera, S. Stepaniants, S. Izrailev, Y. Oono, and K. Schulten, "Reconstructing potential energy functions from simulated force-induced unbinding processes", *Biophys. J.* **73**:3 (1997), 1281–1287.

[Brenner 1990] D. W. Brenner, "Empirical potential for hydrocarbons for use in simulating the chemical vapor-deposition of diamond films", *Phys. Rev. B* **42**:15 (1990), 9458–9471.

[Brenner et al. 2002] D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, "A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons", *J. Phys. Condens. Matter* **14**:4 (2002), 783–802.

[Buehler 2006a] M. J. Buehler, "Atomistic and continuum modeling of mechanical properties of collagen: elasticity, fracture and self-assembly", *J. Mater. Res.* **21**:8 (2006), 1947–1961.

[Buehler 2006b] M. J. Buehler, "Large-scale hierarchical molecular modeling of nanostructured biological materials", *J. Comput. Theor. Nanoscience* **3**:5 (2006), 603–623.

[Buehler 2006c] M. J. Buehler, "Mechanics of protein crystals: atomistic modeling of elasticity and fracture", *J. Comput. Theor. Nanoscience* **3**:5 (2006), 670–683.

[Buehler 2006d] M. J. Buehler, "Nature designs tough collagen: explaining the nanostructure of collagen fibrils", *P. Natl. Acad. Sci. USA* **103**:33 (2006), 12285–12290.

[Buehler and Gao 2006a] M. J. Buehler and H. Gao, "Dynamical fracture instabilities due to local hyperelasticity at crack tips", *Nature* **439**:7074 (2006), 307–310.

[Buehler and Gao 2006b] M. J. Buehler and H. Gao, "Ultra large scale atomistic simulations of dynamic fracture", Chapter 10, pp. 427–467 in *Handbook of theoretical and computational nanotechnology*, vol. 2, edited by M. Rieth and W. Schommers, American Scientific Publishers, Stevenson Ranch, CA, 2006.

[Buehler et al. 2003] M. J. Buehler, F. F. Abraham, and H. Gao, "Hyperelasticity governs dynamic fracture at a critical length scale", *Nature* **426**:6963 (2003), 141–146.

[Buehler et al. 2004] M. J. Buehler, A. Hartmaier, H. Gao, M. Duchaineau, and F. F. Abraham, "Atomic plasticity: description and analysis of a one-billion atom simulation of ductile materials failure", *Comput. Methods Appl. Mech. Eng.* **193**:48–51 (2004), 5257–5282.

[Buehler et al. 2005] M. J. Buehler, A. Hartmaier, H. Gao, M. A. Duchaineau, and F. F. Abraham, "The dynamical complexity of work-hardening: a large-scale molecular dynamics simulation", *Acta Mech. Sin.* **21**:2 (2005), 103–111.

[Buehler et al. 2006] M. J. Buehler, A. C. T. van Duin, and W. A. Goddard, III, "Multiparadigm modeling of dynamical crack propagation in silicon using a reactive force field", *Phys. Rev. Lett.* **96**:9 (2006), 095505–095509.

[Bustamante et al. 1994] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, "Entropic elasticity of lambda-phage DNA", *Science* **265**:5178 (1994), 1599–1600.

[Chenoweth et al. 2005] K. Chenoweth, S. Cheung, A. C. T. van Duin, W. A. Goddard, III, and E. M. Kober, "Simulations on the thermal decomposition of a poly(dimethylsiloxane) polymer using the ReaxFF reactive force field", *J. Am. Chem. Soc.* **127**:19 (2005), 7192–7202.

[Cheung et al. 2005] S. Cheung, W. Q. Deng, A. C. T. van Duin, and W. A. Goddard, III, "ReaxFF(MgH) reactive force field for magnesium hydride systems", *J. Phys. Chem. A* **109**:5 (2005), 851–859.

[Cieplak and Marszalek 2005] M. Cieplak and P. E. Marszalek, "Mechanical unfolding of ubiquitin molecules", *J. Chem. Phys.* **123**:19 (2005), #194903.

[Cieplak et al. 2002] M. Cieplak, T. X. Hoang, and M. O. Robbins, "Thermal folding and mechanical unfolding pathways of protein secondary structures", *Proteins* **49**:1 (2002), 104–113.

[Contera et al. 2005] S. A. Contera, V. Lemaître, M. R. R. Planque, A. Watts, and J. F. Ryan, "Unfolding and extraction of a transmembrane alpha-helical peptide: dynamic force spectroscopy and molecular dynamics simulations", *Biophys. J.* **89**:5 (2005), 3129–3140.

[Courtney 1990] T. H. Courtney, *Mechanical behavior of materials*, McGraw-Hill, 1990.

[Cui et al. 2002] Q. Cui, H. Guo, and M. Karplus, "Combining ab initio and density functional theories with semiempirical methods", *J. Chem. Phys.* **117**:12 (2002), 5617–5631.

[Dao et al. 2003] M. Dao, C. T. Lim, and S. Suresh, "Mechanics of the human red blood cell deformed by optical tweezers", *J. Mech. Phys. Solids* **51**:11–12 (2003), 2259–2280.

[Datta et al. 2005] D. Datta, A. C. T. van Duin, and W. A. Goddard, "Extending ReaxFF to biomacromolecules", 2005. Unpublished.

[van Duin et al. 2001] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, III, "ReaxFF: a reactive force field for hydrocarbons", *J. Phys. Chem. A* **105**:41 (2001), 9396–9409.

[van Duin et al. 2003] A. C. T. van Duin, A. Strachan, S. Stewrnan, Q. Zhang, X. Xu, and W. A. Goddard, III, "ReaxFF SiO: reactive force field for silicon and silicon oxide systems", *J. Phys. Chem. A* **107**:19 (2003), 3803–3811.

[Fainzilber et al. 1994] M. Fainzilber, A. Hasson, R. Oren, A. L. Burlingame, D. Gordon, M. E. Spira, and E. Zlotkin, "New mollusk-specific alpha-conotoxins block aplysia neuronal acetylcholine-receptors", *Biochemistry* **33**:32 (1994), 9523–9529.

[Feig and Brooks 2004] M. Feig and C. L. Brooks, "Recent advances in the development and application of implicit solvent models in biomolecule simulations", *Curr. Opin. Struct. Biol.* **14**:2 (2004), 217–224.

[Fudge and Gosline 2004] D. S. Fudge and J. M. Gosline, "Molecular design of the alpha-keratin composite: insights from a matrix-free model, hagfish slime threads", *P. Roy. Soc. Lond. B Bio.* **271**:1536 (2004), 291–299.

[Gao et al. 2001] M. Gao, H. Lu, and K. Schulten, "Simulated refolding of stretched titin immunoglobulin domains", *Biophys. J.* **81**:4 (2001), 2268–2277.

[Gao et al. 2002] M. Gao, H. Lu, and K. Schulten, "Unfolding of titin domains studied by molecular dynamics simulations", *J. Muscle Res. Cell M.* **23**:5–6 (2002), 513–521.

[Gouldstone et al. 2001] A. Gouldstone, K. J. V. Vliet, and S. Suresh, "Nanoindentation: simulation of defect nucleation in a crystal", *Nature* **411**:6838 (2001), 656.

[Gupta et al. 2004] H. S. Gupta, P. Messmer, P. Roschger, S. Bernstorff, K. Klaushofer, and P. Fratz, "Synchrotron diffraction study of deformation mechanisms in mineralized tendon", *Phys. Rev. Lett.* **93**:12 (2004), #158101.

[Gupta et al. 2005] H. S. Gupta, W. Wagermaier, G. A. Zickler, D. R. B. Aroush, S. S. Funari, P. Roschger, H. D. Wagner, and P. Fratzl, "Nanoscale deformation mechanisms in bone", *Nano Letters* **5**:10 (2005), 2108–2111.

[Han et al. 2005] S. S. Han, A. C. T. van Duin, W. A. Goddard, III, and H. M. Lee, "Optimization and application of lithium parameters for the reactive force field, ReaxFF", *J. Phys. Chem. A* **109**:20 (2005), 4575–4582.

[Harata and Kanai 2002] K. Harata and R. Kanai, "Crystallographic dissection of the thermal motion of protein-sugar complex", *Proteins* **48**:1 (2002), 53–62.

[Hartmaier et al. 2005] A. Hartmaier, M. J. Buehler, and H. J. Gao, "Multiscale modeling of deformation in polycrystalline thin metal films on substrates", *Adv. Eng. Mater.* **7**:3 (2005), 165–169.

[Hulmes et al. 1995] D. J. S. Hulmes, T. J. Wess, D. J. Prockop, and P. Fratzl, "Radial packing, order, and disorder in collagen fibrils", *Biophys. J.* **68**:5 (1995), 1661–1670.

[Humphrey et al. 1996] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics", *J. Mol. Graphics* **14**:1 (1996), 33.

[Jager and Fratzl 2000] I. Jager and P. Fratzl, "Mineralized collagen fibrils: a mechanical model with a staggered arrangement of mineral particles", *Biophys. J.* **79**:4 (2000), 1737–1746.

[Kadau et al. 2004] K. Kadau, T. C. Germann, and P. S. Lomdahl, "Large-scale molecular-dynamics simulation of 19 billion particles", *Int. J. Mod. Phys. C* **15**:1 (2004), 193–201.

[Karplus and McCammon 2002] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules", *Nat. Struct. Biol.* **9**:9 (2002), 646–652.

[Kiss et al. 2006] B. Kiss, A. Karsai, and M. S. Z. Kellermayer, "Nanomechanical properties of desmin intermediate filaments", *J. Struct. Biol.* **155**:2 (2006), 327–339.

[Landis et al. 2002] W. J. Landis, B. L. H. Kraus, and C. A. Kirker-Head, "Vascular-mineral spatial correlation in the calcifying turkey leg tendon", *Connect. Tissue Res.* **43**:4 (2002), 595–605.

[Langer and Tirrell 2004] R. Langer and D. A. Tirrell, "Designing materials for biology and medicine", *Nature* **428**:6982 (2004), 487–492.

[Leach 2001] A. R. Leach, *Molecular modelling: principles and applications*, 2nd ed., Pearson Prentice Hall, 2001.

[Li and Arteca 2005] Z. Y. Li and G. A. Arteca, "Simulated force-induced unfolding of alpha-helices: dependence of stretching stability on primary sequence", *Phys. Chem. Chem. Phys.* **7**:9 (2005), 2018–2026.

[Lorenzo and Caffarena 2005] A. C. Lorenzo and E. R. Caffarena, "Elastic properties, Young's modulus determination and structural stability of the tropocollagen molecule: a computational study by steered molecular dynamics", *J. Biomech.* **38**:7 (2005), 1527–1533.

[Lu et al. 1998] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, "Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation", *Biophys. J.* **75**:2 (1998), 662–671.

[MacKerell et al. 1998] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, and S. Ha, "All-atom empirical potential for molecular modeling and dynamics studies of proteins", *J. Phys. Chem. B* **102**:15 (1998), 3586–3616.

[Marko and Siggia 1995] J. F. Marko and E. D. Siggia, "Stretching DNA", *Macromolecules* **28**:26 (1995), 8759–8770.

[Mayo et al. 1990] S. L. Mayo, B. D. Olafson, and W. A. Goddard, III, "Dreiding: a generic force-field for molecular simula-
tions", *J. Phys. Chem.* **94**:26 (1990), 8897–8909.

[Mehta et al. 1999] A. D. Mehta, M. Rief, and J. A. Spudich, "Biomechanics, one molecule at a time", *J. Biol. Chem.* **274**:21
(1999), 14517–14520.

[Nalla et al. 2003] R. K. Nalla, J. H. Kinney, and R. O. Ritchie, "Effect of orientation on the in vitro fracture toughness of
dentin: the role of toughening mechanisms", *Biomaterials* **24**:22 (2003), 3955–3968.

[Nalla et al. 2005] R. K. Nalla, J. J. Kruzic, J. H. Kinney, and R. O. Ritchie, "Mechanistic aspects of fracture and *R*-curve
behavior in human cortical bone", *Biomaterials* **26**:2 (2005), 217–231.

[Nelson et al. 1996] M. T. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. V. Kalé, R. D. Skeel, and K. Schulten, "NAMD: a
parallel, object oriented molecular dynamics program", *Int. J. High Perform. Comput. Appl.* **10**:4 (1996), 251–268.

[Nielson et al. 2005] K. D. Nielson, A. C. T. van Duin, J. Oxgaard, W. Q. Deng, and W. A. Goddard, III, "Development of
the ReaxFF reactive force field for describing transition metal catalyzed reactions, with application to the initial stages of the
catalytic formation of carbon nanotubes", *J. Phys. Chem. A* **109**:3 (2005), 493–499.

[Ortiz et al. 2005] V. Ortiz, S. O. Nielsen, M. L. Klein, and D. E. Discher, "Unfolding a linker between helical repeats", *J. Mol.
Biol.* **349**:3 (2005), 638–647.

[Papadopoulos et al. 2006] P. Papadopoulos, G. Floudas, I. Schnell, I. Lieberwirth, T. Q. Nguyen, and H. A. Klok, "Thermody-
namic confinement and alpha-helix persistence length in poly (gamma-benzyl-L-glutamate)-b-poly (dimethyl siloxane)-b-poly
(gamma-benzyl-L-glutamate) triblock copolymers", *Biomacromolecules* **7**:2 (2006), 618–626.

[Papamokos and Demetropoulos 2004] G. V. Papamokos and I. N. Demetropoulos, "Biomolecular springs: low-frequency
collective helical vibrations of Ace-Gly(n)-NHMe (n=3-8). A DFT study employing the PW91(XC) functional", *J. Phys.
Chem. A* **108**:40 (2004), 8160–8169.

[Peterlik et al. 2006] H. Peterlik, P. Roschger, K. Klaushofer, and P. Fratzl, "From brittle to ductile fracture of bone", *Nat.
Mater.* **5**:1 (2006), 52–55.

[Petka et al. 1998] W. A. Petka, J. L. Harden, K. P. McGrath, D. Wirtz, and D. A. Tirrell, "Reversible hydrogels from self-
assembling artificial proteins", *Science* **281**:5375 (1998), 389–392.

[Phillips et al. 2005] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé,
and K. Schulten, "Scalable molecular dynamics with NAMD", *J. Comput. Chem.* **26**:16 (2005), 1781–1802.

[Puxkandl et al. 2002] R. Puxkandl, I. Zizak, O. Paris, J. Keckes, W. Tesch, S. Bernstorff, P. Purslow, and P. Fratzl, "Viscoelastic
properties of collagen: synchrotron radiation investigations and structural model", *Philos. T. Roy. Soc. B* **357**:1418 (2002),
191–197.

[Rappe and Goddard 1991] A. K. Rappe and W. A. Goddard, "Charge equilibration for molecular-dynamics simulations", *J.
Phys. Chem.* **95**:8 (1991), 3358–3363.

[Rief et al. 1997] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, "Reversible unfolding of individual titin
immunoglobulin domains by AFM", *Science* **276**:5315 (1997), 1109–1112.

[Rief et al. 2000] M. Rief, M. Gautel, and H. E. Gaub, "Unfolding forces of titin and fibronectin domains directly measured by
AFM", pp. 129–141 in *Elastic filaments of the cell*, 2000.

[Ritchie et al. 2004] R. O. Ritchie, J. J. Kruzic, C. L. Muhlstein, R. K. Nalla, and E. A. Stach, "Characteristic dimensions and
the micro-mechanisms of fracture and fatigue in 'nano' and 'bio' materials", *Int. J. Fract.* **128**:1–4 (2004), 1–15.

[Root et al. 2006] D. D. Root, V. K. Yadavalli, J. G. Forbes, and K. Wang, "Coiled-coil nanomechanics and uncoiling and
unfolding of the superhelix and alpha-helices of myosin", *Biophys. J.* **90**:8 (2006), 2852–2866.

[Sasaki and Odajima 1996] N. Sasaki and S. Odajima, "Elongation mechanism of collagen fibrils and force-strain relations of
tendon at each level of structural hierarchy", *J. Biomech.* **29**:9 (1996), 1131–1136.

[Schwaiger et al. 2002] I. Schwaiger, C. Sattler, D. R. Hostetter, and M. Rief, "The myosin coiled-coil is a truly elastic protein
structure", *Nat. Mater.* **1**:4 (2002), 232–235.

[Springborg 1997] M. Springborg, *Density-functional methods in chemistry and materials science*, Wiley research series in
Theoretical Chemistry, Wiley, 1997.

[Strachan et al. 2003] A. Strachan, A. C. T. van Duin, D. Chakraborty, S. Dasgupta, and W. A. Goddard, III, "Shock waves in high-energy materials: the initial chemical events in nitramine RDX", *Phys. Rev. Lett.* **91**:9 (2003), #098301.

[Strachan et al. 2005] A. Strachan, E. M. Kober, A. C. T. van Duin, J. Oxgaard, and W. A. Goddard, III, "Thermal decomposition of RDX from reactive molecular dynamics", *J. Chem. Phys.* **122**:5 (2005), #054502.

[Strelkov et al. 2002] S. V. Strelkov, H. Herrmann, N. Geisler, T. Wedig, R. Zimbelmann, U. Aebi, and P. Burkhard, "Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly", *EMBO J.* **21**:6 (2002), 1255–1266.

[Stuart et al. 2000] S. J. Stuart, A. B. Tutein, and J. A. Harrison, "A reactive potential for hydrocarbons with intermolecular interactions", *J. Chem. Phys.* **112**:14 (2000), 6472–6486.

[Sun et al. 2001] Y. L. Sun, Z. P. Luo, and K. N. An, "Stretching short biopolymers using optical tweezers", *Biochem. Biophys. Res. Commun.* **286**:4 (2001), 826–830.

[Sun et al. 2004] Y. L. Sun, Z. P. Luo, A. Fertala, and K. N. An, "Stretching type II collagen with optical tweezers", *J. Biomech.* **37**:11 (2004), 1665–1669.

[Tersoff 1988] J. Tersoff, "Empirical interatomic potential for carbon, with applications to amorphous carbon", *Phys. Rev. Lett.* **61**:25 (1988), 2879–2882.

[Tirrell 2002] D. A. Tirrell, "Molecular engineering of protein assembly on surfaces", *Abstr. Pap. Am. Chem. S.* **224** (2002), U408–U408.

[Vaney et al. 1996] M. C. Vaney, S. Maignan, M. Riès-Kautt, and A. Ducruix, "High-resolution structure (1. 33 angstrom) of a HEW lysozyme tetragonal crystal grown in the APCF apparatus. Data and structural comparison with a crystal grown under microgravity from SpaceHab-01 mission", *Acta Crystallogr. Sect. D* **52** (1996), 505–517.

[Wang et al. 2001] W. Wang, O. Donini, C. M. Reyes, and P. A. Kollman, "Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions", *Annu. Rev. Biophys. Biom.* **30** (2001), 211–243.

[Whitesides and Wong 2006] G. M. Whitesides and A. P. Wong, "The intersection of biology and materials science", *MRS Bull.* **31**:1 (2006), 19–27.

MARKUS J. BUEHLER: mbuehler@mit.edu

Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Room 1-272, Cambridge, MA 02139, United States