PHYSICAL REVIEW E, VOLUME 65, 056110

Reversible stretching of random heteropolymers

Phillip L. Geissler and Eugene I. Shakhnovich
Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

(Received 17 July 2001; published 7 May 2002)

We analyze the equilibrium response of random heteropolymers to mechanical deformation. In contrast to homopolymer response, the stress-induced transformation of a heteropolymer from globule to coil need not be sharp. For chain lengths relevant to biological macromolecules, intermediate necklacelike structures dominate over a range of applied force. Stability of these conformations is primarily a consequence of solvation: In a typical necklace, relatively solvophilic regions of the chain are extended, while solvophobic regions remain compact. In the long-chain limit, homopolymeric behavior is recovered. Our results suggest that only select polypeptide sequences should unfold reproducibly at a specific force, explaining recent experimental observations.

DOI: 10.1103/PhysRevE.65.056110 PACS number(s): 64.60.−i, 61.41.+e

Experiments and computer simulations have demonstrated that the native states of certain protein domains can withstand significant stress [1–6]. At strain rates much greater than spontaneous rates of unfolding, these domains become abruptly extended at end-to-end forces $f \sim (10\text{–}100) k_{\mathrm{B}}T/a$. Here, T is temperature, k_{B} is Boltzmann's constant (which we subsequently set to unity), and $a \sim 1$ nm is a typical distance between connected amino acid monomers. The microscopic origin of this tensile strength is not well understood. Specifically, anecdotal evidence suggests that such resistance to stress is not a generic feature of heteropolymers [7–9]. Here we examine this issue in detail, by analyzing the stretching behavior of a model random heteropolymer.

The response of homopolymers to stretching in poor solvent ($T < \theta$) has been established by scaling analysis [10]. At low force, a collapsed homopolymer globule responds linearly, but weakly, to stress. Due to surface tension, the resulting deformation scales only as $fN^{-1/3}$, where N is the number of monomers. For large forces, the chain is essentially coil-like, and highly flexible until near its maximum extension. The transition between these two regimes occurs sharply at intermediate force. Exactly at the transition, globular and coil-like regions coexist, i.e., the chain is a necklace of compact regions connected by extended strings. (An example of such a structure is sketched in Fig. 1.) Away from the transition, these necklace structures are unstable.

A similar scenario is expected for random heteropolymer stretching, with a few important differences. Specifically, the statistics of compact structures are dominated by $O(1)$ conformations at sufficiently low temperature. The associated freezing transition is well understood [11]. More interestingly, the role of necklace structures is enhanced by energetic heterogeneity of globule and coil subunits. Because the local monomer composition varies within a sequence, the susceptibility of a given region of the chain to extension is effectively a random variable. (See Fig. 1.) We show in this article that localization of subunit arrangements due to disorder can stabilize necklaces in a finite interval of force about the globule-coil transition.

Necklaces have been predicted to play an even more prominent role in the stretching of polyelectrolytes and polyampholytes [12,13]. In fact, these molecules are necklacelike even in the absence of stress for sufficient net charge density [14,15]. Although polyelectrolyte necklaces bear structural similarities to heteropolymer necklaces, the physical origins of their stability are fundamentally different. In the case of polyelectrolytes, a globule breaks into smaller, tethered subunits due to the Rayleigh instability of charged droplets. For random heteropolymers, it is instead a combination of stress and disorder that causes structural segregation. The principal consequences of deformation also differ considerably for these two cases. Stretching polyelectrolytes primarily redistributes monomers among compact and extended regions of necklaces. Stretching random heteropolymers, on the other hand, disrupts a fully compact globular state. Smooth apportionment of monomers into compact and extended regions is frustrated in this case by quenched disorder.

Our interest in polymer deformation is motivated by the protein stretching experiments described above, and the cor-

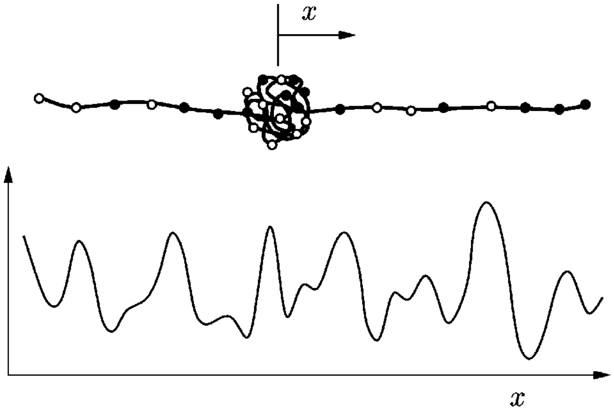

FIG. 1. Energetics of a random heteropolymer necklace. Monomers of two different types are depicted as black and white circles, respectively. Volume interactions of the compact region, as well as solvation energetics of extended regions, depend on the local abundance of each monomer type. Because the local sequence composition is a random variable, the arrangement of globule and coil subunits is effectively determined by a random potential $u(x)$. For the example structure shown, $u(x)$ can be viewed as the potential for "motion" of the single compact region along the sequence.

responding biological functions of mechanical proteins. For this reason, we focus on the effects of sequence disorder. Electrostatic effects *at physiological conditions* are expected to be weak by comparison for the chemical compositions of typical proteins [16,17]. In experiments, modular proteins consisting of several (often identical) domains are elongated at a fixed rate. These domains appear to unfold independently. For extensions at which the protein's restoring force can be reliably measured, the flexibility of unfolded domains mediates the applied strain. Indeed, using a simple elastic model, measured restoring forces can be predicted almost quantitatively from the contour length of unfolded regions [18]. For a particular extension of the entire protein, folded domains are thus effectively subjected to uniform stress. We therefore examine stress-controlled, rather than strain-controlled, deformation of individual domains.

We consider a microscopic model of random heteropolymers, in which monomers are represented as coarse-grained beads on an elastic string. For simplicity, the identity of each monomer is restricted to two possibilities, $\sigma_i = \pm 1$, perhaps denoting hydrophobic and hydrophilic moieties. We take the total composition of the sequence to be fixed, $\Sigma_{i=1}^{N} \sigma_i = 0$. The chain's conformation is specified by the positions $\mathbf{r}_i$ of N monomers. For a given monomer sequence $\{\sigma_i\}$, the energy of a conformation is

$$\mathcal{H} = \mathcal{H}_0 + \Gamma \sum_{i \text{ exposed}} \sigma_i - \mathbf{f} \cdot (\mathbf{r}_N - \mathbf{r}_1), \tag{1}$$

$$\mathcal{H}_0 = \sum_{i,j=1}^{N} \delta(\mathbf{r}_i - \mathbf{r}_j)(B_0 + \chi \sigma_i \sigma_j). \tag{2}$$

Here, $B_0 = T - \theta$ is the mean energy density stabilizing compact conformations for $T < \theta$. The heterogeneity of interactions is determined by χ. We choose $\chi < 0$, so that attractions are strongest between monomers of the same type. The second term in Eq. (1) describes solvation of exposed portions of the chain. The summation thus includes only monomers that are accessible to solvent. For $\Gamma > 0$, solvation of monomers with $\sigma_i = -1$ is favorable, while that for $\sigma_i = 1$ is unfavorable. The final term in Eq. (1) couples the chain's end-to-end vector to a stretching force $\mathbf{f}$.

We analyze first the statistics of fully compact conformations of this model heteropolymer. As in the homopolymeric case, surface tension prevents significant deformation of the globular state for modest forces. We thus consider $\mathbf{f} = \mathbf{0}$. A detailed analysis of this state in the absence of explicit solvation was presented in Ref. [19]. Freezing of the globule into a few low-energy conformations was demonstrated within mean-field theory, employing the replica trick to average over quenched random sequences [11,20]. The one-step replica symmetry breaking shown in that work is consistent with a suitably chosen random energy model. In other words, the average thermodynamics of compact states are reproduced by drawing $(a^3/v)^N$ energy levels at random from a distribution $P(E) = (\pi N \Delta^2)^{-1/2} \exp[-(E-\overline{E})^2/N\Delta^2]$, with width $\Delta = |\chi| \mu^2 \overline{\rho}$ and mean $\overline{E} = B_0 N \overline{\rho}$. Here, v is the volume occupied by a monomer, $\overline{\rho} \sim v^{-1}$ is the mean density in the globule core, and μ^2 is the variance of the monomer distribution. The ground states of particular random sequences are distributed narrowly about $E^*(N) = \overline{E} - N\Delta(\ln a^3/v)^{1/2}$ [21].

Solvation broadens the distribution of globular energies. Following the replica mean-field analysis of Ref. [19], we find that the second term in Eq. (1) increases Δ by $(\Gamma^2 \overline{\rho}/4|\chi|) A/N$ [22]. Here, A is the surface area of the globule. This result may be obtained more simply by assuming that surface and volume contributions to the energy are independent random variables [23]. Within this approximation, the variance in surface energy $\Delta_{\mathrm{surf}}^2 = (\Gamma \mu \overline{\rho})^2 A/N$ adds directly to that of monomer-monomer interactions, yielding the described increase in Δ at lowest order in A/N. Introducing explicit solvation thus lowers the average ground state energy by an amount $(\Gamma^2 \overline{\rho}/4|\chi|)(\ln a^3/v)^{1/2} A$. If χ and Γ are comparable in magnitude, this shift is a significant fraction of the energy gained by exposing only monomers with $\sigma_i = -1$ to solvent. The density of states around E^* is sufficiently large in this case that a ground state with favorable solvation energetics may always be selected.

Coil-like statistics of the chain at large stress are well described by simple models of polymer flexibility. A reasonable choice, the freely jointed chain, has deformation free energy $\mathcal{F}_{\mathrm{coil}}(N) = -T \ln[\sinh(fa/T)/(fa/T)]$ [24]. This model has the advantage that its maximum extension is finite. In an ideal coil, nearly all monomers are exposed to solvent, so that the first two terms in Eq. (1) do not contribute. Sequence disorder is thus unimportant for this fully expanded state.

A heteropolymer necklace is composed of virtually noninteracting globule and coil subunits. (By contrast, interactions between different subunits are important for polyelectrolye necklaces.) Due to the quenched nature of sequence disorder, different arrangements of these subunits within a particular chain have different weights. As illustrated in Fig. 1, this heterogeneity of necklace stabilities can be viewed as the consequence of an effective random potential u. The scale and correlation length of u are determined by the sizes of subunits, and by statistics of the sequence. In detail, a subunit comprising $N_{\mathrm{s}} < N$ monomers, situated at a given location in the sequence, has in general an excess of one monomer type: $q \equiv N_{\mathrm{s}}^{-1} \Sigma_{i \in \mathrm{subunit}} \sigma_i \neq 0$. For uncorrelated sequence statistics, the local composition q is a Gaussian random variable with fluctuations $\mu(1/N_{\mathrm{s}} - 1/N)^{1/2}$. The apparent distribution of monomer types within a subunit reflects this excess, $w(\sigma_i;q) \propto \exp[-(\sigma_i - q)^2/2\mu^2(1-q^2)]$. These local variations in sequence composition affect the free energetics of both globule and coil subunits.

According to replica mean-field theory for globule energetics [19,22], a compact subunit with monomer distribution $w(\sigma_i;q)$ has ground state energy

$$E_{\mathrm{s}}^* = (B_0 - |\chi| q^2) \overline{\rho} N_{\mathrm{s}} - \Gamma q \overline{\rho} N_{\mathrm{s}}^{2/3} - \left(\ln \frac{a^3}{v} \right)^{1/2} \times \left[2|\chi| \mu^2 (1-q^2) \overline{\rho} N_{\mathrm{s}} + \frac{\Gamma^2 \overline{\rho}}{4|\chi|} N_{\mathrm{s}}^{2/3} \right]. \tag{3}$$

This expression extends standard results for a random copolymer [19]: Terms proportional to q^2 account for the possible excess of one monomer type within a subunit, while the final term, proportional to surface area, arises from solvation. This contribution of solvation is consistent with the modified random energy model described above. The energy of a globular subunit varies with its composition q, and thus depends on its location in the sequence. But these variations are weak. The magnitude of fluctuations in E_s^*, averaged over the distribution of q, scale only as $N_s^{1/6}$.

Fluctuations in the energy of coil subunits are more pronounced. Because the composition of a coil subunit is not fixed, its solvation energy $-\Gamma q N_s$ need not vanish. In fact, variations in this energy, averaged over the distribution of q, are of size $\Gamma \mu N_s^{1/2}(1-N_s/N)^{1/2}$, and may be considerable. Solvation effects therefore establish the scale of u. They are enhanced in coil-like subunits because a large fraction of constituent monomers are exposed to solvent.

We focus on necklaces consisting of m globular regions of size M, with free energies

$$\mathcal{F}_{\rm neck}(m,M)=mE_s^*(M)+mM^{2/3}\gamma + \mathcal{F}_{\rm coil}(N-mM)+\mathcal{F}_{\rm rand}(m,M), \qquad (4)$$

where γ is the surface tension of the compact state. The final term in Eq. (4), $\mathcal{F}_{\rm rand}$, arises from fluctuations in the arrangement of subunits within the chain. Statistics of these fluctuations are estimated using a set of Ω states drawn at random from the potential u. Here, $\Omega \approx [M^{-m}(N-M+1)(N-2M+1)\cdots(N-mM+1)/m!]$ is the number of statistically independent subunit arrangements. Identifying the number of arrangements with energy E as $\Omega\sqrt{2\pi\langle u^2\rangle}\exp[-E^2/2\langle u^2\rangle]$, the corresponding free energy may be evaluated directly [25]:

$$\mathcal{F}_{\rm rand}(m,M)=\begin{cases} -sT[1+(T_\ell/T)^2], & T>T_\ell \\ -2sT_\ell, & T\leqslant T_\ell. \end{cases} \qquad (5)$$

Here, $s=\ln\Omega$ and $T_\ell=\Gamma\mu[mM(1-mM/N)/2s]^{1/2}$. For $T<T_\ell$, a few low-energy subunit arrangements dominate. It is this localization in the deepest minima of u that stabilizes heteropolymer necklaces. These minima correspond to structures in which favorably solvated regions of the chain exist in coil-like states, and unfavorably solvated regions are sequestered in the cores of globular subunits. Equation (5) suggests that the distribution of subunit sizes is quite broad. For small globular regions ($M\ll N$), the minimum value of $\mathcal{F}_{\rm rand}$ that does not introduce macroscopic surface tension is of order $N^{1/2}$. For a single, large globular region [$M=O(N)$], $\mathcal{F}_{\rm rand}$ is $O(N^{1/2})$ as well.

This scaling of $\mathcal{F}_{\rm rand}$ indicates that necklaces are stable for stretching forces in an interval of width $N^{-1/2}$ about the globule-coil transition. Deformation behavior of very long chains is therefore effectively homopolymeric. For finite macromolecules, however, a phase of necklace structures modifies the response to stress. Stretching phase diagrams determined from Eq. (5) are plotted in Fig. 2 for $N=10^2$ and $N=10^3$. The general features of these phase diagrams are in good agreement with results of computer simulations. In particular, simulations of a short lattice heteropolymer exhibited partially extended structures, as well as a reentrant coil phase at low temperature [7,26]. This reentrant phase, evident in Fig. 2, includes only small fluctuations about a completely extended structure.

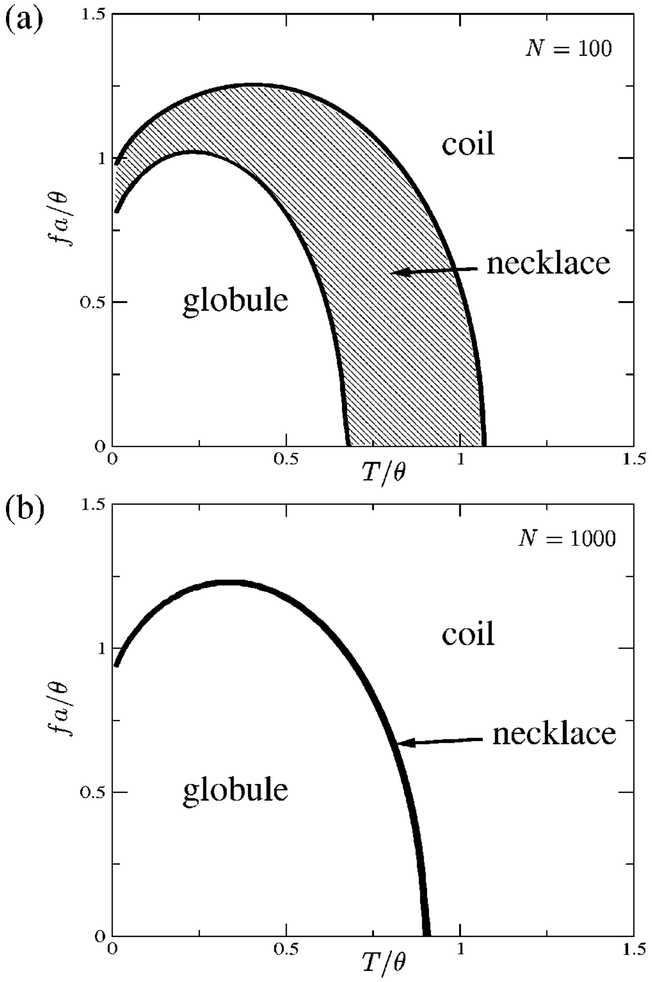

FIG. 2. Stretching phase diagrams of typical random heteropolymers with (a) $N=10^2$ and (b) $N=10^3$. Coexistence lines were determined using Eqs. (4) and (5). Specifically, the average fraction of monomers residing in compact regions of the chain, ϕ, was computed as a function of f and T by summing $mM\exp[-\mathcal{F}_{\rm neck}(m,M)/T]$ over all possible values of m and M. Note that $M=N$ and $M=0$ correspond to fully compact and fully expanded states, respectively. For $\phi>0.95$, we consider the state of the chain to be globular. Similarly, the chain is considered coil-like for $\phi<0.05$. In the intermediate regime, $0.05<\phi<0.95$, the necklace phase (shaded regions) dominates. For these calculations, we have chosen the scale of solvation energetics to coincide with that of monomer attractions, $\Gamma\sim\theta$. We further assume that the chain is flexible on the scale of monomer size, $\overline{\rho a}^3\approx 1$.

We have shown that random heteropolymer necklaces are stabilized by variations in local sequence composition. Necklace stability is therefore sensitive to the nature of sequence statistics. For blocky sequences, in which monomers are likely to be of the same type within a correlation length ξ, variations in local composition are large. If this correlation length is macroscopic, $\xi=O(N)$, or if sequence correlations decay algebraically, then $\mathcal{F}_{\rm rand}\sim N$. In this case, the necklace phase persists in the long-chain limit. By contrast, sequences that are anticorrelated on a scale ξ exhibit small fluctuations in local composition. If the sequence is ''neutral'' (i.e., $q=0$) on a microscopic scale, $\xi \ll N$, fluctuations in solvation energy will be no larger than typical thermal fluctuations. In this case, stretching behavior will be homopolymeric for all chain lengths.

A complete analysis of protein stretching requires consideration of the nonequilibrium conditions at which experiments have been performed. Understanding the corresponding equilibrium behavior is a necessary first step toward this goal. Indeed, the response to reversible stretching determines certain features of the distribution of nonequilibrium dynamics [27]. If we further associate the equilibrium stability of various structures with their kinetic significance, a qualitative picture of nonequilibrium response can be sketched: When the necklace phase is stable over an appreciable range of stretching forces, we expect the globular state to unfold partially at intermediate forces. The breadth we have predicted for the ensemble of necklace structures suggests that fluctuations in extension will be large in this scenario. Such behavior has in fact been observed for the protein barnase [9]. If, on the other hand, a sequence has been designed to suppress partial unfolding, the native state will remain intact until stress favors fully expanded, coil-like conformations. Interestingly, many of the protein domains involved in muscle elasticity and the extracellular matrix are rich in β-sheet secondary structure. Important long-range contacts in these structures discourage local grouping of hydrophobic side chains in the sequence. Evolutionary design of stress-resistant proteins thus appears to reflect the basic relationship we have predicted between sequence statistics and necklace stability.

We thank A. Yu. Grosberg for many helpful conversations regarding this work.

[1] A.F. Oberhauser, P.E. Marszalek, H.P. Erickson, and J.M. Fernandez, Nature (London) **393**, 181 (1998).

[2] M. Rief, M. Gautel, A. Schemmel, and H. Gaub, Biophys. J. **75**, 3008 (1998).

[3] P.E. Marszalek *et al.*, Nature (London) **402**, 100 (1999).

[4] H. Li *et al.*, Proc. Natl. Acad. Sci. U.S.A. **97**, 6527 (2000).

[5] H. Lu and K. Schulten, Proteins: Struct., Funct., Genet. **35**, 453 (1999).

[6] E. Paci and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **97**, 6521 (2000).

[7] D.K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **96**, 6166 (1999).

[8] G. Yang *et al.*, Proc. Natl. Acad. Sci. U.S.A. **97**, 139 (2000).

[9] R. Best *et al.*, Biophys. J. **81**, 2344 (2001).

[10] A. Halperin and E.B. Zhulina, Europhys. Lett. **15**, 417 (1991).

[11] E.I. Shakhnovich and A.M. Gutin, Biophys. Chem. **34**, 187 (1989).

[12] T.A. Vilgis, A. Johner, and J.F. Joanny, Eur. Phys. J. E **2**, 289 (2000).

[13] M.N. Tamashiro and H. Schiessel, Macromolecules **33**, 5263 (2000).

[14] Y. Kantor and M. Kardar, Phys. Rev. E **51**, 1299 (1995).

[15] A.V. Dobrynin, M. Rubinstein, and S.P. Obukhov, Macromolecules **29**, 2974 (1996).

[16] K.A. Dill, Biochemistry **29**, 7133 (1990).

[17] L. Xiao and B. Honig, J. Mol. Biol. **289**, 1435 (1999).

[18] M. Rief, J.M. Fernandez, and H. Gaub, Phys. Rev. Lett. **81**, 4764 (1998).

[19] C.D. Sfatos, A.M. Gutin, and E.I. Shakhnovich, Phys. Rev. E **48**, 465 (1993).

[20] K. Binder and A.P. Young, Rev. Mod. Phys. **58**, 801 (1986).

[21] B. Derrida, Phys. Rev. B **24**, 2613 (1981).

[22] P. L. Geissler and E. I. Shakhnovich, Macromolecules (to be published).

[23] P. L. Geissler and E. I. Shakhnovich (unpublished).

[24] A. Y. Grosberg and A. R. Khokhlov, *Statistical Mechanics of Chain Molecules* (AIP, Woodbury, NY, 1994).

[25] B. Derrida, Phys. Rev. Lett. **45**, 79 (1980).

[26] D.K. Klimov and D. Thirumalai, J. Phys. Chem. B **105**, 6648 (2001).

[27] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).