

Stretching a heteropolymer

Elena Jarkova

Institut Charles Sadron, 67083 Strasbourg Cedex, France

Thijs J. H. Vlugt

Condensed Matter and Interfaces, Debye Institute, Utrecht University, 3508 TA Utrecht, The Netherlands

Nam-Kyung Lee^{a)}

Department of Physics, Institute of Fundamental Physics, Sejong University, Seoul 143-747, South Korea

(Received 16 November 2004; accepted 26 January 2005; published online 23 March 2005)

We study the elastic properties of single heteropolymers. By means of exact enumeration of conformations, Monte Carlo (MC) simulation, and variational principles, we calculate *equilibrium* force-extension curves of heterocopolymers for specific arrangements of the monomer types along the sequence. At a given extension z , the time averaged measured force is the weighted sum of restoring forces for various configurations. Using variational principles, we calculate force-extension (f - z) curves of heteropolymers with fixed extensions z . These results are compared with f - z curves obtained from MC simulations and exact enumeration of all conformations. Typical random sequences manifest several piecewise unfoldings of blocks of various size, which are overlapping due to thermal fluctuations. The shape of the elastic response of a heteropolymer reflects the disorder in the primary block structure and the binding energies of these blocks. © 2005 American Institute of Physics. [DOI: 10.1063/1.1874853]

I. INTRODUCTION

Inspired by a series of remarkable single molecule experiments,¹⁻³ the statics and dynamics⁴ of single polymer chains has drawn much attention in recent years. These experiments using atomic force microscopy¹⁻³ and optical and magnetic tweezers^{5,6} show that molecules unfold domain by domain by applying mechanical stress resulting in a sawtooth shape in the force-extension curve. Essential features of biomolecules such as DNA, RNA, and proteins are heterogeneity in monomer types and disorder in their arrangement. The biological activity and function relies on the spatial conformation of the folded structures mainly determined by the chemical sequence of different monomer types. The chain adapts its conformation to minimize the free energy at a given (quenched) sequence of monomers such as amino acids. The linear sequence of a heteropolymer contains all necessary information for the proper spatial arrangement of the chain.

When an external force is applied, the folded chain undergoes an unfolding transition. The force-extension (f - z) curve seems to reflect detailed knowledge of the information stored in the chain sequence. For example, a chain in its native structure undergoes a very sharp transition, while a heteropolymer with an arbitrary random sequence manifests smooth continuous unfoldings.

Although there exist several theoretical studies of elastic properties of heteropolymers at the coarse grained level,⁷⁻¹¹ the complete connection relating primary structure and f - z curves seems to be still missing especially for heteropolymers whose free energy landscape is rugged. Our main goal

is to study how the sequence of a single chain (which has disorder and heterogeneity) is projected onto its elastic responses. We calculate equilibrium f - z curves of various randomly created sequences and designed proteinlike sequences. The possibility of a reversed analysis is discussed in Ref. 12.

In a previous study,⁹ we minimized the free energy of a self-assembled heteropolymer at a *given force*. This theoretical method uses a Gaussian trial function for a given randomly generated sequence and we obtained distance correlations of all monomer pairs corresponding to the global or local energy minimum. As expected, the characteristic feature of f - z curves is the occurrence of plateaus when domains unravel. The extension z increases abruptly by δz when the elastic energy gain is comparable with increase of the potential energy. We note that this plateau might be related to the unfoldings of several domains of similar binding energies.⁹ Therefore, sequence information is partly washed away under the constant force description. Thus, for the more explicit comparison of f - z curves and primary structure, we are motivated to provide a theoretical framework to describe the elastic response at an imposed extension z . Formally, we incorporate a Lagrangian multiplier in a previously used variational method.^{9,13}

Furthermore, we also evaluate f - z curves using extensive Monte Carlo simulations on a cubic lattice. We use exact enumeration of conformations as well as the density of states (DOS) method in the scheme of Monte Carlo simulations to compute f - z curves and free energies of chain conformations.¹⁴⁻¹⁷ For the systematic comparison between primary sequence information and elastic responses, we investigate simple block copolymers with various block sizes, copolymers with disorder such as HP 2-letter random hetero-

^{a)}Electronic mail: lee@sejong.ac.kr

copolymers (RHP) (Refs. 18 and 19), and a 20-letter amino-acid model²⁰ which resembles model proteins. In the HP model, there are only two types of monomers, hydrophobic H and hydrophilic or polar P groups.^{19,21} In the more elaborate 20-letter matrix we distinguish all 20 different types of amino acids and their mutual interaction energies.²⁰ From the DOS method, we can access to the correct f - z curve in thermodynamic equilibrium.

Several molecular dynamics simulations based on a protein model²²⁻²⁵ report very sharp transitions in f - z curves at least for small globular proteins. Such sharp transitions demonstrate that the energy of the ground state is separated from the energies of the other conformations. When the primary structure consists of random sequences, the density function is a smooth function of z and the f - z curve manifests several smoothed saw-tooth shapes, each of them is originating from the unfolding of a small "domain." Thus the overall shape of the elastic response reflects the degree of disorder in the primary structure as well as the binding energy. In this study, we will discuss in detail the sequence dependence of the shape of f - z curves of heteropolymers. We present our variational calculations of $f(z)$ followed by MC simulations. For simplicity, we assume that the chain is flexible for both the variational method and the MC simulations and we use a fixed bond length in the MC simulations.

II. GAUSSIAN DESCRIPTION OF A HETEROPOLYMER

We consider a polymer chain consisting of N monomers of size b . Formally, the Hamiltonian of a heteropolymer chain is described by

$$\begin{aligned} \frac{\mathcal{H}_{\text{hetero}}}{k_B T} &= \frac{d}{2b^2} \int ds \left(\frac{d\vec{r}(s)}{ds} \right)^2 + \int \int ds ds' \frac{v_{ss'}}{2!} \\ &\times \delta[\vec{r}(s) - \vec{r}(s')] + \frac{w}{3!} \int \int \int ds ds' ds'' \\ &\times \delta[\vec{r}(s) - \vec{r}(s')] \delta[\vec{r}(s') - \vec{r}(s'')], \end{aligned} \quad (1)$$

where $\vec{r}(s)$ is the position of monomer on s along the contour. The first term corresponds to the elastic properties (connectivity) of the polymer chain and is approximated by the Wiener measure²⁶ with d being the dimensionality. The other terms contain all interactions between monomers. In particular, two- and three-body interactions of the virial expansion are included here. The inclusion of three-body interactions prevents the chain from collapsing to a single point. The interaction between each monomer pair is taken into account in the second virial coefficient $v_{ss'}$ in units of b^3 . The two-body interactions can be attractive or repulsive depending on the type of the monomer pairs. The density of the collapsed structure is then determined by the balance between three-body repulsions and two-body attractions. In the HP model, $v_{ss'}$ can take three different values: $v_{PP}, v_{HP} = v_{PH}, v_{HH}$.

Below the critical temperature, the chain is a compact folded globule. When an external force is applied, the chain undergoes a structural transition. The total free energy of a chain with given extension z is

$$F = \langle E \rangle - T \langle S \rangle. \quad (2)$$

Assuming Gaussian statistics for the released segment of length N_r , the entropy of the conformation is $-3k_B z^2/2N_r$. At large imposed extension, the elastic part of the free energy is mainly originating from entropic penalties. To calculate the f - z curve, we should find the proper partition sum at given z . If there are n_r conformations of extension z , each of them with an energy E_r , the partition sum $\mathcal{Z}(z)$ is sum of statistical weights of all contributing conformations, $\mathcal{Z}(z) = \sum_{n_r} e^{-E_r(z)/k_B T}$. The restoring force is thus

$$f = -k_B T \frac{\partial \ln \mathcal{Z}(z)}{\partial z}. \quad (3)$$

However, because of the disorder present in the sequence arrangement, it is not straightforward to calculate this sum. Later we will show that exact enumeration of the configuration space is possible for relatively short chains living on a cubic lattice. In Sec. II A, we present variational principles²⁷⁻³⁰ for heteropolymers which provide the upper bound of the free energy of the disordered system.

If an unfolding occurs at the transition point (z^*), there are two main conformations characterized by different energies and contour lengths, (E_1, N_1) and (E_2, N_2) with statistical weights $\propto \exp[-E_1/k_B T]$ and $\propto \exp[-E_2/k_B T]$, respectively. The $f(z)$ curve is characterized by a saw-tooth shape; a maximum at z^* followed by a force drop. The change of contour length $\delta N = N_2 - N_1$ varies according to the sequence disorder. The elastic response around z^* can be approximated by the overlap of the elastic responses before [denoted as $f_1(z)$] and after [denoted as $f_2(z)$] the unfolding. The suggested fitting function in Ref. 12 is the overlapping elastic responses of both sides, $f_1(N_1, z)$ and $f_2(N_2, z)$,

$$f(z) = f_1(N_1, z) \tilde{\theta}(z^* - z) + f_2(N_2, z) \tilde{\theta}(z - z^*), \quad (4)$$

where $\tilde{\theta}(z^* - z) = 1/(1 + \exp\{-[(z^* - z)/z^*]2\varepsilon/k_B T\})$ and ε being the potential energy difference $E_2 - E_1$. The fitting function is obtained by interpolating free energy difference $E_2(z) - E_1(z)$ as $2\varepsilon(z^* - z)/z^*$ in the vicinity of the transition. This means that the broadening of the saw-tooth shape is proportional to the size of the thermal fluctuations as the crossover in the θ function occurs around $\sim k_B T z^*/2\varepsilon$. Below, we show that the interplay between the sequence disorder and thermal fluctuation determine the shape of the f - z curves.

A. Variational principles

Here we apply the variational method to study stretching experiments of a single heterocopolymer with imposed end-to-end distance z . We prescribe the continuous description of the monomer position and the harmonic bond potential.

To generalize the variational method for the constraint system, the minimization procedure includes the following constraint:

$$|\vec{r}_N - \vec{r}_1| = z. \quad (5)$$

To take this constraint into account we introduce a Lagrangian multiplier l in the Hamiltonian. Then Eq. (1) reads

neighboring H blocks. Thus we focused on the distances between the transition points and their correlation with the connecting P-block size.

A block copolymer forms a self-assembled micellar structure where all H blocks belong to a single globule [Fig. 1(b)]. Upon pulling, a pair of (H-P) blocks becomes separated from the large globule and aligns in the direction of pulling. Separating an end block is more favorable from the point of minimizing the surface energy if all H blocks have equal sizes. Splitting the globule into two smaller globules of similar size will cost more surface energy because of a larger surface area. These conformations are thus less favorable. A separated segment of H monomers still forms a collapsed globule by itself [see Fig. 1(c)]. Neglecting the collapsed size of a H globule, the increase of total contour length is mainly determined by the length of the released P blocks. As the extension increases, the elastic response exhibits a series of smooth saw-tooth shapes. Each peak of the saw-tooth pattern corresponds to a conformational transition.

Without thermal fluctuations, the saw-tooth shape would have a sharp peak and an abrupt force drop. At the point of the n th transition (z_n^*), the most probable state is switched from the conformation of linear length nN_b to $(n+1)N_b$. The size of force-drop is related to the binding energy ε : $\delta f z_n^* = \varepsilon$. For a Gaussian chain, as is assumed in the variational description, elastic responses before and after this transition are linear responses: $f_1 = 3z/N_1 b^2 k_B T$ and $f_2 = 3z/N_2 b^2 k_B T$, respectively. We note that the slope of $f(z)$ before and after the transition [$1/N_i (k_B T/b^2)$] should be inversely proportional to the linear length N_i ($i=1,2$). The position of the n th transition point is $z_n^* = \sqrt{n(n+1)\varepsilon N_b b}$ and the maximum force is $f(z_n^*) = z_n^*/N_b (k_B T/b^2)$. Both z_n^* and $f(z_n^*)$ are increasing functions of n . Thus, from the knowledge on the distance between transitions δz^* and the size of force drop, one can estimate the block size N_b and its binding energy ε .

At finite temperature, the saw-tooth shape of f - z curves is broadened due thermal fluctuations. The width of thermal broadening is $\sim k_B T z_n^*/2\varepsilon$, which is growing with increasing n . The curves of later transitions are strongly overlapping and smoothed out. Therefore, the first few transitions are more pronounced assuming the same binding energy for each unfolding [see Fig. 2(a)]. In the presence of thermal fluctuations, we can access N_b and ε by fitting f - z curves using Eq. (4).

At this point, we consider the influence on disorder in the primary structure. We created several HP sequences by a Markovian process in which ϕ is the probability that monomer i has a different type than monomer $i-1$. This resembles a polymerization process in which the chemical reaction strongly depends on the type of end monomer. The number of created blocks $n_x = \phi N$ follows from a Poisson distribution with mean block size $\langle x \rangle = 1/\phi$. The block size distribution equals $P(x) = (\phi N)^{N/x} \exp[-\phi N]/(N/x)!$ (for large N) with fluctuations in the block size $\delta x \sim (\phi)^{-3/2}/N$. The correlation between monomers decays exponentially with the characteristic length $1/\phi$. If $\phi=0.5$, the sequence is completely random. Small values of ϕ result in large block sizes. If $\phi \approx 1$, the generated sequences are frequently alternating. An alternating sequence can be thought of as a coarse-grained ho-

mopolymer whose elastic response is also expected to be similar to that of a homopolymer. For small ϕ , if the total chain length is comparable to the average block size, $N\phi \sim 1$, the disorder in blocky sequences is not self-averaging. Thus there is a high probability to construct a unique block structure (besides local degeneracy). If ϕ is very small ($\phi \ll 1$), our chain includes only H-type or P-type monomers resulting in a homopolymer. Hence, to form a sequence that folds into a unique structure, the total chain length N should be long enough to include a sufficient number H monomers necessary to form a stable globular structure but also short enough to avoid any repetitions of blocks which causes degeneracy.

The disorder present in the block sizes influences the domain binding energy ε as well as the linear size of domains δl . For homopolymers consisting of all H-type monomers, there is only a single energy scale, $\tau = v_{HH}/b^3$, associated with the collapse of the chain. It is shown by Halperin and Zhulina³² that the unfolding of a homopolymer globule is a first-order phase transition and this transition occurs when the characteristic force equals $f_c = k_B T \tau / b$. Contacts among different monomer types and different segment sizes introduce various spectra in binding energy. If there are several domains with different binding energies, several transitions should appear at different characteristic forces. In the setup of controlled displacement, several peaks appear with increasing peak values at certain displacements z_n^* . Transitions with a large δl and a small $\delta\varepsilon$ occur earlier. For consecutive n th and $(n+1)$ th transitions, the following inequality holds: $l_n/\varepsilon_n > l_{n+1}/\varepsilon_{n+1}$. However, due to the large thermal broadening for small ε , early transitions with small ε are not distinguishable unless δl is large enough [see Fig. 2(b)].

Figure 2(b) shows elastic responses of random chains constructed with different probability ϕ . At each large or small force drop, some part of the chain unfolds. The underlined segments in the caption of Fig. 2(b) show the parts of the chain that unfold easily at small force. For sequences (iv) and (v) ($\phi=0.4$ and 0.3) there are large domains which are clearly separated from each other (by underlined segments). The transition of sequence (iv) at $z=35$ is very pronounced because the associated unraveling segment length is large. Two domains become separated after this transition. All smaller transitions overlap due to thermal fluctuations. For sequence (iii) ($\phi=0.5$) most of the domains are small so the elastic response of these domains overlaps. Sequences (i) and (ii) consist of even smaller blocks. Therefore, the elastic response of these chains does not show the individual breaking of these small blocks.

Figure 2(c) shows the f - z curve and the fitting result using Eq. (4) (Ref. 12) with parameters $N_1=8.6$, $N_2=11.7$. The fitting around the second transition assumes that H monomers form a spherical globule when folded. The best fit predicts a binding energy of $\varepsilon=5.8k_B T$.

Despite the fact that the HP model does not guarantee a unique native structure, we have shown how the magnitude of disorder is reflected in the f - z curves. In the following section, using exact enumeration and MC simulations we explicitly calculate the weight function of each conformation characterized by extension z . We will study interaction types

beyond the HP model focusing on the difference between proteinlike sequences and random heteropolymers.

III. MONTE CARLO SIMULATIONS ON STRETCHING HETEROPOLYMER

A. Simulation technique

To gain further insight into f - z curves at thermodynamic equilibrium, we explore the phase space of all conformations $E_r(z)$ at given extension z . Each conformation of energy E_r contributes with statistical weight $\{\propto \exp[-E_r(z)/k_B T]\}$. Calculation of the f - z curves of a random heterocopolymer requires exploration of the rugged potential energy landscape. The equilibrium force-extension relation for the particular heteropolymer sequence can be obtained via the calculation of DOS $g(E, z)$ as a function of the energy E and extension z via exact enumeration. We have computed this quantity for polymers up to length 18 living on a three-dimensional cubic lattice of lattice constant b , for which exact enumeration of the conformation space is possible on a modern computer.³³ For example, exact enumeration of an 18-mer results in 78 955 042 017 configurations which takes approximately 35 h on a Intel Pentium IV 2.40 GHz PC running Linux.

Once the density of states $g(E)$ is known, ensemble averages that depend on E only [here $\langle A(E) \rangle$] can be computed at all temperatures using

$$\langle A \rangle = \frac{\int d\mathbf{r} A(\mathbf{r}) \exp[-\beta E(\mathbf{r})]}{\int d\mathbf{r} \exp[-\beta E(\mathbf{r})]} = \frac{\sum_E A(E) g(E) \exp[-\beta E]}{\sum_E g(E) \exp[-\beta E]} \quad (10)$$

in which $\beta = 1/(k_B T)$. The brackets $\langle \dots \rangle$ denote an average in the canonical ensemble. Other ensemble averages such as $\langle E \rangle$ can be computed in this way. The entropy S can be computed using $F = \langle E \rangle - TS$ in which the free energy F equals $-k_B T \ln \sum_E g(E) \exp[-\beta E]$.

To compute an ensemble average at given extension z , we have to compute the density of states as a two-dimensional function $g(E, z)$. Sampling of two-dimensional histograms is often computationally more demanding than one-dimensional histograms.¹⁶ Once $g(E, z)$ is obtained, for any applied force f and inverse temperature β , the average extension $\langle z \rangle$ can easily be computed using

$$\langle z \rangle = \frac{\sum_E \sum_z z g(E, z) \exp[-\beta(E - fz)]}{\sum_E \sum_z g(E, z) \exp[-\beta(E - fz)]}. \quad (11)$$

It is important to note that the density of states is symmetric, i.e., $g(E, z) = g(E, -z)$, which implies that $\langle z \rangle_{f=0} = 0$.

The force at a given extension z_0 can be computed by taking the derivative of the free energy F with respect to the extension, which we approximate for a lattice model by using

$$f = \left. \frac{dF(z)}{dz} \right|_{z=z_0} \approx \frac{1}{b} [F(z_0 + b) - F(z_0)], \quad (12)$$

where b is the lattice constant. The free energy $F(z_0)$ is given by

$$F(z_0) = -\frac{1}{\beta} \ln \left[\sum_E \sum_z \delta(z - z_0) g(E, z) \exp[-\beta E] \right]. \quad (13)$$

The specific heat C_v is the derivative of the average energy $\langle E \rangle$ with respect to the temperature T , and this is equal to the fluctuations of the energy E :

$$C_v = \left(\frac{\partial \langle E \rangle}{\partial T} \right)_V = \frac{[\langle E^2 \rangle - \langle E \rangle^2]}{k_B T^2}. \quad (14)$$

For chains with more than 18 monomers living on a cubic lattice, exact enumeration of configurations becomes computationally very expensive. Therefore, we have estimated the density of states using a flat histogram MC scheme first introduced by Wang and Landau.¹⁴⁻¹⁷ For $N=18$ we have checked that the density of states computed using this Monte Carlo scheme accurately reproduces the density of states obtained by exact enumeration of configurations.

We have investigated two types of model polymers all living on a cubic lattice. The bond length between two neighboring monomers is fixed by the lattice constant b . For all systems, only the nearest neighbors of a monomer contribute to the total energy E . However, interactions between consecutive monomers along the chain (monomers i and $i+1$) are excluded. In the HP model, the interaction energies are $\epsilon_{HP} = \epsilon_{PP} = 0$ and $\epsilon_{HH} = -1$ (Refs. 18 and 19) in units of $k_B T$. The interactions between monomers ϵ_{ij} is related to the second virial coefficient by $v_{ij} = b^3 \epsilon_{ij} / k_B T$.

The second model we investigate is the 20-letter amino acid interaction matrix from Ref. 20, Table VI. The amino acid sequences for this interaction matrix that we have investigated have the same unique, native structure. We have also simulated sequences with a G \bar{o} -type interaction potential.³⁴ In the G \bar{o} -model, $\epsilon_{ij} = -1$ only for those monomer pairs ij that are nearest neighbors in the predefined native structure. For all other monomer pairs as well as consecutive monomers, $\epsilon_{ij} = 0$. Although this model is conceptually very simple, it contains many aspects of protein folding.³⁵

The sequences of amino acids that we have investigated are listed in Table I. Sequences 1–12 are HP models, sequences 13–14 are G \bar{o} -models of different compact configurations. Sequences 15–18 are proteinlike sequences (15 monomers, 4 468 911 678 configurations) designed by Betancourt and Thirumalai using a 20-letter interaction matrix.³⁶ E_{\min} is the minimum energy of all configurations of a certain sequence. It turns out that only sequences 13–14 and prot1–prot4 have a unique native state.

B. Results of MC simulations

Density of states. Using the DOS method and exact enumeration, we explore the phase space of various sequences listed in Table I. This list consists of randomly created HP sequences, HP sequences with regular block structures, G \bar{o} -model sequences, and proteinlike sequences. By “protein-

TABLE I. Sequences used in this study. Sequence 1–12 are HP models, sequences 13–14 are Gō models of different compact configurations (18 monomers, 78 955 042 017 configurations). Sequences prot1–prot4 are proteinlike sequences (15 monomers, 4 468 911 678 configurations) designed by Betancourt and Thirumalai using a 20-letter interaction matrix (Ref. 36). E_{\min} is the minimum energy of a sequence. Only Gō-sequences 13–14 and proteinlike sequences prot1–prot4 have a unique native state. β_c is the inverse folding temperature.

Number	Sequence	E_{\min}	$\beta_c=1/k_B T_c$	Remark
1	HHHPPHHHHHHHPPPPPP	-7	1.885	Random HP, $\phi=0.3$
2	HPHHPPHHHHHPHHHHHP	-10	1.953	Random HP, $\phi=0.4$
3	HPRHPPHHHPHHHHHH	-10	2.057	Random HP, $\phi=0.5$
4	HPHHPPHHPPPPPPHPP	-7	2.430	Random HP, $\phi=0.6$
5	HHPHPPHHPPHHPPHPP	-9	2.141	Random HP, $\phi=0.7$
6	HHPHHPPHHPPHHPPHH	-10	2.014	Block HP
7	HHHPPHHHHPPHHHPPPP	-5	2.161	Block HP
8	HHHHHHPPPPPPHHHHHH	-10	1.835	Block HP
9	HHHHHHHHHHHHHHHHHH	-16	2.058	Homopolymer
10	PPPPPPPPPPPPPPPP	0	N.A.	Homopolymer
11	HHHPPHHHPPPPPPPPPPHHPHHHHHPHHH	-17	1.664	Dumbbell HP
12	HHHPPHHHPPPPPPPPPPPPHHPHHHHPHHH	-12	1.989	Dumbbell HP
13	Gō model (18-mer)	-16	1.663	Compact structure $2 \times 3 \times 3$
14	Gō model (18-mer)	-14	1.957	Compact helix
prot1	APSHNYRDNQQKDRC	-13.33	1.380	Proteinlike
prot2	VFSHGYYKGGQQGDKH	-18.60	1.015	Proteinlike
prot3	MPQRKIGDKGIDDAW	-48.98	0.396	Proteinlike
prot4	PCSEWKKDEMKVFC	-75.8725	0.260	Proteinlike

like,” we refer to those sequences which have a unique ground state conformation. The energy spectrum of HP sequences is measured in units of $k_B T$. For proteinlike sequences, the energy resolution of $0.0775k_B T$ is given by the interaction matrix.³⁶

As can be seen in Table I of Ref. 36, the folding temperature increases with the absolute value of the ground state of energy E_{\min} . All ground states of sequences prot1–prot4 are nondegenerate. The inverse folding temperatures (β_c) for proteinlike sequences are well defined and they are below $\beta=2.0$. It is difficult to define the “folding temperature” for the random HP sequences because the folded structure is not unique and the transition between a folded and an unfolded structure is not sharp. We determine the folding temperature (shown in Table I) by the temperature where the specific heat is maximum at zero force [see Fig. 4(a)]. For most of HP sequences, the inverse folding temperature is $\beta_c \approx 2.0$. Table I confirms the correlation between folding temperature and binding energy.

In Fig. 3 we plot the probability $\mathcal{P}(E) \propto g(E)\exp[-\beta E]$ that the system has an energy E at temperature $\beta=2.0$. Proteinlike sequences have a sharp maximum at the ground state which is separated from the higher energy states. Below the folding temperature, this implies a funnel-shaped potential landscape towards the native state. On the other hand, at the same temperature, random HP sequences are close to the transition point and their energy spectrum is broad. A maximum appears at larger value than ground state energy. At lower temperature, the maximum moves toward the ground state but the distribution remains relatively broad.

Force induced transitions and f - z curves. Applying an external force also induces structural transitions of heteropolymers. The unfolding transition occurs when the free energy difference between two conformations $\delta F = F_1 - F_2$ is comparable the elastic energy gain by unfolding, $f_c \delta z$.

Clearly, the peak of the specific heat $C_v(T)$ or $C_v(f)$ captures temperature-induced and force-induced phase transition, respectively. In Fig. 4, we show the C_v as a function of both temperature and force for a few representative sequences. The calculation of the specific heat below the folding temperature shows that the force induced transition is more dramatic than the temperature induced transition.

Various conformations of sequence 12 at extension $z = 0, 4, 8$ ($\beta=2.0$) are shown in Fig. 5. Although only one configuration is shown for each energy, we can see the loss of entropy for larger extensions where fewer conformations are allowed. The force-induced transition of sequence 12 occurs at the characteristic force $f_c = 0.3k_B T/b$ where $C_v(f)$ has the maximum [Fig. 4(b)]. At the first transition, the dominant conformation switches from a single globule structure to a dumbbell-like conformation.

The f - z curves obtained from two different set ups are shown in Figs. 6 and 7. The symbols represent the measured

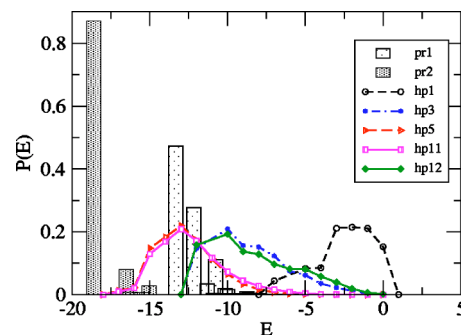


FIG. 3. The probability distribution $\mathcal{P}(E)$ for proteinlike sequences and random sequences at $\beta=2$. For the proteinlike sequences, the probability to be in the ground state energy is dominant and separated from other energy states. The distribution is broad for random sequences for which there is no such separation.

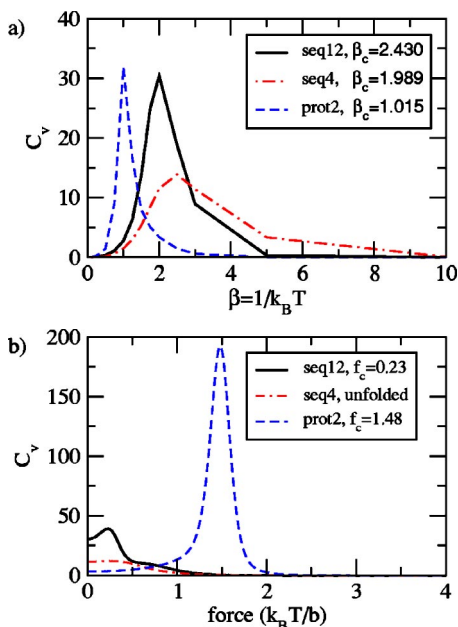


FIG. 4. (a) Specific heat C_v as a function of β . The unfolding temperature is displayed in the legend. (b) C_v during the force induced transition for several sequences; random (sequence 4), dumbbell (sequence 12), and protein-like (prot2) sequences at $\beta=2.0$. Sequence 12 and prot2 are at below the folding temperature and the unfolding force f_c is given in the legend. C_v is represented in units of k_B .

force values $\langle f(z) \rangle$ at the given extension z . The lines are the measured extension $\langle z(f) \rangle$ when the external force is given. The $\langle z(f) \rangle$ curve has the characteristic “plateaus” at the transition—i.e., an abrupt increase of extension at fixed force

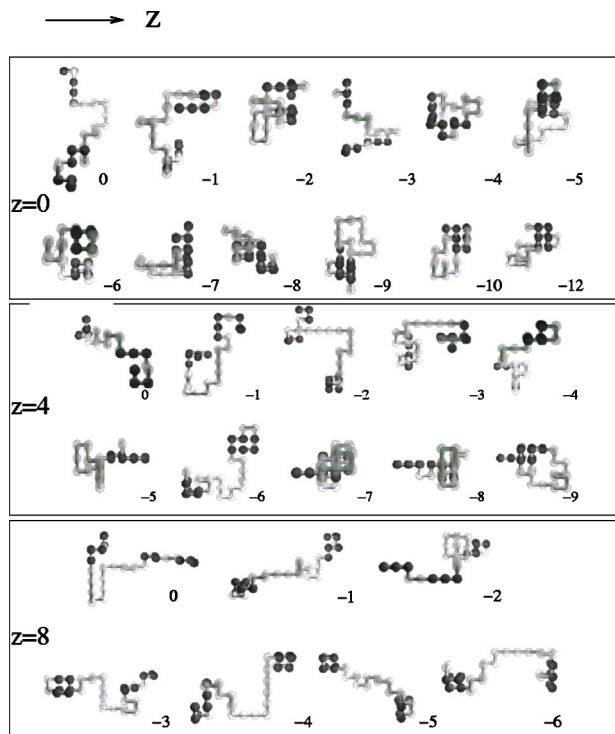


FIG. 5. Various conformations of sequence 12 with extensions $z=0,4,8$ around the folding transition temperature $\beta=2.0$. The energy is specified for each conformation. Dark spheres are hydrophobic units and half-transparent spheres are hydrophilic units.

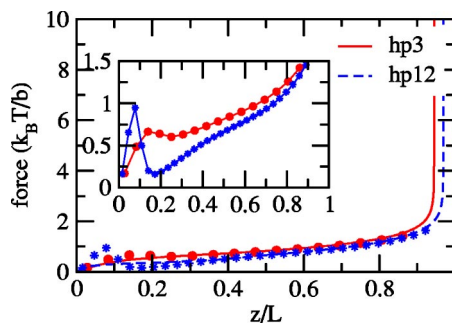


FIG. 6. Force-extension curves of sequences 3 and 12 below the folding temperature ($\beta=2.5$). Symbols stand for the measured force at a given extension [Eq. (12)] and lines represent the measured extension at a given force [Eq. (11)]. The small force regime is blown-up in the inset. The local maximum of sequence 12 depicts the unfolding transition from single globule conformation to a double globule conformation.

while $\langle f(z) \rangle$ is characterized by a “saw-tooth” shape consisting of a “peak” and a “force-drop.” Maxwell’s construction merges both approaches. In the following, we discuss the shape of $\langle f(z) \rangle$ in which we expect to see a saw-tooth shape at conformational transitions.

However, a saw-tooth shape hardly appears for short HP chains (sequence 1–10, $N=18$) even at considerably low temperatures $\beta \leq 2.5$. A random HP sequence normally contains several domains responding to different values of the characteristic forces and their elastic responses often appear to be continuous transitions (see, for example, sequence 3 in Fig. 6). A saw-tooth shape appears only when two domains are clearly separated by a loop (a large block consisting only of P-type monomers), for example, in sequences 11 and 12 (see Fig. 6). Here the distance between transitions is large enough compared to the overlap width $z^* k_B T / 2\epsilon$.

At very low temperatures (Fig. 7), small force-drops appear at the length scale of several monomer sizes. As we showed earlier, the resolution of the saw-tooth shape strongly depends on temperature. The domain size responding to a certain force is also a function of temperature. For a very low temperature ($k_B T < b\epsilon_n/z_n$) where thermal fluctuations are very small, the width of thermal broadening is only of the order of a few monomer sizes. Thus, the elastic response reveals the interaction energy at the monomer level (see Fig. 7). At higher temperature ($k_B T > b\epsilon_n/z_n$), as the overlapping region grows with temperature, force-drops at the monomer

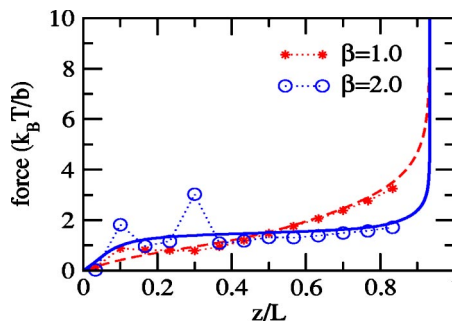


FIG. 7. Force-extension curves of prot2 below and above the folding transition temperature. Symbols are the measured force at a given extension [Eq. (12)] and lines are the measured extension at a given force [Eq. (11)].

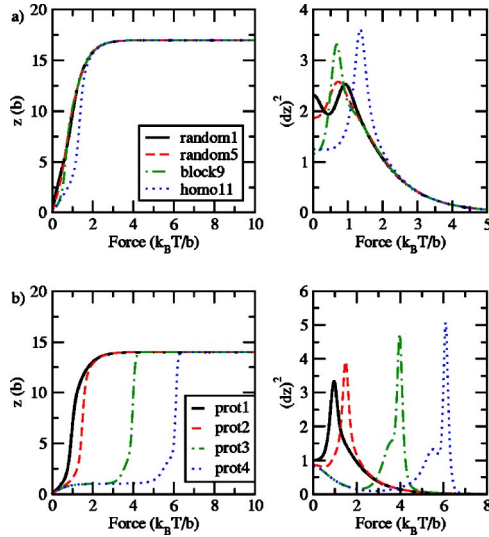


FIG. 8. (a) The z - f curves ($\langle z(f) \rangle$) and fluctuations of the extension $(\delta z)^2$ for HP polymers at $\beta=2.0$ and (b) for proteinlike sequences at $\beta=2.0$.

scale are no longer present. We observe a single sharp unfolding for proteinlike sequences and continuous smooth increase for random random sequences.

In the strong stretching limit, the restoring force is larger at high temperature (Fig. 7). This is because the restoring force at large extension is mainly due to the suppression of entropy upon stretching. For example, the external force to keep a Gaussian chain of length Nb at certain distance $z > \sqrt{Nb}$ is $f = 2k_B Tz/N$ and this increases linearly with temperature.

Finally, Figs. 8(a) and 8(b) demonstrate the difference in elastic responses between proteinlike sequences and random HP sequences. The extension $\langle z \rangle$ is measured for each given force. For proteinlike sequences, the transition in f - z curve is sharp and the fluctuations in z are also sharply peaked at transition and suppressed elsewhere. G \bar{o} -model sequences show similar behavior. The fluctuations in z show a maximum at the characteristic force for the transition. The fluctuations at zero force is already large for random sequences (sequences 1 and 5) which implies that there is no unique native structure. Fluctuations in other HP sequences are also slightly larger than proteinlike sequences.

IV. CONCLUSION

The main issue in this study was to show how different sequences with disorder respond to external forces.

We developed a generalized variational method to describe stretching experiments of heteropolymers with imposed end-to-end distance. This was realized by constraining the imposed end-to-end distance using Lagrangian multipliers. Within a Gaussian approximation, we can access all monomer-monomer correlations of a heteropolymer.

The force-extension curve is also evaluated by the exact enumeration of all conformation with a fixed extension. The obtained saw tooth patterns reveal the size of released blocks as well as the binding energy. The distance between the consecutive transitions and the magnitude of the force-drop provide additional insight into the primary structure of the het-

eropolymers. The resolution of saw-tooth patterns is determined by the interplay between thermal fluctuations, binding energies, and the domain sizes involved in the unfolding transition.

Additional insight into how heteropolymer unfolding depends on the sequence can be obtained from kinetic studies incorporating the pulling speed dependence of the force-extension curves. The hysteresis of these curves will provide a starting point for investigations of the energy barrier of the rugged potential landscape of heteropolymers.

ACKNOWLEDGMENT

E.J. and N.-K.L. acknowledge Sergei Obukhov (University of Florida) for helpful discussions.

APPENDIX: DERIVATION OF THE SELF-CONSISTENT EQUATIONS

The correlation function $R_\lambda(i, j)$ between two monomers i and j is defined by

$$R_\lambda(i, j) = G_\lambda(i, i) + G_\lambda(j, j) - G_\lambda(i, j) - G_\lambda(j, i), \quad (\text{A1})$$

where λ indicates both components of G , namely, parallel and perpendicular components. The square of the mean end-to-end distance $\langle R^2(1, N) \rangle$ is thus determined by the quantity $R_\parallel(1, N)$. We express the imposed distance z as

$$z = R_\parallel^{1/2}(1, N) = |\vec{r}_N - \vec{r}_1|. \quad (\text{A2})$$

The elastic energy term is only related to the parallel component of the correlation function. Straightforward calculations lead to the following expression for the variational free energy in units of $k_B T$:

$$\begin{aligned} F_V(G) = & -\frac{1}{2} [\text{Tr} \ln \hat{G}_\parallel + (d-1) \text{Tr} \ln \hat{G}_\perp] - \frac{d}{2} N^2 \\ & + \frac{d}{2b^2} \sum_{n=1}^{N-1} [R_\parallel(n, n+1) + R_\perp(n, n+1)] + V_2 + V_3 \\ & + \frac{l}{k_B T} [R_\parallel^{1/2}(1, N) - z], \end{aligned} \quad (\text{A3})$$

where V_2 and V_3 are the free energy corresponding to the two-body and three-body interactions, respectively:

$$\begin{aligned} V_2 = & \left(\frac{1}{2\pi} \right)^{d/2} \sum_{m \neq n} \sum_{n \neq m} \frac{v_{nm}}{2} R_\parallel(n, m)^{-1/2} R_\perp(n, m)^{-(d-1)/2} \\ V_3 = & \frac{w}{6} \left(\frac{1}{2\pi} \right)^d \sum_{k \neq n \neq m} \\ & \times R_\parallel(k, n)^{-1/2} R_\perp(k, n)^{-(d-1)/2} R_\parallel(n, m)^{-1/2} \\ & \times R_\perp(n, m)^{-(d-1)/2}. \end{aligned} \quad (\text{A4})$$

In the HP model, the interactions between different type of pairs are described by the second virial coefficients v_{HP} , v_{PP} , and v_{HH} .^{37,38} The Flory mixing parameter χ is related to the second virial coefficients by $\chi = v_{HP} - (v_{HH} + v_{PP})/2$. Here we use attractive interactions between different monomer types ($v_{PP} < v_{HP} < 0$). This allows us to “sense” the demixing transition of P monomers from H

monomers by the applied force. The correlation function $R_\lambda(i, j)$ represents the statistical average of the square of the mutual distance between the i th and j th monomers.

The next step is to compute the Euler equations by minimizing our expression for the free energy in terms of the propagator $G_\lambda(i, j)$,

$$\frac{\delta}{\delta G_\lambda(i, j)} F_V\{G\} = 0, \quad (\text{A5})$$

in which \hat{G}_λ stands for the $N \times N$ matrix with components $G_\lambda(i, j)$. Given that $\delta[\text{Tr} \ln \hat{G}] / \delta G_\lambda(i, j) = G_\lambda^{-1}(i, j)$, we obtain coupled equations for G_\perp and G_\parallel in the form of $N \times N$ matrix equations. For random copolymers with binary interaction v_{ij} , we obtain the following set of the self-consistent equations:

$$\begin{aligned} \frac{1}{2} G_\parallel^{-1}(i, j) = & C(i, j) + F2_\parallel(i, j) + F3_\parallel(i, j) + \frac{1}{2} \lambda R_\parallel^{-1/2}(1, N) \\ & \times (\delta_{i,1} \delta_{j,1} + \delta_{i,N} \delta_{j,N} - \delta_{i,1} \delta_{j,N} - \delta_{i,N} \delta_{j,1}) \end{aligned}$$

$$\frac{d-1}{2} G_\perp^{-1}(i, j) = (d-1)C(i, j) + F2_\perp(i, j) + F3_\perp(i, j), \quad (\text{A6})$$

where $F2_\lambda = (\partial / \partial G_\lambda) V_2(i, j)$ and $F3_\lambda = (\partial / \partial G_\lambda) V_3(i, j)$. The matrix $C(i, j)$ defines the connectivity between the monomers,

$$C(i, j) = \frac{1}{b^2} [cn(i) \delta_{ij} - \delta_{i,j+1} - \delta_{j,i+1}], \quad (\text{A7})$$

where $cn(i) = 1$ if $i = 1, N$ and $cn(i) = 2$ otherwise. Equations (A7) must be solved self-consistently together with the constraint condition Eq. (A2). Clearly, there is an analogy between l and the external force. In contrast to the constant force variational method, here l is not a given constant but a variable which needs to be calculated, i.e., it is the restoring force.

In this construction, the bond length between two consecutive monomers is not fixed. Neighboring monomers are connected by Gaussian springs with spring constant $3k_B T / b^2$. The distance between monomer pairs is determined by minimizing the total free energy.

We start with the initial trial Gaussian matrix as the connectivity matrix and the initial value of l as a constant. At each iteration, all components of $G_\lambda(i, j)$ are evaluated and in turn, we solve l using the standard Newton–Raphson method. We repeat the iteration until all values of G_{ij} converge within the accuracy of 10^{-8} . Each matrix satisfies the

sum rule $\sum_{i \neq j} G_\lambda(i, j) = -G_\lambda(i, i)$. We removed this degree of freedom (zero mode) by eliminating the first row and column.

By solving Eqs. (A7) we obtain correlation functions for all monomer pairs as well as the corresponding force at the imposed z . The resulting f - z curve is the elastic response in thermodynamic equilibrium.

- ¹M. Rief, F. Oesterhelt, B. Heymann, and H. Gau, *Science* **275**, 1295 (1997).
- ²M. Kellermayer, S. Smith, H. Granzier, and C. Bustamante, *Science* **276**, 1112 (1997).
- ³J. Liphardt, B. Onoa, S. Smith, I. Tinoco, Jr., and C. Bustamante, *Science* **292**, 733 (2001).
- ⁴B. J. Haupt, J. Ennis, and E. M. Sevick, *Langmuir* **15**, 3886 (1999).
- ⁵Y. Jiang, A. Lee, J. Chen, M. Cadene, B. T. Chait, and R. MacKinnon, *Nature (London)* **417**, 523 (2002).
- ⁶Y. Jiang, A. Lee, J. Chen, M. Cadene, B. T. Chait, and R. MacKinnon, *Nature (London)* **423**, 33 (2003).
- ⁷P. L. Geissler and E. I. Shakhnovich, *Phys. Rev. E* **65**, 056110 (2002); *Macromolecules* **35**, 4429 (2002).
- ⁸D. Bensimon, D. Dohmi, and M. Mezard, *Europhys. Lett.* **42**, 97 (1998).
- ⁹N.-K. Lee and T. A. Vilgis, *Europhys. Lett.* **57**, 817 (2002); *Eur. Phys. J. B* **28**, 451 (2002).
- ¹⁰O. Borisov and A. Halperin, *Eur. Phys. J. B* **9**, 251 (1999).
- ¹¹N.-K. Lee, A. Johner, and T. A. Vilgis, *Macromolecules* **35**, 6043 (2002).
- ¹²E. Jarkova, N.-K. Lee, and S. Obukhov, *Macromolecules* **38**, 2482 (2005).
- ¹³G. Migliorini, N.-K. Lee, V. Rostiasvili, and T. A. Vilgis, *Eur. Phys. J. E* **6**, 259 (2001).
- ¹⁴F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- ¹⁵F. Wang and D. P. Landau, *Phys. Rev. E* **64**, 056101 (2001).
- ¹⁶Q. L. Yan, R. Faller, and J. J. de Pablo, *J. Chem. Phys.* **116**, 8745 (2002).
- ¹⁷N. Rathore and J. J. de Pablo, *J. Chem. Phys.* **116**, 7225 (2002).
- ¹⁸H. Li, N. Tang, and C. Wingreen, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4987 (1998).
- ¹⁹K. Lau and K. A. Dill, *Macromolecules* **22**, 3986 (1989).
- ²⁰S. Miyazawa and R. L. Jernigan, *Macromolecules* **18**, 534 (1985).
- ²¹S. Ramanathan and E. Shakhnovich, *Phys. Rev. E* **50**, 1303 (1994).
- ²²D. K. Klimov and D. Thirumalai, *J. Phys. Chem. B* **105**, 6648 (2001).
- ²³D. K. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6166 (1999).
- ²⁴E. Paci and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6521 (2000).
- ²⁵H. Lu and K. Schulten, *Biophys. J.* **79**, 51 (2000).
- ²⁶M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986).
- ²⁷J.-P. Bouchaud, M. Mezard, G. Parisi, and J.-S. Yedida, *J. Phys. A* **24**, L1025 (1991).
- ²⁸J. des Cloizeaux, *J. Phys. (Paris)* **31**, 715 (1970).
- ²⁹B. Jönsson, C. Peterson, and B. Söderberg, *Phys. Rev. Lett.* **71**, 376 (1993).
- ³⁰B. Jönsson, C. Peterson, and B. Söderberg, *J. Chem. Phys.* **99**, 1251 (1995).
- ³¹R. R. Netz and H. Orland, *Eur. Phys. J. B* **8**, 81 (1999).
- ³²A. Halperin and E. B. Zhulina, *Europhys. Lett.* **15**, 417 (1991).
- ³³D. K. Klimov and D. Thirumalai, *Proteins: Struct., Funct., Genet.* **26**, 411 (1996).
- ³⁴N. Gō, *J. Stat. Phys.* **30**, 413 (1983).
- ³⁵S. Takada, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11698 (1999).
- ³⁶M. R. Betancourt and D. Thirumalai, *J. Phys. Chem. B* **106**, 599 (2002).
- ³⁷S. P. Obukhov, *J. Phys. A* **19**, 3655 (1986).
- ³⁸C. Sfatos, A. M. Gutin, and E. Shakhnovich, *Phys. Rev. E* **48**, 465 (1993).