Eur. Phys. J. B **28**, 451–465 (2002)
DOI: 10.1140/epjb/e2002-00248-0

THE EUROPEAN PHYSICAL JOURNAL B

Single chain force spectroscopy - Reading the sequence of HP protein models

N.-K. Lee[a] and T.A. Vilgis

Max-Planck-Institut für Polymerforschung, Ackermannweg 10, 55128 Mainz, Germany

Received 5 March 2002 / Received in final form 16 May 2002
Published online 13 August 2002 –

Abstract. We study the elastic properties of single A/B random copolymer chains, with a specific sequence and use them as theoretical model for so called HP proteins. HP proteins carry hydrophilic (P) and hydrophobic (H) monomers. We predict a rich structure in the force-extension relations which can be attributed to the information in the sequence. The variational method is used to probe local minima on the path of stretching and releasing for the chain molecules. At a given force, we find multiple configurations which are separated by energy barriers. A collapsed globular configuration consists of several domains which unravel cooperatively. Upon stretching, the unfolding path shows a stepwise pattern corresponding to the unfolding of each domain. While releasing, several cores can be created simultaneously in the middle of the chain, resulting in a different path of collapse. The long-range interactions and stiffness of the chain simplify the potential landscape given by the disorder in sequence.

PACS. 36.20.Fz Constitution (chains and sequences) – 87.15.Da Physical chemistry of solutions of biomolecules; condensed states – 87.15.-v Molecular biophysics

1 Introduction

Recent advances in nano manipulation allow direct access to single molecule elasticity using Atomic Force Microscopy (AFM) [1–8]. A series of these remarkable experiments show that the elastic response of a molecule is clearly related to the internal structure of the molecule. As a result of interplay between entropy, interactions between monomers and environments (such as solvents, other chains, external fields) the resulting force-extension ($f - z$) relations allow deep insight into the nature of the interactions and conformations [9–14]. In some polymer systems (including biopolymers and proteins), the intra-chain self assembly produces secondary or tertiary structures and the elastic responses reflect this structural hierarchy [12,13]. Furthermore, several single molecule experiments [1–3] and corresponding simulations [15–17] show that on application of force, there is domain by domain unfolding. Such mechanically induced folding/unfolding leads a single molecule to undergo structural transitions following well defined specific paths without being frustrated. Especially in the case of proteins there is considerable hope to gain insight into folding mechanisms and design.

Biomolecules such as DNA, RNA and proteins fall from into the general class of heteropolymers. The biological activity and function is strongly related and determined by the spatial conformation, *i.e.* its globular shape and internal structure. Naively speaking, biopolymers are active when they take specific conformations. Especially proteins are in biological function, when they are in a certain folded state. This folded state itself is mainly determined by the chemical sequence of the amino-acids, which interact locally with the environment (mainly physiological salt solution). The chain adapts its conformation to minimize the free energy with respect to the (quenched) sequence of the amino-acids. The sequence carries relevant information since according to the arrangement of the monomers certain conformations are more preferred than others. Recent studies show that biologically functioning molecules (protein-like molecule) can be distinguished from the random quenched heteropolymers in the folding mechanism. While random heteropolymers have various conformations with similar free energy values, and are therefore often trapped in local minima during the folding process, the protein-like molecules rapidly find the native structure. The folding efficiency suggests that the protein-like molecules have a well defined ground state as a native structure, separate from other conformations. These differences are displayed then naturally in the mechanical response. When external forces are imposed, the elastic properties of protein-like molecules have a sharp transition at a certain critical force f_c. On the other hand, other random sequences show many steps or smooth unfoldings as the external force increases.

There are two possible experimental and theoretical scenarios: either the displacement or the force are

[a] e-mail: lee@mpip-mainz.mpg.de

externally controlled [1,2,4–6,8]. When the displacement is imposed and the force is measured, the unfolding of the sequence of domains leads to repeating "saw tooth" pattern in the force-extension curve [1,2]. In constant force measurements, the folding-unfolding transition appears at a characteristic force, which is monitored as a "plateau" [8] in the force extension curves. In the following theoretical study the second possibility is chosen and the force is imposed whereas the distance is calculated. This choice has technical reasons, simply because the Legendre-transform of the thermodynamic free energy yields these results directly. Therefore, the predictions will provide the corresponding plateaus of the unfolding mechanism.

Here, we are going to provide a microscopic origin relating the force-extension curve and the sequence of a chain. We mainly address the special elastic properties of biopolymers to the heterogeneity of the constituents. Hence, we focus our study on the conformation of random copolymers subject to external forces. This corresponds to the simplest HP protein model where one only distinguish the polarities of amino-acids: hydrophobic and hydrophilic group [18,19]. In view of about twenty different amino-acids this seems to be an oversimplification. Nevertheless, the results given below provide already deep insight about the corresponding mechanisms of folding and unfolding of the chains.

In experiments of stretching of single biomolecules, the information on the sequence is determined and fixed by evolution. The relevant question is how the force-extension profile varies with respect to the differences in sequences. Recently, we investigated the elastic response of a single chain with specific sequences using variational methods with Gaussian trial function [20]. We minimized the free energy at given force (which corresponds to constant force measurement experiments). Minimizing the free energy for a given randomly generated sequence provides an estimate of the corresponding ground state or metastable states. Due to the disorder, some sequences show hysteresis in the stretching/releasing cycle. We predicted for certain sequences rich structures in the force-extension relations reflecting the information in the sequences. A collapsed globular phase consists of several domains and these unravel one by one at a corresponding critical force. Unfolding of each domain is captured as a "plateau". In this article we describe our work on random copolymer stretching in detail. Furthermore, we take into account various types of interactions including electrostatic interactions, as well as partial stiffness and structure of the chain.

Most of the usual theories on random copolymers use a so called disorder average, *i.e.* the free energy is calculated as a function of a certain (unknown) sequence [19,21–24]. Then this free energy is averaged over all possible sequences, by the use of the replica trick, well known for disordered systems [25]. Since the averaged force-extension neglects all specific profiles reflecting the special information of the sequences, these and similar methods are not very useful. Proteins (even if treated in the HP-simplification) respond to external forces according to their sequence (primary structure) which determine the secondary structure. Therefore we need to take into account the knowledge on the (randomly generated) sequence not the average. Then we may learn how to read in the sequence by the use of the single chain force spectroscopy experiment. This will also allow the study on the influence of local defects in sequences or additional stiffness effects.

This article is organized as follows. In Section 2, we formulate the variational method with external force which breaks the spherical symmetry. In Section 3, we show the result of the variational method when it is applied to a homopolymer globule in a poor solvent. In Section 4, we present the force-extension curve of random heteropolymers with short-range correlations under poor solvent conditions. This effect will smoothen out the sharp transition due to the unfolding of each independent domain as we will show. In Section 5, we introduce stiffness in the polymer segments locally which rules out certain configurations. The bending energy cost for a stiff segment is associated with its persistence length $\ell_{\rm p}$. The case of long range electrostatic interactions is discussed in Section 6. When long range interactions are dominant, domains become correlated.

2 Anisotropic variational method

Formally, we consider a polymer chain consisting of N monomers of size b under tension which is described by

$$H = H_{\rm c} + H_{\rm i} - |\boldsymbol{f} \cdot (\boldsymbol{r}_N - \boldsymbol{r}_1)|. \tag{1}$$

The first term corresponds to the elastic properties (connectivity) of the polymer chain and is at first approximated by the Wiener measure [26], whereas below in the paper local stiffness is taken into account at the simplest level. The second term contains 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 to collapse to a point, since depending on the solvent quality certain monomers may attract each other. The two body interactions can be attractive or repulsive depending on the type of the monomer pairs. The density of the collapsed globule (in a poor solvent) is then determined by the balance between three body repulsions and two body attractions. The last term describes the work done by the external force field $\boldsymbol{f}$, where $\boldsymbol{r}_i$ is the position of monomer i. The introduction of the external force in the Hamiltonian corresponds formally to a Legendre transform of the original problem. It is this step, which relates the results obtained from our theory to the constant force measurements.

The variational principle assumes a trial Hamiltonian H_0 with well known properties and makes use of Feynman's inequality

$$F \leq F_V \equiv \langle H - H_0 \rangle_0 + F_0 \tag{2}$$

where $\langle \cdots \rangle_0$ stands for the average over the variational probability distribution:

$$P_V(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) = Z_V^{-1} \exp\{-H_0(\boldsymbol{r}_1, .., \boldsymbol{r}_N)/k_{\rm B}T\}, \tag{3}$$

where Z_V is the normalization constant satisfying $\int P_V = 1$ and $F_0 = -k_{\rm B}T \log Z_V$.

The presence of an external force breaks the spherical symmetry. The deformation in the chain conformation due to the external force can no longer be described by a single order parameter. We need, therefore, to distinguish the deformation in the parallel and perpendicular directions with respect to the external force. In terms of two components of the correlation function, we choose consequently the trial Hamiltonian H_0 as follows:

$$H_0(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N)/k_{\rm B}T = \frac{1}{2}\sum_{j,l=1}^{N}[G_{\parallel}^{-1}(j,l)\boldsymbol{r}_{\parallel}{}^{j}\cdot\boldsymbol{r}_{\parallel}{}^{l} + G_{\perp}^{-1}(j,l)\boldsymbol{r}_{\perp}{}^{j}\cdot\boldsymbol{r}_{\perp}{}^{l}], \quad (4)$$

where $G_{\parallel}(j,l), G_{\perp}(j,l)$ are correlation functions between monomers j and l in the parallel and perpendicular directions with respect to the external force direction (see also [27]). By minimizing the trial free energy with respect to $G_{\perp}, G_{\parallel}$, we obtain two sets of self-consistent equations. These self-consistent equations for both components are coupled, in contrast to systems with spherical symmetry as studied previously [28–33].

The correlation functions $b_\lambda(i,j)$ between two monomers i and j are defined by

$$b_\lambda(i,j) = G_\lambda(i,i) + G_\lambda(j,j) - G_\lambda(i,j) - G_\lambda(j,i) \quad (5)$$

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 $b_{\parallel}(1,N)$. We identify the extension z as

$$z = b_{\parallel}^{1/2}(1,N). \quad (6)$$

We keep further mathematical details brief and defer detailed derivations and solutions of the self-consistent equation in Appendix A. We only mention here that this procedure corresponds to a self-consistent one-loop approximation [34].

3 Stretching a homopolymer in poor solvent

The simplest (theoretical) problem is to stretch a polymer in a poor solvent, where it takes a collapsed conformation. Moreover this case will serve as a crucial test for the quality of the variational calculation employed here. To begin with, we apply this variational principle to calculate the elastic response of a homopolymer globule and compare it to standard results, as this problem has been already studied theoretically (scaling [35], MC simulation [36]), and experimentally [37,38]. The use of poor solvent conditions is a basic input because the native structure of a protein has a hydrophobic core which may be mimicked crudely by a collapsed polymer chain in a poor solvent. To prepare the calculations on the model proteins, we study thus first the effect of an external force on a globular conformation. Indeed this simple problem can be addressed

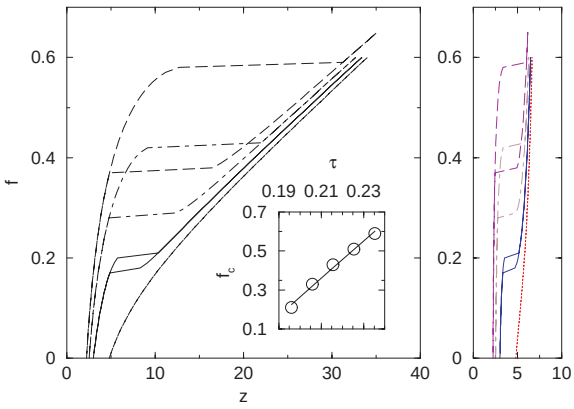

Fig. 1. Force-induced globule to coil transition of a homopolymer in a poor solvent. Each curve corresponds to the force-extension relation in different solvent conditions. From the bottom to the top, $\tau = 0.18, 0.2, 0.22, 0.24$. The transition occurs at the characteristic force $f_{\rm c} \sim \frac{k_{\rm B}T}{b}\tau$ which corresponds to the line tension as shown in the inset. The perpendicular deformation $b_{\perp}^{1/2}$ is shown in the right panel.

by scaling and geometrical arguments to reveal the essential length and force scales.

The relevant length scale for the deformation of a homopolymer globule immersed in poor solvent is determined to be the thermal blob size $\xi_{\rm T} = b/\tau$, where $\tau = |T-\theta|/\theta$. Each blob carries energy of $k_{\rm B}T$. When an external force f is applied, the globule starts to deform to a prolate ellipsoid. The eccentricity is increasing as the force increases. For a sufficiently small force f, the deformation is linear with respect to f. Then, the globule starts to unravel, until for sufficiently large forces, the chain of aligned thermal blobs transforms into a (stretched) string of Pincus blobs. These are of size $\xi_{\rm P} \simeq k_{\rm B}T/f$ [9]. At intermediate forces the stretching induces a first order phase transition involving a coexistence of a collapsed and stretched (thermal blob) chain conformation. A sharp structural transition from the globule to the extended conformation occurs when the Pincus blob size reaches to the thermal blob size $\xi_{\rm P} = \xi_{\rm T}$. Then, the external force is of the order of the line tension $(k_{\rm B}T/b)\tau$ as has been shown in reference [35].

Indeed, the picture developed above can be recovered by the variational technique used here. To see this, we minimize the free energy of a chain of length $N = 60$ in poor solvent condition at a given external force. We obtain the correlation function $b(1,N)$ of end monomers at each given condition. When the external force f reaches to the characteristic line tension $f_{\rm c} = k_{\rm B}T/\xi_{\rm T} \equiv \tau k_{\rm B}T/b$ the globule to coil transition (1st order) occurs and the extension $z = \sqrt{b_{\parallel}(1,N)}$ increases abruptly. This appears as a "plateau" in the force-extension curve. We observe structural transitions at different characteristic forces depending on the solvent conditions for $\tau > 0.19$. The shift of the onset of the globular phase is due to the finite size effect. Figure 1 shows that the poorer the solvent quality is, the larger force required for the structural transition. The linear relation between τ and $f_{\rm c}$ shown in inset of Figure 1 is

in agreement with the scaling picture. When we follow the local minima, we observe a hysteresis effect at the boundary of the first order phase transition. Around f_c, we access to local minima with different extensions. This means that the chain configuration becomes metastable and at the transition point, the stretched string and ellipsoidal conformations coexist. The transition also appears in the deformation in perpendicular direction $b_\perp$ (see Fig. 1). When $f < f_c$, the conformation is nearly isotropic, *i.e.* $b_\parallel \approx b_\perp$. when $f > f_c$, $b_\perp$ shows only little change, mainly due to the extension of the bond length.

4 Stretching a heteropolymer in a poor solvent

So far we dealt with homopolymers with uniform interactions between the monomers. In the next step of complication heteropolymer chains are investigated, which consist of more than one monomer type. We consider a random H/P polymer chain of length N, where we assume that the sequence is determined by a Poisson process. Therefore, the appearance of H and P monomers is completely random. The assumption of complete randomness appears not very satisfactory to model proteins. Nevertheless, we may find sequences which appear most useful and appropriate to discuss several properties of the chains. The interaction Hamiltonian is then simply given by

$$H_{\rm i} = \frac{1}{2!}\sum_{i=1}^{N}\sum_{j=1}^{N} v_{ij}\delta(\boldsymbol{r}_i - \boldsymbol{r}_j) + \frac{w}{3!}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\delta(\boldsymbol{r}_i - \boldsymbol{r}_k)\delta(\boldsymbol{r}_j - \boldsymbol{r}_k), \qquad (7)$$

where the two body interaction is chosen as a random $N \times N$ matrix which includes bond and site disorder simultaneously [21,22,24], *i.e.*,

$$v_{ij} = v_0 - \frac{1}{2}[\alpha(\sigma_i + \sigma_j) + \chi(\sigma_i\sigma_j)]. \qquad (8)$$

The sequence of monomers is described by a set of Ising spin-type variables σ_i with their values $\sigma_i = 1$ if monomer i is of type H (say, hydrophobic) and $\sigma_i = -1$ if it is of type P (hydrophilic, or polar). It is convenient to introduce a Flory like parameter $\chi = v_{\rm HP} - (v_{\rm HH} + v_{\rm PP})/2$ which is positive when similar monomers attract each other. The parameter α is given by the difference of equal interactions, $\alpha = \frac{1}{2}(v_{\rm PP} - v_{\rm HH})$ and is of the same order as χ (we choose $\alpha = \chi$). The negative constant part of the second virial, $v_0 < 0$, provides an average attraction between monomers. The third virial coefficient w is of the order of b^6 and provides a finite density of the globule.

To ensure that the thermodynamic ground state is always a globule, we assume that the average solvent quality is poor compared to the fluctuation due to randomness ($|v_0| > |\chi|$). This is, however, not a very restrictive condition as we will show in a subsequent publication. At first sight it may appear unphysical to force by purpose the polymers into globules, when these are discussed in the context of proteins. Nevertheless we will see on several examples, that the force-extension relations will be able to distinguish between naturally collapsed states and forced collapsed states. In forthcoming work this condition will be relaxed, and only naturally collapsed states will be discussed.

We consider in the following a heteropolymer of finite length N (= 40) at which disorder and the information of a specific choice of sequence plays an important role. As we mentioned before, we propose a theoretical model for protein stretching. Proteins are relatively short ($\sim 10^2$ amino-acids) compared to usual synthetic polymers and have biologically a "reasonable" ratio between total length and their sequences to satisfy requirements of their biological function. In long random chains ($N \to \infty$) polarity fluctuations associated with (quenched) randomness will be averaged out and the polymer behaves as a homopolymer in poor solvent. Nevertheless, segments under certain length scales in certain force regimes respond to stretching where the disorder remains important. This point will be discussed in the next subsection followed by scaling consideration.

4.1 Statistics

Before we discuss the force-extension properties for specific sequences in detail by making use of the variational principle, we discuss the averaged effect of randomness in the model for small and large disorder.

If disorder is small (small χ or large N), the structural transition occurs at large scale in more or less one single step (similar to a homopolymer in a poor solvent) (see Fig. 2). However, if disorder is sufficiently large (as desired in protein-type polymers), the collapsed globule consists of many domains which may respond to the given regimes of the external force separately. More precisely, in average, regardless of the size of disorder, the force for the structural transition is simply shifted by disorder by $\delta f_{\rm c} = k_{\rm B}T\alpha/b^4$. Although the mean net polarity of the polymers is zero (number of A monomers is equal to number of B monomers), a typical random copolymer of size N has excess polarity of order $\pm\sqrt{N}$ and there might be an excess polarity $\sum\sigma_i \neq 0$ for some sequences. The characteristic force $f_{\rm c}$ for each sequence is changing accordingly due to the shift in the effective solvent quality. The mean square fluctuation in $v_{\rm eff}$ is naturally $\sim \alpha/\sqrt{N}$. The correlation between $\delta f_{\rm c}$ and the net polarity is shown in Figure 2b. Most of data points are falling on straight line suggesting linear correlation. Compared to the linear term in equation (8), the contribution of the quadratic term ($\sum\sigma_i\sigma_j$) from equation (8) is minor ($\sim \chi/N$). This is in contrast to the presence of long ranged electrostatic interactions (which will be investigated below) where all pair-wise interactions are determined in terms of quadratic terms, due to the nature of the Coulomb interactions.

When disorder (due to the heterogeneity of the sequences) is large, on average, the structural transitions are

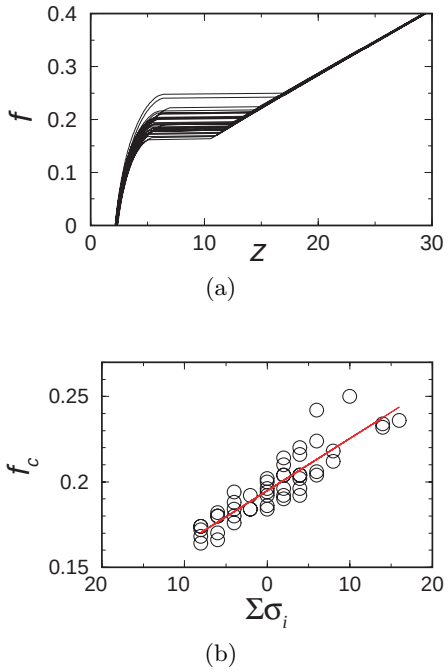

Fig. 2. The force-extension curves of random copolymers with 40 different sequences with small disorder $|\chi/2v_o| = 0.01$ (top panel). The chain length of each polymer is $N = 40$. The correlation between the charateristic forces f_c and the net polarity ($\sum \sigma_i$) for different sequences is depicted at the bottom panel.

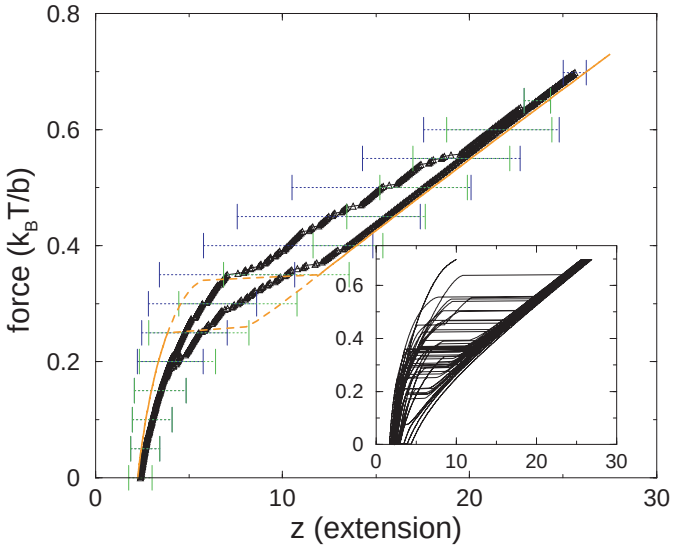

Fig. 3. The averaged force-extension curve of random copolymers over 40 different randomly created sequences with $|\chi/2v_o| = 0.25$. The size of error bars represent the variance in extension at a given external force which reflects the size of disorder. The individual trajectories are shown in the inset.

greatly smoothened (see Fig. 3). This corresponds to the case which Geissler and Shakhnovich considered recently using replica trick [39].

The variance in the characteristic force f_c reflects the size of the disorder, *via* the relation $\sim \alpha\sqrt{N}$. For small values of z ($\xi_P > \xi_T^{\text{largest}}$) a collapsed globule deforms to an ellipsoidal conformation. For large deformation z ($\xi_P < \xi_T^{\text{smallest}}$), the chain corresponds to a chain of thermal blobs which are linearly aligned without further degrees of freedom for rearrangements within the thermal blobs. The only relevant length scale is then given by the Pincus blobs. In the intermediate regime ($\xi_T^{\text{smallest}} < \xi_P < \xi_T^{\text{largest}}$), we expect fluctuations in the extension $z(f)$ to depend on system size and external force f. The intermediate blob sizes respond to the present external force according to the picture developed above.

A more interesting feature is that the unfolding procedure of individual chains shows unique elastic responses reflecting the information in the interactions along the sequence (as are shown in the inset of Fig. 3). Some of the force-extension curves are characterized by several discontinuous steps. In experimental situations, only specific sequences are accessible. Hence, in the following, we focus our study on the elastic response of specific sequences which are created randomly.

4.2 Response to forces by specific sequences

To understand the force-extension relations better we consider first the local structure inside a globular phase of a heteropolymer. Within this collapsed globular phase, any local sequence of "n" chosen monomers may have an excess polarity with typical fluctuations of size $\sqrt{n}$. This creates *e.g.* more hydrophobic (H type dominant) cores. Their sizes are determined by the local disorder. Their positions along the chain are naturally as well random. The resulting force-extension profile is diverse depending on the information each sequence carries. However, the hydrophobic parts of the heteropolymer form local globular phases which become opened upon applying external forces. Therefore we expect to a certain extent similar successive first order transitions as in the collapsed homopolymer [35] or on a pearl-necklace structure in polyelectrolytes in poor solvent [10,11].

Indeed, the appearance of such multiple plateaus indicates that the force-induced unfolding from the globular to the open-string structure occurs *via* intermediate structure formation. Each plateau in the force-extension relation corresponds to the unfolding of a part of the chain which cooperatively defines a "domain". The structural changes are achieved by abrupt unraveling of the corresponding domains. It is clearly visible that the shape and structure of the predictions depend on the sequence.

In Figure 4, we show typical examples for the force-extension relationships which show multiple steps for four different types of sequences with $|\chi/2v_0| = 0.5$. All sequences have a vanishing net polarity, *i.e.* the chains contain as many H as P type monomers. We have chosen these randomly created sequences as they appear most generic for some features presented. In the situation of curves a) and b), several H core-blocks are separated by

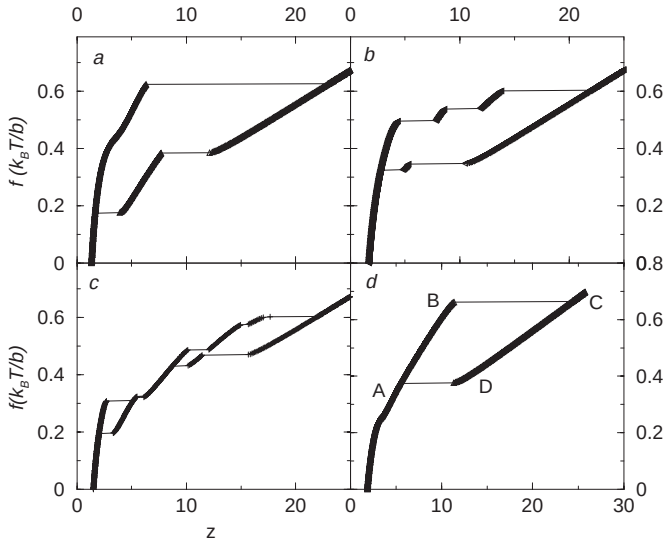

Fig. 4. Typical force-extension curves of random copolymers of $N = 40$ which show multiple steps for four different sequences. $|\chi/2v_0| = 0.5$. The hysteresis during stretch-release cycle is shown at each curve (stretching is always upper line). The sequence of each polymer is
a) HHHHP PPPHP PPPPP HHPPH PHPHH PHHPH PHHHP PHPHH
b) PHPHH PHPPP PPHHH HPPHP PHHHP HPPHH HPPHP PPHHH
c) HPHHP PHHHH HHHHP PHHPP PPHHP PPPHP HPPPP PHPHH
d) PHHPP PHPPH HHPPP HHHPH HPHPH HPPPP HPHHH PPHPH
a) and b) contain blocky structures separated by P-dominant segements, c) has one large H block and an net polarity +6 for the left half and −6 for the right half, d) corresponds to an alternating polymer.

relatively large sized P blocks. These sequences collapse only when the overall poor solvent condition is imposed and thus its core consists of several sub-cores. In consequence, the force-extension relation shows more plateaus. The sequence c) has one large H block and an net polarity +6 for the left half and −6 for the right half. The response to the external force is mainly from the smooth unfolding of the P dominant half chain while H dominant core remains. The curve d) in Figure 4 corresponds to the most homogeneous sequence and resembles an alternating multiblock copolymer sequence. Here, the overall poor solvent dominates, and the conformation is well described by a single globule. The core unfolds at a single characteristic value f_c. Indeed polymers with such sequences are interesting as models for proteins. As the overall hydrophobicity is the same for all the sequences presented, the differences in force-extension curves can be addressed to the local sequence.

In Figure 5, we show the corresponding free energy on the path of the stretching-releasing cycle for the sequences of b) and c) of Figure 4. At intermediate extension, we found more than one free energy minimum solution. Although it is clear that one of the solution has lower energy, it is not obvious to find the path to reach the global minimum solution. The upper curves indicate the existence of metastable states and closed loops indicate that they are separated by energy barriers. We will revisit this point in Section 4.4.

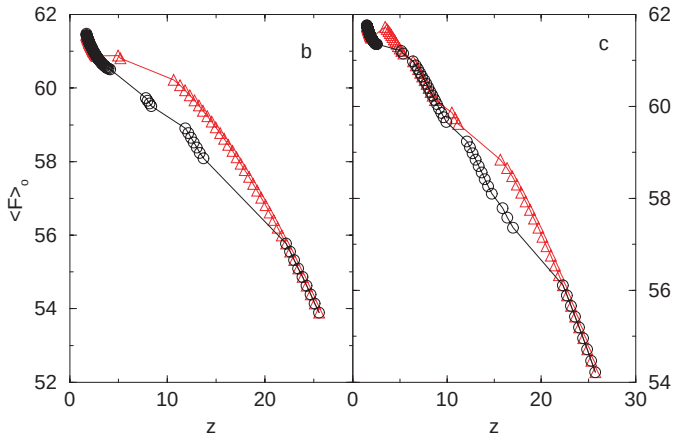

Fig. 5. Free energy of the sequence b and c are plot as a function of extension.

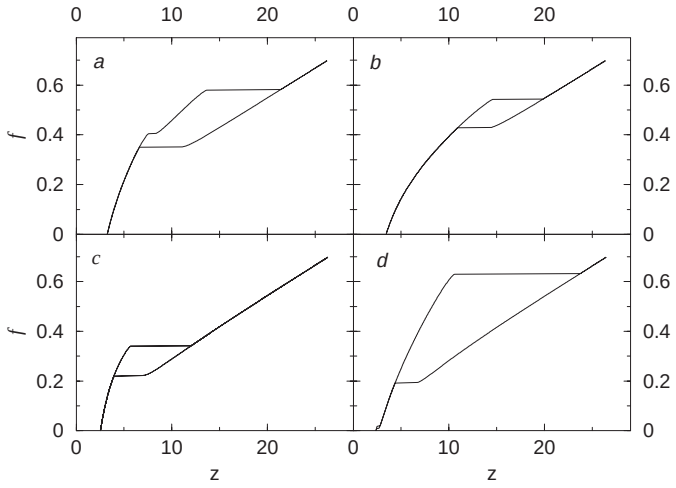

Fig. 6. Typical force-extension curves of random copolymers of $N = 40$ with net polarity of −6 of the hydrophilic monomers.
6-a) PPPPP PHHHP HHPHH PPHHP PHPHH PHPPH PPHHP PPPPH
6-b) HPPHH PHPPH HHPHH PPHPP HPHPP PPPHH PHPPH PHPPP
6-c) HPPPP PPHPP PPPHP HHHPH HHHPH PPHHP HPPHP PPHPH
6-d) PPPHH PPPPP PHHHH HPHPP PPPHP HPPHP PHHHP PHHHP.

In Figure 6 we show the force-extension curves for the sequences with non-vanishing polarity. It can be seen in general that the shape of the force extension curves is mainly determined by the hydrophobic globules. The set of sequences in Figure 6 include more hydrophilic monomers and the excess net polarity is −6. Therefore, the hysteresis loops created during the stretch-release cycle is reduced compared to the case of vanishing net polarity as indicated in Figure 4. The largest H-block of both sequence 6-a) and 6-b) is of size 3. The polymer which corresponds to curve 6-a) is also characterized by the "large" P blocks (5 units) at the beginning and the end of the chain. Between them the sequence is again more alternating. The curve in 6-b) does not contain these larger P blocks, thus the polymer is again dominated by one globule, which is

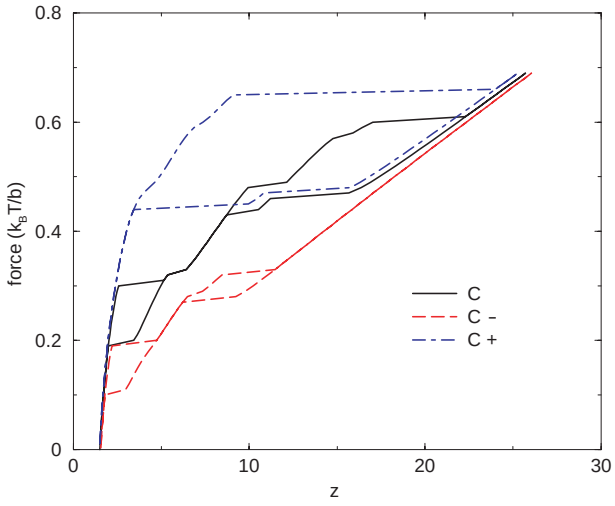

Fig. 7. The force-extension curve of the sequence c) and variations.
c) HPHHP PHHHH HHHHP PHHPP PPHHP PPPHP HPPPP PHPHH
c−) HPHHP PHH*PP* HHHHP PHHPP PPHHP PPPHP HPPPP PHPHH (−4)
c+) HPHHP PHHHH HHHHP PHHPP PPHHP PPPHP HPP*HH* PHPHH (+4).

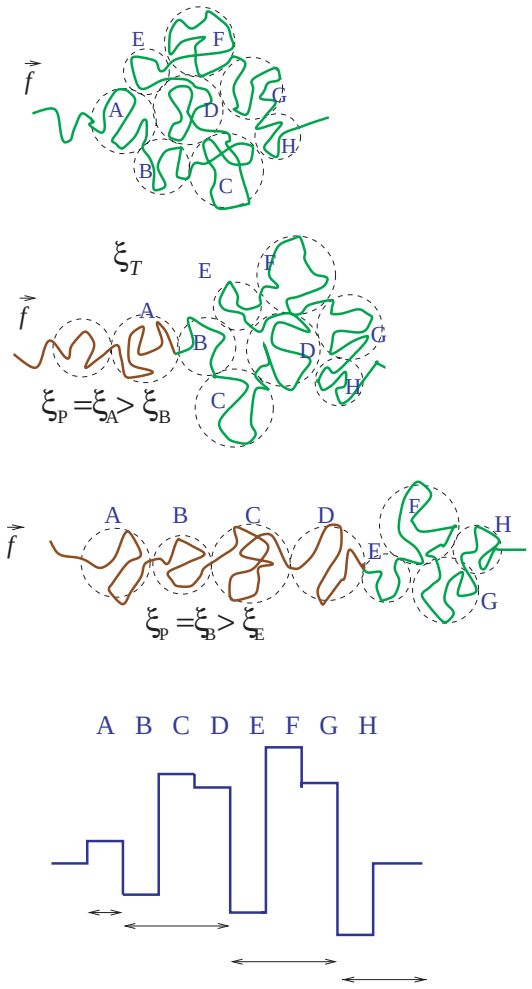

Fig. 8. Blob interpretation of stretching a random copolymer.

determined by the overall poor solvent. The sequence 6-c) is dominated by H blocks in the center and the beginning of the chain is dominated by P monomers (with a point defect of one H monomer). The plateau appears when the weaker part of the core unravels while the center core remains as a globule. The sequence 6-d) includes large sized alternating blocks (both H and P blocks) leading to large characteristic forces where the globule opens.

Figure 6 hints to the importance of "point defects", *i.e.*, simple changes in the sequence for example a replacement between hydrophilic and hydrophobic units. In order to see such effects in more detail in the force-extension shape we allow for such changes in the sequence on small scales. These "point" or "pair" defects will break up several local properties, and yield additional frustration in the local arrangement of the local globules. To see this in more detail we choose the rich sequence c) for Figure 4 as a reference and "perturb" the sequence at specific and characteristic places. The sequence c−) in Figure 7 is obtained by replacing the hydrophobic monomers (H) with hydrophilic monomers (P) at the underlined position. The hydrophobic core (8 consecutive "H"s) in sequence c) is weakened by these inserted hydrophilic monomers. The resulting force-extension curve shows that the critical force f_c is shifted to a smaller value. The largest core size is determined to 4. There are still two discrete unravelings visible. Each part of them is related to the opening of the corresponding domain. For the sequence c+) we replaced some hydrophilic monomers with hydrophobic monomers. Consequently, these monomers combine the two domains which unravel separately in sequence c) and c−). The shift of f_c to the larger value is obvious.

In summary of these results we would like to point out that the domain size which responds to the external force collectively is mainly determined by the distribution of the polarity of the monomers along the sequence. This can be made more clear by realizing that the disorder effect in spatial arrangement for an arbitrary chosen segment of n monomers (after averaging) decays faster ($\sim 1/n$) compared to the polarity of the local chain that decays with segment size as $\sim 1/\sqrt{n}$. Thus, the polarity determines the physical behavior on small (and relevant) scales which in turn determines the shape of the force-extension relation for intermediate forces $f > f_c$.

4.3 Scaling arguments

To get a physical understanding for the problem, we average over the disorder and assume a Gaussian distribution for the disorder part in the two body interaction. We obtain then the first order correction in the effective average second virial coefficient between arbitrary monomer pairs, $\overline{v_{\rm eff}} = v_0 - \alpha^2/\chi$. This effective interaction allows a discussion in terms of a blob picture.

The stretching process without relaxation is illustrated in Figure 8. The external force propagates from the ends to the center of the chain. The extension z is controlled by the smaller thermal blob size along the path of force

propagation. Varying thermal blob sizes ξ_i due to local disorder along the chain are expected. Each blob carries nevertheless a thermal energy $k_{\rm B}T$. Upon application of an external force each thermal blob unravels after another according to their size; *i.e.*, when the force becomes of the same order such that $\xi_{\rm P}(\equiv k_{\rm B}T/f) = \xi_i$ (i =A, B, C, ...). A force with a magnitude which is already sufficient to pull out blob B is also large enough to separate blobs C and D from the globular part, since $\xi_{\rm C}, \xi_{\rm D} < \xi_{\rm B}$. Therefore, as soon as blob B is released, there is no further increase of the force required to pull the blobs named C and D in Figure 8. This means that the B-C-D blobs respond to the external force cooperatively. In a similar way, E-F-G form another domain. After being pulled out, the size of each blob should be rescaled according to the Pincus idea, *i.e.*, its size must be adjusted to the size of the corresponding Pincus blob at the actual force.

More specifically, the thermal blobs vary here in size according to the $v_{\rm eff}$ which is determined by the local sequence, *i.e.* $\xi_{\rm T}^{\rm random} = b^4/v_{\rm eff}$. The average thermal blob size is $\xi_{\rm T}^{\rm random} = \xi_{\rm T}(1 - \alpha/|v_0|)$.

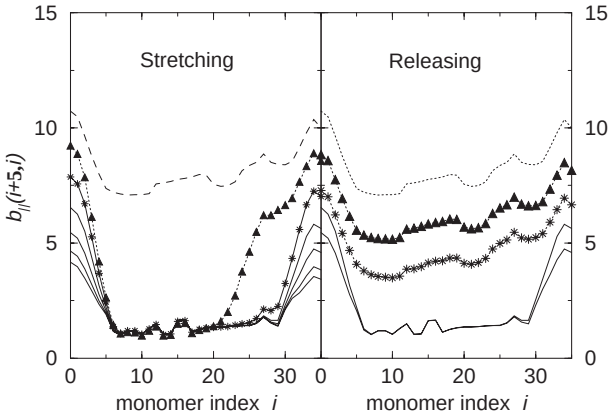

Fig. 9. The internal monomer-monomer correlation functions $b_{\|}(i+5, i)$ of sequence 6-a during the stretch-release cycle. The chain conformation choose different paths for folding and unfolding procedure. Force increases from the bottom to the top. Same symbols represent for the same force in left and right panel capturing the metastability.

4.4 Metastability, hysteresis and energy barriers

In most of the computed force-extension relations, despite that the curves are calculated for equilibrium behavior, *i.e.* for infinitely slow, quasi-static increase of the force, we observe characteristic hysteresis loops which indicate the existence of metastable states. Indeed it is clear that the pathways of stretching and decreasing the force do not have to follow the same path through the complex random free energy landscape due to the multiple interactions between the monomers along the chain. We thus probe the appearance of local minima by the path of increasing/decreasing the force and attribute them as metastable conformations.

These local minima are created by the quenched disorder which is given by the specific monomer positions. Obviously these are stable with respect to fluctuations, as it has been shown already in the early papers by Shaknovich [23]. An unfolding (folding) process of each domain from one phase to the other is always associated with energy barriers of certain heights. The origin of these energy barriers lies on the change of the surface energy of the globule when the monomers are redistributed between the two phases: a folded state (a globular conformation with a possible attached tail) and an unfolded state (a necklace-type conformation consisting of several hydrophobic cores). The unfolding transition resembles a first order phase type transition as discussed above, although the system size (where disorder is effective) is finite. Due to the quenched spin-glass type and heteropolymeric disorder, the chain can be found in various conformations in local energy minima. It is therefore useful to discuss the internal structure of the polymer in more detail. The variational technique proposed here is indeed appropriate to do so, since we do not employ a global variational parameter, but we carry the information of all chain segments in terms of the correlation functions $G(i, j)$ parallel and perpendicular to the force through the analysis. The set of these variational functions can be used to determine the internal structure as individual monomer-monomer correlations under tension. These informations are accessible in $G(i, j)$ as the correlation function $b_{\|}(i, i+d)$ of two monomers which are separated from each other at certain distance, for example $d = 5b$, along the contour.

Two different structures coexisting at a given force are observed. In Figure 9 the correlation function of a specific sequence (a of Fig. 6) is shown during the stretch-release cycle in the left and right panels, respectively. The regions of small values of $b_{\|}(i, j)$ represent the globular phase. As the force increases, a part of the chain is pulled out from one end leading to a tadpole-like conformation. Upon further increase of the force, the remainders in the globular phases are reduced. Then, the unfolding of the globule occurs in stepwise patterns, which have also been predicted for polyelectrolyte necklace chains [10,11]. Different configurations are visited while the chain releases. The internal correlation function for sequence 6-d) is shown in Figure 10. Here the globule unfolds in one step on stretching. The more hydrophobic parts assemble together at the beginning of the release, the necklace-like conformations are formed. These clusters are growing and finally merge to a single globule. Figure 9 also shows that upon force release the structural change from a necklace to a globular structure occurs at smaller characteristic forces than upon stretching.

The detailed theoretical analysis on the hysteresis is certainly beyond the aim of the present paper, since it would require a complete knowledge of the dynamical states of the heteropolymer. We will return to this point in a subsequent publication [40]. Here, we restrict ourselves to more qualitative remarks on the hysteresis problem. The hysteresis in force-extension curves is mainly determined by the energy barriers between two coexisting configurations following a certain force-induced path. The

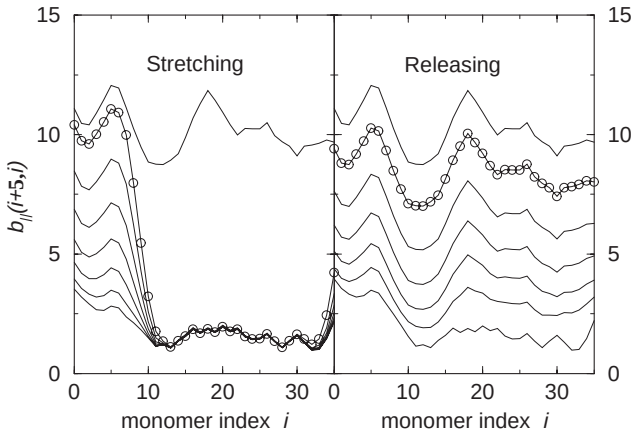

Fig. 10. The internal monomer-monomer correlation functions $b_{\|}(i+5,i)$ of sequence 6-d.

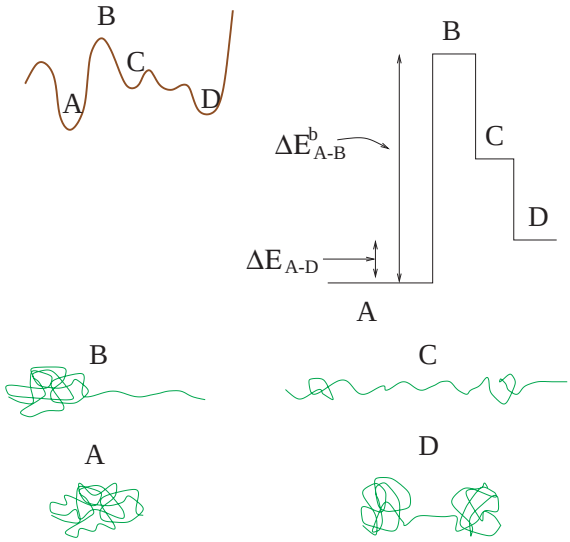

Fig. 11. The sketch of the potential landscape of a random copolymer under a constant force f_0.

hysteresis appears in experimental situations by pulling at finite speed [2] if the pulling speed is fast compared to the time for the relaxation, which depends exponentially on the energy barrier ΔE^b, *i.e.* $t_{\rm r} \sim \exp(\Delta E^b/k_{\rm B}T)$. In a quasi-static limit, the system can hop from one local (global) minimum to the other global (global) minimum after the relaxation time $t_{\rm r}$. In constant force measurements, the conformations in local minima are realized in turns for a finite fraction of time. It has been already seen that for pulling speeds faster than $t_{\rm r}$, the rearrangement of monomers at a given displacement is not immediately associated with the unfolding process of each domain [2]. Therefore, the path of hysteresis must be related to the dynamic properties, and has thus to be associated with the internal time and length scales of the local structures.

We should mention that variational calculation provides only the upper bound of the free energy but not of energy barriers. This is a drawback of the method that we can not describe kinetics – how the actual system will find the free energy minimum. Nevertheless, we observe the local minima appearing depending on the initial conformation and this leads to a hysteresis curve. This might be relevant to the actual experimentally measured hysteresis.

However, it is useful to estimate the energy barrier ΔE^b with a more pictorial argument. To do so we show that more than one conformation with different extension can be addressed to a constant external force in the present heteropolymer problem. In the force-extension curve, these regimes with metastability appear as hysteresis loops. Consider two states A and D which may correspond to the same constant force (see Fig. 11 for a qualitative illustration). The corresponding points are marked in the loop of the force-extension curve d in Figure 4. The conformation A can reach to D by the rearrangement of the monomers. This would require to overcome the energy barriers. When a constant force f_0 ($< f_1$) is applied, one possible route the chain can take in configuration space to transform from its states from A to D would be following the force induced loop, *via* conformation B. The corresponding energy barrier between the two states A–D (see Fig. 11) can be obtained from the force-extension curve, along the path A–B it is $\Delta E^{\rm b}_{\rm A-D} = \int_{\rm A}^{\rm B} f(z){\rm d}z$.

5 Stiffness

Up to now the polymers we investigated have been flexible chains. The chain connectivity was assumed to be completely Gaussian, thus all bare properties of the chains have been ruled by entropic properties only. The heteropolymers span a large number of possible states, since the high level of flexibility allows for a large number of conformational changes without paying a large entropy penalty. In the next step of complication, we overcome this problem to some extent. In addition to flexibility and interactions we introduce stiffness on a specific level. We assume stiffness in special sequences in such a way that only the hydrophilic segment carries stiff segments. In actual biological molecules, hydrophilic segments carry electrostatic charges. The net repulsive interactions between their segments give rise to a significant electrostatic contribution to stiffness and effectively stretch out the chain on local scales. We model such effects by addressing a local bending energy on sequences, whenever a certain number of hydrophilic monomers are sequentially arranged.

We assign the bending energy in any sequence in addition to the short range interactions (polarity) described by equation (8) if more than three consecutive hydrophilic monomers are found. The bending energy gives then rise to an additional control parameter, the persistence length ℓ_p. The corresponding bending term in the original Hamiltonian is then written (in discrete notation) by the second derivative of the chain vectors which corresponds to the total curvature of the stiffer part, *i.e.*,

$$H_{\rm bending}/k_BT = \frac{\ell_p}{b^3} \sum_{hydrophilic} (\boldsymbol{r}_{i+1} - 2\boldsymbol{r}_i + \boldsymbol{r}_{i-1})^2 \quad (9)$$

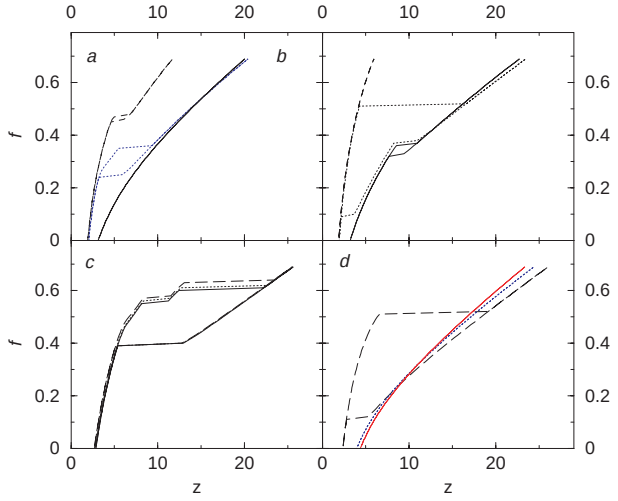

Fig. 12. The force-extension curves with stiff hydrophilic segments. The part of sequences with more than three consecutive "P" s becomes stiff. The dashed lines (top line for each curve) are corresponding force-extension curve for a flexible chain. The dotted lines (middle) are stiff segment with $\ell_p = 1b$ and the solid lines are with $\ell_p = 2b$. The numbers in parenthesis are the net polarity for each sequence.
s-a) HPHHH HPHHP HPPPH PHPPH HPHPH HHPHP PHPPH HPPHP(+2)
s-b) HPHHH HPPHP PPPHP PHPPP HHHPH PHPPP HHHHP HHHPP(0)
s-c) PPPPH HPPHH PPHPP HHPHP PHHPP PHHHP HPHHH HHPPP(−2)
s-d) PPPHP HPHPH HHPPH PPPPP HPHPP HPPHP HHPPH PHHHP(−6).

where ℓ_p is the persistence length related to the bending energy ϵ by $\ell_p = b\epsilon/k_{\rm B}T$. The corresponding matrix expression $\tilde{H}(i)_{\rm bending}$ for the stiff segments, defined for a three consecutive monomers i, $i-1$, $i+1$ is (in $k_{\rm B}T$ units),

$$\tilde{H}(i)_{\rm bending} = \frac{\ell_p}{b}\begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}. \qquad (10)$$

The main role of stiffness is to rule out certain configurations. Some conformations are no longer favorable due to the additional, and depending on the value of ℓ_p, high energy cost. We expect that some metastable conformations are excluded by stiffness. The force-strain relationship profiles in the force-extension curves which had their origin in the great flexibility and in entropy will disappear.

In Figure 12, we show the force-extension calculation of some sequences which include partially stiff segments. Depending on the position of stiff segments, the influence of stiffness on the chain conformation is quite varying. Whenever the stiff part is located at the end of chain, we cannot expect significant effects since the chain ends have anyway an additional degree of freedom. We see indeed, that such a location of the stiff segments does not change the chain conformation (for example the sequence s-c) dramatically. Most serious changes can be seen for cases when the stiff parts are located somewhere in the middle of the chain (s-a, s-b and s-d). Then, the stiff segments provide a significant exclusion of several conformations compared to the fully flexible case. The chain conformation with small external force is a globule. Some hydrophobic parts which are away from each other along the contour belong to the same globule to minimize the surface energy. When the segment in a chain becomes stiffer, holding the hydrophobic cores together costs corresponding bending energy. If the bending of the chain costs more energy than the energy gain by grouping the hydrophobic parts together, the cores prefer to be separated. For example, the sequence s-b in Figure 12 shows three clearly different paths in force-extension relations depending on the stiffness parameter. When $\ell_p = 0$, the chain does not unfold at all from the globular phase. When $\ell_p = 1b$, the unfolding (plateau) is observed as force increases. When $\ell_p = 2b$, we found the chain is not as compact as the flexible chain even without the presence of the force. The extension changes rather smoothly.

To get more physical insight let us consider first a simple case of a triblock copolymer consisting of two hydrophobic chain pieces, which contain $N_{\rm H}$-monomers each. Furthermore, we suppose that the connecting hydrophilic chain piece contains $N_{\rm P}$-monomers. We also assume that only the connecting hydrophilic segment has the above discussed property of stiffness, which can be described by the persistence length ℓ_p. The corresponding bending energy will be $\sim \int {\rm d}s\ell_p\kappa^2(s)$ with $\kappa(s)$, the local curvature. At a given external force, the chain conformation is then adjusted in such a way that it reduces the surface energy and elastic energy. The free energy of a globule with a dangling hydrophilic loop can then be estimated as in $k_{\rm B}T$ units,

$$F_{\rm globule} = \gamma(2N_{\rm A})^{2/3} + \frac{1}{2}\ell_p\int {\rm d}s\kappa^2(s). \qquad (11)$$

If we alternatively assume that hydrophilic segments are stretched out and two cores are separated, the conformation can be described as a dumb-bell whose free energy is written as:

$$F_{\rm dumb-bell} = 2\gamma(N_{\rm A})^{2/3} - fN_{\rm B}. \qquad (12)$$

The free energy difference between the two conformations is $\Delta F = \gamma(2-\sqrt{2})N_A^{2/3} - fN_B - E_{\rm bending}$. For a flexible chain where the persistence length is zero, $\ell_p = 0$, the two conformations become metastable whenever the work due to the external force becomes larger than the surface energy difference between the conformations. Under the condition $\Delta F = 0$, the elastic energy compensates the surface energy difference. The metastable state exists only if an external force satisfying the condition with finite persistence length

$$f = \gamma(2-\sqrt{2})N_{\rm A}^{2/3}/N_{\rm B} - E_{\rm bending}/N_{\rm B} \geq 0. \qquad (13)$$

When the bending energy exceeds a certain value, the single globule state is no longer accessible. In random copolymer systems, some part of the chain can be considered as

a triblock copolymer we considered above. Some metastable conformations become unavailable by adding stiffness to the segments. The hysteresis loops in force-extension curve disappear as the persistence length increases. The number of metastable states decreases and this can be indeed attributed to "forbidden" states, whenever the bending energy exceeds a certain value.

For a flexible chain, we showed that point defects on chain sequence alter the elastic responses. Starting from the sequence s-b) we perturb the part (PPPP) with a point defect by changing one P-unit to one H-unit (PPHP) which destroys the local stiffness. The first half of the sequence becomes flexible. In this case the point defect changes not only the overall net polarity but also the flexibility considerably. The force-extension curve of the sequence with point defect with $\ell_p = 1b$ goes back to the same curve as the flexible chain.

6 Long range interactions

So far we dealt with short range interactions between the monomers to illustrate the principal problems of heteropolymers under external forces. For biological systems these type of interactions are far from being sufficient and long range interactions, such as Coulomb potentials must be added. These will, however, change the energetic contributions to the free energy more than only by their long range nature, because for the electrostatic interaction the site disorder term $\sigma_i\sigma_j$ contribution becomes more relevant. Indeed the variable σ_i corresponds to the electrostatic charge at monomer i.

First, we discuss what we can expect for these situations before we turn to more detailed calculations. In the presence of the long range interactions, the characteristic for a f_c for the structural transition is related to the system size N as well as disorder. For a polyelectrolyte in a poor solvent, interaction is uniform through the chain. Therefore, we expect the structural change to occur through a single-step transition where all chain parts respond together. However, a net charge in the chain provides overall repulsive interactions which makes the globular phase unstable. The resulting conformation of a polyelectrolyte chain in a poor solvent is like a necklace [41] consisting of pearls and connecting strings. The elastic response of a polyelectrolyte necklace shows the step-wise pattern as the number of pearls reduces as the external force increases [10,11].

Here, we consider the simplest case of structural transitions induced by external forces, *i.e.*, from the globular conformation to the 2-pearl-necklace conformation. We assume that charges are uniformly distributed along the polyelectrolyte and the fraction of the charged monomer is ϕ. The total net charge of the polyelectrolyte is $Q = \phi N$. Corresponding Hamiltonian is

$$\frac{H_{\rm Coulomb}}{k_{\rm B}T} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{u_{ij}}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{14}$$

where u_{ij} is the electrostatic interaction strength $u_{ij} = u \equiv l_B\phi^2$ for all pairs of i and j monomers if the chain is uniformly charged with charge fraction ϕ with $l_{\rm B} \equiv e^2/\epsilon k_{\rm B}T$ as the Bjerrum length.

If $Q < Q_{\rm c}$, the overall repulsive interaction is smaller than the surface energy, therefore the chain remains as a globule. Without presence of the electrostatic interactions, in order to stretch the globule, the external force should overcome the line tension of each thermal blob $f_{\rm c} = k_{\rm B}T\tau/b$ which leads to the relation $f_{\rm c}^{HP} = k_{\rm B}T\tau/b$ independent of chain length. The repulsive electrostatic interaction shifts the force required for the structural change to the extended configuration by $\sim Q^2/R^2$

$$f_{\rm c}^{\rm PE} = k_{\rm B}T\left(\frac{\tau}{b} - l_{\rm B}\phi^2\frac{N^2}{R^2}\right) \sim \left(f_c^{HP} - \frac{k_{\rm B}T}{b^2}l_B\phi^2N^{4/3}\right). \tag{15}$$

Since the chain is collapsed, we have $R \sim N^{1/3}$, and the shift in characteristic force due to the electrostatic interactions increases according to the chain size $\sim N^{4/3}$.

In the case of random polyampholytes, we have the interaction $u_{ij} = l_Bq_iq_j$ randomized according to the random charge distribution $q_i = \pm 1$.

Again we impose overall poor solvent conditions first to ensure globular conformations at all net charges, second to compare the cases here with those discussed already above.

In contrast to the case of short range interactions the domains created by randomness become correlated due to the long range interactions between charges and the effective charges of the globular parts. These additional correlations will smoothen out the sharp transitions due to the unfolding of formally independent domains observed above. When all globular regions are unfolded, the characteristic force captures the overall size of the chain. H typical randomly created polyampholyte chain carries a net charge of the order $Q \sim \pm\sqrt{N}$, and the size of compact globule is $R \sim N^{1/3}$. In the globular state, the surface tension and repulsive electrostatic interactions are competing. The transition to the coil state occurs at a force smaller than the characteristic force of the homopolymer transition. The force-balance equation gives the characteristic force for PA transition in a poor solvent as:

$$f_{\rm c}^{\rm PA} = k_{\rm B}T\left(\frac{\tau}{b} - l_{\rm B}\frac{Q^2}{R^2}\right) \sim \left(f_c^{HP} - k_{\rm B}T\frac{l_B}{b^2}N^{1/3}\right). \tag{16}$$

The shift of the characteristic force scales therefore with overall chain size with $N^{1/3}$.

In order to evaluate the force-extension curve for the PA with specific sequence of electrostatic charges, we calculate corresponding variational free energy:

$$\frac{F_V^{\rm Coulomb}}{k_{\rm B}T} = \sum_{i,j}\frac{\tanh^{-1}(u)}{\sqrt{\pi b_\parallel^{1/2}(i,j)u}} \tag{17}$$

where $u = \sqrt{1 - b_\perp/b_\parallel}$. We assume that $b_\parallel \geq b_\perp$ for all monomer pairs. It is worthwhile to mention the difference

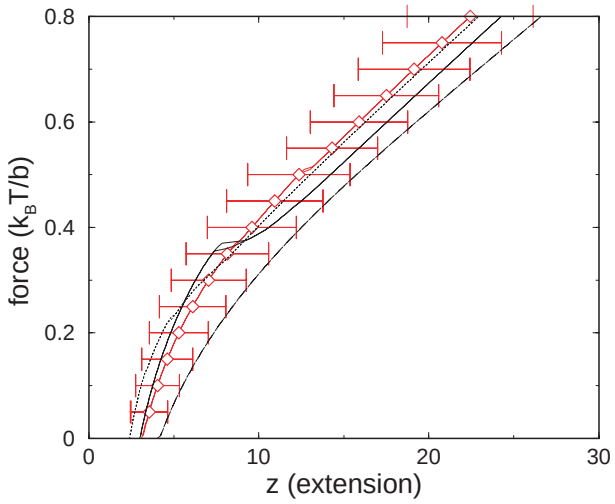

Fig. 13. Force-extension curves of polyampholytes averaged over 40 independent sequences of chain length $N = 40$ (symbols). Error bars indicate the size of disorder. The dotted line is for the sequence with net charge -4, solid line is with vanishing net charge and the dashed line is for relatively large net charge (+6).

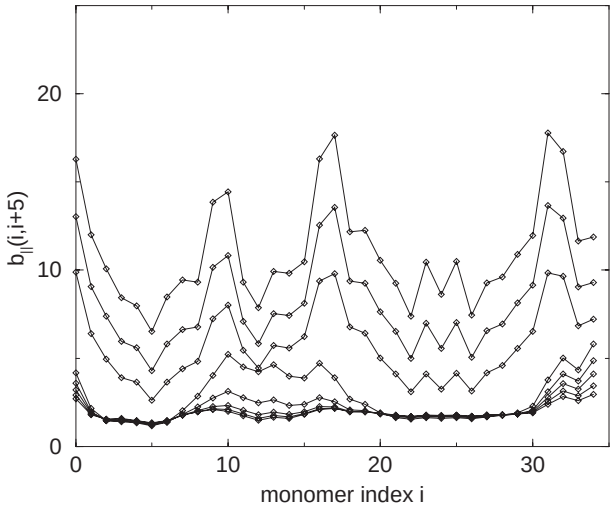

Fig. 14. The internal correlation function of a polyampholyte chain under the influence of external force with charge sequence:
$--+-++--+-\ \ +++--+--++$
$---+-+-+-+\ \ --+---++++$.
The net charge is zero. The force is increasing from the bottom to the top with interval $0.05k_BT/b$.

in the choice of trial Hamiltonian H_o from that of Netz and Orland [27] where the classical path is explicitly included in the trial Hamiltonian in order to describe the spontaneous symmetry breaking in PE. This is not necessary in our system where the external force direction is given. There is no ambiguity in choosing the direction of the symmetry breaking. As we minimize the free energy under the external force, the corresponding shift appears in $b_{\|}$ (see also Appendix B).

The force-extension curve for a typical PA sequence is shown in Figure 13. The hysteresis loop, which is an indication of the metastability, disappear for most of sequences. The correlation function for internal structure (Fig. 14) clearly shows that domains are correlated. For random heteropolymers one domain unfolds independent of the others. The external force unfolds a local structure. In the presence of electrostatic interaction, the large scale structure responds to the external force simultaneously. To give an example, we plot in Figure 14 the correlation function of PA. Without external force, we observe that some parts are more compact than the others. As the external force increases, the globule unfolds into an open structure. The transition is then global.

7 Conclusion

We investigated the elastic properties of a single chain with arbitrary sequences by using a variational method. The main feature of the method is that information on the monomer correlations can be found on different levels. The method allows the computation of monomer-monomer correlations which reveals, apart from the global scaling, the structure of the chain. Thus it was possible to calculate the internal structure of the chains under different circumstances. We note that the accuracy of the method is tested by comparison with Monte Carlo simulation [42].

The essential property which governs the force-extension curve is domain-wise unfolding from globular to open-string conformation. The characteristic force related to each domain shows itself as a "plateau" in the force-extension curve. The main issue here was to show how different sequences respond to external forces. It was shown however, that special types of sequences can be reflected to the force-strain measurements. It became clear that "protein-like" sequence need structured units of monomers but also unstructured units, *e.g.*, largely alternating properties, which prevents the chain of being trapped in a metastable conformation.

We also investigated the case that monomers are interacting with long range interactions. The disorder effect in electrostatic charge sequences are compensated by the long range correlations. When long range interactions are dominant, domains begin to be correlated. This smoothes out the sharp transition due to the unfolding of formally independent domains. It was shown that the correlation *via* the long range potential is essential to rule out several conformations with respect to the short range case. Indeed, the charges and the long range nature of their interactions will drive the polymer into special conformations. The number of states will be less as compared to the short range case. We may state that "protein-like" sequences need electrostatic interactions to rule out many of the physically accessible conformations of random polymers. As in spin glasses there exist $\mathcal{O} \sim e^N$ pure states, whereas proteins have only a few (one in the ideal case).

The influence of partially stiff segments also simplifies the force-extension profile to a certain extent. More

realistic proteins could be also treated in a similar way by assigning special sequences using a binary interaction matrix and by adding more types of monomers which may correspond to the amino-acids.

Of course, we must discuss in more detail the internal structure of the molecules considered in this paper. In a forthcoming work we will provide additional techniques to determine the structure and conformations of the polymers. One direction we want to go is to calculate the effective interaction potentials between the monomers. This is important as the bare interactions (given by the bare potentials v_{ij} and w) will be modified significantly by entropic contributions. These entropic contributions are determined by the local stiffness, the connectivity and the chain length. As we have shown earlier [43] these may also provide insight into the random part of the free energy in the globular phase. Thus, we will be able to provide relevant information on the phase space in globular phases. Moreover, we also hope that the knowledge of the effective potentials will be helpful to select between "protein-like" sequences and "random" sequences.

We thank V. Rostiashvili, S.P. Obukhov and A. Johner for thoughtful discussions. N.L. would like to thank T.J.H. Vlugt for sharing his MC simulation results before the publication. N.L. acknowledges the financial support by the Schwerpunktprogramm Polyelectrolyte of the German Science Foundation, DFG.

Appendix A: Derivation of the self consistent equations

In this appendix we derive the self-consistent equations for describing the chain under tension. The principle of the variational method makes use of the Gibbs Bogoliubov inequality,

$$F \leq F_V = \langle H - H_0 \rangle_0 + F_0 \tag{18}$$

or

$$F \leq F_V = \langle H \rangle_0 - TS_0. \tag{19}$$

We include here contributions corresponding to two and three body interactions. The total Hamiltonian of the system consists of the following terms.

$$\begin{aligned} H_{\text{connectivity}}/k_BT &= \frac{d}{2b^2}\sum_{i=1}^{N}(\boldsymbol{r}_{i+1}-\boldsymbol{r}_i)^2, \\ H_{\text{interaction}}/k_BT &= \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{v_{ij}}{2!}\delta(\boldsymbol{r}_i-\boldsymbol{r}_j) \\ &\quad +\frac{w}{3!}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\delta(\boldsymbol{r}_i-\boldsymbol{r}_j)\delta(\boldsymbol{r}_j-\boldsymbol{r}_k), \\ H_{\text{f}} &= -|\boldsymbol{f}\cdot(\boldsymbol{r}_N-\boldsymbol{r}_1)| \end{aligned} \tag{20}$$

and the trial Hamiltonian is chosen to be

$$\frac{H_0(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N)}{k_BT} = \frac{1}{2}\sum_{j,l=1}^{N}\left[G_{\parallel}^{-1}(j,l)\boldsymbol{r}_{\parallel}{}^{j}\cdot\boldsymbol{r}_{\parallel}{}^{l}+G_{\perp}^{-1}(j,l)\boldsymbol{r}_{\perp}{}^{j}\cdot\boldsymbol{r}_{\perp}{}^{l}\right] \tag{21}$$

$G_{\parallel}(j,l), G_{\perp}(j,l)$ are the correlation functions between j and l monomers in parallel and perpendicular direction with respect to the external force $\boldsymbol{f}$.

Straightforward calculations lead to the following result for the variational free energy in $k_{\text{B}}T$ units:

$$\begin{aligned} F_V(G) = &-\frac{1}{2}[\mathrm{Tr}\log\hat{G}_{\parallel}+(d-1)\mathrm{Tr}\log\hat{G_{\perp}}]-\frac{d}{2}N \\ &+\frac{d}{2b^2}\sum_{n=1}^{N-1}[b_{\parallel}(n,n+1)+(d-1)b_{\perp}(n,n+1)] \\ &+\left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\sum_{m}\sum_{n\neq m}\frac{v_{nm}}{2}b_{\parallel}(n,m)^{-\frac{1}{2}}b_{\perp}(n,m)^{-\frac{d-1}{2}} \\ &+\frac{w}{6}\left(\frac{1}{2\pi}\right)^{d}\sum_{k\neq n\neq m}b_{\parallel}(k,n)^{-\frac{1}{2}} \\ &\times b_{\perp}(k,n)^{-\frac{d-1}{2}}b_{\parallel}(n,m)^{-\frac{1}{2}}b_{\perp}(n,m)^{-\frac{d-1}{2}} \\ &-\frac{|\boldsymbol{f}|}{k_{\text{B}}T}b_{\parallel}^{\frac{1}{2}}(1,N) \end{aligned} \tag{22}$$

where $\hat{G_{\lambda}}$ stands for the $N\times N$ matrix with components $G_{\lambda}(i,j)$. The correlation functions $b(i,j)_{\lambda}$ are defined in terms of G_{λ} in equation (5) and the extension in the direction of the external force is

$$z = \sqrt{b_{\parallel}(1,N)}. \tag{23}$$

The elastic energy term is only related to the parallel component of the correlation function. We proceed to compute the Euler equations by minimizing the above expression for the free energy in terms of the propagator $G_{\lambda}(i,j)$.

$$\frac{\delta}{\delta G_{\lambda}(i,j)}F_V\{G\} = 0. \tag{24}$$

Given that $\delta\left[\mathrm{Tr}\log\hat{G}\right]/\delta\, G_{\lambda}(i,j) = G^{-1}{}_{\lambda}(i,j)$, we obtain coupled equations for $G_{\perp}$ and $G_{\parallel}$ in the form of $N\times N$ matrix equations respectively. These equations should be solved self consistently.

The matrix $C(i,j)$ defines the connectivity between the monomers.

$$C(i,j) = \frac{1}{b^2}[cn(i)\delta_{ij}-\delta_{i,j+1}-\delta_{j,i+1}] \tag{25}$$

where $cn(i) = 1$ if $i = 1, N$ and $cn(i) = 2$ for rest of the monomers.

For random copolymer with binary interaction $v_{i,j}$, we obtain the following set of the self-consistent equations,

$$\begin{aligned} \frac{1}{2}G_{\parallel}^{-1} &= C(i,j)+F2_{\parallel}(i,j)+F3_{\parallel}(i,j)+\frac{1}{2}fb_{\parallel}^{-\frac{1}{2}}(1,N) \\ \frac{d-1}{2}G_{\perp}^{-1} &= (d-1)C(i,j)+F2_{\perp}(i,j)+F3_{\perp}(i,j), \end{aligned} \tag{26}$$

where $F2$ and $F3$ defined as:

$$F2_{\parallel} = \frac{\partial}{\partial G_{\parallel}} V2_{\parallel}(i,j) =$$
$$\frac{v_{ij}}{2}\left(\frac{1}{2\pi}\right)^{\frac{d}{2}} \begin{cases} b_{\parallel}(i,j)^{-\frac{3}{2}} b_{\perp}(i,j)^{-\frac{d-1}{2}} & i \neq j \\ -\sum_n b_{\parallel}(i,n)^{-\frac{3}{2}} b_{\perp}(i,n)^{-\frac{d-1}{2}} & i = j \end{cases}$$

$$F2_{\perp} = \frac{\partial}{\partial G_{\perp}} V2_{\perp}(i,j) = \frac{v_{ij}(d-1)}{2}$$
$$\times \left(\frac{1}{2\pi}\right)^{\frac{d}{2}} \begin{cases} b_{\parallel}(i,j)^{-\frac{1}{2}} b_{\perp}(i,j)^{-\frac{d-1}{2}-1} & i \neq j \\ -\sum_n b_{\parallel}(i,n)^{-\frac{1}{2}} b_{\perp}(i,n)^{-\frac{d-1}{2}-1} & i = j. \end{cases}$$

In order to simplify the expression we define

$$S_m(\nu_1,\nu_2) = \sum_{i\neq m}^{N} b_{\parallel}(i,m)^{-\nu_1} b_{\perp}(i,m)^{-\nu_2}, \qquad (27)$$

$$F3_{\parallel} = \frac{\partial}{\partial G_{\parallel}} V3_{\parallel}(i,j) = \frac{w}{3!}\left(\frac{1}{2\pi}\right)^{\frac{d}{2}}$$
$$\times \begin{cases} b_{\parallel}(i,j)^{-\frac{3}{2}} b_{\perp}(i,j)^{-\frac{d-1}{2}} S_j(\frac{1}{2},\frac{d-1}{2}) \\ +b_{\parallel}(i,j)^{-\frac{3}{2}} b_{\perp}(i,j)^{-\frac{d-1}{2}} S_i(\frac{1}{2},\frac{d-1}{2}) & i \neq j \\ -[\sum_m S_m(\frac{1}{2},\frac{d-1}{2}) b_{\parallel}(i,m)^{-\frac{3}{2}} b_{\perp}(i,m)^{-\frac{d-1}{2}} \\ +S_i(\frac{1}{2},\frac{d-1}{2}) S_i(\frac{3}{2},\frac{d-1}{2}) - S_i(2,d-1)] & i = j \end{cases}$$

$$F3_{\perp} = \frac{\partial}{\partial G_{\perp}} V3_{\perp}(i,j) = \frac{w(d-1)}{3!}\left(\frac{1}{2\pi}\right)^{\frac{d}{2}} \qquad (28)$$
$$\times \begin{cases} b_{\parallel}(i,j)^{-\frac{1}{2}} b_{\perp}(i,j)^{-\frac{d-1}{2}-1} S_j(\frac{1}{2},\frac{d-1}{2}) \\ +b_{\parallel}(i,j)^{-\frac{1}{2}} b_{\perp}(i,j)^{-\frac{d-1}{2}-1} S_i(\frac{1}{2},\frac{d-1}{2}) \\ -[\sum_m S_m(\frac{1}{2},\frac{d-1}{2}) b_{\parallel}(i,m)^{-\frac{1}{2}} b_{\perp}(i,m)^{-\frac{d-1}{2}-1} & i \neq j \\ \quad +S_i(\frac{1}{2},\frac{d-1}{2}) S_i(\frac{1}{2},\frac{d-1}{2}+1) - S_i(1,\frac{d-1}{2})] & i = j. \end{cases}$$

Each matrix satisfies the sum rule $\sum_{i\neq j} M(i,j) = -M(i,i)$. We removed the center of mass degree of freedom (zero mode) by using relative coordinates. The procedure is described in [33].

Appendix B: Remarks on the classical path

In order to describe the spontaneous symmetry breaking, Netz and Orland [27] introduced the classical path in the trial Hamiltonian. When the external force direction is given, the corresponding energy $H_f = -|\boldsymbol{f}\cdot(\boldsymbol{r_N}-\boldsymbol{r_1})|$ shifts the free energy by $\frac{|\boldsymbol{f}|}{k_{\mathrm{B}}T} b_{\parallel}^{\frac{1}{2}}(1,N)$ (see Eq. (22)) when it is averaged with the Wiener measure. Then, equation (26) should be solved self-consistently. The change in free energy influences in turn on the measure. The measure satisfying self consistency eventually will be

$$P_V = \frac{\mathrm{e}^{-(H_o-\gamma z_\gamma)}}{Z_V} \qquad (29)$$

where $\gamma = f/k_BT$ and z_γ stands for the self consistent solution for the extension under the external force f. It can be easily seen that the new measure is identical to that of Netz and Orland by transforming the above measure to the quadratic form. Using $H_o = \int\int \mathrm{d}s\mathrm{d}s' G_{\parallel}^{-1}(s,s') z^2(s.s') + G_{\perp}^{-1}(s,s')\rho^2(s,s')$ (where $z = \sqrt{\boldsymbol{r}_{\parallel}(s)-\boldsymbol{r}_{\parallel}(s')}$ and $\rho = \sqrt{\boldsymbol{r}_{\perp}(s)-\boldsymbol{r}_{\perp}(s')}$), the z component of measure is,

$$\mathrm{e}^{-\int \mathrm{d}s \int \mathrm{d}s' [G_{\parallel}^{-1}(z-\gamma G_{\parallel}/2)^2 - \gamma^2 G_{\parallel}/4]}. \qquad (30)$$

The extension $\langle z\rangle$ is therefore shifted to a non zero value according to the external force f. For a polyelectrolyte, as the Coulomb interaction becomes stronger the repulsive interaction deforms the globule to the stretched (necklace) form. If there is an additional net force, the deformation is naturally induced in the direction of the external force. In the case of spontaneous symmetry breaking, we take the limit of $f \to 0$ starting from a finite value of f. The order parameter then remains at a finite value.

References

1. M. Rief, F. Oesterhelt, B. Heymann, H. Gau, Science **275**, 1295 (1997)
2. M. Kellermayer, S. Smith, H. Granzier, C. Bustamante, Science **276**, 1112 (1997)
3. L. Tskhovrebova, J. Trinick, J.A. Sleep, R.M. Simmons, Nature (London) **387**, 308 (1997)
4. T. Senden, J. diMeglio, P. Auroy, Eur. Phys. J. B **3**, 211 (1998)
5. H. Li, M. Rief, F. Oesterhelt, H. Gaub, Adv. Mater. **10**, 316 (1998)
6. A. Courvoisier, F. Isel, J. François, J. Maaloum, Langmuir **14**, 3727 (1998)
7. H. Li, A.F. Oberhouser, S.B. Fowler, J. Clarke, J.M. Fernandez, Proc. Natl. Acad. Sci. USA **97**, 6527 (2000)
8. J. Liphardt, B. Onoa, S. Smith, I. Tinoco Jr., C. Bustamante, Science **292**, 733 (2001)
9. P. Pincus, Macromolecules **9**, 386 (1976)
10. T.A. Vilgis, A. Johner, J.-F. Joanny, Eur. Phys. J. E **2**, 289 (2000)
11. M.N. Tamashiro, H. Schiessel, Macromolecules **33**, 5263 (2000)
12. O. Borisov, A. Halperin, Eur. Phys. J. B **9**, 251 (1999)
13. N.-K. Lee, A. Johner, T.A. Vilgis, Macromolecules **35**, 6043 (2002)
14. M.N. Tamashiro, P. Pincus, Phys. Rev. E **63**, 021909 (2001)
15. D.K. Klimov, D. Thirumalai, Proc. Natl. Acad. Sci. USA **96**, 6166 (1999)
16. E. Paci, M. Karplus, Proc. Natl. Acad. Sci. USA **97**, 6521 (2000)
17. H. Lu, K. Schulten, Biophys. J. **79**, 51 (2000)
18. K.F. Lau, K.A. Dill, Macromolecules **22**, 3986 (1989)

19. S. Ramanathan, E. Shakhnovich, Phys. Rev. E **50**, 1303 (1994)
20. N.-K. Lee, T.A. Vilgis, Europhys. Lett. **57**, 817 (2002)
21. S.P. Obukhov, J. Phys. A **19**, 3655 (1986)
22. C. Sfatos, A.M. Gutin, E. Shakhnovich, Phys. Rev. E **48**, 465 (1993)
23. E. Shakhnovich, A.M. Gutin, J. Phys. France **50**, 1843 (1989)
24. T. Garel, H. Orland, Europhys. Lett. **6**, 597 (1988)
25. M. Mézard, G. Parisi, M.V. Virasoro, *Spin glasses and beyond* (World Scientific, Singapore, 1986)
26. M. Doi, S.F. Edwards, *The theory of polymer dynamics* (Oxford Clarendon Press, Oxford, 1986)
27. R.R. Netz, H. Orland, Eur. Phys. J. B **8**, 81 (1999)
28. J.-P. Bouchaud, M. Mezard, G. Parisi, J.S. Yedida, J. Phys. A **24**, L1025 (1991)
29. J. Cloizeaux, J. Phys. France **31**, 715 (1970)
30. B. Jönsson, C. Peterson, B. Söderberg, Phys. Rev. Lett. **71**, 376 (1993)
31. B. Jönsson, C. Peterson, B. Söderberg, J. Phys. Chem. **99**, 1251 (1995)
32. G. Migliorini, V. Rostiashvili, T.A. Vilgis, Eur. Phys. J. E **4**, 475 (2001)
33. G. Migliorini, N.-K. Lee, V. Rostiashvili, T.A. Vilgis, Eur. Phys. J. E **6**, 259 (2001)
34. R.P. Feynman, *Statistical Mechanics* (Addison Wesley, The advanced book program, 1972)
35. A. Halperin, E.B. Zhulina, Europhys. Lett. **15**, 417 (1991)
36. P.-Y. Lai, Phys. Rev. E **58**, 6222 (1998)
37. K. Furukawa, K. Ebata, N. Matsumoto, Appl. Phys. Lett. **75**, 781 (1999)
38. K. Furukawa, K. Ebata, N. Matsumoto, Appl. Phys. Lett. **77**, 4289 (2000)
39. P.L. Geissler, E.I. Shakhnovich, Macromolecules **35**, 4429 (2002)
40. N.-K. Lee, T.A. Vilgis, preprint
41. A.V. Dobrynin, M. Rubinstein, S.P. Obukhov, Macromolecules **29**, 2974 (1996)
42. T.J.H. Vlugt, private communication
43. V.G. Rostiashvili, T.A. Vilgis, Europhys. Lett. **49**, 162 (2000)