Correlated energy landscape model for finite, random heteropolymers

Steven S. Plotkin, Jin Wang, and Peter G. Wolynes
Department of Physics and School of Chemical Sciences, University of Illinois, Urbana, Illinois 61801
(Received 26 December 1995)

In this paper, we study the role of correlations in the energy landscape of a finite random heteropolymer by developing the mapping onto the generalized random energy model (GREM) proposed by Derrida and Gardner [J. Phys. C **19**, 2253 (1986)] in the context of spin glasses. After obtaining the joint distribution for energies of pairs of configurations, and by calculating the entropy of the polymer subject to weak and strong topological constraints, the model yields thermodynamic quantities such as ground-state energy, entropy per thermodynamic basin, and glass transition temperature as functions of the polymer length and packing density. These are found to be very close to the uncorrelated landscape or random energy model (REM) estimates. A tricritical point is obtained where behavior of the order parameter q changes from first order with a discrete jump at the transition, to second-order continuous. While the thermodynamic quantities obtained from the free energy are close to the REM values, the Levinthal entropy describing the number of basins which must be searched at the glass transition is significantly modified by correlations. [S1063-651X(96)02406-3]

PACS number(s): 61.41.+e, 64.70.Pf, 05.90.+m, 87.10.+e

I. INTRODUCTION

The statistical characterization of the energy landscape of random and designed heteropolymers has been a major component of the newer theoretical approaches to biological protein folding [1–6]. Completely random heteropolymers have rugged energy landscapes due to the frustration inherent in conflicting interactions between different monomers that are covalently linked in the polymer chain. For appropriate choices for the interactions in native proteins [7,8], or the appropriate sequences of artificially designed polymers [9,10], the effects of frustration can be minimized, leading to a funneled energy landscape [1] with driving forces toward a well-defined native structure, in addition to the generic ruggedness of random heteropolymers. Much of our understanding of the dynamics on these energy landscapes has been derived from the study of the most rugged energy landscape, the so-called random energy model (REM) originally studied by Derrida [11]. This model is very simple because it is characterized by a single energy scale giving the overall energetic randomness, and a configurational entropy. For biological proteins, the funnel aspect of the landscape gives a new energy scale, the stability gap, which determines the average trend of the energy as the protein molecule becomes more similar to its ground-state configuration. In this paper, we wish to address quantitatively the role of correlations in the energy landscape of random heteropolymers. The effect of minimal frustration and the corresponding stability gap on an already correlated energy landscape will be treated in a later paper.

The uncorrelated energy landscape, or random energy model, possesses a glass transition that arises from an entropy crisis. This phase transition is representative of a wide universality class of phase transitions in spin glasses that lack special symmetries [13]. Replica methods from spin glass theory have been applied to the random heteropolymer and confirm that the glass transition of random heteropolymers is also in this universality class [14]. While the random heteropolymer glass transition is of the same type as the random energy model, it is clear that correlations in the energy landscape can play a role in determining quantities such as the glass transition temperature, as well as the characteristics of the basins of attraction into which the system freezes below the glass transition.

Nevertheless, the convenience of the random energy model has made it useful for quantitatively treating the phase transitions of random heteropolymers. Its very simplicity allows it to be used as an approximation for models with elaborate interaction potentials and complex stereochemical constraints that can mimic proteins [9]. It also allows the inclusion of various kinds of partial order in collapsed heteropolymers, such as liquid crystalline ordering and secondary structure formation [15,16]. The more elegant replica methods, while partially taking into account the correlations of the energy landscape, are considerably more cumbersome to use for models with these realistic levels of molecular complexity.

The approach we take in this paper to the correlated energy landscape is simpler than the replica method. It is based on the use of the generalized random energy model (GREM) of Derrida [17]. In this model, one takes into account the energy correlations of a pair of states on the energy landscape as a function of the similarity of the two configurations. The model then visualizes the characterization of the landscape by describing its properties upon the resulting ''triangulation'' on the energy surface. The input to the model consists of two quantities: (1) the number of states within a certain distance of a given molecular configuration, and (2) the expected degree of correlation of their energies with the given molecular configuration. The model reduces to the uncorrelated landscape in the obvious way, but also allows a more quantitative approximation of the glass transition temperatures when the pair correlations are known *a priori*. Derrida and Gardner [17] have shown how the generalized random energy model can be thought of as the beginning of a systematic set of approximations to the thermodynamics of any random system by taking further correlations in energy levels. They have also described how

it can be used to approximate the glass transition temperatures of many standard spin glass models.

The application of the generalized random energy model to random heteropolymers raises some interesting questions in polymer physics. Just as for the random energy model, conventional phase transitions in ordering of the heteropolymer that are analogous to those of a homopolymer (e.g., collapse, or secondary structure formation) can be taken into account in a straightforward way, as in homopolymer physics. The present questions revolve around the counting of structures with a given degree of similarity to other ones. When the underlying energy surface is made up by pair interactions, this counting exercise is very similar to the theory of a rubber vulcanization [18]. Indeed, rubber vulcanization has already been addressed by the replica methods used in spin glass theory [19,20]. As far as the entropic issues are concerned, however, these theories have, in the main, reproduced the results of the much older analysis of Flory. We adopt an analysis in this Flory style here, because we believe it lends itself straightforwardly to generalization by taking into account more molecular details. We find that it is necessary, however, to go beyond the Flory analysis when one must count states that correspond to highly cross-linked structures. To this end we undertake an analysis of the collective process of melting out of local structure in a random heteropolymer.

With these polymer issues under control, it is possible to evaluate the statistical thermodynamics of random heteropolymers within the generalized random energy model approximation. We present results for these thermodynamic properties for mesoscopic random heteropolymers; that is, we study the finite size effects that are quite important for heteropolymers in the size range relevant for protein folding. We present transition temperatures for three-dimensional lattice systems as a function of polymer size. We show that the generalized random energy approximation and the random energy model give closely similar results for the transition temperature. On the other hand, the correlations in the energy landscape do modify the size of the basins of attraction and the effective number of basins that need to be searched through as the glass transition is approached. In a later paper we will show how the information in the correlated energy landscape can also be used to address kinetic issues such as barrier height distributions, but here we limit our discussion to the problem of the number of basins, i.e., the Levinthal entropy [21].

The organization in this paper is as follows: In Sec. II we discuss the general issues related to the correlated energy landscape, and obtain a formula for the energy correlations between states with a given similarity q. In Secs. III and IV we calculate the configurational entropy of a polymer given the existence of weak and strong topological constraints (the log of the total number of states consistent with the constraints imposed). In Sec. V this entropy and the energy-pair correlations are applied to the GREM model to obtain thermodynamic quantities such as entropy and energy for the finite random heteropolymer. The glass transition temperature of partial and total freezing, as well as the ground-state energy and entropy at the glass transition, are obtained for various size polymers and packing fractions. The probability distribution for overlaps q below the glass temperature is also obtained. The GREM results are compared with those of the uncorrelated energy landscape (the REM), and the issues of thermodynamic basins of attraction are discussed. In Sec. VI we discuss the results and conclude with some remarks.

II. BASIC ISSUES

It was shown in a previous paper [22] that the thermodynamic glass transition temperature T_g of a polymer should always be less than its collapse temperature T_θ. For our purposes then, if we define the packing fraction of the polymer, η, by $N\sigma/R_g^3$ where σ is the volume per monomer, and R_g is the radius of gyration of the polymer, it is sufficient to consider only configurations of the polymer that have values of $\eta \gtrsim 1/2$. In considering these collapsed or semicollapsed states, we neglect fluctuations about the mean number of contacts (off-chain pair interactions) $Nz(N)\eta$, where $z(N)\eta$ is then the mean number of contacts per monomer for a collapsed walk of packing fraction η.

Let us consider the most collapsed walk to have $\eta \cong 1$, where $z(N)$ is then the number of bonds made per monomer in a Hamiltonian (dense) walk of N steps, a quantity studied extensively by Douglas and Ishinabe [23]. The dependence of $z(N)$ upon N is clearly due to the fact that monomers on the surface have less contacts than in the bulk. For three-dimensional (3D) systems, $z_3(N)$ is given approximately by

$$z_3(N) \approx \frac{1}{N}\,\mathrm{Int}[2N - 3(N+1)^{2/3} + 3], \tag{2.1}$$

where Int[ ] means the integer part. For 27-mer collapsed cube structures with 28/27 actual contacts per mer, Eq. (2.1) gives 29/27. On the other hand, the effect of the surface on the number of contacts is quite important even for large macromolecules, as $z_3(N)$ approaches its bulk value of two contacts per monomer rather slowly, as $\sim 2 - 3N^{-1/3}$.

For a dense walk of $\eta = 1$, the total number of states is determined by the connectivity constant, which to second order is given by [24]

$$\ln\mu \approx \ln\mu^{(mf)} + \frac{1}{6\zeta^2}, \tag{2.2}$$

where $\mu^{(mf)}$ is the mean field connectivity constant and ζ is the coordination number of a monomer in a polymer chain (4 for a polymer on a cubic lattice), so that to within ~ 0.01 accuracy, we can rely on the mean field result of ν/e states per monomer (neglecting the $\sim \ln N$ translational entropy associated with the number of places the polymer can start its walk from), where ν is the number of states per monomer in an ideal chain (6 on a $d = 3$ cubic lattice).

If the polymer is not completely collapsed ($\eta < 1$) the mean field connectivity constant increases as η decreases [22]

$$\ln\mu^{(mf)} = \ln\frac{\nu}{e} - \left(\frac{1-\eta}{\eta}\right)\ln(1-\eta), \tag{2.3}$$

which just reflects the fact that partly collapsed walks have more sterically allowed states than fully collapsed ones.

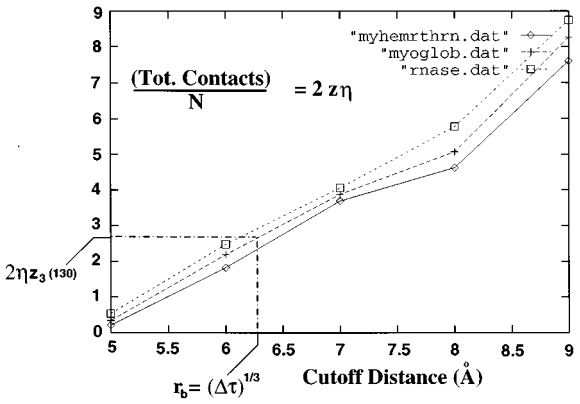

FIG. 1. Plot of the $2\times$(total number of contacts) vs the cutoff distance of a contact for three real proteins of mean sequence length 130. Using the number of expected contacts for a dense walk on a cubic lattice gives a bond radius over Kuhn length (≈ 1.7) reasonably consistent with the cubic lattice value (≈ 1.6).

If the average energy per contact is $\overline{\varepsilon}$, the total average energy is then

$$\langle E(N,R_g)\rangle = Nz(N)\,\eta(R_g)\overline{\varepsilon}, \tag{2.4}$$

which we will set to zero as our zero point of energy. In using formulas like (2.4), we are neglecting the coupling of density (η) with topology (bonds formed, or overlap q). This is a ''van der Waals'' picture appropriate well below the collapse temperature. It is straightforward to simultaneously include this coupling in obtaining a complete phase diagram as obtained earlier in [8] (albeit without correlations), but this is saved for another paper.

For a random heteropolymer, a pair of interacting monomers $\{ij\}$ has an interaction energy ε_{ij} that can be taken to be a random variable. The energy $\mathcal{H}$ for a given total configuration is given by

$$\mathcal{H} = \sum_{i<j} \varepsilon_{ij}\sigma_{ij}, \tag{2.5}$$

where $\sigma_{ij}=1$ when there is a contact made between monomers $\{ij\}$ in the chain, and $\sigma_{ij}=0$ otherwise. Here contact means that the two monomers $\{ij\}$ are within a small distance (bond radius) of each other, or we can equivalently speak of the volume around monomer i which another monomer must be inside for a bond to be present ($\equiv\Delta\tau$). To specify dimensionless quantities such as the entropy in Boltzmann units, another distance scale must enter into the problem, which is the Kuhn length ℓ of the polymer. For a flexible polymer on a 3D cubic lattice ℓ^3 is the volume one lattice site occupies, and $\Delta\tau$ is the volume of four lattice sites.

It is worthwhile to note that the ratio $\Delta\tau/\ell^3=4$ for cubic lattices is roughly that obtained in real proteins by finding what the bond radius would have to be for the protein to have a number of contacts equal to $\eta z_3(N)$, where $\eta\simeq 1$ [25] and $z_3(N)$ is the contacts per monomer for a dense walk of N steps on a cubic lattice. Taking three typical proteins of mean sequence length 130, and using $\eta z_3(130)\cong 1.4$ gives a bond radius of $\cong 6.3$ Å (see Fig. 1). Using a $C_\alpha - C_\alpha$ distance of $\cong 3.6$ Å gives $\Delta\tau/\ell^3\cong 5.3$, in rough agreement with the lattice value.

Since the total energy of the polymer is a sum of random variables, it is a Gaussian random variable with probability distribution

$$P(E|N,R_g) = \frac{1}{(2\pi\Delta E^2)^{1/2}}\exp\left(-\frac{E^2}{2\Delta E^2}\right), \tag{2.6}$$

where the variance in energy $\Delta E^2 = Nz(N)\,\eta(R_g)\varepsilon^2$, where $\sqrt{\varepsilon^2}$ is the width of the effective Gaussian energy distribution of a single bond, which is the roughness energy scale as in the random energy model [11].

If we pick two different states of the polymer, both collapsed with $Nz\,\eta$ total contacts, and ask what the probability is of the states having energies E_a and E_b, respectively, our answer will depend not only on the two energies we have picked, but on how similar the states a and b were to begin with. This similarity can be represented by a single parameter q defined as

$$q = \frac{\text{(number of contacts identical in the two states)}}{\text{(total number of contacts)}}. \tag{2.7}$$

The ''overlap parameter'' q varies from 0 to 1. If we define μ as the number of identical contacts in the two different states, it is clear that

$$\mu = \sum_{i<j}\sigma_{ij}^{(a)}\sigma_{ij}^{(b)} \equiv \sum_{i<j} s_{ij}^{(a,b)} \tag{2.8}$$

$$= Nqz\,\eta, \tag{2.9}$$

where $s_{ij}^{(a,b)}=1$ if i and j are in contact in both states and 0 otherwise. To find the pair energy distribution mentioned above, we can follow the procedure used by Derrida in the generalized random energy model [26]. This is done in Appendix A. The result is

$$P_{a,b}(E_a,E_b) = \text{const}\times\exp\left[-\frac{1}{4Nz\,\eta\varepsilon^2}\times\left(\frac{(E_a+E_b)^2}{1+q}+\frac{(E_a-E_b)^2}{1-q}\right)\right], \tag{2.10}$$

where q is again $\mu/(Nz\,\eta)$. Note that as $q\to 0$ the states share no common bonds, and the pair energy distribution factors into the product of two single (uncorrelated) energy distributions, $P(E_a,E_b)\to P(E_a)P(E_b)$. As $q\to 1$ we are looking at states very similar to each other, and $P(E_a,E_b)\to\delta(E_a-E_b)$. If we integrate $P(E_a,E_b)$ over one of the energies, we obtain the single energy distribution for the remaining energy, as expected.

If interactions of a more collective nature than pair interactions are present in the Hamiltonian describing the system, as, for example, must be present in the interactions facilitating ligand binding when there are cofactors, the interactions involved in side-chain packing, or as results from averaging over the solvent degrees of freedom, then these m-body

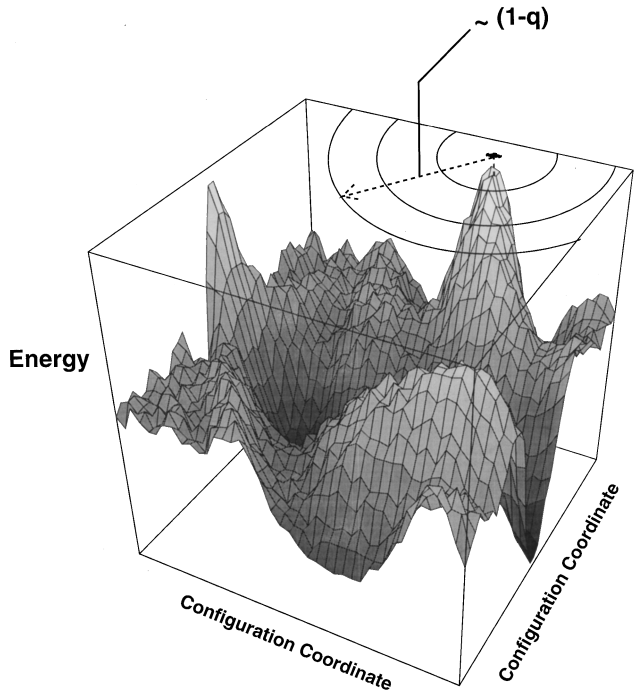

FIG. 2. Qualitative picture of the energy landscape, pictured here as a two-dimensional projection of the multidimensional configuration space. We can speak of a distance radius $\sim(1-q)$ from any given state, which determines how similar or correlated the energies of states at that radius are. The correlations smooth out the energy landscape, which affects the nature of the glass transition and somewhat lowers the temperature at which it occurs.

terms can give rise to a q^{m-1} dependence of the pair energy correlation on the overlap q, defined as above with a suitable decomposition law for each $\sigma_{ijk\cdots m}$ into pair interaction terms $\sigma_{ij}\sigma_{jk}\cdots$, such as in the superposition approximation in the theory of fluids [27]. Whether such explicit cooperative effects which enhance the first-order-like folding transition behavior are necessary to fully describe proteins is an open issue. For states that are mostly collapsed, a REM-like cooperative glass transition is still seen in the present GREM analysis which uses only pair interactions, in that there is a finite jump in the order parameter q, signifying the sudden onset of freezing at discretely different values of overlap [however, the glass transition is still second order, with a thermally averaged overlap that is continuous at T_g (see the comments in Sec. V)].

The pair energy distribution we have obtained has the same form as in the generalized random energy model [17], with the parameter v simply equal to q, so $a(q)=1$ (see Appendix B for a brief review of the GREM). The overlap q smoothes out the energy landscape by making states that are similar to a given state a (i.e., states b with q_{ab} close to 1) more likely to have an energy close to E_a (see Fig. 2). More precisely, given that we have picked state a of energy E_a the probability for another state b to have energy E_b is

$$\frac{P_{ab}}{P_a}=\text{const}\times\exp\left(-\frac{(E_b-qE_a)^2}{2Nz\eta\varepsilon^2(1-q^2)}\right). \qquad (2.11)$$

Note the q-dependent mean and variance of the distribution, and the appropriate limits as $q\to 0$ and $q\to 1$.

Now that we have specified the parameter that determines the roughness of the energy landscape as a function of similarity, $a(q)$, the only remaining quantity needed to describe the thermodynamics (as described in Appendix B) is essentially the rate of decrease in the number of states as we move towards a given state by increasing the similarity q. This quantity is calculated in the next two sections. Those who wish to take the entropy results as given may skip to Sec. V on thermodynamics.

III. ENTROPY OF A WEAKLY CONSTRAINED POLYMER

The remaining quantity needed to apply the GREM is the number of states that have an overlap q with a given state, or equivalently the entropy $s(q)$ of a polymer that has $\mu=Nqz\eta$ bonds in common with a given state. This entropy is given by the sum of several terms [see Eq. (3.9)], the first of which is simply the total entropy before any constraints are imposed (here and throughout the text, unless explicitly stated otherwise, all entropies are in units of Boltzmann's constant):

$$S_o=N\ln\mu^{(mf)},$$

i.e., that of a collapsed random walk, where $\ln\mu^{(mf)}$ is defined in Eq. (2.3).

For low values of q, the polymer is weakly constrained, and the entropy formula is essentially one of entropy reduction due to bond formation, i.e., configurational entropy is lost due to the constraints imposed by the bonds. We should distinguish here between the bonds which are decreasing the total entropy, i.e., those which contribute to the overlap q, and the total $Nz\eta$ bonds which are present in all states of the collapsed heteropolymer [28]. These other bonds do not further constrain the polymer, as their effect has already been taken into account by assigning the collapsed entropy S_o. The polymer can always explore its $(\mu^{(mf)})^N$ states of a collapsed walk, all of which have $Nz\eta$ bonds. We are looking at the fraction of these states consistent with a particular set of $qNz\eta$ bonds being formed. [However, as mentioned below there is a set of $Nz\eta(1-q)$ bonds which *cannot* be formed in addition to the set of $qNz\eta$ bonds that must be formed. This causes a further reduction in the entropy.]

In calculating the decrease in conformational entropy of the chain segments due to the formation of a bond, we want to find the probability that the given bond will be formed, which is equal to the fraction of the total number of states without the bond that are equivalent in configuration to the bond present, assuming effectively a microcanonical search through the possible states. Since the polymer is collapsed, for small q (weakly constrained) any given piece of it behaves as if it were in a melt, i.e., as an ideal chain. This means we can neglect excluded volume effects in our calculations, e.g., the probability distribution of end-to-end distances of a piece of polymer chain has no ''hole'' at small values (as in a self-avoiding walk).

To calculate the bond formation entropy loss we use the approach of Flory's older work on the formation of cross links in polymer chains [18]. The reduction in configura-

tional entropy from the unconstrained polymer due to the addition of $qNz\eta$ cross links was found by Flory to be

$$\frac{1}{N}\Delta S_{bond}(q|N,\eta)=\frac{3}{2}qz\eta[\ln C-1+\ln qz\eta], \quad (3.1)$$

where

$$C=\frac{3}{4\pi}\left(\frac{\Delta\tau}{b^3}\right)^{2/3}. \quad (3.2)$$

For details concerning the derivation of Eq. (3.1) see Appendix C. C is a constant of order 1, and contains the ratio of length scales discussed above in the factor $\Delta\tau/b^3$. It should be mentioned that the entropy term due to bond formation has also been more recently reproduced by replica calculations [29].

Specifying the overlap q introduces an additional entropy reduction due to the fact that $Nz\eta-qNz\eta$ contacts of the reference state must *not* be formed. This ''antibond'' entropy reduction is largest for small overlaps, and goes to zero as $q\to 1$. It is given by (see Appendix D for a derivation)

$$\frac{1}{N}\Delta S_{AB}(q|N\eta)=\frac{1}{C}\int_{Cqz\eta}^{Cz\eta}dx\ln(1-x^{3/2}), \quad (3.3)$$

where C is given by Eq. (3.2).

Another term in the entropy is due to the many ways an overlap of μ bonds can be realized, because these different sets of overlapping bonds are all realizable in the conformational search of the polymer, there being nothing essential to distinguish the bonds in common with the reference state from any others. Neglecting the fact that some overlaps are impossible due to steric constraints, the entropy ''of mixing'' (per monomer) associated with choosing $qzN\eta$ bonds from $zN\eta$ total is

$$\frac{1}{N}S_{mix}(q|N,\eta)\cong -z(N)\eta\ [q\ln q+(1-q)\ln(1-q)\ ]. \quad (3.4)$$

The Flory approach (see Appendix C) of considering the formation of a cross link from four chains defined by the cross link's four neighboring bonds breaks down in a (small-N-dependent) region of very weak constraint, where the entropy loss due to the formation of the cross link is more accurately accounted for by considering the formation of *loops* in a nearly unconstrained chain. Furthermore, as described below, the loops that define a bond or cross link in this nearly unconstrained regime are confined inside a region of the linear size of the polymer, which reduces the rate at which entropy is lost, i.e., the confinement to configurations of the polymer consistent with a collapsed walk inside a molten globule imposes a restriction on the size of a random walk of a section of the polymer chain, which makes it more likely for a loop to be formed. The rms length of a walk of the size of the average chain length $\overline{nb}=N/2\mu b$ cannot be larger than the linear size of the globule $R_g\cong(N/\eta)^{1/3}b$, so

$$\left(\frac{N}{2\mu_c}\right)^{1/2}\approx\left(\frac{N}{\eta}\right)^{1/3} \quad (3.5)$$

defines a critical number of bonds $\mu_c\approx(N^{1/3}\eta^{2/3})/2$ [30], below which random walks ''see'' the boundary of the globule. So the number of the chains' allowable configurations must be reduced, and their distributions modified from the usual Gaussian behavior assumed so far, for values of overlap less than

$$q_c\approx\frac{1}{2z\eta^{1/3}N^{2/3}}. \quad (3.6)$$

Note that confinement effects become less important as N and η increase. (The effect of localization induced by the presence of cross links has also been investigated recently by Bryngelson and Thirumalai [31].)

A straightforward way to introduce the effect of a boundary on the configurations of the polymer is to consider it confined to a box of length $L=(N/\eta)^{1/3}b$ by introducing an external potential $U_e=0$ inside the box, and $U_e=\infty$ outside. The details of the calculation are in Appendix E; a result is that for $\mu<\mu_c$

$$\frac{1}{N}\Delta S_{bond}^{conf}(q|N,\eta)\cong\frac{3}{2}qz\eta(\ln C'-1+\ln qz\eta), \quad (3.7)$$

where

$$C'=\frac{12}{\pi}\left(\frac{\Delta\tau}{b^3}\right)^{2/3}. \quad (3.8)$$

$(1/N)\,\Delta S_{bond}^{conf}$ has the same form as (3.1), the only difference being the constant C' which makes the entropy loss not as great.

Figure 3 is a plot of all the separate entropy contributions and their sum vs q for the 27-mer; the total low q entropy is an interpolation between the confined and unconfined expressions (see Appendix E).

Figure 4(a) is a comparison of the weakly-constrained-polymer theory, where

$$\frac{1}{N}S_{\rm TOT}(q)=\frac{1}{N}[S_o+\Delta S_{bond}(q)+\Delta S_{AB}(q)+S_{mix}(q)], \quad (3.9)$$

with a lattice simulation of $S(q)/N$ for the 27-cube [32]. The theory and simulations of the 27-cube are relevant for collapsed, proteinlike heteropolymers of sequence length ~ 60 with $\sim 65\%$ helicity and liquid crystalline ordering [3,15]. In this sense the molten globule state of the polymer can possess secondary structure, but this structure is renormalized away in coarse-graining the description of the individual monomeric units.

One detail that must be accounted for is that $S(q)$ considered in the simulation is relative to the collapsed cube state, and therefore η varies as a function of q from about 0.7 to 1. To compensate for this effect in our constant-η theory, we have used an initial unconstrained entropy S_o for that of a partially collapsed $\eta=0.7$ polymer, but to obtain accurate values for higher values of q, the fully collapsed $\eta=1$ values for the other entropy terms were used. Later, when we will interpolate between the weakly constrained and strongly constrained entropies to obtain the glass temperature and other quantities, we will hold η constant in all expressions.

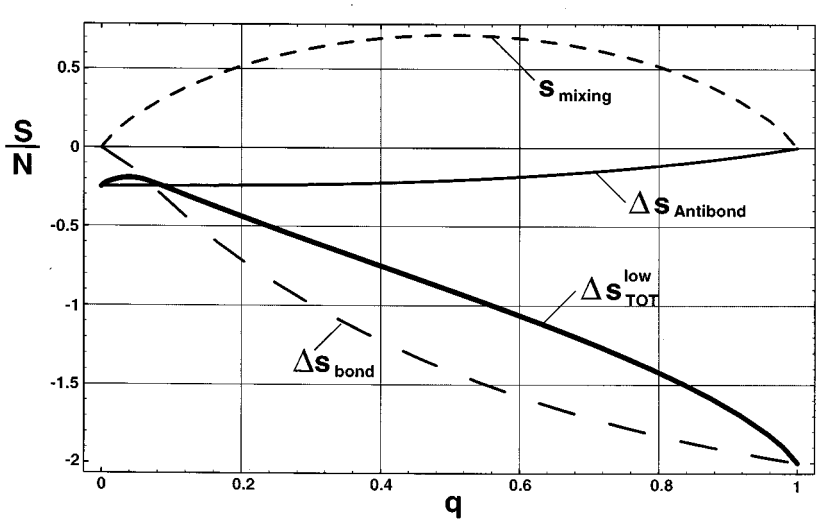

FIG. 3. Entropy contributions (divided by N) to the total low q entropy for $N=27$, $\eta=1$ (all entropies here and throughout are in units of Boltzmann's constant k_B.): Long-dashed line: Entropy loss due to the formation of bonds [Eq. (3.1)]. Short-dashed line: Entropy associated with the complexions consistent with the characterization q [Eq. (3.4)]. Thin solid line: Entropy loss due to forbidden configurations which would cause the overlap to exceed q. This ''antibond'' term becomes larger for more compact polymers and for longer polymers [Eq. (3.3)] Thick solid line: Total entropy loss for the weakly constrained polymer. Adding s_o to this gives formula (3.9). Confinement effects add further modifications to the theory for small values of q (see Sec. III).

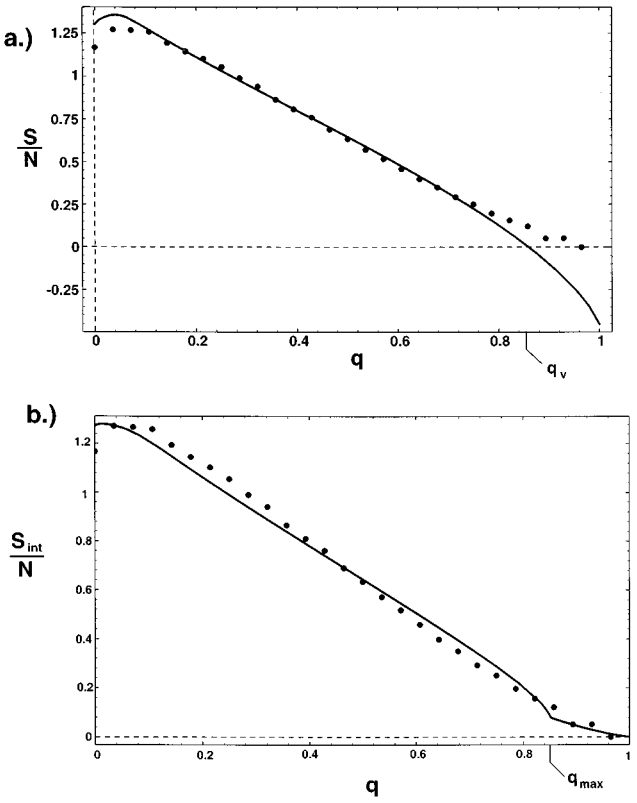

FIG. 4. (a) Theoretical entropy curve for a weakly constrained polymer compared with molecular dynamics simulations for the 27-mer. (b) Interpolated entropy using the weakly constrained formula with $\eta=0.7$ and the strongly constrained formula with $\eta=1$. The small correction for q values above q_{max} is a linear interpolation between $s(q_{max})=(2\ln\mu^{(mf)})/N$ and $s(q=1)=0$. The analytical curves are also qualitatively similar to the 2D simulations carried out by Chan and Dill [56].

One important feature of the $S_{\mathrm{TOT}}(q)$ curve is the existence of a maximum in the entropy ($\cong\ln\mu^{(mf)}$) at a small but nonzero value of $q=q_{min}$, which indicates the statistically most likely value of the overlap as $T\to\infty$, for two states both with $Nz\eta$ bonds. This ''statistical overlap'' signifies the most uncorrelated two states can get as a consequence of the finiteness of the polymer (its corresponding freezing temperature is therefore ∞, as we shall see below), and as $N\to\infty, q_{min}\to 0$.

Note that there is a value of q less than 1, q_v, at which

$$\Delta S_{\mathrm{TOT}}(q_v)=0, \tag{3.10}$$

where $q_v\sim 1/(z(N)\eta)$, i.e., at q_v there is ~ 1 bond per monomer. For 27-mers $q_v\approx 1$, but for larger polymers this problem can become more serious [33]. For $q>q_v$, there is more than one bond on average per monomer, and configurations with finite entropy are highly inhomogeneous, i.e., segments of the polymer which are correctly configured are clustered together, and only a few loops or dangling ends are free and contribute to the configurational entropy. An appropriate formula for high q based on the combinatorics of discrete sections of the polymer chain ''melting out'' of the ''frozen'' constrained medium, rather than a ''gas'' of a few individual formed contacts as in the low q treatment, is obtained in the next section.

IV. ENTROPY OF A STRONGLY CONSTRAINED POLYMER

In the high q limit, we start from a reference state in which all the bonds are formed, and the polymer is ''frozen.'' By switching from the contact representation used at low q to an atomic representation, we can study how certain parts of the ''frozen'' polymer are ''melted out'' by keeping

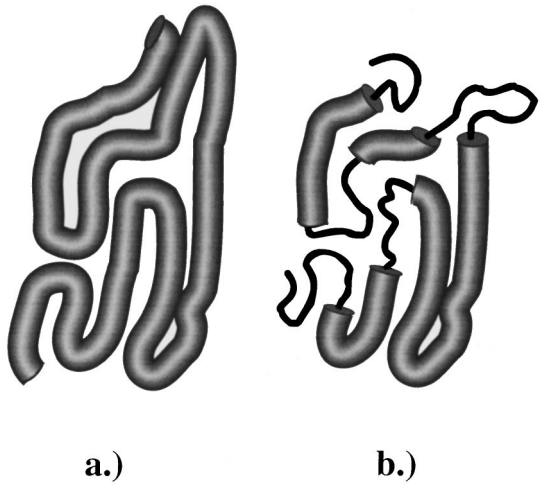

FIG. 5. (a) Diagram of a polymer in the geometrical configuration of the reference state ($q=1$). (b) For large values of the similarity parameter q, the entropy can be considered to come from melted out strands along the sequence which are not in their correct geometrical positions (dark lines), and their combinatorics with the rest of the frozen medium.

track of which residues are still in their correct geometrical positions relative to the three-dimensional structure of the reference state. The melted pieces each carry a certain amount of entropy, and there is also a mixing entropy associated with the different places that the given melted pieces can occur along the sequence of the polymer. The process of melting physically involves the collective freeing up of several monomers at once, i.e., at least some critical number ℓ_c of monomers must be free for the melted strand to have any entropy. Each melted piece of segment length ℓ carries with it an entropy

$$S(\ell) \cong \ln\mu^{(mf)}[\ell-(\ell_c-1)]. \tag{4.1}$$

In addition, the ends are allowed to be freed up in the same fashion, but we expect them to be easier to free up, with a correspondingly smaller value of critical collective length. So the entropy for a free end of length ℓ is

$$S(\ell) \cong \ln\mu^{(mf)}[\ell-(\ell_{\mathrm{EC}}-1)], \tag{4.2}$$

where $\ell_{\mathrm{EC}}<\ell_c$ typically. See Appendix F for arguments giving these two results.

We wish now to express the total number of states of an entire polymer composed of melted and frozen pieces, along with melted or frozen ends (see Fig. 5), and concurrently estimate q. We can characterize a state microscopically by the number distribution of melted pieces of length ℓ, $\{n_\ell\}$, the number distribution of frozen pieces of length ℓ, $\{m_\ell\}$, and of the probability distribution that an end has length ℓ, $\{p_\ell\}$. As a consequence of specifying the total number of states in terms of the given distributions $\{n_\ell\}$, $\{m_\ell\}$, and $\{p_\ell\}$, there must be a combinatorial factor present associated with the permutation degeneracy given the above distributions. There is also a mixing term pertaining to the end length distributions, which is necessary for the end lengths to have a probability distribution rather than just their mean value. The melted pieces, frozen pieces, and end lengths each have their own internal partition function given by (4.1) and (4.2) (the frozen pieces do not have any internal entropy in our model), so the total number of states is given by [35]

$$\Omega_{\mathrm{TOT}} = \sum_{\substack{\{n_\ell\}\\ \{m_\ell\}\\ \{p_\ell\}}} \frac{(Nf)!^2}{\Pi_{\ell_c}^{N} n_\ell! \Pi_1^N m_\ell!} \prod_{\ell=\ell_c}^{N} (\mu^{\ell-(\ell_c-1)})^{n_\ell} \times \prod_{\ell=\ell_{\mathrm{EC}}}^{N} \left(\frac{1}{p_\ell}\mu^{\ell-(\ell_{\mathrm{EC}}-1)}\right)^{2p_\ell}, \tag{4.3}$$

where $Nf \equiv N_{tot}$ is the total number of melted pieces, $\mu=\mu^{(mf)}$, and the sum is over all possible distributions of $\{n_\ell\}$, $\{m_\ell\}$, and $\{p_\ell\}$ (see Appendix G for arguments leading to this expression, specifically the partition function for the ends). Expressions of this form for the number of states have been used in models of the helix-coil transition [34], and in models of polymer adsorption onto a surface [36,37]. Maximizing the log of the largest term in (4.3) subject to constraints (see Appendix H) gives the usual negative binomial distributions for $\{n_\ell\}$, $\{m_\ell\}$, and $\{p_\ell\}$:

$$\frac{n_\ell}{N_{\mathrm{TOT}}} = \frac{f}{r}\left(1-\frac{f}{r}\right)^{\ell-\ell_c},$$

$$\frac{m_\ell}{N_{\mathrm{TOT}}} = \frac{f}{q}\left(1-\frac{f}{q}\right)^{\ell-1},$$

$$p_\ell = s(1-s)^{\ell-\ell_{\mathrm{EC}}}, \tag{4.4}$$

where $N_{\mathrm{TOT}}=Nf$, and

$$r = 1-q-f(\ell_c-1)-2\ell_E/N, \quad s = \frac{1}{\ell_E-(\ell_{\mathrm{EC}}-1)}. \tag{4.5}$$

Substituting the distributions (4.4) back into $\ln W$ [Eq. (H1); see Appendix H] gives the entropy as a function of the macroscopic parameters q, f, and ℓ_E:

$$\begin{aligned}\frac{1}{N}S(q,f,\ell_E) = {} & \ln\mu^{(mf)}\left[1-q-f(\ell_c-1)-\frac{2(\ell_{\mathrm{EC}}-1)}{N}\right] \\ & + q\ln q-(q-f)\ln(q-f)-2f\ln f \\ & + \left(1-q-f(\ell_c-1)-2\frac{\ell_E}{N}\right) \\ & \times \ln\left(1-q-f(\ell_c-1)-2\frac{\ell_E}{N}\right) \\ & - \left(1-q-f\ell_c-2\frac{\ell_E}{N}\right)\ln\left(1-q-f\ell_c-2\frac{\ell_E}{N}\right) \\ & + 2\left(\frac{\ell_E-(\ell_{\mathrm{EC}}-1)}{N}\right)\ln\left(\frac{\ell_E-(\ell_{\mathrm{EC}}-1)}{N}\right) \\ & - 2\left(\frac{\ell_E-\ell_{\mathrm{EC}}}{N}\right)\ln\left(\frac{\ell_E-\ell_{\mathrm{EC}}}{N}\right) - 2\frac{\ln N}{N}. \end{aligned} \tag{4.6}$$

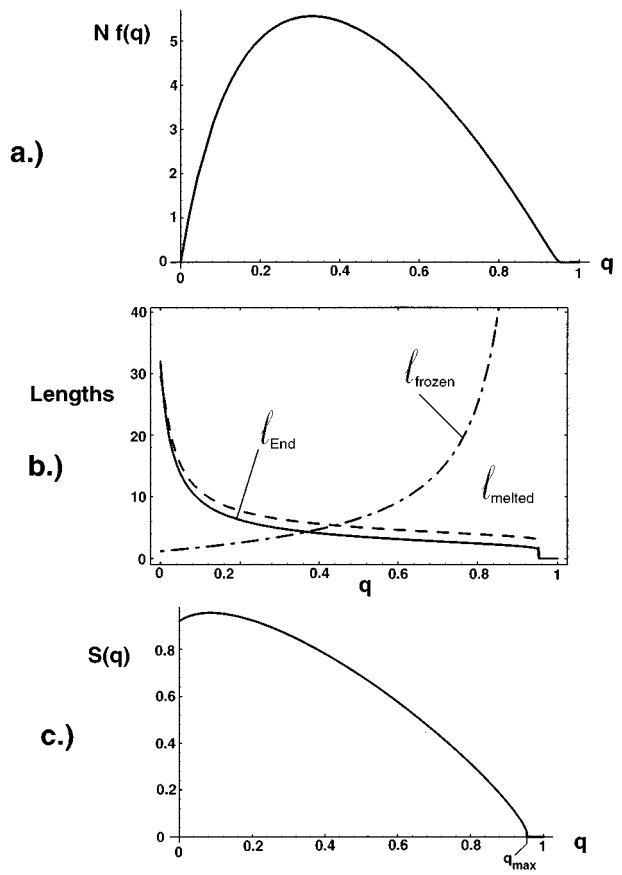

FIG. 6. (a) Plot of the number of internal melted pieces versus similarity parameter q, $Nf(q)$, (with mean length per piece as in (b) in a 64-mer with $\eta=1$, $\ell_c=3$, and $\ell_{\rm EC}=1.5$. (b) Plots of the average free end sequence length $\ell_E(q)$ (solid line), the mean internal melted strand length $\langle \ell_{\rm melted}(q)\rangle$ (dashed), and the mean internal frozen train length $\langle \ell_{\rm frozen}(q)\rangle$ (dot-dashed) as a function of the similarity parameter q, for a 64-mer with $\eta=1$, $\ell_c=3$, and $\ell_{\rm EC}=1.5$ [Eqs. (4.8), (4.9) and (4.10)]. (c) Entropy of a strongly constrained polymer as a function of similarity parameter q, for a 64-mer with $\eta=1$, $\ell_c=3$, and $\ell_{\rm EC}=1.5$.

Finding the most probable end lengths and total number of melted pieces for a given overlap involves maximizing $S(q,f,\ell_E)$ with respect to ℓ_E and f, which gives the equations

$$f(q,\ell_E)=\frac{1-q-2\ell_E/N}{\ell_c+\ell_E-\ell_{\rm EC}} \tag{4.7}$$

and

$$(\ell_E-\ell_{\rm EC})^{\ell_c}\left[q(\ell_c+\ell_E-\ell_{\rm EC})+q-1+2\frac{\ell_E}{N}\right]$$
$$-\left(1-q-2\frac{\ell_E}{N}\right)(\ell_E-\ell_{\rm EC}+1)^{\ell_c-1}\mu^{\ell_c-1}=0. \tag{4.8}$$

Equation (4.8) determines the end length solely in terms of q. For values of $\ell_c\geq 3$ the solution $\ell_E(q)$ is numeric [see Fig. 6(b) for $\ell_E(q)$ with $\ell_c=3$ and $\ell_{\rm EC}=1.5$]. Putting the solution $\ell_E(q)$ into (4.7) gives $f(q)$, where $Nf(q)$ is the number of melted internal pieces as a function of q [see Fig. 6(a)]. Putting both $\ell_E(q)$ and $f(q)$ into Eq. (4.6) for $S(q,f,\ell_E)$ gives the entropy for a strongly constrained polymer solely as a function of q [see Fig. 6(c)]. If the polymer is not completely collapsed ($\eta<1$), then there is still entropy at $q=1$, and the entropy curve is then interpreted as the entropy relative to that of the $q=1$ state.

Note that $f(q)$ shows that there are melted pieces in the interior — it is maximum for moderate values of q because at small q the ends eventually unwind and leave less sequence space for melted pieces. $\ell_E(q=0)=N/2$, and $\ell_E(q)$ drops faster than linearly as it must for there to be interior melted pieces. The mixing term causes there to be a maximum in the entropy for a nonzero value of q. Note also that there is a maximum value of q less than 1 where the entropy essentially runs out as a result of the collectivity of the melting process (the total entropy at q_{max} is just that of two free monomers, $2\ln\mu^{(mf)}$), i.e., the fact that ℓ_c monomers must be melted at once means that the overlap q cannot get infinitesimally close to 1, but has a maximum value at a finite distance from 1. If $2\ell_{\rm EC}<\ell_c$, $q_{max}=1-2\ell_{\rm EC}/N$, otherwise $q_{max}=1-\ell_c/N$. From the above solutions we can also obtain the average length of a typical melted piece, or frozen ''train,'' as a function of q:

$$\langle \ell_{\rm melted}(q)\rangle=\frac{\Sigma_{\ell_c}^{N}\ell n_\ell}{\Sigma_{\ell_c}^{N} n_\ell}=\frac{1-q}{f(q)}-\frac{2\ell_E(q)}{Nf(q)}, \tag{4.9}$$

$$\langle \ell_{\rm frozen}(q)\rangle=\frac{\Sigma_{1}^{N}\ell m_\ell}{\Sigma_{1}^{N} m_\ell}=\frac{q}{f(q)}, \tag{4.10}$$

which are plotted in Fig. 6(b) for the 64-mer.

Note that in our analysis of the polymer entropy we have considered only states associated with different configurations of the backbone and have neglected other contributions to the entropy such as side chain configurations, and entropy due to the solvent.

From the calculation of the entropy $S(q)$ we can now easily obtain the entropic quantity $\ln\alpha(q)=-dS(q)/dq$, which along with the roughness parameter $a(q)$ is sufficient to calculate thermodynamic quantities associated with the correlated energy landscape.

The entropy calculated in the preceding two sections does not consider the energetic dependence of an allowable state's probability of occupation through a Boltzmann factor, and in this respect $s(q)$ is a ''microcanonical'' entropy which just counts the total number of states of all energies with overlap q. The transformation to a canonical entropy and thermodynamic free energy is described in the next section.

Lastly, the entropy theories of the preceding two sections can be easily modified to describe a polymer in dimension d (e.g., $d=2$). The analysis of the often studied $2d$ case parallels the three-dimensional treatment, but the effects of confinement are considerably less. We will not discuss these results in detail here, but mention that the GREM result for $2d$ does not reproduce the replica symmetry breaking found by variational calculations that include vibrational chain entropy [38].

V. THERMODYNAMICS OF THE MODEL

To get a better feel for the GREM results such as freezing temperature and free energy, it is helpful to compare them with those obtained for an uncorrelated landscape, i.e., in the REM [11]. In general, the mean number of states with energies in the interval $(E,E+dE)$ is just

$$\langle n(E)\rangle=(\mu^{(mf)})^N P(E)\sim\exp N\left(s_o-\frac{E^2}{2N\Delta E^2}\right) \quad (5.1)$$

with $P(E)$ given by Eq. (2.6), $\mu^{(mf)}$ given by Eq. (2.3), and $s_o\equiv\ln\mu^{(mf)}$. If $E>E_g=-(2Ns_o\Delta E^2)^{1/2}$ the average number of states is very large for even fairly large N (we will consider only the thermodynamically significant negative energy states here). If the energy landscape is uncorrelated, these states are all statistically independent, and so the relative fluctuations in the number of states at energy E, $\sqrt{\langle[n(E)-\langle n(E)\rangle]^2\rangle}/\langle n(E)\rangle$, are $\sim\langle n(E)\rangle^{-1/2}$ and are thus negligible. So $n(E)\sim\langle n(E)\rangle$ for $E>E_g$, and the microcanonical entropy $S(E)$ is then

$$S(E)=\ln n(E)=N\left[s_o-\frac{E^2}{2N\Delta E^2}\right]. \quad (5.2)$$

On the other hand, if $E<E_g$, the entropy vanishes (the number of states at these low energies is thermodynamically zero), and the system is frozen into one energy state (of energy E_g). Using $dS/dE=1/T=-E/\Delta E^2$ and $\Delta E^2=Nz\eta\varepsilon^2\equiv NJ^2$ where $J=\varepsilon\sqrt{z\eta}$ is a convenient energy scale, we can find the free energy $E(T)-TS(T)$:

$$-\frac{F}{N}=\begin{cases} Ts_o+J^2/2T, & T>T_{rem} \\ E_g/N=J\sqrt{2s_o}, & T<T_{rem}, \end{cases} \quad (5.3)$$

where the freezing temperature T_{rem} is given by

$$T_{rem}=\left(\frac{\Delta E^2}{2S_o}\right)^{1/2} \quad \text{or} \quad \frac{T_{rem}}{J}=(2s_o)^{-1/2}, \quad (5.4)$$

where $S_o=Ns_o$. The entropy as a function of temperature, $-dF(T)/dT$, is

$$\frac{S(T)}{N}=\begin{cases} s_o-J^2/2T^2, & T>T_{rem} \\ 0 & T<T_{rem}. \end{cases} \quad (5.5)$$

Note that the thermodynamic entropy is always less than s_o, the reduction being due to the fact that higher energy states are less likely to be occupied, with a correspondingly small $-p_\alpha\ln p_\alpha$ contribution to the entropy [39]. The energy as a function of temperature, using $E=-T^2\,\partial/\partial T\,(F/T)$, is given by

$$-\frac{E(T)}{NJ}=\begin{cases} J/T, & T>T_{rem} \\ \sqrt{2s_o}, & T<T_{rem}, \end{cases} \quad (5.6)$$

where $-NJ\sqrt{2s_o}$ is the ground-state energy.

The magnitude of the freezing temperature T_{rem} is determined by the competition between the roughness of the energy landscape of states (characterized by ΔE^2), and the entropy which must be lost to be localized to one state. The process can be visualized as a localization to one branch of a one-level GREM ultrametric tree with e^{Ns_o} branches. (Only one level of the Parisi hierarchical replica-symmetry-breaking scheme is necessary to obtain the correct free energy in the REM.) Replica calculations [12] of the order parameter $q(x)$ also show a discrete jump [40] from 0 to 1 at $x=T/T_{rem}$ (see Fig. 7):

$$q(x)=q_{min}\Theta\left(\frac{T}{T_{rem}}-x\right)+\Theta\left(x-\frac{T}{T_{rem}}\right). \quad (5.7)$$

The $q(x)$ curve determines the probability $P(q)$ of seeing a similarity q between two states through

$$P(q)=\frac{dx(q)}{dq}=\left(\frac{T}{T_{rem}}\right)\delta(q-q_{min})+\left(1-\frac{T}{T_{rem}}\right)\delta(q-1) \quad (5.8)$$

(see inset of Fig. 8), where $x(q)$ is the inverse of $q(x)$. For temperatures below T_{rem}, there is a nonzero probability $(1-T/T_{rem})$ of seeing the polymer localized in one state with a ''self-overlap'' of $q=1$. However, this probability is infinitesimal at $T=T_{rem}$ so that the freezing transition is second order in the thermodynamic sense, even though the order

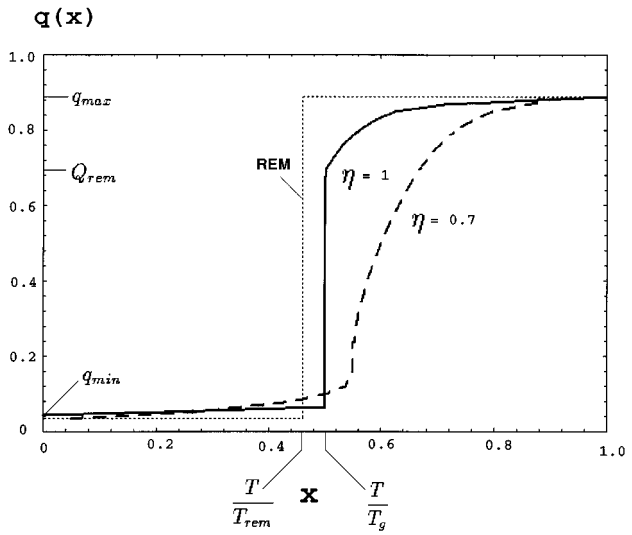

FIG. 7. The order parameter $q(x)$ for a 27-mer. Light-dashed line: The REM $q(x)$ has a jump from q_{min} to q_{max} at $x_{rem}=T/T_{rem}=(0.74/0.81)(T/T_g^o)=0.46$ [using Eq. (5.4) and $T_g^o/\varepsilon=0.74$]. Solid line: For the completely collapsed polymer ($\eta=1$), there is a discrete jump in the order parameter from q_{min} to Q_g^o at $x=T/T_g^o$ (= 0.5 here). Here [for $Q_g^o<q<q_{max}(T)$] the inverse of

$$x(q)=\frac{T}{T_g(q)}=\frac{T}{T_g^o}\left(\frac{ds(q)/dq}{ds(Q_g^o)/dq}\right)^{1/2}$$

is used [see Eq. (5.12)]. One can then show this corresponds to $T/J\cong0.36$ and that $T_g^o\cong0.73J\cong0.74\varepsilon$ (see Fig. 13). Heavy-dashed line: The partially collapsed polymer ($\eta=0.7$) has continuous-type GREM behavior with a more gradual freezing transition. The inverse of Eq. (5.26) is used, with $T/J=0.36$. As the packing density η is increased, the transition from a continuous to a discontinuous order parameter occurs at $\eta_c\cong0.85$.

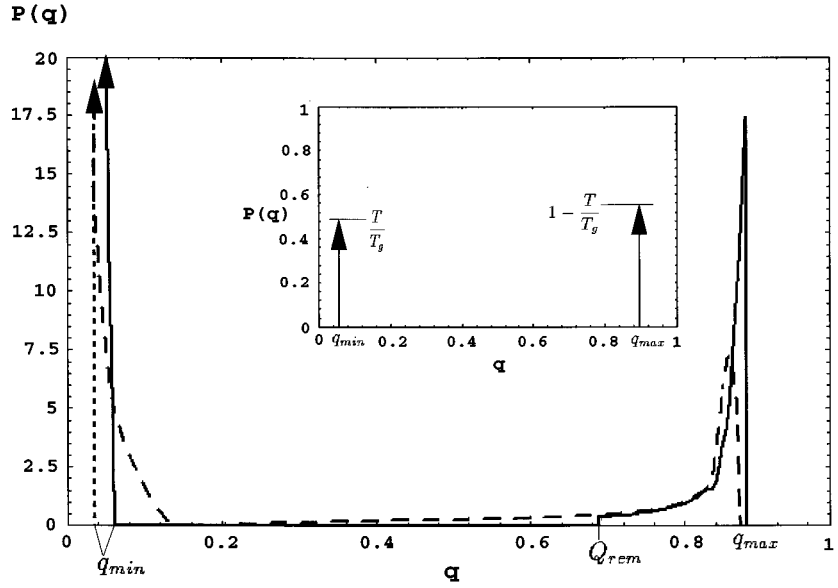

FIG. 8. Probability distribution of similarity parameter q. Inset: for a REM heteropolymer with $T\cong 0.46T_g^o$. Solid line: The completely collapsed ($\eta=1$) heteropolymer (here at $T=\frac{1}{2}T_g^o$) essentially retains a δ function at q_{min}, but the REM spike at q_{max} is spread to a continuum group of states with overlaps $q>Q_g^o$, the total weight of which is slightly less than the REM weight of $1-T/T_{rem}$. Dashed line: In the partially collapsed ($\eta=0.7$) heteropolymer with $T/J\cong 0.36$, there is a finite probability of intermediate q values occurring (more spreading) than in the discrete $\eta=1$ GREM polymer.

parameter $q(x)$ itself undergoes a discrete jump [the entropy (5.5) and energy (5.6) are continuous at T_{rem}] [41].

The derivation of the free energy in the GREM involves essentially the same concepts as in the REM, but with the modification that instead of the *total* energies of different states being uncorrelated, the contributions to the energies $\varepsilon_i^{(\beta)}$ on the branches of the ultrametric tree at the ith level are uncorrelated (see Fig. 9). A freezing temperature associated with each level i involves the same competition between roughness and entropy loss as before, but now the competition is between the decrease in roughness as we move towards a given state an increment dq,

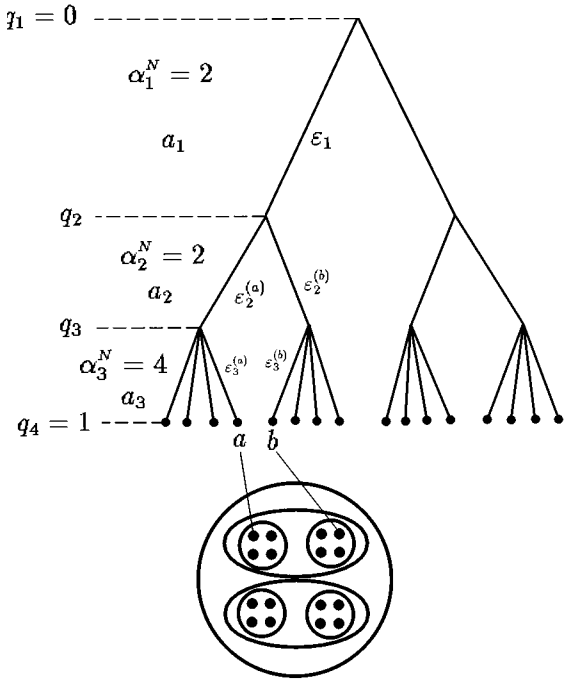

FIG. 9. The ultrametric tree used in GREM — the parameters q_i, α_i^N, a_i, and $\varepsilon_i^{(\beta)}$ are described in Appendix B.

$$\Delta E^2\Big|_{rem} \Rightarrow (\Delta E^2)\frac{dv(q)}{dq}dq,$$

and the loss in entropy for the same increment dq,

$$\Delta S = S_o\Big|_{rem} \Rightarrow -\frac{dS}{dq}dq,$$

so that T_{rem} in (5.4) is replaced by

$$T_g(q)=\left(\frac{(\Delta E^2)\dfrac{dv(q)}{dq}}{-2\dfrac{dS}{dq}}\right)^{1/2}. \tag{5.9}$$

It was shown in Sec. II that in our theory $v(q)=q$, so

$$\frac{T_g(q)}{J}=\left[2\left(-\frac{ds(q)}{dq}\right)\right]^{-1/2}, \tag{5.10}$$

where ds/dq is obtained, numerically if necessary, from the theories of Secs. III and IV.

Take, for example, a fully collapsed ($\eta=1$) polymer of length $N=27$. Using a simple interpolation formula

$$S_{tot}(q)=(1-q)S_{low}(q)+qS_{high}(q) \tag{5.11}$$

between the low q or weakly constrained and the high q or strongly constrained entropy formulas of Secs. III and IV, we obtain an entropy curve as in Fig. 10 and a $T_g(q)$ curve as in Fig. 11. The interpolated entropy formula incidentally gives a crude way to introduce the coupling of collapse and bond formation by considering the weakly constrained polymer to be partially collapsed (say $\eta\cong 0.7$), and the strongly constrained polymer to be completely collapsed ($\eta=1$)— see Fig. 10(b). Notice there is a maximum in the $s(q)$ plot indicating a most probable overlap q_{min} for two globule states, which corresponds to a freezing temperature $T_g(q_{min})=\infty$, i.e., for $q<q_{min}$ states are actually *anticorrelated*, and so this

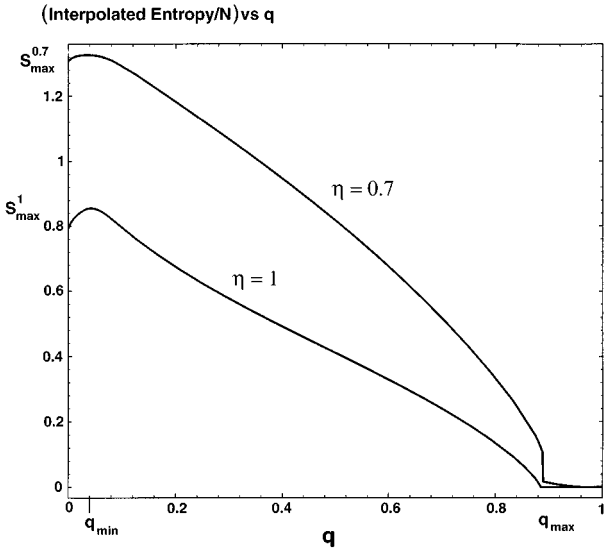

FIG. 10. Interpolated entropies for $\eta=0.7$ and $\eta=1$ [Eq. (5.11)], for the 27-mer. The shape of the $\eta=0.7$ entropy curve at q_{max}, indicating a discontinuity in the freezing temperature there from $T(q_{max})$ to $T=0$, means that a finite temperature $[T(q_{max})]$ will localize the polymer to $\exp S(q_{max})$ states, but there are no lower glass temperatures for the system until $T=0$ (the vertical part of the entropy curve).

region cannot be analyzed by the GREM (in its present form). This situation is analogous to considering p spin models only for the region $t\geqslant 1/2$ in the pair correlation parameter $v=(2t-1)^p$ [17], or for a GREM applied to a spin system in a magnetic field [42], where two states must both have a magnetization m, and thus have an overlap m with the all-spins-up state, and then by ultrametricity have an overlap greater than m with each other (they are on the same branch of the tree). Since any branch of the GREM is still itself a GREM, we can use the analysis of Appendix B to obtain the free energy, but we consider only states that have an overlap of at least q_{min} with each other [which is equivalent to obtaining the free energy at fixed N and η since these parameters determine q_{min} $(q_{min}\approx q_c\sim \eta^{-1/3}N^{-2/3})$] [43]. Notice also that the $T_g(q)$ plot has, in addition to the divergence at q_{min} associated with the statistical overlap between uncorrelated states, a single maximum freezing temperature at $q^{\star}\approx 1/2$. So if we were to cool the system down from high temperatures, the implication is that as the temperature is lowered, the system will undergo a REM-like transition with a discrete jump in the order parameter q from q_{min} to Q_g^o (defined below and in Appendix B), i.e., this is a *discrete-type* GREM. The glass transition in this case is such that the polymer is frozen into basins in the energy landscape which contain collections of states that are all similar at least to the degree Q_g^o.

It was found by Derrida [42] that the Parisi ansatz applied to the GREM reproduced the correct free energy, and the function $x(q)$ which maximized [the equilibrium free energy tends to a maximum in the $\lim_{n\to 0} n(n-1)$ negative dimensionality replica space] the replica-derived expression for the free energy in the *discrete-type* GREM was

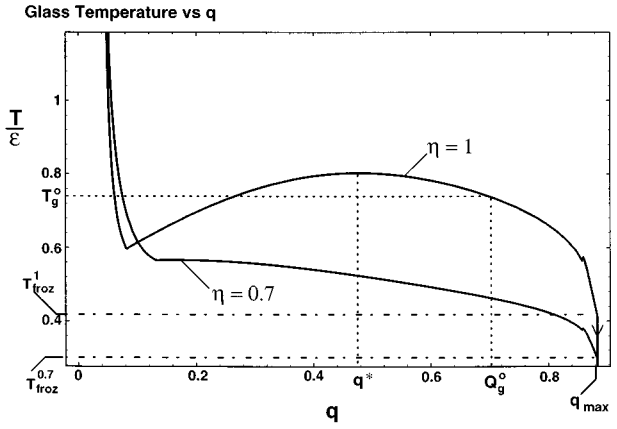

FIG. 11. $T_g(q)$ curves for $\eta=0.7$ and $\eta=1$ [Eq. (5.10)], for the 27-mer. Q_g^o is the similarity parameter where freezing begins at T_g^o. The temperatures $T_{\rm froz}=J/\sqrt{2(-ds(q_{max})/dq)}$, with $J=\sqrt{z\eta}\varepsilon$, where freezing is complete [$q=q_{max}$ and $S=0$ for the $\eta=1$ curve, $q=q_{max}$ and $S=S(q_{max})$ for the $\eta=0.7$ curve], are given for both cases.

$$x(q)=\begin{cases} 0, & 0<q<q_{min} \\ T/T_g^o, & q_{min}<q<Q_g^o \\ T/T_g(q), & Q_g^o<q<q_{max}(T) \\ 1, & q_{max}(T)<q<1. \end{cases} \tag{5.12}$$

[see Fig. 12 for the inverse function $q(x)$], where $T_g(q)$ is Eq. (5.10) in our application, and $q_{max}(T)$ is defined by

$$1=\frac{T}{T_g(q_{max}(T))} \tag{5.13}$$

[$q_{max}(T)$ here is approximately (but not identically) q_{max} as defined earlier in describing the entropy curves]. Q_g^o ($\cong 0.7$ for the collapsed 27-mer) is defined through (see Appendix B)

$$-\frac{ds(Q_g^o)}{dq}=\frac{s_{max}-s(Q_g^o)}{Q_g^o-q_{min}}, \tag{5.14}$$

where $s_{max}=s(q_{min})$, and

$$\frac{T_g^o}{J}=\frac{T_g(Q_g^o)}{J}=\left[\frac{Q_g^o-q_{min}}{2[s_{max}-s(Q_g^o)]}\right]^{1/2}. \tag{5.15}$$

Comparing T_g^o/J with T_{rem}/J in Eq. (5.4), we can see that the decrease in roughness due to correlations lowers T_g^o [since $Q_g^o-q_{min}<1-q_{min}$], but the decrease in the amount of entropy lost at freezing since $Q_g^o<1$ tends to make T_g^o higher (since $s_{max}-s(Q_g^o)<s_{max}$). These two competing effects nearly cancel one another to leave $T_g^o/\varepsilon\cong 0.74$ close to the REM value of $T_{rem}/\varepsilon\cong 0.78$, for the 27-mer. The GREM freezing temperatures increase with N, essentially because of the increase in roughness associated with $z(N)$. Also T_g^o approaches T_{rem} as N increases, since the increase of Q_g^o with N makes the transition more REM-like for larger N (see Fig. 13).

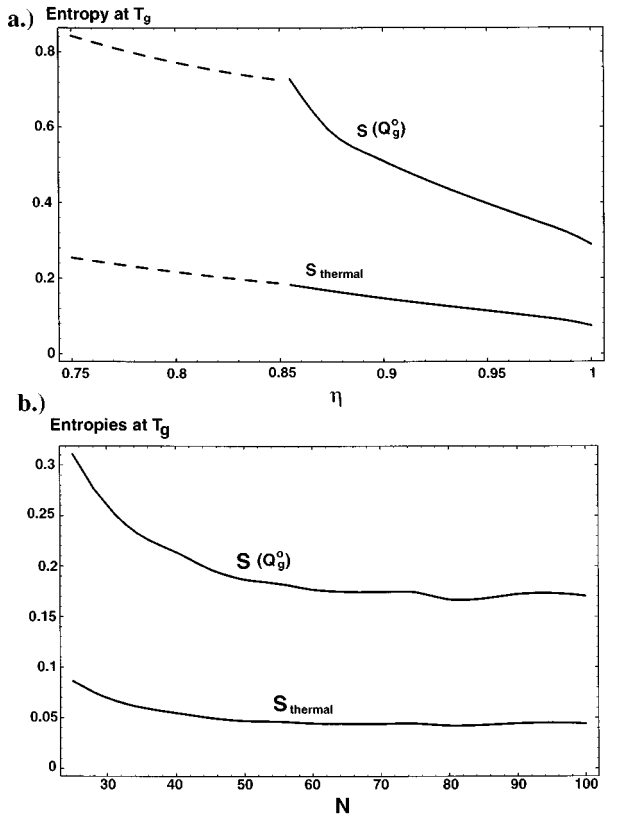

FIG. 12. Entropy of a single basin over total entropy at the freezing temperature, and thermodynamic entropy over total entropy at the freezing temp, plotted here vs packing density η for $N=27$, and vs sequence length N for $\eta=1$. Since Q_g^o increases with N, the basin size decreases as N increases, hence the decrease in entropy. (This is to be compared with the Levinthal entropy for the REM model, which is s_{max}; $s_{thermal}$ for the REM is zero.)

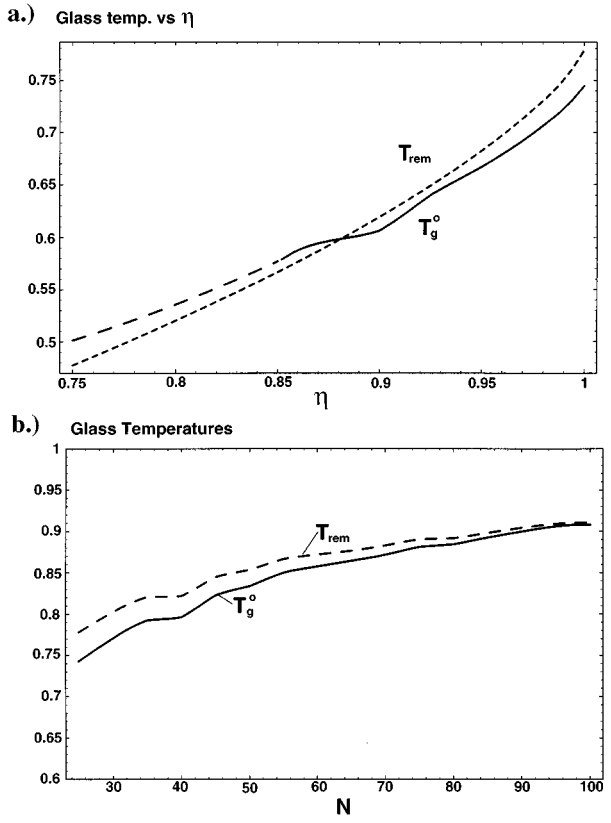

FIG. 13. The temperature T_g^o at the onset of freezing (solid line), plotted with the REM freezing temperature T_{rem} (dashed line), as a function of packing density η and sequence length N. The dashed continuation of the T_g^o curve is the extension below the tricritical point η_c as described in Sec. V. Temperatures are in units of ε. The temperature T_{froz} where total freezing occurs is lower than both T_{rem} and T_g^o (not plotted here).

In addition to the order parameter $q(x)$ for the REM, Fig. 12 shows $q(x)$ for collapse values of $\eta=1$ and $\eta=0.7$, where the order parameter is discontinuous and continuous, respectively. One can then see there must be a critical value of η where the discontinuity in $q(x)$ disappears (analogous to the GREM tricritical point on the de Almeida–Thouless line). This tricritical behavior occurs at $\eta_c \cong 0.85$ for $N=27$ [one can extend this analysis to obtain a tricritical line $\eta_c(N)$].

Down to the tricritical point η_c we can look at the packing density dependence of the REM-like freezing temperature

$$\frac{T_g^o}{\varepsilon}=\left[\frac{z\,\eta(Q_g^o-q_{min})}{2[s_{max}-s(Q_g^o)]}\right]^{1/2}.$$

As η increases, the decrease in entropy lost at T_g^o, the increasingly REM-like behavior of the order parameter characterized by a larger Q_g^o [Figs. 7 and 14], and the fact that the roughness of the landscape increases ($\Delta E^2 \sim \eta$) all cooperate to give a T_g^o that increases with η (see Fig. 13). The behavior closely follows the REM behavior of Eq.(5.4), with a crossover point at $\eta \cong 0.88$.

Below η_c, there is no solution Q_g^o to Eq. (5.14), but the function $T_g(q)$ still has a (weakly peaked) maximum at $q^\star$ ($Q_g^o \cong q^\star$ at $\eta \cong \eta_c$), and we can still compute all the thermodynamic quantities by using this characteristic value q which signals the onset of overlaps of appreciable probability in $P(q)$. However below $\eta \cong 3/4$ the freezing temperatures $T_g(q)$ are monotonically decreasing with q and there is no longer any characteristic value of q to describe REM-like freezing. Throughout these regimes though, the functions $q(x)$ and $P(q)$, as well as thermodynamic quantities, are not significantly different.

The structure of $x(q)$ implies that the order parameter $q(x)$ [the inverse of $x(q)$] has a plateau at $q_{min}(N)>0$, and a discrete jump at $x=T/T_g^o$ from $q_{min} \cong 0.04$ to $Q_g^o \cong 0.7$ indicating $P(q)=0$ in this region. It is fruitful to compare our function $q(x)$ with the GREM analogue applied to a spin glass in a magnetic field (see Ref. [42] and Fig. 7).

All of the thermodynamics obtained above is for a polymer in equilibrium on the time scale(s) during which it is trapped within a single basin. However, there may be ways to escape from a basin kinetically, which leads to an investigation of the search time for the polymer to explore all of its stable basins of attraction. Since $\exp[S(Q_g^o)]$ is the average number of states in a given basin, the (configurational) entropy associated with the total number of basins (i.e., the

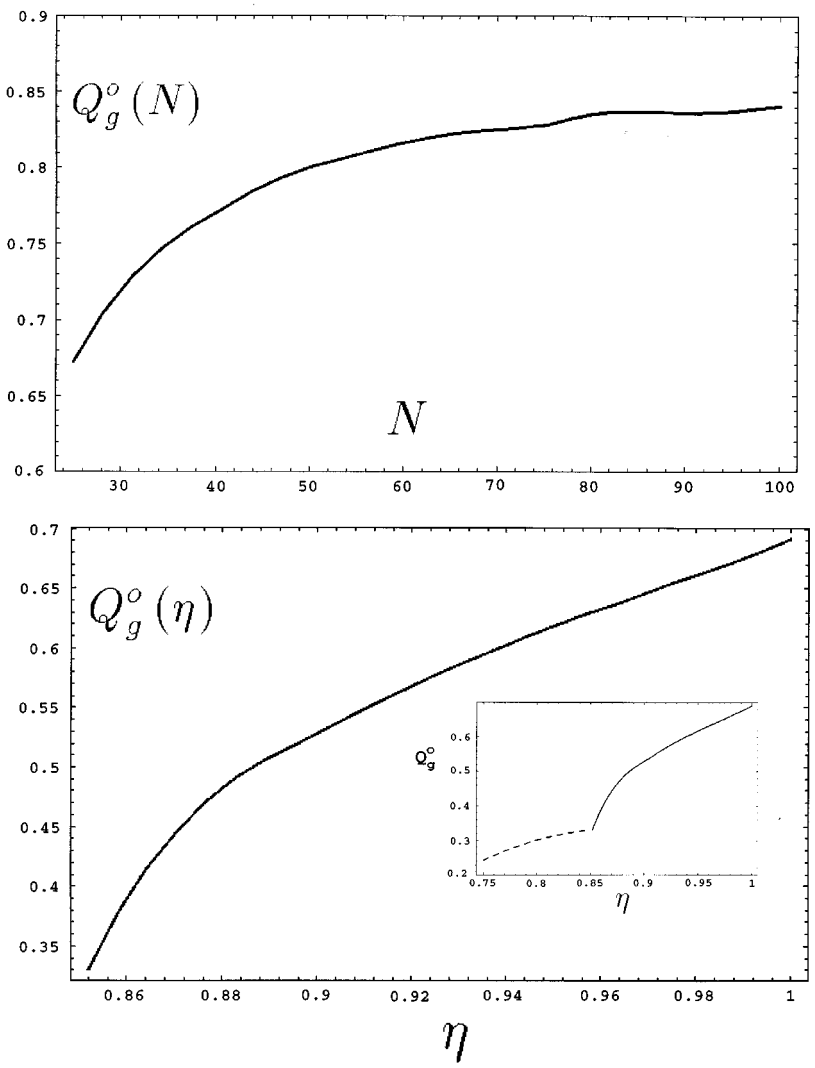

FIG. 14. (a) The minimum similarity Q_g^o within a frozen basin below T_g^o is an increasing function of the sequence length of the polymer N, for the completely collapsed polymer ($\eta=1$). (b) Q_g^o is an increasing function of collapse parameter η for the 27-mer, plotted above the tricritical point $\eta_{27}^\star \cong 0.852$. Below $\eta_{27}^\star$ the nature of the transition changes to continuous freezing, but we can still obtain an approximate Q_g^o value by equating Q_g^o with the broad maximum of the $T_g(q)$ curve (dotted line in inset). The increasing roughness in the landscape as N and η increase leads to more REM-like behavior, with the similarity at freezing getting closer to 1, and the corresponding basin size getting smaller (see Fig. 12).This implies that in the folding model which includes an energy gap, where η is an increasing function of the overlap Q_{nat} with the native state, Q_g^o (and hence T_g^o) will be an increasing function of native similarity Q_{nat}.

Levinthal entropy [21]) is $S_{\text{Levinthal}}=S_{max}-S(Q_g^o)$. A value of $Q_g^o<q_{max}$ indicates a reduction in the number of basins from the REM value S_{max}, to be searched through at the glass transition T_g^o (every state in the uncorrelated landscape is itself a basin). This reduction in basin size on the energy landscape, relevant for a kinetic search below T_g^o, is quite significant for a typical collapsed polymer. For example, a heteropolymer with $N=100$ and 60% helicity has an entropy $S_{max}\approx 70k_{\text{B}}$ [3,15]. Using the interpolated $s(q)$ theory with $\eta\cong 0.7$ for low q and $\eta=1$ for high q (see Fig. 4) $s_{\max}\cong 1.3$ (see Fig. 10) gives an equivalent collapsed cubic lattice polymer of length $N\cong 54$, and a REM Levinthal number of basins $e^{70}\approx 10^{30}$. Using the configurational entropy at T_g^o for a mostly collapsed 54-mer (consistent with Q_g^o) of $\eta\cong 0.9$, the Levinthal entropy $S_{\text{Levinthal}}\cong(2/3)S_{max}=46k_{\text{B}}$, and on a correlated landscape the system must search through $\approx 10^{20}$ basins at the glass transition. Note that the Levinthal entropy increases as η increases, so it is important for the conformational search as to how collapsed the polymer is when it undergoes its glass transtion.

If m-body forces dominate the interaction energies between monomers, we expect that in the limit $m\to\infty$ we should recover the REM results such as $Q_g^o\to q_{max}$, $T_g^o\to T_{rem}=J(2s_{max})^{-1/2}$, $S_{\text{Levinthal}}\to S_{max}=N\ln\mu^{(mf)}$, etc. Using a Kirkwood superposition approximation as mentioned in Sec. II, and using simplifying assumptions such as the probability to form a bond to monomer j is Markovian, overlaps of m-body interactions, in terms of the two-body single-bond overlap q, go as $\sim q^{m-1}$. To see the effect of these m-body forces, we can consider them as the sole contributors to the energy, and modify the pair energy distribution (2.10) by replacing the two-body correlation q with q^{m-1}. This results in a GREM with

$$v(q)=q^{m-1} \quad \text{and} \quad a(q)=(m-1)q^{m-2},$$

which has an m-dependent freezing temperature

$$\frac{T_g^o}{J}=\sqrt{\frac{(m-1)(Q_g^o)^{m-2}}{-2\dfrac{ds(Q_g^o)}{dq}}}=\sqrt{\frac{(Q_g^o)^{m-1}-(q_{min})^{m-1}}{2[s_{max}-s(Q_g^o)]}}$$

and a Levinthal entropy

$$S_{\text{Levinthal}}=S_{max}-S(Q_g^o), \tag{5.16}$$

where Q_g^o is the solution to

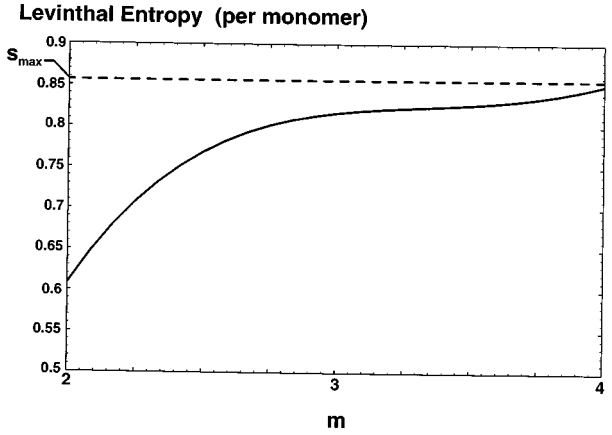

FIG. 15. Levinthal entropy per monomer in units of k_B, as a function of m where the Hamiltonian describing the energetics of the system has m-body interactions. The dotted line is the REM result, the integer values of the solid line are the m-body results. Real proteins may contain both two-body and higher-body interactions. For these cases, the interpolated values of the Levinthal entropy can give some crude idea of the magnitude of many-body effects on the search problem.

$$\frac{(Q_g^o)^{m-1}-(q_{min})^{m-1}}{s_{max}-s(Q_g^o)}=\frac{(m-1)(Q_g^o)^{m-2}}{-\dfrac{ds(Q_g^o)}{dq}}. \qquad (5.17)$$

Q_g^o can be shown to be a monotonically increasing function of m, and above $m\cong 4$, there is no solution to Eq. (5.17), and in this region $Q_g^o=q_{max}$ [42]. The increasing value of Q_g^o gives the order parameter $q(x)$ a larger discontinuity at $x=T/T_g^o$, consistent with the REM description (see Fig. 7). If $q_{max}\cong 1$, in the REM limit $m\to\infty$, T_g^o approaches the REM value of $J(2s_{max})^{-1/2}$. Since Q_g^o increases with m, the configurational entropy per basin $S(Q_g^o)$ decreases, and the Levinthal entropy, measuring the number of thermodynamic basins to be searched below the glass temperature T_g^o, increases to s_{max} above $m\cong 4$ (see Fig. 15).

The general shape of $q(x)$ as well as $P(q)=dx(q)/dq$ can be seen to be ''smoothed out'' versions of the REM results [see Eq. (5.18)] and Fig. 8). $P(q)$ has a δ function at q_{min} as in the REM, but instead of pure states at q_{max} having a finite weight $Y=\Sigma_\alpha P_\alpha^2=1-T/T_g^o$ below T_g^o as in the REM, the weight is spread out among a group of similar ergodically confined states with high overlap ($q>Q_g^o$). (There is also some spreading near q_{min} which increases with decreasing η.) There will be a plateau, however, if $T<T_{\rm froz}=T(q_{max})$, where $s(T)=0$. This can be seen from the $T_g(q)$ curve [Fig. 11], which drops vertically at q_{max}, meaning that $x(q)$ has a vertical step at q_{max} and $q(x)$ has a plateau in this region [see also the inverse function $q(T)$ in Fig. 16]. The finite weight of a pure state for $T<T_{\rm froz}$ is consistent with the vanishing entropy in this region of temperature. The form of $P(q)$ is given by

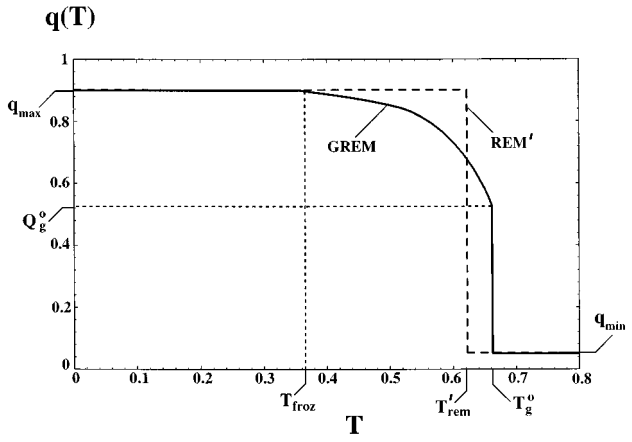

FIG. 16. $q(T)$, defined as the inverse of $T_g(q)$ (see Fig. 11) for the 27-mer at $\eta=0.9$, for the REM and GREM, the REM being defined here with a linear $s(q)$ from s_{max} at q_{min} to 0 at q_{max}. All temperatures are in units of ε. This is similar to the function $q(x)$ (Fig. 7) but here we can see there exists a plateau below $T=T_{\rm froz}$. Note the temperature in the GREM where the freezing begins is higher than T'_{rem}, but the GREM $T=T_{\rm froz}$ is much lower.

$$P(q)=\frac{dx(q)}{dq}$$
$$=\begin{cases}0, & q<q_{min}\\ \text{spike of weight } T/T_g^o, & q=q_{min}\\ 0, & q_{min}<q<Q_g^o\\ \dfrac{TT_g(q)}{J^2}\left(-\dfrac{d^2s(q)}{dq^2}\right), & Q_g^o<q<q_{max}(T)\\ 0, & q_{max}(T)<q<1.\end{cases} \qquad (5.18)$$

As mentioned in Appendix B, a monotonically increasing $T_g(q)$ curve will give a REM-like freezing with a jump in order parameter $q(x)$ as in Fig. 7. A REM $x(q)$ dependence can be obtained from a linearly decreasing $s(q)$ (Fig. 17) as follows. The shape of $x(q)$ for the REM has two steps, one from 0 to T/T_{rem} at q_{min} and another from T/T_{rem} to 1 at q_{max}, with horizontal pieces from 0 to q_{min} at $x=0$ and q_{min} to q_{max} at $x=T/T_{rem}$. Since each state in the REM is itself one basin, $S_{\rm Levinthal}=S_{max}$ at T_{rem}, and therefore in the GREM language $Q_g^o|_{rem}=q_{max}$. So there is no second region between Q_g^o and q_{max} in $x(q)$ in Eq. (5.12), and there is only one freezing temperature, which by Eq. (5.15) is

$$\frac{T_{rem}}{J}=\frac{T_g(q_{max})}{J}=\left(\frac{q_{max}-q_{min}}{2s_{max}}\right)^{1/2},$$

which is just Eq. (5.4) when $q_{max}=1$ and $q_{min}=0$. It also follows that the bilinear (two-slope) approximation for $s(q)$ when N is large (Fig. 17) has two freezing temperatures, and by the above arguments can be seen to be equivalent to a two-tier GREM [26].

For $N=27$ and $\eta=1$, the glass temperature curve $T_g(q)$ and the order parameter $q(x)$ justify that the free energy is

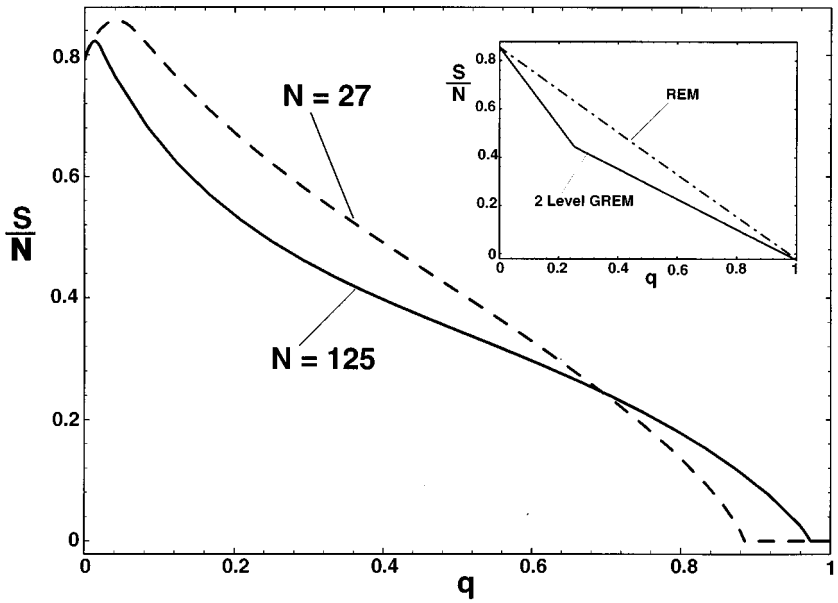

FIG. 17. Interpolated entropies for $N=27$ and $N=125$ [Eq. (5.11)]. The entropy curve imitates a bilinear form for larger N. The small discrepancy between s_{max} for the curves is due to the N-dependent confinement theory: $s_{max}\rightarrow \ln\mu^{(mf)}$ as N approaches the bulk limit. The range in q where confinement is important $\rightarrow 0$ as $N\rightarrow\infty$, while q_{max} (where $s\cong 0$) $\rightarrow 1$ as $N\rightarrow\infty$. Note that q_{max} increases with N approximately in the manner described by the high q formula; the decreasing value of q_v where the entropy crisis occurs in the low q formula only affects the slope of the interpolated entropy for small q.

obtained by applying to the polymer a discrete-type GREM in a ''magnetic field'' of strength q_{min}, with the result

$$
\begin{aligned}
-\frac{F(T)}{N} &= Ts_{max}+\frac{J^2}{2T}(q_{max}-q_{min}), && T>T_g^o,\\
-\frac{F(T)}{N} &= Ts(q(T))+\frac{J^2}{2T}[q_{max}-q(T)]\\
&\quad +\sqrt{2}J\int_{Q_g^o}^{q(T)} dq\left(-\frac{ds(q)}{dq}\right)^{1/2}\\
&\quad +\frac{J^2}{T_g^o}(Q_g^o-q_{min}), && T_{\text{froz}}<T<T_g^o,\\
-\frac{F}{N} &= \sqrt{2}J\int_{Q_g^o}^{q_{max}} dq\left(-\frac{ds(q)}{dq}\right)^{1/2}\\
&\quad +\frac{J^2}{T_g^o}(Q_g^o-q_{min}), && T<T_{\text{froz}},
\end{aligned}
\tag{5.19}
$$

where $q(T)$ is the inverse of $T_g(q)$, $s(q)=S(q)/N$ is the specific entropy (per monomer) as in Fig. 10, q_{max} is the q value where $s(q_{max})=0$, $T_{\text{froz}}=T(q_{max})$, and $J^2=z\eta\varepsilon^2$ (J is defined slightly differently in Appendix B).

Below the temperature T_g^o, overlaps with values of Q_g^o and greater begin to be seen with finite weight, illustrating that the system begins to be confined within basins in the energy landscape containing configurations at least as similar as Q_g^o (see Fig. 14). Since $Q_g^o<q_{max}$, there are still many states within these basins, and a corresponding finite entropy left over below the glass transition temperature T_g^o. Since the glass transition is thermodynamically second order as it was in the REM, we can obtain the thermodynamic entropy left over at the freezing temperature, $s(T_g^o)$, by taking the temperature derivative of either the high temperature or partially frozen phase and evaluating at T_g^o:

$$
\frac{S(T_g^o)}{N}=-\frac{\partial}{\partial T}\left(\frac{F}{N}\right)\Bigg|_{T_g^o}=s(Q_g^o)-\frac{1}{2}\left(\frac{J}{T_g^o}\right)^2(q_{max}-Q_g^o),
\tag{5.20}
$$

which has a reduction from the raw configurational entropy per basin $s_{\text{basin}}=s(Q_g^o)$. The thermodynamic entropy depends on N and η mostly through Q_g^o, T_g^o, and the $z(N)$ and η dependence of J (see Fig. 13). So a feature of the GREM is that there is now a nonzero entropy below the glass temperature T_g^o, where a collective transition takes place to localize states to within similar values, but not immediately to one pure state.

We can also find the energy in the polymer at T_g^o through

$$
\frac{E(T_g^o)}{N}=\frac{1}{N}\frac{\partial}{\partial(1/T)}\left(\frac{F}{T}\right)\Bigg|_{T_g^o}=-\frac{J^2}{T_g^o}(q_{max}-q_{min})
\tag{5.21}
$$

and the ground-state energy

$$
\begin{aligned}
\frac{E_{\text{GS}}}{N}&=\lim_{T\rightarrow 0}\frac{F(T)}{N}=-\frac{J^2}{T_g^o}(Q_g^o-q_{min})\\
&\quad -\sqrt{2}J\int_{Q_g^o}^{q_{max}} dq\left(-\frac{ds}{dq}\right)^{1/2}\\
&=-2T_g^o(s_{max}-s(Q_g^o))\\
&\quad -\sqrt{2}J\int_{Q_g^o}^{q_{max}} dq\left(-\frac{ds}{dq}\right)^{1/2}
\end{aligned}
\tag{5.22}
$$

which is above the REM ground-state energy $E_{\text{GS}}/N=-2T_{\text{rem}}s_{\max}$ (Fig. 18).

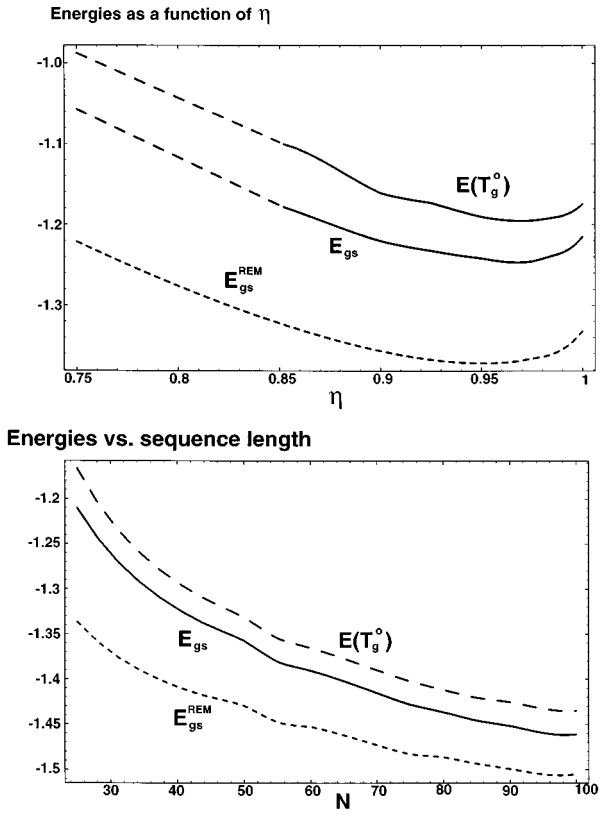

FIG. 18. Ground-state energy for the GREM and REM, as well as the energy at the glass transition, vs packing density (for the 27-mer), and sequence length [for a densely packed ($\eta=1$) polymer]. All energies are per monomer and in units of ε. The minimum at η somewhat less than 1 is due to the competition between the increase in ground-state energy as the polymer becomes more dilute (because the width of the Gaussian random energy distribution becomes narrower as the number of bonds decreases), and the decrease in ground-state energy as the total number of states which must be frozen out at the glass transition increases (with decreasing η).

From the free energy $F(T)$ in the free, partially frozen, and completely frozen regimes, we can obtain the thermodynamic entropy and energy as functions of temperature. Above the freezing temperature T_g^o the results are as in the REM, with overlap $q=q_{min}$ and $S(T)$ and $E(T)$ in the high temperature phase. So that the high temperature results agree, we have modified the REM analysis by making $s(q)$ a linearly decreasing function of q from s_{max} to 0 on the interval $\{q_{min},q_{max}\}$ instead of $\{0,1\}$. This increases the slope ds/dq and thus lowers T_{rem} below T_g^o, so we will call the REM transition temperature T'_{rem} here. In the example in Figs. 19 and 20, $N=27$ and $\eta=0.9$; the freezing begins above T'_{rem} and there is a gradual transition down to T_{froz} below which the system is completely frozen into one state. In the partly frozen regime between T_g^o and T_{froz}, all temperatures T are freezing temperatures $T(q)$ [Eq. (5.10)] :

$$T(q)=\frac{J}{\sqrt{-2\dfrac{ds(q)}{dq}}}, \tag{5.23}$$

as the overlap q monotonically increases the freezing temperatures monotonically decrease, and $q(T)$ in Eq. (5.19) is understood as the inverse of $T(q)$ above (Eq. (5.23) see Fig. 18). Below T_{froz}, $q=q_{max}$, $S(T)=0$, and $E(T)=E_{\text{GS}}$. The form of the entropy as a function of T is

$$\frac{S(T)}{N}=s_{max}-\frac{J^2}{2T^2}(q_{max}-q_{min}), \quad T>T_g^o,$$

$$\frac{S(T)}{N}=s[q(T)]-\frac{J^2}{2T^2}[q_{max}-q(T)], \quad T_{\text{froz}}<T<T_g^o,$$

$$\frac{S(T)}{N}=0, \quad T<T_{\text{froz}} \tag{5.24}$$

and the energy vs T is

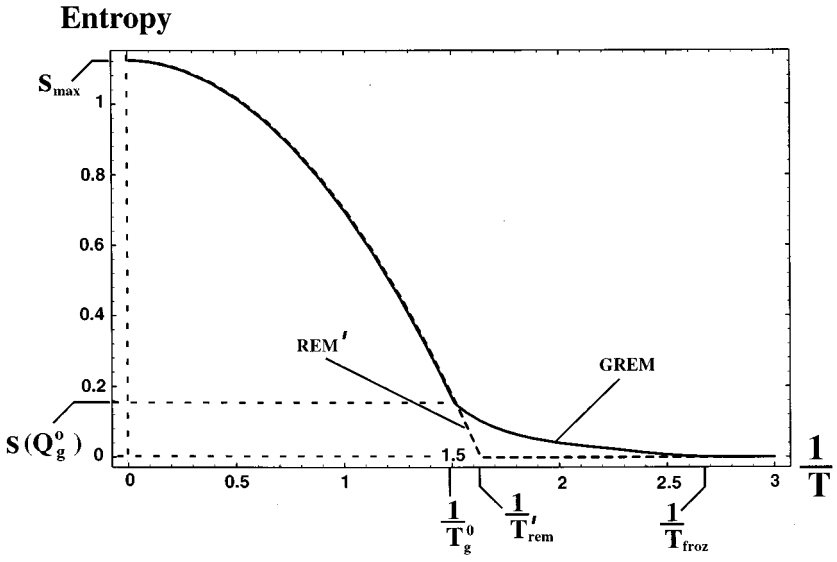

FIG. 19. Entropy per monomer, in units of Boltzmann's constant, of a GREM heteropolymer ($N=27$, $\eta=0.9$) plotted vs $1/T$, where temperatures are in units of ε. In the GREM the polymer is more gradually localized to one state at a lower temperature than in the REM. This is because the energy correlations between states smooth the energy landscape in the vicinity of a state, making glassy localization to that state happen at lower temperatures than in the uncorrelated landscape.

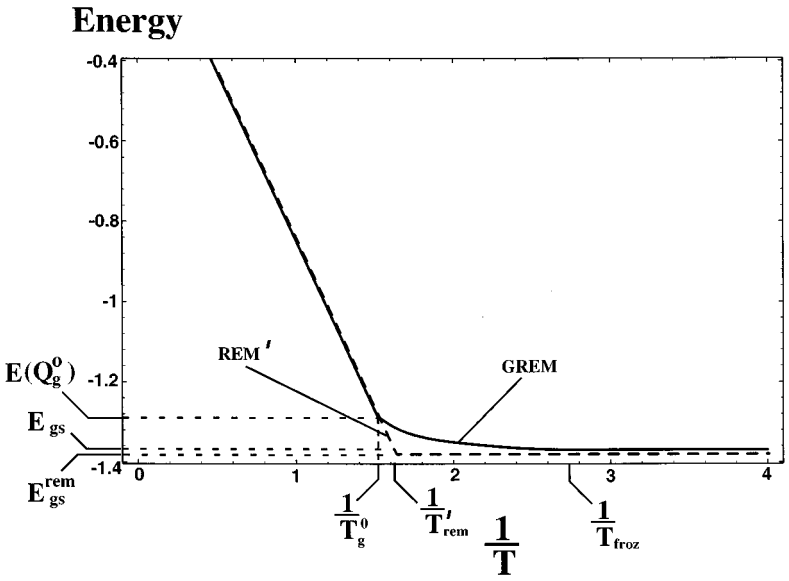

FIG. 20. Energy per monomer, in units of ε, of a GREM heteropolymer ($N=27$, $\eta=0.9$) plotted vs $1/T$ (T also in units of ε). In the GREM, the polymer approaches its ground-state energy more gradually than in the REM, because in the correlated landscape the polymer is not yet trapped to its ground-state at temperatures below T'_{rem}, the smoothness of the energy landscape allowing the polymer to still explore many states at colder temperatures.

$$\frac{E(T)}{NJ}=-\frac{J}{T}(q_{max}-q_{min}),\quad T>T_g^o,$$

$$\frac{E(T)}{NJ}=-\frac{J}{T}[q_{max}-q(T)]-\frac{J}{T_g^o}(Q_g^o-q_{min})-\sqrt{2}\int_{Q_g^o}^{q(T)}dq\left(-\frac{ds(q)}{dq}\right)^{1/2},\quad T_{\rm froz}<T<T_g^o,$$

$$\frac{E}{NJ}=-\frac{J}{T_g^o}(Q_g^o-q_{min})-\sqrt{2}\int_{Q_g^o}^{q_{max}}dq\left(-\frac{ds(q)}{dq}\right)^{1/2},\quad T<T_{\rm froz},\qquad (5.25)$$

where $T_{\rm froz}=T_g(q_{max})$.

We can also apply the GREM to a partially collapsed 27-mer with, say, $\eta=0.7$, and obtain the corresponding quantities as above. The interpolated configurational entropy curve of Eq. (5.11) (see Fig. 10) corresponds to a larger total number of states since the polymer is less compact, and also has a minimum overlap $q_{min}\cong 0.04$ where the entropy is a maximum ($s_{max}\cong\ln\mu^{(mf)}=1.31$). The corresponding freezing temperature curve $T_g(q)$ is indicative of a *continuous-type GREM* with the exception that there is a diverging glass temperature $T_g(q_{min})=\infty$ (see Fig. 11), which we deal with as before by considering this the finite-size analogue of a spin system in a magnetic field. This is seen most clearly by investigating the order parameter $q(x)$ (see Fig. 7), defined as the inverse of

$$x(q)=\begin{cases}0, & 0<q<q_{min}\\ T/T_g(q), & q_{min}<q<q_{max}(T)\\ 1, & q_{max}(T)<q<1,\end{cases}\qquad (5.26)$$

where q_{max} is as defined before, and $T_g(q)$ is Eq. (5.10).

In the random heteropolymer with $N=27$ and $\eta=0.7$, the monotonically decreasing $T_g(q)$ curve (Fig. 11) indicates a continuous-type GREM, with no completely free high temperature phase since $T_g(q_{min})=\infty$, and the free energy (*relative* to that of the fully constrained state) in the remaining two phases (partly frozen and completely frozen) given by

$$-\frac{F(T)}{N}=Ts[q(T)]+\frac{J^2}{2T}[q_{max}-q(T)]+\sqrt{2}J\int_{q_{min}}^{q(T)}dq\left(-\frac{ds(q)}{dq}\right)^{1/2},\quad T_g(q_{max})<T<T_g(q_{min}),$$

$$-\frac{F}{N}=\sqrt{2}J\int_{q_{min}}^{q_{max}}dq\left(-\frac{ds(q)}{dq}\right)^{1/2},\quad T<T_g(q_{max}),\qquad (5.27)$$

where $q(T)$, $s(q)$, q_{max}, and J are defined as before.

VI. CONCLUSION

We have analyzed the thermodynamics of a mesoscopic random heteropolymer by combining the generalized random energy model of a correlated energy landscape with the appropriate polymer physics of a simple collapsed polymer. For higher collapse density the glass transition is a first-order-like random phase transition (with respect to the order parameter) like the transition exhibited by the random energy model. A feature is the emergence of a tricritical point at lower packing density, where the transition becomes continuous. The physical observables such as the probability distribution of the order parameter $P(q)$ are not dramatically different quantitatively on either side of the transition. The transition temperature T_g^o for the first-order REM-like transition is within about 5% of the REM value T_{rem}, which is reassuring for the previous thermodynamic description of real proteins by the REM. In fact, even when the transition is continuous the large values of $P(q)$ occur near the REM δ functions. The continuous REM transition coupled with collapse may be related to the unusual non-self-averaging be-

havior manifest in the sensitivity of collapse to single-site mutations in staphylococcal nuclease observed by Engleman and co-workers [44]. We must also bear in mind that their system is a natural protein and thus is minimally frustrated, which gives an another reason why the collapse would be non-self-averaging.

While the thermodynamic transition temperature is not very different from the REM value, the properties of the basins of attraction are quantitatively modified. The Levinthal number, measuring the number of basins to be searched at the glass transition, is significantly reduced (by a factor of about 1/3 on a logarithmic scale for collapsed protein-sized molecules). This means it may be possible to get thermodynamically proteinlike behavior for a larger fraction of considerably longer random chains than would have been expected. Experiments on sampling random polypeptides and studying their thermodynamics [45,46] have indeed given many more sequences with a first-order-like transition than naively anticipated. According to the GREM, however, it is likely that the basins into which the polypeptide freezes at the transition still have considerable conformational freedom, as manifest by the large entropy left over after the transition. For the 27-mer in low density assemblies ($\eta=0.75$), the residual entropy of a single basin is $\cong 0.8 s_{max}$, while at high density ($\eta=1$) this entropy is $\cong 0.3 s_{max}$. Similarly, we expect the correlations to considerably reduce the size of the barriers between basins, an issue we shall investigate quantitatively within the GREM in a future paper.

One technical point regarding the GREM analysis relates to the fact that we have analyzed polymers of finite size (N). Because of the significant surface to volume ratio of biopolymers, thermodynamic properties within the analysis depend moderately upon N. There can be other specific effects due the finite size that will act to round the transitions, which will come from the introduction of defects in the frozen order. Some of these effects may be correctly handled by the high Q analysis, but higher order correlations, reflecting the possibility of a type of freezing into different low energy reference structures for the melted regions, would have to be taken into account (this effect requires at least triplet correlations between the energy levels). We should also note that the high Q analysis may be useful in describing hydrogen exchange experiments on proteins with low denaturant concentrations [47].

The GREM analysis is only approximate, but it also allows us to address important questions. An especially important issue is the approximate treatment of barriers between local minima, above the glass transition [48]. It is also possible to use it to estimate the fraction of sequences which are sufficiently minimally frustrated to fold kinetically, and to discuss the shape of the free energy surfaces and folding funnels of minimally frustrated random heteropolymers. The quantitative application of the theory to natural proteins also raises additional questions concerning partial order in protein molten globules [15,16]. Not only do such states exhibit local secondary structure (increasing the rigidity of the backbone), but also liquid crystallinity [15,16] and microphase separation [49]. We believe each of these effects can be accommodated within the GREM formalism by modifying the Flory-style analysis of the configurational entropy $S(q)$, a problem which we hope to return to in the future.

ACKNOWLEDGMENTS

We wish to thank B. Derrida, J. Onuchic, J. Saven, and N. Socci for helpful discussions. This work was supported by NIH Grant No. 1RO1 GM44557 and NSF Grant No. DMR-89-20538.

APPENDIX A

Consider two different states of the polymer that have μ bonds in common. Assuming that the μ energies of interaction corresponding to the μ identical bonds in both ''copies'' of the polymer are the same, let us define a parameter $\Phi=\Sigma\varepsilon_{ij}s_{ij}$ which is the contribution to the total energy that is the same for both states. Next, let $E_a=\Phi+\phi_a$ and $E_b=\Phi+\phi_b$, where $\phi_{a,b}$ is the contribution to the total energy from the remaining $Nz\eta-\mu$ bonds. The probability of states a and b having energies E_a and E_b is then

$$P_{a,b}(E_a,E_b)=\int d\phi_a\int d\phi_b\int d\Phi P(\phi_a)P(\phi_b)P(\Phi)\times\delta(E_a-(\Phi+\phi_a))\delta(E_b-(\Phi+\phi_b)), \tag{A1}$$

where $P(\Phi)$ and $P(\phi_{a,b})$ are Gaussian probability distributions with variances $\langle\Phi^2\rangle=\mu\varepsilon^2$ and $\langle\phi_{a,b}^2\rangle=(zN\eta-\mu)\varepsilon^2$. Integrating out ϕ_a and ϕ_b using the δ functions,

$$\begin{aligned}P_{a,b}(E_a,E_b)&=\text{const}'\times\int d\Phi\exp\left[-\frac{(E_a-\Phi)^2}{2(Nz\eta-\mu)\varepsilon^2}-\frac{(E_b-\Phi)^2}{2(Nz\eta-\mu)\varepsilon^2}-\frac{\Phi^2}{2\mu\varepsilon^2}\right]\\&=\text{const}\times\exp\left[-\frac{1}{4Nz\eta\varepsilon^2}\times\left(\frac{(E_a+E_b)^2}{1+q}+\frac{(E_a-E_b)^2}{1-q}\right)\right].\end{aligned} \tag{A2}$$

APPENDIX B

In the GREM, one can consider the $(\mu^{(mf)})^N$ states of a polymer as the end points of an ultrametric tree of n levels (see Fig. 9) [50]. To each level i ($1\leqslant i\leqslant n$) of the tree one associates three quantities α_i, a_i, and q_i. Two configurations a and b have an overlap $q_{ab}=q_i$, where q_i is the level on the tree where the branches coming from a and b join. q_i is an increasing function of i with $0=q_1<q_2\cdots<q_{n+1}=1$. At the ith level one branch divides into α_i^N branches, so at level i there are $(\alpha_1\alpha_2\cdots\alpha_i)^N$ branches, and $(\alpha_1\alpha_2\cdots\alpha_n)^N=(\mu^{(mf)})^N$.

On each branch of the tree at level i, one chooses a random variable $\varepsilon_i^{(b)}$ according to a distribution $\rho_i(\varepsilon_i^{(b)})$ whose width is a_i:

$$\rho_i(\varepsilon_i^{(b)})=\frac{1}{(\pi a_i NJ^2)^{1/2}}\exp\left(-\frac{(\varepsilon_i^{(b)})^2}{a_i NJ^2}\right). \tag{B1}$$

The energy of each configuration b is then given by

$$E_b=\sum_{i=1}^{n}\varepsilon_i^{(b)}, \tag{B2}$$

where the $\varepsilon_i^{(b)}$ are the energies associated with the n branches that connect each state to the top of the tree. States a and b with overlap $q_{ab}=q_i$ have $\varepsilon_j^{(a)}=\varepsilon_j^{(b)}$ for $j\leqslant i-1$ and $\varepsilon_j^{(a)}\neq\varepsilon_j^{(b)}$ for $j\geqslant i$. The model is defined once the two sequences α_i and a_i are given for $1\leqslant i\leqslant n$. If we choose the normalization

$$\sum_{i=1}^{n}a_i=1, \tag{B3}$$

then the energies E_b of the $(\mu^{(mf)})^N$ states are distributed as Gaussian random variables:

$$P_a(E_a)=\frac{1}{(\pi NJ^2)^{1/2}}\exp\left(-\frac{(E_a)^2}{NJ^2}\right). \tag{B4}$$

The probability distribution $P_{ab}(E_a,E_b)$ that two configurational states a and b have energies E_a and E_b is

$$P_{a,b}(E_a,E_b)=\text{ const}\times\exp\left[-\frac{1}{2NJ^2}\times\left(\frac{(E_a+E_b)^2}{1+v_{ab}}+\frac{(E_a-E_b)^2}{1-v_{ab}}\right)\right], \tag{B5}$$

where v_{ab} is a measure of the correlation in energy between two configurations with overlap q_i:

$$v_{ab}=v_i=\sum_{j=1}^{i-1}a_j. \tag{B6}$$

Given a configuration (a), the number of configurations that have an overlap of q_i with (a) is

$$\Omega_i=e^{Ns_i}=(\alpha_i^N-1)(\alpha_{i+1}\cdots\alpha_n)^N. \tag{B7}$$

This is the number of states to which formulas (B5) and (B6) apply. In the thermodynamic limit ($N\to\infty$), the entropy at level i is

$$s_i=\sum_{j=i}^{n}\ln\alpha_j. \tag{B8}$$

We assume this equation holds approximately for fairly large N. For N fairly large there is almost a continuous range of possible overlaps ($0\leqslant q\leqslant 1$) which means the number of levels in the ultrametric tree is large [51]. So s_i and v_i may be treated as continuous quantities $s(q)$ and $v(q)$, which means

$$\ln\alpha(q)=-\frac{ds(q)}{dq}\quad\text{and}\quad a(q)=\frac{dv(q)}{dq}. \tag{B9}$$

The GREM free energy and its derived quantities are discussed in the section on thermodynamics. In brief, there are two cases where the GREM has been solved [a third scenario is if $T(q)$ is monotonically increasing or constant — in this case we just retrieve the REM results [11]].

Continuous-type GREM

If the freezing (glass) temperature as a function of q, defined by

$$T(q)=\frac{J}{2}\left(\frac{a(q)}{\ln\alpha(q)}\right)^{1/2}=\frac{J}{2}\left(\frac{\frac{dv(q)}{dq}}{-\frac{ds(q)}{dq}}\right)^{1/2} \tag{B10}$$

is a monotonically decreasing function of the overlap q, then the freezing occurs from the top of the ultrametric tree downward (most dissimilar states freeze out first), and the thermodynamic free energy is given by

$$\begin{aligned}
-\frac{F}{N}&=Ts_o+\frac{J^2}{4T}, && T>T(0),\\
-\frac{F}{N}&=Ts(q(T))+\frac{J^2}{4T}[v(1)-v(q(T))]\\
&\quad+J\int_0^{q(T)}dq\left(-\frac{ds(q)}{dq}\frac{dv(q)}{dq}\right)^{1/2}, && T(1)<T=T(q),\\
-\frac{F}{N}&=J\int_0^1 dq\left(-\frac{ds(q)}{dq}\frac{dv(q)}{dq}\right)^{1/2}, && T<T(1),
\end{aligned} \tag{B11}$$

where $q(T)$ is the inverse of $T(q)$, $s_o=\ln\mu^{(mf)}$, and $s(q)=S(q)/N$ is the specific entropy (per monomer) obtained from the theories in Secs. III and IV. At the highest temperatures [i.e., those higher than $T(q=0)$ if $T(q=0)<\infty$] the system can freely explore all of its states regardless of dissimilarity. At lower temperatures there is a continuous freezing which gradually causes states to be more localized.

Discrete-type GREM

The function $T(q)$ has a single maximum, say, at $q^\star$. We would expect based on the comments in Sec. V that there will be a REM transition with a discreet jump in the order parameter q, and then a gradual freezing as in the continuous-type GREM above. Define q_g^o such that

$$\frac{v(q_g^o)-v(0)}{s(0)-s(q_g^o)}=\frac{\frac{dv}{dq}(q_g^o)}{-\frac{ds}{dq}(q_g^o)}. \tag{B12}$$

q_g^o is always greater than $q^\star$. Now define what will be a REM-like transition temperature, where the freezing will have a sudden onset at q_g^o:

$$T_g^o=\frac{J}{2}\left(\frac{\frac{dv}{dq}(q_g^o)}{-\frac{ds}{dq}(q_g^o)}\right)^{1/2}=\frac{J}{2}\left(\frac{v(q_g^o)-v(0)}{s(0)-s(q_g^o)}\right)^{1/2}. \tag{B13}$$

Then the thermodynamic free energy is

$$-\frac{F}{N}=Ts_o+\frac{J^2}{4T}, \qquad T>T_g^o$$

$$-\frac{F}{N}=Ts(q(T))+\frac{J^2}{4T}[v(1)-v(q(T))]$$
$$+J\int_{q_g^o}^{q(T)}dq\left(-\frac{ds(q)}{dq}\frac{dv(q)}{dq}\right)^{1/2}$$
$$+J\{[v(q_g^o)-v(0)][s(0)-s(q_g^o)]\}^{1/2},$$
$$T(1)<T<T_g^o, \quad \text{(B14)}$$

$$-\frac{F}{N}=J\int_{q_g^o}^{1}dq\left(-\frac{ds(q)}{dq}\frac{dv(q)}{dq}\right)^{1/2}$$
$$+J\{[v(q_g^o)-v(0)][s(0)-s(q_g^o)]\}^{1/2}, \quad T<T(1).$$

If there is a limit q_{min} as to how uncorrelated two states can be, e.g., an SK spin glass in a magnetic field or a finite polymer, the above formulas are only slightly modified by effectively replacing the lower limits of 0 with q_{min}. This is described in Sec. V.

APPENDIX C

Consider the region of a cross linked polymer around a given cross link about to be formed (see Fig. 21). Each cross link and its two associated monomers [52] has four neighboring cross links C_1, C_2, C_3, C_4. We seek the fraction of allowable states that are consistent with the formation of a specific cross link at A, $\omega_A\Delta\tau$. We assume that if the specific monomers that A joins are within a volume $\Delta\tau$ of each other then a cross link is formed. If we consider the system to be composed of four separate chains, $\omega_A\Delta\tau$ equals the probability that all four chains meet in $\Delta\tau$, divided by the probability that the chains meet in pairs (restoring the allowable configurations in the unbonded initial structure). If the chains are Gaussian, the propagator from the origin to position $\mathbf{r}_i$ for a polymer chain of n_i statistical segments is

$$G(\mathbf{r}_i|n_i)=\left(\frac{3}{2\pi\langle r_i^2\rangle}\right)^{3/2}\exp\left(-\frac{3\mathbf{r}_i^2}{2\langle r_i^2\rangle}\right) \qquad \text{(C1)}$$

with $\langle r_i^2\rangle=n_ib^2$, where b is the length of one segment. So the probability of forming a cross link is then given by

$$\omega_A\Delta\tau=\Delta\tau\frac{\int d\tau G_1(\mathbf{r}_1)G_2(\mathbf{r}_2)G_3(\mathbf{r}_3)G_4(\mathbf{r}_4)}{\int d\tau G_1(\mathbf{r}_1)G_2(\mathbf{r}_2)\int d\tau G_3(\mathbf{r}_3)G_4(\mathbf{r}_4)}, \qquad \text{(C2)}$$

where $G_i(\mathbf{r}_i)$ is a Gaussian function extending from position C_i, and $\mathbf{r}_i$ is the vector from the position of C_i to the volume element $\Delta\tau$. The integrations extend over all space — to perform them set up an origin O at the most probable position of the junction A, defined such that from O

$$\frac{\mathbf{R}_1}{\langle R_1^2\rangle}+\frac{\mathbf{R}_2}{\langle R_2^2\rangle}+\frac{\mathbf{R}_3}{\langle R_3^2\rangle}+\frac{\mathbf{R}_4}{\langle R_4^2\rangle}=\mathbf{0}, \qquad \text{(C3)}$$

where $\mathbf{R}_i$ goes from O to C_i. Let $\mathbf{r}$ be a vector from O to $\Delta\tau$, and using the fact that $\mathbf{r}_i=\mathbf{r}-\mathbf{R}_i$ in (C2), and separating into Cartesian coordinates, we can integrate over the x, y, and z components of $\mathbf{r}$ to obtain $\omega_A\Delta\tau$ as the product of X, Y, and Z factors which look like

$$\omega_{A,x}=\left[\frac{(\beta_1^2+\beta_2^2)(\beta_3^2+\beta_4^2)}{\pi(\beta_1^2+\beta_2^2+\beta_3^2+\beta_4^2)}\right]^{1/2}\exp\left[-\frac{(\beta_1^2X_1+\beta_2^2X_2)^2}{\beta_1^2+\beta_2^2}-\frac{(\beta_3^2X_3+\beta_4^2X_4)^2}{\beta_3^2+\beta_4^2}\right], \qquad \text{(C4)}$$

where X_i is the x component of $\mathbf{R}_i$, and

$$\beta_i^2=\frac{3}{2\langle r_i^2\rangle}=\frac{3}{2n_ib^2}. \qquad \text{(C5)}$$

In the (isotropic) reference state, the position of X_i is randomly distributed over a Gaussian distribution of values

$$P(X_i)=\sqrt{\frac{\beta_i^2}{\pi}}\exp(-\beta_i^2X_i^2). \qquad \text{(C6)}$$

Averaging $\omega_{A,x}$ over the X coordinates of the C_i's, we obtain the mean value of the probability to form a cross link at A,

$$\langle\omega_{A,x}\rangle=\left[\frac{(\beta_1^2+\beta_2^2)(\beta_3^2+\beta_4^2)}{\pi(\beta_1^2+\beta_2^2+\beta_3^2+\beta_4^2)}\right]^{1/2}\left(\frac{1}{2}\right). \qquad \text{(C7)}$$

To a good approximation, we can replace the β_i^2 by their average values, so that

$$\overline{\omega_{A,x}}=\left(\frac{\overline{\beta^2}}{4\pi}\right)^{1/2}, \qquad \text{(C8)}$$

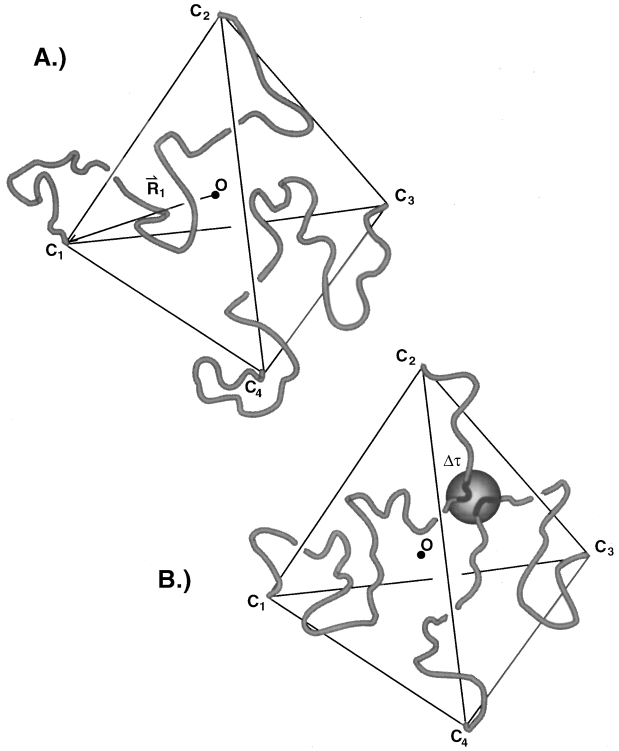

FIG. 21. Allowable configurations of polymer strands (a) before and (b) after bond formation. Each pair of bonded mers has four nearest cross-linked neighbors.

where

$$\overline{\beta^2}=\frac{3}{2\overline{n}b^2}, \tag{C9}$$

where $\overline{n}$ is the average number of statistical segments between cross links given that there are μ cross links present.

To calculate $\overline{n}$, note that the probability p_μ that a given monomer is cross linked is equal to the number of cross-linked monomers over the total number of monomers N,

$$p_\mu=\frac{2\mu}{N}. \tag{C10}$$

The probability $P_{n,\mu}$ of having a chain of length n with μ bonds present is just the negative binomial distribution

$$P_{n,\mu}=p_\mu(1-p_\mu)^{n-1}, \tag{C11}$$

so that the average chain length is

$$\overline{n}=\frac{\Sigma nP_{n,\mu}}{\Sigma P_{n,\mu}}=\frac{1}{p_\mu}=\frac{N}{2\mu}. \tag{C12}$$

Thus we obtain for the average probability to form bond A

$$\overline{\omega_A}\Delta\tau=\overline{\omega_{A,x}}\overline{\omega_{A,y}}\overline{\omega_{A,z}}=\left(\frac{\overline{\beta^2}}{4\pi}\right)^{3/2}\Delta\tau \tag{C13}$$

or

$$\overline{\omega(\mu)}\Delta\tau=\left(\frac{3\mu}{4\pi Nb^2}\right)^{3/2}\Delta\tau, \tag{C14}$$

which is the fraction of states permissible as a result of forming one more cross linkage with μ cross links already present. We can now consider adding one cross link at a time, using formula (C14) in a mean field sense, to obtain the total fraction of states permissible as a result of forming $qNz\eta$ bonds [53]:

$$\frac{\Omega_q}{\Omega_o}=\prod_{\mu=1}^{qNz\eta}\overline{\omega(\mu)}\Delta\tau=\left(\frac{\Delta\tau}{b^3}\right)^{qNz\eta}\left(\frac{3}{4\pi N}\right)^{\frac{3}{2}qNz\eta}[(qNz\eta)!]^{3/2}. \tag{C15}$$

Thus the reduction in entropy (per monomer), $-(S_o-S_q)/N$, associated with the formation of $qNz\eta$ bonds, is

$$\frac{\Delta S(q)}{N}=\frac{1}{N}\ln\frac{\Omega_q}{\Omega_o}=\frac{3}{2}qz\eta\left[\ln\frac{3}{4\pi}\left(\frac{\Delta\tau}{b^3}\right)^{2/3}-1+\ln qz\eta\right]. \tag{C16}$$

APPENDIX D

Given a polymer with μ cross links present, the probability of *not* forming one more cross link is

$$p_{not}(\mu)=1-\overline{\omega(\mu)}\Delta\tau. \tag{D1}$$

The probability of not forming $Nz\eta(1-q)$ more cross links is

$$P_{not}(qN\eta)=\prod_{\mu=qNz\eta}^{Nz\eta}p_{not}(\mu) \tag{D2}$$

and the antibond entropy reduction is then

$$\Delta S_{AB}(q|N\eta)=\ln\prod_{\mu=qNz\eta}^{Nz\eta}[1-\overline{\omega(\mu)}\Delta\tau]\cong\int_{qNz\eta}^{Nz\eta}d\mu\ln[1-(B\mu)^{3/2}], \tag{D3}$$

where

$$B=\frac{3}{4\pi N}\left(\frac{\Delta\tau}{b^3}\right)^{2/3}. \tag{D4}$$

Letting $x=B\mu$, Eq. (3.3) follows. The upper limit in the integral in (3.3) cannot be greater than 1, which sets a limit as to how large N can be for the antibond term in the entropy to be valid. Setting $Cz\eta=1$, using the lattice value of $\Delta\tau/b^3=4$, and using $\eta=1$ gives $z(N_{max})\cong(\pi/3)4^{1/3}$ or $N_{max}\cong 710$, which is much higher than the typical size of the polymers we are concerned with. N_{max} is higher for smaller values of η.

APPENDIX E

To find the probability of bond formation and its associated entropy loss for a polymer in a box, we must seek the Green's function solution to the differential equation

$$\left(\frac{\partial}{\partial N}-\frac{b^2}{6}\nabla_r^2+\frac{U_e(\mathbf{r})}{T}\right)G(\mathbf{r},\mathbf{r}',N)=\delta(\mathbf{r}-\mathbf{r}')\,\delta(N) \tag{E1}$$

with the boundary conditions $U_e=0$ inside the box and $U_e=\infty$ outside. The solution, obtained by an expansion in eigenfunctions [54], is

$$G(\mathbf{r},\mathbf{r}',N)=G_x(x,x',N)G_y(y,y',N)G_z(z,z',N) \tag{E2}$$

with

$$G_x(x,x',N)=\frac{2}{L}\sum_{k_x=1}^{\infty}\sin\left(\frac{k_x\pi x}{L}\right)\sin\left(\frac{k_x\pi x'}{L}\right)\times\exp\left(-\frac{k_x^2\pi^2Nb^2}{6L^2}\right). \tag{E3}$$

Given this propagator, there are several approaches of varying complexity one can use to find the probability of bond formation. We can start by considering the probability of forming a loop of length $\overline{n}=N/2\mu$, and then average over the position of the starting point:

$$\overline{\omega^{conf}}(\mu)\Delta\tau=\frac{1}{L^3}\int_V d\mathbf{r}\int_{\mathbf{r}'=\mathbf{r}}^{\mathbf{r}'=\mathbf{r}+\Delta\boldsymbol{\tau}}d\mathbf{r}'G(\mathbf{r},\mathbf{r}',\overline{n}). \tag{E4}$$

Splitting the integrations over Cartesian coordinates gives

$$\overline{\omega^{conf}}(\mu)\Delta\tau=\overline{\omega_x^{conf}}(\mu)\overline{\omega_y^{conf}}(\mu)\overline{\omega_z^{conf}}(\mu) \tag{E5}$$

with

$$\overline{\omega_x^{conf}}(\mu)=\frac{\Delta x}{L}\sum_{k_x=1}^{\infty} e^{\beta\overline{n}k_x^2} \tag{E6}$$

where

$$\beta=\frac{\pi^2}{6}\left(\frac{\eta}{N}\right)^{2/3}. \tag{E7}$$

So

$$\overline{\omega^{conf}}(\mu)\Delta\tau=\frac{\eta}{N}\left(\frac{\Delta\tau}{b^3}\right)\sum_{k_x,k_y,k_z} e^{\beta\overline{n}(k_x^2+k_y^2+k_z^2)}. \tag{E8}$$

We can (without necessarily assuming ground-state dominance) approximate the sums by integrals to obtain

$$\overline{\omega^{conf}}(\mu)\Delta\tau=\frac{\eta}{N}\left(\frac{\Delta\tau}{b^3}\right)\left(\int_0^{\infty} dk e^{-\beta\frac{N}{2\mu}k^2}\right)^3 \tag{E9}$$

$$=\frac{\Delta\tau}{b^3}\left(\frac{12}{\pi}\frac{\mu}{N}\right)^{3/2}, \tag{E10}$$

which preserves the $\mu^{3/2}$ dependence in the Flory theory [see Eq. (C14)], but gives a probability $\sim 10^2$ times higher to form a bond.

One can obtain comparable values to those from (E10) with a more detailed calculation which takes a flat average of (E8) over loop sizes $\overline{n'}$ from a minimum loop size N_c to $\overline{n}=N/2\mu$:

$$\omega^{conf}(\mu,N_c)\Delta\tau=\frac{\eta}{N}\left(\frac{\Delta\tau}{b^3}\right)\sum_{k_x,k_y,k_z}\frac{1}{\overline{n}-N_c+1}\sum_{\overline{n'}=N_c}^{\overline{n}} e^{-\beta\overline{n'}\mathbf{k}^2}, \tag{E11}$$

where $\mathbf{k}^2=k_x^2+k_y^2+k_z^2$. Carrying out the geometric sum on $\overline{n'}$ and approximating the sums on $\mathbf{k}$ by a volume integral in k space gives

$$\omega^{conf}(\mu,N_c)\Delta\tau\cong\frac{\eta}{N}\left(\frac{\Delta\tau}{b^3}\right)\frac{1}{\overline{n}-N_c+1}\frac{4\pi}{8}\times\int_0^{\infty}k^2dk\frac{e^{-\beta N_c k^2}}{1-e^{-\beta k^2}}. \tag{E12}$$

Changing variables to $\epsilon=\beta k^2$, the integral becomes

$$\frac{\pi}{4\beta^{3/2}}\int_0^{\infty}d\epsilon\epsilon^{1/2}\frac{e^{-N_c\epsilon}}{1-e^{-\epsilon}}=\frac{\pi}{4\beta^{3/2}}\sum_{m=1}^{\infty}\int_0^{\infty}d\epsilon\epsilon^{1/2}e^{-(N_c+m-1)\epsilon}$$

$$=\left(\frac{\pi}{4\beta}\right)^{3/2}\sum_{m=1}^{\infty}(N_c+m-1)^{-3/2}$$

$$=\left(\frac{\pi}{4\beta}\right)^{3/2}\zeta(\tfrac{3}{2},N_c), \tag{E13}$$

where $\zeta(3/2,N_c)$ is the generalized Riemann zeta function, plotted vs N_c in Fig. 22. We can see from the figure that increasing the number of segments needed to make a loop, N_c (i.e., making the chain stiffer), decreases the probability

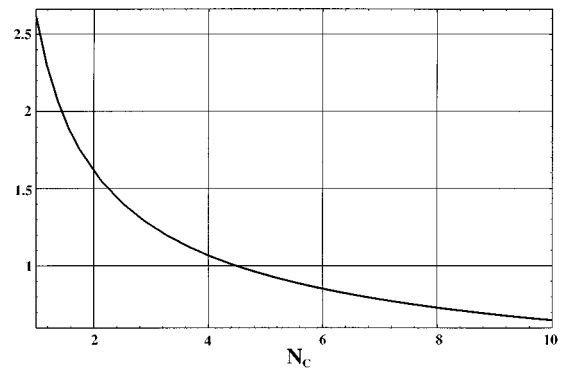

FIG. 22. Riemann zeta function $\zeta(\frac{3}{2},N_c)$ as a function of N_c.

that a bonded loop will be formed. So the probability to form a new bond given μ bonds present is now

$$\omega^{conf}(\mu,N_c)\Delta\tau\cong\left(\frac{\Delta\tau}{b^3}\right)\left(\frac{3}{2\pi}\right)^{3/2}\frac{\zeta\left(\frac{3}{2},N_c\right)}{\frac{N}{2\mu}-(N_c-1)}, \tag{E14}$$

which gives probabilities comparable to (E10) for $N_c\approx 1$. (However, the averaging over loop lengths results in a linear μ dependence, instead of the mean field $\mu^{3/2}$ dependence.) The formula for entropy loss due to bond formation for $\mu<\mu_c$ is then obtained as in Appendix C. (This was the entropy formula used in the comparison with the lattice simulations in Fig. 4.)

The critical value of μ_c divides the entropy formulas up into two regions. For $\mu<\mu_c$, confinement effects are important, and $(1/N)\Delta S_{bond}(q)$ and $(1/N)\Delta S_{\mathrm{AB}}(q)$ have the form:

$$\frac{1}{N}\Delta S_{bond}(q)=\text{ Eq. (3.7)},$$

$$\frac{1}{N}\Delta S_{\mathrm{AB}}(q)=\frac{1}{C'}\int_{C'qz\eta}^{C'q_cz\eta}dx\ln(1-x^{3/2})+\frac{1}{C}\int_{Cqz\eta}^{Cz\eta}dx\ln(1-x^{3/2}). \tag{E15}$$

For $\mu>\mu_c$, the bonds have essentially confined the polymer within its collapsed radius, and $(1/N)\Delta S_{bond}(q)$ and $(1/N)\Delta S_{\mathrm{AB}}(q)$ have the form:

$$\frac{1}{N}\Delta S_{bond}(q)=\frac{1}{N}\Delta S_{bond}^{conf}(q_c)+\frac{1}{N}\ln\prod_{\mu=Nq_cz\eta}^{qNz\eta}\overline{\omega(\mu)}\Delta\tau$$

$$=\frac{3}{2}q_cz\eta\ln\frac{C'}{C}+\frac{3}{2}qz\eta(\ln C-1+\ln qz\eta),$$

$$\frac{1}{N}\Delta S_{\mathrm{AB}}(q)=\text{ Eq. (3.3)}. \tag{E16}$$

The first term in the bond formation entropy of Eq. (E16) is a finite size effect $\sim N^{-2/3}$, and vanishes if $C'=C$, whereupon $(1/N)\,\Delta S_{bond}(q)$ becomes Eq. (3.1).

APPENDIX F

Consider the entropy of one of the melted pieces in a strongly constrained polymer. The average probability for a melted piece of m segments to propagate from position A to position B is (assuming an ideal chain Green's function)

$$\langle G(\mathbf{r}_A,\mathbf{r}_B|m)\rangle \Delta\tau' = \left(\frac{3}{2\pi m b^2}\right)^{3/2} \times \exp\left(-\frac{3\langle(\mathbf{r}_A-\mathbf{r}_B)^2\rangle}{2mb^2}\right)\Delta\tau', \tag{F1}$$

where $\Delta\tau'$ is the volume each of the end points must be localized within, which is $\cong b^3$, and where $\langle(\mathbf{r}_A-\mathbf{r}_B)^2\rangle$ is obtained by sampling $(\mathbf{r}_A-\mathbf{r}_B)^2$ for a melted piece of m segments starting at A and ending at B for all the different end-to-end distances it would have if the melted piece were ''slid'' along the length of the polymer structure, i.e., the average square end-to-end distance of the melted piece over all its possible locations along the polymer. Assuming that the frozen globule is essentially a collapsed random walk (we are not considering any secondary structure formation or other order parameters besides q in this paper), $\langle(\mathbf{r}_A-\mathbf{r}_B)^2\rangle = mb^2$. So the average number of states the melted piece has is

$$\Omega_m = \Omega_{\rm TOT}(m)\langle G(\mathbf{r}_A,\mathbf{r}_B|m)\rangle\Delta\tau' = (\mu^{(mf)})^m\left(\frac{3}{2\pi e}\right)^{3/2}\frac{1}{m^{3/2}}\left(\frac{\Delta\tau'}{b^3}\right), \tag{F2}$$

where $\Delta\tau'/b^3\approx 1$, and $\mu^{(mf)}=\nu/e$ if the walk is completely collapsed ($\eta=1$). So the average entropy of a melted piece of m segments is

$$S(m) = \frac{3}{2}\ln\left(\frac{3}{2\pi e}\right) + (\ln\mu^{(mf)})m - \frac{3}{2}\ln m \tag{F3}$$

which is plotted vs m in Fig. 23. Note that the shape of the curve vs sequence length roughly obeys a linear behavior with a cutoff sequence length m_c of ≈ 5 [Eq. (4.1)], i.e., each monomer freed after the fourth has an entropy of $\ln\mu^{(mf)}$, but at least four segments must be melted for the piece to be free enough to have any entropy.

We suspect that the ends of the polymer should follow the same behavior but with a smaller critical length. To model an end, consider the entropy of a chain confined to a half-plane [see Fig. 24(a)]. This problem can be solved by the same method as used much earlier for the adsorption of molecules onto a surface by Chandrasekhar [55]. We wish to find the number of random walks of n steps that can start anywhere, but must not touch the wall until the nth step (if it touches before it may be considered a melted piece as above, but on the surface). Let a walk start at O, and let the wall be $m_z=z/b$ steps away (if n is odd, m_z must be odd) [see Fig. 24(b)]. *Neglecting* the wall, the total number of paths from O to the wall M in n steps is just

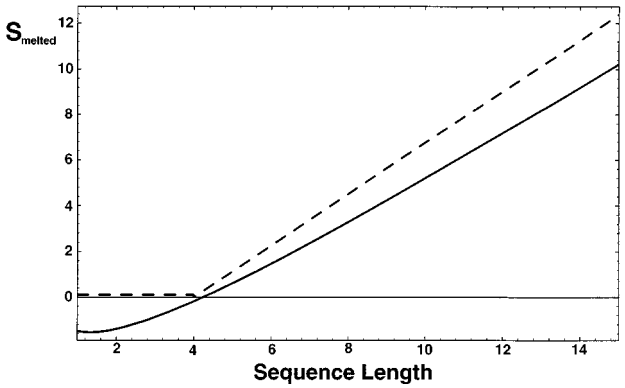

FIG. 23. Entropy of a melted strand or loop along the sequence of the polymer. Dashed line: linear approximation used in the high q analysis. Solid line: formula (F3) with $\eta=0.85$. Both curves display a cutoff sequence length for a melted loop, and are comparable for the length values of typical loops in a polymer.

$$\Omega_{free}(n,m_z) = \frac{n!}{\left(\frac{n+m_z}{2}\right)!\left(\frac{n-m_z}{2}\right)!} \tag{F4}$$

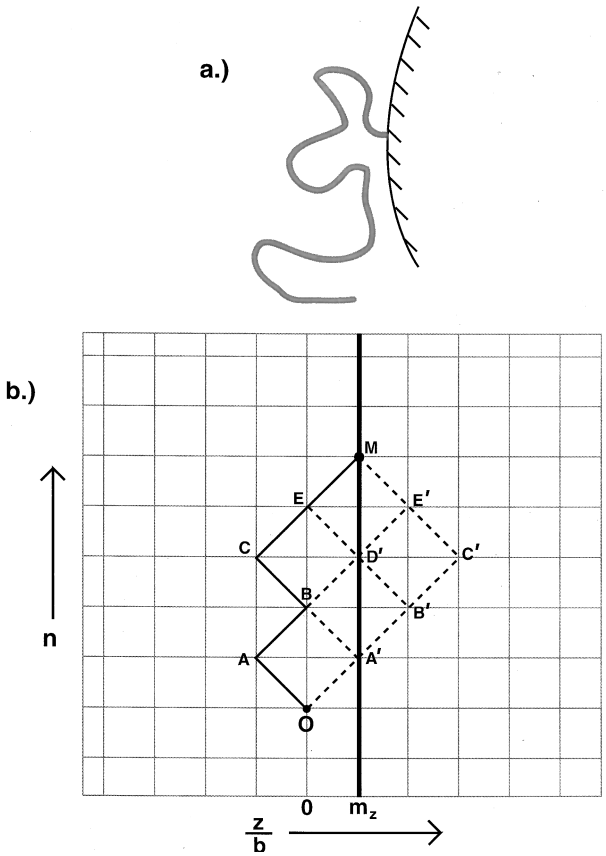

FIG. 24. (a) Polymer end confined to lie on the surface of the globule. (b) The polymer end can be modeled as a random chain confined to, and attached to the surface of, a half-plane, which can be solved by considering a random walk of a particle near an adsorbing wall [55].

We must subtract from this number all the walks that either crossed or touched before. If the walk goes to m_z in the nth step, it must have come from either m_z-1 or m_z+1 in the $(n-1)$th step, but the

$$\frac{(n-1)!}{\left(\frac{n+m_z}{2}\right)!\left(\frac{n-m_z-2}{2}\right)!} \tag{F5}$$

walks to m_z+1 cannot be counted. Any walk that went to m_z+1 must have crossed the barrier at some point (e.g., D' in walk $OABD'E'$, or A' in walk $OA'B'D'E'$). Consider the segments of these walks after they first touched or crossed the wall—each of these walks must have a *unique* reflection about the wall into a walk that went to $(n-1,m_z-1)$, but that either touched and/or crossed before (e.g., $OABD'E$ or $OA'BD'E$). These walks, *both equal in number*, are all the walks that we want to subtract from Ω_{free}. So the true number of walks from O to M without touching or crossing the wall earlier is

$$\begin{aligned}\Omega_{wall}(n,m_z)&=\Omega_{free}(n,m_z)-2\frac{(n-1)!}{\left(\frac{n+m_z}{2}\right)!\left(\frac{n-m_z}{2}-1\right)!}\\&=\frac{m_z}{n}\frac{n!}{\left(\frac{n+m_z}{2}\right)!\left(\frac{n-m_z}{2}\right)!}\\&\cong 2^n\frac{m_z}{n}\sqrt{\frac{2}{\pi n}}e^{-m_z^2/2n}\end{aligned} \tag{F6}$$

using Stirling's approximation. So, letting $z=m_z b$, the probability a walk goes within dz of a wall at position z in n steps without hitting the wall first, or equivalently, the probability that the free end of a one-dimensional polymer of length n fastened to a wall at O is in the interval $(z,z+dz)$ is

$$W(z,n)dz=\left(\frac{dz}{2b}\right)\frac{z}{nb}\sqrt{\frac{2}{\pi n}}\exp\left(-\frac{z^2}{2nb^2}\right). \tag{F7}$$

The factor $dz/2b$ is the number of states in the interval dz, since $2b$ is the distance between allowable states for a given n (if n is even, m_z must be even).

The fraction of free Gaussian states that remain to the outside of the wall is this probability integrated over all z:

$$f_{end}=\int_0^\infty W(z,n)dz=\frac{1}{(2\pi n)^{1/2}}. \tag{F8}$$

This is the reduction in the total number of states due to the fact that the polymer chain must be outside the globule. Using the fact that 1/3 of the steps of a three-dimensional walk would be in the z direction, the total number of states of a polymer end of length ℓ is

$$\Omega_{end}(\ell)=\Omega_{free}f_{end}=\frac{1}{\left(2\pi\frac{\ell}{3}\right)^{1/2}}(\mu^{(mf)})^n \tag{F9}$$

with entropy

$$S(\ell)=\frac{1}{2}\ln\left(\frac{3}{2\pi}\right)+(\ln\mu^{(mf)})\ell-\frac{1}{2}\ln\ell, \tag{F10}$$

which has the same form as Eq. (F3) (but with the connectivity constant probably only about 1 or 2 for small ℓ), so we are led to use the same linear entropy formula as in (4.1), but with a different critical length [formula (4.2)].

APPENDIX G

We can calculate the total number of states in terms of the distributions $\{n_\ell\},\{m_\ell\}$, and $\{p_\ell\}$ from the fundamental formula

$$S=-\sum_\alpha p_\alpha\ln p_\alpha$$

by considering the ensembles of melted internal pieces and ends in the polymer (Fig. 5). For example, let α be the total state of both ends i,j. Since the states of each of the ends are independent,

$$S\Big|_{\ell_E^{tot}}=-{\sum_{ij}}' p_i p_j\ln p_i p_j \tag{G1}$$

is the entropy at fixed total end length, where Σ' is constrained so that $\ell_i+\ell_j=\ell_E^{tot}$. Here

$$p_i=p_{\ell_i}\left(\frac{1}{\gamma_{\ell_i}}\right), \tag{G2}$$

where p_{ℓ_i} is the probability for an end to have length ℓ_i, and γ_{ℓ_i} is the number of configurational states for a chain of length ℓ_i:

$$\gamma_{\ell_i}=(\mu^{(mf)})^{\ell-(\ell_{\rm EC}-1)}. \tag{G3}$$

Using $\Sigma'_{ij}=\Sigma'_{\ell_i,\ell_j}\gamma_{\ell_i}\gamma_{\ell_j}$, and allowing all possible total end lengths, the entropy of the ends is

$$\begin{aligned}S_E&=-\sum_{\ell_E^{tot}}\left({\sum_{\ell_i,\ell_j}}'\gamma_{\ell_i}\gamma_{\ell_j}\right)\left(\frac{p_{\ell_i}}{\gamma_{\ell_i}}\frac{p_{\ell_j}}{\gamma_{\ell_j}}\right)\left(\ln\frac{p_{\ell_i}}{\gamma_{\ell_i}}+\ln\frac{p_{\ell_j}}{\gamma_{\ell_j}}\right)\\&=-\sum_{\substack{\ell_i,\ell_j\\ \text{unconstrained}}}p_{\ell_i}p_{\ell_j}\left(\ln\frac{p_{\ell_i}}{\gamma_{\ell_i}}+\ln\frac{p_{\ell_j}}{\gamma_{\ell_j}}\right)\\&=-2\sum_\ell p_\ell\ln\frac{p_\ell}{\gamma_\ell}.\end{aligned} \tag{G4}$$

So the total number of states has a mixing component, and a ν^N component:

$$\Omega_E=e^{S_E}=\prod_{\ell=\ell_{\rm EC}}^{N}\left(\frac{1}{p_\ell}\right)^{2p_\ell}[(\mu^{(mf)})^{\ell-(\ell_{\rm EC}-1)}]^{2p_\ell}. \tag{G5}$$

A similar derivation for the internal melted pieces is equivalent to simply replacing

$$p_\ell \to \frac{n_\ell}{N_{\rm TOT}}, \quad \ell_{\rm EC} \to \ell_c, \quad 2 \to N_{\rm TOT}, \tag{G6}$$

where n_ℓ is the number of melted pieces of length ℓ, $N_{\rm TOT}=Nf$ is the total number of melted pieces, and ℓ_c is the critical length for internal melted pieces to have entropy. So making the replacements gives

$$\Omega_{\rm melt} = \prod_{\ell=\ell_c}^{N} \left(\frac{N_{\rm TOT}}{n_\ell}\right)^{n_\ell} [(\mu^{(mf)})^{\ell-(\ell_c-1)}]^{n_\ell} \cong \frac{N_{\rm TOT}!}{\Pi_{\ell_c}^{N} n_\ell!} \prod_{\ell_c}^{N} [(\mu^{(mf)})^{\ell-(\ell_c-1)}]^{n_\ell}. \tag{G7}$$

For the internal frozen pieces $\mu^{(mf)}=1$, $N_{\rm TOT}({\rm frozen})=N_{\rm TOT}({\rm free})$ thermodynamically, and $\ell_c=1$, so

$$\Omega_{\rm frozen} = \frac{N_{\rm TOT}!}{\Pi_1^N m_\ell!}, \tag{G8}$$

where m_ℓ is the number of frozen pieces or ''trains'' of length ℓ. So the total number of states (4.3) is just the product of all these factors.

Note that in our analysis we have treated the entropies of the melted strands as independent units, and they will remain energetically independent in calculating the free energy from this entropy (and landscape roughness) using the GREM analysis. More realistic models would include some interaction between melted pieces depending on their proximity.

APPENDIX H

The log of the maximum term which dominates the sum in Eq. (4.3) is

$$\ln W = 2Nf\ln Nf + \ln\mu \sum_{\ell_c}^{N} [\ell-(\ell_c-1)]n_\ell + 2\ln\mu \sum_{\ell_{\rm EC}}^{N} [\ell-(\ell_{\rm EC}-1)]p_\ell - \sum_{\ell_c}^{N} n_\ell \ln n_\ell - \sum_{1}^{N} m_\ell \ln m_\ell - 2\sum_{\ell_{\rm EC}}^{N} p_\ell \ln p_\ell, \tag{H1}$$

which when maximized subject to the constraints

$$\alpha_f : \sum_{\ell_c}^{N} n_\ell = Nf, \beta_f : \sum_{\ell_c}^{N} \ell n_\ell = N(1-q) - 2\ell_E, \quad \alpha_c : \sum_{1}^{N} m_\ell = Nf, \beta_c : \sum_{1}^{N} \ell m_\ell = Nq, \quad \alpha_E : \sum_{\ell_{\rm EC}}^{N} p_\ell = 1, \beta_E : \sum_{\ell_{\rm EC}}^{N} \ell p_\ell = \ell_E \tag{H2}$$

gives the thermodynamic entropy. The constraints (H2) are straightforward to derive, for example,

$$\sum_{1}^{N} \ell m_\ell = Nq$$

comes from the fact that if there are N_F frozen monomers, there are $z\eta N_F$ frozen bonds. Equating this with $qNz\eta$ bonds gives $N_F=Nq$. The other constraints follow from this type of reasoning. Introducing the constraints into $\delta \ln W=0$ with the Lagrange multipliers listed in (H2) gives the usual exponential dependence for the most probable distributions:

$$n_\ell = C_f e^{-\beta_f \ell}, \quad m_\ell = C_c e^{-\beta_c \ell}, \quad p_\ell = C_E e^{-\beta_E \ell}, \tag{H3}$$

where $C_x = e^{-\alpha_x - 1}$. Substituting the distributions (H3) back into the constraints (H2) gives the negative binomial distributions in Eq. (4.4).

[1] J. Bryngelson, J.O. Onuchic, N.D. Socci, and P.G. Wolynes, PROTEINS **21**, 167 (1995).
[2] P.G. Wolynes, J.N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[3] J.N. Onuchic, P.G. Wolynes, Z. Luthey-Schulten, and N.D. Socci, Proc. Natl. Acad. Sci. USA, **92**, 3626 (1995).
[4] H. Frauenfelder and P.G. Wolynes, Phys. Today **47**, (2) 58 (1994).
[5] K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas, and H.S. Chan, Protein Sci. **4**, 561 (1995).
[6] P.E. Leopold, M. Montal, and J.N. Onuchic, Proc. Natl. Acad. Sci. USA **89**, 8721 (1992).
[7] N. Go, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
[8] J. Bryngelson and P.G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).
[9] R.A. Goldstein, Z.A. Luthey-Schulten, and P.G. Wolynes, Proc. Natl. Acad. Sci. USA **89**, 4918 (1992).
[10] E.I. Shakhnovich, Phys. Rev. Lett. **72**, 3907 (1994).
[11] B. Derrida, Phys. Rev. B **24**, 2613 (1981).
[12] D.J. Gross and M. Mezard, Nucl. Phys. B **240**, 431 (1984).
[13] D.J. Gross, I. Kantor, and H. Sompolinsky, Phys. Rev. Lett. **55**, 304 (1985); T.R. Kirkpatrick and P.G. Wolynes, Phys. Rev. B **36**, 8552 (1987).
[14] T. Garel and H. Orland, Europhys. Lett. **6**, 307 (1988); E.I. Shakhnovich and A.M. Gutin, *ibid.* **8**, 327 (1989); Biophys. Chem. **34**, 187 (1989).
[15] Z. Luthey-Schulten, B.E. Ramirez, and P.G. Wolynes, J. Phys. Chem. **99**, 2177 (1995).
[16] J.G. Saven and P.G. Wolynes, J. Mol. Biol. **257**, 199 (1996).
[17] B. Derrida and E. Gardner, J. Phys C **19**, 2253 (1986).
[18] P.J. Flory, J. Am. Chem. Soc. **78**, 5222 (1956).
[19] S.F. Edwards, in *Proceedings of the Third International Conference on Amorphous Materials*, edited by R.W. Douglass and B. Ellis (Wiley, New York, 1972), p. 279.
[20] P. Goldbart and N. Goldenfeld, Phys. Rev. Lett. **58**, 2676 (1987).

[21] C. Levinthal, in *Mossbauer Spectroscopy in Biological Systems* (University of Illinois Press, Urbana, IL, 1969), p. 22.
[22] J. Bryngelson and P.G. Wolynes, Biopolymers **30**, 177 (1990).
[23] J.F. Douglas and T. Ishinabe, Phys. Rev. E **51**, 1791 (1995).
[24] A.M. Nemirovsky, J. Dudowicz, and K.F. Freed, J. Stat. Phys. **67**, 395 (1992).
[25] The packing fraction for real proteins (without cavities) on an atomistic basis is about that of an fcc crystal (0.74), while the packing fraction for molten globules is somewhat less [F.M. Richards and W.A. Lim, Qt. Rev. Biophys. **26**, 423 (1993)]. However, the packing fraction η used here should not be understood as the atomistic packing fraction: each monomer + side chain unit is crudely approximated as a lattice site, e.g., $\eta=1$ corresponds to the highest density of cubes on a cubic lattice.
[26] B. Derrida, J. Phys. Lett. (Paris), **46**, 401 (1993).
[27] J.G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[28] We have not assumed any dependence of the collapse parameter η on the overlap q with the given state, as would be the case if we had picked a particularly high or low energy state. This further complication will be dealt with in a later paper.
[29] A.M. Gutin and E.I. Shakhnovich, J. Chem. Phys. **100**, 5290 (1994).
[30] Note confinement effects as determined by the scaling law (3.5) are not important in two-dimensional systems.
[31] J. Bryngelson and D. Thirumalai, Phys. Rev. Lett. **76**, 542 (1995).
[32] N. Socci and J. Onuchic (private communication).
[33] It is of some importance to note that q_v obtained without the mixing term in the entropy is not significantly lower, so that the range in q where the mixing entropy is used invalidly is not too large.
[34] P.J. Flory and R.R. Matheson, J. Phys. Chem. **88**, 6606 (1984).
[35] We expect the antibond effects to be small for high q overlaps; indeed an excluded bond analysis that weights $\mu^{\ell-(\ell_c-1)}$ by a term $(1-1/8\ell^{1/2})$ to account for the antibond effect gives almost the same result.
[36] A. Silberberg, J. Phys. Chem. **66**, 1872 (1962).
[37] C.A.J. Hoeve, E.A. DiMarzio, and P. Peyser, J. Chem. Phys. **42**, 2558 (1965).
[38] E.I. Shakhnovich and A.M. Gutin, J. Phys. A **22**, 1647 (1989).
[39] Equation (5.5) is also derivable, for the self-averaging region above T_{rem}, from $S(T)=-\Sigma_\alpha p_\alpha \ln p_\alpha$ using $\Sigma_\alpha=\int dE n(E)$, with $n(E)\cong\langle n(E)\rangle$ as before, and $p_\alpha=(e^{-E_\alpha/T})/Z$ with $\ln Z\cong\ln\langle Z\rangle=Ns_o+\Delta E^2/2T^2$.
[40] The jump may be from some nonzero value q_{min} to 1 because of finite-size effects (discussed below), and a maximum value of overlap $q_{max}<1$ can be considered because of collective melting described in Sec. IV. The main important feature is two discretely different values of q.
[41] The average value of the overlap as defined by
$$\langle q(T)\rangle=\frac{\int_{q_{min}}^{q_{max}} dq q P(q,T)}{\int_{q_{min}}^{q_{max}} dq P(q,T)}$$
will continuously rise (linearly for the REM, almost linearly with slightly smaller values for the GREM) from q_{min} at $T=T_g$ to q_{max} at $T=0$.
[42] B. Derrida and E. Gardner, J. Phys. C **19**, 5783 (1986).
[43] One may wish to go further and investigate systems of assembling monomers at constant chemical potential μ by considering a Legendre transformation by μN of the free energy in terms of N [implicit in $q_{min}(N)$] to obtain an analogous de Almeida–Thouless line in the (μ,T) plane.
[44] J.M. Flanagan, M. Kataoka, T. Fujisawa, and D.M. Engelman, Biochemistry **32**, 10 359 (1993).
[45] T.H. LaBean, S.A. Kauffman, and T.R. Butt (unpublished).
[46] A.R. Davidson and R.T. Sauer, Proc. Natl. Acad. Sci. USA **91**, 2146 (1994).
[47] Y. Bal, T.R. Sosnick, L. Mayne, and S.W. Englander, Science **269**, 192 (1995).
[48] J. Wang, S.S. Plotkin, and P.G. Wolynes (unpublished).
[49] K.A. Dill and D. Stigter, Adv. Protein Chem. **46**, 59 (1995).
[50] The ansatz of ultrametricity in the GREM and the resulting restriction on the allowable overlaps between states, for example, two states having overlap q with a third must have an overlap $\geqslant q$ with each other, is certainly an approximation here, and one can conceive of triplets of polymeric states which do not satisfy this ultrametric requirement.
[51] For the collapsed 27-mer there are 28 levels, for the 64-mer there are 82 levels. However, in analyzing small sequence fragments, we would wish to keep the possible overlaps discreet.
[52] We will use the terms monomers and statistical segments interchangeably here. A polymer of length N is really understood to have N statistical segments.
[53] The limits of the product Π in this formula are modified when effects due to confinement of the polymer within its collapsed globule size are taken into account. This slightly modifies the bond-formation entropy formula (C16); see formulas (E15) and (E16).
[54] S.F. Edwards and K.F. Freed, J. Phys. A **2**, 145 (1969).
[55] S. Chandrasakhar, Rev. Mod. Phys. **15**, 1 (1943).
[56] H.S. Chan and K.A. Dill, J. Chem. Phys. **99**, 2116 (1993).