

Solvation versus freezing in a heteropolymer globule

Phillip L. Geissler*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Eugene I. Shakhnovich

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA

Alexander Yu. Grosberg

Department of Physics, University of Minnesota, Minneapolis, Minnesota 55455, USA

(Received 5 May 2003; revised manuscript received 9 April 2004; published 24 August 2004)

We address the response of a random heteropolymer to preferential solvation of certain monomer types at the globule-solvent interface. For each set of monomers that can comprise the molecule's surface, we represent the ensemble of allowed configurations by a Gaussian distribution of energy levels, whose mean and variance depend on the set's composition. Within such a random energy model, mean surface composition is proportional to solvation strength under most conditions. The breadth of this linear response regime arises from the approximate statistical independence of surface and volume energies. Fluctuations play a crucial role in determining the excess of solvophilic monomers at the surface, and for a diverse set of monomer types can be overcome only by very strong solvent preference.

DOI: 10.1103/PhysRevE.70.021802

PACS number(s): 82.35.Lr, 82.35.Jk, 87.15.Aa, 87.15.Cc

I. INTRODUCTION: CAN WE TREAT SURFACE AND VOLUME FREE ENERGIES AS STATISTICALLY INDEPENDENT FOR RANDOM HETEROPOLYMERS?

A polymer chain collapses into a compact globular state in poor solvents. A chain with a quenched sequence of chemically different units can further undergo a freezing transition, in which the freedom of chain shape fluctuations is sacrificed for the choice of optimal spatial contacts between monomers. This freezing, in some ways akin to folding of a protein, is subject to constraints imposed by the chain connectivity, quenched sequence, and excluded volume. The effects of frustration due to these constraints are well understood [1,2]. However well developed, current theories of heteropolymer freezing ignore one obvious fact, namely, that some chain segments are more favorably solvated than others. By contrast, much of the protein literature presumes preferential solvation to be a leading determinant of tertiary structure. It is commonly held that a protein's surface is composed of hydrophilic units, while hydrophobic units are invariably buried in the core.

For heteropolymers in general, it is clear that energy gained through preferential exposure of solvophilic units comes at a cost. Constraining particular units to the globule surface restricts the selection of contacting monomer pairs inside the globule, exacerbating frustration. In other words, when the sequence of units has not been designed in an intelligent way, as is the case for the random sequence heteropolymer, preferential exposure may significantly reduce

the availability of low energy conformations. The question thus arises: what is the effect of solvation on heteropolymer freezing, or, more specifically, how large an excess of solvophilic units at the globule surface is consistent with freezing?

This question was first discussed by two of us in [7]. Using a replica approach, it was found that a solvent preference of strength Γ for particular monomers at the surface lowers the ground state energy E_g by an amount $\sim K\Gamma^2/T_{fr}$. Here, $K \sim N^{2/3}$ is the number of monomers exposed to the solvent, N is the number of monomers comprising the molecule, and T_{fr} is the freezing temperature below which the ground state dominates. For strong solvation this approach apparently fails. In particular, $K\Gamma^2/T_{fr}$ can exceed the maximum possible solvation energy (without distortion of the globular shape) $K\Gamma$ corresponding to a completely solvophilic surface.

This article describes a more comprehensive treatment of solvation. Our approach is based on the peculiar generalization of the random energy model (REM) of Derrida [3]. Originally suggested as the simplest model of freezing in spin glasses, the REM captures the essential features of heteropolymer freezing in three dimensions [4,5]. Although it was emphasized in [6] that the REM is neither exact nor universally applicable for heteropolymers, it is well known [2] to be a very useful first approximation. The conventional application of the REM to heteropolymers considers only mean field effects, and does so only in the volume approximation. In this approximation, the total internal energy of a conformation is just the sum of $\sim N$ contact energies, and it is approximately Gaussian distributed, $w(E) \propto \exp[-(E - \bar{E})^2/2N\Delta^2]$. In the volume approximation, this distribution is determined solely by the mean \bar{E} and variance δB^2 of contact energies, so that $\bar{E} = N\bar{B}$ and $\Delta^2 = \delta B^2$. With these

*Present address: Department of Chemistry, University of California at Berkeley, Berkeley, CA 94720, USA.

parameters the energy spectrum of a typical sequence is constructed by drawing $M=e^{sN}$ values from $w(E)$, where M is the total number of compact conformations, and s is the corresponding conformational entropy per monomer.

By contrast, we wish to examine a finite molecule in a way that explicitly incorporates effects of surface-related fluctuations. The simplest way to incorporate the surface into this picture is to imagine that contacts between surface monomers and the solvent are also, in effect, statistically independent random variables. The variance of surface energy, $K\Gamma^2$, then adds to that of the volume energy. We use the saddle point of the partition function $Z=e^{sN}\int dEw(E)e^{-E/T}$, to estimate the energy of representative conformations, $E=\bar{E}-N\delta B^2/T-K\Gamma^2/T$. The lower bound of the spectrum is reached when $e^{sN}w(E)\approx 1$, yielding a typical ground state energy

$$E_g^{(\text{typ})}\approx\bar{E}-\sqrt{2s}N\Delta\approx\bar{E}-\sqrt{2s}N\delta B-\sqrt{\frac{s}{2}}\frac{\Gamma^2}{\delta B}. \quad (1)$$

Correspondingly, $T_{\text{fr}}\approx\delta B/\sqrt{2s}$. The final term in Eq. (1), i.e., the change in ground state energy due to solvation, is (within a factor of order unity) $-K\Gamma^2/T_{\text{fr}}$, just as found in the replica approach of Ref. [7]. Thus, the result of Ref. [7] is tantamount to the assumption of statistical independence of the surface and the volume in a heteropolymer globule.

II. MODEL AND METHOD: CONVOLUTION OF MANY REMS

There are several reasons to be skeptical of the suggested independence of surface and volume energies. First, removing a specific set of monomers from the globule interior to the surface modifies the distribution of contacting monomers. Second, there are a finite number of solvophilic monomers in a given molecule (possibly fewer than K). When a large fraction is placed on the surface, the supply of solvophilic monomers is strongly depleted, and solvation energy saturates. Finally, certain choices of surface monomers constrain configuration space more strongly than others.

We examine these effects using a model in which each monomer is labeled by a quenched variable σ . When a monomer with label σ resides on the surface, it is assigned a solvation energy $-\Gamma\sigma$. Solvophilic species are thus characterized by $\sigma>0$, while $\sigma<0$ for solvophobic species. In its total effect the solvent preference Γ can be viewed as an external field that couples linearly to a net surface composition $C_{\text{surf}}\equiv\sum_{i\in\text{surf}}\sigma_i$. Within the globule, a contacting pair of monomers, of type σ and σ' , is ascribed energy $B_{\sigma\sigma'}=\bar{B}+\delta B\sigma\sigma'$. For simplicity we restrict attention to distributions of monomer types, $p(\sigma)$, with zero mean and unit variance:

$$\int d\sigma p(\sigma)\sigma=0, \quad \int d\sigma p(\sigma)\sigma^2=1. \quad (2)$$

Imagine that a certain set of monomers is constrained to stay on the surface. We denote this set as G . In our model energetic consequences of such a constraint depend only on

the distribution $f(\sigma)$ of monomer types in G . For example, the effective distribution of contacting monomers (i.e., those remaining inside the globule when G is removed), $p_{\text{eff}}(\sigma)$, may be written as $p_{\text{eff}}(\sigma)=p(\sigma)+(K/N)[p(\sigma)-f(\sigma)]$. The effective mean and variance of the contact energies are then $\bar{B}_{\text{eff}}=\bar{B}+(K/N)\alpha_G$ and $\delta B_{\text{eff}}=\delta B+(K/N)\beta_G$, respectively. For distributions satisfying Eq. (2), $\alpha_G=0$ and $\beta_G=2\delta B^2[1-\int d\sigma\sigma^2f(\sigma)]$. Similarly, the solvation energy per surface monomer is $\gamma_G=-\Gamma\int d\sigma\sigma f(\sigma)$.

We express the number of accessible conformations when all monomers in G are confined to the surface as $M_G\equiv e^{sN-K\omega_G}$. Here, ω_G is the entropy loss per surface monomer for particular G . Although smaller than M , M_G is still exponentially large in N . In general ω_G is not simply a functional of $f(\sigma)$, but is instead a complicated function of G . We will assume that for any specific $f(\sigma)$ the average of ω_G over all consistent realizations of G is a constant independent of $f(\sigma)$. In order to recover the appropriate total number of conformations after summing over G , we choose this constant to be $\bar{\omega}=K^{-1}\ln\binom{N}{K}\approx\ln(Ne/K)$.

We consider a separate REM for each possible choice of surface G . In doing so, we assume that allowed conformations in the corresponding subensembles are sufficiently diverse that their energies are Gaussian distributed, with

$$w_G(E)\propto\exp\left[-\frac{[E-N\bar{B}-K(\alpha_G+\gamma_G)]^2}{2N\delta B^2+2K\beta_G}\right]. \quad (3)$$

Ultimately, we must reconstruct the full ensemble of compact chain fluctuations by superposing all possible subensembles, i.e., by summing over G . This convolution of REMs, each representing a distinct choice of G , constitutes our caricature of a random heteropolymer with a solvated surface. It allows a straightforward treatment of the energy fluctuations accompanying different states of the surface, a feature we will show to be indispensable.

The set of conformations inconsistent with each choice of surface-exposed monomers G will have a corresponding minimum energy $E_g(G)$. The lowest of these minima among all subensembles then yields the ground state energy $E_g=\min_G E_g(G)$. Characterizing the surface of the ground state thus amounts to identifying which set G produces the lowest value of $E_g(G)$. This task is more subtle than might appear at first glance, since the optimal choice of G may vary from one realization of the REM to the next. Interfacial energy clearly favors a solvophilic surface, but does it always yield the ground state?

It is straightforward to determine which set G is more likely than any other to be manifest in the ground state. We use the condition $M_G w_G[E_g^{(\text{typ})}(G)]\approx 1$ to estimate the typical lower bound of a subensemble energy spectrum, giving

$$E_g^{(\text{typ})}(G)\approx N\bar{B}-N\sqrt{2s}\delta B+K\epsilon_{\text{surf}}(G), \quad (4)$$

$$\epsilon_{\text{surf}}(G)=\alpha_G+\gamma_G+\frac{\delta B}{\sqrt{2s}}\bar{\omega}-\sqrt{\frac{s}{2}}\frac{\beta_G}{\delta B}. \quad (5)$$

The choice of surface G_{sol} minimizing $\epsilon_{\text{surf}}(G)$ in Eq. (5) is maximally solvophilic, with $|\epsilon_{\text{surf}}(G_{\text{sol}})|\sim O(K)$. There is no

single choice of G with a lower *typical* minimum energy.

For large N , however, the number of available choices of G (roughly $e^{\bar{\omega}K}$) is enormous, essentially guaranteeing that $E_g(G)$ is *atypically* low (or high) for some subensembles. We must therefore consider the distributions of $E_g(G)$ to assess whether variations in subensemble ground state energies can compete with the optimal solvation energy associated with G_{sol} . According to the statistics of extreme values [8], the probability that the lowest energy in a particular subensemble deviates from $E_g^{(\text{typ})}(G)$ by an amount $\delta E_g(G) = E_g(G) - E_g^{(\text{typ})}(G)$ is

$$\mathcal{W}[\delta E_g(G)] = \exp\left[\frac{\delta E_g(G)}{T_{\text{fr}}} - \exp\left(\frac{\delta E_g(G)}{T_{\text{fr}}}\right)\right]. \quad (6)$$

Compared to the Gaussian distribution of energies within a subensemble, $\mathcal{W}[\delta E_g(G)]$ decays very slowly for $\delta E_g(G) < 0$. This breadth is unimportant for $E_g(G_{\text{sol}})$, since this choice of surface is unique (or at least can be accomplished in a small number number of ways compared to $e^{\bar{\omega}K}$). The number of subensembles with unremarkable surface energy ($|\epsilon_{\text{surf}}(G)| \ll K$), on the other hand, is vast. Within this group we expect significant variations in $E_g(G)$.

In Eq. (6) $\mathcal{W}(\mathcal{E})$ expresses the probability of observing a ground state fluctuation of size \mathcal{E} for a single subensemble. For $|\epsilon_{\text{surf}}(G)| \ll K$ the expected number of subensembles with $\delta E_g(G)$ between \mathcal{E} and $\mathcal{E} + d\mathcal{E}$ is thus approximately $e^{\bar{\omega}K} \mathcal{W}(\mathcal{E})$. Since the tail of $\mathcal{W}[\delta E_g(G)]$ is exponential, we expect $O(1)$ of the subensembles with insignificant surface energy to have $|\delta E_g(G)| = O(K)$. In other words, the variations in volume energy among these subensembles are comparable in magnitude to the largest possible surface energy. This result may be viewed as the consequence of an effective entropy that remains important even at low temperature. The collection of subensembles with appreciable surface energy is much smaller than $e^{\bar{\omega}K}$, and its entropy is correspondingly low. Interestingly, as a consequence of the freezing transition at $T = T_{\text{fr}}$, this entropy remains important even as $T \rightarrow 0$. The ground state surface is uniformly solvophilic only when solvent preference is strong enough to offset the effective entropic cost.

Because $w_G(E)$ depends only on $f(\sigma)$, it is natural to group all subensembles with the same number density of monomer types. We have shown that accounting for the disparity in sizes of these groups is essential. The number of ways to choose K monomers with distribution $f(\sigma)$ from a pool of N monomers with distribution $p(\sigma)$ is $e^{Ns\{f\}}$, where

$$s\{f\} = - \int d\sigma p(\sigma) [\phi \ln \phi + (1 - \phi) \ln(1 - \phi)]. \quad (7)$$

The density $\phi(\sigma) \equiv Kf(\sigma)/Np(\sigma)$ and its corresponding entropy $s\{f\}$ are precisely those relevant for Langmuir adsorption of an ideal gas mixture onto K distinguishable sites.

At and above the freezing temperature, equilibrium of a subensemble group is dominated by the saddle point of the partition function

$$Z\{f\} = e^{Ns - K\bar{\omega} + Ns\{f\}} \int dE w_G(E) e^{-E/T}. \quad (8)$$

The group free energy $F\{f\} = -T \ln Z\{f\}$, is then

$$F\{f\} \approx \bar{F} + KT \left[-\frac{\delta B^2}{T^2} + \bar{\omega} \right] + NT \int d\sigma p(\sigma) \times \left\{ \phi \left[\eta(\sigma) - \ln\left(\frac{1 - \phi}{\phi}\right) \right] + \ln(1 - \phi) \right\}, \quad (9)$$

where

$$\eta(\sigma) = \frac{\delta B^2}{T^2} \sigma^2 - \frac{\Gamma}{T} \sigma. \quad (10)$$

Volume terms independent of $f(\sigma)$ have been collected as $\bar{F}/N = \bar{B} - T_s - \delta B^2/2T$. According to Eqs. (9) and (10), the binding energy in our analogy to Langmuir adsorption varies with particle type σ as $(\delta B^2/T)\sigma^2 - \Gamma\sigma$.

The full partition function of the polymer, a sum over all $Z\{f\}$, is dominated by the subensemble group with lowest free energy:

$$Z = \sum_{f(\sigma)} Z\{f\} \approx Z\{f^*\}. \quad (11)$$

We calculate the optimal surface distribution $f^*(\sigma)$ variationally, using a Lagrange multiplier to enforce proper normalization of $\phi(\sigma)$. We thereby obtain

$$f^*(\sigma) = \frac{(N/K)p(\sigma)}{1 + \Lambda e^{\eta(\sigma)}}, \quad (12)$$

where the constant Λ is determined by the normalization

$$\int d\sigma \frac{(NK)p(\sigma)}{1 + \Lambda e^{\eta(\sigma)}} = 1. \quad (13)$$

Finally, evaluating $F\{f\}$ at $f^*(\sigma)$ yields our approximation for the total free energy $F = \bar{F} + F_{\text{surf}}$, with

$$\frac{F_{\text{surf}}}{K} \approx T \left[\frac{\delta B^2}{T^2} - 1 + \ln\left(\Lambda \int d\sigma \frac{p(\sigma)}{1 + \Lambda e^{\eta(\sigma)}}\right) - \frac{\int d\sigma p(\sigma) \ln[\Lambda e^{\eta(\sigma)}/(1 + \Lambda e^{\eta(\sigma)})]}{\int d\sigma p(\sigma)(1 + \Lambda e^{\eta(\sigma)})} \right]. \quad (14)$$

Equations (12)–(14), appropriate for $T \geq T_{\text{fr}}$, are our principal results. They express the equilibrium distribution of monomer types on the polymer surface and the corresponding interfacial free energy density in terms of model parameters and an effective fugacity for surface monomers, Λ .

III. ANALYSIS OF THE RESULTS

In order to make these results concrete we consider some limiting cases and specific forms of $p(\sigma)$. First, let us assume

that preferential solvation does not lead to a significant depletion of any monomer type inside the globule, so that $Kf(\sigma) \ll Np(\sigma)$ for every σ . Then, Eq. (12) requires that $\Lambda e^{\eta(\sigma)} \gg 1$, simplifying the above expressions to yield $f(\sigma) \propto p(\sigma) \exp[-\eta(\sigma)]$ and

$$\frac{F_{\text{surf}}}{K} \simeq -T \ln \left[\int d\sigma p(\sigma) \exp\left(\frac{\delta B^2}{T^2} - \eta(\sigma)\right) \right]. \quad (15)$$

To simplify this result even further, let us consider a binary distribution $p(\sigma) = (1/2)[\delta(\sigma+1) + \delta(\sigma-1)]$, which corresponds to the minimum of chemical diversity. Such a model has prominence in the literature as the simplest model of proteins [9–11] and designed heteropolymers in general [12]. It has been pointed out though that this model does not possess sufficient chemical diversity to ensure proteinlike properties such as the existence of a unique structure [13,14]. Nevertheless, its consideration is instructive as the simplest heteropolymer model where surface effects in finite molecules may play a role. In this case depletion is invariably weak, since taking K monomers away to the surface cannot exhaust the total stock $N/2$ of either monomer type. Equation (15) then trivially yields

$$\frac{F_{\text{surf}}}{K} = -T \ln \left[\cosh\left(\frac{\Gamma}{T}\right) \right]. \quad (16)$$

For $\Gamma/T \ll 1$, $F_{\text{surf}} \simeq -K\Gamma^2/2T$, precisely as obtained by assuming statistical independence of surface and volume. Since net surface composition is conjugate to solvation strength, its equilibrium value may be computed by differentiating Eq. (16) with respect to Γ , yielding

$$\langle C_{\text{surf}} \rangle_{\Gamma} = K \tanh\left(\frac{\Gamma}{T}\right) \simeq \frac{K\Gamma}{T}. \quad (17)$$

In this limit the net surface composition is proportional to the ‘‘field’’ Γ . The above results may therefore be understood in simple terms as a manifestation of linear response. From Eq. (17) we identify a susceptibility $\chi \simeq K/T$ corresponding to surface fluctuations of size $\langle C_{\text{surf}}^2 \rangle_{\Gamma=0} = K$ in the absence of solvation. In other words, the excess of solvophilic monomers at the surface is governed by K effectively independent random variables. This simple behavior results directly from the prevalence of variations in volume energy over surface interactions. But when $\Gamma \gtrsim T$ solvation wins out. Linear response then breaks down due to saturation, as F_{surf}/K and C_{surf} approach their limiting values of $-\Gamma$ and K .

The properties of the ground state are obtained by evaluating Eqs. (17) and (16) at $T = T_{\text{fr}}$. Dependence on the interaction parameter δB is implicit (through T_{fr}) below the freezing transition. For $T > T_{\text{fr}}$, however, the surface response is insensitive to δB . In particular, β_G vanishes, since no binary choice of $f(\sigma)$ can change the second moment of the contact energy distribution. As a consequence, surface and volume behave independently for arbitrary Γ .

The opposite extreme of monomer diversity is described by a smooth form of $p(\sigma)$, describing a continuous variety of chemical identities. We take a Gaussian distribution $p(\sigma) \propto \exp(-\sigma^2/2)$ as a simple example. For weak solvation,

$\Gamma/T \lesssim 1$, $p(\sigma)$ is nowhere significantly depleted, and Eq. (15) remains an appropriate approximation. Gaussian integration yields

$$\frac{F_{\text{surf}}}{K} = -\frac{\Gamma^2}{2T} - \frac{\delta B^4}{T^3} \quad (18)$$

to leading order in $\delta B/T$. (The basic assumption that monomer contacts are statistically independent is plausible only for $\delta B/T_{\text{fr}} = \sqrt{2}s \ll 1$ [15].) The first term in Eq. (18) again reflects a linear response. The second term describes the benefit in monomer contact energy due to partial removal of some monomer types from the globule interior. This effect is independent of solvation strength to leading order and dominates interfacial free energy for very small Γ .

For such a diverse set of monomer types, surface response saturates only when a molecule’s supply of the most solvophilic type is exhausted. Assuming weak depletion is clearly inappropriate here. A maximally solvophilic surface is obtained when $Kf(\sigma) = Np(\sigma)$ for $\sigma \geq \sigma_{\text{max}}$, and $f(\sigma) = 0$ for $\sigma < \sigma_{\text{max}}$. The cutoff point σ_{max} is determined by normalization:

$$\int_{\sigma_{\text{max}}}^{\infty} d\sigma p(\sigma) = \frac{K}{N}. \quad (19)$$

For a Gaussian $p(\sigma)$, Eq. (19) gives $\sigma_{\text{max}} = \sqrt{2 \ln(N/K)} \simeq \sqrt{(2/3) \ln N}$. Because this choice of surface composition uniquely specifies a monomer set G , the associated entropy $s\{f\}$ vanishes. The free energy is then easily estimated from Eq. (9), giving $F_{\text{surf}}/K \simeq -\Gamma \sigma_{\text{max}}$. Comparing this result with the free energy of linear response, we estimate that saturation occurs around $\Gamma \simeq 2T \sigma_{\text{max}} \sim T \sqrt{\ln N}$. Reaching this crossover may thus require much stronger solvation, and result in more favorable surface energy, than in the binary case. The relevant distinction between these distributions is the existence of extremely solvophilic monomers, whose small numbers entail considerable entropic cost in constraining them to the surface.

Figure 1 summarizes the mechanisms of surface response we have identified. These results support a view of surface solvation energy and volume energy as statistically independent random variables. In particular, the linear response corresponding to this notion is valid over a wide range of temperature and solvation strength. Saturation at large Γ , although a nonlinear effect, does not truly arise from correlation of surface and volume. It is instead a consequence of the finitude of surface area or of the number of solvophilic monomers. The regime of weak response, in which $F_{\text{surf}}/K \sim \delta B^4/T^3$, does reflect coupling of surface and volume. But it involves monomer contact energies alone, as indicated by insensitivity to Γ . Within our model, contributions from more intimate connections between surface and volume are small compared to unity when $K \ll N$.

IV. TESTING THE RESULTS WITH SIMULATIONS

Evaluating the accuracy of our predictions for model polymers is made difficult by the paucity of unexposed

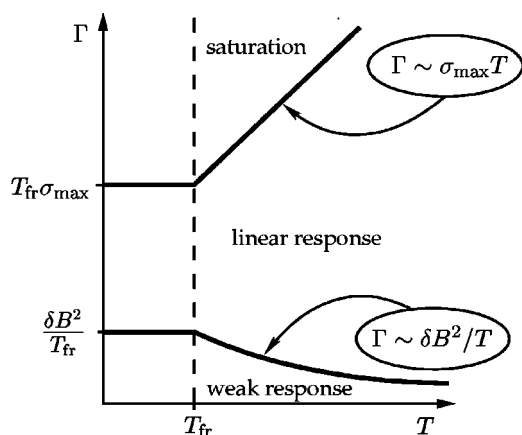


FIG. 1. Response of a random heteropolymer to surface solvation, shown in the plane of temperature T and solvation strength Γ . Crossover lines are the result of equating free energies (or ground state energies) for $T < T_{fr}$. For a binary distribution of monomer types, the weak response regime is absent, and $\sigma_{max}=1$.

monomers in chains of computationally manageable length. Here, we present numerical results for the surface response of a lattice heteropolymer with $N=27$. In any compact conformation of the 27-mer, only one monomer is completely sequestered from the external environment. Nonetheless, there is some variety in the degree of surface exposure of the remaining sites. The eight corners of a $3 \times 3 \times 3$ cube have $n^{(s)}=3$ “solvent” contacts each, while $n^{(s)}=2$ for the 12 edge sites, and $n^{(s)}=1$ for the six faces. For this model we thus define surface composition as $C_{surf} = \sum_{i=1}^{27} n_i^{(s)} \sigma_i$ and surface energy as ΓC_{surf} . For both binary and Gaussian distributions of monomer types, we have determined the ground state conformation of 10^5 random sequences as a function of Γ by enumerating all compact conformations. The deviation of C_{surf} from its maximum value (consistent with polymeric constraints) C_{max} , averaged over these sequences, is plotted in Fig. 2. Results for the two distributions are remarkably similar. Despite the short length of this model chain, the basic features we have described are evident. Surface composition grows roughly linearly for $\Gamma \lesssim \delta B$. Full saturation occurs only for very large Γ and is approached more quickly for the copolymer.

We note that the results of the simulations are consistent with analytical theory despite the small size of the lattice model protein used. Real proteins are longer chains and for them the effect of separation into volume and surface should be more pronounced than for the 27-mers studied here.

V. DISCUSSION AND CONCLUDING REMARKS

The analysis we have presented is based on the mean field view of monomer contact energies underlying the REM. In treating individual contacts as statistically independent quantities, we thus neglect the possibly significant overlap between similar chain conformations. This approximation is justified when only a small fraction of conformational pairs overlap strongly, a condition that depends on the details of a

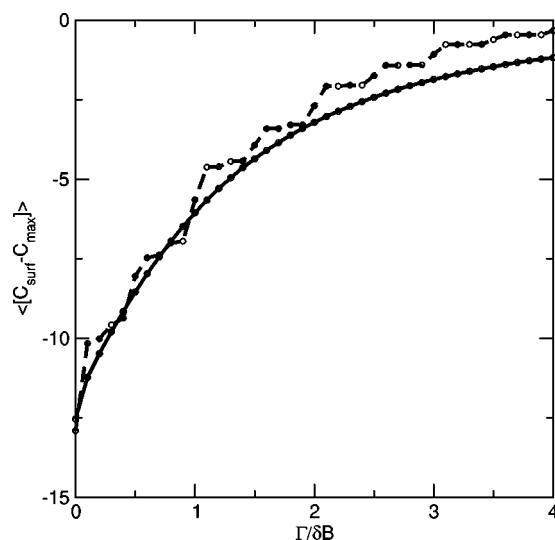


FIG. 2. Ground state surface composition C_{surf} , relative to the maximum possible value C_{max} , averaged over 10^5 random sequences drawn from binary (dashed line) and Gaussian (solid line) distributions.

particular kind of molecule or model chain [6]. Within this model of volume energetics our treatment of surface is, by contrast, not mean field in nature. Indeed, we have shown that variations in subensemble energy minima may be comparable to the most favorable possible solvation energy. It is easily possible that these variations will overwhelm the energetic benefit of a maximally solvophilic surface. A mean field approach disregarding correlated variations in surface and volume energies would instead predict a uniformly solvophilic surface in the ground state.

A potentially more serious shortcoming of our model is the neglect of variations in the number of conformations, ω_G , consistent with different choices of monomers constrained to the surface. A particular choice of G defines the lengths of chain segments that are free to explore the globule interior before returning to the surface. Different “loop” lengths will certainly correspond to differing degrees of conformational entropy. As a first approximation it seems reasonable to assume that these variations are averaged out by grouping choices of G with the same overall collection of monomer types, $f(\sigma)$. A more sophisticated treatment of the heteropolymer surface must account for this more subtle entropic effect of conformational constraints.

The diversity of amino acid monomers comprising proteins lies somewhere between those of binary and Gaussian distributions. The surface behavior we have described should thus be relevant for chains of these units arranged in random sequence. Specifically, we predict that preferential solvation must be much larger than typical thermal excitations in order to stabilize a strictly solvophilic surface. Indeed, the representation of amino acids at the surface of known folded protein structures reflects their relative free energies of transfer from water to a nonpolar solvent [16]. In other words, surface free energy is evidently not optimized in a protein’s native state. Sequences found in nature are not random in at least one respect important to freezing. Their ground states

lie well below the effective continuum of non-native energies. The influence of this energy gap on surface solvation requires a consideration of sequence design that is beyond this discussion.

ACKNOWLEDGMENT

P.L.G. received partial financial support from MIT throughout most of this work.

-
- [1] C. D. Sfatos and E. I. Shakhnovich, *Phys. Rep.* **288**, 77 (1997).
 - [2] V. S. Pande, A. Yu Grosberg, and T. Tanaka, *Rev. Mod. Phys.* **72**, 259 (2000).
 - [3] B. Derrida, *Phys. Rev. Lett.* **45**, 79 (1980).
 - [4] E. I. Shakhnovich and A. M. Gutin, *Biophys. Chem.* **34**, 187 (1989).
 - [5] J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7524 (1987).
 - [6] V. S. Pande, A. Yu. Grosberg, C. Joerg, and T. Tanaka, *Phys. Rev. Lett.* **76**, 3987 (1996).
 - [7] P. L. Geissler and E. I. Shakhnovich, *Phys. Rev. E* **65**, 056110 (2002); *Macromolecules* **35**, 4429 (2002).
 - [8] J. P. Bouchaud and M. Mezard, *J. Phys. A* **30**, 7997 (1997).
 - [9] K. Lau and K. Dill, *Macromolecules* **22**, 3986 (1989).
 - [10] E. Shakhnovich and A. Gutin, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7195 (1993).
 - [11] K. Dill *et al.*, *Protein Sci.* **4**, 561 (1995).
 - [12] S. Kuchanov and A. Khohlov, *J. Chem. Phys.* **118**, 4672 (2003).
 - [13] E. Shakhnovich, *Curr. Opin. Struct. Biol.* **7**, 29 (1997).
 - [14] K. Yue *em et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 325 (1995).
 - [15] C. D. Sfatos, A. M. Gutin, and E. I. Shakhnovich, *Phys. Rev. E* **48**, 465 (1993).
 - [16] C. Chothia, *J. Mol. Biol.* **105**, 1 (1976); R. V. Wolfenden, P. M. Cullis, and C. C. Southgate, *Science* **206**, 575 (1979).