PHYSICAL REVIEW E VOLUME 51, NUMBER 4 APRIL 1995

Freezing transition of random heteropolymers consisting of an arbitrary set of monomers

Vijay S. Pande, Alexander Yu. Grosberg,* and Toyoichi Tanaka
Department of Physics and Center for Materials Science and Engineering,
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
(Received 15 July 1994)

Mean field replica theory is employed to analyze the freezing transition of random heteropolymers comprised of an arbitrary number (q) of types of monomers. Our formalism assumes that interactions are short range and heterogeneity comes only from pairwise interactions, which are defined by an arbitrary $q \times q$ matrix. We show that, in general, there exists a freezing transition from a random globule, in which the thermodynamic equilibrium is comprised of an essentially infinite number of polymer conformations, to a frozen globule, in which equilibrium ensemble is dominated by one or very few conformations. We also examine some special cases of interaction matrices to analyze the relationship between the freezing transition and the nature of the interactions involved.

PACS number(s): 61.41.+e, 64.60.Cn, 87.15.Da, 64.60.Kw

I. INTRODUCTION

The relationship between the sequence and conformation of a heteropolymer is one of the most challenging unsolved problems in biophysics. In the case of proteins, it is widely believed that the native functional conformation is, in a sense, "written" in the sequence of the heteropolymer in the "language" of the interactions between monomer species. This conformation is also believed to be both the ground state from thermodynamic point of view (better to say, it is structurally very close to the ground state, up to some short scale thermal and/or frozen fluctuations) and reliably accessible from the kinetic point of view.

The fact that even chains with random sequences can have a unique frozen ground state was first discussed in terms of phenomenological models [1], where the freezing transition was shown to be similar to that of the random energy model (REM) [2]. The REM-like freezing transition was also derived starting from a microscopic Hamiltonian in which the interactions between pairs of monomers were assumed to be random, independently taken from a Gaussian distribution [3]. In this model, the nature of interactions between species was parametrized in terms of the mean and width of the monomer-monomer interaction distribution. Thus, in this sense, polymer sequence was not explicitly included in this model since it is absent from the Hamiltonian. As for models with polymer sequences explicitly present, two have been considered so far: the two-letter Ising-type model [4] and the so-called p-charge model [5,6]. These models were shown to also exhibit a freezing phase transition for random chains.

Therefore, it is natural to conjecture that any sort of random heteropolymer will have this kind of transition and the question is whether we are able to understand the properties and characteristic temperature of this transition for realistic models of heteropolymers. Indeed, proteins, for example, are comprised of 20 kinds of different monomers, which interact with each other in a complicated manner. There are several relevant types of interactions between different monomers, such as van der Waals interactions, dipole-dipole interactions, hydrogen bonds, and hydrophobic interactions.

As long as we are speaking about short-range interactions, interactions can be described in terms of a matrix: if there are q types of monomers, we have a $q \times q$ matrix, where each (i, j) matrix element represents the energy of interaction between monomers of the types i and j, given that they are in spatial contact. There were several attempts in the literature to derive this kind of "interaction" matrix for real amino acids (see, in particular, [7]). It is rather difficult, however, to derive this kind of matrix. Furthermore, the sensitivity of heteropolymer properties to deviations of the interaction matrix is unclear. For computer simulations, for example, it is important to know how precise one should be in choosing the interaction energies in order to reproduce the native state and to avoid the appearance of some other state, structurally completely different, which may appear as the ground state of a simulated system due to an imperfect interaction matrix. Of course, other nonprotein heteropolymers might be also of interest.

In this paper, we consider the freezing transition for a heteropolymer with an arbitrary interaction matrix. We derive a general formalism for the analysis of the freezing transition of random chains in which only short range interactions are assumed. In addition to the formal benefit that the general treatment establishes a formalism with which other short range species interaction models can be derived as special cases by using specific interaction matrices, this theory can be used to analyze what properties of a species-species interactions matrix effect the freezing transition and in what way.

---

*On leave from Institute of Chemical Physics, Russian Academy of Sciences, Moscow 117977, Russia.

1063-651X/95/51(4)/3381(12)/$06.00 ©1995 The American Physical Society

II. DEVELOPMENT OF THE FORMALISM

A. The model and its Hamiltonian

Consider a heteropolymer chain with a frozen sequence of monomers s_I, where I is the number of monomer along the chain $(1 \leq I \leq N)$ and s_I is the sort of monomer I in the given sequence. Let q be the total number of different monomer species $1 \leq s(I) \leq q$. In the condensed globular state, the spatial structure of the chain is governed by volume interactions between monomers. The disorder and heteropolymer effects of different monomer species comes mainly through pairwise monomer-to-monomer interactions. On the other hand, higher order interactions provide the nonspecific excluded volume effect, while chain connectivity defines the set of available placements of monomers in space. This is clear when one considers the lattice model, where subsequent monomers are nearest neighbors on the lattice (chain connectivity): a site on the lattice can be occupied by only one monomer (excluded volume effect) and the energy is given as a sum of pairwise interactions of the nearest neighbors on the lattice. The complicated set of monomer-monomer interactions, related to the frozen-in sequence, appears then due to the restricted set of pairings of monomers in the space. The interaction part of Hamiltonian can therefore be written in a rather simple way:

$$\mathcal{H} = \sum_{i,j}^{q} \sum_{I,J}^{N} B_{ij}\delta(\mathbf{r}_I - \mathbf{r}_J)\delta(s_I, i)\delta(s_J, j) + \mathcal{H}', \qquad (1)$$

where $B_{ij}\delta(\mathbf{r}_I - \mathbf{r}_J)$ gives the Mayer function of short range interaction between monomers of *species* i and j, placed in space at the distance $\mathbf{r}_I - \mathbf{r}_J$ apart from each other, s_I is the species of monomer number I ("spin" of monomer I), and δ is either the Kroneker or the Dirac delta. Equation (1) has the simple interpretation that monomers number I and J interact based upon their proximity $\delta(\mathbf{r}_I - \mathbf{r}_J)$ and the second virial coefficient of interaction between the species of the two monomers $B_{s_I s_J}$. The $\mathcal{H}'$ contribution contains all higher order interactions of monomers. We assume that it is "homopolymeric" in form, i.e., it does not depend on the monomer species, but only on the overall density ρ. It can be written as $\mathcal{H}' = C\rho^2 + D\rho^3 + \cdots$, where all virial coefficients $C, D, \ldots$ are assumed to be positive (repulsive).

Throughout the paper, we will use the following notation: upper case latin characters label monomer numbers, i.e., bead number along the chain $(1 \leq I \leq N)$; lower case latin characters label monomer species numbers $(1 \leq i \leq q)$; and lower case greek characters are for replica indices $(1 \leq \alpha \leq n)$, which will be defined later. We will be also using the notation for vectors and operators (matrices) with the clear indication of the dimensionality of the corresponding space, as we consider several different spaces simultaneously. For example, the interaction matrix with matrix elements B_{ij} will be denoted as $\widehat{B}^{(q)}$. In this notation, vector $\vec{\rho}^{(q\infty)}$ means the density distribution $\rho_i(\mathbf{R})$ for all species (q) over three-dimensional (3D) $\mathbf{R}$ space (∞).

B. Replicas

The statistical mechanics of a heteropolymer chain is expressed through the partition function, which can be somewhat formally written as

$$Z(\text{seq}) = \sum_{\text{conformations}} \exp\left[-\frac{1}{T}\mathcal{H}(\text{C}, \text{seq})\right], \qquad (2)$$

where we have clearly indicated that our Hamiltonian depends on both conformation C and sequence seq. The standard way to approach the partition function of a system with frozen disorder is to employ, first, the principle of self-averaging of free energy and, second, the replica trick

$$\begin{aligned} F = \langle F(\text{seq})\rangle_{\text{seq}} &= -T\langle \ln Z(\text{seq})\rangle_{\text{seq}} \\ &= \lim_{n\to 0} \frac{\langle Z^n(\text{seq})\rangle_{\text{seq}} - 1}{n}, \end{aligned} \qquad (3)$$

where $\langle\,\rangle_{\text{seq}}$ means average over the set of all possible q^N sequences.

In this paper, we consider random sequences, meaning that the species $1, 2, \ldots, q$ appear independently along the chain with the probabilities $p_1, p_2, \ldots, p_q$ $(\{p_i\} = \vec{p}^{(q)})$, so that the probability of realization of the given sequence $(= s_1, s_2, \ldots, s_I, \ldots, s_N)$ is written as

$$P_{\text{seq}} = p_{s_1} p_{s_2} \cdots p_{s_I} \cdots p_{s_N} = \prod_{I=1}^{N} p_{s_I}. \qquad (4)$$

Collecting the above equations, we can write the key value of the n-replica partition function as

$$\begin{aligned} \langle Z^n(\text{seq})\rangle_{\text{seq}} &= \sum_{\text{seq}} P_{\text{seq}} \left\{ \sum_{\text{conformations}} \exp\left[-\frac{1}{T}\mathcal{H}(\text{C}, \text{seq})\right]\right\}^n \\ &= \sum_{\text{seq}} P_{\text{seq}} \sum_{C_1, \ldots, C_n} \exp\left[-\frac{1}{T}\sum_{\alpha=1}^{n} \mathcal{H}(C_\alpha, \text{seq})\right], \end{aligned} \qquad (5)$$

where $C_\alpha = C_1, \ldots, C_n$ stand for conformations of replica number α.

For each conformation and each replica, we introduce density distributions of all species as

$$m_i^\alpha(\mathbf{R}) = \sum_{I=1}^{N} \delta(s_I, i)\delta(\mathbf{r}_I^\alpha - \mathbf{R}), \quad \{m_i^\alpha(\mathbf{R})\} \equiv \vec{m}^{(qn\infty)}. \qquad (6)$$

For simplicity, we will not explicitly include the sequence independent terms $\mathcal{H}'$ from the original Hamiltonian (1). We then write in terms of the densities

$$\begin{aligned}\langle Z^n(\text{seq})\rangle_{\text{seq}} &= \sum_{\text{seq}} P_{\text{seq}} \sum_{C_1,\ldots,C_n} \exp\left\{-\frac{1}{T}\sum_{\alpha=1}^{n}\sum_{i,j=1}^{q}\int d\mathbf{R}_1 d\mathbf{R}_2\, m_i^\alpha(\mathbf{R}_1)B_{ij}\delta(\mathbf{R}_1-\mathbf{R}_2)m_j^\alpha(\mathbf{R}_2)\right\}\\ &= \sum_{\text{seq}} P_{\text{seq}} \sum_{C_1,\ldots,C_n} \exp\left\{-\frac{1}{T}\left\langle \vec{m}\left|\widehat{B}\right|\vec{m}\right\rangle^{(qn\infty)}\right\}, \end{aligned} \tag{7}$$

where $\langle | | \rangle^{(qn\infty)}$ means a scalar product in which all vectors and operators are supposed to have dimensionality as indicated ($q \times n \times \infty$ in this case). Operator $\widehat{B}^{(qn\infty)}$ is B_{ij} with respect to monomer species and it is diagonal in both replica space and real coordinate space, meaning that it has matrix elements $B_{ij}\delta_{\alpha\beta}\delta(\mathbf{R}_1-\mathbf{R}_2)$. The next step is to perform Hubbard-Stratonovich transformation [4] of the form

$$\langle Z^n(\text{seq})\rangle_{\text{seq}} = \mathcal{N}\sum_{C_1,\ldots,C_n}\int \mathcal{D}\{\phi\}\, \exp\left\{\frac{T}{4}\left\langle\vec{\phi}\left|\widehat{B}^{-1}\right|\vec{\phi}\right\rangle^{(qn\infty)}\right\} \times \sum_{\text{seq}} P_{\text{seq}} \exp\left\{\left\langle\vec{\phi}\,|\,\vec{m}\right\rangle^{(qn\infty)}\right\}. \tag{8}$$

Here $\{\phi_i^\alpha(\mathbf{R})\} = \vec{\phi}^{(qn\infty)}$ are the fields conjugated to the corresponding densities and $\mathcal{N}$ is normalization factor which comes from integration over ϕ.

Note that the sum over sequences enters only in the last "source" term G of this expression

$$\exp\{G\} = \sum_{\text{seq}} P_{\text{seq}} \exp\left\{\left\langle\vec{\phi}\,|\,\vec{m}\right\rangle^{(qn\infty)}\right\}. \tag{9}$$

The summation, or average, over the sequences is easier to describe in nonvector notation:

$$\begin{aligned}\exp\{G\} &= \sum_{s_1,s_2,\ldots,s_N}^{q}\prod_{I=1}^{N}p_{s_I}\prod_{i=1}^{q}\exp\left\{\delta(s_I,i)\sum_{\alpha=1}^{n}\int d\mathbf{R}\;\phi_i^\alpha(\mathbf{R})\delta(\mathbf{r}_I^\alpha-\mathbf{R})\right\}\\ &= \prod_{I=1}^{N}\sum_{s_I=1}^{q}p_{s_I}\prod_{i=1}^{q}\exp\left\{\delta(s_I,i)\sum_{\alpha=1}^{n}\int d\mathbf{R}\;\phi_i^\alpha(\mathbf{R})\delta(\mathbf{r}_I^\alpha-\mathbf{R})\right\}\\ &= \prod_{I=1}^{N}\sum_{i=1}^{q}p_i\exp\left\{\sum_{\alpha=1}^{n}\int d\mathbf{R}\;\phi_i^\alpha(\mathbf{R})\delta(\mathbf{r}_I^\alpha-\mathbf{R})\right\}. \end{aligned} \tag{10}$$

As in the case of a two-letter heteropolymer, to extract the relevant order parameters, we expand over the powers of the fields ϕ (high temperature expansion) and keep terms up to $O(\phi^2)$:

$$\begin{aligned}G &= \sum_{i=1}^{q}\sum_{\alpha=1}^{n}\int d\mathbf{R}\mathbf{R}\rho^\alpha(\mathbf{R})p_i\phi_i^\alpha(\mathbf{R})\\ &+\frac{1}{2}\sum_{i,j=1}^{q}[p_i\delta_{ij}-p_ip_j]\sum_{\alpha,\beta=1}^{n}\int d\mathbf{R}_1\int d\mathbf{R}_2\;\phi_i^\alpha(\mathbf{R}_1)\\ &\times Q_{\alpha\beta}(\mathbf{R}_1,\mathbf{R}_2)\phi_j^\beta(\mathbf{R}_2), \end{aligned} \tag{11}$$

where we use standard definitions [8,3,9]

$$Q_{\alpha_1,\ldots,\alpha_k}(\mathbf{R}_1,\ldots,\mathbf{R}_k) = \sum_{I=1}^{N}\prod_{\kappa=1}^{k}\delta(\mathbf{r}_I^{\alpha_\kappa}-\mathbf{R}_\kappa); \tag{12}$$

$$Q_\alpha(\mathbf{R}) \equiv \rho^\alpha(\mathbf{R}) = \sum_{I=1}^{N}\delta(\mathbf{r}_I^\alpha-\mathbf{R}),$$

$$Q_{\alpha\beta}(\mathbf{R}_1,\mathbf{R}_2) = \sum_{I=1}^{N}\delta(\mathbf{r}_I^\alpha-\mathbf{R}_1)\delta(\mathbf{r}_I^\beta-\mathbf{R}_2). \tag{13}$$

Note that the total density of the polymer chain $\rho^\alpha(\mathbf{R})$ in equilibrium does not depend on replica number and, within a large globule, does not depend on $\mathbf{R}$. Replicas are interpreted as pure states of the polymer chain [10,3,9] and the k-replica order parameter $Q_{\alpha_1,\ldots,\alpha_k}$ is interpreted as the overlap between replicas $\alpha_1,\ldots,\alpha_k$.

The n-replica partition function is now written in the form

$$\begin{aligned}\langle Z^n(\text{seq})\rangle_{\text{seq}} = \mathcal{N}\sum_{C_1,\ldots,C_n}\int\mathcal{D}\{\phi\}\exp\Bigg\{&\left\langle\vec{\phi}\left|\frac{T}{4}B_{ij}^{-1}\delta(\mathbf{R}_1-\mathbf{R}_2)\delta_{\alpha\beta}+\frac{1}{2}Q_{\alpha\beta}(\mathbf{R}_1,\mathbf{R}_2)\Delta_{ij}\right|\vec{\phi}\right\rangle^{(qn\infty)}\\ &+\left\langle\vec{\rho}\middle|\vec{\phi}\right\rangle^{(nq\infty)}\Bigg\}, \end{aligned} \tag{14}$$

where

$$\Delta_{ij} = p_i \delta_{ij} - p_i p_j ,$$

$$\vec{\rho}^{(qn\infty)} \equiv \rho_i^\alpha(\mathbf{R}) = p_i \sum_{I=1}^{N} \delta(\mathbf{r}_I^\alpha - \mathbf{R}) \ . \tag{15}$$

We are left with a Gaussian integral (14) for the n-replica partition function, which is simplified by the argument given in [3,4,9], showing that the $\mathbf{R}$ dependence of $Q_{\alpha\beta}$ is of δ type, so that

$$Q_{\alpha\beta}(\mathbf{R}_1, \mathbf{R}_2) = \rho q_{\alpha\beta} \delta(\mathbf{R}_1 - \mathbf{R}_2) \quad , \tag{16}$$

where diagonal matrix elements of matrix $\widehat{q}^{(n)}$ are 1, while off-diagonal elements are either 0 or 1. This means physically that two replicas α and β might be either uncorrelated (independent), so that $Q_{\alpha\beta} = 0$, or correlated so that one repeats the 3D fold of the other down to the microscopic length scale, so that $Q_{\alpha\beta}(\mathbf{R}_1, \mathbf{R}_2) = \rho\delta(\mathbf{R}_1 - \mathbf{R}_2)$. We do not repeat this argument here, as it is explained elsewhere (see the argument presented in [9], which is slightly different from the original one [3]).

C. Effective energy in replica space

With the simplified form of the Q matrix, we evaluate the Gaussian integral over all ϕ^α variables. This yields

$$\langle Z^n(\text{seq}) \rangle_{\text{seq}} = \sum_{C_1, \ldots, C_n} \exp\left[-NE\{Q\}\right] \tag{17}$$

with energy of the form

$$\begin{aligned} E = &\left\langle \vec{\rho}^{(nq)} \right| \left[T \left(\widehat{B}^{(q)} \right)^{-1} \otimes \widehat{I}^{(n)} \right. \\ &\left. + 2\rho \widehat{q}^{(n)} \otimes \widehat{\Delta}^{(q)} \right]^{-1} \left| \vec{\rho}^{(nq)} \right\rangle \\ &+ \frac{1}{2} \ln\det \left[\frac{T}{4} \left(\widehat{B}^{(q)} \right)^{-1} \otimes \widehat{I}^{(n)} + \frac{1}{2} \rho \widehat{q}^{(n)} \otimes \widehat{\Delta}^{(q)} \right] \\ &+ \frac{1}{2} \ln\det \left(4 \widehat{B}^{(qn)} / T \right) \ . \end{aligned} \tag{18}$$

Here $\otimes$ means the direct product, e.g., for the block matrix $\widehat{B}^{(qn\infty)} = B_{ij}\delta_{\alpha\beta}\delta(\mathbf{R}_1 - \mathbf{R}_2) = \widehat{B}^{(q)} \otimes \widehat{I}^{(n)} \otimes \widehat{I}^{(\infty)}$. In general, $\widehat{A}^{(r)} \otimes \widehat{B}^{(s)}$ produces a block matrix of the total size rs, according to the following rule: instead of each matrix element of $\widehat{A}^{(r)}$ matrix, say A_{uv}, we substitute the block equal to $A_{uv}\widehat{B}^{(s)}$. The last term in (18) comes from the normalization factor $\mathcal{N}$ in Eq. (14); it is easy to check that the normalization factor created by Gaussian integration $\mathcal{N}$ simply eliminates normalization factors first introduced by the Hubbard-Stratonovich transformation. Noting that $\det(\widehat{A})\det(\widehat{B}) = \det(\widehat{A}\widehat{B})$, we can simplify the last relationship as

$$\begin{aligned} E = &\frac{1}{2} \ln\det \left[\widehat{I}^{(qn)} + \frac{2\rho}{T} \widehat{q}^{(n)} \otimes \widehat{\Delta}^{(q)} \widehat{B}^{(q)} \right] \\ &+ \left\langle \vec{\rho}^{(nq)} \right| \frac{1}{T} \widehat{B}^{(q)} \otimes \widehat{I}^{(n)} \left[\widehat{I}^{(qn)} + \frac{2\rho}{T} \widehat{q}^{(n)} \right. \\ &\left. \otimes \widehat{\Delta}^{(q)} \widehat{B}^{(q)} \right]^{-1} \left| \vec{\rho}^{(nq)} \right\rangle , \end{aligned} \tag{19}$$

We have assumed that the Gaussian integral converges and can be calculated. This is guaranteed only by the appropriate form of $\widehat{q}^{(n)}$ matrix, i.e., by replica symmetry breaking. We make an ansatz that $\widehat{q}^{(nq)}$ is of the form of a Parisi matrix with one-step replica symmetry breaking [11,4]. We say that replicas can be gathered into n/x groups, each of which consists of x replicas. The conformations of all of the replicas in a given group coincide with the microscopic scale, i.e., for $\alpha, \beta \in$ group A and $\gamma \in$ group B, then $q_{\alpha\beta} = 1$ and $q_{\alpha\gamma} = q_{\beta\gamma} = 0$. Thus $\widehat{q}^{(nq)}$ can be written as a block matrix (in replica space) which is partitioned into n/x blocks of size $x \times x$ along the diagonal. Inside each diagonal block $q_{\alpha\beta} = 1$ and outside $q_{\alpha\beta} = 0$. In fact, it was recently shown that this form can be derived by energy minimization in the two-letter case [9] and we can easily repeat this argument for the general q-letter case at hand. For the sake of simplicity, however, we omit the derivation and thus formally employ the ansatz.

We can substantially simplify both terms in the energy (19) and convert them into the form

$$\begin{aligned} E = &\frac{n}{2x} \ln\det \left(\widehat{I} + \frac{2\rho x}{T} \widehat{\Delta}\widehat{B} \right) \\ &+ n \left\langle \vec{p} \right| \rho \, \frac{\widehat{B}}{T} \left(\widehat{I} + \frac{2\rho x}{T} \widehat{\Delta}\widehat{B} \right)^{-1} \left| \vec{p} \right\rangle . \end{aligned} \tag{20}$$

Here we have dropped the labels of the dimensionality of the vectors and operators, as all of them are of the same dimensionality (q). This is because we have diagonalized the energy in both $\mathbf{R}$ (∞) and replica (n) spaces, so only the species dimension (q) remains.

The proof of the simplification leading to (20) is given in the Appendix. We now turn to its analysis.

D. Effective entropy in replica space

In order to get the free energy, we must also consider the entropy change due to the constraint on $q_{\alpha\beta}$. Following Refs. [4,12], we argue that due to the polymeric bonds connecting monomers along the chain, once one monomer is fixed in space, the next must be placed within a volume a^3, where a is the distance between monomers along the chain. Since replicas that belong in the same group coincide within a tube of radius $R_t \sim v^{\frac{1}{3}}$, where v is the excluded volume of a single monomer, there are a^3/v ways to place the next monomer and thus the entropy per monomer is just $\ln(a^3/v)$. But since all replica conformations coincide within the group, we must restrict the position of the next monomer to a single place. Following the Parisi ansatz for one-step replica symmetry breaking, for n replicas, there are n/x groups with x replicas per group. The entropy loss is therefore

$$\Delta S = -Ns\frac{n}{x}(x-1), \tag{21}$$

where $s = \ln(a^3/v)$ is related to the flexibility of the chain.

E. Freezing transition

Recall that, for notational convenience, we drop the indication of dimensionality, as all operators and vectors are now assumed to be in species space, i.e., dimensionality q. We optimize the free energy

$$F = \frac{n}{2x}\ln\det\left(\hat{I}+\frac{2\rho x}{T}\hat{\Delta}\hat{B}\right) + n\left\langle \vec{p}\left|\frac{\rho}{T}\hat{B}\left(\hat{I}+\frac{2\rho x}{T}\hat{\Delta}\hat{B}\right)^{-1}\right|\vec{p}\right\rangle + s\frac{n}{x}(x-1) \qquad (22)$$

with respect to x, yielding

$$2s = \ln\det\left(\hat{I}+\frac{2\rho x}{T}\hat{\Delta}\hat{B}\right) - \mathrm{Tr}\left[\frac{2\rho x}{T}\hat{\Delta}\hat{B}\left(\hat{I}+\frac{2\rho x}{T}\hat{\Delta}\hat{B}\right)^{-1}\right] + \left\langle \vec{p}\left|\frac{2\rho x}{T}\hat{B}\frac{2\rho x}{T}\hat{\Delta}\hat{B}\left(\hat{I}+\frac{2\rho x}{T}\hat{\Delta}\hat{B}\right)^{-2}\right|\vec{p}\right\rangle . \qquad (23)$$

As is clear from the very structure of this equation, its solution is of the form $x = T\xi/2\rho$, where ξ is given by

$$2s = \ln\det\left(\hat{I}+\xi\hat{\Delta}\hat{B}\right) - \mathrm{Tr}\left[\xi\hat{\Delta}\hat{B}\left(\hat{I}+\xi\hat{\Delta}\hat{B}\right)^{-1}\right] + \left\langle \vec{p}\left|\xi^2\hat{B}\hat{\Delta}\hat{B}\left(\hat{I}+\xi\hat{\Delta}\hat{B}\right)^{-2}\right|\vec{p}\right\rangle . \qquad (24)$$

Recall that x is the number of replicas in one group, i.e., the number of replicas which have the same conformation down to microscopic fluctuations. This interpretation is clear when n is integer and $n > 1$. While taking the $n \to 0$ limit, we have to consider x to be in between n and 1, so that $x < 1$ means the existence of grouping of replicas, or broken replica permutation symmetry, while $x \to 1$ means the restoration of replica symmetry. Therefore, $x = 1$ defines the point of phase transition between the frozen globular phase with broken replica symmetry and the phase of a random "liquidlike" replica symmetric globule. The corresponding freezing temperature is given by $T_f = 2\rho/\xi$.

Thus, from the n-replica free energy, we obtain the real free energy

$$F = \begin{cases} \frac{T_f}{2}\ln\det\left(\hat{I}+\frac{2\rho}{T_f}\hat{\Delta}\hat{B}\right) + \left\langle \vec{p}\left|\rho\hat{B}\left(\hat{I}+\frac{2\rho}{T_f}\hat{\Delta}\hat{B}\right)^{-1}\right|\vec{p}\right\rangle - s\,(T_f - T) & \text{for } T < T_f \\ \frac{T}{2}\ln\det\left(\hat{I}+\frac{2\rho}{T}\hat{\Delta}\hat{B}\right) + \left\langle \vec{p}\left|\rho\hat{B}\left(\hat{I}+\frac{2\rho}{T}\hat{\Delta}\hat{B}\right)^{-1}\right|\vec{p}\right\rangle & \text{for } T > T_f . \end{cases} \qquad (25)$$

III. DISCUSSION

A. What is $\hat{\Delta}$?

We first examine the physical meaning of the operator $\hat{\Delta}$ and the term $\hat{\Delta}\hat{B}$. From the definition of $\hat{\Delta}$, we have

$$\left(\hat{\Delta}\hat{B}\right)_{ik} = \sum_j \left(p_i\delta_{ij} - p_i p_j\right) B_{jk} = p_i B_{ik} - \sum_j p_i p_j B_{ij}. \qquad (26)$$

We can always write B_{ij} in terms of a sum of a homopolymeric attraction (B_0) and heteropolymeric deviations ($b_{ij} = B_{ij} - \langle B\rangle$). From (26), we see that $\hat{\Delta}$ removes the mean interaction of species k from all matrix elements B_{kj}. In other words, $\hat{\Delta}$ removes all homopolymeric effects.

It is instructive to examine what happens to the energy (20) when one formally takes $\hat{\Delta}\hat{B} = 0$; in this case

$$E = n\,(\rho/T)\left\langle \vec{p}\left|\hat{B}\right|\vec{p}\right\rangle = n\,(\rho/T)\sum_{ij} p_i p_j B_{ij} = n\,(\rho/T)\,\langle B\rangle \ , \qquad (27)$$

which is simply the averaged second virial term. Note that as this term is not coupled to x, $\langle B\rangle$ does not enter into the calculations of the freezing temperature. We note that the terms $n\mathcal{H}' = nC\rho/T + \cdots$ from the original Hamiltonian (1) are not explicitly written, but must be considered when optimizing the free energy. Thus, for $|\langle B\rangle| \gg |b_{ij}|$, we can optimize the free energy with respect to x and ρ independently. However, if this condition is not valid, the coupling between density and the replica overlap order parameter becomes significant; this should lead to other interesting physical phenomena, which are beyond the scope of this paper. The "homopolymeric" attractive second virial term, in competition with the repulsive higher order terms in $\mathcal{H}'$, is responsible for the formation and maintenance of the globular conformation with a reasonably high density. Therefore, $\langle B\rangle$ primarily enters into homopolymer effects, such as the coil to globule transition. Other effects, such as the freezing transition, are purely heteropolymeric and are due to b_{ij} or $\hat{\Delta}\hat{B}$ terms; they are related to the choice of some energetically preferential conformations out of the total vast number of globular conformations.

For the homopolymer case ($q = 1$ or $B_{ij} = B_0$) or the effective homopolymer case (a heteropolymeric interaction matrix is rendered homopolymeric due to the choice of composition $\vec{p}$, say, $p_1 = 1$, while others $p_i = 0$), we immediately see that $\hat{\Delta}\hat{B} = 0$, so $T_f = 0$ and thus there is no freezing transition. (This is of course just a trivial check of consistency of our equations.)

B. Two exactly solvable models

There are some models which can be solved exactly from Eq. (24). We will see that the exact solution of simple models yields insight which will be important in the more general consideration of the next section.

1. Potts model

Potts interactions are defined by the interaction matrix $B_{ij} = b\delta_{ij} + B_0$. The freezing temperature can be found exactly for this model for the case of even composition, i.e., $p_i = 1/q$. From (20), we see that the relevant matrix to address is $\widehat{I} + 2\rho x \widehat{\Delta}\widehat{B}/T$. As the diagonal elements of this matrix are $1+2b\rho x(q-1)/Tq^2$ and all the off-diagonal elements are equal $-2b\rho x/Tq^2$, we find a $(q-1)$-fold degenerate eigenvalue $1 + 2b\rho x/Tq$ and a nondegenerate eigenvalue of 1 (see the Appendix for details). This leads to the energy term of the form

$$\ln\det\left(\widehat{I} + \frac{2\rho x}{T}\widehat{\Delta}\widehat{B}\right) = (q-1)\ln\left(1 + \frac{2b\rho x}{qT}\right). \quad (28)$$

Note that this term vanishes for the homopolymer ($q = 1$) case. As for the other term of the energy (20), it reduces to

$$\left\langle \vec{p}\left|\frac{\rho}{T}\widehat{B}\left(\widehat{I} + \frac{2\rho x}{T}\widehat{\Delta}\widehat{B}\right)^{-1}\right|\vec{p}\right\rangle = \left\langle \vec{p}\left|\frac{\rho}{T}\widehat{B}\right|\vec{p}\right\rangle , \quad (29)$$

i.e., to the average second virial term (27). This term does not contribute to optimization with respect to x. We find the freezing temperature

$$T_f = \frac{-2b\rho}{q\Xi(2s/[q-1])}, \quad (30)$$

where $\Xi(\sigma)$ is given self-consistently by

$$\sigma = \ln(1-\Xi) + \Xi/(1-\Xi) \simeq \begin{cases} \Xi^2/2 & \text{for } \Xi \ll 1 \\ 1/(1-\Xi) & \text{for } \Xi \to 1. \end{cases} \quad (31)$$

We see that the freezing temperature decreases with increasing q. Physically, this corresponds to the fact that in the Potts model, all monomers from differing species interact with each other in the same way, so that the part of the chain without similar monomers is effectively homopolymeric. As q increases, these homopolymerlike regions increase and the freezing temperature consequently decreases. When b is negative (positive), we have physical solutions of T_f for positive (negative) Ξ. We see from Eq. (31) that the nature of the Ξ function is different for positive and negative values of Ξ: there is a singularity at $\Xi = 1$, whereas $\Xi < 0$ is well behaved. Thus there is a fundamental difference between ferromagneticlike ($b < 0$) and antiferromagneticlike ($b > 0$) interactions in terms of the freezing behavior.

Two simplified asymptotic expressions for T_f can be mentioned, coming from the two asymptotics of the $\Xi(\sigma)$ function (31):

$$T_f \simeq \begin{cases} -(\rho b/\sqrt{s})(\sqrt{q-1}/q) & \text{for effectively flexible chains, } 2s/(q-1) \ll 1 \\ -(2\rho b/q)[1 + (q-1)/2s] & \text{for effectively stiff chains, } 2s/(q-1) \gg 1. \end{cases} \quad (32)$$

Recall that the parameter $s = \ln(a^3/v)$ is related to chain flexibility [12], where a and v are the chain spacer size and the monomer excluded volume, respectively; s is small for flexible chains and large for stiff ones. Note that the regions of applicability of the two asymptotics in (31) are controlled by what can be called the effective flexibility $\sigma = s/(q-1)$. Physically, this corresponds again to the specific nature of Potts interactions. Indeed, the main difference between flexible and stiff chains is the number of neighbors along the chain in the interaction sphere in space around a given monomer. This number is large for flexible chains and small for stiff chains. As for Potts interactions, what is relevant is how many neighbors along the chain *attract* a given monomer. This number is obviously reduced by a factor $q-1$ and this explains the appearance of the effective flexibility $s/(q-1)$.

2. p charge

In the p-charge model [5,6], each monomer has a set of p generalized charges, which can be $s_I^k = \pm 1$. The Hamiltonian is defined to be

$$\mathcal{H} = \sum_{I,J}^{N} \delta(\mathbf{r}_I - \mathbf{r}_J) \sum_{k=1}^{p} \chi_k s_I^k s_J^k. \quad (33)$$

In the interaction matrix, we define each possible combination of charges as a different species. Thus there are $q = 2^p$ species in the interaction matrix. For species number i ($1 \leq i \leq q$), the value of charge k is given by $s^k(i) = 2\left(\lfloor i/2^k \rfloor \bmod 2\right) - 1$, where $\lfloor\ \rfloor$ means truncate to the lowest integer. Thus we have an interaction matrix of the form

$$\widehat{B}_{ij} = \sum_k \chi_k \left[2\left(\left\lfloor \frac{i}{2^k} \right\rfloor \bmod 2\right) - 1\right] \times \left[2\left(\left\lfloor \frac{j}{2^k} \right\rfloor \bmod 2\right) - 1\right]. \quad (34)$$

The $\widehat{\Delta}\widehat{B}$ matrix has p nonzero eigenvalues $\chi_1, \chi_2, \ldots, \chi_p$ and a $(2^p - p)$-degenerate eigenvalue of 0. Thus

$$\ln\det\left(\widehat{I}+\frac{2\rho x}{T}\widehat{\Delta B}\right)=\sum_{i=1}^{p}\ln\left(1+\frac{2\chi_i\rho x}{T}\right) \qquad (35)$$

and, as in the Potts case,

$$\left\langle\vec{p}\left|\frac{\rho}{T}\widehat{B}\left(\widehat{I}+\frac{2\rho x}{T}\widehat{\Delta B}\right)^{-1}\right|\vec{p}\right\rangle=\left\langle\vec{p}\left|\frac{\rho}{T}\widehat{B}\right|\vec{p}\right\rangle. \qquad (36)$$

Thus the freezing temperature is determined by

$$2s=\sum_{i=1}^{p}\left[\ln\left(1+\frac{2\chi_i\rho}{T_f}\right)-\frac{2\chi_i\rho/T_f}{1+2\chi_i\rho/T_f}\right]. \qquad (37)$$

For the specific case $\chi_i=\chi$, we have

$$T_f=-\frac{2\rho\chi}{\Xi(2s/p)}\ , \qquad (38)$$

where the $\Xi(\sigma)$ function is defined above by (31). As in Potts interactions, the asymmetry of the Ξ function yields different behavior, depending on the sign of χ. Unlike the Potts case, the behavior of the p-charge model becomes more heteropolymeric, i.e., T_f increases, with the addition of more species.

The two asymptotics, for flexible and stiff chains, in the p-charge model are

$$T_f\simeq\begin{cases}-\rho\chi(2p/s)^{1/2} & \text{for effectively flexible chain, } s/p\ll 1\\ -2\rho\chi(1+p/2s) & \text{for effectively stiff chain, } s/p\gg 1.\end{cases} \qquad (39)$$

Note that effective flexibility is given by $\sigma=s/p$ for the p-charge model, i.e., it is again reduced by the number of species.

We note that our result (38) reproduces automatically what is trivially expected for the homopolymer case ($T_f=0$, i.e., no freezing, when $p=0$) and also at $p=1$ agrees with our previous result (30) at $q=2$ in the case of two-letter Ising heteropolymer. On the other hand, our Eq. (38), or its asymptotics in the first line of Eq. (39), agrees with earlier results of the work [6] in the opposite extreme of $p\gg 1$, i.e., in the region of applicability of that work.

C. Reduction theorems

There are several cases in which the same physical system can be depicted in terms of formally different interaction matrices $\widehat{B}$ and/or composition vectors $\vec{p}$. Clearly, the expression for the freezing temperature, as well as for any other real physical quantity, must not depend on any arbitrary choice.

For example, there might be some monomer species which are formally included in the list and in the interaction matrix, but they are not physically presented in the chain, as the corresponding p vanishes, say, $p_q=0$. It is easy to check that in this case Eq. (24) is reduced to a smaller list of $q-1$ monomer species with a $(q-1)\times(q-1)$ interaction matrix.

Another example is when there are duplicate species, say, species labeled q and $q-1$ are physically identical, i.e., they interact in identical ways to all other species. Physically, we would expect that this problem is identical to the $q-1$ species case, except with the composition $p'_{q-1}=p_{q-1}+p_q$. Even though we skip the proof, Eq. (24) indeed gives this expected reduction. These two statements, which we call "reduction theorems," are not only a good check of consistency of our result (24), but they will also be important in further discussion.

D. Freezing temperature: General consideration

We return to the general analysis of the equation (24) for the freezing temperature and we will show how to implement in the general case the limits of both stiff and flexible chains, similar to how those cases appear in the exact solutions for the Potts and p-charge models.

We first perform an expansion in powers of $\xi=2\rho x/T$. For example,

$$\widehat{B}(I+\xi\widehat{\Delta B})^{-1}=\widehat{B}-\xi\widehat{B}\widehat{\Delta B}+\xi^2\widehat{B}\widehat{\Delta B}\widehat{\Delta B}+\cdots\ . \qquad (40)$$

Note that any term $\widehat{B}\left(\widehat{\Delta B}\right)^k$, where k is a positive integer, is independent of $\langle B\rangle$ and therefore is purely heteropolymeric. The matrix product $\left(\widehat{\Delta B}\right)_{i_1i_2}\left(\widehat{\Delta B}\right)_{i_2i_3}\cdots\left(\widehat{\Delta B}\right)_{i_{k-1}i_k}$ can be interpreted as the propagation of heteropolymeric interactions from monomer species i_1 to i_2, from i_2 to i_3, etc., up to i_k. As we suppose from the very beginning that all of the heterogeneity comes from the second virial coefficient only, so that all higher order virial terms of the original Hamiltonian are in a sense homopolymeric, all heteropolymeric interactions are simply pair collisions of monomers. Each monomer takes part, of course, in a variety of pair collisions during a very long time, i.e., in thermodynamic equilibrium. Those collisions are weighted with the corresponding energies and they form chains of collisions described by $\widehat{B}\left(\widehat{\Delta B}\right)^k$ terms. Depending on both the B_{ij} interaction matrix and the species occurrence probabilities p_i, some of those chains might be more or less favorable than others and this determines freezing transition in the system.

To employ the expansion (40), we first rewrite (24) by noting that $\ln\det\widehat{A}=\mathrm{Tr}\ln\widehat{A}$ and $\left\langle\vec{p}\left|\widehat{A}\right|\vec{p}\right\rangle=\mathrm{Tr}\widehat{P}\widehat{A}$, where $\widehat{P}_{ij} = p_i p_j$:

$$2s = \mathrm{Tr}\left\{ \ln\left(\widehat{I} + \xi\widehat{\Delta}\widehat{B}\right) - \xi\widehat{\Delta}\widehat{B}\left[\widehat{I} + \xi\widehat{\Delta}\widehat{B}\right]^{-1} + \xi^2 \widehat{P}\widehat{B}\widehat{\Delta}\widehat{B}\left[\widehat{I} + \xi\widehat{\Delta}\widehat{B}\right]^{-2} \right\} . \tag{41}$$

Now we are in a position to perform the expansion over the powers of ξ, yielding

$$2s = \sum_{k=2}^{\infty} \xi^k \left\langle B^k \right\rangle_m , \tag{42}$$

where

$$\left\langle B^k \right\rangle_m = \frac{k-1}{k} \mathrm{Tr}\left[\left(-\widehat{\Delta} - k\widehat{P}\right)\widehat{B}\left(-\widehat{\Delta}\widehat{B}\right)^{(k-1)}\right] . \tag{43}$$

The values $\left\langle B^k \right\rangle_m$ can be considered as moments of the $\widehat{B}$ matrix produced by a given $\widehat{\Delta}$ matrix. In fact, we can make the substitution $b_{ij} = B_{ij} - \sum_{kl} p_k p_l B_{kl}$, i.e., remove the "homopolymer" mean from the interaction matrix, and the moments can be rewritten exactly with the exchange $B_{ij} \to b_{ij}$. A consequence of this symmetry is that these moments vanish in the homopolymer case ($b_{ij} = 0$).

We now analyze two opposite extremes in Eq. (41).

1. Freezing temperature: Stiff chain limit

As we are instructed by the examples of Potts and p-charge models, what is important in the high s limit is the singularity of the right-hand side of (41). This is obviously governed by high k terms of power series, which are basically related to $\left(-\widehat{\Delta}\widehat{B}\right)^k$. This is reminiscent of the standard problems of 1D statistical physics, such as the 1D Ising model, the ideal polymer, or other Markovian processes, where $\left(-\widehat{\Delta}\widehat{B}\right)$ plays the role of the transfer matrix. It is well known that highest eigenvalue of the transfer matrix is only relevant in $k \to \infty$ limit ("ground state dominance principle"). In this limit, $\xi \simeq 1/\lambda_{\max}$, where $\lambda_{\max}$ is highest eigenvalue of $\left(-\widehat{\Delta}\widehat{B}\right)$ matrix, and thus

$$T_f \simeq 2\rho\lambda_{\max} . \tag{44}$$

To find the next terms in asymptotic formula for T_f, we note that the most divergent term in Eq. (42) comes from the last term in (41) and is due to $k\widehat{P}$ term in (43); it diverges as $(1 - \xi\lambda_{\max})^{-2}$. We know, however, that this term vanishes for both Potts and p-charge models. Moreover, we can show, that it vanishes also for many other models with some regularities, producing cancellation of correlations and anticorrelations between matrix elements of $\widehat{B}$. For this reason, we keep next to the highest singularity, thus obtaining

$$2s \simeq \frac{c}{(1 - \xi\lambda_{\max})^2} + \frac{1}{(1 - \xi\lambda_{\max})} ,$$
$$c = \left\langle \vec{p} \middle| \vec{\psi} \right\rangle \left\langle \vec{\psi} \middle| -\widehat{B}/\lambda \middle| \vec{p} \right\rangle , \tag{45}$$

where λ and $\left|\vec{\psi}\right\rangle$ are the eigenvalue and the corresponding eigenvector of the $\left(-\widehat{\Delta}\widehat{B}\right)$ operator. This gives finally

$$T_f \simeq 2\rho\lambda_{\max}\left[1 + \frac{1 + \sqrt{1 + 4cs}}{2s}\right]$$
$$\simeq \begin{cases} 2\rho\lambda_{\max}\left[1 + \sqrt{c/s}\right] & \text{for } cs \gg 1 \quad (c \neq 0) \\ 2\rho\lambda_{\max}\left[1 + 1/s\right] & \text{for } cs \ll 1 \quad (c = 0). \end{cases} \tag{46}$$

Note that $\lambda_{\max}$, as an eigenvalue, depends strongly on the arrangement of matrix elements. Therefore the freezing transition for stiff chains is very dependent on the pattern of interactions, not only on their overall heterogeneity. This has clear physical meaning. In the case of stiff chains, real monomers represent the physical units of interaction. In other words, quasimonomers almost coincide with monomers. In terms of propagation, or chains of collisions (see above), it is clear that the highest eigenvalue of the $\left(-\widehat{\Delta}\widehat{B}\right)$ matrix corresponds to the lowest (because of the sign) energy of interaction, while the corresponding eigenvector, in terms of the obvious quantum mechanical analogy, is the linear combination of monomers which realizes this lowest energy and thus controls the freezing temperature.

2. Freezing temperature: Flexible chain limit

The examination of the small s case may be at first glance questionable, as our approach is entirely mean field in nature and therefore it might be applicable for large enough s only. We have seen, however, in the examples of Potts and p-charge models, that the applicability of the flexible chain limit is controlled by the *effective* flexibility, which is considerably smaller than s itself. We therefore consider formally the small s limit, leaving the analysis of applicability for each particular case.

In small s limit, only the first term with $k = 2$ is relevant in the series (42). Omitting all higher order terms, we obtain the remarkably simple result

$$T_f = \frac{2\rho}{\sqrt{s}} \left\langle \widehat{B}^2 \right\rangle_c^{1/2} , \tag{47}$$

where the second cumulant (variance) is defined as $\left\langle \widehat{B}^2 \right\rangle_c \equiv \left\langle \left[\widehat{B} - \left\langle \widehat{B} \right\rangle\right]^2 \right\rangle$ and matrix averages are defined by

$$\left\langle \widehat{B} \right\rangle \equiv \sum_{i,j} p_i p_j B_{ij} . \tag{48}$$

Unlike the stiff chain limit, in the flexible chain case at hand, the freezing transition is controlled mainly by overall heterogeneity of interaction energies B_{ij}. Thus, if one started with an interaction matrix with independent elements and shuffled the matrix elements (even though it is hard to think of real physical experiment of this kind), this transformation would not change the freezing temperature for flexible chains. This is qualitatively a very natural result, as the nature of flexible chains is such that for any given monomer, many of the neighbors in space are neighbors along the chain. In other words, the interaction units are quasimonomers, which are substantially different from the monomers and represent clouds of monomers, where the individuality of each monomer species (with different patterns of energetic preferences to other species) is lost.

In the case of the Potts and p-charge models, the variance of the interaction matrix yields the flexible chain limits for both the Potts (32) and p-charge (38) models. Thus the solution (47) for T_f in this limit is remarkably simple and powerful. To demonstrate this, we show some particular examples.

E. Independent interaction model

In the independent interaction model, all B_{IJ} are taken independently from the Gaussian distribution

$$P\left(B_{IJ}\right) = \left(\frac{\tilde{B}^2}{2\pi}\right)^{1/2} \exp\left[-\frac{\left(B_{IJ}-B_0\right)^2}{\tilde{B}^2}\right] \qquad (49)$$

(recall that capital I and J are related to monomer numbers along the chain and not to species). From the physical point of view, the independence of, say, B_{IJ} and B_{JK} can be realized if and only if the total number of different species is very large, i.e., in the $q \to \infty$ limit. The effective stiffness in this limit is small and we have to use expression (47) for the freezing temperature. Therefore, $T_f = 2\rho\tilde{B}/\sqrt{s}$. This indeed coincides with original result of the work [3].

F. Random sequences of real amino acids

It is of special interest to examine the freezing transition for polymers comprised of real amino acids, i.e., of constituents of real proteins. This can be done using the matrix of interaction energies derived for amino acids by Miyazawa and Jernigan [7]. We are in a position to examine the freezing transition for random sequences (even though real protein sequences might not be random [13–17]). In the work [7], the interaction matrix was given in the form U_{ij}/T_{MJ}, where U_{ij} is the interaction energy and T_{MJ} is a temperature not formally defined in [7]. In some rough approximation, we identify the MJ matrix with our $\rho\widehat{B}$. To avoid rewriting Eq. (24), we substitute the MJ matrix into (24) instead of $\widehat{B}$, meaning that now $\xi = 2T_{MJ}/T_f$. We also assume an equal composition $p_i = 1/q = 1/20$. We can then numerically calculate the ξ vs s dependence. The result is shown in Fig. 1. Note the qualitative similarity of the graph of ξ vs s for the MJ matrix and Ξ vs s given by (31).

Given the realistic value of $s \approx 1.4$ ($v/a^3 \approx 0.25$) for a polypeptide chain, we obtain from Fig. 1 the estimate $\xi \approx 1.6$, or $T_f \approx 1.25T_{MJ}$. By taking a more realistic uneven composition, we arrive at $\xi \approx 1.75$, or $T_f \approx 1.14T_{MJ}$. Note that for the real amino acids system the relevant solution is generally in the high flexibility regime.

To understand these results, recall the way that the MJ matrix was derived in [7]. The protein 3D structures data bank was employed such that if there were $\mathcal{M}_{ij}$ contacts between amino acids labeled i and j in the data bank and the total number of contacts was $\mathcal{M}$, then the ratio $\mathcal{M}_{ij}/\mathcal{M}$ was interpreted as a probability governed by some effective Boltzmann distribution $\mathcal{M}_{ij}/\mathcal{M} = \exp\left[-U_{ij}/T_{MJ}\right]$, thus yielding the MJ matrix of energies U_{ij}. In a later work [18], it was shown that the ratio $\mathcal{M}_{ij}/\mathcal{M}$ indeed obeys a Boltzmann-type formula, if proteins match the random energy model, and

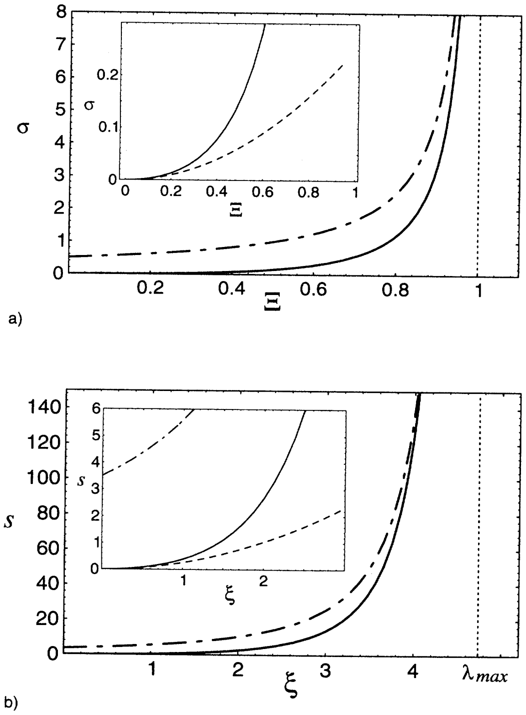

FIG. 1. (a) Plot of the effective flexibility (σ) vs the inverse reduced freezing temperature (Ξ), with the inset of the graph showing the detail of the small σ vs small Ξ regime. The important characteristics of this function are that it is described by Ξ^2 for small Ξ and the existence of a singularity at $\Xi = 1$. The solid line denotes the exact solution, the unevenly dashed line denotes the stiff chain expansion, and the evenly dashed line denotes the flexible chain expansion. (b) For the Miyazawa and Jernigan matrix of amino acid interactions, we plot the flexibility (s) vs the reduced inverse freezing temperature (ξ), with the inset of the graph showing the detail of the small s vs small ξ regime. Qualitatively, this curve is similar to $\Xi(\sigma)$. Further note, however, that most polymers will be described by the small ξ regime. The dashing of curves denotes the same approximations as in (a).

then the parameter of distribution T_{MJ} is nothing but the freezing temperature T_f [19]. From that logic, we expect thus $T_f = T_{MJ}$. Our result is slightly higher. We conclude thus that there is reasonable agreement between the works [8,18] and our results.

IV. CONCLUSION

Starting from a sequence-model Hamiltonian in which interactions between *species* of monomers are expressed in terms of some arbitrary symmetric matrix $\widehat{B}$, we have derived a formalism with which to examine the freezing transition of random heteropolymers. As monomer species interactions are given by some matrix, this formulation is the most general form, assuming that interactions are short range and that heteropolymeric contributions come primarily from two-body interactions.

First, we have related the freezing temperature to the interaction matrix self-consistently. This self-consistent equation can be solved exactly for certain specific systems. For example, models such as the q Potts and p-charge models are important as they describe interesting physical cases, but with only a minimal amount of complexity in their solutions. It is especially interesting that these two simple models have radically different freezing behavior with respect to the number of species. Clearly simply adding different monomer species does not necessarily enhance the freezing transition.

Taking another approach, we can trade the accuracy of an exact result for the generality of the assumption of only some arbitrary symmetric interaction matrix. To this end, we solved perturbatively the exact self-consistent equation. Due to the nature of the Ξ function, there are two regimes of interest: small s (high effective flexibility) where $\Xi \to 0$, and $s \to \infty$, where Ξ approaches a singularity in the self-consistent formulation. Expanding at these two limits, we found

$$T_f \simeq \begin{cases} (2\rho/\sqrt{s}) \left\langle \widehat{B}^2 \right\rangle_c^{1/2} & \text{for small } s \\ \rho\lambda_{\max} & \text{for large } s, \end{cases}$$

where $\lambda_{\max}$ is the largest eigenvalue of the $-\widehat{\Delta}\widehat{B}$ matrix.

The equation above quantitatively details certain descriptions of what one could qualitatively call the "heteropolymeric character" of the interaction matrix $\widehat{B}$ and the species composition $\vec{p}$. Specifically, for flexible chains, one would expect that the physical unit of interactions, or quasimonomers, consists of several monomers. The variance of the interaction matrix gives, in a sense, the heteropolymeric width of interactions. If these interaction energies are ordered in the interaction matrix, however, the correlations between monomer species interactions reduces the heteropolymeric nature of the system and thus reduces the freezing temperature.

In the limit of stiff chains, quasimonomers generally consist of individual monomers. Thus the specific nature of interactions is of paramount importance. In this limit, one can imagine the interactions in space (i.e., not necessarily along the chain) as interactions propagating through the pairwise interactions of monomer species. This chain of interactions, in the stiff polymer limit, becomes very long and thus the system shares characteristics with other one-dimensional systems, such as the 1D Ising model. Specifically, here the freezing temperature is proportional to the largest eigenvalue, which dominates in the long interaction chain limit, of the transfer matrix $-\widehat{\Delta}\widehat{B}$.

To conclude, we comment on the applicability of this theory. Three points are to be mentioned here. First, since we truncate the series (11) to $O(\phi^2)$, we cannot comment on the nature of any phase transitions in the average value of ϕ, such as the microphase segregation seen in two-letter ferromagnetic interaction matrices [4]. Second, all calculations have been mean field and therefore effects due to the fluctuations in the order parameter $Q_{\alpha\beta}$ have been neglected. At present, these effects have been examined in the one-loop approximation only [20,21]. For further details, we refer the reader to the discussions in [4,9]. Even though the issue remains open, to the best of our understanding at this moment, mean field ϕ^2 theory is definitely applicable in the vicinity of the freezing transition, at least when the number of monomer species is more than 2. Third, we have ignored the possibility of liquid crystalline ordering in the stiff chain regime. This issue has not been examined at all; we can only guess qualitatively that orientational ordering should be suppressed in a strongly nonuniform heteropolymer system compared to a homopolymer of comparable stiffness.

Thus, for models with heteropolymeric character, i.e., the interaction matrix and probability distribution cannot be reduced to that of a homopolymer, our theory predicts a freezing transition. Our formalism facilitates the calculation of specific models of interactions, but perhaps most importantly, the direct relationship between the interaction matrix and the freezing transition is demonstrated.

ACKNOWLEDGMENTS

The work was supported by NSF (Grant No. DMR 90-22933) and NEDO of Japan. V.S.P. acknowledges the support of NSF. A.Y.G. acknowledges the support of Kao.

APPENDIX: PROOF OF EQ. (20)

Consider first the auxiliary problem of some $x \times x$ matrix $\widehat{q}^{(x)}$ with diagonal elements $\tilde{q}$ and off-diagonal elements q. This matrix has a $(x-1)$-fold degenerate eigenvalue $\lambda = \tilde{q} - q$, corresponding to the eigenvectors

$$(1 \;\; -1 \;\; 0 \;\; 0 \;\; \cdots \;\; 0),(1 \;\; 0 \;\; -1 \;\; 0 \;\; \cdots \;\; 0),\ldots,$$

$$(1 \;\; 0 \;\; \cdots \;\; 0 \;\; -1 \;\; 0 \;\; \cdots \;\; 0),\ldots,$$

$$(1 \;\; 0 \;\; 0 \;\; \cdots \;\; 0 \;\; -1),$$

and a nondegenerate eigenvalue of $\lambda = \tilde{q} + (x-1)q$, corresponding to the eigenvector $(1 \;\; 1 \;\; 1 \;\; \cdots \;\; 1)$. Of course, there are other ways of choosing eigenvectors; in particular, we can built up an orthonormal basis by choosing

$$\mathcal{R}_{\alpha\beta} = \frac{1}{\sqrt{x}} \exp\left[\frac{2\pi i}{x}(\alpha-1)(\beta-1)\right] \;, \quad 1 \le \alpha, \beta \le x \;. \tag{A1}$$

Here α numerates eigenvectors, while β numerates components of the given eigenvector (or vice versa). We can interpret $\widehat{\mathcal{R}}^{(x)} = \mathcal{R}_{\alpha\beta}$ as the unitary operator transforming $\widehat{q}^{(x)}$ to diagonal form $\widehat{\mathcal{R}}\widehat{q}\widehat{\mathcal{R}}^{-1} = \widehat{\lambda}^{(x)} \equiv \lambda_\alpha \delta_{\alpha\beta}$, with the eigenvalues λ_α given above. (For completeness, we write also the inverse of $\widehat{q}^{(x)}$: it has diagonal elements $(\tilde{q}-q)^{-1} - q\{(\tilde{q}-q)[\tilde{q}+(x-1)q]\}^{-1}$ and off-diagonal elements $-q\{(\tilde{q}-q)[\tilde{q}+(x-1)q]\}^{-1}$.) We will be particularly interested in the case $q = \tilde{q} = 1$. In this case, the nondegenerate eigenvalue is $\lambda = 1$, while all the others are zero.

Consider now some general properties of the "direct product" operation for matrices. We repeat the definition $\widehat{A}^{(r)} \otimes \widehat{B}^{(s)}$ is $rs \times rs$, built up by substitution of $s \times s$ block $A_{uv}\widehat{B}^{(s)}$ instead of each matrix element of $\widehat{A}^{(r)}$.

(i) By matrix row and column operations, it is easy to show that the rule is commutative, i.e.,

$$\widehat{A}^{(r)} \otimes \widehat{B}^{(s)} = \widehat{B}^{(s)} \otimes \widehat{A}^{(r)} \;. \tag{A2}$$

(ii) *The block matrix multiplication rule* states that it is well known that the operation of block matrix multiplication is carried out in the same scheme as normal matrix multiplication, except the multiplication of elements is replaced by the matrix multiplication of blocks. This can be written as

$$\left(\widehat{A}^{(r)} \otimes \widehat{B}^{(s)}\right)\left(\widehat{A}'^{(r)} \otimes \widehat{B}'^{(s)}\right) = \left(\widehat{A}^{(r)}\widehat{A}'^{(r)}\right) \otimes \left(\widehat{B}^{(s)}\widehat{B}'^{(s)}\right) \;. \tag{A3}$$

(iii) The commutation of $\widehat{A}^{(r)} \otimes \widehat{B}^{(s)}$ and $\widehat{A}'^{(r)} \otimes \widehat{B}'^{(s)}$ depends on the commutation of *both* pairs $\widehat{A}^{(r)}, \widehat{A}'^{(r)}$ and $\widehat{B}^{(s)}, \widehat{B}'^{(s)}$ [this directly follows from (ii)].

(iv) The determinant of a block diagonal matrix equals the product of determinants of the diagonal blocks. In particular,

$$\det\left(\widehat{A}^{(r)} \otimes \widehat{I}^{(s)}\right) = \left(\det \widehat{A}^{(r)}\right)^s \;. \tag{A4}$$

(v) The definition of the direct product can be trivially generalized for nonsquare matrices and, in particular, for vectors ($\widehat{A}^{(r\times r')} \otimes \widehat{B}^{(s\times s')}$ is generally the matrix $rs \times r's'$). For example, $\left|\vec{\rho}^{(nq)}\right\rangle = \vec{\rho}^{(n)} \otimes \vec{p}^{(q)}$.

(vi) *The matrix operation with a vector* is shown by

$$\widehat{A}^{(r)} \otimes \widehat{B}^{(s)} \left|\vec{a}^{(r)} \otimes \vec{b}^{(s)}\right\rangle = \widehat{A}^{(r)} \left|\vec{a}^{(r)}\right\rangle \otimes \widehat{B}^{(s)} \left|\vec{b}^{(s)}\right\rangle . \tag{A5}$$

(vii) *The scalar product of vectors* is given by

$$\left\langle \vec{a}^{(r)} \otimes \vec{b}^{(s)} \middle| \vec{a}'^{(r)} \otimes \vec{b}'^{(s)} \right\rangle = \left\langle \vec{a}^{(r)} \middle| \vec{a}'^{(r)} \right\rangle \left\langle \vec{b}^{(s)} \middle| \vec{b}'^{(s)} \right\rangle . \tag{A6}$$

The proof of all the above mentioned properties is straightforward.

Let us return now to the expression of energy (19). We have to address the matrix $\left[\widehat{I}^{(qn)} + \frac{2\rho}{T}\widehat{q}^{(n)} \otimes \widehat{\Delta}^{(q)}\widehat{B}^{(q)}\right]$. We know (or we assume) that $\widehat{q}^{(n)}$ is comprised of n/x $\widehat{q}^{(x)}$ blocks along the diagonal, with $\tilde{q} = q = 1$, that is, $\widehat{q}^{(n)} = \widehat{I}^{(n/x)} \otimes \widehat{q}^{(x)}$. First, this form of $\widehat{q}^{(n)}$ matrix allows us to factor the matrix of our interest:

$$\left[\widehat{I}^{(qn)} + \frac{2\rho}{T}\widehat{q}^{(n)} \otimes \widehat{\Delta}^{(q)}\widehat{B}^{(q)}\right] = \widehat{I}^{(n/x)} \otimes \left[\widehat{I}^{(qx)} + \frac{2\rho}{T}\widehat{q}^{(x)} \otimes \widehat{\Delta}^{(q)}\widehat{B}^{(q)}\right] \;. \tag{A7}$$

This means physically that replicas of different groups are not coupled; they do not interact to each other.

The remainder [in the square brackets on the right-hand side of (A7)] can be diagonalized via the rotation operator $\widehat{\mathcal{R}}^{(xq)} = \widehat{\mathcal{R}}^{(x)} \otimes \widehat{I}^{(q)}$. Indeed, using properties (ii) and (iii) above, we have

$$\left(\widehat{\mathcal{R}}^{(xq)}\right)^{-1} \left[\widehat{I}^{(xq)} + \frac{2\rho}{T}\widehat{q}^{(x)} \otimes \widehat{\Delta}^{(q)}\widehat{B}^{(q)}\right] \widehat{\mathcal{R}}^{(xq)} = \widehat{I}^{(xq)} + \frac{2\rho}{T}\widehat{\lambda}^{(x)} \otimes \widehat{\Delta}^{(q)}\widehat{B}^{(q)} \;. \tag{A8}$$

Recall that there is only one nonzero λ and therefore the last matrix has one $q \times q$ block $(2\rho/T)\,\widehat{\Delta}^{(q)}\widehat{B}^{(q)}$ in the upper-left corner, it has one down this block on the main diagonal, and all other matrix elements are 0.

We are now in a position to simplify the first term of energy (19). First, we apply rule (iv) to this energy term. Then we note that determinant does not change upon rotation (A8), while the determinant of the right-hand side of (A8) is trivially computed, yielding

$$\ln\det\left[\widehat{I}^{(qn)} + \frac{2\rho}{T}\widehat{q}^{(n)} \otimes \widehat{\Delta}^{(q)}\widehat{B}^{(q)}\right] = \frac{n}{x}\ln\det\left[\widehat{I}^{(q)} + \frac{2\rho x}{T}\widehat{\Delta}^{(q)}\widehat{B}^{(q)}\right] . \tag{A9}$$

As for the second term in (19), we first apply rule (vii) to get

$$\left\langle \vec{\rho}^{(nq)} \left| \frac{1}{T}\widehat{B}^{(q)} \otimes \widehat{I}^{(n)} \left[\widehat{I}^{(qn)} + \frac{2\rho}{T}\widehat{q}^{(n)}\widehat{\Delta}^{(q)}\widehat{B}^{(q)} \right]^{-1} \right| \vec{\rho}^{(nq)} \right\rangle$$

$$= \frac{n}{x}\left\langle \vec{\rho}^{(xq)} \left| \left(\frac{1}{T}\widehat{B}^{(q)} \otimes \widehat{I}^{(x)}\right) \left[\widehat{I}^{(qx)} + \frac{2\rho}{T}\widehat{q}^{(x)} \otimes \widehat{\Delta}^{(q)}\widehat{B}^{(q)} \right]^{-1} \right| \vec{\rho}^{(xq)} \right\rangle. \quad \text{(A10)}$$

We then use the rotation (A8) and note that $\widehat{B}^{(q)} \otimes \widehat{I}^{(x)}$ and $\widehat{\mathcal{R}}^{(qx)}$ do commute with each other due to rule (iii). This yields the form

$$\frac{n}{x}\left\langle \vec{\rho}^{(xq)} \left| \left(\widehat{\mathcal{R}}^{(x)} \otimes \widehat{I}^{(q)}\right)^{-1} \left(\frac{1}{T}\widehat{B}^{(q)} \otimes \widehat{I}^{(x)}\right) \left[\widehat{I}^{(qx)} + \frac{2\rho}{T}\widehat{\lambda}^{(x)} \otimes \widehat{\Delta}^{(q)}\widehat{B}^{(q)} \right]^{-1} \left(\widehat{\mathcal{R}}^{(x)} \otimes \widehat{I}^{(q)}\right) \right| \vec{\rho}^{(xq)} \right\rangle . \quad \text{(A11)}$$

We consider, therefore, the rotation of density vector $\left|\vec{\rho}^{(xq)}\right\rangle$. First, we note that $\vec{\rho}^{(xq)} = \vec{\rho}^{(x)} \otimes \vec{p}^{(q)}$. Second, the density, as the physical quantity, is the same for all replicas and does not depend on replica indices. To write it formally, let us define two x-dimensional vectors $\vec{i}^{(x)} = (1 \;\; 1 \;\; 1 \;\; \cdots \;\; 1)$ and $\vec{j}^{(x)} = (1 \;\; 0 \;\; 0 \;\; \cdots \;\; 0)$. Then we see by direct implementation of the formula (A1) $\widehat{\mathcal{R}}^{(x)}\left|\vec{i}^{(x)}\right\rangle = \sqrt{x}\vec{j}^{(x)}$. On the other hand, $\vec{\rho}^{(x)} = \rho\vec{i}^{(x)}$. Therefore, according to rule (v), we have $\widehat{\mathcal{R}}^{(xq)}\left|\vec{\rho}^{(xq)}\right\rangle = \rho\sqrt{x}\vec{j}^{(x)} \otimes \vec{p}^{(q)}$. This yields the energy term in the form

$$\frac{n}{x}\left\langle \rho\sqrt{x}\vec{j}^{(x)} \otimes \vec{p}^{(q)} \left| \left(\frac{1}{T}\widehat{B}^{(q)} \otimes \widehat{I}^{(x)}\right) \right. \right.$$
$$\times \left. \left[\widehat{I}^{(qx)} + \frac{2\rho}{T}\widehat{\lambda}^{(x)} \otimes \widehat{\Delta}^{(q)}\widehat{B}^{(q)} \right]^{-1} \right| \rho\sqrt{x}\vec{j}^{(x)} \otimes \vec{p}^{(q)} \Big\rangle. \quad \text{(A12)}$$

As $\vec{j}^{(x)}$ has only one nonzero component and $\widehat{\lambda}^{(x)}$ has also only one nonzero matrix element, corresponding to the same direction in vector x-dimensional space, we have

$$\left[\widehat{I}^{(qx)} + \frac{2\rho}{T}\widehat{\lambda}^{(x)} \otimes \widehat{\Delta}^{(q)}\widehat{B}^{(q)} \right]^{-1} \left| \vec{j}^{(x)} \otimes \vec{p}^{(q)} \right\rangle$$
$$= \vec{j}^{(x)} \otimes \left[\widehat{I}^{(q)} + \widehat{\Delta}^{(q)}\widehat{B}^{(q)} \right]^{-1} \vec{p}^{(q)}$$
$$+ \left(\vec{i}^{(x)} - \vec{j}^{(x)}\right) \otimes \vec{p}^{(q)} . \quad \text{(A13)}$$

The last step is to implement the scalar product rule (vii), yielding

$$\frac{n}{x}\rho^2 x\left\langle \vec{p}^{(q)} \left| \left(\frac{1}{T}\widehat{B}^{(q)}\right) \left[\widehat{I}^{(q)} + \frac{2\rho}{T}\widehat{\Delta}^{(q)}\widehat{B}^{(q)} \right]^{-1} \right| \vec{p}^{(q)} \right\rangle . \quad \text{(A14)}$$

Combining (A9) with (A14), we arrive at (20).

[1] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).

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

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

[4] C. Sfatos, A. Gutin, and E. Shakhnovich, Phys. Rev. E **48**, 465 (1993).

[5] T. Garel and H. Orland, Europhys. Lett. **6**, 597 (1988).

[6] C. D. Sfatos, A. M. Gutin, and E. I. Shakhnovich, Phys. Rev. E. **50**, 2898 (1994).

[7] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).

[8] T. Garel and H. Orland, Europhys. Lett. **6**, 307 (1988).

[9] V. S. Pande, A. Yu. Grosberg, and T. Tanaka, J. Phys. (France) **4**, 1771 (1994).

[10] M.G. Mezard *et al.*, J. Phys. (Paris) **45**, 843 (1984).

[11] G. Parisi, J. Phys. A **13**, 1887 (1980).

[12] A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP, New York, 1994).

[13] E. I. Shakhnovich and A. M. Gutin, Proc. Natl. Acad. Sci. U.S.A. **90**, 7195 (1993).

[14] V. S. Pande, A. Yu. Grosberg, and T. Tanaka, Proc. Natl. Acad. Sci. U.S.A. **91**, 12 976 (1994).

[15] K. A. Dill, Biochemistry **24**, 1501 (1985).

[16] O. B. Ptitsyn and M. V. Volkenstein, J. Biomol. Struct. Dynamics **4**, 137 (1986).

[17] V. S. Pande, A. Yu. Grosberg, and T. Tanaka, Proc. Natl. Acad. Sci. U.S.A. **91**, 12 972 (1994).

[18] A. V. Finkelstein, A. M. Gutin, and A. Ya. Badretdinov, FEBS Lett. **325**, 23 (1993).

[19] We are indebted to A. Gutin for the discussion of this point.

[20] A. V. Dobrynin and I. Ya. Erukhimovich (private communication).

[21] A. M. Gutin and E. I. Shakhnovich (private communication).