Statistical Mechanics of Simple Models of Protein Folding and Design

Vijay S. Pande,* Alexander Yu. Grosberg,# and Toyoichi Tanaka#
*Physics Department, University of California, Berkeley, California 94720-7300, and #Department of Physics and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA

ABSTRACT It is now believed that the primary equilibrium aspects of simple models of protein folding are understood theoretically. However, current theories often resort to rather heavy mathematics to overcome some technical difficulties inherent in the problem or start from a phenomenological model. To this end, we take a new approach in this pedagogical review of the statistical mechanics of protein folding. The benefit of our approach is a drastic mathematical simplification of the theory, without resort to any new approximations or phenomenological prescriptions. Indeed, the results we obtain agree precisely with previous calculations. Because of this simplification, we are able to present here a thorough and self contained treatment of the problem. Topics discussed include the statistical mechanics of the random energy model (REM), tests of the validity of REM as a model for heteropolymer freezing, freezing transition of random sequences, phase diagram of designed ("minimally frustrated") sequences, and the degree to which errors in the interactions employed in simulations of either folding and design can still lead to correct folding behavior.

INTRODUCTION

In the last decade, there has been remarkable progress in the theoretical understanding of protein folding. This degree of progress is due mainly to the concept of heteropolymer freezing, which is the phase transition of a heteropolymer chain between two compact globular phases, one of which is dominated by an exponentially large number of conformations ($\mathcal{O}(e^N)$), whereas in the other only one or very few of them ($\mathcal{O}(1)$) are thermodynamically relevant. To describe heteropolymer freezing, ideas and concepts were borrowed from the statistical mechanics of spin glasses, such as the random energy model (REM) first suggested by Derrida (1980).

Previous work has centered around two parallel approaches to this problem. Although these two approaches certainly share similar underlying philosophies, they are very different otherwise. One of them, rooted in the seminal work by Bryngelson and Wolynes, postulates the applicability of REM to the heteropolymer freezing problem; one benefit of this approach is that as REM can be solved without recourse to the heavy mathematical formalism of replica field theory, the resulting polymer theory is unencumbered. Not surprisingly, it is extremely difficult, if not altogether impossible, to generalize this approach to different heteropolymeric systems, (for example, systems with long-range interactions, such as random sequence polyampholytes), or to find out its conditions of applicability. Another approach, starting with another seminal contribution by Shakhnovich and Gutin (1989), uses a model that is very clearly formulated, but unfortunately, this theory remains overshadowed by the complexity of theoretical machinery employed (replica theory). Not surprisingly, although this theory is widely considered to be very important, it is hardly known beyond some qualitative conclusions.

In the meantime, both theories are related to REM, which in turn is fairly simple, in both its physical nature and mathematical treatment. It should therefore be possible to formulate a theory that combines the simplicity of the Bryngelson-Wolynes approach with the sophistication of the Shakhnovich-Gutin theory. In this paper we suggest and discuss such an approach. Perhaps the greatest accomplishment of the SG approach is the derivation of REM, starting from the Hamiltonian for polymer-polymer interaction and the nature of the polymer conformation space. In this work we present another derivation that does not rely on replica field theory, but yields the same result. We intentionally try to make this paper self-contained, such that knowledge of some particular works (Derrida, 1980; Bryngelson and Wolynes, 1987; Shakhnovich and Gutin, 1989) is not necessary. Moreover, we review many of the central issues in the statistical mechanics of protein folding.

The paper is organized as follows. We begin with a discussion of models commonly employed in theoretical protein folding studies. We continue with the formulation of REM and discuss why one can hope that REM would be applicable for heteropolymer globules. We then discuss REM in more detail, including the nature of the freezing transition in REM, and arrive at a powerful general relationship that is valid for arbitrary REM-type models. Using this, we derive in a simple yet rigorous way all of the main theoretical results, such as freezing conditions for a heteropolymer with an arbitrary interaction matrix, properties of sequence design, etc.

Received for publication 5 March 1997 and in final form 10 September 1997.

Address reprint requests to Dr. Alexander Grosberg, Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139. Tel.: 617-253-4889; Fax: 617-225-2585; E-mail: shura@gels.mit.edu.

0006-3495/97/12/3192/19 $2.00

COMPACT GLOBULAR HETEROPOLYMER

Energy

We consider heteropolymer chains that are in a maximally compact, globular state. This means, in particular, that the

density of the globule cannot fluctuate and is evenly distributed in space. In the simplest lattice model case, a polymer of N monomers occupies a region with exactly N lattice sites and, therefore, visits every site once and only once. The Hamiltonian that envelops all three important ingredients of the problem, namely, the sequence of the certain set of monomer species, arbitrary interactions between them, and conformations, has the following form:

$$\mathcal{H}(\{s_I\}, \{\mathbf{r}_I\}) = \sum_{I,J}^{N} B_{s_I s_J} \Delta(\mathbf{r}_I - \mathbf{r}_J) \quad (1)$$

Capital Latin indices count the monomers along the chain, $s_I \in \{1, \ldots, q\}$ is the species of monomer I along the chain (and thus $\{s_I\}$ represent the polymer "sequence"), q is the number monomer of species, and $\mathbf{r}_I$ is the position of monomer I ($\{\mathbf{r}_I\}$ represent "conformation"). $\Delta(r)$ is a function concentrated on nearest-neighbor points in space; on the lattice, $\Delta(a) = 1$ and $\Delta(r > a) = 0$, where a is the lattice spacing. Thus our model simply says that the energy of a polymer conformation is determined by the matrix of species-species energies B_{ij} for the monomers in contact. (We assume that heterogeneity comes solely from pairwise interactions (high-order interactions contribute to excluded volume) and the polymer is a dense globule.) In writing the energy in the form of Eq. 1, we implicitly assume that chain connectivity (points $\mathbf{r}_I$ and $\mathbf{r}_{I+1}$ are always next to each other in space for all I), excluded volume ($\mathbf{r}_I \neq \mathbf{r}_J$ for $I \neq J$), and dense packing are all met.

Clearly, Eq. 1, however general, is still an approximation. For example, one could also consider heteropolymeric three-body interactions, which depend on the species of the three monomers in contact, etc. Nevertheless, it does include many essential components of the problem, and we shall restrict ourselves to this model.

Conformations

Different models of conformations have been used in various computational and analytic works, and particular choices of models arise from the desire to capture the essential physics while keeping the model tractable. Protein native states are globules, and therefore any model of folding or design must incorporate maximally compact conformations. Examining only maximally compact conformations renders the problem much more computationally tractable, as the constraint of filling each site on a lattice once and only once greatly reduces the number of possible conformations (Shakhnovich and Gutin, 1990); analytically, this constraint allows one to neglect complications due to density fluctuations. Therefore, it is not surprising that maximally compact conformations have commonly been employed.

Whereas the enumeration of maximally compact conformations is much more feasible than, for example, all conformations of chain with a given length N, exhaustive enumerations of even maximally compact conformations of chains longer than $N = 27$ requires a great deal of computing power (Pande et al., 1994c) (36-mers can easily be enumerated on a massively parallel supercomputer (128 node CM-5) and the maximally compact 48-mer conformations have also been enumerated, but required 2 CPU weeks on the CM-5). Thus 27-mers have become a canonical model for folding and design studies. Recently it has been conjectured that the number of contacts in the space of 27-mers is similar to that found in short proteins (typically 60–80 amino acids) (Onuchic et al., 1995), and thus perhaps 27-mers is not a completely unreasonable model. (For a lattice model, this means nearest neighbors in space; 27-mers, for example, has 28 contacts. For proteins, it has been suggested that contacts be counted in terms of monomers which are neighbors in space and are farther apart than 4 monomers along the chain (Onuchic et al., 1995).) Moreover, in heteropolymer lattice models, a single site is not meant to necessarily represent a single amino acid; instead, one must consider that elements of secondary structure have already been considered in the lattice models, and one is describing some arrangements of these elements. Thus lattice monomers are really renormalized "quasimonomers" (Grosberg and Khokhlov, 1994).

To model larger chains exhaustively, two approaches have been taken. First, polymers have vastly fewer conformations in two dimensions than in three. Thus one can enumerate much longer chains in $2D$ (Chan and Dill, 1993; Dinner et al., 1994). Moreover, it has been argued that enumerable $2D$ chains (25-mers and 36-mers, for example) have a surface-to-volume ratio more similar to proteins than to that of enumerable $3D$ chains (27-mers and 36-mers) (Chan and Dill, 1993). A second approach is to enumerate in $3D$, but restrict the conformation space. For example, "crumpled" 64-mers have been enumerated (a crumpled 64-mer consists of eight $2 \times 2 \times 2$ size 8-mer cubes, strung together to make a single $4 \times 4 \times 4$ cube) (Pande et al., 1996a); they make an interesting model as, unlike shorter chains such as 27-mers and 36-mers, crumpled 64-mers allow small-scale rearrangements and perhaps model multidomain proteins (Pande et al., 1996a; Panchenko et al., 1995).

On the other hand, sampling just the maximally compact conformations may not be sufficient to describe the freezing transition, especially if it is accompanied by a coil-to-globule transition (i.e., the system goes directly from a coil to a frozen globule phase; Pande et al., 1997b; for example, see also Klimov and Thirumalai, 1996). To study all conformations, one must either perform Monte Carlo kinetics and use the Monte Carlo histogram technique (Socci and Onuchic, 1995) to gather information about the density of states or perform a full enumeration of all conformations. Of course, full enumeration is extremely computationally intensive and has only been performed for chains with $N \leq 18$ (Pande et al., 1997a). On the other hand, Monte Carlo simulations have been performed on extremely long chains (in a relative sense), up to $N = 125$ (Abkevich et al., 1994).

In the end, however, differences in conformational spaces are most important for corrections to REM (Pande et al., 1996a) and will not be important for the primary aspects of this work.

Interactions

Natural proteins include $q = 20$ species of monomers, and thus the interaction matrix B_{ij} should be 20×20. Neither the values of its matrix elements nor the aspects of models with a smaller number of species have been agreed upon by the experts. For example, a commonly employed matrix in simple models is the 20×20 MJ matrix, which is extracted from the statistics of the protein data base (Miyazawa and Jernigan, 1985). There have also been other attempts to derive realistic amino acid interaction energies (for example, see Godzik et al., 1995 and references therein). Clearly, these matrices cannot serve as perfect potentials, as the very idea of a "contact" is somewhat approximate or semiqualitative for such bulky molecules as amino acids. On the opposite extreme, as hydrophobicity is believed to be the main driving force of protein collapse, various models are used with just two monomeric species, hydrophobic and polar (Chan and Dill, 1993). The Independent Interaction Model (IIM), where the number of monomer species is as large as the total number of monomers, such that each matrix element B_{ij} enters in the energy of any conformation at most once and never more than once, is somewhat special; matrix elements are then taken independently from a Gaussian distribution. This model is convenient for theorists (as we will see in later sections). The most often used and natural interaction matrices, along with some comments, are given in Table 1.

Furthermore, one may be curious about how the solvent enters into our model Hamiltonian (Eq. 1). We assume that the solvent molecules equilibrate considerably faster than the polymer, and thus we integrate over all solvent degrees of freedom. This leads to an effective interaction between monomer species, i.e., our interaction matrix B_{ij}. For example, the hydrophobic effect explicitly details the interaction between oily molecules and water, but leads to an effective attraction between hydrophobic molecules as they come together to try to avoid the water molecules (Chan and Dill, 1993).

Sequences

The first models of heteropolymers considered just random sequences of monomer species. This was consistent with the fact that real protein sequences statistically look very much like random sequences (Ptitsyn and Volkenstein, 1986; see also Monod, 1971, for the underlying philosophy). However, upon more detailed examination, random sequences were found to be insufficiently protein-like. In particular,

TABLE 1 Commonly employed models of heteropolymer interactions

Name	No. of letters	Matrix	Reference
MJ	20	Realistic amino acid energies	Miyazawa and Jernigan (1985)
IIM	N	Independent random energies	Shakhnovich and Gutin (1989)
Potts	q	$B_{ij} = +1 - 2\delta_{ij}$	Goldenfeld (1992)
Generalized p-charge q-Potts	q^p	$B_{ij} = \sum_{k=1}^{p} b_k \delta\left(\left\lfloor \frac{i}{2^k} \right\rfloor \bmod q, \left\lfloor \frac{j}{2^k} \right\rfloor \bmod q\right)$	Garel and Orland (1988) Pande et al. (1995b)
Generalized BWM	2	$B_{ij} = \sigma_i\sigma_j \sin\theta + \frac{\sigma_i + \sigma_j}{\sqrt{2}} \cos\theta = \begin{pmatrix} -\sqrt{2}\cos\theta + \sin\theta & -\sin\theta \\ -\sin\theta & \sqrt{2}\cos\theta + \sin\theta \end{pmatrix}$	See text below (Geometrical Interpretation)
HP	2	$\theta = -\arccos(\sqrt{2/3}),\ B_{ij} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$	Chan and Dill (1993)
Ising	2	$\theta = \pi/2,\ B_{ij} = -\sigma_i\sigma_j = \begin{pmatrix} +1 & -1 \\ -1 & +1 \end{pmatrix}$	Sfatos et al. (1993)
GLO	2	$\theta = 0,\ B_{ij} = \sigma_i + \sigma_j = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}$	Garel et al. (1994)
LHTW	2	$\theta \approx 0.2,\ B_{ij} = \begin{pmatrix} -2.3 & -1 \\ -1 & 0 \end{pmatrix}$	Li et al. (1996)

For studies of folding and design, 2×2 matrices can be parameterized in terms of a single parameter θ without loss of generality (see Geometrical Interpretation, in text). In this table, we use the following notations. $\sigma_I = \pm 1$ is a "hidden Ising variable" attributed to each monomer in models with just two monomer species. $\lfloor \ldots \rfloor$ means truncate to the lowest integer and $a \bmod b = a - b\lfloor a/b \rfloor$. The q-Potts model of interactions assumes q types of monomer species, with interaction energy between similar and different monomers of -1 and $+1$, respectively. On the other hand, the p-charge model, suggested by Garel and Orland (1988), models the presence of p different physical short-range interactions (an abstraction of screened Coulomb, van der Waals, hydrophobic, etc.). Each monomer is depicted in this model with a set of p generalized "charges," each taking one of two possible values, say 0 or 1. In a generalized Potts model, which generalizes both the q-Potts and p-charge models, each monomer has p different charges, $s^1, \ldots, s^k, \ldots, s^p$, and we allow the values of each charge s^k to range from 0 to $q - 1$. Furthermore, we define the interaction between charges of monomers I and J to be of Potts form: if the value of the charges s_I^k and s_J^k are the same, then the interaction energy is b_k, otherwise it is zero. The total interaction energy between monomers I and J is given as the sum of the interaction energy of the charges of the monomers. Note that the q-Potts model is recovered for $p = 1$, the p-charge model is recovered for $q = 2$, and the Ising model is recovered for $p = 1$ and $q = 2$.

although many chains with sequences taken at random have unique ground states, these ground states are not sufficiently robust such that a minor perturbation of the interaction energies (induced, for example, by the change in the surrounding solution) leads to a complete alteration of the ground-state conformation. Furthermore, for most of the sequences whose respective ground states are unique, folding to these states is neither quick nor reliable (Shakhnovich and Gutin, 1993).

This is, of course, by no means surprising, as protein sequences are known to have undergone evolutionary optimization (Volkenstein, 1994). It was hypothesized (Bryngelson and Wolynes, 1987) that protein evolution has resulted in sequences that obey the "minimal frustration principle." To model the evolutionary-like optimization of proteins, Shakhnovich and Gutin Monte Carlo annealed sequences with the criterion that the sequence minimized the energy of a particular target conformation (Shakhnovich and Gutin, 1993). Another incarnation of the same idea is the so-called Imprinting Model (Pande et al., 1994b). Interestingly, as soon as these models were suggested, the kind of correlations that should exist in protein sequences was predicted, and those correlations were indeed immediately found (Pande et al., 1994a) (see Correlations in Protein Sequences, below, for more details). Later in this work we shall discuss both random and selected sequences.

RANDOM ENERGY MODEL

In this section we digress from the polymer problem and discuss the properties of the REM. The reader who is interested in going straight to polymer freezing can skip to the Summary of REM Properties, below.

What is REM?

Formally, to compute the partition function of an arbitrary system, one needs only the list of all microstates (conformations), 1, 2, . . . , $\mathcal{M}$, with their respective energies E_1, E_2, . . . , $E_{\mathcal{M}}$. Generally, $\mathcal{M}$ is huge, as it scales exponentially with the number of particles (monomers), N:

$$\mathcal{M} \simeq \exp(\omega N) \tag{2}$$

where $\omega \approx 1$ depends on conformations available, i.e., on chain flexibility, packing conditions, lattice geometry in case of lattice models, etc., as discussed above under Conformations. As the system is disordered, all of its energies, E_1, E_2, . . . , $E_{\mathcal{M}}$, depend in general on the realization of disorder, that is, on the sequence. In the REM, one says that the energy of each conformation, say, E_1, is distributed over the realizations of disorder in the same way as energies of all other conformations and is statistically independent of them. If we call $P(E)$ the probability distribution of the energy of some particular conformation over disorder, and $P(E_1, E_2)$ the joint probability distribution that conformations 1 and 2 have energies E_1 and E_2, respectively, then REM dictates that

$$P(E_1, E_2) = P(E_1)P(E_2) \tag{3}$$

It is also usually supposed that the $P(E)$ distribution is Gaussian:

$$P(E) = (2\pi N\mathscr{E}^2)^{-1/2}\exp\left[-\frac{E^2}{2N\mathscr{E}^2}\right] \tag{4}$$

where $\mathscr{E}$ is characteristic width of the distribution. It should be stressed that the Gaussian character of the single-energy distribution (Eq. 4) is less important by far than the statistical independence of states expressed in Eq. 3, which is the hallmark property of REM.

Density of states for REM

To later discuss the REM freezing phase transition, we look at the energy spectrum of a typical realization of disorder (e.g., a random sequence in the polymer problem). It is easy to generate realizations of this spectrum computationally, and two of them are shown as examples in Fig. 1. The figure shows that typical spectra consist of a very dense region, with many states at high and relatively modest energies, and a low-energy part of the spectrum, which is discrete and comprises only a few levels. Furthermore, the continuous part looks identical for all realizations of the disorder, whereas the discrete part is very individual and looks completely different for different realizations.

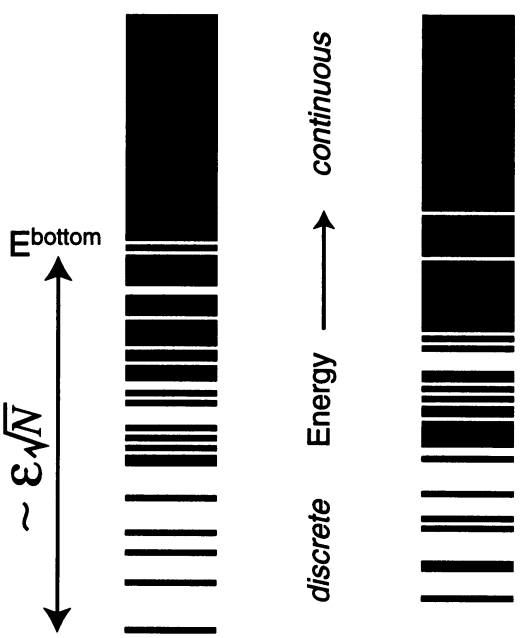

FIGURE 1 Two typical energy spectra for REM. Each consists of continuous and discrete parts. The two realizations demonstrate the property that the continuous part does not depend on realization, whereas the discrete part depends strongly on realization.

We can get an insight into these properties of the REM energy spectrum by examining the density of states, $n(E)$. We remind the reader that $n(E)$ is defined such that $n(E)\Delta E$ is the number of states with energies between E and $E + \Delta E$. It is very easy to write the expectation value for $n(E)$:

$$\langle n(E)\rangle = \mathcal{M}P(E) \tag{5}$$

This value is huge (because of $\mathcal{M}$) whenever E is not far from the central part of the spectrum; in other words, an astronomically large set of states forms an almost continuous spectrum at all energies where the probability in Eq. 4 is not very small. When the density of states is so large, it is essentially the same in all particular realizations, such that $n(E) \simeq \langle n(E)\rangle$. This argument, however, works only as long as $\mathcal{M}P(E) > 1$, or $E > E^{\text{bottom}}$, where

$$\mathcal{M}P(E^{\text{bottom}}) \approx 1 \Rightarrow E^{\text{bottom}} \simeq -N\epsilon\sqrt{2\omega} \tag{6}$$

If we go to low energies $E < E^{\text{bottom}}$, such that this breaks down, then the expectation number of the energy levels in an interval ΔE becomes less then unity. This means, that sometimes, for some realizations, there is one energy level, whereas for others there is not even a single one. Thus we come to the important conclusion that REM in a typical realization has a practically continuous spectrum of states above a certain energy, and a discrete spectrum below it:

$$n(E) = \begin{cases} \mathcal{M}P(E) & \text{when } E > E^{\text{bottom}} \\ \text{random peaks} & \text{when } E < E^{\text{bottom}} \end{cases} \tag{7}$$

It is important that the continuous part of the spectrum is practically independent of the particular realization of the disorder, whereas the discrete part (comprising very few energy levels) is absolutely "individual" for each new realization.

Thermodynamics of REM

Consider now the thermodynamics of REM. While, in principle, one may wish to compute the partition function and the free energy for each particular realization of the disorder, this is clearly impractical for most of the applications. Instead, one notes that the average free energy is dominated by the typical realizations of the disorder. Thus we are first of all interested in the average of the form

$$F(T) = \langle F_{\text{seq}}(T)\rangle = -T\langle \ln Z_{\text{seq}}(T)\rangle \tag{8}$$

To average the logarithmic function is a tedious mathematical task, and this is precisely the reason why disordered systems are so difficult for theoretical examination. This is the place where the famous replica trick (Mezard et al., 1987) enters. The good news about REM is that one does not need to resort to such big theoretical machinery.

Indeed, $Z_{\text{seq}}(T)$, the partition function for the given realization of the disorder, is just the sum over all states $i = 1, 2, \ldots, \mathcal{M}$, and it can always be rewritten in terms of the density of states:

$$Z_{\text{seq}}(T) = \sum_{i=1}^{\mathcal{M}} \exp[-E_i/T] = \int_{-\infty}^{\infty} n(E)e^{-E/T}\,dE \tag{9}$$

At high enough temperatures, this sum is dominated by the states of high entropy (large $n(E)$), where the spectrum is continuous and independent of sequence. This means that all of the complications with the difference between individual realizations of the disorder do not occur in this temperature region, and disorder is, in a way, irrelevant. Indeed, as long as the saddle point of the integral

$$Z = \int_{-\infty}^{\infty} \mathcal{M}P(E)e^{-E/T}\,dE \simeq \mathcal{M}P(E_{\text{saddle}})e^{-E_{\text{saddle}}/T} \tag{10}$$

belongs to the continuous spectrum region $E > E^{\text{bottom}}$, the first line of Eq. 7 is valid, and thus we get a partition function that is independent of the disorder. (The fact that the REM partition function is independent of disorder is very different from what one would expect in a typical disordered system. In extensive quantities (linear in the number of particles N), differences in contributions from each particle are averaged out when we sum up N of them and take the thermodynamic limit (because of the central limit theorem). Because of this property, extensive quantities are called "self-averaging" and are in this sense independent of disorder. Quantities such as the partition function (which scales exponentially in N) typically do not average in this form (for the partition function, terms with lower free energy dominate the average because of the exponential). Thus the fact that the REM partition function is self-averaging is very unusual and results from the assumption of statistical independence of states.) For the Gaussian distribution (Eq. 4), $E_{\text{saddle}} = -N\epsilon^2/T$, we get that $E_{\text{saddle}} > E^{\text{bottom}}$ is valid at $T > T_{\text{glass}} = \epsilon/\sqrt{2\omega}$. Thus, at $T > T_{\text{glass}}$, we can safely average the partition function over disorder (as it does not depend on disorder!) and arrive at the free energy that is also independent of the disorder.

This is not valid at lower temperatures, where just one or a few low energy states dominate the partition function. In principle, one might expect that at this low temperature, the thermodynamics of a particular sample will strongly depend on disorder. Note, however, that typical differences between low energy states are only about $\sqrt{N}$ and they are negligible within thermodynamic limits.

As the free energy is a continuous function of the temperature, we arrive at the following powerful conclusion for REM (see also Koukiou, 1993):

$$F(T) = \begin{cases} -T\ln\langle Z_{\text{seq}}(T)\rangle & \text{if } T > T_{\text{glass}} \\ -T_{\text{glass}}\ln\langle Z_{\text{seq}}(T_{\text{glass}})\rangle & \text{if } T \le T_{\text{glass}} \end{cases} \tag{11}$$

We shall comment in more detail later, that to take the average of the partition function means to take the "an-

nealed average." Thus Eq. 11 shows that for REM, the real quenched average of the free energy coincides, above the temperature T_{glass}, with the annealed average (see Annealed Averaged Free Energy, below). This powerful conclusion is valid for every REM type model, and it will provide us with a tool for further consideration. Finally, we note that as the frozen state has negligible entropy, one can write the simpler relation, $-T_{glass} \ln\langle Z_{seq}(T_{glass})\rangle = \langle E_{gnd}\rangle$.

Summary of REM properties

We summarize here the main properties of REM:

- The defining hallmark property of REM is the statistical independence of states (Eq. 3).
- The REM energy spectrum consists of a continuous part that is independent of disorder and a few discrete energy levels that are placed very individually for each realization of disorder.
- The REM ground state for typical realizations is order $\sqrt{N}$ below the edge of continuous spectrum, which, in turn, is order N below the mean energy. Furthermore, for typical realizations discrete levels are order $\sqrt{N}$ from each other.
- There is a certain temperature for REM, T_{glass}, such that at $T > T_{glass}$ the system explores the high entropy continuous part of its spectrum, whereas at $T < T_{glass}$ it is locked into discrete individual states.
- The free energy of the REM is given by Eq. 11.

Note that these properties are independent of the Gaussian form of the single energy distribution (Eq. 4).

IS REM VALID FOR HETEROPOLYMER FREEZING?

The question posed in the title of this section has been addressed in more detail elsewhere (Pande et al., 1996a). We also acknowledge recent theories that involve corrections to REM, such as those of Franz et al. (1994), Plotkin et al. (1996, 1997), and Pande et al. (1997a). Finally, in Appendix A, we present a nonreplica derivation of REM for heteropolymers; although this derivation does not use the heavy math of replica theory, it does require the concepts of design, which we discuss in later sections. The reader who is ready to trust REM can skip this section and go straight to the next one.

REM cannot be exact for heteropolymers

Bryngelson and Wolynes (1987) postulated the applicability of REM for the folding of protein globules. Shakhnovich and Gutin (1989) showed that REM is applicable for compact heteropolymers with independent interactions (see below). We have to stress that neither Bryngelson and Wolynes (1987) nor Shakhnovich and Gutin (1989) said that the REM was rigorously valid. REM was always considered an approximation. However, the nature of this approximation remained unclear. This is why the statement of REM applicability is often met with understandable distrust. Indeed, REM obviously cannot be exact for heteropolymers. To understand that, let us imagine two conformations, say α and β, each of which represents some small local rearrangement of the other (see Fig. 2). As energies E_α and E_β are given as sums over all pairs of monomers that are in contact (we are speaking now about short-range interactions), they are dominated by identical contributions and differ only because of the small region of difference between α and β. Clearly, these two energies are strongly dependent.

REM and non-REM logic

The REM-like assumption of statistical independence of states is implicit in the motivation of several experimental works; for example, de novo protein design (Hecht et al., 1990) makes an REM-like assumption that the selection of sequences that lower the energy of a desired conformation will not also lower the energies of other conformations.

An opposite intuition is also prevalent in many works, such as the computational generation for a given sequence of a low, but not lowest, energy conformation (Holden, 1995); if REM were valid, then a low but not lowest energy conformation will tell us nothing about the ground state.

How can REM be a good approximation?

Although REM cannot be exact, it is a very good approximation in many cases. Its validity resides in the nature of conformation space, which allows only relatively few local rearrangements of the type shown in the Fig. 2. Typically, this happens because of severe constraints imposed on the conformations when a polymer is maximally compact; this is especially obvious if one thinks of a compact polymer on the lattice, such as 27-mer on a $3 \times 3 \times 3$ cube, where each site is visited once and only once. If that is the case, closely related and thus statistically dependent conformations like those shown in Fig. 2 are rare.

Moreover, as the heteropolymer freezing transition occurs between a phase consisting of exponentially many

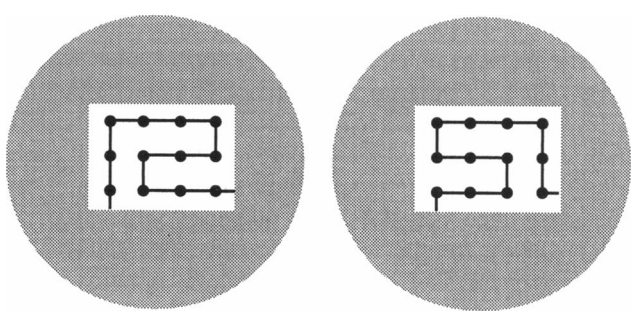

FIGURE 2 Two different though closely related conformations are shown. Everywhere within the shaded area they are identical, and the local difference is seen in the window.

unrelated conformations (i.e., typical overlap between any two vanishes) to a phase consisting of one conformation, any corrections to REM due to the statistical dependence of states will have no effect on the thermodynamics. Such corrections are important for describing protein folding kinetics (Pande et al., 1997a).

Throughout this paper, we assume that REM is applicable. REM violations as well as non-REM generalizations of the theory presented here are discussed elsewhere (Plotkin et al., 1996, 1997; Pande et al., 1996a, 1997a), but in this work we concentrate on the situation when REM is applicable.

ANNEALED HETEROPOLYMER

What is an annealed heteropolymer?

The result (Eq. 11) under Thermodynamics of REM, above, involves the average value of the partition function over all possible sequences. Let us look at this value more closely:

$$\langle Z_{seq}(T)\rangle = \frac{1}{\mathcal{H}} \sum_{seq} Z_{seq}(T) \tag{12}$$

where $\mathcal{H}$ is the total number of sequences. As $Z_{seq}(T)$ itself represents the sum over conformations, the bigger sum $_{seq}Z_{seq}(T)$ has the physical meaning of the partition function of the system in which both conformation and sequence take part in thermal motion on an equal footing. This hypothetical system is called an annealed heteropolymer. Note that the Hamiltonian of annealed heteropolymer is given by the same equation (Eq. 1), but in contrast to the real quenched heteropolymer case, the sequence is not frozen, but represents just another variable defining microstates. To imagine an annealed heteropolymer, one can consider either a polymer whose monomers may change to minimize energy or, equivalently, just a solution of monomers without the polymeric bonds.

An annealed heteropolymer is, in principle, far simpler compared to its quenched counterpart. This is why the relationship in Eq. 11 is so powerful, as it allows one to directly express all of the properties of a real quenched heteropolymer in terms of the much simpler annealed free energy. Although there is no universal exact solution, even for the annealed free energy in terms of B_{ij}, the relationship in Eq. 11 allows the use a variety of approximations or heuristic phenomenological formulae for the annealed free energy. This is similar to the approach of standard polymer theory. Indeed, a polymer fluid is a priori more difficult to study compared to its counterpart of regular small molecules. Given that there is not (and cannot be) a simple and satisfactory theory for the latter, one does not try to create such a theory for the former. Instead, one typically expresses properties of a polymeric liquid in terms of the macroscopic statistical properties of the appropriate low-molecular-weight fluid (which is usually the system of disconnected quasimonomers; Grosberg and Khokhlov, 1994). Similarly, our program here is to employ some qualitatively plausible interpolation for the annealed free energy to gain insight into the freezing behavior of quenched heteropolymers. We stress that the power of the results obtained is not undermined by the approximate character of the annealed free energy that we shall use. By contrast, as long as REM is valid, our method allows for easy incorporation of any potential improvement of the impression for annealed free energy, whether taken from computer simulations, numerical computations for further terms of high temperature expansion, etc.

Accordingly, before proceeding to the quenched case, we first examine the annealed free energy.

Annealed averaged free energy

Annealed averaged free energy of IIM

The only model that allows for an exact solution for annealed free energy is the Independent Interaction Model (IIM), as it is mappable onto the ideal gas problem. In IIM, we assume that there are at least as many monomer species as monomers $q \geq N$, such that interaction energies between the monomers are chosen independently from a Gaussian distribution,

$$P(B_{ij}) = \frac{1}{(2\pi\delta B^2)^{1/2}} \exp\left[-\frac{(B_{ij}-\bar{B})^2}{2\delta B^2}\right] \tag{13}$$

where $\bar{B}$ and δB^2 are the mean and the variance, respectively. In this case, the annealed problem is solved by the following simple argument. To calculate the partition function over both conformations and sequences, let us first fix some arbitrary conformation and consider summation (or average) over sequences. This is illustrated schematically in Fig. 3. In IIM, we do not assign species to every monomer, but rather we assign energy to every possible bond; as we take these energies independently from each other from the probability distribution (Eq. 13), averaging over sequences is reduced to independent averaging over all interaction energies, and this transforms a heteropolymer with a variety of monomers ("colorful pattern") into a homopolymer ("grey background"), with interaction energy given by

$$\exp[-B_{eff}/T] = \int \exp[-B/T]P(B)\mathrm{d}B \tag{14}$$

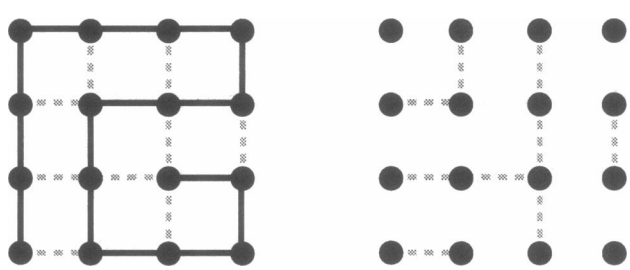

FIGURE 3 In this two-dimensional figure, we illustrate that fixed conformation means a fixed set of bonds between monomers.

Given the Gaussian distribution in Eq. 13, we arrive at

$$B_{\text{eff}} = \bar{B} - \frac{\delta B^2}{2T} \tag{15}$$

As long as we consider only maximally compact conformations, the total number of bonds, each with energy B_{eff}, is the same for all $\mathcal{M}$ conformations, and thus we end up with an annealed average free energy of the form

$$\begin{aligned} F_{\text{ann av}} &= -T \ln\langle Z_{\text{seq}}(T)\rangle \\ &= -T \ln[\mathcal{M} \exp(-\mathcal{Q} B_{\text{eff}}/T)] \\ &= \mathcal{Q}\left[\bar{B} - \frac{\delta B^2}{2T}\right] - TN\omega \end{aligned} \tag{16}$$

where $\mathcal{Q} = \Sigma_{I \neq J} \Delta(\mathbf{r}_I - \mathbf{r}_J)$ is the (independent of the conformation) number of bonds (or contacts) between monomers in any particular compact conformation, and $\omega = -\ln \mathcal{M}/N$ is polymer entropy per monomer. Alternatively, we can arrive at the same answer formally, by calculating the Gaussian integral over B_{ij} in

$$\langle Z_{\text{seq}}(T)\rangle = \prod_{K\,L} P(B_{K\,L}) \sum_{\text{confs}} \exp[-\sum_{I,J}^{N} B_{I\,J} \Delta(\mathbf{r}_I - \mathbf{r}_J)/T] \tag{17}$$

Annealed averaged free energy in terms of high temperature expansion

Unfortunately, for all other models there is no exact solution. Instead, one commonly employs a high temperature expansion to perturbatively calculate the annealed partition function. It may seem unjustified a priori to use a high temperature expansion to study freezing, which seems to be a "low temperature" effect. However, we have to consider that freezing is caused by frustrations that prohibit the system from reaching lower energy microstates of the unfrustrated system. In the polymeric case, the monomers may wish to rearrange themselves into a lower energy configuration, but the polymeric bonds prohibit this. Thus the system is "frozen" at some temperature T_{glass}. The validity of the high temperature expansion in describing freezing resides in the value of T_{glass} compared with the annealed phase transition temperature (i.e., the temperature of the order-disorder transition of the annealed system).

As we did for the IIM above, we begin with averaging over sequences for some given compact conformation and performing a high temperature expansion, keeping terms of order $\mathcal{O}(1/T^2)$ in the annealed average:

$$\begin{aligned} -T \ln W^{\text{conf}}(T) &= -T \ln[\langle \exp[-\mathcal{H}(\text{seq, conf})/T]\rangle] \\ &\simeq \langle \mathcal{H}\rangle - \frac{1}{2T}[\langle \mathcal{H}^2\rangle - \langle \mathcal{H}\rangle^2] \\ &= \mathcal{Q}\left[\bar{B} - \frac{\delta B^2}{2T}\right] \end{aligned} \tag{18}$$

In the last line here, we have performed all averages with the Hamiltonian (Eq. 1). We take the probability for each sequence in the form of the product,

$$P^{(0)}_{\text{seq}} \equiv P^{(0)}(\{s_i\}) = \prod_{I=1}^{N} p_{s_I} \tag{19}$$

which corresponds to monomer species, $\{i\}$, occurring independently with probabilities $\{p_i\}$ (cf. Eq. 12). Averaging is straightforward because, as in the IIM case above, the result does not depend on a particular conformation, as the number of contacts $\mathcal{Q}$ is the same in all compact conformations. In this general case, the mean and variance of the interaction matrix, $\bar{B}$ and δB^2, are defined as

$$\begin{aligned} \bar{B} &= \sum_{ij} p_i B_{ij} p_j \\ \delta B^2 &= \sum_{ij} p_i (B_{ij})^2 p_j - (\sum_{kl} p_k B_{kl} p_l)^2 \\ &= \sum_{ij} p_i (B_{ij} - \bar{B})^2 p_j \end{aligned} \tag{20}$$

When we finally sum (the partition function) over all $\mathcal{M}$ compact conformations, we arrive at

$$F_{\text{ann av}} = \mathcal{Q}\left[\bar{B} - \frac{\delta B^2}{2T}\right] - TN\omega \tag{21}$$

which is pretty much the same as for the IIM (Eq. 16), except for the more general definitions (Eq. 20). Thus, to this order, we are essentially approximating $P(B_{ij})$ by a Gaussian and thus reforming the model into the IIM. Deviations from this behavior will be seen by examining terms in the expansion to higher order. We also note the difference between the free energy of the annealed system and the "annealed average." In the annealed average, the realizations of disorder behave like states, but we average, not sum over them. Thus the entropy of the number of realizations of disorder is present in the annealed system, but not the annealed average.

There are several general properties that derive from aspects of this free energy, which have previously been worked out (Pande et al., 1995a), and we simply repeat here:

- Heteropolymeric effects are independent of the mean of B_{ij}. The mean of B_{ij} only affects the homopolymeric mean attraction. As we assume globular conformations, this mean should be rather attractive for our formalism to be valid. Moreover, small changes such that the mean remains attractive should therefore have no effect.
- Changing the variance of B_{ij} is equivalent to changing the temperature. One can always factor out a constant from the definition of the interaction matrix; this constant only affects the temperature. This fact will be used later under Geometrical Interpretation.

- Reduction theorems. One can choose two matrices that are formally different but physically identical (for example, by simply creating a new "clone" species with identical interaction energies of a previous species). However, as one would expect physically, this does not make a difference in the final calculation, and these cases can be shown to be formally equivalent from the nature of the free energy.

FREEZING TRANSITION

We are now equipped to describe the phase behavior of real quenched heteropolymers with random sequences. Our tools are Eq. 11, which expresses the quenched free energy in terms of the annealed average, and the expression in Eq. 21 for the annealed averaged free energy. For further reference, we here collect these and write the free energy of a real quenched system (averaged over sequences, as in Eq. 8):

$$F(T) \simeq \begin{cases} \mathcal{Q}\left[\bar{B} - \dfrac{\delta B^2}{2T}\right] - TN\omega & \text{if } T > T_{\text{glass}} \\ \mathcal{Q}\left[\bar{B} - \dfrac{\delta B^2}{2T_{\text{glass}}}\right] - T_{\text{glass}}N\omega & \text{if } T \le T_{\text{glass}} \end{cases} \tag{22}$$

Glasslike freezing in REM

Our discussion of REM suggests that something important happens when, because of a temperature decrease, the average energy becomes lower than the boundary of the continuous spectrum. Above the corresponding temperature T_{glass}, the REM-represented system explores many (of order $\mathcal{O}(e^N)$) states and behaves practically independently on the particular realization of disorder. Below this temperature, on the other hand, the equilibrium is dominated by a few discrete states of low energy that vary strongly between each realization of disorder. At $T > T_{\text{glass}}$, the entropy of mixing over the continuous spectrum wins; at $T < T_{\text{glass}}$, the energy of fluctuations involving low-lying energy levels wins. The temperature T_{glass} is called the glass temperature, and the transition is called freezing. Although it is very easy to find that $T_{\text{glass}} = \epsilon/\sqrt{2\omega}$ for the example of Gaussian single energy distribution (Eq. 4), we are now more interested in applying the idea of freezing to the heteropolymer. To this end, we note that the freezing transition is marked by the temperature at which entropy becomes $\mathcal{O}(1)$. In the thermodynamic limit, we can therefore calculate the freezing transition temperature by looking at the point where the entropy vanishes. Thus, to find freezing temperature, we can simply examine the relation

$$S(T_{\text{glass}}) = -\left.\frac{\partial F}{\partial T}\right|_{T=T_{\text{glass}}} = 0 \tag{23}$$

Freezing of random heteropolymers

We use our main relationship (Eq. 22) and find for the entropy of the quenched heteropolymer,

$$S(T) = -\frac{dF}{dT} \simeq N\omega - \mathcal{Q}\frac{\delta B^2}{2T^2} \qquad \text{at } T \ge T_{\text{glass}} \tag{24}$$

Thus $S(T) = 0$ at the temperature T_{glass}, such that

$$T_{\text{glass}}^2 = \frac{\delta B^2}{2s} \tag{25}$$

where $s = \ln a^3/v$ is the conformational entropy per bond (a is the typical distance between monomers and v is the excluded volume) and is therefore related to the entropy per monomer ω by the relation $s = N\omega/\mathcal{Q}$. Higher order corrections can easily be obtained and agree with the previous results of the replica calculation (Shakhnovich and Gutin, 1989; Sfatos et al., 1993; Pande et al., 1995a).

It is instructive to rewrite the expression in Eq. 22 for the free energy of the globule in terms of T_{glass} instead of ω or s; using Eq. 25, we get

$$F(T) \simeq \begin{cases} \mathcal{Q}\left[\bar{B} - \dfrac{\delta B^2}{2T}\left(1 + \dfrac{T^2}{T_{\text{glass}}^2}\right)\right] & \text{at } T \ge T_{\text{glass}} \\ \mathcal{Q}\left[\bar{B} - \dfrac{\delta B^2}{T_{\text{glass}}}\right] & \text{at } T \le T_{\text{glass}} \end{cases} \tag{26}$$

DESIGNED HETEROPOLYMERS

Why should sequences be designed?

When the freezing transition for random heteropolymers was first discovered (Bryngelson and Wolynes, 1987; Shakhnovich and Gutin, 1989), it was believed by some that this was already a good model for protein folding, as it yields a unique ground state with a reasonable (N independent) probability. It was later realized that this ground state, although formally unique, is not sufficiently robust. As the typical energy difference between low energy states scales as $\sqrt{N}$, and in REM systems these low energy states are structurally very different and even unrelated, even a slight change of parameters or solvent conditions leads to a complete alteration of the ground-state conformation (Bryngelson, 1994). Obviously, this is not what happens in nature, where protein native states are remarkably stable.

Another aspect of the problem is that the choice of the ground state of a random sequence cannot be controlled, and thus it is problematic to obtain any desirable properties of the native state out of random choice of sequences.

This discussion suggests that the design of sequences should be directed at choosing atypical realizations from the REM ensemble, such that the ground-state energy is sufficiently below the bottom of the REM continuous spectrum. This was realized in the procedures called sequence selection (Shakhnovich and Gutin, 1993) and Imprinting (Pande et al., 1994b); they can be also viewed as constructive

implementations of the general "principle of minimal frustration" (Bryngelson and Wolynes, 1987).

Microcanonical and canonical design

In principle, one can try to select sequences based directly on their lowest native state (NS) energy, E_{NS}. The corresponding ensemble of sequences is similar to the microcanonical ensemble in the regular statistical mechanics.

In statistical mechanics, it is technically more convenient to use the canonical ensemble, where temperature is fixed instead of energy. A similar idea is also valid for sequence design. We use an analog of the canonical ensemble where the native state energy E_{NS} is not fixed, but rather is controlled through an artificial temperature T_{des}. Equivalently, we choose some "target" conformation $\star$ that we want to be the native state of the designed sequence, and constrain its energy $E_\star = \mathcal{H}(\text{seq}, \star)$ with a Lagrange multiplier $1/T_{des}$. In this canonical ensemble, each sequence appears with Gibbs distributed probability:

$$P^\star_{seq} = \frac{P^{(0)}_{seq} \exp[-\mathcal{H}(\text{seq}, \star)/T_{des}]}{\sum_{seq} P^{(0)}_{seq} \exp[-\mathcal{H}(\text{seq}, \star)/T_{des}]} \tag{27}$$

where $P^{(0)}_{seq}$ is the probability of the sequences made randomly from independent monomer species, with occurrence probabilities p_i; we have already used these probabilities earlier (see Eq. 19).

Thus we characterize a given canonical ensemble of designed sequences by the value of T_{des}: for lower T_{des}, we model sequences whose native states are better optimized energetically, whereas for higher T_{des} we are left with an unaltered ensemble of random sequences.

We note that this prescription for energetic optimization of the native state has many incarnations, including "minimal frustration" (Bryngelson and Wolynes, 1987), "sequence selection" (Shakhnovich and Gutin, 1993), and "Imprinting" (Pande et al., 1995b). Whatever the name, it is important that the procedure be targeted at one particular conformation $\star$, which is completely under control, as it can be chosen arbitrarily. In particular, one can work computationally with target conformations as complex as the ones with a specific "pocket," as in an enzyme active site (Pande et al., 1994b).

There is an interesting question as to whether all possible target conformations $\star$ are equally suitable for design. Computer simulations (Li et al., 1996) indicate that they are not. Currently, however, this effect is beyond what we can understand analytically. It is possible, in particular, that some relatively rare conformations, such as crumpled (Pande et al., 1996a), can be of importance in this respect. In this paper we shall consider all compact $\star$ conformations on an equal basis.

Energy of the target conformation

In this section, we shall compute the energy of the target conformation, averaged over sequences. Using the probability distribution for the designed sequences (Eq. 27), we write

$$\langle E_\star(\text{seq})\rangle = \sum_{seq} P^\star_{seq} E_\star(\text{seq}) = \frac{\sum_{seq} P^{(0)}_{seq} \exp[-\mathcal{H}(\text{seq}, \star)/T_{des}] \cdot \mathcal{H}(\text{seq}, \star)}{\sum_{seq} P^{(0)}_{seq} \exp[-\mathcal{H}(\text{seq}, \star)/T_{des}]} \tag{28}$$

The very structure of this equation suggests the following simple trick (similar to what is done regularly in statistical mechanics). Let us return to the definition in Eq. 18 of the annealed average partition function as a function of temperature:

$$W^\star(T) = \left\langle \exp\left[-\frac{\mathcal{H}(\text{seq}, \star)}{T} \right] \right\rangle = \sum_{seq} P^{(0)}_{seq} \exp[-\mathcal{H}(\text{seq}, \star)/T] \tag{29}$$

Then, in terms of this partition function, we can immediately write

$$\langle E_\star(\text{seq})\rangle = -\frac{1}{W^\star} \frac{\partial W^\star}{\partial(1/T)}\bigg|_{T\to T_{des}} = -\frac{\partial \ln W^\star}{\partial(1/T)}\bigg|_{T\to T_{des}} \tag{30}$$

Up to this point, we have made no approximations. To obtain some concrete result, however, we use our lowest order high temperature expansion for the annealed averaged free energy (Eq. 18) and find that in this approximation, $\ln W^\star(T)$ does not depend on $\star$ and equals

$$-\ln W^\star(T) = \frac{\mathcal{Q}\bar{B}}{T} - \frac{\mathcal{Q}\delta B^2}{2T^2} \tag{31}$$

Therefore $\langle E_\star \rangle$ is given by

$$\langle E_\star\rangle = -\frac{\partial \ln W^\star}{\partial(1/T)}\bigg|_{T\to T_{des}} = \mathcal{Q}\left[\bar{B} - \frac{\delta B^2}{T_{des}}\right] \tag{32}$$

This is a very important result. Here we see directly how the design temperature affects the target state energy. Note that the "susceptibility" of the target state energy to design is proportional to the variance of the interaction matrix: it is natural that it is purely a heteropolymeric effect, as it is based on energy optimization for the target conformation. One can conceptualize this process in terms of the schemes used to computationally model a design: the energy is minimized by swapping monomers in prepolymerized solution, as in Imprinting (Pande et al., 1994b), or by swapping sequences in the sequence space, as in the sequence selection (Shakhnovich and Gutin, 1993).

REM implies a very peculiar spectrum of energies for a heteropolymer with a designed sequence. Note that design affects only the energy of the target conformation, whereas the statistics of all other energies remain unaffected. In this sense, one should understand the common jargon of design,

which describes this as the "pulling down" of just one energy level. This idea is illustrated by Fig. 4. Thus design means selection of very atypical realizations from the REM ensemble.

Note also that the "susceptibility" of the energy level to the design "efforts" to pull this energy down does not depend, to our approximation, on the particular choice of the target state. Thus the effect discussed in recent work (Li et al., 1996) is beyond the approximation employed here.

Folded phase and folding temperature

This peculiar energy spectrum implies a very special freezing behavior for designed sequences. Indeed, when we design at infinite temperature ($T_{des} = \infty$), we are not selective as to the energy of the sequence in the target conformation ★ at all, and the sequences we obtain are random. As we lower T_{des}, we choose sequences in which ★ has lower energy. At the point where T_{des} is sufficiently low such that the energy of ★ is less than that of the REM typical ground state, then ★ is the ground state and we expect freezing to ★. This occurs for $T_{des} < T_{glass}$, as the REM ground state is stable at T_{glass}. Furthermore, freezing of the sequences designed with $T_{des} < T_{glass}$ occurs at the temperature above T_{glass}, because the ground state for those sequences is more stable and of lower energy compared to a typical REM ground state.

Moreover, for $T_{des} < T_{glass}$, the target conformation will be better optimized than the REM ground state and will freeze at some folding temperature T_{fold} that is greater than the glass temperature T_{glass}. We can find T_{fold} by determining which is lower: target energy or random globule free energy.

Thus we compare target state energy $\langle E_\star \rangle$ (Eq. 32) to the random globule free energy expressed most conveniently by the first line of Eq. 26. We find for the folding transition temperature T_{fold}

$$\frac{1}{T_{fold}^2} + \frac{1}{T_{glass}^2} = \frac{2}{T_{fold} T_{des}} \tag{33}$$

Note that this relation is independent of the specific aspects of $\hat{B} = B_{ij}$, although the variance δB^2 enters through T_{glass}. This is an approximation inferred by truncation of the high temperature series, and it is valid for the case in which the number of monomer species is large and the matrix elements of $\hat{B}$ are uncorrelated. For other cases, higher order terms of the high temperature expansion must be used (Pande et al., 1995b).

The results of this section are summarized in the phase diagram (Fig. 5). All phase boundaries are determined directly from the annealed average partition function. This allows quick calculation of even exotic heteropolymer interaction models, within the validity of REM.

We stress that REM is important, not just to calculations, but to the possibility of this simple design scheme (i.e., selecting sequences that minimize the energy in a desired conformation) working at all. Because of the statistical independence of states, we can alter the energy of a particular state without significantly influencing the majority of other states.

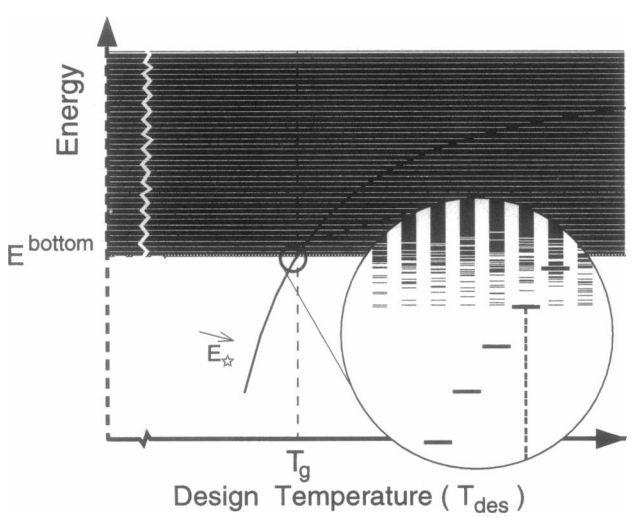

FIGURE 4 Sample energy spectra for sequences imprinted at different polymerization temperatures (T_{des}). The energy of the target conformation ($E_\star$) versus polymerization temperature (T_{des}) is plotted. As T_{des} is increased to T_g, $E_\star$ increases. In the region $T_{des} \approx T_g$ (*magnified section*), we see that $E_\star$ is equal to E^{bottom}, the average ground-state energy of a random chain. This is related to the phase transition between the folded and glassy phases (see phase diagram, Fig. 5). It is instructive to see a realistic representation of the very bottom part of the energy spectrum, as shown here in the magnified section.

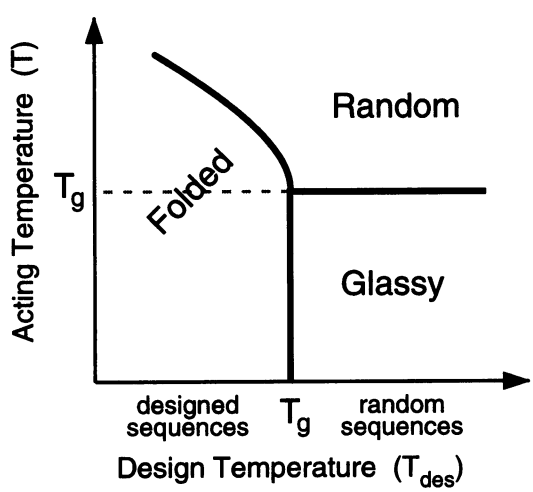

FIGURE 5 For the freezing of globular heteropolymers, there are three phases: 1) Random: an exponential number of globular conformations dominate equilibrium (similar to a homopolymer globular state); 2) Glassy: for sequences that are not well optimized (sufficiently high T_{des}), only order one conformations dominate below the glass temperature T_{glass}, but these conformations are not the target conformation; one can consider random-sequence ground states to be optimized at $T_{des} = T_{glass}$; 3) Folded: the target conformation ★ dominates equilibrium; for $T_{des} < T_{glass}$, ★ is better optimized than the ground state of random sequences.

Computational tests of design

There are at least two reasons to test the conclusions above with computer simulations:

- All of our results rely heavily on REM, although the validity of REM itself may be questionable.
- As in every statistical mechanics approach, our consideration uses the thermodynamic limit. Real polymers are long, but the typical numbers of monomers, often as small as hundreds, are far less than in conventional applications of thermodynamics. Thus it is desirable to test our conclusions for relatively small systems.

To this end, we employ a model that has become a standard for heteropolymer freezing studies: the 27-mer placed on a $3 \times 3 \times 3$ piece of cubic lattice, as shown in Fig. 6. In this paper, we shall not go into the simulations details and only present Fig. 7, which shows a successful comparison between the theoretical and computational phase diagrams. Note that there are no free parameters involved in this comparison.

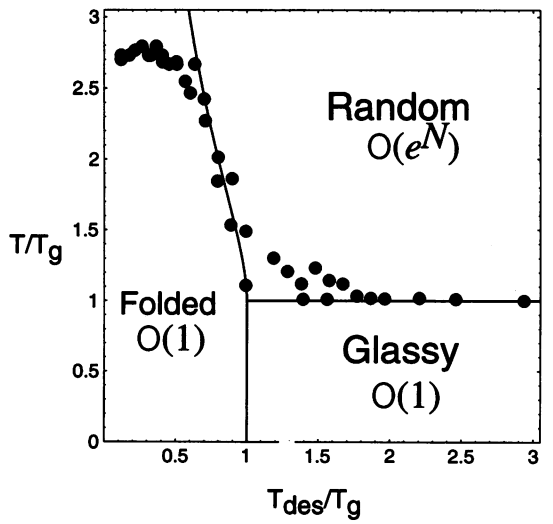

FIGURE 7 Phase diagram for designed 7-Potts model heteropolymers: computer simulation of compact 27-mers on the $3 \times 3 \times 3$ cubic lattice and analytic prediction with no free parameters. The computer simulations generate chains, using Monte Carlo annealing at a given T_{des}. Next, the partition function for maximally compact conformations was exactly calculated. The folding temperature was determined by the temperature at which the REM order parameter $X(T) \equiv 1 - \Sigma_\alpha P_\alpha^2 = 0.9$, where the sum is over all maximally compact conformations α and $P_\alpha = \exp[-\mathcal{H}(\text{conf} = \alpha, \text{seq})/T]$ is the Boltzmann weight of conformation α with the given sequence. We will further discuss the meaning of $X(T)$ below (What Our Approach Cannot Do). Because of finite system effects, one sees that the folding temperature becomes constant for $T_{des} < 0.5$, as we have reached the maximum degree of optimization for 27-mers. Moreover, near $T_{des}/T_{glass} \approx 1$, there are large fluctuations due to small size effects, which lead to small quantitative deviations from our theory. To join the computer simulations with the analytic predictions, we measured the freezing temperature T_{glass} for random sequences directly from the simulation data at high T_{des}. With the measured value of T_{glass} and the calculated value of the variance of the interaction matrix, we plotted the theoretical curves using only the lowest order term $\mathcal{O}(1/T^2)$. The 7-Potts model was used in this demonstration, as it has been shown to be the "most heteropolymeric" (Pande et al., 1994b) for 27-mers. Other models show similar behavior; however, for the 2-Potts model, for example, one must carry further terms in the free energy to get a good quantitative agreement.

Correlations in protein sequences

Design implies that the sequences are not taken at random. One may ask if real protein sequences look random, or if they bear some fingerprint of evolutionary optimization. Even more than that, one can predict, at least qualitatively, the character of sequence correlations that should be expected if proteins are indeed designed in the sense described in the previous sections. This has been done by Pande et al. (1994a), and we briefly summarize the methodology and results below.

From our theoretical models, we expect that evolutionary energy minimization of protein sequences should lead to correlations of monomer species along the chain consistent with energy minimization. For example, we expect that positively charged monomers will be followed predominantly by negatively charged monomers; furthermore, as hydrophobicity induces an effective attraction between hydrophobic monomers, we expect hydrophobic monomers to be followed by other hydrophobic monomers along the protein chain. Moreover, Imprinting predicts much stronger correlations along the chain than other design schemes, as these other schemes do not include the interactions along the chain during design; thus the possible force driving correlations along the chain must come from indirect, multiple body interactions.

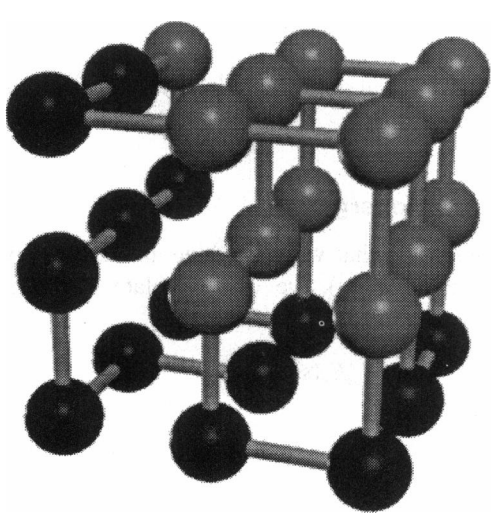

FIGURE 6 27-mers have been commonly employed in lattice computer simulations of folding and design. For example, the 27-mer shown here consists of two types of monomers (black and white), and the conformation shown here is a ground-state conformation for Ising interactions.

To convert from amino acid sequences into a language more convenient for analysis, we used mappings that translate the 20 amino acids into a three-letter code, -1, 0, $+1$. For example, one such mapping translates amino acids

based upon their charge. We next employed a sensitive mathematical apparatus to find correlations in the translated three-letter code sequences. We found correlations in these sequences, but more interestingly, these correlations were consistent with energy minimization. For example, we found anticorrelations in the Coulomb mapping (+1 is typically followed by −1, etc.) and correlations in hydrophobic mappings.

Where are the replicas?

Although we intentionally avoid in this paper the use of the replica trick, some readers may be familiar with it and may be interested in the connections between our approach and the more standard, albeit more heavy, replica treatment. If the reader is not interested in this question, he/she should skip this section and go directly to the next section.

To uncover the parallelism with the replicas, let us consider the annealed average partition function $W^{\star}(T)$ at the special temperature, defined as

$$\frac{1}{T_{\text{eff}}} = \frac{1}{T_{\text{des}}} + \frac{n}{T} \tag{34}$$

and rewrite Eq. 30 in the form

$$\left\langle \frac{E_{\star}(\text{seq})}{T} \right\rangle = -\left. \frac{\partial W^{\star}(T_{\text{eff}})}{\partial n} \right|_{n \to 0} \tag{35}$$

As one would expect, W is mathematically similar to the replicated partition function Z^n in the replica trick. Indeed, we have an annealed average and n replicated Hamiltonians. Here we have no need to interpret n as replicas, but merely as some "external source field" that we eventually set to zero. Therefore, not surprisingly, the results of the nonreplica method agree well with that of the previous replica calculation for design in the matrix formalism (Pande et al., 1995b).

Of course, avoiding difficulties never comes without cost. In this case, we avoid difficult replica calculation by the expense of the use of REM. In the full replica formalism, one can, at least in principle, try to go beyond the REM framework. Some attempts to apply more general models, such as GREM, are discussed by Derrida (1985) and Plotkin et al. (1996). In recent work (Plotkin et al., 1997) researchers have attempted to incorporate sequence design into the GREM type scheme. In the meantime, the approach we suggest here also allows for some generalizations, as we show elsewhere (Pande et al., 1997a,b); interestingly, these works are in many ways very similar to the approach of Plotkin et al. (1997).

DESIGNING AND FOLDING WITH DIFFERENT INTERACTIONS

Why can interactions be different?

Consider, for example, a computer simulation of protein folding or design. We understand now that folding is very sensitive to how well the native state energy is optimized, or how the sequence in question has been designed. On the other hand, one must make some approximations of the nature of the interaction potentials involved. This leads directly to the problem: we believe that the sequence has been designed by Nature, and this natural design was governed, obviously, by natural interactions, and we are now trying to fold this same polymer by using somewhat distorted or "noisy" interactions. Equivalently, we wish to use our approximated potentials to design sequences for (artificial) "proteins," which we want to fold experimentally. If we make a good approximation of potentials, then the results will be good, but if the potentials are sufficiently different, one would expect to arrive at spurious results. Toward this end, there are several questions we can ask:

- How different can these two sets of potentials be and still somewhat accurately model protein folding?
- In what way do we define the "similarity" of set interactions?
- What is the phase behavior of this system?

These questions are also relevant to some other situations. For example, we can imagine that the polymer is folding under somewhat different solvent conditions compared to the environment in which the polymer had been designed.

To examine this problem explicitly, we use two different Hamiltonians, one for design,

$$\mathcal{H}^{\text{d}}(\text{seq}, \text{conf}) = \sum_{\text{I,J}}^{\text{N}} B^{\text{d}}_{s_I s_J} \Delta(\mathbf{r}_I - \mathbf{r}_J) \tag{36}$$

and another for folding,

$$\mathcal{H}^{\text{f}}(\text{seq}, \text{conf}) = \sum_{\text{I,J}}^{\text{N}} B^{\text{f}}_{s_I s_J} \Delta(\mathbf{r}_I - \mathbf{r}_J) \tag{37}$$

We consider that (canonical) design is performed for the target state ⋆ according to Eq. 27, where $\mathcal{H}^{\text{d}}$ is employed for the Gibbs statistical weight of a given sequence to be in the designed ensemble.

Target state energy

In analogy with what we have done above (Energy of the Target Conformation), we can calculate the target state energy, averaged over sequences, by

$$\langle E_{\star}(\text{seq}) \rangle = \sum_{\text{seq}} P^{\star}_{\text{seq}} E_{\star}(\text{seq}) = \frac{\sum_{\text{seq}} P^{(0)}_{\text{seq}} \exp[-\mathcal{H}^{\text{d}}(\text{seq}, \star)/T_{\text{des}}] \cdot \mathcal{H}^{\text{f}}(\text{seq}, \star)}{\sum_{\text{seq}} P^{(0)}_{\text{seq}} \exp[-\mathcal{H}^{\text{d}}(\text{seq}, \star)/T_{\text{des}}]} \tag{38}$$

The only difference from Eq. 28 is that two different Hamiltonians are involved, $\mathcal{H}^{\text{d}}$ and $\mathcal{H}^{\text{f}}$. Nevertheless, we can still

use our approach: we define the annealed average partition function (similar to Eq. 29) for the effective Hamiltonian $\mathcal{H}_{\text{eff}}$,

$$W^\star = \left\langle \exp\left[-\frac{\mathcal{H}_{\text{eff}}(\text{seq}, \star)}{T} \right] \right\rangle = \sum_{\text{seq}} P^{(0)}_{\text{seq}} \exp[-\mathcal{H}_{\text{eff}}(\text{seq}, \star)/T] \quad (39)$$

where

$$\frac{\mathcal{H}_{\text{eff}}(\text{seq}, \star)}{T} = \frac{\mathcal{H}^{\text{d}}(\text{seq}, \star)}{T_{\text{des}}} + n\frac{\mathcal{H}^{\text{f}}(\text{seq}, \star)}{T} \quad (40)$$

Then it is easy to check explicitly that Eq. 38 is reproduced by the following trick:

$$\frac{\langle E_\star(\text{seq})\rangle}{T} = -\frac{1}{W^\star}\frac{\partial W^\star}{\partial n}\bigg|_{n\to 0} = -\frac{\partial \ln W^\star}{\partial n}\bigg|_{n\to 0} \quad (41)$$

(compare with Eq. 35).

This represents the complete REM solution for the problem. Although, in principle, we can use a variety of approximations for the annealed average partition function $W^\star$, to be specific, we use (as we did in previous sections) a high temperature expansion. We proceed as we did in Eq. 18, except with a new interaction matrix,

$$\frac{\hat{\mathcal{B}}_{\text{eff}}}{T} = \frac{\hat{B}^{\text{d}}}{T_{\text{des}}} + n\frac{\hat{B}^{\text{f}}}{T} \quad (42)$$

To lowest order in $1/T$ and $1/T_{\text{des}}$, the annealed average free energy is independent of $\star$ (again, because all compact conformations have the same number of monomer contacts, $\mathcal{Q}$) and it is given by

$$-\ln W^\star = \frac{\mathcal{Q}\overline{\mathcal{B}_{\text{eff}}}}{T} - \frac{\mathcal{Q}\delta\mathcal{B}^2_{\text{eff}}}{2T^2} \quad (43)$$

where

$$\frac{\overline{\mathcal{B}_{\text{eff}}}}{T} = \frac{\overline{B^{\text{d}}}}{T_{\text{des}}} + n\frac{\overline{B^{\text{f}}}}{T} \quad (44)$$

and

$$\frac{\delta\mathcal{B}^2_{\text{eff}}}{T^2} = \sum_{\text{ij}} p_{\text{i}}\left[\frac{B^{\text{d}}_{\text{ij}} - \overline{B^{\text{d}}}}{T_{\text{des}}} + n\frac{B^{\text{f}}_{\text{ij}} - \overline{B^{\text{f}}}}{T}\right]^2 p_{\text{j}} = \frac{\delta B^2_{\text{d}}}{T^2_{\text{des}}} + 2n\sum_{\text{ij}} p_{\text{i}}\frac{\delta B^{\text{d}}_{\text{ij}}}{T_{\text{des}}}\frac{\delta B^{\text{f}}_{\text{ij}}}{T}p_{\text{j}} + n^2\frac{\delta B^2_{\text{f}}}{T^2} \quad (45)$$

Thus

$$\langle E_\star\rangle = \mathcal{Q}\left[\overline{B^{\text{f}}} - \frac{1}{T_{\text{des}}}\sum_{\text{ij}} p_{\text{i}}\delta B^{\text{d}}_{\text{ij}}\delta B^{\text{f}}_{\text{ij}}p_{\text{j}}\right] = \mathcal{Q}\left[\overline{B^{\text{f}}} - \frac{\delta B^2_{\text{eff}}}{T_{\text{des}}}\right] \quad (46)$$

where

$$\delta B^2_{\text{eff}} = \sum_{\text{ij}} p_{\text{i}}\widehat{\delta B}^{\text{d}}_{\text{ij}}\widehat{\delta B}^{\text{f}}_{\text{ij}}p_{\text{j}} \quad (47)$$

Moreover, we see that a particular correlator of the matrix elements is important. We next investigate a geometrical interpretation of this correlator.

Geometrical interpretation

The form of our results suggests the following interpretation. Let us treat the $\hat{B}$ matrices as vectors, albeit with the components B_{ij} numbered with a pair of indices. Subtracting the mean interaction, $\delta B_{\text{ij}} = B_{\text{ij}} - \bar{B}$ geometrically means that the δB_{ij} vector has zero projection along the "main diagonal" (vector $(1, 1, \ldots, 1)$) in this vector space. For any two vectors in this space, say $\mathcal{A}$ and $\mathcal{B}$, we can define the scalar product as $\mathcal{A}\cdot\mathcal{B} = \sum_{\text{ij}} p_{\text{i}}\mathcal{A}_{\text{ij}}\mathcal{B}_{\text{ij}}p_{\text{j}}$. Note that this has nothing to do with the matrix product of the corresponding matrices. From this point of view, both the glass transition temperature T_{glass} (Eq. 25) and the "susceptibility" of the target state energy for design (Eq. 32) are defined by the (squared) length of the δB vector.

Thus Eq. 46 means that, in the general case of two different matrices for folding and design (B^{f} and B^{d}), the situation depends on the angle between δB^{f} and δB^{d}. Indeed, this angle is given by $\cos\phi = g$, where

$$g \equiv \frac{\widehat{\delta B}^{\text{d}}\cdot\widehat{\delta B}^{\text{f}}}{\sqrt{(\widehat{\delta B}^{\text{d}}\cdot\widehat{\delta B}^{\text{d}})(\widehat{\delta B}^{\text{f}}\cdot\widehat{\delta B}^{\text{f}})}} \quad (48)$$

"Parallel" (completely correlated) interactions yield $g = 1$ ($\phi = 0$), and "orthogonal" (completely uncorrelated) interactions yield $g = 0$ ($\phi = \pi/2$).

Models with two types of monomers: generalized BWM (black and white model)

Our geometrical view explains why we wrote the generalized BWM interaction matrix in the form shown in Table 1. Indeed, in the 2×2 symmetrical matrix, there are three independent elements. By setting the mean to 0 and the variance (length of the vector) to 1, we are left with one variable only, which has the natural interpretation as an angle (θ) in the vector space.

In terms of this angle θ, matrices with zero mean, unit variance, and even composition (equal fraction of all monomers, $p_{\text{i}} = 1/2$) have the form

$$B_{\text{ij}}(\theta) = \begin{pmatrix} -\sqrt{2}\cos\theta - \sin\theta & \sin\theta \\ \sin\theta & \sqrt{2}\cos\theta - \sin\theta \end{pmatrix} \quad (49)$$

This result can also be generalized for arbitrary p_{i}, but becomes considerably more cumbersome. (In general, one can express all symmetrical 2×2 matrices with zero mean, unit variance, and fixed composition $p_1 = (1 + \epsilon)/2$ and

$p_2 = (1 - \epsilon)/2$, with the matrix elements

$$B_{11}(\epsilon, \theta) = -\sqrt{\frac{2(1-\epsilon)}{(1+\epsilon)}}\cos\theta - \frac{(1-\epsilon)}{(1+\epsilon)}\sin\theta$$

$$B_{22}(\epsilon, \theta) = \sqrt{\frac{2(1+\epsilon)}{(1-\epsilon)}}\cos\theta - \frac{(1+\epsilon)}{(1-\epsilon)}\sin\theta$$

$$B_{12}(\epsilon, \theta) = \epsilon\sqrt{\frac{2}{(1+\epsilon)(1-\epsilon)}}\cos\theta + \sin\theta$$

For even composition ($\epsilon = 0$), the expression above reduces to Eq. 49.) As one might suspect, θ not only is a means of parameterizing this space of matrices, but also describes the similarity of matrices; in the previous section we have shown that the similarity of interaction matrices is given by the "dot product," $g = (1/4)\,\Sigma_{ij}\,B_{ij}(\theta)B_{ij}(\theta')$. Similar matrices have overlap $g \approx 1$, unrelated ("orthogonal") matrices $g = 0$, and anticorrelated matrices (e.g., ferromagnetic versus antiferromagnetic interactions) have $g = -1$. We find that θ is a metric in 2×2 matrix space, as the distance between matrices is given by $g(\theta, \theta') = \cos(\theta - \theta')$; in other words, $\phi = \theta - \theta'$.

It is interesting to consider properties of these matrices as a function of θ. Whereas we have fixed the mean and variance of the elements of $B_{ij}(\theta)$ to zero and unity, respectively, the skewness η varies with θ: $\eta(\theta) = (1/4)\,\Sigma_{ij}B_{ij}^3 = -3\cos^2\theta\sin\theta$. We note that the skewness never formally enters into any of the formulæ of this work, because we truncate the high temperature expansion of the free energy to lowest order. This results in only the variance entering into our calculations, and thus differences between two letter models are not seen. The next order corrections include the skewness and thus demonstrate differences between interaction matrices.

By examining η versus θ, we see that there are three matrices with zero skewness, $\theta = -\pi/2, 0, \pi/2$; the cases with $\theta = \pm\pi/2$ correspond to the familiar ferromagnetic/antiferromagnetic Ising model $B_{ij} = \pm\delta_{ij}$. $\theta = 0$ corresponds to the matrix $B_{ij} = \{\{1, 0\}, \{0, -1\}\}$, which is similar to a matrix used in some protein folding simulations (Li et al., 1996). The matrices with maximum and minimum skewness, $\theta = \pm\arccos(\sqrt{2/3})$, correspond to the HP model (see Interactions, above), which is intended as a model for amino acids (whose interactions appear to be dominated by whether a given amino acid is hydrophobic or polar). From previous analysis, interaction matrices derived from protein statistics, such as the Miyazawa and Jernigan matrix (MJM), appear to be HPM-like, with simply some "noise" that slightly differentiates different H and P monomer species.

Phase diagram

To find out if a polymer designed with B^d interactions will still fold correctly under B^f interactions, we now compare the target state energy (Eq. 46) with the free energy of the random globule (Eq. 26). They are equal at the folding phase transition temperature T_{fold}, which is given by

$$\frac{\widehat{\delta B}^f \cdot \widehat{\delta B}^f}{T_{fold}^2} + \frac{1}{T_{glass}^2} = 2\frac{\widehat{\delta B}^f \cdot \widehat{\delta B}^d}{T_{fold}T_{des}} \tag{50}$$

The most interesting case is obtained when both δB^f and δB^d are of unitary "length" (unitary variance) (we can always simplify to this, by rescaling (though differently) the temperatures T_{fold}, T_{des}, and T_{glass}). In this case we are left with

$$\frac{1}{T_{fold}^2} + \frac{1}{T_{glass}^2} = \frac{2g}{T_{fold}T_{des}} \tag{51}$$

Thus the original phase diagram (Fig. 5) is modified simply by $T_{des} \rightarrow gT_{des}$, as shown in Fig. 8.

How accurate must potentials be for successful modeling of protein folding?

The results of this section agree with previous replica calculations (Pande et al., 1995c), although we have derived the results in a much simpler way. We emphasize that the error limit for retaining correct folding is independent of the length of the polymer. Previous calculations (Bryngelson, 1994) have made estimates that are based directly upon N

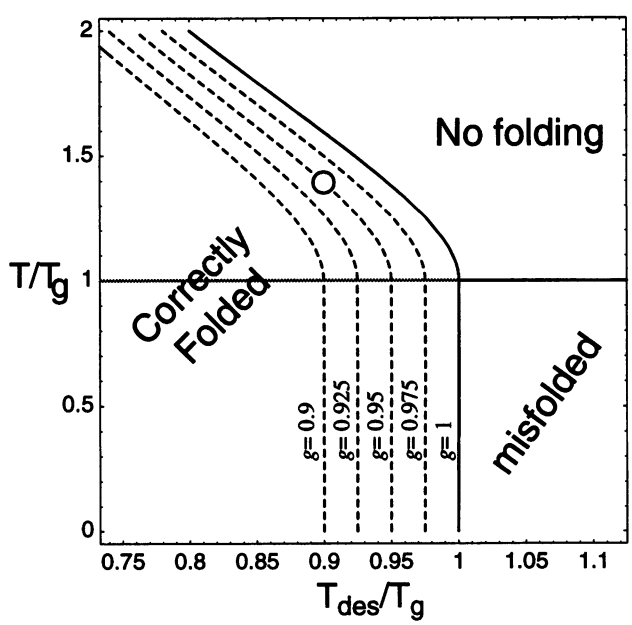

FIGURE 8 Phase diagram for a series of g factors. If the heteropolymer is sufficiently optimized ($T_{des} < gT_{glass}$) and the acting temperature is sufficiently low ($T < T_{fold}$), then even with errors in the interaction matrix, there should be correct folding. However, if the acting temperature is below the glass temperature ($T < T_{glass}$) and there are sufficient errors ($T_{des} > gT_{glass}$), then there will be slow kinetics and the REM states will dominate; this could lead to somewhat protein-like behavior, because there are only $\mathcal{O}(1)$ relevant states, but the ground state is not the designed target conformation. Finally, for sufficiently high temperature ($T > T_{fold}$), there will not be any freezing and many globular conformations will dominate equilibrium.

(i.e., the error must be small compared with $1/\sqrt{N}$). This reflects a fundamental difference between random sequences studied by Bryngelson (1994) and designed sequences. In the work of Bryngelson (1994), neither any kind of design nor the principle of minimal frustration was imposed; accordingly, the ground states were not found to be robust and were typically very unstable with respect to error-based renaturation, especially for long chains. In contrast, our treatment implicitly incorporates design, and thus it is not surprising that we found robust ground states in our analysis of well-designed sequences.

Which of the BWMs is better, and how good is it?

As we know now that the similarity between interaction matrices is given by the value of the angle ϕ between them, or by correlation factor $g = \cos \phi$, we can address the question: Which of the simplified models yields a better approximation for reality? Suppose, for example, that we consider the 20 × 20 MJ matrix (Miyazawa and Jernigan, 1985) to be the "real" one and want to find out if HP or another model with just two types of monomers will yield a reasonable approximation. To address that, we compute the dot product, $g(\theta) = \delta B^{MJ} \cdot \delta B(\theta)$, between the MJ matrix and the 20 × 20 matrix, which is obtained from $\delta B(\theta)$ by using the reduction theorem (see above, Annealed Averaged Free Energy in Terms of High-Temperature Expansion). The best model corresponds to the maximum value of g (or minimum angle ϕ, as $g = \cos \phi$). The results of this computation indicate that, as one would have expected, the best is the HP model $\theta \approx 35°$. However, it is not that good at all: it forms an angle of ~60° with the MJ and thus can hardly yield quantitatively good results for folding predictions.

CONCLUSIONS

What our approach cannot do

The approach we have developed in this article seems to have proved to be very powerful, for it is able to predict various phase diagrams using rather simple calculations. The hallmark of our approach is the use of REM with the relationship in Eq. 11 expressing the free energy of the quenched system above freezing transition in terms of the annealed average. Within the REM framework, one can make the results even more accurate by taking further terms of the high temperature expansion or by using other means to approximate the annealed free energy. Nevertheless, the possibilities of this approach are restricted, in terms of the physics and the applications of the formalism.

In terms of the physics involved, the main restrictions are as follows:

- REM itself is not universally valid; for example, it can be violated in cases of long-range interactions (Pande et al., 1996b).
- Our approach does not allow the examination of the nature of the phases with few conformations in which self-averaging breaks down (i.e., the glassy or folded phases). Thus we cannot calculate parameters such as the number of thermodynamically relevant states, $M = 1/(\sum_\alpha P_\alpha^2)$, where the sum is over conformations α, and P_α is the Boltzmann probability of finding conformation α (note that M is often written in terms of an additional order parameter, $X = 1 - 1/M = 1 - \sum_\alpha P_\alpha^2$). Such calculations are most easily performed by using the replica trick. Replica treatments indicate that

$$X(T) = \begin{cases} 1 & \text{if } T > T_{glass} \\ T/T_{glass} & \text{if } T \le T_{glass} \end{cases} \quad (52)$$

The potentially relevant biological aspects of protein folding that cannot be analyzed with our methods include

- Secondary structure. As discussed in above (Conformations), we assume that secondary structure has already been coarse grained out of the problem and is beyond the analysis discussed here.
- Coulomb interactions. We have been discussing short-range interactions, as proteins appear to be dominated by the role of hydrophobicity (Chan and Dill, 1993). In fact, the inclusion of long-range interactions (Coulomb interactions between charges, for example) leads to the invalidity of REM, and other approaches must be taken (Pande et al., 1996b).
- Validity of the short-range, pairwise interaction Hamiltonian (Eq. 1). In addition to Coulomb interactions, other effects, such as shorter range Lenard-Jones potentials and heteropolymeric three-body interactions, could, in principle, be considered.

Because of the assumptions of short-range interactions, our model leads to freezing to a single conformation (as overlaps are calculated with $\Delta(r)$, short-range interactions lead to overlaps with small scales). In proteins, freezing is only to a set of very similar conformations (Frauenfelder and Wolynes, 1994). To examine this behavior, it is likely that one must go beyond our Hamiltonian (Eq. 1) and perhaps include some of the effects above.

Perspectives

Although the application of REM to the problem of heteropolymer freezing is now quite common, we have found a powerful relationship between the quenched and annealed free energies which, in addition to significantly reducing the complexity of calculations, yields new insights into both the physics involved as well as the obscure and often distrusted replica methods previously used. Within REM validity, we found that the quenched and annealed systems differ only by the entropy of the realizations of disorder. This entropy reduction leads to freezing in the quenched case, but in all other aspects (including change in entropy between two

temperatures), the two systems are thermodynamically identical to REM.

This allows simple calculations of many thermodynamic properties of heteropolymers, including the freezing transition for random sequences and the phase diagram for designed, protein-like heteropolymers. Furthermore, within this formalism, one can easily calculate the effects of designing with one Hamiltonian and folding with another; this models many aspects of the folding problem, including computer simulations of folding, which must use approximated potential functions and the experimental outcome of folding proteins in solutions that have different effects on different amino acids, thereby effectively altering the nature of interactions during folding.

The obvious next step is to develop a non-REM theory of folding and design. However, as the freezing transition goes from a phase in which none of the conformations are similar (on average) directly to a phase in which only one conformation dominates equilibrium (i.e., all of the conformations found in equilibrium overlap completely), these corrections to REM will have no effect on the thermodynamics of freezing. However, to describe the kinetics of this transition, one must consider how the system moves through conformation space, and thus the statistical dependence of states will be of paramount importance (Plotkin et al., 1996; Pande et al., 1997a). It is of special interest, therefore, to examine the effect that deviations from REM may have on the design; for the first steps in this direction, see Plotkin et al. (1997) and Pande et al. (1997b).

Finally, we note that the problem of freezing of heteropolymers is not confined to biopolymers such as proteins. Indeed, the freezing of random, synthetic polymers has attracted great interest, because of potential industrial and biomedical applications. In this light, we have previously suggested "Imprinting," a method of synthesizing heteropolymers with the protein-like properties of renaturability and molecular recognition (Pande et al., 1994b). Just as the applicability of the formalism developed here relies on the speculation that evolution has energetically optimized proteins (Pande et al., 1994a,b), this energetic optimization is a principal aspect of Imprinting. Thus, perhaps in the pursuit of understanding proteins, we have made some progress in the direction of making synthetic protein-like heteropolymers as well.

APPENDIX A: CORRELATOR ANALYSIS OF REM VALIDITY

To investigate the validity of REM statistical independence, the natural quantity to calculate is the energy correlator between two states. Using the Hamiltonian (Eq. 1), we can write for states α and β,

$$\langle E_\alpha E_\beta \rangle_c = \langle E_\alpha E_\beta \rangle - \langle E_\alpha \rangle \langle E_\beta \rangle = \sum_{ij} B_{ij}^2 \mathcal{Q}_{\alpha\beta} + \sum_{ijk} B_{ij} B_{jk} \mathcal{H}_{\alpha\beta} \quad (53)$$

where

$$\mathcal{Q}_{\alpha\beta} \equiv \sum_{I \neq J} \Delta(r_I^\alpha - r_J^\alpha)\Delta(r_I^\beta - r_J^\beta) \quad (54)$$

is the number of contacts conformations α and β have in common and

$$\mathcal{H}_{\alpha\beta} \equiv \sum_{I \neq J \neq K} \Delta(\mathbf{r}_I^\alpha - \mathbf{r}_J^\alpha)\Delta(\mathbf{r}_J^\beta - \mathbf{r}_K^\beta) \quad (55)$$

describes the covariance of the coordination number of typical monomers in conformations α and β; certain types of interactions, such as HP interactions, favor a large variation in the coordination number, and thus low energy states are not necessarily maximally compact (e.g., see Yue et al., 1995).

When $\Sigma_{ijk} B_{ij} B_{jk}$ vanishes, $\mathcal{Q}_{\alpha\beta} = 0$ means statistical independence between states α and β. Physically, this reflects the fact that as these states have no bonds in common, their energies will have no terms in common and should be statistically independent; analogously, for the case $\mathcal{Q}_{\alpha\beta} = \mathcal{Q}_{max} = \mathcal{Q}$, the two conformations are identical and thus are trivially statistically dependent.

When $\Sigma_{ijk} B_{ij} B_{jk}$ does not vanish, there is a residual statistical dependence between states and REM is not valid. The quantity $\Sigma_{ijk} B_{ij} B_{jk}$ can be interpreted as a three-body interaction. This quantity vanishes for IIM, where low energy configurations involve all monomer species equally, but does not vanish for models like HP, in which interactions are determined purely by how many H-H bonds are formed. Indeed, for generalized BWM, we have $\Sigma_{ijk} B_{ij} B_{jk} = 4 \cos^2 \theta$ (see Table 1).

APPENDIX B: DERIVATION OF REM FOR HETEROPOLYMERS

This section is quite technical and can be skipped for those not interested in the details of how to derive REM for the heteropolymer freezing problem. The previous derivation involves replicas and replica-generalized Lifshitz globule theory (Shakhnovich and Gutin, 1989). The following derivation is considerably simpler, but has at its heart the same physics, i.e., one must balance the energetic benefits of having more native (and thus well optimized) bonds and the entropic loss of pinning the polymer.

Our derivation of the validity of REM is based on the following remark. The hallmark property of REM, i.e., the lack of statistical dependence between states, can be considered to be valid to the mean field approximation if there are two and only two free energy minima at $\mathcal{Q}^* = 0$ and $\mathcal{Q}^* = \mathcal{Q}_{max} = \mathcal{Q}$, where $\mathcal{Q}^*$ is the overlap between current conformation and the native one (see the definition in Eq. 54 above). To the mean field approximation, when fluctuations are ignored and thus only minima of free energy are seen, such a two-minima free energy indeed produces REM-type behavior: $\mathcal{Q}^* = 0$ corresponds to no overlap and so to no statistical dependence on the native state, whereas $\mathcal{Q}^* = \mathcal{Q}$ corresponds, of course, to the maximum possible (absolute) statistical dependence.

Thus we have to consider the free energy, $F(\mathcal{Q}^*)$, of the polymer that is fixed to have an overlap $\mathcal{Q}^*$ with its native state conformation NS. With the Hamiltonian (Eq. 1), the free energy of interest can be written in a general

form like

$$F_{\text{seq}}(\mathcal{Q}^\star) = -T \ln\left[\sum_{\text{C}} e^{-\mathcal{H}_C/T}\, \Delta(\mathcal{Q}_{\text{C,NS}} - \mathcal{Q}^\star)\right] \quad (56)$$

where the subscript seq indicates that the free energy is written for a given sequence, summation is performed over conformations C, and $\mathcal{Q}_{\text{C,NS}}$ is the overlap (the number of bonds in common) between conformations C and NS (Eq. 54). The dependence on the sequence enters both through the Hamiltonian and through the ground-state conformation for the given sequence. As above, we resort to high temperature expansion and truncate it at the quadratic terms:

$$F_{\text{seq}}(\mathcal{Q}^\star) \simeq \overline{\langle\mathcal{H}\rangle|_{\mathcal{Q}}} + \frac{1}{2T}[\overline{\langle\mathcal{H}^2\rangle|_{\mathcal{Q}}} - (\overline{\langle\mathcal{H}\rangle|_{\mathcal{Q}}})^2] - T \ln M \quad (57)$$

where $\overline{(\ldots)|_{\mathcal{Q}}}$ denotes averaging over all conformations with the given overlap to the ground state, and M is the total number of conformations (the appearance of the last term is because we average, rather than sum, over conformations).

We now employ the idea of sequence design in a completely new capacity, as a technical tool to elucidate the dependence on the native state conformation. Indeed, we average the free energy over the ensemble of sequences that is designed for a given conformation $\star$, that is, we use the probability distribution

$$P_{\text{seq}} \approx \exp[-E_{\text{NS}}(\text{seq})/T_{\text{des}}] \simeq 1 - \frac{1}{T_{\text{des}}} E_{\text{NS}}(\text{seq}) \quad (58)$$

where in the last transformation we have resorted to high T_{des} expansion (which is justified for the same reason as high temperature expansion above). Of course, in the large T_{des} case, we return to the random sequence limit. Actually, the convenient way to think about random sequences is, in fact, to consider the $T_{\text{des}} = T_{\text{glass}}$ case. Indeed, in this case we still know that the ground state is $\star$, but the statistics of energy levels is unaffected by the design.

When one averages the free energy (Eq. 57) over the probability distribution (Eq. 58), the result appears very simple in the sense that, because of the structure of the Hamiltonian, the energy depends directly on $\mathcal{Q}^\star$; thus it is very easy to implement the condition of the fixed $\mathcal{Q}^\star$, and the result can be written in the form

$$F(\mathcal{Q}^\star) \equiv \sum_{\text{seq}} P_{\text{seq}} F_{\text{seq}}(\mathcal{Q}^\star) = E(\mathcal{Q}^\star) - T\,S(\mathcal{Q}^\star) \quad (59)$$

where the energy $E(\mathcal{Q}^\star)$ is given by

$$E(\mathcal{Q}^\star) \simeq -\frac{\tilde{B}^2}{T_{\text{p}}}\mathcal{Q}^\star - \frac{\tilde{B}^2}{T}(\mathcal{Q}_{\text{m}} - \mathcal{Q}^\star) \quad (60)$$

The first energy term comes from the mean of the density of states $\langle\bar{\mathcal{H}}\rangle$ and describes how the energy is pulled down because of selection; indeed, to this order, the ground-state energy is $E_{\text{NS}} = -\mathcal{Q}_{\text{m}}\tilde{B}^2/T_{\text{p}}$; thus this term says that on average, the energy is given by the energy of the native state times the fraction of native contacts.

The second energy term comes from the width of the density of states, resulting from the $\overline{\langle\mathcal{H}^2\rangle} - \langle\bar{\mathcal{H}}^2\rangle$ terms; the $\mathcal{Q}^\star$ dependence of this width enters from the correlator $\langle\bar{\mathcal{H}}^2\rangle$ (Pande et al., 1996a), as two conformations with a given overlap $\mathcal{Q}^\star$ with NS also have $\mathcal{Q}^\star$ bonds in common, on average. Higher order terms can modify the $\mathcal{Q}^\star$ dependence of the width, depending on the nature of interactions. The $\mathcal{Q}^\star$ dependence of the mean and width of the density of states describes the nature of energy correlations.

As to the entropy (i.e., the number of states there are with the given overlap with the NS), it is technically cumbersome to systematically derive an accurate expression for this quantity (Pande et al., 1997a). However, it is easy to write down the following simple interpolation expression:

$$S(\mathcal{Q}) = \ln \mathcal{M} - \mathcal{Q}\left(s - \frac{d}{2}\ln\frac{\mathcal{Q}}{\mathcal{Q}_{\text{m}}}\right) - \mathcal{Q}\ln(\mathcal{Q}/\mathcal{Q}_{\text{m}}) - (\mathcal{Q}_{\text{m}} - \mathcal{Q})\ln(1 - \mathcal{Q}/\mathcal{Q}_{\text{m}}) \quad (61)$$

The interesting $\mathcal{Q}\ln\mathcal{Q}$ terms come from the loop entropy and the entropy of choosing $\mathcal{Q}$ contacts out of $\mathcal{Q}_{\text{m}}$. It is intriguing that these two terms cancel at $d = 2$ dimensions, where the transition in $\mathcal{Q}$ goes from first ($d > 2$) to second ($d \leq 2$) order (E. I. Shakhnovich, personal communication).

Thus, within the mean field approximation, we find a free energy that has two minima at $\mathcal{Q}^\star = 0$ and $\mathcal{Q}_{\text{m}}$, and REM is therefore applicable for $d > 2$.

We have benefitted greatly from discussions with A. M. Gutin, M. Kardar, D. Nelson, J. N. Onuchic, E. I. Shakhnovich, and P. G. Wolynes. We also thank D. Butts, R. Du, and E. Shender for their critical reading of the manuscript.

The work was supported by NSF (DMR 90-22933) and NEDO of Japan. AYG acknowledges the support of the Kao Fellowship. Computations were performed on Project SCOUT (ARPA contract MDA972-92-J-1032).

REFERENCES

Abkevich, V. I., A. M. Gutin, and E. I. Shakhnovich. 1994. Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry.* 33:10026–10036.

Bryngelson, J. D. 1994. When is a potential accurate enough for structure prediction? Theory and application to a random heteropolymer model of protein folding. *J. Chem. Phys.* 100:6038–6045.

Bryngelson, J. D., and P. G. Wolynes. 1987. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA.* 84: 7524–7528.

Chan, H. S., and K. A. Dill. 1993. The protein folding problem. *Phys. Today.* 46:24–31.

Derrida, B. 1980. Random-energy model: limit of a family of disordered models. *Phys. Rev. Lett.* 45:79–82.

Derrida, B. 1985. A generalization of the random energy model which includes correlations between energies. *J. Phys. Lett.* 46:L401–L407.

Dinner, A., A. Sali, M. Karplus, and E. Shakhnovich. 1994. Phase diagram of a model protein derived by exhaustive enumeration of the conformations. *J. Chem. Phys.* 101:1444–1451.

Franz, S., M. Mezard, and G. Parisi. 1994. On the mean field theory of random heteropolymers. *Int. J. Neural Syst.* 3:195–199.

Frauenfelder, H., and P. G. Wolynes. 1994. Biomolecules: where the physics of complexity and simplicity meet. *Phys. Today.* 47:58–64.

Garel, T., L. Leibler, and H. Orland. 1994. Random hydrophilic-hydrophobic copolymers. *J. Phys. II (France).* 4:2139–2148.

Garel, T., and H. Orland. 1988a. Chemical sequence and spatial structure in simple models of biopolymers. *Europhys. Lett.* 6:597–601.

Garel, T., and H. Orland. 1988b. Mean-field model for protein folding. *Europhys. Lett.* 6:307–310.

Godzik, A., A. Kolinski, and J. Skolnick. 1995. Lattice representations of globular proteins: how good are they? *J. Comput. Chem.* 14:1194–1202.

Goldenfeld, N. 1992. Lectures on Phase Transitions and the Renormalization group. Addison-Wesley, Reading, MA.

Grosberg, A. Yu., and A. R. Khokhlov. 1994. Statistical Physics of Macromolecules. American Institute of Physics, New York.

Gutin, A. M., and E. I. Shakhnovich. 1994. Statistical mechanics of polymers with distance constraints. *J. Chem. Phys.* 100:5290–5293.

Hecht, M. H., J. S. Richardson, and D. C. Richardson. 1990. De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science.* 249:884–888.

Holden, C., editor. 1995. Random samples. *Science.* 269:1821–1822.

Klimov, D. K., and D. Thirumalai. 1996. A criterion that determines the foldability of proteins. *Phys. Rev. Lett.* 76:4070–4073.

Koukiou, F. 1993. Analyticity of the partition function of the random energy model. *J. Phys. A.* 26:L1207–1210.

Li, H., R. Helling, C. Tang, and N. Wingreen. 1996. Why do proteins look like proteins? Emergence of preferred structures in a simple model of protein folding. *Science.* 273:666–669.

Mezard, M., G. Parisi, and M. Virasoro. 1987. Spin Glass Theory and Beyond. World Scientific, Singapore.

Miyazawa, S., and R. Jernigan. 1985. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules.* 18:534–552.

Monod, J. 1971. Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology. Knopf, New York.

Onuchic, J. N., P. G. Wolynes, Z. Luthey-Schulten, and N. D. Socci. 1995. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. USA.* 92:3626–3630.

Panchenko, A., Z. Luthey-Schulten, and P. G. Wolynes. 1995. Foldons, protein structural modules, and exons. *Proc. Natl. Acad. Sci. USA.* 93:2008–2013.

Pande, V. S., A. Yu. Grosberg, C. Joerg, and T. Tanaka. 1996a. Is heteropolymer freezing well described by a random energy model? *Phys. Rev. Lett.* 76:3987–3990.

Pande, V. S., A. Yu. Grosberg, M. Kardar, C. Joerg, and T. Tanaka. 1996b. Freezing transition of compact polyampholytes. *Phys. Rev. Lett.* 77: 3565–3568.

Pande, V. S., A. Yu. Grosberg, and T. Tanaka. 1994a. Nonrandomness in protein sequences: evidence for a physically driven stage of evolution? *Proc. Natl. Acad. Sci. USA.* 91:12972–12975.

Pande, V. S., A. Yu. Grosberg, and T. Tanaka. 1994b. Thermodynamic procedure to synthesize heteropolymers that can renature to recognize a given target molecule. *Proc. Natl. Acad. Sci. USA.* 91:12976–12979.

Pande, V. S., A. Yu. Grosberg, and T. Tanaka. 1995a. Freezing transition of random heteropolymers consisting of an arbitrary set of monomers. *Phys. Rev. E.* 51:3381–3392.

Pande, V. S., A. Yu. Grosberg, and T. Tanaka. 1995b. Phase diagram of heteropolymers with an imprinted conformation. *Macromolecules.* 28: 2218–2227.

Pande, V. S., A. Yu. Grosberg, and T. Tanaka. 1995c. How accurate must potentials be for successful modeling of protein folding? *J. Chem. Phys.* 103:9482–9491.

Pande, V. S., A. Yu. Grosberg, and T. Tanaka. 1997a. On the theory of folding kinetics for short proteins. *Folding Design.* 2:109–114.

Pande, V. S., A. Yu. Grosberg, and T. Tanaka. 1997b. Thermodynamics of the coil to frozen globule transition in heteropolymers. *J. Chem. Phys.* 107: 5118–5124.

Pande, V. S., C. Joerg, A. Yu. Grosberg, and T. Tanaka. 1994c. Enumerations of the Hamiltonian walks on a cubic sublattice. *J. Phys. A.* 27:6231–6236.

Plotkin, S. S., J. Wang, and P. G. Wolynes. 1996. Correlated energy landscape model for finite, random heteropolymers. *Phys. Rev. E.* 53: 6271–6296.

Plotkin, S. S., J. Wang, and P. G. Wolynes. 1997. Statistical mechanics of a correlated energy landscape model for protein folding funnels. *J. Chem. Phys.* 106:2932–2948.

Ptitsyn, O. B., and M. V. Volkenstein. 1986. Protein structures and neutral theory of evolution. *J. Biomol. Struct. Dyn.* 4:137–156.

Ramanathan, S., and E. I. Shakhnovich. 1994. Statistical mechanics of proteins with "evolutionary selected" sequences. *Phys. Rev. E.* 50: 1303–1312.

Sfatos, C., A. Gutin, and E. Shakhnovich. 1993. Phase diagram of random copolymers. *Phys. Rev. E.* 48:465–475.

Shakhnovich, E., and A. Gutin. 1989. Formation of unique structure in polypeptide chains: theoretical investigation with the aid of a replica approach. *Biophys. Chem.* 34:187–199.

Shakhnovich, E., and A. Gutin. 1990. Enumeration of all compact conformations of copolymers with random sequence of links. *J. Chem. Phys.* 93:5967–5971.

Shakhnovich, E. I., and A. M. Gutin. 1993. Engineering of stable and fast-folding sequences of model proteins. *Proc. Natl. Acad. Sci. USA.* 90:7195–7199.

Socci, N. D., and J. N. Onuchic. 1995. Kinetic and thermodynamic analysis of protein-like heteropolymers: Monte Carlo histogram technique. *J. Chem. Phys.* 103:4732–4741.

Volkenstein, M. V. 1994. Physical Approaches to Biological Evolution. Springer Verlag, Berlin and New York.

Yue, K., K. M. Fiebig, P. D. Thomas, H. S. Chan, E. I. Shakhnovich, and K. A. Dill. 1995. A test of lattice protein folding algorithms. *Proc. Natl. Acad. Sci. USA.* 92:325–329.