Universal correlation between energy gap and foldability for the random energy model and lattice proteins

Nicolas E. G. Buchler
Biophysics Research Division, University of Michigan, Ann Arbor, Michigan 48109-1055

Richard A. Goldstein[a)]
Biophysics Research Division and Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1055

(Received 7 May 1999; accepted 14 July 1999)

The random energy model, originally used to analyze the physics of spin glasses, has been employed to explore what makes a protein a good folder versus a bad folder. In earlier work, the ratio of the folding temperature over the glass–transition temperature was related to a statistical measure of protein energy landscapes denoted as the foldability $\mathcal{F}$. It was posited and subsequently established by simulation that good folders had larger foldabilities, on average, than bad folders. An alternative hypothesis, equally verified by protein folding simulations, was that it is the energy gap Δ between the native state and the next highest energy that distinguishes good folders from bad folders. This duality of measures has led to some controversy and confusion with little done to reconcile the two. In this paper, we revisit the random energy model to derive the statistical distributions of the various energy gaps and foldability. The resulting joint distribution allows us to explicitly demonstrate the positive correlation between foldability and energy gap. In addition, we compare the results of this analytical theory with a variety of lattice models. Our simulations indicate that *both* the individual distributions and the joint distribution of foldability and energy gap agree qualitatively well with the random energy model. It is argued that the universal distribution of and the positive correlation between foldability and energy gap, both in lattice proteins and the random energy model, is simply a stochastic consequence of the ''thermodynamic hypothesis.'' © *1999 American Institute of Physics.* [S0021-9606(99)50538-2]

INTRODUCTION

True to their universal aspiration as general models of disordered systems, the success of spin-glass models has not been limited to physics. Since the 1980s, there has been a plethora of applications of such models to biological topics such as neural networks,[1,2] prebiotic evolution,[3,4] evolutionary dynamics,[5–7] immunology,[8,9] and protein structure recognition.[10,11] In particular, spin-glass analogies have been employed to explore protein folding in the context of ''heteropolymer freezing.'' [12–17] The impetus was to provide a model explaining a wide range of hallmark features characteristic of protein folding: (1) all-or-none transitions between the ''unfolded'' and the ''folded'' states, (2) the existence of measurable, discrete intermediates and multiexponential kinetics on folding and/or experimental time scales, and (3) observations of ''misfolds,'' protein drift, and irreversible denaturation.

A central parameter, the equilibrium glass transition temperature T_g (also called the ''heteropolymer freezing'' temperature) is defined as the temperature where the liquid-like protein chain entropy drops below zero (in the thermodynamic limit) and the chain becomes solid-like and ''frozen'' in any one of its low-energy, metastable states.[12–17] Using the random energy model (REM) where correlations between the energies of the various conformations are neglected, Bryngelson and Wolynes demonstrated that the equilibrium glass temperature T_g partitions the kinetic behavior into two regimes.[13] For $T>T_g$, the distribution of escape rates from low-energy metastable states is log-normal and fast rates are dominant. However, for $T<T_g$, the kinetic distribution of escape rates becomes flat and broad so that slow escape rates are as equally likely as fast rates. These prominent slow transition rates between minima lead to multiexponential time dependencies on biologically relevant time scales, lack of self-averaging for many properties of the system, and folding kinetics that are sensitive to the details of the protein sequence and its initial conditions. Using the REM, it can be shown that the equilibrium glass temperature is equal to

$$T_g = \sqrt{\frac{\sigma^2}{2S_0}}, \tag{1}$$

where σ describes the variance or ''roughness'' of the REM energy distribution and S_0 is the conformational entropy of the system. (In this equation and throughout the paper, k_B has been set to one.) A conclusion based on this relation is that rougher energy landscapes lead to higher glass transition temperatures. Thus, at physiological temperatures, the aforementioned hallmark features (2) and (3) would be indicative

a)Author to whom correspondence should be addressed; electronic mail: richardg@umich.edu

of proteins with significantly rougher energy landscapes compared to ones which fold consistently and exhibit single-exponential kinetics.[18]

Given this simple understanding of the glass transition temperature and its role in the lack of self-averaging and slow protein folding rates, what can be said about native state uniqueness and stability in folding proteins? The folded state must thermodynamically dominate the ensemble of other kinetically accessible structures under equilibrium conditions. A measure of the relative stability of the native state is the folding temperature T_f. Based on these ideas of T_f and T_g, Wolynes and co-workers postulated that optimal folding landscapes would seek to maximize T_f and minimize T_g or, equivalently, increase the ratio T_f/T_g.[19,20] Using the REM, it was shown that this ratio is equal to

$$\frac{T_f}{T_g}=\sqrt{\frac{\mathcal{F}^2}{2S_0}}+\sqrt{\frac{\mathcal{F}^2}{2S_0}-1},$$

where

$$\mathcal{F}=\frac{\overline{E}-E_{\mathrm{ns}}}{\sigma}, \tag{2}$$

where $\overline{E}$ is the average energy of the protein chain in all conformations, E_{ns} is the energy of the native structure, and σ describes the variance or ''roughness'' of this REM energy landscape. Clearly, T_f/T_g is a monotonically increasing function of $\mathcal{F}$, which is termed the ''foldability.'' It was shown both with molecular dynamic and Monte Carlo kinetic simulations that faster folders had higher average foldabilities and larger T_f/T_g.[19–22] This connection between foldability and faster folding is intuitive when one notes that $\mathcal{F}$ is increased by: (1) stabilizing native-like interactions which deepens the native state E_{ns} with respect to the bulk of conformational energies $\overline{E}$, and (2) destabilizing non-native interactions and misfolds, which alleviates the roughness σ of the energy landscape. Although REM theory, on which foldability is based, ignores correlations in the energy landscape, the stabilization of these native-state contacts creates a natural ''folding funnel,'' which has been advocated as a requirement for fast folding by a number of investigators.[23,24] The allure of foldability as a simple statistical measure of an energy landscape and as an indicator of folding ability has spawned a flurry of research into areas of protein structure designability[25,26] and evolutionary dynamics.[27,28]

Another thermodynamic quantity related to folding kinetics, referred to as the energy gap Δ, was first introduced by Shakhnovich and co-workers.[29,30] Originally, this energy gap was defined as the difference in energy between the native state E_{ns} and that of the next highest energy E_1. (Throughout this paper, we will define energies by their rank; that is, E_{ns} is concomitant with energy of rank 0, the next highest energy with rank 1, etc.) Upon running Monte Carlo kinetic simulations of lattice proteins and analyzing their energy landscapes, Shakhnovich and co-workers concluded that the relevant statistical feature for fast folding was a large energy gap Δ.[31,32] However, due to criticism that Δ was perhaps too local a measure, the definition has speciated with time. Currently there are three measures of energy gap: Δ_{10}, Δ_g, and $\Delta_{10}^{\mathrm{dis}}$ claimed in the literature to be correlated to fast folding

$$\Delta_{10}=E_1-E_{\mathrm{ns}};\quad \Delta_g=E_g-E_{\mathrm{ns}};\quad \Delta_{10}^{\mathrm{dis}}=E_1^{\mathrm{dis}}-E_{\mathrm{ns}}. \tag{3}$$

Δ_{10} is simply the original energy gap. Δ_g measures the depth of the native state with respect to the glass transition energy E_g, defined by the relation $S(E_g)=0$.[33] The alternative, $\Delta_{10}^{\mathrm{dis}}$, is similar to the original energy gap except that it has been revised to neglect energetic correlations between similar structures.[34] Namely, E_1^{dis} is meant to be the next highest energy of a *dissimilar* structure, defined as having an amount of amino acid pair contacts similar to the native state that is less than or equal to the what is expected between random structures. This construction is due to the fact that the next highest energy is typically a structure sharing $\sim 80\%-95\%$ similarity to the native state and is most likely within its energy basin; it is not a competitive ''misfold.'' Without explicitly having to calculate correlations between structures, one can reword E_1^{dis} and $\Delta_{10}^{\mathrm{dis}}$ in the context of the REM with the following argument: by neglecting those structures with strong similarity to the native state, one is potentially rescaling the energy gap as the difference between a structure of energy rank r, E_r, and the energy of the native state E_{ns}. The actual values of r, with no loss of generality, can be left undetermined until explicit comparison of the REM to lattice proteins. Hence, both Δ_{10} and $\Delta_{10}^{\mathrm{dis}}$ can be simultaneously calculated through a Δ_{r0} framework, where $\Delta_{r0}=E_r-E_{\mathrm{ns}}$. We should point out that to rigorously calculate the joint distribution of $\mathcal{F}$ and $\Delta_{10}^{\mathrm{dis}}$, one explicitly needs to take correlations between structures and energies into account using an analytical approach such as the generalized random energy model (GREM).[35,36] Since the introduction of these various Δ and $\mathcal{F}$ into the literature, there have been peripheral data and heuristic arguments indicating that energy gap and foldability should be related. Unfortunately, since there has never been a unifying demonstration of the relationship between Δ and $\mathcal{F}$, the presence of two loosely associated concepts has created some confusion and controversy in the literature.

For traditional and analytical reasons, in this paper we revisit the REM to derive the various distributions of $\mathcal{F}$, Δ_g, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$. Foldability and Δ_g are shown to be equivalent measures, directly related to one another by the underlying conformational entropy. Using the REM model, we demonstrate that $\mathcal{F}$, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$ are inherently positively correlated. We subsequently compare these REM foldability and energy gap results to a variety of lattice protein simulations. Surprisingly, our simulations show that, despite the cavalier application of the REM to proteins, both the individual distributions and the joint distribution of $\mathcal{F}$, Δ_g, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$ agree qualitatively well with the predictions based on the REM theory. It should be mentioned that there are other measures of good folders, aside from foldability and energy gap, in the literature. One prominent candidate is the $\sigma_\theta=(T_\theta-T_f)/T_\theta$ criterion proposed by Thirumulai and co-workers.[37–39] Unfortunately, as T_θ represents the transition temperature from noncompact to compact conformations, comparison of σ_θ to $\mathcal{F}$ and Δ in all its forms is beyond

the scope of our REM analysis, which focuses only on maximally compact conformations. One could in principle extend our REM model to include noncompact conformations and a hydrophobic driving force, similar to previous work by Bryngelson and Wolynes and Chiu and Goldstein.[17,40] However, recent work by Shakhnovich and co-workers suggests that Thirumalai's calculation of T_θ, by equating it to the maximum in the specific heat C_v, may be erroneous, particularly for protein sequences with strong, hydrophobic driving forces.[41] Thus, the absence of an uncontroversial definition of σ_θ precludes any comparison of $\mathcal{F}$ or Δ to σ_θ for the moment.

REM AND FOLDABILITY

The REM was originally introduced by Derrida to describe the energy landscape of a generic, disordered system.[42,43] Powerful in its analytical simplicity, the REM is based on the key assumption of statistical independence of energy states; namely, the energy of one state is uncorrelated with the energy of another state. We begin where others have before; namely by invoking the REM as a description of the underlying protein conformational energy landscape.[12,13,16] The energies between any two compact conformations are assumed to be independently drawn from a single distribution. In the limit of large proteins, the energy distribution of these compact states has been shown to be Gaussian in form.[44] Consequently, throughout this paper, we will describe the density of states of a REM heteropolymer sequence by $\Omega(E)=n\rho_{\mathrm{REM}}(E)$, where n is the number of compact protein structures and $\rho_{\mathrm{REM}}(E)$ is a normalized REM Gaussian distribution

$$\rho_{\mathrm{REM}}(E)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-(E-\bar{E})^2/2\sigma^2}, \tag{4}$$

where $\bar{E}$ is the average energy of the compact states and σ is the REM roughness and width of the energy density distribution. We start by calculating the foldability of the native state $\mathcal{F}$ as a function of the number of compact protein structures n: similar calculations have been done before with statistical protein models.[13,31,45] The condition of native state uniqueness and thermodynamic dominance imposes the condition that the native state energy E_{ns} be nondegenerate and have the lowest value among all other $n-1$ energies ($E_{\mathrm{ns}}=E_0$), a condition commonly referred to as the ''thermodynamic hypothesis.''[46] (The validity of the thermodynamic hypothesis has been shown to be increasingly likely when proteins have undergone significant amounts of selective evolution to a stable target state.[47]) For the REM, it is analytically impossible for any energy to be exactly degenerate. However, this assumption cannot be taken for granted in the discrete REM, which models the energy landscape of random, block copolymers.[48] Given these constraints, one can formally describe the native state energy distribution in the REM by

$$\rho(E_{\mathrm{ns}}|n)=\rho_{\mathrm{REM}}(E_{\mathrm{ns}})\mathcal{P}(E_{\mathrm{ns}}<n-1). \tag{5}$$

Based on the independence of energies and Eq. (4), these various probability densities are straightforward to calculate

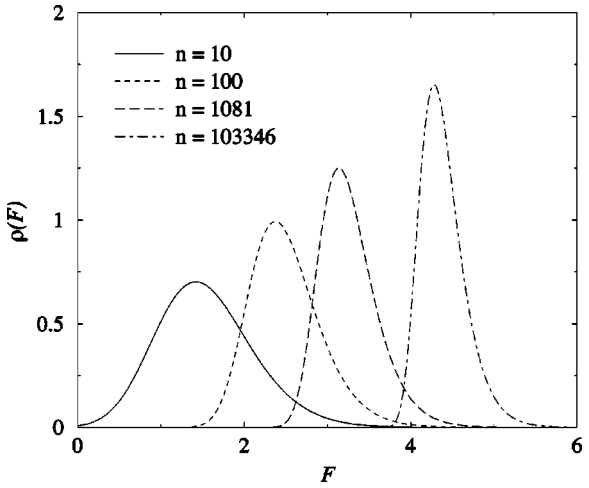

FIG. 1. Plot of the foldability distribution $\rho(\mathcal{F})$ for different numbers of compact states ($n=10$, 100, 1081, 103 346), calculated using the random energy model.

$$\rho_{\mathrm{REM}}(E_{\mathrm{ns}})=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-(E_{\mathrm{ns}}-\bar{E})^2/2\sigma^2}, \tag{6}$$

$$\mathcal{P}(E_{\mathrm{ns}}<n-1)=\left[\int_{E_{\mathrm{ns}}}^{\infty}\rho_{\mathrm{REM}}(E)\,\delta E\right]^{(n-1)}. \tag{7}$$

We combine these probabilities and normalize $\int_{-\infty}^{+\infty}\rho(E_{\mathrm{ns}}|n)\,\delta E_{\mathrm{ns}}=1$ via integration by parts. Notice that the normalization constant reflects the combinatorial nature of how many different ways one can have a lowest energy state given n ''picks.'' The resulting density of native state energies is

$$\rho(E_{\mathrm{ns}}|n)=n\rho_{\mathrm{REM}}(E_{\mathrm{ns}})\left[\int_{E_{\mathrm{ns}}}^{\infty}\rho_{\mathrm{REM}}(E)\,\delta E\right]^{(n-1)}, \tag{8}$$

$$\rho(E_{\mathrm{ns}}|n)=\frac{n}{\sigma\sqrt{2\pi}}\,e^{-(E_{\mathrm{ns}}-\bar{E})^2/2\sigma^2}\times\left[\frac{1}{2}\left(1-\mathrm{Erf}\left(\frac{E_{\mathrm{ns}}-\bar{E}}{\sigma\sqrt{2}}\right)\right)\right]^{(n-1)}. \tag{9}$$

This distribution of native state energies, satisfying the thermodynamic hypothesis, is commonly known as an extreme value distribution. The conversion from E_{ns} to the foldability $\mathcal{F}$, a dimensionless quantity, is simple enough as $\mathcal{F}=(\bar{E}-E_{\mathrm{ns}})/\sigma$

$$\rho(\mathcal{F})=\frac{n}{\sqrt{2\pi}}\,e^{-\mathcal{F}^2/2}\left[\frac{1}{2}\left(1+\mathrm{Erf}\left(\frac{\mathcal{F}}{\sqrt{2}}\right)\right)\right]^{(n-1)}. \tag{10}$$

This distribution of native state foldabilities, for different values of n, is shown in Fig. 1. Admittedly, low values of n are rather unrealistic even for very small proteins. However, because the largest changes occur for small values of n, we have taken the liberty to include them. In addition, we have plotted $n=1081$ and $n=10\,3346$, the number of unique structures for commonly used 5×5 two-dimensional (2D) and $3\times3\times3$ 3D compact lattice proteins. It is noticed that as the number of compact structures swells, several things hap-

TABLE I. Statistical parameters describing the distributions of foldabilities ($\mathcal{F}$), the energy gap between the ground state and glass transition energy (Δ_g), and the energy gap between the ground and first excited state (Δ_{10}), calculated using Eq. (10) for $\mathcal{F}$, Eq. (13) for Δ_g, and Eq. (19) with $r=1$, integrated over $\mathcal{F}$, for Δ_{10}. Except for the mean, Δ_g statistics were identical to $\mathcal{F}$. The sixth column is the percentage of REM heteropolymer sequences that satisfy $\Delta_g>0$ or, equivalently, $T_f/T_g>1$.

REM	$\overline{\mathcal{F}}$	$\sigma_{\mathcal{F}}$	Skew$_{\mathcal{F}}$	$\overline{\Delta_g}$	$\int\rho(\Delta_g>0)$	$\overline{\Delta_{10}}$
$n=10$	1.5388	0.5867	0.4116	-0.6072	14.84%	0.5374
$n=100$	2.5076	0.4294	0.6592	-0.5273	11.34%	0.3594
$n=1081$	3.2637	0.3497	0.7703	-0.4741	9.55%	0.2856
$n=103\,346$	4.3915	0.2715	0.8961	-0.4139	7.67%	0.2180

pen to the foldability distribution: (1) the mean foldability $\overline{\mathcal{F}}$ increases, (2) the variance $\sigma_{\mathcal{F}}$ decreases, and (3) a positive skew emerges. The statistical measures for these REM foldability distribution are shown in Table I.

REM, ENERGY GAP, AND FOLDABILITY

The alternative measure of thermodynamic stability and fast folding is the energy gap, Δ. Due to the heterogeneity in its definition, we must break our calculation of energy gap into two parts: namely, that of Δ_g and those of $\Delta_{r0}\rightarrow\Delta_{10}$ and $\Delta_{10}^{\mathrm{dis}}$. Δ_g is the energy gap between the native state energy E_{ns} and E_g, defined as the transition where the finite energy spectrum goes from discretely and thinly populated ($\mathcal{O}(N)$) to continuously and densely populated ($\mathcal{O}(e^N)$), where N is the amino-acid length of the protein. Thus, similar to foldability, one only needs to know the distribution of E_{ns} because for the REM, E_g is determined by $\overline{E}$, σ, and n. On the other hand, Δ_{r0} will involve calculating the *joint* distribution of two independent energies, E_{ns} and E_r. For simplicity's sake, we begin our calculations by relating Δ_g to $\mathcal{F}$

$$\mathcal{F}=\frac{\overline{E}-E_{\mathrm{ns}}}{\sigma}=\frac{\overline{E}-E_g+\Delta_g}{\sigma}. \tag{11}$$

Since the relation $S(E_g)=\ln\Omega(E_g)=0$ gives $E_g=\overline{E}-\sigma\sqrt{2\ln n}$ (after dropping terms involving $\mathcal{O}(\sqrt{\ln\sigma})$), we can rewrite Eq. (11) as

$$\mathcal{F}=\frac{\sigma\sqrt{2\ln n}+\Delta_g}{\sigma}=\sqrt{2\ln n}+\Delta_g. \tag{12}$$

In order to facilitate comparison with the foldability, we have redefined the energy gap Δ_g as a dimensionless quantity by absorbing σ in the denominator, as will also be done for Δ_{r0}. It has been shown previously that $\overline{E}$ and σ of a random heteropolymer depend on the relative amount of amino acids in the sequence and the details of interaction between these different amino acids.[15,44] Thus, by absorbing σ in the denominator, the dimensionless nature of $\mathcal{F}$, Δ_{r0}, and Δ_g leads to universal and composition-independent relationships between these measures. This said, we need to go no further to demonstrate the inherent correlation between $\mathcal{F}$ and Δ_g as identical measures. The REM distribution of $\rho(\Delta_g)$ is simply given by substituting Eq. (12) into Eq. (10)

$$\rho(\Delta_g)=\frac{n}{\sqrt{2\pi}}e^{-(\sqrt{2\ln n}+\Delta_g)^2/2}\times\left[\frac{1}{2}\left(1+\mathrm{Erf}\left(\frac{\sqrt{2\ln n}+\Delta_g}{\sqrt{2}}\right)\right)\right]^{(n-1)}. \tag{13}$$

As demonstrated in Fig. 2, $\rho(\Delta_g)$ is identical to $\rho(\mathcal{F})$ except that $\overline{\Delta_g}$ is shifted down by an amount equal to $\sqrt{2\ln n}$. The statistics for these REM Δ_g distributions are given in Table I. Interestingly, all curves intersect at $\Delta_g=0$ and a good portion of REM native state energies are actually above the glass–transition energy. What is the physical significance of these negative Δ_g? It is clear that $\Delta_g=0$ is equivalent to $\mathcal{F}=\sqrt{2\ln n}$. Since $S_0=\ln n$, $\Delta_g=0$ is equivalent to $\mathcal{F}^2=2S_0$, which defines the transition where $T_f/T_g=1$. In other words, in order to guarantee that T_f exists and has meaning, it must be larger than T_g or, equivalently, E_{ns} must be lower than E_g. The physical relationship of $\Delta_g>0$ and $E_{\mathrm{ns}}<E_g$ has been discussed before in the protein design literature, although not in relation to foldability.[33,49] Figure 2 emphasizes that a large percentage of REM heteropolymer sequences are not even able to satisfy this weakest foldability criterion. The exact percentage of REM sequences expected to satisfy $T_f/T_g>1$ is shown in Table I; clearly it is a decreasing function of n. Thus, although foldability increases

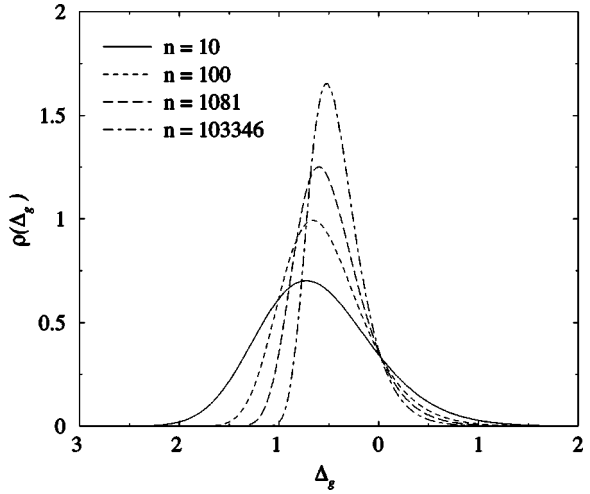

FIG. 2. Plot of the Δ_g distribution $\rho(\Delta_g)$ for different numbers of compact states ($n=10$, 100, 1081, 103 346), calculated using the random energy model.

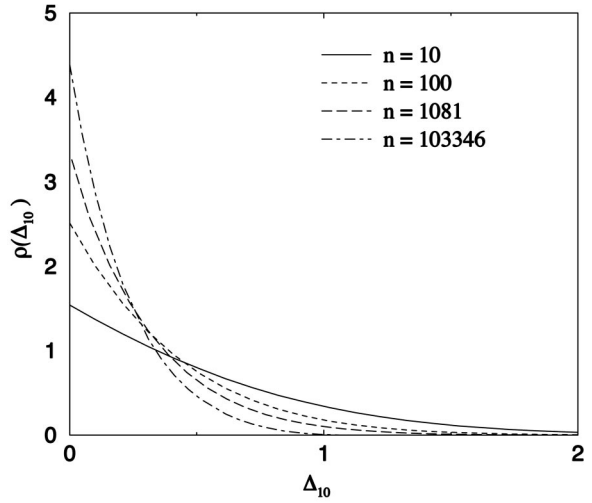

FIG. 3. Plot of the Δ_{10} distribution $\rho(\Delta_{10})$ for different numbers of compact states ($n=10$, 100, 1081, 103 346), obtained by numerically integrating $\rho(\mathcal{F},\Delta_{10})$ over $\mathcal{F}$.

with n, the reality is that a smaller fraction of REM heteropolymer sequences are actually ''foldable'' for larger proteins.

Now that a strict correlation between $\mathcal{F}$ and Δ_g has been demonstrated for REM heteropolymer proteins, we turn to Δ_{r0}. In terms of the REM, we begin by looking at $E_{\rm ns}=E_0$ and E_r, where E_r is the next highest energy of rank r. Formally, this joint distribution is given by

$$\rho(E_{\rm ns},E_r|n,r)=\rho(E_{\rm ns}|n)\rho(E_r|E_{\rm ns},n,r). \tag{14}$$

The first part of the distribution has already been determined by Eqs. (8)–(9). The second part is similar in derivation to the first, except that E_r is dependent on $E_{\rm ns}$, n, and r

$$\rho(E_r|E_{\rm ns},n,r)=\rho_{\rm REM}(E_r)\mathcal{P}(r-1<E_r<n-r-1)\times\Theta(E_{\rm ns}<E_r). \tag{15}$$

$\Theta(E_{\rm ns}<E_r)$ is the Heaviside function, which ensures proper integration via the constraint $E_{\rm ns}<E_r$. The condition that $r-1<E_r<n-r-1$ remaining energies is given by

$$\mathcal{P}(r-1<E_r<n-r-1)=\left[\int_{E_{\rm ns}}^{E_r}\rho_{\rm REM}(E)\,\delta E\right]^{(r-1)}\left[\int_{E_r}^{\infty}\rho_{\rm REM}(E)\,\delta E\right]^{(n-r-1)}. \tag{16}$$

Normalization of $\int_{-\infty}^{+\infty}\rho(E_r|E_{\rm ns},n,r)\,\delta E_r=1$ reduces Eq. (15) to

$$\rho(E_r|E_{\rm ns},n,r)=\frac{(n-1)!}{(n-r-1)!(r-1)!}\rho_{\rm REM}(E_r)\Theta(E_{\rm ns}<E_r)\times\frac{[\int_{E_{\rm ns}}^{E_r}\rho_{\rm REM}(E)\,\delta E]^{(r-1)}[\int_{E_r}^{\infty}\rho_{\rm REM}(E)\,\delta E]^{(n-r-1)}}{[\int_{E_{\rm ns}}^{\infty}\rho_{\rm REM}(E)\,\delta E]^{(n-1)}}. \tag{17}$$

Thus, combining all the appropriate parts and cancelling terms, Eq. (14) describing the joint probability is

$$\rho(E_{\rm ns},E_r|n,r)=\frac{n!}{(n-r-1)!(r-1)!}\rho_{\rm REM}(E_{\rm ns})\rho_{\rm REM}(E_r)\Theta(E_{\rm ns}<E_r)\times\left[\int_{E_{\rm ns}}^{E_r}\rho_{\rm REM}(E)\,\delta E\right]^{(r-1)}\left[\int_{E_r}^{\infty}\rho_{\rm REM}(E)\,\delta E\right]^{(n-r-1)}. \tag{18}$$

Finally, substituting $\mathcal{F}=(\overline{E}-E_{\rm ns})/\sigma$ and $\Delta_{r0}=(E_r-E_{\rm ns})/\sigma$ into the equation

$$\rho(\mathcal{F},\Delta_{r0}|r)=\frac{n!}{(n-r-1)!(r-1)!}\frac{e^{-\mathcal{F}^2/2}e^{-(\Delta_{r0}-\mathcal{F})^2/2}}{2\pi}\Theta(\Delta_{r0}>0)\times\left[\frac{1}{2}\left(\mathrm{Erf}\left(\frac{\Delta_{r0}-\mathcal{F}}{\sqrt{2}}\right)+\mathrm{Erf}\left(\frac{\mathcal{F}}{\sqrt{2}}\right)\right)\right]^{(r-1)}\times\left[\frac{1}{2}\left(1-\mathrm{Erf}\left(\frac{\Delta_{r0}-\mathcal{F}}{\sqrt{2}}\right)\right)\right]^{(n-r-1)}. \tag{19}$$

We begin by analyzing the distribution of the original energy gap $\rho(\Delta_{10})$ obtained by numerically integrating Eq. (19) for $r=1$ over $\mathcal{F}$. As shown in Fig. 3, $\rho(\Delta_{10})$ is a monotonically decreasing function of Δ_{10} and the allowable values of Δ_{10} shrink rapidly for increasing n. The statistical details of $\rho(\Delta_{10})$ are found in Table I. The corresponding REM joint distribution $\rho(\mathcal{F},\Delta_{10})$ is shown graphically for $n=1081$ in Fig. 4 and $n=103\,346$ in Fig. 5. For values larger than $n=2$, the general form of this joint distribution, aside from statistical measures, is similar across the entire range of n. The striking conclusion is that both $\mathcal{F}$ and Δ_{10} are positively correlated.

To calculate the appropriate joint distribution of $\rho(\mathcal{F},\Delta_{10}^{\rm dis})$, we first need to specify the distribution of rank $\mathcal{N}(r)$ for the lowest energy of dissimilar structures; that is, what is the rank r of the first, lowest energy of a structure *dissimilar* to the native state? Admittedly, this distribution $\mathcal{N}(r)$ is lattice model specific and, for comparative purposes, needs to be explicitly calculated ahead of time. Because we are interested in comparing REM $\rho(\mathcal{F},\Delta_{10}^{\rm dis})$ to lattice models, we determined that the normalized distribution of $\mathcal{N}(r)$ for 5×5 2D and $3\times3\times3$ 3D lattice proteins could be well represented by

$$\mathcal{N}^{55}(r)=\frac{e^{-0.1605r}}{5.7439}, \tag{20}$$

$$\mathcal{N}^{333}(r)=\frac{e^{-0.0441r}}{22.1794}. \tag{21}$$

Thus, given an appropriate $\mathcal{N}^{\star}(r)$, the corresponding REM joint distribution $\rho(\mathcal{F},\Delta_{10}^{\rm dis})$ is obtained by

$$\rho(\mathcal{F},\Delta_{10}^{\rm dis})=\sum_{r=1}^{n}\mathcal{N}^{\star}(r)\rho(\mathcal{F},\Delta_{r0}|r). \tag{22}$$

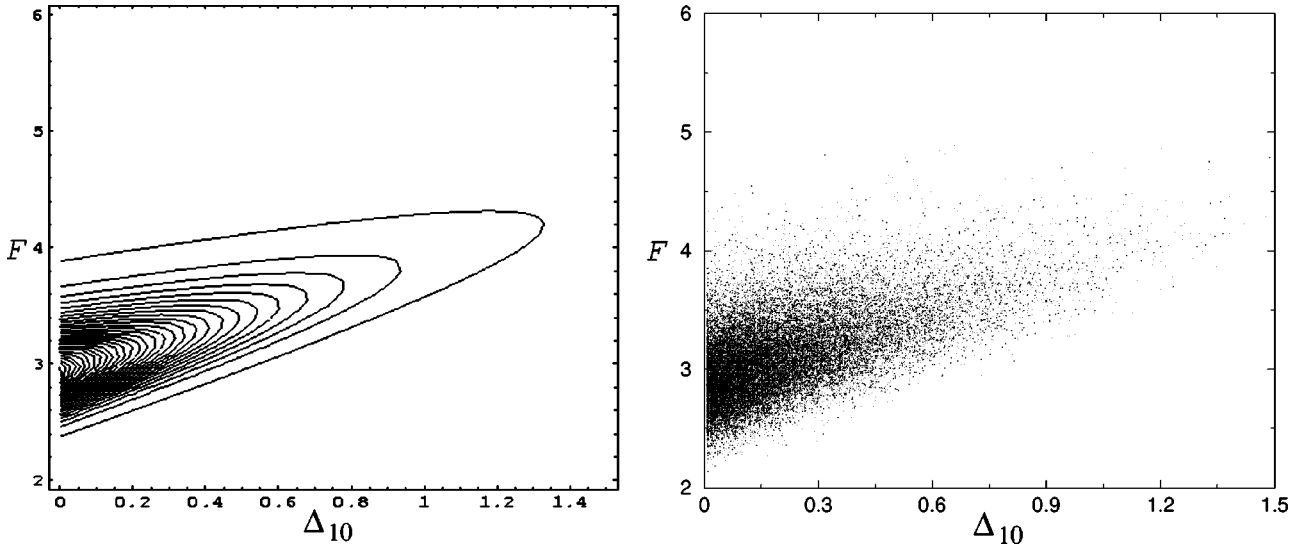

FIG. 4. On the left is a contour plot of the REM joint distribution $\rho(\mathcal{F},\Delta_{10})$ for $n=1081$, which demonstrates the the strong, statistical correlation between foldability and the original energy gap. On the right is a scatter plot of $\mathcal{F}$ and Δ_{10} for the corresponding 5×5 lattice protein simulation.

Similar to Δ_{10}, we start by analyzing the distribution $\rho(\Delta_{10}^{\mathrm{dis}})$, obtained by numerically integrating Eq. (22) over $\mathcal{F}$. As shown in Fig. 6, instead of being a monotonically decreasing function of $\Delta_{10}^{\mathrm{dis}}$ similar to Δ_{10}, $\rho(\Delta_{10}^{\mathrm{dis}})$ is a cropped Gaussian, peaking near ''low–medium'' values of $\Delta_{10}^{\mathrm{dis}}$. The cause of this decreased density of small $\Delta_{10}^{\mathrm{dis}}$ are the $r-1$ energies between E_r and E_{ns} and the large number of *dissimilar* lowest energy structures which have $r\geqslant2$. The REM joint distribution $\rho(\mathcal{F},\Delta_{10}^{\mathrm{dis}})$ is shown graphically for $n=1081$ and $\mathcal{N}^{55}(r)$ in Fig. 7 and $n=10\,3346$ and $\mathcal{N}^{333}(r)$ in Fig. 8. Again, notice the positive correlation between both $\mathcal{F}$ and $\Delta_{10}^{\mathrm{dis}}$. Intuitively, this statistical correlation between energy gap and foldability makes sense as larger values of foldability, where E_{ns} is drastically low, probably leaving more room for larger possible values of all Δ_{r0}. The cause of this positive correlation between $\mathcal{F}$ and Δ_{r0} is independent of the shape of the underlying energy distribution and, hence, universal across all Δ_{10} and $\Delta_{10}^{\mathrm{dis}}$. However, the particular form of the correlation and the joint distributions $\rho(\mathcal{F},\Delta_{10})$ and $\rho(\mathcal{F},\Delta_{10}^{\mathrm{dis}})$ does depend on the details of the underlying energy distribution and $\mathcal{N}(r)$.

Now that we demonstrated the connections between $\mathcal{F}$, Δ_g, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$ for REM heteropolymer sequences, the natural question is whether these REM results have any bearing on real proteins. Namely, the REM was based on the assumption that the energies between structures of a random heteropolymer sequence are uncorrelated. In reality, particularly with the lattice models commonly used by researchers, this independence of energies is false; there are necessarily energetic correlations between structures as they all share, to various degrees, common energetic contacts. Far from being a technical nuisance to REM applicability to proteins, these energetic correlations between conformations are actually

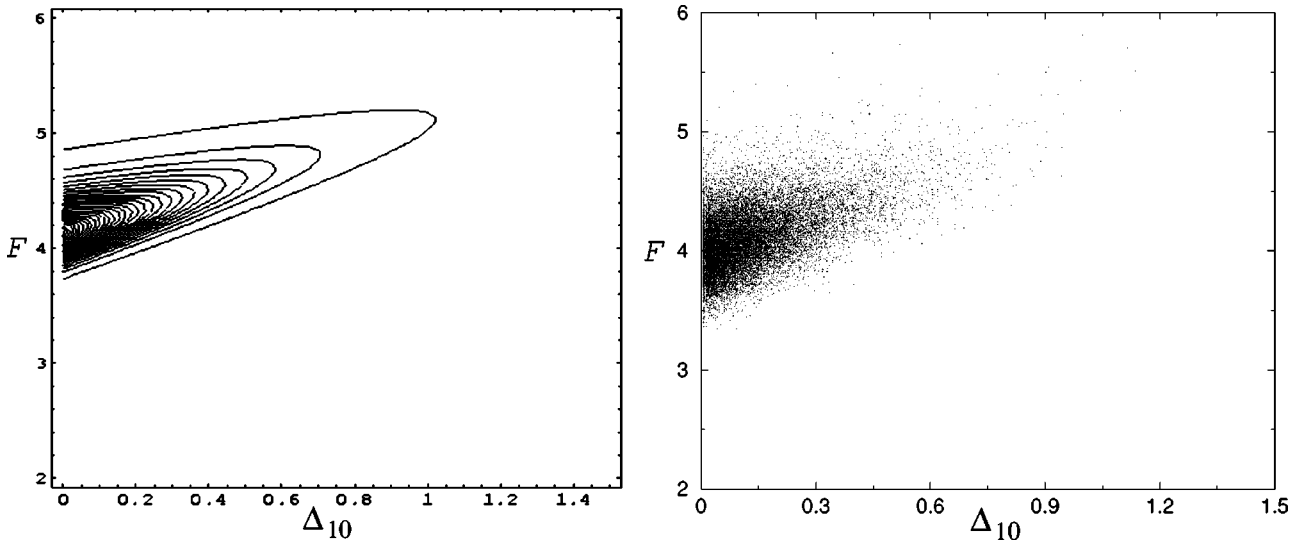

FIG. 5. On the left is a contour REM plot of the joint distribution $\rho(\mathcal{F},\Delta_{10})$ for $n=103\,346$. On the right is a scatter plot of $\mathcal{F}$ and Δ_{10} for the corresponding $3\times3\times3$ lattice protein simulation.

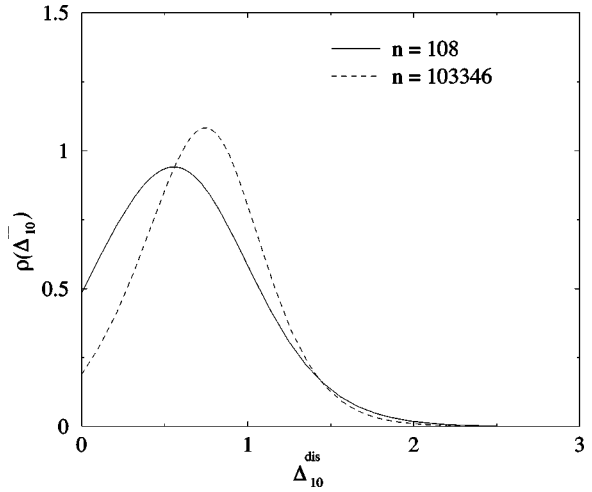

FIG. 6. Plot of the Δ_{10}^{dis} distribution $\rho(\Delta_{10}^{dis})$ for $n=1081$, 103 346, obtained by numerically integrating $\rho(\mathcal{F},\Delta_{10}^{dis})$ over $\mathcal{F}$.

important for the emergence of folding-funnel landscapes of good protein folders[50,51] and answering why some structures are more common or ''designable'' than others.[25,26,52,53] The applicability of the REM to lattice proteins and heteropolymer freezing has been explored by Pande and co-workers.[54] They demonstrated that for large, compact lattice proteins the energetic correlations between structures, although never zero, were generally small enough to ensure that the REM is a good approximate model. In addition, work by Wolynes and co-workers established that previous REM thermodynamic quantities were practically unchanged when redone using the GREM, which explicitly takes correlations into account.[55] However, there are substantial differences between 2D and 3D proteins in terms of replica-symmetry breaking, which directly reflects on REM validity.[14,15,49] Dimension is also of paramount importance for entropy calculations of chain loops in 2D and 3D proteins.[55,50,49] Consequently, given this history, we were interested as to how the distributions of $\mathcal{F}$, Δ_g, Δ_{10}, and Δ_{10}^{dis} for compact lattice proteins in 2D and 3D would compare to those of REM heteropolymer sequences.

LATTICE PROTEINS, ENERGY GAP, AND FOLDABILITY

Lattice proteins are coarse-grained versions of proteins, where the level of detail focuses on amino acids as entities occupying lattice points and protein conformations as self-avoiding walks on these regular lattices. Clearly, this ignores very real aspects of proteins, such as atoms, backbone angles, sidechain packing, etc. Nevertheless, lattice proteins have a rich history in theoretical biophysics because their simplicity manages to capture salient features of biopolymers.[56] In this paper, we used two different compact lattices: a 5×5 2D and a $3\times3\times3$ 3D lattice. Our choice of using only compact lattices is based on the observation that: (1) hydrophobic collapse and excluded volume are dominant forces, (2) compact lattice structures, with a constant amount of pair contacts, exhibit a Gaussian distribution of energies,[44] and (3) a majority of competitive misfolds and glass transitions are expected to occur in collapsed conformations.[17] For the maximally compact 5×5 2D lattice protein there are a total of 1081 possible self-avoiding walks, excluding rotations and reflections. Similarly, there are 103 346 such possible conformations for the maximally compact $3\times3\times3$ 3D lattice protein chain.

The energy for any given sequence S in a conformation k is a linear function of the amino-acid pair contacts that are made

$$E_S^k=\sum_{i<j}\gamma_{ij}^S\Delta_{ij}^k, \tag{23}$$

where the set $\{\gamma_{ij}^S\}$ specifies the residue pair-contact energies of all possible pair contacts that can be formed for S. Δ_{ij}^k is equal to one if nonsequential residues i and j are on adjacent

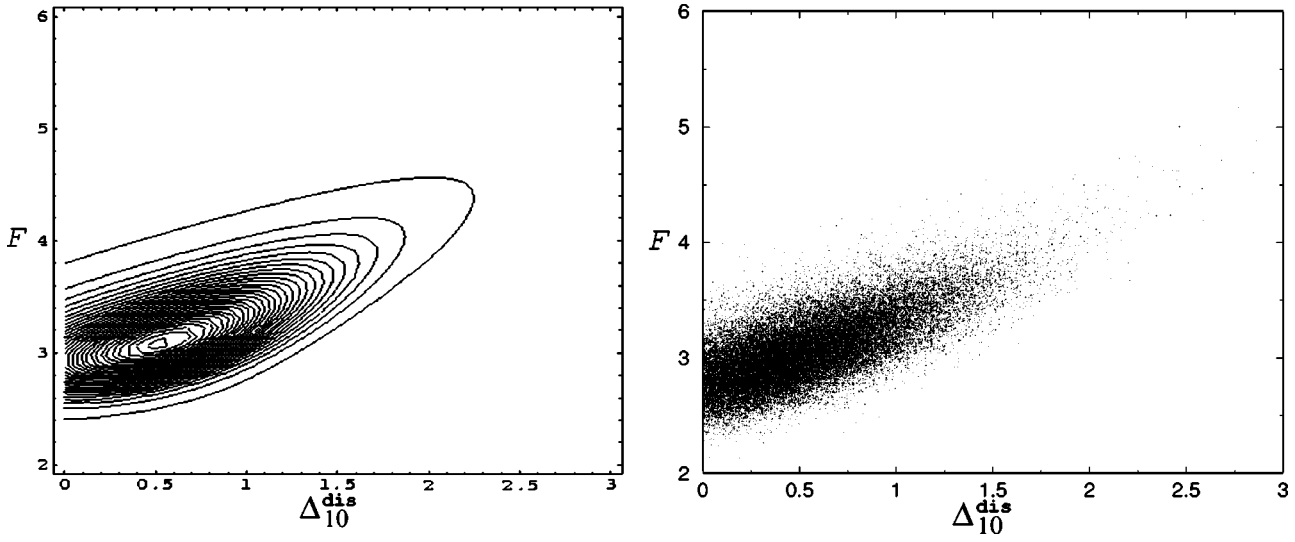

FIG. 7. On the left is a contour plot of the REM joint distribution $\rho(\mathcal{F},\Delta_{10}^{dis})$ for $n=1081$ and $\mathcal{N}^{55}(r)$, which demonstrates the strong, statistical correlation between foldability and energy gap. On the right is a scatter plot of $\mathcal{F}$ and Δ_{10}^{dis} for the corresponding 5×5 lattice protein simulation.

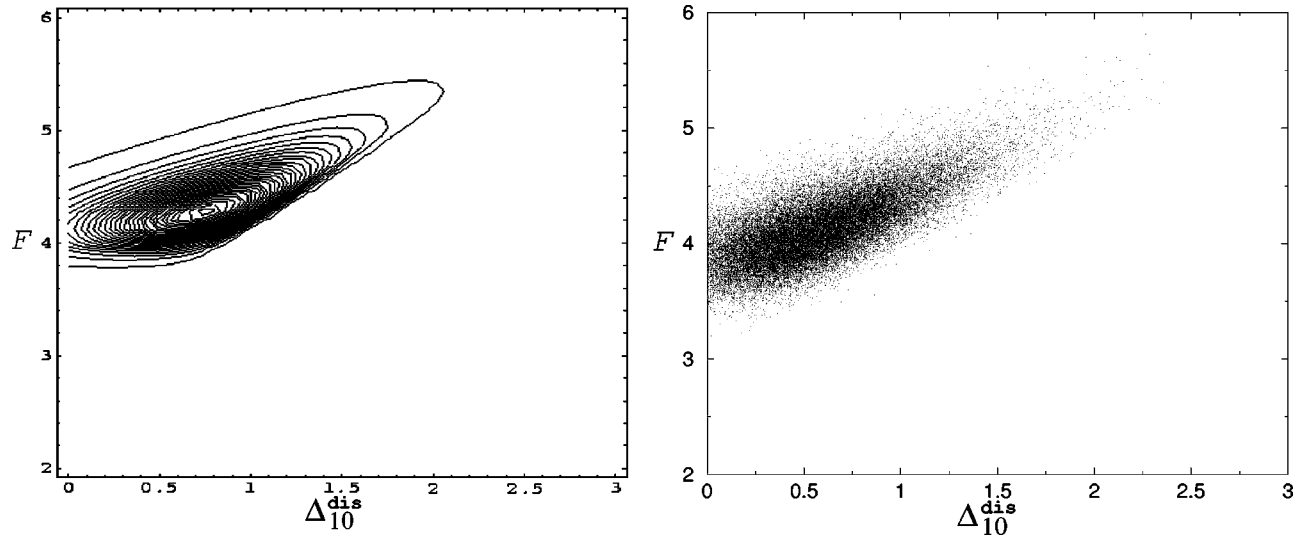

FIG. 8. On the left is a contour REM plot of the joint distribution $\rho(\mathcal{F},\Delta_{10}^{\mathrm{dis}})$ for $n=103\,346$ and $\mathcal{N}^{333}(r)$. On the right is a scatter plot of $\mathcal{F}$ and $\Delta_{10}^{\mathrm{dis}}$ for the corresponding $3\times3\times3$ lattice protein simulation.

lattice sites in conformation k and zero, otherwise. Since all conformations are unique, no set $\{\Delta_{ij}^{k}\}$ is identical. Each pair-contact energy $\gamma_{ij}^{S}=\gamma(\mathcal{A}_i^{S},\mathcal{A}_j^{S})$ is a function of the sequence amino acids $\mathcal{A}_i^{S}$ and $\mathcal{A}_j^{S}$ at positions i and j, where $\gamma(\mathcal{A}_i,\mathcal{A}_j)$ is specified in the definition of the amino-acid alphabet. For both compact lattice models, we constructed our sequences randomly using the standard Miyazawa–Jernigan (MJ) 20-letter alphabet.[57] In the simulation themselves, all sequences had to satisfy similar criteria to what was imposed in our REM calculations: have a unique, nondegenerate, global energy-minimum native state. We generated 50 000 such random MJ heteropolymer sequences for both lattice geometries and kept track of foldability $\mathcal{F}$ and energy gaps Δ_g, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$. As in our REM derivations, all Δ were normalized by σ so as to ensure that distributions of energy gap and foldability remain composition independent. This is in stark contrast to previous simulations which have looked at energy gap statistics. All histograms of $\mathcal{F}$, Δ_g, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$ were binned with a width of 0.05 and normalized to sum to 1.0.

In similar order to our REM calculations, we begin by looking at the single distributions $\rho(\mathcal{F})$, $\rho(\Delta_g)$, $\rho(\Delta_{10})$, and $\rho(\Delta_{10}^{\mathrm{dis}})$ for lattice proteins. Figures 9–11 demonstrate that the $\mathcal{F}$, Δ_g, Δ_{10}, $\Delta_{10}^{\mathrm{dis}}$ histograms are qualitatively similar to their corresponding REM distributions ($n=1081$ and $n=103\,346$). Namely, the distributions of $\rho(\mathcal{F})$ and $\rho(\Delta_g)$ exhibit a Gaussian-like shape with strong positive skew, $\rho(\Delta_{10})$ is an exponentially decreasing function of Δ_{10}, and $\rho(\Delta_{10}^{\mathrm{dis}})$ is a cropped Gaussian, peaking near low–medium values of $\Delta_{10}^{\mathrm{dis}}$. The statistical details of these lattice protein simulations as compared to their REM counterparts are shown in Table II. However, despite these qualitative simi-

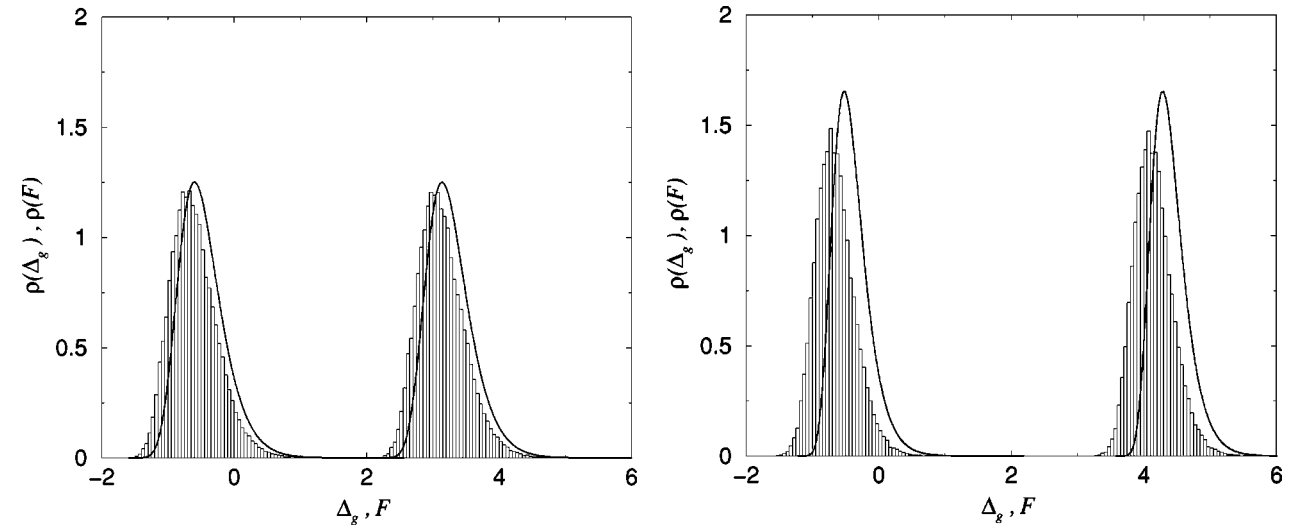

FIG. 9. Histogram of normalized $\mathcal{F}$ and Δ_g distributions for random sequences in different compact lattice geometries (left) for the 5×5 and (right) for the $3\times3\times3$ lattice proteins. We have included the REM $\mathcal{F}$ and Δ_g distributions for comparison.

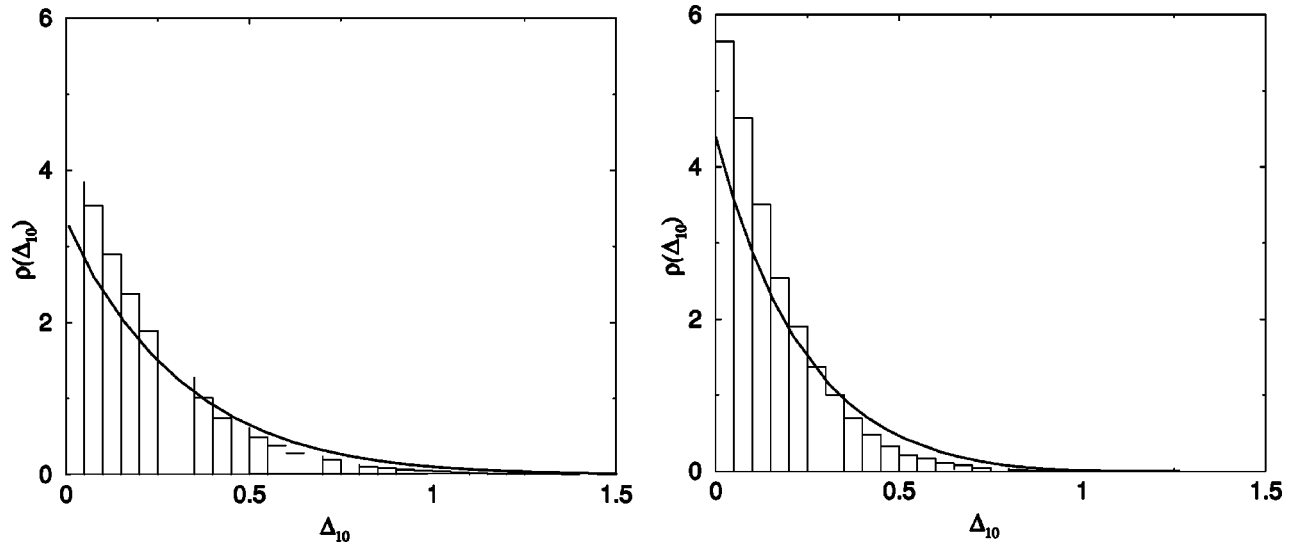

FIG. 10. Histogram of Δ_{10} distribution for random sequences in different compact lattice geometries (left) for the 5×5 and (right) for the $3\times3\times3$ lattice proteins. REM distributions are given by the solid lines.

larities, there are some undeniable quantitative differences: namely, all lattice protein distributions of $\mathcal{F}$, Δ_g, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$ are consistently shifted to lower values. Fortunately, these quantitative differences can be rationalized on the basis of structural and energetic correlations endemic to lattice proteins. Because all structures share a certain amount of similar contacts with other structures, the stochastic consequence of these correlations is an effective decrease in the *possible* energy difference between E_{ns}, $\overline{E}$, and E_r. This results in smaller, average values of foldability and energy gap. Note that the percent of average, similar contacts for 5×5 is 15.44% and 18.79% for $3\times3\times3$ lattice proteins. As shown in Table II, this higher number of average, similar contacts in $3\times3\times3$ lattice proteins conveniently explains the larger drop in $\overline{\mathcal{F}}$, $\overline{\Delta_g}$, and $\overline{\Delta_{10}}$ relative to the REM, when compared to 5×5 lattice proteins. The quick analysis above parallels the discussions of Pande and co-workers concerning the REM breakdown for compact lattice proteins.[54]

What about the correlation between energy gap and foldability in lattice proteins? As expected in their definition, $\mathcal{F}$ and Δ_g were perfectly correlated for lattice proteins (unpublished). The resulting joint distribution of $\rho(\mathcal{F},\Delta_{10})$ is found in the form of a scatter plot in Figs. 4 and 5. Additionally, we plot the joint distribution of $\rho(\mathcal{F},\Delta_{10}^{\mathrm{dis}})$ in the form of a scatter plot in Figs. 7 and 8. We have specifically included them next to their corresponding REM contour plots to effectively highlight the resemblance of lattice distributions and analytical expressions for both $\rho(\mathcal{F},\Delta_{10})$ and $\rho(\mathcal{F},\Delta_{10}^{\mathrm{dis}})$. Again, notice the strong statistical correlation between $\mathcal{F}$, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$ for lattice proteins and the REM. This conclusively demonstrates that $\mathcal{F}$, Δ_g, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$ are all correlated,

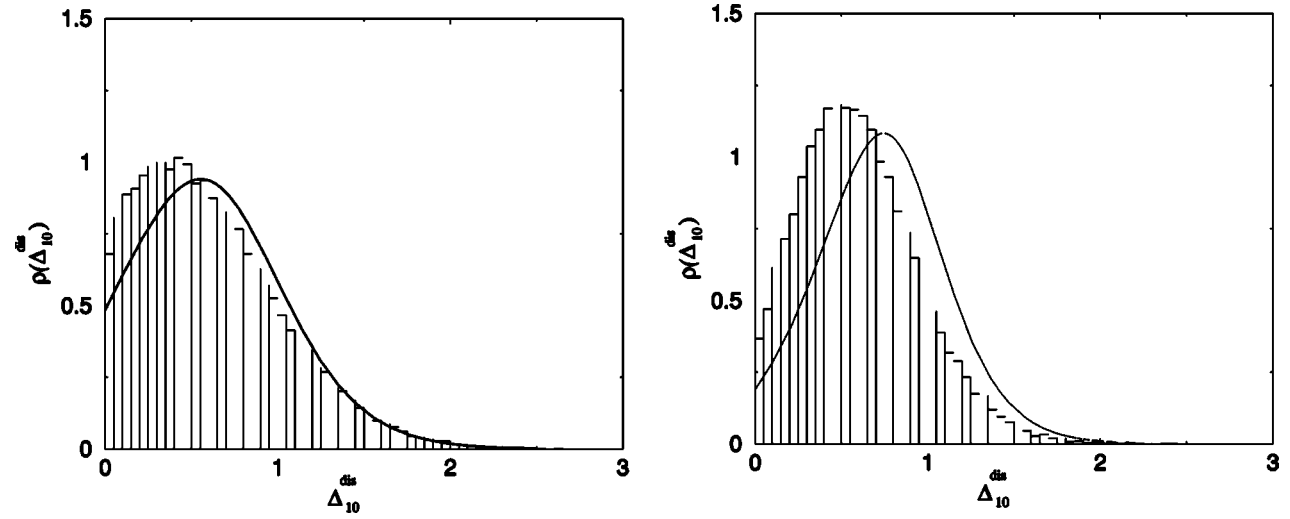

FIG. 11. Histogram of $\Delta_{10}^{\mathrm{dis}}$ distribution for random sequences in different compact lattice geometries (left) for the 5×5 and (right) for the $3\times3\times3$ lattice proteins. REM distributions are given by the solid lines.

TABLE II. A comparison of $\mathcal{F}$, Δ_g, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$ statistics for lattice protein simulations and their corresponding REM heteropolymer sequences.

	$\overline{\mathcal{F}}$	$\sigma_{\mathcal{F}}$	$\mathrm{Skew}_{\mathcal{F}}$	$\overline{\Delta_g}$	$\int\rho(\Delta_g>0)$	$\overline{\Delta_{10}}$	$\overline{\Delta_{10}^{\mathrm{dis}}}$
REM $n=1081$	3.2637	0.3497	0.7703	−0.4741	9.55%	0.2856	0.7066
5×5 lattice	3.1328	0.3527	0.6447	−0.6051	5.58%	0.2230	0.6194
REM $n=103\,346$	4.3915	0.2715	0.8961	−0.4139	7.67%	0.2180	0.6825
$3\times3\times3$ lattice	4.1474	0.2923	0.5466	−0.6580	2.34%	0.1553	0.6069

either perfectly or statistically to one another, both in REM theory and in lattice protein simulation.

CONCLUSIONS

One of the central topics of theoretical biophysics has been to understand how proteins fold so quickly. Interdisciplinary insights quickly focused on protein energy landscapes to extract features relevant to faster folding. In particular, with advances in computational power, it was established using molecular dynamic and Monte Carlo kinetic simulations of proteins that foldability $\mathcal{F}$, energy gap Δ, and sigma σ_θ were well correlated to fast folding. Unfortunately, there has been no theoretical attempt to relate these disparate measures. For reasons mentioned earlier, it was the purpose of this paper to elucidate the inherent connection between $\mathcal{F}$ and energy gap, Δ_g, Δ_{10}, $\Delta_{10}^{\mathrm{dis}}$, using the REM. Our analytical calculations demonstrate that $\mathcal{F}$ and Δ_g are identical measures and that, as shown by the joint distributions $\rho(\mathcal{F},\Delta_{10})$ and $\rho(\mathcal{F},\Delta_{10}^{\mathrm{dis}})$, $\mathcal{F}$ and energy gap are statistically correlated measures. Consequently, *a posteriori* it comes as no surprise that both foldability and energy gap are all highly correlated to fast and reliable folding.

Despite these REM results, we found it necessary to run our own lattice protein simulations to explore whether $\mathcal{F}$ and Δ were indeed correlated. All lattice protein $\mathcal{F}$, Δ_g, Δ_{10}, and $\Delta_{10}^{\mathrm{dis}}$ distributions and joint distribution were qualitatively similar to that predicted using the REM. There were quantitative differences between these two models, but these differences could be explained on the basis of correlations between lattice proteins. It might be worthwhile to calculate $\mathcal{F}$, Δ_{10}, and Δ_g for the GREM, which takes the relationship between energetic and structural correlations into account, but qualitatively little is expected to change.

Given that the stochasticity of the REM and energetic correlations in lattice proteins can be reconciled in the universal GREM, why are Δ and $\mathcal{F}$ positively correlated? We argue that in a stochastic framework, the correlation between foldability and energy gap arises because both measures are defined by E_{ns}, the lowest energy, which is a constraint of the thermodynamic hypothesis. The fact that energy gap, in all its forms, and foldability are a measure of the depth of the native state with respect to *something*, either E_r, E_g, or $\overline{E}$, leads to a positive correlation between these measures. The particular *form* of the correlation or joint distribution, however, is sensitive to the details of the underlying distribution of energies and our particular definition of energy gap. Thus, given that $\overline{E}$ and E_g are constants of the underlying distribution, Δ_g and $\mathcal{F}$ should be less fickle measures of fast folding, as compared to Δ_{10} and $\Delta_{10}^{\mathrm{dis}}$.

ACKNOWLEDGMENTS

The authors would like to thank Tom Weinacht for technical assistance and Todd Raeker for computational support. Financial backing was provided by NIH Grant Nos. LM05770 and GM08270, and NSF shared-equipment Grant No. BIR9512955.

[1] D. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).
[2] D. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985).
[3] P. Anderson, Proc. Natl. Acad. Sci. USA **80**, 3386 (1983).
[4] D. Rokhsar, P. Anderson, and D. Stein, J. Mol. Evol. **23**, 119 (1986).
[5] S. Kauffman and S. Levin, J. Theor. Biol. **128**, 11 (1987).
[6] C. Amitrano, L. Peliti, and M. Saber, J. Mol. Evol. **29**, 513 (1989).
[7] C. A. Macken and A. S. Perelson, Proc. Natl. Acad. Sci. USA **86**, 6191 (1989).
[8] S. Kauffman, E. Weinberger, and A. Perelson, *Santa Fe Institute Studies in the Sciences of Complexity* (Addison-Wesley, Reading, MA, 1988).
[9] S. Kauffman and E. Weinberger, J. Theor. Biol. **141**, 211 (1989).
[10] M. S. Friedrichs and P. G. Wolynes, Science **246**, 371 (1989).
[11] M. S. Friedrichs, R. A. Goldstein, and P. G. Wolynes, J. Mol. Biol. **222**, 1013 (1991).
[12] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).
[13] J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. **93**, 6902 (1989).
[14] E. I. Shakhnovich and A. M. Gutin, J. Phys. A **22**, 1647 (1989).
[15] E. I. Shakhnovich and A. M. Gutin, Biophys. Chem. **34**, 187 (1989).
[16] E. I. Shakhnovich and A. V. Finkelstein, Europhys. Lett. **9**, 569 (1989).
[17] J. D. Bryngelson and P. G. Wolynes, Biopolymers **30**, 171 (1990).
[18] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins **21**, 167 (1995).
[19] R. A. Goldstein, Z. A. Luthey-Schulten, and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **89**, 4918 (1992).
[20] R. A. Goldstein, Z. A. Luthey-Schulten, and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **89**, 9029 (1992).
[21] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **101**, 1519 (1994).
[22] M. R. Betancourt and J. N. Onuchic, J. Chem. Phys. **103**, 773 (1995).
[23] P. E. Leopold, M. Montal, and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **89**, 8721 (1992).
[24] H. Nymeyer, A. Garcia, and J. Onuchic, Proc. Natl. Acad. Sci. USA **95**, 5921 (1998).
[25] S. Govindarajan and R. A. Goldstein, Biopolymers **36**, 43 (1995).
[26] S. Govindarajan and R. A. Goldstein, Proc. Natl. Acad. Sci. USA **93**, 3341 (1996).
[27] S. Govindarajan and R. A. Goldstein, Biopolymers **42**, 427 (1997).
[28] S. Govindarajan and R. A. Goldstein, Proteins **29**, 461 (1997).
[29] E. I. Shakhnovich and A. M. Gutin, J. Chem. Phys. **93**, 5967 (1990).
[30] E. I. Shakhnovich and A. M. Gutin, Nature (London) **346**, 773 (1990).
[31] A. Sali, E. I. Shakhnovich, and M. J. Karplus, J. Mol. Biol. **235**, 1614 (1994).
[32] A. Sali, E. I. Shakhnovich, and M. J. Karplus, Nature (London) **369**, 248 (1994).

[33] E. I. Shakhnovich and A. M. Gutin, Proc. Natl. Acad. Sci. USA **90**, 7195 (1993).
[34] A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. USA **92**, 1282 (1995).
[35] B. Derrida and G. Toulouse, J. Phys. (France) Lett. **46**, L401 (1985).
[36] B. Derrida and E. Gardner, J. Phys. C **19**, 2253 (1986).
[37] C. J. Camacho and D. Thirumalai, Proc. Natl. Acad. Sci. USA **90**, 6369 (1993).
[38] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. **76**, 4070 (1996).
[39] D. K. Klimov and D. Thirumalai, J. Chem. Phys. **109**, 4119 (1998).
[40] T. L. Chiu and R. A. Goldstein, J. Chem. Phys. **107**, 4408 (1997).
[41] A. R. Dinner, V. Abkevich, E. I. Shakhnovich, and M. Karplus, Proteins **35**, 34 (1999).
[42] B. Derrida, Phys. Rev. Lett. **45**, 79 (1980).
[43] B. Derrida, Phys. Rev. B **24**, 2613 (1981).
[44] W. Wilbur and J. Liu, Macromolecules **27**, 2432 (1994).
[45] W. Wilbur, F. Major, J. Spouge, and S. Bryant, Biopolymers **38**, 447 (1996).
[46] C. Anfinsen, Science **181**, 223 (1973).
[47] S. Govindarajan and R. A. Goldstein, Proc. Natl. Acad. Sci. USA **95**, 5545 (1998).
[48] A. M. Gutin and E. I. Shakhnovich, J. Chem. Phys. **98**, 8174 (1993).
[49] V. S. Pande, A. Y. Grosberg, and T. Tanaka, Rev. Mod. Phys. (in press).
[50] S. S. Plotkin, J. Wang, and P. G. Wolynes, J. Chem. Phys. **106**, 2932 (1997).
[51] B. A. Shoemaker, J. Wang, and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **94**, 777 (1997).
[52] A. V. Finkelstein and O. B. Ptitsyn, Prog. Biophys. Mol. Biol. **50**, 171 (1987).
[53] A. V. Finkelstein, A. M. Gutin, and A. Y. Badretdinov, FEBS Lett. **325**, 23 (1993).
[54] V. S. Pande, A. Y. Grosberg, C. Joerg, and T. Tanaka, Phys. Rev. Lett. **76**, 3987 (1996).
[55] S. S. Plotkin, J. Wang, and P. G. Wolynes, Phys. Rev. E **53**, 6271 (1996).
[56] H. S. Chan and K. A. Dill, Phys. Today **46**, 24 (1993).
[57] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).