

Sridhar Govindarajan
Department of Chemistry
University of Michigan
Ann Arbor, MI 48109-1055

Richard A. Goldstein
Department of Chemistry
and Biophysics Research
Division
University of Michigan
Ann Arbor, MI 48109-1055

Searching for Foldable Protein Structures Using Optimized Energy Functions

During evolution, the effective interactions between residues in a protein can be adjusted through mutations to allow the protein to fold to its native structure on an adequate time scale. We seek to address the question: Are there some structures that can be better optimized than others? Using exhaustive enumeration of the compact conformations of short proteins confined to simple lattices, we find that the best structures are those that contain contacts rare in random structures, indicating the importance of nonlocal contacts for assisting the folding process. Certain structural motifs such as long β -hairpins, Greek-key motifs, and jelly rolls, commonly found in proteins of known structure, have a high degree of optimizability. Contrary to what might be expected, positive correlations between the various interactions reduce optimizability. The optimization procedure produces a correlated energy landscape, which might assist folding. © 1995 John Wiley & Sons, Inc.

INTRODUCTION

Understanding how a protein obtains its biologically active conformation is one of the central problems of molecular biophysics. Much of the theoretical and experimental study of protein folding has been directed toward a resolution of the so-called Levinthal paradox, understanding how a protein can find the relatively few states classified as native in the midst of a vast space of possible conformations.¹ In the absence of a detailed understanding of the interactions between various parts of the amino acid sequence, and given the lack of computational power necessary to model the folding process in detail, valuable insight has come from the results of simple models such as lattice calculations, where proteins are represented as

polymers with random interactions, with interactions designed to capture the salient features of the true energy function, or with interactions designed to stabilize the folded conformation. (For reviews, see Refs. 2 and 3.)

Work in this area has been dominated by two different approaches. According to the first view, folding proceeds rapidly because of the existence of a "folding code" embedded in the sequence, which directs the folding along a well-defined pathway. This approach is often combined with a view that the final conformation is determined by kinetic barriers, and that the folded conformation is not necessarily the conformation of lowest free energy. Recently, a second approach has concentrated on the thermodynamic requirements necessary for rapid folding. In this view, folding is rapid precisely

because there is no single pathway, but instead a multiplicity of routes and a large number of possible transition states. Some recent work along these lines has been based on ideas borrowed from the physics of spin glasses, systems dominated by an extremely rough free-energy landscape, yet whose analysis has been made tractable by the development of simple models.^{4–15} This analytical work in combination with computer simulations^{16–19} suggests that proteins should be able to fold rapidly as long as there is a large energy difference between the folded state and other possible conformations. In particular, molecular dynamics and Monte Carlo simulations have demonstrated successful simulations of protein folding when the energy landscape has been suitably prepared, either through optimization of an energy function for a fixed sequence,¹³ or optimization of the sequence for a fixed energy function.^{20–23} In contrast, simulations have shown that successful folding to a consistent final structure is rather rare among random sequences, even among sequences with only one thermodynamically dominant state.^{16,24–28}

This optimization approach takes advantage of the relative robustness of structure compared with the plasticity of sequences during evolution. Sequences mutate, changing the interactions between the various residues, all in a way consistent with foldability of the structure. While on the folding time scale, it is the sequence that determines the structure; on the evolutionary time scale, it is the structure that determines the sequence, and thus the nature of the interactions between various parts of the protein. Evolution provides biology with a mechanism for optimizing the energy landscape of the protein, so as to facilitate folding.

In this paper, we take advantage of this separation in time scales and consider the folding properties of various structures, assuming that the inter-residue interactions have long adapted to that particular structure through sequence mutations, in the same way that electrons are considered to instantaneously adapt to the position of nuclei when analyzing nuclear motion in the Born–Oppenheimer approximation. By considering only these optimized interactions, it is then possible to consider how easy it is to produce a foldable protein as a function of its structure.

This assumption of optimized interactions is, of course, simplistic. There are many constraints on the ways that proteins can modulate the interactions between the various residues—for instance, due to the availability of only a limited number of possible amino acids, a limitation explicitly in-

cluded in the Shakhnovich model.^{20–23} Protein sequences have to evolve to fulfill many functions, only one of which is the ability to fold; excessive native state stability may conflict with the mobility required for these functions. Finally, any finite mutation rate will decrease the ability of a protein to hold its interactions at an optimal level. Experimentally, the ease at which natural proteins can be mutated in order to increase their stability argues that biological proteins have not been so optimized, at least not for maximum stability.^{29,30} The presence of long-range multibody interactions, and extra degrees of freedom furnished by variations in side-chain positions, however, may actually give the proteins more flexibility in adjusting the interaction parameters than the simple model discussed here. In addition, as Finkelstein has pointed out, small differences in the ability of various structural motifs to have the favorable thermodynamics necessary for rapid folding can have a large effect on the likelihood that such structures would be found among biological proteins.^{31,32}

This study considers short model proteins, where each residue is represented as a point confined to a two-dimensional or three-dimensional lattice. For such proteins, it is possible to do an exhaustive enumeration of all compact conformations.^{33–36} We use the methods developed by Wolynes and co-workers to independently optimize each conformation.^{13–15} Different structures have, in general, different abilities to be optimized for folding. Structures are considered to be “good” or “bad” folders according to how well their interactions can be optimized. In contrast to what would be expected given a number of models of protein folding,^{37–47} we find that the structures that can be best optimized for folding have many contacts that are unlikely in random structures. In the two-dimensional lattice models, this results in a preference for long β -sheet structures compared to sheet structures with more numerous but shorter β -strands, and a preference for Greek-key and jelly-roll motifs compared to the β -meander. Interactions that are statistically positively correlated are less favorable for optimization than uncorrelated interactions. The optimization procedure naturally stabilizes structures that are similar in structure, increasing the basin of attraction of the folded state, possibly forming “folding funnels” as has been proposed based on other lattice model simulations.²⁸

METHODS

Three different lattice models were used in this study. The first model consisted of a chain of 36 monomers,

confined to a 6×6 two-dimensional square lattice. The second model elaborated on the first model by considering all compact rectangular shapes, including conformations confined to 9×4 , 12×3 , and 18×2 square lattices, in addition to the 6×6 lattice. The third model consisted of a chain of 27 monomers, confined to a $3 \times 3 \times 3$ three-dimensional cubic lattice. The bonds were all of unit length, with adjacent residues existing at adjacent sites. In all three cases, it is possible to enumerate all of the possible self-avoiding walks, not counting reflections and rotations: there are a total of 57,337 self-avoiding walks for the 6×6 square lattice, 133,104 walks for the set of rectangular two-dimensional lattices, and 103,346 walks for the cubic lattice.

The energy function is of the simple contact form:

$$E = \sum_{i < j} \gamma_{ij} \delta_{ij} \quad (1)$$

where δ_{ij} is equal to one if residues i and j are not adjacent in sequence but are on adjacent lattice sites, and zero otherwise, and γ_{ij} represents the energy contribution for contact between residues i and j . The values of $\{\gamma_{ij}\}$ are the adjustable parameters determined by the optimization process.

Due to the nature of the lattice, the only contacts possible are between odd and even residues, so the total number of possible contacts are 289 for the two-dimensional lattice and 156 for the three-dimensional cubic lattice. There are a total of 25 contacts in any conformation of the 6×6 two-dimensional square lattice, and 28 contacts for the $3 \times 3 \times 3$ three-dimensional cubic lattice. As each conformation in these two particular lattices has the same number of contacts, and only relative energy differences matter, there are one fewer adjustable parameters than possible contacts, i.e., 288 for the square lattice and 155 for the cubic lattice, and there is an adjustable parameter that determines the conformation of zero energy. In the expanded two-dimensional lattice model, the number of contacts depends on the shape of the protein, so the number of adjustable parameters equals 289, the total number of possible contacts. In this case, all energies are relative to an extended conformation where no contacts are made.

Energy functions were optimized using the spin-glass optimization methodology developed by Wolynes and co-workers,¹³⁻¹⁵ based on the random-energy model of Derrida,^{48,49} as applied to the protein-folding problem by Bryngelson and Wolynes.⁵ In this model, there is a competition between two phase transitions, the first to an ordered state, representing the folded conformation, at a temperature T_f , and the second to a glassy state, representing unfoldability, at a temperature T_g . The glassy state corresponds to the situation when the dynamics are dominated by the roughness of the energy landscape, due to a lack of thermally accessible transition states between minima. The optimal condition for folding is when the protein can fold at a temperature high with respect to the glass transition temperature. This leads to a measure of

optimality for the energy function, as that energy function that maximizes the ratio of T_f/T_g .

Wolynes and co-workers showed that, for an energy function linear with respect to a set of adjustable parameters, as is the case with the energy function represented in Eq. (1), the optimal values of these parameters can be solved for in closed form, as follows.¹³⁻¹⁵ In the random energy model, T_f/T_g is given by

$$\frac{T_f}{T_g} = \sqrt{\frac{\mathcal{R}^2}{2S_0}} + \sqrt{\frac{\mathcal{R}^2}{2S_0} - 1} \quad (2)$$

for $\mathcal{R} = \Delta/\Gamma$, where Γ is the width of the distribution of energy values in the ensemble of random states, Δ is the average energy difference between these states and the correctly folded state, and S_0 is the configurational entropy of the protein. As T_f/T_g is a monotonically increasing function of \mathcal{R} , T_f/T_g will be maximized when \mathcal{R} is maximized. For the Hamiltonian in Eq. (1), we can express the energy of the protein in its native state by $E_T = \sum_{i < j} \gamma_{ij} \delta_{ij}^T$ and in random state k by $E_k = \sum_{i < j} \gamma_{ij} \delta_{ij}^k$. Δ and Γ are then given by $\Delta = \mathbf{A}\boldsymbol{\gamma}$ and $\Gamma^2 = \boldsymbol{\gamma}\mathbf{B}\boldsymbol{\gamma}$, for vector \mathbf{A} and matrix \mathbf{B} , with

$$A_{ij} = \delta_{ij}^T - \langle \delta_{ij}^k \rangle_k \quad (3)$$

$$B_{ij, mn} = \langle \delta_{ij}^k \delta_{mn}^k \rangle_k - \langle \delta_{ij}^k \rangle_k \langle \delta_{mn}^k \rangle_k \quad (4)$$

where the averages are over the random states k . Maximization of T_f/T_g leads to:

$$\boldsymbol{\gamma} = -\mathbf{B}^{-1}\mathbf{A} \quad (5)$$

In general, 27-residue proteins with optimized interactions on the $3 \times 3 \times 3$ cubic lattice folded readily in Monte Carlo simulations, often in as few as 100,000 time steps. Wolynes and co-workers showed the ability of this procedure to produce optimized energy functions for use in tertiary structure prediction.¹³⁻¹⁵ A discretized form of this optimization procedure was used by Shakhnovich and co-workers in order to produce sequences that would fold readily in Monte Carlo calculations.²⁰⁻²³ A criterion related to \mathcal{R} was found by Chan and Dill and by Karplus and co-workers to distinguish random sequences that fold easily from those that do not.¹⁶⁻¹⁸

All of the compact structures are used to compute \mathbf{B} and $\langle \delta_{ij}^k \rangle_k$. Each of the compact structures in turn are used to compute \mathbf{A} , and an optimal set of $\{\gamma_{ij}\}$ are found for each possible compact conformations using Eq. (5). The structures were judged by the resulting value of \mathcal{R}_{opt} , the value of \mathcal{R} calculated with the optimized set of $\{\gamma_{ij}\}$.

The true thermodynamic description would include all of the possible conformations; this is currently unfeasible for even the short polymers used in this study. The use of compact conformations models the situation where all of the contact energies were large enough so that the extended conformations are thermodynamically

irrelevant. Use of a random ensemble of conformations generated by infinite-temperature Monte Carlo simulation confined to a moderately sized three-dimensional volume did not qualitatively change the optimal interactions nor the relative optimizability of the various structures.

RESULTS

The resulting distributions of \mathcal{R}_{opt} values for the three lattice models are shown in Figure 1.

A number of properties distinguish good structures from poor structures. The best structures are those that have the maximum number of contacts that are unlikely in random structures. Figure 2 shows the conformational entropy change ΔS_0 for formation of the various native contacts in highly and poorly optimizable structures on the 6×6 square lattice, where ΔS_0 is defined as the logarithm of the number of structures where that contact is made relative to the total number of structures. As can be seen, proteins where the native contacts individually most decrease the conformational entropy have a higher degree of optimizability. When the off-diagonal terms of \mathbf{B} are small, γ_{ij} corresponding to native contacts is given by

$$\gamma_{ij} \simeq -\frac{1}{\langle \delta_{ij}^k \rangle_k} \quad (6)$$

Smaller values of $\langle \delta_{ij}^k \rangle_k$, corresponding to infrequent interactions and large negative values of ΔS_0 , cause larger interactions stabilizing the native state, resulting in a larger value of \mathcal{R}_{opt} .

Figure 3 shows the change in conformational entropy of forming any given contact in the 6×6 square lattice.³³ In general, the most common interactions are between the residues closest in sequence that can possibly interact, between i and $i + 3$. These interactions are common in the poorly optimizable structures. In both the two-dimen-

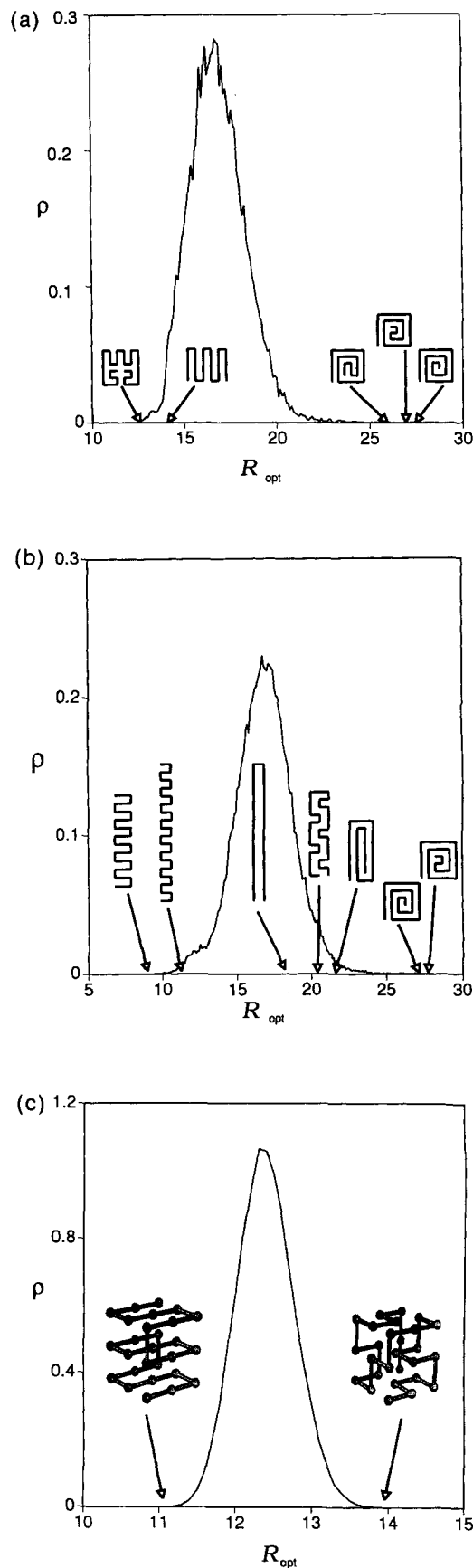


FIGURE 1 Distribution of \mathcal{R}_{opt} values for the various conformations of (A) a 36-residue protein confined to a two-dimensional 6×6 square lattice; (B) a 36-residue protein confined to a set of possible two-dimensional square lattices, including 6×6 , 9×4 , 12×3 , and 18×2 ; and (C) a 27-residue protein confined to a $3 \times 3 \times 3$ three-dimensional cubic lattice. The highly and poorly optimizable structures are shown in the figure.

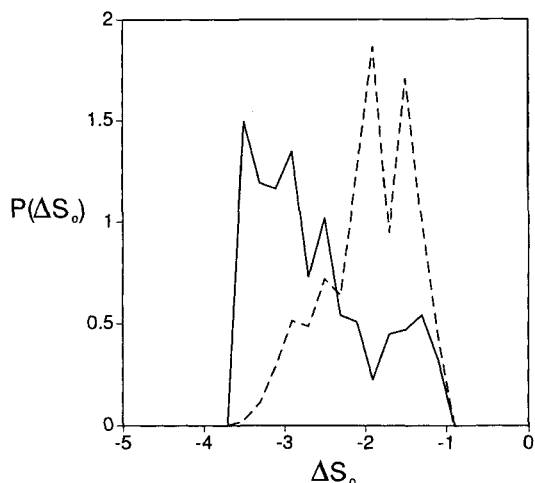


FIGURE 2 Probability density [$P(\Delta S_0)$] for a native contact to have conformational entropy change (ΔS_0) for the best hundred (—) and worst hundred (---) structures on the 6×6 square lattice. The conformational entropy change ΔS_0 for a contact ij is defined as $\Delta S_0 = \ln(\Omega_{ij}/\Omega_0)$, where Ω_{ij} is the number of conformations with contact ij , and Ω_0 is the total number of conformations.

sional and the three-dimensional case, the most unlikely contacts do not correspond to the residues that are furthest from each other along the sequence. In fact, the least likely contacts are between residues 8 and 25 and between 12 and 29 for the 6×6 square lattice, and between residues 6 and 13 and between 15 and 22 for the $3 \times 3 \times 3$ cubic lattice. This is likely the result of confining the ensemble of random structures to compact conformations. If random conformations during *in vitro* folding followed random-flight statistics, the most unlikely contacts would be the long-range contacts, suggesting that in real proteins foldability would be increased by increasing the number of these long-range contacts. It is, however, not clear how valid random-flight statistics are for proteins in the denatured state, especially under folding conditions.⁵⁰

As shown in Figure 1C, the most optimizable structure in the cubic lattice has an 11-residue α -helix. More study of larger three-dimensional lattice models would be necessary to see if the optimizability of this structure is due to the properties of the α -helix or due to the peculiar statistics of the possible compact conformations on the $3 \times 3 \times 3$ cubic lattice. On the other hand, interesting correlations can be made between the various square lattice conformations and the occurrences of various β -sheet structures. For instance, all of the possible

compact β -meanders in the extended two-dimensional square lattices are shown as examples (a)–(g) in Table I. In general, as the length of the β -strands gets larger, longer range contacts are made, and the optimizability of the structure increases. The β -hairpin motif [example (g)] is quite optimizable, even compared with other structures with a larger total number of contacts. It is also interesting to note the greater degree of optimizability of more complex β -sheet structures, such as the Greek-key motif [example (h)] and a reduced representation of the jelly-roll motif [example (i)].

Positive statistical correlations between contacts in random conformations, represented by off-diagonal terms in \mathbf{B} , reduce the significance of these contacts if they are both present in the native structure. For instance, assuming the off-diagonal terms are small, γ_{ij} , the energy contribution due to the native contact between residues i and j is given by

$$\gamma_{ij} \simeq -\frac{1 - \langle \delta_{ij}^k \rangle_k}{B_{ij,ij}} + \sum_{mn} \frac{(\delta_{mn}^T - \langle \delta_{mn}^k \rangle_k) B_{ij,mn}}{B_{ij,ij} B_{mn,mn}} \quad (7)$$

As $B_{ij,ij}$ and $B_{mn,mn}$ are always positive, positive values of $B_{ij,mn}$ decrease the contribution of γ_{ij} to sta-

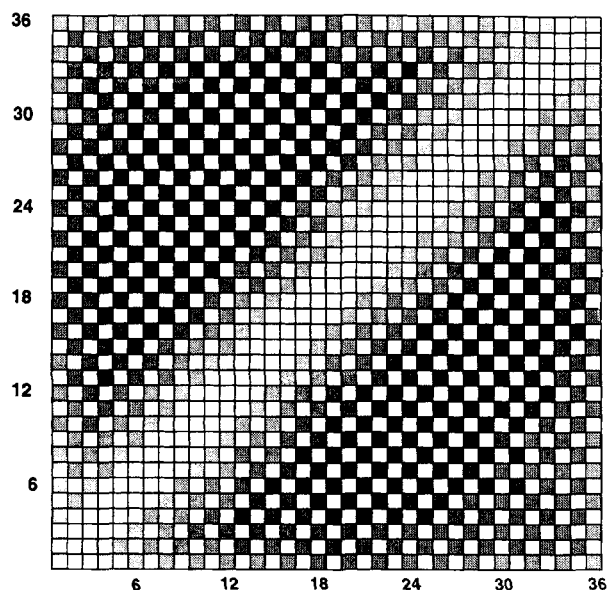




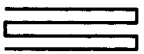
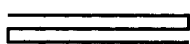
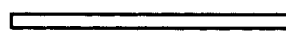
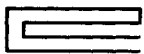



FIGURE 3 Configurational entropy change ΔS_0 in forming a particular contact for proteins confined to a 6×6 square lattice. White squares represent contacts that are not possible due to the nature of the lattice model. The darker the square the larger the entropy change in establishing that contact. The greatest entropy change is for contact between residues 8 and 25 and between residues 12 and 29.

Table I List of R_{opt} Values for various 36-Residue β -Strand Conformations Possible in the Extended Set of Square Lattices

a.		$R_{\text{opt}} = 12.01$
b.		9.58
c.		11.09
d.		14.38
e.		16.13
f.		16.59
g.		18.17
h.		18.74
i.		19.09

bility when the native state contains contact mn and $\delta_{mn}^T = 1$. Conversely, the presence of contacts in the native structure that are negatively correlated in random structures should cause an even greater degree of optimizability. Some contacts have to be highly correlated, due to the nature of the lattice model: the presence of an $i - i + 5$ contact increases the probability of an $i + 1 - i + 4$ contact. It is the large number of these correlated contacts that cause the lack of optimizability of example (b) compared with example (a), even though the total number of contacts made in (b) is larger.

In contrast to these *statistical* correlations between the various interactions that reduce optimizability, the optimization criterion induces correlations in the energy landscape, by stabilizing conformations similar to the correct conformation. This can be seen in Figure 4, which shows the energies of a number of such conformations relative to the overall distribution of energy levels, for interactions optimized for the most optimizable 6×6 square lattice structure.

These induced correlations can also be seen in plots of $P_T(Q)$, defined by

$$P_T(Q) = \sum_{k,l} P_i(k)P_i(l)\delta(Q - Q_{k,l}) \quad (8)$$

where $P_T(k)$ is the probability of the system being in conformation k , given as the Boltzmann probability at temperature T , and $Q_{k,l}$ is the proportion of the contacts that are the same in structure k and l .^{10,16} $P_T(Q)$ measures the probability that two structures picked at random from an ensemble of such conformations will be structurally similar. $P_T(Q)$ is plotted in Figure 5, for the $3 \times 3 \times 3$ cubic lattice, computed at the temperature at which the probability of a structure being in the native conformation is 0.5. A number of statistical tendencies are observable. The energy landscape becomes well correlated, causing a large value of $P_T(Q)$ between 0.8 and 1.0, relative to the case where interactions are chosen at random from a Gaussian distribution. Although the optimization procedure assumes a *lack* of such correlations, the correlations so produced in the energy landscape could be important in increasing the rate of folding by increasing the size of the basin of attraction of the folded state, as discussed by Onuchic and co-workers.²⁸ This effect is even more pronounced for the 6×6 square lattice, as the more restrained nature of the lattice induces greater correlations in the energy landscape.

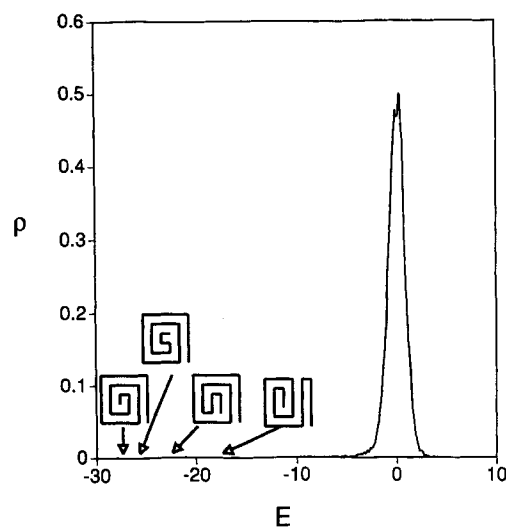


FIGURE 4 Distribution of energies of the various conformations on the 6×6 two-dimensional square lattices, for the set of $\{\gamma_{ij}\}$ values optimized for the structure shown at the left in the figure, the most optimizable conformation on the lattice. As shown, many similar structures are also stabilized by the optimization procedure.

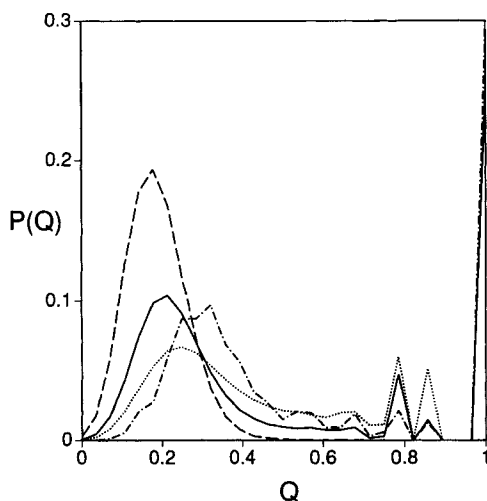


FIGURE 5 $P_{T_f}(Q)$ curves for the 100 best structures (—), 100 worst structures (···), and 100 structures with random γ values (---), compared with the distribution of Q values for random pairs of proteins (-.-), for structures on the $3 \times 3 \times 3$ three-dimensional cubic lattice.

In addition to the increase in $P_{T_f}(Q)$ for high values of Q , there is also an increase in $P_{T_f}(Q)$ for low values of Q , especially for highly optimizable structures. This reflects the fact that as \mathcal{R}_{opt} increases, as well as T_f , more states in the continuous distribution will be thermally accessible at the folding temperature. The distribution of these lower Q values becomes closer to that for random pairs of structures. Again, it is the large number of thermally accessible states that would contribute to folding, as the protein has more routes accessible to escape from local minima in the energy landscape.

DISCUSSION

There have been a number of models of protein folding that have emphasized the building up of local contacts.³⁷⁻⁴⁷ These models often focus on the interactions that can be readily made and the correlations between these interactions, with an emphasis on determining what generic features of polymers can be used to direct the folding process into a limited set of folding pathways. The present model in some ways represents the opposite extreme, where the evolutionary process is able to dominate the more generic polymeric features of the protein. The most optimizable structures are the ones with the largest number of interactions that are rare in random structures, presumably

those that are nonlocal. In addition, the presence of strong, positive correlations between the various interactions, rather than assisting in the folding process, reduces the stabilization due to those interactions, leading to less optimizable structures.

As discussed in the introduction, there are a number of reasons why this extreme limit of perfectly optimized interactions is unrealistic. Even in the absence of total optimization, however, biological proteins are still likely to have highly optimizable structures. Consider a fitness space, where the independent variables are the strengths of the various interactions, and the dependent variable is foldability, represented by the value of \mathcal{R} . The optimal values of the various interactions correspond to a fitness value of $\mathcal{R} = \mathcal{R}_{opt}$, with any shift away from the optimal values corresponding to a decrease in \mathcal{R} . We can consider that there is a critical value of \mathcal{R} , \mathcal{R}_{crit} , below which proteins cannot fold in an adequate time scale. The question then arises: How large is the region of interaction space around a given optimal value where $\mathcal{R} > \mathcal{R}_{crit}$? If we expand the value of \mathcal{R} around $\mathcal{R} = \mathcal{R}_{opt}$, the volume of parameter space for $\mathcal{R} > \mathcal{R}_{crit}$ will scale as $(\mathcal{R}_{opt} - \mathcal{R}_{crit})$ to a large power, due to the large dimension of the parameter space. For biological proteins, where the adjustable parameters are the residues at each position in the chain, the dimension of the parameter space is on the order of the number of residues. This means that the neighborhood of interactions around higher values of \mathcal{R}_{opt} will be much larger than for lower values of \mathcal{R}_{opt} . Structures with larger values of \mathcal{R}_{opt} will be more likely to originate from evolution, as well as to be more robust to the effect of mutations. The plasticity of sequences, the ability of the sequence to change while preserving the coarser aspects of the protein structure, may reflect the predominance of larger values of \mathcal{R}_{opt} and the concomitant large range of interaction parameters preserving foldability. The importance of this effect will depend on the relative values of \mathcal{R}_{opt} and \mathcal{R}_{crit} .

The result that some structures might be more “designable” than others has been suggested by Dill and co-workers,^{51,52} but is in conflict with the results of other work that found little correlation between structure and foldability.¹⁶ This latter result may be a consequence of the use of inappropriate measures for monitoring the structural properties of folding and nonfolding sequences, especially given the few foldable sequences found.

Despite the differences in emphasis between the local contact model and the optimized interaction model developed here, there are some structures,

such as long β -hairpins, Greek-key motifs, and jelly-roll structures, that are suggested by both models as well as being commonly observable in proteins of known structure.⁵³ In addition, different models may be relevant at different stages in the folding process. Initially, local structures, such as α -helices, can form due to local interactions, followed by an assembly of these structures into maximally optimizable tertiary structures following the principles derived here. For instance, in cytochrome C, hydrogen-exchange studies suggest that the first tertiary contacts formed are between the two terminal α -helices.⁵⁴ This result, hard to understand using the local interaction picture, would be quite predictable, as the interactions between the ends would be the strongest after optimization. It is also conceivable that understanding the folding process of different proteins may require different considerations. For instance, large differences are observed in the intermediate states seen in the folding of α -helical proteins⁵⁴⁻⁵⁷ and β -sheet proteins,⁵⁸ as detected using hydrogen-exchange measurements.

Although the random-energy model that forms the basis for the optimization procedure neglects correlations between the energies of neighboring states, it would not be possible for the protein to avoid the Levinthal paradox if this were rigorously true: a high degree of correlation must exist in the energy landscape in order for the protein to be able to efficiently search the possible conformation space. While the optimized-interaction theory suggests that structures with highly correlated interactions are less optimizable than structures without such correlations, the energy surface will itself become correlated by such an optimization procedure; alternative states will also be stabilized to the extent that they contain interactions found also in the native state. Understanding these correlations may be central to understanding the folding process and may have important ramifications in the dynamics of protein folding.

We would like to thank Gordon Crippen and William Jockusch for helpful comments, and Kurt Hillig for computational assistance. Financial support was provided by the College of Literature, Science, and the Arts, the Horace H. Rackham School of Graduate Studies, and the Program in Protein Structure and Design at the University of Michigan.

REFERENCES

1. Levinthal, C. (1969) in *Mossbauer Spectroscopy in Biological Systems*, eds. Debrunner, P., Tsbiris, J. C. M. & Munck, E., Eds., University of Illinois Press, Urbana, pp. 22-24.
2. Skolnick, J. & Kolinski, A. (1989) *Ann. Rev. Phys. Chem.* **40**, 207-235.
3. Karplus, M. & Shakhnovich, E. (1992) in *Protein Folding*, Creighton, T., Ed., W. H. Freeman and Co., New York, pp. 127-195.
4. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7524-7528.
5. Bryngelson, J. D. & Wolynes, P. G. (1990) *Biopolymers* **30**, 171-188.
6. Garel, T. & Orland, H. (1988) *Europhys. Lett.* **6**, 307-310.
7. Garel, T. & Orland, H. (1988) *Europhys. Lett.* **6**, 597-601.
8. Shakhnovich, E. I. & Gutin, A. M. (1988) *Europhys. Lett.* **8**, 327-332.
9. Shakhnovich, E. I. & Gutin, A. M. (1989) *Stud. Biophys.* **132**, 47-56.
10. Shakhnovich, E. I. & Gutin, A. M. (1989) *Biophys. Chem.* **34**, 187-199.
11. Sasai, M. & Wolynes, P. G. (1990) *Phys. Rev. Lett.* **65**, 2740-2743.
12. Wolynes, P. G. (1992) in *Spin Glasses and Biology*, Stein, D., Ed., World Scientific Press, New York, pp. 225-259.
13. Goldstein, R. A., Luthey-Schulten, Z. A. & Wolynes, P. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4918-4922.
14. Goldstein, R. A., Luthey-Schulten, Z. A. & Wolynes, P. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9029-9033.
15. Goldstein, R. A., Luthey-Schulten, Z. A. & Wolynes, P. G. (1993) in *Proceedings of the 26th Annual Hawaii International Conference on System Sciences*, Vol. 1, Mudge, T. N., Milutinovic, V. & Hunter, L., Eds., IEEE Computer Society Press, Los Alamitos, pp. 699-707.
16. Šali, A., Shakhnovich, E. I. & Karplus, M. J. (1994) *J. Mol. Biol.* **235**, 1614-1636.
17. Šali, A., Shakhnovich, E. I. & Karplus, M. J. (1994) *Nature (London)* **369**, 248-251.
18. Chan, H. S. & Dill, K. A. (1994) *J. Chem. Phys.* **100**, 9238-9257.
19. Fukugita, M., Lancaster, D. & Mitchard, M. G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6365-6368.
20. Shakhnovich, E. I. & Gutin, A. M. (1993) *Protein Eng.* **6**, 793-800.
21. Shakhnovich, E. I. & Gutin, A. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7195-7199.
22. Shakhnovich, E. I. (1994) *Phys. Rev. Lett.* **72**, 3907-3910.
23. Shakhnovich, E. I. (1994) in *Protein Structure by Distance Analysis*, Bohr, H. & Brunak, S., Eds., IOS Press, Amsterdam, pp. 201-212.
24. Shakhnovich, E. I., Farztdinov, G., Gutin, A. M. & Karplus, M. (1991) *Phys. Rev. Lett.* **67**, 1665-1668.

25. Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. (1994) *J. Chem. Phys.* **101**, 6052–6062.
26. Honeycutt, J. D. & Thirumalai, D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3526–3529.
27. Honeycutt, J. D. & Thirumalai, D. (1992) *Biopolymers* **32**, 695–709.
28. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8721–8725.
29. Matthews, B. W. (1993) *Curr. Op. Struct. Biol.* **3**, 589–593.
30. Goldenberg, D. P. (1992) in *Protein Folding*, Creighton, T., Eds., W. H. Freeman and Co., New York, pp. 353–403.
31. Finkelstein, A. V. & Ptitsyn, O. B. (1987) *Prog. Biophys. Mol. Biol.* **50**, 171–190.
32. Finkelstein, A. V. (1994) *Curr. Biol.* **4**, 422–428.
33. Chan, H. S. & Dill, K. A. (1989) *Macromolecules* **22**, 4559–4573.
34. Chan, H. S. & Dill, K. A. (1990) *J. Chem. Phys.* **92**, 3118–3135.
35. Shakhnovich, E. I. & Gutin, A. M. (1990) *J. Chem. Phys.* **93**, 5967–5971.
36. Covell, D. G. & Jernigan, R. L. (1990) *Biochemistry* **29**, 3287–3294.
37. Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. USA* **70**.
38. Levitt, M. & Chothia, C. (1976) *Nature (London)* **261**, 552–557.
39. Kanehisa, M. I. & Tsong, T. Y. (1979) *Biopolymers* **18**, 1375–1388.
40. Kanehisa, M. I. & Tsong, T. Y. (1979) *Biopolymers* **18**, 2913–2928.
41. Gō, N., Abe, H., Mizuno, H. & Taketomi, H. (1980) in *Protein Folding*, Jaenicke, N., Ed., Elsevier, Amsterdam, pp. 167–181.
42. Gō, N. & Abe, H. (1981) *Biopolymers* **20**, 991–1011.
43. Gō, N. (1983) *Ann. Rev. Biophys. Bioeng.* **12**, 183–210.
44. Lesk, A. M. & Rose, G. D. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4304–4308.
45. Moult, J. & Unger, R. (1991) *Biochem.* **30**, 3816–3824.
46. Dill, K. A., Fiebig, K. M. & Chan, H. S. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1942–1946.
47. Bahar, I. & Jernigan, R. L. (1994) *Biophys. J.* **66**, 467–481.
48. Derrida, B. (1980) *Phys. Rev. Lett.* **45**, 79–82.
49. Derrida, B. (1981) *Phys. Rev. B* **24**, 2613–2626.
50. Dill, K. A. & Shortle, D. (1991) *Ann. Rev. Biochem.* **60**, 795–825.
51. Chan, H. S. & Dill, K. A. (1991) *J. Chem. Phys.* **95**, 3775–3787.
52. Yue, K. & Dill, K. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4153–4167.
53. Brändén, C. & Tooze, J. (1991) *Introduction to Protein Structure*, Garland Publishing, Inc., New York.
54. Roder, K., Elove, G. A. & Englander, S. W. (1988) *Nature (London)* **335**, 700–704.
55. Udgaonkar, J. B. & Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8197–8201.
56. Bycroft, M., Matouschek, A., Kellis, Jr., J. T., Serano, L. & Ferscht, A. R. (1990) *Nature (London)* **346**, 488–490.
57. Radford, S. E., Dobson, C. M. & Evans, P. A. (1992) *Nature (London)* **358**, 302–307.
58. Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H. & Clore, G. M. (1993) *Science* **260**, 1110–1113.