

Optimal Local Propensities for Model Proteins

Sridhar Govindarajan¹ and Richard A. Goldstein^{1,2}

¹Department of Chemistry and ²Biophysics Research Division, University of Michigan, Ann Arbor, Michigan 48109-1055

ABSTRACT Lattice models of proteins were used to examine the role of local propensities in stabilizing the native state of a protein, using techniques drawn from spin-glass theory to characterize the free-energy landscapes. In the *strong* evolutionary limit, optimal conditions for folding are achieved when the contributions from local interactions to the stability of the native state is small. Further increasing the local interactions rapidly decreases the foldability. © 1995 Wiley-Liss, Inc.

Key words: protein folding, lattice models, protein energetics, local interactions, spin-glass theory

INTRODUCTION

The relative importance of the various interactions stabilizing the native state of a protein is a matter of some dispute. In particular, there has been significant interest in trying to clarify the contributions made by local propensities and tertiary interactions, not only in determining the final folded state, but also in directing the folding process so that state can be achieved in a reasonable time-scale. Addressing this issue is of obvious importance in discriminating between various proposed models for protein folding dynamics. Studies on the role of local and tertiary interactions by Gō and co-workers show the essential nature of specific tertiary interactions in the stabilization of the native state.^{1, 2} Monte Carlo simulations of the folding process with tertiary interactions and a small amount of bias toward a native local structure have been shown to produce successful folding of proteins in adequate time.^{3–6} Zwanzig and co-workers claimed that proteins could fold rapidly if there is a rather large local bias towards the correct conformation.⁷ Karplus and Shakhnovich pointed out, however, that in such a case, protein folding would not be cooperative and a significant fraction of the population of proteins will not be in the ground state at reasonable temperatures.⁸ Dill and co-workers used simple lattice models with nonlocal interactions and helical propensities to show that the strength of the local propensities has to be small with respect to the tertiary interactions in order to obtain the average length and number of helices observed in biological proteins.^{9, 10} Dill and Gō and co-workers have sug-

gested that the limited success of the secondary structure prediction methods argues against the primacy of local propensities.^{2, 11} With such evidence for the primacy of tertiary interactions, it is then important to question whether folding is in fact enhanced by large local propensities.

In contrast to many studies that try to understand proteins by considering how they reflect the more generic properties of heteropolymers, we assume that it is difficult for proteins to be able to fold rapidly, resulting in significant evolutionary pressure for proteins to have properties that may be quite different from that of more random heteropolymers. In particular, we explore the limit where proteins have been optimized for rapid folding. While there is evidence that suggests that proteins have not been so optimized, this model can help to explain the folding process and rationalize the observed structure of biological proteins.¹² This is because there is a direct connection between the most optimizable structures and those that are most likely to arise through evolution, even in the absence of complete optimization. Here we seek to determine what combinations of interactions would give the most optimal folding behavior and how this behavior is modulated by changes in the magnitude of these interactions, in order to try to unravel the magnitude of local propensities versus tertiary interactions in the *strong* evolutionary limit.

Folding proteins must avoid the slow dynamics characteristic of rough energy landscapes. We characterize the roughness of the energy landscape using concepts applied to proteins originally borrowed from the physics of spin glasses.^{12–24} In this methodology, we consider that there are two possible phase transitions, the first to the folded state of the protein at a temperature T_f , the second to a glassy state at a temperature T_g . At temperatures close to T_g , the dynamics of the system become dominated by the roughness of the energy landscape, due to a lack of thermally accessible transition states between minima. In the thermodynamic limit, the sys-

Received January 19, 1995; revision accepted March 21, 1995.

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

tem becomes nonergodic at temperatures below T_g . Although there has been some indication that proteins do have glassy properties at low temperature,²⁵ in this model, the question is how do proteins evolve so as to *avoid* such glassy behavior. It has been observed that protein structures evolve much more slowly than protein sequences.^{26–30} The strong evolutionary limit involves the assumption that the sequence has evolved to optimize the interactions in the protein chain in order to facilitate rapid folding, by maximizing the ratio of T_f/T_g .

We can take advantage of the fact that for proteins confined to a lattice, we can do an exhaustive enumeration of all compact conformations for short proteins. This allows us to calculate the thermodynamic properties of these proteins exactly, assuming that the non-compact conformations are thermodynamically irrelevant. In particular, we can use the random energy model (REM) for determining how the values of T_f and T_g vary with the strength of the various interactions; and if the energy function is a linear function of a set of adjustable parameters, we can determine the values of those parameters that maximize T_f/T_g in closed form.^{12,22–24}

In this work we include both nonlocal and local interactions and optimize the interaction parameters to maximize the ratio T_f/T_g . We find that the sets of interactions with optimal foldability have quite modest contributions from local propensities. While T_f/T_g is somewhat increased by the addition of local propensities to the energy function, the relative optimizability of various structures does not appreciably change. Further increasing the magnitude of the local interactions causes a large decrease in T_f/T_g .

METHODS

We use a simple model where each protein consists of a chain of 36 monomers, confined to a 6×6 two-dimensional square lattice. The bonds are all of unit length, with adjacent residues existing at adjacent sites. There are a total of 114,674 self-avoiding walks for the 6×6 square lattice, not counting rotations. With the inclusion of energetic parameters representing angles, it is necessary to consider conformations and their mirror images separately.

The energy function consists of a simple contact form combined with an additional term that is a function of the local angle:

$$E = \sum_{i < j} \gamma_c(i, j) \delta_{ij} + \sum_i \gamma_d(i, \theta_i) \quad (1)$$

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 ϕ_i is the angle formed by residues $i - 1$, i , and $i + 1$, with possible values $\{-90^\circ, 0^\circ, +90^\circ\}$. $\gamma_c(i, j)$ is the adjustable parameter reflecting the energy contribution due to contact between residues i and j , while $\gamma_d(i, \phi_i)$ is the local propensity for residue i to have angle ϕ_i . The sets of

$\gamma_c(i, j)$ and $\gamma_d(i, \phi_i)$ are the adjustable parameters supposedly optimized through evolution. These energy parameters include entropic contributions due to the side chain and the solvent degrees of freedom.

Due to the nature of the lattice, the only contacts possible are between odd and even residues, so the total number of possible contacts is 289. As each compact conformation has exactly 25 contacts, and only relative energies matter, there is one fewer adjustable parameters for the contact energies. Likewise, for similar reasons, there are only two independent parameters for each of the 34 nonterminal residue angles, bringing the total number of adjustable parameters to 356.

The set of energy parameters was optimized for each compact conformation to maximize the ratio of T_f/T_g , using the method developed by Wolynes and co-workers to create optimal energy functions for tertiary structure prediction,^{22–24} later extended to the study of evolutionarily-optimized proteins by Govindarajan and Goldstein.¹² To summarize this approach, in the context of 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 energies of the ensemble of random states, Δ is the average energy difference between these states and the native state, and S_0 is the conformational entropy of the protein.²⁴ As T_f/T_g increases monotonically with \mathcal{R} , maximizing \mathcal{R} will therefore maximize T_f/T_g . As long as the energy is a linear function of the energetic parameters $\{\gamma_i\}$, as is true with the energy function in Eq. (1), we can express the energy of any target state, E_T , and the energy of any random state k , E_k , as

$$E_T = \sum_i \xi_i^T \gamma_i \quad (3)$$

$$E_k = \sum_i \xi_i^k \gamma_i \quad (4)$$

where ξ_i is equal to one if the interaction corresponding to γ_i is present in target structure T or k , and zero otherwise. In this case, Δ and Γ are given by

$$\Delta = \mathbf{A} \boldsymbol{\gamma} \quad (5)$$

$$\Gamma^2 = \boldsymbol{\gamma} \mathbf{B} \boldsymbol{\gamma} \quad (6)$$

with

$$A_i = \xi_i^T - \langle \xi_i^k \rangle_k \quad (7)$$

$$B_{i,j} = \langle \xi_i^k \xi_j^k \rangle_k - \langle \xi_i^k \rangle_k \langle \xi_j^k \rangle_k \quad (8)$$

where the averages are over all the k random states, in this work the total ensemble of compact conformations. The multiplication of the energetic parameters by a constant does not change the value of \mathcal{R} or

T_f/T_g , \mathcal{R} , and therefore T_f/T_g , is maximized for the set of energy parameters given by

$$\gamma = -\mathbf{B}^{-1}\mathbf{A} \quad (9)$$

Under these conditions, \mathcal{R}_{opt} , the value of \mathcal{R} with the optimal set of γ values, is given by

$$\mathcal{R}_{\text{opt}} = -\mathbf{A} \gamma \quad (10)$$

In addition to optimizing the interactions for all of the compact conformations, and calculating the relative contributions of local and tertiary interactions to the stabilization of the native state, optimizations were also carried out under constraints which fixed the relative strength of local interactions and non-local interactions. In general, it is not insightful to maximize T_f/T_g while constraining the value of $\phi = (E_T^{\text{loc}} - \langle E^{\text{loc}} \rangle_k) / (E_T - \langle E \rangle_k)$ where E^{loc} is the energy of the conformation due to local propensities only; even with $\phi = 0$, it is possible for large and cancelling local contributions to the ground state energy to increase \mathcal{R} through their influence on Γ . We instead optimized the interactions for fixed values on η , defined by

$$\eta^2 = \frac{\gamma_{\text{loc}} B_{\text{loc}} \gamma_{\text{loc}}}{\gamma_{\text{tert}} B_{\text{tert}} \gamma_{\text{tert}}} \quad (11)$$

where the subscripts loc and tert refer to the local propensities and tertiary contacts, respectively. This serves the purpose of modulating the total effect of the local interactions over the entire ensemble of structures. The relationship between η and ϕ is monotonic and well-behaved, and does not differ significantly between the various structures studied.

RESULTS

All the compact structures of the 36-mer confined to a 6×6 square lattice were enumerated, and the energetic parameters for the local and nonlocal interactions of the energy function of Eq. (1) were optimized for each structure using the spin-glass methodology described above. The distribution of \mathcal{R}_{opt} values for the optimized structures are shown in Figure 1. The results are compared with our previous study¹² where only the tertiary interactions were optimized and local interactions are ignored. As expected the \mathcal{R}_{opt} values increased with the number of adjustable parameters. The *relative* optimizability of the various structures did not change significantly, as shown in Figure 2: the highly optimizable structures are the same whether or not local interactions are included. There was no observable tendency for highly optimizable structures to be either more or less affected by the inclusion of local interactions than poorly optimizable structures.

The ϕ values, representing the percentage of the stabilization of the folded state due to local interactions, were calculated for all of the structures under the condition of optimized interactions. The distri-

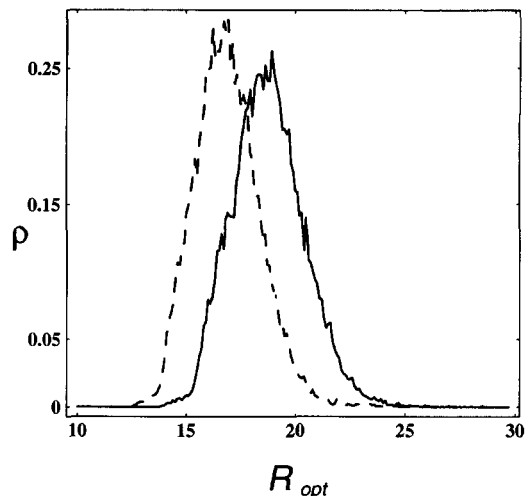


Fig. 1. The distribution of \mathcal{R}_{opt} values for the optimized structures with (—) and without (---) local interactions.

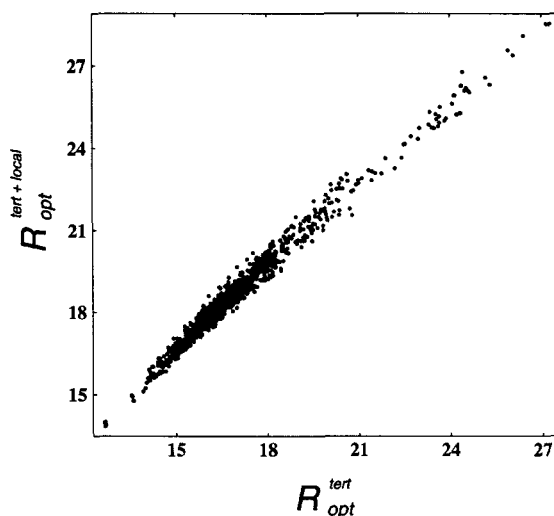


Fig. 2. Correlation plot for \mathcal{R}_{opt} values for the optimized structures with local and tertiary interactions ($\mathcal{R}_{\text{opt}}^{\text{tert+local}}$) and tertiary interactions only ($\mathcal{R}_{\text{opt}}^{\text{tert}}$).

bution of these ϕ among the various structures is shown in Figure 3. The contribution from the local propensities to the stability of the native state of the protein is small under optimal conditions, with a mean ϕ value of approximately 0.19. There were again no correlations observed between the \mathcal{R}_{opt} values and the fractional stability due to local interactions.

The dependence of the foldability on the extent of stabilization of the native state due to the local interactions was studied by optimizing the interactions under the constraint of a fixed value of η , defined by Eq. (11). Optimization of interaction parameters was carried out, for various structures,

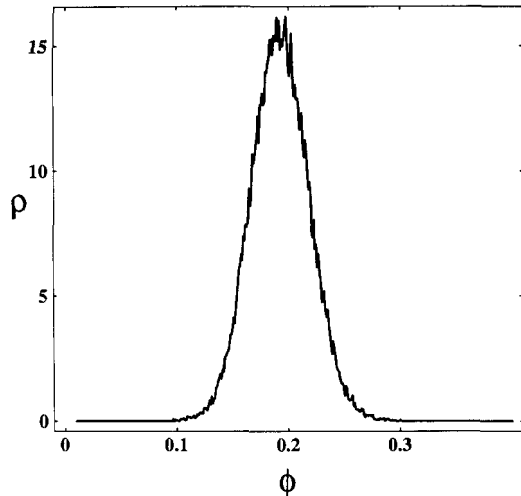


Fig. 3. The distribution of ϕ , the fractional stability of the native state due to local propensities, for various structures with optimized local propensities and tertiary interactions.

for several values of η between zero and infinity, using the Lagrange's method of undetermined multipliers. As the constraint is a nonlinear one, the solution was obtained numerically. For each set of interaction parameters, ϕ values were also calculated. Figure 4 shows the resulting variation of \mathcal{R} with ϕ for a number of different structures. As the strength of local propensities increases, the optimizability increases until $\mathcal{R} = \mathcal{R}_{\text{opt}}$ value is reached, and then \mathcal{R} decreases rapidly as the local interactions begin to dominate. In a previous paper, we showed how correlations between various energetic terms can affect the optimal set of interaction parameters and \mathcal{R}_{opt} .¹² When we neglect correlations between the tertiary contacts and local propensities, there is a small decrease in the \mathcal{R} values in general, but the dependence of \mathcal{R} on ϕ does not appreciably change.

DISCUSSION

In an attempt to explain how a protein folds into an unique structure from a random unfolded state in a short time several models have been developed. Some of these models emphasize the importance of local propensities for directing the folding process. Others involve more balance between the tertiary interactions and local propensities. We find, using our model, that the optimal folding interactions occur when the local contributions are small.

There are obviously many simplifications and approximations made in these calculations, involving both our model for the protein and our optimization procedure. It is important to understand how much the perspectives gained from these models can be extrapolated to biological proteins. We consider the nature of the lattice model for proteins first.

In order to be able to do the exhaustive enumera-

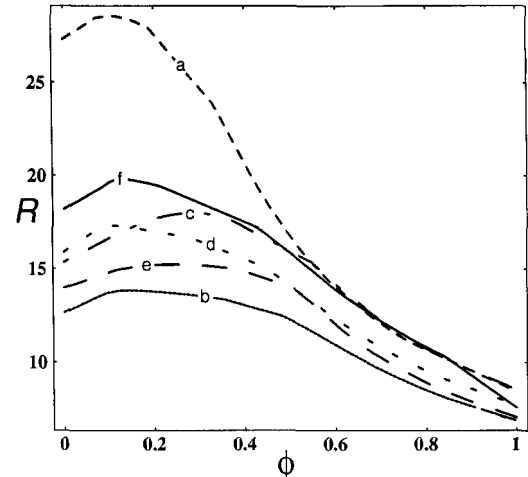


Fig. 4. \mathcal{R} values for optimized interactions, under a fixed strength of local propensities, as a function of ϕ , the fractional stability of the native state due to local propensities for (a) the most optimizable structure, (b) the least optimizable structure, (c) the structure with highest ϕ under optimized conditions without constraints, (d) the structure with lowest ϕ under optimized conditions without constraints, (e) a β meander, (f) packed two-dimensional α -helices.

tion for the 36-residue protein, we restricted our consideration to the compact conformations on the 6×6 two-dimensional square lattice. There is experimental evidence that the search for the native conformation takes place in a compact space, after an initial collapse of the unfolded chain into a compact one.³¹ In our previous work, we observed that qualitatively similar results were obtained by using an ensemble of random states generated by infinite-temperature Monte Carlo simulations of the protein confined to a semicompact space. Even so, the relative contributions of local and nonlocal interactions will be dependent on the set of random structures used. Using more noncompact conformations should decrease the importance of local interactions. If we assume that the various interactions are uncorrelated so that B is diagonal, then the optimal value of each native interaction is given by¹²

$$\gamma_i \approx -\frac{1}{\langle \xi_i^k \rangle_k} \quad (12)$$

so that native interactions that are less probable are in general stronger. As there are three possible values for each of the angles, the average local propensity will be approximately 3. With 25 contacts per compact conformation out of 289 possible contacts, the average interaction strength of any contact is approximately $289/25$, or about 11.6. With 34 angles and 25 contacts, that means that the fraction of the stability due to local interactions will be approximately $(34 \times 3)/(34 \times 3 + 25 \times 289/25) = 0.26$, close to what we observe with our calculations which

include correlations and the varying probabilities of various interactions in the random conformations. If noncompact conformations are included, that will make contacts more infrequent in the ensemble of random conformations, causing their magnitude to increase. The relative likelihood of the various angles will still remain approximately constant. As a result, the relative importance of the local interactions will decrease.

The various parameters of the model are appropriate for 36-residue two-dimensional lattice proteins. We can try to relate these parameters to biological proteins, again neglecting correlations. Again, the optimal interaction energy for each local conformation is inversely proportional to the probability of such a conformation being found in a random structure, i.e., $1/(1/k)$, where k is the total number of possible angles. The total energy due to local interactions in a N -residue protein would then be Nk . If each amino acid makes contacts with l other residues in random states, the probability of a given native contact being formed would be l/N , so the optimal strength of these interactions would be N/l . Assuming that the same number of contacts are formed in the folded state as in random states, the total stabilization energy due to the $Nl/2$ contacts would be $N^2/2$. The resulting ratio of the local interactions to the total stabilization energy, $\phi = 2k(2k + N)$ is plotted in Figure 5 with $k = 3$ (representing the three regions of angles populated in Ramachandran plots) and with $k = 10$ (representing the approximate number of local conformations derived from studies of entropy change during folding). As shown, local interactions are a rather small proportion of the stabilization energy; especially for larger proteins. Note that as long as the number of contacts in the folded state is equal to number of contacts in the random state, the optimal interaction strengths are independent of the actual number of contacts formed by the individual amino acids. More extended random states would further decrease the importance of local interactions by decreasing the probability of a native tertiary contact. This argument also neglects the correlations between tertiary contacts and local propensities, as well as the dependence of interaction probability on the location of the amino acids on the chain. In particular, α -helical proteins have larger numbers of contacts between residues close in the sequence, contacts that would be more likely in random conformations, compared with more β -sheet proteins. This suggests that local interactions would play a more prominent role in these α -helical proteins.

Another, deeper approximation of this method is that there is a connection between optimal interactions for folding and what is characteristic of biological proteins. As emphasized in our previous paper and mentioned above, there are many reasons to expect that proteins have not been optimized for fold-

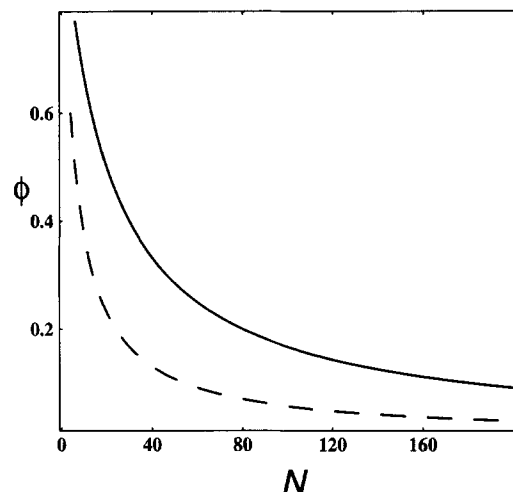


Fig. 5. The ratio of local interactions to total stabilization energy, for a simple model of biological proteins, ϕ , for $k=3$ (—) and $k=10$ (---), as a function of the number of residues N .

ing. On the other hand, proteins would have been more likely to occur through evolution in a form that is highly optimizable. As a tall mountain generally has a wider base than a short mountain, any random point in space is more likely to be on a taller mountain, especially if that point has to be at a relatively high elevation. In a similar way, highly optimizable structures, meaning structures with approximately the optimal ratio of local to tertiary interaction strengths, will be more likely to originate from evolution, as well as to be more robust to sequence mutation. 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 highly optimizable structures and the concomitant large range of interaction parameters resulting in foldability. This argument depends on the relative values of T_f and T_g in biological proteins and random amino acid sequences, an issue that is likely to be best resolved experimentally.

The last approximation involves the use of T_f/T_g , and through the use of the random energy model, \mathcal{R} , as a measure of the ability of a structure with a given set of interactions to fold. While the complete story must, of course, be more complicated, there is much evidence that dynamic measures of foldability are highly dependent on these thermodynamic measures. Wolynes and co-workers used this procedure to produce optimized energy functions for tertiary structure prediction.²²⁻²⁴ Shakhnovich and co-workers used a similar form of optimization procedure to produce sequences that would fold rapidly during Monte Carlo simulations.³²⁻³⁵ Chan and Dill and Karplus and co-workers found that a criterion similar to \mathcal{R} distinguished foldable and nonfoldable sequences.³⁶⁻³⁸

What we have demonstrated in this paper is that, for these model systems, optimal folding is obtained by limited energetic contributions from local propensities. This result is in agreement and supports other work that has reached similar conclusions through completely different analyses. Conversely, such an analysis can be used to justify and rationalize the dominance of tertiary interactions given the independent evidence of modest local propensities, and provide insight into the nature of protein evolution, biological optimization, and biological proteins.

ACKNOWLEDGMENTS

We thank Kurt Hillig for computational assistance. Financial support was provided by the College of Literature, Science, and the Arts, the Program in Protein Structure and Design, and the Horace H. Rackham School of Graduate Studies.

REFERENCES

- Go, N., Taketomi, H. Respective roles of short- and long-range interactions in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 75:559–563, 1978.
- Go, N., Abe, H., Mizuno, H., Taketomi, H. Local structures in the process of protein folding. In: "Protein Folding." Jaenicke, N., ed. Amsterdam: Elsevier, 1980:167–181.
- Skolnick, J., Kolinski, A. Simulations of the folding of a globular protein. *Science* 250:1121–1125, 1990.
- Skolnick, J., Kolinski, A. Dynamic Monte Carlo simulations of a new lattice model of globular protein folding, structure, and dynamics. *J. Mol. Biol.* 221:499–531, 1991.
- Godzik, A., Skolnick, J., Kolinski, A. Simulations of the folding pathway of triose phosphate isomerase-type $\alpha\beta$ barrel proteins. *Proc. Natl. Acad. Sci. U.S.A.* 89:2629–2633, 1992.
- Rey, A., Skolnick, J. Computer modeling and folding of four-helix bundles. *Proteins* 16:8–28, 1993.
- Zwanzig, R., Szabo, A., Bagchi, B. Levinthal's paradox. *Proc. Natl. Acad. Sci. U.S.A.* 89:20–22, 1992.
- Karplus, M., Shakhnovich, E. Protein folding: Theoretical studies. In: "Protein Folding." Creighton, T., ed. New York: W. H. Freeman, 1992:127–195.
- Dill, K. A., Fiebig, K. M., Chan, H. S. Cooperativity in protein-folding kinetics. *Proc. Natl. Acad. Sci. U.S.A.* 90:1942–1946, 1993.
- Thomas, P. D., Dill, K. A. Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation. *Protein Sci.* 2:2050–2065, 1993.
- Dill, K. A. Dominant forces in protein folding. *Biochemistry* 29:7133–7155, 1990.
- Govindarajan, S., Goldstein, R. A. Searching for foldable protein structures using optimized energy functions. *Biopolymers*, 36:43–51, 1995.
- Bryngelson, J. D., Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 84:7524–7528, 1987.
- Bryngelson, J. D., Wolynes, P. G. A simple statistical field theory of heteropolymer collapse with application to protein folding. *Biopolymers* 30:171–188, 1990.
- Garel, T., Orland, H. Mean-field model for protein folding. *Europhys. Lett.* 6:307–310, 1988.
- Garel, T., Orland, H. Chemical sequence and spatial structure in simple models of bio-polymers. *Europhys. Lett.* 6:597–601, 1988.
- Shakhnovich, E. I., Gutin, A. M. The nonergodic (spin-glass like) phase of heteropolymers with quenched disordered sequence of links. *Europhys. Lett.* 8:327–332, 1988.
- Shakhnovich, E. I., Gutin, A. M. Protein folding as pattern recognition. *Studia Biophys.* 132:47–56, 1989.
- Shakhnovich, E. I., Gutin, A. M. Formation of unique structure in polypeptide chains: Theoretical investigation with the aid of a replica approach. *Biophys. Chem.* 34:187–199, 1989.
- Sasai, M., Wolynes, P. G. Molecular theory of associative memory hamiltonian models of protein folding. *Phys. Rev. Lett.* 65(21):2740–2743, 1990.
- Wolynes, P. G. Spin glass ideas and the protein folding problem. In: "Spin Glasses and Biology." Stein, D., ed. New York: World Scientific Press, 1992:225–259.
- Goldstein, R. A., Luthey-Schulten, Z. A., Wolynes, P. G. Optimal protein folding codes form spin glass theory. *Proc. Natl. Acad. Sci. U.S.A.* 89:4918–4922, 1992.
- Goldstein, R. A., Luthey-Schulten, Z. A., Wolynes, P. G. Protein tertiary structure recognition using optimized hamiltonians with local interactions. *Proc. Natl. Acad. Sci. U.S.A.* 89:9029–9033, 1992.
- Goldstein, R. A., Luthey-Schulten, Z. A., Wolynes, P. G. Protein tertiary structure recognition using optimized associative memory hamiltonians. In: "Proceedings of the 26th Annual Hawaii International Conference on System Sciences," Vol. 1. Mudge, T. N., Milutinovic, V., Hunter, L., eds. Los Alamitos: IEEE Computer Society Press, 1993: 699–707.
- Frauenfelder, H., Sligar, S. G., Wolynes, P. G. The energy landscape and motions of proteins. *Science* 254:1598–1603, 1991.
- Chothia, C., Lesk, A. M. The relationship between the divergence of sequence and structure in proteins. *EMBO J.* 5:823–826, 1986.
- Hubbard, T. J. P., Blundell, T. L. Comparison of solvent-inaccessible cores of homologous proteins: Definitions useful for protein modeling. *Protein Eng.* 1:159–171, 1987.
- Sander, C., Schneider, R. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* 9:56–68, 1991.
- Flores, T. P., Orengo, C. A., Moss, D. S., Thornton, J. M. Comparison of conformational characteristics in structurally similar protein pairs. *Protein Sci.* 2:1811–1826, 1993.
- Aronson, H. G., Royer, W. E., Thornton, J. M. Quantification of tertiary structural conservation despite primary sequence drift in globin fold. *Protein Sci.* 3:1706–1711, 1994.
- Ptitsyn, O. B. The molten globule state. In: "Protein Folding." Creighton, T., ed. New York: W. H. Freeman, 1992: 243–300.
- Shakhnovich, E. I., Gutin, A. M. A new approach to the design of stable proteins. *Protein Eng.* 6:793–800, 1993.
- Shakhnovich, E. I., Gutin, A. M. Engineering of stable and fast-folding sequences of model proteins. *Proc. Natl. Acad. Sci. U.S.A.* 90:7195–7199, 1993.
- Shakhnovich, E. I. Proteins with selected sequences fold into unique native conformation. *Phys. Rev. Lett.* 72: 3907–3910, 1994.
- Shakhnovich, E. I. Design of model fast-folding proteins. In: "Protein Structure by Distance Analysis." Bohr, H., Brunak, S., eds. Amsterdam: IOS Press, 1994:201–212.
- Sali, A., Shakhnovich, E. I., Karplus, M. J. Kinetics of protein folding: A lattice model study of the requirements for folding to the native state. *J. Mol. Biol.* 235:1614–1636, 1994.
- Sali, A., Shakhnovich, E. I., Karplus, M. J. How does a protein fold. *Nature (London)* 369:248–251, 1994.
- Chan, H. S., Dill, K. A. Transition states and folding dynamics of proteins and heteropolymers. *J. Chem. Phys.* 100:9238–9257, 1994.