

Competition between protein folding and aggregation: A three-dimensional lattice-model simulation

D. Bratko^{a)} and H. W. Blanch

Department of Chemical Engineering, University of California at Berkeley, Berkeley, California 94720-1462

(Received 7 August 2000; accepted 11 October 2000)

Aggregation of protein molecules resulting in the loss of biological activity and the formation of insoluble deposits represents a serious problem for the biotechnology and pharmaceutical industries and in medicine. Considerable experimental and theoretical efforts are being made in order to improve our understanding of, and ability to control, the process. In the present work, we describe a Monte Carlo study of a multichain system of coarse-grained model proteins akin to lattice models developed for simulations of protein folding. The model is designed to examine the competition between intramolecular interactions leading to the native protein structure, and intermolecular association, resulting in the formation of aggregates of misfolded chains. Interactions between the segments are described by a variation of the Go potential [N. Go and H. Abe, *Biopolymers* **20**, 1013 (1981)] that extends the recognition between attracting types of segments to pairs on distinct chains. For the particular model we adopt, the global free energy minimum of a pair of protein molecules corresponds to a dimer of *native* proteins. When three or more molecules interact, clusters of *misfolded* chains can be more stable than aggregates of native folds. A considerable fraction of native structure, however, is preserved in these cases. Rates of conformational changes rapidly decrease with the size of the protein cluster. Within the timescale accessible to computer simulations, the folding-aggregation balance is strongly affected by kinetic considerations. Both the native form and aggregates can persist in metastable states, even if conditions such as temperature or concentration favor a transition to an alternative form. Refolding yield can be affected by the presence of an additional polymer species mimicking the function of a molecular chaperone.

© 2001 American Institute of Physics. [DOI: 10.1063/1.1330212]

I. INTRODUCTION

The biological function of a protein or a peptide is intimately related to its conformation, or tertiary structure, determined by complex interactions among protein residues.¹ Varying degrees of residue hydrophobicity, hydrogen-bonding groups, and the presence of partial charges result in specific potentials among the different residue pairs. Competition between these interactions and topological constraints imposed by chain connectivity and rigidity leads to the characteristic conformational behavior of native proteins. The conformational energy spectra of protein molecules display an energy gap between the continuous part and a small set of low-energy conformations.² These conformations or folds optimize the binding between attracting residues while minimizing contacts between constituents with unfavorable interactions. The native form of a protein is believed to correspond to the global minimum on the complex free energy landscape of a single chain. However, this situation can change³⁻⁵ in the presence of additional protein molecules. Forces among residues from distinct polymers offer opportunities for favorable couplings, which can lead to the formation of intermolecular clusters or aggregates. In this process, termed abnormal protein aggregation,⁴ they acquire a different tertiary structure incompatible with the biological

function of the protein. To a lesser extent, the secondary structure is also modified. At sufficient concentration, aggregation results in protein precipitation.⁵

Protein aggregation leading to reduced biological activity and insoluble or poorly soluble forms is of great concern in a variety of biotechnological processes and pharmaceutical applications, including the formation and renaturation of inclusion bodies, the formulation of protein drugs, and the aging and storage of protein drugs. In addition, the formation of aggregated states is associated with a number of diseases, well-known examples including the Alzheimer's disease, Bovine Spongiform Encephalopathy (BSE), or mad cow disease, and amyotrophic lateral sclerosis (ALS). The enormous impact of these disease states and the growing number of protein drugs have given rise to intense research of both equilibrium and dynamic aspects of protein aggregation, with the ultimate objective of learning to manipulate the protein chemistry and system conditions in ways that will preclude or slow down the process.

In the past few years, a number of groups have analyzed the aggregation mechanism using computer modeling employing a lattice representation. Dynamic Monte Carlo simulations of two-dimensional multichain models with proteins represented as two-letter copolymers,⁶⁻⁹ as well as Monte Carlo^{10,11} and enumeration studies³ of 20-letter models of protein dimers in three dimensions have been reported. Most recently, Giugliarelli *et al.*¹² designed an exactly solvable

^{a)}Electronic mail: db@lolita.cchem.berkeley.edu

model for the phase behavior of a concentrated heteropolymer solution. Calculations for a two-letter square-lattice representation captured transitions between monomeric or associated folds and misfolded prionlike aggregates. While each of these works has enhanced our understanding of aggregation principles, a *combination* of a multichain simulation and the more realistic, three-dimensional representation is clearly desirable for a more detailed analysis. In the present article, we describe Monte Carlo simulations of systems comprising several proteinlike heteropolymer chains on a cubic lattice. The model is designed to study the competition between refolding and aggregation of denatured chains. For this purpose, we consider situations with initially denatured model proteins and monitor the relaxation toward either of the two relatively stable forms: an assembly of folded chains, and clusters of partly misfolded polymers stabilized by intermolecular contacts. We aim to characterize the structure of the aggregated form in terms of the extent of native structure preserved upon association.

Besides the question as to which form represents the actual stable state, we are interested in kinetic factors that can favor the formation of either of the two forms. Dynamic Monte Carlo simulations are employed to consider aggregation kinetics. To mitigate computational difficulties associated with a remarkably increased number of degrees of freedom and slowed configurational relaxation, we rely on the simplest form of pair interactions known to capture the essential physics of the folding process in the case of an isolated protein. We monitor the behavior of concentrated 27-mers with Go and Abe¹³ residue-residue pair potentials that stabilize only bonds between the types of residues that are in contact in the native protein conformation. Standard segment moves,¹⁴ supplemented by chain translations,⁶ are used to consider conformational evolution for a group of adjacent, initially unfolded model protein molecules. In Sec. II, we describe the details of the model and simulation. In Sec. III, we present simulation results for isolated model proteins and for systems containing up to six protein chains per simulation box. Results of these calculations provide insights into the energetics of interprotein association, characterized by increased ruggedness of the free energy landscape. For three or more chains, the global minimum involves partially misfolded chains. The conformation dynamics of a cluster of adjacent molecules displays features of glasslike behavior characterized by trapping of the system in either of a multitude of the local free-energy minima that correspond to distinct aggregate configurations. Once in the folded form, however, the model proteins display considerable stability against aggregation pointing to the existence of significant activation barriers between the two favorable forms. Finally, we examine the effect of introducing an additional proteinlike species that mimics the action of a molecular chaperone. In Sec. IV, our findings are summarized and directions of future work are outlined.

II. MESOSCOPIC SIMULATIONS OF AGGREGATING PROTEIN SYSTEMS

The mechanism of aggregation is considered by Monte Carlo simulations in an ensemble of coarse-grained model

proteins. Mesoscopic models have proved extremely helpful in studies of protein folding, where they continue to provide valuable insights into both mechanics and dynamics of the transition to the native conformation.^{2,3,14–25} Advances in computational techniques and facilities are making similar simulations possible in multichain systems.^{6,9} In what follows, we describe an on-lattice simulation study of model protein solutions comprising several chains, placed on a cubic lattice with step length equal to the segment length of the protein. Chain segments can represent individual amino-acid residues, or statistical segments comprising several residues.^{6,9} Due to excluded volume interactions, each lattice site can accommodate only one protein segment. Empty sites are assumed to be occupied by the solvent that is not explicitly considered in the model. Instead, we treat segment-segment interactions as McMillan–Mayer potentials²⁶ obtained by integration over degrees of freedom associated with solvent molecules. As usual in studies of folding of a single protein, we assume the residue-residue potentials to be pairwise additive and short-ranged, effectively limited to nearest neighbor interactions. The energy function of a system comprising M chains of length N is of the form

$$U(\{\mathbf{r}\}) = \sum_{m=1}^M \sum_{i=1}^{N-2} \sum_{j=i+2}^N U_{ij} \Delta(\mathbf{r}_{i,m} - \mathbf{r}_{j,m}) + \sum_{m=1}^{M-1} \sum_{n=m+1}^M \sum_{i=1}^N \sum_{j=1}^N U_{ij} \Delta(\mathbf{r}_{i,m} - \mathbf{r}_{j,n}), \quad (1)$$

where subscripts $m, n = 1, 2, 3, \dots, M$ denote a chain, and subscripts $i, j = 1, 2, 3, \dots, N$ specify the contour positions of chain segments. The configuration of the system is completely described by a $3MN$ -dimensional vector $\{\mathbf{r}\}$ that comprises all segment positions $\mathbf{r}_{i,m}$. Function $\Delta(\mathbf{r}_{i,m} - \mathbf{r}_{j,n})$ equals unity if segments (i,m) and (j,n) are in contact and zero otherwise. Intrachain interactions among consecutive segments ($|i-j|=1$) are excluded from $U(\{\mathbf{r}\})$. Infinite energy is ascribed to overlapping configurations. Chain composition determines coupling constants, $U_{i,j}$, between segments at contour positions i and j . Equation (1) is easily generalized to a mixture of chains differing in lengths and segment sequences, where coupling constants $U_{i,m;j,n}$ depend on four indices specifying contour positions of the two interacting segments (i,j) and the identities of the two chains, (m,n) . For simplicity and to secure relatively fast conformational dynamics, we employed the inter-residue pair potentials of Go and Abe,¹³ with all *types* of segments that form native contacts being ascribed an equally strong, attractive interaction, $U_{ij} = U_0 < 0$. Apart from excluded volume, interactions among other residue pairs are assumed to be negligibly weak. While the teleological Go potential lacks a clear physical interpretation for the presumed recognition effects, it has helped in obtaining many valuable insights into folding mechanisms²⁵ and it appears natural to examine its applicability in studies of multichain events. For convenience, we absorb the coupling strength and the temperature in the reduced temperature, $T^* = k_B T / U_0$. In these units, the interaction of native residue-residue contacts, $U_{ij} / k_B T$, equals $-1/T^*$. Interchain potentials are calculated using the

same pair potentials, U_{ij} , that determine intrachain interactions for identical residue pairs. The complete interaction map, $\{U_{ij}/k_B T\}$, is therefore determined by the temperature, T^* , and the native structure of the model protein. According to our operational definition, a protein is in the native structure when all inter-residue distances, r_{ij} , in the protein acquire exactly prescribed values. By symmetry, a mirror image of the original structure is stabilized by identical potentials. This degeneracy does not appear relevant to the aggregation behavior of our model system since it does not enhance opportunities for favorable interactions among protein folds.

To account for a finite concentration of the solution, we employed Minimum Image (MI) boundary conditions^{26,27} that treat every segment as a center of a cubic Monte Carlo box of size L , interacting with each of the remaining segments or its nearest image located at a distance not exceeding $L/2$ in any of the three directions. While satisfactory for studies of interactions among MN nearest segments, the method suppresses density fluctuations at length scales exceeding L . Applications of open ensemble techniques would be needed to capture large-scale phenomena such as the phase separation or protein precipitation.

Equilibrium properties are calculated as canonical averages over configurations $\{\mathbf{r}\}$ that we compute using two different approaches. In one, we employed a dynamic Monte Carlo method with Metropolis acceptance criteria,²⁷ which samples polymer configurations as a sequence of states generated by performing standard types of segment moves.^{14,28} Attempted moves include the displacement of either of the end beads in the chain to one of the four available sites next to the nearest contour neighbor in the chain, corner flips of beads whose bonds with nearest contour neighbors form the right angle, and crankshaft moves that involve a simultaneous displacement of the second and the third bead of a four-member loop by loop rotation around the axis through the first and the fourth bead of the loop. The above moves have been found satisfactory in numerous studies of folding of a single chain.^{2,14,15} To facilitate the translation of the chains as a whole, we also included collective translation moves^{6,9} in which all segments of a randomly selected chain, or several chains,⁹ undergo a parallel translation along a random direction. Only translations of length equal to a single lattice step at a time were attempted. In addition, the chains were allowed to undergo reptations corresponding to slithering snakelike moves along the chain contour.²⁹ While relatively efficient in early stages of runs initiated with all chains in random configurations, the acceptance of reptation moves is very low in compact states with partly or completely folded chains and aggregated states and hence did not significantly contribute to conformational evolution. Once a move type was selected, one of the MN segments was picked randomly. A move of the segment, loop, or chain to which the segment belonged was then attempted when consistent with local chain configuration, and rejected otherwise. The different move types were attempted at random with different *a priori* probabilities: 25% for single-particle moves, 50% for crankshaft moves, 20% for attempted chain translations, and 5% for attempted reptations. These choices

follow from empirical considerations and should, in principle, have no effect on thermodynamic averages. The choice of move probabilities will, however, affect the apparent dynamics of the system. Only qualitative comparisons between simulated dynamics and real time kinetics are therefore possible.

In addition to the dynamic simulation, we also performed a number of runs using a version of nondynamic, Ensemble Growth Monte Carlo technique (EGMC) from earlier works.^{30,31} This method determines configurational statistics by averaging over a large ensemble of systems grown from monomers to a final number of particles. During the growth of the ensemble, the systems multiply in competition with each other in accordance with Metropolis criteria, which secure the canonical distribution of states at all stages of the growth.³² A particular advantage of this technique is that it produces valid results for different system sizes at all steps of the growth procedure. A deficiency of the method, however, is a rapid increase in the ensemble size required for satisfactory statistics as the number of chains increases. This increase results from the degrees of freedom introduced upon the addition of new chains whose positions have to be adequately sampled over the whole volume. Bigger ensemble must therefore be used, leading to a rapid increase in computer memory we need. The method turned out to be very efficient in studies of single chain folding, where its results coincide with those from dynamic simulations. Increased computational requirements, however, made it hard to apply the technique in multichain systems without prior refinements of the algorithm. For this reason, and because of our interest in kinetic properties, in the present article we discuss only results obtained from dynamic Monte Carlo simulations.

The pair potentials $\{U_{ij}\}$ used in our calculations correspond to the native structure of the model 27-mer depicted in Fig. 1. Three additional native structures, including the structure presented in Fig. 1 of Ref. 14 were also used in a set of calculations. All four native structures corresponded to the most compact, cubelike form with 28 native contacts among the residues not linked by covalent bonds ($|i-j|>1$). The runs performed using pair potentials $\{U_{ij}\}$ for the three alternative native structures revealed qualitatively identical behavior as those corresponding to the structure presented in Fig. 1. Only results for this case are presented in the next section. A systematic study of the effect of sequence mutations³³ and more realistic choices of inter-residue potentials^{34–36} will be described at a later date.

III. RESULTS AND DISCUSSION

Although we are primarily interested in the interchain effects, we first characterize the behavior of a single model protein that we use as a reference for comparison with multichain systems. In this and the following examples, we visualize the distribution over structural states and their evolution by relying on a set of order parameters; specifically, we determine the numbers, n_{nc} , or fractions, $f_{nc} = n_{nc}/n_{nc}^{\max}$, of native contacts among protein segments. We also determine the probability of the completely folded state, f_n . In mul-

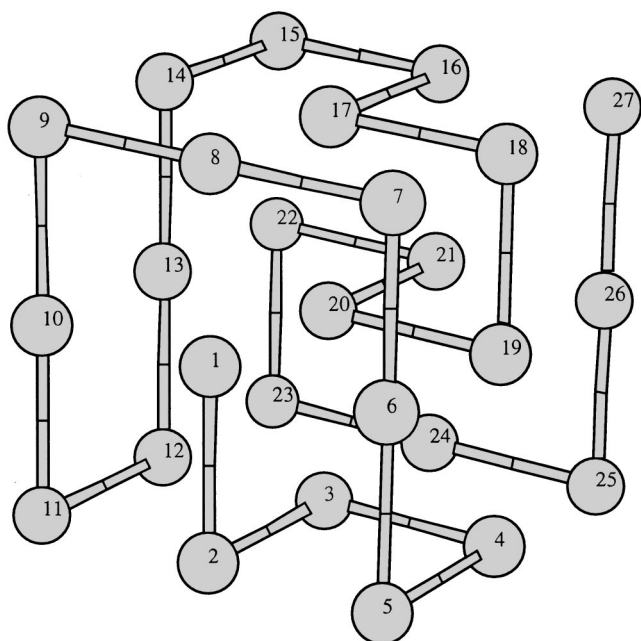


FIG. 1. The native structure of a model protein.

tichain systems, the numbers of intermolecular contacts, n_i , are also monitored. In this way, we obtain simultaneous measures for the resemblance to the native form, and for the degree of interprotein association. Figure 2 illustrates the temperature dependence of the average structure for a single model protein. It shows the temperature variation of the fraction of time a protein spends in its native form, f_n . At low reduced temperature, $T^* \sim 0.4$ ($U_0/k_B T = -2.5$), f_n is close to 95%. As we increase T^* , a slow decrease in f_n is observed at first. Upon a further rise in the temperature, the slope gradually increases with rapid unfolding observed in the interval $0.6 < T^* < 0.75$. The native form is practically extinct above $T^* \sim 0.8$. The temperature dependence of the fraction of native contacts, f_{nc} , is similar to that for the

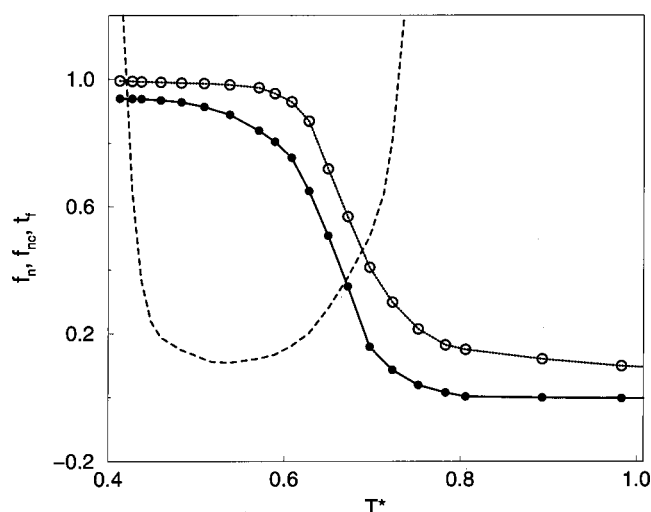


FIG. 2. Fraction of native contacts, f_{nc} (open circles), fraction of time the protein spends in the native state, f_n (solid symbols), and average refolding time, t_f (dashed) as functions of the reduced temperature, T^* . The minimal refolding time at $T^* \sim 0.56$ corresponding to $\sim 8 \times 10^3$ passes.

probability of the native form of the protein as a whole, f_n , however, the fraction of native contacts, f_{nc} , always exceeds f_n and approaches a nonzero limit at high T^* . A temperature (or interaction strength) interval of interest for our present purposes has to include the range where both the native and denatured forms are well represented. Ideally, it should also provide for rapid transition kinetics. In Fig. 2, the average refolding time of an isolated model protein from a random initial conformation is shown as a function of reduced temperature, T^* . The minimum at $T^* \sim 0.56$ ($U_0 = -1.8$) corresponds to the optimal compromise between two opposing effects: the increased role of activation barriers at low temperatures, and the shift of the equilibrium structure toward the denatured form at high temperatures. In view of higher activation barriers encountered in processes of aggregation, the optimal temperature we choose for studies of multichain systems is somewhat higher. $T^* = 0.69$ was used in most of our calculations for multichain systems.

The speed of structural rearrangements, including the rate of folding or refolding, will generally decrease with increasing opportunities for intersegment interactions. At finite protein concentrations, interactions among the segments from distinct protein molecules give rise to additional low-energy domains of configurational space, $\{\mathbf{r}\}$, and contribute to potential barriers between them. In Table I, we list the mean first passage times (in simulation passes), i.e., the average times it takes to fold M chains, each comprising 27 segments, from initially random configurations to a completely folded state. In Table I we present results for four different volume fractions $\phi = 5\%$, 10% , 16% , 21% , and 32% , corresponding to 1, 2, 3, 4, and 6 chains per simulation box of size $L = 8$ and volume $V = L^3$ at $T^* = 0.69$.

Relaxation of initially random systems is of interest since it involves passing through intermediate, partly folded states believed to represent the critical stage from which adjacent polymers can either evolve to folded chains with weak interprotein interactions, or form an aggregate of misfolded molecules. In view of the apparent stability, or metastability (especially pronounced at lower T^*), of the aggregates or fully folded chains, the intermediate stage is believed to represent the point where the system behavior can be influenced by relatively mild intervention. For the chosen temperature, $T^* = 0.69$, and moderate concentrations, $\phi \sim 10\% - 16\%$, we observe a trend toward reversible association, while chains in highly concentrated systems often lock-in to either of the many favorable aggregated states. To illustrate the kinetics of these processes, in Fig. 3 we present the time dependencies of the numbers of native intrachain contacts, n_{nc} , inter-

TABLE I. Average refolding times, t_f , fractions of native and interprotein contacts, n_{nc}/n_{nc}^{\max} and n_i/n_{nc}^{\max} , for systems containing M protein chains of length N in a periodic simulation box of size $L = 8$, corresponding to different volume fractions $\phi = MN/L^3$.

M	1	2	3	4	6
ϕ (%)	5	10	16	21	32
t_f (10^3 passes)	30	67	147	$\sim 10^3$	$> (7 \times 10^6)$
n_{nc}/n_{nc}^{\max}	0.62	0.69	0.71	0.81	0.82
n_i/n_{nc}^{\max}	...	0.10	0.23	0.33	0.39

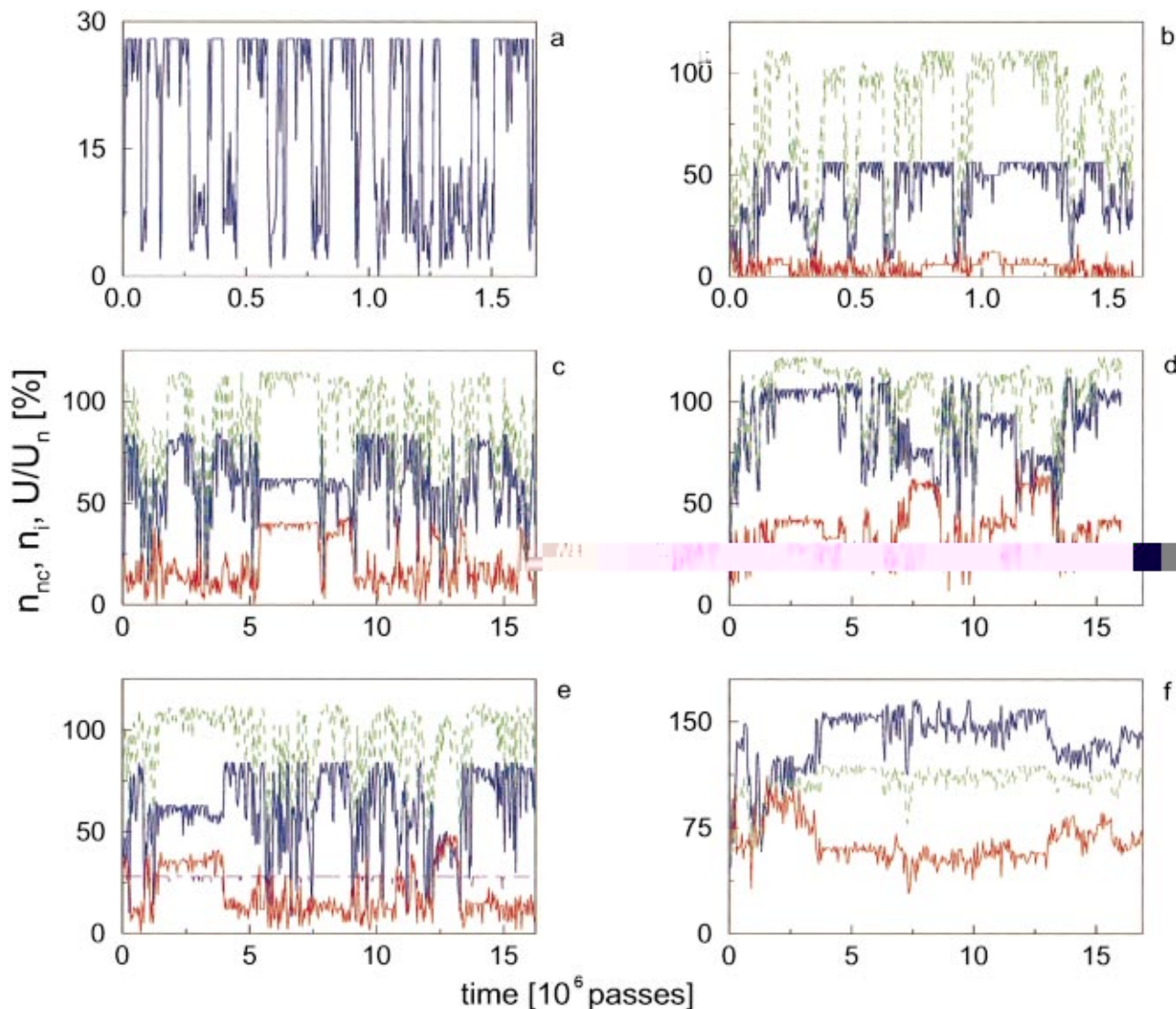


FIG. 3. (Color) Number of native contacts, n_{nc} , for protein molecules (blue) or chaperonelike species (indigo, dot-dashed), number of interprotein contacts, n_i , (red), energy relative to a completely folded unassociated system, U/U_n (green) for systems with a single protein molecule and volume fraction $\phi = 5\%$ (a), two molecules per simulation box, $\phi = 10\%$ (b), three molecules, $\phi = 16\%$ (c), four molecules, $\phi = 21\%$ (d), three protein molecules ($\phi = 16\%$), and a chaperonelike molecule (e), and six protein molecules per box, $\phi = 32\%$ (f), all at $T^* = 0.69$.

protein bonds, n_i , and the total energy of simulated chains normalized with respect to the minimum energy of M isolated folded polymers, U/U_n . In each of the runs shown, the initial state was a random configuration of denatured molecules.

Figure 3(a) shows the number of native contacts, n_{nc} , for a single protein molecule as a function of the simulation time in a typical run. For the energy function we use, $U/k_B T$ equals n_{nc}/T^* , and there are no interprotein bonds. The protein spends similar amounts of time in the low energy native ($n_{nc} = 28$) or nearly native state, and in the almost entirely unfolded form, ($n_{nc} < 10$), which has high energy but favorable conformational entropy. Intermediate states between the two extremes are rare and short-lived, showing that folding and refolding have characteristics of barrier crossing events. This behavior is consistent with a recent interpretation of the enthalpy distribution and cooperatively effects for essentially identical model situations.²⁵ According to Fig. 2, lowering the temperature, or equivalently, increasing the bonding en-

ergy U_0 , shifts the equilibrium toward the folded form, while a concomitant increase in activation barriers between the two forms also reduces the frequency of unfolding events. An example of a much more stable structure (corresponding to $U_0/k_B T = -1.9$) is given in Fig. 3(e). At $T^* = 0.69$, Fig. 3(a), a typical lifetime of either the folded or an unfolded state is of the order of 10^5 passes (one pass corresponds to a cycle of N attempted moves, where N is the number of segments in the simulation box). This time is appreciably longer than the refolding time from an entirely random initial (state shown in Fig. 2) due to the presence of stabilizing contacts (typically five to ten) present in the denatured form.

In Fig. 3(b), we show the time dependencies of the number of native contacts, n_{nc} , interchain contacts, n_i , and the reduced energy, U/U_n , for a typical run in a two-chain system ($\phi \sim 10\%$) at the same temperature. Here, U_n is the energy of fully folded isolated chains. In this system, through dimerization of folded molecules, the presence of a neigh-

boring protein stabilizes the native form increasing its lifetime by about a factor of 2–3 at the expense of misfolded states. This behavior resembles the association-enhanced folding observed experimentally in certain systems.³⁷ Low populations of the transition states between the two extreme forms, again, point to significant free-energy barriers between the native and the considerably unfolded form, but short-lived, partly folded aggregates are also detected. Model chains can form up to six favorable intermolecular bonds per dimer and still remain completely folded. According to Fig. 3(b) the number of interchain contacts rarely exceeds six. The energy curve in Fig. 3(b) shows that the global minimum of the system corresponds to a dimer of completely folded chains, $n_{nc}=56$. This behavior reflects the absence of distinct homodimeric native states characteristic of model prions³ for the sequence we use. The same holds true for the remaining model sequences that we have considered (not shown). We subsequently show that the global energy minimum can correspond to an aggregated state as the number of interacting molecules is increased *above* two. Clearly, our observations may be somewhat affected by the particular choice for the intersegment potentials. While the nature and shortcomings of the Go potential¹³ have been reviewed in the context of protein folding,²⁵ less is known about its suitability for studies of interchain effects. Calculations with alternative model potentials³⁴ will be needed to assess the generality of our findings.

Figure 3(c) describes the behavior of a three-chain system ($\phi \sim 16\%$) over an essentially longer time of observation (exceeding 16 million passes). This was necessary because of notably slowed dynamics of the system. Here, long-lived aggregates persist over prolonged time intervals characterized by a significantly reduced number of native contacts. These contacts are replaced by an approximately equivalent number of intermolecular bonds. Because of this trade-off, the total energies of the two forms do not differ

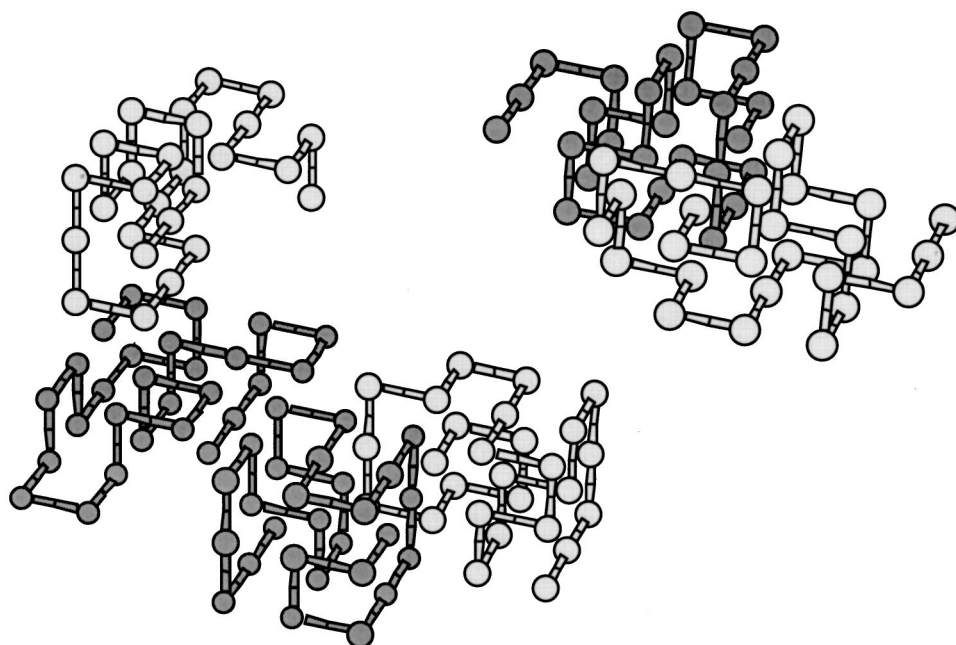


FIG. 4. A snapshot of a misfolded six-chain system shown in Figs. 3(f) and 5(f).

times for refolding presented in Table I, the correlation times display a similar increase with ϕ . Unlike the reported refolding times of single proteins,³⁸ the size dependence of the mean refolding times, t_f , and correlation times, t_c , in our multichain systems does not follow a simple scaling relation. Despite a limited statistical accuracy, the trends revealed by estimated relaxation rates indicate that glasslike behavior with a multitude of metastable states is approached upon increasing the number of associated polymers.

Following the initial decay, $c_s(t)$ sometimes displays oscillations associated with repeated conformational rearrangements. These oscillations average out when considering several runs, or by sufficiently increasing the simulation length.

When present, the oscillations in $c_s(t)$ are accompanied by compensating variations in $c_c(t)$. This observation is consistent with the perpetual trade-off between intra- and intermolecular bonds while allowing relatively small temporal variations in the total number of bonds and concomitant energy.

In an attempt of influencing the refolding yield, we have considered the effect of introducing an additional molecular species that might affect the propensity toward the native state of model proteins, especially at elevated concentrations. For the sake of simplicity, the added species that was intended to mimic the function of a chaperone molecule, had the same native structure as the remaining protein molecules, but with about 31% stronger intramolecular interactions. The

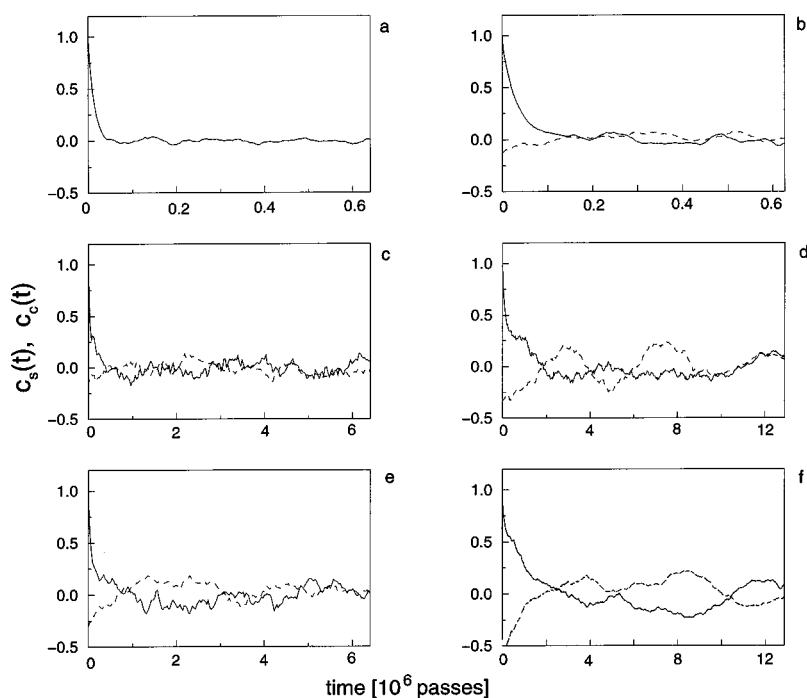


FIG. 5. Self- (solid line) and cross-correlation functions (dashed lines) for the numbers of native contacts, or native and interprotein contacts, respectively. (a)–(f) correspond to the same systems as in Fig. 3.

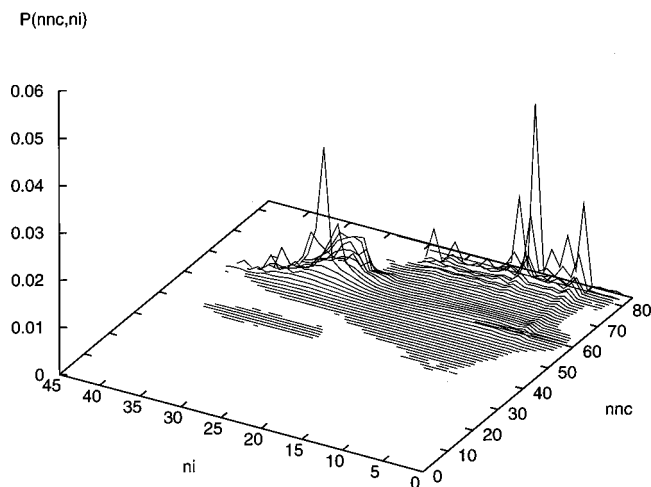


FIG. 6. Population densities of states of a three-chain system ($\phi = 16\%$) in the order parameter space (n_{nc}, n_i) . n_{nc} and n_i are the numbers of native and interprotein contacts, respectively. $n_{nc} = 84$ corresponds to fully folded proteins and the peak located at $n_{nc} \sim 68$ and $n_i \sim 35$ corresponds to aggregates of misfolded proteins.

additional strength was sufficient to keep these chains in the native state, allowing only small and short-lived fluctuations around it during any length of the simulation. The molecule is expected to promote folding of an adjacent model protein by nonspecific shielding from other, partly unfolded intermediates, and by serving as a structural template as we have shown that dimerization can stabilize the native form through favorable surface binding. Peptide binding to hydrophobic sites of the apical domain of GroEL provides an example of a similar interaction.³⁹ Figure 3(e) depicts a system containing three proteins from previous examples, plus a strongly folded molecule of the type described above. The nearly horizontal line in Fig. 3(e) shows the number of native contacts for the second species. Interprotein residue-residue interactions were of the same strengths as the intramolecular interactions in the original protein, i.e., $U_0 = -1/T^*$ for residue pairs of the types forming native contacts, and 0 otherwise. A comparison between the fractions of preserved native contacts of protein molecules considered in Fig. 3(e), and in the one-component protein system with an identical protein concentration illustrated in Fig. 3(c), reveals some improvement in the refolding yield upon the addition of the second component. The effect is further illustrated in Figs. 6 and 7, where we compare population densities, $P(n_{nc}, n_i)$, in the order parameter space, $\{n_{nc}, n_i\}$, for two typical systems containing three protein molecules in the absence, and in the presence, of a chaperone-like molecule. Out of 15 pairs of repeated runs, a comparison of folded population densities showed an improvement in the refolding yield upon the introduction of the second species in 11 cases. Population densities $P(k, l)$ are defined as

$$P(k, l) = \frac{\int d\{\mathbf{r}\} P(\{\mathbf{r}\}) \delta(k - n_{nc}(\{\mathbf{r}\})) \delta(l - n_i(\{\mathbf{r}\}))}{\sum_{m_{nc}} \sum_{m_i} \int d\{\mathbf{r}\} P(\{\mathbf{r}\}) \delta(m_{nc} - n_{nc}(\{\mathbf{r}\})) \delta(m_i - n_i(\{\mathbf{r}\}))} \quad (5)$$

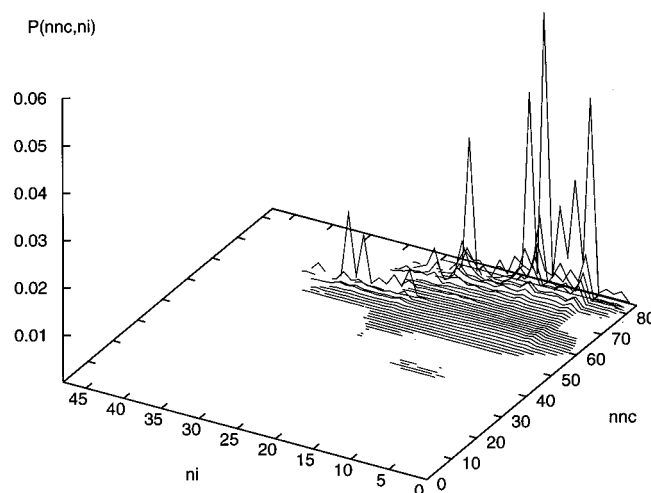


FIG. 7. Population densities of states of a four-chain system comprised of three proteins and a chaperone-like molecule in the order parameter space (n_{nc}, n_i) .

In view of slow configuration dynamics, the calculated populations do not correspond to the true canonical probabilities $P(\{\mathbf{r}\}) \propto \exp(-U(\{\mathbf{r}\})/k_B T)$ for given energy landscape, $U(\{\mathbf{r}\})$. During accessible simulation times, a typical system samples native or nearly folded states along with a limited number of favorable domains corresponding to aggregates of partly misfolded chains. The general pattern of presented populations has been reproduced in many repeated runs (not shown) for different concentrations. An important characteristic observed in the majority of these runs was the lack of any significant populations for intermediate states corresponding to the transition paths *between* the native and aggregated forms. The position of the aggregate peak or peaks, however, varied from run to run revealing the presence of a multitude of comparable but separated free energy minima located within the aggregated regime. The use of the Go potential¹³ characterized by pronounced cooperativity effects²⁵ may partly contribute to the low intermediate populations. However, a limited set of calculations we performed using a two-letter heteropolymer model³⁰ with both attractive and repulsive interactions revealed a very similar qualitative behavior.

The above observations point to an aggregation-induced crossover to a glasslike behavior of concentrated model proteins, a notion affirmed in a limited number of runs made at lower temperatures (or using stronger residue-residue interactions). For multichain systems at temperatures below $T^* \sim 0.57$, initially established states, corresponding to either the native form or to aggregates of misfolded chains, had a tendency to resist further changes. Computational restrictions, however, prevented a more systematic analysis of temperature variation within the frame of the present work.

IV. CONCLUDING REMARKS

A simple lattice model for protein folding has been extended to simulate protein association in multichain systems. Attractive interactions among the residues from neighboring chains significantly alter both the structural and dynamic be-

havior of the solution. We consider a regime where single proteins exist in an equilibrium with the denatured form. At finite protein concentrations, we observe the formation of disordered clusters of partly misfolded chains that represent an onset of abnormal protein aggregation. Interestingly, the aggregates retain and even stabilize a significant fraction of native residue–residue contacts while forming a smaller number of interprotein bonds. In the two-chain system, a dimer of native folds represents the most stable form while global energy minimum corresponds to an aggregated state when more than two chains are involved.

The association of protein molecules results in an increased ruggedness of the free energy landscape and concomitant slowing of configuration dynamics. This slowdown is escalated with increasing concentration and cluster size suggesting an extrapolation to a glasslike regime, even at relatively high temperatures. At low temperatures, the system tends to persist in either of the initially occupied free energy wells resulting in the apparent stability of the existing folded or aggregated form. The refolding yield of initially denatured proteins can be affected by the presence of an additional protein component that loosely resembles the function of a molecular chaperone. Using experimentally derived residue–residue potentials,^{34,35} the effects of protein sequence and its mutations will be considered in the following works.

The separation of time scales between elementary segment moves and global structural rearrangements possess a serious challenge in protein simulations. The situation is aggravated upon increasing association. Hierarchical treatments, along with advanced simulation techniques^{23,40} and algorithms designed for studies of rare, barrier crossing events²⁷ appear promising for future studies.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. BES-9901054.

¹Y. A. R. Leach, *Molecular Modeling* (Longman, United Kingdom, 1996), p. 453.

²A. Sali, E. I. Shakhnovich, and M. Karplus, *Nature* (London) **369**, 248 (1994).

³P. M. Harrison, H. S. Chan, S. B. Prusiner, and F. E. Cohen, *J. Mol. Biol.* **286**, 593 (1999).

⁴A. L. Fink, *Folding Des.* **3**, R9 (1998).

⁵N. Asherie, J. Pande, A. Lomakin, O. Ogun, S. R. A. Hanson, J. B. Smith, and G. B. Benedek, *Biophys. Chem.* **75**, 213 (1998).

- ⁶P. Gupta, C. K. Hall, and A. C. Voegler, *Protein Sci.* **7**, 2642 (1998).
- ⁷P. Gupta, C. K. Hall, and A. Voegler, *Fluid Phase Equilibria* **160**, 87 (1999).
- ⁸S. Istrail, R. Schwartz, and J. King, *J. Comput. Biol.* **6**, 143 (1999).
- ⁹L. Toma and S. Toma, *Biomacromolecules* **1**, 232 (2000).
- ¹⁰R. A. Broglia, G. Tiana, S. Pasquali, H. E. Roman, and E. Vigezzi, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12 930 (1998).
- ¹¹R. A. Broglia, G. Tiana, S. Pasquali, H. E. Roman, and E. Vigezzi, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10 943 (1999).
- ¹²G. Giugliarelli, C. Micheletti, J. R. Banavar, and A. Maritan, *J. Chem. Phys.* **113**, 5072 (2000).
- ¹³N. Go and H. Abe, *Biopolymers* **20**, 1013 (1981).
- ¹⁴A. Šali, E. Shakhnovich, and M. Karplus, *J. Mol. Biol.* **235**, 1614 (1994).
- ¹⁵H. S. Chan and K. A. Dill, *Phys. Today* **46**, 24 (1993).
- ¹⁶J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, *Proteins* **21**, 167 (1995).
- ¹⁷K. A. Dill, S. Bromberg, S. Yuc, K. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, *Protein Sci.* **4**, 561 (1995).
- ¹⁸B. H. Park and M. Levitt, *J. Mol. Biol.* **249**, 493 (1995).
- ¹⁹D. Thirumalai, *J. Phys. (France)* **5**, 1455 (1995).
- ²⁰E. I. Shakhnovich, *Folding Des.* **1**, R50 (1996).
- ²¹H. A. Scheraga, *Biophys. Chem.* **59**, 329 (1996).
- ²²N. D. Socci, J. N. Onuchic, and P. G. Wolynes, *J. Chem. Phys.* **104**, 5860 (1996).
- ²³H. S. Chan and K. A. Dill, *Proteins: Struct., Funct., Genet.* **30**, 2 (1998).
- ²⁴V. S. Pande, A. Y. Grosberg, and T. Tanaka, *Rev. Mod. Phys.* **72**, 259 (2000).
- ²⁵H. Kaya and H. S. Chan, *Proteins* **40**, 637 (2000).
- ²⁶H. L. Friedman, *A Course in Statistical Mechanics* (Prentice–Hall, Englewood Cliffs, NJ, 1985).
- ²⁷D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic, New York, 1996).
- ²⁸A. Baumgartner, “Simulations of macromolecules,” in *The Monte Carlo Method in Condensed Matter Physics*, edited by K. Binder (Springer-Verlag, Berlin, 1992), p. 285.
- ²⁹F. T. Wall and F. Mandel, *J. Chem. Phys.* **63**, 4592 (1975).
- ³⁰D. Bratko, A. K. Chakraborty, and E. I. Shakhnovich, *Phys. Rev. Lett.* **76**, 1844 (1996).
- ³¹D. Bratko, A. K. Chakraborty, and E. I. Shakhnovich, *J. Chem. Phys.* **106**, 1264 (1997).
- ³²P. G. Higgs and H. Orland, *J. Chem. Phys.* **95**, 4506 (1991).
- ³³G. Tiana, R. A. Broglia, H. E. Roman, E. Vigezzi, and E. I. Shakhnovich, *J. Chem. Phys.* **108**, 757 (1998).
- ³⁴S. Miyazawa and R. L. Jernigan, *Macromolecules* **18**, 5340 (1985).
- ³⁵O. Keskin, I. Bahar, A. Y. Badretdinov, O. B. Ptitsyn, and R. L. Jernigan, *Protein Sci.* **7**, 2578 (1998).
- ³⁶R. I. Dima, G. Sattanni, C. Micheletti, J. R. Banavar, and A. Maritan, *J. Chem. Phys.* **112**, 9151 (2000).
- ³⁷V. N. Uversky, D. J. Segel, S. Doniach, and A. L. Fink, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5480 (1998).
- ³⁸K. Dimitrievski, B. Kasemov, and V. P. Zhidanov, *J. Chem. Phys.* **113**, 883 (2000).
- ³⁹A. M. Buckle, R. Zahn and A. R. Fersht, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3571 (1997).
- ⁴⁰H. S. Chan and K. A. Dill, *J. Chem. Phys.* **100**, 9238 (1994).