

MONTE CARLO APPROACHES TO THE PROTEIN FOLDING PROBLEM

JEFFREY SKOLNICK

*Department of Molecular Biology, The Scripps Research Institute, La Jolla,
CA 92037*

ANDRZEJ KOLINSKI

Department of Chemistry, University of Warsaw, Warsaw, Poland

CONTENTS

- I. Introduction
- II. Protein Representation Force Field, and Sampling Protocols
 - A. Sampling Protocols
 - 1. General Considerations
 - 2. Monte Carlo Sampling Methods
 - 3. Use of Dynamic Monte Carlo to Simulate Protein Dynamics
 - B. Simple, Exact Models
 - C. Protein-Like Models
 - 1. Lattice Models of Protein-Like Systems
 - 2. Continuous Models of Protein-Like Systems
 - D. Models of Real Proteins
 - 1. Continuous Models
 - 2. Discretized, Lattice Models
 - 3. Complementary Insights from Various Models
- III. Protein Folding Thermodynamics
 - A. Nature of the Interactions
 - B. Extent of Frustration in Native Conformations
 - C. Origin of Cooperativity in Protein Folding and Two-State Thermodynamic Behavior in Protein Folding
- IV. Tertiary Structure Prediction
 - A. Peptides and Cyclic Peptides
 - B. Proteins

Advances in Chemical Physics, Volume 105, Monte Carlo Methods in Chemical Physics, edited by David M. Ferguson, J. Ilja Siepmann, and Donald G. Truhlar. Series Editors I. Prigogine and Stuart A. Rice.
ISBN 0-471-19630-4 © 1999 John Wiley & Sons, Inc.

1. De Novo Approaches
2. Prediction Using Known Secondary Structure and Correct Tertiary Restraints
3. Method for Prediction of Tertiary Structure from Predicted Secondary Structure and Tertiary Restraints
- V. Quaternary Structure Prediction
 - A. Coiled Coil, De Novo Simulations
 - B. Method for Prediction of Equilibrium Constants of Multimeric Helical Proteins
- VI. Simulation of Peptide Insertion into Membranes
 - A. Small Proteins and Peptides
 - B. Translocation across Model Membranes
- VII. Perspectives
 - A. Where Are We Now?
 - B. Shortcomings of the Present Force Fields, Possible Refinements, and Reformulations
 - C. Promises of New Sampling Methods
- VIII. Conclusions

I. INTRODUCTION

Under the appropriate solvent and temperature conditions, globular proteins adopt a well-defined three-dimensional structure [1-3] that is capable of performing a specific biochemical or structural task not only in a living cell [4] but also in vitro. This unique three-dimensional structure is encoded in the protein amino acid sequence. The question of how to relate the sequence to the native structure is commonly referred to as the protein folding problem [5,6]. It is widely believed that proteins obey the "thermodynamic hypothesis" that says that the native conformation of a protein state corresponds to a global free-energy minimum, which is dictated by the various interactions present in a system comprised of proteins and solvent [7,8]. Unfortunately, due to the complexity of the interactions within proteins and between proteins and the surrounding solvent, the task of finding this free-energy minimum in the myriad of multiple minima in the free-energy landscape [9,10] is extremely difficult.

While molecular dynamics (MD) simulations have proven to be very powerful for studying numerous aspects of protein dynamics and structure [11-13], this technique cannot yet access the millisecond-to-second time-scales required for folding even a small protein. To address this timescale gap, one has to simplify the protein model by reducing the number of degrees of freedom. Such approaches assume that the basic physics could be reproduced in model systems that employ "united atoms" and effective solvent models. On the basis of recent work, it has become apparent that the crux of the solution to the protein folding problem does not lie in whether a reduced protein model is used, but rather in the development of

potentials that can recognize a native-like state from a misfolded state, and techniques that can explore the relevant regions of conformational space [14-20]. In this review chapter, we describe approaches that address these difficulties, and focus in particular on Monte Carlo based approaches to conformational sampling.

II. PROTEIN REPRESENTATION, FORCE FIELD, AND SAMPLING PROTOCOLS

Various levels of simplification of the protein's structure and energy surface have been assumed. As a result, different aspects of the protein folding problem can be addressed by computer simulations. Somewhat arbitrarily, we divide protein models into three classes: simple, exact models; protein-like models; and realistic protein models. Roughly speaking, the level of structural detail taken into account increases in parallel with the classification scheme. At the same time, the resulting simulations become more and more expensive, and analysis of the particular model is less comprehensive. In what follows, we start from the description of simple, exact models that can be studied in great detail. Then, we outline studies of somewhat more complex models of protein-like systems. Finally, realistic protein models are described. Since this issue is devoted to the application of Monte Carlo methods in chemical physics, we focus our attention on studies that employed Monte Carlo techniques as the conformational sampling protocol. However, for completeness, models studied by means of different sampling methods, but which are amenable to investigation by a Monte Carlo-based methodology, will be also mentioned.

A. Sampling Protocols

1. General Considerations

Exploration of the conformational space of protein models could be done using different computational techniques. These include MD [21], Brownian dynamics [22,23], Monte Carlo methods [24-27], and other simulation or optimization techniques such as genetic algorithms [25,28-31].

2. Monte Carlo Sampling Methods

To address the conformational sampling problem, reduced protein models use techniques that allow for a faster and more efficient exploration of protein conformational space than can be provided by MD. Here, various Monte Carlo techniques are commonly used [32,33]. In a well-designed Monte Carlo sampling method, relatively large conformational changes occur in a single iteration of the simulation algorithm; therefore, the numerous local minima of the energy surface are more easily surmounted.

The well-known Metropolis Monte Carlo (MMC) procedure randomly samples conformational space according to the Boltzmann distribution of (distinguishable) conformations [34]:

$$P_i = \exp\left(\frac{-E_i}{kT}\right) \quad (2.1a)$$

In order to generate this distribution in asymmetric MMC, the transition probability $P_{i,j}$ from an "old" conformation i to a "new" conformation j is controlled by the energy difference $\Delta E_{ij} = E_j - E_i$ via

$$p_{i,j} = \min\left\{1, \exp\left(\frac{-\Delta E_{ij}}{kT}\right)\right\} \quad (2.1b)$$

Obviously, this technique is very sensitive to the presence of energy barriers. To ensure adequate sampling, typically a collection of elemental moves involving end moves and collective motions of two to four bonds are performed randomly. In addition, small distance motions of a large, randomly selected part of the chain are employed. The key to a successful dynamic Monte Carlo protocol is to include a sufficiently large move set so that no element of structure is artificially frozen in space.

To enhance the sampling efficiency, Hao and Scheraga have employed the entropy-sampling Monte Carlo method (ESMC) in their study of simplified protein models [35-37]. ESMC was originally proposed by Lee [38] in the context of a simple Ising model, which is closely related to the multi-canonical MC technique of Berg and Neuhaus [39]. Since the formulation of Hao and Scheraga [35-37] is the most straightforward, we briefly review their approach; additional details are found in the chapter "Entropy Sampling Monte Carlo for Polypeptides and Proteins." Unlike MMC, ESMC generates an artificial distribution of states that is controlled by the conformational entropy as a function of the energy of a particular conformation E_i :

$$p_i^{\text{ESMC}} = \exp\left(\frac{-S(E_i)}{k}\right) \quad (2.2a)$$

The transition probability can be formally written as

$$p_{i,j}^{\text{ESMC}} = \min\left\{1, \exp\left(\frac{-\Delta S_{i,j}}{k}\right)\right\} \quad (2.2b)$$

where $\Delta S_{i,j}$ is the entropy difference between energy levels i and j , respectively. At the beginning of the simulation, the entropy is unknown;

however, from a density-of-states energy histogram, $H(E)$, an estimate, $J(E)$, for the entropy $S(E)$ can be iteratively generated. After a sufficient number of runs, all states are sampled with the same frequency. Then, the histogram of $H(E)$ becomes flat, and the curve of $J(E) + \text{constant}$ approaches the true $S(E)/E$ curve. ESMC offers the advantage that one can discern whether the simulation has converged over the sampled energy range, and, if so, one has directly determined the free-energy of the system (within a constant). However, a potentially serious problem with ESMC is its computational cost. Kolinski et al. [40] have described a means of increasing the rate of convergence of the simulation. The basic idea is to first run a standard dynamic Monte Carlo folding simulation to provide a library of "seed" structures, and then, periodically, randomly select one of these conformations and introduce it into the conformational pool used to construct the $S(E)/E$ curve. In practice, this technique works quite well.

3. Use of Dynamic Monte Carlo to Simulate Protein Dynamics

As demonstrated by Orwoll and Stockmayer [41], a proper sequence of small, random conformational changes of lattice polymers constitutes a solution of a stochastic equation of motion. Of course, the same move sets could be generalized to off-lattice models. However, in all cases, the problem of ergodicity has to be taken into consideration. This problem could be especially dangerous for low-coordination-number lattices. With a high-coordination-number lattice and correspondingly large number of allowed local conformational transitions, the risk of a serious ergodicity problem becomes negligible [42]. A properly designed scheme for local conformational transitions allows simulations of the long-time dynamics of model polymeric systems [43-45], with results consistent with MD [46].

The design of an efficient model of Monte Carlo dynamics for off-lattice models is a nontrivial task. When using all-atom models with the standard semiempirical force fields, the local conformational transitions usually introduce very small changes in chain geometry and are usually uncorrelated with the nature of the energy surface. Consequently, the efficiency of such MC algorithms is not much better than standard MD. An effective set of local moves has to be able to pass over (or rather to neglect) the intervening narrow-width, local energy maxima (which may be of considerable height). Such "smart" moves have to take into account and exploit specific correlations of rotational degrees of freedom [medium-range phi-psi (ϕ - ψ) correlations] over some portion of the polypeptide backbone. One such algorithm has been recently elaborated on by Elofsson et al. [31]. An even more radical approach to local moves has been proposed by Knapp et al. [25,26,30], who assumed a rigid-body representation of peptide bond plates and developed an algorithm that allows for the collective motion of a few

peptide units. The moves can interpolate between deep energy minima for ϕ - ψ correlated potentials. When the tertiary interactions are neglected, the model reproduces Rouse dynamics [47,48] of denatured polypeptide chains, that is, the correct long-time dynamics of a polymer in the free draining limit [25,26,30]. Of course, long-range interactions can be introduced, and the model could be employed in folding studies [49]. Efficient local continuous-space deformations of polypeptide chains also have been investigated by Baysal and Meirovitch [50]. These developments are expected to extend the applicability of Monte Carlo off-lattice models in computational studies of protein dynamics and protein folding processes.

Lattice Monte Carlo dynamics of proteins could be studied at various levels of generalization. The dynamics of simple lattice models and equivalent continuous protein-like models led to a reasonable approximation of the main features of the denatured state and gives some insight into protein folding pathways [51-53]. More detailed models can provide additional information on the effect of short-range interactions on polypeptide dynamics at various temperatures, and consequently at various distances from the native state. It has been shown that a high-coordination-lattice discretization of protein conformational space [16,17,20,40,42] leads to Rouse-like dynamics [48] in the denatured state [54]. It was possible to investigate the effect of the short- and long-range cooperativity on the dynamic properties of denatured proteins [55]. Also, the dynamics near the transition state of medium-resolution protein models have been analyzed [16,42,56]. These experiments indicate that high-resolution models could be a powerful tool for model studies of the long-time dynamics of real proteins. The approach is an alternative to the off-lattice models of protein long-time dynamics.

In summary, the recent developments allow Monte Carlo simulations of protein dynamics in their denatured state, in the intermediate states and in the folded state. In the latter case, where the fine details are of major interest, standard MD techniques are usually superior to the Monte Carlo reduced model approaches.

B. Simple, Exact Models

Simple, exact models [51] consist of simple lattice polymers or heteropolymers where each amino acid in the hypothetical protein is represented by a bead occupying a single lattice point. Because of their simplicity, these models have generated considerable attention [57,58].

Short homopolymeric chains restricted to a simple square lattice and to a simple cubic lattice have been studied by Chan and Dill [59-61]. For very small systems, all compact conformations could be enumerated. Not surprisingly, significant intrachain entropy loss occurs on collapse. Expand-

ed conformations could be sampled by a simple Monte Carlo scheme employing various local and global conformational transitions [62]. Chan and Dill suggested that some short-range backbone correlations, similar to protein secondary structure, may be induced solely due to chain compactness [59,60]. This postulate was later questioned [55,63-66]. Also, an earlier computer study of the collapse transition in long flexible lattice homopolymers did not indicate any local ordering of the chain segments [44]. Interestingly, it has been shown that the collapse transition of semiflexible polymers leads to substantial and cooperative local ordering in the compact state [43,67,68].

Much could be learned from computer studies of simple, exact heteropolymeric models of proteins. A classic example is the HP model introduced by Dill and co-workers [61,69,70]. In this model, a simple cubic lattice chain consists of two kinds of beads, H (P), corresponding to non-polar hydrophobic (polar) residues. Nonbonded nearest neighbor contacts of two hydrophobic (HH) residues contribute a constant negative energy. HP and PP contacts are assumed to be inert. A similar model (AB model) has been investigated by Shakhnovich and co-workers [71,72]. In this case, all residues are attractive; however, the energy of attractive interactions was lower for pairs of the same type (AA or BB) than for different residue pairs (AB). Most simulations were done on 27-mer systems with the target structure being a $3 \times 3 \times 3$ compact cube.

Both models have similar properties and mimic some general features of globular proteins. Not all sequences undergo a collapse transition to a well-defined compact state (neglecting the existence of mirror-image structures that emerge as a result of lack of chiral energy terms in these simple models). This is in accord with the idea that only a small fraction of possible amino acid sequences form well defined globular structures [70,72-76]. Moreover, only a small fraction of HP or AB sequences are "good folders." In cases of such "good folders," the collapse transition is very cooperative. Various sequences of simple, exact models exhibit different folding pathways [76-78]. The folding pathways may suggest (on a very general level) how real proteins reduce their conformational space during the folding process [79-82].

Recent studies of larger-size, simple cubic lattice heteropolymers [83] show that with increasing protein size, the folding pathway becomes more complex. In these models, the initial hydrophobic collapse stage has been followed by a finer structure fixation involving the rearrangement of the surface residues. Some signatures of such a two-stage folding mechanism also have been observed in 27-mer AB- or HP-type systems [81,84,85]. These larger-model systems permit a more meaningful albeit, due to lattice restrictions, still very biased modeling of protein secondary structure.

Longer cubic lattice heteropolymeric chains may also mimic some aspects of multidomain protein folding [86].

Camacho and Thirumalai [87] investigated the effect of strong covalent interactions (disulfide bonds) on the folding pathway of a minimal lattice model. They found that the folding rate is fastest when a proper balance is achieved between effective hydrophobic interactions and the strength of disulfide bonds. This rationalizes how weak denaturing conditions can accelerate the folding process of real crosslinked proteins.

Finally, it should be pointed out that simple, exact models could be used to test new sampling methodologies [88-90] or optimization techniques [84,91]. A good example is the work by Hao and Scheraga [92], where pairwise contact energy parameters have been optimized to maximize the stability of the native state with respect to other conformations for a 27-mer chain composed of 10 different residues types [92].

The tolerance of the protein folding process to noise in the interaction parameters and sequence modifications [93] was also studied using a $3 \times 3 \times 4$ target structure [94]. The conclusions were in agreement with the qualitative picture of real protein folding dynamics and thermodynamics [93]. Such studies are important because they may provide some insights into how specific the protein sequence has to be in order to maintain the structural uniqueness of the native conformation.

It is also possible to design a simple, exact model with side chains [95]. As would be expected, the presence of side chains increases the entropy difference between folded and unfolded states. The possible effects of somewhat more complex and cooperative interactions have also been investigated [96].

Extensive studies of simple, exact models have provided new insights into general principles governing protein folding dynamics, thermodynamics, and the relationship between protein sequence and protein structure. The importance of hydrophobic collapse in the initial stage of the protein folding process is clearly visible in a majority of the preceding studies. As in real proteins, foldable sequences in simple, exact models exhibit specific amino acid patterns of hydrophobic and hydrophilic residues. The sequences of these models control not only the stability of the globular state but also the dynamics of folding. Very likely these and other important generalizations of computer studies stimulated the development of new theoretical descriptions of the protein folding process [58,85,97-102].

C. Protein-Like Models

The simple, exact models described in the previous section employ nearest-neighbor interactions as the only amino acid sequence specific contribution to the conformational energy. Consequently, the dominance of nonpolar

tertiary interactions in protein folding is implicitly assumed in these models. Short-range (i.e., local down the chain) interactions are a priori neglected, or more precisely, they implicitly arise from inherent restrictions imposed by the use of a very low-coordination-number lattice. In this section, we describe a class of more complex, protein-like models that tries to address the interplay between secondary and tertiary interactions, a feature characteristic of real proteins. Thus, they account for a wider variety of interactions that control protein folding dynamics, thermodynamics, and the structural uniqueness of the native state. They also attempt to more realistically mimic aspects of the geometry of the major secondary structure motifs that occur in proteins. However, these are not faithful models of real proteins because they lack the ability to reproduce the detailed chain geometry and packing of supersecondary structural elements. In this respect, such protein-like models are similar to simple, exact models, but here the secondary structure and interaction scheme are better defined. We separately review lattice and continuous space models of such protein-like systems.

1. Lattice Models of Protein-Like Systems

In their pioneering investigations, Go and co-workers [103-105] investigated the dynamics and thermodynamics of square lattice chains as highly idealized protein-like systems. Three types of interactions were considered: short-range secondary structure conformational propensities, long-range contact potentials (similar to that employed in simple, exact models), and hydrophobic potentials. The sampling scheme consisted of various types of conformational transitions controlled by the Metropolis Monte Carlo algorithm. The results of these and other studies [106,107] showed that long-range interactions consistent with the assumed target structure increase folding cooperativity. In contrast, short-range (but also native-like) interactions decrease folding cooperativity, but accelerate the folding process. All interactions inconsistent with the target structure decrease the folding rate. Similar conclusions were drawn from geometrically more complex models with a similar interaction scheme [108].

Systematic diamond lattice, Monte Carlo studies of various types of interactions and their effects on folding thermodynamics, structure and stability, and folding mechanism have been done by Kolinski et al. [43,67,68] and Skolnick et al. [109-117]. The interaction scheme was gradually modified, when necessary, so as to reproduce various features of protein folding. In the simplest model, they investigated the interplay between short- and long-range interactions for various homopolymeric chain lengths. Short-range interactions were modeled as a preference for the *trans* conformation. Long-range interactions (i.e., those that are local in space, but far apart down the chain) were simulated by an energetic preference for contacts

between nonbonded chain units. Flexible polymers (having the same a priori probability of all rotational isomeric states) of any finite length undergo a continuous transition to a compact, disordered globular state [44,67]. This is typical for all flexible polymer systems [118,119]. Semiflexible polymers exhibited a qualitatively different behavior. When the ratio of short-range interactions (the stiffness parameter) to long-range interactions exceeds some critical (but moderate) value, the collapse transition for moderate length chains (50–200 units) becomes discontinuous (pseudo-first-order). There is a cooperative increase of the apparent chain stiffness on collapse, and the globular state exhibits almost perfect orientational ordering [67]. Under the same conditions, longer polymers form several ordered domains [68]. This closely resembles the formation of multiple domains in real proteins. It could be expected that for sufficiently long chains, the collapse transition is again continuous. In all cases, the compact globular state of homopolymers was not unique [43].

Introducing neutral (flexible and neutral with respect to the long-range contacts) turn regions leads to a unique, minimal, four-member (β -barrel globular state. Furthermore, the conformational transition is of the all-or-none type [109,111]. When robust patterns of hydrophobic and hydrophilic residues were introduced and/or when the native conformation of the turns was favored, in all folding simulations, it was possible to reproduce the nontrivial, six-member, Greek key, β -barrel topology [110,112]. The folding scenario could be described as on-site assembly, where already folded fragments of native structure serve as a folding scaffold for the remaining portions of the model chain [110,113].

In a similar way, the minimal requirements for the folding of all typical topologies of four-helix bundles were elucidated [114–116]. In all these studies, the folding process was accelerated when an orientational coupling of long-range interactions had been implemented to model hydrogen bond interactions. The effects of hydrogen bonds and other polar interactions incorporated into a very simple protein-like model have also been investigated by O'Toole et al. [120]. Finally, the diamond lattice protein model has also been used to investigate the possible differences between *in vivo* and *in vitro* folding of typical protein motifs [117].

The diamond lattice representation of a protein, while reasonable for idealized models of β and α proteins, cannot reproduce the chain geometry of more complex motifs such as α/β proteins. To remove this fundamental limitation, the chess-knight model [54] of the protein backbone was developed. In contrast to a diamond lattice model, it allows all possible protein folding motifs to be represented at low resolution [121]. The force field requirements for a unique native state are essentially the same in this finer lattice representation as for diamond lattice models.

The aforementioned studies indicated the crucial importance of the proper balance between short-range versus long-range interactions in order to obtain the requisite cooperativity of the folding process, as well as a unique native state. They also show that the sequence information required for a given structure is rather robust. In addition, a sequential, on site, mechanism of folding was preferred in all cases. Qualitatively, these findings are in agreement with conclusions from simple, exact models and even more so with those of Go and co-workers [104]; however, they are more detailed because of the use of more complex models.

Brower et al. [122] investigated the chain-length dependence of the time required to find the global energy minimum in the framework of various high-coordination-number lattice models and a simple interaction scheme. It has been found that the folding time when simple simulated annealing is used increases exponentially with the chain length. This points out the necessity of more efficient sampling algorithms for the conformational search in nontrivial protein models. In this respect, Finkelstein et al. [123–125] have proposed a very interesting Monte Carlo model of protein assembly, where larger polypeptide fragments are moved around to find an optimal structure.

2. Continuous Models of Protein-Like Systems

Valuable insights into the thermodynamics and mechanisms of protein assembly come from Monte Carlo and Brownian dynamics studies of simple, continuous protein-like models. For example, Rey and Skolnick [126] studied a series of reduced models of four-helix bundles. They investigated the role of the balance between short- and long-range interactions in the folding pathways, and the stability and structural uniqueness of the four-helix bundle folding motif. Monte Carlo simulations for various types of models enabled a qualitative assessment of the role of side chains in protein structure fixation. The results have been generalized into a phase diagram describing various regimes of polypeptide behavior. Possible implications for real protein folding and the design of artificial proteins have also been discussed. In related work, Rey and Skolnick [127] demonstrated that the dynamics and folding pathways of various simple folding motifs studied via Brownian dynamics and Monte Carlo dynamics are qualitatively the same. Moreover, the general dynamical characteristics of these continuous models were essentially the same as for related lattice models studied by means of lattice Monte Carlo [127].

Continuous models of minimal β sheets have been investigated by Monte Carlo simulated annealing [128] and by Brownian dynamics [23,52,53,129]. The design of the β -barrel geometry and the essential features of the force field are complementary to the diamond lattice models of Skolnick

et al. that were studied much earlier [109,111]]. Again, the behavior of lattice and continuous models are qualitatively the same [23,52,53,129]. More recently, the coil-to- β -sheet transition and its cooperativity and relation to existing theories has been studied via Monte Carlo simulations on a reduced continuous model [130].

D. Models of Real Proteins

1. Continuous Models

To make protein folding simulations computationally tractable, the protein model must be simplified [131]. Among the earliest examples of such models is the original approach of Levitt and Warshel [22]. There the polypeptide main chain was reduced to its α -carbon trace, and side chains modeled as single united atoms. Long-range interactions were counted only for pairs of side chains and approximated by a Lennard-Jones potential. A constant value of the planar angle between the two consecutive pseudobonds of the α -carbon trace has been assumed. This is a simplification since real proteins exhibit a bimodal distribution of this angle. The local degrees of freedom were the dihedral angles of the α -carbon trace. The torsional potentials associated with these degrees of freedom were derived from a conformational analysis of several "representative" dipeptides. The model has been used in folding simulations of bovine pancreatic trypsin inhibitor. Simulations were done by means of Brownian dynamics; however, the Monte Carlo method could be used as well as for conformational sampling. Similar models have been developed and studied by Kuntz et al. [132], Robson and Osguthorpe [133], and Hagler and Honig [134]. The work of Hagler and Honig showed that similar quality structures of BPTI could be obtained using a sequence code reduced to just two types of amino acids.

Another example of a continuous reduced model is due to Wilson and Doniach [24]. They assumed a fixed planar conformation of peptide bonds, and the main chain was allowed to change its conformation by rotation around the phi (ϕ) and psi (ψ) angles. A single united atom representation of each side group also has been assumed. The importance of their work lies in the systematic application of a knowledge-based force field, an idea applied previously to other aspects of the protein folding problem [135,136]. Short-range interactions were simulated by restricting the allowed values of the phi-psi angles to those commonly observed conformations in known folded protein structures. For long-range (tertiary) interactions, they derived distance dependent pairwise potentials based on a statistical analysis of the regularities in known globular protein structures. Conformational sampling was done using simulated thermal annealing within the framework of a Metropolis-type Monte Carlo scheme. They performed folding simulations of crambin both with and without assumed knowledge

of the native secondary structure. Overall, the accuracy of the predicted structures was very low. However, the predicted secondary structure was to a large extent in agreement with that of the native protein and elements of a protein-like hydrophobic core were formed. Also, the native-like pattern of cystine crosslinks was observed.

The accuracy of a reduced protein representation could be improved by taking into account some internal degrees of freedom of the side chains [137]. For example, larger side chains could be represented by two united atoms [138]. Alternatively, an all-atom representation of the main chain could be employed with a reduced representation of the side chains [28]. Using this kind of representation and more elaborate statistical potentials, the structure of short peptides such as melittin, pancreatic polypeptide inhibitor, apamin [28], PPT, and PTHrP [138] have been predicted with an accuracy ranging from 1.7-Å root-mean-square deviation, (RMSD) (measured for the α -carbon positions) for the small single helix of melittin to 4.5-Å RMSD for larger peptides.

Reduced continuous-space models were also employed in studies of the various aspects of protein [27,139] or polypeptide [49,140,141] folding.

2. Discretized, Lattice Models

Lattice models have proved to be extremely useful in studies of simple, exact models and somewhat more complex models of protein-like systems. Similarly, the conformational space of a real protein could be discretized and explored in a very efficient way by various versions of Monte Carlo sampling techniques. Depending on the assumed level of discretization and model of force field, various levels of accuracy can be achieved. For example, simple lattice models of real proteins were studied by Ueda et al. [108], Krigbaum and Lin [142], Dashevskii [143], Covell [144], Covell and Jernigan [145], Hinds and Levitt [146], and others.

As mentioned before, the studies of protein-like models by Go et al. [103,104,107] provided a plausible explanation for the origin of protein folding cooperativity and the role played by short- and long-range interactions in both folding and stability. Monte Carlo studies of a lattice model of lysozyme [108] provided a very nice demonstration of these findings applied to a real protein. The target structure of this 129-residue protein was represented by 116 simple cubic lattice main-chain units plus an additional 15 lattice points for some larger side chains. Thus, there was no one-to-one correspondence between the model and the real protein structural units, but the overall geometry and dense packing of the native structure was reproduced with reasonable accuracy.

Krigbaum and Lin [142] studied a bcc (body-centered cubic) lattice model of PTI. Their simulations demonstrated that some overall features of

the native fold (at least for very small structures) could be reproduced using only a one-body, centrosymmetric potential that describes the burial preference of various amino acids.

Covell and Jernigan [145] enumerated all possible compact conformations of several small proteins restricted to an fcc (face-centered cubic) lattice (one residue-per-lattice site). Using pairwise interactions, they found that the native-like conformation could be always found within 2% of the lowest-energy lattice structures. Very much in this spirit, Hinds and Levitt [146] used a diamond lattice to search compact chain conformations where the lattice defines a chain tracing, but each lattice vertex need not correspond to a single residue. Subsequently, Covell [144] investigated simple cubic lattice chains. However, now a more elaborate interaction scheme was used, which consisted of pairwise interactions, surface exposure terms and a packing density regularizing term. These potentials were derived from a statistical analysis of a protein structural database. During Monte Carlo simulations, which employed various local and global moves, the model chains rapidly collapsed to compact states having some features of native proteins.

Kolinski and Skolnick [42,147] developed a series of high-coordination lattice protein models that are capable of describing protein structures at various levels of resolution. The simplest of these, the chess-knight model [54], assumes an α -carbon representation of the polypeptide chain. The model enabled Monte Carlo folding pathway simulations of complex real protein structures, including plastocyanin and two TIM barrels. In these simulations [148,149], the model force field consisted of short-range interactions, biased toward the target native secondary structure, and pairwise interactions between side chains defined according to the Miyazawa-Jernigan [139] hydrophobicity scale.

One problem with low-resolution lattice models is their spatial anisotropy [15]. That is, a given secondary structural element may be represented by substantially different lattice structures as the element assumes various orientations on the lattice. In the hybrid 210-lattice model of protein backbone, this disadvantage has, to a large extent, been removed, mostly due to the larger set of basis vectors (56) employed in building the α -carbon trace. This lattice representation has about a 1.0-Å coordinate RMSD from the corresponding native α -carbon coordinates. The relatively accurate main-chain representation enabled the construction of a side-chain rotamer library. Side chains are represented as single spheres and their various positions, with respect to the main chain, simulated various rotamers of each amino acid. The model force field has been derived from the appropriate statistics of high-resolution crystallographic structures. Short-range interactions consisted of several types of terms that reflect secondary propen-

sities of various amino acids. A somewhat similar potential has been recently investigated by DeWitte and Shakhnovich [150]. Tertiary interactions are described by amino acid-specific burial terms, pair terms, and several kinds of multibody, knowledge-based contributions that act to regularize protein structure. These consisted of a cooperative model of a hydrogen bond network (employing geometric criteria similar to that of Levitt and Greer [151]), and several types of amino acid pair-specific and four-body packing correlations based on side-chain contact patterns seen in globular proteins [152]. This force field was accurate enough to enable the reproducible de novo folding of several simple proteins, starting from a random expanded conformation [16,18].

A finer 310-hybrid lattice model employs a 90-vector basis set for α -carbon trace representation. This model can reproduce the geometry of native proteins with an average α -carbon RMSD of 0.6–0.7 Å. The general principles of the force field were similar to that of the 210-hybrid lattice model. However, the more accurate geometric representation enabled a better knowledge-based modeling of short-range interactions [20] and the more straightforward design of hydrogen bonding [17,40]. As a result, the folding of more complex protein structures became possible [19,40,42,56,153–155]. Single secondary-structure elements, such as long helices, assembled during the Monte Carlo simulations with an accuracy in the range of 1.0–2.0 Å RMSD. The model also predicted some simple protein structures whose coordinate RMSD from native ranged from 2.0 to 5.0 Å, depending on protein size.

3. Complementary Insights from Various Models

The preceding outline shows that very different continuous and discretized models of proteins and protein-like systems have been used to address closely related problems of protein physics. Which are better for the study of protein folding dynamics, thermodynamics or folding patterns? The answer depends on the particular aspects of the protein folding problem to be addressed. Certainly, the fine geometric details of protein conformation could be better reproduced by continuous models. Thus, when dealing with small peptides, a properly designed continuous, reduced protein model should be more useful than a lattice model. This is especially true for short-range correlations that are much easier to account for in continuous models. What about larger systems, such as proteins? When small-scale conformational transitions are of interest, a continuous-space model is again the natural choice. However, when dealing with large-scale conformational transitions, especially in simulations of folding from the denatured state, discretized, lattice models offer several distinct advantages. First, because of the lattice structure itself, the conformational space is a priori

partitioned, and one avoids many local, and perhaps irrelevant, energy barriers. Second, local conformational transitions could be predefined and used during simulations in a very effective way. Moreover, the conformational energy computation could be accelerated by precalculating many components just once and storing the results in a table. Because of the last two attributes of the lattice models, lattice Monte Carlo simulations could be a couple of orders of magnitude faster than simulations in an otherwise equivalent, continuous-space model [42,127]. However, when designing a lattice model, it is always necessary to carefully analyze the possible effects of the lattice anisotropy and the discrete structure of conformational space [15]. Recently, much progress has been achieved in various methods of discretization [156,157] of protein conformational space [42] and in the analysis of the fidelity of such discretizations [14,156,158-160].

Reduced models of proteins and protein-like models utilize various levels of generalization. On one side there are the simple, exact models that can be analyzed in great detail. The simple cubic lattice representation of polypeptide chains reflects the very fundamental role of the close, crystal-like packing of protein structures. It has been shown that the folding transition of cubic lattice copolymers responds in a reasonable way to sequence variation. For some protein-like sequences, unique compact states represented the minimum of conformational energy. If so, it is very likely that simple, exact models properly mimic the most general physics of hydrophobic collapse and packing complementarity of globular proteins. With a proper model of chain dynamics, the most fundamental features of protein folding pathways also could be studied [41].

However, in spite of their many virtues, it appears that the simple, exact models do not sufficiently account for a rather important feature of polypeptides and proteins, specifically, the complex interplay between short- and long-range interactions associated with the presence of a main chain and side chains. The importance of this aspect of protein structure and dynamics has been addressed in an excellent short review by Honig and Cohen [161]. On the most trivial level, it should be pointed out that proteins are rather stiff polymers. In the range of chain length and conformational stiffness typical of proteins, semiflexible lattice homopolymers undergo a first-order collapse transition to a highly ordered, globular state [67]. This is perhaps the most fundamental aspect of protein folding physics that has to be demonstrated by any protein-like system. Furthermore, the "decoration" of protein-like models with sequence specific short- and long-range interactions (and also with some more complex protein-like geometrical correlations) leads to a structural fixation of compact states characteristic of real proteins. Finally, if the tertiary structure of a real protein is to be predicted, the model must be capable of representing the

geometry of the native state at an acceptable level of resolution.

What, then, is the relationship between models of various levels of generalization? They are complementary for several reasons. First, they address various aspects of the protein folding problem. The general insights from studies of the simple, exact models and protein-like models provide guidelines for designing reduced force fields in more detailed models. On the other hand, conformational space cannot be explored with high accuracy when many structural details are taken into account. In such cases, only a limited number of folding simulations could be performed. Consequently, it is more difficult to derive a quantitative description of protein folding dynamics and thermodynamics. However, one can address numerous important aspects of protein physics using these more complex models. When appropriate, the predictions of such models also can be tested experimentally.

III. PROTEIN FOLDING THERMODYNAMICS

A. Nature of the Interactions

A key question is what types of interactions are important for protein stability. Protein backbones contain polar groups that either form hydrogen bonds with water or with other backbone or side-chain atoms; the free energy cost for the absence of hydrogen bonds is substantial [162]. Thus, hydrogen bonds play an important structural regularizing role in proteins [42]. In addition, individual residues exhibit differential secondary structural preferences. The secondary structure found in protein fragments, although small, is nonetheless present and can be significant [163]. Such intrinsic secondary structural preferences act both to assist in the early states of folding and to reduce the configurational entropy of compact states [110]. Another important interaction is the hydrophobic effect that acts to sequester hydrophobic molecules within the interior of the protein [8]. Furthermore, amino acid pair-specific interactions and higher-order tertiary and quaternary interactions may be responsible for the selection of the native fold or the destabilization of an alternative structure [164]. Such interactions may be electrostatic in origin or arise from the complex interplay of potentials of mean force. Thus, there are a variety of interactions present in a globular protein, and the native structure is the result of a balance of such terms.

B. Extent of Frustration in Native Conformations

The thermodynamic hypothesis [1,7] that native proteins are in a global minimum of free energy is now generally accepted. Since the native state

has a relatively low conformational entropy, they are in a minimum of conformational energy; however, the larger corresponding entropy of solvent molecules also has to be kept in mind [165]. Of course, this does not mean that all interactions in a globular protein stabilize the native state. There is always competition between some interactions. For example, some pentapeptide fragments [166] can adopt different secondary structures in various proteins. This means that the observed native structure is a compromise between short-range secondary propensities and tertiary interactions. A similar analysis could be performed for other interactions in proteins. Hydrophobic residues are not always buried, and polar residues could be found in otherwise mainly hydrophobic core proteins. Sometimes, such "defects" are essential for structural uniqueness and can cause a "compensation" effect—the destabilization of one part of a protein increases the stability in another part. This leads us to very important question: What are the molecular reasons for protein structural uniqueness and the cooperativity of the folding process?

C. Origin of Cooperativity and Two-State Thermodynamic Behavior in Protein Folding

Typically, the *in vitro* folding of a single domain globular protein resembles a first-order phase transition in the sense that the thermodynamic properties undergo an abrupt change, and the population of intermediates at equilibrium is very low. In other words, the process is cooperative and is well described by a two-state model [8]. The first attempts to explain protein folding cooperativity focused on the formation of secondary structure. Theoretical and experimental analysis of coil-helix transitions indeed proved that the process is cooperative [167]. However, the helix-coil transition is always continuous [168], and thus it cannot explain the two-state behavior of the protein folding transition.

Simulations of protein-like models with target short- and long-range interactions [103,104,107,108,110] demonstrated that the abrupt transition is due to strong native-like tertiary interactions. Native-like short-range interactions increased the stability of the globular state; however, they appeared to decrease folding cooperativity. The dominant role of the long-range interactions for folding cooperativity also has been demonstrated in studies of simple, exact models. Shakhnovich et al. [57,71–73,169,170], Bryngelson et al. [85,99,100], and others [65,75,78,84,171,172] pointed out the importance of the entire energy landscape for the protein folding cooperativity. In this context, a cooperative all-or-none transition also could be expected for systems with many competing interactions provided that there is a well-pronounced energy minimum corresponding to the native state [65,75,78,84,172].

An interesting realization of such an energy landscape is provided by semiflexible polymer models. As was suggested by mean field considerations [173,174], semiflexible polymers should undergo a pseudo-first-order phase transition. This has been quantitatively illustrated in Monte Carlo studies of homopolymeric lattice systems by Kolinski et al. [43,67,68]. Thus, a cooperative collapse transition may also result from the interplay between non-specific short-range and uniform long-range interactions.

A very convincing explanation of protein folding cooperativity has been given recently by Hao and Scheraga [35]. They used the entropy-driven Monte Carlo (ESMC) method [38] to investigate the collapse transition in the chess-knight model [54] of a minimal β -barrel protein (see the chapter "Entropy Sampling Monte Carlo for Polypeptides and Proteins"). Their model force field consisted of short-range interactions, a qualitative model of hydrogen bonding and pairwise side-chain interactions. The side chains lacked internal degrees of freedom. Computer experiments demonstrated that some amino acid sequences exhibit an abrupt first-order transition from a random coil state to a well-defined globular state. The cooperativity of the folding process emerged from a characteristic depletion of the system's conformational entropy as a function of conformational energy as one moves away from the native state. More random sequences exhibited a continuous collapse transition [36]. Sequence-specific long-range interactions were responsible for the protein-like, all-or-none folding, while the short-range interactions only contributed to the stability of the globular state [37].

A similar study on a somewhat more complex 310-hybrid lattice model of an idealized Greek-key fold has been performed by Kolinski et al. [40,42]. The model force field consisted of knowledge-based, short-range interactions that quantitatively reproduced secondary propensities of various sequences, a cooperative model of hydrogen bonds, and several (also knowledge-based) types of tertiary interactions. The model employs a single united-atom multiple-rotamer representation of the side chains. A first-order transition has been obtained only when the explicit cooperativity of the side-chain interactions was taken into account. These cooperative terms had the form of four-body potentials preferring protein-like side-chain packing arrangements. The transition-state of such a cooperative model has an overall conformation very similar to that of the native state with most of the native secondary structure (in a sense of loose interactions of expanded β strands); however, the average volume is larger. The folding transition was associated with the fine-tuning of side-chain packing accompanied by small rearrangements of the entire structure. Such "side-chain fixation" increased the relative contribution of the multibody interactions and produced a unique native state. Since the cooperative interactions con-

tribute very little to the transition-state energy, they acted by providing a sufficient energy gap from native for the all-or-none transition to emerge. Interestingly, the native state still had a substantial energetic degeneracy (and thereby a substantial entropy) with a large number of conformations having essentially the same hydrophobic core packing. However, these simply reflect small structural differences that arise mostly in the loop and turn regions. The first-order transition was due to the change of the conformational entropy as a function of energy. The entropy changed most rapidly near the transition-state. The conformational entropy of the side chains may play a very important role in the late stages of the folding process [40,42]. This picture of the transition state is consistent with many experimental facts and with the theoretical postulate that the molten globule is a general transition state in protein assembly [169,175-184].

All these studies seem to indicate that the cooperativity of the folding process is due to the specific pattern of tertiary interactions and/or the specific interplay between short- and long-range interactions. This may appear to be a trivial statement, but detailed analysis of the results from the simple, exact model, protein-like models and reduced models of real proteins show several specific requirements for the protein folding cooperativity. It is very encouraging that over the entire spectrum of theoretical model studies of the protein folding process, these requirements essentially overlap [35-37,51,83,92,95]. Naturally, various models place different stress on the specific interactions that may control protein folding and structural uniqueness.

IV. TERTIARY STRUCTURE PREDICTION

Knowledge of protein structure is important for understanding their biological functions. However, the number of known protein sequences is many times larger than the number of solved three-dimensional structures. Thus, the ability to predict protein structure from amino acid sequence is one of the most challenging goals of contemporary molecular biology. Considerable progress toward a solution of the protein folding problem has been made recently, although a complete solution is not yet in hand. One area of such progress has been through the use of threading approaches that are designed to match a sequence to the most compatible structure in a library of known protein structures [185]. Here, we limit ourselves to methods based on the direct search of protein conformational space, and focus mainly on Monte Carlo and closely related methods.

A. Peptides and Cyclic Peptides

At first glance, it may appear that the problem of structure prediction in

small peptides should be much simpler than in globular proteins. To some extent this is true. On other hand, a minimum-energy conformation of a peptide may differ very little from the manifold of competing conformations. Thus, the accuracy requirements are greater for the force field. In this respect, Scheraga and co-workers [186] have developed a semiempirical force field (ECLIPP) [187] that is particularly well suited for studies of oligopeptides, where the short-range interactions may dominate the structural properties. Using various minimization [9,10,29,91,188-190] and simulation techniques [191,192], the conformational space of (Met)enkephalin [191,193], the cyclic decapeptide gramicidin S, cyclic hexaglycine, and the 20-residue membrane-bound fragment of melittin [194], alamethicin, and other peptides and small proteins [195-198] have been investigated in detail. Possible effects of the solvent environment also have been analyzed. Good-quality structure predictions of various coiled-coil and peptide crystal structures also have been achieved [186,199,200]. Although not all of these works have employed a Monte Carlo search methodology, we list them to provide an overview of the present status of the oligopeptide structure prediction field.

Using a statistical potential and a genetic algorithm as the optimization procedure, Sun has achieved a relatively high resolution (1.66-Å) RMSD for the 20-residue membrane fragment of melittin) prediction of a few oligopeptides [28]. However, as demonstrated by Hansmann and Okamoto [201], peptide conformational space could be effectively explored by a Monte Carlo multicanonical algorithm. Okamoto [202] employed the ECEPP potential and a simulated annealing Monte Carlo procedure to estimate the helix-forming tendencies of several nonpolar amino acids within the context of homooligopeptides. Such an approach can perhaps be used to obtain better semiempirical short-range interaction potentials for use in protein simulations. Finally, a very interesting Monte Carlo minimization procedure has been employed by Meirovitch et al. [203-205] for the determination of peptide conformations in solution from averaged NMR (nuclear magnetic resonance) spectra.

This brief overview shows that for particular applications, Monte Carlo methods and other minimization methods could be complementary to standard molecular dynamics methods (see some reviews on MD techniques and applications to proteins and peptides) [11,12,206].

B. Proteins

1. *De Novo Approaches*

Early attempts at *de novo* predictions of protein structure focused on a few small proteins. In spite of the rather simple test cases, the obtained struc-

tures were of low resolution [22,24,132]. It is interesting that the simple cubic lattice simulations of Covell [144] led to a similar accuracy. Also, a very simple lattice representation has been employed in the work of Hinds and Levitt [146]. These results suggest that the proper design of the interaction scheme is no less important than the accuracy of the geometrical representation [207,208]. The more elaborate Monte Carlo model of Wallqvist and Ullner [138] allowed for the prediction of the 36-residue PPT structure with an accuracy of 4.5-Å RMSD. This is the same accuracy as obtained by Sun [28] via a completely different search procedure.

An interesting hierarchical Monte Carlo procedure for the prediction of protein structures has been proposed recently by Rose et al. [209]. Their model employs an all-atom representation of the main chain and a crude representation of the side groups interacting via a simple contact potential. The method seems to be quite accurate in the prediction of protein secondary and supersecondary structure; however, the overall global accuracy of the folded structures is rather low.

Kolinski, Skolnick, and co-workers [16-18,42,56,147,153,155,210] have developed high-coordination-number lattice models of protein conformation and dynamics. The method has been tested in de novo structure predictions of several small proteins, including protein A fragments, two designed four-helix bundle proteins, a monomeric version of ROP, crambin and several coiled-coil motifs (see Section V.A). The Monte Carlo folding procedure employed a simulated annealing protocol and a knowledge-based force field derived from protein structural regularities [152,211-213]. For globular proteins, the accuracy of these predictions (measured as the coordinate RMSD from the α -carbon trace of known structures) ranged from 2.25 Å for the B domain of protein A and 4.0 Å for crambin to about 5 Å for the 120-residue ROP monomer. The obtained lattice folds were consistent with an all-atom representation, and detailed atomic models could always be reconstructed at virtually identical accuracy as the corresponding lattice models [18,155,210]. This high-coordination-number lattice model also has been used in de novo computer-aided design (CAD) [19,40] and in the redesign of protein topology [153,154]. However, the method fails for larger proteins of complex topology. In these cases, fragments of the protein under consideration usually form native-like motifs; however, topological errors (wrong strand reversals, etc.) lead to globally misfolded structures. The knowledge-based force field of these models seems to be qualitatively correct (it recognizes most native folds in threading experiments), but the native energy minimum is not very deep. Therefore, the applicability of the model to structure prediction could be substantially extended with the help of some theoretically or experimentally derived restraints.

2. Prediction Using Known Secondary Structure and Correct Tertiary Restraints

One possible way of improving tertiary structure prediction is to use either known or predicted secondary structural information. In that regard, using an off-lattice model and exact knowledge of the secondary structure, Friesner et al. have successfully folded two four-helix bundle proteins, cytochrome b562 and myohemerythrin [214]. Furthermore, assuming known native secondary structure, Dandekar and Argos [215,216] have reported encouraging results for simple helical and β proteins using a genetic algorithm to search conformational space. Mumenthaler and Braun [217] have developed a self-correcting distance geometry method that also assumes known secondary structure. They successfully identified the native topology of six of eight helical proteins. There also have been a number of studies that incorporate the known native secondary structure and a limited number of known and correct long-range restraints to predict the global fold of a globular protein. In particular, Smith-Brown, Kominos, and Levy [218] have modeled the protein as a chain of glycine residues and used Monte Carlo as the search protocol. Restraints are encoded via a biharmonic potential, with folding forced to proceed sequentially via successive implementation of the restraints. A number of proteins were examined. By way of example, flavodoxin, a 138-residue α/β protein, was folded to a structure whose backbone RMSD from native was 3.18 Å for 147 restraints. Another effort to predict the global fold of a protein from a limited number of distance restraints is due to Aszodi et al. [219]. This approach is very much in the spirit of the work of Mumenthaler and Braun and is based on distance geometry, where a set of experimental tertiary distance restraints are supplemented by a set of predicted interresidue distances. Here, these distances are obtained from patterns of conserved hydrophobic amino acids that have been extracted on the basis of multiple sequence alignments. In general, they find that to assemble structures below 5-Å RMSD, on average, typically more than $N/4$ restraints are required, where N is the number of residues. Even then, the Aszodi et al. [219] method has problems selecting out the correct fold from competing alternatives. An advantage of the Aszodi et al. approach is that it is very rapid, with a typical calculation taking on the order of minutes on a typical contemporary workstation.

Turning to lattice-based methods that sample conformations via Monte Carlo, the MONSSTER (*m*odeling of *n*ew structures from secondary and *t*ertiary restraints) method for folding of proteins using a small number of tertiary distance restraints and loosely defined knowledge of the secondary structure of regular fragments has been developed [220]. The method incorporates potentials reflecting statistical preferences for secondary struc-

ture, side-chain burial and pair interactions, and a hydrogen bond potential based on the α -carbon coordinates. Using this algorithm, several globular proteins have been folded to moderate resolution native-like, compact states. For example, flavodoxin, with 35 restraints, has been folded to structures whose average RMSD is 4.28 Å. Plastocyanin with 25 long-range restraints adopts conformations whose average RMSD is 5.44 Å. These results compare very favorably with the Aszodi et al. and Smith-Brown et al. approaches [218,219]. In general, preliminary indications are that helical proteins can be folded with roughly $N/7$ restraints, while β and $\alpha\beta$ proteins require about $N/4$ restraints, where N is the number of residues. Of course, for any particular case, the accuracy depends on the location of the restraints. This point requires additional investigation.

3. Method for Prediction of Tertiary Structure from Predicted Secondary Structure and Tertiary Restraints

Recently, a procedure has been developed that does not rely on a priori knowledge of the native conformation to fold a number of single-domain proteins [221,222]. The procedure can be derived into two parts: restraint derivation and structure assembly using the MONSSTER algorithm. With respect to restraint derivation, using multiple sequence alignment, protein secondary structure is predicted using the PHD method [223]. Regions where the chain reverses global direction (U-turns) are predicted using the method of Kolinski et al. [224] and override PHD predictions because they have been shown to be highly accurate. Then, the multiple sequence alignment and the secondary structure prediction for the topological elements between U-turns are combined to predict side-chain contacts. Such contact map prediction is carried out in two stages. First, an analysis of correlated mutations is carried out to identify pairs of topological elements of secondary structure (regions of the chain between U-turns) that are in contact [225]. The rationale is that in this way it is possible to restrict the predictions to rigid elements of the core, for which the assumption of closeness in space as reflected in the covariance of their mutational behavior is, in principle, more valid. Then, inverse folding [226] is used to select compatible fragments in contact, thereby enriching the number and identity of predicted side-chain contacts. Parenthetically, we note that for topological elements that are known to touch, this procedure produces contacts of which 67% are correct within ± 1 residue. This is of comparable accuracy to the situation where the contacts are predicted. Of course, they do not employ any information about the native structure in the prediction protocol. The final outcome of the prediction protocol is a set of noisy secondary and tertiary restraints.

The predicted secondary and tertiary restraints are inputted to the

MONSSTER method described above. However, now restraint information is predicted rather than being extracted from the native structure. Implementation of the restraints is carried out so as to take their low resolution into account. Low-energy structures are then searched for by simulated annealing, followed by isothermal refinement. Those structures having the lowest average and minimum energies are assigned as native.

The above protocol has been applied to the set of 17 proteins. All are extrinsic to the set of proteins employed in the derivation of the potentials. The average RMSD of the lowest-energy set of structures ranges from 2.8 Å for small helical proteins to roughly 6 Å for β and mixed-motif proteins. In all cases, the global topology is recovered. It is very important to emphasize that all structural predictions use the identical parameter set and folding protocol. Furthermore, they also considered one case, Ictf, where both the Rost-Sander prediction and the U-turn prediction protocol were poor. Because of the failure to identify the core regions of the molecule, the correlated mutation analysis yielded only two tertiary contacts (i.e., the protocol failed). More generally, the range of validity of this protocol remains to be established. Nevertheless, it is very encouraging that in the test case where the algorithm was grossly incorrect, the result yielded by the protocol was that the molecule is not foldable. It is very important to have an algorithm that indicates when it cannot be used; this preliminary result provides some encouragement that this might be the case.

V. QUATERNARY STRUCTURE PREDICTION

A. Coiled Coil, De Novo Simulations

Theoretical studies of coiled-coil systems are surprisingly scarce. Nevertheless, coiled coils represent an important test bed for quaternary structure prediction methods. Their quaternary structure consists simply of the side-by-side association of α helices wrapped around each other with a slight supertwist [227]. In the 1980s, Skolnick and Holtzer generalized the Zimm Bragg helix-coil theory [167] to include interhelical interactions [228]. The resulting phenomenological theory was able to predict the quaternary structure of tropomyosin and its fragments over a broad range of temperature and pH. However, these studies were mainly limited to dimeric coiled coils and required that the native structure be assumed. More detailed atomic modeling of coiled coils commenced in 1991 when Nilges and Brunger developed an automated algorithm for the prediction of coiled-coil structure, based on the assumption that the helices are parallel and in register [200,229]. They applied this approach to predict a structure of the GCN4 leucine zipper, whose backbone atom RMSD from the subsequently solved crystal structure is 1.2 Å. They concluded that given the

correct registration, the best packing of the hydrophobic residues in the core dictates the detailed geometry. This observation was first made by Crick in 1953. At about the same time, Novotny et al. [230] estimated the stability of GCN4, fos, and jun leucine zippers from molecular mechanics calculations. Their conclusions suggest that Leu in the *d* position of the canonical coiled-coil heptet makes a major contribution to the stability of dimers, whereas residues in the *a* positions are far less important. These conclusions are supported by the studies of Zhang and Hermans on a model leucine zipper [231]. Most recently, Harbury and Kim have developed a very simple and fast algorithm for predicting structures with ideal coiled-coil geometry. Unfortunately, it will not work if the coiled coil geometry is distorted [232]. To date, the only de novo simulations of coiled-coil assembly that starts from a pair of random chains and results in structures that are at the level of experimental resolution are those of Vieth et al. [17,155]. Starting from random denatured states and using their high-coordination-number lattice protein model [17], they produced parallel, in-register folding conformations on lattice. Full atomic models are built and refined using a molecular dynamics annealing protocol. This produced structures whose backbone RMSD from the crystal structure is 0.81 Å.

B. Method for Prediction of Equilibrium Constants of Multimeric Helical Proteins

Harbury et al. [233] have examined the shift in equilibrium among coiled-coil dimers, trimers, and tetramers associated with varying the identity of the residues of the hydrophobic residues in the coiled-coil heptad repeat of the GCN4 leucine zipper. Vieth and co-workers have developed a method to calculate the equilibrium constant among a manifold of assumed conformations (using a hybrid Monte Carlo-transfer matrix approach) [234]. When applied to the eight mutants studied by Harbury et al., their calculations are in agreement with experiment over the entire concentration range for five of eight sequences, and over a portion of the concentration-range, they are in agreement for an additional two sequences. They find that local, intrinsic secondary structure preferences and side-chain entropy act to favor lower order multimers, whereas tertiary and quaternary interactions act to favor higher-order multimers. The model also correctly predicted the state of association of a number of GCN4 leucine zipper fragments [235], as well as the quaternary structure of the DNA binding proteins fos and jun. Finally, they examined the coil-ser sequence designed by Eisenberg, DeGrado, and co-workers. In agreement with the experiment, an antiparallel, three-helix bundle is predicted to be the native conformation [236].

VI. SIMULATION OF PEPTIDE INSERTION INTO MEMBRANES

A. Small Proteins and Peptides

An off-lattice Monte Carlo dynamics simulation has been used to investigate the behavior of filamentous bacteriophage coat proteins pfl and fd and a number of peptides such as m2 δ , melittin, and magainin2 in a model membrane environment [237,238]. The objective is to predict the orientation, location and conformation of the proteins with respect to a model lipid bilayer. The protein is represented by a chain of balls with centers at the α carbons and whose radii are residue-dependent. Here, a continuous-space model is used; that is, the α -carbon positions are not restricted to lattice sites. The membrane is treated as a z-coordinate-dependent effective potential representing the environments of water, lipid-water interface and lipid. The hydrophobic interaction energy depends not only on amino acid identity but also on its position in space. In this set of simulations, the membrane is modeled as an effective medium. The simulations begin with a random conformation in the water portion of the box, and sampling occurs using Monte Carlo dynamics. These simulations predict that despite the very low sequence similarity between the major coat proteins of Pfl and fd bacteriophages, their structures in a membrane environment are very similar. This suggests that the hydrophobic effect exerts an important influence on membrane protein structure. Focusing on fd by way of example, the simulations predict that an amphipathic helix runs from residues 5 to 19, and that a transbilayer helix runs from residues 25 to 41. This is to be compared with the experiment that indicated that the *N*-terminal amphipathic helix runs from residues 7 to 20, and the transbilayer helix is located between residues 23 and 42. Thus, good agreement with experiment is found [239,240]. Similar agreement with experiment was found for the other systems studied.

B. Translocation Across Model Membranes

A legitimate question concerns the limitations of a model where the membrane is treated as an effective medium. To explore this issue, a series of simulations were done where each lipid molecule in the bilayer was represented by a dumbbell, and the effect of curvature on the transport of a structureless polymer was explored [241]. Interestingly, when the membrane is highly curved, there is almost irreversible transport from the outside to the inside of the spherical vesicle. This arises from entropic effects. Basically, because the interior leaflet is of lower density, after penetrating into the membrane, the polymer prefers to remain in this region. In

contrast, this effect vanishes when the membrane is treated as a structureless, hydrophobic medium. Thus, there may well be situations where the structure of the membrane exerts important effects.

The translocation of a polymer driven by an external field bias of strength ϵ across a model membrane has also been investigated using Monte Carlo methods [242]. Again, each lipid molecule is modeled as a dumbbell. It is found that below a characteristic field strength, ϵ^* , the membrane is practically impermeable. In the high-field limit of $\epsilon > \epsilon^*$, the permeability is independent of the field strength. At low fields, the permeability decreases according to $\exp(-\epsilon^*/\epsilon)$. The translocation mechanism can be understood as a Kramers process describing the escape of a particle over potential barriers.

VII. PERSPECTIVES

A. Where Are We Now?

Computer studies of simple, exact models [51,57,78], protein-like models, and more realistic protein models have greatly increased our understanding of the protein folding process and the stability of protein structures. For example, Levinthal's paradox [243-245] seems to be rationalized by the simulation results. It is now more or less clear how proteins partition their conformational space, thereby sampling only a very small fraction of all possible states [58]. Studies of simple, exact models have demonstrated that one such way of reducing the available conformational space is by hydrophobic collapse. Also, simple, exact models and protein-like models provide a plausible explanation of why some sequences fold rapidly, whereas others fold very slowly or not at all. Different folding scenarios can be now rationalized by analogy to a spin-glass [85,99,100,246].

From studies on somewhat more complex models, we have learned about the interplay between the short- and long-range interactions [35-37,40,42,103,104,107] and their possible role in both the thermodynamics and kinetics of protein folding. The formation of secondary structure nuclei can also drastically reduce available conformational space, providing a somewhat different (although in principle equivalent) explanation of Levinthal's paradox. Actually, an opposite paradox, termed here the *proximity paradox*, should be addressed: Why does a confined (by chain connectivity) molecular system consisting of thousands of atoms (whose degrees of conformational freedom are locally strongly coupled) require times ranging from milliseconds to seconds to adopt its native state, whereas folding into an "almost correct" structure from the denatured state is very rapid. The type of collective molecular mechanism responsible for the "side-chain fixation" or molten globule-native state transition is not yet

known. Simulation studies [131], physical experiments on natural proteins [247], and the investigation of artificial proteins [177,248-252] have started to address this issue. Recent studies seem to indicate that the satisfactory solution of this problem will require the analysis of the various types of physical interactions in proteins, including explicit multibody approximations to tertiary interactions.

Protein structure prediction [253,254] is one of the major goals of computational molecular biology. Up to now, homology based and threading methods have been the most successful [185,226,255-264]. However, due to the increasing ratio of the number of known protein sequences to the number of solved protein structures, the development of de novo (or related) methods would be extremely valuable. To date, only limited, nevertheless encouraging, progress has been achieved in such direct approaches [42,254]. Purely de novo predictions are now possible only for peptides [186] and very small and structurally simple proteins [42,200,229]. It appears that the most promising are the methods that employ knowledge-based factorization of protein interaction and structural regularities [28,135,139,152,207,211,212,214,265,266]. Over the short term, hierarchical techniques that exploit evolutionary information and fragmentary experimental data should become a standard tool for low-resolution protein structure prediction [222,267].

B. Shortcomings of the Present Force Fields, Possible Refinements, and Reformulations

The work of Novotny et al. [268] and others [269,270] has shown that the classic, detailed atomic force fields are locally, but not necessary globally, correct. It is unlikely that for these force fields the global minimum of conformational energy corresponds to the native state. On other hand, in many cases, even an extremely simplified interaction scheme can match a sequence to its native structure in a threading experiment. In other words, when a sequence is fitted to various protein-like structures, the demands on the force field are much less. In such cases, the numerous interactions that make a polypeptide chain a protein become to a large extent irrelevant, and only sequence specific components have to be correctly accounted for in threading experiments. The conclusion is rather simple—an efficient force field for simulation of the protein folding process and, consequently, for the de novo protein structure prediction must discriminate against a majority of conformations that never occur in folded proteins. Otherwise, a conformational search procedure would mostly explore rather irrelevant portions of the model conformational space. In the early attempts to build reduced models of proteins, this important fact was not sufficiently appreciated.

Within the context of the above, let us try to formulate some necessary requirements for the design of a moderate resolution reduced model (based on united-atom representation and knowledge-based potentials) of real proteins and its force field:

1. The resolution of the model has to be at least on the level of 2 Å. Otherwise, packing complementarity of the side groups would be lost.
2. Short-range interactions (knowledge-based or semiempirical) must account for the polypeptide chain's conformational stiffness [20,42]. In other words, protein-like correlations enforced by the potential should extend over several residues. This would considerably narrow the available conformational space.
3. At least for main-chain units, a highly directional model of hydrogen bond interactions (perhaps with an explicit cooperative term) has to be designed [16,17,19,40,42,55]. Globular proteins have a regular network of hydrogen bonds. The energy difference between a peptide group hydrogen-bonded to water molecules and the same group hydrogen-bonded to another peptide group is minor. However, the free energy price for having non saturated hydrogen bonds in a protein interior is large. Thus, the hydrogen bond network plays a very important structure-regularizing role, eliminating the majority of otherwise possible compact conformations of polypeptide chains.
4. Side chains (even when modeled as single-interaction spheres) should have some conformational freedom that reflects their internal mobility in real proteins. Pairwise interaction potentials should be as specific as possible, and in the absence of explicit solvent, a burial potential that reflects the hydrophobic effect may be necessary [17,42].
5. Somehow, higher-order than pair, multibody interactions have to be built into the model. Since unrestrained MD simulations with all atom potentials lead to liquid-like (instead of native-like) packing of side chains [269], it is rather unlikely that a protein-like pattern of side-group contacts would emerge in a reduced model context without tertiary structure-regularizing interactions. Such side-chain multibody interactions provide a bias toward a protein-like packing. The possible role of multibody correlations for protein structure modeling and recognition is starting to receive some attention [16,17,40,42,56,153,154,213,271,272].

Of course, the specific realization of these proposed minimal requirements may differ in various models. Some other features may be necessary.

The practical realization of the preceding set of requirements allowed the design of protein models that were capable of folding a subset of small, relatively simple proteins [16,42,56,155,210]. Very likely, further progress will depend on the development, refinement and consistent implementation of more complete knowledge-based force fields.

C. Promises of New Sampling Methods

Several recent developments in sampling methodology have increased the accuracy and reliability of Monte Carlo simulations of protein folding processes. In standard Metropolis-type algorithms, sampling efficiency can be immensely improved by the application of "smart" local conformational updates. This is true for both off-lattice continuous-space models [25,30,31] and for lattice models of protein chains [17,20,42]. These local moves, in contrast to purely random conformational updates, are designed in such a way that the model system visits only the local conformational energy minima; thus, it ignores all conformations that would be rejected as a result of steric overlaps, large distortions of valence angles, or other factors. Local minimization techniques [189] could be combined with the Metropolis sampling scheme.

Recently developed high-coordination-number lattice models of protein dynamics and structure have several methodological advantages. Such models have an acceptable resolution, no worse than that of most off-lattice reduced models. However, due to the use of a lattice, the simulation algorithms are much faster than those otherwise equivalent continuous-space models. It also should be noted that the resolution of recently developed reduced models is sufficient for hierarchical simulations. In these cases, large-scale relaxation can be performed via Monte Carlo dynamics on a reduced model and the intervening fragments of trajectory could be simulated via the MD technique. It is easy to do fast projections of all-atom models onto reduced models. Conversely, rebuilding all-atom structures from a reduced representation could also be done rapidly with reliable accuracy.

A novel approach to protein conformation is the entropy-sampling Monte Carlo method (ESMC), which is described in detail in another contribution to this volume. The method provides a complete thermodynamic description of protein models, but it is computationally quite expensive. However, because of the underlying data-parallel structure of ESMC algorithms, computations could be done on massively parallel computers essentially without the communication overhead typical for the majority of other simulation techniques. This technique will undoubtedly be applied to numerous systems in the near future.

VIII. CONCLUSIONS

Over the last decade, Monte Carlo-based approaches have proved to be a powerful tool for addressing various aspects of the protein folding problem. To date, their greatest impact has been in the application to reduced protein models designed either to explore general questions of protein folding or to predict low to moderate resolution native structures. Because dynamic Monte Carlo can effectively simulate dynamic processes, it has provided numerous insights into the mechanism by which proteins fold. Both dynamic and entropy-sampling Monte Carlo have permitted the investigation of the panoply of interactions responsible for the unique native structure characteristic of globular proteins. More recently, these methods have been applied with increased success to the prediction of protein tertiary and quaternary structure. This is especially true when knowledge-based, statistical potentials are combined with evolutionary information provided by multiple sequence alignments. Such combined approaches are very likely to yield additional progress in the protein folding problem in the near future.

This review has emphasized the application of reduced models to problems of protein folding. Such models combined with Monte Carlo sampling also could be used in the study of other biological processes, such as the mechanism of virus assembly [273]. Furthermore, Monte Carlo-driven, reduced protein models could provide effective reaction coordinates, which are then sampled by molecular dynamic simulations of atomic models. This would combine the best of both approaches: the ability to simulate large-scale, long-time processes and the ability to examine such processes at atomic detail. Thus, by being employed in novel contexts, Monte Carlo-based sampling approaches are likely to continue to provide numerous valuable insights into the behavior of biological systems.

REFERENCES

1. C. B. Anfinsen and H. A. Scheraga, *Adv. Prot. Chem.* **29**, 205-300 (1975).
2. T. E. Creighton, *Proteins: Structures and Molecular Properties*, Freeman, New York, 1993.
3. R. L. Jernigan, *Curr. Opin. Struct. Biol.* **2**, 248-256 (1992).
4. M. J. Gething and J. Sambrook, *Nature* **355**, 33-45 (1992).
5. O. B. Ptitsyn, *J. Prot. Chem.* **6**, 273-293 (1987).
6. T. E. Creighton, *Biochem. J.* **270**, 131-146 (1990).
7. C. B. Anfinsen, *Science* **181**, 223-230 (1973).
8. P. L. Privalov and S. J. Gill, *Adv. Prot. Chem.* **39**, 191-235 (1988).
9. L. Piela, J. Kostrowicki, and H. A. Scheraga, *J. Phys. Chem.* **93**, 3339-3346 (1989).

10. D. R. Ripoll, L. Piela, M. Velasquez, and H. A. Scheraga, *Proteins* **10**, 188-198 (1991).
11. C. L. I. Brooks, M. Karplus, and B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamic Structure and Thermodynamics*, Wiley, New York, 1988.
12. M. Karplus and G. A. Petsko, *Nature* **347**, 631-639 (1990).
13. J. A. McCammon, *Rep. Prog. Phys.* **47**, 1-46 (1984).
14. B. H. Park and M. Levitt, *J. Mol. Biol.* **249**, 493-507 (1995).
15. A. Godzik, A. Kolinski, and J. Skolnick, *J. Comp. Chem.* **14**, 1194-1202 (1993).
16. A. Kolinski, A. Godzik, and J. Skolnick, *J. Chem. Phys.* **98**, 7420-7433 (1993).
17. A. Kolinski and J. Skolnick, *Proteins* **18**, 338-352 (1994).
18. J. Skolnick, A. Kolinski, C. Brooks III, A. Godzik, and A. Rey, *Curr. Biol.* **3**, 414-423 (1993).
19. A. Kolinski, W. Galazka, and J. Skolnick, *J. Chem. Phys.* **103**, 10286-10297 (1995).
20. A. Kolinski, M. Milik, J. Rycombel, and J. Skolnick, *J. Chem. Phys.* **103**, 4312-4323 (1995).
21. Y. Zhou, C. K. Hall, and M. Karplus, *Phys. Rev. Lett.* **77**, 2822-2825 (1996).
22. M. Levitt and A. Warshel, *Nature* **253**, 694-698 (1975).
23. J. D. Honeycutt and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **87**, 3526-3529 (1990).
24. C. Wilson and S. Doniach, *Proteins* **6**, 193-209 (1989).
25. E. W. Knapp, *J. Comp. Chem.* **13**, 793-798 (1992).
26. D. Hoffmann and E. W. Knapp, *Phys. Rev. E* **53**, 4221-4224 (1996).
27. B. Lee, N. Kurochkina, and H. S. Kang, *FASEB J.* **10**, 119-125 (1996).
28. S. Sun, *Prot. Sci.* **2**, 762-785 (1993).
29. A. A. Rabow and H. A. Scheraga, *Prot. Sci.* **5**, 1800-1815 (1996).
30. E. W. Knapp and A. Irgens-Defregger, *J. Comp. Chem.* **14**, 19-29 (1993).
31. A. Elofsson, S. M. Le Grand, and D. Eisenberg, *Proteins* **23**, 73-82 (1996).
32. K. Binder, *Monte Carlo Methods in Statistical Physics*, Springer-Verlag, Berlin, 1986.
33. K. Binder, paper, Institut Für Physik, Johannes Gutenberg-Universität, Mainz, 1991.
34. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **51**, 1087-1092 (1953).
35. M.-H. Hao and H. A. Scheraga, *J. Phys. Chem.* **98**, 4940-4948 (1994).
36. M.-H. Hao and H. A. Scheraga, *J. Phys. Chem.* **98**, 9882-9893 (1994).
37. M.-H. Hao and H. A. Scheraga, *J. Chem. Phys.* **102**, 1334-1348 (1995).
38. J. Lee, *Phys. Rev. Lett.* **71**, 211-214 (1993).
39. B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9-12 (1991).
40. A. Kolinski, W. Galazka, and J. Skolnick, *Proteins* **26**, 271-287 (1996).
41. R. A. Orwoll and W. H. Stockmayer, *Adv. Chem. Phys.* **15**, 305-324 (1969).
42. A. Kolinski and J. Skolnick, *Lattice Models of Protein Folding, Dynamics and Thermodynamics*, Landes, Austin, TX, 1996.
43. A. Kolinski, J. Skolnick, and R. Yaris, *Biopolymers* **26**, 937-962 (1987).
44. A. Kolinski, J. Skolnick, and R. Yaris, *Macromolecules* **20**, 438-440 (1987).
45. A. Kolinski, J. Skolnick, and R. Yaris, *J. Chem. Phys.* **86**, 7164-7173 (1987).
46. A. Kolinski and J. Skolnick, *J. Phys. Chem.* **97**, 3450 (1993).
47. P. E. J. Rouse, *J. Chem. Phys.* **21**, 1272-1278 (1953).

48. R. Zwanzig, *J. Chem. Phys.* **60**, 2717-2720 (1974).
49. D. Hoffmann and E. W. Knapp, *Eur. Biophys. J.* **24**, 387-403 (1996).
50. C. Baysal and H. Meirovitch, *J. Chem. Phys.* **105**, 7868-7871 (1996).
51. K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, *Prot. Sci.* **4**, 561-602 (1995).
52. Z. Guo and D. Thirumalai, *J. Chem. Phys.* **97**, 525-536 (1992).
53. Z. Guo and D. Thirumalai, *Biopolymers* **36**, 83-102 (1995).
54. A. Kolinski, M. Milik, and J. Skolnick, *J. Chem. Phys.* **94**, 3978-3985 (1991).
55. A. Kolinski and J. Skolnick, *J. Phys. Chem.* **97**, 9412-9426 (1992).
56. A. Kolinski and J. Skolnick, *Proteins* **18**, 353-366 (1994).
57. E. I. Shakhnovich, *Folding & Design* **1**, R50-R54 (1996).
58. M. Karplus and A. Sali, *Curr. Opin. Struct. Biol.* **5**, 58-73 (1995).
59. H. S. Chan and K. A. Dill, *Macromolecules* **22**, 4559-4573 (1989).
60. H. S. Chan and K. A. Dill, *Proc. Natl. Acad. Sci. USA* **87**, 6388-6392 (1990).
61. H. S. Chan and K. A. Dill, *J. Chem. Phys.* **95**, 3775-3787 (1991).
62. H. S. Chan and K. A. Dill, *J. Chem. Phys.* **99**, 2116-2127 (1993).
63. L. M. Gregoret and F. E. Cohen, *J. Mol. Biol.* **219**, 109-122 (1991).
64. M.-H. Hao, S. Rackovsky, A. Liwo, M. R. Pinkus, and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **89**, 6614-6618 (1992).
65. N. D. Socci and J. N. Onuchic, *J. Chem. Phys.* **100**, 1519-1528 (1994).
66. G. N. Hunt, L. M. Gregoret, and F. E. Cohen, *J. Mol. Biol.* **241**, 214-225 (1994).
67. A. Kolinski, J. Skolnick, and R. Yaris, *J. Chem. Phys.* **85**, 3585-3597 (1986).
68. A. Kolinski and J. Skolnick, *Proc. Natl. Acad. Sci. USA* **83**, 7267-7271 (1986).
69. K. A. Dill, *Biochemistry* **24**, 1501-1509 (1985).
70. K. F. Lau and K. A. Dill, *Macromolecules* **22**, 3986-3997 (1989).
71. E. Shakhnovich, G. Farztdinov, and A. M. Gutin, *Phys. Rev. Lett.* **67**, 1665-1668 (1991).
72. E. I. Shakhnovich and A. M. Gutin, *Proc. Natl. Acad. Sci. USA* **90**, 7195-7199 (1993).
73. E. I. Shakhnovich and A. M. Gutin, *Protein Eng.* **6**, 793-800 (1993).
74. M. E. Kellman, *J. Chem. Phys.* **105**, 2500-2508 (1996).
75. A. Sali, E. Shakhnovich, and M. Karplus, *J. Mol. Biol.* **235**, 1614-1636 (1994).
76. N. D. Socci, J. N. Onuchic, and P. G. Wolynes, *J. Chem. Phys.* **104**, 5860-5868 (1996).
77. V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, *J. Chem. Phys.* **101**, 6052-6062 (1994).
78. A. Sali, E. Shakhnovich, and M. Karplus, *Nature* **369**, 248-251 (1994).
79. P. E. Leopold, M. Montal, and J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **89**, 8721-8725 (1992).
80. L. A. Mirny, V. Abkevich, and E. I. Shakhnovich, *Folding & Design* **1**, 103-116 (1996).
81. H. S. Chan and K. A. Dill, *J. Chem. Phys.* **100**, 9238-9257 (1994).
82. H. S. Chan and K. A. Dill, *Proteins* **24**, 335-344 (1996).
83. A. R. Dinner, A. Sali, and M. Karplus, *Proc. Natl. Acad. Sci. USA* **93**, 8356-8361 (1996).
84. N. D. Socci and J. N. Onuchic, *J. Chem. Phys.* **103**, 4732-4744 (1995).
85. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, *Proteins* **21**, 167-195 (1995).

86. V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, *Prot. Sci.* **4**, 1167-1177 (1995).
87. C. J. Camacho and D. Thirumalai, *Proteins*, **22**, 27-40 (1995).
88. E. M. O'Toole and A. Z. Panagiotopoulos, *J. Chem. Phys.* **97**, 8644-8652 (1992).
89. T. C. Beutler and K. A. Dill, *Prot. Sci.* **5**, 2037-2043 (1996).
90. K. Yue, K. M. Fiebig, P. D. Thomas, H. S. Chan, E. I. Shakhnovich, and K. A. Dill, *Proc. Natl. Acad. Sci. USA* **92**, 325-329 (1995).
91. M.-H. Hao and H. A. Scheraga, *J. Phys. Chem.* **100**, 14540-14548 (1996).
92. M.-H. Hao and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **93**, 4984-4989 (1996).
93. V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, *Folding & Design* **1**, 221-230 (1996).
94. A. F. P. de Araujo and T. C. Pochapsky, *Folding & Design* **1**, 299-314 (1996).
95. S. Bromberg and K. A. Dill, *Prot. Sci.* **3**, 997-1009 (1994).
96. S.-I. Segawa and T. Kawai, *Biopolymers* **25**, 1815-1835 (1986).
97. M. Karplus and E. Shakhnovich, in *Protein Folding*, T. E. Creighton, ed., Freeman, New York, 1992, pp. 127-196.
98. H. Frauenfelder and P. G. Wolynes, *Phys. Today* **47**, 58-64 (1994).
99. J. D. Bryngelson and P. G. Wolynes, *Biopolymers* **30**, 117-188 (1990).
100. J. D. Bryngelson, *J. Chem. Phys.* **100**, 6038-6045 (1994).
101. J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, and N. D. Socci, *Proc. Natl. Acad. Sci. USA* **92**, 2626-2630 (1995).
102. A. V. Finkelstein, A. Y. Badretdinov, and A. M. Gutin, *Proteins* **23**, 142-150 (1996).
103. N. Go and H. Taketomi, *Proc. Natl. Acad. Sci. USA* **75**, 559-563 (1978).
104. N. Go, H. Abe, H. Mizuno, and H. Taketomi, eds., *Protein Folding*, Elsevier/North Holland, Amsterdam, 1980.
105. H. Taketomi, F. Kano, and N. Go, *Biopolymers* **27**, 527-559 (1988).
106. H. Taketomi, Y. Ueda, and N. Go, *Int. J. Pept. Prot. Res.* **7**, 445-449 (1988).
107. N. Go, *Ann. Rev. Biophys. Bioeng.* **12**, 183-210 (1983).
108. Y. Ueda, H. Taketomi, and N. Go, *Biopolymers* **17**, 1531-1548 (1978).
109. J. Skolnick, A. Kolinski, and R. Yaris, *Proc. Natl. Acad. Sci. USA* **85**, 5057 (1988).
110. J. Skolnick and A. Kolinski, *Annu. Rev. Phys. Chem.* **40**, 207-235 (1989).
111. J. Skolnick, A. Kolinski, and R. Yaris, *Biopolymers* **28**, 1059-1095 (1989).
112. J. Skolnick, A. Kolinski, and R. Yaris, *Proc. Natl. Acad. Sci. USA* **86**, 1229-1233 (1989).
113. J. Skolnick and A. Kolinski, *J. Mol. Biol.* **212**, 787-817 (1989).
114. A. Sikorski and J. Skolnick, *Biopolymers* **28**, 1097-1113 (1989).
115. A. Sikorski and J. Skolnick, *Proc. Natl. Acad. Sci. USA* **86**, 2668-2672 (1989).
116. A. Sikorski and J. Skolnick, *J. Mol. Biol.* **212**, 819-836 (1990).
117. A. Sikorski and J. Skolnick, *J. Mol. Biol.* **215**, 183-198 (1990).
118. A. Kolinski and A. Sikorski, *J. Polym. Sci. Polym. Lett. Ed.* **20**, 177-179 (1982).
119. A. Kolinski and A. Sikorski, *J. Polym. Sci. Polym. Chem. Ed.* **22**, 97-106 (1984).
120. E. O'Toole, R. Venkataramani, and A. Z. Panagiotopoulos, *AIChE J.* **41**, 954-958 (1995).
121. J. Skolnick and A. Kolinski, *J. Mol. Biol.* **221**, 499-531 (1991).
122. R. C. Brower, G. Vasmatiz, M. Silverman, and C. Delisi, *Biopolymers* **33**, 329-334 (1993).
123. A. V. Finkelstein and B. A. Reva, *Nature* **351**, 497-499 (1991).

124. O. V. Galzitskaya and A. V. Finkelstein, *Prot. Eng.* 8, 883-892 (1994).
125. A. V. Finkelstein and B. A. Reva, *Prot. Eng.* 9, 387-397 (1996).
126. A. Rey and J. Skolnick, *Proteins* 16, 8-28 (1993).
127. A. Rey and J. Skolnick, *Chem. Phys.* 158, 199-219 (1991).
128. D. G. Garrett, K. Kastella and D. M. Ferguson, *J. Am. Chem. Soc.* 114, 6555-6556 (1992).
129. J. D. Honeycutt and D. Thirumalai, *Biopolymers* 32, 695-709 (1992).
130. F. Kano and H. Maeda, *Mol. Simul.* 16, 261-274 (1996).
131. J. Skolnick and A. Kolinski, in *Encyclopedia of Computational Chemistry*, P. von Rauge Schleyer and P. Kollman, eds., Wiley, New York, (in press) 1998.
132. I. D. Kuntz, G. M. Crippen, P. A. Kollman, and D. Kimelman, *J. Mol. Biol.*, 106, 983-994 (1976).
133. B. Robson and D. J. Osguthorpe, *J. Mol. Biol.* 132, 19-51 (1979).
134. A. T. Hagler and B. Honig, *Proc. Natl. Acad. Sci. USA* 75, 554-558 (1978).
135. S. Tanaka and H. A. Scheraga, *Macromolecules* 9, 945-950 (1975).
136. G. M. Crippen and V. N. Viswanadhan, *Int. J. Pept. Prot. Res.* 25, 487-509 (1985).
137. F. Fogolari, G. Esposito, P. Viglino, and S. Cattarinussi, *Bophys. J.* 70, 1183-1197 (1996).
138. A. Wallqvist and M. Ullner, *Proteins* 18, 267-289 (1994).
139. S. Miyazawa and R. L. Jernigan, *Macromolecules* 18, 534-552 (1985).
140. J. A. McCammon, S. H. Northrup, M. Karplus, and R. M. Levy, *Biopolymers* 19, 2033-2045 (1980).
141. D. M. Ferguson and D. G. Garret, in *Adaptation of Simulated Annealing to Chemical Optimization Problems*, Vol. 15, by J. H. Kalivas, ed., Elsevier, New York, 1995.
142. W. R. Krigbaum and S. F. Lin, *Macromolecules* 15, 1135-1145 (1982).
143. V. G. Dashevskii, *Molekulyarnaya Biologiya (Engl. transl.)*, 14, 105-117 (1980).
144. D. G. Covell, *Proteins* 14, 409-420 (1992).
145. D. G. Covell and R. L. Jernigan, *Biochemistry* 29, 3287-3294 (1990).
146. D. A. Hinds and M. Levitt, *Proc. Natl. Acad. Sci. USA* 89, 2536-2540 (1992).
147. J. Skolnick and A. Kolinski, in *Computer Simulations of Biomolecular Systems. Theoretical and Experimental Studies*, W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, eds., ESCOM Science Publishers B. V., Leiden, 1997.
148. J. Skolnick and A. Kolinski, *Science* 250, 1121-1125 (1990).
149. A. Godzik, J. Skolnick, and A. Kolinski, *Proc. Natl. Acad. Sci. USA* 89, 2629-2633 (1992).
150. R. S. DeWitte and E. I. Shakhnovich, *Prot. Sci.* 3, 1570-1581 (1994).
151. M. Levitt and J. Greer, *J. Mol. Biol.* 114, 181-293 (1977).
152. A. Godzik, J. Skolnick, and A. Kolinski, *Prot. Eng.* 6, 801-810 (1993).
153. K. A. Olszewski, A. Kolinski, and J. Skolnick, *Prot. Eng.* 9, 5-14 (1996).
154. K. A. Olszewski, A. Kolinski, and J. Skolnick, *Proteins* 25, 286-299 (1996).
155. M. Vieth, A. Kolinski, C. L. Brooks III, and J. Skolnick, *J. Mol. Biol.* 237, 361-367 (1994).
156. D. S. Rykunov, B. A. Reva, and A. V. Finkelstein, *Proteins* 22, 100-109 (1995).
157. A. Monge, E. J. P. Lathrop, J. R. Gunn, P. S. Shenkin, and R. A. Friesner, *J. Mol. Biol.* 247, 995-1012 (1995).
158. B. A. Reva, D. S. Rykunov, A. J. Olson, and A. V. Finkelstein, *J. Comp. Biol.* 2, 527-535 (1995).
159. B. A. Reva, A. V. Finkelstein, M. F. Sanner, and A. J. Olson, *Proteins* 25, 379-388 (1996).
160. B. A. Reva, A. V. Finkelstein, D. S. Rykunov, and A. J. Olson, *Proteins* 26, 1-8 (1996).
161. B. Honig and F. E. Cohen, *Folding & Design* 1, R17-R20 (1996).
162. G. D. Rose and R. Wolfenden, *Annu. Rev. Biophys. Biomol. Struct.* 22, 381-415 (1993).
163. J. H. Dyson and P. E. Wright, *Curr. Biol.* 3, 60-65 (1993).
164. K. Lumb, J. and P. Kim, S., *Biochemistry* 34, 8642-8648 (1995).
165. R. A. Abagyan, *FEBS Lett.* 325, 17-22 (1993).
166. P. Argos, *J. Mol. Biol.* 197, 331-348 (1987).
167. B. H. Zimm and J. K. Bragg, *J. Chem. Phys.* 31, 526-535 (1959).
168. D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers*, Academic Press, New York, 1970.
169. E. I. Shakhnovich and A. V. Finkelstein, *Biopolymers* 28, 1667-1680 (1989).
170. E. I. Shakhnovich, *Phys. Rev. Lett.* 72, 3907-3910 (1994).
171. I. Bahar and R. L. Jernigan, *Biophys. J.* 66, 467-481 (1994).
172. C. J. Camacho and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* 90, 6369-6372 (1993).
173. C. B. Post and B. H. Zimm, *Biopolymers* 18, 1487-1501 (1979).
174. I. C. Sanchez, *Macromolecules* 12, 980-988 (1979).
175. S. F. Betz, D. P. Raleigh, and W. F. DeGrado, *Curr. Opin. Struct. Biol.* 3, 601-610 (1993).
176. C. L. I. Brooks, *Curr. Opin. Struct. Biol.* 3, 92-98 (1993).
177. T. Handel and W. F. DeGrado, *Biophys. J.* 61, A265 (1992).
178. K. Kuwajima, *Proteins* 6, 87-103 (1989).
179. O. B. Ptitsyn, R. H. Pain, G. V. Semisotnov, E. Zerovnik, and O. I. Razgulyaev, *FEBS Lett.* 262, 20-24 (1990).
180. D. P. Raleigh and W. F. DeGrado, *J. Am. Chem. Soc.* 114, 10079-10081 (1992).
181. J. Skolnick, A. Kolinski, and A. Godzik, *Proc. Natl. Acad. Sci. USA* 90, 2099-2100 (1993).
182. A. V. Finkelstein and E. I. Shakhnovich, *Biopolymers* 29, 1681-1694 (1989).
183. R. L. Baldwin and H. Roder, *Curr. Biol.* 1, 219-220 (1991).
184. O. B. Ptitsyn, *Curr. Opin. Struct. Biol.* 5, 74-78 (1995).
185. A. Godzik, A. Kolinski, and J. Skolnick, *J. Comp. Aided Mol. Design* 7, 397-438 (1993).
186. M. Vásquez, G. Nemethy, and H. A. Scheraga, *Chem. Rev.* 94, 2183-2239 (1994).
187. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.* 79, 2361-2381 (1975).
188. M. H. Lambert and H. A. Scheraga, *J. Comp. Chem.* 10, 817-831 (1989).
189. Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* 84, 6611-6615 (1987).
190. H. Meirovitch, M. Vasquez, and H. A. Scheraga, *Biopolymers* 27, 1189-1204 (1988).
191. M. Vásquez and H. A. Scheraga, *Biopolymers* 24, 1437-1447 (1985).
192. H. A. Scheraga, M.-H. Hao, and J. Kostrowicki, in *Methods in Protein Structure Analysis*, M. Z. Atassi and E. Appela, eds., Plenum Press, New York, 1995.
193. M. Vasquez, E. Meirovitch, and H. Meirovitch, *J. Phys. Chem.* 98, 9380-9382 (1994).
194. D. R. Ripoll and H. A. Scheraga, *Biopolymers* 30, 165-176 (1990).
195. A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, *Prot. Sci.* 2, 762-785 (1993).
196. A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, *Prot. Sci.* 2,

- 2725-1731 (1993).
197. A. Liwo, S. Oldziej, J. Ciarkowski, G. Kupryszewski, M. R. Pinkus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, *J. Prot. Chem.* **13**, 375-380 (1994).
198. B. Cheng, A. Nayeem, and H. A. Scheraga, *J. Comp. Chem.* **17**, 1453-1480 (1996).
199. R. J. Wawak, K. D. Gibson, A. Liwo, and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **93**, 1743-1746 (1996).
200. M. Nilges and A. T. Brunger, *Proteins* **15**, 133-146 (1993).
201. U. H. E. Hansmann and Y. Okamoto, *J. Comp. Chem.* **14**, 1333-1338 (1993).
202. Y. Okamoto, *Proteins* **19**, 14-23 (1994).
203. H. Meirovitch and E. Meirovitch, *J. Phys. Chem.* **99**, 4847-4854 (1995).
204. E. Meirovitch and H. Meirovitch, *Biopolymers* **38**, 68-88 (1996).
205. H. Meirovitch and E. Meirovitch, *J. Phys. Chem.* **100**, 5123-5133 (1996).
206. W. F. van Gunsteren and P. K. Weiner, *Computer Simulations of Biomolecular Systems. Theoretical and Experimental Applications*, ESCOM Science Publishers B.V., Leiden, 1989.
207. G. M. Crippen and M. E. Snow, *Biopolymers* **29**, 1479-1489 (1990).
208. G. M. Crippen, *Biochemistry* **30**, 4232-4237 (1991).
209. R. Srinivasan and G. D. Rose, *Proteins* **22**, 81-99 (1995).
210. M. Vieth, A. Kolinski, C. L. Brooks, III, and J. Skolnick, *J. Mol. Biol.* **251**, 448-467 (1995).
211. A. Godzik, A. Kolinski, and J. Skolnick, *Prot. Sci.* **4**, 2107-2117 (1995).
212. A. Godzik, *Current Biol.* **4**, 363-366 (1996).
213. M. Milik, A. Kolinski, and J. Skolnick, *Prot. Eng.* **8**, 225-236 (1995).
214. R. Friesner and J. R. Gunn, *Annu. Rev. Biophys. Biomol. Struct.* **25**, 315-342 (1996).
215. B. Persson and P. Argos, *J. Mol. Biol.* **237**, 182-192 (1994).
216. T. Dandekar and P. Argos, *J. Mol. Biol.* **256**, 645-660 (1996).
217. C. Mumenthaler and W. Braun, *Prot. Sci.* **4**, 863-871 (1995).
218. M. J. Smith-Brown, D. Kominos, and R. M. Levy, *Prot. Eng.* **6**, 605-614 (1993).
219. A. Aszodi, M. J. Gradwell, and W. R. Taylor, *J. Mol. Biol.* **251**, 308-326 (1995).
220. J. Skolnick, A. Kolinski, and A. R. Ortiz, *J. Mol. Biol.* **265**, 217-241 (1997).
221. A. Ortiz, A. Kolinski, and J. Skolnick, *J. Mol. Biol.* **277**, 419-448 (1998).
222. A. Ortiz, A. Kolinski, and J. Skolnick, *Proc. Natl. Acad. Sci. USA* **95**, 1020-1025 (1998).
223. B. Rost and C. Sander, *J. Mol. Biol.* **232**, 584-599 (1993).
224. A. Kolinski, J. Skolnick, A. Godzik, and W-P. Hu, *Proteins* **27**, 290-308 (1997).
225. U. Gobel, C. Sander, R. Schneider, and A. Valencia, *Proteins* **18**, 309-317 (1994).
226. A. Godzik, J. Skolnick, and A. Kolinski, *J. Mol. Biol.* **227**, 227-238 (1992).
227. F. H. C. Crick, *Acta Cryst.* **6**, 689-697 (1953).
228. A. Holtzer, M. E. Holtzer, and J. Skolnick, in *Protein Folding. Deciphering the Second Half of the Genetic Code*, Vol. 4, L. M. Gierash, and J. King AAAS, Washington, DC, 1990, pp. 2401-2411.
229. M. Nilges and A. T. Brunger, *Prot. Eng.* **4**, 649-659 (1991).
230. S. R. Krystek Jr., R. E. Bruccoleri, and J. Novotny, *Int. J. Pept. Prot. Res.* **38**, 229-236 (1991).
231. L. Zhang and J. Hermans, *Proteins* **16**, 384-392 (1993).
232. P. Harbury, II., B. Tidor, and P. S. Kim, *Proc. Natl. Acad. Sci. USA* **92**, 8408-8412 (1995).
233. P. B. Harbury, T. Zhang, P. S. Kim, and T. Alber, *Science* **262**, 1401-1407 (1993).
234. M. Vieth, A. Kolinski, and J. Skolnick, *Biochemistry* **35**, 955-967 (1996).
235. K. J. Lumb, C. M. Carr, and P. S. Kim, *Biochemistry* **33**, 7361-7367 (1994).
236. B. Lovejoy, S. Choe, D. K. McRorie, W. F. DeGrado, and D. Eisenberg, *Science* **259**, 1288-1293 (1992).
237. M. Milik and J. Skolnick, *Proteins* **15**, 10-25 (1993).
238. M. Milik and J. Skolnick, *Biophys. J.* **69**, 1382-1386 (1995).
239. P. McDonnell, K. Shon, Y. Kim, and S. Opella, *J. Mol. Biol.* **233**, 447-463 (1993).
240. K. J. Shon, Y. Kim, L. A. Colnago, and S. J. Opella, *Science* **252**, 1303-1305 (1991).
241. A. Baumgartner and J. Skolnick, *Phys. Rev. Lett.* **74**, 2142-2145 (1995).
242. A. Baumgartner and J. Skolnick, *J. Phys. Chem.* **98**, 10655-10658 (1994).
243. C. Levinthal, *J. Chim. Phys.* **65**, 44-45 (1968).
244. R. Zwanzig, A. Szabo, and B. Bagchi, *Proc. Natl. Acad. Sci. USA* **89**, 20-22 (1992).
245. P. G. Wolynes, in *Protein Folds. A Distance-Based Approach*, H. Bohr and S. Brunak, ed., CRC Press, New York, 1996, pp. 3-17.
246. P. G. Wolynes, in *Physics of Biomaterials: Fluctuations, Selfassembly and Evolution*, T. Riste and D. Sherrington, eds., Kluwer Academic Publishers, Netherlands, 1996, pp. 235-248.
247. D. Eliezer, P. A. Jennings, P. E. Wright, S. Doniach, K. O. Hodgson, and H. Tsuruta, *Science* **270**, 487-488 (1995).
248. W. F. DeGrado, Z. R. Wasserman, and J. D. Lear, *Science* **243**, 622-628 (1989).
249. T. M. Handel, S. A. Williams, and W. F. DeGrado, *Science* **261**, 879-885 (1993).
250. D. P. Raleigh, S. F. Betz, and W. F. DeGrado, *J. Am. Chem. Soc.* **117**, 7558-7559 (1995).
251. C. Sander, G. Vriend, F. Bazan, A. Horovitz, H. Nakamura, L. Ribas, A. V. Finkelstein, A. Lockhart, R. Merkl, L. J. Perry, *Proteins* **12**, 105-110 (1992).
252. M. H. Hecht, J. S. Richardson, D. C. Richardson, and R. C. Ogden, *Science* **249**, 884-891 (1990).
253. G. Bohm, *Biophys. Chem.* **59**, 1-32 (1996).
254. M. Levitt, *Curr. Opin. Struct. Biol.* **1**, 224-229 (1991).
255. J. W. Ponder and F. M. Richards, *J. Mol. Biol.* **193**, 775-791 (1987).
256. H. Flockner, M. Braxenthaler, P. Lackner, M. Jaritz, M. Ortner, and M. J. Sippl, *Proteins* **23**, 376-386 (1996).
257. D. T. Jones, W. R. Taylor, and J. M. Thornton, *Nature* **358**, 86-89 (1992).
258. R. Luethy, J. U. Bowie, and D. Eisenberg, *Nature* **356**, 83-85 (1992).
259. S. H. Bryant and C. E. Lawrence, *Proteins* **16**, 92-112 (1993).
260. S. J. Wodak and M. J. Rooman, *Curr. Opin. Struct. Biol.* **3**, 247-259 (1993).
261. J. M. Thornton, T. P. Flores, D. T. Jones, and M. B. Swindells, *Nature* **354**, 105-106 (1991).
262. P. D. Thomas and K. A. Dill, *J. Mol. Biol.* **257**, 457-469 (1996).
263. M. J. Sippl and S. Weitckus, *Proteins* **13**, 258-271 (1992).
264. C. Chothia and A. Finkelstein, *Annu. Rev. Biochem.* **59**, 1007-1039 (1990).
265. J. Richardson and D. C. Richardson, in *Prediction of Protein Structure and the Principles of Protein Conformation*, G. D. Fasman, ed., Plenum Press, New York, 1989, pp. 1-99.

266. V. N. Maiorov and G. M. Crippen, *Proteins* **20**, 167-173 (1994).
 267. A. R. Ortiz, W.-P. Hu, A. Kolinski, and J. Skolnick, *J. Mol. Graph.* (in press).
 268. J. Novotny, R. Brucolleri, and M. Karplus, *J. Mol. Biol.* **177**, 787-818 (1984).
 269. A. Elofsson and L. Nilsson, *J. Mol. Biol.* **223**, 766-780 (1993).
 270. S. DeBolt and J. Skolnick, *Prot. Eng.* **9**, 637-655 (1996).
 271. R. K. Singh, A. Tropsha, and I. I. Vaisman, *J. Comput. Biol.* **3**, 213-221 (1996).
 272. A. Tropsha, R. K. Singh, I. I. Vaisman, and W. Zheng, in *Pacific Symposium on Bio-computing'96*, L. Hunter and T. E. Klein, eds., World Scientific, Singapore, 1996, pp. 614-623.
 273. D. C. Rapaport, J. E. Johnson, and J. Skolnick, *Biophys. J.*, submitted.

ENTROPY SAMPLING POLYPEPTIDES

HAROLD A. SCHERAGA

Baker Laboratory of Chemistry, Cornell University

CONTENTS

- I. Introduction
- II. The Microcanonical Approach to Protein Folding
- III. The Entropy Sampling Monte Carlo Method
- IV. The Quasi-Ergodicity Problem in Entropy Sampling
- V. Effective Sampling Techniques for Entropy Sampling
- VI. Applications of the ESMC Method to Protein Folding
- VII. Comparisons with Related Monte Carlo Methods
- VIII. Conclusions

I. INTRODUCTION

Monte Carlo (MC) methods have been used in simulation studies of polypeptides and globular proteins. Our laboratory has developed and improved these methods for treating different problems concerning protein folding. These MC techniques can be categorized into two groups: those for finding the native state of a peptide and those for studying the statistical properties of the native state.

In searching for the lowest-energy conformation, a major technical requirement for an MC method is to find an energy conformation as quickly as possible. The ESMC method for finding the global-energy minimum is a Monte Carlo method that samples the conformational space in a target region of the conformational space.

Advances in Chemical Physics, Volume 105, Monte Carlo Simulation of Protein Folding, edited by David M. Ferguson, J. Ilja Siepmann, and David A. McQuibban, and Stuart A. Rice.
 ISBN 0-471-19630-4 © 1999 John Wiley & Sons