Available online at www.sciencedirect.com
SCIENCE DIRECT®
Polymer 45 (2004) 511–524
polymer
www.elsevier.com/locate/polymer

Reduced models of proteins and their applications

Andrzej Kolinski[a,b,*], Jeffrey Skolnick[b]

[a]*Laboratory of Theory of Biopolymers, Faculty of Chemistry, Warsaw University, Pasteura 1, 02-093 Warsaw, Poland*
[b]*UB Center of Excellence in Bioinformatics, University at Buffalo, 901 Washington Street, Buffalo, NY 14203, USA*

Received 12 March 2003; received in revised form 11 June 2003; accepted 12 June 2003

Abstract

Reduced computer modeling of proteins now has a history of about 30 years. In spite of the enormous increase in computing abilities, reduced models are still very important tools for theoretical studies of protein structure, dynamics and thermodynamics. Very simple, highly idealized lattice (and recently also off-lattice) models could be studied in great detail, providing valuable insight into the most general factors governing structure stability, folding kinetics and interactions responsible for characteristic two-state behavior near the folding temperature. More complex models now enable modeling of real proteins on the level of low to moderate resolution, allowing us to address more detailed questions. Ab initio protein structure predictions, still being far from a routine task, have become feasible. When supported by evolutionary information from multiple sequence alignments and potential local and/or global structural similarity to known structures, reduced modeling opens up new areas of comparative modeling, thereby complementing contemporary structural genomics.

Keywords: Reduced protein models; Lattice proteins; Protein folding

1. Introduction

Proteins are fascinating molecular objects. In spite of a very complex composition of 20 amino acids of various size and shape, most proteins can adopt a unique, three dimensional structure that is necessary for their biological function [1]. The ability to achieve specific, on some levels quite regular, packing of their side chains has been developed during the course of evolution [2,3]. Indeed, the vast majority of random copolymers of amino acids do not behave as proteins; they do not collapse to a unique native-like conformation. The sequences of natural polypeptides frequently look 'almost' random. Therefore, proteins could be described as 'edited' (by evolution) random copolymers. The free energetic stabilization of the folded native structure of proteins in respect to the denatured states is small—equivalent to the energy of a few hydrogen bonds [1]. However, the free energy barrier between the native state and the manifold of denatured conformations is huge. Consequently, the denaturation–renaturation process has many properties of a first-order phase transition [4]. This is a very striking feature of so small objects, composed of tens of hundreds of atoms. The folding transition of single domain globular proteins is highly cooperative [5] and is frequently abbreviated as an all-or-none transition to avoid referring to the first-order phase transition in the case where the thermodynamic limit is certainly not satisfied. It is postulated that the folded state of proteins corresponds to the global minimum of their free energy. This hypothesis, first articulated by Anfinsen [2,3], has been proven to be true for the majority of proteins; however, there seem to be some exceptions. The folding process, although slow, is much faster than a random search of the enormous conformational space of polypeptide chains. This so-called Levinthal's [6,7] 'paradox' could be easily explained by a very rapid sequential reduction of available conformational space—for instance, via a loosely hierarchical formation of elements of secondary [8–14] (and supersecondary) structure.

Various purely theoretical models of proteins [15–19] are quite limited in their ability to explain the unique properties of proteins. This is due to the complexity of intramolecular interactions, sequence effects, and the difficulty in separating the effects of physical forces from the evolutionary selected specific patterns and regularities. Consequently, computer modeling plays a very important

* Corresponding author. Address: Laboratory of Theory of Biopolymers, Faculty of Chemistry, Warsaw University, Pasteura 1, 02-093 Warsaw, Poland.
E-mail addresses: kolinski@chem.uw.edu.pl (A. Kolinski), skolnick@buffalo.edu (J. Skolnick).

0032-3861/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2003.10.064

role in understanding the nature of proteins. There are various purposes of protein modeling, as outlined below.

Very important is to understand the basic physics of protein dynamics and thermodynamics. In particular, the elucidation of the various forces which drive the folding process and determine the uniqueness and stability of the native state. For this purpose, the models should be rather simple to make the interpretation of computer experiments feasible and as unambiguous as possible [20–26]. On the other hand, too simple models may miss important aspects of physics. A possible answer to this problem may come from comparison of various models of increased fidelity [22,27] (and consequently increasing complication).

Another reason for studying reduced models is to build algorithms which could be used for protein structure prediction [28–46] and possibly for prediction of the protein folding pathways [47,48]. The second task seems to be generally more difficult. While in many cases of close homology modeling (or comparative modeling) the prediction of protein structure may solely rely on the detailed atomistic models [49], the reduced models seem to be necessary tools for more challenging ab initio structure predictions [28,34,35,39,40,42,46,50–55] or in structure refinement from distant homology modeling or threading based comparative modeling [55–57]. Reduced models require the explicit treatment of much smaller numbers of degrees of conformational freedom and usually have a less rugged energy landscape then detailed atomic models [58,59]. Therefore, the cost of computations decreases by orders of magnitude, and the task of global minimization [42,60,61] of the protein conformational energy becomes tractable. The role of reduced models is probably even more fundamental in the prediction of protein folding pathways and the mechanism of protein–protein association [62–71]. It is still impractical (due to enormous computational cost) to simulate a full folding process of even relatively small proteins using detailed molecular models [72]. Related to ab initio protein structure prediction is model building based on fragmentary experimental data [73]—for example starting from NMR assigned secondary structure (partial or complete) and/or from a small number of known side chain contacts from NMR [74–76] or other experiments [77] as crosslink walks or tryptophan fluorescence. Building good molecular models from electron microscopy still remains a big challenge and may also benefit from the application of properly designed reduced models in the near future.

The design and application of reduced protein models have several, to some extent separable (and exchangeable between the models), conceptual levels. The first is the choice of the representation of the polypeptide chain. There are at least two choices that have to be made at this stage. The most important is the level of detail intended for reproduction. The spectrum of explored possibilities is large—from a single interaction center per residue (or even a small number of residues treated as a single unit [37,78,79]) to several united atoms per residue. In the last case, various choices are being made. In some models, all main chain atoms are treated in an explicit way while in others single united atoms replace the side groups. Sometimes, the main chain could be replaced by the alpha-carbon trace and the side chains partitioned (depending on their size) into several interacting units. Models that use an all-atom representation of the main chain and several united atoms per side group are rare, because they are too close in the cost of computations to the all-atom models and their advantages become marginal. Nevertheless, applications exist which employ the all-atom representation of the polypeptide chains with a simplified, highly reduced force field and a specific sampling scheme. The second important decision that needs to be made during the design of a protein model is the choice between a continuous space and a discretized lattice representation of the conformational space. Continuous models, as a more straightforward method, do not require any general comment. As for the lattice-type models, there are a variety of possibilities. Many studies have been done for simple cubic lattice models and other low coordination lattice models that highly idealize the nature of proteins. However, it is possible to design quite detailed lattice-confined polypeptide chains. The resolution of lattice models [22,80,81] can vary from a very crude shape of the main chain to a resolution similar to that of good experimental structures [56]. Usually, the main chain is restricted to a lattice. The side chain, if explicitly treated, could be restricted to a lattice or could be allowed to occupy off-lattice positions. It is possible to reverse the situation—restricting the positions of the side chains to a lattice and allowing off-lattice positions of the main chain atoms. The main disadvantages of the lattice models are related to distortions of local geometry of polypeptide chains, the limited resolution of the model, the possibility of anisotropy effects (depending on the resolution of particular lattices), some restrictions on the available selections of the sampling methods and the necessity of discretization of the force field. For high coordination lattice models, these apparent disadvantages are negligible, and such models can achieve significantly better performance than that of otherwise equivalent continuous space models.

Strictly connected to the representation of protein geometry is the range of available schemes for conformational updating. The advantage of the continuous models is the possibility of adaptation of various schemes typical for classical molecular dynamics (MD). On the other hand, lattice models allow for the easy design of local conformational transitions that avoid local energy barriers, thereby increasing sampling efficiency. Moreover, lattice models enable the precalculation of entire sets of some conformational transitions and some elements of the force fields in the form of 'prefabricated' reference tables. This additionally speeds-up the sampling, and the dynamics of the model system over a period of time can be simulated.

It is obvious that the design of the model force field is to some extent dictated by the choice of geometrical

representation of the polypeptide chains. Having a single interacting center per residue requires a potential function that simultaneously accounts for a variety of physical interactions including hydrogen bonds, averaged Van der Waals interactions, hydrophobic effects, etc. When the main chain is represented in a more explicit fashion, a directional potential that mimics hydrogen bonds can be separated from other interactions—for instance, from the pairwise interactions between the side chains. The optimal shape of a potential function also depends on the representation [82]. For continuous models, a Lennard–Jones type could be appropriate [36]. On the contrary, for lattice models, the bottom of the potential curve needs to be flat, with the width adjusted to the lattice spacing. Otherwise, lattice anisotropy artifacts could be severe [83].

The next and final stage of the model design is the choice of a sampling or a conformational search scheme (or schemes). Again, to some extent, it has to be a consequence of the selected representation and the force field design. Obviously, classical MD algorithms are not applicable to the lattice models.

The type of representation, design of the force field and the sampling scheme employed determine the range of applicability of a given reduced model. It would be nonsensical to use cubic lattice chains to model the loops in comparative modeling tasks. Similarly nonsensical would be an attempt to enumerate all conformations of a polypeptide restricted to a high coordination lattice with tens or hundreds of possible orientations of the virtual bonds between the Cαs.

The remainder of this review follows the line of the above general remarks on the design of reduced models. We start from a review of various representations of the conformational space. This is followed by a discussion of different methods of designing interaction schemes for various models. Afterwards, various sampling techniques are described. Finally, we outline some typical applications of reduced models and their meaning for the theoretical and practical study of proteins.

Due to the very large number of publications on reduced protein models, it is impossible to make references to all of them in a short review. The main purpose of this contribution is to cover as diverse as possible a spectrum of reduced models and simulation techniques. Thus, for various approaches, we have attempted to choose a representative set of references that might not necessarily be proportional to the actual impact of particular studies on protein science. We hope that some techniques developed for studies of protein (and outlined here) can also be useful in theoretical studies of polymers in general, similar to the flow of ideas from polymer sciences into the more specific field of protein theory.

2. Protein representation in reduced models

2.1. Simple lattice models of protein-like polypeptides

Probably the simplest models of protein-like systems employ a simple cubic lattice (sc) representation of the polypeptide chain. An example of a short lattice chain with two types of residues is shown in Fig. 1A. Two types of chain beads reflect a crude approximation of the two types of residues: hydrophobic (H) and polar (P). This so called HP model [84] has been extensively studied by Chan and Dill [85–94], Sali and coworkers [95–97], Shakhnovich and coworkers [27], Karplus and coworkers [21,96,97] and many others [24,98–106]. Modifications of this model include different codes for sequence [23,95,107–121] (from a simple collapsing homopolymer to full twenty-residue types of sequences [117,122]), more complex (and more realistic) interaction schemes [23,25,114,116,118,122–128] effects of other molecules [66,90,129–135] a simple lattice representation of the side chains [62,128,133,136–140], and modifications of the geometry of the HP model [98,141, 142]. An excellent review of earlier studies of this model can be found in work of Dill et al. [20]. The simplicity of the HP type models is an obvious advantage, making it possible to study the model systems in great detail [20,143]. Probably the main weakness of the model is related to the lack of a clear notion of secondary structure [144]; however some studies were done with a symbolically defined secondary structure in simple cubic lattice chains [23–25, 41,144–148]. Secondary structure is an important feature of proteins, enabling the fast partitioning of the conformational

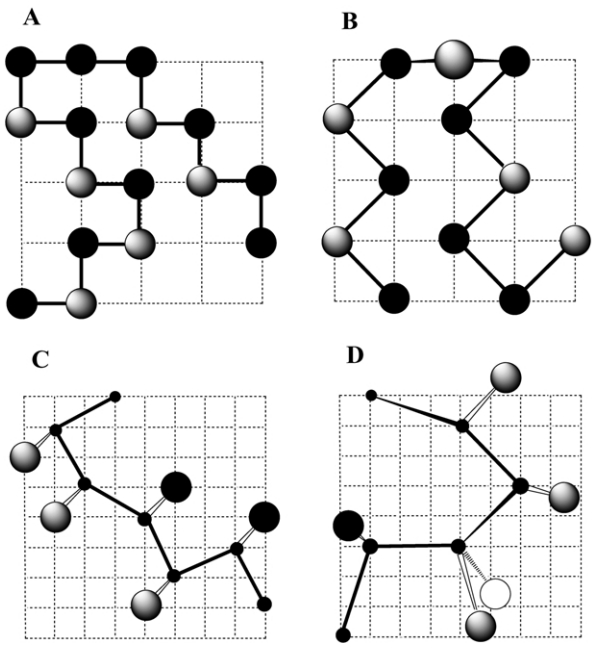

Fig. 1. Examples of simple lattice and moderate resolution lattice models of proteins. (A) Simple cubic lattice (a planar example) chain with two types of residues. (B) Lattice chain restricted to the face centered cubic lattice (fcc) with two types of residues (almost planar example, one residue in the loop is placed out of plane). (C) 'Chess knight' 210 lattice model (24 possible orientations of the Cα-trace vectors) of polypeptide (planar example of the main chain) where the side chain centers are also restricted to the underlying cubic lattice and the side chain vectors are type of (±1, ±1, ±1). (D) 'Hybrid' 310 lattice model (90 possible orientations of the Cα-trace vectors) with off-lattice side chains and multiple rotamers.

space during the folding process [9,16]. Nevertheless, the idea of hydrophobic collapse [116] certainly describes an important aspect of protein folding. Indeed, studies by Go et al. [136,149–153], Crippen [154] and others [155,156], clearly show that even some overall geometrical resemblance to real proteins could be achieved.

Other simple lattices were employed in the modeling of real proteins [141] (Dashevskii [157], Levitt et al [78], Krigbaum and Lin [158], Reva et al [138]) as well as in studies of protein like systems (Kolinski and Skolnick [159–163], Sikorski and Skolnick [164–166]). Diamond, body-centered-cubic (bcc) and face-centered-cubic (fcc) lattices enable a somewhat better description of the local geometry of polypeptides than was possible for sc-lattice models. Consequently, it becomes possible to build helices and β-sheets with a reasonable (but crude) similarity to the patterns seen in real proteins. A qualitative account of the hydrogen bonding is also feasible in this type of model. An example of a short fcc-restricted polypeptide is schematically shown in Fig. 1B. Actually, the fcc lattice could be the best choice among the simple regular lattices [26,141,167,168]. This lattice describes relatively well the local packing in folded proteins [169]. It is noteworthy that some models do not follow a one-to-one correspondence between the lattice beads and residues of modeled proteins [78,136,170]. For instance, Levitt and coworkers used diamond lattice compact chains, with more than one residue per lattice unit [78,170], to model decoys that were subject to further analysis in their hierarchical methodology of ab initio protein structure prediction [79,171].

2.2. Moderate resolution lattice models

It is possible to design relatively simple lattice models that allow the moderate resolution modeling of the local geometry of polypeptides while maintaining most of the computational simplicity of the simple lattice models. One such model is the three-dimensional 'chess knight' model [9,172–174] illustrated in Fig. 1C. The main chain Cα trace is represented by a chain of vectors type (±2, ±1, 0). The number of such virtual Cα–Cα bonds is equal to 24 (4 permutations of the signs times six permutations of the digits). The side chains of this model are also restricted to the underlying cubic lattice and the centers of the side chains are separated from the main chain vertices by diamond lattice vector type of (±1, ±1, ±1). The side chain vectors are selected to be as close as possible to the orientation of average rotamers in the database. Of course, the direction of the side chain is defined by the geometry and orientation of the two corresponding main chain vectors. The model allows a very simple control of excluded volume and various contact interactions [172]. The lattice excluded volume of a model residue included the Cα vertex, six neighboring points of the underlying lattice and the three closest lattice points to the center of the side chain (except glycine, which lacks of side group). The 'chess knight' model enables the low resolution study of real proteins [9,174,175] and a quite detailed (due to protein like geometry) study of idealized protein-like geometrical motifs [173] (helical bundles, complex topologies of β-type proteins and mixed α/β motifs). Unfortunately, the 'chess knight' model exhibits quite significant spatial anisotropy [83]. For instance, acceptably regular helices propagate only along a few directions of the lattice. Of course, the effect is much less severe than for regular low coordination lattices [83].

For the purpose of studying real proteins, the chess-knight model has been modified. Namely, in order to breakdown the effect of anisotropy, additional possible main chain vectors were added. The hybrid 210 model [28,176,177] employs 56 basis vectors of the type (±1, ±1, ±1), (±2, ±1,0) and (±2, ±1, ±1). The main chain excluded volume is modeled in the same fashion as was done for the 'chess knight' model. This is one of the reasons why vectors type (±2, 0, 0) are not included in the set of the allowed Cα–Cα orientations. The side chains are spherical, off lattice and represent the coarse-grained library of rotamers. The best fit of the experimental structures to the 210 hybrid lattice chains is obtained when the spacing of the underlying sc-lattice is assumed to be equivalent to 1.7 Å. As a result, the average fitting accuracy is better than 1.0 Å [83,178]. Of course, this does not mean that the accuracy of structures folded with this model is this good. The accumulation of inaccuracies from various potentials (impossible to avoid in lattice models [179]) significantly decrease the practical fidelity of the model.

Using a similar philosophy, a higher resolution 310 hybrid lattice model [29,31–33,53,54,60,68,69,180–186] with 90 basis vectors for the Cα–Cα trace was developed and employed in several studies of protein stability, dynamics and thermodynamics. The idea of this polypeptide representation is depicted in Fig. 1D. Similar moderate resolution models were later designed in several laboratories [124,187,188].

2.3. High coordination lattice models

The majority of reduced protein models assume that the most convenient reference frame for the geometry of the entire model is the main chain. Indeed, the main chain segments in proteins exhibit higher regularity than the mutual orientations of the side chains. On the other hand, the interactions between the main chain units are, to a large extent, generic—i.e. they do not depend too much on sequence. The sequence specificity of intraprotein interactions is much higher for the side chains. Packing of side chains is probably the main factor responsible for the native structure of a protein. Taking the above into consideration, a model that focuses on the side chains has been developed (see Fig. 2A). SICHO [71,74,81,189–192] (Side Chain Only) uses an extremely simple representation in the form of strings and beads chains connecting the centers of mass of the side groups in their actual rotational isomeric

conformations. The model allows a large number (646) of possible virtual 'bonds' between side groups. The lengths of these bonds cover a broad distribution of intraprotein distances, from 3.8 Å between two consecutive glycines to over 10 Å between centers of large side groups. The distributions are controlled by appropriate statistical potentials. Such a defined model is very flexible and easily driven by the interactions between the side chains. The positions of the Cαs are defined in an approximate fashion, using the positions of the three consecutive side groups. The lattice spacing of the underlying cubic lattice is equal to 1.45 Å. Thus, the limit of accuracy is about 0.8 Å for the cRMSD (coordinate root mean square deviation after the best superimposition of the centers of mass of the side groups). The test of spatial proximity of the interacting groups is done via a local lattice search algorithm. Thus, the algorithm scales nicely with the chain length. Consequently, relatively large systems could be effectively simulated (a few hundred of residues). Interestingly, the overall accuracy of this model (including accuracy of the main chain) is higher than the accuracy of the 310 hybrid lattice model, for which the spacing of the lattice is equal to 1.22 Å. This means that, due to statistical cancellation of errors, the approximation of the Cα trace is actually more accurate than

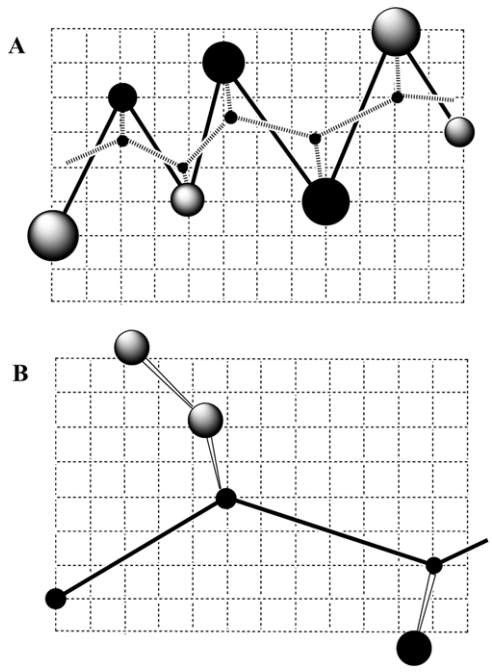

Fig. 2. Two types of high coordination lattice models. (A) Side CHain Only (SICHO) model with the centers of side groups (SG) restricted to the underlying cubic lattice. The number of possible SG-SG vectors is equal to 646 and covers the distribution observed in real proteins, with the assumption that the lattice constant is equal to 1.45 Å. The broken lines and small black dots show an implicit (calculated from the position of the side chains) Cα-trace (off-lattice). (B) CABS model (Cα–Cβ–SG) with lattice representation of the main chain trace (800 possible Cα–Cα vectors) and the beta carbons and centers of mass of the side chains placed off-lattice. Different rotamers are used for expanded and compact conformations of the main chain and the lattice spacing is equal to 0.61 Å.

the accuracy of the side chain positions. In other words, the reconstruction of the main chain atoms from the known positions of the centers of the side chains is a better defined task than the reconstruction of the side groups from a given main chain. Thus, the SICHO model is a valuable element of multiscale simulation tools in molecular biology [193]. The actual fidelity of the SICHO model (due to square well potentials used and other approximations in the interaction scheme) is obviously lower and varies from 2 Å for small proteins to 5–6 Å for larger ones. This is probably satisfactory for ab initio fold prediction or refinement of a low resolution threading models [34,35,55,57,73]. The resolution is sometimes satisfactory for applications in homology modeling [194]; however, when a high resolution is necessary, a model of higher accuracy is needed.

The recently developed CABS (Cα–Cβ–Side groups) has three interaction centers per residue and 800 allowed Cα–Cα vectors (see Fig. 2B) The positions of up to two side chain united atoms are average rotameric states for a given residue type and a given conformation of the main chain fragment. This is probably a much better approximation than it might at first appear [38]. The lattice spacing is equal to 0.61 Å. The resulting average cRMSD for the Cα-trace of PDB structures fitted to this lattice is in the range of 0.35 Å. The actual resolution of the model is in the range of 1–2 Å depending on protein size, which makes it an appropriate tool for fold refinement in comparative modeling, fold assembly based on sparse NMR data (and experimental data from other sources) and ab-initio folding supported by evolutionary derived restraints. While the algorithm is much less computationally demanding than equivalent continuous space models, it is still too expensive for purely ab initio folding of larger proteins, assuming the force field would be adequate to the task, which at present, it is not.

2.4. Continuous space models

Similar to the lattice models, continuous space models employ various levels of generalization. The number of structural details and the selection of explicitly treated degrees of freedom differ significantly in various models. There are two general classes of continuous space models (similarly to the lattice models): idealized models of protein like systems and reduced models of real proteins. The idealized models resemble very much some lattice models except they are sampled in a continuous space [195–203].

The majority of reduced continuous space models are aimed at the study of real proteins of various levels of detail. In the classic work of Levitt and Warshel [36,204] the polypeptide chain has been reduced to the Cα trace and spherical, single united atom, side chains (see Fig. 3A). The planar angle was kept at a single fixed value for all residues. This is probably an unnecessary simplification since the distribution of this angle exhibits two well defined maxima that correspond to compact (for instance helical) and

expanded conformations (β-strands or expanded coil fragments), respectively. On the other hand, it is a low resolution model and somewhat adjusted dihedral angle preferences employed in the model (and derived from conformational analysis of representative dipeptides) should correct for this approximation. This model and its variants were subsequently used in other studies of protein folding [205–208].

A completely different approach to protein chain modeling in a continuous space is represented by the model developed by Hoffmann and Knapp [209,210]. Here (see Fig. 3B), the main chain is pictured as a string of rigid peptide bond plates in their *trans*-conformations. Actually, an independent unit of the chain consists of three successive peptide bond plates. The authors have shown that there are collective transformations of such units (involving simultaneous and coordinated changes of their $\Phi-\Psi$ angles) that leave unchanged the remaining (outside a 3-bond window) portions of the chain. This led to an impressive speed-up of the sampling process with respect to models with independent, more local, conformational transitions. Interestingly, such types of moves are typical for almost all of the more complex lattice-confined models [22,81,183]. Thus, besides other interesting results, the Hoffmann and Knapp work provides a qualitative explanation for one of the aspects responsible for much faster sampling of the lattice models.

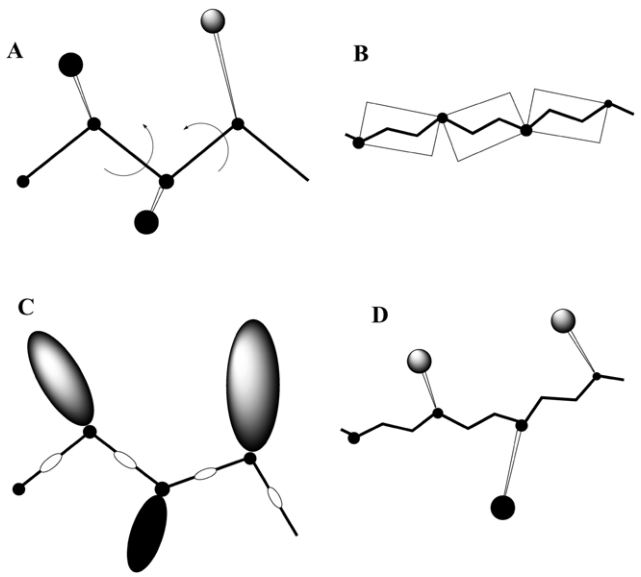

Fig. 3. Examples of continuous space reduced models of polypeptide chains. (A) Model proposed by Levitt and Warshel. Two united atoms per residue, with the center of a side chain equivalent to the center of mass of its most probable rotamer. The planar angle for the Cα-trace assumed constant. (B) Model proposed by Hoffmann and Knapp with rigid peptide bond plates moving in a collective fashion (three plates has to be rotated simultaneously). (C) UNRES (UNited RESidues) model of Liwo, Scheraga and co-workers. Two centers of interactions per residue; center of peptide bond (open ellipsoids) and centers of side chains. Ellipsoids symbolize asymmetric character of interactions. (D) A model with all atom representation of the main chain and reduced representation of the side chains.

A very elegant UNRES (UNited RESidues) model developed by Liwo, Scheraga and coworkers (see Fig. 3C) assumes only two centers of interaction per residue. One is located in the center of the peptide bond; the second is an ellipsoidal side chain. Due to the asymmetry of the united atoms, and asymmetry or/and directionality of their interactions, the model accounts surprisingly well for most of the physical interactions observed in proteins [41,211,212] and allows for high geometrical fidelity. The concept of ellipsoidal side chains was also explored by others [208].

Frequently, the main chain is modeled on the all-atom level, while a reduced representation is assumed for the side groups [144,213–216]. An example of this is the model studied by Sun [217,218] (Fig. 3D), with a single sphere approximation of the side chains.

A separate class of reduced models of proteins are the continuous space, off-lattice models in which some degrees of freedom are locally discretized [219–222]. Usually, the purpose of such a discretization is related to the discretization of potentials or/and is aimed at achieving better computational efficiency [221,223].

Finally, there are models that employ an all atom representation of the conformational space; however their force fields are simplified [224–228]. Such models should probably be also considered as reduced models. The most well known works focus on protein structure assembly from polypeptide fragments excised from the experimentally solved protein structures [50]. The ROSETTA algorithm of Baker and coworkers [37,46,51], being one of the most innovative and successful methods of ab initio structure prediction, is a good example of such an approach to protein modeling.

3. Interaction schemes in reduced models

The level of detail in the design of a force field strongly depends on the complexity of protein representation in reduced models. In simple lattice models, there is always some simplified potential of the long range interactions. Numerous studies have been done with the binary code for residues as in the HP model of Chan and Dill [20], where H are hydrophobic residues and P are polar, hydrophilic residues. They interact via a contact potential (nearest nonbonded lattice neighbors) with different strengths of interactions for HH, PP and HP pairs [20]. In other simple lattice models, a Go [149–151] type potential is used, and only the residues that are ‘in contact’ in the assumed ‘native’ conformation interact. In various modifications of the HP-type models, the pairwise long-range potentials were expanded, accounting for more than two types of residues [23] (up to 20 amino acids [122]) or a protein like distribution of the strength of binary interactions was assumed (as in the random energy model [114]). In some studies of simple lattice models, knowledge-based, statistical potentials were applied in order to generate low

resolution (cartoon) folds of real proteins [155,156,158]. In the majority of studies, the short range conformational propensities were ignored [155], or enforced via some patterns of the low energy contacts along the chain [229]. A significant contribution to understanding some aspects of the folding mechanism comes from the random energy model [230] (REM) applied to the simple cubic lattice representation of the protein [114,231,232].

A more complex, and perhaps more realistic, interaction scheme was also explored in the framework of simple lattice models (sc, bcc, diamond and fcc lattices). First, it is relatively easy to reproduce the characteristic interplay between the short (between residues close in sequence) and long range interactions, especially for lattices that have geometry somewhat similar to the geometry of polypeptide chains. For instance, on the diamond lattice one can mimic 'helical' turns via the gauche conformations of the three consecutive chain links and expanded conformations via the trans conformation [160,161,164–166,233,234]. A more complex fcc lattice enables a cartoon-type representation of all structural classes of real proteins [26], with a clear definition of right-handed helices. It is also possible to mimic the directional effect of hydrogen bonds in these simplified models, thereby accounting for the three main classes of interactions: long range pairwise interactions (with implicitly averaged hydrophobic effect), short range conformational propensities and hydrogen bonding [23,235] that provides a bias towards secondary structure ordering. Based on the results of several recent studies, it is apparent that the presence of these short range and directional contributions [235,236] is necessary for modeling the basic physics of proteins. Of course, the details may vary. For instance, the short range conformational propensities may be, to some extent, replaced by a more complex (than two letter) code for the long range pairwise interactions [23].

A very similar 'minimal' force field has proven very effective (in the sense of reproducing the basic physics of protein folding) in application to the somewhat more complex chess-knight model of Kolinski and Skolnick [172,173]. The interaction scheme designed for this type of model by Hao and Scheraga [237–239] had three contributions; pairwise interactions with a three letter code (polar, nonpolar, and inert residues), local preferences towards expanded conformations of the putative β-strands and directional interactions of the model Cα–Cα bonds mimicking the effect of hydrogen bonds.

Similar designs of the force fields of continuous idealized protein like systems also enabled us to address more specific problems of protein folding, dynamics and thermodynamics. In the model studied by Klimov and Thirumalai [240,241], a quite detailed force field was proposed for the model with just two united atoms per residue; the first corresponding to the Cα and the second to the spherical side chain. Particular components of the force field modeled the bond angle potential, dihedral angle potential, pairwise long range interactions of both types of united atoms and a model of hydrogen bonds [242,243].

The lack of some directional interactions mimicking intraprotein hydrogen bonds might be the main reason for the relatively low predictive power in early continuous reduced models of real proteins. More advanced force fields in reduced continuous models (such as that of Klimov and Thirumalai [243]) try to model hydrogen bonds in a more explicit fashion. Also, a very good example is the UNRES force field [40–42,211,212], where the interactions between the united atoms located in the centers of peptide bonds are of a complex form that mimics the directional character of hydrogen bonds. In contrast to the majority of force field designs in reduced models, the UNRES interaction scheme is carefully derived from physical principles, and a great deal of effort has been made to clearly separate various contributions [41,211,212]. For this purpose, higher order multibody potentials were introduced. As a result, the derived force field has its global minimum of potential energy at the native (or, more precisely, a near-native) conformation of a significant fraction of small, topologically simple globular proteins.

Perhaps some interesting features are included in force fields for high coordination lattice models. The force field of the SICHO model [189] is an example of a relatively complex, knowledge based interaction scheme [74]. A very important part of this force field consists of a number of protein sequence independent biases that provide protein-like conformational stiffness of the model chain. The generic SICHO chain is extremely flexible due to the large number of allowed orientations and bond lengths of the virtual 'bonds' connecting the side chain centers of mass [189]. Thus, the local geometry has to be restricted to the ranges of distances and angles that are observed in real proteins. For instance, too narrow (or completely open) planar angles are forbidden. Additionally, regular protein-like (helical, β-strands) conformations of short fragments (from three bond to five bonds) are energetically rewarded. As a result, the observed distributions of the short range distances and angles become similar to average distributions extracted from protein structures [189]. Sequence dependent secondary structure propensities are encoded in several statistical potentials controlling the distances between ith and $i+1$, $i+2$, $i+3$ and $i+4$th residues. The three-bond term is chiral. The idea is very similar to the concept of reduced backbone dihedral potentials used in several other models [10,244]. All short-range potentials depend on the identity of two residues. Main chain hydrogen bonds are modeled as a directional potential between Cα atoms. Pairwise interactions of the side groups also depend on the mutual orientation of the fragments involved [34,35,56]. Statistical potential describing these interactions was derived for three types of side chain contacts: parallel, intermediate and antiparallel. Contacts are parallel when the angle between the bisectors of the two corresponding two-bond virtual chain vectors is smaller than 60°. This feature

of the pairwise potential is very important. For instance, the statistical potential for two oppositely charged residues has a large negative value for the parallel contact and a large positive value for antiparallel contacts [56]. This reflects the fact that when charged residues are close to each other, they are almost always located on the surface of protein; therefore, their side chains point in the same direction. Such a design of the knowledge-based potential provides strong an average force for segregation of surface and buried residues in the model chain. Finally, the SICHO force field contains a couple of terms mimicking the averaged hydrophobic effect in form of single amino acid dependent potentials controlling number and types of contacts and location of a residue in respect to the center of mass of the model structure [56].

Multibody potentials seem to play an important role in more complex reduced models [28,41,211,245,246]. An extreme example is to use a neural network to evaluate structure packing based on large fragments of the side-chain contact map [184,185,247].

Various terms in the force fields designed for reduced models usually account for an averaged effect of several types of actual physical interactions in real systems. Consequently, it is usually unclear what should be the weighting of particular model potentials. An optimization of the force field becomes necessary [124,187,248–252]. Frequently, force fields are optimized to recognize a selected native-like (or near native) state from all other conformations of the model protein and to ensure fast folding [253]. A very interesting optimization procedure has been proposed recently by Liwo and coworkers [42,211,212]. By dividing the conformational space according to a certain measure of the 'distance' from the native structure, they actually optimized the free-energy landscape, biasing it towards a funnel-like shape.

4. Sampling conformational space of reduced models

The method of sampling depends on the complexity of a model and on the purpose of studies [250,254]. For very simple discrete models, all conformations could be enumerated [20,76,96,141,255–260] and an exact analysis of the model thermodynamics performed [20,261]. Structure and thermodynamics of more complex models are studied via classical molecular dynamics [262–267] (MD), Monte Carlo (MC) methods [26,109,129,158–160,167,168,237,239,268–274], genetic algorithms and hybrid combinations of these methods [275]. The dynamics of continuous reduced models could be studied via classical MD [36,201,204,265–267,276,277] or its variants [263] (for instance, Brownian dynamics [265,278]), but also via various MC schemes [209,210] or combinations of various methods of global minimization [275]. Long time dynamics of discrete models could be studied using Monte Carlo dynamics (MCD) schemes [22,60,69,146,163,166,172–175,233,268,279]. The simplest example is application of Verdier–Stockmayer algorithm [280] to the cubic lattice models. It should be noted, however that this algorithm is not ergodic (except for very short chains), and therefore, the results of such studies need to be carefully analyzed. With decreasing symmetry of discrete models, the problem of ergodicity disappears. A properly designed MCD scheme provides a numerical solution of a stochastic equation of motion (Master Equation) and for time intervals significantly larger than the elementary jump (a local conformational change) time coincides with MD of equivalent continuous models. When the only purpose is to find the lowest energy conformation various 'multicopy' methods seem be the most effective [168,273]. The most typical examples are: Replica Exchange Monte Carlo (REMC) method [281], genetic algorithms [217,282–285], Conformational Space Annealing (CSA)—a combination of genetic algorithm (GA) with local minimization [42], or hybrid algorithms combining GA and MC [286,287]. Very powerful in study of thermodynamics of protein models are methods based on multicanonical or generalized ensembles approaches [61,237–239,288–291].

5. Some applications of reduced models

5.1. Study of dynamics and thermodynamics of idealized protein-like systems

Go and coworkers did very early studies of protein like lattice systems. They addressed the role of various types of interactions on stability and folding mechanism of idealized model systems [149–153].

Highly idealized protein models, as in the HP model of Chan and Dill [84–89,292] or the random energy model of Shakhnovich [114], are usually studied from the point of view of the most basic physical properties that could be common for a large subset of proteins [20,92,93]. An important virtue of these models is that they could be treated in an exact fashion, or in an almost exact fashion. It is even possible to enumerate all compact conformations of such models [96,255]. Recently, a very sophisticated method for such enumerations has been developed [256–259]. Sometimes the enumeration of conformational states could be extended onto somewhat more complex models [76,141]. The problems addressed in the context of 'simple exact' [20] models include: the fundamentals of hydrophobic collapse [293] and some elements of folding kinetics [115,145–147,232,255,294–298], design of foldable sequences with a unique ground state and the requirements of the two-state cooperativity of the folding process [23–25,299,300]. Recently it has been shown that in order to have a true all-or-none folding transition [25] within these simple models, it is important to allow for some degeneracy of the ground state of the model [24,100], an account for secondary structure and hydrogen bonds and perhaps more than a two letter code for the sequence [23,248]. The presence of even a cartoon side chain may also increase cooperativity of the

folding process [63,139]. For somewhat more complex diamond lattice models, it has been shown that when the degeneracy of the ground state is allowed, even homopolymeric chain with short range interactions simulating protein-like stiffness undergo a first-order collapse transition [159,271]. Therefore the interplay between the long-range (long range distance along the chain) contact interactions and the local conformational stiffness seems be one of the simplest and probably most fundamental requirements for protein-like cooperative folding. The structural uniqueness of the folded state in these idealized diamond lattice models could be achieved by introducing a proper sequence with differentiated long range interactions (polar and nonpolar residues) and differentiated short range interactions along the chain [161–166,233,234,279]. Increasing complexity of the idealized lattice models [22,172,173,176] leads to a larger number of allowed conformations and, consequently, complicates the interaction scheme required for protein-like folding [167,168,237–239,249]. Recently, it has been shown that in order to fold a Greek-key six-member β-barrel restricted to the fcc lattice in a cooperative fashion, the following properly balanced interactions are necessary [26]: the sequence dependent secondary structure propensities (a two letter code—expanded and flexible loop), a proper pattern of polar and nonpolar residues, and orientation-dependent interactions between the polar residues. The last could be considered as an ersatz of the ordering effect of hydrogen bonds. Probably such a type of model [26] constitutes a minimalist system that exhibits the most characteristic features of proteins—a well defined, however geometrically degenerate, ground state of a relatively complex topology, fast folding with two-state thermodynamics, a clearly defined, but simplified notion of secondary structure and a funnel-like energy landscape.

In parallel, significant work has been done with simple continuous space models [242,301,302]. These focused on dynamics and thermodynamics of idealized protein motifs, as single helices [243], small helical bundles [10,11,276,278,303,304], small β-sheets [240,241,277], β-barrels [197,203,305], and idealized models of real proteins [306,307]. These studies contributed significantly to our understanding of the role of various interactions in polypeptides [196,199,243,275,303], the origin of the two-state folding transition, the folding pathways and nucleation of the process [264,305], and also effects of external restraints [241,308], including models of chaperone–protein systems [309,310] on the protein folding mechanism. Possibly, such simplified models could also provide a guideline for the design of artificial proteins [276,304].

5.2. Low resolution modeling of real proteins

Probably one of the very first attempts to model real proteins using a reduced representation was done by Levitt and Warshel [204]. The model had two centers of interaction per residue (Cα and side-group united atom), a simple interaction scheme for short-range and long-range interactions, and conformational space was explored by means of conventional MD technique [36,311]. The main purpose was to predict the three-dimensional native structure from the sequence of amino acids alone. The folding of a small globular protein, bovine pancreatic trypsin inhibitor (BPTI) led in some runs to conformations resembling the crystallographic structure, with a cRMSD in the range of 6.5 Å. Other studies of similar models with various modifications of the interaction scheme and sampling techniques did not improve prediction quality [205–207,312]. Nevertheless, taking into consideration the enormous size of the conformational space to be searched, these models certainly catch some fundamental properties of proteins. Significantly better accuracy of predicted structures was achieved later using related continuous space models [214,217,275], however, with a more exact representation of the side chains and statistical potentials derived from the analysis of structural regularities in known protein structures [217,313–315].

Interestingly, there seems be little difference in prediction accuracy (overall fidelity of fold, not local geometry) between the continuous space models and simple lattice models of real proteins [76,155–158]. It could also be demonstrated that the long time dynamics of continuous space reduced models and properly designed lattice models are also very similar [314,315].

Simple lattice and off lattice models of real proteins were studied not only for the sake of test structure predictions but also to elucidate effects of various interactions on protein folding dynamics [9,174] and thermodynamics [136,266] or the role of structural restraints on the modeled structures [76].

5.3. Applications of moderate resolution and 'high resolution' reduced models in the study of protein structure, dynamics and thermodynamics

The level of actual resolution of reduced models depends on two major factors: the accuracy of the geometrical representation [22,27,80,219,220] and the design of the interaction scheme [22,81,124,211,221,224,245,251,252,254,316].

Probably the most common aim of high resolution reduced models is the prediction of protein three dimensional structure from sequence of amino acids [52]. It is hoped that with sufficiently accurate geometry, a fast and efficient sampling technique and sufficiently specific interaction scheme, one can fold proteins 'in silico'. Indeed, it has been demonstrated by many that higher resolution reduced models can indeed find conformations close to the native one, at least for small and topologically simple single domain proteins [28,29,40,42,177,211,217,254,275,287,289,291,307,317]. For very small systems, approaches that use an all-atom representation of the protein structure and

dynamics and simplified (and therefore computationally efficient) force field [227,228,285,318] could be quite successful.

The most successful ab initio protein structure prediction methods could be roughly divided onto two broad categories. In the first category could be placed various approaches that start from random conformations and simulate the folding process or minimize the conformational energy. Typical examples for this kind of approach are high coordination lattice based methods developed by Kolinski and Skolnick and coworkers [56,74,81,189,273,316,319] (the previously described SICHO and CABS models) and continuous space models [208], where a good example is the UNRES model of Scheraga and coworkers [42,133,212]. The high coordination lattice models heavily rely on knowledge-based statistical potentials [316] (that generalize structural regularities seen in real proteins) while the UNRES force field was designed and optimized basing on more rigorous physical principles [41,42,133,212]. A successful application of the above mentioned lattice models to the nontrivial cases of protein structure predictions requires substantial support from threading that provides a prediction of some fraction of native side chain contacts and short range native distances along the polypeptide chain [34,35,55–57,273,316,319]. The best known example of the second category of ab initio folding approaches is the ROSETTA method [43–46,75,77] developed by Baker et al. This method (and related methods [320–322]) employs protein fragments excised from the structural database [50,321,323]. These fragments are then used in an iterative process of query structure assembly that is controlled by a set of simplified potentials, knowledge based biases [46] and structure regularizing filters [14,45].

Methods exist that combine these two types of approaches. A very interesting hierarchical 'bootstrapped' strategy of fold prediction has been proposed by Levitt et al [37,79,171]. In the first stage, a large number of compact lattice polymers are generated. The diamond lattice is used, and a single lattice unit can correspond to more than one protein residue [78]. Subsequently, the fragments excised from the database of known structures are fitted to the lattice chains with a bias superimposed towards predicted secondary structure. The best structures are selected basing on a simplified interaction scheme and finally refined using molecular mechanics for all-atom structures.

Other applications of reduced models of real proteins (or peptides) include the study of folding mechanisms [191,192,295,324–327] and thermodynamics [27,60,191,219], in silico protein design [185,304] and redesign [182,326,328,329], assembly of protein structures based on sparse experimental data [73–77,330–333], force-induced unfolding [126,334,335] effects of spatial confinement on the protein behavior [241], and structure prediction using restraints derived from template (templates) structures identified via sequence based alignments or by threading methods [30–33,55–57,186,317,319]. The last application, providing medium to high resolution structures [34,35,55,57], opens the possibility of a genomic scale distant homology comparative modeling [35] with a large potential impact on structural genomics [336]. The reduced models could be effectively refined by all-atom reconstruction and subsequent energy minimization [337].

6. Summary

Conformational space and the associated energy landscape of detailed atomistic models of proteins are of enormous complexity. Only small peptides could be effectively simulated over the time period corresponding to the longest relaxation time of such a system, or the time corresponding to the time required for folding into a relatively unique native state from an arbitrary random coil conformation. Thereby, the protein representation or the model of interactions (or both) needs to be simplified in order to make computational study practical. In this short review, we attempted to outline some approaches to this problem. Various levels of generalization are assumed in reduced protein models. Levels of detail in geometrical representation vary from a single degree of freedom per amino acid residue (although some models were studied with more than one residue per explicitly treated degree of freedom) through two or three degrees of freedom per residue up to the full atom representation, with simplified models of motion and/or a reduced interaction scheme. On a different level, reduced models could be divided onto continuous space models, continuous space models with discretized internal coordinates and lattice models. Lattice models span a broad range from idealized simple lattice chains to high coordination lattice discretization of the conformational space.

The range of applications of reduced models depends on their complexity and the level of structural detail. Simple exact model could be studied in great detail, providing insight into the most general aspects of protein folding dynamics and thermodynamics. More complex models allow the study of more subtle effects and provide a more complete picture. Particular physical interactions and their effect on protein behavior could be modeled and analyzed in more straightforward way. The most complex models become complementary to detailed atomistic models and enable a moderate resolution study of proteins dynamics, thermodynamics and (with still a lot of limitations) theoretical predictions of protein three-dimensional structure.

Very likely, future applications of reduced models will address more complex problems of interactions between proteins and small molecules, protein–protein interactions and interactions with other types of biomolecules, especially DNA and biological membranes. This field of biomolecular research with reduced molecular models is now emerging, as signalized several times in this short review.

Acknowledgements

This research has been supported in part by NIH grants GM-37408 and RR-12255, as well as the Oishei Foundation.

References

[1] Creighton TE. Biochem J 1990;270:131–46.
[2] Anfinsen CB. Science 1973;181:223–30.
[3] Anfinsen CB, Scheraga HA. Adv Protein Chem 1975;29:205–300.
[4] Ptitsyn OB. J Protein Chem 1987;6:273–93.
[5] Brooks III CL, Gruebele M, Onuchic JN, Wolynes PG. Proc Natl Acad Sci USA 1998;95:11037–8.
[6] Levinthal C. J Chim Phys 1968;65:44–5.
[7] Zwanzig R, Szabo A, Bagchi B. Proc Natl Acad Sci USA 1992;89: 20–2.
[8] Skolnick J, Kolinski A, Godzik A. Proc Natl Acad Sci USA 1993;90: 2099–100.
[9] Skolnick J, Kolinski A. Science 1990;250:1121–5.
[10] Zhou Y, Karplus M. J Mol Biol 1999;293:917–51.
[11] Zhou Y, Karplus M. Nature 1999;401:400–3.
[12] Alm E, Baker D. Proc Natl Acad Sci USA 1999;96:11305–10.
[13] Alm E, Baker D. Curr Opin Struct Biol 1999;9:189–96.
[14] Baker D. Nature 2000;405:39–42.
[15] Karplus M, Shakhnovich E. Protein folding: theoretical studies of thermodynamics and dynamics. In: Creighton TE, editor. Protein folding. USA: Freeman; 1992. p. 127–96.
[16] Karplus M, Sali A. Curr Opin Struct Biol 1995;5:58–73.
[17] Bohr HG, Wolynes PG. Phys Rev A 1992;46:5242–8.
[18] Wolynes PG, Onuchic JN, Thirumalai D. Science 1995;267: 1619–20.
[19] Shoemaker BA, Wang J, Wolynes PG. J Mol Biol 1999;287:675–94.
[20] Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Protein Sci 1995;4:561–602.
[21] Sali A, Shakhnovich E, Karplus M. Nature 1994;369:248–51.
[22] Kolinski A, Skolnick J. Lattice models of protein folding, dynamics and thermodynamics. Austin, TX: Landes; 1996.
[23] Kaya H, Chan HS. J Mol Biol 2002;315:899–909.
[24] Kaya H, Chan HS. Proteins 2000;40:637–61.
[25] Kaya H, Chan HS. Phys Rev Lett 2000;85:4823–6.
[26] Pokarowski P, Kolinski A, Skolnick J. Biophys J 2003;84:1518–26.
[27] Mirny L, Shakhnovich E. Annu Rev Biophys Biomol Struct 2001;30: 361–96.
[28] Kolinski A, Godzik A, Skolnick. J Chem Phys 1993;98:7420–33.
[29] Skolnick J, Kolinski A, Brooks C, Godzik A, Rey A. Curr Biol 1993; 3:414–23.
[30] Ortiz AR, Hu WP, Kolinski A, Skolnick J. Pac Symp Biocomput 1997;316–27.
[31] Ortiz AR, Kolinski A, Skolnick J. Proteins 1998;30:287–94.
[32] Ortiz AR, Kolinski A, Skolnick. J Mol Biol 1998;277:419–48.
[33] Ortiz AR, Kolinski A, Skolnick J. Proc Natl Acad Sci USA 1998;95: 1020–5.
[34] Kihara D, Lu H, Kolinski A, Skolnick J. Proc Natl Acad Sci USA 2001;98:10125–30.
[35] Kihara D, Zhang Y, Kolinski A, Skolnick J. Proc Natl Acad Sci USA 2002;99:5993–8.
[36] Levitt M. J Mol Biol 1976;104:59–107.
[37] Samudrala R, Xia Y, Levitt M, Huang ES. Pac Symp Biocomput 1999;505–16.
[38] Samudrala R, Huang ES, Koehl P, Levitt M. Protein Engng 2000;13: 453–7.
[39] Samudrala R, Levitt M. BMC Struct Biol 2002;2:3.
[40] Saunders JA, Scheraga HA. Biopolymers 2003;68:300–17.
[41] Pilardy J, Czaplewski C, Liwo A, Wedemeyer WJ, Lee J, Ripoll DR, Arlukowicz P, Oldziej S, Arnautova YA, Scheraga HA. J Phys Chem 2001;105:7299–311.
[42] Lee J, Liwo A, Scheraga HA. Proc Natl Acad Sci USA 1999;96: 2025–30.
[43] Bonneau R, Tsai J, Ruczinski I, Chivian D, Rohl C, Strauss CE, Baker D. Proteins 2001;Suppl 5:119–26.
[44] Bonneau R, Strauss CE, Rohl CA, Chivian D, Bradley P, Malmstrom L, Robertson T, Baker D. J Mol Biol 2002;322:65–78.
[45] Bonneau R, Ruczinski I, Tsai J, Baker D. Protein Sci 2002;11: 1937–44.
[46] Simons KT, Bonneau R, Ruczinski I, Baker D. Proteins 1999;Suppl 3:171–6.
[47] Takada S. Proc Natl Acad Sci USA 1999;96:11698–700.
[48] Plaxco KW, Riddle DS, Grantcharova V, Baker D. Curr Opin Struct Biol 1998;8:80–5.
[49] Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Proteins 1995;23:318–26.
[50] Simons KT, Kooperberg C, Huang E, Baker D. J Mol Biol 1997;268: 209–25.
[51] Simons KT, Strauss C, Baker D. J Mol Biol 2001;306:1191–9.
[52] Schonbrun J, Wedemeyer WJ, Baker D. Curr Opin Struct Biol 2002; 12:348–54.
[53] Kolinski A, Skolnick J. Proteins 1994;18:338–52.
[54] Kolinski A, Skolnick J. Proteins 1994;18:353–66.
[55] Skolnick J, Kolinski A, Kihara D, Betancourt M, Rotkiewicz P, Boniecki M. Proteins 2001;Suppl 5:149–56.
[56] Kolinski A, Betancourt M, Kihara D, Rotkiewicz P, Skolnick J. Proteins 2003; in press.
[57] Skolnick J, Kolinski A. Adv Chem Phys 2002;1209:131–92.
[58] Brooks III CL, Onuchic JN, Wales DJ. Science 2001;293:612–3.
[59] Levitt M, Gerstein M, Huang E, Subbiah S, Tsai J. Annu Rev Biochem 1997;66:549–79.
[60] Kolinski A, Galazka W, Skolnick. J Chem Phys 1998;108:2608–17.
[61] Hansmann UH, Okamoto Y. Curr Opin Struct Biol 1999;9:177–83.
[62] Harrison PM, Chan HS, Prusiner SB, Cohen FE. J Mol Biol 1999; 286:593–606.
[63] Dima RI, Thirumalai D. Protein Sci 2002;11:1036–49.
[64] Mohanty D, Kolinski A, Skolnick J. Biophys J 1999;77:54–69.
[65] Sikorski A, Kolinski A, Skolnick J. Biophys J 1998;75:92–105.
[66] Tiana G, Broglia RA. Proteins 2002;49:82–94.
[67] Vieth M, Kolinski A, Brooks III CL, Skolnick J. Pac Symp Biocomput 1996;653–62.
[68] Vieth M, Kolinski A, Brooks ICL, Skolnick J. J Mol Biol 1995;251: 448–67.
[69] Vieth M, Kolinski A, Brooks III CL. J Mol Biol 1994;237:361–7.
[70] Vieth M, Kolinski A, Skolnick J. Biochemistry 1996;35:955–67.
[71] Vinals J, Kolinski A, Skolnick J. Biophys J 2002;83:2801–11.
[72] Skolnick J, Kolinski A. Monte Carlo lattice dynamics and the prediction of lattice folds. In: van Gunsteren WF, Weiner PK, editors. Computer simulations of biomolecular systems. Theoretical and experimental applications. Leiden: ESCOM Science Publishers BV; 1989.
[73] Bujnicki JM, Rotkiewicz P, Kolinski A, Rychlewski L. Protein Engng 2001;14:717–21.
[74] Kolinski A, Skolnick J. Proteins 1998;32:475–94.
[75] Bowers PM, Strauss CE, Baker D. J Biomol NMR 2000;18:311–8.
[76] DeWitte RS, Michnick SW, Shakhnovich EI. Protein Sci 1995;4: 1780–91.
[77] Rohl CA, Baker D. J Am Chem Soc 2002;124:2723–9.
[78] Hinds DA, Levitt M. J Mol Biol 1994;243:668–82.
[79] Xia Y, Huang ES, Levitt M, Samudrala R. J Mol Biol 2000;300: 171–85.
[80] Park BH, Levitt M. J Mol Biol 1995;249:493–507.
[81] Kolinski A, Rotkiewicz P, Ilkowski B, Skolnick J. Prog Theor Phys (Kyoto) 2000;Suppl 138:292–300.
[82] Reva BA, Sanner MF, Olson AJ, Finkelstein AV. J Comput Chem 1996;17:1025–32.

[83] Godzik A, Kolinski A, Skolnick. J Comp Chem 1993;14:1194–202.
[84] Lau KF, Dill KA. Macromolecules 1989;22:3986–97.
[85] Chan HS, Dill KA. Macromolecules 1989;22:4559–73.
[86] Chan HS, Dill KA. Proc Natl Acad Sci USA 1990;87:6388–92.
[87] Chan HS, Dill KA. J Chem Phys 1991;95:3775–87.
[88] Chan HS, Dill KA. J Chem Phys 1993;99:2116–27.
[89] Chan HS, Dill KA. J Chem Phys 1994;100:9238–57.
[90] Chan HS, Dill KA. Proteins 1996;24:345–51.
[91] Chan HS, Dill KA. Proteins 1996;24:335–44.
[92] Dill KA. Biochemistry 1985;24:1501–9.
[93] Dill KA. Curr Biol 1993;3:99–103.
[94] Dill KA, Yue K. Protein Sci 1996;5:254–61.
[95] Sali A, Shakhnovich E, Karplus M. J Mol Biol 1994;235:1614–36.
[96] Dinner AR, Sali A, Karplus M, Shakhnovich EJ. Chem Phys 1994; 101:1444–51.
[97] Dinner AR, Sali A, Karplus M. Proc Natl Acad Sci USA 1996;93: 8356–61.
[98] Agarwala R, Batzoglou S, Dancik V, Decatur SE, Hannenhalli S, Farach M, Muthukrishnan S, Skiena S. J Comput Biol 1997;4: 275–96.
[99] Yue K, Fiebig KM, Thomas PD, Chan HS, Shakhnovich EI, Dill KA. Proc Natl Acad Sci USA 1995;92:325–9.
[100] Crippen GM, Chhajer M. J Chem Phys 2002;116:2261–8.
[101] Crippen GM. J Mol Biol 1996;260:467–75.
[102] Chan HS. Proteins 2000;40:543–71.
[103] Irback A, Peterson C, Potthast F, Sandelin E. Structure 1999;7: 347–60.
[104] Camacho CJ, Thirumalai D. Proc Natl Acad Sci USA 1993;90: 6369–72.
[105] Klimov DK, Thirumalai D. J Mol Biol 1998;282:471–92.
[106] Betancourt MR, Thirumalai D. J Mol Biol 1999;287:627–44.
[107] Shakhnovich EI, Finkelstein AV. Biopolymers 1989;28:1667–80.
[108] Shakhnovich EI, Gutin AM. Biophys Chem 1989;34:187–99.
[109] Shakhnovich E, Farztdinov G, Gutin AM. Phys Rev Lett 1991;67: 1665–8.
[110] Shakhnovich EI, Gutin AM. Proc Natl Acad Sci USA 1993;90: 7195–9.
[111] Shakhnovich EI, Gutin AM. Protein Engng 1993;6:793–800.
[112] Shakhnovich EI. Phys Rev Lett 1994;72:3907–10.
[113] Shakhnovich EI. Fold Des 1996;1:R50–4.
[114] Gutin AM, Shakhnovich EIJ. J Chem Phys 1993;98:8174–7.
[115] Gutin AM, Abkevich VI, Shakhnovich EI. Fold Des 1998;3:183–94.
[116] Hao M-H, Scheraga HA. J Phys Chem 1996;100:14540–8.
[117] Xia Y, Levitt M. J Chem Phys 2000;113:9318–30.
[118] Mirny LA, Shakhnovich EI. J Mol Biol 1996;264:1164–79.
[119] Cieplak M, Banavar JR. Fold Des 1997;2:235–45.
[120] Leopold PE, Montal M, Onuchic JN. Proc Natl Acad Sci USA 1992; 89:8721–5.
[121] Locker CR, Hernandez R. Proc Natl Acad Sci USA 2001;98: 9074–9.
[122] Li H, Tang C, Wingreen NS. Proteins 2002;49:403–12.
[123] de Araujo AFP, Pochapsky TC. Fold Des 1996;1:299–314.
[124] Hao M-H, Scheraga HA. Proc Natl Acad Sci USA 1996;93:4984–9.
[125] Segawa S-I, Kawai T. Biopolymers 1986;25:1815–35.
[126] Socci ND, Onuchic JN, Wolynes PG. Proc Natl Acad Sci USA 1999; 96:2031–5.
[127] Konig R, Dandekar T. Protein Engng 2001;14:329–35.
[128] Kussell E, Shakhnovich EI. Phys Rev Lett 2002;89:168101.
[129] Sfatos CD, Gutin AM, Abkevich VI, Shakhnovich EI. Biochemistry 1996;35:334–9.
[130] Zhdanov VP, Kasemo B. Proteins 2000;39:76–81.
[131] Broglia RA, Tiana G, Pasquali S, Roman HE, Vigezzi E. Proc Natl Acad Sci USA 1998;95:12930–3.
[132] Harrison PM, Chan HS, Prusiner SB, Cohen FE. Protein Sci 2001;10: 819–35.
[133] Hart WE, Istrail S. J Comput Biol 1997;4:241–59.
[134] Gupta P, Hall CK, Voegler AC. Protein Sci 1998;7:2642–52.
[135] Istrail S, Schwartz R, King J. J Comput Biol 1999;6:143–62.
[136] Ueda Y, Taketomi H, Go N. Biopolymers 1978;17:1531–48.
[137] Bromberg S, Dill KA. Protein Sci 1994;3:997–1009.
[138] Reva BA, Rykunov DS, Olson AJ, Finkelstein AV. J Comput Biol 1995;2:527–35.
[139] Klimov DK, Thirumalai D. Fold Des 1998;3:127–39.
[140] Li L, Shakhnovich EI. J Mol Biol 2001;306:121–32.
[141] Will S. Pac Symp Biocomput 2002;661–72.
[142] Meirovitch H. J Comput Chem 2002;23:166–71.
[143] Shakhnovich EI. Fold Des 1998;3:R45–R58.
[144] Honig B, Cohen FE. Fold Des 1996;1:R17–R20.
[145] Dinner AR, Karplus M. J Mol Biol 1999;292:403–19.
[146] Dinner AR, Karplus M. J Phys Chem B 1999;103:7976–94.
[147] Dinner AR, Abkevich V, Shakhnovich E, Karplus M. Proteins 1999; 35:34–40.
[148] Wind AF, Kemp JP, Ermoshkin AV, Chen JZ. Phys Rev E Stat Nonlin Soft Matter Phys 2002;66:031909.
[149] Go N, Taketomi H. Proc Natl Acad Sci USA 1978;75:559–63.
[150] Go N, Abe H, Mizuno H, Taketomi H. In: Jaenicke N, editor. Protein folding. Amsterdam: Elsevier; 1980. p. 167–81.
[151] Go N. Ann Rev Biophys Bioeng 1983;12:183–210.
[152] Taketomi H, Kano F, Go N. Biopolymers 1988;27:527–59.
[153] Taketomi H, Ueda Y, Go N, Int. J Pept Protein Res 1988;7:445–9.
[154] Crippen GM. Biochemistry 1991;30:4232–7.
[155] Covell DG. Proteins 1992;14:409–20.
[156] Covell DG, Jernigan RL. Biochemistry 1990;29:3287–94.
[157] Dashevskii VG. Molekulyarnaya Biologiya (Translation from) 1980; 14:105–17.
[158] Krigbaum WR, Lin SF. Macromolecules 1982;15:1135–45.
[159] Kolinski A, Skolnick J, Yaris R. J Chem Phys 1986;85:3585–97.
[160] Kolinski A, Skolnick J, Yaris R. Biopolymers 1987;26:937–62.
[161] Skolnick J, Kolinski A, Yaris R. Proc Natl Acad Sci USA 1988;85: 5057.
[162] Skolnick J, Kolinski A. Annu Rev Phys Chem 1989;40:207–35.
[163] Skolnick J, Kolinski A, Sikorski A. Chem Des Automat News 1990; 5:1–20.
[164] Sikorski A, Skolnick J. Biopolymers 1989;28:1097–113.
[165] Sikorski A, Skolnick J. Proc Natl Acad Sci USA 1989;86:2668–72.
[166] Sikorski A, Skolnick. J Mol Biol 1990;212:819–36.
[167] Gront D, Kolinski A, Skolnick. J Chem Phys 2000;113:5065–71.
[168] Gront D, Kolinski A, Skolnick. J Chem Phys 2001;115:1569–74.
[169] Raghunathan G, Jernigan RL. Protein Sci 1997;6:2072–83.
[170] Hinds DA, Levitt M. Proc Natl Acad Sci USA 1992;89:2536–40.
[171] Samudrala R, Xia H, Huang E, Levitt M. Proteins 1999;Suppl 3: 194–8.
[172] Kolinski A, Milik M, Skolnick. J Chem Phys 1991;94:3978–85.
[173] Skolnick J, Kolinski A. J Mol Biol 1991;221:499–531.
[174] Godzik A, Skolnick J, Kolinski A. Proc Natl Acad Sci USA 1992;89: 2629–33.
[175] Godzik A, Kolinski A, Skolnick J. J Mol Biol 1992;227:227–38.
[176] Kolinski A, Skolnick. J Phys Chem 1992;97:9412–26.
[177] Godzik A, Kolinski A, Skolnick. J Comp Aid Mol Des 1993;7: 397–438.
[178] Rykunov DS, Reva BA, Finkelstein AV. Proteins 1995;22:100–9.
[179] Reva BA, Finkelstein AV, Sanner MF, Olson AJ. Proteins 1996;25: 379–88.
[180] Kolinski A, Galazka W, Skolnick. J Chem Phys 1995;103: 10286–97.
[181] Kolinski A, Milik M, Rycombel J, Skolnick. J Chem Phys 1995;103: 4312–23.
[182] Kolinski A, Galazka W, Skolnick J. Proteins 1996;26:271–87.
[183] Kolinski A, Skolnick. J Chem Phys 1997;107:953–64.
[184] Olszewski KA, Kolinski A, Skolnick J. Protein Engng 1996;9:5–14.
[185] Olszewski KA, Kolinski A, Skolnick J. Proteins 1996;25:286–99.
[186] Skolnick J, Kolinski A, Ortiz AR. J Mol Biol 1997;265:217–41.
[187] Sun Z, Xia X, Guo Q, Xu D. J Protein Chem 1999;18:39–46.
[188] Kolinski A, Madziar P. Biopolymers 1997;42.

[189] Kolinski A, Jaroszewski L, Rotkiewicz P, Skolnick. J Phys Chem 1998;102:4628–37.
[190] Kolinski A, Rotkiewicz P, Ilkowski B, Skolnick J. Proteins 1999;37: 592–610.
[191] Kolinski A, Ilkowski B, Skolnick J, Biophys J. Biophys J 1999;77: 2942–52.
[192] Ilkowski B, Skolnick J, Kolinski A. Macromol Theor Simul 2000;9: 523–33.
[193] Feig M, Rotkiewicz P, Kolinski A, Skolnick J, Brooks CLI. Proteins 2000;41:86–97.
[194] Rotkiewicz P, Sicinska W, Kolinski A, DeLuca HF. Proteins 2001; 44:188–99.
[195] Honeycutt JD, Thirumalai D. Proc Natl Acad Sci USA 1990;87: 3526–9.
[196] Shea J-M, Nochomovitz YD, Guo Z, Brooks CLI. J Chem Phys 1998;109:2895–903.
[197] Jang H, Hall CK, Zhou Y. Biophys J 2002;82:646–59.
[198] Irback A, Potthast F. J Chem Phys 1995;103:10298–305.
[199] Irback A, Peterson C, Potthast F, Sommelius O. J Chem Phys 1997; 107:273–82.
[200] Guo Z, Thirumalai D. J Chem Phys 1992;97:525–36.
[201] Guo Z, Thirumalai D. Biopolymers 1995;36:83–102.
[202] Berriz GF, Gutin AM, Shakhnovich EI. J Chem Phys 1997;106: 9276–85.
[203] Nymeyer H, Garcia AE, Onuchic JN. Proc Natl Acad Sci USA 1998; 95:5921–8.
[204] Levitt M, Warshel A. Nature 1975;253:694–8.
[205] Hagler AT, Honig B. Proc Natl Acad Sci USA 1978;75:554–8.
[206] Wilson C, Doniach S. Proteins 1989;6:193–209.
[207] Kuntz ID, Crippen GM, Kollman PA, Kimelman D. J Mol Biol 1976; 106:983–94.
[208] Fogolari F, Esposito G, Viglino P, Cattarinussi S. Biophys J 1996;70: 1183–97.
[209] Hoffmann D, Knapp EW. Eur Biophys J 1996;24:387–403.
[210] Hoffmann D, Knapp EW. Phys Rev E 1996;53:4221–4.
[211] Lee J, Ripoll DR, Czaplewski C, Pilardy J, Wedemeyer WJ, Scheraga HA. J Phys Chem 2001;105:7291–8.
[212] Liwo A, Arlukowicz P, Czaplewski C, Oldziej S, Pilardy J, Scheraga HA. Proc Natl Acad Sci USA 2002;99:1937–42.
[213] Fain B, Xia H, Levitt M. Protein Sci 2002;11:2010–21.
[214] Wallqvist A, Ullner M. Proteins 1994;18:267–89.
[215] Sun S, Brem R, Chan HS, Dill KA. Protein Engng 1995;8:1205–13.
[216] Hardin C, Luthey-Schulten Z, Wolynes PG. Proteins 1999;34: 281–94.
[217] Sun S. Protein Sci 1993;2:762–85.
[218] Monge A, Friesner RA, Honig B. Proc Natl Acad Sci USA 1994;91: 5027–9.
[219] Monge A, Lathrop EJP, Gunn JR, Shenkin PS, Friesner RA. J Mol Biol 1995;241:995–1012.
[220] Friesner RA, Gunn JR. Annu Rev Biophys Biomol Struct 1996;25: 315–42.
[221] Park BH, Levitt M. J Mol Biol 1996;258:367–92.
[222] Toma L, Toma S. Protein Sci 1999;8:196–202.
[223] de la Cruz XF, Mahoney MW, Lee B. Fold Des 1997;2:223–34.
[224] Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D. Proteins 1999;34:82–95.
[225] Melo F, Sanchez R, Sali A. Protein Sci 2002;11:430–48.
[226] Park BH, Huang ES, Levitt M. J Mol Biol 1997;266:831–46.
[227] Shimada J, Kussell EL, Shakhnovich EI. J Mol Biol 2001;308: 79–95.
[228] Shimada J, Shakhnovich EI. Proc Natl Acad Sci USA 2002;99: 11175–80.
[229] Cieplak M, Hoang TX, Robbins MO. Proteins 2002;49:104–13.
[230] Derrida B. Phys Rev Lett 1980;45:79–82.
[231] Pande VS, Grosberg AY, Joerg C, Tanaka T. Phys Rev Lett 1996;76: 3987–90.
[232] Pande VS, Grosberg A, Tanaka T. Fold Des 1997;2:109–14.
[233] Skolnick J, Kolinski A, Yaris R. Proc Natl Acad Sci USA 1989;86: 1229–33.
[234] Skolnick J, Kolinski A, Yaris R. Biopolymers 1989;28:1059–95.
[235] Borg J, Jensen MH, Sneppen K, Tiana G. Phys Rev Lett 2001;86: 1031–3.
[236] Chen H, Zhou X, Ou-Yang ZC. Phys Rev E 2001;64:041905.
[237] Hao M-H, Scheraga HA. J Phys Chem 1994;98:4940–8.
[238] Hao M-H, Scheraga HA. J Phys Chem 1994;98:9882–93.
[239] Hao M-H, Scheraga HA. J Chem Phys 1995;102:1334–48.
[240] Klimov DK, Thirumalai D. Proc Natl Acad Sci USA 2000;97: 2544–9.
[241] Klimov DK, Newfield D, Thirumalai D. Proc Natl Acad Sci USA 2002;99:8019–24.
[242] Veitshans T, Klimov D, Thirumalai D. Fold Des 1997;2:1–22.
[243] Klimov DK, Betancourt MR, Thirumalai D. Fold Des 1998;3: 481–96.
[244] DeWitte RS, Shakhnovich EI. Protein Sci 1994;3:1570–81.
[245] Betancourt MR, Thirumalai D. Protein Sci 1999;8:361–9.
[246] Gan HH, Tropsha A, Schlick T. Proteins 2001;43:161–74.
[247] Milik M, Kolinski A, Skolnick J. Protein Engng 1995;8:225–36.
[248] Abkevich VI, Gutin AM, Shakhnovich EI. Fold Des 1996;1:221–30.
[249] Chiu TL, Goldstein RA. Proteins 2000;41:157–63.
[250] Hardin C, Eastwood MP, Prentiss M, Luthey-Schulten Z, Wolynes PG. J Comput Chem 2002;23:138–46.
[251] Tobi D, Elber R. Proteins 2000;41:40–6.
[252] Vendruscolo M, Mirny LA, Shakhnovich EI, Domany E. Proteins 2000;41:192–201.
[253] Samudrala R, Levitt M. Protein Sci 2000;9:1399–401.
[254] Hardin C, Pogorelov TV, Luthey-Schulten Z. Curr Opin Struct Biol 2002;12:176–81.
[255] Sali A, Shakhnovich E, Karplus M. J Mol Biol 1994;235:1614–36.
[256] Kloczkowski A, Jernigan RL. Macromolecules 1997;30:691–4.
[257] Kloczkowski A, Jernigan RL. Comput Theor Polym Sci 1997;7: 163–73.
[258] Kloczkowski A, Jernigan RL. J Chem Phys 1998;109:5134–46.
[259] Kloczkowski A, Jernigan RL. J Chem Phys 1998;109:5147–59.
[260] Backofen R. Pac Symp Biocomput 1998;389–400.
[261] Ozkan SB, Dill KA, Bahar I. Biopolymers 2003;68:35–46.
[262] Zhou Y, Karplus M, Wichert JM, Hall CK. J Chem Phys 1997;107: 10691–708.
[263] Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Fold Des 1998;3:577–87.
[264] Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. J Mol Biol 2000;296:1183–8.
[265] Honeycutt JD, Thirumalai D. Biopolymers 1992;32:695–709.
[266] Cieplak M, Hoang TX, Robbins MO. Proteins 2002;49:114–24.
[267] Cieplak M, Hoang TX. Biophys J 2003;84:475–88.
[268] Abkevich VI, Gutin AM, Shakhnovich EI. Biochemistry 1994;33: 10026–36.
[269] Abkevich VI, Gutin AM, Shakhnovich EI. Protein Sci 1995;4: 1167–77.
[270] Abkevich VI, Gutin AM, Shakhnovich EI. Proteins 1998;31: 335–44.
[271] Kolinski A, Skolnick J, Yaris R. Proc Natl Acad Sci USA 1986;83: 7267–71.
[272] Shakhnovich E, Abkevich V, Ptitsyn O. Nature 1996;379:96–8.
[273] Zhang Y, Kihara D, Skolnick J. Proteins 2002;48:192–201.
[274] Zhang Y, Skolnick J. J Chem Phys 2001;115(11):5027–32.
[275] Gibbs N, Clarke AR, Sessions RB. Proteins 2001;43:186–202.
[276] Guo Z, Thirumalai D. J Mol Biol 1996;263:323–43.
[277] Baumketner A, Hiwatari Y. Phys Rev E Stat Nonlin Soft Matter Phys 2002;66:011905.
[278] Berriz GF, Shakhnovich EI. J Mol Biol 2001;310:673–85.
[279] Skolnick J, Kolinski A. J Mol Biol 1990;212:787–817.
[280] Verdier PH, Stockmayer WH. J Chem Phys 1962;36:227–35.
[281] Hukushima K, Nemoto K. J Phys Soc (Jpn) 1996;65:1604–8.
[282] Dandekar T, Argos P. Protein Engng 1992;5:637–45.

[283] Unger R, Moult. J Mol Biol 1993;231:75–81.
[284] Sugita Y, Okamoto Y. Chem Phys Lett 1999;314:141–51.
[285] Pedersen JT, Moult J. Proteins 1997;Suppl 1:179–84.
[286] Backofen R, Will S, Clote P. Pac Symp Biocomput 2000;95–106.
[287] Standley DM, Gunn JR, Friesner RA, McDermott AE. Proteins 1998;33:240–52.
[288] Berg BA, Neuhaus T. Phys Rev Lett 1991;68:9–12.
[289] Hansmann UH, Masuya M, Okamoto Y. Proc Natl Acad Sci USA 1997;94:10652–6.
[290] Hansmann UH, Okamoto Y. Phys Rev E 1996;54:5863–5.
[291] Hansmann UH, Okamoto Y, Onuchic JN. Proteins 1999;34:472–83.
[292] Chan HS, Dill KA. Annu Rev Biophys Chem 1991;20:447–90.
[293] Dill KA, Fiebig KM, Chan HS. Proc Natl Acad Sci USA 1993;90: 1942–6.
[294] Pande VS, Grosberg A, Tanaka T, Rokhsar DS. Curr Opin Struct Biol 1998;8:68–79.
[295] Pande VS, Rokhsar DS. Proc Natl Acad Sci USA 1999;96:1273–8.
[296] Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. Trends Biochem Sci 2000;25:331–9.
[297] Gutin AM, Abkevich VI, Shakhnovich EI. Proc Natl Acad Sci USA 1995;92:1282–6.
[298] Gutin AM, Abkevich VV, Shakhnovich EI. Phys Rev Lett 1996;77: 5433–6.
[299] Chen H, Zhou X, Ou-Yang ZC. Phys Rev E 2001;63:031913.
[300] Jewett AI, Pande VS, Plaxco KW. J Mol Biol 2003;326:247–53.
[301] Thirumalai D, Klimov DK. Curr Opin Struct Biol 1999;9:197–207.
[302] Nymeyer H, Socci ND, Onuchic JN. Proc Natl Acad Sci USA 2000; 97:634–9.
[303] Shea JE, Onuchic JN, Brooks III CL. Proc Natl Acad Sci USA 1999; 96:12512–7.
[304] Miller J, Zeng C, Wingreen NS, Tang C. Proteins 2002;47:506–12.
[305] Guo Z, Thirumalai D. Fold Des 1997;2:377–91.
[306] Sorenson JM, Head-Gordon T. Proteins 2002;46:368–79.
[307] Sorenson JM, Head-Gordon T. J Comput Biol 2002;9:35–54.
[308] Hillson N, Onuchic JN, Garcia AE. Proc Natl Acad Sci USA 1999; 96:14848–53.
[309] Gorse D. Biopolymers 2001;59:411–26.
[310] Gorse D. Biopolymers 2002;64:146–60.
[311] Levitt M. Annu Rev Biophys Bioeng 1982;11:251–71.
[312] Kuntz ID. J Am Chem Soc 1975;97:4362–6.
[313] Miyazawa S, Jernigan RL. Macromolecules 1985;18:534–52.
[314] Rey A, Skolnick J. Proteins 1993;16:8–28.
[315] Rey A, Skolnick. J Chem Phys 1994;100:2267–76.
[316] Skolnick J, Kolinski A, Ortiz A. Proteins 2000;38:3–16.
[317] Skolnick J, Kolinski A, Ortiz AR. J Biomol Struct Dyn 1998;16: 381–96.
[318] Kussell E, Shimada J, Shakhnovich EI. Proc Natl Acad Sci USA 2002;99:5343–8.
[319] Zhang Y, Kolinski A, Skolnick J. Biophys J 2003;85:1145–64.
[320] Bystroff C, Baker D. Proteins 1997;Suppl 1:167–71.
[321] Kolodny R, Koehl P, Guibas L, Levitt M. J Mol Biol 2002;323: 297–307.
[322] Kolodny R, Levitt M. Biopolymers 2003;68:278–85.
[323] Kolodny R, Koehl P, Guibas L, Levitt M. J Mol Biol 2003;326:337.
[324] Clementi C, Garcia AE, Onuchic JN. J Mol Biol 2003;326:933–54.
[325] Clementi C, Jennings PA, Onuchic JN. Proc Natl Acad Sci USA 2000;97:5871–6.
[326] Clementi C, Jennings PA, Onuchic JN. J Mol Biol 2001;311: 879–90.
[327] Clementi C, Nymeyer H, Onuchic JN. J Mol Biol 2000;298:937–53.
[328] Nauli S, Kuhlman B, Baker D. Nat Struct Biol 2001;8:602–5.
[329] Kuhlman B, O'Neill JW, Kim DE, Zhang KY, Baker D. J Mol Biol 2002;315:471–7.
[330] Fain B, Levitt M. J Mol Biol 2001;305:191–201.
[331] Kolinski A, Skolnick J. Acta Biochim Pol 1997;44:389–422.
[332] Brunger AT, Clore GM, Gronenborn AM, Karplus M. Proc Natl Acad Sci USA 1986;83:3801–5.
[333] Brunger AT, Karplus M. Acc Chem Res 1991;24:54–61.
[334] Klimov DK, Thirumalai D. Proc Natl Acad Sci USA 1999;96: 6166–70.
[335] Klimov DK, Thirumalai D. Proc Natl Acad Sci USA 2000;97: 7254–9.
[336] Skolnick J, Fetrow JS, Kolinski A. Nat Biotech 2000;18:283–7.
[337] Lee MR, Tsai J, Baker D, Kollman PA. J Mol Biol 2001;313: 417–30.