Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 16, Issue Number 2, (1998)

Reduced Protein Models and their Application to the Protein Folding Problem†

http://www.albany.edu/chemistry/sarma/jbsd.html

Jeffrey Skolnick[1]*, Andrzej Kolinski[1,2] and Angel R. Ortiz[1]

[1]Department of Molecular Biology,
The Scripps Research Institute,
10550 N. Torrey Pines Rd.,
La Jolla, CA. 92037 USA

[2]Department of Chemistry
University of Warsaw
02-093 Warsaw, Poland

Abstract

One of the most important unsolved problems of computational biology is prediction of the three-dimensional structure of a protein from its amino acid sequence. In practice, the solution to the protein folding problem demands that two interrelated problems be simultaneously addressed. Potentials that recognize the native state from the myriad of misfolded conformations are required, and the multiple minima conformational search problem must be solved. A means of partly surmounting both problems is to use reduced protein models and knowledge-based potentials. Such models have been employed to elucidate a number of general features of protein folding, including the nature of the energy landscape, the factors responsible for the uniqueness of the native state and the origin of the two-state thermodynamic behavior of globular proteins. Reduced models have also been used to predict protein tertiary and quaternary structure. When combined with a limited amount of experimental information about secondary and tertiary structure, molecules of substantial complexity can be assembled. If predicted secondary structure and tertiary restraints are employed, low resolution models of single domain proteins can be successfully predicted. Thus, simplified protein models have played an important role in furthering the understanding of the physical properties of proteins.

Introduction

The question of how to relate a protein's sequence to its native structure is referred to as the protein folding problem (1). It is widely believed that proteins obey the "thermodynamic hypothesis", which asserts that the protein's native conformation corresponds to a global free energy minimum that is dictated by the various interactions present in the protein-solvent system (2-4). Unfortunately, due to the complexity of the interactions, finding this free energy minimum in the myriad of multiple minima on the free energy landscape (5,6) is extremely difficult.

While molecular dynamics simulations of all atom models have provided many insights into protein dynamics and structure (7-9), they cannot yet explore the millisecond-to-second time scales required for protein folding. One way to reduce this time scale gap is to simplify the protein representation by employing "united atoms" and effective solvent models. Such approaches assume that the essential physics is retained, and both continuous space (10-12) and lattice-based protein representations have been developed (13-15). For highly idealized treatments, lattice models offer obvious computational advantages (16), but the situation is less clear-cut when one wishes to model real proteins (17). A lattice representation provides the important advantage that many structurally related quantities (including numerous energy contributions) can be precomputed. Thus, for two equivalent models, calculations on a lattice model are about 100 times faster than for an off-lattice model (18,19), and yet their equilibrium and dynamic properties are the same (19-21). There is an immense qualitative difference when a computer simulation is completed overnight instead of in several months. However, recent work

†This article is dedicated to David Beveridge on the occasion of his 60th birthday.

*Author to whom correspondence should be addressed. Phone: 1-619-784-8821; Fax: 1-619-784-8895; E-mail: skolnick@scripps.edu

suggests that the crux of the solution to the protein folding problem does not lie in the particular choice of protein model, but in the development of potentials that can recognize a nativelike state from a misfolded state and the development of techniques that can explore the relevant regions of conformational space (17,19-30).

Use of Simplified Models to Elucidate General Features of Protein Folding

a. *Highly Simplified Lattice Models*

In an important series of papers, Go *et al.* examined a lattice model (13,31-33) designed to elucidate the major features of the protein folding process rather than to predict a specific protein's tertiary structure *ab initio*. Here, "native" short- and long-range pairwise interactions, consistent with the target lattice structure, were defined. The effect of nonnative interactions inconsistent with the target structure was also studied. They concluded that the model system exhibited an all-or-none folding transition (33) provided that the long-range tertiary interactions were specific for the target structure. Certainly, these studies captured some important features of protein folding and inspired many other lattice models of proteins (35).

A very simple lattice model that could be treated exactly by enumeration of all compact states in short chains (36) is the HP model (16,37-39). The polypeptide is represented as a string of hydrophobic (H) and hydrophilic (P) beads (residues) on a simple square or cubic lattice. Hydrophobic residues attract each other, while the remaining possible pairwise interactions are equal to zero, except for excluded volume interactions. The model assumes that hydrophobic interactions play the predominant role in protein folding. This view was recently questioned by Honig and Cohen (40) who argued that interactions involving backbone hydrogen bonding are also important. In related models studied by Shakhnovich *et al.* (41-58), the same interaction strength was assumed for HH and PP pairs, while interactions for HP pairs were somewhat weaker. Eventually, the model was generalized to include all 20 amino acid types.

General questions related to the protein folding process could be addressed within the framework of this class of models, including the origin of the uniqueness of the native state (38,48,49). For some sequences, the collapse transition was very cooperative, while for others, it was continuous (59). The folding pathway(s) changed as well (49,60). Uniqueness is facilitated by incorporating a larger number of amino acids (48,49,56,61). The HP model has been used to search sequence space and to "design" optimal sequences (54,58,59). In this context, a variety of reduced models that stress various aspects of the physical forces governing protein folding (60,62-64) and dynamics (60,65-71) were proposed. These investigations have provided interesting insights into the protein folding process and have motivated the idea of a folding funnel. The development of the folding funnel theory has led to a considerable interest in the study of these *minimalistic* protein models, for which excellent reviews are available (72,73).

b. *Lattice Models of Intermediate Complexity*

The HP model assumes that protein folding is driven by long-range interactions and that short-range conformational propensities are only involved in structural fine tuning (74). Earlier, a different viewpoint emerged from studies of simple diamond lattice systems and the chess knight lattice model (75-88). These studies concluded that the native conformation emerges from an interplay of secondary structural preferences and tertiary interactions. As Go *et. al.* found (13,31-33), folding is more cooperative when the long- and short-range interactions are consistent with the native fold. This qualitative picture has persisted even when far more sophisticated versions employing knowledge-based potentials describing statistical secondary structure preferences and tertiary interactions were developed (35).

c. *High Coordination Number Lattice Models*

Reduced protein models with an increased number of lattice basis vectors were developed to improve their geometric fidelity with respect to real protein structures. For example, Kolinski & Skolnick employ a Cα virtual bond model where Cα atoms are connected by 90 possible virtual bond vectors (23). Regardless of protein size and orientation on the lattice, the Cα representation has about a 0.6-0.7Å coordinate root mean square deviation, cRMSD, with respect to experimental structures (22). The fact that space is essentially isotropic and that all structures (essentially independent of size) can be represented at comparable geometric resolution are the major reasons why this high coordination lattice has been developed. Side chains, which are not restricted to lattice points, are built from a set of rotamers, each at the side chain center of mass. Excluding Gly, Pro and Ala, each amino acid has multiple rotamers chosen so that the center of mass of any real protein side chain is no farther than 1Å from a member of the rotamer library.

i. Interaction Scheme

Recently, it has become popular to consider a simple interaction scheme (89,90) where amino acids are divided into two types: hydrophobic and hydrophilic. Such an approach is appealing because the number of energy parameters is small. However, when ranked by their energy, nativelike states are found in the best several hundred structures (90). While these results are better than random, there are too many alternative conformations for use in a practical prediction scheme. Thus, while complexity for complexity's sake is certainly to be avoided, to reduce the number of alternative low energy topologies and to reproduce essential features of the physics of proteins, more complicated interaction schemes have been developed (23,26,91,92). The empirical potential must capture both generic (sequence independent) and sequence specific features of proteins. An example of a generic term is hydrogen bonding (the free energy cost for the backbone residues not being hydrogen bonded is very high), whereas the preference of Glu and Lys to interact favorably is an example of a sequence specific contribution. As a example of an empirical potential, we describe our present realization. The empirical potential, E_{empir}, is given by

$$E_{empir} = E_{hb} + E_{sec} + E_{bur} + E_{pair} + E_{multi} \quad (1)$$

where E_{hb} is the hydrogen bond energy, E_{sec} accounts for statistical conformational preferences for secondary structure and contains both generic and sequence specific components, E_{bur} is a centrosymmetric burial potential that describes the preference of a given residue to lie a given distance from the center of mass of the molecule, E_{pair} accounts for the tertiary pair interactions, and E_{multi} describes the higher order multibody terms selected to reproduce the preferred side chain packing patterns of pairs of supersecondary elements (23,25,27,35,91,93-95). All parameters are available on the Internet (96) .

ii. Interplay of Intrinsic Secondary Structure Propensities and Tertiary Interactions

This class of models has been used to explore a number of the general features of protein folding thermodynamics as well as the prediction of tertiary structure. Without tertiary interactions, but with potentials reflecting local, amino acid pair specific, secondary structural preferences, on average the model correctly predicts 57% of the secondary structure of the native state (97). When tertiary interactions as approximated by a one-body centrosymmetric burial potential are included (98), the accuracy of secondary structure prediction increases to 66%, which is comparable to standard methods without multiple sequence alignments (99). Most importantly, this calculation provides support for the idea that the observed secondary

structure in a native protein results from the interplay between the short-range conformational propensities and tertiary interactions.

iii. Origin of Structural Uniqueness in Protein Models

For a putative 45-residue, six-stranded β-barrel, the factors responsible for the structural uniqueness of the native state were studied for a set of model protein sequences (26). Among possible origins of this uniqueness are hydrophilic/hydrophobic amino acid patterns (100) and the role of polar amino acids in destabilizing misfolded conformations (101). Thus, the first sequence studied had an alternating pattern of valines and serines in the putative β strand regions that were punctuated by appropriate turn forming residues. For this sequence, numerous low energy, six-stranded, β-barrel topologies resulted. In systems whose conformational entropy is comparable to real proteins (102), a simple HP pattern, even when punctuated by appropriate turn forming regions, does not have a structurally unique native state. To enhance the uniqueness of the hydrophobic core, four Phe residues were introduced. This diminished the number of distinct low energy topologies, but none was energetically favored. Substitution of Asp for Ser at positions designed to destabilize incorrect topologies yielded a sequence that adopted both the desired and mirror image topology (where the chirality of the turns is reversed, but the chirality of the intervening secondary structural elements is not, *e.g.*, helices would remain right-handed). Analysis of the energetic contributions indicated that tertiary interactions favored the desired fold, but that the turn residues favored the mirror image topology. Substitution with Gly linkers in the turns yielded the desired fold as being the most stable. Thus, this study suggested that the structural uniqueness of a globular protein requires a variety of polar and non polar residue types and that it is just as important to destabilize alternative conformations as to stabilize the native conformation.

iv. Origin of the Cooperativity of Protein Folding

Experimentally, many proteins exhibit a highly cooperative, two-state conformational transition (34) where the cooperativity often occurs on passage from the molten globule to the native state and is accompanied by the fixation of tertiary contacts (1,103-105). A minimal requirement for any protein model is that it qualitatively reproduce protein folding thermodynamics. The question is: what interactions are responsible for the origin of the cooperativity of protein folding? To explore this issue, the entropy sampling Monte Carlo method as developed by Scheraga and coworkers (87,88,106,107) was employed to investigate the folding thermodynamics of the model six-strand, Greek key, β protein described above (27). Two distinct models were considered. In model Type I, only pair potentials were used, whereas Type II models also included higher order side chain packing terms. The scale factors for the pair and higher order multibody interactions were adjusted so that the total tertiary energy in the putative native fold in Type I and II models was essentially the same. The two models give rise to qualitatively different behavior. Type I, lacking any high order multibody interactions, essentially had a continuous thermodynamic transition. On inclusion of higher order multibody packing interactions in Type II models, the conformational transition became all-or-none. Interestingly, the lowest energy states corresponded to the same structures in the two models. Proteins exhibiting two-state thermodynamic behavior have a decreased density of states (lower entropy) in that part of the energy spectrum near to, but higher in energy than the native conformation. These simulations suggest that the cooperativity of protein folding arises from cooperative tertiary interactions. Similar conclusions about the origin of two-state thermodynamic behavior have been presented recently by Hao and Scheraga (102). Finally, a series of studies exploring the relationship of the folding thermodynamics and kinetics for this designed sequence, protein A and the α/β protein G (108) has been undertaken. Similar qualitative conclusions emerge.

a. *Off-lattice Models*

The pioneering attempt to use a simplified protein model to predict protein structure is due to Levitt and Warshel (109) who succeeded in predicting a structure of BPTI whose cRMSD from native was about 6.5Å. The significance of these predictions was later questioned by Hagler and Honig (110), who obtained comparable quality structures using a glycine and alanine heteropolymer model of the BPTI sequence. Wilson and Doniach (111) subsequently developed a similar model that, when applied to crambin, yielded low resolution structures with several proteinlike features. More elaborate statistical potentials were used to predict the structure of short peptides (112,113). Such studies achieved better prediction accuracy, with errors ranging from 1.66Å cRMSD for the mellitin single helix to 4.5Å cRMSD for some larger polypeptides. Interestingly, Sun used a genetic algorithm for the conformational search (112), an idea subsequently employed by others (114,115). Srinivisan and Rose have employed a hierarchical approach based on the staged accretion of structure (89).

b. *Low Coordination Number Lattice Models*

Low coordination number lattice models have been used for tertiary structure prediction for almost 20 years (15,116-118). For example, Covell and Jernigan (116) enumerated all possible conformations of five small proteins restricted to fcc and bcc lattices. They found that the nativelike conformation always had an energy within 2% of the lowest energy. Virtually simultaneously, Hinds and Levitt (119) used a diamond lattice model where a single lattice unit represents several residues. While such a representation could not reproduce the geometric details of helices or β-sheets, the topology of native folds could be recovered with moderate accuracy.

c. *High Coordination Number Lattice Models*

Using a 56-neighbor (coarse) lattice to describe the Cα positions, Kolinski and Skolnick performed *ab initio* folding simulations (91) on two 73-residue sequences designed by DeGrado *et al.* (120-122). One sequence contained an all-leucine core, and the simulations predicted that the right- and left-handed four-helix bundles were isoenergetic, a prediction subsequently confirmed by experiment (122). They also explored the origin of protein folding cooperativity and found that including cooperative side-chain packing terms was necessary to mimic the process of side chain fixation associated with passage from the molten globule to the native state (103-105,123,124). The folding of a number of other single domain proteins was examined on both 56- and 90-neighbor lattices (23,24,92-94). For example, folding simulations of the B domain of protein A (24,25) yielded structures whose cRMSD from native in the ordered parts of the molecule is 3.3Å. The folding of crambin (without assuming the identity of the disulfide crosslinks) produced low energy conformations having an average Cα cRMSD below 4Å. A similar treatment of the folding of BPTI on a high coordination number lattice has also been done (125). Finally, the 90-neighbor lattice model was applied to predict the quaternary structure of the GCN4 leucine zipper (126) starting from two isolated, random coil chains (92). The lowest energy lattice structures have a Cα cRMSD from native ranging from 2.3 to 3.7Å. Using these structures, detailed atomic models were then built and relaxed using CHARMM with explicit water (127). The resulting average structure has a cRMSD of 0.8Å for the backbone atoms, 1.31Å for the heavy atoms in the dimerization interface, and 2.29Å for all heavy atoms, respectively. These studies demonstrate the compatibility of the reduced protein representation on a high coordination lattice with models at atomic detail.

a. *Use of Known Secondary Structure*

One way to improve the quality of tertiary structure predictions is to use known secondary structural information. In that regard, using an off-lattice model and exact knowledge of the secondary structure, Friesner *et al.* successfully folded two four-helix bundle proteins, cytochrome b562 and myohemerythrin, the large helical protein myoglobin and the relatively complicated fold of the α/β L7/L12 ribosomal protein (11,12,128). Furthermore, assuming known secondary structure and using a genetic algorithm to search conformational space, Dandekar and Argos (114) reported encouraging results on a test set of 19 small helical and β proteins where they succeeded in predicting a significant portion of these proteins at about 5Å resolution. However, use of predicted (rather than known) secondary structure information substantially degrades the performance of their prediction algorithm. Mumenthaler and Braun (129) have developed a self-correcting distance geometry method that assumes known secondary structure and successfully identified the native topology for 6 of 8 helical proteins.

b. *Folding with Correct Secondary Structure and Tertiary Restraints*

There have also been a number of studies that incorporate correct secondary structure and a limited number of correct tertiary restraints to predict the global fold. One of the very early studies is due to Vasquez and Scheraga (130). In addition, Smith-Brown *et al.* (131) have modeled a protein as a chain of glycine residues with restraints encoded via a biharmonic potential. Unfortunately, they find that a considerable number of restraints is required to assemble the native structure, thereby rendering the approach impractical for realistic situations. Another effort to predict the global fold of a protein from a limited number of tertiary restraints is due to Aszodi et al. (132). Their approach is very much in the spirit 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. These distances are obtained from patterns of conserved hydrophobic amino acids extracted from multiple sequence alignments. In general, they find that to assemble structures below 5Å cRMSD, on average, typically more than N/4 restraints are required, where N is the number of residues. Again, a key problem with all these approaches is the relatively large number of exact tertiary restraints required for successful topology assembly.

Using their 90-neighbor lattice model, Skolnick and coworkers have developed the MONSSTER (MOdeling of New Structures from Secondary and TErtiary Restraints) program for folding proteins using loosely defined knowledge of the correct secondary structure of regular fragments and a small number of exact tertiary distance restraints (133). The method also incorporates the empirical potentials described above reflecting statistical preferences for secondary structure, side-chain burial and pair interactions, and hydrogen bond contributions (131,132). Helical proteins can be folded with roughly N/7 tertiary restraints, while β and α/β proteins require about N/4 restraints, with N being the number of residues in the protein. However, if the empirical potentials are turned off, then with this level of restraint information, essentially random, compact structures result. Thus, there is an important synergism between the empirical contributions to the potential and the restraints. Of course, for any particular case, the accuracy depends on the restraint distribution (133,135). Most recently, Kolinski and coworkers have reported an approach that reduces the requirement for the folding of β and α/β proteins to N/7 known tertiary restraints (135). These studies served as calibration studies to develop our current protocol to structure prediction, which we summarize in what follows.

c. *Prediction of Tertiary Structure from Predicted Secondary Sructure and Tertiary Restraints*

a. *Overview of Methodology*

As depicted in Figure 1, our current approach to the prediction of protein structure can be conceptually divided into three stages (1). *Restraint derivation*, (2). *Structure assembly*, and (3). *Selection of the native conformation.* In addition, we present *objective, rigorous validation criteria* that are applied in order to judge the success of the prediction technique.

For *restraint derivation*, a multiple sequence alignment with the sequence of interest is generated (136). Then, predicted secondary structure restraints are obtained from a standard secondary structure prediction scheme. The predicted secondary structural elements define the predicted core regions of the molecule. Next, tertiary contacts (restraints), termed *seeds*, between these core elements are predicted from multiple sequence alignments. Multiple sequence information is used to derive such *seed* side chain contacts based on patterns of residue covariation in a set of homologous sequences (139-141). These *seed* contacts between predicted topological elements are then *enriched* by threading fragments of the test sequence through a structural database that typically produces about N/4 contacts, the number required for successful topology assembly (133,142).

In the *structure assembly* step, the set of predicted restraints is used in the MONSSTER method (133) to drive the conformational search. A series of up to 1000 independent, simulated annealing structure assembly runs are performed. Low energy structures are selected, typically the lowest 1% set of the complete pool of structures, and the resulting structures are clustered on the basis of their pairwise cRMSD. From these, we select representative structures from each of the families obtained, and proceed to the *native structure selection* stage, which consists of long isothermal runs from which the putative native topology is chosen on the basis that it has the lowest average energy (14,15). If the differing topologies cannot be selected on this basis, then the prediction consists of several lowest average energy

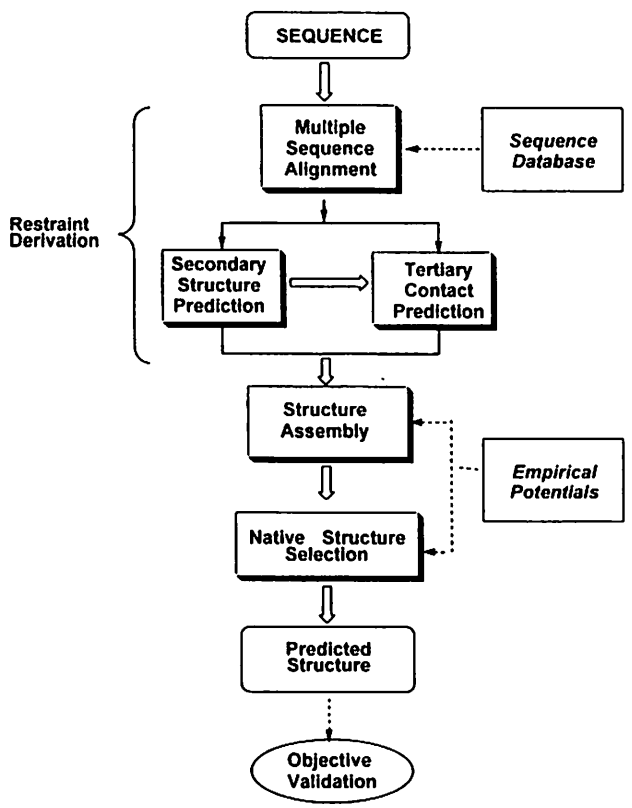


Figure 1: Schematic overview of the procedure for tertiary structure prediction.

representatives of the various generated topologies.

b. *Validation of the Tertiary Structure Prediction Method on Proteins of Known Structure*

We have applied this method to a set of 19 different non-homologous small proteins listed in Table IA. On applying this protocol, the following results were obtained: on average, the predicted secondary structure is 69% correct; this is slightly less than the reported average of this technique, which is 72 ± 9% (99,137). As to the contact prediction, and leaving out proteins for which known disulfides made a substantial contribution to the total set of contacts used in the simulations, about 75% of the predicted contacts are correct within ±3 residues. Often, there were also a number of grossly incorrect restraints that could potentially lead to non native topologies. Turning to the assembly stage, for a particular sequence, about 10-30% of the assembly runs with MONSSTER produced nativelike topologies, independent of the protein class.

Table Ia

Summary of Prediction Results

A. Proteins with known structure in advance of prediction.

Protein[a]	Type	N[b]	Q_3[c]	Nc[d]	Np[e]	Nw[f]	d=0[g]	d=2[g]	Native Topology			Lowest Energy, Nonnative Topology			Final Score[n]
									rms_n[h]	E_n[i]	rs_n[j]	rms_w[k]	E_w[l]	rs_w[m]	
3cti	small	29	50.5	39	6	0	83.3	100.	*3.8*	-107	6	6.7	-103	6	U
1ixa	small	39	70.5	48	5	0	100.	100.	*5.6*	-130	5	7.7	-131	5	S
1gpt	small	47	70.6	70	13	0	46.1	100.	*5.9*	-276	9	6.6	-142	10	S
1tfi	small	50	60.0	84	37	0	21.6	88.8	*5.9*	-202	28	7.0	-191	31	U
prota[o]	α	47	77.6	91	17	0	0	70.5	*3.1*	-246	2	9.4	-240	10	S
1ftz	α	56	63.5	149	12	1	25.0	58.3	*5.1*	-277	11	10.1	-270	15	S
1c5a	α	66	85.8	105	43	1	24.4	73.3	*4.2*	-194	20	9.8	-182	26	S
1pou	α	71	78.8	122	49	0	28.6	89.8	*3.5*	-418	18	11.9	-364	22	S
3icb	α	75	82.3	154	25	0	28.0	68.0	*4.5*	-406	21	12.6	-342	11	S
1hmd	α	85	90.6	157	20	2	10.0	65.0	*4.6*	-458	3	9.3	460	13	PS
1shg	β	57	67.1	109	39	0	28.2	100.	*4.5*	-420	19	6.7	-397	18	S
1fas	β	61	67.1	98	25	1	26.3	78.9	*6.2*	-330	19	9.37	-284	20	S
6pti	αβ	56	58.8	92	19	0	68.4	100.	*4.7*	-410	19	9.7	-397	18	U
1cis	αβ	66	64.7	144	23	0	8.6	78.2	*6.4*	-240	7	7.6	-232	7	S
1lea	αβ	73	63.5	131	41	2	9.7	75.6	*6.1*	-136	26	9.4	-115	27	U
1ubi	αβ	76	62.3	153	17	0	23.5	94.1	*6.1*	-238	9	11.5	-203	8	S
1poh	αβ	85	65.9	162	36	3	8.3	55.5	*6.5*	-336	42	11.7	-299	23	U
1ego	αβ	85	60.0	223	33	0	15.1	93.9	*5.7*	-417	20	9.0	-396	16	S
1ife	αβ	100	75.3	148	21	3	14.2	38.0	*6.7*	-419	15	8.2	-482	16	PS

Table Ib

B. Results for blind predictions.

T42[p]	α	78	80.8	150	24	1	29.1	58.3	*5.2-5.5*	*-362*	15	11.7	-360	8	*S*
KIX	α	81	87.7	320	37	11	26.3	57.9	*5.8*	*-477*	19	10.7	-479	26	*PS*

[a] Prot refers to the PDB access number of the protein studied.

[b] N is the number of residues in the protein in the PDB file.

[c] Q_3 is the percent of correctly predicted secondary structure. All proteins have a Q_3 within one standard deviation of the average.

[d] N_c is the number of contacts in the native structure.

[e] N_p is the number of predicted contacts.

[f] N_w is the number of contacts that are incorrect when no native contact is found within ±5 residues of a predicted contact.

[g] % of predicted contacts within d residues of a native contact.

[h] rms_n is the average cRMSD deviation in Å from the native structure.

[i] E_n is the lowest average energy (in kT) after refinement for the nativelike topology.

[j] r_n is the number of restraints satisfied in the nativelike topology.

[k] rms_w is the average cRMSD deviation from native in Å of the alternative topology of lowest energy.

[l] E_w is the lowest average energy (in kT) in the alternative topology after refinement runs.

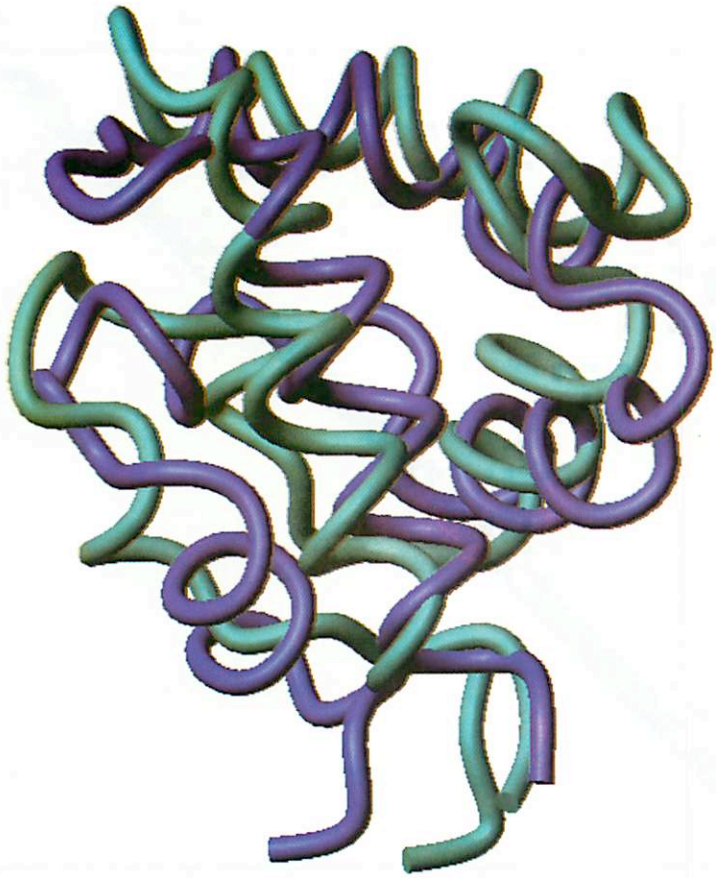
[m] rs_w is the number of restraints satisfied in the alternative topology.

[n] Relationship of lowest average energy structure to the native conformation if known. *S* indicates that the full structural selection criterion as assessed by the energy and DALI are "successful", *PS* indicates that the tertiary structure prediction is "partially successful", and *U* indicates that the tertiary structure prediction is "unsuccessful".

[o] The B domain of protein A (152).

[p] T42 is target 42 of the CASP2 meeting.

2a



2b

Figure 2: Blindly predicted structures of a) T42 superimposed on its native conformation; b) the left turning, three-helix bundle topology of KIX, superimposed on the experimental structure. The predicted structures are in cyan and the experimental structures are in blue. All figures were produced with MOLMOL (153).

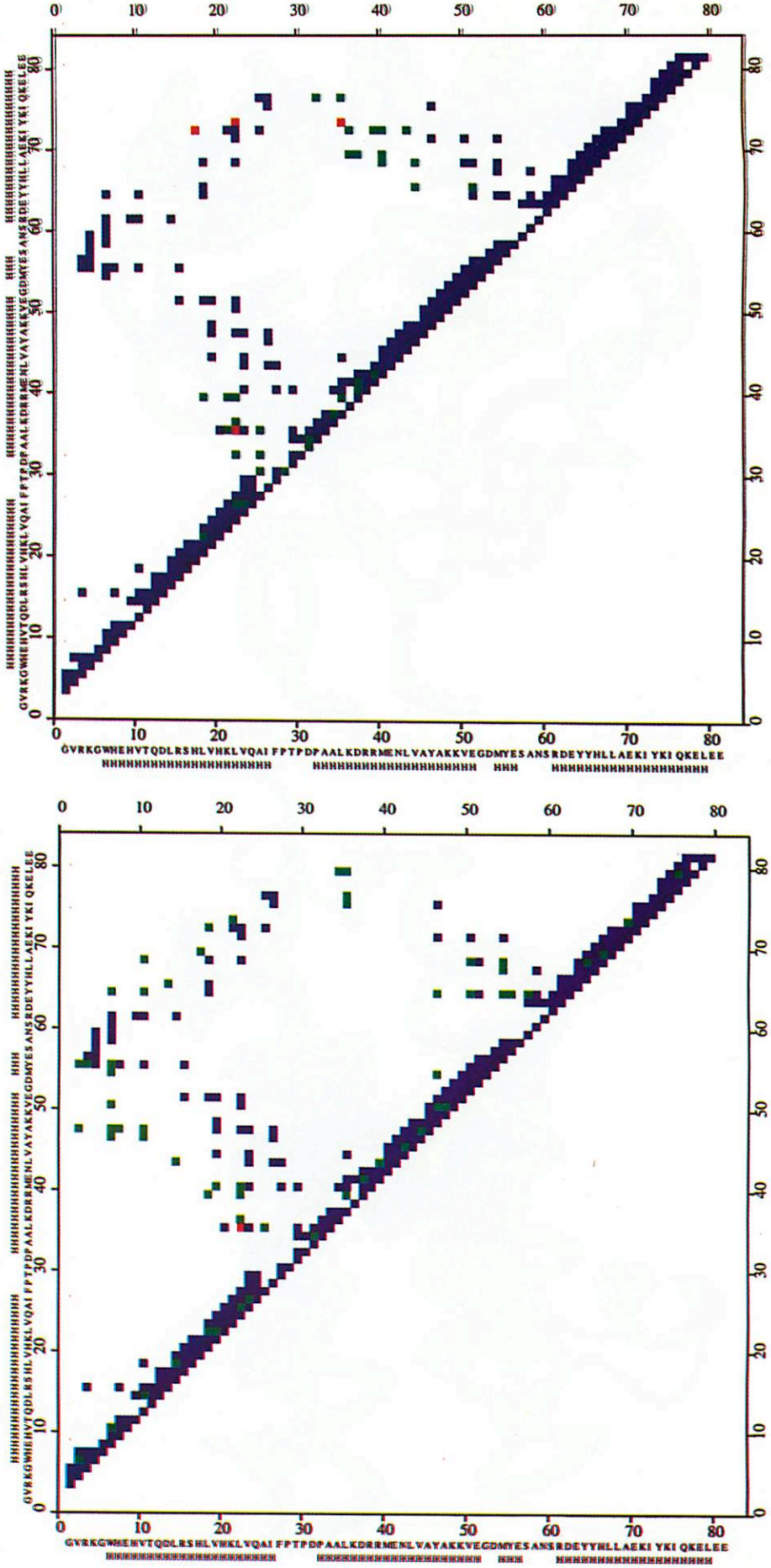


Figure 3: Contact map of the KIX domain. a) Predicted seeds (red), enriched contacts (green) and experimental contact map (blue); b) Contact map of the correct quirality of the predicted model (green) together with the contact map of the experimental structure (blue).

We *have judged success* as follows: The predicted folds were subjected to a structural similarity search over a representative database using the DALI (143) structural superimposition program. The prediction was a "success" if the predicted lowest energy structure, when used as a query with DALI, found the target structure or a homologue as a first hit in the search. On the other hand, if two or more of the selected topologies were isoenergetic, they were subjected to the same protocol, and if one matched the native topology, this was considered a "partial success". If the next lowest average energy matched the native fold rather than the lowest average energy structure, this was also considered to be a "partial success"; otherwise, the prediction was "unsuccessful". We also used DALI to generate the best structural superposition between the predicted and native structures so as to examine whether the differences between the predicted and actual structures reflect relatively minor shifts in registration due to errors in the predicted secondary and tertiary restraints, but not topological defects.

In 14 of 19 cases, "success" or "partial success" was obtained with the lowest average Cα cRMSD values ranging from 3.5 to 6.7Å. For the five "unsuccessful" cases, 3cti,1tfi, 6pti, 1lea and 1poh, DALI failed to find any structure that was significantly related to the lattice model; thus, the prediction is labeled as being "unsuccessful". Furthermore, in spite of the relatively high RMSD of the predicted folds, good topological predictions were made. Thus, DALI produced structural alignments between 2.7 and 4.0Å over about 75% of the sequence, on average. These results suggest that, when successful, this folding algorithm provides low resolution structures of comparable quality to those generated by threading techniques (145).

c. *Blind Predictions of Tertiary Structure*

We also examined the results of predictions on those proteins whose structures were not known at the time the predictions were made. Due to its inherent complexity, these studies are required for validation of any structure prediction technique. Here, we review two documented cases of blind predictions that illustrate both the virtues and flaws of the technique previously described.

i. Target 42 of the CASP2 Competition

At the second meeting on the Critical Assessment of Techniques for Protein Structure Prediction (CASP2), a variety of protein prediction targets were provided (146). We made a prediction on target 42 (T42) outside the competition when the experimental structure was still unavailable. This protein was selected because it was the most popular target sequence chosen by the *ab initio* protein folding groups.

Here, some details are given about the prediction of this target to give the reader a feeling about how the method performs in practice. From the multiple sequence alignment, a prediction of secondary structure was carried out using PHD. The prediction wrongly merges helices 3 and 4, and does not predict the last helix (Figure 2A). The correlated mutation analysis (139) provides three *seed* contacts (6-56, 7-70, 34-61), one involving a known disulfide crosslink (7-70). In this regard, since the CASP2 organizers provided the location of three disulfide bridges, we used this information in the subsequent topology assembly and refinement stages of the algorithm. The three *seeds* along with the three crosslinks were expanded to give 24 predicted restraints.

As indicated in Table IB, subsequent comparison of the blindly predicted and native structures indicates a backbone cRMSD ranging from 5.2-5.5Å from the set of NMR conformations, with a superposition of the predicted and native folds shown in Figure 2A. Although absent from the secondary structure prediction, the turn between the third and fourth helices is present in the predicted native structure.

However, predicted helix four extends from residues 59-62 as compared to residues 57-61 in the native conformation. Even though the C-terminal helix is entirely missed by the secondary structure prediction scheme, it partially forms in the predicted structure. These results, while quite encouraging, also demonstrate some of the shortcomings of this approach. Most salient is the fact that (see Table IB) the alternative, non native topology is only 2 kT higher in average energy than the native topology. If the energy is dissected into its components, the pair potential strongly favors the native fold by about 10-20 kT, whereas the restraint energy favors the alternative topology by about 10 kT.

ii. KIX Domain of the CREB Binding Protein

We also undertook the prediction of the 81-residue KIX domain of the CREB binding protein, which is involved in gene expression as mediated by AMPc (147,148). The secondary structure prediction scheme suggested that KIX should adopt a three-helix bundle fold. Correlated mutation analysis provided four seed contacts (22-35, 22-73, 35-73, 17-72), which when enriched yielded 38 predicted tertiary contacts. Figure 3A shows the predicted seeds together with the enriched set of contacts superimposed onto the experimental contact map.

Fold assembly simulations yielded either a left- or right-handed three-helix bundle. As indicated in Table IB, on the basis of their average energies, the two topologies are essentially isoenergetic. Decomposing the energy into its constituent contributions (149), the pair interactions, secondary structure preferences and hydrogen bond terms favor the right-handed bundle, whereas the burial energy and terms designed to generate proteinlike densities favor the left-handed bundle. The difficulty in distinguishing topological mirror images is a problem that this method often experiences with helical proteins, and indicates that improvements in the empirical potential are necessary. Figure 2B shows the predicted left-turning bundle superimposed on the experimentally determined NMR solution structure, which also adopts a left-turning, three-helix bundle fold. For the correct topology, a RMSD of 5.5Å with the experimental fold is obtained. This is in spite of the fact that one small helix was missed in the secondary structure prediction and there were shifts in registration in the contact prediction. Figure 3B shows the contact map of the left-turning bundle together with the experimental contact map. Note that although there is a general agreement, there are also considerable shifts in registration in the positions of the contacts. Since both the left- and right-handed bundle topologies are isoenergetic, our prediction criteria judge this to be a "partial success".

d. *Conclusions from Restraint Assisted Folding*

Based on these studies of small proteins, the following conclusions emerge. First, the current level of accuracy of existing secondary structure prediction schemes is adequate, in most cases, for this approach to tertiary structure prediction to work. However, if an element of secondary structure is entirely missed, then depending on its location in the native conformation, its absence might or might not prohibit successful tertiary structure prediction. Second, these low resolution models of small proteins can be assembled from rather inaccurate predictions of a subset (25%) of the total number of tertiary, side chain contacts. Third, helical proteins are predicted with higher accuracy than α/β proteins, and these are predicted with higher accuracy than β proteins. Fourth, the method usually predicts a handful of global folds, one of which is native, but selection of the native fold on the basis of its average energy is problematic.

Summary and Outlook for Future Progress

In the past decade, reduced models have provided an increased understanding of

globular protein systems. Nowadays, simplified lattice models are routinely used to elucidate the qualitative, very general features of the protein folding process. Furthermore, more complex lattice models, because of their increased detail and better treatment of the configurational entropy of compact, near native states can provide qualitative insights into the factors responsible for the observed two-state thermodynamic behavior of proteins. At least at the level of reduced protein models, it is becoming increasingly apparent that the specificity of the protein for its native fold arises from higher order than pair, multibody interactions. The derivation of such interactions will remain a challenge, but with the increasing size of the sequence and structural databases, it is quite likely that progress in this area will be made.

Turning to the problem of tertiary structure prediction for small single domain proteins, the combination of predicted secondary and tertiary restraints resulting from multiple sequence information should in many cases allow for the prediction of a handful of topologies, one of which is native. Aside from the obvious improvements in the method of contact map prediction, further progress demands the development of better conformational sampling approaches and improved energy functions, and efforts to achieve these goals are being made by a number of groups. Given such low resolution models of the native structure, a key question is what information can such models provide? Or put another way, how close does a model have to be to the native state for it to be useful? Recent work suggests that such low resolution models as produced by the current state-of-the-art can be used to identify protein active sites (150), but obviously they cannot be used to identify binding ligands. Methods that refine these types of low resolution models to higher resolution will have to be developed, and encouraging preliminary progress has been reported using molecular dynamics in explicit water (151). In the near future, studies of models at varying levels of resolution should prove to be a very powerful means of retaining the advantages of atomic detail where appropriate while simultaneously exploiting the advantages of conformational sampling provided by reduced representations. Thus, reduced models are likely to play an increasing role in computational biology.

Acknowledgments

This paper is dedicated with respect to Dr. David Beveridge for his contributions to the field of molecular modeling. This research was supported in part by NIH Grants GM37408 and P41 RR12255. AK is an International Scholar of the Howard Hughes Medical Institute. ARO acknowledges partial support from the Spanish Ministry of Education.

References and Footnotes

1. O.B. Ptitsyn, *Journal of Protein Chemistry* 6, 273-293 (1987).
2. B. Anfinsen, *Science* 181, 223-230 (1973).
3. L. Privalov and S.J. Gill, *Advances in Protein Chemistry* 39, 191-235 (1988).
4. B. Anfinsen and H.A. Scheraga, *Adv. Prot. Chem.* 29, 205-300 (1975).
5. L. Piela, J. Kostrowicki and H.A. Scheraga, *J. Phys. Chem.* 93, 3339-3346 (1989).
6. R. Ripoll, L. Piela, M. Vasquez and H.A. Scheraga, *Proteins* 10, 188-198 (1991).
7. C.L. Brooks, M. Karplus and B.M. Pettitt, *Proteins: A theoretical perspective of dynamics structure and thermodynamics* (Wiley, New York, 1988).
8. A. McCammon, *Rep. Prog. Phys.* 47, 1-46 (1984).
9. M. Karplus and G.A. Petsko, *Nature* 347, 631-639 (1990).
10. M. Levitt, *J. Mol. Biol.* 104, 59-107 (1976).
11. A. Monge, R.A. Friesner and B. Honig, *Proc. Natl. Acad. Sci. USA* 91, 5027-5029 (1994).
12. A. Monge, E.J.P. Lathrop, J.R. Gunn, P.S. Shenkin, and R.A. Friesner, *J. Mol. Biol.* 247, 995-1012 (1995).
13. N. Go and H. Taketomi, *Proc. Natl. Acad. Sci. USA* 75, 559-563 (1978).
14. Y. Ueda, H. Taketomi and N. Go, *Biopolymers* 17, 1531-1548 (1978).
15. V.G. Dashevskii, *Molekulyarnaya Biologiya (Translation from)* 14, 105-117 (1980).
16. 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).

17. B.H. Park and M. Levitt, *J. Mol. Biol.* 249, 493-507 (1995).
18. A. Sikorski and J. Skolnick, *J. Mol. Biol.* 215, 183-198 (1990).
19. A. Rey and J. Skolnick, *Proteins* 16, 8-28 (1993).
20. A. Rey and J. Skolnick, *Chemical Physics* 158, 199-219 (1991).
21. A. Rey and J. Skolnick, *J. Chem. Phys.* 100, 2267-2276 (1994).
22. A. Godzik, A. Kolinski and J. Skolnick, *J. Comp. Chem.* 14, 1194-1202 (1993).
23. A. Kolinski and J. Skolnick, *Proteins* 18, 338-352 (1994).
24. A. Kolinski and J. Skolnick, *Proteins* 18, 353-366 (1994).
25. J. Skolnick, A. Kolinski, C. Brooks III, A. Godzik and A. Rey, *Curr. Biol.* 3, 414-423 (1993).
26. A. Kolinski, W. Galazka and J. Skolnick, *J. Chem. Phys.* 103, 10286-10297 (1995).
27. A. Kolinski, W. Galazka and J. Skolnick, *Proteins* 26, 271-287 (1996).
28. R.L. Jernigan and I. Bahar, *Curr. Opin. Struct. Biol.* 6, 195-209 (1996).
29. S. Miyazawa and R.L. Jernigan, *Macromolecules* 18, 534-552 (1985).
30. J. Skolnick, L. Jaroszewski, A. Kolinski and A. Godzik, *Prot. Sci.* 6, 676-688 (1997).
31. H. Taketomi, F. Kano and N. Go, *Biopolymers* 27, 527-559 (1988).
32. *Protein folding*, Vol., edited by N. Go, H. Abe, H. Mizuno, and H. Taketomi (Elsevier/North Holland, Amsterdam, 1980).
33. H. Taketomi, Y. Ueda, and N. Go, *Int. J. Pept. Protein Res.* 7, 445-449 (1988).
34. J.S. Richardson and D.C. Richardson, *Science* 240, 1648-1652 (1988).
35. A. Kolinski and J. Skolnick, *Lattice models of protein folding, dynamics and thermodynamics* (R. G. Landes, Austin, TX., 1996).
36. R.C. Brower, G. Vasmatiz, M. Silverman and C. Delsi, *Biopolymers* 33, 329-334 (1993).
37. E.E. Lattman, K.M. Fiebig and K.A. Dill, *Biochemistry* 33, 6158-6166 (1994).
38. K.F. Lau and K.A. Dill, *Macromolecules* 22, 3986-3997 (1989).
39. P.D. Thomas and K.A. Dill, *J. Mol. Biol.* 257, 457-469 (1996).
40. B. Honig and F.E. Cohen, *Folding & Design* 1, R17-R20 (1996).
41. V.I. Abkevich, A.M. Gutin, and E.I. Shakhnovich, *J. Chem. Phys.* 101, 6052-6062 (1994).
42. V.I. Abkevich, A.M. Gutin, and E.I. Shakhnovich, *Protein Sci.* 4, 1167-1177 (1995).
43. V.I. Abkevich, A.M. Gutin, and E.I. Shakhnovich, *Folding & Design* 1, 221-230 (1996).
44. R.S. DeWitte and E.I. Shakhnovich, *Protein Sci.* 3, 1570-1581 (1994).
45. A.V. Finkelstein and E.I. Shakhnovich, *Biopolymers* 29, 1681-1694 (1989).
46. M. Karplus and E.I. Shakhnovich, in *Protein Folding*, edited by T. E. Creighton (W.H.Freeman, 1992), pp. 127-196.
47. L.A. Mirny, V. Abkevich, and E.I. Shakhnovich, *Folding & Design* 1, 103-116 (1996).
48. A. Sali, E.I. Shakhnovich, and M. Karplus, *J. Mol. Biol.* 235, 1614-1636 (1994).
49. A. Sali, E.I. Shakhnovich, and M. Karplus, *Nature* 369, 248-251 (1994).
50. E.I. Shakhnovich and A.V. Finkelstein, *Biopolymers* 28, 1667-1680 (1989).
51. E.I. Shakhnovich and A.M. Gutin, *Biophys. Chem.* 34, 187-199 (1989).
52. E.I. Shakhnovich, and A. Finkelstein, V., *Biopolymers* 26, 1681-1694 (1989).
53. E.I. Shakhnovich, G. Farztdinov, and A.M. Gutin, *Phys. Rev. Lett.* 67, 1665-1668 (1991).
54. E.I. Shakhnovich and A.M. Gutin, *Proc. Natl. Acad. Sci. USA* 90, 7195-7199 (1993).
55. E.I. Shakhnovich and A.M. Gutin, *Protein Engng.* 6, 793-800 (1993).
56. E.I. Shakhnovich, *Phys. Rev. Lett.* 72, 3907-3910 (1994).
57. E.I. Shakhnovich, *Folding & Design* 1, R50-R54 (1996).
58. E.I. Shakhnovich, *Curr. Opin. Struct. Biol.* 7, 29-40 (1997).
59. H.S. Chan and K.A. Dill, *J. Chem. Phys.* 95, 3775-3787 (1991).
60. C.J. Camacho and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* 90, 6369-6372 (1993).
61. E. O'Toole, R. Venkataramani and A.Z. Panagiotopoulous, *AIChE J.* 41, 954-958 (1995).
62. C.J. Camacho and D. Thirumalai, *Proteins* 22, 27-40 (1995).
63. J.D. Honeycutt and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* 87, 3526-3529 (1990).
64. J.D. Honeycutt and D. Thirumalai, *Biopolymers* 32, 695-709 (1992).
65. Z. Guo and D. Thirumalai, *J. Chem. Phys.* 97, 525-536 (1992).
66. Z. Guo and D. Thirumalai, *Biopolymers* 36, 83-102 (1995).
67. D. Hoffmann and E.W. Knapp, *Eur. Biophys. J.* 24, 387-403 (1996).
68. D. Hoffmann and E.W. Knapp, *Phys. Rev. E* 53, 4221-4224 (1996).
69. E.W. Knapp, *J. Comput. Chem.* 13, 793-798 (1992).
70. E.W. Knapp and A. Irgens-Defregger, *J. Comput. Chem.* 14, 19-29 (1993).
71. D. Thirumalai and S.A. Woodson, *Accts. Chem. Res.* 29, 433-439 (1996).
72. P.G. Wolynes, in *Physics of Biomaterials: Fluctuations, Self assembly and evolution*, edited by T. Riste and D. Sherrington (Kluwer Academic Publ., Netherlands, 1996), pp. 235-248.
73. K.A. Dill and H. Chan, *Nature Struct. Biol.* 4, 1-19 (1997).
74. H.S. Chan and K.A. Dill, *Macromolecules* 22, 4559-4573 (1989).
75. A. Sikorski and J. Skolnick, *Biopolymers* 28, 1097-1113 (1989).
76. A. Sikorski and J. Skolnick, *Proc. Natl. Acad. Sci. USA* 86, 2668-2672 (1989).
77. A. Sikorski and J. Skolnick, *J. Mol. Biol.* 212, 819-836 (1990).
78. A. Kolinski, M. Milik and J. Skolnick, *J. Chem. Phys.* 94, 3978-3985 (1991).
79. A. Kolinski and J. Skolnick, *J. Phys. Chem.* 97, 9412-9426 (1992).
80. J. Skolnick and A. Kolinski, *Annu. Rev. Phys. Chem.* 40, 207-235 (1989).
81. J. Skolnick, A. Kolinski and R. Yaris, *Biopolymers* 28, 1059-1095 (1989).
82. J. Skolnick, A. Kolinski and R. Yaris, *Proc. Natl. Acad. Sci. USA* 86, 1229-1233 (1989).
83. J. Skolnick and A. Kolinski, *J. Mol. Biol.* 212, 787-817 (1989).
84. J. Skolnick and A. Kolinski, *Science* 250, 1121-1125 (1990).
85. J. Skolnick, A. Kolinski and A. Sikorski, *Chemical Design Automation News* 5, 1-20 (1990).

86. J. Skolnick and A. Kolinski, *J. Mol. Biol.* 221, 499-531 (1991).
87. M.-H. Hao and H.A. Scheraga, *J. Phys. Chem.* 98, 9882-9893 (1994).
88. M.-H. Hao and H.A. Scheraga, *J. Chem. Phys* 102, 1334-1348 (1995).
89. R. Srinivasan and G.D. Rose, *Proteins* 22, 81-99 (1995).
90. K.A. Dill and K. Yue, *Prot. Sci.* 5, 254-261 (1996).
91. A. Kolinski, A. Godzik and J. Skolnick, *J. Chem. Phys.* 98, 7420-7433 (1993).
92. M. Vieth, A. Kolinski, C.L. Brooks III and J. Skolnick, *J. Mol. Biol.* 237, 361-367 (1994).
93. K.A. Olszewski, A. Kolinski and J. Skolnick, *Protein Eng.* 9, 5-14 (1996).
94. K.A. Olszewski, A. Kolinski and J. Skolnick, *Proteins* 25, 286-299 (1996).
95. A. Godzik, J. Skolnick and A. Kolinski, *Protein Eng.* 6, 801-810 (1993).
96. http://bioinformatics.scripps.edu
97. A. Kolinski, M. Milik, J. Rycombel and J. Skolnick, *J. Chem. Phys.* 103, 4312-4323 (1995).
98. A. Kolinski and J. Skolnick, *J. Chem. Phys.* 107, 953-964 (1997).
99. B. Rost and C. Sander, *J. Mol. Biol.* 232, 584-599 (1993).
100. K.A. Dill, *Biochemistry* 29, 7133-7155 (1990).
101. B. Harbury, T. Zhang, P.S. Kim and T. Alber, *Science* 262, 1401-1407 (1993).
102. M.-H. Hao and H.A. Scheraga, *J. Mol. Biol.* , in press (1998).
103. Kuwajima, *Proteins* 6, 87-103 (1989).
104. Kuwajima, M. Mitani and S. Sugai, *J. Mol. Biol.* 206, 547-561 (1989).
105. B. Ptitsyn, R.H. Pain, G.V. Semisotnov, E. Zerovnik and O.I. Razgulyaev, *FEBS Lett.* 262, 20-24 (1990).
106. M.-H. Hao and H.A. Scheraga, *J. Phys. Chem.* 98, 4940-4948 (1994).
107. H.A. Scheraga, M.-H. Hao and J. Kostrowicki, in *Methods in Protein Structure Analysis*, edit ed by M.Z. Atassi and E. Appela (Plenum Press, New York, 1995).
108. A. Kolinski, W. Galazka and J. Skolnick, *J. Chem. Phys.* 108, 2608-2617 (1998).
109. M. Levitt and A. Warshel, *Nature* 253, 694-698 (1975).
110. T. Hagler and B. Honig, *Proc. Natl. Acad. Sci. USA* 75, 554-558 (1978).
111. Wilson and S. Doniach, *Proteins* 6, 193-209 (1989).
112. Sun, *Protein Sci.* 2, 762-785 (1993).
113. Wallqvist and M. Ullner, *Proteins* 18, 267-289 (1994).
114. T. Dandekar and P. Argos, *J. Mol. Biol.* 256, 645-660 (1996).
115. A. Rabow and H.A. Scheraga, *Protein Sci.* 5, 1800-1815 (1996).
116. G. Covell and R.L. Jernigan, *Biochemistry* 29, 3287-3294 (1990).
117. R. Krigbaum and S.F. Lin, *Macromolecules* 15, 1135-1145 (1982).
118. G. Covell, *Proteins* 14, 409-420 (1992).
119. A. Hinds and M. Levitt, *Proc. Natl. Acad. Sci. USA* 89, 2536-2540 (1992).
120. Handel and W.F. DeGrado, *Biophysical J.* 61, A265 (1992).
121. Handel and W.F. DeGrado, *Biophysical J.* 61, A265 (1992).
122. M. Handel, S.A. Williams and W.F. DeGrado, *Science* 261, 879-885 (1993).
123. B. Ptitsyn, *Curr. Opinion Struct. Biol.* 5, 74-78 (1995).
124. *The molten globule state*, Vol. , edited by O.B. Ptitsyn (W.H. Freeman and Co., New York, 1992).
125. M.-H. Hao and H.A. Scheraga, *Proc. 1995 ACM/IEEE Supercomputer Conference*, 57 (1995).
126. Alber, *Curr. Opin. Genet. Develop.* 2, 205-210 (1992).
127. B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan and M. Karplus, *J. Comp. Chem.* 4, 187-217 (1983).
128. A. Friesner and J.R. Gunn, *Annu. Rev. Biophys. Biomol. Struct.* 25, 315-342 (1996).
129. Mumenthaler and W. Braun, *Prot. Sci.* 4, 863-871 (1995).
130. Vasquez and H.A. Scheraga, *J. Biomolecular Structure & Dynamics* 5, 757-784 (1988).
131. J. Smith Brown, D. Kominos and R. M. Levy, *Prot. Engn.* 6, 605-614 (1993).
132. Aszodi, M.J. Gradwell and W.R. Taylor, *J. Mol. Biol.* 251, 308-326 (1995).
133. J. Skolnick, A. Kolinski and A.R. Ortiz, *J. Mol. Biol.* 265, 217-241 (1997).
134. Wako and H. A. Scheraga, *Macromolecules* 14, 961-969 (1981).
135. A. Kolinski and J. Skolnick, *Proteins 32*, 475-494 (1998).
136. C. Sander and R. Schneider, *Proteins* 9, 56-68 (1991).
137. Rost, R. Schneider and C. Sander, *TIBS* 18 (1993).
138. A. Kolinski, J. Skolnick, A. Godzik, and W.P. Hu, *Proteins* 27, 290-308 (1997).
139. Goebel, C. Sander, R. Schneider and A. Valencia, *Proteins* 18, 309-317 (1994).
140. J. Thomas, G. Cesari and C. Sander, *Prot. Engn.* 11, 941-948 (1996).
141. Olmea and A. Valencia, *Folding & Design* 2, S25-S32. (1997).
142. R. Ortiz, A. Kolinski and J. Skolnick, *J. Mol. Biol.* , 277, 419-448 (1998).
143. Holm and C. Sander, *Nucleic Acids Res.* 25, 231-234 (1997).
144. J. Wodak and M. J. Rooman, *Current Opinion in Structural Biology* 3, 247-259 (1993).
145. T. Miller, D.T. Jones and J.M. Thornton, *FASEB* 10, 171-178 (1996).
146. http://iris4.carb.nist.gov/casp2/.
147. K. Brindle, and M. Montminy, R., *Curr. Opin. Genet. Develop.* 2, 199-204 (1992).
148. K. Brindle, S. Linke and M. Montminy, *Nature* 364, 821-824. (1993).
149. R. Ortiz, A. Kolinski and J. Skolnick, *Proteins* 30, 287-294 (1998).
150. S. Fetrow and J. Skolnick, *J. Mol. Biol.* in press (1998).
151. C. Simmerling, M. Lee, A.R. Ortiz, A. Kolinski, J. Skolnick and P.A. Kollman, *J.A.C.S.*, sub mitted (1998).
152. H. Gouda, H. Torigoe, A. Saito, M. Sato, Y. Arata and I. Shimada, *Biochemistry* 40, 9665-9672 (1992).

153. R. Koradi, M. Billeter, and K. Wuethrich, *J. Mol. Graph.* 14, 51-55 (1996).

Date Received: April 17, 1998

Communicated by the Editor Ramaswamy H. Sarma