Effective energy functions for protein structure prediction

Themis Lazaridis* and Martin Karplus†

Protein structure prediction, fold recognition, homology modeling and design rely mainly on statistical effective energy functions. Although the theoretical foundation of such functions is not clear, their usefulness has been demonstrated in many applications. Molecular mechanics force fields, particularly when augmented by implicit solvation models, provide physical effective energy functions that are beginning to play a role in this area.

Addresses
*Department of Chemistry, City College of CUNY, Convent Ave and 138th Street, New York, NY 10031, USA; e-mail: themis@sci.ccny.edu
†Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA and Laboratoire de Chimie Biophysique, ISIS, Institut le Bel, Université Louis Pasteur, 67000 Strasbourg, France; e-mail: marci@tammy.harvard.edu

Current Opinion in Structural Biology 2000, **10**:139–145

0959-440X/00/$ – see front matter

Abbreviations

CASP	Critical Assessment of Structure Prediction
EMBL	European Molecular Biology Laboratory
GB	Generalized Born
PB	Poisson–Boltzmann
PDB	Protein Data Bank
PEEF	physical effective energy function
SEEF	statistical effective energy function

Introduction

Approaches to protein structure prediction are based on the thermodynamic hypothesis, which postulates that the native state of a protein is the state of lowest free energy under physiological conditions. Normally, this state corresponds to the lowest basin of the effective energy surface. The term 'effective energy' or 'potential of mean force' refers to the free energy of the system (protein plus solvent) for a fixed protein conformation; that is, it consists of the intramolecular energy of the protein plus the solvation free energy [1,2]. Although the vibrational entropy of the folded state is large [3], it is approximately equal to that of a single unfolded conformer [4,5]. The large stabilizing entropy of the unfolded 'state' arises from the multitude of conformers of similar energy that contributes to it. It has been suggested that the native state is surrounded by an ensemble of similar conformations [6]. This appears to be true for a protein like α-lactalbumin, which has a molten globule intermediate that is stable at low pH [7] and consists of an ensemble of different conformers ([8]; E Paci *et al.*, personal communication). Although the existence of many such low energy conformers may aid in the search for a low-resolution model of the native state, it does not remove the requirement that the native (active) state be in a deep energy well; otherwise, it would not be the predominant species at physiological temperatures. In fact, the large increase in entropy between the native state and conformations that are significantly different (where 'significantly different' is defined by the fact that the protein is inactive in these non-native conformations) gives rise to the 'effective energy gap' requirement for the stability of the native state [9,10].

A counterexample to the full generality of the thermodynamic hypothesis has been given recently for α-lytic protease, whose native (functional) state appears to be kinetically trapped. It has been shown to be less stable than a molten-globule-like intermediate and the denatured state [11]. As the stability is determined by the free energy of unfolding (which equals the change in effective energy between the native state and the denatured states, plus the increase in conformational entropy on unfolding), this result is not inconsistent with the conclusion that the native state is that of lowest effective energy. For α-lytic protease, the measured enthalpy of the native state is lower than that of the misfolded intermediate by about 18 kcal/mol. Thus, the stabilization of the latter arises from the increase in conformational and solvation entropy. As has been suggested previously [12], the native conformation is likely to be lower in effective energy than any other *single* conformation, as it is even for this unusual system. Thus, the demonstration of kinetic control of folding for some proteins does not necessarily invalidate efforts to predict protein structures via minimization of an appropriate energy function.

Given the thermodynamic hypothesis, studies of protein folding (i.e. structure prediction, fold recognition, homology modeling and design) generally make use of some form of effective energy function. There are two very different types of energy function that are in use. The first is based on the true effective energy function, which can be obtained, in principle, from a fundamental analysis of the forces between the particles; the second is an energy function based on data derived from known protein structures (often statistics concerning pair contacts and surface area burial). We refer to the first type of energy function as a physical effective energy function (PEEF) and the second type as a statistical effective energy function (SEEF). The advantage of PEEFs is that we know that such functions exist, even if they may be very complicated and not necessarily convenient for protein structure prediction. As such functions are usually based on an atomic model, they result in a rough energy surface that does not decrease smoothly as the native state is approached [13]. However, a number of optimization techniques for such surfaces have been developed [14–17]. Several of these try to overcome the rough energy surface problem by generating a surface with the same or fewer minima and reduced energy barriers between them.

SEEFs are based on information from known protein structures, as mentioned above. Most often, the frequency

distribution of pair distances is used to extract a set of effective potentials between residues and, as the database of protein structures has grown, between atom pairs. In most studies, it has been assumed that a pairwise effective potential is sufficient. In certain cases, the pairwise terms have been augmented by introducing additional interactions (e.g. three- and four-body terms, surface area terms, probabilities of mainchain and sidechain dihedral angles, and/or whatever appears useful to the people constructing the SEEF). This lack of constraints on what to include in a SEEF is both a strength and a weakness of such functions. Some attempts have been made to optimize the SEEF not only by using known protein structures, but also by creating misfolded structures and choosing parameters to maximize the energy gap between them and the native state ([18–20]; H Holley, personal communication). For many applications of these functions, it is not necessary that the whole energy surface is reproduced correctly, only that the native state has the lowest effective energy. If the functions are physically meaningful, however, one might expect that, when they are expressed on an atom–atom basis, it should be possible to compare the SEEF effective pair potentials with PEEF values.

PEEFs typically consist of a molecular mechanics energy function and a model for the effect of solvation on the free energy. Thus, PEEFs are approximations to the (unknown) true energy function. They are empirical in that experimental information is used in the parameterization. However, the important difference from SEEFs is that the information used in parameterization of PEEFs is generally not extracted from protein structures, but rather from physical measurements of simple systems (small-molecule crystal and solvation data, as well as *ab initio* calculations); see MacKerell *et al*. [21] for a recent example of such parameterization for the CHARMM energy function [22] and Lazaridis and Karplus [23] for an empirical implicit solvation correction. Although widely used in molecular dynamics simulations of proteins in their native and denatured states [24], empirical energy functions and their PEEF extensions have been out of favor in protein structure prediction because of their greater computational cost and the impression that they cannot recognize native folds; the latter reason is based, in part, on the widely misinterpreted paper by Novotny *et al*. [25] (see, also, Lazaridis and Karplus [26]). As described below, this situation appears to be changing.

The primary emphasis of this review is on effective energy functions that recognize the native state and can therefore be used as part of structure prediction, whether by *ab initio*, homology-based or threading-based methodologies. Protein design is somewhat different. It is not important if some sequences that might work are excluded by the design process; rather, it is essential that most sequences that do not work are excluded. As packing considerations are a major factor in the latter, the primary use of energy functions has been concerned with excluded volume effects. For that purpose, PEEF-like functions with van der Waals terms and core exclusion terms have been used, augmented secondarily by hydrogen bonding and other stabilizing terms; such functions have been reviewed recently by Gordon *et al*. [27] and we do not consider them further here.

Reviews of energy functions for protein folding have appeared almost yearly in this journal [12,19,28–31]. This review covers recent developments in energy functions that are intended to describe proteins, rather than energy functions designed to endow a particular model with protein-like properties (e.g. Shakhnovich and Gutin [32]; Mirny and Shakhnovich [33]).

Statistical effective energy functions

As outlined above, SEEFs generally use the database of known protein structures to extract 'pseudo-potentials' for predicting unknown structures. One advantage of SEEFs over PEEFs, particularly when comparing residue-based SEEFs with atom-based PEEFs, is that the former tend to be less sensitive to small displacements. This makes them more robust for low-resolution protein structure prediction, for which small errors are inevitable. On the other hand, the price for this robustness may be a lower discriminatory power. Another advantage of SEEFs is that, because of their statistical nature, they can, in principle, include all known, and possibly unrecognized, physical effects. For example, the extended-atom model for aromatic rings often used in PEEFs does not capture cation–π interactions, although this is accounted for in all-atom models. Also, most PEEFs do not include polarization effects, which could give rise, for example, to hydrogen-bond cooperativity; polarizable empirical energy functions for proteins are under development ([34]; CL Brooks, personal communication). SEEFs, using known structures as the database, effectively include such effects, at least approximately.

The theoretical basis of SEEFs has been questioned based on theory [35] and lattice models [36], although some of the conclusions of the latter study have been challenged as being possible artifacts of the simple two-dimensional lattice that was used [37]. It has been shown also that a set of pairwise contact potentials cannot be constructed to distinguish the true energy minimum for even a single protein [38]. One source of possible artifacts in SEEFs is the lack of independence of the pair potentials. A striking example is an attractive interaction at medium range between aspartic acid residues [39]. This appears to be due to the clustering of aspartic acid residues around metals, so that the true potential for this case would be an attractive term between aspartic acid residues and metal ions. However, the interaction appears as an attraction between aspartic acid residues in the SEEF and favors incorrectly close positions of aspartic acid residues, even when a metal is not present. Once understood, such effects can be corrected easily (in this case by using a database that excludes metal ions, for example), but it is not straightforward to eliminate all such artifacts unless they are recognized. This type of criticism has led to

some rethinking and reformulation of SEEF potentials, but the problems have not been eliminated.

Most SEEFs employ a reduced representation of the protein. Many use a single interaction center at Cα or Cβ for each residue [30]. Distance-dependent energy functions employing a two-site representation of amino acid residues, one site for the backbone and one for the sidechain, are used by Bahar and Jernigan [37]. Their results indicate that charged and polar sidechain interactions are more important at close range than thought previously. The function was tested on decoys for eight small proteins constructed by sampling a few Cα–Cα virtual dihedral angles and was found to be useful for discriminating the native state in several cases [40]. It would be interesting to test this function on the Park and Levitt decoys [41]. The same SEEF has been used for a Monte Carlo simulation of the internal motions of ribonuclease H in its native state [42]. An alternative potential has been generated by combining a tertiary structure potential [43] with a repulsive contact potential and a short-range (local) secondary structure potential [44]. The paper also proposed a scheme to estimate the absolute stability of a protein for use in sequence recognition.

Eyrich *et al.* [45] have used a representation consisting of the backbone and a Cβ sphere sidechain model plus a van der Waals, hydrophilic and core repulsion potential determined from protein structures to predict protein folds. Using the experimental secondary structure and a branch and bound optimization [14], they obtained structures within 6 Å for a series of proteins.

Baker and co-workers [46] updated their earlier SEEF [39] by incorporating more heuristic contributions, such as terms for the relative orientation of secondary structure elements and sidechains and for hard sphere repulsion, compactness and the number of strands in β sheets; as before, the protein is represented by the heavy atoms of the mainchain plus Cβ to represent the sidechains. The authors found redundancies among various terms, which is expected as the same physical effects are included in more than one term. The modified function improved substantially in the discrimination of Park and Levitt [41] decoys, although 100% success was not achieved. Simulated annealing Monte Carlo simulations with this function, in conjunction with a library of small fragments from known protein structures with sequence similarity to the modeled fragment, produced encouraging results in the third Critical Assessment of Structure Prediction (CASP3), in which 'real' predictions are made based on protein sequences before the structure is known [47]. The same function has been used for predicting folding pathways [48]. In this case, additional requirements are placed on the SEEF; they must provide a physically meaningful description of the entire effective energy surface, which is not essential for structure prediction, *per se*.

Scheraga and co-workers [49–51] developed a united-residue force field based on a combination of database statistics and averaging of the ECEPP/2 force field; the function is thus a PEEF/SEEF hybrid. The chain is represented by the Cα and united-atom groups for the peptide bond and the sidechains (a total of three sites per residue). The interaction sites are located on the peptide bond and sidechain sites. This function is sometimes augmented by a cooperative (many-body) interaction potential for stabilizing regular secondary structure elements. The function has been successfully tested for decoy discrimination [49] and has performed reasonably well in *ab initio* folding tests for α-helical proteins [50,51].

SEEFs have begun to be used in physical applications, including the study of protein folding (see Riddle *et al.* [48] and Haliloglu [42], above). Skolnick and co-workers [52] used a SEEF potential to study the unfolding thermodynamics of the GCN4 leucine zipper, assuming that the potential can describe not only native states, but also non-native states. They used a lattice representation of the protein, with explicit consideration of only the Cα mainchain atom and the center of mass of the sidechains. The energy function includes terms for hydrogen bonding, Ramachandran torsions, sidechain orientational coupling, rotamer energy, burial energy, pair potential and (with a high weight) a cooperative pair potential, which is intended to correct the missing atomic detail. The extensive sampling performed in this study found that the lowest energy conformations are 2–4 Å away from the known crystal structure of GCN4. Furthermore, it was shown that, when applied to conformations generated by high-temperature molecular dynamics simulations in explicit water, the energies of the SEEF correlated well with the results obtained by use of the CHARMM all-atom potential function [53].

A PEEF-inspired potential with a reduced representation and a simplified energy function has been used by Osguthorpe [54] for structure prediction with molecular dynamics and simulated annealing. The protein is represented by one site per residue for the backbone and up to three sites for the sidechain, depending on sidechain size and shape. The energy function includes terms for the virtual bonds, angles and torsional angles, secondary structure propensities, nonbonded van der Waals repulsions, electrostatic interactions between the charged sidechains and a hydrophobic and an electrostatic solvation self-energy term. The model did well for a two-helix bundle structure in CASP3, but less so for more complex proteins.

Realization of the limitations of reduced protein representations (e.g. Mirny and Shakhnovich [33]) has led to the development of more detailed representations of proteins. Samudrala and Moult [55] implemented a SEEF that includes all protein heavy atoms and residue-specific atom types, rather than using a reduced set of atom types, as in many molecular mechanics programs (e.g. only one type of

CH_2 group). This leads to a large number of atom types (167) and pairwise potentials (167x168/2), some of which may be redundant; it is likely, for example, that the CD methyl group of leucine behaves similarly to the CG methyl group of valine. Samudrala and Moult [55] found that their SEEF performs significantly better than reduced representations with amino acid residues composed of one or two centers. Also, they showed that a distance-dependent potential is more effective in decoy discrimination than a step function contact potential. The SEEF was applied to several decoy sets from the PROSTAR web site with good results and has been applied also to comparative modeling [56] and sidechain rotamer prediction [57]. The CASP3 predictions [58] suggested that the function may not be able to discriminate near-native topologies from the native state. Another atom-based SEEF [59] uses a reduced set of atom types and seems to do equally well in decoy discrimination. As these functions [55,59] are closer than any other to a molecular mechanics function, it would be interesting to perform molecular dynamics simulations with it after adding the necessary bonded terms. Although such an all-atom function takes more time to evaluate than other statistical potentials, it appears to be less than for molecular mechanics energy functions; comparisons of calculation times would be of interest. A similar, all-atom SEEF was developed and employed to distinguish current from earlier (less accurate) protein structures in the Protein Data Bank (PDB), but it has not yet been tested on grossly misfolded decoys [60].

A different way of employing knowledge from the protein structure database is to adjust the parameters of a simple energy function so as to maximize the gap between the target native structure and a set of misfolded structures generated by threading onto known protein structures [19]. So far, it appears that the optimal parameter values are not transferable from one protein to another, that is, no universal function for protein folding has been discovered. A related strategy of adjusting parameters so as to maximize recognition of a set of training proteins without regard to either the physical nature of the interactions or known structure statistics has been proposed [61,62]; the resulting function has been used for protein fold recognition with some success.

An original approach to using the database of protein structures has been developed in the 'associative memory' Hamiltonian method of Wolynes and co-workers [63]. As it used molecular dynamics with simulated annealing to fold proteins, a molecular-mechanics-like backbone representation was added to the Hamiltonian. The approach has been applied recently with some success to predicting the structures of a series of proteins [20].

Physical effective energy functions

It has been assumed that molecular mechanics energy functions are not appropriate for protein structure prediction. A function that describes the energy of a protein in a vacuum should not be expected to always discriminate native from misfolded proteins, even though in many cases it does [25]. In recent calculations on the European Molecular Biology Laboratory (EMBL) decoy set [26], it was found that, if energy minimizations are employed to relax the structures, the vacuum energy function discriminated the native state in most cases, but when molecular dynamics was performed, it was rarely successful. This is due largely to the strong interactions between ionic sidechains.

The effect of the solvent has to be included in the energy function, not so much for the hydrophobic effect (the solvation free energy of nonpolar groups is small in magnitude), but mostly for the large desolvation cost of the polar groups. This is done in PEEFs by using the effective energy (the sum of the vacuum energy and the solvation free energy). The word 'effective' in PEEFs is well defined and means integrating out the solvent degrees of freedom, as pointed out in the Introduction. By contrast, effective energy SEEFs 'integrate out' not only the solvent, but also often the structural context, and approximate the full effective energy as a sum of pairwise residue–residue interactions plus heuristic terms.

Over the past few years, significant progress has been made in developing descriptions of the solvation free energy by using implicit models to replace an atomic representation of the solvent. Much of the work has been based on the dielectric continuum approach and the Poisson–Boltzmann (PB) equation, which is still computationally expensive, however. In addition, it has been difficult to perform molecular dynamics simulations with the PB equation. Recently developed analytical expressions for the derivatives of the PB solvation free energy will facilitate such applications [64]. A useful overview of implicit solvent models has been given by Roux and Simonson [65] as part of a special issue devoted to implicit solvent models for biomolecular simulations.

A PEEF based on a molecular mechanics energy function and a PB solvation free energy augmented by a surface-area-dependent hydrophobic term has been shown to discriminate a subset of the EMBL set of misfolded proteins [5,66]. Molecular dynamics simulations in explicit solvent were used to generate an ensemble of conformations for the native and the misfolded states and then the average effective energy of each ensemble was evaluated with the PEEF. The vibrational entropy was also estimated, but was found not to make an important contribution. Interestingly, the ensemble average effective energy discriminated the native state, although the function does not necessarily discriminate all individual conformations of the two ensembles. It should be noted that this hybrid approach does not provide a complete characterization of the effective energy function, as the conformations were sampled using a different function (explicit water simulations). A more stringent test would be to search conformational space with the function used in the evaluation to find minima on its energy

hypersurface and compare them with those corresponding to native proteins.

Because solving the PB equation is computationally demanding, simplified approaches have been developed. A promising method is based on the Generalized Born (GB) model [67–70]. The most time-consuming component in the original GB model is the calculation of the Born radii for the self energy by solving the PB equation. To reduce the computational cost, Schaefer and Karplus [71] proposed an analytical approximation to the self energy (ACE) that has been coupled with the CHARMM energy function and used in molecular dynamics simulations of peptides. It was shown that the combined functions can predict the correct structures of several peptides (e.g. α-helical versus β hairpin) and evaluate their thermodynamic properties in agreement with experiment [72]. In this work, the self energy was obtained from an integral of the dielectric displacement of each atom over the solute volume. As the solute volume can be broken into atomic contributions, the self energy is decomposed into a sum of pairwise terms. The dielectric displacement of each atom is approximated by the Coulomb field. The ACE potential has not yet been systematically tested on decoys, but the results so far indicate that it can recognize some wrong folds, such as those produced by threading (M Schaefer, personal communication).

An alternative analytical approximation to the self energy was proposed by Still and co-workers [73]. In this approach, the electric polarization is first estimated by starting with the polarization energy of an isolated site and, for each of the other solute atoms, adding a term proportional to V/r^4, where V is the volume of the atom. The Born radii are back-calculated from the estimated polarization energy. This model was parameterized for proteins and nucleic acids for use with the CHARMM energy function [74]. It has been employed in a molecular dynamics simulation of a protein and gave deviations from the crystal structure smaller than those of explicit solvent simulations. The computational cost was five times that of a vacuum simulation and 1/177 times that of an explicit solvent simulation. This solvation model when combined with the CHARMM energy function was also successful in discriminating the EMBL decoy set [53].

One approach to solvation that is not based on continuum electrostatics is the Gaussian solvent exclusion model [13,23]. When combined with a modified form of the CHARMM polar hydrogen energy function, the resulting PEEF is referred to as EEF1. Because it does not use surface areas, as in the work of Eisenberg and McLachlan [75], there is a large saving in computer time and simulations with EEF1 take only about 1.5 times longer than a vacuum simulation. The solvation model in EEF1 is based on a one-body parameterization. The pairwise desolvation interactions are obtained from a combination of individual group properties (in the same way that the van der Waals interactions can be obtained from the Lorentz–Berthelot combining rules); that is, there are no genuine pairwise parameters. This leads to a very small number of parameters, all of which are obtained from measurements of solvation free energies of small molecules.

EEF1 has been tested in several ways. Molecular dynamics simulations of the native state of several proteins at room temperature showed deviations from the experimental structures comparable to those of explicit solvent simulations. The energy function was applied to three standard decoy sets; they are the EMBL set, the Park and Levitt four-state reduced set and the CASP1 homology models. Perfect discrimination of the native state was obtained in all cases (except for two homology models with structures close to the native state) [26]. In a recent application to a set of 'hard' decoys generated by Baker and co-workers, EEF1 performed better than SEEFs of the type described by Simons *et al.* [46] (D Baker *et al.*, personal communication), although limitations of the energy function in distinguishing the native state also became apparent.

The original purpose in developing EEF1 was to obtain a realistic energy function for protein folding and unfolding simulations, rather than a convenient function for rapidly distinguishing native folds. The misplacement of a single sidechain in CI2 in an otherwise correct structure could increase the energy by several tens of kilocalories per mole, suggesting that EEF1 may be too sensitive to details for the study of low-resolution folds. Also, short simulations starting from unfolded states were trapped in compact non-native states, suggesting that the surface generated by EEF1 has multiple barriers [76].

EEF1, ACE and the analytical GB have all been implemented in the CHARMM program. More tests are needed to determine the relative strengths and weaknesses of these models. We hope to offer the possibility of evaluating protein models through a web server for CASP4 protein structure predictions.

Concluding remarks

It has been demonstrated that both SEEFs and PEEFs can be useful for protein fold evaluation. It is likely that a combination of the two will be employed in the future. For example, rapid structure generation based on SEEFs could be followed by their evaluation with a PEEF. Although the lack of a theoretical basis for SEEF remains a concern, it seems clear that many of the SEEFs now in use have physically correct features. Some of them may guide improvements in PEEFs.

With the large amount of work being invested in the development of effective energy functions of both types, it is important to submit them to well-defined tests. Decoy sets have become standardized and at least some of them are easily available, mainly on the World Wide Web. It should be required that, prior to publication, any function proposed for protein structure prediction of either the SEEF or PEEF type be tested on at least some of these sets; such a test with

the Levitt and Park [41] decoys was, in fact, suggested by a referee of the EEF1 paper and led to the encouraging results reported in [26]. Use of standard misfolded structures or other threading-type tests appears to be less demanding. Tests on decoys not used in constructing the SEEF are particularly important. As more knowledge-based terms are introduced, it is necessary to check the possibility that the function is simply learning to distinguish certain structures and can generalize to structures not in the training set. It is also worthwhile to search conformational space, at least partially, with the function being tested, as opposed to merely evaluating structures generated by other methods, to ensure that there are no false positives [77].

In the field of protein structure prediction, there is a tendency to regard an experimental structure as the ultimate truth. One has to keep in mind protein flexibility (e.g. alternative conformations with different sidechain orientations can have similar energies) and the effect of the environment, including crystal contacts and interactions with other molecules, such as cofactors or metal clusters. For example, a domain interacting with another protein or forming dimers may not be an autonomously folding unit. Some care is required, therefore, in evaluating predictions. Also, most SEEFs and PEEFs have been developed for proteins in aqueous solution. Different functions may be required for proteins in a membrane environment; in PEEFs, a different solvation contribution would be involved. The focus of this review, as of most research on SEEFs and PEEFs, has been on evaluating the structures of soluble, monomeric proteins that fold without extraneous factors.

Although much progress has been made recently, we close this review by reminding ourselves and the reader where we are in *ab initio* protein structure prediction. The assessment in CASP3 [78] indicates significant progress over CASP2; however, it was pointed out that "for the 'harder' targets more than 90% of the submissions gave a global rmsd greater than 10 Å". This makes clear that *ab initio* structure prediction remains a goal for the future, rather than a present reality.

Acknowledgements

Support by the National Science Foundation (MK) and by an RCMI grant from NIH to CCNY (TL) is gratefully acknowledged.

References

1. Karplus M, Shakhnovich E: **Protein folding: theoretical studies of thermodynamics and dynamics.** In *Protein Folding*. Edited by Creighton T. New York: WH Freeman; 1992:127-195.
2. Lazaridis T, Archontis G, Karplus M: **Enthalpic contribution to protein stability: atom-based calculations and statistical mechanics.** *Adv Protein Chem* 1995, **47**:231-306.
3. Brooks BR, Karplus M: **Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor.** *Proc Natl Acad Sci USA* 1983, **80**:6571-6575.
4. Karplus M, Ichiye T, Pettitt BM: **Configurational entropy of native proteins.** *Biophys J* 1987, **52**:1083-1085.
5. Vorobjev YN, Almagro JC, Hermans J: **Discrimination between native and intentionally misfolded conformations of proteins: ES/IS.** *Proteins* 1998, **32**:399-413.
6. Shortle D, Simmons KT, Baker D: **Clustering of low energy conformations near the native structures of small proteins.** *Proc Natl Acad Sci USA* 1998, **95**:11158-11162.
7. Ptitsyn OB: **Molten globule and protein folding.** *Adv Protein Chem* 1995, **47**:83-229.
8. Smith LJ, Dobson CM, van Gunsteren WF: **Side-chain conformational disorder in a molten globule: molecular dynamics simulations of the A-state of human α-lactalbumin.** *J Mol Biol* 1999, **286**:1567-1580.
9. Bryngelson JD, Wolynes PG: **Intermediates and barrier crossing in a random energy model (with applications to protein folding).** *J Phys Chem* 1989, **93**:6902-6915.
10. Sali A, Shakhnovich E, Karplus M: **Kinetics of protein folding: a lattice model study of the requirements for folding to the native state.** *J Mol Biol* 1994, **235**:1614-1636.
11. Sohl JL, Jaswal SS, Agard DA: **Unfolded conformations of α-lytic protease are more stable than its native state.** *Nature* 1998, **395**:817-819.
12. Moult J: **Comparison of database potentials and molecular mechanics force fields.** *Curr Opin Struct Biol* 1997, **7**:194-199.
13. Lazaridis T, Karplus M: **'New view' of protein folding reconciled with the old through multiple unfolding simulations.** *Science* 1997, **278**:1928-1931.
14. Westerberg KM, Floudas CA: **Locating all transition states and studying the reaction pathways of potential energy surfaces.** *J Chem Phys* 1999, **110**:9259-9295.
15. Wales DJ, Scheraga HA: **Global optimization of clusters, crystals, and biomolecules.** *Science* 1999, **285**:1368-1372.
16. Berne BJ, Straub JE: **Novel methods of sampling phase space in the simulation of biological systems.** *Curr Opin Struct Biol* 1997, **7**:181-189.
17. Foreman KW, Phillips AT, Rosen JB, Dill KA: **Comparing search strategies for finding global optima on energy landscapes.** *J Comput Chem* 1999, **20**:1527-1537.
18. Maiorov VN, Crippen GM: **Contact potential that recognizes the correct folding of globular proteins.** *J Mol Biol* 1992, **227**:876-888.
19. Hao M-H, Scheraga HA: **Designing potential energy functions for protein folding.** *Curr Opin Struct Biol* 1999, **9**:184-188.
20. Koretke KK, Luthey-Schulten Z, Wolynes PG: **Self-consistently optimized energy functions for protein structure prediction by molecular dynamics.** *Proc Natl Acad Sci USA* 1998, **95**:2932-2937.
21. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S *et al.*: **All-atom empirical potential for molecular modeling and dynamics studies of proteins.** *J Phys Chem B* 1998, **102**:3586-3616.
22. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M: **CHARMM: a program for macromolecular energy, minimization, and dynamics calculations.** *J Comput Chem* 1983, **4**:187-217.
23. Lazaridis T, Karplus M: **Effective energy function for proteins in solution.** *Proteins* 1999, **35**:133-152.
24. Brooks CL III, Karplus M, Pettitt BM: *Proteins: A Theoretical Perspective of Dynamics, Structure, Thermodynamics.* New York: John Wiley and Sons; 1998.
25. Novotny J, Bruccoleri R, Karplus M: **An analysis of incorrectly folded protein models. Implications for structure predictions.** *J Mol Biol* 1984, **177**:787-818.
26. Lazaridis T, Karplus M: **Discrimination of the native from misfolded protein models with an energy function including implicit solvation.** *J Mol Biol* 1999, **288**:477-487.
27. Gordon DB, Marshall SA, Mayo SL: **Energy functions for protein design.** *Curr Opin Struct Biol* 1999, **9**:509-513.
28. Jernigan RL, Bahar I: **Structure-derived potentials and protein simulations.** *Curr Opin Struct Biol* 1996, **6**:195-209.
29. Jones DT, Thornton JM: **Potential energy functions for threading.** *Curr Opin Struct Biol* 1996, **6**:210-216.
30. Sippl MJ: **Knowledge-based potentials for proteins.** *Curr Opin Struct Biol* 1995, **5**:229-235.

31. Vajda S, Sippl M, Novotny J: **Empirical potentials and functions for protein folding and binding.** *Curr Opin Struct Biol* 1997, **7**:222-228.

32. Shakhnovich EI, Gutin AM: **Engineering of stable and fast-folding sequences of model proteins.** *Proc Natl Acad Sci USA* 1993, **90**:7195-7199.

33. Mirny LA, Shakhnovich EI: **How to derive a protein folding potential?** *J Mol Biol* 1996, **264**:1164-1179.

34. Stern HA, Kaminski GA, Banks JL, Zhou R, Berne BJ, Friesner RA: **Fluctuating charge, polarizable dipole, and combined models: parameterization from *ab initio* quantum chemistry.** *J Phys Chem B* 1999, **103**:4730-4737.

35. Ben-Naim A: **Statistical potentials extracted from protein structures: are these meaningful potentials?** *J Chem Phys* 1997, **107**:3698-3706.

36. Thomas PD, Dill KA: **Statistical potentials extracted from protein structures: how accurate are they?** *J Mol Biol* 1996, **257**:457-469.

37. Bahar I, Jernigan RL: **Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation.** *J Mol Biol* 1997, **266**:195-214.

38. Vendruscolo M, Domany E: **Pairwise contact potentials are unsuitable for protein folding.** *J Chem Phys* 1998, **109**:11101-11108.

39. Simons KT, Kooperberg C, Huang E, Baker D: **Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions.** *J Mol Biol* 1997, **268**:209-225.

40. Ozkan B, Bahar I: **Recognition of native structures from complete enumeration of low-resolution models with constraints.** *Proteins* 1998, **32**:211-222.

41. Park B, Levitt M: **Energy functions that discriminate X-ray and near-native folds from well-constructed decoys.** *J Mol Biol* 1996, **258**:367-392.

42. Haliloglu T: **Characterization of internal motions of *E. coli* ribonuclease H by Monte Carlo simulations.** *Proteins* 1999, **34**:533-539.

43. Miyazawa S, Jernigan RL: **Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term for simulation and threading.** *J Mol Biol* 1996, **256**:623-644.

44. Miyazawa S, Jernigan RL: **An empirical energy potential with a reference state for protein folding and sequence recognition.** *Proteins* 1999, **36**:357-369.

45. Eyrich VA, Standley DM, Felts AK, Friesner RA: **Protein tertiary structure prediction using a branch and bound algorithm.** *Proteins* 1999, **35**:41-57.

46. Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D: **Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins.** *Proteins* 1999, **34**:82-95.

47. Simons KT, Bonneau R, Ruczinski I, Baker D: ***Ab initio* protein structure prediction of CASP III targets using ROSETTA.** *Proteins* 1999, **(suppl 3)**:171-176.

48. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D: **Experiment and theory highlight role of native topology in SH3 folding.** *Nat Struct Biol* 1999, **6**:1016-1023.

49. Liwo A, Pillardy J, Kazmierkiewicz R, Wawak RJ, Groth M, Czaplewski C, Oldziej S, Scheraga HA: **Prediction of protein structure using a knowledge-based off-lattice united-residue force field and global optimization methods.** *Theor Chem Acc* 1999, **101**:16-20.

50. Lee J, Liwo A, Ripoll DL, Pillardy J, Scheraga HA: **Calculation of protein conformation by global optimization of a potential energy function.** *Proteins* 1999, **(suppl 3)**:204-208.

51. Lee J, Liwo A, Scheraga HA: **Energy-based *de novo* protein folding by conformational space annealing and an off-lattice united-residue force field.** *Proc Natl Acad Sci USA* 1999, **96**:2025-2030.

52. Mohanty D, Kolinski A, Skolnick J: ***De novo* simulation of the folding thermodynamics of the GCN4 leucine zipper.** *Biophys J* 1999, **77**:54-69.

53. Mohanty D, Dominy BN, Kolinski A, Brooks CL III, Skolnick J: **Correlation between knowledge-based and detailed atomic potentials.** *Proteins* 1999, **35**:447-452.

54. Osguthorpe DJ: **Improved *ab initio* predictions with a simplified, flexible geometry model.** *Proteins* 1999, **(suppl 3)**:186-193.

55. Samudrala R, Moult J: **An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction.** *J Mol Biol* 1998, **275**:895-916.

56. Samudrala R, Moult J: **A graph-theoretic algorithm for comparative modeling of protein structure.** *J Mol Biol* 1998, **279**:287-302.

57. Samudrala R, Moult J: **Determinants of side chain conformational preferences in protein structures.** *Protein Eng* 1998, **11**:991-997.

58. Samudrala R, Xia Y, Huang E, Levitt M: ***Ab initio* protein structure prediction using a combined hierarchical approach.** *Proteins* 1999, **(suppl 3)**:194-198.

59. Melo F, Feytmans E: **Assessing proteins structures with a nonlocal atomic interaction energy.** *J Mol Biol* 1998, **277**:1141-1152.

60. Rojnuckarin A, Subramaniam S: **Knowledge-based interaction potentials for proteins.** *Proteins* 1999, **36**:54-67.

61. Ayers DJ, Huber T, Torda AE: **Protein fold recognition score functions. Unusual construction strategies.** *Proteins* 1999, **36**:454-461.

62. Huber A, Torda AE: **Protein sequence threading, the alignment problem and a two step strategy.** *J Comput Chem* 1999, **20**:1455-1467.

63. Goldstein RA, Luthey-Schulten ZA, Wolynes PG: **Optimal protein-folding codes from spin-glass theory.** *Proc Natl Acad Sci USA* 1992, **89**:4918-4922.

64. Friedrichs M, Zhou R, Edinger R, Friesner RA: **Poisson-Boltzmann analytical gradients for molecular modeling calculations.** *J Phys Chem B* 1999, **103**:3057-3061.

65. Roux R, Simonson T: **Implicit solvent models.** *Biophys Chem* 1999, **78**:1-20.

66. Vorobjev YN, Hermans J: **ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model.** *Biophys Chem* 1999, **78**:195-205.

67. Klopman G: **Solvations: a semi-empirical procedure for including solvation in quantum mechanical calculations of large molecules.** *Chem Phys Lett* 1967, **1**:200-202.

68. Constanciel R, Contreras R: **Self consistent field theory of solvent effects representation by continuum models: introduction of desolvation contribution.** *Theor Chim Acta* 1984, **65**:1-11.

69. Tucker SC, Truhlar DG: **Generalized Born fragment charge model for solvation effects as a function of reaction coordinate.** *Chem Phys Lett* 1989, **157**:164-170.

70. Still WC, Tempczyk A, Hawley RC, Hendrickson T: **Semianalytical treatment of solvation for molecular mechanics and dynamics.** *J Am Chem Soc* 1990, **112**:6127-6129.

71. Schaefer M, Karplus M: **A comprehensive analytical treatment of continuum electrostatics.** *J Phys Chem* 1996, **100**:1578-1599.

72. Schaefer M, Bartels C, Karplus M: **Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model.** *J Mol Biol* 1998, **284**:835-847.

73. Qiu D, Shenkin PS, Hollinger FP, Still WC: **The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii.** *J Phys Chem A* 1997, **101**:3005-3014.

74. Dominy BN, Brooks CL III: **Development of a generalized Born model parameterization for proteins and nucleic acids.** *J Phys Chem B* 1999, **103**:3765-3773.

75. Eisenberg D, McLachlan AD: **Solvation energy in protein folding and binding.** *Nature* 1986, **319**:199-203.

76. Lazaridis T, Karplus M: **Heat capacity and compactness of denatured proteins.** *Biophys Chem* 1999, **78**:207-217.

77. Abagyan R: **Towards protein folding by global energy optimization.** *FEBS Lett* 1993, **325**:17-22.

78. Orengo CA, Bray JE, Hubbard T, LoConte L, Sillitoe I: **Analysis and assessment of *ab initio* three-dimensional prediction, secondary structure and contacts prediction.** *Protein* 1999, **(suppl 3)**:149-170.