

# Empirical potentials and functions for protein folding and binding

Sandor Vajda\*, Manfred Sippl† and Jiri Novotny‡

Simplified models and empirical potentials are being increasingly used for the analysis of proteins, frequently augmenting or replacing molecular mechanics approaches. Recent folding simulations have employed potentials that, in addition to terms assuring proper polypeptide geometry, include only two noncovalent effects – hydrogen bonding and hydrophobicity, with extremely simple approximations to the latter. The potentials that have been used in the free-energy ranking of protein–ligand complexes have generally been more involved. These potentials have more detailed solvation models and account for both local (hydrophobic and polar) solute–solvent phenomena and long range electrostatic solvation effects. The models of solvation that have been used most frequently are surface area related atomic parameters, knowledge-based models extracted from protein-structure data, and continuum electrostatics with an additional area-related parameter. The knowledge-based approaches to solvation, although convenient and accurate enough, are suspect of double counting certain free-energy terms.

## Addresses

\*Department of Biomedical Engineering, Boston University, 44 Cummington St, Boston, MA 02215, USA;

e-mail: vajda@enga.bu.edu

†Department for Applied Molecular Engineering, Institute for Chemistry and Biochemistry, University of Salzburg, Jacob Haringer Strasse 1, A-5020, Salzburg, Austria

‡Department of Macromolecular Structure, Bristol-Myers Squibb Research Institute, Princeton, NJ 08543-4000, USA

Current Opinion in Structural Biology 1997, 7:222–228

Electronic identifier: 0959-440X-007-00222

© Current Biology ISSN 0959-440X

## Abbreviations

2D	two-dimensional
3D	three-dimensional
ASP	atomic solvation parameter
MC	Monte Carlo
rmsd	root mean square deviation

## Introduction

Simplified theoretical treatments are being increasingly applied to problems of protein folding and complexation. A common feature of all these formulas (effective interparticle potentials, Gibbs/Helmholtz free-energy functions and potentials of mean force) is the representation of complex solute–solvent interactions and solute entropic effects (e.g. hydrophobicities) in an implicit manner. The empirical approaches have emerged as a response to various limitations that are inherent in computational methods established in the past. Molecular mechanics, itself a linear combination of simple empirical terms,

has served well wherever effects of covalent bonding, excluded volumes, and coulombic electrostatics needed to be investigated *in vacuo*; however, it has proven inadequate for a thermodynamical description of stable, compact protein folds that appear to be determined by the nature of their solvent-exposed surfaces and, implicitly, by phenomena such as hydrophobicity and electrostatics at dielectric boundaries [1]. The same can be said of macromolecular complexes, of practical problems of biological specificity, and tasks such as discriminating between correct and incorrect protein folds and protein–protein interactions (protein–ligand docking) [2]. It is still widely believed that the correct answers to all this biology may emerge from applying molecular mechanics force fields to explicit models of solvated macromolecules, for example, by running extensive molecular dynamics or Monte Carlo (MC) simulations. The sheer computational burden of such an exercise remains prohibitive, however, notwithstanding the problems of convergence, adequate conformational sampling, and proper parametrization that continue to loom very large over this field. Molecular mechanics/dynamics is also at a serious disadvantage because of its extreme sensitivity to small perturbations of structures embedded in a phase space with very many local minima and characterized by a rugged potential landscape.

Of course, disadvantages to empirical potentials and functions of state also occur. Simplifications beget conceptual problems and often lead to answers which may be only approximate and of uncertain generality. A discussion of such conceptual issues is the main subject of our review.

## Background

The main goal of molecular mechanics has been to provide meaningful molecular structures, and its formalisms have been based on the mental picture of atoms connected by rotatable (rigid or flexible) bonds, and being acted on by pairwise potentials. Protein structures, however, present themselves as complicated, solvophilic/solvophobic polymers at equilibrium with the solvent, and calling for the use of thermodynamics—a theory that is preoccupied with macroscopic observables. Protein folding and complexation can be regarded as quasistatic processes between two equilibrium states. At a constant temperature and pressure, the system is best described by Gibbs or Helmholtz free energy, which are essentially identical to each other for condense systems. Thus, any effective potential, however simplified, may be usefully viewed as an approximation to the free-energy change  $\Delta G = G - G_0$ , where  $G_0$  is the free energy in a reference state (i.e. an unfolded or unbound conformation). Keeping this in mind, we search for free-energy expressions with a minimum

number of terms, each calculated by the most convenient, fast and accurate method [3]. One generally applicable decomposition is given by

$$G = \Delta G_{\text{conf}} + \Delta G_{\text{solv}} + \Delta G_{\text{other}} \quad (1)$$

where the sum  $\Delta G_{\text{conf}} + \Delta G_{\text{other}}$  represents the free-energy change calculated in a reference medium, most frequently in vacuum or a nonpolar liquid (i.e. the molecular mechanics paradigm), and  $\Delta G_{\text{solv}}$  is the solvation term, which describes the transfer of the folding or binding reaction from the reference medium into water [4,5]. The conformational free energy,  $\Delta G_{\text{conf}}$ , is defined by  $\Delta G_{\text{conf}} = \Delta E_{\text{conf}} - T\Delta S_{\text{conf}}$ , where  $\Delta E_{\text{conf}}$  is the conformational energy change in the reference medium. The entropy-change term,  $\Delta S_{\text{conf}}$ , is present because, due to the mean field character of the potential, a particular conformation (actually a subset of the phase space) may represent an ensemble of equienergetic structures such as different sidechain rotamers. In the most general case,  $\Delta G_{\text{other}}$  includes translational, rotational, vibrational, cratic and protonation/deprotonation effects (see also Brady and Sharp, pp 215–221, in this section).

The solvation term,  $\Delta G_{\text{solv}}$ , includes both local solute–solvent interactions, which are often assumed to be proportional to the solvent accessible surface area [5], and long range solvation effects due to dielectric screening and polarization [6]. Sometimes the hydrophobic term (surface tension) captures the work of forming a solute-containing cavity in solvent as applied to polar and nonpolar atoms alike [4]. Dielectric solvation effects, proportional to partial atomic charges, are then separately ascribed to all the polar atoms. At other times, solvation has been considered only for nonpolar atoms, thereby reducing  $\Delta G_{\text{solv}}$  to a (short range) hydrophobic term,  $\Delta G_{\text{hydro}}$ . Implicit representations of solvation effects yield equations that formally resemble *in vacuo* interaction potentials, and deciding whether we are practising simplified thermodynamics, amended molecular mechanics, or an empirical mean force potential is often difficult.

### Folding simulations on lattice models

The main advantage of lattice models has been the finite number of discrete conformational states that allow an extensive, and often exhaustive, exploration of the conformational and sequence spaces [7]. The potentials employed in lattice simulations have had an appearance of ideal interaction potentials, accounting for both solute–solute and bulk solvent effects without any clear, formal distinction between them. Such functions have proved to be successful in capturing some characteristics of protein folding, and hence similar potentials are now being used in off-lattice, nearly real, simulations.

The simplest potential used to fold 2D, lattice-bound polymers is the HP (hydrophobic–polar) model, a step function that accounts for a pairwise hydrophobic at-

traction between nonpolar particles [7]. The two-letter force-field alphabet can be extended, with strengths of interaction made more varied [8] or, better still, based on an empirical contact potential [9••]. The potential may include baseline attractions among all monomers (perturbed homopolymer model), or it may include repulsions [10•,11••]. When phase spaces of polymers governed by these interactions are fully enumerated, one may obtain ensemble quantities: degeneracy, that is, the number of sequences that fold to a given number of structures; encodability, the number of folds that are unique lowest energy structures of certain monomer sequences [10•]; and foldability, the ability to fold rapidly to the global energy minimum [7,8,11••,12]. Degeneracies and encodabilities of real proteins are poorly known, however, and it is not clear which lattice potential yields the most protein-like behavior [10•]. Another problem is the accuracy of energy calculations [13,14•]. Furthermore, without a realistic representation of the backbone, lattice models are unable to describe hydrogen bonding, which is an important structural constraint in off-lattice simulations [15••]. Thus, despite their elegance and undeniable usefulness as an ideal metaphor for folding, the value of lattice representations has occasionally been questioned [15••].

### Off-lattice folding simulations

Folding simulations on idealized proteins have a long history full of excitement and controversy [16–18]. In the recent past, the formation of secondary structures has been given much attention [19,20], and various types of folding functions have been experimented with [21•,22–26]. Success in predicting the folded conformations of small proteins or protein fragments has been claimed using several *de novo* methods [27••,28,29••,30•] that use simple 3D models of polypeptides and interaction potentials akin to those used with lattices. The protein models generally maintain the planarity of the peptide bonds, have fixed bond lengths and angles, and simplified sidechains. The potentials used have the general form

$$\Delta G = \Delta E_{\text{conf}} + \Delta G_{\text{hydro}} \quad (2)$$

where the energy change  $\Delta E_{\text{conf}}$  usually includes a hydrogen bond term  $\Delta E_{\text{H-bond}}$ , and possibly other directional, pairwise contributions.

Srinivasan and Rose [27••] have represented sidechains by one, two or three virtual atoms and use the energy change

$$\Delta E_{\text{conf}} = \Delta E_{\text{H-bond}} + \Delta E_{\text{tors}} + \Delta E_{\text{excl}} \quad (3)$$

where  $\Delta E_{\text{tors}}$  and  $\Delta E_{\text{excl}}$  are torsional and excluded volume energy terms, and the only favorable contribution is the hydrogen-bond energy  $\Delta E_{\text{H-bond}}$ . The hydrophobic contribution,  $\Delta G_{\text{hydro}}$ , accounts for short range interactions among three types of residues: hydrophobic (H), amphipatic (A), and polar. Only H–H and A–H contacts

contribute to the free energy at a ratio of 2:1, both scaled by the distance of the contacts. Dill and coworkers [28,29••] have developed two simulation programs on the basis of simple peptide geometry and potential. Each sidechain is represented by a single virtual atom fixed at the sidechain centroid in the position of its average rotamer. Lee *et al.* [30•] employ a similar model for MC simulations. In the peptide simulations by Sung [31], the solvation free-energy terms include both short range hydrophobicity (group-based hydrophobicity parameters) and long range electrostatic hydration, which are obtained from an approximate solution of the Poisson–Boltzmann equation. The energy contribution,  $\Delta E_{\text{conf}}$ , covers both hydrogen bonding and excluded volume terms. Although the simple functions used in [27•,28,29••,30•,31] seem to provide better results than the more complex ones used in the past, successful folding still remains somewhat anecdotal and difficult to reproduce and judge independently. Also difficult is a comparison of the above computer experiments, which are based on simplifications, with those employing full molecular mechanics force-fields that have added solvation terms (e.g. [32,33,34•,35•]). It may be interesting to note that at the Second Asilomar Meeting on the Critical Assessment for Protein Structure Prediction (CASP-2; [36]), methods directly exploiting protein structure data performed better than any of the *ab initio* folding methods.

### Structure-based (database-derived) potentials

A popular approach for deriving effective potentials from an ensemble of experimentally determined protein structures consists of computing frequencies of structural features ('structural frequencies'), and converting these frequencies into free energies [37•,38–41]. The main application of structure-based potentials is in fold recognition, which is reviewed by Torda (pp 200–205) in this section. These potentials have frequently been combined with other terms to form effective free-energy functions for folding and binding, however, and hence we will briefly discuss them.

The observed frequencies (e.g. pairwise residue distributions) can be transformed, in a straightforward way, into contact or pairwise potentials that, in principle, should include all thermodynamically important contributions. In practice, however, some contributions can be underestimated due to the selection of the reference state, or the neglect of chain connectivity and correlations between interactions [41,42••,43]. Thus, distinctions among structural frequencies, particle–particle contact potentials, and the underlying physical phenomena (e.g. hydrophobicity) are by no means clear. For example, it has been suggested that the Miyazawa–Jernigan-type [9••,41,44•] and other [45••] structure-based potentials correlate with solvation free energies and can be used as implicit solvation terms in free-energy expressions [44•,45••], in other words, in combination with other terms [45••]. Although such combined potentials may be useful in applications, their

theoretical basis is very weak and assuring that no contribution to the free energy is double counted is difficult.

### Docking and binding guided by free-energy potentials

Docking algorithms generate many alternative juxtapositions of a ligand (often a small organic molecule) and a receptor (always a protein) while ranking each of them by a scoring function [46•]. At current computer clock speeds, an oversimplified scoring function is essential for efficient and fast docking search [47•,48•]. The identification of near-native conformations, however, usually requires target functions that have a fairly faithful representation of the free energy of the system [49•]. The need for balancing computational speed, on the one hand, and numerical accuracy, on the other hand, is particularly acute in applications such as structure-based drug design [50••].

In the rigid, lock-and-key approximation to binding, surface complementarity has emerged as a single, plausible scoring-function candidate [46•], which is often used at a resolution coarser than the atomic scale [47•,48•]. Shape complementarity alone often yields false positives [2], however, and attempts have been made to amend it with pairwise potentials that capture hydrogen bonding and electrostatic complementarity [46•] or to expand it into an approximate free-energy expression [49•].

The general formalism given by Equation 1 describes the binding free energy of the system whenever the unbound but solvated proteins define the reference state [4,5]. Thus,  $\Delta G = \Delta G_{rl} - \Delta G_r - \Delta G_l$ , where  $\Delta G_{rl}$ ,  $\Delta G_r$ , and  $\Delta G_l$  are the free energies of the receptor–ligand complex, the free receptor, and the free ligand, respectively. Similarly, the energy component is defined by  $\Delta E_{\text{conf}} = \Delta E_{rl} - \Delta E_r - \Delta E_l$ . For rigid body association,  $\Delta E_{\text{conf}}$  reduces to the receptor–ligand interaction energy  $\Delta E_{r-l}$ . Furthermore, assuming a rigid backbone implies that the conformational entropy change is restricted to the sidechains. Thus, we can rewrite Equation 1 in the form

$$\Delta G = \Delta E_{r-l} - T\Delta S_{\text{sc}} + \Delta G_{\text{solv}} + \Delta G_{\text{other}} \quad (4)$$

where  $\Delta S_{\text{sc}}$  denotes the sidechain entropy loss upon binding [4,5]. Notice that using a (nonpolar) liquid as reference medium,  $\Delta G_{\text{solv}}$  includes only a small differential van der Waals contribution, and hence the van der Waals terms should also be removed from the solute–solute interaction energy  $\Delta E_{r-l}$ , which is thereby reduced to the electrostatic component. In this instance,  $\Delta G_{\text{other}}$  becomes a constant.

Practical free-energy functions used in docking are of three different kinds: partition methods [4,5,49•,50••,51–53], which are based on a 'master' thermodynamic equation [50••] (e.g. Equation 4) that has free-energy terms

formulated in the most practical way; structure-based, contact or pairwise-particle potentials in which free energies may be obtained from structural frequencies observed in receptor–ligand complexes [54<sup>••</sup>,55<sup>•</sup>,56]; and regression-type methods in which the scoring function is a linear combination of quantities, such as conformational energy, hydrophobic and hydrophilic surface areas, number of hydrogen bonds etc., that are fitted to experimental binding data [57–61]. The structure-based and regression-type methods have been relatively accurate when applied to a set of homologous molecules [59,60] but may not be transferable among dissimilar systems. By contrast, the various terms in free-energy functions based on partitioning (Equation 4) can be independently validated from thermodynamic and structural data. These functions are expected to be quite general and their parameters should be widely transferable.

The partition-type methods may be grouped according to the solvation models used, that is, atomic solvation parameter (ASP) models [4,5,49<sup>•</sup>,50<sup>••</sup>,51,52], structure-based solvation terms [44<sup>•</sup>,45<sup>••</sup>,53], and continuum electrostatic models of solvation [62<sup>•</sup>,63<sup>•</sup>]. The hydrophobic solvation or ASP models are empirical relationships between the solvation free energy and either solvent accessible surface areas or first hydration-shell volumes. The use of different reference media (nonpolar liquids or a vacuum) explains many of the differences among the various sets of ASP values reported in the literature for the same atom type [64,65<sup>•</sup>,66<sup>•</sup>]. Both the receptor–ligand interaction energy and the sidechain conformational entropy loss can be included within atomic surface area terms, and the ASPs have also been adjusted to take, for example, hydrogen bonding [51] or atomic charges [52] into account. These approaches have yielded good estimates of binding free energy, but at the expense of the ASP values becoming conceptually rather ill defined. As we noted above, the main problem with the structure-based solvation terms is that they probably include other contributions to the free energy, which are therefore double counted in the potential.

The atomic parameter models capture only local solvation, and the structure-based solvation terms are dominated by short range phenomena. Long range effects of electrostatic solvation and dielectric screening have sometimes been modeled by introduction of a distance-dependent dielectric constant into the Coulomb equation. Finite difference solution of the Poisson–Boltzmann equation is known to give a much more accurate description of electrostatic free energy [6], however, the technical problems associated with discrete space representation (gridding), infinite point charge self-energy, and the steepness of dielectric boundaries have not been easy to overcome [67,68]. The method has been used, with success, for estimating relative binding free energies of complexes with mutant

proteins [62<sup>•</sup>], and also for distinguishing native from nonnative complex conformations [63<sup>•</sup>]. An analysis of the relative binding free energies of unrelated protein–protein complexes, however, was less successful [63<sup>•</sup>] (for a discussion of the potential sources of this difficulty, see [49<sup>•</sup>]).

## Conclusions

We are currently witnessing the applications of simple thermodynamic potentials in off-lattice, near-real protein folding simulations. The simplifications that are used involve, in essence, the retention of only two meaningful physical effects: maximization of the number of hydrophobic interactions; and minimization of the number of buried polar atoms that do not participate in hydrogen bonds [27<sup>••</sup>,28,29<sup>••</sup>]. Although they are often included to assure proper polypeptide geometry, other terms (e.g. excluded volume, disulfide bridge), do not seem to affect the overall results of folding simulations that appear to be governed mainly by the trade-off between satisfying hydrogen bonding and affecting the complete hydrophobic collapse [15<sup>••</sup>,69].

The current methods of docking and binding use free-energy formulas that are based on more complex potentials. Most authors recognize that the major effects contributing to the specificity and strength of binding are solvation (including hydrophobicity, which implies an approximate surface-to-surface complementarity, minimizing solvent-exposed nonpolar surfaces, and long range electrostatic effects modified by dielectric boundaries) and, possibly, a loss of sidechain conformational entropy.

The balancing needs of computational speed and numerical accuracy are being approached in so many different ways, and the thermodynamical empirical short cuts are conceptually so complex, that to foretell the best avenues of future development is nearly impossible. In folding, the very simple potentials and low resolution models perform at least as well as the ones based on molecular mechanics that have been used in the past. Binding, by contrast, probably requires relatively precise solvation scales, that take a good account of long range effects. In addition, the most general treatment of binding would have to include energy terms associated with flexible (e.g. bond-angle and torsional) deformations due to induced fits. Thus, molecular mechanics, when amended by solvation terms and equipped with algorithms for estimating conformational entropy changes (e.g. CONGEN uniform conformational sampling [70,71]) may well remain an important tool in the analysis of receptor–ligand interactions.

## Acknowledgements

Sandor Vajda thanks Charles DeLisi, Temple Smith, Andrew Torda, and Adam Godzik for helpful discussions. The work of Sandor Vajda on this project has been supported by research grants from the Petroleum Research Fund, the National Science Foundation, and the Department of Energy.

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
  - of outstanding interest
1. Novotny J, Rashin AA, Brucoleri RE: **Criteria that discriminate between native proteins and incorrectly folded models.** *Proteins* 1988, 4:19–30.
  2. Shoichet BK, Kuntz ID: **Protein docking and complementarity.** *J Mol Biol* 1991, 221:327–346.
  3. Brady GP, Sharp KA: **Decomposition of interaction free energies in proteins and other complex systems.** *J Mol Biol* 1995, 17:77–85.
  4. Novotny J, Brucoleri RE, Saul FA: **On the attribution of binding energy in the antigen–antibody complexes McPC 603, D1.3 and HyHEL-5.** *Biochemistry* 1989 28:4735–4749.
  5. Vajda S, Weng Z, Rosenfeld R, DeLisi C: **Effect of conformational flexibility and solvation on receptor–ligand binding free energies.** *Biochemistry* 1994, 33:13977–13988.
  6. Honig B, Nicholls A: **Classical electrostatics in biology and chemistry.** *Science* 1995, 268:1144–1149.
  7. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thoma PD, Chan HS: **Principles of protein folding—a perspective from simple exact models.** *Protein Sci* 1995, 4:561–602.
  8. Sali A, Shakhovich 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.
  9. Miyazawa S, Jernigan R: **Residue–residue potentials with a favorable contact pair term and unfavorable high packing density term, for simulation and threading.** *J Mol Biol* 1996, 256:623–644.
- This paper shows that the re-evaluation of the classical Miyazawa–Jernigan potential, using a significantly larger set of protein crystal structures, leaves the interaction parameters almost unchanged. The authors make an important observation that the estimates of hydrophobicity from the contact energies for nonpolar sidechains agree well with the experimental values, in other words, the potential can be used to calculate hydrophobic contributions to the free energy. Whereas the original potential included only attractive terms, repulsive packing energy terms have now been added that are operative at high densities.
10. Chan HS, Dill KA: **Comparing folding codes for proteins and polymers.** *Proteins* 1996, 24:335–344.
- Simple potentials (folding codes) are studied on 2D lattice models. The properties of interests are degeneracy (the number of sequences that fold to a given number of structures) and encodability (the number of folds that are unique lowest energy structures of certain monomer sequences). This paper also emphasizes the role of the correlations in the folding code. Correlation in a lattice model means that all instances of an interaction between residue types *i* and *j* must have the same energy.
11. Unger R, Moulit J: **Local interactions dominate folding in a simple protein model.** *J Mol Biol* 1996, 259:988–994.
- This paper shows that the full characterization of simple potentials, including thermodynamic and kinetic consequences, is far from simple even on 3D lattice models. In particular, the question of foldability on a 3D lattice with a random pairwise-interaction potential is re-examined. One of the main results of Sali *et al.* [8] is that the sufficient and necessary condition for folding in this model is the existence of a pronounced energy minimum. Further simulations by Unger and Moulit in this paper reveal that this conclusion is true if, and only if, a specific folding temperature is selected for each individual protein. The second problem studied is the effect of local (i.e. between residues close in the sequence) versus nonlocal interactions. Sequences with many possible strong local interactions are shown to be easy to fold. This result also contradicts some conclusions that have appeared in the literature [12].
12. Abkevics VI, Gutin AM, Shakhovich EI: **Impact of local and non-local interactions on thermodynamics and kinetics of protein folding.** *J Mol Biol* 1995, 252:460–471.
  13. Godzik A, Kolinski A, Skolnick J: **Lattice representations of globular proteins: how good are they?** *J Comp Chem* 1993, 10:1194–1202.
  14. Reva BA, Sanner MF, Olson AJ, Finkelstein AV: **Lattice modeling: accuracy of energy calculations.** *J Comp Chem* 1996, 17:1025–1032.

This paper emphasizes that energy calculations based on lattice models are always approximate, because the distances between chain links, and hence

the interaction energies, are distorted. The distortions can be reduced by using a fine enough lattice, or by adjusting the lattice spacing.

15. Honig B, Cohen FE: **Adding backbone to protein folding: why proteins are polypeptides?** *Fold Des* 1996, 1:R17–R20.
- This review recalls the importance of the hydrogen-bonding structure, which may have been lost because the energetic contribution of hydrogen bonds is relatively minor. Burying polar groups, however, is strongly penalized unless they can form hydrogen bonds. Sidechain-only models fail to account for the penalty associated with the burial of backbone atoms, and hence have difficulty capturing essential features of protein folding.
16. Levitt M, Warshel A: **Computer simulations of protein folding.** *Nature* 1975, 253:694–698.
  17. Hagler AT, Honig B: **On the formation of protein tertiary structure.** *Proc Natl Acad Sci USA* 1978, 2:554–558.
  18. Kuntz ID, Crippen GM, Kollman PA, Kimelman D: **Calculation of protein tertiary structure.** *J Mol Biol* 1976, 106:983–994.
  19. Bowie JU, Eisenberg D: **An evolutionary approach to folding small alpha-helical proteins that uses sequence information and an empirical guiding fitness function.** *Proc Natl Acad Sci USA* 1994, 91:4436–4440.
  20. Monge A, Friesner RA, Honig B: **An algorithm to generate low-resolution tertiary structures from knowledge of secondary structure.** *Proc Natl Acad Sci USA* 1994, 91:5027–5029.
  21. Brasseur R: **Simulating the folding of small proteins by use of the local minimum energy and the free solvation energy yields native-like structures.** *J Mol Graph* 1995, 13:312–322.
- Simulations are based on a potential that includes a complete molecular mechanics energy function and two solvation terms: the first representing the effect of solvation on the interactions between solute atoms; the second representing the effect of solvation on the interactions between solute and solvent. These terms are very probably not independent, and thus the model includes more parameters than necessary. Its application to small proteins yields good results.
22. Wilson C, Doniach SA: **A computer model to dynamically simulate protein folding: studies with crambin.** *Proteins* 1989, 6:193–209.
  23. Koretke KK, Luthey-Schulten Z, Wolynes PG: **Self-consistently optimized statistical mechanical energy functions for sequence structure alignment.** *Protein Sci* 1996, 5:1043–1059.
  24. Dandekar T, Argos P: **Ab initio tertiary-fold prediction of helical and non-helical protein chains using a genetic algorithm.** *Int J Biol Macromol* 1996 18:14.
  25. Kolinski A, Skolnick J: **Monte-Carlo simulations of protein folding. 1. Lattice model and interaction scheme.** *Proteins* 1994, 18:338–352.
  26. Kolinski A, Skolnick J: **Monte-Carlo simulations of protein folding. 2. Application to protein-A, ROP, and crambin.** *Proteins* 1994, 18:353–366.
  27. Srinivasan R, Rose GD: **LINUS: a hierarchical procedure to predict the fold of a protein.** *Protein Sci* 1995, 22:81–99.
- This paper received tremendous publicity due to the reported progress in predicting the folds of protein fragments of up to 50 residues. A simplified geometry and a very simple potential have been used. The search is based on a MC-type method in which a single move alters the torsional angles of three consecutive residues. Whenever a chain segment adopts a conformation that is persistent in a number of moves, it is constrained to remain in the same conformation throughout the simulation.
28. Sun S, Thomas PD, Dill KA: **A simple protein folding algorithm using binary code and secondary structure constraints.** *Protein Eng* 1995, 8:769–778.
  29. Yue K, Dill KA: **Folding proteins with a simple energy function and extensive conformational searching.** *Protein Sci* 1996, 5:254–261.
- This paper extends the method of folding small proteins described in [28]. The potential includes only two terms, representing favorable interactions due to hydrophobic contacts and unfavorable interactions due to unsatisfied polar burials. Native-like structures of four small proteins are determined as a compromise between these two effects. The advantages of using folding models with very few parameters are discussed.
30. Lee B, Kurochikina N, Kang HS: **Protein folding by a biased Monte Carlo procedure in the dihedral angle space.** *FASEB J* 1996, 10:119–125.
- The search in the dihedral-angle space based on a potential that includes only hydrogen-bond and solvation contributions with various relative weights between them. Solvation is approximated by several methods, mainly by the Miyazawa–Jernigan potential. As in [29••], the balance between hydrophobic and hydrogen-bond contributions is emphasized. Application to the folding

of crambin yields, as one of the lowest energy structures, a conformation that has 3 Å rmsd from the crystal structure. The disulfide-bond information is not exploited in this example.

31. Sung SS: **Folding simulations of alanine-based peptides with lysin residues.** *Biophys J* 1995, **68**:826–834.
32. Chiche L, Gregoret LM, Cohen FE, Kollman PA: **Protein model structure evaluation using the solvation free energy of folding.** *Proc Natl Acad Sci USA* 1990, **87**:3240–3243.
33. Cardozo T, Totrov M, Abagyan R: **Homology modeling by the ICM method.** *Proteins* 1995, **23**:403–414.
34. Sun S: **A genetic algorithm that seeks native states of peptides and proteins.** *Biophys J* 1995, **69**:340–355.  
This simulation is based on simple polypeptide geometry and a potential that includes structure-based nonlocal terms, and a local term among nearest neighbor amino acids. The nonlocal terms are contact potentials representing backbone-to-backbone and sidechain-to-sidechain interactions. The local torsional term is extracted from the results of detailed molecular mechanics calculations. Applications include predicting the structure of crambin with 3 Å rmsd from the crystal structure.
35. Fraternali F, Van Gunsteren WF: **An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution.** *J Mol Biol* 1996, **256**:939–948.  
Two atomic solvation terms are added to the GROMOS potential, one to reward exposed polar area, the other to penalize exposed nonpolar area. The extension of the potential significantly improves the agreement with the properties obtained in molecular dynamics simulations with explicit water. In particular, the main artefacts typically encountered in *in vacuo* simulations are considerably reduced.
36. **Second meeting on the critical assessment of techniques for protein prediction on World Wide Web URL:** <http://iris4.carb.nist.gov/casp2>
37. Gilis D, Rooman M: **Stability changes upon mutation of solvent-accessible residues in proteins evaluated by database-derived potentials.** *J Mol Biol* 1996, **257**:1112–1126.  
This paper includes a general discussion of the formalism for deriving effective potentials from known structures.
38. Sippl MJ, Ortner M, Jaritz M, Lackner P, Floeckner HL: **Helmholtz free energy of atom pair interactions in proteins.** *Fold Des* 1996, **1**:289–298.
39. Sippl MJ: **Calculation of conformational ensembles from potentials of the main force.** *J Mol Biol* 1990, **213**:167–180.
40. Sippl MJ: **Helmholtz free energy of peptide hydrogen bonds in proteins.** *J Mol Biol* 1996, **220**:644–648.
41. Jernigan RL, Bahar I: **Structure-derived potentials and protein simulations.** *Curr Opin Struct Biol* 1996, **6**:195–209.
42. Thomas PD, Dill KA: **Statistical potentials extracted from protein structures: how accurate are they?** *J Mol Biol* 1996, **257**:457–469.  
Interaction energies are extracted from a set of conformations generated on lattices for short chains using simple contact potentials. Unlike in the case of real proteins, the underlying true potential is known for the generated lattice conformations and hence can be compared to the potential extracted. The significant differences found primarily stem from the use of randomized globules as the reference state, thereby neglecting the effects of excluded volume and chain connectivity. Since the burial of nonpolar groups is the dominant property of folded proteins, most of the information contained in statistical potentials with a random reference state is simply hydrophobic-clustering propensity, as has been observed previously [29•].
43. Thomas PD, Dill KA: **An iterative method for extracting energy-like quantities from protein structures.** *Proc Natl Acad Sci USA* 1996, **93**:11628–11633.
44. Zhang C, Vasmatzis G, Cornette JL, DeLisi C: **Determination of atomic desolvation energies from the structures of crystallised proteins.** *J Mol Biol* 1997, in press.  
The authors extend the Miyazawa–Jernigan potential to the atomic level, and determine contact energies for 18 different atom types. The free energies of solvating amino acid sidechains have been calculated using this method and are found to correlate to a very high degree ( $r=0.975$ ) with experimentally determined free energies of transferring the sidechains between octanol and water. Exploiting this property, the atomic contact energy has been used for the calculation of the solvation contribution to the binding free energy.
45. DeBolt SE, Skolnick J: **Evaluation of atomic level mean force potentials via inverse refinement of protein structures: atomic burial position and pairwise non-bonded interactions.** *Protein Eng* 1996, **8**:637–655.

Structure-based potentials are evaluated in terms of the problem of distinguishing native and slightly misfolded conformations. The ability of combining

knowledge-based and molecular mechanics terms are studied, and the use of such combined potentials in minimization is discussed.

46. Gschwend DA, Good AC, Kuntz ID: **Molecular docking toward drug design.** *J Mol Recog* 1996, **9**:175–186.  
A well written review of both rigid and flexible docking procedures, with some emphasis on the successful DOCK program. The problem of ranking docked complexes and the development of empirical target functions are discussed.
47. Norel R, Lin SL, Wolfson HJ, Nussinov R: **Molecular surface complementarity at protein-protein interfaces – the critical role played by surface normals at well placed, sparse points in docking.** *J Mol Biol* 1995, **252**:263–273.  
This paper describes the use of normals to the surface at specific points as shape descriptor for efficient rigid-body docking.
48. Vakser I: **Low-resolution docking: prediction of complexes for undetermined structures.** *Biopolymers* 1996, **39**:455–464.  
This paper extends the correlation technique introduced by the author. Increasing the range of interactions in a simple potential, defined in terms of attractive and repulsive step functions, is shown to remove the local minima in an exhaustive search.
49. Weng Z, Vajda S, DeLisi C: **Prediction of complexes using empirical free energy functions.** *Protein Sci* 1996, **5**:614–626.  
A partition-type binding free-energy evaluation model [5], which has an atomic parameter model of the solvation, is shown to discriminate native and nonnative docked conformations and, to yield binding free energies that are within 5% of the experimentally determined values. General aspects of the effective evaluation of binding free energy and the problems associated with the use of continuum electrostatics methods are discussed.
50. Ajay, Murcko MA: **Computational methods to predict binding free energy in ligand–receptor complexes.** *J Med Chem* 1995, **38**:4953–4967.  
A thoughtful review of current approaches to the evaluation of binding free energy including a classification of various methods is presented.
51. Horton N, Lewis M: **Calculation of the free energy of association for protein complexes.** *Protein Sci* 1992, **1**:169–181.
52. Nauchitel V, Villaverde MC, Sussman F: **Solvent accessibility as a predictive tool for the free energy of inhibitor binding to the HIV-1 protease.** *Protein Sci* 1995, **4**:1356–1364.
53. Viswanadhan VN, Reddy MR, Wlodawer A, Varney MD, Weinstein JN: **An approach to rapid estimation of relative binding affinities of enzyme inhibitors: application to peptidomimetic inhibitors of the human immunodeficiency virus type 1 protease.** *J Med Chem* 1996, **39**:705–712.
54. Wallqvist A, Jernigan RL, Covell DG: **A preference-based free energy parameterization of enzyme-inhibitor binding. Applications to HIV-1 protease inhibitor design.** *Protein Sci* 1995, **4**:1881–1903.  
The experimentally determined frequencies of pairwise atomic surface interactions have been used to derive atom–atom preferences of the receptor–ligand interface. The binding free energy is calculated by a linear combination of atomic contact areas, using parameters depending on the preference for the given atom pair. The expression includes three additional parameters, obtained from fitting the model to free-energy data. According to the authors, the method of converting the frequencies into preferences and then into free energy retains a rich variety of specific interactions, and unlike with the Miyazawa–Jernigan potential [9•] and similar methods [30•,44•], the potential is not dominated by the overall hydrophobic attraction.
55. Wallqvist A, Covell DG: **Docking enzyme–inhibitor complexes using a preference-based free energy surface.** *Proteins* 1996, **25**:403–419.  
A two-step docking scheme is presented, which uses a surface complementarity screen followed by a free energy based evaluation. The potential is a distance-scaled version of the structure-based free-energy function described in [51•].
56. DeWitte RS, Shakhovich EI: **SMoG: De novo design method based on simple, fast, an accurate free energy estimates. 1. Methodology and supporting evidence.** *J Am Chem Soc* 1996, **118**:11733–11744.
57. Bohm HJ: **The development of a simple empirical scoring function to estimate the binding constant for a protein ligand complex of known three-dimensional structure.** *J Comp Aid Mol Des* 1994, **8**:243–256.
58. Laskowski RA, Thornton JM, Humblet C, Singh J: **X-SITE – use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins.** *J Mol Biol* 1996, **259**:175–201.
59. Verkhivker G, Appelt K, Freer ST, Villafranca JE: **Empirical free energy calculations of ligand–protein crystallographic**

- complexes. 1. Knowledge-based ligand–protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng* 1995, 8:677–691.
60. Holloway MK, Wai JM, Halgren TA: *A priori* predictions of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J Med Chem* 1995, 38:305–317.
61. Head RD, Smyte ML, Oprea TI, Waller CL, Green SM, Marshall GR: VALIDATE: a new method for receptor-based prediction of binding affinities of novel ligands. *J Am Chem Soc* 1996, 118:3959–3969.
62. Zhang T, Koshland DE Jr: Computational method for relative binding energies of enzyme substrate complexes. *Protein Sci* 1996, 5:348–356.
- The continuum electrostatic approach has been applied to 63 pairs of nine mutant proteins with seven different substituted substrates. Only relative binding free energies are given for each complex using the complex of the natural protein and the unsubstituted substrate as the reference state. Thus, the results, which are in good agreement with the relative free energies observed, show the effects of substitutions in the substrate, but not the effects of mutations in the protein.
63. Jackson RM, Sternberg MJE: A continuum model for protein–protein interactions: application to the docking problem. *J Mol Biol* 1995, 250:258–275.
- The application of the continuum electrostatic method to the docked conformations of three protease–inhibitor complexes that were generated in [2]. The same conformations have been studied using a free-energy expression with an atomic solvation parameter model [49\*]. The continuum model distinguishes near-native from nonnative conformations very well but ranks the binding free energies of the three different complexes incorrectly.
64. Cummings MD, Hart TN, Read, RJ: Atomic solvation parameters in the analysis of protein–protein docking results. *Protein Sci* 1995, 4:2087–2099.
65. Juffer AH, Eisenhaber F, Hubbard SJ, Walther D, Argos P: Comparison of atomic solvation parametric sets: applicability and limitations in protein folding and binding. *Protein Sci* 1995, 4:2499–2509.
- A detailed account of the various atomic solvation parameter values that have been published is given.
66. Vajda S, Weng Z, DeLisi C: Extracting hydrophobicity parameters from solute partition and protein mutation/unfolding experiments. *Protein Eng* 1995, 8:1081–1092.
- A rigorous study of the free energy associated with the transfer of a non-polar group from water into a nonpolar environment. These free energies are shown to correlate with the solvent accessible surface, and their dependence on the volume of the solute is negligibly small. As in folded proteins, any hydrophobic group is surrounded by a largely nonpolar environment, the hydrophobicity parameter of  $30 \text{ cal mol}^{-1} \text{ \AA}^{-2}$  derived from hydrocarbon-to-water transfer, is a better choice than the  $16 \text{ cal mol}^{-1} \text{ \AA}^{-2}$  value based on octanol-to-water transfer.
67. Bruccoleri RE, Novotny J, Davis ME, Sharp KA: Finite difference Poisson–Boltzmann electrostatic calculations: increased accuracy achieved by harmonic dielectric smoothing and charge antialiasing. *J Comp Chem* 1997, 18:268–276.
68. Novotny J, Bruccoleri RE, Davis ME, Sharp KA: Empirical free energy calculations: a blind test and further improvements to the method. *J Mol Biol* 1997, in press.
69. Smith TF: Models of protein folding. *Science* 1995, 268:958–959.
70. Bruccoleri RE, Karplus M: Prediction of folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* 1987, 26:137–168.
71. Bruccoleri RE: Application of systematic conformational search to protein modeling. *Mol Simulat* 1993, 10:151–174.