Theor Chem Acc (2000) 104:1–6
DOI 10.1007/s002149900095

Theoretical
Chemistry Accounts

Overview

Ab initio polypeptide structure prediction

Philippe Derreumaux

Laboratoire de Biochimie Théorique, UPR CNRS 9080, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France

Received: 23 June 1999 / Accepted: 20 September 1999 / Published online: 15 December 1999
© Springer-Verlag 2000

Abstract. The biological activity of a polypeptide strongly depends on its 3D structure. Ab initio prediction of the native structure from the sequence of amino acids has long motivated the development of an optimum energy model such that interactions present in the native conformation are stronger than those present in nonnative conformations and of algorithms capable of finding the basin of lowest free energy among an astronomically large number of possible conformations. Despite recent progress in our understanding of the factors responsible for both polypeptide stability and formation, computer simulations of polypeptide models are still far from being practical software tools for biologists. In this work, state-of-the-art computer simulations aimed at ab initio structure prediction in aqueous solution are reviewed and their strengths and weaknesses are highlighted.

Key words: Peptide structure prediction – Chain representation – Effective potential – Global optimization approaches

1 Introduction

Predicting the 3D structure of a polypeptide is necessary for understanding its biological utility, in inhibiting its dysfunctions and in designing effective drugs. This is no simple task for computer simulations. As a result of the cooperative unfolding–folding (UF) mechanism, the bioactive native structure, which is commonly at the minimum free energy [1, 2], must be located among a number of possible conformations much larger than Avogadro's number. The second difficulty is to determine the minimum model that reproduces the essential structural and energetic features of polypeptides and allows an extensive search of conformational space. A variety of methods including molecular dynamics (MD) simulations, Monte-Carlo-based methods (MC), multicanonical MD and MC simulations and conformational search strategies have been applied to protein models; however, such calculations are still of limited structural use for biologists without additional experimental data. The factors responsible for both polypeptide formation and stability are nevertheless now better understood and computer efficiency has increased by several orders of magnitude. The goal of this review is to put ab initio approaches into the context of experimental work and to clarify a number of aspects that can bring theoretical predictions closer to experiment.

2 Stability and formation: insights from experimental data

Short peptides, fewer than 40 amino acids, have only marginally stable conformations in solution. Peptide α-helices and α-helical hairpins (two helices connected by a turn) do not show 100% helix content in circular dichroism (CD) measurements [3]. De novo designed β-hairpins (two β-strands connected by a turn), $\beta_2\alpha$ (two antiparallel β-strands packed against a helix) and β_3 (three antiparallel strands) peptides have been shown by NMR spectroscopy to adopt the correct folded structures in equilibrium with unfolded structures [4–6].

The rate of formation of several secondary structural elements of size and composition comparable with those found in proteins has been investigated by nanosecond-resolved kinetic methods. α-helices can form within about 100 ns, turns within about 500 ns and β-hairpins within about 1 μs [7]. No timescale for more complex topologies such as β_3 and $\beta\alpha\beta$ motifs has been reported as yet. The denaturation process of peptides has also been investigated by CD and fluorescence spectroscopy. The lack of strong cooperativity in the unfolding transition is common to all small peptides [3–6], but some cooperativity is detected. As an example, Schenck and Gellman [8] used a designed β_3 peptide to study antiparallel β-sheet formation.

Natural proteins are more stable than short peptides; however, they are rarely stabilized by more than 5–10 kcal/mol compared to their unfolded forms. Stability results from a complex balance between entropy, which favors unfolding, and enthalpic components, which favor folding, including salt bridges [9, 10], hydrogen

bonds, backbone–backbone interactions [11], disulfide bonds [12], propensities of amino acids to form α-helices and β-sheets [13, 14] and hydrophobic effects [9–15]. Other factors such as ion binding, pH and concentration contribute as well [16]. Direct experimental information on the magnitude of each of these various terms is not yet available.

Proteins can fold to their average native states much faster than previously thought (millisecond-to-second range) [17]. Reduced cytochrome c has been found to fold in 130 μs [18] and the helical λ repressor protein in only 20 μs [19]. Proteins exhibit a strong cooperative UF transition and the UF process can be described by either the two-state (denatured-native) model with a unique transition state or a multiple-state model with an ensemble of transition states having either weak or strong similarity to the native state [20]. Thus, Zaidi et al. [21] detected two pathways for unfolding the α/β protein barstar. One unfolding intermediate has nativelike secondary structure, the other has lost most of its secondary structure. Similarly, Grantchavora et al. [22] and Martinez et al. [23] found that the transition states of two β-sheet proteins were polarized versions of the native structure, i.e. nativelike in a specific region of the protein and largely unstructured elsewhere.

3 Off-lattice versus lattice protein models

Ideally one would like to calculate exactly the density of conformational states of a solvated polypeptide chain in order to clarify the effects of parameters (e.g. the temperature, the energy function) and of the amino acid sequence on the thermodynamic and kinetic properties. Such studies are feasible for highly simplified models with each amino acid represented by a bead. This is possible for 2D and 3D models restricted to moves on a regular lattice with a number of particles $np < 27$ and 15, respectively, because all states can be enumerated [24, 25], and for models with np up to 50 by combining sampling and histogram techniques [26]. (The interest in using lattice models is to limit the discrete number of conformational states.) Although such simulations with simplified chains have contributed to the design of potential-energy functions (see later) and the emergence of possible protein UF mechanisms [24–28], the lack of explicit side chains is an obstacle to reproducing the cooperative nature of the folding transition [29] and experimental structures.

As a compromise between structural precision and computational cost, two types of representations are employed according to the length, N, of the polypeptide chain. One type, used for proteins ($N > 40$), includes coarse-grained models, restricted or not to moves on a lattice, where each residue is represented by two beads: an α-carbon [30] or united peptide group bead for the main chain [31] and a bead for any side chain, with its van der Waals radius and position derived from an analysis of side chains in high-resolution protein structures. The solvent is considered implicitly. Another type of model, used for short peptides, comprises off-lattice detailed models where the main-chain atoms are treated explicitly and the side-chain representation ranges from a bead [32–34] to two beads [35, 36] or an all-atom representation (excluding nonpolar hydrogens) [37, 38]. The solvent is most often considered implicitly in such calculations.

All models do not lead to the same structural precision. Simple lattice models, based on cubic or face-centered cubic grids, generate low-resolution protein structures with little α-helix content. Lattice models of higher complexity, for instance with 56 or 90 orientations of the virtual α-carbon–α-carbon bonds, allow resolutions of 1.0 Å [30]. Off-lattice models fit protein structures to an average of 1.6 Å when using an optimized set of discrete main-chain conformations [39], but can otherwise reproduce experimental structures exactly.

4 Analytic form of the potential

The design of a satisfactory potential is critical to prediction. Such design requires that the lowest energy structures deviate little from the experimental ones and that these native structures are stable under physiological conditions [1]. Based on MC simulations [40] and analytic theory [41] with simple protein models, the procedure for designing an optimum potential has been clarified: maximize the quantity $Z = E_N - E_M/\delta$, where E_N and E_M are the average energies of the native and misfolded (nonnative) states and δ is the fluctuation of the energy of the misfolded states. By definition, misfolded states are taken to exclude states that correspond to fluctuations around the native state. This quantity was also found to promote fast folding of short peptide sequences [42].

In practice, there are two factors which limit the development of the analytic form of the energy function and the set of parameters for detailed protein models. One factor is related to uncertainties in the experimental structure of folded polypeptides. This uncertainty is much higher for short peptides than for proteins (Sect. 2). It is also important to recall that NMR-determined structures are time and space averages and that both polypeptides and proteins may contain flexible or disordered regions. The connection between these thermally populated states and the function of proteins is still not resolved in detail. Another limiting factor is the generation of a representative sample of nonnative, low-energy states with little structural resemblance to the native state, given the size of the conformational space to be sampled. This is no simple task since the misfolded states themselves depend on the energy function.

Different methods have been used to generate an ensemble of nonnative states:

1. Threading the sequences into a library of possible folds constructed from the structures in the protein data bank [43–45].
2. MD simulations at 500 K for about 1 ns starting from the native state [46, 47].
3. Search strategies with constraints on the native secondary elements [39, 48].
4. MC simulations at 700 K starting from fully extended structures [49].
5. Exhaustive searches with a restricted volume [50].

Some of these methods fail to generate a representative ensemble of competing low-energy unfolded forms. The threading method generally provides noncompact high-energy forms, deviating substantially from the observed ones. Methods 2 and 3 are not guaranteed to cover an adequate ensemble of nonnative states because of length restrictions (2), rigid-body approximations (3) or sampling from a single starting point (2 and 3). In contrast to these approaches which generate a significantly biased distribution of nonnative states, methods (4) and (5) can provide a representative ensemble of misfolded states, at least for short peptides.

As a result of differences in the ensemble of misfolded structures, the chain representation and the treatment of solvent effects, there is no universal analytic form for the potential [30–39, 44–57]. Rather, current potentials include a variety of all-atom molecular mechanics potentials combined with implicit solvent models [37, 51–54], knowledge-based functions (which derive all the parameters from a statistical analysis of known protein structures) [44, 47, 55], functions that combine knowledge-based terms and other components related to hydrogen-bond (H-bond) formation and side-chain positions [30–32] and, finally, simple functional forms based on H-bond and hydrophobic interactions [35, 50].

All these potentials perform differently with respect to the structure discrimination problem. Most potentials work for low-resolution fold recognition but not for high-resolution fold recognition or for ab initio folding, because misfolded forms with energies lower than the native form are often found in test cases [34, 39, 44, 45]. Their performances also differ according to the proteins studied [57]. Fold recognition involves the identification of the native form among a list of forms, while ab initio folding involves locating the native form starting from random conformations and thus is sensitive to the absolute value of the energy. However, recent progress in elaborating effective potentials has been achieved for a small number of proteins [53] and short peptides adopting elementary topologies in aqueous solution [49, 50]. Their performances remain to be evaluated on a larger subset of models.

5 Strengths and weaknesses of current computer simulations

Ab initio structure prediction. We have selected to review methods which build a structure for the target sequence without using specific template structures, a priori location of secondary structure elements or experimental distance restraints. The use of such information is described in Refs. [58–61].

5.1 Classical MD simulations

MD simulations, which are based on the integration of Newton's equations of motion, have a long history for the study of structural and dynamic properties of biomolecules in solution [62]. Their utility to simulate reversible UF processes of very short duration (nanosecond timescale) has been demonstrated for various β-peptides in methanol at room temperature [63]. Attempts to fold short peptides were also conducted in water at high temperature (e.g. Ref. [64]); however, since typical timescales covered by MD simulations (about 10 ns) are many orders of magnitude shorter than the experimental folding times (see Sect. 2), folding a small protein with an all-atom model in explicit solvent was not feasible until recently at room temperature.

Efficient MD simulations on 256 parallel processors (CRAY T3D) have recently opened new perspectives. Duan and Kollman [65] have presented a 1 μs trajectory for the villin headpiece subdomain, a 36-residue peptide consisting of three short helices held together by a loop, a turn and a hydrophobic core. Starting from a partly unfolded state (the turn region was in its native position), they found structures deviating by 5.0 Å root mean square from the NMR structure with 50% of the side-chain contacts detected by NMR. (The root mean square deviation between corresponding α-carbons is a commonly used measure of similarity between two structures). Although much longer computer runs are necessary for this small protein to reach its native state, several simulations still need to be carried out starting from unbiased conformations so as to have a clear picture of the dynamics and the average structure on the microsecond timescale.

5.2 MC-based methods

In many cases, it is more convenient to solve a set of coupled equations of motion by reducing it to a corresponding random process which can itself be simulated. This is the essence of the MC method. The Metropolis MC method, based on the idea of importance sampling and originally developed to simulate the equilibrium properties of fluids, is not efficient for structure prediction of detailed polypeptide models because the probability of visiting energy barriers (much greater than kT) is very small. The Metropolis criterion accepts a trial configuration if its energy is lower than the previous starting conformation ($\Delta E < 0$) or with a probability $\exp(-\Delta E/kT)$ if the energy increases [66].

In order to enhance the Metropolis acceptance ratio, i.e. the ratio of the number of accepted configurations to the total number of generated ones, several MC-based methods have been developed. They can be classified into (1) the methods that are not guaranteed to generate a Boltzmann-distributed (canonical) ensemble of conformations at temperature T and (2) those which are by bound the canonical ensemble generation requirement. (The multicanonical MC approach is described in the next section).

1. As a first step towards enhanced sampling of conformational space, Li and Scheraga [67] developed the MC-minimization (MCM) method. The trial configuration, generated by random changes of one residue at a time, is minimized before application of the Metropolis criterion. This technique has been useful in

locating the lowest-energy regions of the five-residue metenkephalin, but scales exponentially with the number of dihedral angles.

In order to accelerate the search efficiency, several variants of the MCM method have been proposed. They differ in the choice of the simulation temperature, T_m, and in the way of generating the trial configuration. Choices include guided moves from the reorientations of the permanent dipoles in the local field generated by the system [68], probability-biased moves from experimentally derived dihedral angle distributions for each residue and changes of one or several residues at a time [37, 38]. Such MCM methods, in which T_m is momentarily increased if a given number of steps has been rejected, have been shown to be useful in predicting helices [37, 68], β-hairpins [37, 38] and a $\beta_2\alpha$ motif [37] with 10–23 residues; however, their applicability to longer all-atom peptide fragments is still unclear. A number of authors have preferred to combine their MCM methods and other techniques. As an example, a combination of the MCM method and a potential-smoothing approach has been tested on the 10–55 fragment of staphylococcal protein A and on apo calbindin D9K using a two bead per residue representation. The concept behind the potential-smoothing approach is to reduce the ruggedness of the energy landscape. Nativelike structures and their mirror images were found. Given the simplicity of the chain representation, it is unfortunate that other proteins of similar chain lengths with more complex topologies than the three-helix and four-helix bundles have not been studied [31].

Recently, another variant of the MCM method, based on a kinetic requirement and diffusion-process-controlled moves, has been proposed [32, 49]. The idea in this case, is to search for trial conformations that are thermodynamically and kinetically accessible from the current conformation in a reasonable time. Applications of the combined use of this technique and the Optimized Potential for Efficient peptide-structure Prediction (OPEP) potential to 24 polypeptides with 7–38 residues have been reported. Using on average six beads per residue and starting from fully extended configurations, all OPEP-MC runs generate nativelike structures independently of the complexity of the fold (e.g. helices, β-hairpins and β_3, $\beta_2\alpha$, α-helical hairpin motifs) [49]. This approach is also found to predict the three-helix bundle structure of the 10–55 fragment of staphylococcal protein A and is currently being applied to small proteins.

Other MC-based approaches, without energy minimization, have also been described and tested. Pedersen and Moult [52] have explored the combination of the MC method and genetic algorithms for all-atom protein fragments up to 14 residues long. The idea of the genetic algorithm is to give better chances of survival and reproduction to the good individuals (i.e. the low-energy structures) within a population. Again, accurate structural prediction has been presented for most peptides studied.

In order to go beyond such a chain length, alternative approaches have been used. Simulated annealing, based on a MC procedure with T_m gradually decreasing during the simulation, has been applied to several peptides of increasing complexity (e.g. β-sheet [69] and α/β [34] structures). Although this method is often used, convergence to the global minimum is problematic because of its extreme sensitivity to the cooling procedure.

An alternative hierarchical MC approach developed by Srinivasan and Rose [35] is based on the idea that secondary structures form early in the folding process and subsequently self-assemble to form tertiary structures. This algorithm uses on average six beads per residue and generates trial conformations by perturbing three consecutive residues using four possible structures: α-helix, β-strand, turn and coil. It imposes hierarchy by constraining the conformational regions that form in the earliest steps of folding to persist throughout the simulation. Excellent results have been achieved for six proteins in terms of secondary and supersecondary structures, but the algorithm has failed in one case. Attempts to predict other protein structures have not been reported. This approach can be assumed to be limited by the fact that the sequential folding mechanism is known to be valid only for certain proteins.

Another hierarchical MC approach starts with a coarse lattice model and ends with a finer lattice model, using two beads per residue [30]. Applications to three proteins with 46–120 residues show accurate predictions for both three-helix and four-helix bundles, but inaccurate results for a 46-residue α/β structure [30] and for larger proteins with more complex folds [70].

2. A limited number of MC-based approaches are aimed at finding the global energy minimum while directly generating a canonical ensemble for detailed polypeptide models.

The so-called window MC algorithm changes the conformation by local jumps in a randomly chosen window involving a few residues of the polypeptide chain. Outside the window, the chain is fixed and the effect of these constraints on jumps within the window is taken into account. If this effect is neglected, the configurations will not be generated with the correct limiting distribution [33]. This method has been applied to a 26-residue peptide designed to adopt a helix–turn–helix conformation. The turn region is modeled by five glycine residues and the α-helical parts are modeled by alanine residues. Starting from β-strand or loop structures, four simulations find nativelike conformations, but two end in misfolded conformations (e.g. long helices). Since the impact of glycine residues on the search efficiency was not discussed, the real performance of this method remains to be evaluated for natural sequences and realistic potential-energy functions.

Another possibility involves the MC growth method which generates a Boltzmann ensemble of chains. In this method, the chains are grown atom by atom and are replicated or deleted according to Boltzmann statistics. The growth process can easily be modified to include experimentally derived dihedral angle distributions for each residue. This method has only been used to study either short peptides which have a strong preference for helical conformations or small loops in immunoglobulins [71].

Finally, the combination of the rigid element algorithm and the Metropolis MC procedure has been tested on short model peptides [54]. The algorithm keeps the

amide CONH elements rigid and rotates one or two amide elements at a time. The energy function has the property of lowering energy barriers. Two important questions are whether this procedure, which allows folding of helix and β-hairpin models at 274 K, works for more complex motifs and whether its energy function is not so smoothed that it could destabilize the native folded structure.

5.3 Multicanonical MC and MD simulations

The advantage of multicanonical MC or MD simulations is that any energy barrier can be crossed, while the canonical distribution at any temperature can be calculated from the multicanonical ensemble using reweighting techniques. As a result, the lowest-energy state will be detected and any thermodynamic property or any observable, such as the $^3J_{HN\alpha}$ spin–spin coupling constants of amide and α protons, can be checked against experimental data. The price which must be paid involves running multiple simulations until the energy histogram becomes flat. The reader is referred to Refs. [51, 72, 73] for algorithmic details and the applications of multicanonical MC and MD simulations to various all-atom peptides of limited chain lengths (fewer than 15 residues) in aqueous solution. Excellent agreement between prediction and experiment has been found concerning the thermodynamics of helix–coil transitions [72], the stability of two short peptides known experimentally to adopt an α-helix and a β-hairpin [51] and the position of one region of an antibody heavy chain [73]. It remains to be determined whether these simulations can be applied to peptides with 20–40 amino acids, especially with an accurate implicit solvent model [51]. Such studies would provide very useful insights into the thermodynamics underlying the most elementary tertiary motifs.

5.4 Conformational search strategies

While early work in this field coupled build-up procedures and energy minimization, recent work uses constraint-based conformational searching. In the build-up procedure, starting conformations are constructed by combining minima of conformational building blocks (e.g. tripeptides). This strategy has performed rather well on the 36-residue avian pancreatic polypeptide, but is not guaranteed to find the global minimum. Furthermore, the number of starting conformations is still exponential in the number of building blocks [74]. In the constraint-based Geocore method, the polypeptide is grown residue by residue allowing four ϕ,ψ choices for every residue and all conformations with a near-maximum number of nonpolar contacts are constructed [50]. As an example, 3×10^8 conformations are calculated for a 17-residue peptide model. Although, this method finds nativelike conformations for 15 different peptides containing 17–30 residues using an extremely simple energy function, there is evidence that a larger set of ϕ,ψ choices improves the quality of the prediction and other constraints limiting the conformational search will need to be used in studies of longer chain lengths [50].

6 Future directions

This summary of the strengths and limitations of current computer simulations aimed at ab initio structure prediction for single-chain polypeptides in aqueous solution shows that while many methods work for a few amino acid sequences, promises of reliable structural predictions have not yet materialized for a wider variety of sequences adopting both simple and complex topologies.

These methods share the use of an energy function which attempts to model the various energetic components contributing to both polypeptide stability and specificity. It is well established that the key factor for structure prediction is thermodynamics (existence of a stability gap between the native and misfolded forms). While energy errors are unavoidable [75], current research is devoted to the development of energy functions including energetic subtleties that specify well-defined structures at a detailed level. Certainly, many factors are likely to affect the stability gap and the details of native structures. Factors include multibody interactions, position-dependent effects on amino acid structural propensities [76, 77], side-chain–side-chain interactions [78] and the balance between backbone and long-range side, chain terms [11, 79, 80]. However, it also remains to be determined whether our inability to make native forms more stable than misfolded forms results from the chain representation, the energy function or a combination of both factors.

In order to address this issue, it is necessary to study small monomeric proteins with 20–40 residues which fold to simple topologies (without disulfide bonds, ion binding, metal ligation and nonnatural amino acids) and, furthermore, which exhibit a cooperative two-state UF transition, which is the hallmark of larger proteins. Such a study requires combined effort in three fields:

1. Protein engineering. There is a need for de novo sequences designed to adopt stable elementary topologies such as $\alpha\beta_2$, $\alpha_2\beta$, α_3, $\beta\alpha\beta$, $\alpha\beta\alpha$, $\beta\alpha_2$, $\beta_2\alpha$, β_3 motifs. The β_3 fold has already been designed, but is only 80% folded at low temperature [5]. Determination of the stabilization free energy as a function of point mutations is also desirable. An essential aspect of this library of sequences is that all 20 amino acids must have a high probability of occurrence.

2. Multidimensional NMR studies. Structural characterization of the unfolded forms in equilibrium with the folded forms is a major problem. Unfolded forms are never described because the sequence-specific assignment of resonances, other than the backbone ^{15}N and $^{13}C(O)$ signals, is difficult; however, these less structured forms are essential for checking the theoretical energy landscape.

3. Molecular modeling. For relatively small proteins, extensive sampling of conformational space is tractable on fast computers and determination of the chain representations and energy functions allowing discrimination of the native forms with their correct Boltzmann weights should be feasible.

Acknowledgements. I thank Richard Lavery for his helpful comments on the manuscript.

References

1. Anfinsen CB (1973) Science 181: 223
2. Sohl JL, Jaswai SS, Agard DA (1998) Nature 395: 817
3. Fezoui Y, Connolly PJ, Osterhout JJ (1997) Protein Sci 6: 1869
4. Dahiyat BI, Mayo SL (1997) Science 278: 82
5. Kortemme T, Ramirez-Alvarado M, Serrano L (1998) Science 281: 253
6. de Alba E, Santoro J, Rico M, Jimenez MA (1999) Protein Sci 8: 854
7. Eaton WA, Munoz V, Thompson PA, Chan C-K, Hofrichter J (1997) Curr Opin Struct Biol 7: 10
8. Schenck HL, Gellman SH (1998) J Am Chem Soc 120: 4869
9. Bryson JW, Desjarlais JR, Handel TM, DeGrado WF (1998) Protein Sci 7: 1414
10. Lumb KJ, Kim PS (1996) Science 271: 1137
11. Zhang C, Cornette JL, Delisi C (1997) Protein Sci 6: 1057
12. Price-Carter M, Hull MS, Goldenberg DP (1998) Biochemistry 37: 9851
13. Malakauskas SM, Mayo SL (1998) Nat Struct Biol 5: 470
14. Reddy BVB, Datta S, Tiwari S (1998) Protein Eng 11: 1137
15. Takano K, Yamagata Y, Yutani K (1998) J Mol Biol 280: 749
16. (a) Kohn WD, Kay CM, Hodges RS (1998) J Peptide Res 51: 9; (b) Kaushik JK, Bhat R (1999) Protein Sci 8: 222
17. Jackson SE, Fersht AR (1991) Biochemistry 30: 10428
18. Pascher T, Chesick JP, Winkler JR, Gray HB (1996) Science 271: 1558
19. Burton RE, Huang GS, Daugherty MA, Calderone TL, Oas TG (1997) Nat Struct Biol 4: 305
20. Baldwin RL (1994) Nature 369: 183
21. Zaidi FN, Nath U, Udgaonkar JB (1997) Nat Struct Biol 4: 1016
22. Grantcharova VP, Riddle DS, Santiago JV, Baker D (1998) Nat Struct Biol 5: 714
23. Martinez JC, Pisabarro MT, Serrano L (1998) Nat Struct Biol 5: 721
24. Shakhnovich EI (1997) Curr Opin Struct Biol 7: 29
25. Chan HS, Dill KA (1998) Proteins 30: 2
26. Nymeyer H, Garcia AE, Onuchic JN (1998) Proc Natl Acad Sci USA 95: 5921
27. Wolynes PG, Onuchic JN, Thirumalai D (1996) Science 267: 1619
28. Dobson C, Karplus M (1999) Curr Opin Struct Biol 9: 92
29. (a) Klimov DK, Thirumalai D (1998) Fold Des 3: 127; (b) Hao M-H, Scheraga HA (1998) J Mol Biol 277: 973
30. Kolinsli A, Skolnick J (1994) Proteins 18: 338
31. (a) Liwo A, Pillardy J, Kazmierkiewicz R, Wawak RJ, Groth M, Czaplewski C, Oldziej S, Scheraga HA (1999) Theor Chem Acc 1–3: 16; (b) Lee J, Liwo A, Scheraga HA (1999) Proc Natl Acad Sci USA 96: 2025
32. (a) Derreumaux P (1998) J Chem Phys 109: 1567; (b) Derreumaux P (1997) J Chem Phys 107: 1941
33. Hoffmann D, Knapp E-W (1997) J Phys Chem B 101: 6734
34. Wilson C, Doniach S (1989) Proteins 6: 193
35. Srinivasan R, Rose GD (1995) Proteins 22: 81
36. Wallqvist A, Ullner M (1994) Proteins 18: 267
37. Abagyan R, Totrov M (1999) J Comput Phys 151: 402
38. Evans JS, Mathiowetz AM, Chan SI, Goddard WA III (1995) Protein Sci 4: 1203
39. Park B, Levitt M (1996) J Mol Biol 258: 367
40. Go N, Abe H (1981) Biopolymers 20: 991
41. Goldstein R, Luthey-Schulten ZA, Wolynes P (1992) Proc Natl Acad Sci USA 89: 4918
42. Sali A, Shakhnovich E, Karplus M (1994) Nature 369: 248
43. Jones DT, Taylor WR, Thornton JM (1992) Nature 358: 86
44. Miyazawa S, Jernigan RL (1996) J Mol Biol 256: 623
45. Mirny LA, Shakhnovich EI (1996) J Mol Biol 264: 1164
46. Wang Y, Zhang H, Li W, Scott RA (1995) Proc Natl Acad Sci USA 92: 709
47. Huang ES, Subbiah S, Tsai J, Levitt M (1996) J Mol Biol 257: 716
48. Ozkan B, Bahar I (1998) Proteins 32: 211
49. Derreumaux P (1999) J Chem Phys 111: 2301
50. (a) Yue K, Dill KA (1996) Protein Sci 5: 224; (b) Ishikawa K, Yue K, Dill KA (1999) Protein Sci 8: 716
51. Schaefer M, Bartels C, Karplus M (1998) J Mol Biol 284: 835
52. Pedersen JT, Moult J (1997) J Mol Biol 269: 240
53. Lazaridis T, Karplus M (1999) Proteins 35: 133
54. Sung S-S (1999) Biophys J 66: 1796
55. Sippl MJ (1990) J Mol Biol 213: 859
56. Maiorov VN, Crippen GM (1992) J Mol Biol 227: 876
57. Rooman M, Gilis D (1998) Eur J Biochem 254: 135
58. Ortiz AR, Kolinski A, Skolnick J (1998) Proc Natl Acad Sci USA 95: 1020
59. Sanchez R, Sali A (1998) Proc Natl Acad Sci USA 95: 13957
60. Dandekar T, Argos P (1996) J Mol Biol 236: 844
61. Standley DM, Gunn JR, Friesner RA, McDermott AE (1998) Proteins 33: 240
62. Cheatham TE III, Brooks BR (1998) Theor Chem Acc 99: 279
63. Daura X, van Gunsteren WF, Mark AE (1999) Proteins 34: 269
64. Prevost M, Ortmans I (1997) Proteins 29: 212
65. Duan Y, Kollman PA (1998) Science 282: 740
66. Metropolis NS, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) J Chem Phys 21: 1087
67. Li Z, Scheraga HA (1987) Proc Natl Acad Sci USA 84: 6611
68. Ripoll DR, Scheraga HA (1988) Biopolymers 27: 1283
69. Okamoto Y, Masuya M, Nabeshima M, Nakazawa T (1999) Chem Phys Lett 299: 17
70. Skolnick J, Kolinski A, Ortiz (1997) J Mol Biol 265: 217
71. Bascle J, Garel T, Orland H, Velikson B (1993) Biopolymers 33: 1843
72. Okamoto Y, Hansmann UHE (1995) J Phys Chem 99: 11276
73. Shirai H, Nakajima N, Higo J, Kidera A, Nakamura H (1998) JMol Biol 278: 481
74. Lambert MH, Scheraga HA (1989) J Comput Chem 10: 817
75. Finkelstein AV (1997) Curr Opin Struct Biol 7: 60
76. Minor DL Jr, Kim PS (1994) Nature 371: 264
77. Petukhov M, Munoz V, Yumoto N, Yoshikawa S, Serrano L (1998) J Mol Biol 278: 279
78. Thomas PD, Dill KA (1996) J Mol Biol 257: 457
79. Honig B, Cohen FE (1996) Fold Des 1: 17
80. Koretke KK, Luthey-Schulten Z, Wolynes PG (1998) Proc Natl Acad Sci USA 95: 2932