Volume 325, number 1,2, 17–22 FEBS 12543 June 1993

Minireview

Towards protein folding by global energy optimization

Ruben A. Abagyan

European Molecular Biology Laboratory, Meyerhofstraße 1, 6900 Heidelberg, Germany

Received 18 April 1993

Different components of the theoretical protein folding problem are evaluated critically. It is argued that: (i) as a rule, small- and medium-sized proteins are in the free energy minimum; (ii) long-living metastable states may either appear occasionally with growing protein size, or be selected by evolution for a specific function; (iii) functions discriminating against incorrect folds would fail if they were used directly in the global optimization, unless they approximate the true free energy accurately; (iv) surface and electrostatic free energies should be treated separately; (v) conformational entropy (of side chains in particular) should be taken into account; (vi) Monte Carlo procedures considering all free energy terms and combining global knowledge-based random moves with local optimization have the largest potential for success.

Protein folding; Global optimization; Free energy; Electrostatics; Solvation; Entropy

1. THE THEORETICAL PROTEIN FOLDING PROBLEM

Prediction of the native three-dimensional structure of a protein from the amino acid sequence remains an unsolved problem despite numerous efforts to solve it for more than a quarter of a century. A physical approach to the problem in its pure form is based on the assumption that the native conformation corresponds to the structure with the lowest free energy and is thus in a state of thermodynamic equilibrium. This view is backed by in vitro observations that many proteins can spontaneously and successfully refold from a variety of denatured states [1–3]. The biological factors of in vivo folding, such as peptidyl-prolyl-isomerase, protein disulphide isomerase, molecular chaperonins and translocation control, clearly influence the kinetics of protein folding, assembly and transport [4,5] by reducing energy barriers and protecting intermediates from aggregation, however, these facts do not invalidate the hypothesis that the native conformation corresponds to the global free energy minimum.

The alternative assumption (backed by Levinthal's argument [6]) is that the native state is the lowest kinetically accessible free energy minimum which is separated from the true global minimum by a large kinetic barrier (more than 25–30 kcal/mol). The first experimental evidence supporting this view has been provided recently [7–9]. α-Lytic protease was shown to have two conformational forms, active and inactive [7]. The protein needs a catalyst, which is normally covalently attached to it, to bypass a kinetic barrier of more than about 27 kcal/mol separating the intermediate metastable conformation from its stable active form. It was also suggested that β-sheet rearrangements in serpins (serine protease inhibitors) imply the existence of a metastable kinetically trapped five-stranded β-sheet conformation which can be rearranged slowly to the thermodynamically stable six-stranded conformation [8,10].

These interesting observations raise two questions: (i) could a metastable state have all the features of the normal protein? and if it could, (ii) is this behaviour typical for all proteins? So far there is no clear experimental evidence for a positive answer to the first question; absence of tertiary interactions and the expanded radius of the inactive form of α-lytic protease were indicative of the 'molten globule' [11] rather than of the normal protein. The answer to the second question is definitely negative. We know that small, single-domain proteins do fold into their lowest energy state. Maybe the strongest argument in favour of the thermodynamic hypothesis is an *evolutionary* one. Indeed, protein sequences have evolved under pressure to perform certain functions, which for most known functions requires a stable, unique and compact structure. It seems that there was *no* evolutionary pressure forcing all proteins to hide their global minimum behind a kinetic barrier, unless this were needed for a specific function. Therefore we may conclude that although the stable native-like kinetic intermediates, or at least kinetic blocks, may occur, they should be either specially discovered by evolution only for certain functions or else be random

Correspondence address: R.A. Abagyan, European Molecular Biology Laboratory, Meyerhofstraße 1, 6900 Heidelberg, Germany.

(with the probability growing with the protein size). Having backed the theoretical formulation of the protein folding problem as a search for the free energy minimum, we must immediately admit that three decades of effort have not yet solved the problem.

2. WHAT DO WE NEED TO FOLD A PROTEIN?

Two major components are needed to solve the problem theoretically. First, we should be capable of calculating all essential terms of the free energy of a trial protein conformation with an accuracy sufficient to ensure the uniqueness of the native conformation. Second, we need a procedure to locate the global minimum of this energy function in the giant space of conformational possibilities using a limited number of function evaluations. Unfortunately, even a quick look at the state of both sides of the problem is enough to understand how much further work is required.

The free energy of a protein consists of a potential energy in vacuo (which can be evaluated using empirical potentials e.g. [12–14]), the free energy of solvation and a term proportional to the conformational entropy of the polypeptide. The solvation energy consists in turn of the electrostatic free energy [15–17] and 'surface free energy' related to hydrophobicity [18]. Whilst the vacuum potential energy and its derivatives may be computed relatively easily, efficient and accurate evaluation of the solvation energy and the entropic term still poses a serious problem. This has led to the opinion that 'the true Hamiltonian (energy) of a protein is so unimaginably complicated that we have neither means nor hope of determining it properly' [19] as well as some attempts to obviate the difficulties.

3. DISCRIMINATING RIGHT FROM WRONG OR PLATO'S MAN

When Plato gave a popular definition, 'Man is a two-legged featherless animal', Diogenis plucked a cock, took it to his school and proclaimed: 'This is Plato's man!'. After this the definition was extended by 'with wide fingernails' (Diogenis Laertis 'Life, Teaching and Apophthegms of Famous Philosophers'). The obvious incapacity of the potential energy in vacuo to replace the true free energy in structure prediction calculations led to an attempt to design functions (or extra definitions) discriminating between the native conformation and incorrect models [20–30]. Typically these functions are either directly related to the missing solvation free energy term [20,21,31] or are derived from statistics of atomic (residue) contacts [22–27] or accessibilites [29] in the database of known three-dimensional structures.

Representing only a part of the free energy (usually with rather low accuracy), however, does not guarantee (as in the Diogenis story) that no false conformations fitting the function may be generated, simply because it is not the true free energy. The following pattern seems to be typical. Suppose we generate a set of misfolded conformations by optimizing only one part (E_{part}) of the free energy; for example, the van der Waals energy plus hydrogen bonding energy. Not surprisingly, the native conformation is not recognized since some of the misfolded conformations have comparable or lower energies, E_{part}. Now let us build a function F_{discr} which represents one or several energy terms omitted in E_{part}, for example, the cubed difference between the non-polar and the polar accessible surface (it is sufficient if F_{discr} correlates only approximately with the omitted term(s)). It is possible that this 'wide finger-nails' function patching the initially incomplete definition, E_{part}, will be capable of discriminating between the misfolded and the native conformations, however, two unpleasant things may happen. First, there may be some obvious exceptions: e.g. the solvation energy from [32] taken as F_{discr} will certainly prefer a completely unfolded conformation over the native one (one may argue here that it may be easily patched by adding '.. and compact'). Second and more seriously, if we were to *generate* conformations with optimal $E = E_{part} + F_{discr}$ rather than only E_{part} using a global optimization procedure, most probably many 'false positives' will be found.

Any hope of finding a filter which can be applied to a limited set of conformations at the very end of the procedure globally optimizing the incomplete function E_{part} is based on the belief that this set is small. In reality an enormous number of conformations with E_{part} energies comparable with the E_{part}(native) can be found even for medium-sized peptides if essential parts of the free energy are omitted (Fig. 1). Therefore a new energy term or discriminating function should be tested by subjecting it to the global optimization procedure and proving that no conformation better than the native may be generated. It seems that the discriminating functions will eventually converge to the true free energy.

4. SIMPLIFIED MODELS

If we are to simplify the system and the free energy function, how should we do it and how far can we go without losing the ability to predict the protein three-dimensional structure from its sequence? To answer these questions let us briefly list the two main classes of simplified protein models [33].

Lattice models are the most schematic models of the protein structure [33–36]. In these models simple functions of the nearest neighbours, along with some constraints (or filters), represent the free energy. A major advantage of lattice models is computational efficiency, which in some cases allows us to find the globally optimal distribution on a lattice of the chain of nodes representing the polypeptide chain [37,38]. Two-dimensional and a variety of three-dimensional lattices have been used. Usually this level of simplification does not allow

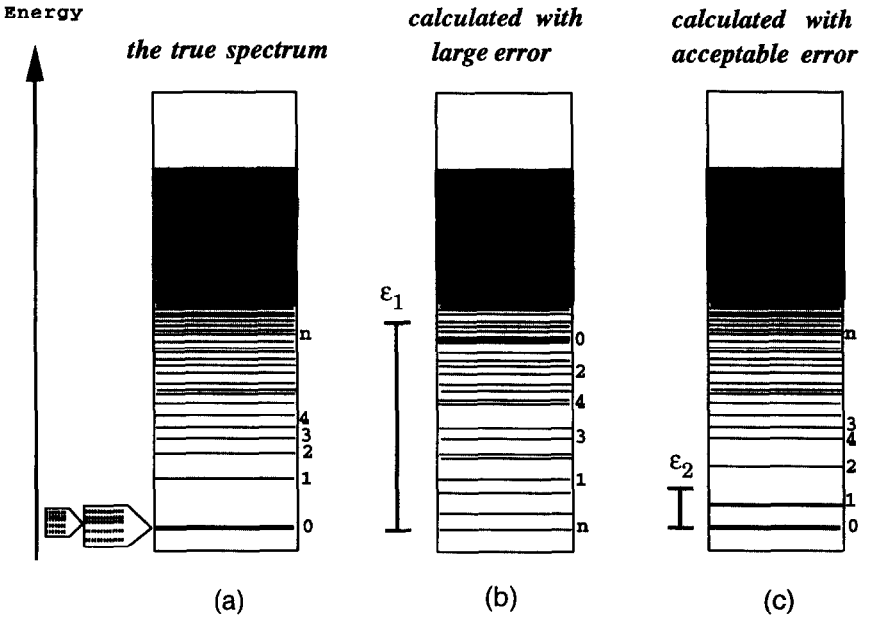

Fig. 1. Consequences of the energy inaccuracies on the protein structure prediction (folding, docking, etc.). Lines in these schematic energy spectra represent different conformations (in reality each line is a series of lines corresponding to the local conformational substates organized hierarchically [51]). The lower part of the true distribution (a) may be approximated by a gaussian because of the complexity and quasi-random nature of the individual contributions bounded below. If the energy difference between two conformations is evaluated with an error, ε_1 (b), one should accumulate all conformations in the range from the lowest energy E_0 to $E_0+\varepsilon_1$ in order to catch the native one among the false positives. The number of false positives, n, however, grows *exponentially* as the relative error, ε_1 increases. If the relative error is less than $\varepsilon_2 \sim E_1 - E_0$ (c) the global minimum can be recognized.

modelling of a particular protein with its particular sequence, although the most elaborate methods are aiming at this goal [35,38]. General questions about principles of protein structure, importance of different types of energy contributions, and folding kinetics, however, may be addressed [36,39–41].

Simplified models in a continuous space present a next step towards a more realistic description of a polypeptide chain [42,43]. Typically, in these models a residue is represented by one or two interaction points, and either the backbone torsion angles [42,44], or Cartesian coordinates [28,43,45].

The reduced representation logically leads to the idea of a mean inter-residue potential. Deriving a reliable mean potential by actual averaging of a detailed potential over an appropriate ensemble [42] is rather difficult. Neither the ensemble nor an adequate interaction energy can be calculated easily. Instead, the mean potentials from database statistics have become increasingly popular [22,27,28,46–49]. Simplified models may be subjected to a global search procedure such as minimization from random starting conformations, molecular dynamics, Monte Carlo, simulated annealing or combinations of these [42,44,45,50].

Two arguments against simplified models may be put forward. First, the relative error in the free energy evaluation strongly influences the number of false positives (Fig. 1). Therefore, if the accuracy of simplified models appears to be insufficient for the unambiguous prediction of the native conformation (which seems to be the case [44]), then, most probably, the number of false positives will be exceedingly large. Second, even if the native conformation and the predicted simplified model do share some common features, the further use of the approximate model seems rather problematic. A procedure to search around the initial model in order to reach and recognize a native conformation 3–4 Å C^{α} RMSD away, does not exist yet and may be nearly as complicated and computationally intensive as a detailed ab initio protein folding algorithm, which hence remains the main objective.

5. SOLVATION ENERGY

Over the years the idea that residues pack densely and uniquely like pieces of a three-dimensional jigsaw puzzle [52,53] has been strongly exaggerated. This has led to another delusion that the standard empirical potentials (which to some extent reflect the protein compactness) should be sufficient for protein structure prediction (e.g. [50]). Analysis of deliberately misfolded protein models [20], however, reveals the weakness of this assumption. It has become clear that additional energy terms (such as solvation) are necessary to model large conformational changes. Since introduction of explicit water molecules is far too demanding computationally, semi-em-

pirical approaches, in which solvent is treated as a continuous medium [18,21,31,32,54,55], are preferred. These kinds of terms can be calculated efficiently [56–58], differentiated analytically [56,59] and used in molecular simulations [32,60,61].

The solvation free energy consists of two major components. The first one represents cavity formation, solute–solvent dispersion interactions and solvent structure changes [55], and is linearly related to solvent-accessible surface areas. The second component is an electrostatic polarization energy which is a complex and computationally expensive function of the protein shape and positions of the charges [17]. This complexity created a temptation to mix both components into one function of atomic accessible surfaces and empirical surface parameters [21,31,32]. Indeed, (i) it is computationally simple and (ii) most of the fully charged atoms (in Lys,Arg,Glu and Asp) are exposed anyway and hence at least their 'self-energy' may be represented by atomic accessibilites. Ignoring contributions from partial charges buried in the protein interior [62] and omitting the cross-terms of the electrostatic polarization energy, however, are dangerous and may introduce a significant error in the calculations.

Published parameter sets are rather inconsistent [21,31,32]. For example, a set of atomic solvation parameters derived from the octanol-to-water transfer energies for C, N/O, O^-, N^+, and S was 16, −6, −24, −50 and 21 cal/Å^2·mol [21], whereas the vapour-to-water transfer energies [32] led to very different parameters: 12, −116, −175, −186 and −18, respectively. Which data should be used to derive parameters for the energy calculations? Mixing hydrophobicity and electrostatics creates another conundrum: while the standard empirical force potentials correspond to the energy in vacuo and hence data on vacuum (or vapour)-to-water transfer would be pertinent for the derivation of solvation parameters [32], the electrostatic polarization energy should rather be evaluated from the protein interior-like medium-to-water transfer data since in the surface method the charge gets *zero* solvation energy once it is buried.

The separate treatment of surface energy and electrostatics has been successful for small organic molecules [18,55] where electrostatic polarization energy can be evaluated relatively easily. For peptides and protein, however, the separate treatment terms will displace the mixed approaches [31,32] only when electrostatic polarization calculations are actually introduced in the molecular simulations and the surface solvation parameters are re-evaluated appropriately. Such an attempt has been undertaken recently [61].

6. ELECTROSTATICS

Methods to calculate the electrostatic free energy of biological molecules in solution have been described in a number of excellent reviews (e.g. [15–17,63]). The methods cover a wide range of possible applications. However, simulations of large structural changes impose severe time limitations on the energy evaluation. The fastest way to calculate the electrostatic energy, which is more reasonable than the Coulomb energy but is still not physically justified, is to use the Coulomb formula with the distance-dependent dielectric constant $\varepsilon(r) = 4.5\,r$ [64] or a more complex function [65]. Unfortunately, all effective dielectric models lack the self-energy of the charge, which is clearly an important component of the energy. The next level methods, still fast and analytical, use either image charges with planar boundaries [66] or more accurate spherical approximation [67–69]. The spherical solution gives an infinite series, so further approximations are necessary. In the spherical image method only the first term of the series is retained [68], but omitting higher terms leads to noticeable inaccuracies. Recently an analytical correction term complementing the spherical image energy was derived [61]. The modified image method is fast and has been included in a Monte Carlo simulation of the folding of a 12-residue peptide.

One problem of the electrostatic calculations is their strong dependence on the radii of the charges and the dielectric constant inside the protein. Optimization of these parameters may be carried out together with empirical solvation parameters (see previous section) using experimental free energies of solvation.

7. ENTROPIC TERM SHOULD BE INCLUDED IN THE SIMULATION

What is better energetically: to expose the lysine side chain completely, or to bury its hydrophobic part and expose only the charged NH_3 group? This kind of question is being addressed by the Monte Carlo or molecular dynamics program many times during a simulation and, obviously, entropy is vital to make the right decision. Actually, at room temperature the enthalpic and entropic changes upon exposure of the side chain are well balanced, e.g. an exposure of one CH_2 group of a chain costs about 0.88 kcal/mol [70], which is comparable with the entropic gain of ca. $-R\ln 3 \approx -0.66$ kcal/mol, resulting from three additional rotation states in the exposed group. The importance of the entropy has been recognized in a number of publications [19,71–74], however, a practical method to evaluate the conformational entropy remains to be developed.

Because of the close packing, the variation of the main chain entropies among different folded states seems to be less than the variation of side chain entropies. By determining probabilities (P_i) of different rotamers (i) of the side chain in the folded and denatured states, one can calculate the entropy difference as:

$$\Delta S = S_U - S_F$$

where S is given by the Boltzmann formula:

$$S = -R\Sigma P_i \ln(P_i)$$

Probabilities may be determined either from Monte Carlo or molecular dynamics simulation [74] or from statistical distributions [75]. An attempt to incorporate the calculations of entropy in the global optimization procedure was made in [61]. In this work the number of states contributing to the entropic term was related to the solvent-accessible areas of the side chain atoms. More accurate and computationally efficient evaluations of side chain entropies may also be possible.

8. WHAT PROCEDURE WILL FOLD THE PROTEIN?

The biggest challenge, however, is to find the global minimum of the energy function which is of unprecedented dimensionality [76,77]. Integrating equations of motion in either Cartesian coordinate space (e.g. [78,79]) or in torsion angle space [80], and thus following one of the possible dynamic trajectories until the minimum is reached, would be ideal, however, there are two arguments against it. First, the time step of integration is far too small to calculate a trajectory of sufficient length. Some increase is possible but there seem to be important limiting factors [80]. Second, there is no need to follow a *continuous* dynamic trajectory if we believe that the native conformation is in the free energy minimum and is thus trajectory independent.

Build-up methods combining the low-energy conformations of polypeptide segments [81–83] rely on an assumption about predominance of the local interactions. They are very attractive since they bypass the problem of exponential growth of conformational space with the number of residues, however, the non-local interactions are not that small. For, say, a 5-residue fragment they may easily contribute 10–15 kcal/mol, which necessitates storage of the enormous number of pentapeptide conformations for the further build-up (Fig. 1). It is also well known that pentapeptides with identical sequence may have many kinds of local conformations, ranging from the all-helical to the extended ones [84].

Deterministic methods of global optimization smoothing either the energy surface (diffusion–equation method [85,86]) or the probability distribution represented by gaussians (packet annealing [87]) have been developed and applied to peptides. They are promising but there are two difficulties. Firstly, being dependent on the expansion in terms of gaussians, these methods can not currently deal with complex functions like solvation or electrostatic free energy terms. Secondly, in the process of annealing, the objective function is not the same as the energy, so one has to keep track of many minima to guarantee that the future global one is not lost; however, the argument presented on Fig. 1 implies that this number should be exceedingly large.

9. A SMART RANDOM MOVE IS WANTED

The family of Monte Carlo methods can be subdivided into local- and global-step methods. The former methods [88,89] follow some approximation of the energy surface in the vicinity of the current point, and hence suffer from the same difficulties as calculations of molecular dynamics when dealing with large structural rearrangements, although to a somewhat lesser extent. The global-step methods attempt to hop between minima, and benefit greatly from the power of the local minimization which finds the nearest minimum after each random step in the most efficient way [90,91], however, two key problems need to be solved for this method to fold even small proteins: (i) how to incorporate the missing free energy terms, which are computationally expensive, particularly when the analytical energy derivatives are required; and (ii) how to make a smart random movement and avoid sampling of irrelevant parts of the conformational space. One way to add the solvation, electrostatics, and, possibly, the entropy in the simulation is to evaluate them only once per step when the selection criterion is applied. Omission of these terms in local minimization can be justified by their relatively weak dependence on the local adjustments. The smart move may use electrostatics [92], statistical distributions [61] and local chain deformations.

The large number of combinations is not that detrimental; as has been shown recently [93], side chain conformations can be placed relatively independently of each other, and the vast majority of the main chain torsion angle combinations have exceedingly large energies and can be skipped. I believe that the folding problem will be solved for small proteins by a highly realistic calculation in the near future.

Acknowledgments: The author is grateful to Maxim Totrov for helpful discussions, and Dr. David Thomas for critical reading of the manuscript.

REFERENCES

[1] Anfinsen, C.B. (1973) Science 181, 223–230.
[2] Kim, P.S. and Baldwin, R.L. (1990) Annu. Rev. Biochem. 59, 631–666.
[3] Creighton, T.E. (1992) Protein Folding, Freeman, NY.
[4] Gething, M.-J. and Sambrook, J. (1992) Nature 355, 33–45.
[5] Hartl, F.U. and Martin, J. (1992) Annu. Rev. Biophys. Biomol. Struct. 21, 293–322.
[6] Levinthal, C. (1968) Chim. Phys. 65, 44–45.
[7] Baker, D., Sohl, J.L. and Agard, D.A. (1992) Nature 356, 263–265.
[8] Mottonen, J., Strand, A., Symersky, J., Sweet, R.M., Danley, D.E., Geoghegen, K.F., Gerard, R.D. and Goldsmith, E.J. (1992) Nature 355, 270–273.
[9] Creighton, T.E. (1992) Nature 356, 194–195.
[10] Carrell, R.W. and Evans, D.L.I. (1992) Curr. Opin. Struct. Biol. 2, 438–336.
[11] Kuwajima, K. (1989) Proteins 6, 87–103.
[12] Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A. (1975) J. Phys. Chem. 79, 2361–2381.

[13] Weiner, S.J., Kollman, P.A., Case, D.A., Chandra Singth, U., Ghio, C., Alagona, G., Prefeta Jr., S. and Wiener, P. (1984) J. Am. Chem. Soc. 106, 765–783.
[14] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) J. Comput. Chem. 4, 187–217.
[15] Harvey, S. (1989) Proteins 5, 78–92.
[16] Davis, M.E. and McCammon, J.A. (1990) Chem. Rev. 90, 509–521.
[17] Sharp, K.A. and Honig, B. (1990) Annu. Rev. Biophys. Biophys. Chem. 19, 301–332.
[18] Still, W.C., Tempczyk, A., Hawley, R.C. and Hedrickson, T. (1990) J. Am. Chem. Soc. 112, 6127–6129.
[19] Thomas, D. (1992) FEBS Lett. 307, 10–13.
[20] Novotny, J., Bruccoleri, R. and Karplus, M. (1984) J. Mol. Biol. 177, 787.
[21] Eisenberg, D. and McLachlan, A.D. (1986) Nature 316, 199–203.
[22] Miyazawa, S. and Jernigan, R.L. (1985) Macromolecules 18, 534–552.
[23] Hendlich, M., Lackner, P., Weitckus, S., Floeckner, H., Froschauer, R., Gottsbacher, K., Casari, G. and Sippl, M.J. (1990) J. Mol. Biol. 216, 167–180.
[24] Maiorov, V.N. and Crippen, G.M. (1992) J. Mol. Biol. 227, 876–888.
[25] Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) Nature 358, 86–89.
[26] Vriend, G. and Sander, C. (1993) J. Appl. Cryst. 26, 47–60.
[27] Casari, G. and Sippl, M.J. (1992) J. Mol. Biol. 224, 725–732.
[28] Friedrichs, M.S. and Wolynes, P.G. (1990) Science 246, 371–373.
[29] Holm, L. and Sander, C. (1992) J. Mol. Biol. 225, 92–105.
[30] Godzik, A., Kolinski, A. and Skolnick, J. (1992) J. Mol. Biol. 227, 227–238.
[31] Vila, J., Williams, R.L., V'asquez, M. and Scheraga, H.A. (1991) Proteins 10, 199–218.
[32] Wesson, L and Eisenberg, D. (1992) Protein Sci. 1, 227–235.
[33] Troyer, J.M. and Cohen, F.E. (1991) Reviews in Computational Chemistry, vol. II (Lipkowitz, K.B. and Boyd, D.B. eds.) pp. 57–97, VCH Publishers.
[34] Go, N. and Taketomi, H. (1978) Proc. Natl. Acad. Sci. USA 75, 559–563.
[35] Skolnick, J. and Kolinski, A. (1991) J. Mol. Biol. 221, 499–531.
[36] Chan, H.S. and Dill, K.A. (1991) Annu. Rev. Biophys. Chem. 20, 447–490.
[37] Finkelstein, A.V. and Reva, B.A. (1991) Nature 351 , 497–499.
[38] Hinds, D.A. and Levitt, M. (1992) Proc. Natl. Acad. Sci. USA 89, 2536–2540.
[39] Go, N. (1983) Annu. Rev. Biophys. Bioeng. 12, 183–210.
[40] Shakhnovich, E.I. and Gutin, A.M. (1990) Nature 346, 773–775.
[41] Gregoret, L.M. and Cohen, F.E. (1991) J. Mol. Biol. 219, 109–122 .
[42] Levitt, M. and Warshell, A. (1975) Nature 253, 694–698.
[43] Kuntz, I.D., Crippen, G.M. Kollman, P.A. and Kimelman, D. (1976) J. Mol. Biol. 106, 983–994.
[44] Wilson, C. and Doniach, S. (1989) Proteins 6, 193–209.
[45] Friedrichs, M.S., Goldstein, R.A. and Wolynes, P.G. (1991) J. Mol. Biol. 222, 1013–1034.
[46] Tanaka, S. and Scheraga, H.A. (1976) Macromolecules 9, 945–950.
[47] Crippen, G.M. and Snow, M.E. (1990) Biopolymers 29, 1479–1489.
[48] Gregoret, L.M. and Cohen, F.E. (1990) J. Mol. Biol. 211, 959–974.
[49] Sippl, M.J. (1990) J. Mol. Biol. 213, 859–883.
[50] Levitt, M. (1983) J. Mol. Biol. 170, 723–764.
[51] Noguti, T. and Go, N. (1989) Proteins 5, 97–138.
[52] Richards, F.M. (1977) Annu. Rev. Biophys. Bioeng. 6, 151–176.
[53] Desmet, J., Maeyer de M., Hazes, B. and Lasters, I. (1992) Nature 356, 539–542.
[54] Ooi, T., Oobatake, M., Nemethy, G. and Scheraga, H.A. (1987) Proc. Natl. Acad. Sci. USA 84, 3084–3090.
[55] Cramer, C.J. and Truhlar, D.G. (1992) Nature 256, 213–217.
[56] Perrot, G., Cheng, B., Gibson, K.D., Vila, J., Plamer, K.A., Nayeem, A., Maigret, B. and Scheraga, H.A. (1992) J. Comp. Chem. 13, 1–11.
[57] Eisenhaber, F. and Argos, P. (1993) J. Comp. Chem. (in press).
[58] Abagyan, R.A., Totrov, M.M. and Kuznetsov, D.A. (1993) J. Mol. Biol. (submitted).
[59] Richmond, T.J. (1984) J. Mol. Biol. 178, 63–89.
[60] Williams, R.L., Vila, J., Perrot, G. and Scheraga, H.A. (1992) Proteins 14, 110–119.
[61] Abagyan, R.A. and Totrov, M.M. (1993) J. Mol. Biol. (submitted).
[62] Yang, A.-S., Sharp, K.A. and Honig, B. (1992) J. Mol. Biol. 227, 889–900.
[63] Warshel, A. and Russell, S.T. (1984) Quart. Rev. Biophys. 17, 283–422.
[64] Pickersgill, R.W. (1988) Protein Eng. 2, 247–248.
[65] Mehler, ElL. and Eichele, G. (1984) Biochemistry 23, 3887–3891.
[66] Schaefer, M. and Froemmel, C. (1990) J. Mol. Biol. 216, 1045.
[67] Tanford, C. and Kirkwood, J.G. (1957) J. Am. Chem. Soc. 79, 5333–5339.
[68] Friedman, H.L. (1975) Mol. Phys. 29, 1533–1543.
[69] Imoto, T. (1983) Biophys. J. 44, 293–298.
[70] Tanford, C.H. (1980) Hydrophobic Effect, Wiley, New York.
[71] Go, N. and Scheraga, H.A. (1969) J. Chem. Phys. 51, 4751–4767.
[72] Paine, G.H. and Scheraga, H.A. (1985) Biopolymers 25, 1391–1436
[73] Finkelstein, A.V. and Janin, J. (1989) Protein Eng. 3, 1–3.
[74] Creamer, T.P. and Rose, G.D. (1992) Proc. Natl. Acad. Sci. USA 89, 5937–5941.
[75] Ponder, J.W. and Richards, F.M. (1987) J. Mol. Biol. 193, 775–791.
[76] Leach, A.R. (1991) Reviews in Computational Chemistry, vol. II (Lipkowitz, K.B. and Boyd, D.B., eds) pp. 1–55, VCH Publishers.
[77] Dill, K.A. (1992) Curr. Opin. Struct. Biol. 3, 99–103.
[78] Bruccoleri, R.E. and Karplus, M.A (1990) Biopolymers 29, 1847–1862.
[79] van Gunsteren, W.F. and Berendsen, H.J.C. (1990) Angew Chem. Int. Ed. Engl. 29, 992–1023.
[80] Mazur, A.K., Dorofeev, V.E. and Abagyan, R.A. (1991) J. Comput. Phys. 92, 261–272.
[81] Vásquez, M. and Scheraga, H.A. (1985) Biopolymers 24, 1437–1447.
[82] Vajda, S. and Delisi, C. (1990) Biopolymers 29, 1755–1772.
[83] Simon, I., Glasser, L. and Scheraga, H.A. (1991) Proc. Natl. Acad. Sci. USA 88, 3661–3665.
[84] Kabsch, W. and Sander, C. (1984) Proc. Natl. Acad Sci. USA 81 , 1075.
[85] Piela, I., Kostowicki, J. and Scheraga, H.A. (1989) J. Phys. Chem 93, 3339–3346.
[86] Kostrowicki, J. and Scheraga, H.A. (1992) J. Phys. Chem. 96, 7442–7449 .
[87] Shalloway, D. (1992) in: Recent Advances in Global Optimization, vol. 1 (Floudas, C.A. and Pardalos, P.M. eds.) pp. 433–477, Princeton University Press.
[88] Noguti, T. and Go, N. (1985) Biopolymers 24, 527–546.
[89] Shin, J.K. and Jhon, M.S. (1991) Biopolymers 31, 177–185.
[90] Li, Z. and Scheraga, H.A. (1987) Proc. Natl. Acad. Sci. USA 84, 6611–6615.
[91] Abagyan, R.A. and Argos, P. (1992) J. Mol. Biol. 225, 519–532.
[92] Ripoll, D.R. and Scheraga, H.A. (1988) Biopolymers 27, 1283–1303.
[93] Eisenmenger, F., Argos, P. and Abagyan, R.A. (1993) J. Mol. Biol. (in press).