Conformational Energy Calculations on Polypeptides and Proteins

Maximiliano Vásquez

Protein Design Labs, Inc., 2375 Garcia Avenue, Mountain View, California 94043

George Némethy[†]

Department of Biomathematical Sciences, Box 1023, Mount Sinai School of Medicine, New York, New York 10029

Harold A. Scheraga[*]

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

Received March 5, 1993 (Revised Manuscript Received May 19, 1994)

Contents

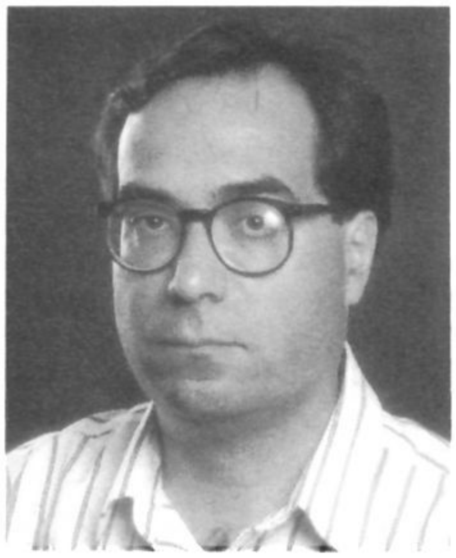
I. Introduction

One of the main goals of conformational energy calculations on polypeptides and proteins is the analysis and prediction of their three-dimensional structures. These structures are the result of the balance of intra- and intermolecular interactions, which in turn can be expressed in terms of interatomic potential energy functions. In order to gain an understanding of the physical reasons for the folded structures of these molecules, it is necessary to elucidate how the potential energy determines the structures themselves, their stability, and their dynamic properties.[1-3]

Small flexible molecules exist in an ensemble of low-energy conformations.[4-6] By contrast, a globular protein in its native, biologically active state exists in a well-defined and recognizable average conformation, with small, but not necessarily harmonic, fluctuations around this average.[7] There is considerable evidence indicating that proteins can fold spontaneously into their native conformations, both in vivo and in vitro. It follows that the native conformation is determined by the amino acid sequence, together with the characteristic properties of the physical environment (solvent medium, temperature) in which the molecule folds into its preferred spatial structure.[1,2,5,8-11] The latter idea provides the fundamental assumption that underlies many conformational energy computations, viz. that the native conformation is the one for which the *free* energy of the system (protein and surroundings) is a minimum. This would mean that, at a finite temperature, the occupancy of conformational states will be significantly different from zero for only one low potential energy minimum, or perhaps for a collection of spatially close minima. Although this principle, commonly termed the "thermodynamic hypothesis", provides the most direct justification for calculations of protein structure, it is by no means universally accepted. Starting with the statement of the Levinthal paradox,[12,13] which argues for the impossibility of a complete search of conformational space for even a small protein, a number of pathway-oriented justifications have been made.[14-16]

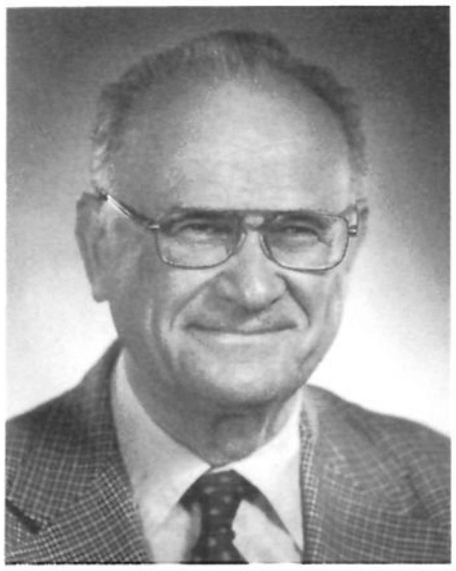
[†] Deceased May 18, 1994.

0009-2665/94/0794-2183$14.00/0



Maximiliano Vásquez was born in San José, Costa Rica, in 1958. He received his B.Sc. in Chemistry in 1979 from the Universidad de Costa Rica, and his Ph.D. in Biophysical Chemistry from Cornell University in 1987. At Cornell, he worked on computational aspects of peptide and protein structural problems with Professor Harold A. Scheraga. In 1988, he joined Tripos Associates Inc. in St. Louis, MO, where he led the development of modeling software for biopolymers. In 1990, he moved to Protein Design Labs in Mountain View, CA, where he has continued to pursue his research interests in peptide and protein structure, protein engineering, and general issues in computational studies of molecular recognition. In particular, he has participated in the design process of over a dozen "humanized" antibodies.

George Némethy was born in Budapest, Hungary, in 1934. He left Hungary in 1951. He received a B.A. degree (magna cum laude) in chemistry at Lincoln University in 1956 and a Ph.D. degree in physical chemistry at Cornell University in 1962. He wrote his Ph.D. thesis, entitled "The Structure of Water and Hydrophobic Bonding in Proteins", under the direction of Professor Harold A. Scheraga. He was on the faculty of the Rockefeller University, initially affiliated with the laboratory of Professor Daniel E. Koshland, with whom he developed the KNF model of allosteric interactions, and later as head of a laboratory in biophysical chemistry. He held visiting professorships in the laboratory of Professor Jeannine Yon at the Université de Paris-Sud in Orsay and at the State University of New York at Binghamton. From 1975 to 1989, he was senior research associate at Cornell University, where he collaborated with Professor Scheraga on the development of theoretical methods for protein conformational studies. Subsequently, he was professor in the Department of Biomathematical Sciences of the Mount Sinai School of Medicine in New York. He was also a visiting lecturer at the Instituto Superiore de Sanità in Rome and the Università Federico II in Napoli. His research interests were in the theory of protein structure and interactions, the physical chemistry of proteins, the theoretical conformational analysis of peptides and proteins, including the structure and assembly of collagen, and the structure of water and aqueous solutions. In 1968, he was the recipient of the Lincoln University Alumni Award. In 1972, he received the "Pius XI Gold Medal" award of the Pontifical Academy of Sciences. He was the author of over 160 scientific publications. Sadly, George Némethy died of cancer on May 18, 1994.



Harold A. Scheraga was born in Brooklyn, NY, in 1921. He attended the City College of New York, where he received his B.S. degree in 1941, and went on to graduate work at Duke University, receiving his Ph.D. degree in 1946 and an Sc.D. degree (Hon) in 1961. Following postdoctoral work at Harvard Medical School with Professor John T. Edsall, he joined the faculty at Cornell University in 1947, where he is now Todd Professor of Chemistry Emeritus. He has received numerous National and International awards including election to the U.S. National Academy of Sciences in 1966, the Linderstrøm–Lang medal of the Carlsberg Laboratory in 1983, and most recently the International Society of Quantum Chemistry and Quantum Pharmacology Award in Theoretical Biology in 1993, and he will receive the Stein and Moore Award of the Protein Society in 1995. He has given numerous invited lectures, has served on government, industrial and academic advisory committees, and is a member of the editorial boards of several scientific Journals. His research interests are in the experimental and theoretical aspects of the structure and function of proteins and other macromolecules, the physical chemistry of blood clotting, and the structure of water and dilute aqueous solutions.

Historically, the computational analysis of polypeptide conformations started with simple considerations of stereochemistry, based on a hard-sphere potential.[17,18] Very soon, more realistic potential functions were developed, leading to the force fields that are in use today (Section II.A). The development of the field of conformational energy computations on polypeptides and proteins has been surveyed in earlier reviews.[1,2,5,8,19−32]

A very general consideration of calculations of polypeptide or protein structures reveals the complexity of these systems and the difficulties that may be encountered in describing the most structurally relevant interactions accurately. The systems of interest are described by a large number of degrees of freedom, which in turn lead to astronomical numbers of possible conformations, even if restricted to true (local) minimum-energy conformations. In analyzing polypeptide or protein systems with computational techniques, one is thus confronted by an inherent conflict: better sampling versus more accurate descriptions of energetics (the best, in theory, including a large number of explicit solvent molecules). A number of studies concentrate primarily on one of these issues, while limiting the scope of conclusions governed by the second issue. For example, most of the early molecular dynamics (MD) studies of proteins concentrated on sampling the neighborhood of the native structure as observed by X-ray diffraction (or some other suitable starting conformation). Only short times, at best on the order of nanoseconds, can be simulated and thus the methodological emphasis is placed mainly on the accuracy of the description of the potential energy. This involves issues such as solvation and accurate but efficient treatment of long-range electrostatic interactions. On the other hand, other studies use a simplified form of the potential function, for example, one without explicit solvent, but a more ambitious exploration of conformational space is attempted. Clearly, the philosophy of these two types of studies is quite different, even though some of the methodological aspects may at times be very similar if not identical. An understanding of these differences is necessary for a fair and informative evaluation of conformational studies of polypeptides and proteins in general. It should also be pointed out that progress in methodology for efficient conformational search, and in particular, for global optimization, would immediately lead to a better study of more complex descriptions of potential energy.

Basic aspects of the computational methodology, together with a discussion of some of the current problems and of the approaches used to address them, are reviewed in section II. The application of these methods to a variety of oligopeptide, polypeptide, and protein structures is summarized in section III. Computations on known systems are an important part of a test and of a verification of the basic assumptions of new and established computational approaches. In conjunction with experimental methods for structural determination, such as X-ray crystallography and solution NMR spectroscopy, computations are also an essential tool of structural refinement.

II. Methodology

A. Force Fields

1. Atomic Potential Functions

Various parameterizations of molecular mechanics potential functions have been derived for computations on polypeptides. These include those used in the AMBER program,[33,34] CHARMM,[35] DISCOVER,[36] ECEPP,[37,38] ENCAD,[39,40] GROMOS,[41] AMBER/OPLS,[42] etc. The basic structure of these force fields is similar in all of them, as well as in force fields used in computations of small organic molecules, e.g., MM2[43,44] and MM3.[45-47] The MM2 force field has also been generalized to treat polypeptides,[48] but its use has not been extensive.

It is important to distinguish between the force fields themselves and the computer programs used to implement them, because a one-to-one correspondence does not always exist. For example, the DISCOVER program also has the option of implementing the force field of Kollman and co-workers, often termed the AMBER force field. The program Macromodel of Still and co-workers[49] uses versions of the MM2, AMBER/OPLS, and AMBER* [50] force fields, some of which can be applied to polypeptide systems. The program X-PLOR[51,52] has used versions of the CHARMM force field. Similarly, there are several implementations of the ECEPP force field, including those in the KONF90,[53] ICM,[54-56] and FANTOM[57] computer programs.

The force fields used in all of these programs are expressed in terms of classical empirical potential functions, and are written as a sum of several components. The mathematical form of each term is based largely on a phenomenological concept concerning the nature of the energy term, expressed in a manner that renders the function efficient for computational programming. The constants that describe molecular geometry (i.e., bond lengths and bond angles) and the strength of particular interatomic interactions are generally parameterized on empirical structural (e.g., crystallographic), spectroscopic (e.g., force constants), and thermodynamic information available from small organic molecules or relatively simple solutions (as, for example, in the OPLS[42] potential). The functions may also include some data derived from quantum mechanical computations (most often the partial atomic charges). The potential energy is expressed in the form of atom-centered potentials, with the energy of the molecule computed as a sum over all pairwise interactions. This description is generic and applies to some, but by no means all, of the aspects of the force fields mentioned above. Alternative formulations, where fairly new forms of the interaction terms have been proposed, were reported recently.[58,59] Hagler and Ewig have reviewed the field extensively, including the most recent advances.[60]

As an illustrative example, we summarize here the formulation of the ECEPP force field[37,38,61,62] derived at Cornell University. Very similar forms, but with different parameters, are used in the other potential energy functions mentioned. The intramolecular energy is given by the expression:

$$U = \sum_{i\neq j}\epsilon_{ij}\left[\left(\frac{r_{ij}^{0}}{r_{ij}}\right)^{12} - \left(\frac{r_{ij}^{0}}{r_{ij}}\right)^{6}\right] + \sum_{i\neq j}\epsilon_{ij}\left[5\left(\frac{r_{ij}^{0}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij}^{0}}{r_{ij}}\right)^{10}\right] + \sum_{i\neq j}\frac{q_iq_j}{Dr_{ij}} + \sum_{k}\frac{A_k}{2}(1 \pm \cos n\theta_k) \quad (1)$$

where ϵ_{ij} and r_{ij}^0 are the potential depth and position of the minimum of the pair interaction (the 12-6 expression pertains to the nonbonded energy, and the 12-10 expression to the hydrogen-bonding energy), q_i is the partial atomic charge on atom i, D is the dielectric constant, r_{ij} is the distance between the two interacting atoms, A_k is the barrier height for rotation around the kth bond, $\theta_{\rm k}$ is the dihedral or torsion angle, and n is the n-fold degeneracy of the torsional potential. The variable dihedral angles [(ϕ,ψ,ω) for the backbone and χs for the side chains] are defined in ref 5. In ECEPP, the bond lengths and bond angles are fixed at experimental values. The original form of ECEPP was published in 1975.[37] It was upgraded in 1983 and 1984 (ECEPP/2)[61,62] and in 1992 (ECEPP/3),[38] as improved experimental data became available for its parameterization. Independently, Momany and co-workers developed their own updated version (ECEPP83).[63]

Other formulations of the potential energy function (e.g., those used in AMBER, CHARMM, or DISCOVER) include terms that allow for bond stretching and bond angle bending, and possibly for other deformations, such as the out-of-plane deformation of the peptide bond, i.e. for flexible geometry. Hence, eq 1 may be augmented by terms of the form:

$$U = \frac{1}{2}\sum K_b(b - b_0)^2 + \frac{1}{2}\sum K_\tau(\tau - \tau_0)^2 + \sum K_\lambda(\lambda - \lambda_0)^2 \quad (2)$$

where the terms refer to bond stretching, bond angle bending, and out-of-plane deformation (often implemented as improper dihedral angle terms[33,35]), respectively.

In principle, it is more rigorous to include the terms in eq 2 because molecules are not completely rigid (in their bond lengths and bond angles, for example). Allowing for flexibility is a necessity when details of the molecular geometry must be determined in small, highly substituted organic molecules, because most local steric overlaps can be relieved (fully or partially) only by distortions of the geometry. On the other hand, the consideration of flexible geometry may be less helpful in the static structural analysis of oligopeptide and protein structures, because constraints may be relieved by other means. A discussion of this issue can be found elsewhere.[64-67] However, molecular dynamics trajectories that use rigid bond lengths and bond angles appear qualitatively distorted when compared with those using full flexibility or only rigid bond lengths.[68]

The complete description of the energy of a polypeptide must also include its interaction with the solvent. The most direct approach to accomplish this is by use of a number of explicit water or, in general, solvent molecules. These solvent molecules are often described by potentials of identical form to those in eqs 1 and 2; this applies to both the solvent–solvent and the peptide–solvent interactions. A number of water models have been developed and used in simulations of hydrated polypeptides and proteins.[40,69,70] These models give a fair description of some structural and thermodynamic properties of bulk water, although usually within a narrow range of temperature and pressure,[71] while remaining relatively simple. In spite of its simplicity, explicit inclusion of solvent molecules greatly increases the computational time needed to study a polypeptide or protein system. Thus, most studies of this type have been limited to simulations of restricted conformational explorations. Important exceptions are calculations of free energy differences by solution simulation using MD or Monte Carlo (MC) algorithms; these have been reviewed recently,[72] and only a few examples will be examined in this review in section II.C. The computational complexity of explicit solvent calculations has also motivated the development of *implicit* models for solvation, and these are described in section II.A.3.

The potential forms discussed above usually include all atoms of a polypeptide molecule explicitly, while some may even include additional lone pair or dummy atoms to improve aspects of the interaction, e.g., the extra interaction centers around sulfur atoms in the force field developed by Kollman and co-workers,[33,34] and the early work on the empirical potential based on the interactions of electrons and nuclei or EPEN.[73] However, it is often possible to treat some of the atoms, most frequently the nonpolar hydrogen atoms, implicitly as part of modified "united atoms". The 1975 version of ECEPP was adapted to this form in the UNICEPP force field.[74] The original force field in AMBER had this form,[33] but a full-atom version was introduced later;[34] both versions are commonly used by Kollman and co-workers as well as by other researchers.[72,75] The AMBER/OPLS force field of Jorgensen and Tirado-Rives exists only in a united atom form.[42] These simplifications lead to some decrease in computational time, but, of course, also represent more approximate forms.

Another common approximation is the use of a cutoff distance for the nonbonded and electrostatic interactions. For large molecules, the leading term in computational time is given by the first three expressions in eq 1. The number of such terms, and thus the computational effort, grows with the square of the number of atoms. When using interaction distance cutoffs, the scaling of the computation with system size is close to linear, resulting in significant savings for the largest systems.[35,49,76,77] The use of cutoffs represents yet another approximation in the description of polypeptide conformational behavior.

Recently, the results of a number of high quality *ab initio* quantum mechanical calculations on amino acids and small peptides have become available.[78-80] These are important for calibration and further refinement of the potentials described above. Critical comparisons between the *ab initio* results and empirical calculations of terminally blocked alanine (the alanine "dipeptide") have been presented.[75]

It should be clear from the discussion in this section that a number of approximations are often made in computational studies of polypeptides and proteins. The appropriate balance of approximations is often a matter of computational expediency modulated by the experience of the researcher. It also depends on the goal of the calculation, and the interpretation that one wishes to make of its results. The recurrent issue of accuracy versus sampling discussed in the Introduction is also relevant here.

2. Simplified Forms of the Potential Function

The astronomical size of the conformational space of even the smallest proteins has motivated the development of very simplified potential functions in which each amino acid residue is represented by one or two interaction points. The parameters describing interresidue interactions have come from two sources: averaging over atomic level potentials and extracting energetics from observed protein structures.

In their early work on simplified models, Levitt and Warshel [81,82] represented residues in terms of two centers of interaction that describe the backbone and the centroid of the side chain, respectively. The parameters describing intra- and interresidue interactions were derived by examination of full-atom potentials. Their force field also included an approximate description of solvation. A similar approach has been described recently by Liwo et al.[83,84]

The method of Pincus and Scheraga used averaged residue–residue potentials in the context of implementing a cutoff scheme for the ECEPP potential.[76] A similar algorithm, using truncated Fourier series fitted to a full-atom force field, was described recently and termed peptide mechanics.[85] However, peptide mechanics is intended to be used by itself, and not necessarily as an alternative to treat long-range interactions in a full-atom force field. Thus, this simplified model attempts to provide a proper description of local peptide energetics, and interactions between nonsequential amino acids use united-residue forms even at short distances.

A somewhat different approach was described by Head-Gordon and Brooks,[86] who developed a reduced representation starting from the CHARMM[35] force field. They represent each amino acid residue by a small number of virtual interaction points. To estimate interactions between these virtual points, they use a multipole expansion with coefficients derived from an analysis of point charges in the reference, CHARMM, full-atom description. In addition, they have presented a treatment for dynamic simulation of peptides using this approach, which they name virtual rigid-body dynamics.

A second important source for residue–residue effective potentials is the analysis of structures in the protein structural databank.[87,88] Most often, this is done by obtaining an empirical free energy function or potential of mean force, based on Boltzmann statistics derived from the spatial distribution of residues in proteins of known structure. In one version of this formalism, the change in free energy corresponds to the formation of a contact between a given pair of residues. Its numerical values are derived from the relative contact frequencies of pairs (or even triplets[89]) of various residues in known protein structures, using the quasichemical approximation.[90] These contact potentials are particularly useful in, but not limited to, studies of lattice models of proteins. In other representations, the energy is expressed as a function of the interresidue distance, and its parameters are derived from the observed distance distributions between given residue types in proteins.[89,91–93]

It is also possible to derive potentials of mean force in terms of variables other than distances. To the best of our knowledge, the first extraction of effective energies for proteins from frequency information (a process termed the inverse Boltzmann approach[11]) was made by Pohl using dihedral angles ϕ and ψ.[94] This approach was clearly premature because less than a dozen three-dimensional structures of proteins were known at the time. However, the value of this work cannot be underestimated, and its influence can be found in many of the more statistically robust approaches of this nature being used today.[11] More recently, others[95–97] have also started to use ϕ–ψ-dependent potentials of mean force in their studies.

In addition to reducing the number of variable coordinates, residue–residue potentials derived from protein structural information will contain, in an averaged manner, solvation and other effects that are difficult to account for explicitly.

Crippen and co-workers have developed a series of simplified potentials for proteins by explicitly fitting the parameters in an attempt to make sure that the native structure possesses the lowest energy out of a given ensemble of alternatives.[98–103] The most recent approach,[103] based on a scheme for efficient solution of large systems of coupled inequalities, seems as effective as the best in the literature.

An alternative procedure to extract energy information from observed structures has been presented by Wolynes and co-workers, who derived energy functions for simplified protein models by use of associative memory Hamiltonians. The resulting form of these energy functions can be modified using ideas from spin-glass theory, to reduce as much as possible the number of false local minima.[104,105] This is expected to make the multiple-minima problem less severe in subsequent calculations using this model.

The tremendous growth of the amount of experimental information available from protein structures, coupled with the application of more computationally rigorous approaches for its analysis, constitutes one of the major and most promising recent developments in the field of protein structure computations. The application of some of these approaches is reviewed in section III.E.3.

3. Interactions with Solvent

As outlined earlier, the energy of solute–solvent and solvent–solvent interactions can be expressed in terms of the same force fields as those used to calculate the intramolecular conformational energies of proteins.[106] In addition to the large computational cost of including solvation explicitly in conformational energy calculations, there are also some conceptual problems with a direct generalization of local or global energy minimization procedures to the complete system of solute and bath of solvent molecules (see, for example, the discussion by Schiffer and co-workers[107]). These conceptual difficulties arise mainly because the interactions between a protein molecule in solution and the surrounding water molecules are generally short-lived. In other words, the positions of most water molecules, relative to the protein, are not fixed for a specified protein conformation, and therefore no unique set of pairwise interactions can be computed between atoms of the protein and of the solvent. One cannot simply minimize the energy with respect to the positions of the solvent molecules and effectively "freeze" the water; a thermal average is required. Thus, it is not only necessary to compute the potential energy of *one* water bath configuration for each conformation of the solute, but the free energy, or potential of mean force, of hydration must be computed by averaging over a *large number* of configurations of the solvent, using MC or MD techniques.[25,69,108] Simulations of proteins in baths of explicit water molecules can in principle yield the free energy of hydration, together with information regarding the dynamics of solutes and solvent. The use of simulation methods to study solvent effects on polypeptide structure and dynamics has been reviewed elsewhere.[75]

The computational demands of explicit solvent simulations preclude their use in studies involving extensive searches of the conformational space of a polypeptide or protein, e.g., in the search for (global) minimum-energy conformations in the absence of restraining information. In such cases, a practical solution is the use of a mean-field approximation for the free energy of solvation.[109] The solute–solvent interaction is computed by expressing it as an appropriate average over all positions and orientations of the solvent molecules for a given conformation of the solute, without taking into account the molecular nature of the solvent and of details of motions of the solvent molecules. This average value may be obtained by using simpler implicit hydration models, instead of an actual computation over the possible solvent configurations. The use of potentials of mean force to describe peptide–solvent interactions is, in principle, sufficient to study static or thermodynamic aspects of peptide structure. However, when studying kinetic or dynamic effects, it is also necessary to consider the frictional effects of solvation. Coupling to a stochastic bath by way of Langevin dynamics is one possible method to accomplish the implicit description of this aspect of solvation.[25]

It is also important to realize that, in creating an implicit model of solvation, the potential energy function acquires a hybrid character, because the solvent-dependent terms become temperature dependent, and thus they have some free energy character. More explicitly, the full potential energy function[110]

$$E(\mathbf{X}_\mathrm{p},\mathbf{Y}_\mathrm{w}) = U_\mathrm{p}(\mathbf{X}_\mathrm{p}) + U_\mathrm{wp}(\mathbf{X}_\mathrm{p},\mathbf{Y}_\mathrm{w}) + U_\mathrm{ww}(\mathbf{Y}_\mathrm{w}) \quad (3)$$

is replaced by averaging over the solvent coordinates:

$$F(\mathbf{X}_\mathrm{p},T) = -RT \ln \int \mathrm{e}^{-E/RT}\, \mathrm{d}\mathbf{Y}_\mathrm{w} = U_\mathrm{p}(\mathbf{X}_\mathrm{p}) + \mathrm{PMF}(\mathbf{X}_\mathrm{p},T) \quad (4)$$

where $\mathbf{X}_\mathrm{p}$ and $\mathbf{Y}_\mathrm{w}$ correspond to solute and solvent coordinates, respectively. The direct and indirect effects of the solvent are all now included in the potential of mean force term, PMF($\mathbf{X}_\mathrm{p}$,T), and it has been assumed that the $U(\mathbf{X}_\mathrm{p})$ term corresponds to a true vacuum potential (e.g., dielectric constant D is 1). In many applications, solvent effects are separated into local effects, included in some form of potential of mean force, plus electrostatic effects, introduced as a modification of the electrostatic component of the intramolecular potential. Often, it is better not to carry out the partition into terms of eq 3, and thus to refer to the *full* function $F(\mathbf{X}_\mathrm{p},T)$ as the potential of mean force; i.e. it is not always possible or practical to separate out a true vacuum intramolecular potential. One reason why this separation is not always helpful is the strong effect of highly polarizable solvents, such as water, on electrostatic interactions.

a. **Empirical Hydration Models**. The principal assumption of these models is that the hydration free energy can be expressed for every functional group of the solute in terms of an averaged free energy of interaction of the group with a layer of nearby water molecules that form a hydration shell. It is assumed that this interaction is specific for each of the various types of atoms of the protein or other solute, and that the total solute–solvent interaction can be expressed additively, as a sum of the free energies of solvation for each of the component groups of the solute molecule. Several alternative forms have been developed for this purpose.

i. Accessible Volume Shell Model. In this type of the model, the free energy of hydration of any group is taken to be proportional to the water-accessible volume of a hydration layer surrounding it. This layer usually is considered to have the thickness of one water molecule. Interactions with the rest of the solvent are implicitly included in the empirical parameters of the model. The total free energy of hydration thus becomes

$$G_\mathrm{hyd} = \sum_{i=1}^{N} V_i g_i \quad (5)$$

where V_i is the solvent-accessible fraction of the hydration shell of group i, g_i is an empirically derived free energy parameter for group i, representing the "average free energy density", and N is the number of groups in the solute molecule.

Several forms of this model have been derived, starting from the early period of conformational energy computations, up to recent times.[111-115] This representation has a simple intuitive physical interpretation in terms of molecular nearest-neighbor interactions. On the other hand, it contains a number of approximations which may be sources of error and reduce the speed of computations.[116,117] These difficulties have led to the development of alternative forms.

ii. Accessible Surface Area Model. In this formulation, the free energy of hydration is assumed to be proportional to the solvent-accessible surface area of a group,[117-120] and is obtained as

$$G_\mathrm{hyd} = \sum_{i=1}^{N} A_i \sigma_i \quad (6)$$

where A_i is the solvent-accessible surface area, and σ_i is an empirically derived free energy density parameter. According to Richards,[121,122] A_i is defined as the surface traced by the center of a spherical test probe, with a radius equal to that of a solvent molecule, as the probe is rolled over the van der Waals surface of the molecule. Several recently developed algorithms for the rapid calculation of accessible surface areas and their derivatives are starting to make this approach very efficient[123-126] (two of these[124,125] are recent adaptations of analytical algorithms first developed by Richmond[127]).

Also, a number of approximations to the exact surface area have been tried, starting with the statistical procedure of Wodak and Janin,[128] later generalized by Still and co-workers.[129] Le Grand and Merz have also developed a very fast numerical approximation to surface areas and their derivatives.[130]

iii. Weighted Contact Model. The key assumptions in volume and surface area-based models are the proportionality of free energy to the degree of solvent exposure and the additive form of the solute–solvent potential of mean force, cf. eqs 5 and 6. It has recently been proposed that this simplified "physics" can just as easily be described by much simpler weighted contact models similar to the early ones by Gibson and Scheraga[131] and by Levitt.[82] Sander and co-workers have developed two versions of this type of model, and they have applied them in modeling and simulation of protein structure.[132-134]

In one version of the procedure,[134] the free energy density parameters, σ_i, are derived by fitting to experimental solvation free energies of amino acid side-chain analogs. The quality of this fit is at least as good as that achieved with a surface area-based model.[124] The resulting hydration model was then combined with a stochastic dynamics procedure[135] to simulate the behavior of bovine pancreatic trypsin inhibitor (BPTI) in water solution.

In a second version, parameters were derived by the inverse Boltzmann device[11,92] from solvent accessibility data on protein structures. The resulting model was used to discriminate between correct and incorrect protein folds.[133]

iv. Free Energy Parameters. In most forms of the hydration shell models, the proportionality constant representing the free energy density, σ_i (or g_i) is derived as an empirical parameter from G_{hyd} for small organic molecules, obtained in turn from observed free energies of hydration. A_i for these molecules is calculated from the molecular geometry. The experimental data are usually obtained from solution thermodynamic measurements,[117,119] but may also be derived from conformation-dependent physical properties, such as NMR coupling constants.[120] As noted above, the protein structural data base itself has been used as a source to parameterize a solvation preference model that uses weighted atomic contacts.[133]

Care must be exercised in the selection of the thermodynamic data used in the parameterization, as discussed by Ooi et al.,[117] because there are two possible ways in which transfer free energies can be obtained, and they refer to distinctly different processes. In one approach, numerical values are obtained from the distribution coefficients of solutes between an organic solvent and water.[119] These parameters model the free energy of transfer of amino acid side chains from the interior of the protein to aqueous environment, and hence they should be used by themselves, without any additional energy terms, to describe the total free energy of this process. In the second approach, on the other hand, if the expression for G_{hyd} is to be used to supplement the solute–solute interaction energies computed by intramolecular molecular mechanics forms [i.e., the $U(\mathbf{X}_p)$ given above], then G_{hyd} represents only the *additional* interactions of the atoms of the solute with water over and above intramolecular solute–solute interactions. Consequently, the G_{hyd} used in this manner must be derived from observed free energies of transfer of model solutes from gas to aqueous solution.[117,124] Furthermore, these model solutes should be small molecules (for example, CH_4, CO_2, $(CH_3)_2CO$, etc.) or, in general, rigid molecules, so that the assumption that the conformation does not change during the transfer (from gas to water) remains reasonable. Free energy data from larger, flexible molecules would contain contributions due to averaging over the conformations of the molecule and not just over the positions of the water molecules, as is ideal for this type of approach (see eqs 3 and 4 above and the accompanying discussion.)

Mention should also be made of the controversial reevaluation of solvation and transfer data carried out by Sharp et al.[136] These authors suggest that a correction term to account for differences in the sizes of solute and solvent needs to be added to reported values of the free energy of transfer. More recently, several groups have presented very strong arguments against the use of this correction.[137,138]

b. Integral Equation Theory. Integral equation theories have been used to calculate potentials of mean force in water between pairs of chemical groups. Several applications have started from extensions of the reference interaction site model (RISM) theory of Chandler and Andersen.[139] For the larger solutes, it is often necessary to invoke a superposition approximation. Using this approach, Pettitt and Karplus have built hydrated $\phi-\psi$ maps that compare favorably with those obtained by explicit simulation.[140] More recently, the superposition approximation has been removed in several studies, most notably in an analysis of a simulation of melittin, a small protein with 436 atoms.[141] A comparison of integral equation theory, with and without invoking the superposition approximation, with explicit solvent simulations has also been carried out for the *N*-methylacetamide peptide model compound.[142] An extensive review of the integral equation approach for the estimation of solvation effects on polypeptide conformation was presented recently.[143] This review[143] also presents an interesting comparison between these theories and continuum models such as those based on the Poisson–Boltzmann equation (see next section).

c. Electrostatic Interactions and Solvation. The solvent environment contributes to the conformational energy of a polypeptide not only through local solvation effects, as approximated in some of the treatments above, but also by its influence on the intramolecular force field, because of the dependence of the electrostatic term on the dielectric constant used. The proper treatment of electrostatic interactions in proteins, especially that of the dielectric medium, gives rise to a variety of physical and mathematical questions. These problems, together with current approaches, have been evaluated and discussed in several recent reviews.[144-148] Only a few aspects are summarized here briefly, because of their relevance to conformational energy computations.

The dielectric constant is a macroscopic quantity, yet it is applied to a microscopic situation in protein computations.[144] In the force fields listed in section II.A, it has been assumed that the dielectric constant in the interior of a folded macromolecule is low, in analogy to an organic solvent medium. Accordingly, D of eq 1 is usually chosen to lie between 2.5 and 4. This range has been justified by a theoretical analysis[149] using the Kirkwood–Fröhlich dielectric theory,[150] and it gives satisfactory results in many computations of the intramolecular energy.

The choice of the dielectric constant is more ambiguous for charges and dipoles located near the surface of the protein, i.e. in an inhomogeneous environment. It has been pointed out by Warshel and Aqvist[148] that the value of a dielectric constant depends on the property used to define it, so that one must use operational definitions that suit the problem at hand. In the case of a rigid body, the interactions of buried charges near a dielectric boundary can be treated by classical electrostatic theory, in terms of image charges.[151] The total electrostatic energy of a macromolecular system, including the change in this energy upon binding, can be calculated numerically in terms of thermodynamic cycles.[152]

Charges exposed on the surface of a protein and nearby free ions in the solution are located in a high-dielectric medium, with $D = 80$ or more in the case of an aqueous solution. If the Coulombic law form is used for the electrostatic energy (eq 1), there should be a transition from the low value of D in the interior to the high D in the free solvent. Therefore, the use of a distance-dependent dielectric constant has been proposed in numerous investigations. A simple assumption, used frequently, is that D depends linearly on the interatomic distance, i.e. $D = kr_{ij}$. It has been shown[153] that this form results in an overestimation of electrostatic forces, especially in the case of weak interactions.[154] Harvey[144] has pointed out that "distance-dependent dielectric constants and scaled partial charges can mimic solvent screening effects on electrostatic *energies*, at least qualitatively, ... but they cannot correctly treat electrostatic *forces*". This limits their usefulness both in energy minimization and in simulations.[144,151]

A rigorous and computationally feasible approach to the treatment of electrostatic interactions in folded proteins has been provided by the development of algorithms for the numerical solution of the Poisson–Boltzmann equation by means of a finite difference method.[154] The method was first used by Warwicker and Watson for the calculation of the electric potential in the active-site cleft of phosphoglycerate mutase.[155] A general algorithm has been developed by Honig and co-workers to solve the linearized finite-difference Poisson–Boltzmann equation (FDPB method), allowing for the treatment of intramolecular electrostatic interactions as well as for the effects of ionic strength.[146,152,154,156] It has been implemented in the program package DELPHI.[156]

In this method, the protein is placed in a cubic box divided into a regular grid. The Poisson–Boltzmann equation is solved iteratively at every point of the grid to find the electric field. A local charge density and ionic strength are assigned to each grid point, while the dielectric constant is assigned at midpoints along the grid lines. Thus, the variation of the dielectric response can be presented at atomic resolution if the grid is chosen sufficiently fine.[146] This eliminates the dilemma of the distance dependence of the dielectric constant.

The FDPB method has been used for numerous computations of charge effects in proteins.[146] Its performance has undergone severe testing,[156] in a comparison with other methods, for computing the interactions of ionized residues in subtilisin BPN′. The method reproduces the dependence on ionic strength better than other methods. In another test, it has been shown that the FDPB method and the use of a screened Coulomb potential worked similarly in the computation of pK shifts.[157]

Vorobjev et al.[158] have presented a combined iterative and boundary element approach, and a rapidly computable approximation thereof,[159] for solving the nonlinear Poisson–Boltzmann (NLPB) equation. The essence of the method is the separation of the calculation of the solvent reaction potential from that of the potential due to the ion distribution. The solvent reaction potential is calculated by using an efficient boundary element method. The ion-induced potential is then calculated by means of an efficient volume integration procedure using an iterative solution of the NLPB equation, coupled to the fixed molecular and solvent electrostatic potential. The method has been tested on oligopeptides and model protein solutes. Another recent improvement for the solution of the Poisson–Boltzmann equation is the use of multiple grids.[160]

Methods based on solution of the Poisson–Boltzmann equation are attractive because they are firmly grounded in classical continuum electrostatics. However, they are still an approximation to the system of a protein in solution, which is essentially discontinuous, and where the boundary between dielectric regions[161] is not always unambiguously defined. (See also section II.A.3.b and the discussion by Marlow et al.[143]) Thus, alternative methodologies have been developed, most notably the Langevin dipole–protein dipole (LDPD) methods of Warshel and co-workers, that attempt to reintroduce some of the discrete nature of the solvated protein system. This approach has been refined and applied to a number of problems, and an extensive review has also appeared recently.[148]

d. Generalized Born Models. Solvation treatments based on some form of continuum electrostatics are in some way complementary to empirical shell models. However, a simple merger of these models cannot be made directly because, in their parameterization and development, several effects may have been included independently, and thus simply adding them up would lead to incorrect weighting of these interactions. An early but promising attempt to achieve this merger is the generalized Born–surface area (GB/SA) method of Still and co-workers.[162] These authors start with the Born equation for the free energy of solvation of a simple ion:

$$\Delta G = -\frac{1}{2}\frac{q^2}{r_B}\left(1 - \frac{1}{D}\right) \tag{7}$$

where q is the charge and r_B the effective radius of the ion, and D is the dielectric constant of the solvent. Still et al. generalize the model for polyatomic molecules whose electrostatic potential is described by point charges. The "local" part of the solvation, also known as the cavity term together with effective solute–solvent van der Waals interactions, has the same form as surface area-based shell models. However, in this treatment the *same* free energy density parameter, in a given solvent, is used for all atom types (in this and other treatments, hydrogen atoms are often not considered explicitly, except, perhaps, in the "vacuum" intramolecular potentials). The differences between different atom types will ultimately be given by the electrostatic component. In more detail, the total solvation free energy is expressed as

$$G_{\mathrm{sol}} = G_{\mathrm{cav}} + G_{\mathrm{vdW}} + G_{\mathrm{pol}} \qquad (8)$$

and the contribution of the first two terms has the same form as eq 6 above with a single value of σ. The solute–solvent polarization term is then given by

$$G_{\mathrm{pol}} = -\frac{1}{2}\left(1 - \frac{1}{D}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{f_{\mathrm{GB}}} \qquad (9)$$

The generalized Born factor is in turn given by

$$f_{\mathrm{GB}} = [r_{ij}^2 + \alpha_{ij}^2 \exp\{-r_{ij}^2/(2\alpha_{ij})^2\}]^{1/2} \qquad (10)$$

where $\alpha_{ij} = (\alpha_i\alpha_j)^{1/2}$ is the geometric mean of the Born radii. The Born radii are computed numerically for each charged atom of the solute by equating the polarization free energy of solvation of each atom i, from the expression:

$$G_{\mathrm{pol}-i} = -\frac{1}{2}\left(1 - \frac{1}{D}\right)\frac{q_i^2}{\alpha_i} \qquad (11)$$

with the value obtained by numerical integration over spherical shells:

$$G_{\mathrm{pol}-i} = -\frac{1}{2}\left(1 - \frac{1}{D}\right)q_i^2\left[\sum_{k=1}^{k=k_{\max}}\frac{A_k}{4\pi r_k^2}\left(\frac{1}{r_k - (T_k/2)} - \frac{1}{r_k + (T_k/2)}\right) + \frac{1}{r_{k_{\max}+1}}\right] \qquad (12)$$

In this expression, A_k is the area of the kth shell outside the van der Waals envelope created by all atoms $j \neq i$, while T_k and r_k are the thickness and radius of the kth shell, respectively. $k_{\max}$ is defined such that a shell of radius r_k with $k = k_{\max} + 1$ would be large enough to surround the entire molecule. If it is assumed that α_i varies slowly with conformation, the more time-consuming numerical integration given in eq 12 can be executed every 10 to 20 iterations of a minimization or molecular dynamics run. Tests of the procedure have been reported for relatively small molecules, including terminally blocked alanine (the "alanine dipeptide").[162]

More recently, Abagyan and Totrov described a version more suitable for polypeptide calculations.[163] Sitkoff et al. have parameterized a model that combines the FDPB method with a generic surface area term, and report accurate reproduction of hydration free energies for a set of 67 small molecules.[164] Finally, several groups have merged quantum mechanical procedures with computations of the reaction field arising from a continuum, high dielectric, solvent.[165–168]

4. Entropy

The free energy and, thus, the entropy of a *single* conformation is not a well-defined concept in statistical mechanics. However, it is possible to *associate* a statistical weight (and thus, a free energy) to a local minimum-energy conformation, or more generally, to a *region* of conformational space. This statistical weight is a function not only of the potential energy (i.e. the depth of the minimum) but also of the conformational entropy, which depends on the shape and width of the bottom of the potential well. The conformational entropy arises from the fluctuations of the independent coordinates used to describe the spatial structure of the molecule. Another contribution may arise from the distribution of the polypeptide among various local minima in the same state, i.e. from the existence of conformational microstates.[5,7] These entropic effects occur above and beyond those entering from the possible use of temperature-dependent potentials of mean force as discussed in the previous section. For thermodynamic analysis, what follows is equally applicable to a pure intramolecular potential energy as well as to the augmented functions that include implicit solvation.

The free energy contribution of *small* fluctuations in dihedral angles (or other variables) around an energy minimum can be computed by means of a harmonic approximation. The normalized statistical weight w_i, which expresses the probability of the ith conformation, is given by[110,169,170]

$$w_i = (1/Z)(2\pi RT)^{k/2}(\det \mathbf{F}_i)^{-1/2}\exp(-\Delta U_i/RT) \quad (13)$$

where ΔU_i is the conformational energy at the ith minimum (relative to the lowest energy), R is the gas constant, T is the temperature, k is the number of variable dihedral angles (degrees of freedom), and $\mathbf{F}_i$ is the matrix of second derivatives of the energy[110] at the ith minimum. The partition function Z is given by

$$Z = (2\pi RT)^{k/2}\sum_{i=1}^{N}(\det \mathbf{F}_i)^{-1/2}\exp(-\Delta U_i/RT) \qquad (14)$$

where N is the number of low-energy minima (say, ΔU < 3–5 kcal/mol).

The conformational free energy and the relative free energy at the ith minimum are defined as

$$G_i = -RT\ln w_i \qquad (15a)$$

$$\Delta G_i = G_i - G_0 \qquad (15b)$$

where G_0 is the free energy of the conformation of lowest potential energy (i.e. the one at $\Delta U = 0$). The relative entropy is

$$\Delta S_i = (1/T)(\Delta U_i - \Delta G_i) \qquad (16)$$

This is equivalent to the definition of librational entropy given by Gō et al.[110,169,170] This method has also been generalized for ring structures.[171]

The method outlined above can assign a reasonable value of conformational entropy to a conformation if the fluctuations around the energy minimum are small enough, so that the harmonic approximation will remain accurate. The basic idea of the method has also been adapted to the computation of entropy differences between various conformations by means of simulation techniques,[172,173] in a manner that removes some of the limitations of the original approach. Most notably, these authors suggest that, at room temperature, the conformational fluctuations around a minimum energy conformation can be highly unharmonic, and thus, the use of just the second derivative will underestimate the entropy. They suggest the use of simulation data to construct a covariance matrix to serve as an effective harmonic, but temperature dependent, replacement of the inverse of the second derivative matrix:

$$S_i = \ln(\det \sigma) + C \qquad (17)$$

where C does not depend on the coordinates, and the covariance matrix collects the fluctuations of the coordinates q

$$\sigma_{kl} = \langle (q_k - \langle q_k \rangle)(q_l - \langle q_l \rangle) \rangle \qquad (18)$$

Clearly the magnitude of these fluctuations depends on the temperature at which the simulation is carried out; i.e., $\sigma = \sigma(T)$. This treatment reduces to the Gō–Scheraga procedure when the temperature of the simulation is very low, or when the energy surface is indeed purely harmonic. The analysis of Karplus and Kushick still relies on a harmonic form of the fluctuations; hence, Di Nola et al. generalized their approach by collecting full statistical distribution information about the fluctuations of the coordinates during a simulation.[174] In the approach of Di Nola et al., correlations between variables, corresponding to the off-diagonal elements of the covariance matrix, are still treated as effective harmonics:

$$S_i = -R \sum_k \int \mathrm{d}q_k \, P(q_k) \ln P(q_k) + \frac{R}{2} \ln\left(\frac{\det \sigma}{\prod_k \sigma_{kk}} \right) \qquad (19)$$

where $P(q_k)$ is a histogram of the variable k obtained from the simulation. Rojas et al. have proposed another procedure to correct the quasiharmonic approximation of Karplus and Kushick, by self-consistently incorporating higher order correlations or moments of the coordinate fluctuations into the entropy expressions.[175]

Similar procedures as those outlined above have been used to analyze normal mode vibrational motions of a protein[176] and to estimate the vibrational entropy of folded proteins.[177]

All the above approaches tend to break down when the definition of a "conformation" reasonably includes more than one local minimum. An attempt to address this issue is the scanning simulation technique,[178-180] and the related local states method to calculate the conformational entropy from a sample of conformations generated by either Monte Carlo[181] or molecular dynamics simulation techniques.[182]

For the limit of very flexible chains, a reasonable estimation of the conformational entropy can be obtained by neglecting long-range interactions completely.[183-186] This limit is often used to estimate properties of the unfolded or statistical coil state of globular proteins.

The methods discussed thus far give *relative* estimates of entropy and free energy. An interesting method to estimate the absolute entropy, is the procedure of Stoessel and Nowak.[187] These authors use a harmonic system as a reference and compute free energy and entropy by stepwise perturbation of this system (whose thermodynamic quantities can be evaluated analytically) to the system of interest. This algorithm, thus, is closely related in concept to the free energy perturbation simulation techniques (reviewed by Kollman[72]).

B. Treatment of the Multiple-Minima Problem

According to the thermodynamic hypothesis that followed from Anfinsen's experiments on ribonuclease,[188] a stable protein conformation corresponds to the lowest minimum of the empirical conformational "free" energy (including the effect of hydration), termed the global minimum. The multidimensional surface that describes the energy of the polypeptide chain as a function of the internal variables, however, has an astronomically large number of local minima. Except for very simple systems with only a few variables, e.g. oligopeptides with less than about four residues (without explicit solvent and in the rigid-geometry approximation), it is not possible to explore the entire surface systematically in order to find the lowest minimum. Therefore, a major problem of conformational energy computations remains the efficient search of conformational space to locate the region that contains a potential well around the global minimum, or at least reduce the number of low-energy regions under consideration to a computationally manageable number.[189] Once such a potential well is located, it is possible to determine the position of its local minimum efficiently, using any of a great variety of function-optimizing computer algorithms.[190] Alternatively, if one is able to select a relatively small number of different low-energy regions to be considered further, energy minimization can be carried out for all of them, and the relative energies of the minima can be compared directly. It should be noted that, for small, flexible, peptides, finding and analyzing, in principle, all the relevant, low-energy, local minima, in addition to the global one, is also an important part of a correct description of their conformational behavior.

In an interesting theoretical analysis of the folding process of proteins, Zwanzig et al.[191] have studied the Levinthal paradox,[12,13] and pointed out that a biased random search, with a small but realistic bias favoring correctly folded local states, can enormously reduce the number of conformations that have to be explored by the polypeptide chain during folding. Possible numerical evidence for this proposal is contained in the lattice simulations of Skolnick and Kolinski[192] described in section III.E.3. On the other hand, it remains to be seen whether the results of the analysis of Zwanzig et al. can be exploited in practice in the design of general computational algorithms for search of folded structures. Recently, the stochastic model of Zwanzig et al. has been used to analyze the time scale of protein conformational changes.[193]

A large variety of procedures have been developed to search multidimensional conformational space, in order to focus the effort in the region containing the global energy minimum, and thereby to provide an approximate structure to which local energy minimization can be applied. Usually these procedures are based on one or more of the following approaches: (a) reorganization and breakdown of the computation into smaller steps, (b) the use of statistical or heuristic information, derived from known protein structures, (c) a simplified formulation of the potential, (d) stochastic search procedures using Monte Carlo or molecular dynamics, or (e) a mathematical transformation of the potential surface. Some of these procedures are described here. In most instances, atomic level descriptions of globular proteins cannot be used because of their large size; thus, either an atomic level model of a smaller peptide (or a much simplified model of an entire protein) tends to be the test system of choice.

1. Chain-Growth Procedures

The most direct conformational search method would be a systematic and exhaustive enumeration of all possible conformations, perhaps given a finite discretization of each variable (for example, see the early work of Némethy and Scheraga[18]). Clearly, this procedure is practical only when one has a very small number of variables, or when there is a sufficient number of constraints. The build-up method[194,195] starts conceptually as a truncated systematic search. The first approximation is that the possible states of each individual amino acid are fairly represented initially by a small set of single-residue low-energy minima. The number of these, in the ECEPP description, varies from less than 10 for alanine or proline, up to a few hundred for lysine or arginine.[196,197] Even this approximation could not prevent an exponential explosion in the number of possible states as one constructs a long polypeptide; thus, the second key assumption is that it is possible to ignore partial conformations that have energies higher than some preassigned value above that of the current lowest one; i.e., by using an energy cutoff.

In the usual version of the build-up procedure, the polypeptide is broken into small fragments for each of which the number of conformations is sufficiently small, viz. at most a few hundred or thousand, and the energies of all these fragments are minimized. These small fragments are then combined into larger ones, and the energies of the conformations of the latter are in turn minimized. At each stage, an ensemble of low-energy structures, within a cutoff of the global minimum, is retained. The method is based on the assumption that short-range interactions play a dominant role in determining the conformation of a polypeptide or protein.[5,189] As the fragments become larger and larger, more and more of the long-range interactions are built into the computations. Detailed descriptions of several versions of the method have been given.[198-200]

As detailed in section III, this procedure has been applied in a number of cases. The method appears to work well for small oligopeptides, and for very special cases of larger ones, but its application quickly becomes unmanageable for polypeptides with 10 or more amino acids in the absence of constraining information.

A modified approach to the build-up procedure relies on a generalization of dynamic programming.[201,202] It makes use of the combinatorial optimization of a finite number of discrete states that correspond to local minima of single residues.[197] Local optimization is not always applied at every stage of polypeptide chain growth, and this results in considerable speed—up over the usual build-up implementations.[201,202] As in the build-up procedure itself, however, long-range interactions cannot be introduced at an early stage, and thus the procedure will not work when these interactions eventually overwhelm short-range ones. This is a feature of all build-up-like or dynamic programming methods and limits their applicability in unconstrained problems to relatively small oligopeptides, and to polypeptides interacting with a mostly rigid protein (for an example, see the GROW procedure for analysis of peptide—protein complexes[203]). Pincus and co-workers have used their version of the method extensively and their work has recently been reviewed.[200]

In similar work, Marshall and co-workers used a constrained systematic search[204] to derive conformations of the cyclic undecapeptide cyclosporin A from NMR information.[205] Their computations show how elimination of incompatible conformations early during the search can result in a practical procedure, if enough constraining information is available. Applications of the build-up method, including a combination with a version of the variable target function procedure,[206] have also been made to small proteins using distance information derivable from NMR experiments.[207,208] Bruccoleri and co-workers report using a procedure similar to the build-up in their program CONGEN to treat protein loops too long to be considered by a true systematic search;[209] further description of this work is given in section III.E.1.

Another way to use step-by-step procedures applicable to analysis of relatively long polypeptide models is the replacement of exhaustive enumeration by Monte Carlo procedures. The scanning method of Meirovitch already mentioned in section II.A.4 is an example of such a procedure.[178,180] A similar method, which, like the scanning procedure, is also based on early work of Rosenbluth and Rosenbluth,[210] has been used in studies of lattice proteins.[211,212] Premilat has also presented a Monte Carlo step-by-step procedure with different treatments for short- and long-range interactions.[213] Finally, Garel and co-workers have developed a novel chain-growth Monte Carlo method that appears to compare favorably with molecular dynamics studies on the same polypeptide system.[214-216]

2. Deformation plus Minimization Procedures

Direct application of Monte Carlo methods to polypeptide systems has been limited by the strong coupling between variables and by the highly anisotropic surfaces typical of most commonly used force fields. These features force the use of very small (torsion angle or Cartesian coordinate) steps if a practical acceptance rate is to be achieved. In order to enable use of very large steps, Li and Scheraga introduced the Monte Carlo minimization (MCM) method[217,218] in which a large random deformation is followed by local minimization of the potential energy, and by application of the Metropolis test[219] for acceptance or rejection of the new conformation. The deformation usually consists of a variation of one dihedral angle chosen over the entire 0 to 2π range; but two or more dihedral angles may be altered with decreasing probability. The procedure is then iterated. Thus, this method carries out an *approximate* Metropolis walk over the discrete space of local energy minima, rather than on the continuum representing the entire conformational space. Because the probability of reaching a particular energy minimum depends on the shape of its energy basin and not only on the energy value at the bottom, MCM does not generate a rigorous Boltzmann sample in the discrete space of local minima.

Saunders independently described a method called stochastic search, in which large deformations in Cartesian coordinates are also followed by local minimization; this work does not make explicit reference to Monte Carlo or Metropolis methods.[220] The random incremental pulse search (RIPS) method of Ferguson and Raber[221] is conceptually very similar to the stochastic search method of Saunders; they have also applied it to a number of organic molecules. More similar to MCM is the procedure of Chang et al.;[222] in this method, large deformations are made to dihedral angles but local minimization is carried out in Cartesian coordinates using a fully flexible-geometry potential function. A comparison of the performance of some of these procedures for calculations on cycloheptadecane was conducted, and the deformation-plus-minimization methods seem to exhibit the best behavior.[223] Most of the published tests of these three methods have been carried out on hydrocarbon molecules; however some applications to polypeptides have been reported.[224]

A generalization of the original MCM method for calculations involving isolated peptides as well as peptide–protein complexes has been reported by Caflisch and co-workers.[225,226] These authors have also exploited the use of variable temperatures (see section on simulated annealing, section II.B.3) for a further speed-up of their procedure. As in the work of Chang et al., they carry out large deformations in torsion angle space, but conduct the local minimization stage in full Cartesian coordinate space.

One possible criticism of the original MCM method is the totally random fashion in which new deformations are chosen. This has been addressed by Ripoll et al.[227–229] who use the self-consistent electrostatic field (SCEF) procedure of Piela and Scheraga[230] as a source of advice for large torsional angle deformations in their electrostatically driven Monte Carlo (EDMC) method. The choice of dihedral angle jumps is based on the assumption that low-energy conformations must also have favorable electrostatic interactions. Therefore, the peptide dipoles should be optimally oriented in the local electrostatic field created by the rest of the polypeptide molecule in its current conformation. If the dipoles are not aligned, then their orientations are changed in a way that will decrease the electrostatic energy within the mean-field approximation. Since this involves a local movement, it is computationally very fast. In EDMC, the SCEF diagnosis is combined with random sampling and thermalization, and, as in MCM, deformed structures are energy minimized before application of the Metropolis test. An alternative way to bias steps has been described by Abagyan and Totrov,[163] who used distribution functions derived from analysis of protein structures to sample preferentially in well-populated areas of $\phi-\psi$ and χ space.

In summary, MCM, RIPS, and related procedures combine the global sampling features of stochastic procedures, which makes them less likely to become trapped in local energy minima, with the excellent local exploration properties of numerical minimization methods, which move quickly to the bottom of the energy well, without wasting time in fluctuations inside the well. In this sense, the earlier GLOBEX method of Robson and co-workers,[231,232] which combines nonlinear SIMPLEX with local minimization, presents some similarity to this class of methods.

The implementation of the simplest of these methods, e.g. MCM, is quite straightforward, and generalization for use with a variety of systems has been carried out. Their major limitation remains the rapid increase of computational resources as the size of the system increases. Just the time needed for each local optimization increases at least as the square of the number of variables describing the system (this assumes use of a nonbonded distance cutoff for large systems, otherwise, this increase would vary at least with the cube of the system size), while the number of steps needed for global convergence is also expected to increase with system size.

3. Simulated Annealing

Simulated annealing techniques[233–236] have been applied in numerous studies to optimization problems. In its usual implementation, as an extension of the Metropolis Monte Carlo algorithm,[219] the procedure involves Monte Carlo searches at an initial elevated temperature, with subsequent similar searches at progressively lower temperatures.[236] The key feature of the method is the existence of a cooling schedule, that is, of a method to manipulate the values of successively lower temperatures as the algorithm progresses. Most often, this cooling schedule amounts to a gentle lowering of the temperature that, if carried out appropriately, assures one that statistically the system will eventually become trapped ("frozen") into the conformation of lowest energy. For this purpose, simulated annealing is an improvement over a constant-temperature Metropolis Monte Carlo search, because it biases the acceptance criterion in a way that favors convergence to the global minimum. In its simplest versions, the method avoids time-consuming energy minimization steps. The main weakness of this method, for the purpose of global optimization, is that its performance depends strongly on the choice of the cooling schedule. The determination of an optimal cooling schedule is very problem dependent, and although useful solutions have been found for certain types of optimization problems,[237] its tuning remains a matter of trial and error.

The simulated annealing technique can also be used in the context of molecular dynamics (MD) methodology. Here, the temperature enters the calculation through the kinetic energy. In more detail, the atomic velocities have to conform to a Maxwell–Boltzmann distribution at a given temperature. Several algorithms have been proposed to carry out the thermal coupling; perhaps one of the simplest and most commonly used is that of Berendsen et al.;[238] a comparison of different methods to accomplish this coupling was presented recently.[239] An embryonic version of the simulated annealing technique in molecular simulations can be found in the quenched dynamics protocols of Levitt[39] and of Brooks et al.[35] In these early applications, MD runs were started at room temperature, but then the system was cooled to close to 0 K. The observation was then made that this protocol enabled the system to escape from shallow local minima and that runs starting from the same initial conformation eventually led to lower energies than to those obtained by direct energy minimization.

More recently, temperature-setting protocols have become more complex, and the initial temperatures are now in the thousands of degrees Kelvin. One of the most successful applications of dynamic simulated annealing (DSA) has been in macromolecule calculations using restraints derived from X-ray crystallography or from NMR.[240–242] Most applications of DSA start the calculations with an initial guess already close to the expected correct structure; hence, the role of the procedure is mostly refinement. However, the power of the method has been demonstrated in two sets of calculations that used NMR-derived distance restraints, and that started from random structures. In the first,[242] randomly generated polypeptide chains could be refined by a procedure that combined DSA with dynamic adjustment of the objective (energy plus restraints) function (this is similar in concept to the variable-target function optimization method[206,243]). In a second[244] even more impressive test, the starting "conformation" was a random arrangement of atoms, and again a DSA protocol succeeded in producing conformations of low energy that also satisfied the input restraints (this procedure has also served as the basis for model building operations useful in modeling proteins by homology; see section III.E.1). Because the relative ease with which DSA protocols can be implemented in packages in general use such as AMBER, CHARMM, DISCOVER, and others, this method has been used as a general "global" optimization tool in a large number of studies. A number of these are reviewed in section III.B.

Using the Metropolis algorithm, a number of researchers have carried out simulated annealing calculations of peptide conformation.[245–248] In an application to enkephalin, however, the annealing trajectory did not necessarily proceed toward the global minimum.[249] Possible reasons for the failure of the method, in problems in which restraining information is not available in advance, have been discussed elsewhere.[249] However, one sometimes sees qualitative agreement between simulated annealing and more sophisticated techniques, often at a much lower computational cost for the former.[250] Moreover, it should be noted that several SA calculations[250–252] of [Met]enkephalin have used slightly different versions of the ECEPP potential and/or different sets of variable dihedral angles (e.g. keeping ω's fixed[250]). When the conformations reported by these authors are reminimized using a program entirely compatible with that used by Nayeem et al.,[249] we find that the correct global minimum is achieved in each case (for the case of fixed ω's, this "global minimum" is the same as the lowest energy structure obtained by several runs of MCM using the same set of variables; M. V. unpublished). Recently, a new Monte Carlo procedure, consisting of simulations in the multicanonical ensemble,[253] has been applied to the enkephalin problem using the ECEPP potential;[254] as in the SA calculations of Okamoto et al.,[250] the method converged to conformations that upon minimization led to the same global minimum found by MCM (keeping ω's fixed in *all* stages of the algorithm); in addition, the new method enables estimation to be made of thermodynamic functions.

A number of new versions of the Metropolis Monte Carlo procedure have been applied to biomolecular calculation problems. These new procedures have in common inclusion of techniques that enable optimization of the direction and magnitude of each proposed step. Bouzida et al.[255,256] introduced the acceptance-ratio method (ARM), and the dynamically optimized Monte Carlo (DOMC) procedure. With ARM, the magnitude of the maximum step size is adjusted depending on the acceptance rate monitored over a number of moves; with DOMC, it is possible to include directionality in the procedure according to the local anisotropy. The authors have applied the method to polypeptides, conducting moves directly in Cartesian coordinate space. Their new procedures seem to be more efficient in exploring conformational space than molecular dynamics. If of general applicability, this is an important result since the influential work of Northrup and McCammon[257] has been taken for years as convincing evidence that MC procedures could not be competitive with MD in Cartesian space simulations of complex biomolecules. More recently,[256] Bouzida et al. have combined their ARM and DOMC improvements with simulated annealing to conduct conformational searches of the small protein glucagon.

Higo et al. have proposed an extended simulated annealing process (ESAP) to obtain conformational ensembles of peptide loops in a protein.[258,259] This procedure combines simulated annealing with the scaled collective variables method of Noguti and Gō.[260] The original method used a fast analytical calculation of the second derivative of the energy with respect to torsional variables, to deduce optimized normal-mode like variables along which optimized steps can be taken. In this manner, one can consider the local anisotropy of the energy surface. In their first applications,[258a] Higo et al. compute the second derivative numerically, but only after every 2000 steps of the simulation. More recently,[258b] they have incorporated an analytical calculation of the second derivatives.

A similar goal is pursued by Shin and Jhon in their high directional Monte Carlo (HDMC) procedure.[261] In HDMC, the matrix that defines the collective variables is not obtained from a second derivative calculation, but instead by the covariance matrix accumulated over a segment of the run. As in ESAP, this results in highly optimized anisotropic steps. The method, used with simulated annealing, was applied to an investigation of [Met]enkephalin described by the AMBER force field. In spite of the use of a different search procedure, and most importantly a different force field, their best conformation was reported to be similar to that of the global minimum found by Li and Scheraga using MCM.[217] The authors estimate that the sampling efficiency in HDMC is at least 20 times greater than in standard Metropolis Monte Carlo.

Both HDMC and DOMC appear very similar to each other, and in turn to the extension of simulated annealing to continuous variables first outlined by Vanderbilt and Louie.[234] It is important to point out that some of these procedures may not be consistent with the detailed balance condition sufficient to ensure Boltzmann sampling;[219] by using additional information, the simulation may no longer correspond strictly to a Markov process. Thus, in the absence of detailed balance, some of these methods may or may not lead to a Boltzmann distribution. Bouzida et al. have carefully tuned their procedure in an effort to preserve detailed balance and thus correct sampling.[255] Clearly, if the objective of the calculation is to find a set of very low-energy structures, and not necessarily to generate a correctly equilibrated sample, then this issue is not important. It is also interesting to point out the clear analogy between the HDMC and DOMC methods (and the ESAP) with several of the algorithms for estimation of conformational entropy described in section II.A.4.

Simulated annealing concepts have also been applied successfully in combination with an MCM-like procedure by von Freyberg and Braun;[262] these authors generated an extensive list of low-energy conformations of [Met]enkephalin. They showed that constant temperature MCM was able to find the global minimum consistently, confirming the original work of Li and Scheraga,[217] but they needed to vary the temperature in order to find suboptimal conformations more efficiently. As mentioned in the Introduction, most small linear peptides exist in an ensemble of conformations, which one cannot hope to describe by just examining the structure of the global minimum; thus, this work represents an interesting methodological improvement. Caflisch and co-workers[226] and Abagyan and Argos[56] have also used simulated annealing in conjunction with MCM procedures. In general, the simulated annealing idea can be, and has been, used in the context of methods other than simple Monte Carlo. In some cases, use of annealing, or in general, variable temperature, has led to improved protocols. We close this section by mentioning the work of Snow,[263,264] who has developed a highly optimized annealing/thermalization protocol in a novel scheme to compute the structures of proteins using homology (see also section III.E.1).

4. Use of Statistical Information

In this family of procedures, an attempt is made to select a limited number of starting conformations for energy minimization, by establishing a hierarchy of progressively more complex descriptions of the polypeptide structure. The method used to carry out this selection may be superficially similar to traditional local backbone structure prediction schemes. Thus, they use information derived from the protein structural databank, either in the form of statistical preferences,[265-267] or of rules of pattern matching.[268-270a] It is important to realize, however, that the usual predictions, in addition to being at best 70% accurate,[270b] are not very informative, especially, about what they term the "coil" state. Thus, in the hierarchical optimization procedures reviewed in this section, one attempts to predict values for structurally informative geometrical or topological "variables". These variables may be backbone dihedral angles, interresidue distances, or information about residue–residue contacts.

In the pattern-recognition importance sampling minimization (PRISM) procedure of Lambert and Scheraga,[265-267] the polypeptide chain is built up from probabilities of occurrences in actual regions of the $\phi-\psi$ map. This is divided into four regions (α, ϵ, α^*, and ϵ^*), and all possible three-residue segments from a nonredundant set of X-ray structures are collected and grouped according to each of the possible 4^3 or 64 conformations. The pattern-recognition procedure uses amino acid properties[271] to map peptide sequences into a multivariate continuous property space. These regions are represented by multivariate Gaussian distributions, with parameters derived from the three-residue segments mentioned above. These data are then used to build sequence-dependent probabilities.

In PRISM, the chain is built from the N-terminus, joining the most probable three-residue segment conformations together, allowing for proper overlaps. As the construction proceeds, only the 1000 most probable are retained, up until the end of the chain. Thus, this part of the procedure ends with 1000 different predictions of the backbone structure of the protein. These predictions of conformational regions can be converted to actual sets of $\phi-\psi$ pairs, which can in turn be investigated by standard energy minimization. The approach was applied to the 36-residue avian pancreatic polypeptide. The structure of this molecule is quite simple, consisting of a polyproline-like helix packed against an α-helix, with both helices being connected by a sharp turn structure. The relative success of PRISM in this problem can be traced back to a roughly correct prediction of the location of the turn, and to a mostly correct prediction of the α-helix. The presence of several proline residues in the N-terminal portion seems to be enough to induce a polyproline-like helical structure upon energy minimization. Application of the full implementation of the method (local structure predictions plus minimization) has not been extended to more complex proteins. The dependence on detailed predictions of very local features makes it possible that it may not be useful in the absence of information about long-range interactions.

A different procedure[272] has been described by Head-Gordon and Stillinger in an application of their "ant lion" global optimization procedure (see section II.B.5). Predictions of local structure are carried out by using neural networks, and the results are expressed as angular restraint terms to be added to a standard empirical energy function. This has the effect of deforming the original energy surface defined by the empirical potential in a way that dramatically reduces the number of local minima. Optimization of the smoothed function, followed by eventual minimization of the empirical energy alone, leads to structural predictions. They applied the method to the bee venom peptide melittin. This is a 26-residue polypeptide and its native conformation is an α-helix with a slight kink near the middle.[273] As in the PRISM application, there is very good agreement between calculated and experimental[272] structures, but the reliance on prediction of local features only makes it difficult to see how the approach can be generalized to work effectively with more complex protein folds.

The weaknesses of the two representative approaches described above suggests that perhaps one should explore ways of predicting long-range features of protein structure. Such an approach has been pursued by Cohen and collaborators since the early 1980's in their formalization and extension[274,275] of the "sausage-packing" method of Ptitsyn and Rashin.[276] They developed a series of rules that identify possible sites for packing of α-helices and β-strands. In their first applications, they assumed exact knowledge of the location of regular structure elements, and tested their packing rules and their algorithms to produce low-resolution structures. More recently, they have applied their methods, including full-atom optimization of the best models, for blind predictions of a number of proteins. Although they have not achieved a perfect prediction yet, one can see steady progress in how close they are getting to that goal, when comparing their work on interleukin-2, growth hormone, and interleukin-4.[269,277,278] For the first two, the predictions of the location of the α-helices were far from perfect, and the proposed three-dimensional folds were largely incorrect. For interleukin-4, the location of the α-helices was predicted with almost perfect accuracy, and the differences between the predicted and observed folds were far more subtle: a change in the handedness of the 4-helix bundle.[270a] Interestingly, the initial NMR study could not reliably establish the handedness of the bundle and thus the correctness or not of the prediction.[279] Definitive confirmation had to await more extensive NMR[280,281] and, later, X-ray crystallography[282,283] work. More recently, it has been shown how data derived from mutagenesis and epitope mapping experiments can be used to screen and evaluate three-dimensional models generated by the techniques of Cohen and co-workers.[284]

In summary, the goal of this family of methods is to carry out a global exploration of conformational space using statistical information (or logical rules) derived from analysis of known protein structures; then refine the best choices using progressively more complex energy functions. It seems that, for all but the simplest protein folds, predictions of variables other than (or in addition to) backbone torsion angles may offer better possibilities for success. In this respect, one should mention the efforts of several groups to use neural network and other pattern recognition approaches to predict interresidue distance ranges or contact distance matrices.[285]

5. Surface Deformation

These algorithms are based on a deformation of the original potential energy hypersurface in such a way that high-energy minima disappear, until the deformed surface contains only a single minimum that, in most cases, is related to the global minimum of the original surface. This single minimum can then easily be obtained on the deformed surface by a standard local minimization procedure, starting from any point on the surface. The position of this minimum with respect to the global one in the original hypersurface may have been changed during deformation. Therefore, a procedure may have to be applied in which the global minimum is attained by gradually reversing the deformation.

One way of achieving a smoothing surface deformation transform is by increasing the number of dimensions of the problem. Crippen proposed the energy embedding method in which the minimization problem is solved in $(N - 1)$-dimensional space, where N is the number of atoms in the molecule, and thus, each atom has $N - 1$ "coordinates". In this space, energy functions of the form described in section II.A usually have only one minimum. The tracing back to the original three-dimensional space is done by progressive elimination of the $N - 4$ extra dimensions.[286-288] Several procedures have been used to carry out this reduction of dimensions. In the original description, the force constant-weighted inertial tensor $\mathbf{T}$ of the interatomic separation vectors was calculated. In this procedure, $\mathbf{T}$ is a d-by-d matrix, where d is the number of dimensions, initially $d = N - 1$; $\mathbf{T}$ is defined by

$$T_{ij} = 2\sum_{k=1}^{N} \sum_{l=k+1}^{N} K_{kl}(x_{ki} - x_{li})(x_{kj} - x_{lj}) \qquad (20)$$

where x_{ki} is the ith coordinate of the kth atom and K_{kl} is the force constant for the interaction between atoms k and l. The eigenvalues and eigenvectors of $\mathbf{T}$ are then computed and the coordinates $\boldsymbol{x}_k$ expressed in terms of them. The dimensionality is reduced by setting to zero components corresponding to the smallest eigenvalues (i.e., project these out). Local minimization in the resulting lower dimensional space was then carried out, and the process was iterated until the dimensionality of the problem was three. In other applications,[100] energy embedding was formulated as a constrained optimization problem solved by minimization of an augmented Lagrangian function, which is the sum of the target energy $U(\mathbf{X})$ and penalty function-like contributions of a vector that represents the unwanted coordinates. Crippen and co-workers have also introduced rotational embedding and dimensional oscillation procedures as variations on the general energy embedding theme.[289-291]

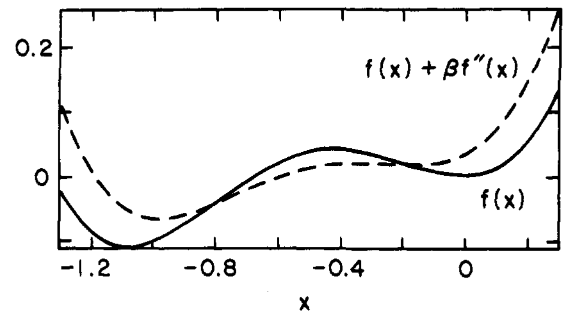


Figure 1. Illustration of the basic principle of the diffusion equation method (DEM). The original double-minimum potential energy curve $f(x)$ (solid line) is transformed, according to eq 21 into a curve $f^{[1]}(x)$ with only a single minimum (dashed line). The values of the transformed function at the inflection points do not change. The particular function used in the figure is $f(x) = x^4 + ax^3 + bx^2$, with $a = 2.0$ and $b = 0.9$. The dashed curve corresponds to $\beta = 0.02$. (Reprinted from ref 296. Copyright 1989 American Chemical Society.)

A similar procedure was proposed by Purisima and Scheraga, using Cailey–Menger determinants as dimensionality-enforcing penalty functions.[292] An interesting combination of dimensional relaxation and variable-temperature molecular dynamics was described recently in the context of refinement of NMR-derived structures.[293] These authors conduct restrained MD in four dimensions (4D-MD) and, after a number of simulation steps, conduct a gradual projection to three-dimensional space using a penalty function and a quenched MD protocol. The use of 4D refinement of NMR structures had been proposed earlier, but only static minimization calculations had been carried out.[294,295]

Another example of surface smoothing is the diffusion equation method, or DEM,[296-298] which is illustrated in Figure 1 by a simple one-dimensional function with two minima. The original function, $f(x)$, can be deformed, in the first iteration, to $f^{[1]}(x)$ by adding its second derivative, $f''(x)$, which is zero at the inflection points, viz.

$$f^{[1]}(x) = f(x) + \beta f''(x) = \left(1 + \beta\frac{d^2}{dx^2}\right)f(x) \qquad (21)$$

where β is a small positive constant. Since the second derivative is positive at minima and negative at maxima, the original curve is smoothed to some extent. Repeated applications of this procedure lead to the following result in the Nth iteration:

$$f^{[N]}(x) = \left(1 + \frac{t}{N}\frac{d^2}{dx^2}\right)^N f(x) \qquad (22)$$

where t/N has been written for β, with the parameter t being positive. Destabilization of the surface is most effective when $N \to \infty$. Taking this limit, we may write

$$F(x,t) = \lim_{N\to\infty}\left(1 + \frac{t}{N}\frac{d^2}{dx^2}\right)^N f(x) = \exp\left(t\frac{d^2}{dx^2}\right)f(x) \qquad (23)$$

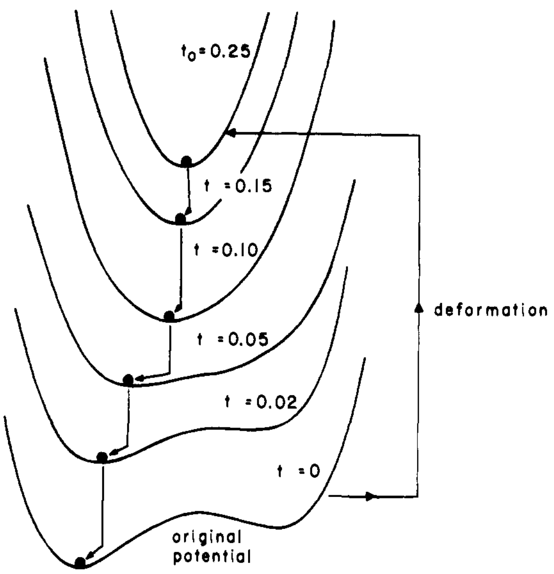


Figure 2. Illustration of the stages of deformation of the function $f(x)$ of Figure 1 according to eq 22, followed by the reversing procedure. The deformation at t_0 leads to the curve on the top with a unique minimum that is attainable from any point in space by a simple minimization. Then, the reversing procedure (shown by the arrows directed downward) is applied by considering a sequence of the deformed curves at the successive values of t indicated. Each step of the procedure is followed by a minimization symbolized by a ball moving downhill from the minimum position of the upper curve and always reaching the position of the minimum on the lower curve. In the final step, the global minimum is found. (Reprinted from ref 296. Copyright 1989 American Chemical Society.)

It can be shown that, equivalently, $F(x,t)$ is a solution of the diffusion equation

$$\frac{\partial^2 F}{\partial x^2} = \frac{\partial F}{\partial t} \qquad (24)$$

where the parameter t takes on the meaning of "time", with the initial condition being $F(x,0) = f(x)$.

In higher dimensions, $\partial^2/\partial x^2$ is replaced by the Laplacian, $\Delta = \sum_{i=1}^{m}\partial^2/\partial x_i^2$, so that the diffusion equation becomes

$$\Delta F = \frac{\partial F}{\partial t} \qquad (25)$$

The successive deformations of the one-dimensional function of Figure 1 from $t = 0$ to $t_0 = 0.25$, and the reversal from $t_0 = 0.25$ to $t = 0$, are illustrated in Figure 2. It can be seen how the global minimum of the original function is attained.

In the diffusion equation method, the original potential surface is the analog of a varying concentration that becomes uniform as $t \to \infty$. Thus, as $t \to \infty$, all minima would disappear, and the surface would become uniformly flat. However, if the deformation procedure is stopped at an earlier time, t_0, then only one minimum (a descendant of the global minimum) remains.

If there are minima with different widths, the method takes the entropic contribution (see section II.A.4) into account correctly in the classical approximation. The minimum that is attained after reversal of the deformation is the one with largest

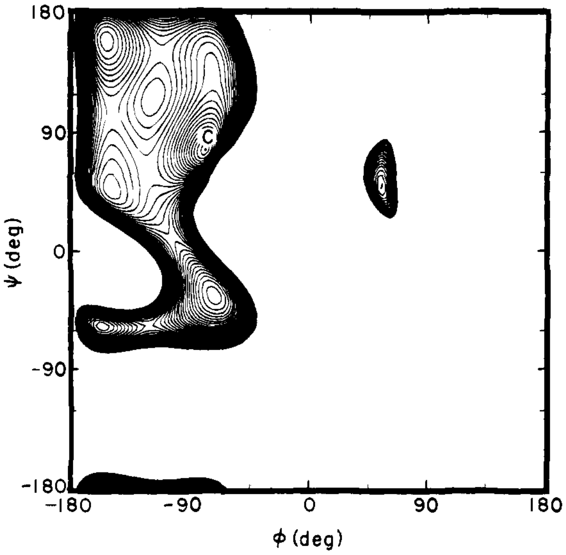


Figure 3. Original (ϕ,ψ) map of terminally blocked alanine, calculated with ECEPP/2, used as the starting point of the deformation procedure ($t = 0$). Point C is the position of the global minimum. (Reprinted from ref 298. Copyright 1992 American Chemical Society.)

statistical weight, not necessarily the deepest minimum, in contrast to most other methods of minimization.[297,298] Therefore, this method locates the minimum that satisfies the thermodynamic hypothesis in the limit of a *correct* potential for a polypeptide or protein in solution.

The method has been tested on various systems. In the case of a cluster of 55 Lennard-Jones atoms, with 159 degrees of freedom, it has been estimated that the total number of local minima is $\approx 10^{45}$. Using the DEM, the global minimum was attained in $\approx$400 s on one processor of an IBM 3090 computer.[297] In an application to terminally blocked alanine, by using the ECEPP/2 potential function, the original (ϕ,ψ) potential surface (Figure 3) was deformed until only one minimum remained[298] (Figure 4). Figure 5 shows the trajectory of the global minimum in the course of the reversal procedure, leading to the global minimum, located in conformational region C (according to the letter code of Zimmerman et al.[196]) of the original surface. Application to [Met]enkephalin[298] led to practically the same global-minimum backbone structure as obtained by the MCM method (Figure 6 and section III.B). The DEM found the global minimum for terminally blocked alanine in <1 min and for the pentapeptide in $\approx$10 min, using one processor of an IBM 3090 supercomputer. Since the DEM scales as n^3, where n is the number of residues, it should take (10 min) × $10^3 = 10^4$ min or $\approx$7 days, to scale up by a factor of 10, i.e. to go from a pentapeptide to a 50-residue protein, using one processor of the IBM 3090 computer. The method has also been applied[299] to the folding of the 36-residue C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*.[300] In a preliminary run on a Fujitsu (FACOM VP 2600) supercomputer, starting from a partially deformed structure that did not resemble the native one, the overall fold of the native protein was obtained. The good performance of this method makes the development of more accurate potentials a more pressing issue.

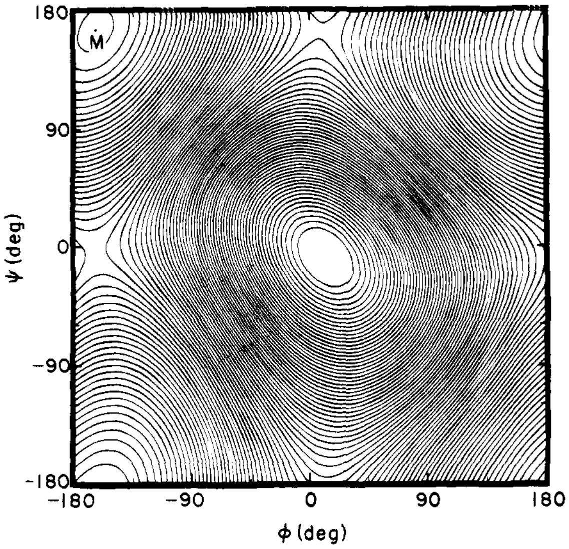


Figure 4. Deformed potential surface for terminally blocked alanine at $t_0 = 10$ Å^2 where only one minimum remains. Point M is the position of the unique minimum. (Reprinted from ref 298. Copyright 1992 American Chemical Society.)

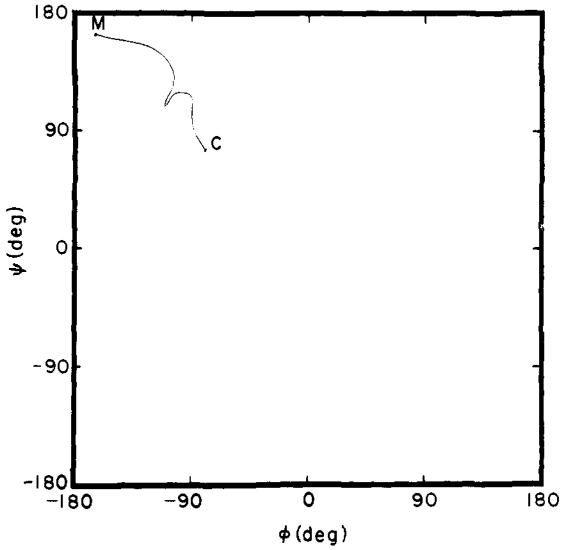


Figure 5. Trajectory of the local minima in the course of the reversal procedure, leading from the unique minimum M at $t_0 = 10$ Å^2 (Figure 4) to point C, which is identical to the global minimum C of the original surface at $t = 0$. (Reprinted from ref 298. Copyright 1992 American Chemical Society.)

The ant lion method is another procedure for deforming the energy surface; here, alteration of the components of the force field has been suggested.[301] In the initial description of the procedure, it appeared that it required prior detailed knowledge of the energy surface.[302] In a later application, however, this apparent limitation was addressed and a test on melittin was presented.[272] (See also section II.B.4.) In the ant lion strategy, the energy surface is deformed by addition of penalty functions that favor a preselected minimum:

$$\Psi(\mathbf{X}) = U(\mathbf{X}) + \sum k_\phi[1 - \cos(\phi - \phi_0)] + k_\psi[1 - \cos(\psi - \psi_0)] \quad (26)$$

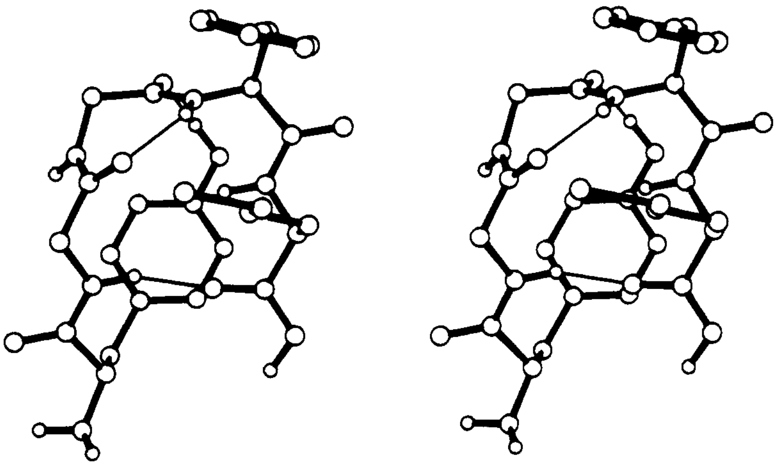


Figure 6. Stereoview of the global minimum-energy structure of [Met]enkephalin in the absence of water. (Reprinted from ref 217. Copyright 1987 Z. Li and H. Scheraga.)

In addition to the backbone dihedral angle restraints, the procedure also utilizes penalty functions that favor hydrogen-bonding patterns characteristic of helices, turns, and sheets. To resolve the apparent circularity in the choice of preselected minima, the ant lion method is adapted to use neural networks as a guide for designing appropriate penalty function parameters from amino acid sequence information alone.[272]

6. Probability Methods

In this type of procedure, one constructs and then examines distribution functions that describe the possible values that important structural variables can attain. Then, an examination of the global energy surface is carried out in order to obtain updated distribution functions. The simplest of these methods is the classical version of the self-consistent mean-field method. Here one computes mean field potentials $V(\phi_i)$ by sampling or enumerating over all variables ϕ_j other than ϕ_i. This process is repeated for a number of possible values of ϕ_i and for all the other variables. Probability density functions can then be derived by way of the Boltzmann law, and the self-consistent protocol for the kth iteration is established as

$$\varrho^{[k]}(\phi_i) = \exp\left(-\frac{V^{[k]}[\phi_i]}{RT}\right)$$
$$V^{[k+1]}(\phi_i) = \langle V(\phi_{j\neq i})\rangle_{\varrho^{[k]}} \qquad (27)$$

where $\langle ... \rangle_{\varrho^{[k]}}$ refers to averaging or sampling using the probability information obtained in the previous iteration. Finkelstein and Reva[303,304] presented a version of this approach that includes nearest-neighbor correlations in a manner similar to helix–coil transition theories.[305] Thus, their probability functions for the variable i are contingent on the state of variable $i - 1$. They applied their approach to study the ways in which an amino acid sequence could be mounted on a preexisting three-dimensional template.[303,304]

Rabow and Scheraga recently outlined a lattice neural network minimization procedure[306] that is virtually identical to the classical self-consistent scheme outlined above. Their key equations (eqs 11 and 12 in the original reference) are

$$V_{xi} = \frac{e^{-U_{xi}/T}}{\sum_j e^{-U_{xj}/T}}$$
$$U_{xi} = \frac{\partial E}{\partial V_{xi}} \qquad (28)$$

Here, E is the energy of the system, and (xi) refers to the lattice position x for atom i. If one expresses the derivative for a pairwise potential with $f(xi,yj)$ representing the interaction between two atoms i and j, at lattice positions x and y, respectively, one would obtain

$$\frac{\partial E}{\partial V_{xi}} = \sum_{y,j} V_{yj} f(xi,yj) \qquad (29)$$

Reinterpreting V_{xi} as the probability $\varrho_i(x)$ of atom i being at position x, the identification of $\sum_{y,j} V_{yj} f(xi,yj) = \sum_{y,j} \varrho_j(y) f(xi,yj) \equiv \langle V_i(x)\rangle_\varrho$ with a mean-field potential becomes transparent (an explicit connection between mean-field optimization and neural networks has been given[307]).

A quantum mechanical version of the self-consistent mean field optimization procedure was presented by Somorjai.[308,309] He outlined a possible way in which the method could be used to calculate the three-dimensional structures of proteins, but presented actual numerical results only for one-dimensional tests. Shortly thereafter, Olzewski et al. described their self-consistent mean-torsional field, SCMTF, method.[310-312] Like Somorjai's procedure, this method is based on the idea that the ground-state wave function ψ of a system of nuclei in a molecule spreads over the entire potential energy surface, irrespective of the number of potential wells. Hence, the maximum of ψ^2 should lie close to the global minimum of the potential energy. In contrast to Somorjai's outline, however, their implementation of the method uses the dihedral angles, θ_i, of the polypeptide as the independent variables of the problem. This modification may result in a potentially better scaling with problem size than the Cartesian coordinate-based method of Somorjai.

By application of the variational principle, one obtains a set of N coupled *one*-dimensional Schrödinger equations, one for each $\phi_i(\theta_i)$, where $\Psi(\theta_1,...,\theta_N) = \Pi_{i=1}^{N}\phi_i(\theta_i)$, and N is the number of dihedral angles.

$$\hat{H}_i\phi_i = \epsilon_i\phi_i \tag{30}$$

with

$$\hat{H}_i = \hat{T}_i + \hat{V}^{\text{eff}}(\theta_i) \tag{31}$$

$\hat{T}_i$ is the kinetic energy operator, and $\hat{V}^{\text{eff}}(\theta_i)$ is the potential energy operator. The N-coupled one-dimensional Schrödinger equations are solved iteratively in N-dimensional dihedral angle space. In analogy to the classical procedure of eqs 27, each equation describes the variation of a *single* dihedral angle in the *averaged field* of the others. The Hamiltonian is

$$\hat{H}_i = \frac{-\hbar^2}{2I_i}\frac{\partial^2}{\partial\theta_i^{\,2}} + \hat{V}^{\text{eff}}(\theta_i) \tag{32}$$

where I_i is a moment of inertia, and the effective potential $\hat{V}^{\text{eff}}(\theta_i)$ depends on the mean field created by averaging over the other dihedral angles, $\theta_l(l \neq i)$, according to the probability distribution $\varrho_l^0 = |\phi_l^0|^2$. In order to calculate $\hat{V}^{\text{eff}}(\theta_i)$, a Monte Carlo procedure[310] is used, i.e.

$$\hat{V}_{\text{eff}}(\theta_i) = \frac{1}{M^c}\sum_m^{M^c} V(\theta_1^{m*},...,\theta_i,...,\theta_N^{m*}) \tag{33}$$

where the summation extends over M^c locally minimized trial points in the $(N - 1)$-dimensional space of all dihedral angles θ_l except θ_i. First, each point θ^m in the space is selected randomly according to some preassumed one-dimensional distribution ϱ_l^0. Then, whenever a θ^m is chosen, the potential energy $V(\theta)$ is minimized with respect to *all* θs. The minimization gives the new θ^{m*}, which is then used in the definition of the mean field. Solution of the set of one-dimensional Schrödinger equations gives a new set of ϱ_l^0. The procedure is repeated iteratively until self-consistency of the distributions is achieved.

Application of this method to [Met]enkephalin (with side-chain torsion angles fixed to values corresponding to the global minimum determined by other procedures) located the global minimum in a time comparable to that required by the DEM. The SCMTF method scales like the DEM. Hence, both of these procedures, in principle, should be able to treat implicit or no-solvent models of polypeptides containing about 50 residues in ≈7 days on an IBM 3090 computer or ≈2−3 days on the faster IBM 9000 computer. The SCMTF method has also been applied to homopolypeptides[311] and to melittin.[312]

Given the limited number of applications to date, it is difficult to compare classical versus quantum mechanical versions; nearest-neighbor versus no-coupling, or Cartesian coordinate versus torsion angles as independent variables in the type of methods described in this section. The use of Boltzmann weighting in some of these procedures also opens the possibility of introducing some level of annealing or thermalization,[306,310] again, this has not been explored systematically to date.

A number of other interesting approaches, which are similar to the ones in this section, have been outlined but not applied yet to polypeptide systems. Straub and co-workers have presented two alternative ways of establishing the self-consistency equations for iterative calculation of the probability density functions. One approach[313] is based on an approximate solution of the time-dependent Schrödinger equation, which has the form

$$\frac{\partial}{\partial\tau}\Psi(\mathbf{X},\tau) = -H\psi(\mathbf{X},\tau) \tag{34}$$

where τ plays the role of imaginary time. The formal solution of this equation is

$$\Psi(\mathbf{X},\tau) = \exp(-H\tau)\Psi(\mathbf{X},0) \tag{35}$$

and the authors note the similarity with the equilibrium Boltzmann distribution, $\exp(-H/RT)$, establishing the analogy between imaginary time τ and inverse temperature T. Their approximate solution leads to a self-consistent procedure using a product of single-particle wave functions and requiring evaluation of a mean-field potential. Interestingly, their implementation in terms of Gaussian packets also presents a strong similarity to the diffusion equation method described in the previous section. These authors tested their algorithm on a series of Lennard-Jones N-mer clusters, with N ranging from 2 to 19, and obtained results that appear as good as those of the DEM on the same systems.

The second approach developed in Straub's group uses approximate solutions of the Liouville equation with Gaussian phase packet (GPP) dynamics.[314] They also tested this procedure on Lennard-Jones clusters, and pointed out the formal similarities of this method with DEM, SCMTF, and their own procedure based on the time-dependent Schrödinger equation.

A procedure that appears to combine probability concepts with smoothing of the potential energy surface has been described recently by Shalloway.[315,316] The procedure has been outlined in detail but no numerical applications to polypeptide models have been reported yet.

The locally enhanced sampling (LES) approach of Elber and co-workers[32,317,318] derived from the time-dependent Hartree approximation[319] also presents some similarities with the methods described in this section. In this protocol, a bundle of classical trajectories move on the *average force field* generated by each other. In the original applications,[319,320] multiple copies of a ligand interacting with a protein are generated. The interactions between different copies are turned off, while the protein interacts with the average field of all the ligands. The whole system is simulated by a largely standard molecular dynamics or energy minimization protocol. Using the multiple trajectories, it is thus possible to obtain information that is almost equivalent to that obtainable in many independent simulations, while using computer time comparable to that of one individual simulation. It has been shown that the procedure is not limited to ligand–protein systems, but it can be generalized to parts of proteins such as side chains[317] or even loops.[321] Another extension involves using the mean-field methodology with chemically different "copies", which may be two or more types of different side-chain residue types in an attempt to exploit multiple sequence information in homology modeling.[321]

It should become apparent by now that there is not always a great difference between some of the methods in this section and those in the previous one dealing with deformation of the energy surface. The smoothing properties of the mean-field approaches are well known. In many of these methods one has all but abandoned the idea of a single discrete conformation "traveling" over the energy surface (cf. Monte Carlo and related procedures), and thinks instead of a probability density carrying out this "bundle" trajectory (in surface deformation procedures one can recover the idea of a sharpening probability density by using the Boltzmann law, for example).

7. Genetic Algorithms

Genetic algorithms represent an increasingly popular family of methods that also abandon the idea of a single conformation "trajectory" in exchange for evolution of a population, and depend on an analogy between biological evolution and optimization problems.[322] These procedures have been applied in a number of peptide and protein model calculations, including lattice representations,[323–325] simplified continuum proteins,[96,326–328] placement of side chains on a fixed backbone,[329] and full-atom models of oligopeptides[330] or proteins;[331] a review of applications and methodology has appeared recently.[332]

The main idea of genetic algorithms (GA) is to generate a population of individuals (e.g., conformations) to which one can assign a *fitness score* (a decreasing function of the energy) and some kind of *string* (most often binary) representation. This population then "evolves" by a series of elementary biological-like operations. Operations found in most GA applications are mutation, recombination by cross-over, and selection. Mutation and selection have clear counterparts in, for example, Monte Carlo methods: a mutation is analogous to a deformation in one or a set of coordinate variables; while selection is related to the idea of accepting or rejecting a move. Recombination by cross-over has no obvious counterpart in "classical" optimization procedures, and this operator generates new conformations by mixing parts of two old ones. In most polypeptide applications, these variables are related to torsion angles or some other residue-based local variables; thus, Unger and Moult see this aspect of GAs as implementing an *implicit* build-up procedure,[324] where locally dominant interactions create initiation structures[333] that are later combined with others far along the chain to create globally better solutions. Procedures based on GAs compare favorably with other methods for conformational analysis of small molecules.[334]

C. Simulation

Numerical computer simulation methods are among the most powerful theoretical tools for the analysis of the dynamical behavior of chemical systems. They have been used extensively in recent years for the study of dynamical properties of water and aqueous solutions, polypeptides, and proteins, including the interactions of proteins with water. Detailed discussions of various aspects of simulations are reviewed elsewhere; viz. conformational dynamics and thermodynamics by Brooks and Case,[75] and free energy calculations by Kollman.[72] For further detailed discussions of the methodology, the reader is referred to several books[25,335–337] and reviews.[3,338]

The computation of conformational energies underlies the use of most of the simulation procedures in the study of molecular structure and dynamics. Thus, the discussion, section II.A, of force fields and of interactions with the solvent applies to these studies.

A recent exciting development is simulation of unfolding of globular proteins in explicit solvent baths. At present some of these studies have required somewhat unrealistic settings, such as temperatures above 200 °C, to accelerate conformational changes. Nevertheless, the resulting partially unfolded states seem to have properties compatible with what is expected experimentally.[40,339–343] A novel method to study unfolding has been presented recently by Hao et al.[344] This procedure uses a backward Euler scheme (see below) to integrate the equations of motion, and incorporates a shape- and size-enforcing term to drive the transition.

A major problem in the application of molecular dynamics simulations to many conformational changes in polypeptides and proteins is the time scale. Current numerical methods for integrating the equations of motion proceed in such small time steps that the limitation of computer time precludes realistic simulations of significant conformational changes. The required size of the time step is a consequence of the high-frequency bond stretching and bending modes; thus, a possible method to increase the step size would be to freeze those modes. Fixing of bond lengths has been successfully achieved by use of the SHAKE algorithm, but this increases the time step only by a factor of 2 or so.[345,346] The same algorithm is inefficient in simulations with fixed-bond angles.[68]

One method that may result in faster simulations is the use of multiple time steps. Components of the total energy expected to vary rapidly are recalculated at every time step. Other, slowly varying, terms are assumed to remain constant for a larger number of steps, and thus recalculated with an effectively larger step size. A way of dividing these two types of terms could be to consider bond length and bond angle vibrations, in addition to very short-ranged nonbonded interactions within a very short cutoff, as rapidly varying, and thus computed every one or two femtoseconds. Interactions due to nonbonded terms operating between the very short cutoff, around 4–5 Å, and the more realistic cutoff value, approaching 10 Å, are computed less often (and assumed constant in between). For large systems, the second type of interactions may account for a much larger portion of the computational effort. Thus, from the point of view of computational efficiency, the calculations proceed with "effective" time steps that are 5–10 times longer than usual.[347,348] It should be noted, however, that the earliest algorithms of this type were not rigorously time-reversible and showed poor energy conservation; thus, the resulting trajectories were not very accurate. The most recent formulations have much better properties and represent rigorous, but more efficient, alternatives to standard MD treatments.[348–350]

Several groups have attempted to generalize and implement MD directly in torsional space, but the resulting methods do not seem to have widespread use yet.[176,351–354]

Schlick and co-workers derived an alternative approach that combines an implicit backward-Euler scheme to solve the differential equations of motion, with a method based on Langevin dynamics for achieving thermal equilibrium. The method enables the choice of a critical cutoff frequency, ω_c, such that vibrational modes with frequencies above that value become frozen.[355–357] More recently, Zhang and Schlick[358] have combined normal mode analysis with the backward-Euler implicit Langevin method, in an effort to produce realistic trajectories, while still allowing for the use of larger time steps. The idea of the new method is to treat the high-frequency modes analytically using normal mode analysis, instead of just freezing those modes as done in earlier applications of this methodology. They applied their procedure in a simulation of liquid butane and reported achieving stable simulations with time steps as large as 50 fs. As indicated earlier, Hao et al.[344] have also used the backward-Euler method in a simulation of unfolding of BPTI and achieved a significant degree of unfolding in a relatively short time.

Jain et al.[359] have described a recursive procedure to carry out MD in terms of internal variables; e.g. torsion angles. They solve the equations of motion using a spatial operator formalism originally developed for analysis and simulation of multibody systems as encountered in robotics. This work is mostly an outline and no actual numerical data on polypeptide or other molecular system is presented. Thus, the actual increase in the time step achievable for protein simulations using this method is presently unknown.

A very interesting approach to enforce bond length and bond angle restraints in MD has been introduced by Grønbech-Jensen and Doniach.[360] They extend an earlier Langevin dynamics procedure used for semimicroscopic simulation,[361] and apply it in simulations of an OPLS[42] rigid-valence model of polyalanine. Using this procedure, time steps as large as 1 ps (i.e. a 1000-fold higher than in conventional MD) seem possible.

It is not clear at the present time how general and useful any of these new approaches will eventually become. In particular, it is not known if the resulting dynamics trajectories generated by some of these methods may become too distorted when compared with those produced by more conventional methods.[68] Nevertheless, they may still remain useful as tools for efficient conformational sampling.

A number of variants of MD have also been implemented in ways that are expected to enhance the rate at which conformational space is searched and barriers are crossed. The most common procedure involves non-physical settings of the temperature of the system (described in section II-B-3). Other innovations include potential energy annealing for conformational search (or PEACS),[362] atomic mass weighting,[363] and rush dynamics;[364] in addition to the previously described 4D-MD[293] and LES procedures[317,319] (see section II.B).

Monte Carlo simulation forms the basis of several procedures used to overcome the multiple-minima problem, as described in section II.B. In a related approach, the use of a statistical mechanical procedure with an adaptive importance sampling Monte Carlo algorithm (SMAPPS) has been developed and tested on [Met]enkephalin.[365,366] Other simulation work on peptides using MC includes the calculations of Hagler and co-workers on a number of simple model peptides,[367] and the very long computations of Rapaport and Scheraga on short poly(Ala) and poly(Gly) peptides.[368]

III. Applications

The foregoing methods have been applied to a variety of model structures, small linear and cyclic oligopeptides (as individual molecules and in crystals), fibrous proteins, and globular proteins. Other applications, not reviewed here, include docking (e.g. enzyme–substrate interactions), refinement of X-ray and NMR structures, and the influences of site-specific mutations on protein structure.

A. Model Structures

The main commonly occurring structural features of proteins are the α-helix, β-sheet, and β-turn. Interatomic interactions dictate these features, the packing of these structures, and the transitions between these regular structures and the disordered form, the statistical coil. Conformational energy calculations with interatomic potentials should be able, in principle, to account for the observed structures and stabilities. As discussed earlier, however, a number of approximations are usually made to make the calculations feasible, most importantly incomplete sampling and neglect, or at best approximation, of solvation effects.

1. Regular Structures

a. Handedness of α-Helices. An early application of this methodology was to the computation of the relative stabilities of the right- and left-handed α-helical forms of several homopolymers of amino acids.[369,370] Calculations that did not include solvent effects were capable of estimating qualitatively correct trends in a number of cases. Moreover, the origin of the preference for right or left handedness could be attributed to specific interatomic interactions. As an illustration, in the series of poly(*o*-, *m*-, and *p*-chlorobenzyl aspartate)s, the *ortho, meta* and unsubstituted polymers adopt left-handed α-helical

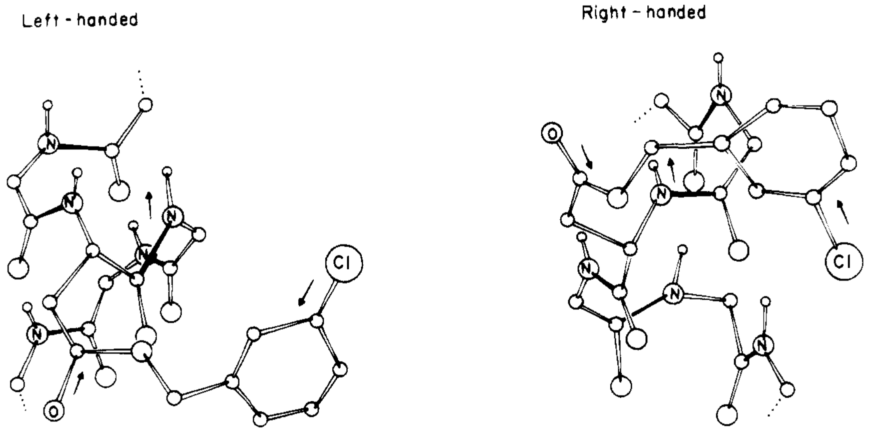


Figure 7. Orientation of the side chains of the left- and right-handed α-helices of poly(*m*-chlorobenzyl L-aspartate). The solid arrows represent the direction of the C−Cl, ester, and amide dipoles, respectively. (Reprinted from ref 29. Copyright 1991 Ellis Horwood.)

forms, whereas the *para* polymer adopts a right-handed α-helical form. In these polymers, the side chain takes on a transverse or longitudinal orientation with respect to the backbone, bringing the chlorine atom close enough to the backbone to influence its helical twist. Figure 7 illustrates the lowest energy conformations of the *meta* polymer, showing a favorable, attractive interaction between the C−Cl dipole and the dipole of the closest peptide group in the left-handed form; the corresponding interaction in the right-handed form is repulsive. Thus, this dipole−dipole interaction plays a dominant role in leading to the preference for left handedness. These preferences have been verified experimentally for all three chloro-substituted poly(benzyl aspartates).[371]

b. 3_{10} to α-Helix Transition. If the α-hydrogen is substituted as, for example, in α-aminoisobutyric acid (Aib), then the conformational energy (ϕ,ψ) map is very restricted, and the preferred form of Aib peptides is computed to be the 3_{10}- rather than the α-helical form.[372] This prediction has been verified by NMR and infrared spectroscopic measurements on solutions of oligomers of Aib.[373] The stability of the 3_{10}-helix for short poly(Aib-L-Ala) polypeptides and the increased stabilization of the α-helical form with a lengthening of the chain has been demonstrated by X-ray analysis of peptide crystals.[374]

The problem of the 3_{10} to α-helix equilibrium has been revisited recently by Tirado-Rives et al.[375] and by Smythe et al.,[376] who used explicit solvent free energy simulations to estimate the relative stability of the two helical forms for several model peptides. These calculations were partially motivated by the surprising findings by Millhauser and co-workers[377,378] on a series of alanine-based peptides long thought to be α-helical in aqueous solution.[379] Presently, there appears to be a discrepancy between the experimental results, based on electron spin resonance (ESR) of doubly spin labeled peptides, which favor the 3_{10} helices for short (16-amino acid) peptides, and the MD and MC results of Tirado-Rives et al., which favor the α-helix for a blocked form of undecaalanine. The origins of this discrepancy are not clear, although the systems studied by each group are not exactly identical, and the equilibrium between the two forms may be due to rather subtle reasons as α-helical character is indeed observed by ESR techniques when the peptide length increases to 21.[378] The calculations by Smythe et al. supplement their earlier results,[380] by including simulations in mixed solvents. Their results also seem to question the interpretation[377] of the ESR experiments. The dependence on chain length may be related to the observation of end effects wherein the α-helix changes its form at the termini.[381]

c. Twist of β-Sheets. Chothia[382] observed that the β-sheets in globular proteins have a right-handed twist. Salemme and co-workers[383,384] accounted for the right-handed twist from studies of models of β-sheets; they attributed the twist to a partially tetrahedral deformation of bonds about the peptide nitrogen atoms that was adopted to diminish the extent of distortion of hydrogen bonds in the twisted sheet. Computations by Chou et al.[385−387] on model β-sheets, e.g. the parallel and antiparallel structures of poly(L-valine) sheets,[386] illustrated in Figure 8, suggest an alternative explanation, in terms of interatomic interactions, for the observed right-handed twist. The conformations of the chains in β-sheets are found in the upper left-hand corner of the conventional (ϕ,ψ) map. In general, side chain−backbone interactions within each strand result in a preference for a right-handed twist for L-amino acids, although there are exceptions. In addition, interstrand side chain−side chain interactions also make significant contributions. Thus, intrastrand interactions in an isolated extended poly(L-Ile) strand energetically favor the left-handed twist, but interstrand interactions result in the stabilization of a poly(L-Ile) β-sheet with a right-handed twist. Poly(L-Ser)[388] is exceptional, in that it is computed to favor a left-handed β-sheet. This prediction is consistent with the observed behavior of Ser residues in proteins: even though Ser occurs relatively infrequently in β-sheets, it usually imparts a local deformation to the polypeptide chain that corresponds to reduced or left-handed local twisting. These calculations did not include solvation effects, and in most cases used idealized sequences; nevertheless, good qualitative correspondence with experimental findings was observed.

The computed energies have been used to predict the relative stabilities of parallel and antiparallel β-sheets of poly(amino acid)s.[29,388] The antiparallel

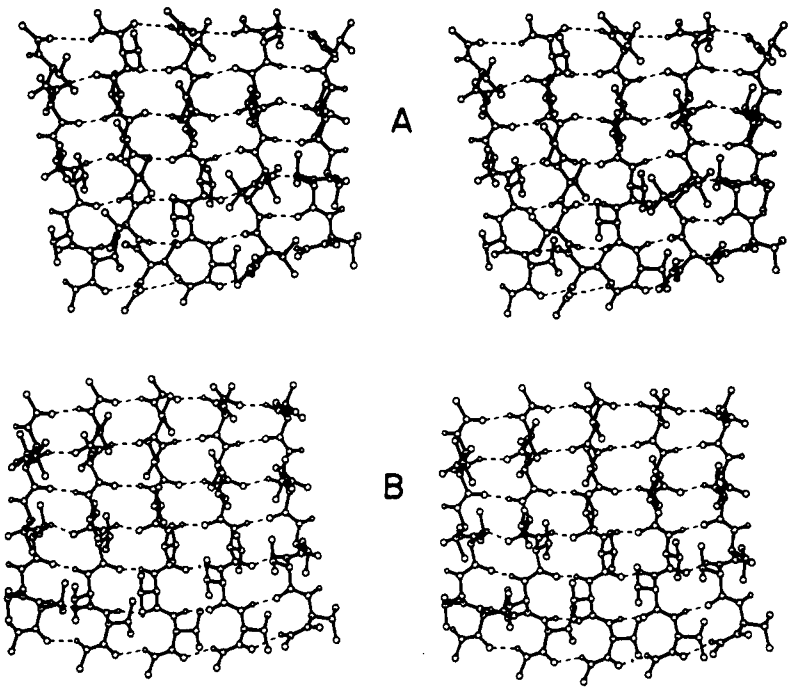


Figure 8. Stereodrawings of the minimum-energy β-sheets with five CH_3CO-(L-Val)$_6$-$NHCH_3$ chains: (A) antiparallel structure and (B) parallel structure. (Reprinted from ref 386. Copyright 1982 K.-C. Chou and H. Scheraga.)

form was predicted to be favored for sheets formed by residues with small unbranched (or γ-branched) side chains (Gly, Ala, Leu), while the parallel form is favored for residues such as Val, Ile, Lys, Ser, Thr, Phe, Tyr. All of these predictions agree with experimental observations on oligopeptides, wherever data are available.[388]

d. Reverse Turns. Another commonly observed local conformational state in proteins cannot be described in terms of a single unique region of the (ϕ,ψ) map. This is the bend or chain reversal, also called a reverse or β-turn, formed by two successive residues. The two residues usually are in two different conformational states, or regions of the (ϕ,ψ) map. Moreover, the dihedral angles in bends can take on a variety of values. Many bends can be classified into 11 types, according to the values of the dihedral angles of the two residues.[389,390] A revised nomenclature, based on analysis of the growing protein structural data base, was proposed recently.[391] Bends usually are defined in terms of the geometry of the backbone chain.[390,392,393] The most common definition requires that the distance R between the C^α atoms of residues i and $i + 3$ (for a bend involving residues $i + 1$ and $i + 2$) be less than a certain distance R_b. This limiting distance was taken as $R_b = 7$ Å in a number of studies. It was shown by Zimmerman et al.[394] that bends are distinguishable in a natural way from other nonregular structures in terms of their distances R and their average relative stabilities. These authors found that the probability distribution function, $P(R)$, for 20 globular proteins has two distinct peaks separated by a minimum at $R \sim 7$ Å, indicating the existence of a natural division between bend and nonbend structures in terms of the values of R (Figure 9). Potential energy computations on a variety of dipeptides showed that the relative stabilities of bend conformations can be correlated with intraresidue interactions.[395] These calculations did not include solvation effects.

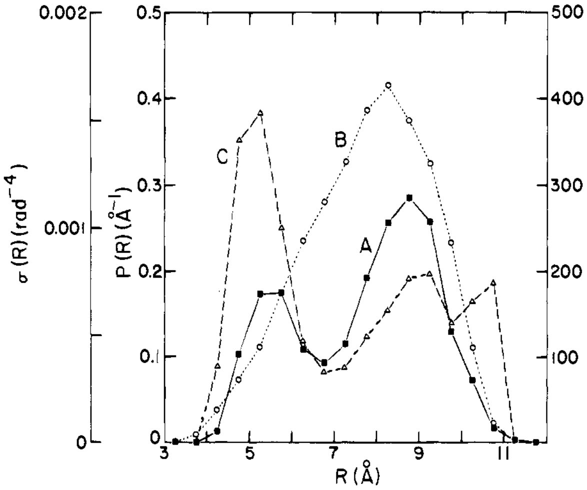


Figure 9. (A) $P(R)$ vs R for dipeptide sequences (excluding those in which both residues are in an α-helix or in an extended structure) in globular proteins; (B) $V(R)$ vs R for N-acetyl-N'-methylalanylalanineamide, from a 30° map over all conformational space, where $V(R)$ is the volume of conformational space which has a value of R in the interval between R and $R + dR$; (C) $\sigma(R)$ vs R obtained by dividing curve A by curve B at each point; i.e., $\sigma(R)$ is a mean probability density distribution function. (Reprinted from ref 394. Copyright 1977 American Chemical Society.)

More recently, Brooks and co-workers have used free energy calculations with explicit solvent to estimate bend stability.[396] Because of the computationally intensive nature of this methodology, they did not explore sequence effects as extensively as in the earlier work. Nevertheless, they concluded that bend conformations are intrinsically unstable in water.[396] They have also carried out a nanosecond-scale simulation[397] of a pentapeptide known[398] to form a high population of bend structures in aqueous solution. There is qualitative agreement between

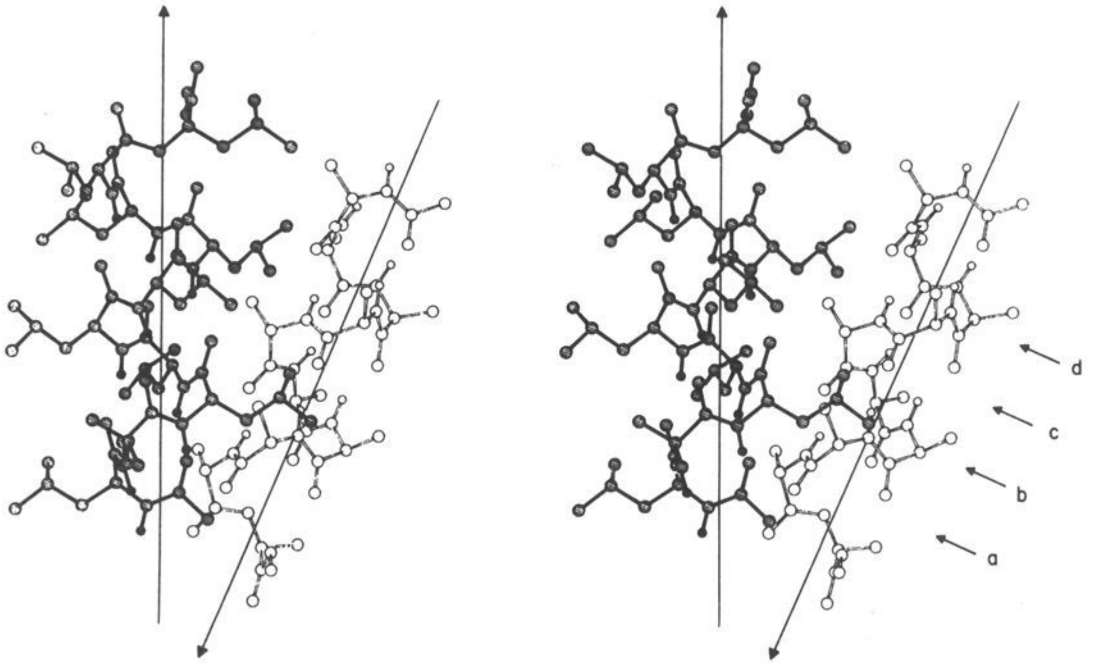


Figure 10. Stereoscopic pictures of an α[CH_3CO-(L-Leu)$_{10}$-$NHCH_3$] (shaded atoms and bonds) and an α[CH_3CO-(L-Ala)$_{10}NHCH_3$] (open atoms and bonds) α-helix in the lowest energy packing state. The two helices are nearly antiparallel. The helix axes are indicated by arrows, with the head of the arrow pointing in the direction of the C-terminus of each helix. Hydrogen atoms are omitted, except for the amide hydrogens. The arrows marked a–d indicate regions in which the surfaces of the two α–helices are complementary. (Reprinted from ref 401. Copyright 1984 American Chemical Society.)

simulation and experiment, but the authors point out the difficulty of carrying out a thorough comparison between simulations and experiment.

These two sets of studies illustrate the conflict alluded to earlier between complex explicit solvent descriptions, where sampling issues are always difficult to address, and simplified models without explicit solvent, where thorough sampling of conformational space may be achieved, but some questions remain about the accuracy of the *total* potential. In either case, it is not always easy to find the most relevant experimental system for validation. As pointed out above (and earlier by Tobias et al.[396]), comparisons of calculations of isolated small peptides with local trends in native protein structures can have only a *qualitative* character. Further discussion of some of these calculations has been given in a recent review by Brooks and Case.[75]

2. Packing of Regular Structures

The computational methodology has also accounted for various types of packing of α-helices and β-sheets: α/α packing, coiled-coil packing of helices, α/β packing, β/β packing, βαβ crossover packing, β-barrels, and so on. Most of the calculations conducted with the ECEPP potentials have been reviewed recently,[387] and only a selection of these are described briefly here.

a. α/α Packing. Two α-helices can pack efficiently against each other in only a limited number of ways, i.e. with a small number of relative orientations of the helix axes. Some of this restriction on the packing arrangements arises from the geometrical shape of the surfaces of the helices.[387,399] In general, the side chains of one helix intercalate into the spaces between the side chains of the other helix. In addition to this geometrical complementarity, however, interaction energies are also important in helix/helix packing. Energy minimization computations, without inclusion of solvation effects, on the packing of two poly(L-Ala) α-helices and of a poly(L-Ala) helix with a poly(L-Leu) helix have shown that only about 10 low-energy packing arrangements can occur in each of these cases.[400,401] The helices are nearly antiparallel in the energetically most favorable packing arrangements. The lowest energy structure, with an orientation angle of about −154° between the helix axes (Figure10), is the most frequently observed α/α packing arrangement in globular proteins.[387,399] Apparently, the basic patterns of packing are established by the overall geometrical and energetic features of the interacting α-helices, even though sequence-specific side chain–side chain interactions may lead to some alteration of the preferences in actual packings found in specific proteins.[402]

The packing orientation and the interhelix energy are influenced by both nonbonded (van der Waals-like) and electrostatic interactions. The electrostatic interactions arise primarily from the net dipoles of the two α-helices. In an α-helix, the individual peptide dipoles are oriented nearly parallel to the axis of the helix. Their alignment gives rise to a dipole of considerable strength.[403–405] As a result, the electrostatically most favorable orientation of two α-helices would be one with an antiparallel orientation of the helix axes. In calculations using ECEPP, the nonbonded interaction (arising from the 12-6 term of the potential) dominates the total energy of an *assembly* of α-helices. At the same time, both types of interactions, nonbonded and electrostatic, contribute comparably to the energy *differences* between various ways of packing, i.e. both participate significantly in choosing preferential orientations.[401] It should be noted, however, that the presence of a solvent with a high dielectric constant (such as water) is expected to weaken electrostatic interactions between α-helices, by reducing the magnitude of dipole interactions, and because of the unfavorable desolvation of the helical dipoles upon association; these effects have been observed in calculations of helix packing energetics using numerical solutions to the Poisson–Boltzmann equation.[406] The presence of water may also provide a stabilizing effect for the association, however, by way of hydrophobic interactions involving buried nonpolar helical faces.[127,274]

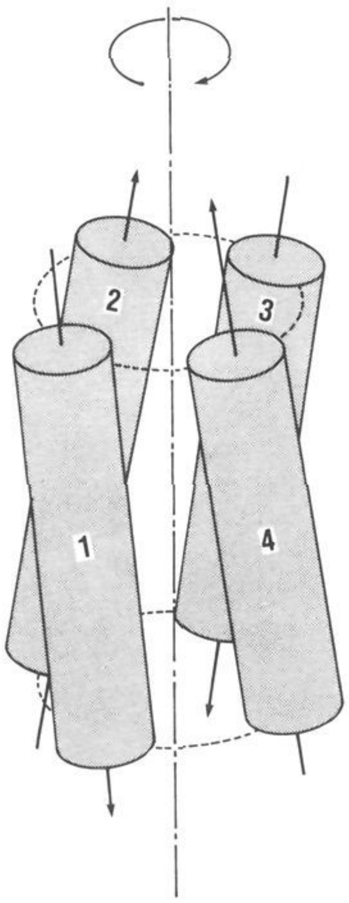


Figure 11. Schematic illustrations of an antiparallel four α-helix bundle. The helices are shown schematically as cylinders. Arrows indicating the helix axes point from the N- to the C-terminus. (Reprinted from ref 409. Copyright 1988 K.-C. Chou et al.)

A bundle of four α-helices is a frequently occurring structural pattern in globular proteins.[402,407,408] The main structural features of this bundle are the near-antiparallel orientation of neighboring pairs of helices and a tilting of the helix axes that corresponds to a left-handed twisting of the entire bundle (Figure 11). Both features are qualitatively consistent with an analysis of the nonbonded and electrostatic interactions between the constituent helices.[409] In a bundle in which neighboring α-helices are sequentially connected by polypeptide links in nonhelical conformations, these links also seem to contribute significantly to the stabilization of the bundle.[410,411] Carlacci et al. have also introduced simulated annealing algorithms in these packing calculations to improve the sampling of different packings;[412] earlier work was limited to local minimization from a series of strategically chosen starting arrangements. Another interesting variation introduced recently in helix packing calculations is the method of Tufféry and Lavery,[413] where side chain-side chain interactions are globally optimized at each tested relative spatial arrangement with a rapid rotamer-based procedure.[329] Improved algorithms for packing nonregular[414] and regular[415] structures by energy minimization have been presented recently.

b. Packing of Helices into Coiled Coils. Several fibrous proteins, such as α-keratin,[416] collagen,[417-419] and tropomyosin,[420] exist in the form of coiled coils that are formed by several closely interacting polypeptide chains. The coiling of two α-helices around a common axis is also the main feature of the leucine zipper dimerization motif observed in a large class of DNA-binding proteins.[421-423]

Coiled coils can be described in terms of major and minor helices. The minor helix is formed by the individual polypeptide chains. The axis of each minor helix is not straight, in contrast to the axis of simple helices, but it follows a helical path, denoted as the major helix, around the common axis of the coiled-coil structure. In simple helices, every residue along the chain must have identical backbone dihedral angles. If the residues within one repeat unit do not all have the same backbone dihedral angles, a coiled coil is formed.[424] An example of this coiled coiling of strands in β-sheets has been mentioned in section III.A.l.

Crick has derived a general formula which relates the parameters of the major and minor helices of a coiled-coil polypeptide chain.[425] Geometrical relations have also been derived between the dihedral angles of polypeptide chains and the parameters that characterize the minor and major helices of coiled coils, together with the establishment of relationships between the major helix and the averaged structure of the minor helix.[424] It has been shown that severe geometrical restrictions exist for the formation of coiled-coil structures that correspond to a given, rigid-geometry, backbone conformation of a polypeptide repeat unit.[424]

Better understanding of coiled-coil α-like helical conformations can be achieved by examination of high-resolution structures (see review by Cohen and Parry[426]). Of particular interest is the structural work on the leucine zipper portion of the GCN4 transcription factor. This apparently simple system has shown the dramatic effect of amino acid changes in the association surfaces in determining the oligomerization state of these peptides. Harbury et al. have found that mutant versions of GCN4 with different sets of hydrophobic residue types (along the hydrophobic face of the helix) can lead to dimers, trimers, or even tetramers, as confirmed by equilibrium analytical centrifugation, and, in some cases, by high-resolution X-ray crystallography.[427] These results extend earlier, surprising, findings[428] with a peptide designed to form dimeric coiled coil structures, and thus expected to be quite similar to GCN4, but which turned out to be a *triple*-stranded coiled coil. This conclusion was based on both high-resolution X-ray crystallography and sedimentation equilibrium measurements; thus, trimer formation is not an effect of crystal packing. These authors also conducted interesting model calculations that indicated that a simple hydrophobic potential[119,429] could qualitatively account for the observed results. It would be interesting to see if this type of calculation could also account for the results on GCN4 mutants obtained by Harbury et al.[427] Slightly more complex calculations had been carried out earlier on GCN4 and on models of other leucine zipper dimers by Krystek et al.[430]

An interesting calculation on the GCN4 system was carried out by Nilges and Brünger,[431] before the high-resolution structural work on the same system had been completed. For these calculations they used the information that the 37-residue peptide would dimerize in a coiled coil form. Thus, while this study cannot be considered a global "folding" of this molecule, the parameters of the coiled coil, such as crossing angle and helix−helix separation, were determined by the computational procedure. Their method included extensive restrained dynamic simulated annealing (DSA) runs without solvation, followed by free MD in explicit solvent.[431] The results have been compared[432] with the subsequently determined X-ray structure,[423] and found to be in very fair agreement: the rms difference was 1.26 Å for backbone atoms. Perhaps more significant, the DSA-MD protocol clearly moved the model coordinates toward closer agreement with experiment, when compared with the initial crude model. The authors conclude[432] that "with the current state of methodology and availability of computing power it is possible to make (accurate) *low-resolution* predictions for structures when the folding motif is known".

More recently, Skolnick and co-workers have approached the calculation of the structure of GCN4 using a hierarchical method[433] (cf. section II.B.4). Their computations started with completely random conformations of the two monomers, and approximately correct coiled coils were obtained after simulations on a lattice (see section III.E.3) that included energy evaluation based on statistical potentials[89,434] (see also section II.A.2). In analogy with the work of Nilges and Brünger,[432] refinement of these crude structures by molecular dynamics, including explicit solvent, led to quite accurate predictions for the structure of GCN4: the rms deviation from the X-ray structure[423] for the backbone atoms was 0.81 Å. It should be stressed that, in this work, the authors did not include *any* information beyond the amino acid sequence (and the knowledge that there is dimerization in the system).

In the triple-stranded collagen molecule,[417-419] the minor helix formed by each polypeptide chain is left-handed. The axes of the minor helices follow right-handed major helices, winding around the major helical axis.[435,436] The coiled coil may be described in terms of the translational repeat D and the azimuthal angular repeat Θ per tripeptide repeat unit along the major helical axis z,[424,435,436] as shown in Figure 12. Furthermore, in triple-stranded (or higher-stranded) coiled coils with screw symmetry, there exist two kinds of disposition of the equivalent repeat units in neighboring strands, viz. "clockwise" and "counterclockwise", depending on whether these repeat units are related to each other by right-handed or left-handed screw symmetry operations, respectively.[435,436]

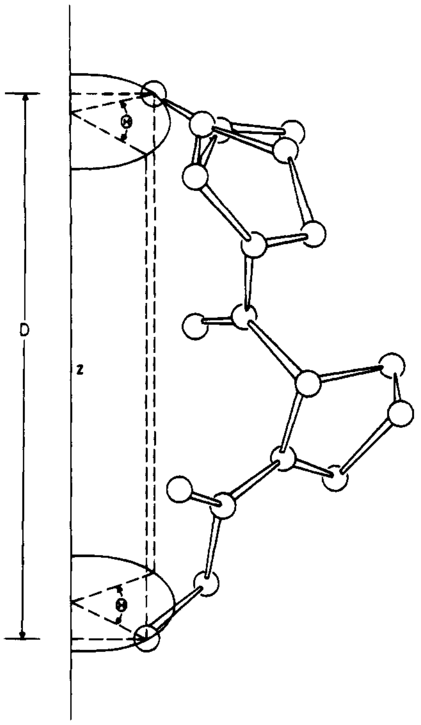


Figure 12. Definition of the coiled-coil helical parameters, using the collagen-like poly(Gly-Pro-Pro) triple helix as an example. One tripeptide unit (Gly-Pro-Pro) is shown in a major helical coordinate system. The major helical axis (z), the translational repeat per tripeptide (D), and the angular repeat per peptide (Θ) are indicated. (Reprinted from ref 435. Copyright 1976 Wiley.)

c. *α/β* Packing. The association of an α-helix with a β-sheet is a frequently occurring structural motif in proteins. The packing is related to the properties of the helix and the sheet. Because of the twisting of the β-sheet, its surface is saddle-shaped, i.e. it can be described as a hyperboloid, while the α-helix is essentially a rigid rod. An energy computation carried out for a poly(L-Val) sheet interacting with a poly(L-Ala) helix predicted that there are four classes of low-energy arrangements.[437] The structures may be characterized by the angle Ω between the axes of the helix and the sheet. In the most favorable arrangements, the helix is nearly parallel or nearly perpendicular to the direction of the strands, because then the helix lies along a tangent line to the curved surface, so that it can interact with the sheet along its entire length (Figure 13). The two remaining arrangements (not shown in Figure 13) can be described as diagonal packings, in which only either the middle or the two ends of the helix are in contact with the sheet (with Ω near −60° or +60°, respectively). Of these two, the first one has a low energy, while the second one is less favorable.

A histogram of observed distributions of Ω in 163 α/β packings in 39 proteins shows a large peak near $\Omega = 0°$, a broad distribution in the range of −30 to -60°, a smaller peak near ±90°, and a few structures near +60°. The positions of the maxima in the distribution correspond to the preferred orientations in the computed structures.[437] As a result of sequence differences and of packing interactions with the rest of the protein molecule, the observed peaks are much broader than the computed distribution for the one computed model structure that was studied by Chou et al,[437] but the grouping into well-defined classes is evident.

d. *β/β* Packing. Because of the hyperboloid shape of twisted β-sheets, as discussed above, two β-sheets can be packed efficiently in only two distinct classes of low-energy arrangements.[438] In the energetically most favored class, the strands of the two sheets are nearly parallel or antiparallel to each other, so that the two curved structures are complementary over most of their surfaces (Figure 14). This class is seen frequently in protein crystal structures, where it has also been termed "aligned packing".[399,439] In the other class, with 1−4 kcal/mol higher energies, the strands are nearly perpendicular to each other, and good packing occurs between the corner of one sheet and the interior of the other sheet. In observed structures

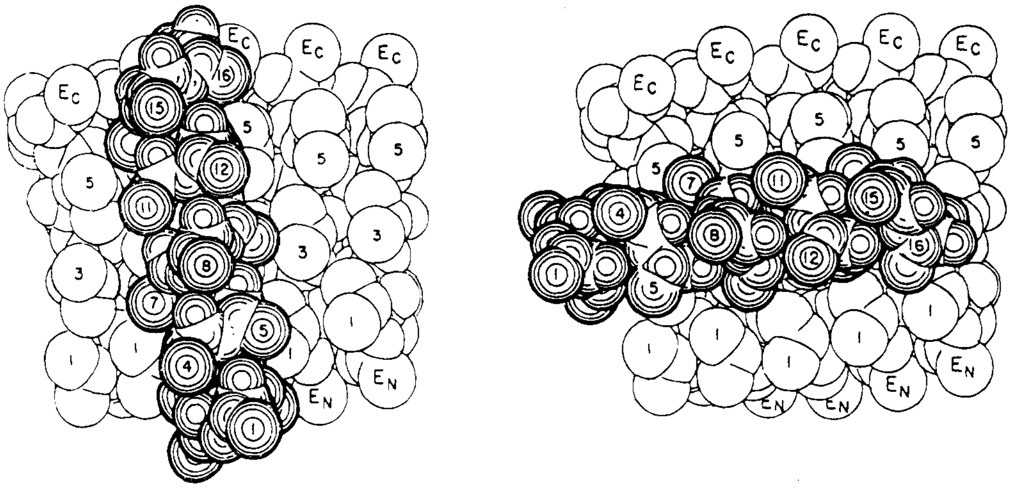


Figure 13. Space-filling representations of a right-handed α-helix of CH_3CO-(L-Ala)$_{16}$-$NHCH_3$ and a right-twisted parallel β-sheet of [CH_3CO-(L-Val)$_6$-$NHCH_3$]$_5$ in low-energy packing arrangements: (left) Ω = 4.4°; (right) Ω = −89.4°. Only heavy atoms are shown, with approximate van der Waals radii. E_N and E_C indicate the N- and C-terminal blocking groups, respectively, of the β-strands. (Reprinted from ref 437. Copyright 1985 Academic.)

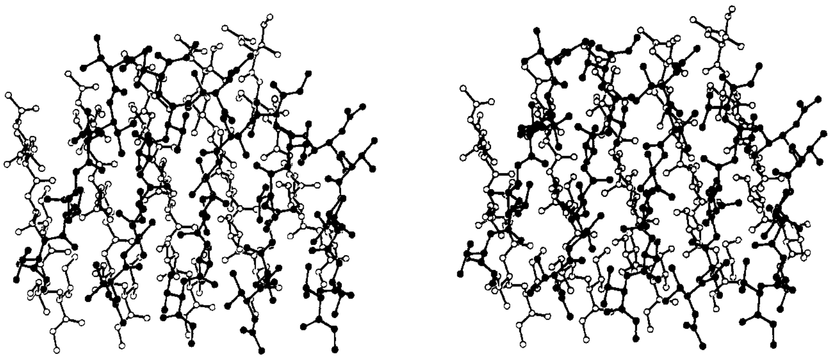


Figure 14. Stereoscopic picture of the low-energy near-parallel (aligned) packing of a [CH_3CO-(L-Ile)$_6$-$NHCH_3$]$_5$ parallel β-sheet (open atoms) on a [CH_3CO-(L-Val)$_6$-$NHCH_3$]$_5$ antiparallel β-sheet (filled atoms). All hydrogen atoms are omitted. (Reprinted from ref 438. Copyright 1986 Academic.)

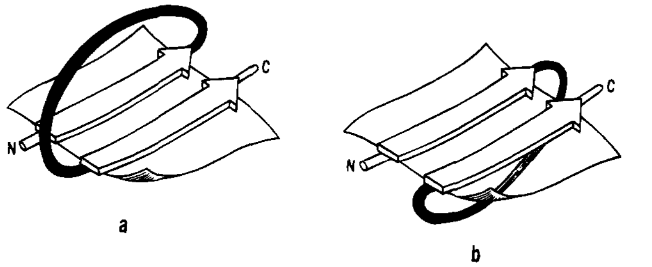


Figure 15. Schematic representation of (a) a right-handed and (b) a left-handed βαβ crossover structure. (Adapted, with permission, from Richardson, ref 408.)

of this type, termed "orthogonal packing", there is usually a covalent connection between a corner of the two sheets.[399,439]

e. *βαβ* Crossover Packing. Two parallel strands of a β-sheet often are connected by a peptide chain that contains an α-helix. This crossover connection could, in principle, be either right or left handed, as shown in Figure 15.[408,440] Actually, crossovers in globular proteins are always right-handed, with very few exceptions.[408,441] This strong preference has been interpreted as a consequence of the right-handed twist of the β-sheet, resulting in a reduction of strain in the connecting chain between the strands for a right-handed crossover.[441] A comparison of the two forms of the crossover by means of conformational energy computations has established that the right-handed form is energetically much more favorable than the left-handed form.[440] Its low energy arises from favorable interactions between the α-helix and the β-sheet, just as in α/β packing, and from the absence of conformational strain in the nonhelical parts of the connecting chain.

This crossover also occurs in many proteins in the doubled form βαβαβ, called the Rossmann fold.[408,442] Although the two α-helices usually connect neighboring pairs of strands in the β-sheet, other connectivities may occur. A computation of the conformational energy of a structure composed of two α-helices and a three-stranded β-sheet, with various connectivities between the five elements, has established that right-handed crossovers are also favored in βαβαβ structures.[443]

f. *β*-Barrels. Two types of β-barrels are observed in proteins, one with parallel and one with antiparallel chains.[408,444] The parallel-chain barrel involves α-helical segments in βαβαβ form, whereas the antiparallel-chain barrel usually does not involve intervening α-helical segments. β-Barrels usually are not straight, with strands running approximately parallel to the axis of the barrel (as in Figure 16a), but the strands are right tilted, as shown schematically in Figure 16b. This tilting is in part a consequence of the intrinsic right-handed twisting of the β-sheet, but the following effect also contributes to

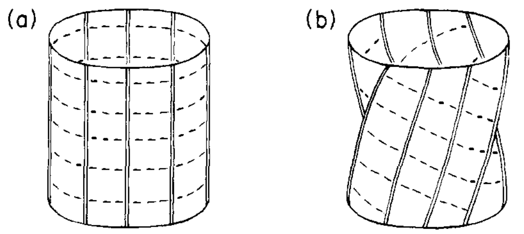


Figure 16. Schematic drawing of an eight-strand β-barrel: (a) nontilted barrel, with strands running parallel to the axis of the barrel, and (b) right-tilted barrel, with strands inclined with respect to the direction of the axis of the barrel. Dashed lines schematically indicate the interstrand hydrogen-bonding pattern. (Reprinted from ref 445. Copyright 1990 Liss.)

it. Tilting improves the packing of the side chains in a β-barrel with bulky side chains, such as Val, on the outside. As a result, the energy of an eight-strand antiparallel β-barrel, in which L-Val and Gly residues alternate along the chains, is 8.6 kcal/mol lower in the right-tilted form than in the absence of tilting.[445] Conversely, left-handed tilting is energetically very unfavorable, because it would require the unfavorable left-handed twisting of the β-sheet.

The computations have also demonstrated the necessity for numerous residues with small side chains or Gly in alternating positions along the strands, because too many large residues cannot be packed into the inside of the barrel or their presence would lead to a severe distortion of the barrel.[445] On the other hand, it is favorable to have large side chains on residues in the alternating positions, i.e. pointing to the outside of a tilted barrel. This is the pattern seen frequently in antiparallel β-barrels in globular proteins.[445,446] Tilting increases the diameter of a β-barrel slightly, so that there is more room to accommodate the internal side chains in a tilted β-barrel.[447]

3. Transitions

Conformational transitions, for example, the helix–coil transition, for which a large literature exists,[305] have also been treated.[29] The helix–coil transition has been studied by statistical mechanics, making use of the one-dimensional Ising model, and has been addressed both for homopolymers[448,449] and binary[450–452] and multicomponent[453–455] random copolymers of amino acids. It has also been treated by molecular mechanics with computational methodology described earlier in this review.[185,456–459]

a. Molecular Theory of the Helix–Coil Transition. A molecular theory has been formulated to compute the phenomenological Zimm–Bragg[448] parameters σ and s of the thermally induced helix–coil transition, using empirical potential functions.[456] This required the computation of the free energy of both the helix and coil forms *in water*.[185,459] The effect of water in these calculations was introduced by a continuum model that considers changes in the number of water molecules in the first hydration layer.[460] The calculation of the free energy of the helix was based on the small-vibration harmonic approximation, and that of the coil was based on the nearest-neighbor approximation, i.e. ignoring long-range interactions (see section II.A.4). The computations were carried out for polyglycine,[458] poly(L-alanine),[458] poly(L-valine)[459] and poly(L-isoleucine).[185] Figure 17 shows the agreement between the computed and experimental results for poly(L-valine), wherein the dominant effect of hydrophobic interactions in the helix accounts for the *increase* in s with increasing temperature. A dominant feature that determines the s vs T behavior in Figure 17 is the *difference* in hydration between the helical and coil forms. Poly(L-valine) and poly(L-isoleucine) differ considerably in their hydration properties, due to the extra methyl group in the isoleucine side chain, so that the s vs T curve for poly(L-isoleucine) has the opposite slope of that of poly(L-valine).[185] Simulations of helix formation that explicitly include a large number of water molecules have been reviewed recently by Hermans.[461] These simulations are carried out on a particular model system, most often a small polypeptide and, because of the high computational complexity, have not included temperature effects. Thus, they are in many respects complementary to the calculations of s in homopolypeptide systems described above.

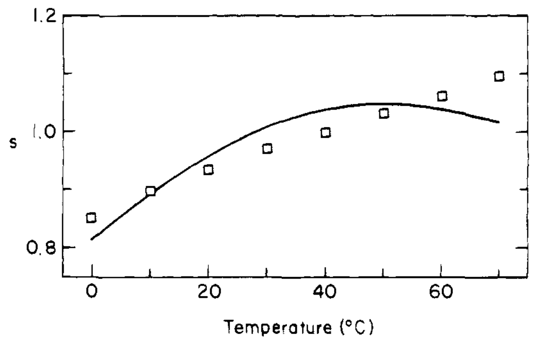


Figure 17. s vs T curves for poly(L-valine) in water. The squares are the experimental results, and the line is the calculated result. (Reprinted from ref 459. Copyright 1974 American Chemical Society.)

In addition to the thermally induced helix–coil transition, the pH-induced transition [in poly(L-lysine)] has been treated by similar computational methods.[462] To obtain a physically realistic estimate of the electrostatic contribution to the free energy, it was necessary to introduce the screening effect of the ionic atmosphere (with a Debye–Hückel screening function). It was thus possible to compute the effect of both pH and ionic strength on the transition curve. For example, the computed midpoint of the helix–coil transition in poly(L-lysine) in 0.1 M salt occurs at a degree of ionization of ~0.5, compared to an experimental value of 0.35.[462] This problem is being reconsidered by making use of recent treatments of the Poisson–Boltzmann equation to compute the electrostatic free energy of a biomolecule in an aqueous solution containing a supporting electrolyte.[158,159,406]

b. Molecular Theory of Helix–Helix Interconversion. The stable crystal arrangements of several homopolymers, including the thermally induced conversion between α- and ω-helical forms of polyamino acids, have been computed.[463,464] These involve intermolecular, crystal-packing degrees of freedom, as well as the internal ones. Whereas poly(p-chlorobenzyl L-aspartate) exists as a right-handed

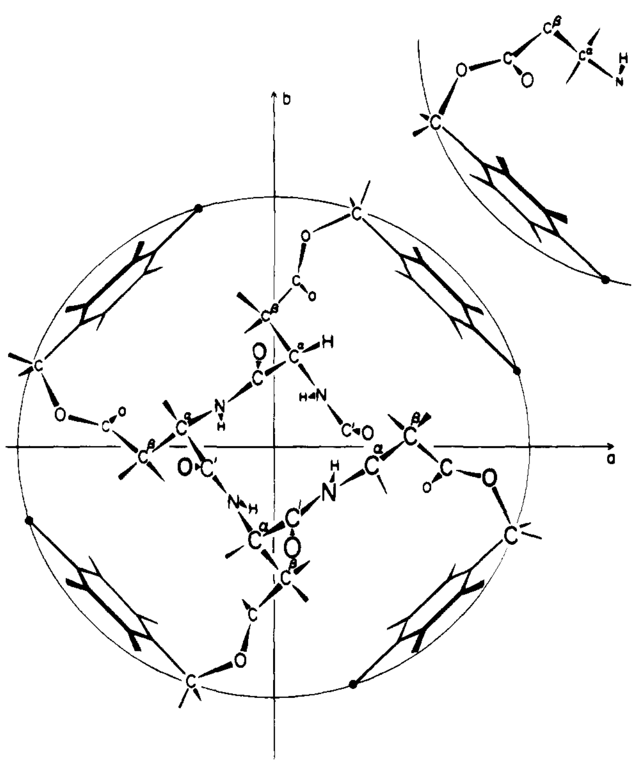


Figure 18. View of the minimum-energy crystalline ω-helical form of poly(p-chlorobenzyl L-aspartate). (Reprinted from ref 464. Copyright 1974 American Chemical Society.)

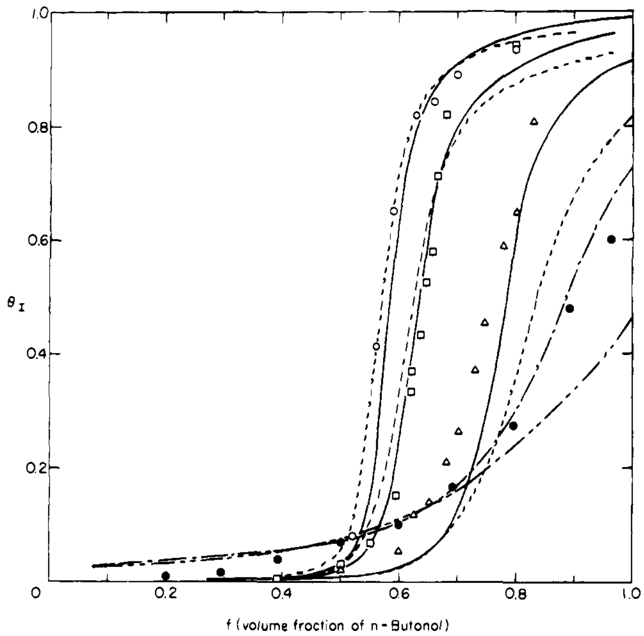


Figure 19. Calculated transition curves of poly(L-proline) in n-butyl alcohol/benzyl alcohol at 70 °C. The fraction of form I helix, θ_I, is plotted against the volume fraction of n-butyl alcohol. The experimental points (from ref 467) are shown for degrees of polymerization of 217 (○), 90 (□), 33 (△), and 14 (●). The various solid and dashed curves are based on different assumptions made in the computations. (Reprinted from ref 468. Copyright 1975 American Chemical Society.)

α-helix in solution and in the crystal, it is possible to convert it to an ω-helix in the crystal. Figure 18 illustrates how favorable interchain interactions enable the backbone to adopt the ω-helical form in the crystal.[464] Entropy effects play a role in the α to ω conversion as the temperature is raised.[464] X-ray fiber-diffraction studies have demonstrated the presence of the ω-helix in crystals of this polymer.[465]

Poly(L-proline) can also exist in two helical forms, I and II, with all-cis and all-trans peptide groups, respectively.[466] Interconversion between these two helical forms can be induced by changes of solvent.[466] For example, form I is stable in 1-butanol, and form II in benzyl alcohol. The transition curves (shown in Figure 19), computed from empirical potential functions, also taking the effect of these solvents into account (using specific experimental binding constants for forms I and II, respectively), match experimental data[467] fairly well.[468]

B. Small Linear and Cyclic Oligopeptides

The results of energy-based computations on model systems show at least qualitative reasonable agreement with experiment in a good number of cases. This relative success suggests that the overall approach is reasonable We may therefore attempt to apply this methodology to polypeptides and proteins. However, as pointed out in section II.B, one of the main obstacles to be overcome is the multiple-minima problem. Therefore, the discussion of computational methods for treating polypeptides and proteins will include frequent references to the multiple-minima problem. For this purpose, we divide polypeptides into three categories: (a) small open-chain and cyclic oligopeptides, (b) fibrous proteins, and (c) globular proteins. The approach to the multiple-minima problem is different for each category and, indeed, to some extent it has been solved for categories a and b.

This does not mean that any of the methods reviewed here is able, consistently and with generality, to offer a complete and correct description of, say, small peptides in aqueous solution without any reference to experimental data. What it means is that it is probably possible to find a fairly complete set of low-energy conformations (perhaps including an excellent approximation to the global minimum) of small peptides described by a potential energy model *without explicit solvation*. Thus, it is probably fair to say that the multiple-minima problem has been solved for the *ECEPP model* of [Met]enkephalin, a five amino acid polypeptide, as strongly suggested by a number of studies in a growing number of laboratories.[56,217,262,292,469] Its structure is shown in Figure 6.

However, this is still a far cry from a complete theoretical description of the conformational behavior of the [Met]enkephalin *real molecule* in a given environment, especially in a complex solvent such as water. The relevance of this "correct" theoretical result to a real system depends on how good the potential functions are in representing it, and in how well the real system can be understood by looking only at the lowest energy conformation: the latter point is especially critical for flexible peptides. For example, calculations pertaining to nonpolar environments (e.g. those on melittin[195,228,272]) tend to be more successful than those applied to systems where polarization and other complex effects are more

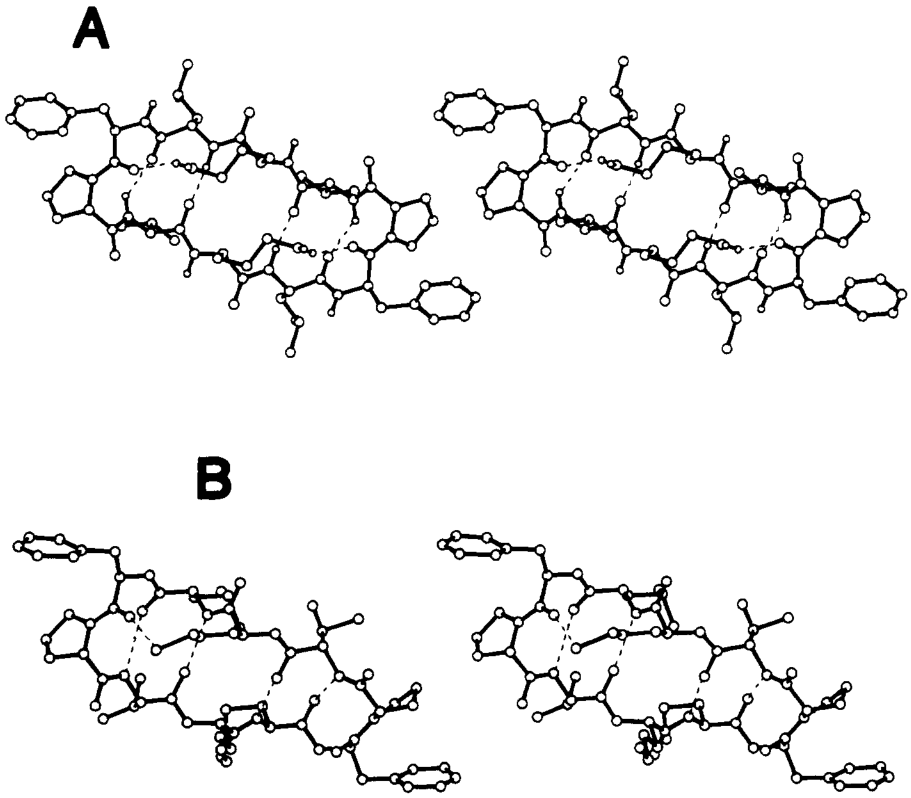


Figure 20. Stereoviews of (A) computed (refs 470 and 471) and (B) X-ray (ref 472) structures of gramicidin S, showing the hydrogen bonds connecting the antiparallel chains and a hydrogen bond between the ornithine side chain and the phenylalanine backbone carbonyl group.

dominant. For fibrous proteins we may have a similar situation; for example, using the fact that one should obtain a coiled coil triple-helical structure for collagen, it is possible to achieve a result in reasonable agreement with experiment, which probably reflects a global energy minimum of the model energy. However, it is not clear if the reviewed methodology would be capable of predicting the *triple helix itself* from sequence information alone. We cite the recent findings by Harbury et al.[427] (described in detail in section III.A.2) as a reminder of the complexity of real polypeptide systems.

The build-up procedure (section II.B.1) has been used to treat a number of linear and cyclic peptides. We will describe a limited number of examples, including applications to gramicidin S, melittin, and alamethicin.

1. Gramicidin S

Gramicidin S is a cyclic decapeptide, whose computed structure[470,471] consists of two antiparallel extended chains, cross-linked with four hydrogen bonds, and connected by β-turns at each end. A comparison of this structure (shown in Figure 20A) indicated that it was in good agreement with a subsequently determined[472] X-ray crystal structure of a hydrated gramicidin S−urea complex (shown in Figure 20B). The small distortion in the lower right-hand portion of the observed structure (Figure 20B) is most likely due to a nearby urea molecule which cocrystallized with the decapeptide. Subsequently, Mirau and Bovey[473] carried out a 2D NMR ROESY experiment on gramicidin S in solution and compared the experimental spectrum with a theoretical spectrum calculated from the published atomic coordinates[470] of the energy-minimized structure; close agreement was obtained for the backbone protons. Differences that were observed for the side-chain protons were attributed to motion in solution. More recently, Xu et al.[474] have also obtained NMR evidence that supports the calculated structures.[470,471]

The computed result of Figure 20 is not a fortuitously obtained one, but was arrived at (in the last stage of the build-up procedure) by energy minimization from 10 541 starting conformations, with the imposition of C_2 symmetry. To provide an idea of the magnitude of the number of possible conformations, we refer to Figure 21 which shows some of the computed low-energy conformations of cyclo-hexaglycine.[475] These are conformations that have some kind of symmetry, and there are still others (obtained by a Monte Carlo procedure[476]) with nonsymmetric conformations. A cyclic decapeptide such as gramicidin S would have many more low-energy structures than the cyclic hexapeptide of Figure 21. When viewed in this light, it is seen how well the computational methodology leads to the fairly unique low-energy conformation of Figure 20.

2. Melittin

Because of the extremely large number of low-energy minima for each component fragment of a polypeptide, the size of gramicidin S at first appeared to be the upper limit that could be treated by this methodology, and hence defined what was meant by "small" in the title of this section. However, recently, by not only eliminating high-energy structures of intermediate fragments, but also by reducing the

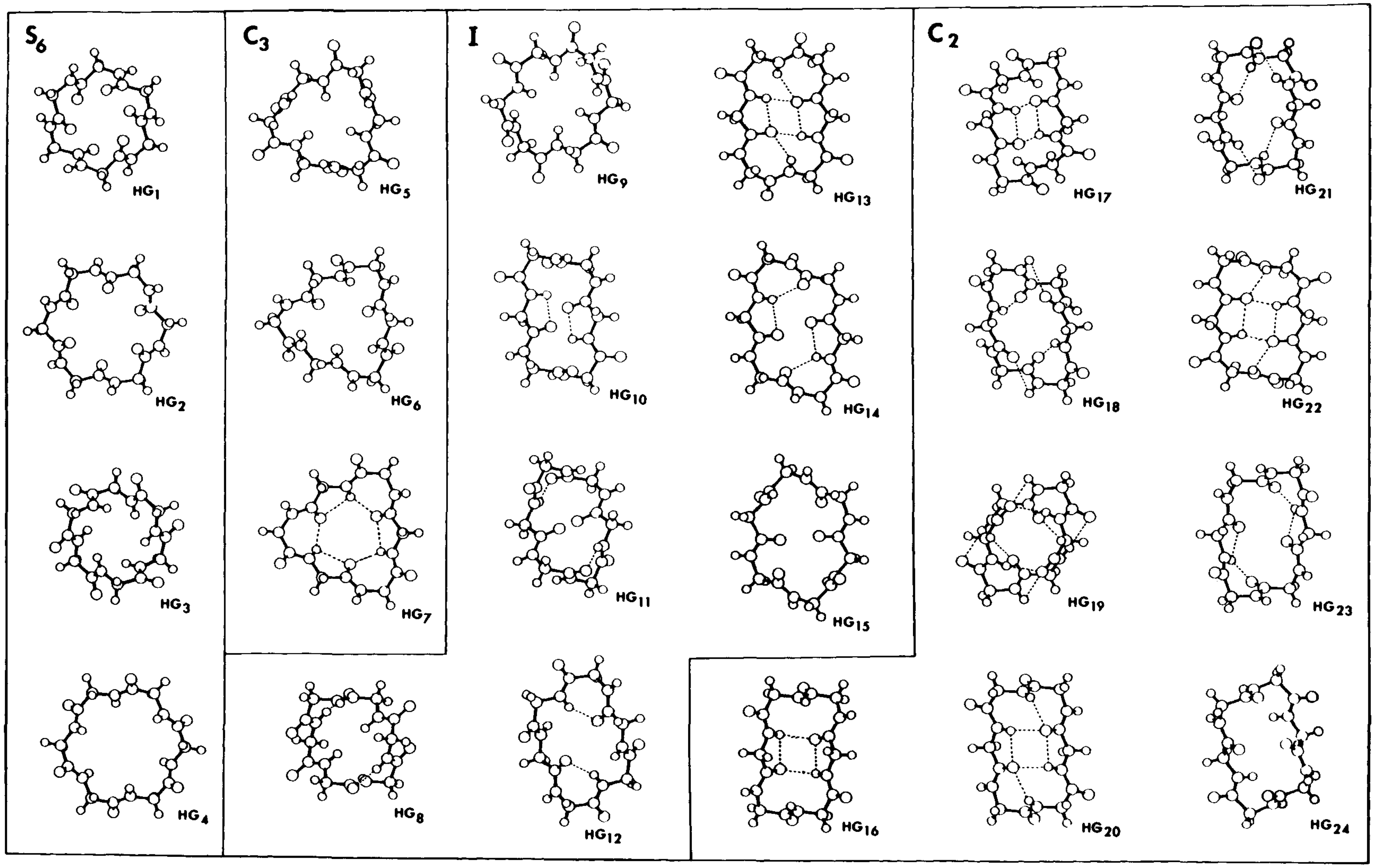


Figure 21. Computed low-energy structures of cyclic hexaglycine, with various kinds of symmetry as indicated. (Reprinted from ref 475. Copyright 1973 American Chemical Society.)

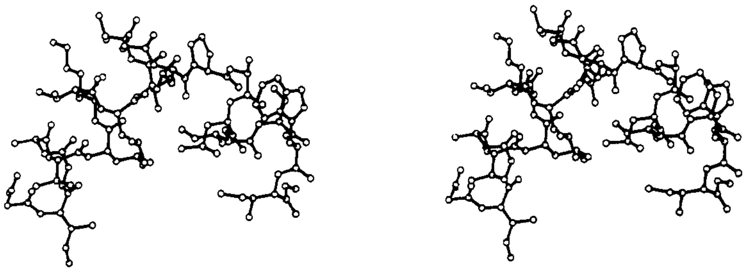


Figure 22. Stereoview of the lowest energy structure calculated for residues 1–20 of melittin. (Reprinted from ref 195. Copyright 1982 M. R. Pincus et al.)

number of low-energy minima further by selecting only those having different backbone conformations (designated as nondegenerate minima[477]), it appears to be possible to treat larger structures in this manner. Thus, the buildup of large fragments from small ones may be possible. In fact, such a calculation has been carried out for the 20-residue membrane-bound portion of a relatively small globular protein, melittin.[195] Only two (very similar) low-energy, largely α-helical structures were found;[195] one of them is shown in Figure 22. Very similar structures were also obtained by the EDMC,[228] SCMTF,[312] and ant lion plus neural network[272] methods. X-ray[273] and NMR[478] structural information is available for melittin either as a tetramer in a crystal or as a monomer bound to micelles, respectively. Considering possible environmental effects in either of these forms, the general qualitative agreement with experiment is satisfactory.

3. Alamethicin

A recent calculation[479] of the effects of electric field on the alamethicin molecule bound at the lipid–water interface used a version of the build-up method, and a volume-shell model[112,480] to estimate solvation effects. This study attempted to examine the entire conformational ensemble of alamethicin in water and at the lipid boundary, and a significant shift in conformational populations was observed when going from one environment to the other. The authors also included a simple model of the effect of the electrostatic potential, which indicated the possibility that it may influence both the orientation and the conformational equilibrium of the alamethicin molecules. Their results provide insights into the possible membrane-binding and voltage-gating mechanisms mediated by this peptide.

4. Design of Constrained Peptides

Dynamic simulated annealing (DSA) and quenched MD (QMD) have been used in a large number of studies of peptides in solution. As representative examples, we mention the work of Hagler, Rivier, and co-workers on analogs of gonadotropin-releasing hormone (GnRH),[481,482] and that of Pettitt, Hruby, and their colleagues on a number of biologically active peptides.[483–485] In both sets of studies, the computational methodology was applied first to fairly flexible molecules, and analysis of the resulting conformational ensemble suggested possible ways in which the molecules could be made more rigid. A number of these have been synthesized, and some of them have interesting biological activities, while being rigid enough to be studied by NMR techniques. For example, analysis of QMD simulations of α-melanotropins, led to the design, synthesis, and testing of a new family of potent cyclic analogs.[484] In a very similar fashion, the conformational studies of GnRH have facilitated the production of potent dicyclic antagonists.[486]

5. Multiconformational NMR

In this section we describe an emerging new, significant, role of energy calculations in the interpretation of NMR experiments on flexible peptides. Because of the large conformational heterogeneity present in these systems, one cannot simply extend the computational methodology successfully used to analyze NMR spectra of proteins in solutions.[242,295,487–491] It has become apparent that new procedures are needed to understand conformational ensembles.[492]

An interesting approach has been described by Brüschweiller et al. in their MEDUSA algorithm, which stands for multiconformational evaluation of distance information using a stochastically constrained minimization algorithm.[4] The essence of the method is to generate a large number of conformations (usually in the hundreds), each of which satisfies a *subset* of all the distance constraints derived from the NMR NOESY spectrum. Then these structures are combined in pairs, triplets, or larger groups to produce an ensemble that satisfies all the experimental data. The procedure has been described in detail and used to analyze conformational equilibrium of the cyclic decapeptide antamanide, *cyclo*(-Val-Pro-Pro-Ala-Phe-Phe-Pro-Pro-Phe-Phe-).[493] The results have also been compared with detailed explicit-solvent, chloroform, simulations of the same molecule.[494–496]

A different approach was used by Nikiforovich et al.[6] to analyze the conformational equilibrium of dermenkephalin in DMSO solution. Their procedure starts by generating a large statistical sample of conformations by carrying out a room temperature Metropolis Monte Carlo[219] simulation in torsion angle space. The experimental data are then used to estimate a set of optimal statistical weights for each conformer, chosen in a way to minimize differences between measured properties (e.g. NOE cross-peaks, or vicinal coupling constants) and calculated ensemble averages. The result of the procedure is a set of conformers and associated statistical weights consistent with the experimental data.

It is clear that other procedures able to generate conformational ensembles for a peptide may be useful in this context; the version of MCM[217] described by von Freyberg and Braun[262] is of particular interest, given the apparent superiority of MCM over conventional Metropolis Monte Carlo in analysis of peptide conformations.[56,249]

C. Crystals

The structure derived by several "global" optimization procedures, shown in Figure 6, for [Met]enkephalin (which would be presumably the same for [Leu]enkephalin) does not agree with any of the polymorphic forms obtained from X-ray crystallographic studies.[497] The latter are either extended chains in hydrogen-bonded β-sheets[498-501] or hydrogen-bonded hairpin conformations,[502,503] but with the bend in a different region of the backbone. Clearly, *inter*molecular interactions such as hydrogen bonds in the crystal influence the conformation of this small linear oligopeptide. Hence, calculations were carried out[497] to minimize the energies of three types of crystals, the two observed ones and the one of Figure 6 in different crystal packings. The minimized energies of the two observed crystal structures were lower than that of any of the packing arrangements of the structure in Figure 6. If, however, an individual enkephalin molecule is removed from any of the energy-minimized crystals, thereby depriving it of favorable *inter*atomic interactions, then the lowest energy conformation of the isolated molecule is indeed that of Figure 6. Presumably, this latter structure undergoes a conformational change upon crystallization to make optimal use of intermolecular interactions. Hence, the structure of Figure 6 is considered to be the global-minimum one for the isolated molecule. Small cyclic molecules, such as gramicidin S, discussed in section III.B, are presumably less influenced by intermolecular interactions.

Kitson and Hagler[504] have carried out a detailed study of the cyclic peptide *cyclo*(-Ala-Pro-D-Phe-)$_2$ in the isolated and crystal states. They used both static analysis by energy minimization, and molecular dynamics. The MD calculations of the isolated molecule led to a structure with an energy that is about 8.5 kcal/mol lower than that of the isolated peptide in the crystal conformation. Thus, although they do not claim to have explored the conformational space of the free peptide extensively, they can safely conclude that crystal interactions are capable of shifting the conformational equilibrium of small flexible peptides. Information about the conformational behavior of the free peptide is available from NMR experiments in a variety of solvents.[505] The MD simulations are at least qualitatively consistent with the NMR measurements. The calculations on the crystalline state included interactions within a unit cell (that contains two peptide molecules and 16 water molecules), plus interactions with symmetry-related unit cells up to a 15 Å distance cutoff. They started the calculations with the water molecules in random but sterically acceptable positions; interestingly, their positions after energy minimization or after MD averaging came closer to the X-ray ones.

Among others, Hall and Pavitt[506] and Jorgensen and Tirado-Rives[42] have carried out static energy minimization calculations of peptide crystals in order to test the adequacy of potential energy functions. There have also been a number of MD calculations of protein crystals by van Gunsteren and Karplus,[507] and by Hagler and co-workers.[508] These simulations are important for tests of the accuracy of potential energy functions, because direct structural comparisons with experiment can be made.

Many simulations of proteins in solution have been compared with experimental data derived from X-ray crystallography.[40] Discrepancies arising from errors in the potentials cannot be easily distinguished from legitimate differences between crystal and solution. Crystal simulations, although perhaps more difficult in execution, do not suffer in principle from this limitation. Nonetheless, as shown recently by Hagler and co-workers,[508] details of the computations, such as the size of the cutoff distance for nonbonded and electrostatic interactions, can have profound effects on the quality of the simulations. They report that a change from 15 to 25 Å in the cutoff distance is sufficient to improve the rms deviations between simulation and X-ray structure from 2 to less than 1 Å in MD calculations of *S. griseus* protease A in the crystal. A very detailed and carefully analyzed simulation of subtilisin in the crystal was described recently.[509] This work includes a comparative survey of previous simulations of proteins in crystalline environments.

D. Fibrous Proteins

1. Collagen

The structure of collagen is a direct consequence of its unique amino acid sequence in which every third residue is glycine, and the intervening two residues of each triplet (denoted X and Y) are frequently proline or hydroxyproline, respectively. The triple-stranded coiled-coil helical structure of natural collagen has been derived from fiber X-ray diffraction.[417-419] Synthetic Gly-X-Y poly(tripeptides) form helical fibrous structures, some of which are collagen-like.[510-512] The main structural features of collagen have been elucidated by means of conformational energy computations in a systematic series of investigations which dealt with various levels of structures, starting with the conformational analysis of the Gly-X-Y repeat unit, through the structure and stability of the triple-helical molecule, to molecular assemblies in microfibrils.[512,513] Advances in instrumentation and in peptide synthesis technology have enabled a series of high-resolution NMR studies of collagen-like peptides of well-defined composition and sequence to be carried out;[514,515] these investigations, still in their early stages, are expected to add significantly to our current understanding of these systems. Jones and Miller[516] have carried out an interesting study linking sequence and X-ray diffraction data by an analysis using computer graphics and solvent accessibility calculations.[121,517] They describe, and attempt to rationalize, a number of observations on triple-helix formation and suggest a fibril model for triple-helix association.

Poly(Gly-Pro-Pro) serves as the simplest general model structure for collagen.[435,512,513,518,519] Three equivalent poly(Gly-Pro-Pro) chains can be assembled

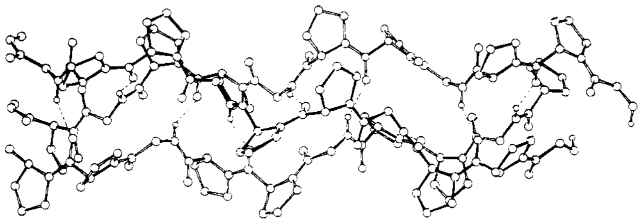


Figure 23. Calculated coiled-coil triple helix of poly(GlyProPro) of lowest energy. (Reprinted from ref 435. Copyright 1976 Wiley.)

into a three-chain structure according to various symmetry arrangements, including coiled coils with screw symmetry, and having either of two dispositions of the strands and parallel-chain complexes formed by helices that are packed with either screw or rotational symmetry.[435,436] The energy was computed for every possible symmetry arrangement for assembling three identical (Gly-Pro-Pro)$_4$ chains in regular conformations. The lowest energy structure turned out to be a coiled-coil triple helix with screw symmetry (Figure 23). Its helical parameters are close to those of the collagen models derived from fiber X-ray diffraction measurements.[417-419] Its atomic coordinates agree, to within an rms deviation of 0.3 Å, with the coordinates obtained subsequently for a single crystal of (Pro-Pro-Gly)$_{10}$ by high-resolution X-ray diffraction.[520] The difference between the computed energy of the triple helix and the energies of the component polypeptide strands accounts closely for the observed enthalpy of the triple helix-to-statistical coil transition of poly(Gly-Pro-Pro), if the contribution of the free energy of hydration, estimated by a volume-based shell model,[115] is included in the computation.[521]

In the first application of conformational energy computations to a real sequence in a collagen molecule (a short triple-helical fragment of type I bovine skin collagen), it was shown that large side chains, such as Arg and Met, can fold tightly against the triple-helical backbone. The side chains contribute significantly to the energy of stabilization of the triple helix, by means of nonbonded and hydrogen-bond interactions.[522] The role of intrachain salt bridges in collagen stability between Arg or Lys with Glu or Asp has been studied[523] with the AMBER[33] force field, both in vacuum and with explicit solvation. Although quantitative stability estimates are hard to provide, these calculations suggest an important role for these salt linkages.

The same collagen-like triple helix is the most stable structure for the assemblies formed by three poly(Gly-Pro-Hyp)[436] or three poly(Gly-Pro-Ala)[524] molecules, in agreement with X-ray powder diffraction experiments.[510] Poly(Gly-Ala-Pro), however, behaves differently.[525] Several coiled-coil triple-helical packing arrangements with low energy have been computed for this polymer, together with a parallel-chain triple-stranded complex in which the polypeptide chains take up conformations that are similar to those found in solid polyproline II[526] or polyglycine II.[527] These results agree qualitatively with X-ray diffraction measurements on poly(Gly-Ala-Pro) films, which can contain either collagen-like or polyproline II-like chain assemblies, depending on the solvent used to prepare the film.[528,529] The computations have provided a qualitative explanation of these observations because they suggest that poly(Gly-Ala-Pro) has several low-energy structures, and one can expect that its relative energy is modified by interactions with various solvents. However, the calculations have not yet included the effects of these solvents explicitly.

The collagen-like triple helices can be assembled into microfibrillar structures. As the first step in the theoretical analysis of the energetics of fibril formation, the geometry and energy of packing of two triple helices has been computed.[530,531] The preferred orientation depends on the amino acid sequence. In the computed lowest energy packing of two [CH_3CO-(Gly-Pro-Pro)$_5$-$NHCH_3$]$_3$ triple helices, the two molecules are arranged nearly parallel to each other, with an orientation angle of $-10°$ between the two helix axes (Figure 24). On the other hand, both near-parallel and near-antiparallel packings with low energies have been computed for [CH_3CO-(Gly-Pro-Ala)$_5$-$NHCH_3$]$_3$ triple helices.[530] This result suggests that the observed preference for the near-parallel packing of molecules in collagen fibrils is not merely a nonspecific packing effect but it is an energetic consequence of specific residue—residue interactions between the triple helices and, in particular, it is due to the frequent presence of imino acids in the sequence.[530]

The substitution of Hyp for Pro in position Y, where 4-Hyp (4-hydroxyproline) is found exclusively in natural collagen as a result of post-translational hydroxylation, enhances the stability of the near-parallel packing structure.[531] Exactly the same computed packing arrangement of two triple-helical molecules is favored by the interaction energy for both [CH_3CO-(Gly-Pro-Pro)$_5$-$NHCH_3$]$_3$ and [CH_3CO-(Gly-Pro-Hyp)$_5$-$NHCH_3$]$_3$, but the stability of the packing for the latter is enhanced by nearly 2 kcal/mol per Hyp residue, because of the formation of an intermolecular hydrogen bond between the side-chain hydroxyl group of Hyp in one molecule and a backbone carbonyl group in the second molecule (Figure 25). This computation has provided at least a qualitative explanation for the observed stabilization of collagen fibrils by the presence of Hyp.[532] An analysis of the change of thermostability upon proline hydroxylation and of the denaturation temperature of several naturally occurring collagens (with varying Hyp content) has been given recently, in an attempt to provide support for a water-bridged mechanism in Hyp stabilization of collagen structure.[533]

The triple helices in an observed collagen microfibril present a 5-fold symmetry of packing.[534] Computations on bundles of poly(Gly-Pro-Hyp) triple helices indicate that optimal packing, with the maintenance of most of the favorable nonbonded and hydrogen-bonding interactions (as in Figure 25), can be achieved easily in a bundle formed by five triple helices.[535] Chen et al.[536] have also carried out an energy minimization study, with the AMBER force field,[33,34] of a five-stranded model using (Gly-Pro-Pro)$_n$, (Gly-Pro-Hyp)$_n$, or (Gly-Ala-Ala)$_n$; with $n = 4$, 12, and 16. They report that the structural characteristics of the computed microfibril are consistent

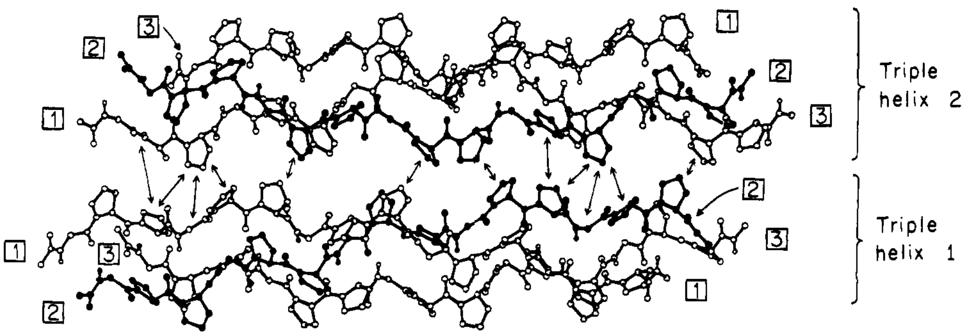


Figure 24. Computed lowest energy packing arrangement of two [CH_3CO-(Gly-Pro-Pro)$_5$-$NHCH_3$]$_3$ triple helices, showing the near-parallel alignment of the two triple helices. Arrows point to residues which are in van der Waals contact. (Reprinted from ref 530. Copyright 1984 Wiley.)

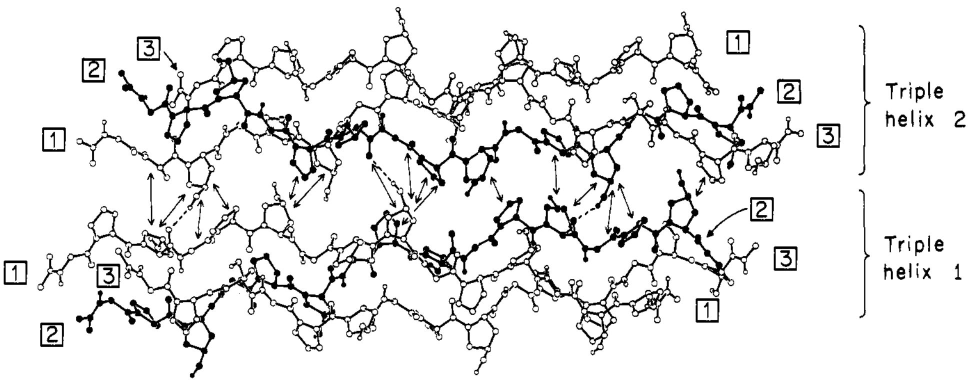


Figure 25. Computed lowest energy packing arrangement of two [CH_3CO-(Gly-Pro-Hyp)$_5$-$NHCH_3$]$_3$ triple helices, showing the near-parallel alignment of the two triple helices and the O—H· · ·O=C hydrogen bonds between the triple helices (shown with dashed lines). The arrows point to residues which are in van der Waals contact. (Reprinted from ref 531. Copyright 1986 American Chemical Society.)

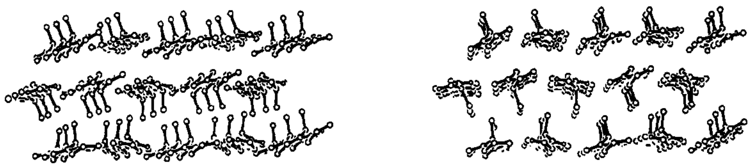


Figure 26. Computed model structure for the crystalline domain of silk II, with "in-register" arrangement of the sheets, viewed along the direction of the polypeptide chains. The figure shows the lowest energy packing for three stacked antiparallel five-stranded β-sheets formed by CH_3CO-(L-Ala-Gly)$_3$-$NHCH_3$ chains. The sheets are perpendicular to the plane of the drawing. Hydrogen bonds within each sheet (not shown explicitly) are horizontal. (Reprinted from ref 541. Copyright 1991 Wiley.)

with those obtained for collagen by X-ray diffraction and electron microscopy. More recently, a role for a structural pattern of charged amino acids has been suggested to account for the periodicity and chiral appearance of skin collagen fibrils.[537] These calculations were carried out with the AMBER force field,[33,34] and some aspects were confirmed with X-PLOR[51] calculations using the CHARMM[35] force field. Thus, conformational energy computations can lead to an increased understanding of the structure of supramolecular assemblies of a fibrous macromolecule.

2. Silk Fibroin

Silk fibroin is a block copolypeptide of crystalline domains (with high Gly and Ala content) and less crystalline domains. The crystalline domains of *Bombyx mori* silk contain the repeating amino acid sequence Gly-Ser-Gly-Ala-Gly-Ala (with some repeating Gly-Ala sequences at their ends), and they can exist in two morphologies, for which models have been proposed on the basis of fiber or powder X-ray diffraction studies.[538] The more stable form is known as silk II. A detailed structural model, consisting of packed pleated β-sheets, was first proposed for silk II by Marsh et al.[539] and refined by Fraser et al.[540] The structure of the less stable silk I form is less well understood.

Conformational energy computations have been carried out on packed sheet model structures, composed of poly(L-Ala-Gly) chains as a simplified model.[541] Such chains can be assembled into two kinds of β-sheets, viz. those in which all Ala side chains project from the same side of the sheet (termed in-register sheets) and those in which the Ala side chains point alternately to the two sides of the sheet (termed out-of-register sheets). The computations have confirmed that the structure with the lowest energy is formed of antiparallel in-register β-sheets, packed in such a way that the Ala-containing sides of the sheets face each other and the Gly-containing sides also face each other (Figure 26), as in the models proposed earlier for silk II.[539,540] The unit cell parameters of the computed structure agree closely with the observed values for silk II. A second, higher

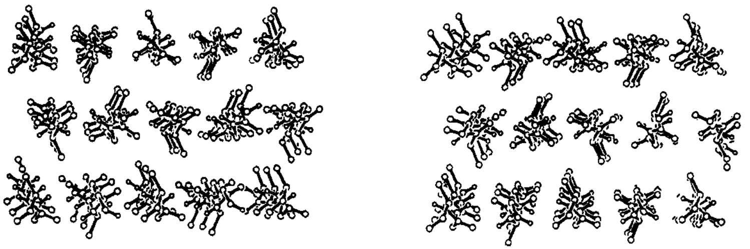


Figure 27. Computed model structure for the crystalline domain of silk I, with "out-of-register" arrangement of the sheets, viewed along the direction of the polypeptide chains. The figure shows the lowest energy packing for three stacked antiparallel five-stranded sheets. Each sheet is formed of alternating CH_3-CO-(L-Ala-Gly)$_3$-$NHCH_3$ and CH_3CO-(Gly-L-Ala)$_3$-$NHCH_3$ chains. The sheets are perpendicular to the plane of the drawing. Hydrogen bonds within each sheet (not shown explicitly) are horizontal. (Reprinted from ref 541. Copyright 1991 Wiley.)

energy computed structure has also been found, in which each strand forms a coiled coil, with residue conformations (for Ala and Gly, respectively) that correspond to a local right-handed and left-handed twist of a strand alternating along the chains. The strands are assembled into antiparallel out-of-register hydrogen-bonded sheets. These, in turn, stack into a structure in which every pair of adjacent sheets forms the same kinds of contacts (Figure 27). The computed unit cell dimensions of the structure are consistent with observed powder X-ray reflections and the observed density of silk I.[542] Therefore, this structure is proposed as a model for the crystalline form of silk I.

E. Globular Proteins

Despite the apparent formidability of extending the methodology described in section II to globular proteins, it now appears to be feasible to address at least some important aspects of protein structure. With presently available computer time, several of the foregoing methods for surmounting the multiple-minima problem can be applied to small globular proteins. As pointed out in section II.B.5, the performance of some of the best sampling procedures has started to shift the emphasis of the problem toward the design and development of more accurate representations of protein conformational energetics.

1. Modeling by Homology

If the X-ray structure of a homologous protein is known and available, then it can be used as a template to compute the structure of the globular protein of interest for which only the amino acid sequence is known. Presumably, the template provided by the known homologous structure will lead to a potential energy well that is close to that characteristic of the global (free) energy minimum of the protein of interest. How close these wells are will depend to a large degree on the percentage of identical or conserved amino acids between the two sequences. In most cases, most of the variation in amino acid sequence in a family of proteins will be on the surface rather than in the interior core.[543,544] The amino acid sequence of the unknown structure and the X-ray coordinates of the homologous structure provide a starting point that usually is of high energy because of steric overlaps. However, these are often easily relieved during the course of energy minimization. For example, this procedure, with the ECEPP potentials or its precursors, has been applied to compute the structure of α-lactalbumin[545] from that of lysozyme (and compared with the subsequently determined X-ray structure[546]), those of three snake venom inhibitors[547] from that of bovine pancreatic trypsin inhibitor, and angiogenin[548] from that of ribonuclease A.

Greer[549,550] has developed a general scheme for homology modeling by utilizing the information from several members of a family of proteins homologous to the one of interest. By achieving maximal overlap of members of the family, these substructures are used to build the molecule of interest. This method has been applied to a variety of proteins, e.g. to deduce a model for human renin based on structures from the family of aspartic proteinases,[551] and a model of the complement protein C5a from the structure of C3a.[552] Eventually one or more proposed models are also refined by energy minimization, or by restrained MD. Blundell and co-workers have developed a very similar protocol, which they call COMPOSER,[553-556] that progressively builds the conserved structural core (structurally conserved regions, or SCRs in Greer's terminology) from a structural consensus of a protein family (achieved by a multiple structural alignment procedure), followed by a loop-building algorithm (see below).

The approaches of Blundell and Greer extend the applicability of homology modeling to situations where the identity may be at or below 30%. This is due largely to the efficient use of information on a protein structural family, rather than on an individual protein. Recently, using very similar procedures, Ring et al.[270,557] and Sudarsanam et al.[558] independently have shown that these model structures can be useful in discovery of lead compounds for drug design. The success of these calculations, both involving proteases, depends to a large degree on the fact that the target active sites are among the most structurally conserved regions in these proteins. Thus, even large errors that may possibly occur in some loop regions may not seriously affect the results of their applications.

Protein modeling by homology can be divided into a number of steps. First, one has to identify one or more proteins of known structure that are likely to possess similar three-dimensional structures; i.e. find a set of structures that are indeed "homologous". Second, it is necessary to establish an alignment between the sequence of the target protein and that (those) of the known structure(s). In the limit of high (about 50% or higher) amino acid sequence identity (or homology), these two steps are rather straightforward, and good alignments may be achieved with few or no insertions and deletions. When one has weaker sequence identity, special procedures may have to be applied as outlined below. Third, after a reliable alignment is obtained, it is possible to build a model for the conserved core of the target protein. Fourth, nonconserved structural segments or loops are built in. At different stages of the overall process, it may be necessary to use specialized partial rebuilding tools to generate side-chain conformations on a largely fixed backbone or on an incomplete side chain (e.g. modeling a substitution of leucine from alanine), or full backbone coordinates from only α-carbon coordinates. Some of these tools are reviewed below. It is also possible or desirable to use energy minimization or MD refinement tools at different stages of the procedure, most crucially at the final stage after a complete model has been built. A final important step in a complete building-by-homology study is validation of the resulting model.

In the subsections below, we review in detail approaches and applications to address the most important steps in modeling by homology. The emphasis and importance of each of these steps depends largely on the sequence similarity between target and template(s). In the limit of very high similarity, say point mutations, side-chain packing calculations and energy refinement may be the most important steps. In these cases, one usually expects very small structural displacements, and thus the success or failure of the calculation ought to be assessed by very detailed comparison of atomic positions. On the other hand, in the limit of very weak homology, a lot of emphasis should be placed in justifying the use of a particular template, and later, on obtaining a good alignment. Clearly, subsequent modeling can do little to recover from serious mistakes at this stage. In the present state of technology, efforts of this nature aim to produce qualitatively correct folds and approximate residue–residue relations; the loop regions may contain some errors, and a few side chains may be misplaced.

a. Alignment. The first step in a typical homology modeling calculation is the choice of one or more proteins of known structure that are likely to be similar to the protein of interest or target. When the percentage of identical residues between template and target sequences is above about 30%, there is usually little doubt that the overall folds will be indeed very similar. When no protein of known structure satisfies this criterion, the identification of a useful template becomes a problem in pattern recognition. The realization that proteins with virtually undetectable sequence homology may actually have very similar three-dimensional structures has encouraged the development of more sensitive methods. Some of these are discussed in section III.E.2.

Even when structural similarity is well established, calculation of the correct alignment between target and template(s) may be a difficult and error-prone process if the percentage identity is much lower than 50–60%. The use of multiple target sequences and/or multiple templates may increase the accuracy of the alignment. Another idea is to include as much structural information as possible in the alignment step. A number of solutions have been proposed and some of the methods of section III.E.2 may also be useful here.

An illustration of the difficulty and importance of this step may be found in the predictions and subsequent structural determinations of the N-terminal domains of CD4.[559,560] Sequence comparisons suggested that CD4 may have strong structural similarity with other members of the immunoglobulin (Ig) family,[561,562] and this observation encouraged several groups to calculate somewhat detailed three-dimensional models of CD4 using antibody structures as templates.[563–566] The experimental X-ray determination confirmed that CD4 had an Ig fold;[559,560] however, none of the published predictions obtained even a correct alignment. (Bowman et al.[566] perhaps obtained the least incorrect result; interestingly, these authors included multiple sequence information in constructing their alignment.) As a consequence, subsequent refinement and loop modeling were rendered almost meaningless by the serious alignment errors.

b. Modeling of Loop Conformations. In aligning the sequence of a protein for which a homology-built model is desired, one often encounters insertions and/or deletions in relation to the template structure(s). Thus, the backbone structure of these regions has to be rebuilt. This may also be the case when, even in the absence of insertions or deletions, the degree of sequence similarity in a region may be so low as to render inapplicable the assumption of structural similarity. This problem can be formulated as the generation of loop segments between two fixed points of the polypeptide chain.

A knowledge-based approach to solve this problem was proposed some time ago by Jones and Thirup,[567] and applied widely since.[555,568,569] In this method, one searches the data base of protein crystal structures for polypeptide segments with the same number of residues as the loop to be built, and with compatible end-to-end geometry. Chothia and Lesk proposed a similar, specialized, protocol to model loop regions in antibody-variable domains.[570] Their method uses the so-called canonical hypothesis, which says that the length of the loop and the presence of a small number of *key* amino acids are the main conformational determinants of these regions; this work extended early observations on antibody crystal structures by Padlan and Davies.[571] The procedure was tested quite successfully in a comparison with subsequently determined X-ray structures of antibodies.[572]

Blundell and co-workers[555,556,573] tried to generalize the key-residue idea to cover other protein families; their procedure reduces to a variant of the Jones–Thirup method when no compatible loops can be found in the set of homologous structures. Knowledge-based procedures of this type are limited by the relatively small conformational repertoire available in the protein structural data base. When there are good biochemical and structural reasons to expect such a limited repertoire, as appears to be the case for five out of the six hypervariable loops in antibody variable domains, this is actually quite an advantage, because the methods are usually very fast and relatively simple to program. Most importantly, grossly incorrect conformations may never be considered. However, there are cases, e.g. the third hypervariable loop in the variable heavy chain domain in antibodies (CDR-H3), where this type of methods fails. In any case, it is important to supplement geometric building procedures with robust energy minimization strategies, in order to generate useful models. Such an integrated antibody modeling approach, which includes an automated implementation of the Chothia–Lesk rules,[570] has been developed by Levitt and co-workers,[574,575] and often applied in design and analysis of engineered antibody molecules.[576–578] Other protocols, also based on the Chothia–Lesk hypothesis, have similarly been found useful in a number of antibody engineering projects.[579–582] Broader overviews of antibody structure and engineering may be found in recent reviews.[583–585]

To construct backbone structures for loops computationally, a number of numerical procedures have been developed over the years. Gō and Scheraga proposed a method for exactly closed rings, without and with symmetry, in molecules whose bond lengths and bond angles are fixed.[586,587] Loops in proteins are special cases of rings in which the ends of the loop do not coincide. In a ring with n bonds around which rotation is possible, there are only $n - 6$ independent variables. The values of the six dependent variables are determined by the conditions of ring closure. These conditions are mathematically equivalent to that of finding a series of operations that will transform a coordinate system at one endpoint of the loop into that at the second endpoint. Gō and Scheraga derived a set of equations expressing the dependent variables as functions of the independent ones. They showed that there exist correlations between the dependent variables, permitting the evaluation for ring closure as a one-dimensional numerical search over the allowed range of one of the variables.

Bruccoleri and Karplus[588,589] used the Gō–Scheraga algorithm to model loops in proteins. They modified the original formulation by allowing bond-angle bending whenever the simultaneous set of equations did not have a solution. Their flexible-geometry method has been incorporated into the program CONGEN,[589] which is designed to execute loop searches in homology-modeling applications. In treating loops larger than three residues, CONGEN carries out searches over additional dihedral angles; however, the final closures are achieved by applying the modified Gō–Scheraga procedure to three residues only. This method has been quite successful as a main component of antibody modeling projects.[209,590]

Dudek and Scheraga developed an alternative formulation of the equations, involving a local minimization procedure with respect to subsets of degrees of freedom, to improve computational efficiency.[591] Palmer and Scheraga[592] modified the original formulation[586] to take into account differences in the backbone geometry of various amino acids, and they developed a systematic procedure, based on localized conformational searches, to generate an ensemble of short loops, followed by energy minimization, applicable to homology-based modeling.[593] The results of Palmer and Scheraga from generating rigid-geometry closures for five-residue segments[592,593] demonstrate, however, that bond-angle bending is not necessary to achieve satisfactory closures for chain segments five residues or longer in length. Instead of adding one or more bond angles[588] as independent variables, these authors increase the number of variable dihedral angles to find solutions to the closure problem. In either case, the fit to α-carbon chains from experimental protein structures appears satisfactory.

Another approach to loop searches was developed by Moult and James and applied to known crystal structures.[594] Their method involves the selection of a representative set of (ϕ,ψ) values for each residue from low-energy regions of the (ϕ,ψ) map. A series of loop fragments having various combinations of these (ϕ,ψ) dihedral angles are then generated. Those fragments that come close to satisfying the loop closure requirements (as determined from a set of empirical rules) are then refined by energy minimization in the presence of the rest of the protein.

A third approach, the "random tweak" method, was developed by Levinthal and co-workers,[595] and applied to protein structures.[596] The random tweak approach differs from others in that a systematic search of the loop conformations is not carried out. Instead, loop fragments are generated randomly and then constrained to meet the distance and orientation requirements for closure. This is achieved quite efficiently by solving the constraint equations using a method based on Lagrange multipliers. Most of the structures thus generated satisfy the closure constraints and are then evaluated using energy-based methods. The side chains are truncated in this procedure, and no attempt is made to find a stable conformation for them until a later stage in the modeling.

The advantage of the random tweak method is that it avoids the exponential increase in computing time with loop size, which is inherent in systematic searches (see the discussion in section II.B.1). In an application of the method to protein loops, Fine et al.[596] report that reasonably correct backbone conformations for a number of loops were indeed located by this procedure. A much larger data base of predicted structures will be needed to show that this result is generally attainable. More recently, Smith and Honig have combined the random tweak method with an evaluation of the energy of each candidate loop conformation using a finite-difference solution of the Poisson–Boltzmann equation;[597] thus, they have attempted to include the effect of solvent on conformational preferences.

An interesting application of the random tweak method, in combination with the loop procedure of Chothia and Lesk,[570] has been presented recently to model an antibody of potential therapeutic value.[598] Earlier, Rees and co-workers[590] have similarly combined the Chothia–Lesk method, a variant of the Jones–Thirup spare parts[567] procedure, and the modified Gō–Scheraga[586] algorithm used in the CONGEN program,[589] in a protocol to model antibody hypervariable loops.

A perturbation approach to loop closure has also been developed by Braun.[599] This method requires that an initial chain closure be located; this starting structure is then perturbed by making small changes in the initial dihedral angles, and then carrying out a minimization to produce another closure. A problem with this method is that the "allowed" changes usually involve only small alterations of the dihedral angles. Thus, only a fraction of the conformational space of the loop is searched, since larger changes in the dihedral angles rarely lead to closure.

Other loop-closure procedures have been described recently. These include the bond-scaling algorithm of Zheng et al.,[600] and the directed Monte Carlo procedure[258,259] of Higo et al. (see also section II.B.3). The bond-scaling method has also been combined with the multiple-copy procedure of Elber and co-workers,[32,317,319] and applied to the six hypervariable loops of an antibody.[601] Summers and Karplus[602] also described a specialized loop-closing protocol to treat single insertions or deletions, in addition to changes involving proline residues; a more general procedure based on energy fingerprints has also been proposed by these authors,[603] as a significant extension of the spare-parts approach of Jones and co-workers.[567,604]

An interesting application of the biased Monte Carlo minimization method[56,163,217] has been reported recently to model a loop of a designed monomeric version of the normally oligomeric trypanosomal triosephosphate isomerase (TIM) protein.[605] The modeled loop was then compared with the subsequently determined crystal structure, and the resulting rms deviation was 0.6 Å.[606]

A general problem with the computational construction of protein loops is the choice of "the" correct conformation (or small set thereof) from the many usually generated in most of the protocols outlined here. Thus, as in the more general problems of complete peptide and protein calculations, the issues of sufficient sampling and correct evaluation of conformational energies are relevant here. There are aspects of the loop problem, however, that make them more attractive as test systems for new search protocols and energy models. Because of the constrained nature of at least small loops (say, up to eight or a few more amino acid residues) the conformational space available can be much smaller than that of a free peptide of the same length. Also, it is easier in principle to include the most important environment influences, mainly the rest of the protein, although solvation effects will be important, especially in solvent-exposed loops. Therefore, as in *full* globular proteins, one expects a more or less unique structure, enabling more meaningful comparisons between calculations and experiments (of course, there will be cases where particular loops in proteins may be flexible enough to render meaningless the notion of a "correct" conformation). Unlike *full* globular proteins, of course, the conformational space is drastically reduced, perhaps to the point where some of the multiple minima-solving procedures described in this review could reliably find excellent approximations to the "global minimum" for a given potential energy model.

The methods outlined in this section concentrate on searches of the conformational space of the backbone; clearly, full side-chain structures will be required for correct evaluation of the relative energies of alternative loop conformations when using atomic potentials. Several of the protocols reviewed here do include treatment of side-chain conformations to some degree. Specialized side-chain building protocols have also been developed and are described below. Finally, we mention that protocols for protein loop calculations may also be adapted successfully for computation of protein–peptide complexes; in fact, the bond-scaling algorithm of Zheng and co-workers[600] has been applied recently to study MHC protein–peptide complexes.[607]

c. **Partial Rebuilding Tools**. In homology- and other protein-modeling studies, it is often necessary to reconstruct full atomic models partially from incomplete coordinate sets. For example, partial rebuilding is used when carrying out nonconservative amino acid substitutions, or when modeling full peptide segments from backbone models of loops. Also, when starting with simplified models of proteins (or protein loops) in which only α-carbons are initially treated, one may need tools to construct full backbone models. Refinement of X-ray structures from low-resolution incomplete data is a special case of this problem.[608,609] In this section, we describe methods to generate full backbones from C^α traces and full side chains from backbone or from nonconserved side chains.

i. Full Backbones from α-Carbons. Protocols to compute full main-chain coordinates from information on the C^α's can be divided into knowledge-based or "spare-parts" methods, and computational procedures. The first type were inspired by the Jones–Thirup method[567] already mentioned above. In this procedure, a data base of high-resolution protein structures is searched for full segments that have the best possible match to the input set of C^α coordinates. As shown by Jones and Thirup[567] as well as by others,[610] good matches between arbitrary pairs of protein segments may always be achieved for up to size five or six residues. Thus, a parameter of these methods is the fixed or variable size of the segments one tries to match. An automated procedure was described by Claessens et al.[568] and found to be able to match experimental backbone coordinates with rms deviations of a little over 0.5 Å, when the correct C^α coordinates were used as input. This protocol used a number of screens, starting with end-to-end distance matching, and ending with full optimal superpositions, in an effort to find the largest possible segment in the data base with good geometric match to the input coordinates.

An improved procedure was described in detail by Levitt.[611] In this method, which is also capable of building entire side-chain conformations, segments of fixed size are matched to the input coordinates. An interesting feature is the use of stochastic averaging whenever arbitrary decisions seemed to creep into the design of the algorithm, followed by a restrained energy minimization procedure. Another interesting result is the rebuilding of full models from even a partial list of C^α coordinates. The performance for correct C^α input is 0.3–0.5 Å rms deviation over a set of eight well-resolved proteins.

Jones et al. have recently updated their original methodology to develop a protocol useful in model building during crystallographic analysis.[612] Holm and Sander[613] have also developed a version of the spare-parts method, and achieved rms deviations of around 0.4 Å, starting from experimental C^α coordinates. Wendoloski and Salemme[569] have given a very brief description of a segment-matching backbone reconstruction algorithm within their PROBIT computer program. They report rms deviations of 0.3–0.4 Å in a test limited to three small proteins.

Among computationally based procedures, the method of Purisima and Scheraga[614] seems to be the first one described in the literature. In their procedure, they looked for solutions in terms of dihedral angles $\phi-\psi$ in an effort to fit standard-geometry full backbone models to input C^α coordinates. The procedure was quite successful in recreating full backbones from coordinates consistent with the same standard geometry; however, it failed with experimental C^α's that were incompatible with standard values of bond lengths and angles. Some of these problems were addressed by Palmer and Scheraga,[592] by using their modified version of the Gō–Scheraga loop-closing method. Neither of these studies presented systematic comparisons between experimental full backbone coordinates (e.g. N, C^β, C′, O) and those computed from knowledge of the α-carbon coordinates alone. Thus, although it does seem quite possible to fit standard-geometry backbones to a wide range of experimental α-carbon arrangements, it is not known how much the resulting full backbone coordinates would deviate from the experimental ones.

An alternative formulation of the Purisima–Scheraga algorithm was developed by Rein and coworkers and termed the λ-algorithm.[615] They use the auxiliary λ_1 and λ_2 variables of Nishikawa et al.[616] to describe the orientation of the two (assumed planar) peptide groups about a central α-carbon. Changes in these variables do not perturb the overall path of the α-carbon trace. Thus, in the context of fixed C^α coordinates, these variables present a more convenient description of backbone geometry than the usual dihedral angles ϕ and ψ. Vicinal λ_1 and λ_2 are coupled by the constraint of fixed bond angle about the C^α. These authors did not present tests of their method with experimental C^α coordinates.

A second procedure based on peptide plane orientations was described by Luo et al.[617] Their T variables are the same as the λ's described above, but unlike previous work, they allowed for some degree of bond angle deformation about the central α-carbon. They used a scaling of the $C^\alpha-C^\alpha$ virtual bond distance to treat small deviations from the ideal 3.8 Å value. They generated sequences of peptide orientations for segments of eight residues and chose the best by selecting combinations with resulting $\phi-\psi$ values in the allowed regions of the map.[17,594] They tested their procedure with the experimental C^α coordinates of a small number of proteins and achieved rms deviations, for the full backbone, of 0.36 to 0.47 Å. They also recorded the number of peptide plane "flips"; i.e., cases in which the calculated and experimental T values differed by more than 90°, and found these to occur in about 2–5% of the residues tested. They included extensive tests for rebuilding the backbone of trypsin inhibitor[618] from partially randomized input α-carbon coordinates, in order to simulate perhaps more realistic model-building situations. In these cases, the rms deviations deteriorated from 0.36 Å for the exact coordinates, to 0.41–0.54 for mean random shifts of 0.2 Å, up to 0.64–0.84 Å for mean shifts of 0.5 Å; and from no peptide flips to 1–6 flips for the largest mean random shifts (of 0.5 Å).

A perhaps more satisfying and, certainly more accurate, realization of the idea of rotation of peptide groups was described recently by Payne.[619] There are several interesting features in this procedure, including allowance for peptide plane deformation (thus, his W_M and W_P variables refer to orientations of *mean* peptide planes, and are not strictly identical to the λ's and T's of the other procedures) and use of an efficient dynamic programming algorithm to carry out the calculation of optimal W_M and W_P combinations. The most notable point , however, may be the use of a potential of mean force approach as the key criterion to choose preferred peptide orientations for given input C^α coordinates. As in the schemes of Pohl[94] and Sippl[11,92] described in section II.A.2, this potential of mean force is derived by way of the Boltzmann law from a careful analysis of the populations of peptide plane orientations in loop regions (i.e. neither α-helix nor β-sheet) in a large set of high-resolution protein structures. This procedure achieves rms backbone atom deviations of 0.2–0.3 Å for most proteins tested, and peptide plane flips occur in less than 1% of the cases. Interestingly, rms deviations increase only linearly with mean random shifts in the input α-carbon coordinates of up to 1 Å; thus, the procedure seems quite robust. The method is quite rapid and appears to be the most accurate one described in the literature to date.

Rey and Skolnick proposed a novel procedure to reconstruct full backbone coordinates from only α-carbons.[620] A key element in their procedure is the observation of an empirical correlation between the orientation of the $C^\alpha-C^\beta$ bond and the local geometry of the virtual α-carbon chain. After the β-carbon positions are obtained, a series of trigonometric relations provide enough information to compute N, C′, and O coordinates. In a test with six proteins, the rms deviations between computed and experimental coordinates were around 0.7 Å, which is much worse than some of the procedures described above. After energy minimization with fixed α-carbons, the rms deviations improved to around 0.4 Å.

Bassolino-Klimas and Bruccoleri[621] have adapted the CONGEN program[589] to produce full backbone models from α-carbons; the rms deviations over a set of six test proteins are quite large, 0.5–0.99 Å, and the method is very time consuming, 50–100 h on a workstation. Also, the described implementation does not allow a very close fit to the α-carbon coordinates themselves, which deviate from the X-ray values by 0.30–0.87 Å. Thus, in its present state of development, this method does not look very competitive with the best in the literature.

Liwo et al. have recently described a backbone rebuilding algorithm.[83] Their procedure, the dipole-path method, is based on determining an extensive hydrogen-bond network between the peptide groups of the backbone. A single test of reconstructing the entire backbone of trypsin inhibitor led to rms deviations of about 1 Å between calculated and experimental coordinates. The deviations for residues involved in the "dipole path" are much lower, averaging about 0.4–0.5 Å.

Methods based on MD simulation techniques have been described by Correa[622] and by Nilges and Brünger.[431,432] In both procedures, the backbone (and side-chain) atoms are initially positioned in a random or arbitrary fashion, and their positions are later optimized during restrained MD. Correa[622] presented a test on α-lytic protease in which he achieved a very low, 0.19 Å, rms deviation between calculated and experimental backbone coordinates, but at the cost of many hours of computer time (these calculations also yielded coordinates for the side-chain atoms). The procedure of Nilges and Brünger uses an optimized[244] sequence of changes in the energy function after essentially random positioning of the missing atoms; they applied their procedure in the calculation of the structure of GCN4[431,432] described earlier in this review (see section III.A.2.b). They make allusion to more extensive tests, but neither general timing nor rms deviation information has yet been given for their protocol.

ii. Side Chains from Backbones. A large and growing number of procedures for calculating side-chain coordinates from backbone, or even just α-carbon, coordinates have been described recently. Some of these methods are integral parts of larger programs for protein homology modeling or peptide modeling. A number of them have been tested in a somewhat artificial, but nevertheless useful, system. This test system consists of generating side-chain conformations on a fixed, full, backbone usually taken directly from an experimental structure; in most studies, β-carbons are model built from the N, C^α, and C′ atoms using standard bond lengths and bond angles. True homology-based protocols will include rules for how to use the structural information available in the template structure(s).

Like other protein modeling tasks, side-chain modeling can be divided into knowledge-based and energy-based methods. It is also possible to devise effective protocols that combine aspects of both kinds of methodology. A key idea in several of these methods is that of the rotamer library; i.e. the approximation that side-chain conformations can be represented fairly well by a limited number of canonical combinations of dihedral angles χ that correspond closely to the trans, gauche-plus, and gauche-minus values expected from simple stereochemical considerations. Early observations on side-chain conformations in proteins[623] and peptides[624] appear to support the applicability of this approximation; subsequent evaluations[625,626] agreed with their conclusions, but most recently the rotamer approach has been seriously questioned.[627] For calculations using a particular energy function, theoretically derived libraries can also be used.[196,197,589,628–630]

A prototypical rule/knowledge-based algorithm for side-chain modeling has been proposed by Sutcliffe et al.[554] In this procedure, explicit rules are given for how to change from one amino acid type in the template structure(s) to the desired one in the target structure. When there is no correspondence (i.e. in loop regions) or when changing from Gly or Ala to longer residues, the most common rotamer in the backbone-dependent library of McGregor et al.[631] is used. Steric clashes between newly built side chains and backbone, or between pairs of side chains may exist, and are expected to be relieved later by minimization or MD procedures. An application of a similar strategy was reported by Reid and Thornton, starting from the α-carbon trace of flavodoxin.[632] Steric conflicts were resolved by a combination of computer graphics-assisted human intervention and limited use of refinement techniques.

Summers and Karplus[603,633,634] have refined the simplest knowledge-based protocols, by explicitly integrating conformational energy-based techniques at several steps. A detailed review of their methodology has been presented recently,[603] and we describe the highlights here. Perhaps the key innovation of their method is the use of what they called "energy fingerprints", which for side chains are based on rigid-rotor (RR) energy maps. As in work described above, they start with a series of rules about which atoms to keep from the template structures, but then energy criteria are used to study ways in which missing atoms (including polar hydrogens) can be added to the model. RR maps are computed in turn for all the side chains being built, and localized refinements are carried out whenever these RR's are not compatible with what is expected from full experimental structures. (Some of this follows the earlier work of Gelin and Karplus.[635]) They have described a complete application to the C terminal domains of aspartic proteinases.[634]

Independently, Levitt[611] and Wendoloski and Salemme[569] have presented similar data-based procedures to model missing portions of proteins; they call their methods SEGMOD and PROBIT, respectively. In particular, these procedures could be used to construct side chains, given the backbone; or to implement the necessary substitutions while modeling by homology. In both approaches, searches for appropriate segments are conducted in a data base of high-resolution protein structures. In Levitt's procedure, averaging over random choices is followed by full energy minimization with tight constraints on the α-carbons. In a test calculation of eight high-resolution proteins, the SEGMOD method[611] was able to build full-atom models from experimental α-carbons with rms deviations over all atoms of 0.93–1.73 Å (0.3–0.4 Å for backbone, and 1.4–2.4 Å for side chains). A particularly attractive feature of SEGMOD is the totally automated, and thus unbiased, fashion in which the calculations are carried out. (The test proteins were, of course, removed from the data base in these calculations.) By working directly on the protein data base, these two procedures, PROBIT and SEGMOD, do not deal explicitly with rotamer definitions or substitution rules. A very similar computer program, XCHAIN, was used to assist model building in the X-ray analysis of the actin: Dnase I complex.[636] These methods can be seen as important generalizations, to include side chains, of the original template building procedures of Jones and co-workers.[567,604]

Partial rebuilding protocols that use extensive energy evaluations include the single point mutant calculations of Shih et al.,[637] the coupled perturbation procedure of Snow and Amzel,[638] and the iterative minimization calculations, LECS (local energy conformational searches), of Schiffer et al.;[107] the latter uses an empirical solvation procedure[119] to help choose the best rotamers. Other protocols proposed and tested for peptide segment modeling are those of Bruccoleri and Karplus as part of the CONGEN program,[589] and of Gibson and Scheraga as a component of their implementation of the build-up procedure.[199] The CONGEN method was used later to refine[639] models of misfolded proteins,[640] and to build complete models of antibody loops,[209,590] as mentioned earlier.

Several studies have attempted to calculate side-chain conformations by using variants or refinements of the rotamer library idea. There are three major ways in which the search space can be defined. The first and most common defines sets of discrete rotamers, and some kind of combinatorial search is conducted in the resulting (discrete) space. A second way, introduced by Lee and Subbiah,[641] uses a fine, 10°, grid for each side-chain torsion angle. The major, practical, difference between these two approaches is that the first one allows precalculation of all the possible residue–residue pairwise interactions. Thus, the energy evaluation during the combinatorial search reduces to look-up and summation over table entries. This is in general not possible or practical when using fine torsional grids. A third approach abandons discretization schemes altogether and, in principle, examines the continuum of torsion-angle values. Published approaches also differ in the kind of energy functions used, as well as in the details of the combinatorial algorithms used.

In the work by Ponder and Richards[625] on tertiary templates, the main objective was to generate all possible sequences compatible with a given three-dimensional structure; thus, exhaustive enumerations had to be carried out. This limited their application to small sets of amino acids within buried hydrophobic cores. By relaxing these requirements, and by redirecting the goal of the calculations toward producing low-energy packing arrangements for a *single* sequence, a number of groups have developed algorithms for side-chain calculations. Tufféry et al.[329,642] extended the rotamer libraries of Ponder and Richards by considering a larger number of protein structures; they then proposed and tested several algorithms to solve the side-chain packing problem. Holm and Sander finetuned a very rapid simulated annealing algorithm for side-chain packing.[613,643] They used a simple nonbonded energy function, consisting of a Lennard-Jones 9-6 potential. The softer $1/r^9$ (versus the more common $1/r^{12}$) repulsive term appears to correct approximately for possible errors introduced by the discretization simplification implicit in rotamer-based procedures. More recently, they have added an empirical solvation function procedure for ranking and evaluation of protein models.[133]

Wilson et al.[644] presented a heuristic search procedure to analyze the side-chain optimization problem in "rotamer space". Perhaps the most interesting aspect of this work is the use of an empirical functional form that includes an approximate treatment of solvation. The parameters for this function had been optimized in earlier work[645] that built an empirical correlation between calculated and experimental data on the relative catalytic efficiency of variants of α-lytic protease on a number of peptide substrates.

A considerable extension of the backbone-dependent rotamer libraries of McGregor et al.[631] has been carried out by Dunbrack and Karplus,[629] who also presented an integrated protocol for calculation of side-chain structures. Their procedure includes extensive energy minimization with the CHARMM force field.

A very interesting rotamer-based optimization scheme was described by Desmet et al., as the dead-end elimination (DEE) theorem.[646] Unlike many of the procedures outlined above, which rely on Monte Carlo or, in general, stochastic approaches to tackle the combinatorial problem, in DEE a more systematic, yet practical, approach is pursued. The main idea is to eliminate hierarchically rotamer conformations that cannot possibly be included in the global minimum (of course, within the discrete space spanned by all rotamer combinations). In more detail, the total energy of the protein can be expressed as

$$E_{\text{total}} = E_{\text{template}} + \sum_i E(i_r) + \sum_i \sum_{j>i} E(i_r j_s) \quad (36)$$

where E_{template} is the self-energy of the backbone (including β-carbons) plus fixed side chains (if any), $E(i_r)$ is the interaction energy between residue i in its r rotameric conformation and the fixed template, and $E(i_r j_s)$ is the pairwise interaction between conformation r of residue i and conformation s of residue j. As mentioned above, it is feasible, for relatively small numbers (three to a few hundred *per residue*) of rotamer possibilities, to precompute all possible values of $E(i_r)$ and $E(i_r j_s)$. Having these results, one can identify rotamers i_r incompatible with the global minimum-energy conformation (GMEC). Thus, conformations including such incompatible rotamers can be categorized as dead-ending, and eliminated from a subsequent systematic or stochastic combinatorial search. The single rotamer DEE theorem can be expressed as follows: if for two rotamers i_r, i_t of residue i the inequality

$$E(i_r) + \sum_j \min_s E(i_r j_s) > E(i_t) + \sum_j \max_s E(i_t j_s) \quad (37)$$

holds true, then i_r is incompatible with the GMEC. The original paper presented extensions of DEE to pairs of rotamers, but the formulation for these cases was modified somewhat by the same group and renamed the *fuzzy-end* elimination procedure.[647]

Realizing that the rotamer approximation represents an idealization of the conformational possibilities of side-chain conformations, Lee and Subbiah[641] proposed a new model in which torsion angles are mapped to a 10° grid. In this model, amino acid side chains described by only one dihedral angle χ can have up to 36 possible conformations (compared to just three for most rotamer libraries), while longer ones will have well over 1000 conformations (e.g. 36^2 or 1296 for residues with two χ's). This precludes straightforward generalization of the precalculation strategy, which is key for the efficient implementation of rotamer-packing calculations. However, some degree of precalculation, e.g. for side chain–backbone interaction terms, is still possible. In the original analysis of this model, Lee and Subbiah used an MC simulated annealing protocol followed by statistical averaging over the set of the few thousand lowest energy conformations found during a long run. A modification based on self-consistent mean-field ideas (see section II.B.6) has resulted in considerable speed up of the algorithm.[648]

Argos and co-workers[627] have questioned discretization approaches by showing that only 70–95% of amino acids in proteins have conformations within 20° of idealized rotamer definitions. They later proposed and tested a method[649] that combines the simplest stereochemical considerations of side-chain torsion angles (i.e., they use −60°, 60°, 180° for most χ's, and 0°, 90° for aromatic χ^2's) with extensive continuum optimization in side-chain torsional space using the ECEPP/2 potential.[61,62] They presented tests for six proteins, and also carried out a homology modeling exercise. They argued that the combinatorial problem in side-chain calculations is nonexistent and, indeed, their procedure examines only one side chain at a time, except in the last step in which a *local* optimization, varying all individually placed side chains, is carried out.

Interesting procedures for side-chain computations have been described by Fine and colleagues in a report of the modeling of an anti-CEA antibody.[598] In their protocol for side-chain modeling, they first apply a scheme in which an original set of typical rotamers is "customized" for the particular three-dimensional environment of a given residue by optimization in torsion space. Rotamers that converged to conformations that were the same within a 10° window were collapsed into one. Thus, for example, two leucines in different parts of the sequence may end up with different sets of optimized rotamers at this stage. They then applied one of three procedures to tackle the resulting combinatorial problem, while the method of Argos and co-workers[649] seems to stop right after the equivalent of the customization stage of Fine and co-workers. The three procedures outlined by Fine's group include a systematic search (applicable for only small numbers of side chains), MC simulated annealing, and, third, a method they call NOVOSIDE, which performs an iterative self-consistent heuristic search. Unfortunately, they did not conduct test calculations on known structures, and thus we cannot presently compare the performance of their protocol with other methods. In any case, the idea of customizing the initial rotamer library is quite attractive, and in theory it should be able to supersede the use of backbone-dependent rotamer libraries.

Other procedures have used MD to "sprout" side-chain atoms from backbones or from α-carbon chains. The methods of Nilges and Brünger[431] and of Correa[622] are direct extensions of the procedures described in the previous section. David[650] has used ideas originally presented by Nilges et al.[244] (for protein structure determination from NMR) to develop a general method for side-chain calculation. Finally, we mention here again the mean field-related LES procedure of Roitberg and Elber[317] (see also section II.B.6).

d. Validation of Models. Experimental reports of protein structures derived by either X-ray crystallography or NMR very often include a series of validation tests in addition to measures of agreement with experimental data (e.g. R factors). These tests include plotting of $\phi-\psi$ pairs to look for possible occurrences in "forbidden" regions of the map, number and quality of hydrogen bonds, deviations from ideal bond lengths and bond angles, etc. It thus follows that homology-modeling exercises should include at least these tests. The number and kind of tests of this nature depends on the percent of identical amino acids between sequence of the target and that (those) of the template structure(s). In the limit of very high sequence homology, the key assumption of similar overall fold is justified, and thus the emphasis in the validation part of the study should be placed on the stereochemical quality of the results. Careful energy refinement involving sequential restrained MD or minimization will often result in high-quality three-dimensional models, although the effects of even single mutations in cores of proteins are still hard to predict in detail. (See for example the X-ray structural determinations of λ-repressor mutants by Lim et al.[651] and their comparison with previous calculations.)

When homology is low, one cannot rely on stereochemical quality alone. As the intriguing study of Novotný et al.[640] clearly shows, it is quite possible to achieve very good potential energies (especially in vacuum) even when homology is nonexistent. In their influential study, they took two protein domains of identical length and known structures, the VL domain of the antibody REI and the hemerythrin molecule, and modeled one on the backbone of the other, thus producing REI-like hemerythrin and hemerythrin-like REI. In both cases, they obtained final energies comparable to those of the optimized correct structures.

Other properties of model structures that one should consider include distribution of hydrophobic and hydrophilic groups in the core and surface,[120,639,652] polarity of the protein surface,[653] packing of regular structure elements, etc. The 3D-1D Profiles[654,655] and similar[656] methods are often very useful here as well. Simple solvation potentials[119,639] can distinguish the correctly from the wrongly folded structures in the REI–hemerythrin example mentioned above. Many of these tests can evaluate parts of the model, and thus suggest relative quality for different segments along the sequence. It is also important to realize that, in the limit of low homology, one is usually forced to consider more than one alternative model, and thus validation protocols become more critical.

e. Consensus Approaches. An alternative to the step-by-step homology building strategy outlined in sections III.E.1.a–c is to use features of the template protein(s) to define an objective or pseudo energy function, which is then optimized for the sequence of the target. The simplest and oldest expression of this idea is in terms of distances, and indeed is a possible way to generate models consistent with standard covalent geometry.[545,548,591]

A significant generalization is the use of information from homologous families to define different distance ranges; the original description of this kind of approach seems to be that of Srinivasan et al.;[657] Havel and Snow presented a very elegant procedure using distance geometry,[658] and shortly thereafter Srinivasan and co-workers described a refinement based on similar ideas.[659,660] Snow has also described an alternative formulation for parameterization[264] accompanied by an effective simulated annealing for optimization.[263] More recently, Šali and Blundell developed the program MODELLER,[661,662] which generalizes this approach to consider many other structural features beyond only distances. The resulting, complex, optimization problem is tackled by a technique similar to the variable-target–function method of Braun and Gō.[206]

While presently more *computer* time-consuming than more conventional methods, the consensus rebuilding strategies described in this section tend to be more amenable to automation. Thus, they may in many cases be less "*human*" time-consuming. Also, their computational structure describes more rigorously the assumptions behind the entire homology modeling protocol; this may lead to a more straightforward evaluation of the faults and merits of the final models. Finally, when used with a large data base of homologous structures, these procedures may still produce reliable results when the percent of sequence identity goes below about 25%. Using the classical methods of Greer[549,550] and of Blundell and co-workers,[555,556] the modeling may be reliable only if the percent identity between target and template sequences is above the 25–30% range.

2. Homology Extension

Perhaps the most reliable way available today to predict overall protein folds is by homology modeling. It is thus not surprising that a great deal of effort has been spent in attempts to extend these methods to cases with no detectable sequence homology. This approach is justified by the observation made by Chothia that the hundreds of proteins of known three-dimensional structure present a limited number of unique folds.[663] This is especially true if one looks only at protein domains: most larger proteins have multidomain structures. The analysis of Dorit et al. of the protein *sequence* data base[664] lends further support to this view. In this, limited, formulation of the protein folding problem, one has to test only the likelihood that a given sequence will attain a particular known fold (the "inverse" folding problem). The main weakness of this formulation, of course, is the dependence on previously observed three-dimensional structures. A more extensive discussion of the limitations of this approach was given recently by Sippl.[11]

The realization that the number of unique folds may be limited, together with the rapid pace at which protein structural information is becoming available, has spawned an entire subfield whose goal is the rapid and objective characterization of folds beyond the classic taxonomy and anatomy of Richardson[408] to provide computationally more useful definitions. A key component of this work is methodology for comparison of protein structures; examples include the approaches of Taylor and Orengo[665] and Šali and Blundell[666,667] that use generalizations of the dynamic programming method that include multiple comparisons of structural features, the graph theoretical methods of Willet and co-workers,[668] and the recent work of Rose and Eisenmenger[669] and Zuker and Somorjai.[670] The last two most closely resemble the initial approaches for structural comparison pioneered by Rossmann,[442,671] and by Remington and Matthews.[672,673] Another useful methodology in this context is the differential geometry description of polypeptide α-carbon chains of Rackovsky and Scheraga.[674,675] It is possible to use this structural description, based on concepts of the differential geometry of space curves, to characterize polypeptide chains at the length scale of four or five residues. More recently, the method has been generalized to make it applicable to arbitrary length scales, and it has been used for the classification of known protein structures.[676]

These homology extension procedures are not completely different from more *de novo* procedures, and in fact Taylor[677,678] has mapped a continuum of methodologies ranging from regular local structure predictions to the more favorable cases of homology modeling reviewed above.

One avenue to carry out the fold-sequence recognition is by analysis of many aligned sequences. The promise of this general approach may be illustrated by the success of Bazán in predicting the structures of cytokines and cytokine receptors[679–681] (most recently in the correction[682,683] of the originally incorrect[684] interleukin-2 structure), and the early accurate predictions of Benner and co-workers.[685,686] Other examples of somewhat successful application of these principles include the prediction of the overall fold of the HIV protease by Pearl and Taylor[687] from the structures of eukaryotic aspartic proteases, and by Barton and co-workers of SH2 homology domains.[688] Both of these were subsequently confirmed by high-resolution structural data.[689–693]

Related in philosophy is the 3D-1D Profiles method of Bowie, Lüthy, and Eisenberg,[654] which detects *structural* similarity between actins and 70 kDa heat shock proteins, although there is no detectable *sequence* similarity between these two classes of proteins. This work can be seen as a three-dimensional extension of the *sequence profile* method of Gribskov et al.,[694] which in turn is a key component of the pattern matching approaches of Bazán[680,695] and Benner.[686] (It should be noted, however, that the methodology used by Benner and co-workers to derive structural predictions from sets of aligned sequences is quite different from that of other workers.[696]) A procedure that combined multiple sequence information with solvent accessibility had been proposed earlier by Bowie et al.[697] while Lüthy et al.[698] developed a similar approach using information about regular local structures. In this context, the 3D-1D Profiles method can be seen as the natural combination of these three approaches.[694,697,698] Specialized early versions of these ideas may be found in the definitions of tertiary templates for immunoglobulins by Taylor[699] and for globins by Bashford et al.[700]

A scheme related to the 3D-1D Profiles method is the template-matching procedure of Overington and co-workers.[701,702] This is an extension of the amino acid comparison matrices commonly used in protein homology work. Their modification aims to exploit aspects of three-dimensional structure known in at least one of the proteins being compared. In fact, an appropriate summation over amino acid types may reduce their matrices to the type of parameterization developed in the Profiles method of Bowie et al.[654]

After the initial publication of the methods of Bowie et al.[654,655] and Overington et al.,[701,702] several laboratories have proposed refinements: Godzik et al.[89] included pairwise and three-body terms in the profile calculation, Jones et al.[703] presented a combination of the Profiles idea with the potential of mean force approach of Sippl[11,92] (see also Ouzounis et al.[704]); and Maiorov and Crippen[103] extended earlier work by use of a solution of a system of coupled inequalities that ensures that the derived parameters will indeed favor observed protein structures. (This work should be considered in the context of previous efforts by Crippen and his co-workers starting with design of potential functions for use with the energy embedding procedure; see sections II.A.2 and II.B.5). A detailed examination of different types of effective "knowledge-based" potentials has been given recently by Kocher and co-workers.[705]

Another interesting approach is the hidden Markov model of Stultz et al.[706] for estimating the probability that each amino acid in a protein sequence is part of a particular type of regular local structure. Other applications of hidden Markov models to proteins have appeared recently.[707,708]

We close this section by mentioning the recent prediction of a three-dimensional model for the hormone-binding domains of steroid receptors.[709] Using a combination of techniques developed earlier by Wolynes and co-workers,[104,105,710,711] these researchers have proposed that there is structural homology between the steroid receptors and subtilisin-like proteases, even though there is no detectable sequence similarity between these two families of proteins. Of course, final evaluation of this interesting proposal would have to await further experimental work on the system.

3. Simplified Protein Models

Simplified computational methods can be used to investigate some of the overall features of the conditions for compact folding of the polypeptide chain and of the course of the folding process. Useful insights have been gained from a variety of computations in which a self-avoiding polymer is generated on a two- or three-dimensional lattice, and occasionally in off-lattice calculations. The polypeptide chain is represented by virtual bonds between points that correspond to the centroids, or to the α-carbons, of amino acid residues. In the case of lattice calculations, the restriction to discrete positions makes exhaustive modeling computationally feasible, because the number of conformations remains finite and countable. The use of simplified forms of the interaction between residues in contact, instead of an actual distance-dependent force field, helps to reduce computational time. This approach lends itself particularly well to the modeling of the folding process by Monte Carlo simulation and, for short enough chains, by complete enumeration of states. A general review of simplified protein models has appeared recently.[712]

Although the simplest lattice models do not reproduce details of specific interactions or the folding of particular proteins, they have been very useful because they provide general insights into several types of constraints that are presumably significant in determining which folded structures are feasible in a protein. Constraints investigated in this manner have included the compact filling of space, the preferential localization of polar and nonpolar amino acid residues on the surface or in the interior, respectively, the balance of long- and short-range interactions, and the cooperativity of collapse into a compact structure.

Currently these simplified models are being studied with two major purposes. One category of studies expects to reveal physical issues of thermodynamics and kinetics of protein folding; it is not the immediate goal of this work to provide a practical protocol for prediction of three-dimensional structures from amino acid sequences. Dill and co-workers,[31,713] Shakhnovich et al.[714,715] and, at the earlier stages, Skolnick and collaborators[716,717] have attempted to achieve this goal. A second group of studies has used slightly more complex models that at least in theory attempt to describe "real" proteins. The early work of Levitt and Warshel,[81,82] Tanaka and Scheraga,[718-720] and others, as well as some of the more recent calculations by Skolnick and co-workers,[10,434] fall in this category. These studies focus more sharply on the prediction of structure from sequence, and may or may not address the problem of folding pathways. (See the recent perspective by Levitt[28] for a similar discussion on the definition of the "protein-folding problem".)

Back in the early and mid 1970s, several simplified protein models were proposed and studied. Levitt and Warshel[81,82] proposed a united-residue potential, which they then used to simulate protein folding by an *ad hoc* optimization procedure in continuous space (i.e., not on a lattice). This approach, which included use of pulling and pushing potentials, was criticized by Némethy and Scheraga[1] and by Hagler and Honig.[721] The criticisms were largely justified, but in retrospect, the pioneering value of the approach of Levitt and Warshel is reflected in a major part of the recent progress in this area.

Tanaka and Scheraga[718-720,722] studied a hierarchical model that made use of statistical trends derived from the then tiny structural protein data base: one portion dealt with short-range interactions and was very similar to secondary structure prediction procedures; the long-range part consisted of a contact potential of mean force derived from data on 25 proteins; the model was driven by a Monte Carlo procedure that produced changes in local conformation with probabilities given by the short-range prediction scheme. (A very similar method has been described and applied in a folding simulation of myoglobin.[723]) The procedure, like other similar ones developed independently at about the same time, was largely unsuccessful, partly because not enough computational power was available, and the potentials derived from the small protein data base were noisy and unreliable. However, one should point out the many similarities between these approaches and the newer ones of Gregoret and Cohen,[724] Skolnick and co-workers,[10,434,725] Sippl and co-workers,[92,726-728] Sun,[96] etc.

Other early work includes the study of Kuntz et al.,[729] which was very similar to that of Levitt and Warshel, and introduced a few novel points. Havel et al.[730] studied a similar problem, but this time used the then-emerging tools of distance geometry. The latter work is important in two respects: first, it anticipated the eventual success of structure determination by NMR, and second, it suggested useful avenues for prediction schemes. Similar work followed by Wako and Scheraga,[731-733] and by Goel and co-workers.[734,735]

The use of lattice models for the study of protein folding was introduced by Gō and co-workers,[736-738] who generated models of polypeptide chains on a two-dimensional square lattice, using the Metropolis Monte Carlo method. The specificity of interaction between various residues and the balance of long- and short-range interactions was varied as a parameter of the computations. It was shown that a cooperative transition between unfolded and compactly folded forms requires favorable specificity between at least some of the amino acid residues. As described later in this section, more recent work, using better lattice protein models, has shown that it is possible to obtain cooperative transitions in the absence of artificial specific interactions. Specific long-range interactions are essential for highly cooperative stabilization of the "native" compact conformations, while short-range interactions primarily accelerate folding and unfolding transitions. Attempts to extend these lattice calculations to three dimensions were not very successful mainly because their model lacked any built-in handedness preferences.[739] This caused, for example, formation of right- and left-handed α-helices in the same simulation of a model protein. Eventually, poor packing of these incompatible types of structures prevented observation of protein-like folding transitions in the more interesting three-dimensional lattice.

Work on simplified models was not very active throughout the early to mid 1980s: real progress of models that rely on pseudopotentials derived from statistics of protein structure would have to await the explosion of experimental structural information that has occurred in the last five years. Interest in lattice models of proteins was largely revived by the independent efforts from the laboratories of Dill[713,740] and Skolnick.[716,717] Both of these approaches are reviewed in some detail below.

Dill and co-workers have proposed a two-dimensional square lattice polymer as a model for protein folding.[713,740-742] Two types of amino acids are allowed in this model. By examining compact conformations only, it is possible to enumerate exhaustively all possibilities up to chain lengths in which protein-like behavior can be observed. One of the most interesting conclusions of this work is that, at least in this model, regular local structure formation, i.e., helices and sheets, is a consequence of compactness, a result which is consistent with theories about the driving forces for folding in proteins.[713,742-745] In a square or cubic lattice the "helices" and "sheets" are only analogs and not exact representations of these structures in real globular proteins. This conclusion, however, may be a consequence of the square lattice formulation as simulations with an equally simple but continuum model by Gregoret and Cohen did not detect such a strong correlation between compactness and formation of regular local structure.[746] However, compactness plus short-range interactions (chain stiffness) do lead to correct average local chain conformations observed in protein structures.[745]

The simplicity of Dill's model allows a fairly complete exploration of sequence space, as well as of conformational space. It was found that only some sequences have globular protein-like behavior in the sense that they have a nearly unique preferred conformation. Similarly, the effect of mutations can also be treated in the lattice model by switching the character of some residues between polar and nonpolar. According to the model, mutations may be neutral in some positions along the chain, i.e. not affect the conformation of the most favorable "native" fold on the lattice, but in other positions they may alter the "native" conformation either by changing the balance of interaction energies or even by changing the degeneracy of compact states, i.e. by affecting the entropy of unfolding.[747,748] Some of the "denatured" conformations are also highly compact, because of numerous contacts between nonpolar residues. Consequently, replacing a polar residue on the surface of the "native" state by a nonpolar one may shift the equilibrium away from the "native" state.[748] These lattice computations required exhaustive enumeration and thus have been limited to short chains of less than 30 amino acids; hence, the validity of the conclusions for longer polypeptide chains has yet to be tested. It is thus important to point out that extensions of the model to longer sequences have been pursued by O'Toole and Panagiotopoulos using a scanning Monte Carlo algorithm,[211,212] and by Unger and Moult[324,325] using genetic algorithms (see also section II.B.7).

More recently, Dill and co-workers have outlined a hydrophobic zipper (HZ) hypothesis of protein-folding cooperativity, and used their simplified two-dimensional lattice model proteins to test the efficiency with which an algorithm based on this hypothesis searches conformational space.[749,750] The authors have proposed the HZ as a method of assembly, not as a process of physical kinetics, because their scheme makes the assumption that there is no reverse rate for the formation of hydrophobic contacts along the "pathway". Thus, this idea may be more useful as a basis for the design of heuristic protein-folding algorithms[750] than as a correct physical description of the folding process. The HZ idea itself is similar to the hydrophobicity-driven "nucleation site" formation proposed earlier by Matheson and Scheraga.[751]

Shakhnovich et al.[715] and Leopold et al.[14] have carried out further explorations of lattice proteins, including the extension to three dimensions and to longer chains. They suggest that protein-like sequences not only have thermodynamic preferences, but also encoded kinetic pathways. Thus, failure to "fold" within a reasonable time (using their MC algorithm) is interpreted as a property of the particular "amino acid" sequence (they use a simplified two-residue-type alphabet). The work of Shakhnovich et al. has been criticized by Skolnick and co-workers,[10] who argue that the MC procedure used in the former work may not be sampling space efficiently enough to justify the generality of their conclusions. More recently, Šali et al.[752] have studied the kinetic behavior of the same[715] simple cubic lattice model for proteins. They first carry out exhaustive enumeration over the compact states of a 27-residue model protein. In this manner they can identify the global energy minimum, and call the corresponding structure the "native" state. They then attempt to carry out a refolding simulation from random conformations using a Metropolis Monte Carlo method, and thus record how often, and how quickly the simulation leads to the native state of the model. The most interesting conclusion is that the ease with which a given sequence can fold increases with the energy gap between the energy of the global-minimum structure and that of the structure with the next lowest value. In other words, they state that the necessary and sufficient condition for a sequence to fold efficiently is that the "native" state be a pronounced energy minimum.

Leopold and Shakhnovich have presented an alternative philosophical justification of protein-folding calculations, which is based not on the traditional thermodynamic hypothesis, but on a kinetic point of view.[16] Also of interest is the work of Shakhnovich and Gutin on establishing how the heteropolymeric nature of proteins plays a major role in the existence of unique preferred conformations.[714,753] They studied maximally compact structures of a 27-residue chain on a three-dimensional cubic lattice and concluded that the heterogeneity of interactions between amino acids was crucial in favoring the possibility of one "native" conformation over all the others.

Shakhnovich and Gutin have also studied in more detail the necessary sequence conditions for globular protein-like behavior.[754,755a] In particular, by carrying out Metropolis Monte Carlo in the space of variable sequences but fixed composition, they searched for sequences that presented a much preferred compact conformation. They used only two types of amino acids and very simple forms of interactions. Detailed conformational study of these "ultra specific" sequences showed behavior similar to what one expects of real proteins, but without any special parameters in the potential function. Thus, it is possible to build more realistic models of protein folding by proper choices of sequences, while keeping the interaction potentials simple and generic. These results were obtained from analysis of 27-monomer chains. More recently, working with a more realistic 80-monomer model protein, Shakhnovich[755b] has determined that two-monomer heteropolymers are not specific enough to have a unique structure (for a particular compact structure defined as "native"), while it is possible to have sequences of 20-monomer heteropolymers that do have a unique preferred structure, and thus protein-like behavior. In this work, a compact target structure is chosen, and then the "sequence design" procedure developed earlier[754,755a] is applied to find an optimal sequence (for a given, fixed composition) for that structure. This designed sequence is shown to have a pronounced energy gap between its "native" conformation and all the other ones (compare with the work in ref 752), and is also shown to fold rapidly in simple lattice Monte Carlo simulations starting from random coil conformations. The author remarks that the thermodynamically oriented sequence design procedure at the same time solved the kinetic protein folding problem by making the "native" structure kinetically accessible.

Some of the work described above has used the analogy between protein folding and spin glasses first studied by Bryngelson and Wolynes.[756,757] Karplus and Shakhnovich recently reviewed the more theoretically oriented work in the field, and offered an interesting and challenging perspective.[30]

Skolnick, Kolinski, and co-workers have developed a set of more complex three-dimensional lattice models, with several versions differing by lattice coordination and by the number of interaction centers representing each amino acid residue. In particular, they have introduced a lattice in which successive residues occupy nonneighboring lattice sites selected in the manner of the "knight's walk" in chess,[758] together with the use of an additional lattice point to represent the side chain. This lattice provides a more realistic stereochemistry for virtual bonds. A critical evaluation of the behavior of several lattice models has been given.[759]

Recently, Hao and Scheraga[760] have used an entropy sampling Monte Carlo procedure[761] (formally related to the multicanonical ensemble[253]) to treat a 38−residue model protein on a Skolnick−Kolinski type lattice[758] with three amino acid types (neutral, polar, and nonpolar) to demonstrate that the folding/unfolding involves a first-order transition. Their simulation procedure included the set of local Monte Carlo moves used by Skolnick and co-workers,[758,762] augmented by more global moves similar to those used in the chain-building method of Rosenbluth and Rosenbluth.[210] Using selected sequences, they further showed[763] that optimized sequences exhibit long-range cooperative behavior in the folding transition, while random sequences reflect only short-range (i.e.

nearest-neighbor) cooperative phenomena. Furthermore, optimized sequences fold to unique lowest energy structures, but random sequences, when cooled in Monte Carlo-simulated annealing simulations, fold to compact but random structures. The numerical simulation results[763] were also fit to an analytical formalism based on the mean-field theory of Bryngelson and Wolynes,[756,757] thereby providing further insights into the folding behavior of polypeptides with different sequences.

The earlier simulations by Skolnick and co-workers were geared toward deducing generic aspects of protein folding with simplified sequences, and showed folding-like behavior. The conformational possibilities of these models have been explored by a very efficient dynamic Metropolis-like Monte Carlo algorithm.[762] Perhaps more interestingly, proper statistical choice of moves leads to dynamic behavior consistent with polymer Rouse dynamics in the limit of weak long-range interactions[764,765] (see also recent work by Knapp[766]); this enables one to establish at least an approximate correspondence between real-time and sample number in the Monte Carlo simulation. The simplicity of the model coupled with the efficiency of the method allows observation of a number of folding–unfolding transitions within a single simulation.

Some of the most interesting observations in the calculations by Skolnick's group described above required imposition of weak local conformational preferences taken from an actual protein structure;[192] however, the formation of well-defined three-dimensional folds is still of great interest. Moreover, folding trajectories, with stable intermediates, could be generated, and structures resembling the actual structures of apoplastocyanin and of triose phosphate isomerase were obtained, with 6 and 3 Å rms deviations, respectively, for the C^α positions.[192] It should be stressed that these results, however, cannot be considered indicative of a general solution of the protein folding problem, because they used specific local preferences derived from actual knowledge of the experimental structures.

From a more practical point of view, these results of Skolnick and co-workers seem to suggest that more accurate prediction of local structure preferences may be exploited in generation of topologically correct three-dimensional structures to a greater extent than it was thought possible. In this context, we point out that the proposed "solution" of the Levinthal paradox by Zwanzig et al. is consistent with these results:[191] biases in local conformational preferences toward the "native" states may accelerate the folding process significantly. This does not, of course, mean that the existence of these biases in "real" proteins has been demonstrated, or that the Zwanzig model represents the folding behavior of proteins accurately; in fact, this model has been criticized recently by Karplus and Shakhnovich.[30] These authors argue that Zwanzig's approach is more descriptive of helix–coil transitions where long-range interactions do not play an essential role, and as such, it misses the fundamental physics of protein folding. In the apoplastocyanin calculations as well as in more recent work, Skolnick and co-workers have started using pseudopotentials derived from the protein structural data base, initially with the parameters of Miyazawa and Jernigan.[90] Thus, their work has taken a more practical angle, and their models are now in principle capable of handling real amino acid sequences.[10]

Recently, Skolnick and co-workers have removed the dependence on preset local preferences and have proposed a more general approach.[434] Although it is not clear if the many-body terms in their potential functions can be defined for general sequences with no previous knowledge of the type of structure, their recent predictions of a pair of designed α-helical proteins are remarkable. In accordance with experiments reported later by DeGrado and co-workers,[767] they proposed that an earlier version of the designed protein would not show preference for one handedness of a four-helix bundle over the other.[434] They also proposed that a second version of the designed sequence would show a definite preference for right-handed over left-handed four-α-helix bundles. We point out that the designed sequences have a large degree of structural redundancy built in, and in some sense may be "easier" to fold than an arbitrarily chosen natural sequence. In fact, these designed sequences may be the real-life counterparts of the "optimal folding sequences" studied by Shakhnovich and co-workers.[754,755] An alternative view, perhaps more consistent with the recent work of Shakhnovich,[755b] may be that most natural sequences are "easy to fold" but do so with some "exact" set of potentials, while simulations are forced to use *approximate* energy forms. By conforming to the best understood protein structural principles, DeGrado's proteins may be seen then as "easy to simulate" (and thus, to fold) with approximate potentials that reflect those same principles.

An innovative approach to modeling on a lattice has been presented by Covell and Jernigan.[768] They generate all conformations of a self-avoiding virtual-bond chain on a lattice that is restricted in size and shape to the volume occupied by a specific protein in its native state. A face-centered cubic lattice is used, because it is possible to fit the virtual-bond representations of the actual structure of five small proteins, having 36–62 amino acids, respectively, on this lattice with an rms deviation of about 1 Å. The total number of lattice conformations for these proteins ranges from about 800 to 15 000. The resulting conformations were ranked according to energy, using the empirical potential energy function of Miyazawa and Jernigan[90] for contact pairs. In all five cases, the conformation closest to the actual native structure of the protein fell within the lowest 2% of all conformations generated.

Covell and Jernigan propose that this procedure can be used to determine a small group of favorable ways of folding a specific amino acid sequence within a restricted space, for use as starting points in more detailed computations. The search method is quite efficient, but its utility for the prediction of an unknown protein structure depends on the possibility of delineating the restricted space narrowly. It was suggested that preliminary X-ray crystallographic data might supply this information (see, for example, the work of Subbiah on solution of the phase problem in protein crystallography[769]). The size restriction of the lattice is critical; however, the insertion of a mere two additional lattice points increased the number of conformations generated by an order of magnitude.[768]

More recently, Covell has tested a novel Monte Carlo procedure for folding of lattice proteins without imposing such strong shape constraints.[770] He still uses the Miyazawa–Jernigan contact potentials, together with the constraints imposed by polypeptide chain connectivity and by excluded volume considerations. Tests with real protein sequences led to relatively high rms deviations from the native conformations; in terms of these deviations, these results are not much better than those obtained much earlier by Kuntz et al.[729] and others.

A similar lattice approach is that of Hinds and Levitt,[771] where mappings containing two or more amino acids per lattice point are explored, together with empirical contact pair potentials. At the very least, the methodologies of Hinds and Levitt, and of Covell and Jernigan[768] will be of great utility in testing and improving different versions of these pair potentials, which, as described below, are still in great need of improvement.

Rabow and Scheraga have formulated the protein-folding problem in neural network optimization terms, using a method called lattice neural network minimization (LNNM).[306] The conformation of a protein is represented as an array of the amino acid sequence versus position on a three-dimensional face-centered cubic lattice (see also section II.B.6), with the Miyazawa–Jernigan energy function[90] defined in terms of the array variables. The energy function is minimized to locate the global minimum energy for the conformation of the protein. The LNNM method found the global minimum for a seven-residue peptide in all of the 15 runs carried out. The time for each run was ~30 s on one processor of an IBM 3090 computer. For a nine-residue peptide, the global minimum was found in seven out of 15 runs in ~50 s per run; the global minimum or the second lowest minimum was found in 10 of the runs. Starting from a uniform array for the protein crambin (46 residues) on the lattice, the energy of the crambin array was minimized and a compact low-energy structure was found in ~25 min of CPU time. Its energy was much lower than that of the native protein, suggesting that there are inadequacies in the Miyazawa–Jernigan potential. The LNNM method was also applied to the prediction of what were previously called nucleation sites[751] but are more properly called chain-folding initiation sites (CFIS)[333] of a protein. LNNM correctly predicted the CFIS for the two proteins examined, RNase S and T4 lysozyme.

Also of interest is the work of Head-Gordon and Stillinger on a two-dimensional "toy" model of protein folding.[772,773] They have started to use this model to develop a deeper understanding of local regular structure prediction schemes, especially methods based on neural networks. In this toy model some of the confounding factors in predictions of real proteins are removed. For example, one can start with a hypothetical data base that contains *all* the possible examples of conformations for short segments. As in other work with simplified models, the hope is that insights obtained in this artificial system could be of use in studies of more realistic models for proteins.

Thirumalai and co-workers have studied dynamical aspects of simplified off-lattice protein models.[774,775] In their most recent work, they have examined a series of two-dimensional lattice models. The novel aspect of this work is the characterization of the energy landscape by exact enumeration, and the partial elucidation of connecting pathways by extensive Monte Carlo pseudo-dynamics using the Metropolis algorithm.[776]

Results from work on simplified continuum, off-lattice, models can provide new insights into the protein-folding process, but they can also help to validate (or invalidate) some of the observations recorded in lattice protein calculations. In this respect, it is important to mention again the *quantitative* differences between off- and on-lattice calculations regarding the correlation between compactness and formation of local regular structures.[745,746] By the same token, the simulations by Fukugita et al.[777] of a *continuum* two-dimensional protein-like polymer confirm one of the main results obtained by Lau and Dill[747] and by Shakhnovich and co-workers[715,754,755] using two- and three-dimensional *lattice* calculations, respectively. This result is that folding is a rare phenomenon in the sense that only a very limited number of sequences will fold given generic interaction potentials. Fukugita et al. remark that one would expect a large difference in the magnitude of energy barriers between lattice and continuum models,[777] with the former being higher; this fact makes the agreement on this issue even more significant.

IV. Concluding Remarks

The past 40 years have witnessed a dramatic transition in the view of a protein molecule from a colloidal particle, described as a rigid ellipsoid of revolution, to a flexible organic molecule with a specifiable covalent and three-dimensional structure.[2] By the development and application of X-ray crystallographic and experimental physical methods, such as NMR and other spectroscopical techniques, it has been possible to elucidate the structures of protein molecules and functional complexes thereof in the solid state and in solution. Theoretical physical methods, including conformational energy computations, have contributed in an equally useful way to an understanding of the intra- and intermolecular interactions that give rise to polypeptide and protein structure. On the basis of the combination of experiment and theory, considerable knowledge has been acquired as to how polypeptide chains fold into the native conformations of proteins and then interact with other molecules to express their biological function.

The basic methodology for calculating structures of peptides and proteins is now well developed, as summarized in section II. Nevertheless, there is still significant room for improvement of potential functions, not only in the numerical values of the parameters used but also in the forms of the functions themselves. A number of significant refinements have been introduced and tested in model systems; these include use of anharmonicity in bond angle bending, the introduction of nonadditive potentials, particularly by the use of polarization to treat electrostatics, and improved treatment of hydration, especially by efficient but accurate implicit models. Effective generalization of these refinements for use in calculations of polypeptide and protein structure can be foreseen in the very near future.

With the recent development of various efficient approaches to overcome the multiple-minima problem, the problem has been in some sense solved for small oligopeptides, as well as for regular-repeating structures and assemblies of fibrous proteins. It can reasonably be expected that the current extension of these methods to globular proteins will result in efficient searches of their conformational space in the near future. Advances in computer hardware and software, especially the wider use of parallelism, will speed up computations, making practical the application to larger molecules. The point may have been reached where wider applicability of these techniques is becoming limited by the accuracy of the potential energy functions used to describe the energetics of polypeptide and protein structures.

A number of the computational methods reviewed here have been developed first by applying them to various model systems, where the computed results could be checked against known structures, in order to provide a verification of the algorithms and of the parameters. Model systems used for this purpose have included poly(amino acid)s that form regular structures, small linear and cyclic oligopeptides, and fibrous proteins such as collagen (sections III.A–D). In addition to providing necessary tests, computations on these model systems have resulted in some new predictions, and they have increased our understanding of the interactions that play a significant role in the folding of proteins.

It has been possible to extend the computations to globular proteins in various ways, viz. by using them in conjunction with experimental methods for structural refinement, by applying them in a predictive way such as in the calculation of structures of homologous proteins, or by using them to gain information about the role of short- and long-range interactions in the folding process (section III.E). Computations have led to useful insights into the reasons underlying the properties of proteins.

While dramatic results have been achieved in the last decade, there is still much more to be learned about the details of protein-folding pathways and of molecular interactions. In particular, many dynamic processes important in protein function, most notably the folding process itself, remain beyond the time scales that can be studied with present simulation techniques. The further development of the theoretical methods that will undoubtedly occur in the foreseeable future should further increase our understanding of protein structure and function.

Acknowledgments. We thank Drs. H. Meirovitch, U. H. E. Hansmann, Z. Zhou, and H. Qian for valuable discussions, and Drs. M.-H. Hao, D. R. Ripoll, and J.-Y. Trosset for reading the manuscript and offering helpful comments. We also thank Dr. N. Grønbech-Jensen for access to material prior to publication. This work was supported by research grants from the National Institute on Aging (AG-00322) and the National Institute of General Medical Sciences (GM-14312) of the National Institutes of Health, U. S. Public Health Service, and from the National Science Foundation (DMB90-15815).

Note Added in Proof

Since this review was completed, several reports on efficient molecular dynamics simulations of polypeptides in explicit solvent have appeared. They are listed in refs 778–780.

References

(1) Némethy, G.; Scheraga, H. A. *Q. Rev. Biophys.* **1977**, *10*, 239.
(2) Scheraga, H. A. *Carlsberg Res. Commun.* **1984**, *49*, 1.
(3) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631.
(4) Brüschweiler, R.; Blackledge, M.; Ernst, R. R. *J. Biomol. NMR* **1991**, *1*, 3.
(5) Scheraga, H. A. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1992; Vol. 3; p 73.
(6) Nikiforovich, G. V.; Prakash, O.; Gehrig, C. A.; Hruby, V. J. *J. Am. Chem. Soc.* **1993**, *115*, 3399.
(7) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598.
(8) Anfinsen, C. B.; Scheraga, H. A. *Adv. Protein Chem.* **1975**, *29*, 205.
(9) Abagyan, R. A. *FEBS Lett.* **1993**, *325*, 17.
(10) Godzik, A.; Kolinski, A.; Skolnick, J. *J. Comput. Aided Mol. Des.* **1993**, *7*, 397.
(11) Sippl, M. J. *J. Comput. Aided Mol. Des.* **1993**, *7*, 473.
(12) Levinthal, C. *J. Chim. Phys.* **1968**, *65*, 44.
(13) Levinthal, C. In *Mossbauer Spectroscopy in Biological Systems*; Debrunner, P., Tsibris, J. C. M., Munck, E., Eds.; University of Illinois Press: Urbana, Illinois, 1969; p 22.
(14) Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8721.
(15) Thomas, D. J. *FEBS Lett.* **1992**, *307*, 10.
(16) Leopold, P. E.; Shakhnovich, E. I. In *26th Hawaii International Conference on System Sciences*; IEEE Computer Society Press: Hawaii, 1993; p 726.
(17) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan., V. *J. Mol. Biol.* **1963**, *7*, 95.
(18) Némethy, G.; Scheraga, H. A. *Biopolymers* **1965**, *3*, 155.
(19) Ramachandran, G. N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283.
(20) Scheraga, H. A. *Adv. Phys. Org. Chem.* **1968**, *6*, 103.
(21) Karplus, M. *Ber. Bunsen-Ges. Phys. Chem.* **1982**, *86*, 386.
(22) Levitt, M. *Annu. Rev. Biophys. Bioeng.* **1982**, *11*, 251.
(23) Kollman, P. *Acc. Chem. Res.* **1985**, *18*, 105.
(24) Kollman, P. *Annu. Rev. Phys. Chem.* **1987**, *38*, 303.
(25) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics Structure and Thermodynamics*; Interscience Publ.: New York, 1988; Vol. 71, p 259.
(26) Scheraga, H. A. *Chem. Scr.* **1989**, *29A*, 3.
(27) Némethy, G.; Scheraga, H. A. *FASEB J.* **1990**, *4*, 3189.
(28) Levitt, M. *Curr. Opinion Struct. Biol.* **1991**, *1*, 224.
(29) Scheraga, H. A.; Némethy, G. In *Molecules in natural science and medicine - an encomium for Linus Pauling*; Maksic, Z. B., Maksic, M. E., Eds.; Ellis Harwood: Chichester, 1991; p 147.
(30) Karplus, M.; Shakhnovich, E. I. In *Protein Folding*; Creighton, T., Ed.; W. H. Freeman: San Francisco, 1992; p 127.
(31) Chan, H. S.; Dill, K. A. *Phys. Today* **1993**, *February*, 24.
(32) Elber, R. *Curr. Opinion Struct. Biol.* **1993**, *3*, 260.
(33) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(34) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(35) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(36) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. *Proteins: Struct. Funct. Genet.* **1988**, *4*, 31.
(37) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361.

(38) Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472.

(39) Levitt, M. *J. Mol. Biol.* **1983**, *170*, 723.

(40) Daggett, V.; Levitt, M. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 353.

(41) Gunsteren, W. F. v.; Berendsen, H. J. C. *The GROMOS Package*; Biomos: Groningen, 1987.

(42) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(43) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.

(44) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington DC, 1982; p 339.

(45) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(46) Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566.

(47) Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576.

(48) Lii, J.-H.; Gallion, S.; Bender, C.; Wilkstrom, H.; Allinger, N. L.; Flurchick, K. M.; Teeter, M. M. *J. Comput. Chem.* **1989**, *10*, 503.

(49) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.

(50) McDonald, D. Q.; Still, W. C. *Tetrahedron Lett.* **1992**, *33*, 7743.

(51) Brünger, A. T.; Kuriyan, J.; Karplus, M. *Science* **1987**, *235*, 458.

(52) Brünger, A. T.; Karplus, M.; Petsko, G. A. *Acta Crystallogr.* **1989**, *A45*, 50.

(53) Okamoto, Y.; Kikuchi, T.; Nakazawa, T.; Kawai, H. *Int. J. Pept. Protein Res.* **1993**, *42*, 300.

(54) Mazur, A. K.; Abagyan, R. A. *J. Biomol. Struct. Dyn.* **1989**, *6*, 815.

(55) Abagyan, R. A.; Mazur, A. K. *J. Biomol. Struct. Dyn.* **1989**, *6*, 833.

(56) Abagyan, R.; Argos, P. *J. Mol. Biol.* **1992**, *225*, 519.

(57) Schaumann, T.; Braun, W.; Wüthrich, K. *Biopolymers* **1990**, *29*, 679.

(58) Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827.

(59) Amodeo, P.; Barone, V. *J. Am. Chem. Soc.* **1992**, *114*, 9085.

(60) Hagler, A. T.; Ewig, C. S. *Comput. Phys. Commun.* **1994**, in press.

(61) Némethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 1883.

(62) Sippl, M. J.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1984**, *88*, 6231.

(63) Chuman, H.; Momany, F. A.; Schäfer, L. *Int. J. Pept. Protein Res.* **1984**, *24*, 233.

(64) Roterman, I. K.; Lambert, M. H.; Gibson, K. D.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1989**, *7*, 421.

(65) Roterman, I. K.; Gibson, K. D.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1989**, *7*, 391.

(66) Kollman, P. A.; Dill, K. A. *J. Biomol. Struct. Dyn.* **1991**, *8*, 1103.

(67) Gibson, K. D.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1991**, *8*, 1109.

(68) Gunsteren, W. F. v.; Karplus, M. *Macromolecules* **1982**, *15*, 1528.

(69) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(70) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

(71) Kozack, R. E.; Jordan, P. C. *J. Chem. Phys.* **1992**, *96*, 3120.

(72) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395.

(73) Snir, J.; Nemenoff, R. A.; Scheraga, H. A. *J. Phys. Chem.* **1978**, *82*, 2497.

(74) Dunfield, L. G.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1978**, *82*, 2609.

(75) Brooks, C. L., III; Case, D. A. *Chem. Rev.* **1993**, *93*, 2487.

(76) Pincus, M. R.; Scheraga, H. A. *J. Phys. Chem.* **1977**, *81*, 1579.

(77) Brooks, C. L., III. *J. Chem. Phys.* **1987**, *86*, 5156.

(78) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L., III; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989.

(79) Frey, R. F.; Coffin, J.; Newton, S. Q.; Ramek, M.; Cheng, V. K. W.; Momany, F. A.; Schäfer, L. *J. Am. Chem. Soc.* **1992**, *114*, 5369.

(80) Schäfer, L.; Newton, S. Q.; Cao, M.; Peeters, A.; Alsenoy, C. V.; Wolinski, K.; Momany, F. A. *J. Am. Chem. Soc.* **1993**, *115*, 272.

(81) Levitt, M.; Warshel, A. *Nature* **1975**, *253*, 694.

(82) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59.

(83) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1697.

(84) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715.

(85) Gerber, P. R. *Biopolymers* **1992**, *32*, 1003.

(86) Head-Gordon, T.; Brooks, C. L., III. *Biopolymers* **1991**, *31*, 77.

(87) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.

(88) Abola, E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. In *Crystallographic Databases - Information Content, Software Systems, Scientific Applications*; Allen, F. H.; Bergerhoff, G., Sievers, R., Eds.; Data Commission of the International Union of Crystallography: Bonn, 1987; p 107.

(89) Godzik, A.; Kolinski, A.; Skolnick, J. *J. Mol. Biol.* **1992**, *227*, 227.

(90) Miyazawa, S.; Jernigan, R. L. *Macromolecules* **1985**, *18*, 534.

(91) Wilson, C.; Doniach, S. *Proteins: Struct. Funct. Genet.* **1989**, *6*, 193.

(92) Sippl, M. J. *J. Mol. Biol.* **1990**, *213*, 859.

(93) Bryant, S. H.; Lawrence, C. E. *Proteins: Struct. Funct. Genet.* **1993**, *16*, 92.

(94) Pohl, F. M. *Nature (New Biology)* **1971**, *234*, 277.

(95) Rooman, M. J.; Kocher, J.-P. A.; Wodak, S. J. *Biochemistry* **1992**, *31*, 10226.

(96) Sun, S. *Protein Sci.* **1993**, *2*, 762.

(97) Kang, H. S.; Kurochkina, N. A.; Lee, B. *J. Mol. Biol.* **1993**, *229*, 448.

(98) Oobatake, M.; Crippen, G. M. *J. Phys. Chem.* **1981**, *85*, 1187.

(99) Crippen, G. M.; Viswanadhan, V. N. *Int. J. Pept. Protein Res.* **1985**, *25*, 487.

(100) Crippen, G. M.; Ponnuswamy, P. K. *J. Comput. Chem.* **1987**, *8*, 972.

(101) Crippen, G. M.; Snow, M. E. *Biopolymers* **1990**, *29*, 1479.

(102) Crippen, G. M. *Biochemistry* **1991**, *30*, 4232.

(103) Maiorov, V. N.; Crippen, G. M. *J. Mol. Biol.* **1992**, *227*, 876.

(104) Friedrichs, M. S.; Wolynes, P. G. *Science* **1989**, *246*, 371.

(105) Goldstein, R. A.; Luthey-Schulten, Z. A.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4918.

(106) Mezei, M.; Mehrotra, P. K.; Beveridge, D. L. *J. Am. Chem. Soc.* **1985**, *107*, 2239.

(107) Schiffer, C. A.; Caldwell, J. W.; Kollman, P. A.; Stroud, R. M. *Proteins: Struct. Funct. Genet.* **1990**, *8*, 30.

(108) Anderson, A.; Carson, M.; Hermans, J. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 51.

(109) Némethy, G.; Peer, W. J.; Scheraga, H. A. *Annu. Rev. Biophys. Bioeng.* **1981**, *10*, 459.

(110) Gō, N.; Scheraga, H. A. *J. Chem. Phys.* **1969**, *51*, 4751.

(111) Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 420.

(112) Hopfinger, A. J. *Macromolecules* **1971**, *4*, 731.

(113) Hodes, Z. I.; Némethy, G.; Scheraga, H. A. *Biopolymers* **1979**, *18*, 1565.

(114) Kang, Y. K.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4118.

(115) Kang, Y. K.; Gibson, K. D.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1988**, *92*, 4739.

(116) Paterson, Y.; Némethy, G.; Scheraga, H. A. *Ann. N.Y. Acad. Sci.* **1981**, *367*, 132.

(117) Ooi, T.; Oobatake, M.; Némethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086.

(118) Chothia, C. *Nature* **1974**, *248*, 338.

(119) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199.

(120) Vila, J.; Williams, R. L.; Vásquez, M.; Scheraga, H. A. *Proteins: Struct. Funct. Genet.* **1991**, *10*, 199.

(121) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(122) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.

(123) Perrot, G.; Cheng, B.; Gibson, K. D.; Vila, J.; Palmer, K. A.; Nayeem, A.; Maigret, B.; Scheraga, H. A. *J. Comput. Chem.* **1992**, *13*, 1.

(124) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227.

(125) Freyberg, B. v.; Braun, W. *J. Comput. Chem.* **1993**, *14*, 510.

(126) Eisenhaber, F.; Argos, P. *J. Comput. Chem.* **1993**, *14*, 1272.

(127) Richmond, T. J. *J. Mol. Biol.* **1984**, *178*, 63.

(128) Wodak, S. J.; Janin, J. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1736.

(129) Hasel, W.; Hendrickson, T. F.; Still, W. C. *Tetrahedron Comput. Methodol.* **1988**, *1*, 103.

(130) Le Grand, S. M.; Merz, K. M., Jr. *J. Comput. Chem.* **1993**, *14*, 349.

(131) Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 1317.

(132) Colonna-Cesari, F.; Sander, C. *Biophys. J.* **1990**, *57*, 1103.

(133) Holm, L.; Sander, C. *J. Mol. Biol.* **1992**, *225*, 93.

(134) Stouten, P. F. W.; Frömmel, C.; Nakamura, H.; Sander, C. *Mol. Simul.* **1993**, *10*, 97.

(135) Gunsteren, W. F. v.; Berendsen, H. J. C. *Mol. Simul.* **1988**, *1*, 173.

(136) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. *Science* **1991**, *252*, 106.

(137) Ben-Naim, A.; Mazo, R. M. *J. Phys. Chem.* **1993**, *97*, 10829.

(138) Holtzer, A. *Biopolymers* **1994**, *34*, 315.

(139) Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1972**, *57*, 1930.

(140) Pettitt, B. M.; Karplus, M. *J. Phys. Chem.* **1988**, *92*, 3994.

(141) Kitao, A.; Hirata, F.; Gō, N. *J. Phys. Chem.* **1993**, *97*, 10231.

(142) Yu, H.-A.; Pettitt, B. M.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 2425.

(143) Marlow, G. E.; Perkyns, J. S.; Pettitt, B. M. *Chem. Rev.* **1993**, *93*, 2503.

(144) Harvey, S. C. *Proteins: Struct. Funct. Genet.* **1989**, *5*, 78.

(145) Rogers, N. K. In *Prediction of Protein Structure and the Principles of Protein Conformation*; Fasman, G. D., Ed.; Plenum Press: New York, 1989.

(146) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301.
(147) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509.
(148) Warshel, A.; Aqvist, J. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267.
(149) Gilson, M.; Honig, B. H. *Biopolymers* **1986**, *25*, 2097.
(150) Bottcher, C. J. F.; vanBelle, O. C.; Bordewijk, P.; Rip, A. *Theory of Electric Polarization*; Elsevier: Amsterdam, 1973; Vol. 1, p 214.
(151) Gilson, M. K.; Rashin, A.; Fine, R.; Honig, B. *J. Mol.Biol.* **1985**, *184*, 503.
(152) Gilson, M. K.; Honig, B. *Proteins: Struct. Funct. Genet.* **1988**, *4*, 7.
(153) Guida, W. C.; Bohacek, R. S.; Erion, M. D. *J. Comput. Chem.* **1992**, *13*, 214.
(154) Gilson, M. K.; Honig, B. *Proteins: Struct. Funct. Genet.* **1988**, *3*, 32.
(155) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671.
(156) Gilson, M. K.; Sharp, K. A.; Honig, B. *J. Comput. Chem.* **1988**, *9*, 327.
(157) Mehler, E. L.; Solmajer, T. *Protein Eng.* **1991**, *4*, 903.
(158) Vorobjev, V. N.; Grant, J. A.; Scheraga, H. A. *J. Am. Chem. Soc.* **1992**, *114*, 3189.
(159) Vorobjev, Y. N.; Scheraga, H. A. *J. Phys. Chem.* **1993**, *97*, 4855.
(160) Oberoi, H.; Allewell, N. M. *Biophys. J.* **1993**, *65*, 48.
(161) Gerstein, M.; Lynden-Bell, R. M. *J. Mol. Biol.* **1993**, *230*, 641.
(162) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(163) Abagyan, R.; Totrov, M. *J. Mol. Biol.* **1994**, *235*, 983.
(164) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
(165) Miertuˇs, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
(166) Grant, J. A.; Williams, R. L.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 929.
(167) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213.
(168) Negre, M.; Orozco, M.; Luque, F. J. *Chem. Phys. Lett.* **1992**, *196*, 27.
(169) Gō, N.; Gō, M.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 137.
(170) Gō, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535.
(171) Gō, N. *Macromolecules* **1986**, *19*, 2054.
(172) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325.
(173) Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules* **1984**, *17*, 1370.
(174) Di Nola, A.; Berendsen, H. J. C.; Edholm, O. *Macromolecules* **1984**, *17*, 2044.
(175) Rojas, O. L.; Levy, R. M.; Szabo, A. *J. Chem. Phys.* **1986**, *85*, 1037.
(176) Gō, N.; Noguti, T.; Nishikawa, T. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3696.
(177) Tidor, B.; Karplus, M. *Proteins: Struct. Funct. Genet.* **1993**, *15*, 71.
(178) Meirovitch, H. *J. Phys. A* **1982**, *15*, L735.
(179) Meirovitch, H.; Vásquez, M.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1189.
(180) Meirovitch, H.; Vásquez, M.; Scheraga, H. A. *J. Chem. Phys.* **1990**, *92*, 1248.
(181) Meirovitch, H.; Vásquez, M.; Scheraga, H. A. *Biopolymers* **1987**, *26*, 651.
(182) Meirovitch, H.; Kitson, D. H.; Hagler, A. T. *J. Am. Chem. Soc.* **1992**, *114*, 5386.
(183) Miller, W. G.; Brant, D. A.; Flory, P. J. *J. Mol. Biol.* **1967**, *23*, 67.
(184) Flory, P. J. *Macromolecules* **1974**, *7*, 381.
(185) Gō, M.; Scheraga, H. A. *Biopolymers* **1984**, *23*, 1961.
(186) Hermans, J.; Anderson, A. G.; Yun, R. H. *Biochemistry* **1992**, *31*, 5646.
(187) Stoessel, J. P.; Nowak, P. *Macromolecules* **1990**, *23*, 1961.
(188) Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 1309.
(189) Gibson, K. D.; Scheraga, H. A. In *Structure and Expression: From Proteins to Ribosomes*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: Guilderland, 1988; Vol. 1, p 67.
(190) Schlick, T. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1992; Vol. 3, p 1.
(191) Zwanzig, R.; Szabo, A.; Bagchi, B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 20.
(192) Skolnick, J.; Kolinski, A. *Science* **1990**, *250*, 1121.
(193) Jackson, M. B. *J. Chem. Phys.* **1993**, *99*, 7253.
(194) Simon, I.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 797.
(195) Pincus, M. R.; Klausner, R. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5107.
(196) Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1.
(197) Vásquez, M.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 1043.
(198) Vásquez, M.; Scheraga, H. A. *Biopolymers* **1985**, *24*, 1437.
(199) Gibson, K. D.; Scheraga, H. A. *J. Comput. Chem.* **1987**, *8*, 826.
(200) Pincus, M. R. *Biopolymers* **1992**, *32*, 347.
(201) Vajda, S.; Kataoka, R.; DeLisi, C.; Margalit, H.; Berzofsky, J. A. *Annu. Rev. Biophys. Bioeng.* **1990**, *19*, 69.
(202) Vajda, S.; DeLisi, C. *Biopolymers* **1990**, *29*, 1755.
(203) Moon, J. B.; Howe, W. J. *Proteins: Struct. Funct. Genet.* **1991**, *11*, 314.
(204) Dammkoehler, R. A.; Karasek, S. F.; Shands, E. F. B.; Marshall, G. R. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 3.
(205) Beusen, D. D.; Iijima, H.; Marshall, G. R. *Biochem. Pharmacol.* **1990**, *40*, 173.
(206) Braun, W.; Gō, N. *J. Mol. Biol.* **1985**, *186*, 611.
(207) Vásquez, M.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1988**, *5*, 705.
(208) Vásquez, M.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1988**, *5*, 757.
(209) Bruccoleri, R. E.; Haber, E.; Novotný, J. *Nature* **1988**, *335*, 564.
(210) Rosenbluth, M. N.; Rosenbluth, A. W. *J. Chem. Phys.* **1955**, *23*, 356.
(211) O'Toole, E. M.; Panagiotopoulos, A. Z. *J. Chem. Phys.* **1992**, *97*, 8644.
(212) O'Toole, E. M.; Panagiotopoulos, A. Z. *J. Chem. Phys.* **1993**, *98*, 3185.
(213) Premilat, S. *J. Chem. Phys.* **1988**, *88*, 3385.
(214) Garel, T.; Niel, J. C.; Orland, H.; Velikson, B. *J. Chim. Phys.* **1991**, *88*, 2473.
(215) Garel, T.; Niel, J. C.; Orland, H.; Smith, J.; Velikson, B. *J. Chim. Phys.* **1991**, *88*, 2479.
(216) Bascle, J.; Garel, T.; Orland, H.; Velikson, B. *Biopolymers* **1993**, *33*, 1843.
(217) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611.
(218) Li, Z.; Scheraga, H. A. *J. Mol. Struct. (Theochem.)* **1988**, *179*, 333.
(219) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(220) Saunders, M. *J. Comput. Chem.* **1989**, *10*, 203.
(221) Ferguson, D. M.; Raber, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 4371.
(222) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379.
(223) Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419.
(224) Richards, N. G. J.; Williams, P. B. *Chem. Des. Automation News* **1994**, *9*, 1.
(225) Caflisch, A.; Niederer, P.; Anliker, M. *Proteins: Struct. Funct. Genet.* **1992**, *13*, 223.
(226) Caflisch, A.; Niederer, P.; Anliker, M. *Proteins: Struct. Funct. Genet.* **1992**, *14*, 102.
(227) Ripoll, D. R.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1283.
(228) Ripoll, D. R.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 165.
(229) Ripoll, D. R.; Vásquez, M. J.; Scheraga, H. A. *Biopolymers* **1991**, *31*, 319.
(230) Piela, L.; Scheraga, H. A. *Biopolymers* **1987**, *26*, S33.
(231) Robson, B.; Platt, E. *J. Mol. Biol.* **1986**, *188*, 259.
(232) Robson, B.; Platt, E.; Fishleigh, R. V.; Marsden, A.; Millard, P. *J. Mol. Graphics* **1987**, *5*, 8.
(233) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. *Science* **1983**, *220*, 671.
(234) Vanderbilt, D.; Louie, S. G. *J. Comput. Phys.* **1984**, *56*, 259.
(235) Černý, V. *J. Opt. Theory Appl.* **1985**, *45*, 41.
(236) Aarts, E.; Korst, J. *Simulated Annealing and Boltzmann Machines*; John Wiley & Sons: Chichester, 1989.
(237) Basu, A.; Frazer, L. N. *Science* **1990**, *249*, 1409.
(238) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(239) Tobias, D. J.; Martyna, G. J.; Klein, M. L. *J. Phys. Chem.* **1993**, *97*, 12959.
(240) Weis, W. I.; Brünger, A. T.; Skehel, J. J.; Wiley, D. C. *J. Mol. Biol.* **1990**, *212*, 737.
(241) Brünger, A. T. *J. Mol. Biol.* **1988**, *203*, 803.
(242) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27.
(243) Ooi, T.; Nishikawa, K.; Oobatake, M.; Scheraga, H. A. *Biochim. Biophys. Acta* **1978**, *536*, 390.
(244) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* 1988, *239*, 129.
(245) Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. *Tetrahedron Lett.* **1988**, *29*, 4373.
(246) Kawai, H.; Kikuchi, T.; Okamoto, Y. *Protein Eng.* **1989**, *3*, 85.
(247) Wilson, S. R.; Cui, W. *Biopolymers* **1990**, *29*, 225.
(248) Kawai, H.; Okamoto, Y.; Fukugita, M.; Nakazawa, T.; Kikuchi, T. *Chem. Lett.* **1991**, 213.
(249) Nayeem, A.; Vila, J.; Scheraga, H. A. *J. Comput. Chem.* **1991**, *12*, 594.
(250) Okamoto, Y.; Kikuchi, T.; Kawai, H. *Chem. Lett.* **1992**, 1275.
(251) Morales, L. B.; Garduño-Juárez, R.; Romero, D. *J. Biomol. Struct. Dyn.* **1991**, *8*, 721.
(252) Morales, L. B.; Garduño-Juárez, R.; Romero, D. *J. Biomol. Struct. Dyn.* **1992**, *9*, 951.
(253) Berg, B. A.; Neuhaus, T. *Physics Lett. B* **1991**, *267*, 249.
(254) Hansmann, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1993**, *14*, 1333.

(255) Bouzida, D.; Kumar, S.; Swendsen, R. H. *Phys. Rev. A* **1992**, *45*, 8894.

(256) Bouzida, D.; Kumar, S.; Swendsen, R. H. In *26th Hawaii International Conference on System Sciences*; IEEE Computer Society Press: Hawaii, 1993; p 736.

(257) Northrup, S. H.; McCammon, J. A. *Biopolymers* **1980**, *19*, 1001.

(258) (a) Higo, J.; Collura, V.; Garnier, J. *Biopolymers* **1992**, *32*, 33. (b) Gibrat, J.-F.; Collura, V.; Higo, J.; Garnier, J. In *Immunomethods*; Padlan, E. A., Ed.; Academic Press: New York, 1992; p 107.

(259) Collura, V.; Higo, J.; Garnier, J. *Protein Sci.* **1993**, *2*, 1502.

(260) Noguti, T.; Gō, N. *Biopolymers* **1985**, *24*, 527.

(261) Shin, J. K.; Jhon, M. S. *Biopolymers* **1991**, *31*, 177.

(262) Freyberg, B. v.; Braun, W. *J. Comput. Chem.* **1991**, *12*, 1065.

(263) Snow, M. E. *J. Comput. Chem.* **1992**, *13*, 579.

(264) Snow, M. E. *Proteins: Struct. Funct. Genet.* **1993**, *15*, 183.

(265) Lambert, M. H.; Scheraga, H. A. *J. Comput. Chem.* **1989**, *10*, 770.

(266) Lambert, M. H.; Scheraga, H. A. *J. Comput. Chem.* **1989**, *10*, 798.

(267) Lambert, M. H.; Scheraga, H. A. *J. Comput. Chem.* **1989**, *10*, 817.

(268) Cohen, F. E.; Abarbanel, R. M.; Kuntz, I. D.; Fletterick, R. J. *Biochemistry* **1986**, *25*, 266.

(269) Curtis, B. M.; Presnell, S. R.; Srinivasan, S.; Sassenfeld, H.; Klinke, R.; Jeffery, E.; Cosman, D.; March, C. J.; Cohen, F. E. *Proteins: Struct. Funct. Genet.* **1991**, *11*, 111.

(270) (a) Ring, C. S.; Cohen, F. E. *FASEB J.* **1993**, *7*, 783. (b) Rackovsky, S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 644.

(271) Kidera, A.; Konishi, Y.; Oka, M.; Ooi, T.; Scheraga, H. A. *J. Protein Chem.* **1985**, *4*, 23.

(272) Head-Gordon, T.; Stillinger, F. H. *Biopolymers* **1993**, *33*, 293.

(273) Terwilliger, T. C.; Weissman, L.; Eisenberg, D. *Biophys. J.* **1982**, *37*, 353.

(274) Cohen, F. E.; Richmond, T. J.; Richards, F. M. *J. Mol. Biol.* **1979**, *132*, 275.

(275) Cohen, F. E.; Sternberg, M. J. E.; Taylor, W. R. *Nature* **1980**, *285*, 378.

(276) Ptitsyn, O. B.; Rashin, A. A. *Biophys. Chem.* **1975**, *3*, 1.

(277) Cohen, F. E.; Kosen, P. A.; Kuntz, I. D.; Epstein, L. B.; Ciardelli, T. L.; Smith, K. A. *Science* **1986**, *234*, 349.

(278) Cohen, F. E.; Kuntz, I. D. *Proteins: Struct. Funct. Genet.* **1987**, *2*, 162.

(279) Redfield, C.; Smith, L. J.; Boyd, J.; Lawrence, G. M. P.; Edwards, R. G.; Smith, R. A. G.; Dobson, C. M. *Biochemistry* **1991**, *30*, 11029.

(280) Powers, R.; Garrett, D. S.; March, C. J.; Frieden, E. A.; Gronenborn, A. M.; Clore, G. M. *Science* **1992**, *256*, 1673.

(281) Smith, L. J.; Redfield, C.; Boyd, J.; Lawrence, G. M. P.; Edwards, R. G.; Smith, R. A. G.; Dobson, C. M. *J. Mol. Biol.* **1992**, *224*, 899.

(282) Walter, M. R.; Cook, W. J.; Zhao, B. G.; Cameron, R. P., Jr.; Ealick, S. E.; Walter, R. L., Jr.; Reichert, P.; Nagabhushan, T. L.; Trotta, P. P.; Bugg, C. E. *J. Biol. Chem.* **1992**, *267*, 20371.

(283) Wlodaver, A.; Pavlovsky, A.; Gustchina, A. *FEBS Lett.* **1992**, *309*, 59.

(284) Jin, L.; Cohen, F. E.; Wells, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 113.

(285) Bohr, H.; Bohr, J.; Brunak, S.; Cotterill, R. M. J.; Fredholm, H.; Lautrup, B.; Petersen, S. B. *FEBS Lett.* **1990**, *261*, 43.

(286) Crippen, G. M. *J. Comput. Chem.* **1982**, *3*, 471.

(287) Crippen, G. M. *Biopolymers* **1982**, *21*, 1933.

(288) Crippen, G. M. *J. Phys. Chem.* **1987**, *91*, 6341.

(289) Crippen, G. M. *J. Comput. Chem.* **1989**, *10*, 896.

(290) Crippen, G. M.; Havel, T. F. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 222.

(291) Snow, M. E.; Crippen, G. M. *Int. J. Pept. Protein Res.* **1991**, *38*, 161.

(292) Purisima, E. O.; Scheraga, H. A. *J. Mol. Biol.* **1987**, *196*, 697.

(293) Schaik, R. C. v.; Berendsen, H. J. C.; Torda, A. E.; Gunsteren, W. F. v. *J. Mol. Biol.* **1993**, *234*, 751.

(294) Lautz, J.; Kessler, H.; Blaney, J. M.; Scheek, R. M.; Gunsteren, W. F. v. *Int. J. Pept. Protein Res.* **1989**, *33*, 281.

(295) Havel, T. F. *Prog. Biophys. Molec. Biol.* **1991**, *56*, 43.

(296) Piela, L.; Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem.* **1989**, *93*, 3339.

(297) Kostrowicki, J.; Piela, L.; Cherayil, B. J.; Scheraga, H. A. *J. Phys. Chem.* **1991**, *95*, 4113.

(298) Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 7442.

(299) Kostrowicki, J.; Kidera, A.; Scheraga, H. A. Work in progress.

(300) Kraulis, P. J.; Clore, G. M.; Nilges, M.; Jones, T. A.; Pettersson, G.; Knowles, J.; Gronenborn, A. M. *Biochemistry* **1989**, *28*, 7241.

(301) Stillinger, F. H.; Weber, T. A. *J. Stat. Phys.* **1988**, *52*, 1429.

(302) Head-Gordon, T.; Stillinger, F. H.; Arrecis, J. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 11076.

(303) Finkelstein, A. V.; Reva, B. A. *Biofizika (English. Trans.)* **1990**, *35*, 407.

(304) Finkelstein, A. V.; Reva, B. A. *Nature* **1991**, *351*, 497.

(305) Poland, D.; Scheraga, H. A. *Theory of Helix-Coil Transitions in Biopolymers*; Academic Press: New York, 1970.

(306) Rabow, A. A.; Scheraga, H. A. *J. Mol. Biol.* **1993**, *232*, 1157.

(307) Hertz, J.; Krogh, A.; Palmer, R. G. *Introduction to the theory of neural computation*; Addison-Wesley: Redwood City, CA, 1991; p 327.

(308) Somorjai, R. L. *J. Phys. Chem.* **1991**, *95*, 4141.

(309) Sylvain, M.; Somorjai, R. L. *J. Phys. Chem.* **1991**, *95*, 4147.

(310) Olszewski, K. A.; Piela, L.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 4672.

(311) Olszewski, K. A.; Piela, L.; Scheraga, H. A. *J. Phys. Chem.* **1993**, *97*, 260.

(312) Olszewski, K. A.; Piela, L.; Scheraga, H. A. *J. Phys. Chem.* **1993**, *97*, 267.

(313) Amara, P.; Hsu, D.; Straub, J. E. *J. Phys. Chem.* **1993**, *97*, 6715.

(314) Ma, J.; Hsu, D.; Straub, J. E. *J. Chem. Phys.* **1993**, *99*, 4024.

(315) Shalloway, D. *J. Global Opt.* **1992**, *2*, 281.

(316) Shalloway, D. In *Recent Advances in Global Optimization*; Floudas, C. A., Pardalos, P., Eds.; Princeton University Press: Princeton, 1992; p 433.

(317) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277.

(318) Ulitsky, A.; Elber, R. *J. Chem. Phys.* **1993**, *98*, 3380.

(319) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161.

(320) Miranker, A.; Karplus, M. *Proteins: Struct. Funct. Genet.* **1991**, *11*, 29.

(321) Zheng, Q.; Rosenfeld, R.; Kyle, D. J. *J. Chem. Phys.* **1993**, *99*, 8892.

(322) Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley: Reading, MA, 1989.

(323) Dandekar, T.; Argos, P. *Protein Eng.* **1992**, *5*, 637.

(324) Unger, R.; Moult, J. *J. Mol. Biol.* **1993**, *231*, 75.

(325) Unger, R.; Moult, J. In *26th Hawaii International Conference on System Sciences*; IEEE Computer Society Press: Hawaii, 1993; p 715.

(326) Judson, R. S.; Colvin, M. E.; Meza, J. C.; Huffer, A.; Gutierrez, D. *Int. J. Quantum Chem.* **1992**, *44*, 277.

(327) Judson, R. S. *J. Phys.Chem.* **1992**, *96*, 10102.

(328) Sun, S.; Luo, N.; Ornstein, R. L.; Rein, R. *Biophys. J.* **1992**, *62*, 104.

(329) Tufféry, P.; Etchebest, C.; Hazout, S.; Lavery, R. *J. Biomol. Struct. Dyn.* **1991**, *8*, 1267.

(330) McGarrah, D. B.; Judson, R. S. *J. Comput. Chem.* **1993**, *14*, 1385.

(331) Le Grand, S. M.; Merz, K. M., Jr. *J. Global Optimization* **1993**, *3*, 49.

(332) Forrest, S. *Science* **1993**, *261*, 872.

(333) Montelione, G. T.; Scheraga, H. A. *Acc. Chem. Res.* **1989**, *22*, 70.

(334) Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M.; Peterson, M. L. *J. Comput. Chem.* **1993**, *14*, 1407.

(335) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.

(336) Gunsteren, W. F. v.; Weiner, P. K. *Computer Simulation of Biomolecular Systems*; ESCOM Publ.: Leiden, 1989.

(337) Gunsteren, W. F. v.; Weiner, P. K.; Wilkinson, A. J. *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*; ESCOM Publ.: Leiden, 1993; Vol. 2.

(338) Gunsteren, W. F. v.; Mark, A. E. *Eur. J. Biochem.* **1992**, *204*, 947.

(339) Daggett, V.; Levitt, M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5142.

(340) Mark, A. E.; Gunsteren, W. F. v. *Biochemistry* **1992**, *31*, 7745.

(341) Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1993**, *32*, 4175.

(342) Daggett, V.; Levitt, M. *J. Mol. Biol.* **1993**, *232*, 600.

(343) Caflisch, A.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1746.

(344) Hao, M.-H.; Pincus, M. R.; Rackovsky, S.; Scheraga, H. A. *Biochemistry* **1993**, *32*, 9614.

(345) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(346) Gunsteren, W. F. v.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.

(347) Teleman, O.; Jönson, B. *J. Comput. Chem.* **1986**, *7*, 59.

(348) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1991**, *94*, 6811.

(349) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990.

(350) Watanabe, M.; Karplus, M. *J. Chem. Phys.* **1993**, *99*, 8063.

(351) Noguti, T.; Gō, N. *J. Phys. Soc. Jpn.* **1983**, *52*, 3283.

(352) Higo, J.; Seno, Y.; Gō, N. *J. Phys. Soc. Jpn.* **1985**, *54*, 4053.

(353) Gibson, K. D.; Scheraga, H. A. *J. Comput. Chem.* **1990**, *11*, 468.

(354) Dorofeyev, V. E.; Mazur, A. K. *J. Biomol. Struct. Dyn.* **1993**, *11* 143.

(355) Peskin, C. S.; Schlick, T. *Commun. Pure Appl. Math.* **1989**, *42*, 1001.

(356) Schlick, T.; Figueroa, S.; Mezei, M. *J. Chem. Phys.* **1991**, *94*, 2118.

(357) Nyberg, A. M.; Schlick, T. *Chem. Phys. Lett.* **1992**, *198*, 538.

(358) Zhang, G. H.; Schlick, T. *J. Comput. Chem.* **1993**, *14*, 1212.

(359) Jain, A.; Vaidehi, N.; Rodriguez, G. *J. Comput. Phys.* **1993**, *106*, 258.

(360) Grønbech-Jensen, N.; Doniach, S. *J. Comput. Chem.* **1994**, *15*, 997.

(361) Deutsch, J. M.; Madden, T. L. *J. Chem. Phys.* **1989**, *90*, 2476.

(362) Schaik, R. C. v.; Gunsteren, W. F. v.; Berendsen, H. J. C. *J. Comput.-Aided Mol. Design* **1992**, *6*, 97.

(363) Mao, B.; Friedman, A. R. *Biophys. J.* **1990**, *58*, 803.

(364) Li, J.; Platt, E.; Waszkowycz, B.; Cotterill, R.; Robson, B. *Biophys. Chem.* **1992**, *43*, 221.

(365) Paine, G. H.; Scheraga, H. A. *Biopolymers* **1985**, *24*, 1391.

(366) Paine, G. H.; Scheraga, H. A. *Biopolymers* **1986**, *25*, 1547.

(367) Hagler, A. T.; Stern, P. S.; Sharon, R.; Becker, J. M.; Naider, F. *J. Am. Chem. Soc.* **1979**, *101*, 6842.

(368) Rapaport, D. C.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 1238.

(369) Ooi, T.; Scott, R. A.; Vanderkooi, G.; Scheraga, H. A. *J. Chem. Phys.* **1967**, *46*, 4410.

(370) Scheraga, H. A. *Harvey Lect.* **1969**, *63*, 99.

(371) Erenrich, E. H.; Andreatta, R. H.; Scheraga, H. A. *J. Am. Chem. Soc.* **1970**, *92*, 1116.

(372) Paterson, Y.; Rumsey, S. M.; Benedetti, E.; Némethy, G.; Scheraga, H. A. *J. Am. Chem. Soc.* **1981**, *103*, 2947.

(373) Paterson, Y.; Stimson, E. R.; Evans, D. J.; Leach, S. J.; Scheraga, H. A. *Int. J. Pept. Protein Res.* **1982**, *20*, 468.

(374) Pavone, V.; Benedetti, E.; Di Blasio, B.; Pedone, C.; Santini, A.; Bavoso, A.; Toniolo, C.; Crisma, M.; Sartore, L. *J. Biomol. Struct. Dyn.* **1990**, *7*, 1321.

(375) Tirado-Rives, J.; Maxwell, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1993**, *115*, 11590.

(376) Smythe, M. L.; Huston, S. E.; Marshall, G. R. *J. Am. Chem. Soc.* **1993**, *115*, 11594.

(377) Miick, S. M.; Martinez, G. V.; Fiori, W. R.; Todd, A. P.; Millhauser, G. L. *Nature* **1992**, *359*, 653.

(378) Fiori, W. R.; Miick, S. M.; Millhauser, G. L. *Biochemistry* **1993**, *32*, 11957.

(379) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95.

(380) Marshall, G. R.; Hodgkin, E. E.; Langs, D. A.; Smith, G. D.; Zabrocki, J.; Leplawy, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 487.

(381) Némethy, G.; Phillips, D. C.; Leach, S. J.; Scheraga, H. A. *Nature* **1967**, *214*, 363.

(382) Chothia, C. *J. Mol. Biol.* **1973**, *75*, 295.

(383) Weatherford, D. W.; Salemme, F. R. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 19.

(384) Salemme, F. R. *J. Mol. Biol.* **1981**, *146*, 143.

(385) Chou, K.-C.; Pottle, M.; Némethy, G.; Ueda, Y.; Scheraga, H. A. *J. Mol. Biol.* **1982**, *162*, 89.

(386) Chou, K.-C.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 7047.

(387) Chou, K.-C.; Némethy, G.; Scheraga, H. A. *Acc. Chem. Res.* **1990**, *23*, 134.

(388) Chou, K.-C.; Némethy, G.; Scheraga, H. A. *Biochemistry* **1983**, *22*, 6213.

(389) Venkatachalam, C. M. *Biopolymers* **1968**, *6*, 1425.

(390) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Biochim. Biophys. Acta* **1973**, *303*, 211.

(391) Wilmot, C. M.; Thornton, J. M. *Protein Eng.* **1990**, *3*, 479.

(392) Kuntz, I. D. *J. Am. Chem. Soc.* **1972**, *94*, 4009.

(393) Crawford, J. L.; Lipscomb, W. N.; Schellman, C. G. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 538.

(394) Zimmerman, S. S.; Shipman, L. L.; Scheraga, H. A. *J. Phys. Chem.* **1977**, *81*, 614.

(395) Zimmerman, S. S.; Scheraga, H. A. *Biopolymers* **1978**, *17*, 1885.

(396) Tobias, D. J.; Sneddon, S. F.; Brooks, C. L., III. *J. Mol. Biol.* **1990**, *216*, 783.

(397) Tobias, D. J.; Mertz, J. E.; Brooks, C. L., III. *Biochemistry* **1991**, *30*, 6054.

(398) Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1988**, *201*, 161.

(399) Chothia, C. *Annu. Rev. Biochem.* **1984**, *53*, 537.

(400) Chou, K.-C.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 2869.

(401) Chou, K.-C.; Némethy, G.; Scheraga, H. A. *J. Am. Chem. Soc.* **1984**, *106*, 3161.

(402) Presnell, S. R.; Cohen, F. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 6592.

(403) Wada, A. *Adv. Biophys.* **1976**, *9*, 1.

(404) Hol, W. G. J.; Halie, L. M.; Sander, C. *Nature (London)* **1981**, *294*, 532.

(405) Sheridan, R. P.; Levy, R. M.; Salemme, F. R. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4545.

(406) Gilson, M. K.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 1524.

(407) Weber, P. C.; Salemme, F. R. *Nature* **1980**, *287*, 82.

(408) Richardson, J. S. *Adv. Protein Chem.* **1981**, *34*, 167.

(409) Chou, K.-C.; Maggiora, G. M.; Némethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4295.

(410) Carlacci, L.; Chou, K.-C. *Protein Eng.* **1990**, *3*, 509.

(411) Chou, K.-C.; Maggiora, G. M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7315.

(412) Carlacci, L.; Chou, K.-C.; Maggiora, G. M. *Biochemistry* **1991**, *30*, 4389.

(413) Tufféry, P.; Lavery, R. *Proteins: Struct. Funct. Genet.* **1993**, *15*, 413.

(414) Gibson, K. D.; Scheraga, H. A. *J. Comput. Chem.* **1994**, in press.

(415) Gibson, K. D.; Scheraga, H. A. *J. Comput. Chem.* **1994**, in press.

(416) Crick, F. H. C. *Acta Crystallogr.* **1953**, *6*, 689.

(417) Ramachandran, G. N.; Kartha, G. *Nature* **1955**, *176*, 593.

(418) Rich, A.; Crick, F. H. C. *Nature* **1955**, *176*, 915.

(419) Traub, W.; Yonath, A.; Segal, D. M. *Nature* **1969**, *221*, 914.

(420) Parry, D. A. D.; Squire, J. M. *J. Mol. Biol.* **1973**, *75*, 33.

(421) Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. *Science* **1988**, *240*, 1759.

(422) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Science* **1989**, *243*, 538.

(423) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539.

(424) Nishikawa, K.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 395.

(425) Crick, F. H. C. *Acta Crystallogr.* **1953**, *6*, 685.

(426) Cohen, C.; Parry, D. A. D. *Proteins: Struct. Funct. Genet.* **1990**, *7*, 1.

(427) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401.

(428) Lovejoy, B.; Choe, S.; Cascio, D.; McRorie, D. K.; DeGrado, W. F.; Eisenberg, D. *Science* **1993**, *259*, 1288.

(429) Eisenberg, D.; Wesson, M.; Yamashita, M. *Chemica Scr.* **1989**, *29A*, 217.

(430) Krystek, S. R., Jr.; Bruccoleri, R. E.; Novotný, J. *Int. J. Pept. Protein Res.* **1991**, *38*, 229.

(431) Nilges, M.; Brünger, A. T. *Protein Eng.* **1991**, *4*, 649.

(432) Nilges, M.; Brünger, A. T. *Proteins: Struct. Funct. Genet.* **1993**, *15*, 133.

(433) Vieth, M.; Kolinski, A.; Brooks, C. L., III; Skolnick, J. *J. Mol. Biol.* **1994**, *237*, 361.

(434) Kolinski, A.; Godzik, A.; Skolnick, J. *J. Chem. Phys.* **1993**, *98*, 7420.

(435) Miller, M. H.; Scheraga, H. A. *J. Polym. Sci. Symp.* **1976**, *54*, 171.

(436) Miller, M. H.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 470.

(437) Chou, K.-C.; Némethy, G.; Rumsey, S.; Tuttle, R. W.; Scheraga, H. A. *J. Mol. Biol.* **1985**, *186*, 591.

(438) Chou, K.-C.; Némethy, G.; Rumsey, S.; Tuttle, R. W.; Scheraga, H. A. *J. Mol. Biol.* **1986**, *188*, 641.

(439) Chothia, C.; Janin, J. *Biochemistry* **1982**, *21*, 3955.

(440) Chou, K.-C.; Némethy, G.; Pottle, M.; Scheraga, H. A. *J. Mol. Biol.* **1989**, *205*, 241.

(441) Sternberg, M. J. E.; Thornton, J. M. *J. Mol. Biol.* **1976**, *105*, 367.

(442) Rao, S. T.; Rossmann, M. G. *J. Mol. Biol.* **1973**, *76*, 241.

(443) Carlacci, L.; Chou, K.-C. *Biopolymers* **1990**, *30*, 135.

(444) Salemme, F. R. *Prog. Biophys. Mol. Biol.* **1983**, *42*, 95.

(445) Chou, K.-C.; Heckel, A.; Némethy, G.; Rumsey, S.; Carlacci, L.; Scheraga, H. A. *Proteins: Struct. Funct. Genet.* **1990**, *8*, 14.

(446) Tainer, J. A.; Getzoff, E. D.; Beem, K. M.; Richardson, J. S.; Richardson, D. S. *J. Mol. Biol.* **1982**, *160*, 181.

(447) Chou, K.-C.; Carlacci, L.; Maggiora, G. G. *J. Mol. Biol.* **1990**, *213*, 315.

(448) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.

(449) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963.

(450) Lifson, S.; Allegra, G. *Biopolymers* **1964**, *2*, 65.

(451) Poland, D.; Scheraga, H. A. *Biopolymers.* **1969**, *7*, 887.

(452) Dreele, P. H. V.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 396.

(453) Kidera, A.; Mochizuki, M.; Hasegawa, R.; Hayashi, T.; Sato, H.; Nakajima, A.; Fredrickson, R. A.; Powers, S. P.; Lee, S.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 162.

(454) Miki, T.; Kidera, A.; Oka, M.; Hayashi, T.; Nakajima, A.; Meinwald, Y. C.; Thannhauser, T. W.; Scheraga, H. A. *Macromolecules* **1985**, *18*, 1069.

(455) Wojcik, J.; Kidera, A.; Leed, A. R.; Nakajima, A.; Scheraga, H. A. *Macromolecules* **1990**, *23*, 3655.

(456) Gō, N.; Gō, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *59*, 1030.

(457) Gō, M.; Gō, N.; Scheraga, H. A. *J. Chem. Phys.* **1970**, *52*, 2060.

(458) Gō, M.; Gō, N.; Scheraga, H. A. *J. Chem. Phys.* **1971**, *54*, 4489.

(459) Gō, M.; Hesselink, F. T.; Gō, N.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 459.

(460) Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1962**, *66*, 1773.

(461) Hermans, J. *Curr. Opin. Struct. Biol.* **1993**, *3*, 270.

(462) Hesselink, F. T.; Ooi, T.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 541.

(463) McGuire, R. F.; Vanderkooi, G.; Momany, F. A.; Ingwall, R. T.; Crippen, G. M.; Lotan, N.; Tuttle, R. W.; Kashuba, K. L.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 112.

(464) Fu, Y.-C.; McGuire, R. F.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 468.

(465) Takeda, Y.; Iitaka, Y.; Tsuboi, M. *J. Mol. Biol.* **1970**, *51*, 101.

(466) Mandelkern, L. In *Poly-α-amino-Acids*; Fasman, G. D., Ed.; Marcel Dekker: New York, 1967; p 675.

(467) Ganser, V.; Engel, J.; Winklmair, D.; Krause, G. *Biopolymers* **1970**, *9*, 329.

(468) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 516.

(469) Ripoll, D. R.; Scheraga, H. A. *J. Protein Chem.* **1989**, *8*, 263.

(470) Dygert, M.; Gō, N.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 750.

(471) Némethy, G.; Scheraga, H. A. *Biochem. Biophys. Res. Commun.* **1984**, *118*, 643.

(472) Hull, S. E.; Karlsson, R.; Main, P.; Woolfson, M. M.; Dodson, E. J. *Nature* **1978**, *275*, 206.

(473) Mirau, P. A.; Bovey, F. A. In *Book of Abstracts*, 199th National Meeting of the American Chemical Society, Boston, Spring 1990; American Chemical Society: Washington, DC, 1990; POLY 58.

(474) Xu, Y.; Sugar, I. P.; Rama Krishna, N. *J. Biomol. NMR* **1994**, in press.

(475) Gō, N.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 525.

(476) Gō, N.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 552.

(477) Pincus, M. R.; Klausner, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 3413.

(478) Brown, L. R.; Braun, W.; Kumar, A.; Wüthrich, K. *Biophys. J.* **1982**, *37*, 319.

(479) Galaktionov, S. G.; Marshall, G. R. *Biophys. J.* **1993**, *65*, 608.

(480) Némethy, G.; Hodes, Z. I.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5760.

(481) Struthers, R. S.; Hagler, A. T.; Rivier, J. *ACS Symp. Ser.* **1984**, *251*, 239.

(482) Bienstock, R. J.; Rizo, J.; Koerber, S. C.; Rivier, J. E.; Hagler, A. T.; Gierasch, L. M. *J. Med. Chem.* **1993**, *36*, 3265.

(483) Hruby, V. J.; Kao, L.-F.; Pettitt, B. M.; Karplus, M. *J. Am. Chem. Soc.* **1988**, *110*, 3351.

(484) Al-Obeidi, F.; Hadley, M. E.; Pettitt, B. M.; Hruby, V. J. *J. Am. Chem. Soc.* **1989**, *111*, 3413.

(485) O'Connor, S. D.; Smith, P. E.; Al-Obeidi, F.; Pettitt, B. M. *J. Med. Chem.* **1992**, *35*, 2870.

(486) Rivier, J. E.; Rivier, C.; Vale, W.; Koerber, S.; Corrigan, A.; Porter, J.; Gierasch, L. M.; Hagler, A. T. In *Peptides: Chemistry, Structure and Biology*; Rivier, J. E., Marshall, G. R., Eds.; ESCOM: Leiden, 1990; p 33.

(487) Havel, T.; Wüthrich, K. *Bull. Math. Biol.* **1984**, *46*, 673.

(488) Braun, W. *Q. Rev. Biophys.* **1987**, *19*, 115.

(489) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *229*, 317.

(490) Güntert, P.; Braun, W.; K. Wüthrich *J. Mol. Biol.* **1991**, *217*, 517.

(491) Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **1991**, *1*, 447.

(492) Kessler, H.; Griesinger, C.; Lautz, J.; Muller, A.; Gunsteren, W. F. v.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393.

(493) Blackledge, M. J.; Brüschweiler, R.; Griesinger, C.; Schmidt, J. M.; Xu, P.; Ernst, R. R. *Biochemistry* **1993**, *32*, 10960.

(494) Brüschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. *J. Am. Chem. Soc.* **1992**, *114*, 2289.

(495) Schmidt, J. M.; Brüschweiler, R.; Ernst, R. R.; Dunbrack, R. L., Jr.; Joseph, D.; Karplus, M. *J. Am. Chem. Soc.* **1993**, *115*, 8747.

(496) Brunne, R. M.; Gunsteren, W. F. v.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1993**, *115*, 4764.

(497) Glasser, L.; Scheraga, H. A. *J. Mol. Biol.* **1988**, *199*, 513.

(498) Karle, I. L.; Karle, J.; Mastropaolo, D.; Camerman, A.; Camerman, N. *Acta Crystallogr.* **1983**, *B39*, 625.

(499) Camerman, A.; Mastropaolo, D.; Karle, I.; Karle, J.; Camerman, N. *Nature* **1983**, *306*, 447.

(500) Griffin, J. F.; Langs, D. A.; Smith, G. D.; Blundell, T. L.; Tickle, I. J.; Bedarkar, S. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3272.

(501) Mastropaolo, D.; Camerman, A.; Camerman, N. *Biochem. Biophys. Res. Commun.* **1986**, *134*, 698.

(502) Smith, G. D.; Griffin, J. F. *Science* **1978**, *199*, 1214.

(503) Blundell, T. L.; Hearn, L.; Tickle, I. J.; Palmer, R. A.; Morgan, B. A.; Smith, G. D.; Griffin, J. F. *Science* **1979**, *205*, 220.

(504) Kitson, D. H.; Hagler, A. T. *Biochemistry* **1988**, *27*, 5246.

(505) Kopple, K. D.; Schamper, T. J.; Gō, A. *J. Am. Chem. Soc.* **1974**, *96*, 2597.

(506) Hall, D.; Pavitt, N. *J. Comput. Chem.* **1984**, *5*, 441.

(507) Gunsteren, W. F. v.; Karplus, M. *Biochemistry* **1982**, *21*, 2259.

(508) Kitson, D. H.; Avbelj, F.; Moult, J.; Nguyen, D. T.; Mertz, J. E.; Hadzi, D.; Hagler, A. T. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8920.

(509) Heiner, A. P.; Berendsen, H. J. C.; Gunsteren, W. F. v. *Proteins: Struct. Funct. Genet.* **1992**, *14*, 451.

(510) Yonath, A.; Traub, W. *J. Mol. Biol.* **1969**, *43*, 461.

(511) Brown, F. R., III; DiCorato, A.; Lorenzi, G. P.; Blout, E. R. *J. Mol. Biol.* **1972**, *63*, 85.

(512) Némethy, G. In *Collagen: Vol. 1: Biochemistry*; Nimni, M. E., Ed.; CRC Press: Boca Raton, 1988; p 79.

(513) Némethy, G.; Scheraga, H. A. *Bull. Inst. Chem. Res., Kyoto Univ.* **1989**, *66*, 398.

(514) Daragan, V. A.; Ilyina, E.; Mayo, K. H. *Biopolymers* **1993**, *33*, 521.

(515) Li, M.-H.; Fan, P.; Brodsky, B.; Baum, J. *Biochemistry* **1993**, *32*, 7377.

(516) Jones, E. Y.; Miller, A. *J. Mol. Biol.* **1991**, *218*, 209.

(517) Richmond, T. J.; Richards, F. M. *J. Mol. Biol.* **1978**, *119*, 537.

(518) Bhatnagar, R. S.; Rapaka, R. S. In *Biochemistry of Collagen*; Ramachandran, G. N., Reddi, A. H., Eds.; Plenum Press: New York, 1976; p 479.

(519) Engel, J.; Chen, H.-T.; Prockop, D. J.; Klump, H. *Biopolymers* **1977**, *16*, 601.

(520) Okuyama, K.; Tanaka, N.; Ashida, T.; Kakudo, M. *Bull. Chem. Soc. Jpn.* **1976**, *49*, 1805.

(521) Némethy, G.; Scheraga, H. A. *Biopolymers* **1989**, *28*, 1573.

(522) Vitagliano, L.; Némethy, G.; Zagari, A.; Scheraga, H. A. *Biochemistry* **1993**, *32*, 7354.

(523) Katz, E. P.; David, C. W. *Biopolymers* **1990**, *29*, 791.

(524) Miller, M. H.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 910.

(525) Némethy, G.; Miller, M. H.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 914.

(526) Cowan, P. M.; McGavin, S. *Nature* **1955**, *176*, 501.

(527) Crick, F. H. C.; Rich, A. *Nature* **1955**, *176*, 780.

(528) Segal, D. M.; Traub, W. *J. Mol. Biol.* **1969**, *43*, 487.

(529) Doyle, B. B.; Traub, W.; Lorenzi, G. P.; Blout, E. R. *Biochemistry* **1971**, *10*, 3052.

(530) Némethy, G.; Scheraga, H. A. *Biopolymers* **1984**, *23*, 2781.

(531) Némethy, G.; Scheraga, H. A. *Biochemistry* **1986**, *25*, 3184.

(532) Burjanadze, T. V. *Biopolymers* **1982**, *21*, 1489.

(533) Burjanadze, T. V. *Biopolymers* **1992**, *32*, 941.

(534) Piez, K. A.; Trus, B. L. *Biosci. Rep.* **1981**, *1*, 801.

(535) Ballesteros, J.; Némethy, G.; Pottle, M. S.; Scheraga, H. A. To be published.

(536) Chen, J. M.; Kung, C. E.; Feairheller, S. H.; Brown, E. M. *J. Protein Chem.* **1991**, *10*, 535.

(537) Katz, E. P.; David, C. W. *J. Mol .Biol.* **1992**, *228*, 963.

(538) Fraser, R. D. B.; MacRae, T. P. *Conformation in Fibrous Proteins and Related Synthetic Polypeptides*; Academic Press: New York, 1973; Chapter 13.

(539) Marsh, R. E.; Corey, R. B.; Pauling, L. *Biochim. Biophys. Acta* **1955**, *16*, 1.

(540) Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C.; Suzuki, E. *J. Mol. Biol.* **1965**, *11*, 706.

(541) Fossey, S. A.; Némethy, G.; Gibson, K. D.; Scheraga, H. A. *Biopolymers* **1991**, *31*, 1529.

(542) Asakura, T.; Kuzuhara, A.; Tabeta, R.; Saito, H. *Macromolecules* **1985**, *18*, 1841.

(543) Chothia, C.; Lesk, A. M. *EMBO J.* **1986**, *5*, 823.

(544) Hubbard, T. J. P.; Blundell, T. L. *Protein Eng.* **1987**, *1*, 159.

(545) Warme, P. K.; Momany, F. A.; Rumball, S. V.; Tuttle, R. W.; Scheraga, H. A. *Biochemistry* **1974**, *13*, 768.

(546) Acharya, K. R.; Stuart, D. I.; Phillips, D. C.; Scheraga, H. A. *J. Protein Chem.* **1990**, *9*, 549.

(547) Swenson, M. K.; Burgess, A. W.; Scheraga, H. A. In *Frontiers in Physicochemical Biology*; Pullman, B., Ed.; Academic Press: New York, 1978; p 115.

(548) Palmer, K. A.; Scheraga, H. A.; Riordan, J. F.; Vallee, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 1965.

(549) Greer, J. *J. Mol. Biol.* **1981**, *153*, 1027.

(550) Greer, J. *Proteins: Struct. Funct. Genet.* **1990**, *7*, 317.

(551) Hutchins, C.; Greer, J. *Crit. Rev. Biochem. Mol. Biol.* **1991**, *26*, 77.

(552) Greer, J. *Science* **1985**, *228*, 1055.

(553) Sutcliffe, M. J.; Haneef, I.; Carney, D.; Blundell, T. L. *Protein Eng.* **1987**, *1*, 377.

(554) Sutcliffe, M. J.; Hayes, F. R. F.; Blundell, T. L. *Protein Eng.* **1987**, *1*, 385.

(555) Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E.; Thornton, J. M. *Nature* **1987**, *326*, 347.

(556) Blundell, T.; Carney, D.; Gardner, S.; Hayes, F.; Howlin, B.; Hubbard, T.; Overington, J.; Singh, D. A.; Sibanda, B. L.; Sutcliffe, M. *Eur. J. Biochem.* **1988**, *172*, 513.

(557) Ring, C. S.; Sun, E.; McKerrow, J. H.; Lee, G. K.; Rosenthal, P. J.; Kuntz, I. D.; Cohen, F. E. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3583.

(558) Sudarsanam, S.; Virca, G. D.; March, C. J.; Srinivasan, S. *J. Comput. Aided. Mol. Des.* **1992**, *6*, 223.

(559) Wang, J.; Yan, Y.; Garrett, T. P. J.; Liu, J.; Rodgers, D. W.; Garlick, R. L.; Tarr, G. E.; Husain, Y.; Reinherz, E. L.; Harrison, S. C. *Nature* **1990**, *348*, 411.

(560) Ryu, S.-E.; Kwong, P. D.; Truneh, A.; Porter, T. G.; Arthos, J.; Rosenberg, M.; Dai, X.; Xuong, N.; Axel, R.; Sweet, R. W.; Hendrickson, W. A. *Nature* **1990**, *348*, 419.

(561) Williams, A. F.; Barclay, A. N. *Annu. Rev. Immunol.* **1988**, *6*, 381.

(562) Williams, A. F.; Davis, S. J.; He, Q.; Barclay, A. N. *Cold Spring Harbor Symp. Quant. Biol.* **1989**, *54*, 637.

(563) Arthos, J.; Deen, K. C.; Chaikin, M. A.; Fornwald, J. A.; Sathe, G.; Sattentau, Q. J.; Clapham, P. R.; Weiss, R. A.; McDougal, J. S.; Pietropaolo, C.; Axel, R.; Truneh, A.; Maddon, P. J.; Sweet, R. W. *Cell* **1989**, *57*, 469.

(564) Bates, P. A.; McGregor, M. J.; Islam, S. A.; Sattentau, Q. J.; Sternberg, M. J. E. *Protein Eng.* **1989**, *3*, 13.

(565) Ashkenazi, A.; Presta, L. G.; Marsters, S. A.; Camerato, T. R.; Rosenthal, K. A.; Fendly, B. M.; Capon, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7150.

(566) Bowman, M. R.; MacFerrin, K. D.; Schreiber, S. L.; Burakoff, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9052.

(567) Jones, T. A.; Thirup, S. *EMBO J.* **1986**, *5*, 819.

(568) Claessens, M.; Cutsem, E. V.; Lasters, I.; Wodak, S. *Protein Eng.* **1989**, *2*, 335.

(569) Wendoloski, J. J.; Salemme, F. R. *J. Mol. Graph.* **1992**, *10*, 124.
(570) Chothia, C.; Lesk, A. M. *J. Mol. Biol.* **1987**, *196*, 901.
(571) Padlan, E. A.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 819.
(572) Chothia, C.; Lesk, A. M.; Tramontano, A.; Levitt, M.; Smith-Gill, S. J.; Air, G.; Sheriff, S.; Padlan, E. A.; Davies, D.; Tulip, W. R.; Colman, P. M.; Spinelli, S.; Alzari, P. M.; Poljak, R. J. *Nature* **1989**, *342*, 877.
(573) Topham, C. M.; McLeod, A.; Eisenmenger, F.; Overington, J. P.; Johnson, M. S.; Blundell, T. L. *J. Mol. Biol.* **1993**, *229*, 194.
(574) Levy, R.; Assulin, O.; Scherf, T.; Levitt, M.; Anglister, J. *Biochemistry* **1989**, *28*, 7168.
(575) Zilber, B.; Scherf, T.; Levitt, M.; Anglister, J. *Biochemistry* **1990**, *29*, 10032.
(576) Queen, C.; Schneider, W. P.; Selick, H. E.; Payne, P. W.; Landolfi, N. F.; Duncan, J. F.; Avdalovic, N. M.; Levitt, M.; Junghans, R. P.; Waldmann, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 10029.
(577) Glaser, S. M.; Vásquez, M.; Payne, P. W.; Schneider, W. P. *J. Immunol.* **1992**, *149*, 2607.
(578) Co, M. S.; Scheinberg, D. A.; Avdalovic, N. M.; McGraw, K.; Vásquez, M.; Caron, P. C.; Queen, C. *Mol. Immunol.* **1993**, *30*, 1361.
(579) Kettleborough, C. A.; Saldanha, J.; Heath, V. J.; Morrison, C. J.; Bendig, M. M. *Protein Eng.* **1991**, *4*, 773.
(580) Carter, P.; Presta, L.; Gorman, C. M.; Ridgway, J. B. B.; Henner, D.; Wong, W. L. T.; Rowland, A. M.; Kotts, C.; Carver, M. E.; Shepard, H. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4285.
(581) Presta, L. G.; Lahr, S. J.; Shields, R. L.; Porter, J. P.; Gorman, C. M.; Fendly, B. M.; Jardieu, P. M. *J. Immunol.* **1993**, *151*, 2623.
(582) Kolbinger, F.; Saldanha, J.; Hardman, N.; Bendig, M. M. *Protein Eng.* **1993**, *6*, 971.
(583) Winter, G.; Milstein, C. *Nature* **1991**, *349*, 293.
(584) Winter, G.; Harris, W. J. *Immunol. Today* **1993**, *14*, 243.
(585) Padlan, E. A. *Mol. Immunol.* **1994**, *31*, 169.
(586) Gō, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 178.
(587) Gō, N.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 273.
(588) Bruccoleri, R. E.; Karplus, M. *Macromolecules* **1985**, *18*, 2767.
(589) Bruccoleri, R. E.; Karplus, M. *Biopolymers* **1987**, *26*, 137.
(590) Martin, A. C. R.; Cheetham, J. C.; Rees, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9268.
(591) Dudek, M. J.; Scheraga, H. A. *J. Comput. Chem.* **1990**, *11*, 121.
(592) Palmer, K. A.; Scheraga, H. A. *J. Comput. Chem.* **1991**, *12*, 505.
(593) Palmer, K. A.; Scheraga, H. A. *J. Comput. Chem.* **1992**, *13*, 329.
(594) Moult, J.; James, M. N. G. *Proteins: Struct. Funct. Genet.* **1986**, *1*, 146.
(595) Shenkin, P. S.; Yarmush, D. L.; Fine, R. M.; Wang, H.; Levinthal, C. *Biopolymers* **1987**, *26*, 2053.
(596) Fine, R. M.; Wang, H.; Shenkin, P. S.; Yarmush, D. L.; Levinthal, C. *Proteins: Struct. Funct. Genet.* **1986**, *1*, 342.
(597) Smith, K. C.; Honig, B. *Proteins: Struct. Func. Genet.* **1994**, *18*, 119.
(598) Mas, M. T.; Smith, K. C.; Yarmush, D. L.; Aisaka, K.; Fine, R. M. *Proteins: Struct. Funct. Genet.* **1992**, *14*, 483.
(599) Braun, W. *Biopolymers* **1987**, *26*, 1691.
(600) Zheng, Q.; Rosenfeld, R.; Vajda, S.; DeLisi, C. *J. Comput. Chem.* **1993**, *14*, 556.
(601) Zheng, Q.; Rosenfeld, R.; DeLisi, C.; Kyle, D. J. *Protein Sci.* **1994**, *3*, 493.
(602) Summers, N. L.; Karplus, M. *J. Mol. Biol.* **1990**, *216*, 991.
(603) Summers, N. L.; Karplus, M. *Methods Enzymol.* **1991**, *202*, 156.
(604) Kraulis, P. J.; Jones, T. A. *Proteins: Struct. Funct. Genet.* **1987**, *2*, 188.
(605) Borchert, T. V.; Abagyan, R.; Jaenicke, R.; Wierenga, R. K. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1515.
(606) Borchert, T. V.; Abagyan, R.; Radha-Kishan, K. V.; Zeelen, J. P.; Wierenga, R. K. *Structure* **1993**, *1*, 205.
(607) Rosenfeld, R.; Zheng, Q.; Vajda, S.; DeLisi, C. *J. Mol. Biol.* **1993**, *234*, 515.
(608) Diamond, R. *Acta Crystallogr.* **1966**, *21*, 253.
(609) Fitzwater, S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 2133.
(610) Unger, R.; Harel, D.; Wherland, S.; Sussman, J. L. *Proteins: Struct. Funct. Genet.* **1989**, *5*, 355.
(611) Levitt, M. *J. Mol. Biol.* **1992**, *226*, 507.
(612) Jones, T. A.; Zou, J.-Y.; Cowan, S. W. *Acta. Crystallogr.* **1991**, *A47*, 110.
(613) Holm, L.; Sander, C. *J. Mol. Biol.* **1991**, *218*, 183.
(614) Purisima, E. O.; Scheraga, H. A. *Biopolymers* **1984**, *23*, 1207.
(615) Luo, N.; Shibata, M.; Rein, R. *J. Math. Chem.* **1991**, *8*, 217.
(616) Nishikawa, K.; Momany, F. A.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 797.
(617) Luo, Y.; Jiang, X.; Lai, L.; Qu, C.; Xu, X.; Tang, Y. *Protein Eng.* **1992**, *5*, 147.
(618) Deisenhofer, J.; Steigemann, W. *Acta Crystallogr.* **1975**, *B31*, 238.
(619) Payne, P. W. *Protein Sci.* **1993**, *2*, 315.
(620) Rey, A.; Skolnick, J. *J. Comput. Chem.* **1992**, *13*, 443.
(621) Bassolino-Klimas, D.; Bruccoleri, R. E. *Proteins: Struct. Funct. Genet.* **1992**, *14*, 465.
(622) Correa, P. *Proteins: Struct. Funct. Genet.* **1990**, *7*, 366.
(623) Janin, J.; Wodak, S.; Levitt, M.; Maigret, B. *J. Mol. Biol.* **1978**, *125*, 357.
(624) Benedetti, E.; Morelli, G.; Némethy, G.; Scheraga, H. A. *Int. J. Pept. Protein Res.* **1983**, *22*, 1.
(625) Ponder, J. W.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 775.
(626) Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. *Proteins: Struct. Funct. Genet.* **1992**, *12*, 345.
(627) Schrauber, H.; Eisenhaber, F.; Argos, P. *J. Mol. Biol.* **1993**, *230*, 592.
(628) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Israel J. Chem.* **1973**, *11*, 121.
(629) Dunbrack, R. L., Jr.; Karplus, M. *J. Mol. Biol.* **1993**, *230*, 543.
(630) Nayeem, A.; Scheraga, H. A. *J. Protein Chem.* **1994**, in press.
(631) McGregor, M. J.; Islam, S. A.; Sternberg, M. J. E. *J. Mol. Biol.* **1987**, *198*, 295.
(632) Reid, L. S.; Thornton, J. M. *Proteins: Struct. Funct. Genet.* **1989**, *5*, 170.
(633) Summers, N. L.; Carlson, W. D.; Karplus, M. *J. Mol. Biol.* **1987**, *196*, 175.
(634) Summers, N. L.; Karplus, M. *J. Mol. Biol.* **1989**, *210*, 785.
(635) Gelin, B. R.; Karplus, M. *Biochemistry* **1979**, *18*, 1256.
(636) Kabsch, W.; Mannherz, H. G.; Suck, D.; Pai, E. F.; Holmes, K. C. *Nature* **1990**, *347*, 37.
(637) Shih, H. H.-L.; Brady, J.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1697.
(638) Snow, M. E.; Amzel, L. M. *Proteins: Struct. Funct. Genet.* **1986**, *1*, 267.
(639) Novotný, J.; Rashin, A. A.; Bruccoleri, R. E. *Proteins: Struct. Funct. Genet.* **1988**, *4*, 19.
(640) Novotný, J.; Bruccoleri, R.; Karplus, M. *J. Mol. Biol.* **1984**, *177*, 787.
(641) Lee, C.; Subbiah, S. *J. Mol. Biol.* **1991**, *217*, 373.
(642) Tufféry, P.; Etchebest, C.; Hazout, S.; Lavery, R. *J. Comput. Chem.* **1993**, *14*, 790.
(643) Holm, L.; Sander, C. *Proteins: Struct. Funct. Genet.* **1992**, *14*, 213.
(644) Wilson, C.; Gregoret, L. M.; Agard, D. A. *J. Mol. Biol.* **1993**, *229*, 996.
(645) Wilson, C.; Mace, J. E.; Agard, D. A. *J. Mol. Biol.* **1991**, *220*, 495.
(646) Desmet, J.; Maeyer, M. D.; Hazes, B.; Lasters, I. *Nature* **1992**, *356*, 539.
(647) Lasters, I.; Desmet, J. *Protein Eng.* **1993**, *6*, 717.
(648) (a) Lee, C.; Levitt, M. *Nature* **1991**, *352*, 448. (b) Lee, C. *J. Mol. Biol.* **1994**, *236*, 918.
(649) Eisenmenger, F.; Argos, P.; Abagyan, R. *J. Mol. Biol.* **1993**, *231*, 849.
(650) David, C. W. *J. Comput. Chem.* **1993**, *14*, 715.
(651) Lim, W. A.; Hodel, A.; Sauer, R. T.; Richards, F. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 423.
(652) Chiche, L.; Gregoret, L. M.; Cohen, F. E.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3240.
(653) Baumann, G.; Frömmel, C.; Sander, C. *Protein Eng.* **1989**, *2*, 329.
(654) Bowie, J. U.; Lüthy, R.; Eisenberg, D. *Science* **1991**, *253*, 164.
(655) Lüthy, R.; Bowie, J. U.; Eisenberg, D. *Nature* **1992**, *356*, 83.
(656) Sippl, M. J. *Proteins: Struct. Funct. Genet.* **1993**, *17*, 355.
(657) Srinivasan, S.; Shibata, M.; Rein, R. *Int. J. Quantum Chem. QBS* **1986**, *13*, 167.
(658) Havel, T. F.; Snow, M. E. *J. Mol. Biol.* **1991**, *217*, 1.
(659) Srinivasan, S.; March, C. J.; Sudarsanam, S. *Protein Sci.* **1993**, *2*, 277.
(660) Srinivasan, S.; March, C. J.; Sudarsanam, S. *Bio/Technology* **1993**, *11*, 1579.
(661) Šali, A.; Matsumoto, R.; McNeil, H. P.; Karplus, M.; Stevens, R. L. *J. Biol. Chem.* **1993**, *268*, 9023.
(662) Šali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779.
(663) Chothia, C. *Nature* **1992**, *357*, 543.
(664) Dorit, R. L.; Schoenbach, L.; Gilbert, W. *Science* **1990**, *250*, 1377.
(665) Taylor, W. R.; Orengo, C. A. *J. Mol. Biol.* **1989**, *208* 1.
(666) Šali, A.; Blundell, T. L. *J. Mol. Biol.* **1990**, *212*, 403.
(667) Šali, A.; Overington, J. P.; Johnson, M. S.; Blundell, T. L. *TIBS* **1990**, *15*, 235.
(668) Artymiuk, P. J.; Rice, D. W.; Mitchell, E. M.; Willett, P. *J. Inf. Sci.* **1989**, *15*, 287.
(669) Rose, J.; Eisenmenger, F. *J. Mol. Evol.* **1991**, *32*, 340.
(670) Zuker, M.; Somorjai, R. L. *Bull. Math. Biol.* **1989**, *51*, 55.
(671) Rossmann, M. G.; Argos, P. *J. Mol. Biol.* **1976**, *105*, 75.
(672) Remington, S. J.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 2180.
(673) Remington, S. J.; Matthews, B. W. *J. Mol. Biol.* **1980**, *140*, 77.
(674) Rackovsky, S.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 1168.
(675) Rackovsky, S.; Scheraga, H. A. *Acc. Chem. Res.* **1984**, *17*, 209.
(676) Rackovsky, S. *Proteins: Struct. Funct. Genet.* **1990**, *7*, 378.
(677) Taylor, W. R. *Protein Eng.* **1988**, *2*, 77.
(678) Taylor, W. R. *Protein Eng.* **1991**, *4*, 853.
(679) Bazán, J. F. *Biochem. Biophys. Res. Comm.* **1989**, *164*, 788.
(680) Bazán, J. F. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6934.

(681) deVos, A. M.; Ultsch, M.; Kossiakoff, A. A. *Science* **1992**, *255*, 306.

(682) Bazán, J. F. *Science* **1992**, *257*, 410.

(683) McKay, D. B. *Science* **1992**, *257*, 412.

(684) Brandhuber, B. J.; Boone, T.; Kenney, W. C.; McKay, D. B. *Science* **1987**, *238*, 1707.

(685) Benner, S. A.; Cohen, M. A.; Gerloff, D. *Nature* **1992**, *359*, 781.

(686) Benner, S. A.; Gerloff, D. L. *FEBS Lett.* **1993**, *325*, 29.

(687) Pearl, L. H.; Taylor, W. R. *Nature* **1987**, *329*, 351.

(688) Russell, R. B.; Breed, J.; Barton, G. J. *FEBS Lett.* **1992**, *304*, 15.

(689) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C. T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. *Nature* **1989**, *337*, 615.

(690) Lapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S., Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F.; Hawrylik, S. J.; Lee, S. E.; Scheld, K. G.; Hobart, P. M. *Nature* **1989**, *342*, 299.

(691) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. H. *Science* **1989**, *245*, 616.

(692) Waksman, G.; Kominos, D.; Robertson, S. C.; Pant, N.; Baltimore, D.; Birge, R. B.; Cowburn, D.; Hanafusa, H.; Mayer, B. J.; Overduin, M.; Resh, M. D.; Rios, C. B.; Silverman, L.; Kuriyan, J. *Nature* **1992**, *358*, 646.

(693) Overduin, M.; Rios, C. B.; Mayer, B. J.; Baltimore, D.; Cowburn, D. *Cell* **1992**, *70*, 697.

(694) Gribskov, M.; McLachlan, A. D.; Eisenberg, D. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4355.

(695) Bazán, J. F.; Fletterick, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7872.

(696) Benner, S. A.; Badcoe, I.; Cohen, M. A.; Gerloff, D. L. *J. Mol. Biol.* **1994**, *235*, 926.

(697) Bowie, J. U.; Clarke, N. D.; Pabo, C. O.; Sauer, R. T. *Proteins: Struct. Funct. Genet.* **1990**, *7*, 257.

(698) Lüthy, R.; McLachlan, A. D.; Eisenberg, D. *Proteins: Struct. Funct. Genet.* **1991**, *10*, 229.

(699) Taylor, W. R. *J. Mol. Biol.* **1986**, *188*, 233.

(700) Bashford, D.; Chothia, C.; Lesk, A. M. *J. Mol. Biol.* **1987**, *196*, 199.

(701) Overington, J.; Johnson, M. S.; Šali, A.; Blundell, T. L. *Proc. R. Soc. London B* **1990**, *241*, 132.

(702) Overington, J.; Donnelly, D.; Johnson, M. S.; O(S,ˆ)ali, A.; Blundell, T. L. *Protein Sci.* **1992**, *1*, 216.

(703) Jones, D. T.; Taylor, W. R.; Thornton, J. M. *Nature* **1992**, *358*, 86.

(704) Ouzounis, C.; Sander, C.; Scharf, M.; Schneider, R. *J. Mol. Biol.* **1993**, *232*, 805.

(705) Kocher, J.-P. A.; Rooman, M. J.; Wodak, S. J. *J. Mol. Biol.* **1994**, *235*, 1598.

(706) Stultz, C. M.; White, J. V.; Smith, T. F. *Protein Sci.* **1993**, *2*, 305.

(707) Baldi, P.; Chauvin, Y.; Hunkapiller, T.; McClure, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1059.

(708) Krogh, A.; Brown, M.; Mian, I. S.; Sjölander, K.; Haussler, D. *J. Mol. Biol.* **1994**, *235*, 1501.

(709) Goldstein, R. A.; Katzenellenbogen, J. A.; Luthey-Schulten, Z. A.; Seielstad, D. A.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9949.

(710) Friedrichs, M. S.; Goldstein, R. A.; Wolynes, P. G. *J. Mol. Biol.* **1991**, *222*, 1013.

(711) Goldstein, R. A.; Luthey-Schulten, Z. A.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9029.

(712) Troyer, J. M.; Cohen, F. E. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; Vol. 2, p 57.

(713) Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 447.

(714) Shakhnovich, E. I.; Gutin, A. M. *Nature* **1990**, *346*, 773.

(715) Shakhnovich, E.; Farztdinov, G.; Gutin, A. M.; Karplus, M. *Phys. Rev. Lett.* **1991**, *67*, 1665.

(716) Kolinski, A.; Skolnick, J.; Yaris, R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 7267.

(717) Skolnick, J.; Kolinski, A. *Annu. Rev. Phys. Chem.* **1989**, *40*, 207.

(718) Tanaka, S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 3802.

(719) Tanaka, S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 1320.

(720) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 9.

(721) Hagler, A. T.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 554.

(722) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 305.

(723) Gunn, J. R.; Monge, A.; Friesner, R. A.; Marshall, C. H. *J. Phys. Chem.* **1994**, *98*, 702.

(724) Gregoret, L. M.; Cohen, F. E. *J. Mol. Biol.* **1990**, *211*, 959.

(725) Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **1992**, *97*, 9412.

(726) Hendlich, M.; Lackner, P.; Weitckus, S.; Floeckner, H.; Froschauer, R.; Gottsbacher, K.; Casari, G.; Sippl, M. J. *J. Mol. Biol.* **1990**, *216*, 167.

(727) Casari, G.; Sippl, M. J. *J. Mol. Biol.* **1992**, *224*, 725.

(728) Sippl, M. J.; Weitckus, S. *Proteins: Struct. Funct. Genet.* **1992**, *13*, 258.

(729) Kuntz, I. D.; Crippen, G. M.; Kollman, P. A.; Kimelman, D. *J. Mol. Biol.* **1976**, *106*, 983.

(730) Havel, T. F.; Crippen, G. M.; Kuntz, I. D. *Biopolymers*, **1979**, *18*, 73.

(731) Wako, H.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 961.

(732) Wako, H.; Scheraga, H. A. *J. Protein Chem.* **1982**, *1*, 5.

(733) Wako, H.; Scheraga, H. A. *J. Protein Chem.* **1982**, *1*, 85.

(734) Y̌cas, M.; Goel, N. S.; Jacobsen, J. W. *J. Theor. Biol.* **1978**, *72*, 443.

(735) Goel, N. S.; Y̌cas, M. *J. Theor. Biol.* **1979**, *77*, 253.

(736) Gō, N.; Taketomi, H. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 559.

(737) Gō, N.; Taketomi, H. *Int. J. Pept. Protein Res.* **1979**, *13*, 235.

(738) Gō, N.; Taketomi, H. *Int. J. Pept. Protein Res.* **1979**, *13*, 447.

(739) Ueda, Y.; Taketomi, H.; Gō, N. *Biopolymers* **1978**, *17*, 1531.

(740) Lau, K. F.; Dill, K. A. *Macromolecules* **1989**, *22*, 3986.

(741) Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1989**, *90*, 492.

(742) Dill, K. A.; Shortle, D. *Annu. Rev. Biochem.* **1991**, *60*, 795.

(743) Chan, H. S.; Dill, K. A. *Macromolecules* **1989**, *22*, 4559.

(744) Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1991**, *95*, 3775.

(745) Hao, M.-H.; Rackovsky, S.; Liwo, A.; Pincus, M. R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6614.

(746) Gregoret, L. M.; Cohen, F. E. *J. Mol. Biol.* **1991**, *219*, 109.

(747) Lau, K. F.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 638.

(748) Shortle, D.; Chan, H. S.; Dill, K. A. *Protein Sci.* **1992**, *1*, 201.

(749) Dill, K. A.; Fiebig, K. M.; Chan, H. S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1942.

(750) Fiebig, K. M.; Dill, K. A. *J. Chem. Phys.* **1993**, *98*, 3475.

(751) Matheson, R. R.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 819.

(752) (a) Šali, A.; Shakhnovich, E.; Karplus, M. *J. Mol. Biol.* **1994**, *235*, 1614. (b) Šali, A.; Shakhnovich, E.; Karplus, M. *Nature* **1994**, *269*, 248.

(753) Shakhnovich, E. I.; Gutin, A. M. *Biophys.Chem* **1989**, *34*, 187.

(754) Shakhnovich, E. I.; Gutin, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7195.

(755) (a) Shakhnovich, E. I.; Gutin, A. M. *Protein Eng.* **1993**, *6*, 793. (b) Shakhnovich, E. I. *Phys. Rev. Lett.* **1994**, *72*, 3907.

(756) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524.

(757) Bryngelson, J. D.; Wolynes, P. G. *Biopolymers* **1990**, *30*, 177.

(758) Skolnick, J.; Kolinski, A. *J. Mol. Biol.* **1991**, *221*, 499.

(759) Godzik, A.; Kolinski, A.; Skolnick, J. *J. Comput. Chem.* **1993**, *14*, 1194.

(760) Hao, M.-H.; Scheraga, H. A. *J. Phys. Chem.* **1994**, *98*, 4940.

(761) Lee, J. *Phys. Rev. Lett.* **1993**, *71*, 211.

(762) Kolinski, A.; Skolnick, J.; Yaris, R. *J. Chem. Phys.* **1986**, *85*, 3585.

(763) Hao, M.-H.; Scheraga, H. A. *J. Phys. Chem.* **1994**, in press.

(764) Skolnick, J.; Kolinski, A. *Adv. Chem. Phys.* **1990**, *78*, 223.

(765) Rey, A.; Skolnick, J. *Chem. Phys.* **1991**, *158*, 199.

(766) Knapp, E. W. *J. Comput. Chem.* **1992**, *13*, 793.

(767) Handel, T. M.; Williams, S. A.; DeGrado, W. F. *Science* **1993**, *261*, 879.

(768) Covell, D. G.; Jernigan, R. L. *Biochemistry* **1990**, *29*, 3287.

(769) Subbiah, S. *Science* **1991**, *252*, 128.

(770) Covell, D. G. *Proteins: Struct. Funct. Genet.* **1992**, *14*, 409.

(771) Hinds, D. A.; Levitt, M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2536.

(772) Stillinger, F. H.; Head-Gordon, T.; Hirshfeld, C. L. *Phys. Rev. E* **1993**, *48*, 1469.

(773) Head-Gordon, T.; Stillinger, F. H. *Phys. Rev. E* **1993**, *48*, 1502.

(774) Honeycutt, J. D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3526.

(775) Honeycutt, J. D.; Thirumalai, D. *Biopolymers* **1992**, *32*, 695.

(776) Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6369.

(777) Fukugita, M.; Lancaster, D.; Mitchard, M. G. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6365.

(778) Simmerling, C.; Elber, R. *J. Am. Chem. Soc.* **1994**, *116*, 2534.

(779) Tsujishita, H.; Moriguchi, I.; Hirono, S. *Biophys. J.* **1994**, *66*, 1815.

(780) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *101*, 5032.