Evolution of Physics-Based Methodology for Exploring the Conformational Energy Landscape of Proteins

HAROLD A. SCHERAGA,¹ JAROSLAW PILLARDY,¹ ADAM LIWO,^{1,2} JOOYOUNG LEE,^{1,*} CEZARY CZAPLEWSKI,^{1,2} DANIEL R. RIPOLL,³ WILLIAM J. WEDEMEYER,^{1,4} YELENA A. ARNAUTOVA¹

¹Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301
 ²Department of Chemistry, University of Gdańsk, Gdańsk, Poland
 ³Cornell Theory Center, Cornell University, Ithaca, NY 14853-3801
 ⁴National Institutes of Health Postdoctoral Fellow

Received 2 February 2001; Accepted 18 April 2001

Abstract: The evolution of our physics-based computational methods for determining protein conformation without the introduction of secondary-structure predictions, homology modeling, threading, or fragment coupling is described. Initial use of a hard-sphere potential captured much of the structural properties of polypeptide chains, and subsequent more refined force fields, together with efficient methods of global optimization provide indications that progress is being made toward an understanding of the interresidue interactions that underlie protein folding.

© 2002 John Wiley & Sons, Inc. J Comput Chem 23: 28–34, 2002

Key words: physics-based methodology; protein conformation; protein folding

Introduction

Our motivation for developing methods for conformational energy calculations on proteins followed from our experimental work¹ to determine the conformation of a protein, bovine pancreatic ribonuclease A (RNase A), in aqueous solution, before X-ray and NMR methods had led to protein structures. Physical chemical methods were used to identify noncovalent interactions, i.e., distance constraints,¹ that would place restrictions as to how the polypeptide chain could fold, thereby enabling the three-dimensional structure to be determined. Such an approach led to the identification of three specific tyrosyl-aspartate interactions,¹ whose existence was confirmed by the subsequent determination of the X-ray structure of RNase A.^{2, 3}

Recognizing that more than three such distance constraints would be required to fold a protein within any predetermined degree of precision,⁴ we began to develop computational methodology⁵ that could ultimately make use of such distance constraints for an efficient search of conformational space. Initially, the computations were based only on a hard-sphere potential,⁵ similar to the approach of Ramachandran and coworkers,⁶ who treated a terminally blocked amino acid residue. Several interesting conclusions were derived about the role of such steric effects in influencing the conformations of polypeptide chains.^{7–9} It is remarkable that much of the structural character of proteins (e.g., the distribution of the dihedral angles ϕ , ψ , and χ^1 for various residues) results from simple steric repulsion.

Clearly, a hard-sphere potential is inadequate to determine stable conformations of a macromolecule,¹⁰ and, in fact, Liquori

and coworkers¹¹ had already introduced a more detailed potential function to treat synthetic polymers. A series of attempts by Levitt and Lifson,¹² Hagler and coworkers,¹³ and Brant and Flory¹⁴ followed to derive improved potential functions. Our efforts in this regard led to our empirical conformational energy program for peptides, ECEPP,¹⁵ which was subsequently upgraded twice as ECEPP/2^{16, 17} and ECEPP/3.¹⁸ Several other force fields, for example, AMBER¹⁹ and CHARMM²⁰ have since been introduced. All of these force fields are augmented by either explicit or continuum treatments of hydration. Efforts continue in many laboratories to improve the current force fields. One of these efforts involves the coupling between conformational changes and ioniza-

Contract/grant sponsor: National Foundation for Cancer Research

Correspondence to: H. A. Scheraga; e-mail: has5@cornell.edu

^{*}Present address: Program of Computational Sciences, Korea Institute for Advanced Study, Seoul, Korea

Contract/grant sponsor: National Institutes of Health; contract/grant number: GM-14312

Contract/grant sponsor: National Science Foundation; contract/grant number: MCB95-13167

Contract/grant sponsor: Fogarty Foundation; contract/grant number: TW1064

Contract/grant sponsor: NIH National Center for Research Resources; contract/grant number: P41RR-04293

Contract/grant sponsor: Polish State Committee for Scientific Research, KBN; contract/grant numbers: 3 T09A 111 17 and 127/E-335/S/2000

tion equilibria.²¹ Simultaneously, the fundamental physical aspects underlying such computations were elucidated.²²⁻²⁴

The other essential ingredient of conformational energy calculations, besides a good force field, is an efficient procedure to search the conformational space for the global minimum of the conformational energy, according to the hypothesis that followed from Anfinsen's²⁵ classical experiment. Such search methods involve energy minimization, and Monte Carlo (MC) and molecular dynamics (MD) procedures. Aside from some preliminary exercises,^{26–29} we have not used MD procedures because the required femtosecond time step does not enable computations to be carried out on the experimental time scale of folding, which is typically milliseconds or seconds. Our efforts at global optimization involved an evolution of minimization and MC procedures,^{30–36} ultimately leading to our present hierarchical one described later in this article.

This article briefly recounts the evolution of energy functions and the conformational search procedures that have been developed over the past 40 years in our laboratory. It is not designed as a comprehensive review of the literature in the field.

Initial Applications

We initially applied the ECEPP force field to gain an understanding of the interatomic interactions that lead to the basic structures from which proteins are built. First, the interactions leading to the preferences for the right-handed twist of α -helices, ^{37–39} β -sheets,^{40–43} and the β - α - β crossover,⁴⁴ and the packing of these structures, viz., α - α ,^{45–51} α - β ,⁵² β - β ,⁵³ and the β -barrel⁵⁴ were identified. Conformational fluctuations were treated in the context of an exact loop-closure algorithm,^{55–58} with applications to the cyclic decapeptide gramicidin S⁵⁹ and cyclo-hexaglycyl.^{60, 61}

Some initial applications were made to globular and fibrous proteins. For globular proteins, use was made of homology modeling.⁶²⁻⁶⁶ Collagen-like polytripeptides were examined as models of fibrous proteins.^{67–73} The energetics of conformational transitions, specifically the helix-coil transition, were also examined.^{74–78} At the same time, early ideas about the mechanism of protein folding^{79–83} and enzyme–substrate interactions^{84–91} were developed.

Statistical Mechanics of Folding Transitions

Following up on earlier considerations,^{22–24} and influenced by work of Gõ,⁹² Shakhnovich,⁹³ Thirumalai,⁹⁴ and Wolynes,⁹⁵ recent efforts in our laboratory were devoted to identify the characteristic features of short- and long-range interactions, and amino acid sequences, that determine the first-order character of the folding transition and its cooperativity in globular proteins.^{96–102} This effort was facilitated¹⁰³ by use of Lee's entropy sampling Monte Carlo (ESMC) method,¹⁰⁴ a general Monte Carlo technique characterized by sampling all energy states with equal probability. ESMC defines the probability density ρ_m in the Metropolis algorithm as $\rho_m = \exp[-S(E_m)/k]$, where $S(E_m)$ is evaluated by an iterative procedure and converges to the microscopic entropy of conformations at energy level E_m . Such an algorithm samples the low-entropy conformations in the same manner as the conventional Monte Carlo samples the energetically important conformations. The ESMC method avoids a local energy-minimum problem, and is quite efficient and accurate in computing the microscopic entropy function. An exact knowledge of the microcanonical entropy of a protein model in both the native and nonnative states provides a precise characterization of the folding process of the model.

Global Optimization

Our most recent work has been devoted to surmounting the multiple-minima problem, i.e., to trying to identify the global minimum in the multidimensional conformational energy space. For this purpose, a menu of such procedures, previously summarized, ¹⁰⁵ was developed. Some of the more recent ones, a few of which played a role in the hierarchy discussed later, are discussed briefly here.

Our earlier build-up procedure, based on combining low-energy fragments, was augmented with a build-up based on probabilities, the pattern-recognition importance-sampling minimization, PRISM,^{106–108} procedure, which makes use of the properties of the individual amino acid residues.^{109, 110} PRISM uses statistical information collected from the PDB to focus the search on the most probable regions of the conformational space. It was applied successfully to predict the overall fold of the 36-residue avian peptide.¹⁰⁸

Subsequent efforts involved various Monte Carlo procedures combined with energy minimization. The simplest of these methods is Monte Carlo with minimization, MCM,111-113 which is a Metropolis Monte Carlo algorithm in which every trial state is first energy-minimized before the Boltzmann acceptance criterion is applied. This modification of the accept/reject criterion does not satisfy the condition of detailed balance and, hence, the MCM algorithm does not produce a thermal Boltzmann distribution. However, our experience indicates that this method is effective in finding low-energy protein conformations,^{111, 112} and almost all of our methods now employ an energy-minimization step. The effectiveness of energy minimization may derive in part from its ability to overcome the entropic barrier to finding low-energy conformations. Such barriers arise naturally in the conformational space of proteins, because the high dimensionality makes it far more likely that a random move near an energy minimum will increase the energy rather than decrease it. The MCM procedure has been applied successfully to study the conformational preferences of the pentapeptide Met-enkephalin,^{111, 112} but it has been implemented with other methods in optimizing protein structures, as described below.

One such method is Electrostatically Driven Monte Carlo (EDMC),^{114–116} which at present is our preferred method for refining all-atom structures of proteins. The EDMC method employs a move set in which individual peptide groups are selected at random and rotated "in place" (i.e., the conformational change is localized to the peptide group as much as possible) so as to optimize the alignment of its dipole moment with the local electric field. [Such a move set was first used in the Self-Consistent Electrostatic Field method.¹¹⁷] The resulting conformation is then energy minimized before applying the Metropolis criterion, as in the MCM method. The basic EDMC method has recently been augmented with new search techniques to improve its efficiency.¹¹⁶ The EDMC method has proven effective in finding the global energy minimum of *all-atom* models of polypeptides consisting of up to 20 amino acid residues,^{114–120} most notably in studying the pH dependence of the conformational properties of polypeptides.^{21, 121, 122}

Our procedure to work in a space of high dimensionality and then relax back to three dimensions¹²³ evolved into a methodology involving deformation of the potential energy surface to eliminate unwanted minima. This deformation was based on a solution of the diffusion equation in cartesian space with a diffusion equation method, DEM.¹²⁴⁻¹²⁷ For application of the DEM in dihedral angle space,¹²⁷ a method was introduced to calculate exact end-to-end distance distributions for *finite* freely rotating chains.¹²⁸ The basic idea of the DEM is to deform the multivariable function that represents the potential energy in such a manner as to make the shallow wells disappear gradually, while other potential wells grow at their expense. Under the assumption that the shallower wells will disappear more easily than the deeper wells, it is possible to envision an iterative procedure which, applied to the potential function, will change its shape, making most of the minima become shallower until they disappear, while leaving a single absorbing minimum related to the lowest minimum of the original function. At this point of the deformation process, a simple local minimization algorithm should be able to retrieve the position of the unique minimum from any starting point.

Simultaneously, an alternative procedure, which scales a variable instead of scaling a function, the distance scaling method, DSM,¹²⁹ and different reversing procedures, were introduced leading to the self-consistent basin-to-deformed-basin mapping (SCB-DBM) as a more sophisticated example of a deformation-based method.¹³⁰⁻¹³² The positions of minima of the deformed function are, in general, different from those of the original function, and a proper reversing procedure is as important as the deformation itself. The simplest approach, consisting of a series of local minimizations carried out on gradually less deformed surfaces, is successful only for simple systems. A multiple-trajectory perturbation approach, which tracks more than one minimum back during deformation and tries to detect branching of a trajectory by using a local search, was applied successfully for more demanding systems. The underlying principle of SCBDBM is the location of large regions of conformational space containing low-energy minima by coupling them to some of the greatly reduced number of minima on the highly deformed surface. This is achieved by iterating cycles, each consisting of reversing the deformation and deforming the newly found low-energy structures. SCBDBM has been applied successfully to predict lowest-energy structures of polyalanine chains of length up to 100 amino acid residues, to locate global minima of Lennard-Jones atomic clusters containing up to 100 atoms in a cluster, and in the theoretical prediction of crystal structures (discussed in the next section).

At present, our most effective procedures for the global optimization of protein structures appear to be conformational space annealing (CSA)¹³³⁻¹³⁵ and conformational-family Monte Carlo (CFMC).¹³⁶ The CSA method¹³³⁻¹³⁵ combines essential aspects of the build-up procedure and a genetic algorithm. The CSA method enforces a broad conformational search in its early stages and gradually allows the search to become focused into smaller regions with low energy. Specifically, the CSA method maintains a minimum distance between conformations (usually defined by their deviation in dihedral angles), which is gradually reduced ("annealed"). The CSA method resembles a genetic algorithm in that it starts with a bank of randomly generated and subsequently energy-minimized conformations separated by the minimum distance. This bank is meant to represent a sparse sampling of the conformational space that captures much of the low-energy shortrange structure of the protein. The CSA method then generates new trial conformations by recombining conformations of the present bank with segments of various sizes drawn from the current and the original bank; these trial conformations are then energy minimized. The trial conformation may then replace a conformation in the present bank, depending on its energy and the present minimum distance cutoff. This method has been successful in obtaining the global minimum of peptides containing up to 20 amino acid residues¹³³⁻¹³⁵ using all-atom models of polypeptides and the ECEPP/3 force field. The CSA method is also our principal method for optimizing the united-residue (UNRES) energy in our hierarchical procedure for protein structure prediction discussed later.

Another efficient global optimization method, Conformation-Family Monte Carlo (CFMC),¹³⁶ was also introduced recently in our laboratory. The CFMC method can be considered as an extension of the original MCM method, because at each iteration of the method a conformation is perturbed, locally minimized, and then subjected to an accept/reject criterion. However, the CFMC method maintains a database of low-energy conformations that are clustered into families, simulating an ensemble of states, rather than the single state characteristic of Metropolis Monte Carlo simulations. Clustering conformations into families helps to "coarse grain" the conformational space, and allows the CFMC method to exploit information about the local structure of the energy landscape to guide the global exploration. In this regard, the CFMC method resembles other ensemble-oriented simulation methods such as CSA and the SCBDBM methods. They all maintain a database of conformations, which is initialized to a set of randomly generated conformations and gradually "pruned" into shape by successive random moves, local minimization, and accept/reject criteria. However, the families of the CFMC method constitute an additional level of organization in the database of conformations; in effect, CFMC moves are made not between conformations, but between families of conformations. The CFMC method relies heavily on perturbations of one or a few adjacent dihedral angles, whereas the CSA method employs a recombination scheme in which pieces of candidate conformations are combined and minimized.

The efficiency of Monte Carlo-type methods (such as simulated annealing, MCM, and CFMC) is enhanced by a good set of moves that produce a relatively high acceptance ratio while favoring a broad search of the conformational space. One method for producing such good moves was introduced by Noguti and Gõ, who proposed taking steps in the space of the normal modes of the protein.¹³⁷ Thus, in the Noguti–Gõ method, the Hessian of the energy function is evaluated at the present conformation and analyzed into its normal modes; Monte Carlo steps are then scaled by the relative dimensions along the eigenvectors. A similar method was proposed by Vanderbilt and Louie, who proposed diagonalizing the covariance matrix of already accepted states.¹³⁸ Two novel methods that have the advantages of the Noguti–Gõ and Vanderbilt–Louie methods but avoid the need for eigenanaly-

sis (which can be computationally expensive) have been developed recently.¹³⁹ These move sets may be combined with novel acceptance criteria (such as that of the ESMC method¹⁰⁴) and with energy minimization (as in MCM).

The interactions within proteins are typically short ranged; hence, *localized* conformational changes are likely to result in another conformation of low energy. However, local rearrangements of the polypeptide backbone are difficult in rigid-geometry models, because the change in a single dihedral angle can move a large number of atoms, thus altering a large number of interaction energies. Therefore, we have developed several methods for sampling all possible rigid-geometry loop conformations consistent with a given set of constraints.^{55–58} The most recent of these solves the loop closure problem by finding the real roots of a simple polynomial,⁵⁸ and may be applied to sampling the conformations of even long loops (greater than seven residues).

The determination of side-chain dihedral angles has been facilitated in recent years by the development of rotamer libraries, ^{140, 141} which couple backbone and side-chain dihedral angles. Several efficient methods have been developed in recent years for finding the rotamers of lowest energy such as the dead-end elimination algorithm.¹⁴² However, our experience suggests that the optimal dihedral angles of *buried* side chains can usually be found by simple methods, for example, several sweeps of energy minimization of the dihedral angles of each side chain in turn.

Most of these procedures benefited from parallelization,¹⁴³ and some of them foreshadowed the development of procedures to compute crystal structures and the three-dimensional structures of proteins, discussed in the next two sections.

Prediction of Crystal Structures

The prediction of crystal structures from the structure of its constituent molecules and their interaction energies (without using information about the space group) is a problem in global optimization.144,145 Several of the optimization methods described above (such as the DEM, DSM, and SCBDBM methods) have been used for crystal structure prediction of small rigid molecules. Specifically, the DEM and DSM predicted the crystal structures of hexasulfur and benzene successfully,146,147 while the SCBDBM method has been used to predict the crystal structures of several heteroatomic organic molecules.132 The more efficient CFMC method is now being used to predict crystal structures of both rigid and flexible molecules in a blind benchmarking test similar to the CASP (Critical Assessment of Techniques for Protein Structure Prediction) experiment.¹⁴⁸ As noted by several groups, such crystal-structure prediction methods offer an important tool for refining energy functions.¹³² However, this is not as simple as it appears, because the experimentally observed structure may be determined by its growth conditions as well as by thermodynamic considerations.149-151

Hierarchical Approach to the Prediction of Protein Structure

As with crystal packing studies, our efforts to predict the threedimensional structures of proteins are driven by the desire to gain an understanding as to how such structures arise solely by global optimization of a potential energy function, without the use of ancillary aides such as secondary structure prediction, homology modeling, threading, fragment coupling, etc. Recognizing the impossibility of searching conformational space with an all-atom potential function, a hierarchical procedure was developed whose two main features are the initial use of a united-residue, UNRES, potential function^{100, 152–155} and an efficient procedure, conformational space annealing, CSA,¹³³ to explore the UNRES space. The protein is first optimized with the low-resolution UNRES model and the CSA method.

In UNRES, the backbone is represented as a virtual-bond chain of C^{α} atoms, and the side chains are depicted as ellipsoids that interact through a Gay-Berne potential.¹⁵⁶ The interaction sites are the united-atom side chains, and the centers of the peptide groups between C^{α} atoms, which interact through empirical terms augmented by correlation terms (multiple-body interactions). The multiple-body interactions among peptide groups in the UNRES potential are represented in terms of a cumulant expansion of the free energy,¹⁵⁷ following Kubo,¹⁵⁸ and includes multibody cooperative terms whose relative weights are determined by a Z-score-and-gap optimization.^{159, 160} The purpose of the last procedure is to maximize both the gap (between the average energy of the native-structure family and that of the nonnative structures) and the ratio of this gap to the standard deviation of the energy distribution of nonnative structures for the chosen test proteins, to obtain a folding potential. The CSA method, resembling elements of build-up and genetic algorithms, searches the UNRES space to narrow the region of the possible location of the global minimum.

The lowest-energy UNRES conformation, as well as a set of distinct low-energy conformations, are then converted to all-atom models. The united-residue chains are first converted to all-atom poly(L)-alanine backbones using the dipole-path algorithm,¹⁶¹ i.e., the specific side chains are neglected. The side-chain conformations are then determined by several sweeps of a simple grid search of successive side-chain dihedral angles. Finally, the all-atom models are refined with the EDMC method, with inclusion of the SRFOPT^{162, 163} continuum hydration model.

Using only low-order terms of the cumulant expansion, the hierarchical approach performed well in predicting α -helical protein structures in the blind CASP3 exercise. ¹⁶⁴⁻¹⁶⁸ A particularly good result was obtained for target T0061, an 89-residue protein (PDB entry 1bg8); the structure of the core, representing 80% of the experimentally observed structure, was predicted with an rmsd of 4.2 Å for the C^{α} atoms. However, the CASP3 force field was not successful in predicting protein structures with β -type secondary structures. This was remedied in the CASP4 force field, which included higher order terms of the cumulant expansion, 155, 157, 169 and which successfully predicted large fragments of α , β , and α/β proteins, as demonstrated in the CASP4 exercise.¹⁷⁰ Our best α -helical prediction corresponds to target T0102, a 70-residue cyclic polypeptide from *Enterococcus faecalis* (PDB code: 1e68). The whole structure was predicted within an rmsd of 4.3 Å for the C^{α} atoms. In addition, for the 163-residue target T0126, an α/β protein, fragments involving 61 residues of model 3 match the experimental structure within 6.0 Å for the C^{α} atoms and correctly predicted the contact between noncontiguous β -strands.¹⁷⁰

Concluding Remarks

With efficient methods to search conformational space, and continually improving potential functions, progress is being made in the prediction of protein structure, based only on the physics of interresidue interactions. The current methodology is the present phase of a series of evolving procedures that began with a simplified hardsphere potential, and led to more complete potential functions and efficient procedures for global optimization.

Acknowledgments

The computations in this work were carried out at the Cornell Theory Center, which receives funding from Cornell University, New York State; the National Center for Research Resources at the National Institutes of Health (NIH), and members of the Theory Center's Corporate Partnership Program; with the computing resources provided by the National Partnership for Advanced Computational Infrastructure at the San Diego Supercomputer Center supported in part by the NSF cooperative agreement ACI-9619020; with the resources of the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk; and with our own array of 55 dual-processor PC computers.

References

- 1. Scheraga, H. A. Fed Proc 1967, 26, 1380.
- Wlodawer, A.; Svensson, L. A.; Sjölin, L.; Gilliland, G. L. Biochemistry 1988, 27, 2705.
- 3. Scheraga, H. A. Carlsberg Res Commun 1984, 49, 1.
- 4. Wako, H.; Scheraga, H. A. Macromolecules 1981, 14, 961.
- 5. Némethy, G.; Scheraga, H. A. Biopolymers 1965, 3, 155.
- Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. J Mol Biol 1963, 7, 95.
- Scheraga, H. A.; Leach, S. J.; Scott, R. A.; Némethy, G. Disc Faraday Soc 1965, 40, 268.
- Némethy, G.; Leach, S. J.; Scheraga, H. A. J Phys Chem 1966, 70, 998.
- Leach, S. J.; Némethy, G.; Scheraga, H. A. Biopolymers 1966, 4, 369, 887.
- 10. Scheraga, H. A. Adv Phys Org Chem 1968, 6, 103.
- DeSantis, P.; Giglio, E.; Liquori, A. M.; Ripamonti, A. J Polym Sci A 1963, 1, 1383.
- 12. Levitt, M.; Lifson, S. J Mol Biol 1969, 46, 269.
- 13. Hagler, A. T.; Huler, E.; Lifson, S. J Am Chem Soc 1974, 96, 5319.
- 14. Brant, D. A.; Flory, P. J. J Am Chem Soc 1965, 87, 2791.
- Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. J Phys Chem 1975, 79, 2361.
- Némethy, G.; Pottle, M. S.; Scheraga, H. A. J Phys Chem 1983, 87, 1883.
- 17. Sippl, M. J.; Némethy, G.; Scheraga, H. A. J Phys Chem 1984, 88, 6231.
- 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.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179.
- Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.

- Ripoll, D. R.; Vorobjev, Y. N.; Liwo, A.; Vila, J. A.; Scheraga, H. A. J Mol Biol 1996, 264, 770.
- 22. Gibson, K. D.; Scheraga, H. A. Physiol Chem Phys 1969, 1, 109.
- 23. Gõ, N.; Scheraga, H. A. J Chem Phys 1969, 51, 4751.
- 24. Gõ, N.; Scheraga, H. A. Macromolecules 1976, 9, 535.
- 25. Anfinsen, C. B. Science 1973, 181, 223.
- 26. Gibson, K. D.; Scheraga, H. A. J Comput Chem 1990, 11, 468.
- 27. Gibson, K. D.; Scheraga, H. A. J Comput Chem 1990, 11, 487.
- 28. He, S.; Scheraga, H. A. J Chem Phys 1998, 108, 271.
- 29. He, S.; Scheraga, H. A. J Chem Phys 1998, 108, 287.
- Gibson, K. D.; Scheraga, H. A. Proc Natl Acad Sci USA 1969, 63, 9.
 Gibson, K. D.; Scheraga, H. A. Proc Natl Acad Sci USA 1969, 63,
- 242.
- Crippen, G. M.; Scheraga, H. A. Proc Natl Acad Sci USA 1969, 64, 42.
- 33. Gibson, K. D.; Scheraga, H. A. Comput Biomed Res 1970, 3, 375.
- Crippen, G. M.; Scheraga, H. A. Arch Biochem Biophys 1971, 144, 453.
- Crippen, G. M.; Scheraga, H. A. Arch Biochem Biophys 1971, 144, 462.
- 36. Crippen, G. M.; Scheraga, H. A. J Comput Phys 1973, 12, 491.
- Ooi, T.; Scott, R. A.; Vanderkooi, G.; Scheraga, H. A. J Chem Phys 1967, 46, 4410.
- Yan, J. F.; Vanderkooi, G.; Scheraga, H. A. J Chem Phys 1968, 49, 2713.
- Yan, J. F.; Momany, F. A.; Scheraga, H. A. J Am Chem Soc 1970, 92, 1109.
- Chou, K.-C.; Pottle, M.; Némethy, G.; Ueda, Y.; Scheraga, H. A. J Mol Biol 1982, 162, 89.
- 41. Chou, K.-C.; Scheraga, H. A. Proc Natl Acad Sci USA 1982, 79, 7047.
- 42. Chou, K.-C.; Némethy, G.; Scheraga, H. A. J Mol Biol 1983, 168, 389.
- Chou, K.-C.; Némethy, G.; Scheraga, H. A. Biochemistry 1983, 22, 6213.
- 44. Chou, K.-C.; Némethy, G.; Pottle, M.; Scheraga, H. A. J Mol Biol 1989, 205, 241.
- Chou, K.-C.; Némethy, G.; Scheraga, H. A. J Phys Chem 1983, 87, 2869; Erratum: J Phys Chem 1983, 87, 4772.
- Chou, K.-C.; Némethy, G.; Scheraga, H. A. J Am Chem Soc 1984, 106, 3161; Erratum: J Am Chem Soc 1985, 107, 2199.
- Gerritsen, M.; Chou, K. C.; Némethy, G.; Scheraga, H. A. Biopolymers 1985, 24, 1271; Erratum: Biopolymers 1985, 24, 2177.
- Chou, K.-C.; Maggiora, G. M.; Némethy, G.; Scheraga, H. A. Proc Natl Acad Sci USA 1988, 85, 4295.
- Chou, K.-C.; Maggiora, G. M.; Scheraga, H. A. Proc Natl Acad Sci USA 1992, 89, 7315.
- 50. Gibson, K. D.; Scheraga, H. A. J Comput Chem 1994, 15, 1403.
- 51. Gibson, K. D.; Scheraga, H. A. J Comput Chem 1994, 15, 1414.
- Chou, K.-C.; Némethy, G.; Rumsey, S.; Tuttle, R. W.; Scheraga, H. A. J Mol Biol 1985, 186, 591.
- 53. Chou, K.-C.; Némethy, G.; Rumsey, S.; Tuttle, R. W.; Scheraga, H. A. J Mol Biol 1986, 188, 641.
- Chou, K. C.; Heckel, A.; Némethy, G.; Rumsey, S.; Carlacci, L.; Scheraga, H. A. Proteins Struct Funct Genet 1990, 8, 14.
- 55. Gõ, N.; Scheraga, H. A. Macromolecules 1970, 3, 178.
- 56. Gõ, N.; Scheraga, H. A. Macromolecules 1973, 6, 273.
- 57. Gibson, K. D.; Scheraga, H. A. J Comput Chem 1997, 18, 403.
- 58. Wedemeyer, W. J.; Scheraga, H. A. J Comput Chem 1999, 20, 819.
- 59. Dygert, M.; Gõ, N.; Scheraga, H. A. Macromolecules 1975, 8, 750.
- Gõ, N.; Scheraga, H. A. Macromolecules 1973, 6, 525; Erratum: Macromolecules 1974, 7, 148.
- 61. Gõ, N.; Scheraga, H. A. Macromolecules 1978, 11, 552.

- Warme, P. K.; Momany, F. A.; Rumball, S. V.; Tuttle, R. W.; Scheraga, H. A. Biochemistry 1974, 13, 768.
- Swenson, M. K.; Burgess, A. W.; Scheraga, H. A. In Frontiers in Physicochemical Biology; Pullman, B., Ed.; Academic Press: New York, 1978, p. 115.
- Palmer, K. A.; Scheraga, H. A.; Riordan, J. F.; Vallee, B. L. Proc Natl Acad Sci USA 1986, 83, 1965.
- Acharya, K. R.; Stuart, D. I.; Phillips, D. C.; Scheraga, H. A. J Protein Chem 1990, 9, 549.
- Allen, S. C.; Acharya, K. R.; Palmer, K. A.; Shapiro, R.; Vallee, B. L.; Scheraga, H. A. J Protein Chem 1994, 13, 649.
- Miller, M. H.; Scheraga, H. A. J Polym Sci Polym Symp 1976, 54, 171.
- Miller, M. H.; Némethy, G.; Scheraga, H. A. Macromolecules 1980, 13, 470.
- Miller, M. H.; Némethy, G.; Scheraga, H. A. Macromolecules 1980, 13, 910.
- Némethy, G.; Miller, M. H.; Scheraga, H. A. Macromolecules 1980, 13, 914.
- 71. Némethy, G.; Scheraga, H. A. Biochemistry 1986, 25, 3184.
- 72. Némethy, G.; Scheraga, H. A. Biopolymers 1989, 28, 1573.
- 73. Lee, J.; Scheraga, H. A.; Rackovsky, S. Biopolymers 1997, 40, 595.
- 74. Gõ, M.; Gõ, N.; Scheraga, H. A. Proc Natl Acad Sci USA 1968, 59, 1030.
- 75. Gõ, M.; Gõ, N.; Scheraga, H. A. J Chem Phys 1970, 52, 2060.
- 76. Gõ, M.; Gõ, N.; Scheraga, H. A. J Chem Phys 1971, 54, 4489.
- 77. Gõ, M.; Hesselink, T. F.; Gõ, N.; Scheraga, H. A. Macromolecules 1974, 7, 459.
- 78. Gõ, M.; Scheraga, H. A. Biopolymers 1984, 23, 1961.
- 79. Burgess, A. W.; Scheraga, H. A. J Theor Biol 1975, 53, 403.
- 80. Tanaka, S.; Scheraga, H. A. Proc Natl Acad Sci USA 1975, 72, 3802.
- 81. Tanaka, S.; Scheraga, H. A. Macromolecules 1977, 10, 291.
- 82. Matheson, R. R., Jr.; Scheraga, H. A. Macromolecules 1978, 11, 819.
- Némethy, G.; Scheraga, H. A. Proc Natl Acad Sci USA 1979, 76, 6050.
- Pincus, M. R.; Burgess, A. W.; Scheraga, H. A. Biopolymers 1976, 15, 2485; Erratum: Biopolymers 1977, 16, 468.
- Pincus, M. R.; Zimmerman, S. S.; Scheraga, H. A. Proc Natl Acad Sci USA 1976, 73, 4261.
- Pincus, M. R.; Zimmerman, S. S.; Scheraga, H. A. Proc Natl Acad Sci USA 1977, 74, 2629.
- 87. Pincus, M. R.; Scheraga, H. A. Macromolecules 1979, 12, 633.
- 88. Pincus, M. R.; Scheraga, H. A. Biochemistry 1981, 20, 3960.
- 89. Pincus, M. R.; Scheraga, H. A. Acc Chem Res 1981, 14, 299.
- Scheraga, H. A.; Pincus, M. R.; Burke, K. E. In Structure of Complexes Between Biopolymers and Low Molecular Weight Molecules; Bartmann, W.; Snatzke, G., Eds.; John Wiley: Chichester, 1983, p. 53.
- Smith-Gill, S. J.; Rupley, J. A.; Pincus, M. R.; Carty, R. P.; Scheraga, H. A. Biochemistry 1984, 23, 993.
- 92. Gõ, N. Int J Pept Protein Res 1975, 7, 313.
- 93. Shakhnovich, E.; Gutin, A. M. Biophys Chem 1989, 34, 187.
- 94. Klimov, D. K.; Thirumalai, D. Folding Design 1998, 3, 127.
- Bryngelson, J. D.; Wolynes, P. G. Proc Natl Acad Sci USA 1987, 84, 7524.
- 96. Hao, M. H.; Scheraga, H. A. J Phys Chem 1994, 98, 4940.
- 97. Hao, M. H.; Scheraga, H. A. J Phys Chem 1994, 98, 9882.
- 98. Hao, M. H.; Scheraga, H. A. J Chem Phys 1995, 102, 1334.
- 99. Hao, M.-H.; Scheraga, H. A. Physica A 1997, 244, 124.
- 100. Liwo, A.; Kazmierkiewicz, R.; Czaplewski, C.; Groth, M.; Oldziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. J Comput Chem 1998, 19, 259.
- 101. Hao, M.-H.; Scheraga, H. A. J Mol Biol 1998, 277, 973.
- 102. Hao, M.-H.; Scheraga, H. A. Acc Chem Res 1998, 31, 433.

- 103. Scheraga, M. A.; Hao, M.-H. Adv Chem Phys 1999, 105, 243.
- 104. Lee, J. Phys Rev Lett 1993, 71, 211; Erratum: Phys Rev Lett 1993, 71, 2353.
- 105. Scheraga, H. A.; Lee, J.; Pillardy, J.; Ye, Y.-Y.; Liwo, A.; Ripoll, D. R. J Global Optimizat 1999, 15, 235.
- 106. Lambert, M. H.; Scheraga, H. A. J Comput Chem 1989, 10, 770.
- 107. Lambert, M. H.; Scheraga, H. A. J Comput Chem 1989, 10, 798.
- 108. Lambert, M. H.; Scheraga, H. A. J Comput Chem 1989, 10, 817.
- 109. Kidera, A.; Konishi, Y.; Oka, M.; Ooi, T.; Scheraga, H. A. J Protein Chem 1985, 4, 23.
- 110. Kidera, A.; Konishi, Y.; Ooi, T.; Scheraga, H. A. J Protein Chem 1985, 4, 265.
- 111. Li, Z.; Scheraga, H. A. Proc Natl Acad Sci USA 1987, 84, 6611.
- 112. Li, Z.; Scheraga, H. A. J Mol Struct (Theochem) 1988, 179, 333.
- 113. Wales, D. J.; Scheraga, H. A. Science 1999, 285, 1368.
- 114. Ripoll, D. R.; Scheraga, H. A. Biopolymers 1988, 27, 1283.
- 115. Ripoll, D. R.; Scheraga, H. A. J Protein Chem 1989, 8, 263.
- 116. Ripoll, D. R.; Liwo, A.; Scheraga, H. A. Biopolymers 1998, 46, 117.
- 117. Piela, L.; Scheraga, H. A. Biopolymers 1987, 26, S33.
- 118. Faerman, C. H.; Ripoll, D. R. Proteins 1992, 12, 111.
- 119. Ripoll, D. R. Int J Pept Protein Res 1992, 40, 575.
- 120. Liwo, A.; Tempczyk, A.; Ołdziej, A.; Shenderovich, M. D.; Hruby, V. J.; Talluri, S.; Ciarkowski, J.; Kasprzykowski, F.; Lankiewicz, L.; Grzonka, Z. Biopolymers 1996, 38, 157.
- 121. Vila, J. A.; Ripoll, D. R.; Villegas, M. E.; Vorobjev, Y. N.; Scheraga, H. A. Biophys J 1998, 75, 2637.
- 122. Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. Proc Natl Acad Sci USA 2000, 97, 13075.
- 123. Purisima, E. O.; Scheraga, H. A. J Mol Biol 1987, 196, 697.
- 124. Piela, L.; Kostrowicki, J.; Scheraga, H. A. J Phys Chem 1989, 93, 3339.
- 125. Kostrowicki, J.; Piela, L.; Cherayil, B. J.; Scheraga, H. A. J Phys Chem 1991, 95, 4113.
- 126. Kostrowicki, J.; Scheraga, H. A. J Phys Chem 1992, 96, 7442.
- 127. Kostrowicki, J.; Scheraga, H. A. In Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding; Pardalos, P. M.; Shalloway, D.; Xue, G., Eds.; DIMACS Series in Discrete Mathematics and Theoretical Computer Science (Am Math Soc) 1996, 23, 123.
- 128. Kostrowicki, J.; Scheraga, H. A. Comput Polym Sci 1995, 5, 47.
- 129. Pillardy, J.; Piela, L. J Phys Chem 1995, 99, 11805.
- 130. Pillardy, J.; Liwo, A.; Scheraga, H. A. J Phys Chem 1999, A103, 9370.
- Pillardy, J.; Liwo, A.; Groth, M.; Scheraga, H. A. J Phys Chem 1999, B103, 7353.
- Pillardy, J.; Wawak, R. J.; Arnautova, Y. A.; Czaplewski, C.; Scheraga, H. A. J Am Chem Soc 2000, 122, 907.
- 133. Lee, J.; Scheraga, H. A.; Rackovsky, S. J Comput Chem 1997, 18, 1222.
- 134. Lee, J.; Scheraga, H. A.; Rackovsky, S. Biopolymers 1998, 46, 103.
- 135. Lee, J.; Scheraga, H. A. Int J Quantum Chem 1999, 75, 255.
- Pillardy, J.; Czaplewski, C.; Wedemeyer, W. J.; Scheraga, H. A. Helv Chim Acta 2000, 83, 2214.
- 137. Noguti, T.; Gõ, N. Biopolymers 1985, 24, 527.
- 138. Vanderbilt, D.; Louie, S. G. J Comput Phys 1984, 56, 259.
- 139. Wedemeyer, W. J. Ph.D. thesis, Cornell University, 1998, p. 236.
- 140. Cheng, B.; Nayeem, A.; Scheraga, H. A. J Comput Chem 1996, 17, 1453.
- 141. Dunbrack, R. L., Jr.; Cohen, F. E. Protein Sci 1997, 6, 1661.
- 142. Pierce, N. A.; Spriet, J. A.; Desmet, J.; Mayo, S. L. J Comput Chem 2000, 21, 999.

- 143. Lee, J.; Pillardy, J.; Czaplewski, C.; Arnautova, Y.; Ripoll, D. R.; Liwo, A.; Gibson, K. D.; Wawak, R. J.; Scheraga, H. A. Comput Phys Commun 2000, 128, 399.
- 144. Gibson, K. D.; Scheraga, H. A. J Phys Chem 1995, 99, 3752.
- 145. Gibson, K. D.; Scheraga, H. A. J Phys Chem 1995, 99, 3765.
- 146. Wawak, R. J.; Gibson, K. D.; Liwo, A.; Scheraga, H. A. Proc Natl Acad Sci USA 1996, 93, 1743.
- 147. Wawak, R. J.; Pillardy, J.; Liwo, A.; Gibson, K. D.; Scheraga, H. A. J Phys Chem 1998, 102, 2904.
- 148. Pillardy, J.; Arnautova, Y. A.; Czaplewski, C.; Gibson, K. P.; Scheraga, H. A. Proc Natl Acad Sci USA, submitted.
- 149. Verwer, P.; Leusen, F. J. J. Rev Comput Chem 1998, 12, 327.
- 150. van Eijck, B. P.; Kroon, J. J Phys Chem 2000, B104, 8089.
- 151. Wedemeyer, W. J.; Arnautova, Y. A.; Pillardy, J.; Wawak, R. J.; Czaplewski, C.; Scheraga, H. A. J Phys Chem 2000, B104, 8090.
- Liwo, A.; Oldziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. J Comput Chem 1997, 18, 849.
- Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Ołdziej, S.; Scheraga, H. A. J Comput Chem 1997, 18, 874.
- Liwo, A.; Pillardy, J.; Kaźmierkiewicz, R.; Wawak, R. J.; Groth, M.; Czaplewski, C.; Ołdziej, S.; Scheraga, H. A. Theor Chem Acc 1999, 101, 16.
- 155. Liwo, A.; Pillardy, J.; Czaplewski, C.; Lee, J.; Ripoll, D. R.; Groth, M.; Rodziewicz–Motowidło, S.; Kaźmierkiewicz, R.; Wawak, R. J.; Ołdziej, S.; Scheraga, H. A. RECOMB 2000, Proceedings of the Fourth Annual International Conference on Computational Molecular Biology; Shamir, R.; Miyano, S.; Istrail, S.; Pevzner, P.; Waterman, M., Eds.; April 8–11, 2000, Tokyo, Japan. ACM: New York, 2000, p. 193.
- 156. Gay, J. G.; Berne, B. J. J Chem Phys 1981, 74, 3316.
- 157. Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. J Chem Phys 2001, 115, 2323.

- 158. Kubo, R. J Phys Soc Jpn 1962, 17, 1100.
- Lee, J.; Ripoll, D. R.; Czaplewski, C.; Pillardy, J.; Wedemeyer, W. J.; Scheraga, H. A. J Phys Chem 2001, B105, 7291.
- 160. Pillardy, J.; Czaplewski, C.; Liwo, A.; Wedemeyer, W. J.; Lee, J.; Ripoll, D. R.; Arlukowicz, P.; Ołdziej, S.; Arnautova, Y. A.; Scheraga, H. A. J Phys Chem 2001, B105, 7299.
- 161. Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. Protein Sci 1993, 2, 1697.
- Vila, J.; Williams, R. L.; Vasquez, M.; Scheraga, H. A. Proteins Struct Funct Genet 1991, 10, 199.
- Williams, R. L.; Vila, J.; Perrot, G.; Scheraga, H. A. Proteins Struct Funct Genet 1992, 14, 110.
- 164. Lee, J.; Liwo, A.; Scheraga, H. A. Proc Natl Acad Sci USA 1999, 96, 2025.
- Liwo, A.; Lee, J.; Ripoll, D. R.; Pillardy, J.; Scheraga, H. A. Proc Natl Acad Sci USA 1999, 96, 5482.
- 166. Lee, J.; Liwo, A.; Ripoll, D. R.; Pillardy, J.; Scheraga, H. A. Proteins Struct Funct Genet 1999, Suppl. 3, p. 204.
- 167. Lee, J.; Liwo, A.; Ripoll, D. R.; Pillardy, J.; Saunders, J. A.; Gibson, K. D.; Scheraga, H. A. Int J Quantum Chem 2000, 77, 90.
- Orengo, C. A.; Bray, J. E.; Hubbard, T.; LoConte, L.; Sillitoe, I. Proteins Struct Funct Genet 1999, Suppl. 3, p. 149.
- 169. Liwo, A.; Czaplewski, C.; Pillardy, J.; Lee, J.; Ripoll, D. R.; Scheraga, H. A. 5th World Congress of Theoretically Oriented Chemists, WATOC '99, Imperial College, London, August 1–6, 1999, Abstr. 556.
- 170. Pillardy, J.; Czaplewski, C.; Liwo, A.; Lee, J.; Ripoll, D. R.; Kaźmierkiewicz, R.; Ołdziej, S.; Wedemeyer, W. J.; Gibson, K. D.; Arnautova, Y. A.; Saunders, J.; Ye, Y.-J.; Scheraga, H. A. Proc Natl Acad Sci USA 2001, 98, 2329.