DIMACS Series in Discrete Mathematics
and Theoretical Computer Science

Global Optimization Approaches in Protein Folding and Peptide Docking

C. A. Floudas*, J. L. Klepeis, and P. M. Pardalos

ABSTRACT. The recent advances in genetic engineering, high powered computing and global optimization continue to stimulate interest in the area of molecular modeling and protein structure prediction. The goal of these efforts is the ability to correctly predict native protein conformations and the binding interactions of macromolecules. These two problems currently dominate the field of computational chemistry and, through the use of detailed molecular models, they have also greatly influenced research in the area of global optimization. This article examines some aspects related to conformational energy modeling and reviews a variety of global optimization approaches developed for the protein folding and peptide docking problems.

1. Introduction

Proteins are undoubtedly the most complex and vital molecules in nature. This complexity arises from an intricate balance of intra- and inter-molecular interactions which define the native three-dimensional structure of the system, and subsequently its biological functionality. Anfinsen's hypothesis [**12**] suggests that this native structure is in a state of thermodynamic equilibrium corresponding to the conformation with the lowest free energy. Experimental studies have since shown that, under native physiological conditions and after denaturation, globular proteins spontaneously refold to their unique, native structure [**84**]. Understanding the transition of a protein from a disordered state to its native state defines the protein folding problem.

A natural extension of the protein folding problem is the related problem of peptide docking. Here equilibrium structures must be identified for a macromolecule-ligand complex. One aspect involves identifying the conformation of the binding site for the "pocket" molecule, which can be accomplished experimentally or approached as an independent protein folding problem. Additionally, equilibrium

1991 *Mathematics Subject Classification.* Primary 54C40, 14E20; Secondary 46E25, 20C20.
Key words and phrases. Global optimization, protein folding, peptide docking.

* Author to whom all correspondence should be addressed.

C.A.F. acknowledges support from National Science Foundation, National Institutes of Health and Air Force Office of Scientific Research; P.M.P. acknowledges support from National Science Foundation and DIMACS.

structures must be identified for a number of candidate "docking" molecules, and a measure of their relative binding affinity must be quantified and compared.

Recent advances in genetic engineering have heightened the interest in research related to predicting native protein folding and docking conformations. The ability to predict these structures is of great theoretical interest, especially in the fields of biophysics and biochemistry. Moreover, the applications of such knowledge also promise to be exciting. For example, the ability to predict these structures would greatly increase our understanding of hereditary and infectious diseases and aid in the interpretation of genome data. In addition to these advances, the ability to understand peptide docking would likely revolutionize the process of de novo drug design.

The use of computational techniques and simulations in addressing the protein folding and peptide docking problems became possible through the introduction of qualitative and quantitative methods for modeling these systems. The development of realistic energy models also established a link to the field of global optimization, where, based on Anfinsen's hypothesis, the quantity to be optimized is the energy of the system. However, because of the computational complexity associated with these problems, only the most efficient global optimization strategies will be successful. In fact, because the number of local minima is vast, the corresponding problem formulation has earned the simple yet suggestive title of "multiple-minima" problem. The basis for these difficulties is best summarized by Levinthal's paradox [**102**]. This paradox suggests a contradiction between the almost infinite number of possible stable states that the system may sample and the relatively short time scale required for actual protein folding. In contradiction to Anfinsen's thermodynamic hypothesis, Levinthal's observations suggest that the native state is the lowest kinetically accessible free energy minimum, which may be different from the true global minimum. These principles have been used to develop computational techniques for predicting protein folding pathways [**27, 101, 155**]. Such techniques attempt to map the shape of the energy hypersurface and determine whether this surface "funnels" a protein towards a dominant conformational basin. By invoking the thermodynamic hypothesis, the overall shape of the energy hypersurface is neglected and the problem can be formulated in terms of global minimization, which requires the use of effective global optimization techniques. If this formulation is to reproduce the behavior of realistic systems, the folding of actual proteins should not be kinetically hindered. This has been verified for various systems by performing denaturation–refolding experiments. In addition, by introducing structural characteristics whose formation may act as kinetic barriers, such as the formation of disulfide bonds, the performance of the thermodynamic equilibrium model should be improved.

Based on the complexity of the energy hypersurface, there is an obvious need for the development of efficient global optimization techniques. Although the energy can be expressed analytically, exhaustive searches are possible for only the smallest of systems. These observations, and the importance of the protein folding and peptide docking problems, have propelled the introduction of new global search strategies specifically designed for these problems. In addition to describing the fundamental modeling issues for the protein folding and peptide docking problems, this review critically assesses a variety of global optimization approaches applied to these problems. A number of proceedings and reviews have also focused on global optimization issues in molecular systems [**95, 132, 138, 139, 156, 177**]

2. Energy Modeling

2.1. Potential Energy. In a rigorous sense, the dynamics of molecular motion are represented by Schrödinger's equation, which, when solved, yields an exact description of the energy as a function of structure. This complexity can be reduced by invoking the Born-Oppenheimer approximation, under which nuclear and electron motions can be considered separately. In addition, because molecular configuration problems require only relative energies, a further simplification has been the development of empirical models. However, in cases where experimental data is lacking, quantum mechanical ab-initio calculations must be performed and used in combination with experimental results to effectively describe the molecular system.

A number of empirically-based molecular mechanics models have been developed for protein systems, including AMBER [**74, 187, 188**], CHARMM [**19**], DISCOVER [**37**], ECEPP [**121, 122, 123**], ECEPP/2 [**130**], ECEPP/3 [**131**], ENCAD [**36, 103**], GROMOS [**175**], MM2 [**7**], and MM3 [**8, 106, 107**]. These models, also known as force fields, are typically expressed as summations of several potential energy components, with the mathematical form of individual energy terms based on the phenomenological nature of that term. A general total energy equation, such as Equation 2.1, should include terms for bond stretching (E_{bond}), angle bending (E_{angle}), torsion (E_{tor}), nonbonded (E_{nb}) and coupled (E_{cross}) interactions.

$$E_{tot} = E_{bond} + E_{angle} + E_{tor} + E_{nb} + E_{cross} \tag{2.1}$$

Bond stretching and angle bending energies are included in those force fields that allow flexible geometries. A simple representation for both terms is based on the harmonic approximation, which corresponds to the classical description of the movement of a spring (by Hooke's law). The simplest approach, based on the fact that most bonds are near the minimum of their respective energy well, employs a quadratic term to model bond stretching and angle bending energies, as shown in Equations 2.2 and 2.3.

$$E_{bond} = \frac{k_{bond}}{2}(l - l_o)^2 \tag{2.2}$$

$$E_{angle} = \frac{k_{angle}}{2}(\theta - \theta_o)^2 \tag{2.3}$$

These equations act as penalty functions to force bond distances and bond angles, l and θ, to reference bond lengths and distances, l_o and θ_o, whose values depend on the specific atoms involved. In actuality, these energy terms are more complicated. For bond energies cubic terms are often introduced, and angle energy terms usually include higher power expansions.

Torsional terms are used to describe the internal rotation energy of torsion angles, which exist between all atoms with a 1–4 relationship (separated by two atoms). For rigid geometry force fields, these torsion angles can be used to define a set of independent variables that effectively describe any protein conformation. This approximately reduces the number of variables by a factor of 3 over those force fields that use a Cartesian coordinate system to describe flexible molecular geometries. In addition, bond and angle energies can be neglected for rigid geometry force fields.

The torsion energy expression is typically represented by a Fourier series expansion which, as shown in Equation 2.4, includes three terms.

$$E_{tor} = E_1(1 - cos\phi) + E_2(1 - cos2\phi) + E_3(1 - cos3\phi) \tag{2.4}$$

The parameters involved in this expansion, namely E_1, E_2 and E_3, are torsional barriers that are usually specified for the pair of atoms around which the torsion occurs. Each term can be interpreted physically. The 1-x ($cos\phi$) symmetry term accounts for those nonbonded interactions not included in general nonbonded terms. The 2-x ($cos2\phi$) symmetry term is related to the interactions of orbitals, while the 3-x ($cos3\phi$) symmetry term describes steric contributions.

Nonbonded energy terms attempt to model electrostatic and van der Waals interactions between those atoms that are not connected to each other or through a common atom. Typically, a Coulombic term is used to represent electrostatic energies based on atomic point charges, as shown in Equation 2.5.

$$E_{elec} = \frac{Q_i Q_j}{\epsilon R_{ij}} \tag{2.5}$$

Here Q_i and Q_j represent the two point charges, while R_{ij} equals the distances between these two points. In some force fields, Coulombic interactions are modified by changing the dependence of the dielectric constant, ϵ. In general, van der Waals interactions are modeled using a 6-12 Lennard-Jones potential energy term. This expression, shown in Equation 2.6, consists of a repulsion and attraction term.

$$E_{vdw} = \epsilon_{ij}\left[\left(\frac{R^*_{ij}}{R_{ij}}\right)^{12} - 2\left(\frac{R^*_{ij}}{R_{ij}}\right)^{6}\right] \tag{2.6}$$

The energy minimum for a given atomic pair is described by the potential depth, ϵ_{ij}, and position, R^*_{ij}. Other force fields model van der Waals interactions using a modified Hill equation, which replaces the twelfth power term in Equation 2.6 with an exponential term [**7, 8**]. Different approaches are also used to describe nonbonded interactions between those atoms that may form hydrogen bonds. Some force fields model these interactions using only Coulombic terms, while other force fields employ special functions, such as a modified 10-12 Lennard-Jones type potential term [**113**], as shown in Equation 2.7.

$$E_{hbond} = \epsilon_{ij}\left[5\left(\frac{R^*_{ij}}{R_{ij}}\right)^{12} - 6\left(\frac{R^*_{ij}}{R_{ij}}\right)^{10}\right] \tag{2.7}$$

The cross term, shown in Equation 2.1, accounts for interactions due to the inherent coupling between bonds, angles and torsions. Generally, these terms are small, and in many force fields they are neglected. Correction terms, which vary for each force field, are also typically added to the general energy equation. For example, the formation of disulfide bridges can be enforced by adding a penalty term to constrain the values of specified bond angles and bond lengths. Correction terms have also been used to adjust conformational energies according to the configurations of proline and hydroxyproline residues [**131**].

2.2. Solvation Energy. In general, the energetic description of a protein should also include solvent effects. A theoretically simple approach would be to explicitly surround the peptide with solvent molecules and compute potential energy contributions for intra- and inter-molecular interactions. Although several methods have been developed to model hydration (water) effects, each simulation is valid only for specific conditions. Explicit calculations also tend to greatly increase the computational cost of the simulation. In addition, solvent configurations are not rigid, so these calculations must consider an average solvent-peptide configuration, which is typically generated by a number of Monte Carlo (MC) or Molecular Dynamics (MD) simulations [**20, 90, 168**].

An alternative way for effectively considering average solvent effects is to use implicit solvation models. Some applications of the Reference Interaction Site Model (RISM) theory have been used to estimate these interactions for hydrated groups [**25, 85**]. One complication involves the solvent's influence on electrostatic interaction energies because of the implicit relationship between dielectric constants and solvation. In some cases this problem has been addressed by modifying the representation of the dielectric constant. In reality, however, the rigorous treatment of electrostatic interactions involves the solution of the Poisson-Boltzmann equation [**53, 183**]. A number of reviews consider the treatment of electrostatic terms and solvation effects [**38, 186**].

Other simple and computationally feasible implicit solvation models are based on empirical representations of the solvation energy. For these models the solvation energy of each functional group is related to the interaction of the solvent with a hydration shell for the particular group. These terms are then summed together to provide a total solvation energy for the system. One form of such an empirical model is based on the assumption that the solvation energy is proportional to the solvent accessible surface of each functional group, as shown in Equation 2.8.

$$E_{solv} = \sum_{i=1}^{N} A_i \sigma_i \tag{2.8}$$

The solvent accessible surface area, A_i, is multiplied by a free energy density parameter, σ_i, to give the solvation energy. A number of algorithms have been developed for calculating these surface areas [**29, 44, 45, 141, 181**]. Although several of these are relatively efficient, the appearance of discontinuities has been one complication in considering solvent accessible surface areas. In addition, a large number of parameterization strategies have been used to derive appropriate σ_i [**43, 136, 182, 190, 191**]. Some generalized Born models are also based on solvent accessible surface area calculations, although electrostatic and solvent contributions are subsequently adjusted [**1**].

A second method, given in Equation 2.9, assumes that the solvation energy is proportional to the solvent accessible volume of a hydration layer.

$$E_{solv} = \sum_{i=1}^{N} V_i \delta_i \tag{2.9}$$

Several methods have been developed for calculating the hydration volumes, V_i, and the free energy parameters, δ_i [**13, 68, 77, 78, 79, 80**]. A recent and computationally inexpensive method has been based on a Gaussian approximation for the volume of a hydration layer [**13**]. This method also inherently avoids numerical problems associated with possible discontinuities. It should be noted that when any of these empirical methods are used in conjunction with other force field computations, free energy parameters should be based on experimental free energies for gas to aqueous solution transfer, rather than for transfer between organic solvent and aqueous solution.

2.3. Free Energy. From a thermodynamic point of view, the calculation of the minimum free energy must include potential energy, as well as entropic contributions. This can be accomplished through detailed MD simulations. However, these search methods are not feasible for full global optimization searches of peptide systems. An alternative treatment of this problem has been based on the development of classical statistical mechanical partition functions. Although there has been some debate related to the development of the appropriate partition function [**46, 54, 55**], it is generally accepted that, for small fluctuations, the free energy contribution can be represented by a harmonic approximation [**24, 81**]. Using this approximation, the normalized statistical weight of a given conformation, w_i, can be calculated using Equation 2.10.

$$w_i = \frac{(2\Pi RT)^{k/2} \exp \frac{-\Delta U_i}{RT}}{Z\,|F_i|^{-1/2}} \tag{2.10}$$

Here k refers to the dimensionality of the problem, R is the gas constant, T is the temperature, $|F_i|$ is the matrix of second derivatives of the potential energy evaluated at the i^{th} minimum, and ΔU_i is the difference in energy between the i^{th} and the global minimum potential energy conformation. A statistical weight, w_i can be associated with a conformational free energy (of the i^{th} minimum), G_i, using the following equation:

$$G_i = -RT \ln w_i \tag{2.11}$$

The rigorous evaluation of Equations 2.10 and 2.11 requires the identification of an ensemble of low energy conformers within a given ΔU of the global minimum conformer. Once this ensemble has been identified, relative free energies can be calculated using Equation 2.12.

$$\Delta G_i = G_i - G_o = -RTln\left(\frac{w_i}{w_o}\right) \tag{2.12}$$

Here G_o and w_o represent the free energy and statistical weight of the global minimum potential energy conformation.

2.4. Peptide Docking. The complexity associated with the prediction of optimum macromolecular-ligand conformations also complicates the task of modeling these interactions. First, the binding site of the target globular protein must be correctly characterized, along with the general orientation of the potential ligands. The second task is to select potential binders (ligands), dock these ligands to the active site, and assign a "score" to each complex. These "scores" may then be used to rank binding affinities for a given list of ligands.

The first task usually requires experimental structure determination of the binding site. One particular example that has received much attention has been the major histocompatibility complex (MHC), which plays a major role in immune response [**171**]. Crystallographic studies have been performed for the two major classes of MHC molecules, class I [**51, 163**] and II [**166**]. This information is invaluable because it can be used to approximate rigid binding sites, which greatly reduces the translational space that must be explored in a conformational search.

The other fundamental feature of the peptide docking problem is the development of accurate scoring functions. Due to the computational complexity of rigorous energy calculations, many methods have relied on qualitative modeling of the peptide docking interactions. In the case of a rigid binding approximation, the use of shape complementarity has had some limited success [**93**]. These algorithms model the ligand and macromolecule according to their surface topology, and attempt to identify which complexes exhibit the best "fit". Here scoring functions are based on the complementarity of the molecules, which, in most cases, is related to their solvent accessible surface areas [**29, 30, 97**]. The obvious strength of these methods is that they can be made computationally efficient and used to screen large databases of potential ligands. However, studies comparing computational results to experimentally derived, native complexes indicate that many non-native low energy structures are identified. Other methods attempt to refine the scoring function by adding additional criteria, such as conformational energy and solvation energy.

On the other hand, it has been demonstrated that exact modeling of binding free energies provides results in nearly exact quantitative agreement with experimental results [**59, 74, 162**]. In contrast to the rigid description of docking, these methods allow for flexibility of both the ligand and receptor molecules. However, for general peptide docking problems, these thermodynamic integration and free energy perturbation methods are computationally infeasible with current computing power. These problems are only tractable when approximate structures are known and relatively small. More detail on these methods can be found elsewhere [**120, 148**]. A comprehensive theoretical treatment of the thermodynamics of binding processes in macromolecules is also available [**41**].

More computationally feasible methods are based on calculating binding free energies using empirically derived free energy functions. Some methods of approximating free energy functions involve structure based potentials [**184**]. Other approximations utilize parameterization of experimental data to construct scoring functions based on conformational energy, hydrophobic and hydrophilic surface areas, and hydrogen bonding geometries [**66, 179**]. However, these methods are generally nontransferable. A general scoring function is simply represented in Equation 2.13.

$$\Delta G = \Delta G_{complex} - \Delta G_{ligand} - \Delta G_{pocket} \tag{2.13}$$

Here $\Delta G_{complex}$, ΔG_{binder} and ΔG_{pocket} are the free energies of the peptide docking complex, the free ligand and the free pocket. A more universal approach, applicable to flexible ligands, is to base the free energy calculation on general force field models, which involve potential energy functions similar to those described in the preceding sections. This free energy function must also account for solvation energy, which can be calculated from structure based solvation terms or continuum based models of solvation. Rigorously, entropic effects of side chain rotations should also be considered. Reviews of methods used to evaluate binding free energies can be found elsewhere [**71, 174**].

3. Global Optimization

Before exploring global optimization approaches for the protein folding and peptide docking problems, it is helpful to review some general local optimization methods used in optimizing conformational energy functions. By far, energy minimization techniques have been the most successful in identifying local minima. Nonderivative minimization techniques are appealing because they can be easily implemented and do not require derivative computations; however, they are generally inefficient and not reliable. Gradient methods, which provide detailed objective function information through first and second derivative information, are very successful in identifying local minima. These methods include steepest descent, conjugate gradient, and a number of modified Newton's methods. A detailed review of these techniques is available [**157**].

Other local optimization methods include molecular dynamics and Monte Carlo searches [**117**]. Although molecular dynamics theoretically can provide all accessible states, the need for small time scales limits the applicability of this technique for computing global minima. However, it has been used extensively for refinement of crystal structures. Similarly, Monte Carlo methods are typically more successful when used in combination with energy minimization techniques.

Since a local minimum can be effectively identified using the aforementioned local minimization techniques, the problem becomes one of locating the conformation with global minimum free energy. Complete exploration of the complicated energy hypersurface is only feasible for small systems. For example, it has been estimated that the conformational space of the five residue oligopeptide met-enkephalin exhibits 10^{11} local minima [**104**]. This illustrates the need for global optimization algorithms that efficiently search this space and/or identify low energy regions. Information on the application of global optimization techniques to solve NP-hard problems can be found elsewhere [**48, 69**].

Fundamentally, global optimization approaches can be classified as being deterministic or stochastic. Stochastic methods rely on probabilistic descriptions to aid in locating the global minimum energy, whereas deterministic techniques provide a certain level of assurance in locating the global minimum. With few exceptions, the methods used to address protein folding and peptide docking problems can be classified as stochastic. Two smoothing methods, namely the diffusion equation and packet annealing techniques, claim to be deterministic. However, as will be shown, a number of approximations must be made to practically implement these methods. Another exception, the αBB (branch and bound) approach [**2, 3, 4, 5, 6, 9**], has specifically been developed to deterministically treat global optimization problems such as those found in protein folding and peptide docking.

The following discussion is based on a classification of the global optimization approaches into five areas - combinatorial, genetic algorithms, simulated annealing and sampling, smoothing/deformation and branch and bound. The first subsections focus on the application of these approaches to the area of protein folding, while the sixth subsection provides a summary of computational requirements for a sampling of these approaches. The seventh subsection comments on the use of global optimization techniques when considering solvation and entropic effects. Finally, the last section concentrates on the use of global optimization approaches in the area of peptide docking.

3.1. Combinatorial Methods. A number of approaches decompose the overall large scale protein folding problem into smaller subproblems. The basic idea is similar to fragment assembly; that is, an ensemble of structures is identified for small fragments of the original protein and then joined to form overall conformations which can be minimized. In many cases fragment conformations are chosen to be local minima or low energy conformers of the particular fragment. For example, a fragment may consist of a single amino acid residue for which the number and location of minima can easily be identified [**195**].

A basic implementation of this fragment assembly methodology has been referred to as the build-up method [**52, 176**]. The basic idea relies on successively combining smaller fragments into larger ones after retaining and minimizing a number of low energy conformers for each smaller fragment. Even this simplification results in a combinatorial explosion of possible conformations, and an appropriate energy cutoff must be employed. Note that in these searches the role of short range interactions is overemphasized and many important subsequences may be overlooked. However, the application has been successful for predicting global minimum energy conformations for proteins that possess special secondary structure, such as α–helices.

A modified build-up type algorithm is based on the principles of dynamic programming [**172, 173**]. As before, a finite number of discrete states are used to define each residue, with these states corresponding to local minima of the energy function. The number of intermediate conformations is reduced by making stagewise decisions, as in dynamic programming. However, because the principle of optimality is not directly applicable to the protein folding problem, more than just the best path to each state is retained. As in the buildup method, the number of retained states is related to a cutoff energy parameter. The general procedure has been tested on two formulations. In the first formulation, conformations were built only by discrete search followed by an overall local minimization. The second formulation, which gave improved results, uses limited local minimization within the build-up stages of the algorithm.

3.2. Genetic Algorithms. Genetic algorithms can be classified as optimization techniques that are qualitatively based on the principles of evolutionary theory [**67**]. For these methods, the variables of the optimization problem (e.g., dihedral angles) correspond to the gene sequence of a given chromosone. The algorithm relies on the manipulation of a population of chromosones through processes known as mutation, selection and recombination.

In general, the initial population is generated by selecting random values for a specified number of chromosones. The fitness of a given chromosone is assigned based on the objective function value (e.g., potential energy) represented by that chromosone. In order to create a child generation, a mating population must first be selected from the parent population. This selection is typically based on a combination of fitness and probability information. Children can then be created via crossover and mutation processes. Crossover is achieved by combining chromosone segments (sequences of variable values) from each of the parent chromosones. A number of random and probability based methods for choosing crossover variables have been developed [**50, 56, 99, 100**]. Once a child generation has been created, further alterations are achieved through mutation. Many methods can be envisioned [**50, 56, 99, 100**]. For example, the effects of point mutations can be

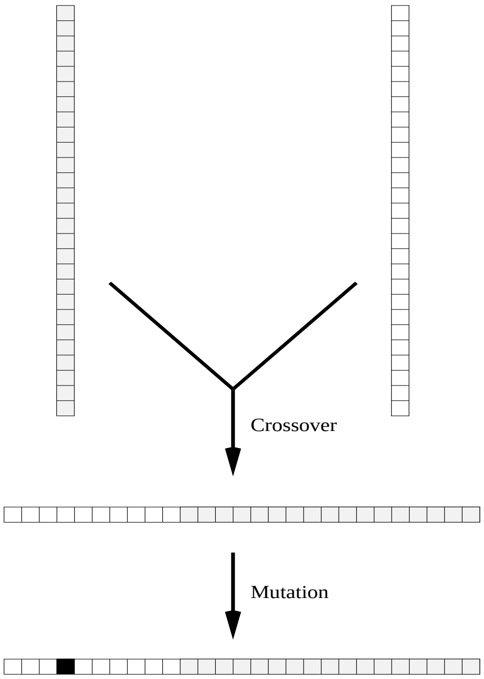

FIGURE 1. Pictorial representation of simple crossover and mutation in genetic algorithms

mimicked by randomly changing the value of a single variable. A simple example of crossover and mutation is given in Figure 1. Finally, the fitness of each child is evaluated, and a subset of the children are used to replace the parent population. This evolutionary process is repeated until the imposed termination criteria are achieved.

In recent years, genetic algorithms have been applied to the protein folding problem. By using a rotamer library of amino acid side chain conformations and entropically based criteria, low energy conformations with native-like structure have been identified for a number of proteins [**170**]. Other implementations using rotamer libraries have been relatively successful in identifying global minimum conformations for polypeptides with special structure (e.g., α-helical) [**100**]. A modified genetic algorithm, which evaluates fitness based on locally minimized conformations, has been shown to improve performance [**75**]. Prediction of low energy conformations via genetic algorithms has also been attempted for larger peptides ($>$ 20 residues). In most cases, these methods employ knowledge based fitness functions [**164**] and conformational libraries from databases of experimentally derived structures. In some of these applications experimental structures were included in the library, which can obviously bias the prediction of near native conformations. The prediction of near native structures worsened for both larger examples and for those not contained in the database [**100, 169**].

In general, there tend to be significant differences in both the formulation of the fitness function and the exact implementation of the algorithmic techniques. The development of these criteria and conditions greatly influences the ability to predict global minimum energy conformations. In addition, when variables are highly interdependent, as in the protein folding problem, the specifications of crossover and mutation become even more important. For these reasons, the application of genetic algorithms will most likely be limited to specific classes of the protein folding problem.

3.3. Simulated Annealing and Sampling Methods. Before describing sampling and simulated annealing techniques, it is beneficial to examine the Metropolis Monte Carlo approach more closely. The fundamental Metropolis enhancement over basic Monte Carlo sampling is the introduction of a Boltzmann distribution in directing the conformational search. The approach is implemented by choosing an initial arbitrary conformation and obtaining its conformational energy, E_i. A new conformation can then be generated by randomly or probabilistically altering some variables, and its energy, E_n, is subsequently evaluated. This new conformation is accepted according to the condition given in Equation 3.1.

$$P_B(T,E) \geq RND \tag{3.1}$$

Here, $P_B(T,E)$ is a probability factor, and RND represents a randomly generated number between 0 and 1. If the condition is met, the conformation is accepted. For the Metropolis technique the probability factor is based on a Boltzmann distribution, as shown in Equation 3.2.

$$p_B(T,E) = e^{\frac{-(E_n - E_i)}{k_B T}} \tag{3.2}$$

k_B is the Boltzmann's constant, in the appropriate units, and T is the temperature. Note that all lower energy conformers are accepted because $p_B(T,E) \geq 1 \; \forall \; E_n \leq E_i$. Although the search is directed toward lower energy regions, uphill moves are randomly allowed, which provides an escape from local minima.

Obviously, the success of a Metropolis Monte Carlo search is highly dependent on the choice of the temperature, T. If the value is excessively large, many high energy conformations will be accepted at the expense of searching low energy regions. On the other hand, low temperatures will direct the search towards the local minima nearest to the starting conformation. This major drawback was overcome by introducing the concept of a simulated annealing cooling schedule [**86**]. Physically, this method is analogous to a crystallization process. The system begins at a high temperature, at which all states are accessible, and brought down to a temperature low enough so that the system is essentially "frozen" at a low energy conformation. If the temperature decrease is applied correctly, and sufficiently slow, the global minimum energy conformation can be identified.

The efficiency of simulated annealing techniques to locate global minimum energy conformations has been evaluated through several studies, and in some cases conflicting results have been observed. For met-enkephalin, simulated annealing algorithms were shown to identify the lowest energy structure using both the AMBER [**192**] and ECEPP [**82, 134**] force fields. For example, one study found convergence to the global minimum energy conformation in 11 out of 40 simulated annealing runs [**134**]. In other studies, the ability of simulated annealing to locate global minimum energy conformations has been questioned. A search for the global minimum of met-enkephalin using simulated annealing could not locate the global minimum energy conformation [**129**]. Although this disagreement may, in part, be attributed to differences in modeling (e.g., keeping certain variables fixed), it also demonstrates the stochastic nature of this technique. Nevertheless, the general simulated annealing approach, due to its simplistic form, has been applied to a large variety of peptide examples [**125, 128, 133, 135, 165**].

In order to overcome some of the inherent limitations of the basic simulated annealing technique, a combined Monte Carlo and energy minimization approach has also been proposed [**104, 105**]. This Monte Carlo Minimization (MCM) technique employs a local minimization of each conformation before application of the Metropolis condition shown in Equation 3.1. A modified version of this technique, known as the electrostatically driven Monte Carlo method [**149, 150**], uses electrostatic interaction energies to direct the random changes in variable values before minimizing and applying the Metropolis condition. Similarly, the self-consistent electrostatic field method [**143**] is used within a MCM framework to enforce favorable electrostatic interactions. Another MCM technique has been used to locate low energy conformers of met-enkephalin [**180**]. Although the MCM modifications (over MC methods) tend to increase the probability of convergence to the global minimum energy conformation, the combined costs of local minimizations and stochastic searching lead to large computational times.

The concept of simulated annealing has also been applied to the area of molecular dynamics. Here the temperature is related to the kinetic energy, which can be obtained from the velocities of the atoms [**19, 103**]. Therefore, molecular dynamics annealing is accomplished by scaling atomic velocities. At high temperatures the conformational space is totally accessible due to large atomic velocities. As temperature is decreased, local conformations are defined because atomic velocities have been restricted. This technique has been most successful in refining peptide structures using restraints derived from crystallography and NMR studies [**21, 189**].

In general, the confidence of locating the global minimum energy conformation via simulated annealing can only be evaluated by performing multiple runs with varying annealing schedules. More recent Monte Carlo based simulation techniques attempt to eliminate these uncertainties by developing generalized rather than canonical ensemble of conformations [**61**]. In the regular canonical ensemble, the probability distribution can be described by Equation 3.3.

$$PD_B(T,E) \propto n(E)p_B(T,E) \tag{3.3}$$

In Equation 3.3 the probability distribution is proportional to the number of states at a given energy, $n(E)$, multiplied by the Boltzmann weighting factor, as given in Equation 3.2. For a given temperature, PD_B resembles the familiar Boltzmann bell-like shape distribution in energy. In order to generate truly random energy walks, and therefore overcome all energy barrier, the generalized ensemble algorithms employ different weighting factors. For example, in the case of the multicanonical algorithm (i.e., multicanonical ensemble) [**15, 16**], the probability distribution is defined to be constant in energy, as shown in Equation 3.4.

$$PD_M(E) \propto n(E)p_M(E) \; = \; C \tag{3.4}$$

Here C represents a constant, and $p_M(E)$ is the multicanonical weight factor. For Equation 3.4 to hold, the multicanonical weight factor must satisfy the following equation:

$$p_M(E) \propto \frac{1}{n(E)} \tag{3.5}$$

Since such weight factors are not know a priori, the major drawback of this procedure is the development of appropriate weight factors. For example, some information can be obtained by initially performing a number of canonical Monte Carlo simulations. Theoretically, the multicanonical algorithm also provide information which can be used to develop canonical distributions at any temperature; information that can be used in the calculation of thermodynamic quantities. Comparative studies have shown the multicanonical method to be more effective than general simulated annealing algorithms [**62**] and comparable to simulations involving simulated tempering and in the 1/k-ensemble [**63**].

3.4. Smoothing/Deformation Methods. These global optimization methods rely on the assumption that the global minimum of a deformed energy hypersurface can be traced back to the global minimum of the original function [**167**]. The method and extent of smoothing are crucial in defining the deformed surface. However, more important are the procedures for reversing the deformation and interpreting the changes in the deformed landscape in order to locate the true global minimum.

One technique for smoothing the energy surface is conceptually related to the process of diffusion (or conduction) [**144**]. This diffusion equation method (DEM) alters the original energy function $E(x)$ by adding the parameter t, which is physically analogous to time in the diffusion problem. The transformed function, $\hat{E}(x,t)$, can be obtained through solution of the diffusion equation, shown in Equation 3.6.

$$\frac{\partial^2 \hat{E}(x,t)}{\partial x^2} = \frac{\partial \hat{E}(x,t)}{\partial t} \tag{3.6}$$

The boundary condition recovers the original function, $E(x)$, at initial time.

$$\hat{E}(x,0) = E(x) \tag{3.7}$$

One consideration in using the diffusion equation method is the choice of an initial value for the parameter t. Obviously, as t goes to infinity Equation 3.6 becomes

$$\frac{\partial^2 \hat{E}(x,t \to \infty)}{\partial x^2} = 0 \tag{3.8}$$

Solution of Equation 3.8 yields a unimodal surface. For sufficiently large values of t the surface should provide one minimum, although the number of steps required for reversing the deformation may become exceedingly large. Once a starting value for t has been selected and an initial minimum has been located, t is gradually decreased and a sequence of local minima can be traced back to a local minimum of the original function at $t = 0$. The choice of the initial minimum and the decrease of t determine whether this final minimum corresponds to the actual global minimum.

The original approach has been extended up to oligopeptide systems [**91**]. In order to decrease the computational cost of this method, the energy function is based on a Gaussian adaptation of the ECEPP force field [**130, 131**]. The diffusion equation can then be obtained by evaluating the Fourier-Poisson integral in the Cartesian coordinate space. These solutions are used as starting points for minimizations in the dihedral angle space of the original force field. The required approximations limit the applicability of this method, although some modifications have been recently proposed to reduce these limitations [**92**].

A number of other spatial smoothing techniques, known as effective energy methods, are jointly motivated by the concepts of simulated annealing and smoothing. In particular, the packet annealing method attempts to identify appropriate spatial scales as the system is annealed, where these spatial scales are complicated functions of temperature and the energy landscape. This is accomplished by representing the energy surface by characteristic (Gaussian) packets, which physically represent metastable macrostates of the system [**160**]. Each packet, α, is represented by a center of mass x_α, spatial scale Λ_α and a probability of occupation p_α. An effective energy equation governs the behavior of these packets, and is used to derive packet annealing equations which can be iteratively solved to identify the location and formation of new packets [**160**]. The procedure is initialized at high temperature, at which spatial scales are large and the energy landscape is defined by one or a small number of packets. As the temperature is decreased, the packet annealing equations are solved and the packets with highest probability are traced towards low energy minima. As with the DEM, successful implementation of this method requires approximations for the force field terms in order to perform the Gaussian integrations. In addition, effective algorithms for identifying bifurcation points in large scale problems must still be developed. The packet annealing method has been tested on met-enkephalin after employing a number of simplifications [**161**]. The effective energy technique has been used in conjunction with simulated annealing [**28**], and general Gaussian integral transforms have been applied to molecular conformation and distance geometry problems [**127**].

A different approach for deforming the energy surface is to increase the dimensionality of the original problem. Initial applications of these energy embedding techniques were based on distance geometry approaches. Typically, the number of dimensions is initially expanded to N-1 dimension, where N is the total number of atoms. One method reduces the dimensionality by progressively projecting out components based on eigenvalue analysis, and subsequently minimizing in the reduced space [**32, 33**]. Another technique uses a constrained problem formulation and a penalty term to eliminate vectors of coordinates [**34**]. A second penalty function method augments the objective function with terms consisting of Cayley-Menger determinants [**146, 147**]. The increased dimensionality methods are generally most effective in maintaining only short distances. A modified rotational energy embedding technique attempts to overcome these limitations (i.e., high energy barriers) by performing internal rotations when reducing dimensionality [**35**].

A number of recent smoothing techniques avoid the use of costly integral transforms in deforming the energy hypersurface. One method modifies the energy terms by adding two new parameters, γ and P, to the energy equation. A new smoothing function is constructed by replacing r^6 from the Lennard-Jones type potential of Equation 2.6 with the term $(r^P + \gamma)/(1 + \gamma)$. The smoothness of the modified function can be be controlled by varying the two parameters, P and γ. As with other smoothing methods, the global search involves tracking local minima from highly deformed states to the original energy hypersurface. This approach has only been tested on the molecular cluster problem [**158**]. A qualitatively similar technique has also been used in molecular dynamics simulations [**145**]. An alternative method smoothes the energy hypersurface using local minima to construct separable quadratic underestimating functions. The local minima of these fitted functions are then used to initiate searches on the original energy surface [**42, 142**].

3.5. Branch and Bound Methods. Typically, branch and bound methods have been associated with, and are efficient methods for, solving combinatorial global optimization problems [**47, 49**]. The protein folding problem possesses inherently combinatorial aspects, and some explicit combinatorial approaches have already been presented. In general, the use of these probabilistic and heuristic methods in a branch and bound framework should prove to increase the efficiency in locating global minimum energy conformations.

A deterministic branch and bound method, αBB, has been developed and applied to general global optimization problems involving twice-differentiable functions [**2, 3, 4, 5, 6, 9**]. The application of the αBB algorithm to the global minimization of energy functions was first introduced for microclusters [**108, 109**], and small acyclic molecules [**110, 111**]. The approach has also been extended to general constrained optimization problems [**3, 4, 5, 6, 9**].

The αBB global optimization algorithm effectively brackets the global minimum energy conformation by developing converging lower and upper bounds. These bounds are refined by iteratively partitioning the initial domain. Upper bounds on the global minimum are obtained by function evaluations or local minimizations of the original energy function, E. Lower bounds belong to the set of solutions of the convex lower bounding functions, which are constructed by augmenting E with the addition of separable quadratic terms. For a protein folding problem defined in dihedral angle space, the lower bounding function, L, is described by Equation 3.9.

$$L = E + \sum_{i=1}^{N} \alpha_{\phi,i} \left(\phi_i^L - \phi_i\right) \left(\phi_i^U - \phi_i\right) \tag{3.9}$$

Here N is equal to the total number of dihedral angles in the protein. ϕ_i^L and ϕ_i^U represent the lower and upper bounds on the dihedral angle ϕ_i in the current domain. The $\alpha_{\phi,i}$ represent nonnegative parameters which control the shape of the underestimating function, L. To maintain convexity of this function, the $\alpha_{\phi,i}$ must be greater or equal to the negative one-half of the minimum eigenvalue of the Hessian of E over the defined domain. A number of methods have been developed to determine these parameters [**2, 3, 5, 6**]. The overall effect of these terms is to overpower the nonconvexities of the original terms by adding the value of 2α to the eigenvalues of the Hessian of E.

The convex lower bounding functions, L, possess a number of important properties which define the deterministic nature of the algorithm and guarantee convergence to the global minimum energy conformation:

(i) L is a valid underestimator of E;
(ii) L matches E at all corner points of the box constraints;
(iii) L is convex in the current box constraints;

(iv) the maximum separation between L and E is bounded and proportional to α and to square of the diagonal of the current box constraints. This property ensures that feasibility, ϵ_f, and convergence, ϵ_c, tolerances can be reached for a finite size partition element;

(v) the underestimators L constructed over supersets of the current set are always less tight than the underestimator constructed over the current box constraints for every point within the current box constraints.

Once solutions for the upper and lower bounding problems have been established, the next step is to modify these problems for the next iteration. This is accomplished by successively partitioning the initial domain into smaller subdomains. A possible partitioning strategy involves successive subdivision of the original hyper-rectangle by halving on the midpoint of the longest side (bisection). In order to ensure non-decreasing lower bounds, the hyper-rectangle to be bisected is chosen by selecting the region which contains the infimum of the minima of lower bounds. A non-increasing sequence for the upper bound is found by solving the nonconvex problem, E, locally (or by function evaluation) and selecting it to be the minimum over all the previously recorded upper bounds. Obviously, if the single minimum of L for any hyper-rectangle is greater than the current upper bound, this hyper-rectangle can be discarded because the global minimum cannot be within this subdomain (fathoming step). In the worst case the full branch and bound tree would need to be expanded and searched. However, in the protein folding problem and in many problems (in general), the ability to fathom significant portions of the domain space greatly accelerates convergence to the global minimum. A one-dimensional example of this algorithm is illustrated in Figure 2.

In recent work, the αBB algorithm has been shown to be successful for isolated peptide systems using the ECEPP/3 potential energy model [**10, 112**]. These results not only illustrate the efficiency of the algorithm to locate global minimum energy conformations, but they also demonstrate its ability to elucidate low energy regions and conformations. A novel approach was recently proposed for oligopeptides based on a domain partitioning scheme that relies on probability data derived from the Brookhaven PDB [**10**].

3.6. Computational Results. The computational requirements of a sampling of global optimization techniques were compared by considering the results for the oligopeptide met–enkephalin, a 5 residue benchmark molecular conformation problem. In the most general case, this problem possesses 24 independent internal coordinate variables. As previously stated, it is believed that the potential energy hypersurface for this protein contains on the order of 10^{11} local minima. Most of the methods reviewed were able to find the global minimum conformation, although issues of consistency are more difficult to assess. Some methods, most notably those grouped under build–up techniques, were less successful in locating the global minimum conformation for met–enkephalin. However, for most of these techniques, qualitative agreement (i.e., backbone structure) was observed.

Table 1 gives computational results for the met–enkephalin protein folding problem. A number of studies reduced the complexity by fixing the ω variables, which results in a 19 variable problem. The best CPU times are on the order of 1 CPU hour. It should also be pointed out that unlike the genetic algorithm, simulated annealing and Monte Carlo approaches, the αBB approach provides information on upper and lower bounds for the global minimum energy. An additional

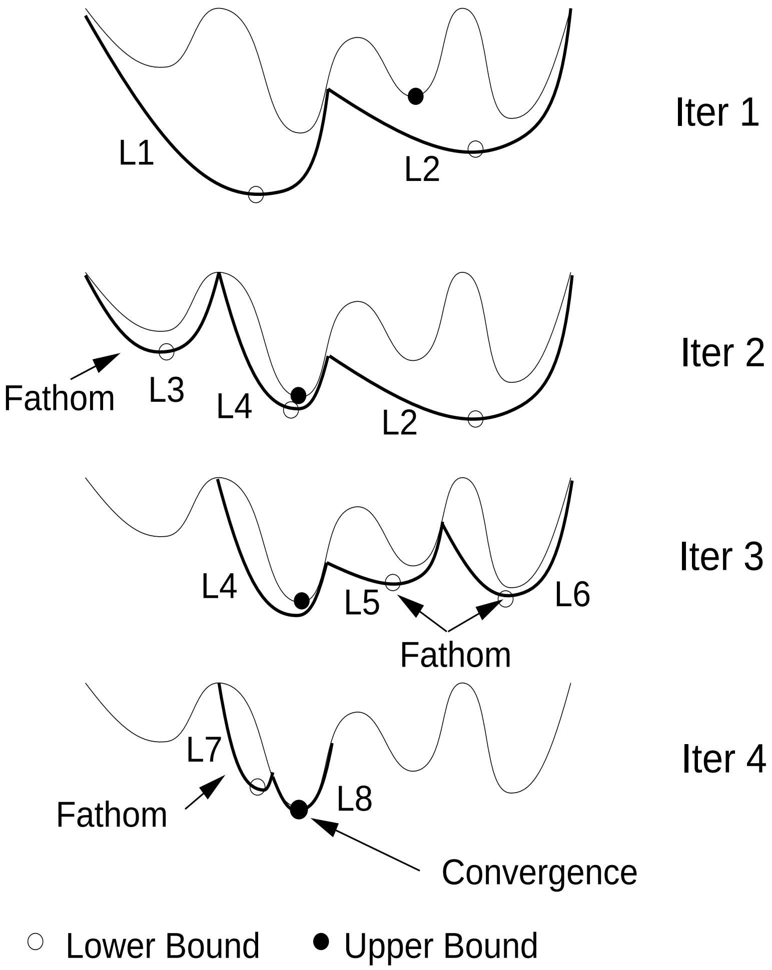

FIGURE 2. One-dimensional illustrative example of the αBB approach. In iteration 1 the overall domain is bisected, the two convex lower bounding functions are created and their unique minima ($L1$ and $L2$) are identified. An upper bound is also identified. Since $L1$ is less than $L2$, the region containing $L1$ is further bisected in iteration 2, while the other region is stored. The minimum of one region ($L3$) is greater than the new upper bound, so this region can be fathomed. The other region is stored. In iteration 3 the region with the next lowest lower bound ($L2$) is bisected and since both new lower bound minima ($L5$ and $L6$) are greater than the current best upper bound, the entire region is fathomed. Finally, by iteration 4, the region containing $L4$ is bisected which results in a region that can be fathomed (containing $L7$) and a convex region whose minimum ($L8$) equals the current upper bound and is the global minimum.

TABLE 1. Computational results for met–enkephalin protein folding problem. The third column indicates the number of free variables (internal coordinates).

Method	Ref	N_{var}	CPU	Platform
Genetic Algorithm	[**100**]	24	2 hrs	SGI IRIS 4D/220
Simulated Annealing	[**125**]	24	2.5 hrs	Apollo DN1000
Simulated Annealing	[**126**]	24	1.5 hrs	Apollo DN1000
Monte Carlo Minimization	[**104**]	19	2-3 hrs	IBM 3090
		24	10 hrs	IBM 3090
Simulated Annealing / Monte Carlo Minimization	[**180**]	24	2 hrs	CRAY X–MP
Monte Carlo Minimization / Simulated Annealing	[**129**]	24	1.5-4 hrs	IBM 3090
Electrostatically Driven Monte Carlo	[**151**]	19	2-3 hrs	IBM 3090
		24	10 hrs	IBM 3090
Multicanonical Algorithm	[**61**]	19	6 hrs	IBM RS600
Diffusion Equation	[**91**]	19	20 min	IBM 3090
αBB	[**10**]	24	1.3 hrs	HP–730

difference is the fact that the αBB algorithm is a domain based, rather than a point based, method. In some cases, these approaches have also been applied to small peptide problems. A general size limit for tractable problems is on the order of 20 to 40 residues, although this value is highly dependent on the form of the energy function and the number of free variables. Obviously, at these limits, the ability to prove global optimality and consistency is debatable.

3.7. Considering Solvation and Free Energy. The problem of considering solvation effects in global conformational energy searches has been made tractable by the development of implicit solvation models and algorithms to efficiently calculate solvent accessible areas and volumes. However, results for such formulations are essentially nonexistent, and those that have appeared are for limited searches only. One recent method utilizes extensive Monte Carlo simulations along with NMR data to predict the hydration structure of leu-enkephalin [**114**]. In this study, hydration energies are modeled using a solvent accessible solvation energy term. Recently, both solvent accessible area and volume effects have been considered in the context of a global optimization search based on the αBB algorithm. Results for all terminally blocked residues and a number of oligopeptide examples are analyzed and compared [**87, 88**].

Similarly, results from global optimization searches including entropic effects have also been lacking. This is understandable because free energy calculations not only require the location of the global minimum potential energy conformation, but also a complete list of low energy conformers. One method, based on the Monte Carlo minimization technique, proposes a free energy directed search to locate low free energy conformers [**178**]. Relative free energy calculations are accomplished using the harmonic approximation outlined above. A comparison between the proposed technique and regular Monte Carlo minimization has also appeared [**115**]. A recent approach, based on a modified version of the αBB algorithm, emphasizes this approach's ability to identify low energy regions. A complete free energy analysis using this method is tested on a number of oligopeptide examples [**89**].

3.8. Peptide Docking Methods. Once a method for "scoring" the binding affinity has been selected, the exact form of the optimization problem must be developed. Several general approaches have been employed. The most obvious and most difficult approach would be to optimize the entire system of the two interacting peptides. To accomplish this, the relative position of the two peptides, which is defined by six degrees of freedom (three translation and three rotation), along with the total number of internal degrees of freedom for the two molecules must be considered. This problem becomes intractable for all but the smallest systems. Alternative approaches have decomposed the problem by considering the binding affinities of shorter subsequences at different binding sites of the macromolecule. The full binding ligand can then be constructed based on the optimally docked subsequences. This approach relies on the ability to build a suitable ligand. An alternative method is based on independently generating conformations of the isolated ligand. Binding affinities for a number of these rigid conformations can then be calculated and compared. Here the drawback is that intermediates with higher binding affinities may be overlooked.

The following discussion classifies the peptide docking approaches according to their treatment of the internal flexibility of the ligand molecule. However, it should be noted that some approaches combine aspects of both rigid and flexible methods. In addition, the choice of scoring function is often closely related to these classifications. For example, it is implicitly difficult for shape based approaches to capture internal flexibility due to their simplified description of the molecular surface. Nevertheless, for almost rigid docking problems, these methods can be used to systematically search a large number of potential ligands. In contrast, detailed energy based approaches better represent the free energy of the system and can deal with internal conformational changes, although their increased dimensionality make these methods more computationally expensive. Therefore, there is a need for rigorous global optimization approaches to address the peptide docking problem.

3.8.1. *Rigid Models.* The first, and most common, methods used to address the peptide docking problems were based on the concept of shape complementarity. These methods employ, at least initially, rigid approximations for both the ligand and receptor molecules. In the most general case, six degrees of freedom, three translational and three rotational, must be optimized to determine the best "fit" for the receptor-ligand complex. In practice, approximations are used to effectively reduce the number of degrees of freedom. In addition, the alignment of each ligand must be optimized within the binding site. Typically, several screening stages are used to reduce these optimizations to a manageable number.

One shape based method utilizes a simplified protein model, which is generated by representing each amino acid by a single sphere. The scoring function is based on interfacial areas and a simplified nonbonded potential energy term. Potential ligand structures are screened by systematically rotating the ligand and then translating the structure, along only one dimension, into the pocket [**72, 193, 194**]. These approximations and simplifications are necessary in order to make the problem tractable, especially in the context of a systematic search. A recent modification attempts to overcome these computational limitations by using a simulated annealing, rather than a systematic, search to screen the ligand structures [**26**].

Distinctive characteristics of molecular surfaces have also been used to reduce the number of degrees of freedom for shape based docking problems. One study considers local shape functions, which are generated by placing spheres at surface points along the ligand and receptor surfaces. The volume within the surface and the unit vector that extends from the center of the sphere to the surface characterize these functions. A combinatorial algorithm can then be used to compare these local shape functions at "knobs and holes" [**98**] on the ligand and receptor surfaces so that the best alignments of the two molecules can be identified [**31**].

More detailed descriptions of molecular surfaces have also been used in determining shape complementarity. One procedure creates a webbed surface for the ligand and receptor by using a local coordinate system to define the surface points for each molecule. After setting the ligand position, a least squares method is used to align the surface points of the two molecules. The method also screens ligands according to a Coulombic scoring function [**14**].

An alternative approach transforms the problem from identifying complementary shapes for the receptor and ligand proteins into one of matching similar shapes for these two molecules. This is accomplished by characterizing the binding site by a collection of spheres that lie on the outside of the receptor surface, and the ligand by a collection of spheres that lie on the inside of the ligand surface [**39, 40, 93**]. Potential matches are identified by grouping and comparing distances between the center of spheres for each molecule. Local refinement of translation and rotation vectors is used for the highest ranking matches. The complexity of the problem is, to some degree, obscured because it also depends on the choice of location, size and number of spheres used to model the receptor molecule. Other modifications of this procedure include the addition of hydrogen bonding criteria, and the use of local minimization of the potential energy in order to relax the rigidity of the ligand molecule [**96, 159**].

The "soft docking" model represents the two molecules as a collection of cubes rather than spheres. This method combines aspects of surface complementarity, grid search and soft potential modeling. The "cubic" representation along with a grid search makes the translational and rotational searches much more efficient. In addition, the cubes implicitly allow for some volume overlap, which can be used in combination with surface complementarity to screen docked complexes [**70**].

In general, when considering a rigid receptor, the concept of a grid search can be used to reduce the computational requirements of evaluating scoring functions. This is accomplished by precomputing values for the receptor based on points of a three dimensional grid [**57**]. The concept is similar to cubic lattice model approaches in molecular conformation problems, for which a recently proposed algorithm using a tabu search has been highly effective [**140**]. This approach has been the basis of a number of recent studies [**116, 185**], including one that employs a Monte Carlo search in the context of "knobs and holes" docking [**185**].

3.8.2. *Flexible Models.* In the most general case, flexible docking approaches attempt to optimize the free energy of the entire macromolecule-ligand complex, which is described by translational, rotational and internal variables of the system. In contrast to most rigid modeling approaches, these methods typically do not require prior knowledge of ligand conformations. As a result, their success in predicting ligand binding is highly dependent on the use of detailed scoring functions to evaluate free energy changes. In addition, although some studies have considered full macromolecular-ligand systems, most approaches also depend on effective decomposition strategies of the overall docking problem.

It should also be noted that some simple approaches have been implemented in an attempt to model flexible docking. For example, a number of methods have incorporated ligand flexibility by considering databases of multiple ligand conformations [**83, 118**]. However, these methods require reliable databases and methods for developing appropriate ligand conformations, and these are typically not available. On the other hand, thermodynamic integration and free energy perturbation methods allow for full flexibility and detailed modeling of binding free energies. However, these simulations, usually accomplished by molecular dynamics, effectively explore only single low energy minima. This has led to the need for global optimization methods that efficiently search the conformational energy hypersurface associated with peptide docking problems.

One of the most common, approaches has been based on Monte Carlo simulating annealing algorithms. This method was first applied to flexible ligand docking using molecular affinity potentials [**58**]. Molecular affinity potentials increase the computational efficiency of the search by employing precomputed energy grids [**57**]. In this case, flexibility was introduced by allowing internal rotations of torsion angles, along with translational and rotational movement. However, for each docking example, a set of simulated annealing runs were necessary in order to increase the confidence of the reported structures.

A second method, also based on simulated annealing, involves a two-step procedure to dock flexible oligopeptide ligands [**22**]. In the first step, a modified potential energy force field was used to reduce unfavorable intermolecular contacts. This energy model was employed in local energy minimizations of arbitrarily docked ligands, which were needed in order to generate an initial set of ligand conformations. The scoring function for the second step described energy interactions between both the flexible ligand and rigid receptor molecules. The set of minimized conformations was then used to generate starting points for a Monte Carlo minimization procedure. Although experimental results were not initially available, later comparison has shown that this method did not correctly predict MHC binding. These discrepancies are most likely attributable to incorrect energy modeling (e.g., no inclusion of solvation), along with the inherent inefficiencies associated with simulated annealing searches.

Another MC based method employs a multiple-start technique in an attempt to reproduce the results of a systematic search. The first step involves a Monte Carlo search with a grid based scoring function in order to limit steric overlaps of the ligand and receptor molecules. A second, energy directed, simulated annealing search uses a pairwise potential energy function. Rather than rely on a single search, this method employs a large number of short simulated annealing runs. Although initial results were based on both rigid receptor and ligand conformations [**64**], more recent work has addressed the issue of flexible ligand docking [**65**].

In a similar way, genetic algorithms have recently been used to dock flexible ligands. In some cases, scoring functions have been based on potential energy force fields [**137**], although some modified potentials have also been used [**73**]. The results of one method [**76**], which includes solvation effects, have emphasized the need for developing reliable scoring functions. In general, as with simulated annealing, the ability to model flexibility is limited as ligand size increases. The coupling of these effects with the implicit unreliability of both the genetic algorithm and simulated annealing search techniques must be closely considered when approaching large scale docking problems such as de novo drug design.

Combinatorial methods have also been used to address the difficulties of modeling full ligand flexibility. In theory, these methods are similar to the build-up methods outlined for the protein folding problem, although in peptide docking intermolecular interactions are also included. An initial application to the peptide docking problem was based on rigid ligand models generated from a database of chemical structures [**40**]. A more detailed implementation uses libraries of low energy conformations for single amino acid residues. These conformations are subsequently joined and grouped according to scoring functions based on the intra- and inter-molecular energies of the macromolecular-ligand complex [**124**]. More recent methods have employed databases developed for smaller ligand fragments such as functional groups [**153**] or even atoms [**154**]. In general, these ligand build-ups are initialized by selecting a starting point within the macromolecular pocket. As with the protein folding approaches, such combinatorial techniques must employ effective reduction schemes in order to limit the number of generated conformations.

Similar approaches combine the ideas of fragment assembly and site mapping. In contrast to the single anchor requirement of simple build-up methods, these techniques attempt to identify a number of anchor fragments or residues which can be joined through a process of fragment assembly. The first step, site mapping, is equivalent to docking probe fragments at specific sites of the macromolecule. Some methods have screened the binding affinities of these probes using shape based modeling [**94**], while others have relied on other energy based descriptions, such as hydrogen bonding interactions [**17, 18**]. In general these site maps are constructed by local minimization, grid or library searches of the probe conformations. Other techniques employ a multiple copy simultaneous search [**119, 152**]. Once anchor positions have been determined using one of these methods, the resulting segments must be joined by fragment assembly. Bridges can be formed by searching through molecular libraries, or in some cases using an exhaustive search over all connections [**23**]. A recently proposed technique applies a dynamic programming approach, as discussed above, to the fragment assembly phase of a nonameric ligand in an MHC HLA-A2 complex [**60**].

Recently, a novel decomposition based approach has been proposed for predicting peptide docking to the MHC HLA-DR1 protein [**11**]. The general approach is along the lines of performing site mappings for the five polymorphic pockets of HLA-DR1 that accommodate peptide docking [**166**]. Specifically, each naturally occurring amino acid is treated as a probe molecule for each of the five pockets. A deterministic global optimization search technique is then used to identify the best binding conformation for each residue. The scoring function accounts for both intra- and inter-molecular interactions using a detailed potential energy force field along with a surface accessible solvation model. The global optimization search, based on the αBB algorithm, is used to identify the global minimum energy conformation for both the bound and free residues. The corresponding energy differences are then used to provide rank ordered lists of the best binders for each pocket. Results for pocket 1 of the HLA-DRB1 macromolecule have exhibited good agreement with experimental binding assays [**11**]. Current research has focused on the prediction of the full binding ligand.

4. Conclusions

The importance of the protein folding and peptide docking problems is evidenced by the large amount of experimental and theoretical research conducted in these areas. Although experimental studies of these systems will always be needed, the ability to theoretically predict folded proteins and macromolecule-ligand complexes would greatly aid in advancing the applications of such knowledge. Both molecular modeling and global optimization are important factors in the overall equation that will eventually provide a solution to these problems.

In this review, issues related to the modeling of protein folding and peptide docking systems have been analyzed and discussed. In addition, a critical assessment of a wide variety of global optimization approaches has been provided. These observations have highlighted the extreme difficulty of these problems and the crucial interdependence of modeling and global optimization approaches.

References

[1] R. A. Abagyan and M. M. Totrov, Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins, J. Mol. Biol., **235**, (1994), 983-1002.
[2] C. S. Adjiman, I. P. Androulakis and C. A. Floudas, Global optimization of MINLP problems in process synthesis and design, Comput. Chem. Eng. **21**, (1997), S445-S450.
[3] C. S. Adjiman, S. Dallwig, C. A. Floudas and A. Neumaier, A global optimization method, αBB, for general twice-differentiable NLPs - I. Theoretical advances, Comput. Chem. Eng., (in press), (1998).
[4] C. S. Adjiman, I. P. Androulakis and C. A. Floudas, A global optimization method, αBB, for general twice-differentiable NLPs - II. Implementation and computational results, Comput. Chem. Eng., (in press), (1998).
[5] C. S. Adjiman, I. P. Androulakis, C. D. Maranas and C. A. Floudas, A global optimization method, αBB, for process design, Comput. Chem. Eng., **20**, (1996), S419-S424.
[6] C. S. Adjiman and C. A. Floudas, Rigorous convex underestimators for general twice-differentiable problems, J. Global Optim., **9**, (1996), 23-40.
[7] N. L. Allinger, Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V_1 and V_2 torsional terms, J. Am. Chem. Soc., **99**, (1977), 8127-8134.
[8] N. L. Allinger, Y. H. Yuh and J-H. Lii, Molecular Mechanics. The MM3 force field for hydrocarbons, J. Am. Chem. Soc., **111**, (1989), 8551-8566.
[9] I. P. Androulakis, C. D. Maranas and C. A. Floudas, αBB: A global optimization method for general constrained nonconvex problems, J. Global Optim., **7**, (1995), 337-363.
[10] I. P. Androulakis, C. D. Maranas and C. A. Floudas, Global minimum potential energy conformations of oligopeptides, J. Global Optim., **11**, (1997), 1-34.
[11] I. P. Androulakis, N. N. Nayak, M. G. Ierapetritou, D. S. Monos and C. A. Floudas, A predictive method for the evaluation of peptide binding in pocket 1 of HLA-DRB1 via global minimization of energy interactions, Proteins, **29**, (1997), 87-102.
[12] C. B. Anfinsen, E. Haber, M. Sela and F. H. White, Jr., The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain, Proc. Natl. Acad. Sci. USA, **47** (1961), 1309-1314.
[13] J. D. Augspurger and H. A. Scheraga, An efficient, differentiable hydration potential for peptides and proteins, J. Comput. Chem., **17**, (1996), 1549-1558.
[14] D. J. Bacon and J. Moult, Docking by least-squares fitting of molecular surface patterns, J. Mol. Biol., **225**, (1992), 849-858.
[15] B. A. Berg and T. Neuhaus, Multicanonical algorithms for $1st$ order phase-transitions, Phys. Lett., **B267**, (1991), 249-253.
[16] B. A. Berg and T. Neuhaus, Multicanonical ensemble - a new approach to simulate 1^{st} order phase transitions, Phys. Rev. Lett., **68**, (1992), 9-12.
[17] H. J. Böhm, The computer program Ludi: A new method for the de novo design of enzyme inhibitors, J. Comput. Aided Mol. Design, **6**, (1992), 61-78.
[18] H. J. Böhm, Ludi: Rule-based automatic design of new substituents for enzyme inhibitor leads, J. Comput. Aided Mol. Design, **6**, (1992), 593-606.

[19] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, CHARMM : A program for macromolecular energy minimization and dynamics calculations, J. Comp. Chem., **4**, (1983), 187-217.

[20] C. L. Brooks III and D. A. Case, Simulations of peptide conformational dynamics and thermodynamics, Chem. Rev., **93**, (1993), 2487-2502.

[21] A. T. Brünger, Crystallographic refinement by simulated annealing. Applications to a 2.8 $\AA$ resolution structure of aspartate aminotransferase, J. Mol. Biol., **203**, (1988), 803-816.

[22] A. Calfisch, P. Niederer and M. Anliker, Monte Carlo docking of oligopeptides to proteins, Proteins, **13**, (1992), 223-230.

[23] A. Calfisch, A. Miranker and M. Karplus, Multiple copy simultaneous search and construction of ligands in binding sites: application of inhibitors of HIV-1 aspartic proteinase, J. Med. Chem., **36**, (1993), 2142-2164.

[24] D. A. Case, Normal mode analysis of protein dynamics, Curr. Opinion Struct. Biol. **4**, (1994), 285-290.

[25] D. Chandler and H. C. Andersen, Optimized cluster expansions for classical fluids. I. Theory of molecular liquids, J. Chem. Phys., **57**, (1972), 1930-1937.

[26] J. Cherfils, S. Duquerroy and J. Janin, Protein-protein recognition analyzed by docking simulation, Proteins, **11**, (1991), 271-280.

[27] B. W. Church, M. Orešič and D. Shalloway, Tracking metastable states to free-energy global minima, In DIMACS Series in Discrete Mathematics and Theoretical Computer science, **23**, Amer. Math. Soc. (1996), 41-64.

[28] T. Coleman, D. Shalloway and Z. Wu, A parallel build-up algorithm for global energy minimization of molecular clusters using effective energy simulated annealing, J. Global Optim., **4**, (1994), 171-185.

[29] M. L. Connolly, Analytical molecular surface calculation, J. Appl. Cryst., **16**, (1983), 548-558.

[30] M. L. Connolly, Solvent accessible surfaces of proteins and nucleic acids, Science, **221**, (1983), 709-713.

[31] M. L. Connolly, Shape complementarity at the hemoglobin $\alpha_1\beta_1$ subunit interface, Biopolymers, **25**, (1986), 1229-1247.

[32] G. M. Crippen, Energy embedding of trypsin inhibitor, Biopolymers, **21**, (1982), 1933-1943.

[33] G. M. Crippen, Why energy embedding works, J. Phys. Chem., **91**, (1987), 6341-6343.

[34] G. M. Crippen and P. K. Ponnuswamy, Determination of an empirical energy function for protein conformational analysis by energy embedding, J. Comput. Chem., **8**, (1987), 972-981.

[35] G. M. Crippen and T. F. Havel, Global energy minimization by rotational energy embedding, J. Chem. Inf. Comput. Sci., **30**, (1990), 222-227.

[36] V. Daggett and M. Levitt, Realistic simulations of native-protein dynamics in solution and beyond, Annu. Rev. Biophys. Biomol. Struct., **22**, (1993), 353-380.

[37] P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest and A. T. Hagler, Structure and energetics of ligand-binding to proteins - Escherichia-coli dihydrofolate reductase trimethoprim, a drug-receptor system, Proteins, **4**, (1988), 31-47.

[38] M. E. Davis and J. A. McCammon, Electrostatics in biomolecular structure and dynamics, Chem. Rev., **90**, (1990), 509-521.

[39] R. L. DesJarlais, G. L. Seibel, I. D. Kuntz, P. S. Furth, J. C. Alvarez, P. R. Ortiz de Montellano, D. L. Decamp, L. M. Babe and C. S. Craik, Structure Based Design of Nonpeptide Inhibitors Specific for The Human Immunodeficiency Virus-1 Protease, Proc. Natl. Acad. Sci. USA, **87**, (1990), 6644-6648.

[40] R. L DesJarlais, R. P. Sheridan, G. L. Seibel, J. S. Dixon, I. D. Kuntz and R. Venkataraghavan, Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure, J. Med. Chem., **31**, (1988), 722-729.

[41] E. Di Cera, Thermodynamic Theory of Site-Specific Binding Processes in Biological Macromolecules, Cambridge Univ. Press, (1995).

[42] K. A. Dill, A. T. Phillips and J. B. Rosen, CGU : An algorithm for molecular structure prediction, In IMA Volumes in Mathematics and its Applications, **94**, Springer-Verlag, (1997), 1-21.

[43] D. Eisenberg and A. D. McLachlan, Solvation energy in protein folding and binding, Nature, **319**, (1986), 199-203.

[44] F. Eisenhaber and P. Argos, Improved strategy in analytic surface calculation for molecular systems : Handling of singularities and computational efficiency, J. Comp. Chem., **14**, (1993), 1272-1280.

[45] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander and M. Scharf, The double cubic lattice method : Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies, J. Comp. Chem., **16**, (1995), 273-284.

[46] P. J. Flory, Foundations of rotational isomeric state theory and general methods for generating configurational averages, Macromolecules, **7**, (1974), 381-392.

[47] C. A. Floudas, Deterministic global optimization in design, control, and computational chemistry, In IMA Volumes in Mathematics and its Applications, **93**, Springer-Verlag, (1997), 129-184.

[48] C. A. Floudas and P. M. Pardalos (editors), State of the Art in Global Optimization : Computational Methods and Applications, Kluwer Academic Publishers, (1996).

[49] C. A. Floudas, Nonlinear and Mixed-Integer Optimization : Fundamentals and Applications, Oxford University Press, (1995).

[50] S. Forrest, Genetic algorithms : Principles of natural selection applied to computation, Science, **261**, (1993), 872-878.

[51] D. H. Fremont, M. Matsumura, E. A. Stura, P. A. Peterson and I. A. Wilson, Crystal structures of two viral peptides in complex with murine MHC class I H-2 K^b, Science, **257**, (1992), 919-927.

[52] K. D. Gibson and H. A. Scheraga, Revised algorithms for the build-up procedure for predicting protein conformations by energy minimization, J. Comput. Chem., **8**, (1987), 826-834.

[53] M. K. Gilson, K. A. Sharp and B. H. Honig, Calculating the electrostatic potential of molecules in solution : Method and error assessment, J. Comput. Chem., **9**, (1988), 327-335.

[54] N. Gō and H. A. Scheraga, Analysis of the contribution of internal vibrations to the statistical weight of equilibrium conformations of macromolecules, J. Chem. Phys., **51**, (1969), 4751-4767.

[55] N. Gō and H. A. Scheraga, On the use of classical statistical mechanics in the treatment of polymer chain conformation, Macromolecules, **9**, (1976), 535-542.

[56] D. Goldberg, Genetic algorithms in search, optimization and machine learning, Addison-Wesley, (1989).

[57] P. J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important molecules, J. Med. Chem., **28**, (1985), 849-857.

[58] D. S. Goodsell and A. J. Olson, Automated docking of substrates to proteins by simulated annealing, Proteins, **8**, (1990), 195-202.

[59] P. D. J. Grootenhuis and P. A. Kollman, Crown ether-neutral molecule interactions studied by molecular mechanics, normal mode analysis, and free energy perturbation calculations. Near quantitative agreement between theory and experimental binding free energies, J. Am. Chem. Soc., **111**, (1989), 4046-4051.

[60] K. Gulukota, S. Vajda and C. DeLisi, Peptide docking using dynamic programming, J. Comput. Chem., **17**, (1996), 418-428.

[61] U. H. E. Hansmann and Y. Okamoto, Prediction of peptide conformation by multicanonical algorithm : New approach to the multiple-minima problem, J. Comput. Chem., **14**, (1993), 1333-1338.

[62] U. H. E. Hansmann and Y. Okamoto, Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem, Physica A, **212**, (1994), 415-437.

[63] U. H. E. Hansmann and Y. Okamoto, Numerical comparisons of three recently proposed algorithms in the protein folding problem, J. Comput. Chem., **18**, (1997), 920-933.

[64] T. N. Hart and R. J. Read, A multiple-start Monte Carlo docking method, Proteins, **13**, (1992), 206-222.

[65] T. N. Hart and R. J. Read, Multiple-start Monte Carlo docking of flexible ligands, In The Protein Folding Problem and Tertiary Structure Prediction, Birkhäuser, (1994), 71-108.

[66] R. D. Head, M. L. Smyte, T. I. Oprea, C. L. Waller, S. M. Green and G. R. Marshall, VALIDATE : a new method for receptor-based prediction of binding affinities of novel ligands, J. Am. Chem. Soc., **118**, (1996), 3959-3969.

[67] J. Holland, Genetic algorithms and the optimal allocation of trials, SIAM J. Computing, **2**, (1973), 88-105.

[68] A. J. Hopfinger, Polymer-solvent interactions for homopolypeptides in aqueous solution, Macromolecules, **4**, 1971, 731-737.

[69] R. Horst and P. M. Pardalos, Handbook of Global Optimization, Kluwer Academic Publishers, (1995).

[70] F. Jiang and S-H. Kim, "Soft docking" : matching of molecular surface cubes, J. Mol. Biol., **219**, (1991), 79-102.

[71] A. N. Jain and M. A. Murcko, Computational methods to predict binding free energy in ligand-receptor complexes, J. Med. Chem., **38**, (1995), 4953-4967.

[72] J. Janin and S. J. Wodak, Reaction pathway for the quartenary structure change in hemoglobin, Biopolymers, **24**, (1985), 509-526.

[73] G. Jones, P. Willett and R. C. Glen, Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation, J. Mol. Biol., **245**, (1995), 43-53.

[74] W. L. Jorgensen and J. Tirado-Rives, The OPLS potential functions for Proteins. Energy minimizations for crystals of cyclic peptides and crambin, J. Am. Chem. Soc., **110**, (1988), 1657-1666.

[75] R. S. Judson, M. E. Colvin, J. C. Meza, A. Huffer and D. Gutierrez, Do intelligent configuration search techniques outperform random search for large molecules ?, Intl. J. Quant. Chem., **44**, (1992), 277-290.

[76] R. S. Judson, Y. T. Tan, E. Mori, C. Melius, E. P. Jaeger, A. M Treasurywala and A. Mathiowetz, Docking flexible molecules: a case study of three proteins, J. Comput. Chem., **16**, (1995), 1405-1419.

[77] Y. K. Kang, G. Némethy and H. A. Scheraga, Free energies of hydration of solute molecules 1. Improvement of hydration shell model by exact computations of overlapping volumes, J. Phys. Chem., **91**, (1987), 4105-4109.

[78] Y. K. Kang, G. Némethy and H. A. Scheraga, Free energies of hydration of solute molecules 2. Application of the hydration shell model to nonionic organic molecules, J. Phys. Chem., **91**, (1987), 4109-4117.

[79] Y. K. Kang, G. Némethy and H. A. Scheraga, Free energies of hydration of solute molecules 3. Application of the hydration shell model to charged organic molecules, J. Phys. Chem., **91**, (1987), 4118-4120.

[80] Y. K. Kang, K. D. Gibson, G. Némethy and H. A. Scheraga, Free energies of hydration of solute molecules 4. Revised treatment of the hydration shell model, J. Phys. Chem., **92**, (1988), 4739-4742.

[81] M. Karplus and J. N. Kushick, Method for estimating the configurational entropy of macromolecules, Macromolecules, **14**, (1981), 325-332.

[82] H. Kawai, T. Kikuchi and Y. Okamoto, A prediction of tertiary structures of peptide by the Monte Carlo simulated annealing method, Protein Eng. **3**, (1989), 85-94.

[83] S. K. Kearsley, D. J. Underwood, R. P. Sheridan and M. D. Miller, Flexibases: a way to enhance the use of molecular docking methods, J. Comput. Aided Mol. Design, **8**, (1994), 565-582.

[84] P. S. Kim and R. L. Baldwin, Intermediates in the folding reactions of small proteins, Annu. Rev. Biochem., **59**, (1990), 631-660.

[85] M. Kinoshita, Y. Okamoto and F. Hirata, Calculation of hydration free energy for a solute with many atomic sites using the RISM theory: A robust and efficient algorithm, J. Comput. Chem., **18**, (1997), 1320-1326.

[86] S. Kirkpatrick, C. D. Gelatt Jr. and M. P. Vecchi, Optimization by simulated annealing, Science, **220**, (1983), 671-680.

[87] J. L. Klepeis, I. P. Androulakis, M. G. Ierapetritou and C. A. Floudas, Predicting solvated peptide conformations via global minimization of energetic atom-to-atom interactions, Comput. Chem. Eng., (in press), (1998).

[88] J. L. Klepeis and C. A. Floudas, A comparative study of global minimum energy conformations of solvated peptides, (submitted).

[89] J. L. Klepeis and C. A. Floudas, Free energy calculations for peptides using a deterministic global optimization algorithm, (submitted).

[90] P. A. Kollman, Free energy calculations : Applications to chemical and biochemical phenomena, Chem. Rev., **93**, 1993, 2395-2417.

[91] J. Kostrowicki and H. A. Scheraga, Application of the diffusion equation method for global optimization to oligopeptides, J. Phys. Chem., **96**, (1992), 7442-7449.

[92] J. Kostrowicki and H. A. Scheraga, Some approaches to the multiple-minima problem in protein folding, In DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **23**, Amer. Math. Soc., (1996), 123-132.
[93] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Landgridge and T. E. Ferrin, A geometric approach to macromolecule-ligand interactions, J. Mol. Biol., **161**, (1982), 269-288.
[94] M. C. Lawrence and P. C. Davis, CLIX: A search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure, Proteins, **12**, (1992), 31-41.
[95] A. R. Leach, A survey of methods for searching the conformational space of small and medium-sized molecules, In Reviews in Computational Chemistry, **II**, VCH Publishers, (1991), 1-55.
[96] A. R. Leach and I. D. Kuntz, Conformational analysis of flexible ligands in macromolecular receptor sites, J. Comput. Chem., **13**, (1992), 733-748.
[97] B. Lee and F. M. Richards, The interpretation of protein structures : estimation of static accessibility, J. Mol. Biol., **55**, (1971), 379-400.
[98] R. H. Lee and G. D. Rose, Molecular recognition. I. Automatic identification of topographic surface features, Biopolymers, **24**, 1985, 1613-1627.
[99] S. M. Le Grand and K. M. Merz, Jr., The genetic algorithm and protein tertiary structure prediction, In The Protein Folding Problem and Tertiary Structure Prediction, Birkhäuser, (1994), 109-124.
[100] S. M. Le Grand and K. M. Merz, Jr., The application of the genetic algorithm to the minimization of potential energy functions, J. Global Optim. **3**, (1993), 49-66.
[101] P. Leopold, M. Montal and J. Onuchic, Protein folding funnels : A kinetic approach to the sequence-structure relationship, Proc. Natl. Acad. Sci. USA **89**, (1992), 8721-8725.
[102] C. Levinthal, Are there pathways to protein folding ?, J. Chem. Phys., **65**, (1968), 44-45.
[103] M. Levitt, Protein folding by restrained energy minimization and molecular dynamics, J. Mol. Biol., **170**, (1983), 723-764.
[104] Z. Li and H. A. Scheraga, Structure and free energy of complex thermodynamic systems, J. Mol. Struct. (Theochem.), **179**, (1988), 333-352.
[105] Z. Li and H. A. Scheraga, Monte Carlo minimization approach to the multiple-minima problem in protein folding, Proc. Natl. Acad. Sci. USA, **84**, (1987), 6611-6615.
[106] J-H. Lii and N. L. Allinger, Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics, J. Am. Chem. Soc., **111**, (1989), 8566-8575.
[107] J-H. Lii and N. L. Allinger, Molecular mechanics. The MM3 force field for hydrocarbons. 3. The van der Waals' potentials and crystal data for aliphatic and aromatic hydrocarbons, J. Am. Chem. Soc., **111**, (1989), 8576-8582.
[108] C. D. Maranas and C. A. Floudas, A global optimization approach for Lennard-Jones microclusters, J. Chem. Phys., **97**, (1992), 7667-7677.
[109] C. D. Maranas and C. A. Floudas, Global optimization for molecular conformation problems, Ann. Oper. Res., **42**, (1993), 85-117.
[110] C. D. Maranas and C. A. Floudas, A deterministic global optimization approach for molecular structure determination, J. Chem. Phys., **100**, (1994), 1247-1261.
[111] C. D. Maranas and C. A. Floudas, Global minimum potential energy conformations of small molecules, J. Global Optim., **4**, (1994), 135-170.
[112] C. D. Maranas, I. P. Androulakis and C. A. Floudas, A deterministic global optimization approach for the protein folding problem, In DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **23**, Amer. Math. Soc., (1996), 133-150.
[113] R. F. McGuire, F. A. Momany and H. A. Scheraga, Energy parameters in polypeptides. V. An empirical hydrogen bond potential function based on molecular orbital calculations, J. Phys. Chem., **76**, (1972), 375-393.
[114] H. Meirovitch and E. Meirovitch, New theoretical methodology for elucidating the solution structure of peptides from NMR data. 3. Solvation effects, J. Phys. Chem., **100**, (1996), 5123-5133.
[115] H. Meirovitch and E. Meirovitch, Efficiency of Monte Carlo minimization procedures and their use in analysis of NMR data obtained from flexible peptides, J. Comput. Chem., **18**, (1997), 240-253.
[116] E. C. Meng, B. K. Shoichet and I. D. Kuntz, Automated docking with grid-based energy evaluation, J. Comput. Chem., **13**, (1992), 505-524.

[117] N. Metropolis, A. W. Rosenbluth and M. N. Rosenbluth and A. H. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys., **21**, (1953), 1087-1092.
[118] M. D. Miller, S. K. Kearsley, D. J. Underwood and R. P. Sheridan, FLOG : A system to select "quasi-flexible" ligand complementary to a receptor of known three-dimensional structure, J. Comput. Aided Mol. Design, **8**, (1994), 153-174.
[119] A. Miranker and M. Karplus, Functionality maps of binding sites: a multicopy simultaneous search method, Proteins, **1**, (1991), 29-34.
[120] S. Miyamoto and P. A. Kollman, What determines the strength of noncovalent association of ligands to proteins in aqueous solutions ?, Proc. Natl. Acad. Sci. USA., **90**, (1993), 8402-8406.
[121] F. A. Momany, L. M. Carruthers, R. F. McGuire and H. A. Scheraga, Intermolecular potentials from crystal data. III. Determination of empirical potentials and application to the packing configurations and lattice energies in crystals of hydrocarbons, carboxylic acids, amines, and amides, J. Phys. Chem., **78**, (1974), 1595-1620.
[122] F. A. Momany, L. M. Carruthers and H. A. Scheraga, Intermolecular potentials from crystal data. IV. Application of empirical potentials to the packing configurations and lattice energies in crystals of amino acids, J. Phys. Chem., **78**, (1974), 1621-1630.
[123] F. A. Momany, R. F. McGuire, A. W. Burgess and H. A. Scheraga, Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids, J. Phys. Chem., **79**, (1975), 2361-2381.
[124] J. B. Moon and J. Howe, Computer design of bioactive molecules: a method for receptor-based de novo ligand design, Proteins, **11**, (1991) 314-328.
[125] L. B. Morales, R. Garduño-Juárez and D. Romero, Applications of simulated annealing to the multiple-minima problem in small peptides, J. Biomol. Struct. Dyn., **8**, (1991), 721-735.
[126] L. B. Morales, R. Garduño-Juárez and D. Romero, The multiple-minima problem in small peptides revisited. The threshold accepting approach, J. Biomol. Struct. Dyn., **9**, (1992), 951-957.
[127] J. Moré and Z. Wu, ϵ-optimal solutions to distance geometry problems via global continuation, In DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **23**, Amer. Math. Soc., (1996), 151-168.
[128] T. Nakazawa, H. Kawai, Y. Okamoto and M. Fukugita, β-sheet folding of fragment (16-36) of bovine pancreatic trypsin inhibitor as predicted by Monte Carlo simulated annealing, Protein Eng., **5**, (1992), 495-503.
[129] A. Nayeem, J. Vila and H. A. Scheraga, A comparative study of the simulated annealing and Monte Carlo with minimization approaches to the the minimum energy structures of polypeptides : [met]-enkephalin, J. Comput. Chem., **12**, (1991), 594-605.
[130] G. Némethy, M. S. Pottle and H. A. Scheraga, Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids, J. Phys. Chem., **87**, (1983), 1883-1887.
[131] G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey and H. A. Scheraga, Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm with application to proline-containing peptides, J. Phys. Chem., **96**, (1992), 6472-6484.
[132] A. Neumaier, Molecular modeling of proteins and mathematical prediction of protein structure, SIAM Rev., **39**, (1997), 407-460.
[133] M. Nilges, A. M. Gronenborn, A. T. Brünger and G. M. Clore, Determination of three-dimensional structure of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxy peptidase inhibitor and barley serine proteinase inhibitor 2, Protein Eng., **2**, (1988), 27-38.
[134] Y. Okamoto, T. Kikuchi and H. Kawai, Prediction of low-energy structures of met-enkephalin by Monte Carlo simulated annealing, Chem. Lett., **7**, (1992), 1275-1278.
[135] Y. Okamoto, M. Fukugita, T. Nakataka and H. Kawai, α-helix folding by Monte Carlo simulated annealing in isolated C-peptide of ribonuclease A, Protein Eng., **4**, (1991), 639-647.
[136] T. Ooi, M. Oobatake, G. Némethy and H. A. Scheraga, Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides, Proc. Natl. Acad. Sci. USA, **84**, (1987), 3086-3090.
[137] C. M. Oshiro, I. D. Kuntz and J. S. Dixon, Flexible ligand docking using a genetic algorithm, J. Comput. Aided Mol. Design, **9**, (1995), 113-130.

[138] P. M. Pardalos, D. Shalloway and G. Xue, Optimization methods for computing global minima of nonconvex potential energy functions, J. Global Optim., **4**, (1994), 117-133.

[139] P. M. Pardalos, D. Shalloway and G. Xue (editors), Global minimization of nonconvex energy functions : Molecular conformation and protein folding, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **23**, Amer. Math. Soc., (1996).

[140] P. M. Pardalos, X. Liu and G. L. Xue, Protein conformation of a lattice model using tabu search, J. Global Optim., **11**, (1997), 55-68.

[141] G. Perrot, B. Cheng, K. D. Gibson, J. Vila, K. A. Palmer, A. Nayeem, B. Maigret and H. A. Scheraga, **13**, (1992), 1-11.

[142] A. T. Phillips, J. B. Rosen and V. H. Walke, Molecular structure determination by convex global underestimation of local energy minima, In DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **23**, Amer. Math. Soc., (1996), 181-198.

[143] L. Piela and H. A. Scheraga, On the multiple-minima problem in conformational analysis of polypeptides. I. Backbone degrees of freedom for a perturbed α-helix, Biopolymers, **26**, (1987), S33-S58.

[144] L. Piela, J. Kostrowicki and H. A. Scheraga, The multiple-minima problem in the conformational analysis of molecules. Deformation of the potential energy hypersurface by the diffusion equation method, J. Phys. Chem., **93**, (1989), 3339-3346.

[145] J. Pillardy and L. Piela, Molecular dynamics on deformed potential energy hypersurfaces, J. Phys. Chem., **99**, (1995), 11805-11812.

[146] E. O. Purisima and H. A. Scheraga, An approach to the multiple-minima problem by relaxing dimensionality, Proc. Natl. Acad. Sci. USA, **83**, (1986), 2782-2786.

[147] E. O. Purisima and H. A. Scheraga, An approach to the multiple-minima problem in protein folding by relaxing dimensionality, J. Mol. Biol., **196**, (1987), 697-709.

[148] C. A. Reynolds, P. M. King and W. G. Richards, Free energy calculations in molecular biophysics, Mol. Phys. **76**, (1992), 251-275.

[149] D. R. Ripoll and H. A. Scheraga, On the multiple-minima problem in the conformational analysis of polypeptides. IV. Application of the electrostatically driven Monte Carlo method to the 20-residue bound portion of melittin, Biopolymers, **30**, (1990), 165-176.

[150] D. R. Ripoll, M. J. Vásquez and H. A. Scheraga, The electrostatically driven Monte Carlo method - Application to conformational analysis of decaglycine, Biopolymers, **31**, (1991), 319-330.

[151] D. R. Ripoll and H. A. Scheraga, The multiple minima problem in the conformational analysis of polypeptides. III. An electrostatically driven Monte Carlo method – tests on enkephalin, J. Protein Chem., **8**, (1989), 263-287.

[152] R. Rosenfeld, Q. Zheng, S. Vajda and C. DeLisi, Computing the structure of bound peptides. Application to antigen recognition by class I major histocompatibility complex receptors, J. Mol. Biol., **234**, 1993, 515-521.

[153] S. H. Rotstein and M. A. Murcko, GroupBuild: a fragment-based method for de novo drug design, J. Med. Chem., **36**, (1993), 1700-1710.

[154] S. H. Rotstein and M. A. Murcko, GensStar: a program for de novo drug design, J. Comput. Aided Mol. Design, **7**, (1993), 23-43.

[155] A. Šali, E. Shaknovich and M. Karplus, Thermodynamics and kinetics of protein folding, In DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **23**, Amer. Math. Soc., (1996), 199-213.

[156] H. A. Scheraga, Predicting three-dimensional structures of oligopeptides, In Reviews in Computational Chemistry, **III**, VCH Publishers, (1992), 73-142.

[157] T. Schlick, Optimization methods in computational chemistry, In Reviews in Computational Chemistry, **III**, VCH Publishers, (1992), 1-71.

[158] C. S. Shao, R. H. Byrd, E. Eskow and R. B. Schnabel, (Manuscript).

[159] B. K. Shoichet and I. D. Kuntz, Protein docking and complementarity, J. Mol. Biol., **221**, (1991), 327-346.

[160] D. Shalloway, Packet annealing : A deterministic method for global minimization. Applications to molecular conformation, In Recent Advances in Global Optimization, Princeton Univ. Press, (1992), 433-477.

[161] D. Shalloway, Application of the renormalization group to deterministic global minimization of molecular conformation energy functions, J. Global Optim., **2**, (1992), 281-311.

[162] J. Shen and F. A. Quiocho, Calculation of binding energy differences for receptor-ligand systems using the Poisson-Boltzmann methods, J. Comput. Chem., **16**, (1995), 445-448.
[163] M. L. Silver, H. C. Guo, J. L. Strominger and D. Wiley, Atomic structure of a human MHC molecule presenting an influenza virus peptide, Nature, **360**, (1992), 367-368.
[164] M. J. Sippl, Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins, J. Mol. Biol., **213**, (1990), 859-883.
[165] M. E. Snow, Powerful simulated annealing algorithm locates global minima of protein folding potentials from multiple starting conformations, J. Comput. Chem., **13**, (1992), 579-584.
[166] L. Stern, J. Brown, T. Jardetsky, J. Gorga, R. Urban, L. Strominger and D. Wiley, Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide, Nature, **368**, (1994), 215-221.
[167] F. H. Stillinger and T. A. Weber, Nonlinear optimization simplified by hypersurface deformation, J. Stat. Phys., **52**, (1988), 1429-1445.
[168] T. P. Straatsma and J. A. McCammon, Computational alchemy, Annu. Rev. Phys. Chem., **43**, (1992), 407-435.
[169] S. Sun, Reduced representation model of protein structure prediction : Statistical potential and genetic algorithms, Protein Sci., **2**, (1993), 762-785.
[170] P. Tuffery, C. Etchebest, S. Hazout and R. Lavery, A new approach to the rapid determination of protein side chain conformations, J. Biomol. Struct. Dynam., **8**, (1991), 1267-1289.
[171] R. G. Urban and R. M. Chicz, MHC Molecules: Expression, Assembly and Function, R. G. Landes Company and Chapman & Hall, (1996).
[172] S. Vajda and C. De Lisi, Determining minimum energy conformations of polypeptides by dynamic programming, Biopolymers, **29**, (1990), 1755-1772.
[173] S. Vajda, M. S. Jafri, U. Sezerman and C. De Lisi, Necessary conditions for avoiding incorrect polypeptide folds in conformational search by energy minimizations, Biopolymers, **33**, (1993), 173-192.
[174] S. Vajda, M. Sippl and J. Novotny, Empirical potentials and functions for protein folding and binding, Curr. Opinion Struct. Biol., **7**, (1997), 222-228.
[175] W. F. van Gunsteren and H. J. C. Berendsen, GROMOS, Groningen Molecular Simulation, (1987).
[176] M. Vàsquez and H. A. Scheraga, Use of buildup and energy-minimization procedures to compute low-energy structures of the backbone of enkephalin, Biopolymers, **24**, (1985), 1437-1447.
[177] M. Vásquez, G. Némethy and H. A. Scheraga, Conformational energy calculations on polypeptides and proteins, Chem. Rev., **94**, (1994), 2183-2239.
[178] M. Vásquez, E. Meirovitch and H. Meirovitch, A free energy based Monte Carlo minimization procedure for biomolecules, J. Phys. Chem., **98**, (1994), 9380-9382.
[179] G. Verkhivker, K. Appelt, S. T. Freer and J. E. Vilafranca, Empirical free energy calculations of ligand-protein crystallographic complexes. 1. Knowledge based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity, Protein Eng., **8**, (1995), 677-691.
[180] B. von Freyberg and W. Braun, Efficient search for all low energy conformations of polypeptides by Monte Carlo methods, J. Comp. Chem., **12**, (1991), 1065-1076.
[181] B. von Freyberg and W. Braun, Minimization of empirical energy functions in proteins including hydrophobic surface area effects, J. Comp. Chem., **14**, (1993), 510-521.
[182] J. Vila, R. L. Williams, M. Vásquez and H. A. Scheraga, Empirical solvation models can be used to differentiate native from near-native conformations of bovine pancreatic trypsin inhibitor, Proteins, **10**, (1991), 199-218.
[183] V. N. Vorobjev, J. A. Grant and H. A. Scheraga, A combined iterative and boundary element approach for solution of the nonlinear Poisson-Boltzmann equation, J. Am. Chem. Soc., **114**, (1992), 3189-3196.
[184] A. Wallquist, R. L. Jernigan and D. G. Covell, A preference-based free energy parameterization of enzyme-inhibitor binding. Applications to HIV-1 protease inhibitor design, Protein Sci., **4**, (1995), 1881-1903.
[185] H. Wang, Grid-search molecular accessible surface algorithm for solving the protein docking problem, J. Comput. Chem., **12**, (1991), 746-750.

[186] A. Warshel and J. Aqvist, Electrostatic energy and macromolecular function, Annu. Rev. Biophys. Biophys. Chem., **20**, (1991), 267-298.

[187] S. J. Weiner, P. A. Kollman, D.A. Case , U.C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins, J. Am. Chem. Soc., **106**, (1984), 765-784.

[188] S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, An all atom force field for simulations of proteins and nucleic acids, J. Comp. Chem., **7**, (1986), 230-252.

[189] W. I. Weis, A. T. Brünger, J. J. Skehel and D. C. Wiley, Refinement of the influenza virus hemagglutinin by simulated annealing, J. Mol. Biol., **212**, (1990), 737-761.

[190] L. Wesson and D. Eisenberg, Atomic solvation parameters applied to molecular dynamics of proteins in solution, Protein Sci., **1**, (1992), 227-235.

[191] R. L. Williams, J. Vila, G. Perrot and H. A. Scheraga, Empirical solvation models in the context of conformational energy searches : Application to bovine pancreatic trypsin inhibitor, Proteins, **14**, (1992), 110-119.

[192] S. R. Wilson, W. Cui, J. W. Moskowitz and K. E. Schmidt, Conformational analysis of flexible molecules : Location of the global minimum energy conformation by the simulated annealing method, Tetrahedron Lett., **29**, (1988), 4373-4376.

[193] S. J. Wodak and J. Janin, Computer analysis of protein-protein interactions, J. Mol. Biol., **124**, (1978), 323-342.

[194] S. J. Wodak, M. De Crombrugghe and J. Janin, Computer studies of interactions between macromolecules, Prog. Biophys. Mol. Biol., **49**, (1987), 29-63.

[195] S. S. Zimmerman, M. S. Pottle, G. Nèmethy and H. A. Scheraga, Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP, Macromolecules, **10**, (1977), 1-9.

Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544-5263
E-mail address: `floudas@titan.princeton.edu`

Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544-5263
E-mail address: `john@titan.princeton.edu`

Center for Applied Optimization, University of Florida, Gainesville, FL 32611
E-mail address: `pardalos@ufl.edu`