Predicting Solvated Peptide Conformations via Global Minimization of Energetic Atom–to–Atom Interactions

J. L. Klepeis, I. P. Androulakis[1], M. G. Ierapetritou, and C. A. Floudas[2]
Department of Chemical Engineering
Princeton University
Princeton, N.J. 08544-5263

Submitted: March 1997/Revised: June 1997

Abstract

A global optimization method is described for identifying the global minimum energy conformation, as well as lower and upper bounds on the global minimum conformer of solvated peptides. Potential energy contributions are calculated using the ECEPP/3 force field model. In considering the effects of hydration, two implicit free energy models are compared. One method is based on the calculation of solvent–accessible surface areas, while the other uses information on the solvent–accessible volume of hydration shells. Detailed information on the potential and solvation energy contributions is presented for the terminally blocked single residue peptides. In addition, based on a procedure that allows the exclusion of domains of the (ϕ, ψ) space, a number of oligopeptide structure prediction problems are considered, and the role of the solvation model in defining global minimum conformations is addressed.

Keywords : Protein folding, Solvation, Global optimization.

[*] Current Address: Corporate Research Science Laboratories, Exxon Research & Engineering Co., Annandale, NJ 08801.

[†] Author to whom all correspondence should be addressed.

1 Introduction

The protein folding problem is one of the most challenging problems in current biochemistry. Advances in genetic engineering already allow us to produce proteins with specific amino acid sequences. The next step is to understand how these proteins fold, both in the static and dynamic sense. To do this, we must learn to predict how the information provided by a particular amino acid sequence controls the formation of the three-dimensional structure of the biologically active protein. Once the folded structure is known, biological and chemical properties can be predicted and adjusted. Success in this area would have important ramifications in both the academic and industrial worlds. Design of proteins and drugs with specific therapeutical properties motivates a large sector of the biotechnology industry. Protein folding information would also advance research in the areas related to the human genome project and understanding the mechanism of hereditary and infectious diseases.

The native protein conformation is defined by the amino acid sequence and environmental conditions, of which the solvent choice has major importance. By employing the thermodynamic hypothesis (Anfinsen et al., 1961), the folded conformation can be found by identifying the structure exhibiting the global minimum free energy. Strictly speaking, this involves the determination of energetic and entropic contributions to the free energy. However, at constant temperature the identification of a number of unique low energy conformations, along with the global minimum energy conformation, should be sufficient for identifying the native conformation. That is, proteins in their biologically active state exist in a well-defined conformation with small fluctuations around this average. The inclusion of the free energy of solvation in these calculations further defines the correct ensemble of structures. In addition, it has been demonstrated that solvation free energy models aid in discerning the native among near–native conformations (Vila et al., 1991).

The identification of the global minimum energy structure, with or without solvation, requires the use of efficient methods to search the nonconvex multi–dimensional conformation space. A large number of techniques have been developed, with varying degrees of success, to treat the multiple–minima problem, including stochastic search methods (e.g. simulated annealing, genetic algorithms), smoothing methods (e.g. diffusion equation, packet annealing), and simulation methods (e.g. molecular dynamics, Monte Carlo methods). In most cases, these techniques have been used to treat only the potential energy terms of the conformational energy. When solvation energy is considered, as is the case in many MD and MC simulations, the search is strictly local. A number of recent review papers have surveyed the treatment of the protein conformation problem in terms of the global minimization of nonconvex energy functions (Neumaier, 1997; Pardalos et al., 1996; Vásquez et al., 1994; Scheraga, 1992).

This work addresses the protein folding problem, including the effects of solvation, through the use of a deterministic global optimization algorithm. This branch–and–bound based global optimization algorithm, known as αBB, is applicable to a large class of nonlinear optimization problems that have twice–differentiable functions (Adjiman et al., 1997b,c,a, 1996; Androulakis et al., 1995). In the protein folding problem the objective function is defined by the potential energy and solvation models, which are described in Section 2. The αBB implementation is then detailed in Section 3. The approach identifies an ϵ-global minimum, along with a number of low energy conformers. Furthermore, upper and lower bounds on the global minimum energy are obtained. In this paper, the αBB algorithm is used to predict the solvated global minimum conformers of the naturally occurring amino acids, as well as a selection of larger oligopeptide problems.

2 Mathematical Modeling

2.1 Protein Representation

From a chemical point of view, a protein is essentially a polymer chain composed of a sequence of various amino acid residues connected by peptide bonds. Naturally occurring proteins are composed of 20 different amino acid residues, where the form of the side chain (e.g., methyl, butyl, benzoic, etc.) defines these residues. This basic structure is slightly different only in the case of proline residues.

The repeating unit -NC$^{\alpha}$C'- connected by peptide bonds defines the backbone of the protein. In addition, the protein possesses amino and carboxyl end groups. Covalent bond angle requirements and interatomic forces bend and twist the chain in a characteristic way for each protein. The protein chain "curls up" into a unique three-dimensional geometric conformation called the folded state of the protein. It is this configuration which defines the shape of the protein surface, as well as the specific chemically active groups present on the surface, which in turn sets the biological functionality of the protein.

Instead of specifying the coordinate vector for all atoms in a protein molecule, one can specify all bond lengths, covalent bond angles and dihedral angles. Under biological conditions, the bond lengths and bond angles are fairly rigid and thus can be assumed to be fixed at their equilibrium values. Using this assumption, the backbone dihedral angles fully determine the geometric shape of the folded protein.

The names of the dihedral angles of a protein chain follow a standard nomenclature. The dihedral angle between the normals of the planes formed by atoms $C'_{i-1}N_iC_i^\alpha$ and $N_iC_i^\alpha C'_i$ respectively, is called ϕ_i where $i-1$ and i are two adjacent amino acid residues. The angle defined by the planes $R_iC_i^\alpha C'_i$ and $C_i^\alpha C'_i N_{i+1}$ respectively, is called ψ_i where i and $i+1$ are two adjacent amino acid residues. Also, ω_i is the dihedral angle defined by the planes $C_i^\alpha C'_i N_{i+1}$ and $C'_i N_{i+1} C_{i+1}^\alpha$. The letter χ is utilized to denote the dihedral angles which are associated with the side groups R_i. Finally, the letter θ is used to name the dihedral angles associated with the two end groups.

2.2 Potential Energy Model

In reality, the dynamics of atoms in a molecule are governed by the quantum theory of its participating electrons. Using the Born-Oppenheimer approximation, one can determine the energy for fixed atomic nuclei from the smallest eigenvalue of the Hamiltonian of the electron system. These approximations and their derivatives are calculated using ab–initio methods. However, due to their computational complexity, such calculations are limited to extremely small molecules.

As a result, many models have been developed using a classical description of molecules in terms of atomic bonds and effective interactions. Some of these parameterizations of molecular potential functions include ECEPP (Momany et al., 1975, 1974a,b), AMBER (Weiner et al., 1986, 1984), CHARMM (Brooks et al., 1983), DISCOVER (Dauber-Osguthorpe et al., 1988), GROMOS (van Groningen and Berendsen, 1987), MM3 (Allinger et al., 1989), ENCAD (Levitt, 1983), ECEPP/2 (Némethy et al., 1983) and ECEPP/3 (Némethy et al., 1992). In general, these models, also known as force fields, are expressed as summations of empirically derived potential functions, with the mathematical form of individual energy terms based on the phenomenological nature of that term. Constants describing molecular geometry, such as bond lengths and bond angles, are parameterized on empirical structural information. In addition, thermodynamic data from small molecules and spectroscopic data are used to derive the parameters describing the relative strengths of particular interatomic interactions. In most cases, these force fields are atom centered potentials from which the total molecular energy is computed as a sum over all pairwise interactions.

In this work, the ECEPP/3 (Empirical Conformational Energy Program for Peptides) potential model is utilized. In this force field, it is assumed that the covalent bond lengths and bond angles are fixed at their equilibrium values. It has been observed that variations in bond lengths and bond angles depend mostly on short range interactions; that is, those between the side chain and backbone of the same residue. Under this assumption, all residues of the same type have essentially the same geometry in various proteins (Momany et al., 1975). Therefore, a chain of any sequence can be generated using the fixed geometry specific to each type of amino acid residue in the sequence.

Based on these approximations, the conformation is only a function of the dihedral angles. That is, ECEPP/3 accounts for energy interaction terms which can be expressed solely in terms of the dihedral angles. The total conformational energy is calculated as the sum of the electrostatic, nonbonded, hydrogen bonded, and torsional contributions. There is also a pseudo–potential for loop closing if the polypeptide contains two or more sulfur–containing residues. More recent work by includes a revised treatment of prolyl and hydroxyprolyl residues (Némethy et al., 1992). For each prolyl or hydroxyprolyl residue contained in the polypeptide a fixed internal conformational energy for the pyrolidine ring is added. The main energy contributions (electrostatic, nonbonded, hydrogen bonded) are computed as the sum of terms for each atom pair (i,j) whose interatomic distance is a function of at least one dihedral angle. The contributing terms to the total potential energy of ECEPP/3 are shown in Figure 1, and the development of the appropriate parameters is discussed and reported in ECEPP/3 (Némethy et al., 1992).

2.3 Solvation Models

A complete description of the total energy of a polypeptide must also include its interactions with the solvent. Explicit methods can be used by actually surrounding the polypeptide with solvent molecules and calculating solvent–peptide and solvent–solvent interactions using potentials similar to those previously described. Although these methods are conceptually simple, explicit inclusion of solvent molecules greatly increases the computational time

$$
\begin{aligned}
E = & \sum_{(ij)\in ES} \frac{q_i\, q_j}{r_{ij}} && \text{(Electrostatic)} \\
& + \sum_{(ij)\in NB} F_{ij} \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} && \text{(Nonbonded)} \\
& + \sum_{(ij)\in HX} \frac{A'_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{10}} && \text{(Hydrogen bonded)} \\
& + \sum_{k\in TOR} \left(\frac{E_{o,k}}{2}\right)(1+c_k \cos n_k \theta_k) && \text{(Torsional)} \\
& + \sum_{l\in SS} B_l \sum_{i=1}^{i=3} (r_{il} - r_{io})^2 && \text{(Cystine Loop-Closing)} \\
& + \sum_{l\in SS} \left(\frac{E_{o,l}}{2}\right)(1+c_l \cos n_l \chi_l) && \text{(Cystine Torsional)} \\
& + \sum_{p\in PRO} E_p && \text{(Proline Internal)}
\end{aligned}
$$

Figure 1: Potential energy terms in ECEPP/3 force field. r_{ij} refers to the interatomic distance of the atomic pair (ij). Q_i and Q_j are dipole parameters for the respective atoms, in which the dielectric constant of 2 has been incorporated. F_{ij} is set equal to 0.5 for 1–4 interactions and 1.0 for 1–5 and higher interactions. A_{ij}, C_{ij}, A'_{ij} and B_{ij} are nonbonded and hydrogen bonded parameters specific to the atomic pair. $E_{o,k}$ and $E_{o,l}$ are parameters corresponding to torsional barrier energies for a given dihedral angle. θ_k represents any dihedral angle, while χ_l refers to those dihedral angles involved in cystine loop–closing. c_k and c_l take the values -1,1, and n_k and n_l refer to the symmetry type for the particular dihedral angle. The cystine loop–closing term is calculated as a penalty term of three distances involved in loop–closing, where r_{il} represents the actual distance and r_{io} represents the required distance. B_i, the penalty parameter, is set equal to 100. Finally, E_p is a fixed internal energy that is added for each proline residue in the protein.

needed to simulate the polypeptide system. Therefore, most simulations of this type are limited to restricted conformational searches. In addition, the interactions between the protein molecule and the surrounding water molecules are not fixed for a given peptide configuration. In reality, a large number of solvent configurations must be considered, and the free energy can then be calculated by averaging over these configurations.

Simpler methods for estimating solvent free energies have been developed using both integral equations and continuum models. Integral equation methods can be used to evaluate solvent structure and thermodynamic properties. Typically, molecular dynamics and Monte Carlo simulations are used to calculate ensemble averages from which free energy differences can be obtained. A number of methods have been proposed to estimate these solvation free energies from simulations based on molecular dynamics and Monte Carlo averages (Dejaegere and Karplus, 1996; Kollman, 1993; Straatsma and McCammon, 1992). The integral equation method has also been used to analyze the solvent structure of a protein system (Kitao et al., 1993). In contrast, continuum models use a simplified representation of the solvent environment by neglecting the molecular nature of the water molecules. Calculations of solvation free energies using electrostatic continuum models rely on numerical solutions to the Poisson–Boltzmann equation from which dielectric and ionic strength effects are obtained (Honig et al., 1993). Other continuum models estimate free energies of solvation as a function of surface areas and volumes.

In this work, solvation contributions are included implicitly using empirical correlations with both surface area and volume. The main assumption of these models is that, for each functional group of the peptide, a hydration free energy can be calculated from an averaged free energy of interaction of the group with a layer of solvent known as the hydration shell. In addition, the total free energy of hydration is expressed as a sum of the free energies of hydration for each of the functional groups of the peptide, that is, an additive relationship is assumed.

2.3.1 MSEED – Accessible Surface Area Model

Accessible surface area methods assume that the free energy of hydration is proportional to the solvent–accessible surface area of the peptide, as described by the following equation:

$$E_{HYD} = \sum_{i=1}^{N}(A_i)(\sigma_i) \qquad (1)$$

In Equation (1), an additive relationship for N individual functional groups is assumed. (A_i) represents the solvent–accessible surface area for the functional group, and (σ_i) is an empirically derived free energy density parameter.

There are a number of ways to define the surface of a peptide. In developing these surfaces the peptide is represented by a union of spheres, with the radii of the spheres set by the van der Waals radii of the constituent atoms. A spherical test probe is then rolled over these spheres, thereby tracing out a surface. The molecular surface is set by direct contact between the probe sphere and the peptide spheres. In areas where the probe cannot make direct contact, the closest part of the probe is used. The solvent–accessible surface is defined by the surface traced by the center of the probe as the probe rolls over the peptide spheres. Of course, these areas depend on the radius of the probe sphere; when this radius is set to zero both the molecular and solvent–accessible surface areas become the van der Waals surface of the peptide.

In this work, solvent–accessible surface areas are calculated using the MSEED (Perrot et al., 1992) program, which employs algorithms developed by Connolly (Connolly, 1983). MSEED eliminates many unnecessary computations by considering only those convex faces that are on the accessible surface. Rigorous implementation of Connolly's method requires the calculation of interior surface areas, which are ultimately found to be zero. A full description of the MSEED algorithm is given by (Perrot et al., 1992). A number of other methods for calculating surface areas are also available (von Freyberg and Braun, 1993; Eisenhaber et al., 1995; Eisenhaber and Argos, 1993).

There are some limitations to these surface area calculations. Each convex face on the surface is defined by the points of intersections for three spheres and by the set of arcs of intersecting circles for two spheres. These points and arcs define a curvilinear polygon on the surface of one sphere, and only contiguous polygons are included in the accessible surface area of the peptide. However, in searching for surface accessible polygons MSEED uses points of intersection of three spheres (vertices). Obviously, in the case of two intersecting spheres or a completely free sphere MSEED will not include this solvent–accessible surface area. It will also not search domains connected by only two overlapping spheres. This could be overcome by using multiple starting points to search for other domains. Despite these limitations, it was found that for water, with an effective probe sphere of 1.4 $\mathring{A}$, the error in calculating the solvent-accessible area is less than 2 % for a number of test problems (Perrot et al., 1992).

Another problem may occur during minimization of the total energy. Specifically, due to conformational changes during minimization, the area of each atomic surface changes continuously but the gradients may have discontinuities. This occurs when a new vertex or edge appears on the surface. If the discontinuity is large, minimization techniques requiring gradients may fail to converge to the local minimum conformation. A full description of all the situations in which the gradient of the molecular surface area becomes discontinuous has been reported (Wawak et al., 1994).

Once the solvent–accessible surface areas have been calculated, these values must be multiplied by the appropriate (σ_i) parameters as shown in Equation (1). There are a number of models available, including JRF, OONS, SRFOPT, which provide estimates for these parameters based on interactions between water and the functional groups of peptides. It has been shown that minimum energy solvated conformations predicted by the JRF model provided the best correspondence to native (crystallographic) structures when compared with other models (Williams et al., 1992). These parameters were derived from NMR studies of low energy solvated configurations of 13 tetrapeptides. An ensemble of low energy structures for these tetrapeptides was also developed using the ECEPP/2 potential function, and a nonlinear least–squares system was optimized for the best set of atomic solvation parameters. Because it was developed from minimum energy conformations of peptides, the JRF parameter set has been shown to produce undesirable perturbations during local minimizations if the solvation energy contributions are added at every iteration. Therefore, the surface–accessible solvation energies are only included at local minimum conformations. The JRF parameters and atomic radii used in computing solvation energies with the MSEED model are given in Table 1.

2.3.2 RRIGS – Accessible Volume Shell Model

In these models, the free energy of hydration is assumed to be proportional to the water–accessible volume of a hydration layer surrounding the peptide. This can be represented in a form very similar to Equation (1):

$$E_{HYD} = \sum_{i=1}^{N} (VHS_i)(\delta_i) \tag{2}$$

An additive relationship for the N individual atoms of the peptide is assumed, and (VHS$_i$) represents the solvent–accessible volume of hydration shell for each atom i which is exposed

Table 1: JRF parameters employed in this work. The first column is the atom type. The JRF solvation parameters in the second column are given in cal/(mol $\mathring{A}^2$), and the last column corresponds to atomic radii ($\mathring{A}$).

Atom Type	**JRF**	**Radii**
C C_a, CH, CH_2, CH_3	216	2.00
C carboxyl, carbonyl	-732	1.55
C aromatic	-678	1.75
N all	-312	1.55
O carboxyl, carbonyl	-262	1.40
O other (e.g., hydroxyl)	-910	1.40
S all	-281	2.00

to water. Finally, the (δ_i) are empirically determined free energy of hydration densities for these atoms.

The hydration shell is defined by the volume inside a sphere of radius R_i^h but outside a sphere of radius R_i^v, with both radii centered on atom i. The larger radius, R_i^h, corresponds to the radius of the first hydration shell of atom i, while R_i^v is equal to the van der Waals radius. In order to calculate (VHS_i), the volume of a collection of overlapping hard spheres must be computed using:

$$V(\mathbf{R}) \;=\; \sum_i a_i S_i - \sum_{ij} b_{ij} D_{ij} + \sum_{ijk} c_{ijk} T_{ijk} - \sum_{ijkl} d_{ijkl} Q_{ijkl} \qquad (3)$$

In Equation (3), S_i signifies the volume of a single sphere, while D_{ij}, T_{ijk} and Q_{ijkl} represent the volume of intersection of two, three and four spheres, respectively. This is sufficient because all higher order overlaps can be decomposed into the three types of intersections included in Equation (3). Therefore, the solvent–accessible volume of hydration can be written as:

$$(VHS_i) \;=\; V(R_i^h) - V(R_i^v) \qquad (4)$$

The first term in Equation (4) is calculated using Equation (3) with the radii of all atoms set equal to their van der Waals radii, while the second term is calculated with the radius of atom i equal to R_i^h and the van der Waals radii of all the other atoms. Although the calculation of the (VHS$_i$) is explicit, the form of Equation (3) is not suitable for force field models using pairwise intramolecular potential, such as ECEPP/3. Furthermore, direct truncation at the double–overlap term would lead to large errors.

In this work, the RRIGS (Reduced Radius Independent Gaussian Sphere) approximation is used to efficiently calculate the exposed volume of the hydration shell (Augspurger and Scheraga, 1996). This method uses a truncated form of Equation (3) but also artificially reduces the van der Waals radii of all atoms other than atom i when calculating (VHS$_i$). These reductions effectively decrease the contribution of the double overlap terms, leading to a cancellation of the error which results from neglecting the triple and higher overlap terms. In addition, the characteristic density of being inside the overlap volume of two intersecting spheres is not represented as a step function, but as a Gaussian function which provides continuous derivatives of the hydration potential. Therefore, the solvation energy contributions can easily be added at every step of local minimizations.

The reduced van der Waals radii are fixed and do not need to be recalculated for every configuration because the fixed geometry assumption was used (as in ECEPP/3). The contribution of covalently bonded atoms to (VHS$_i$) is also fixed because of these assumptions. Therefore, the RRIGS approximation has the same set of interactions as the ECEPP/3 potential. Finally, using the exact (VHS$_i$) expressions, values for the empirical free energy densities, (δ_i), were developed by a least square fitting of experimental free energy of solvation data for 140 small organic molecules (Augspurger and Scheraga, 1996). The empirical free energy density of solvation parameters and the corresponding van der Waals and hydration radii are given in Table 2.

3 Global Optimization

The energy minimization problem can be formulated as a nonconvex nonlinear optimization problem. Let $i = 1, \ldots, N_{RES}$ be an indexed set describing the sequence of amino acid residues in the peptide chain. There are $\phi_i, \psi_i, \omega_i,\ i = 1, \ldots, N_{RES}$ dihedral angles along the backbone of this peptide. In addition, let $k = 1, \ldots, K^i$ denote the dihedral angles of the side chains for the i^{th} residue and $j = 1, \ldots, J^N$ denote the dihedral angles of the amino end group and $j = 1, \ldots, J^C$ of the carboxyl end group, respectively. Therefore, these angles can be defined in following manner : $\chi_i^k,\ i = 1, \ldots, N_{RES},\ k = 1, \ldots, K^i$ for the side chain dihedral angles; $\theta_j^N,\ j = 1, \ldots, J^N$ and $\theta_j^C,\ j = 1, \ldots, J^C$ for the amino and carboxyl end

Table 2: Free energy density of solvation parameters employed in the RRIGS solvation model. The first column describes the atom type, and the second column provides the solvation parameters in cal/(mol $\mathring{A}^2$). The last two columns correspond to the van der Waals and hydration radii ($\mathring{A}$), respectively.

Atom Type	δ	R^v	R^h
H hydroxyl, amino	-10.35	1.415	4.17
H acid	-3.206	1.415	4.17
H amide	-7.714	1.415	4.17
H thiol	2.709	1.415	4.17
C aliphatic CH_3	1.319	2.125	5.35
C aliphatic CH_2	0.2374	2.225	5.35
C aliphatic CH	-1.271	2.375	5.35
C other aliphatic	-2.297	2.060	5.35
C cyclic CH	0.2890	2.375	5.35
C aromatic CH	-0.2137	2.100	5.35
C aromatic CR	-1.713	1.850	5.35
C branched aromatic C	-1.910	1.850	5.35
C aromatic COH	-0.6063	1.850	5.35
C carbonyl	2.696	1.870	5.35
N primary amine	-1.149	1.755	5.05
N secondary amine	-10.28	1.755	5.05
N aromatic	-10.48	1.755	5.05
N amide	-7.332	1.755	5.05
O hydroxyl, ether	-7.396	1.620	4.95
O acid, ester	0.07897	1.620	4.95
O ketone, carbonyl	-15.70	1.560	4.95
O acid, amide carbonyl	-15.56	1.560	4.95
S thiol, disulfide	-4.706	2.075	5.37

group dihedral angles, respectively. Using these definitions the optimization problem takes the following form:

$$\min \quad E(\phi_i, \psi_i, \omega_i, \chi_i^k, \theta_j^N, \theta_j^C) \tag{5}$$

$$\begin{aligned}
\text{subject to} \quad -\pi &\leq \phi_i \leq \pi, \;\; i = 1, \ldots, N_{RES} \\
-\pi &\leq \psi_i \leq \pi, \;\; i = 1, \ldots, N_{RES} \\
-\pi &\leq \omega_i \leq \pi, \;\; i = 1, \ldots, N_{RES} \\
-\pi &\leq \chi_i^k \leq \pi, \;\; i = 1, \ldots, N_{RES}, \;\; k = 1, \ldots, K^i \\
-\pi &\leq \theta_j^N \leq \pi, \;\; j = 1, \ldots, J_N \\
-\pi &\leq \theta_j^C \leq \pi, \;\; j = 1, \ldots, J_C
\end{aligned}$$

In general, E represents the total potential energy function and the free energy of solvation. For accessible volume shell hydration (RRIGS) this is the exact formulation because both energetic and gradient contributions can be added at each step of the minimization. However, in the case of surface–accessible hydration (MSEED and JRF parameters), the potential energy function is minimized before adding the hydration energy contributions. In other words, gradient contributions from solvation are not considered (see Section 2.3.1). This approach is represented by the following equation:

$$E_{Total} = E_{Min}^{Unsol} + E^{Sol} \tag{6}$$

Even after reducing this optimization problem to a function of internal variables (dihedral angles), the multidimensional surface that describes the energy function has an astronomically large number of local minima. This has become known as the multiple–minima problem. In addition, evaluation of the potential, especially with the addition of solvation, is computationally expensive, which makes even local minimization slow. Because the objective function has many local minima, using local optimization techniques depends on the initial points selected. Therefore, global optimization algorithms are needed to effectively locate the global minimum corresponding to the native state of the protein. A large number of techniques have been developed to search this nonconvex conformational space. Many methods employ stochastic search procedures, while others rely on simplifications of the potential model and/or mathematical transformations. In addition, the use of statistical and/or heuristic conformational information is often required. In general, the major limitation is that these methods depend heavily on the supplied initial conformation. As a result, many initial points are necessary to search the conformational space, and there is also no guarantee for global convergence because large sections of the domain space may be overlooked.

In this work, the global optimization approach αBB has been extended to identifying global minimum energy conformations of solvated peptides. The development of this branch and bound method was motivated by the need for an algorithm that could guarantee convergence to the global minimum of nonlinear optimization problems with twice–differentiable functions (Floudas, 1997). The application of this algorithm to the minimization of potential energy functions was first introduced for microclusters (Maranas and Floudas, 1993, 1992), and small acyclic molecules (Maranas and Floudas, 1994a,b). The αBB approach has also been extended to constrained optimization problems (Adjiman et al., 1997c,a, 1996; Androulakis et al., 1995). In more recent work, the algorithm has been shown to be successful for isolated peptide systems using the realistic ECEPP/3 potential energy model (Maranas et al., 1996; Androulakis et al., 1997).

3.1 Minimization of Potential Energy using αBB

The αBB global optimization algorithm effectively brackets the global minimum solution 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 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. The lower bounding function (L) of the energy hypersurface can be expressed in the following manner:

$$\begin{aligned} L \; = \; E + \{ \quad & \sum_{i=1}^{N_{RES}} \alpha_{\phi,i} \left(\phi_i^L - \phi_i\right)\left(\phi_i^U - \phi_i\right) + \qquad (7) \\ & \sum_{i=1}^{N_{RES}} \alpha_{\psi,i} \left(\psi_i^L - \psi_i\right)\left(\psi_i^U - \psi_i\right) + \\ & \sum_{i=1}^{N_{RES}} \alpha_{\omega,i} \left(\omega_i^L - \omega_i\right)\left(\omega_i^U - \omega_i\right) + \end{aligned}$$

$$\sum_{i=1}^{N_{RES}} \sum_{k=1}^{K^i} \alpha_{\chi,i,k} \left(\chi_i^{k,L} - \chi_i^k\right) \left(\chi_i^{k,U} - \chi_i^k\right) +$$
$$\sum_{j=1}^{J^N} \alpha_{j,\theta^N} \left(\theta_j^{N,L} - \theta_j^N\right) \left(\theta_j^{N,U} - \theta_j^N\right) +$$
$$\sum_{j=1}^{J^C} \alpha_{j,\theta^C} \left(\theta_j^{C,L} - \theta_j^C\right) \left(\theta_j^{C,U} - \theta_j^C\right) \}$$

Here $\phi_i^L, \psi_i^L, \omega_i^L, \chi_i^{k,L}, \theta_j^{N,L}, \theta_j^{C,L}$ and $\phi_i^U, \psi_i^U, \omega_i^U, \chi_i^{k,U}, \theta_j^{N,U}, \theta_j^{C,U}$ represent lower and upper bounds on the dihedral angles $\phi_i, \psi_i, \omega_i, \chi_i^k, \theta_j^N, \theta_j^C$. The α represent nonnegative parameters which must be greater or equal to the negative one–half of the minimum eigenvalue of the Hessian of E over the defined domain. These parameters can be estimated by the solution of an optimization problem or by using the concept of the measure of a matrix (Adjiman et al., 1997b,c; Adjiman and Floudas, 1996; Maranas and Floudas, 1994a). The overall effect of these terms is to overpower the nonconvexities of the original nonconvex terms by adding the value of 2α to the eigenvalues of the Hessian of E. The convex lower bounding functions, L, possesses a number of important properties which guarantee global convergence (Maranas and Floudas, 1994b):

(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 an ϵ_f feasibility and ϵ_c convergence 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. The default partitioning strategy used in the algorithm 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 and selecting it to be the minimum over all the previously recorded upper bounds. Obviously, if the single minimum of E 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).

The computational requirement of the αBB algorithm depends on the number of variables (global) on which branching occurs. Therefore, these global variables need to be chosen carefully. Obviously, in a qualitative sense, the branching variables should correspond to those variables which substantially influence the nonconvexity of the surface and the location of the global minimum. With this in mind, principles have been developed to help identify the important variables (Adjiman et al., 1997c,a, 1996).

In terms of the protein folding problem, it is generally accepted that the back–bone dihedral angles (ϕ and ψ) are the most influential variables. Therefore, in larger problems involving oligopeptides, the global variable set includes only the ϕ and ψ variables. In this formulation, the dihedral angles associated with the peptide bond (ω) and the side chains (χ) are treated as local variables.

3.2 Algorithmic Description

The determination of the global minimum energy conformation, and thus the native conformation, for a given peptide using αBB requires the interfacing of several programs: αBB, PACK (Scheraga, 1996), NPSOL (Gill et al., 1986) and the potential and solvation energy modules. PACK, a peptide generation program, is called once directly by αBB in order to initialize the current problem. In subsequent steps PACK is called through NPSOL (Gill et al., 1986), a local nonlinear optimization solver used to solve both the upper and lower bounding problems. PACK internally transforms to and from Cartesian and internal coordinate systems, and provides potential energy and gradient contributions for the ECEPP/3 potential model at every step of the local minimizations. When considering surface–accessible solvation, surface–areas, and thus the JRF solvation energy, are calculated using MSEED (Perrot et al., 1992). This module is called from αBB, through PACK, once a local minimum has been found. The accessible volume shell model for solvation, RRIGS (Augspurger and Scheraga, 1996), which has been interfaced with PACK, is also called from αBB through PACK. In this case solvation energy and gradient contributions are provided at every step of the local minimizations. Finally, an additional module, UBC (Upper Bound Check), is used to verify the quality of the upper bound solutions. The overall interface is shown schematically in Figure 2.

The basic steps of the algorithm are as follows:

(1) The initial best upper bound is set to an arbitrarily large value (e.g., $+\infty$). The original domain is partitioned (e.g., bisection) along one of the global variables.

(2) A convex function (L) is constructed in each hyper–rectangle and minimized using NPSOL, with calls (through PACK) to both ECEPP/3 and one of the two solvation modules. For the accessible volume shell model, both ECEPP/3 and RRIGS energy and gradient contributions are provided at every step of the local minimizations. In the case of surface–accessible solvation, the MSEED hydration energy is added only at the corresponding minima. If a solution is greater than the best upper bound the entire subregion can be fathomed, otherwise the solution is stored.

(3) The local minima solutions for L are used as initial starting points for local minimizations of the upper bounding function (E) in each hyper–rectangle. Again, the appropriate calls are made to PACK and the potential and solvation energy modules. In solving the upper bounding problems, all variable bounds are expanded to [-180,180]. These solutions are upper bounds on the global minimum solution in each hyper–rectangle.

(4) The current best upper bound is updated to be the minimum of those thus far stored. If a new upper bound (from step 3) is selected, the upper bound check, UBC, module is called. UBC checks that the absolute value of each gradient in the objective function gradient vector is below a specified tolerance (kcal/mol/deg). If a gradient does not satisfy this check the corresponding variable bounds are incrementally increased and the problem is resolved with the previous point used as the initial starting point. This process is repeated until the gradient constraints are satisfied or an iteration limit is exceeded. UBC also employs algorithms to calculate the second derivative matrix

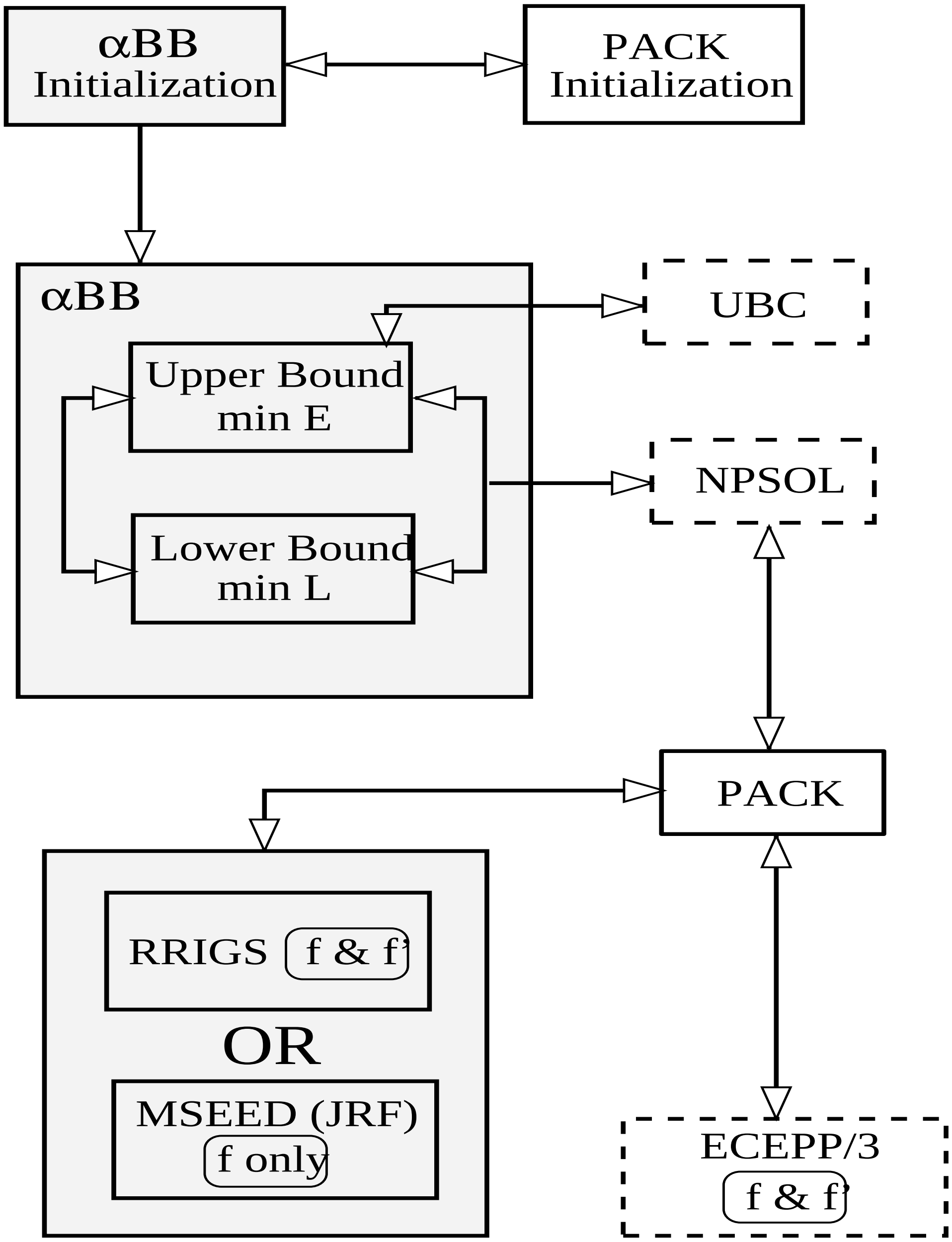

Figure 2: Interface for global optimization

(Noguti and Go, 1983) which is used to verify that the upper bound solution is a local minimum, that is the Hessian matrix is positive semi–definite. If the matrix is not positive semi–definite or the gradient checks are not satisfied, the upper bound solution is rejected.

(5) The hyper–rectangle with the current minimum value for L is selected and partitioned along one of the global variables.

(6) If the best upper and lower bounds are within ϵ the program will terminate, otherwise it will return to Step 2.

3.3 Probability–Based Partitioning

In the original problem formulation the dihedral angles were allowed to vary over the entire $[-\pi, \pi]$ domain. However, as the issue of unsolvated oligopeptide conformations was addressed, it was found that the problem required more intensive computational effort (Androulakis et al., 1997). Therefore, a reduction of the domain space was proposed based on dihedral angle distributions. Obviously, for the algorithm to be successful these reductions must not exclude the region of the global minimum conformation.

The analysis of ϕ,ψ space of amino acids was first proposed by (Ramachandran and Saisekharan, 1968). Using a hard–sphere potential model, plots of allowable ϕ,ψ space exhibit similar patterns for all naturally occurring amino acids. Similar results were obtained when considering low energy amino acid conformations using the ECEPP/3 force–field (Vásquez et al., 1983). In order to extend this analysis to polypeptide conformations, ϕ,ψ plots were also obtained using experimentally obtained conformational data of polypeptides (Lambert and Scheraga, 1989). The results were similar to the original Ramachandran plots, but the plots also identified reduced subdomains which in turn identified specific patterns of polypeptide configurations.

A similar approach was followed when defining reduced subdomains for the initialization of the αBB algorithm (Androulakis et al., 1997). Specifically, an analysis of 98 proteins from the Brookhaven X–ray data bank provided dihedral angle distributions in the form of histograms from $-\pi$ to π for each dihedral angle of each of the naturally occurring amino acids. Based on these one–dimensional distributions, a number of reduced multi–dimensional domains were identified.

Table 3: Bounds on dihedral angles. 1: The ϕ value for the down puckering of proline that ECEPP/3 uses is -68.8

Res	ϕ	ψ	ω	χ_1	χ_2	χ_3	χ_4
Ala	-180,-50	-75,-25	160, 200	-180, 180			
	-180,-50	50,175	160, 200	-180, 180			
Gly	-180,-30	-180,0	160,200				
	30,180	0,180					
Leu	-180,-50	-75,50	160,200	-180,180	-180,180	-180,180	-180,180
		50,175					
Tyr	-180,0	-75,50	160,200	-180,180	-180,180	-180,180	
		50,175					
Met	-180,-50	-75,50	160,200	-180,180	-180,180	-180,180	-180,180
		50,175					
Phe	-180,-50	-75,50	160,200	-180,180	-180,180		
		50,175					
Pro	-68.8[1]	-75,0	160, 200				
		150,200	160, 200				

Based on this procedure, a set of reduced domains can be defined for every dihedral angle of every residue in the oligopeptide sequence. The original optimization problem can be reformulated as:

$$\min \quad E(\phi_i, \psi_i, \omega_i, \chi_i^k, \theta_j^N, \theta_j^C) \tag{8}$$

$$\begin{array}{rrcl}
\text{subject to} & \phi_i & \in & \Phi_{i\phi},\ i_\phi = 1, \ldots, N_{i\phi},\ i = 1, \ldots, N_{RES} \\
& \psi_i & \in & \Psi_{i\psi},\ i_\psi = 1, \ldots, N_{i\psi},\ i = 1, \ldots, N_{RES} \\
& \omega_i & \in & \Omega_{i\omega},\ i_\omega = 1, \ldots, N_{i\omega},\ i = 1, \ldots, N_{RES} \\
& \chi_i^k & \in & X_{i\chi}^k,\ i_\chi = 1, \ldots, N_{i\chi}^k,\ i = 1, \ldots, N_{RES},\ k = 1, \ldots, K^i \\
& \theta_j^N & \in & \Theta_{j_\theta^N},\ j_\theta^N = 1, \ldots, N_{\theta N}^j,\ j = 1, \ldots, J_N \\
& \theta_j^C & \in & \Theta_{j_\theta^C},\ j_\theta^C = 1, \ldots, N_{\theta C}^j,\ j = 1, \ldots, J_C
\end{array}$$

Here $\Phi_{i\phi}, \Psi_{i\psi}, \Omega_{i\omega}, X_{i\chi}^k$ define the reduced subdomains of each dihedral angle for residue i. The number of subdomains are indicated by $N_{i\phi}, N_{i\psi}, N_{i\omega}, N_{i\chi}^k$. The allowable dihedral angles for each residue must belong to one of the following domains:

$$\begin{aligned}
D_{ji} &= \Phi_{i_\phi} \times \Psi_{i_\psi} \times \Omega_{i_\omega} \times X_{i_\chi}^k \\
ji &= 1, \ldots, N_i \\
N_i &= |\Phi_{i\phi}||\Psi_{i\psi}||\Omega_{i\omega}||X_{i\chi}^k|
\end{aligned} \tag{9}$$

Using these definitions, the total number of initial domains is given by:

$$N = (\prod_{i=1}^{N_{RES}} N_i)|\Theta_{j_\theta^N}||\Theta_{j_\theta^C}| \tag{10}$$

These domains correspond to the Cartesian products of all the sub–domains $D_{ji},\ ji = 1, \ldots, N_i$. The reduced domains for the residues in the studied oligopeptides, namely met–enkephalin, leu–enkephalin, Ac–Ala$_4$–Pro–NHMe, and decaglycine are defined in Table 3. For example, the partitioning for met–enkephalin results in 128 subdomains. Each of these domains is included in the αBB implementation.

This approach maintains the guarantee of global optimality over the considered search space of the reduced domains, and is deterministic in those subdomains that possess convex underestimators. In addition, all variable bounds are expanded to the [-180,180] when solving the upper bounding problem. Although the initial point of an upper bounding minimization is restricted to the search space of the corresponding lower bounding problem, the solution may lie outside the original subdomain.

4 Computational Studies

The proposed approach was first tested on the set of 20 uncharged, naturally occurring residues. A number of oligopeptides, including Ac–Ala$_4$–Pro–NHMe, met–enkephalin, leu–enkephalin, and decaglycine, were then tested using the partitioning scheme of Section 3.3. In these cases, the effects of both hydration models are reported.

4.1 Terminally–Blocked Residues

The single residue examples were defined as terminally blocked by using acetyl (amino) and methyl (carboxyl) end groups. All dihedral angles were treated as global variables, excluding the three θ angles of the end groups. The relative convergence was set to 10^{-2}, and the computational requirements are reported in seconds for a HP-C110. The results for MSEED and RRIGS are summarized in Tables 4 and 5, respectively.

A comparison of computational efficiency indicates that the number of required iterations and overall computational times are generally similar for both models. One would expect the overall computational effort for the RRIGS model to be greater than for the MSEED model because function and gradient evaluations are used at each step of local minimization. However, because the MSEED solvation energy is added only at local minima, the UBC routine is performed for all upper bound solutions within 10 kcal/mol of the current best upper bound. In the RRIGS model the UBC is performed only for those upper bounds that are new candidates for the best upper bound. This results in a increase in the average CPU time required for each iteration, especially for the smaller residues. As the residues become larger, and the number of total iterations increase, the computational effort (overall CPU times and CPU/iter) of the two methods are similar.

For a number of residues, the MSEED global minimum solutions listed in Table 4 possess ω angles in the range of [-30,30] with the corresponding ϕ and ψ angles near the [-150,80] region. Additional results in which the ω angles were constrained to the range of [160,200], are presented in Table 6. In all cases, with the exception of serine, this constraint led to

Table 4: Global minimum energies of terminally blocked peptides using the MSEED solvation model. The amino end group is specified as N–Acetyl–amino; the carboxyl end group is specified as Carboxyl–CONHCH$_3$. The total energy, E$_{TOT}$, is provided along with the contributions from hydration, E$_{HYD}$, nonbonded interactions (including hydrogen bonding), E$_{NB}$, electrostatic interactions, E$_{ES}$, and torsion, E$_{TOR}$.

Residue	**# DA**	**E$_{TOT}$**	**E$_{HYD}$**	**E$_{NB}$**	**E$_{ES}$**	**E$_{TOR}$**	**Iter**	**CPU**
Pro	5	28.48	47.98	-4.48	-15.36	0.34	22	8.0
Gly	6	15.99	14.29	3.01	-1.54	0.23	37	9.1
Ala	7	29.71	24.81	2.36	-0.24	2.78	82	23.7
Cys	7	0.26	-4.22	2.51	-0.21	2.18	105	32.6
His	8	-50.22	-42.99	-6.87	-0.46	0.10	109	60.0
Phe	8	-83.47	-86.53	0.12	-0.82	3.76	155	93.8
Ser	8	-5.72	-9.62	2.68	-1.39	2.61	315	98.4
Trp	8	-105.88	-98.00	-7.91	0.03	0.00	193	145.5
Asn	9	-20.76	-20.86	8.13	-16.55	8.52	412	170.1
Asp	9	-41.14	-31.91	2.31	-12.95	1.41	342	133.8
Thr	9	6.56	2.82	0.31	-2.16	5.59	387	173.0
Tyr	9	-102.43	-105.52	1.02	-1.43	3.50	425	254.7
Val	9	46.54	39.84	2.53	-0.86	5.03	505	221.0
Gln	10	-13.89	-6.55	1.93	-12.69	3.42	437	237.0
Glu	10	-33.55	-19.61	-5.18	-8.93	0.17	402	230.4
Ile	10	53.61	56.15	-2.80	-0.52	0.78	458	279.7
Leu	10	47.62	29.61	8.37	-0.54	10.18	526	301.5
Met	10	26.33	21.35	2.10	-1.61	4.49	457	251.0
Lys	11	26.65	22.85	0.40	-1.45	4.85	534	379.1
Arg	13	-34.88	-4.57	-6.12	-24.39	0.20	535	473.9

Table 5: Global minimum energies of terminally blocked peptides using the RRIGS solvation model. The amino end group is specified as N–Acetyl–amino; the carboxyl end group is specified as Carboxyl–CONHCH$_3$. The total energy, E$_{TOT}$, is provided along with the contributions from hydration, E$_{HYD}$, nonbonded interactions (including hydrogen bonding), E$_{NB}$, electrostatic interactions, E$_{ES}$, and torsion, E$_{TOR}$.

Residue	**# DA**	**E**$_{TOT}$	**E**$_{HYD}$	**E**$_{NB}$	**E**$_{ES}$	**E**$_{TOR}$	**Iter**	**CPU**
Pro	5	-32.76	-12.96	-3.73	-16.47	0.40	16	2.8
Gly	6	-22.46	-16.14	-3.71	-2.62	0.01	73	8.8
Ala	7	-20.82	-15.64	-3.92	-1.28	0.02	124	19.4
Cys	7	-23.51	-17.67	-4.66	-1.21	0.03	143	26.3
His	8	-34.47	-25.57	-6.78	-2.21	0.09	183	59,9
Phe	8	-24.72	-16.55	-7.23	-0.94	0.00	248	108.9
Ser	8	-28.32	-20.47	-5.40	-2.49	0.04	253	58.1
Trp	8	-31.48	-21.92	-8.99	-0.59	0.02	239	112.5
Asn	9	-49.07	-26.47	-5.16	-17.47	0.03	241	78.5
Asp	9	-39.96	-20.94	-6.29	-12.74	0.01	298	96.23
Thr	9	-29.18	-19.59	-5.74	-4.19	0.34	324	116.0
Tyr	9	-30.11	-21.90	-6.65	-1.57	0.01	297	133.9
Val	9	-18.92	-14.74	-3.11	-1.16	0.09	271	109.9
Gln	10	-46.49	-27.70	-5.38	-13.49	0.08	373	173.1
Glu	10	-36.11	-20.92	-5.42	-9.85	0.08	337	138.5
Ile	10	-17.11	-14.57	-2.80	-0.52	0.78	345	162.0
Leu	10	-20.22	-14.53	-4.16	-1.88	0.35	530	223.5
Met	10	-23.93	-17.02	-4.62	-2.40	0.11	510	210.3
Lys	11	-28.15	-20.17	-5.91	-2.17	0.10	572	365.9
Arg	13	-63.84	-32.38	-6.21	-25.36	0.11	602	544.8

Table 6: Local minimum energies of terminally blocked peptides using the MSEED solvation model with constrained ω bounds [160,200]. The amino end group is specified as N–Acetyl–amino; the carboxyl end group is specified as Carboxyl–CONHCH$_3$. The total energy, E$_{TOT}$, is provided along with the contributions from hydration, E$_{HYD}$, nonbonded interactions (including hydrogen bonding), E$_{NB}$, electrostatic interactions, E$_{ES}$, and torsion, E$_{TOR}$.

Residue	**# DA**	**E**$_{TOT}$	**E**$_{HYD}$	**E**$_{NB}$	**E**$_{ES}$	**E**$_{TOR}$
Ala	7	32.97	37.44	-4.00	-0.47	0.00
Cys	7	2.34	7.27	-5.17	0.24	0.00
Ser	8	-4.75	-11.02	0.69	-0.28	5.86
Asn	9	-19.13	3.43	-6.13	-16.43	0.00
Asp	9	-39.35	-20.76	-6.13	-12.47	0.01
Val	9	46.71	50.62	-3.26	-0.77	0.12
Gln	10	-13.51	4.20	-5.17	-12.74	0.20
Leu	10	49.68	41.57	1.58	-0.66	7.19
Met	10	27.04	32.32	-4.41	-1.65	0.78
Lys	11	26.96	33.71	-5.51	-1.34	0.10

increases in solvation energies and decreases in potential energy terms while the structures became either β–sheet–like or α–helical. Without exception the ω angles for the RRIGS global minimum energy solutions in Table 5 were within the [160,200] range. The remaining analysis in this section refers to the constrained (ω) minima. This is appropriate not only in comparing the MSEED and RRIGS results, but it also makes the analysis relevant for the oligopeptide studies because the same ω bounds are used.

The disparate results of the two solvation models are more clearly evaluated in Table 7. ΔE^{POT} refers to the change in potential energy of the MSEED and RRIGS global minimum solutions. This difference is positive in all cases, which indicates that the potential energy of the RRIGS structure is always lower and provides more stabilization at the global minimum solution. For five peptides, namely phenylalanine, serine, threonine, tyrosine and leucine, the MSEED potential energy is more than 10 kcal/mole less stabilizing. The response of ΔE^{HYD}, which refers to the change in hydration energy of the MSEED and RRIGS global minimum solutions, is more varied. This difference is also positive for most examples, which again indicates that the hydration energy of the RRIGS structure is lower. However, ΔE^{HYD} is negative for four examples, namely histidine, phenylalanine, tryptophan and tyrosine. Excluding the special case of proline, these four residues correspond to the naturally occurring residues which possess ringed side chain structures. The indole, aromatic side chain of tryptophan is the largest side chain of proteins, and this residue provides the second most negative value (-76.08 kcal/mole) for ΔE^{HYD}. The other two aromatic residues, tyrosine and phenylalanine, have ΔE^{HYD} of -83.62 and -69.98 kcal/mole, respectively. The imidazole ring of histidine provides the least negative ΔE^{HYD} of the four residues, with a value of -17.42 kcal/mole. Other trends are also apparent. The most positive ΔE^{HYD} values are provided by the aliphatic residues; that is, isoleucine, valine, leucine and alanine have values of 70.72, 65.36, 56.10 and 53.08 kcal/mole respectively. The acidic residues, glutamic and aspartic acid, have comparable values of ΔE^{HYD} (1.31 and 0.18 kcal/mole, respectively). In addition, the ΔE^{HYD} for the amide forms of these residues, glutamine and asparagine, are also comparable (31.90 and 29.90 kcal/mole, respectively).

A more detailed analysis was performed by calculating adiabatically relaxed ϕ–ψ maps. The maps for N–acetyl–N'–methyl–alanineamide are shown in Figures 3–5. The adiabatic curves define regions within a given energy of the global minimum value. The first map corresponds to an adiabatically relaxed map for the unsolvated peptides. This was calculated by fixing the ϕ and ψ angles at 3 degree increments and using the NPSOL local minimization

Table 7: Comparison of unsolvated and hydration components for MSEED and RRIGS global minimum solutions of terminally blocked peptides. $\Delta E^{POT} = E^{POT}_{MSEED}$ - E^{POT}_{RRIGS} and $\Delta E^{HYD} = E^{HYD}_{MSEED}$ - E^{HYD}_{RRIGS}, at the corresponding global minimum solutions.

Residue	$\mathbf{\Delta E}^{POT}$	$\mathbf{\Delta E}^{HYD}$
Pro	0.30	60.94
Gly	8.02	30.43
Ala	0.71	53.08
Cys	0.91	24.94
His	1.67	-17.42
Phe	11.23	-69.98
Ser	14.12	9.45
Trp	1.68	-76.08
Asn	0.04	29.90
Asp	0.43	0.18
Thr	13.33	22.41
Tyr	11.30	-83.62
Val	0.27	65.36
Gln	1.08	31.90
Glu	1.25	1.31
Ile	0.00	70.72
Leu	13.80	56.10
Met	1.63	49.34
Lys	1.23	53.88
Arg	1.15	27.81

Table 8: Approximate dihedral angles and nomenclature for ϕ–ψ regions.

Conformer	ϕ,ψ	**Protein structure**
C_5	-150, 150	β–sheet
P_{II}	-80, 150	polyproline II
C_7	-80, 80	γ–turn
α_R	-80, -50	α–helix (right)
α_L	80, 50	α–helix (left)

solver to minimize the ECEPP/3 potential energy by varying the remaining dihedral angles. The second map was constructed by a similar procedure, although the minimized energy now included both ECEPP/3 and the hydration free energy of the RRIGS model. In the third case, the ECEPP/3 energy was first minimized in the absence of solvent at each point and the map was generated by adding the solvation free energy of MSEED (JRF parameters) for the minimized conformation.

Comparison of these maps reveals several important effects of including solvation (notation for the peptide regions are given in Table 8). Experimental data for this peptide suggests that more than one conformation is present in solution, and NMR coupling constants indicate a large population of conformations with $-70 > \phi > -80$ (Madison and Kopple, 1980). It is also expected that hydration weakens intrapeptide hydrogen bonding. The unsolvated map, shown in Figure 3, indicates well defined regions for intramolecular hydrogen bonding (C_7) and for right–handed α–helices (α_R). The global minimum occurs within the C_7 region. The RRIGS map (Figure 4) is strongly dominated by the ECEPP/3 potential, with the global minimum in the C_7 region and a very strong α_R region. However, there is a broadening of the β–sheet (C_5) region as well as a less distinct C_7 minimum. The MSEED map (Figure 5) shows a considerable shift away from the C_7 minimum towards the C_5 region, which contains the global minimum. However, there is also a decreased dominance of the α_R region, which contradicts the prediction of NMR coupling constants. This is also evidenced by the ϕ–ψ distribution of the global minimum energy structures for all terminally blocked peptides, which is shown in Figure 6.

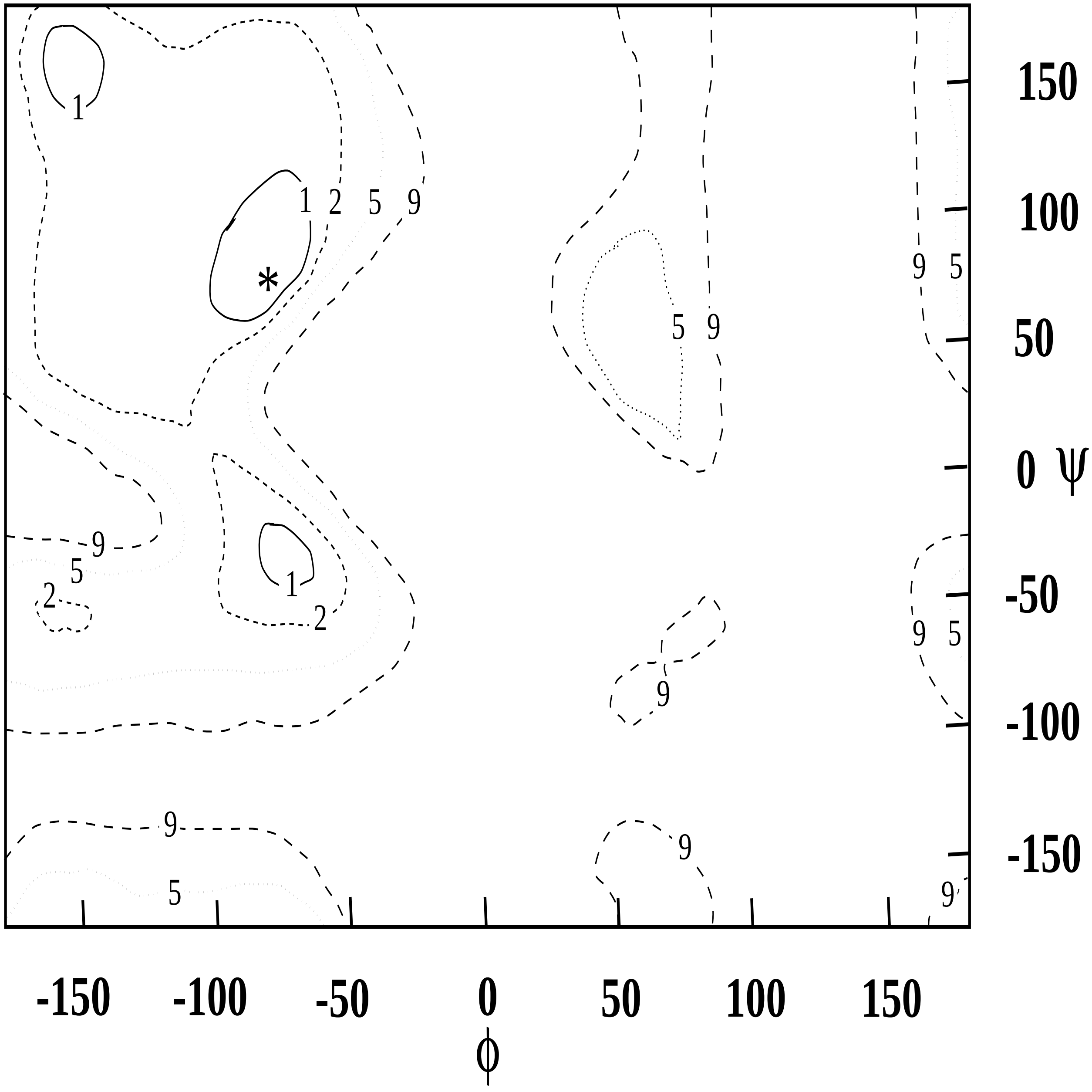

Figure 3: Adiabatic ϕ–ψ map for unsolvated N–acetyl–N'–methyl–alanineamide. The adiabatic curves define regions within a given energy (1, 2, 5, 9 kcal/mole) of the global minimum value, and the (*) represents the location of the global minimum.

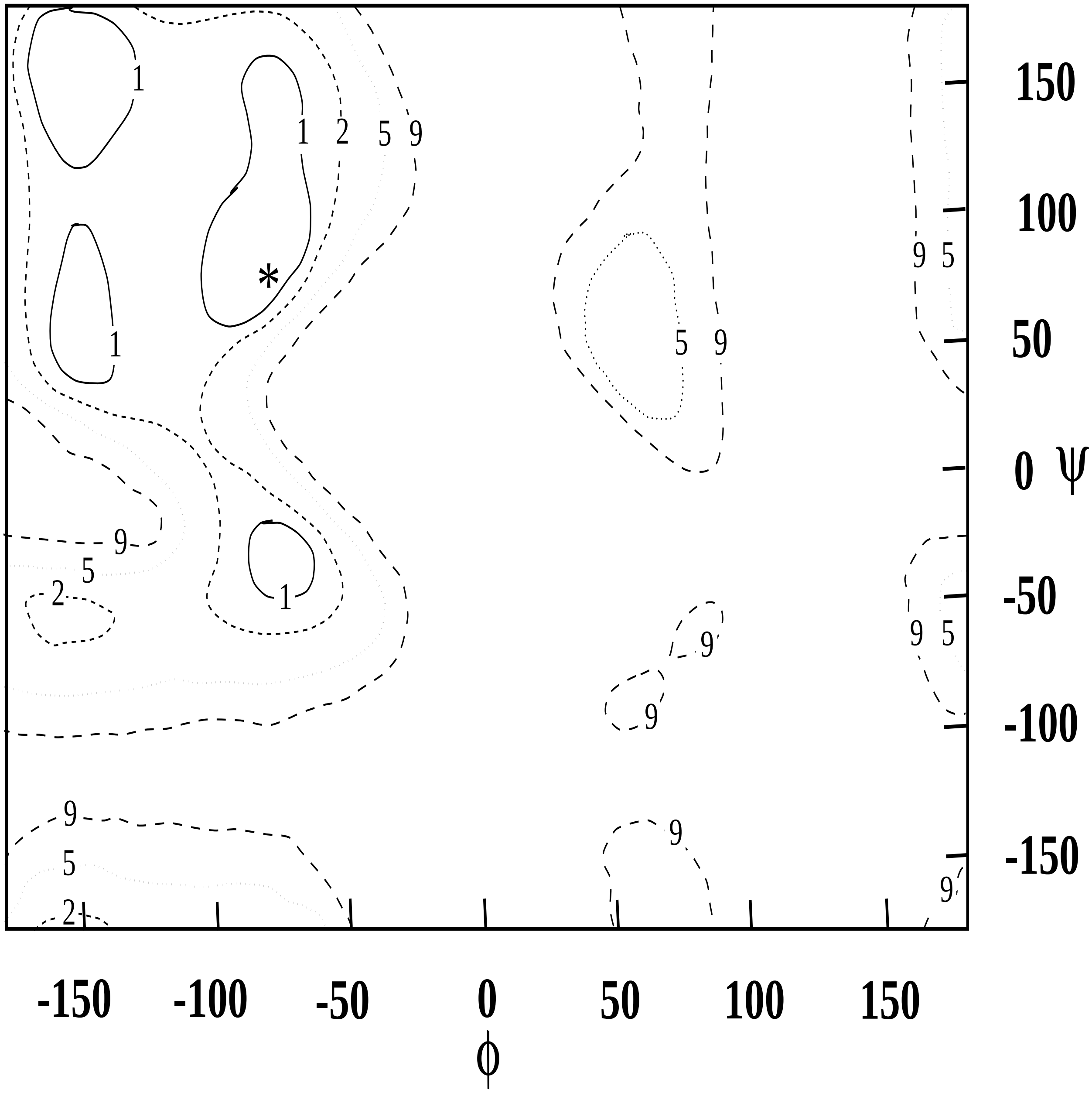

Figure 4: Adiabatic ϕ–ψ map for solvated N–acetyl–N'–methyl–alanineamide, using the RRIGS solvation model. The adiabatic curves define regions within a given energy (1, 2, 5, 9 kcal/mole) of the global minimum value, and the (*) represents the location of the global minimum.

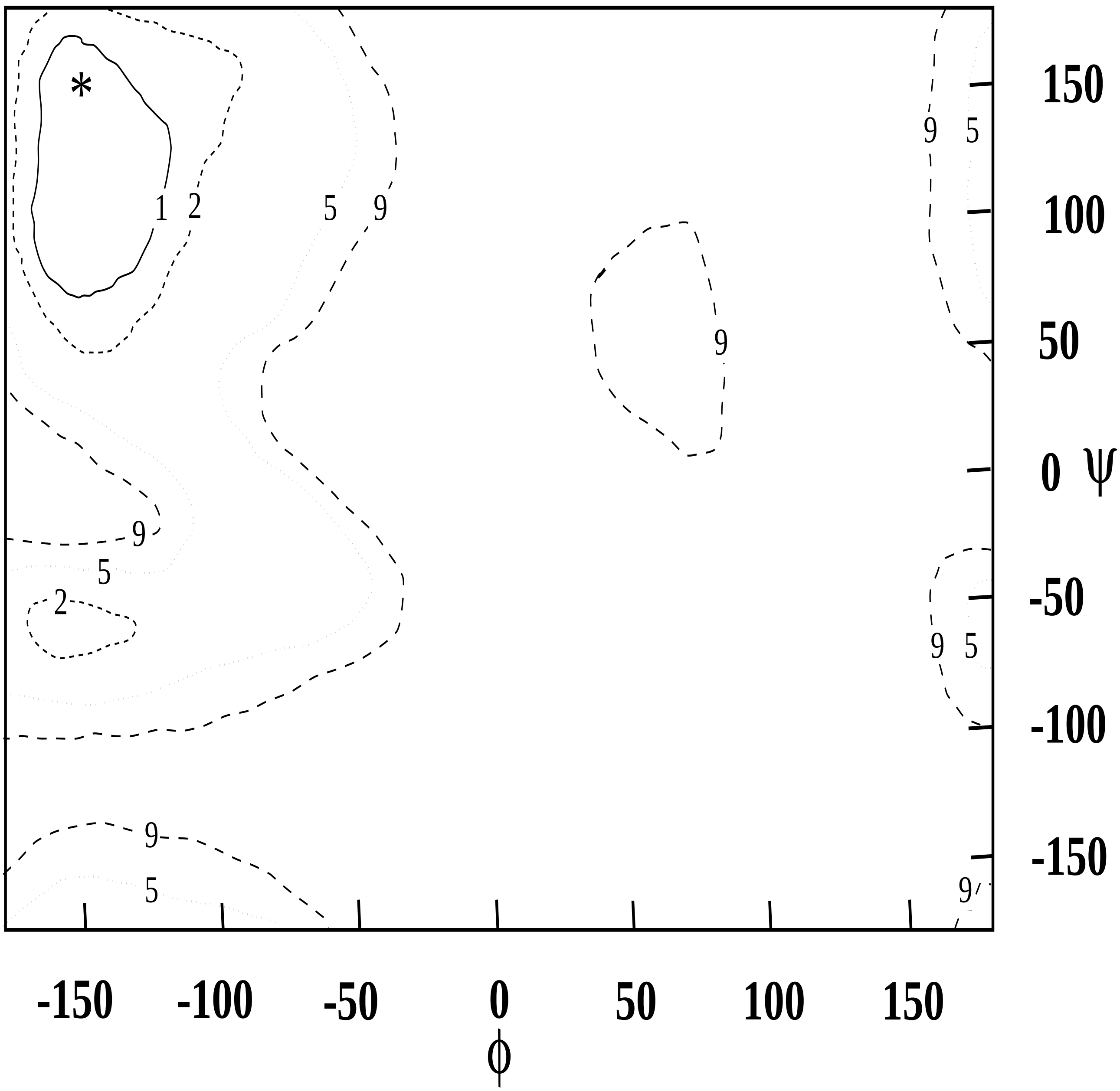

Figure 5: Adiabatic ϕ–ψ map for solvated N–acetyl–N'–methyl–alanineamide, using the MSEED solvation model. The adiabatic curves define regions within a given energy (1, 2, 5, 9 kcal/mole) of the global minimum value, and the (*) represents the location of the global minimum.

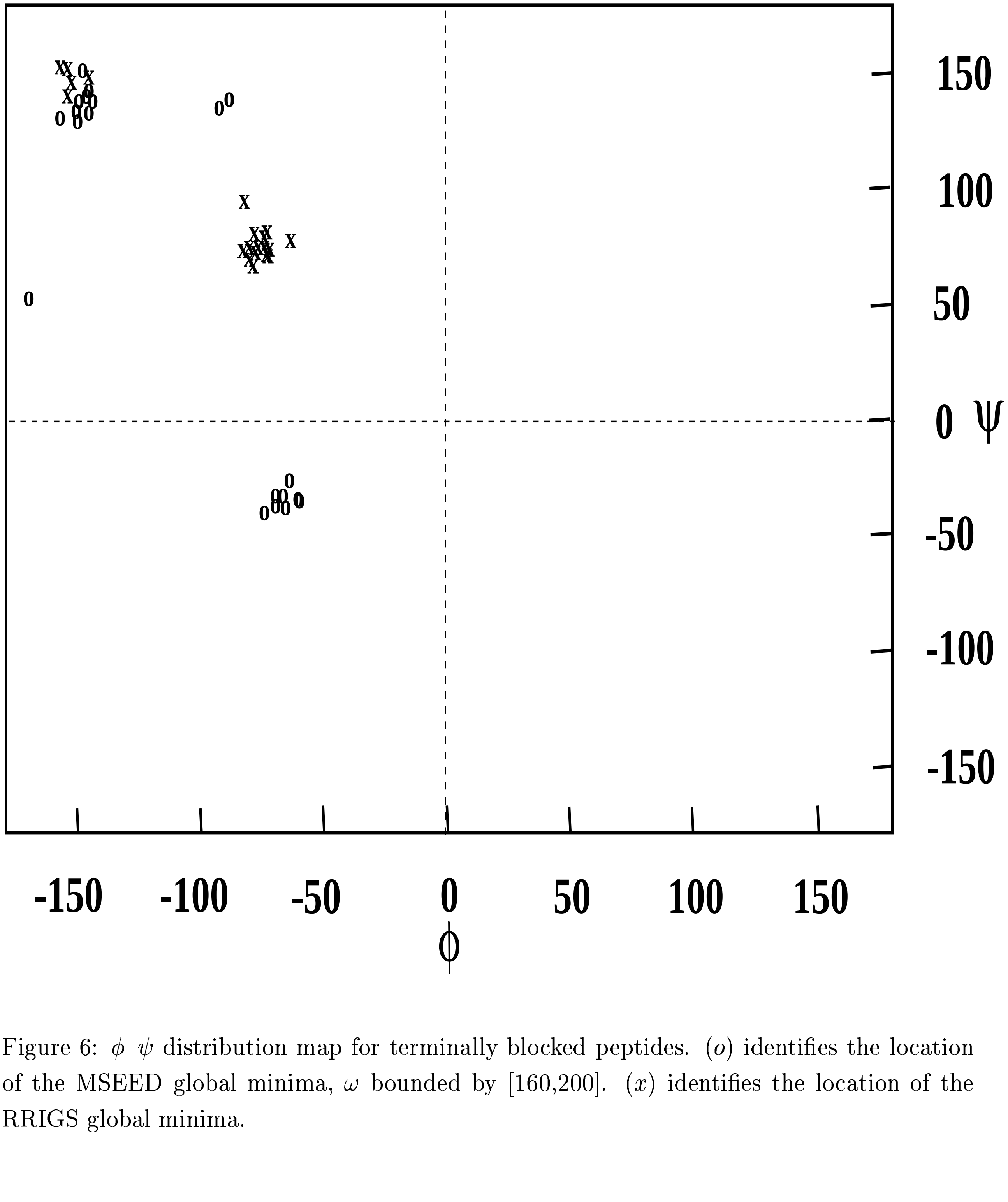

Figure 6: ϕ–ψ distribution map for terminally blocked peptides. (o) identifies the location of the MSEED global minima, ω bounded by [160,200]. (x) identifies the location of the RRIGS global minima.

Table 9: Global minimum for Ac–Ala_4–Pro–NHMe using the MSEED model for hydration.

	ϕ	ψ	ω	χ_1
Ac	62.52	179.13		
Ala	-158.45	-56.49	173.38	51.70
Ala	-155.49	-67.74	179.78	49.37
Ala	-164.24	-64.18	180.10	49.42
Ala	-63.63	-59.04	172.51	32.77
Pro	-68.80	-23.79	177.30	
NHMe	54.50			

4.2 Oligopeptides

4.2.1 N–Acetyl–N'–methylamide of Ala_4–Pro

This example, which was first proposed for evaluating the performance of the ECEPP/3 force field, involves 5 residues and 21 dihedral angles. The peptide was modeled using the acetyl terminal group (CH_3CO–) on the N–terminus and the methylamide group (–$NHCH_3$) on the C–terminus. As described in Section 3.3 single residue patterns were considered in partitioning the domain space, which resulted in an initial partitioning of 32 domains.

Using MSEED, a global minimum energy of 48.23 kcal/mole was located after 592 iterations and 2,503 seconds (HP-C110). The structure, defined by the GGGAA conformational code (Zimmerman et al., 1977), possessed no hydrogen bonds. That is, no potential hydrogen bonding pairs were closer than 2.3 $\AA$. Table 9 summarizes the values of the 21 dihedral angles. A plot of the global minimum structure is given in Figure 7.

The RRIGS model predicted a partially right–handed α–helical conformation, with the first three alanine residues within the A region of the ϕ-ψ map (Zimmerman et al., 1977). It should be noted that the hydration free energy was more stabilizing (as compared to the MSEED predictions), resulting in a contribution of -33.03 kcal/mole to the overall global minimum energy of -51.88 kcal/mole. In addition, nonbonded interactions at the global minimum provided an additional 8 kcal/mole of stabilization, when compared to MSEED contributions at its corresponding global minimum. This is in part due to the formation of a strong hydrogen bond (1.9 $\AA$) between the CO of the acetyl end group and the NH proton

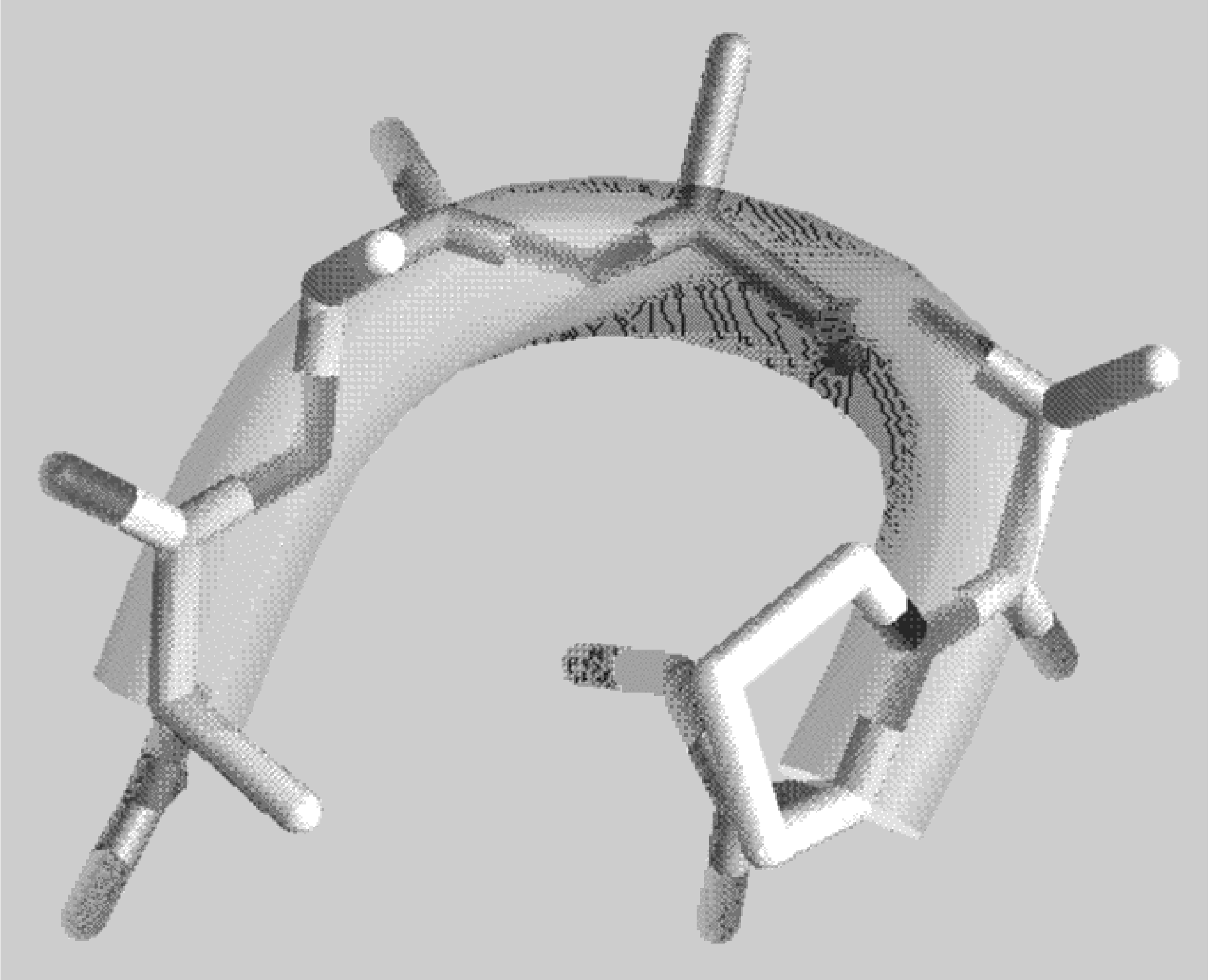

Figure 7: Plot of Ac–Ala$_4$–Pro–NHMe conformation. Global minimum energy of 48.23 kcal/mole using the MSEED model for hydration. The structure corresponds to the conformational code of GGGAA (Zimmerman et al., 1977). A C$^\alpha$ worm is used to highlight the backbone structure.

Table 10: Dihedral angles at global minimum for Ac–Ala$_4$–Pro–NHMe, using the RRIGS model for hydration.

	ϕ	ψ	ω	χ_1
Ac	-61.20	-178.23		
Ala	-71.68	-30.23	180.26	-178.39
Ala	-75.26	-32.05	184.21	61.77
Ala	-79.89	-40.23	184.95	-58.00
Ala	-135.70	71.54	176.92	59.09
Pro	-68.80	-25.20	180.10	
NHMe	-60.14			

of the fourth alanine residue. The algorithm required 623 iterations and 3,233 seconds (HP-C110) to converge to the global minimum structure plotted in Figure 8. The values of the corresponding dihedral angles are given in Table 10.

Although an experimentally derived structure does not exist for this test molecule, comparisons can be made between the two solvated and a previously determined unsolvated global minimum energy structure (Androulakis et al., 1997). When considering C$^\alpha$ carbons, the (rms) deviation between the MSEED and unsolvated structures is calculated to be 1.429. In contrast, the unsolvated structure, which also exhibits a distorted right–handed α helical structure, has a (rms) deviation of only 0.096 from the RRIGS global minimum structure. This difference is also illustrated by comparing energy contributions for the MSEED and RRIGS global minimum structures, as given in Table 11. In addition, a number of function evaluations were performed at the global solution of the other hydration model and the unsolvated global minimum. For both cases, the hydration energy is more stabilizing at the MSEED solution. Although the RRIGS solution provides more stabilizing nonbonded interactions, the change in hydration energy dominates the MSEED model. In contrast, the difference in hydration free energies is smaller when using the RRIGS model, which causes the large difference in nonbonded energies to set the global solution. The correspondence between the RRIGS and unsolvated global minimum energy conformations also become apparent through these function evaluations.

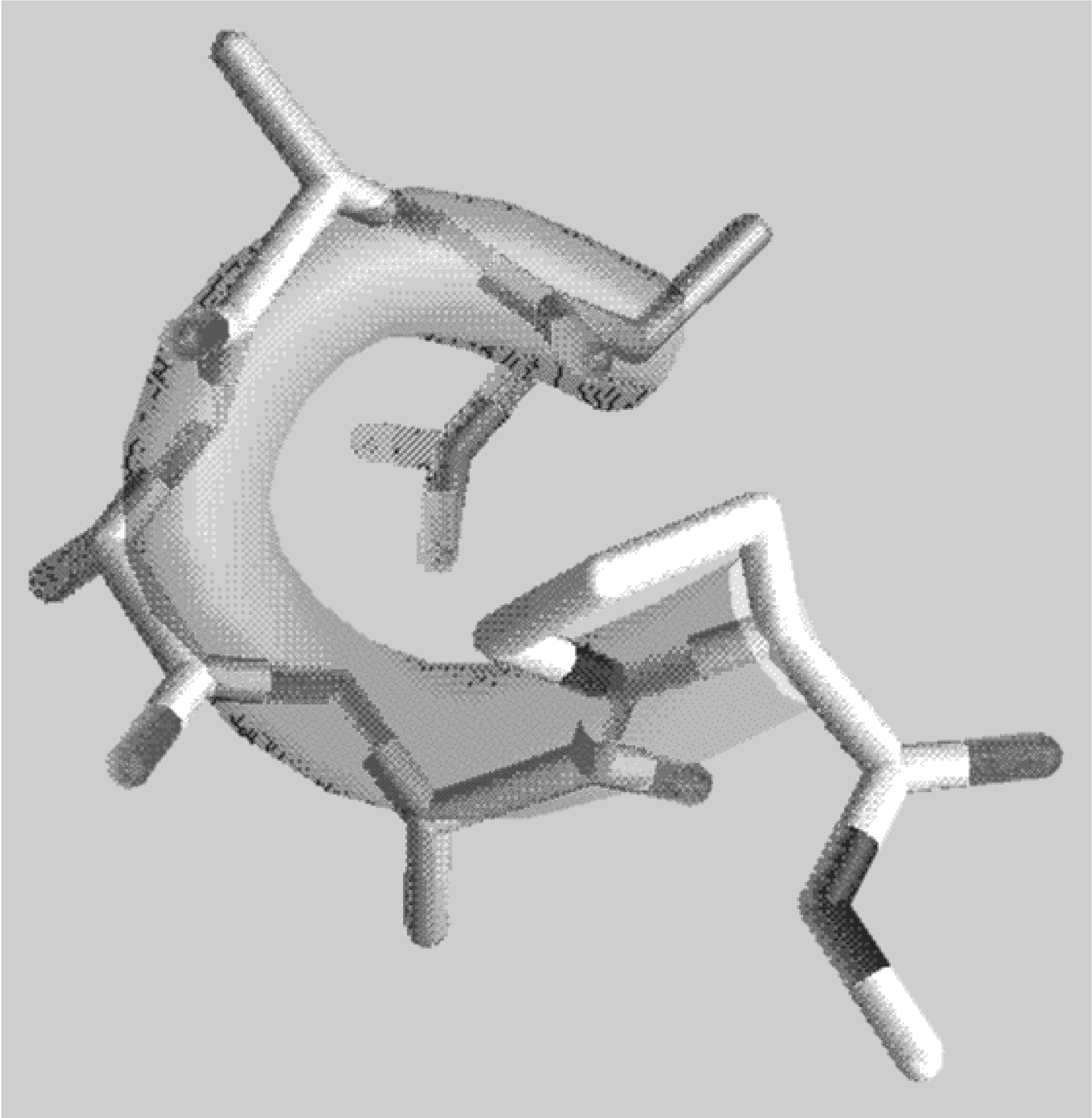

Figure 8: Plot of Ac–Ala$_4$–Pro–NHMe conformation. Global minimum energy of -51.88 kcal/mole using the RRIGS model for hydration. The structure corresponds to the conformational code of AAADA (Zimmerman et al., 1977) A C$^\alpha$ worm is used to highlight the backbone structure.

Table 11: Comparison of hydration energies for Ac–Ala_4–Pro–NHMe. The first column refers to the hydration model used in the function evaluations, which are performed at the global solutions given in the second column. The total energy, E_{TOT}, is provided along with the contributions from hydration, E_{HYD}, nonbonded interactions (including hydrogen bonding), E_{NB}, electrostatic interactions, E_{ES}, and torsion, E_{TOR}.

	Global of	$\mathbf{E}_{TOT}$	$\mathbf{E}_{HYD}$	$\mathbf{E}_{NB}$	$\mathbf{E}_{ES}$	$\mathbf{E}_{TOR}$
MSEED	MSEED	48.23	52.68	-18.53	11.41	2.67
	RRIGS	73.50	92.35	-26.78	7.24	0.69
	UNSOL	73.97	93.02	-26.70	7.01	0.64
RRIGS	RRIGS	-51.88	-33.03	-26.78	7.24	0.69
	MSEED	-39.12	-34.67	-18.53	11.41	2.67
	UNSOL	-51.69	-32.64	-26.70	7.01	0.64

4.2.2 Met–enkephalin

Met–enkephalin (H–Tyr–Gly–Gly–Phe–Met–OH) is an endogenous opioid pentapeptide found in the human brain, pituitary, and peripheral tissues. Its biological function involves a large variety of physiological processes, most notably the endogenous response to pain. The peptide consists of 24 dihedral angles and a total of 75 atoms, and has played the role of a benchmark molecular conformation problem. The energy hypersurface is extremely complex with the number of local minima estimated on the order of 10^{11} (Li and Scheraga, 1988). Based on a previous study, the unsolvated global minimum potential energy conformation was shown to exhibit a type II' β–bend along the N–C' peptidic bond of Gly^3 and Phe^4 (Androulakis et al., 1997).

Experimental results have indicated that met–enkephalin in aqueous solution does not possess an unique structure (Graham et al., 1992). In general, the experimentally determined aqueous conformations were found to exhibit characteristics of extended random–coil polypeptide with no discernible secondary structure. When considering the effects of hydration, the competition for backbone hydrogen bonding (with water), which contributes to the bending of the unsolvated conformation, should result in a more extended structure. These qualitative arguments have been confirmed by the analysis of hydrated met–enkephalin using the MSEED model. The plot of the global minimum energy structure, given in Figure

Table 12: Dihedral angles at the global minimum energy conformation of met–enkephalin, using the MSEED model for hydration.

	ϕ	ψ	ω	χ_1	χ_2	χ_3	χ_4
Tyr	-84.96	160.74	179.09	-59.83	100.80	-179.29	
Gly	-160.26	151.83	-177.53				
Gly	159.50	-157.94	178.71				
Phe	-76.55	76.23	-178.05	-61.87	108.68		
Met	-132.90	147.47	-179.83	-65.17	-175.99	-84.91	59.38

9, shows that the residues near the N–terminus are almost fully extended, although there is slight bending near the C–terminus. This bending is stabilized by the formation of 2.10 $\AA$ hydrogen bond between the CO of the second glycine residue and the NH proton of the methionine residue. In addition, the structure displays a large 17.00 $\AA$ separation between the centroids of the Phe and Tyr aromatic rings.. The values of dihedral angles corresponding to the global minimum energy of -283.76 kcal/mole are given in Table 12. Locating this solution required 1033 iterations and 5,082 seconds (HP-C110).

The RRIGS method also predicts a more extended structure than the global minimum structure reported for the unsolvated case (Androulakis et al., 1997). In fact, although a slight bend occurs near the N–terminus, the structure possesses no hydrogen bonds ($<$ 2.3 $\AA$). In addition, unlike the MSEED structure, there exists close proximity of the Tyr and Phe aromatic rings, as shown in Figure 10. The centroids of these rings are separated by 4.16 $\AA$, which is slightly closer than the preferential aromatic–aromatic interaction distance of 4.5 to 7 $\AA$ (Burley and Petsko, 1985). Furthermore, the aromatic rings are essentially in a parallel, as opposed to the more common orthogonal, orientation. This suggests an attempt to substantially reduce the hydrophobic exposure of the aromatic side chains. The global minimum conformation, with an energy of -50.01 kcal/mole, was located in 1058 iterations and 8,695 seconds (HP-C110). The values of the dihedral angles are given in Table 13.

Additional analysis was performed by calculating (rms) deviation values between both the MSEED and RRIGS global minimum structures and experimentally determined conformations. C^α positions for 5 structures of met–enkephalin in aqueous solution (Graham et al., 1992) were generated and compared to both global minima. Average values for the

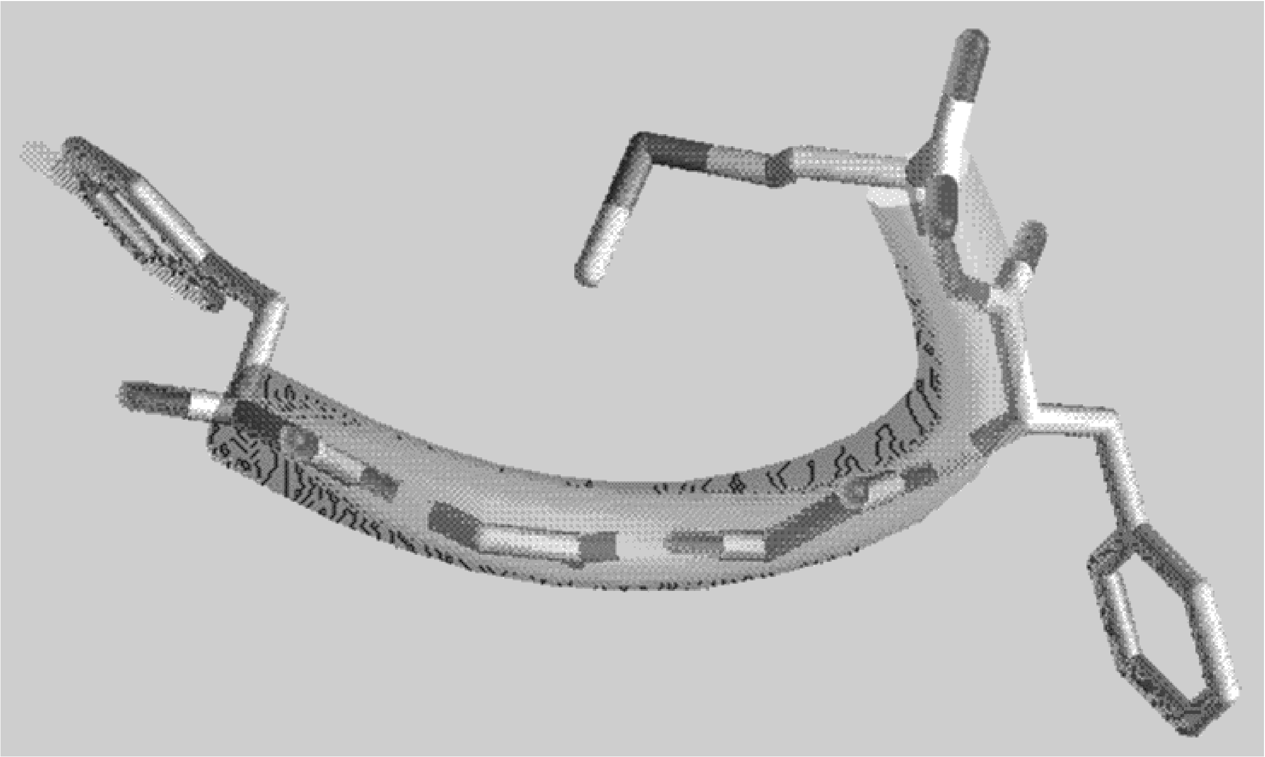

Figure 9: Plot of met–enkephalin conformation. Global minimum energy of -283.76 kcal/mole using the MSEED model for hydration. The structure corresponds to the conformational code of FEE*CE (Zimmerman et al., 1977) A C^{α} worm is used to highlight the backbone structure.

Table 13: Dihedral angles at the global minimum energy conformation of met–enkephalin, using the RRIGS model for hydration.

	ϕ	ψ	ω	χ_1	χ_2	χ_3	χ_4
Tyr	-168.32	-30.81	178.52	-173.58	-101.26	-171.75	
Gly	78.83	-86.96	182.73				
Gly	162.94	91.72	172.83				
Phe	-150.72	162.32	181.50	66.66	92.68		
Met	-77.80	106.79	181.63	-67.82	178.91	180.01	-60.01

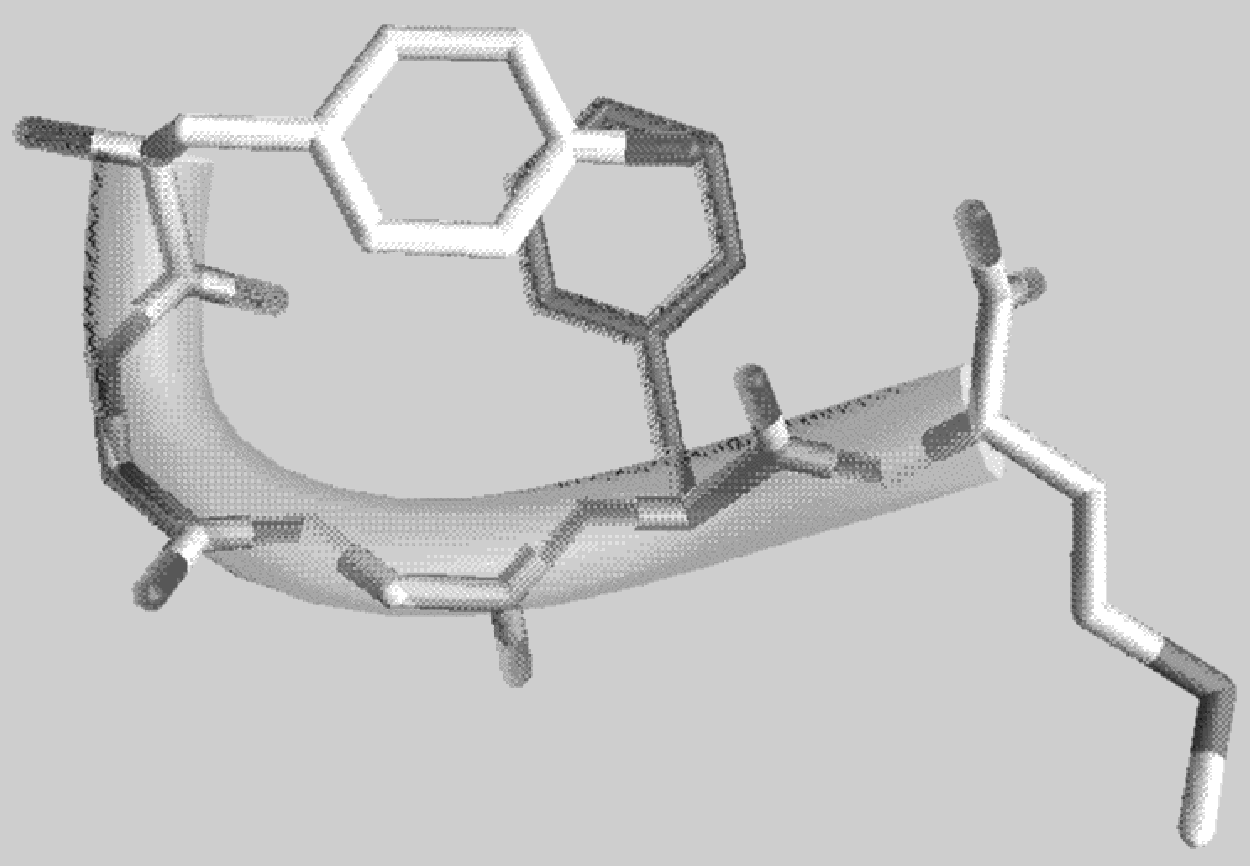

Figure 10: Plot of met–enkephalin conformation. Global minimum energy of -50.01 kcal/mole using the RRIGS model for hydration. The structure corresponds to the conformational code of BC*H*EC (Zimmerman et al., 1977) A C^{α} worm is used to highlight the backbone structure.

Table 14: Comparison of hydration energies for met–enkephalin. The first column refers to the hydration model used in the function evaluations, which are performed at the global solutions given in the second column. The total energy, E_{TOT}, is provided along with the contributions from hydration, E_{HYD}, nonbonded interactions (including hydrogen bonding), E_{NB}, electrostatic interactions, E_{ES}, and torsion, E_{TOR}.

	Global of	$\mathbf{E}_{TOT}$	$\mathbf{E}_{HYD}$	$\mathbf{E}_{NB}$	$\mathbf{E}_{ES}$	$\mathbf{E}_{TOR}$
MSEED	MSEED	-283.77	-288.83	-19.13	23.29	0.90
	RRIGS	-139.35	-130.75	-31.47	21.84	1.03
	UNSOL	-170.88	-159.17	-35.26	21.46	2.09
RRIGS	RRIGS	-50.01	-41.41	-31.47	21.84	1.03
	MSEED	-41.63	-46.69	-19.13	23.29	0.90
	UNSOL	-47.49	-35.78	-35.26	21.46	2.09

(rms) deviations from the 5 experimentally determined structures were 1.534 and 1.139 for the MSEED and RRIGS global minimum energy structures, respectively. Although these values do not reflect extremely good correspondence, the results suggest that the RRIGS structure may be more "extended". Qualitatively, this agrees with the absence of hydrogen bonding in the RRIGS structure.

It is also interesting to compare energy evaluations at corresponding global minimum solutions. This information is given in Table 14. It is apparent that the MSEED model predicts large stabilizing hydration free energies. In addition, these contributions tend to dominate the prediction of the global minimum structure. Specifically, energy evaluations at the RRIGS and unsolvated solutions indicate a substantial increase in overall energy, which can be directly correlated to the increase in hydration free energy. In contrast, this correlation does not hold for the RRIGS model. In fact, the RRIGS model, like the MSEED model, predicts a more stabilizing hydration free energy at the MSEED solution. However, nonbonded interactions are less favorable at this solution, resulting in an overall energy increase. In addition, although the solvation free energy becomes less stabilizing at the unsolvated solution, an increase in the number of favorable nonbonded interactions causes the overall energy to be near the global minimum solution.

It should also be noted that the αBB algorithm is able to inherently identify low energy

Table 15: Low energy conformers (within 0.5 kcal of global minimum energy) for RRIGS model. Total energies and conformational codes (Zimmerman et al., 1977) are given.

Conformer	**E$_{TOT}$**	**Conformational Code**
1	-49.97	BC*G*AG
2	-49.89	BC*H*EG
3	-49.67	BC*H*EB
4	-49.61	BC*H*EA
5	-49.57	BC*GEF

conformers, along with the global minimum energy conformation. Table 15 lists five local minimum energy conformations within 0.5 kcal of the RRIGS global minimum energy. The structures are related to the global minimum energy conformation as evidenced by their similar conformational codes (Zimmerman et al., 1977). Such information has important ramifications for more detailed free energy calculations, and work along these lines is currently in progress.

4.2.3 Leu–enkephalin

Like met–enkephalin, leu–enkephalin (H–Tyr–Gly–Gly–Phe–Leu–OH) is an endogenous pentapeptide in which the methionine residue has been replaced by a leucine residue. The unsolvated global minimum conformation exhibits a type II' β–bend around the Gly^3–Phe^4 backbone region (Androulakis et al., 1997). As expected, the inclusion of hydration effects, using the MSEED model, produces an extended solvated conformation. Again, as the plot in Figure 11 shows, the residues near the N–terminus are almost fully extended with a slight bend near the C–terminus. This bending is also stabilized by a hydrogen bond; in this case a 2.13 $\AA$ hydrogen bond between the CO of the second glycine residue and the NH proton of the leucine residue. The aromatic rings are also widely separated with a distance of 14.44 $\AA$ between the two centroids. The values of the 24 dihedral angles, resulting in a global minimum energy of -263.14 kcal/mole, are given in Table 16. This solution was found after 1131 iterations and 5,597 seconds (HP-C110). A plot of this structure is shown in Figure 11.

The RRIGS model also predicted a global minimum structure that was more extended

Table 16: Dihedral angles at the global minimum energy conformation of leu–enkephalin, using the MSEED model for hydration.

	ϕ	ψ	ω	χ_1	χ_2	χ_3	χ_4
Tyr	-84.88	160.00	178.30	-60.54	100.49	-179.21	
Gly	-160.78	140.99	-178.01				
Gly	144.14	-152.83	177.03				
Phe	-79.95	71.30	-176.06	-60.97	108.26		
Leu	-83.99	138.62	-179.24	-53.91	176.56	-178.84	69.81

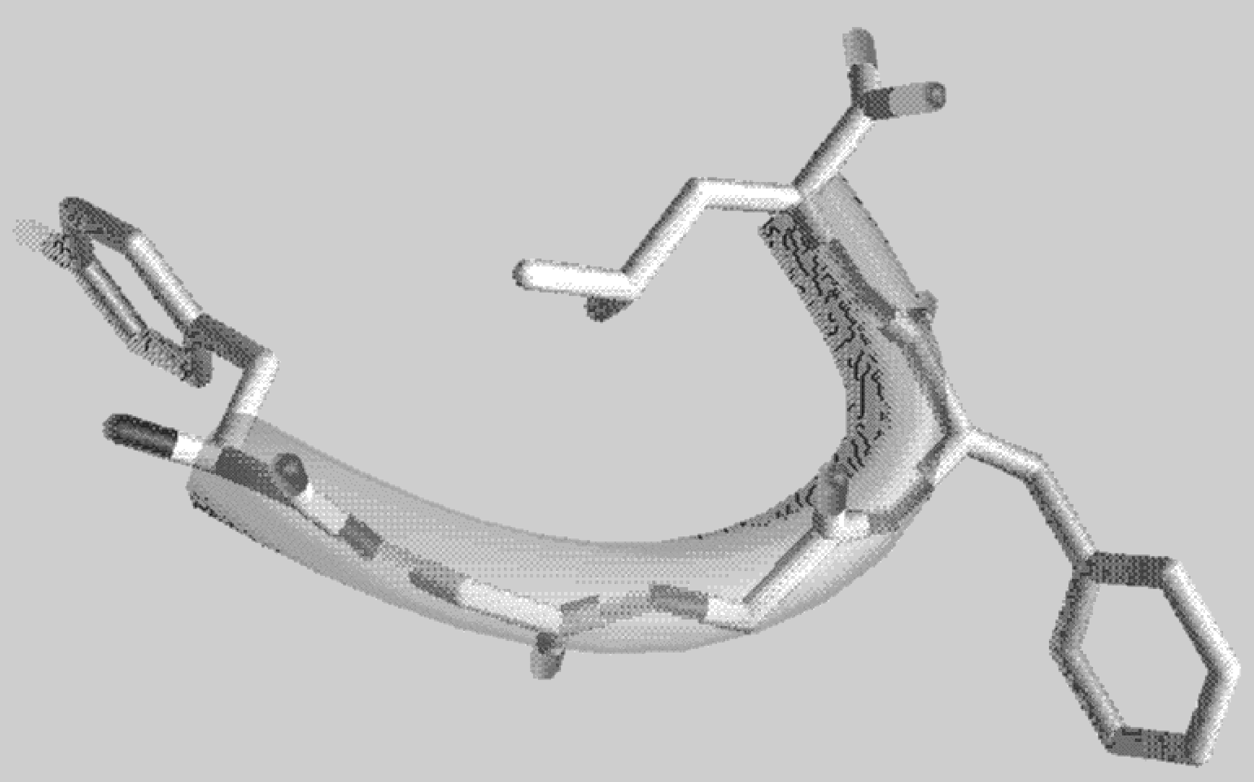

Figure 11: Plot of leu–enkephalin conformation. Global minimum energy of -263.14 kcal/mole using the MSEED model for hydration. The structure corresponds to the conformational code of FEE*CF (Zimmerman et al., 1977) A C$^\alpha$ worm is used to highlight the backbone structure.

Table 17: Dihedral angles at the global minimum energy conformation of leu–enkephalin, using the RRIGS model for hydration.

	ϕ	ψ	ω	χ_1	χ_2	χ_3	χ_4
Tyr	-168.37	-30.66	178.49	-173.40	78.69	-161.13	
Gly	78.92	-87.17	182.69				
Gly	163.20	91.51	172.72				
Phe	-150.66	161.54	181.57	66.75	-86.85		
Leu	-75.45	105.32	181.75	179.5	63.84	172.22	179.31

than the unsolvated global minimum structure. The backbone structure is essentially identical to the met–enkephalin global minimum energy conformation, with a (rms) deviation of only 0.005. As with met–enkephalin the structure exhibits a bending near the N–terminus with no significant hydrogen bonding (< 2.3 $\mathring{A}$). The centroid separation distance of the two aromatic rings is 4.15 $\mathring{A}$ with a nearly parallel orientation. This conformation, shown in Figure 12, corresponds to an energy value of -46.57 kcal/mole, and was found after 1137 iterations and 9,243 seconds (HP-C110). The corresponding values of the dihedral angles are given in Table 17.

Experimental evidence suggests that leu–enkephalin conformations also possess extended peptide backbones (Camerman et al., 1983). Using the same set of experimentally determined conformations, average C^{α} (rms) deviations were found to be 1.440 and 1.137 for the MSEED and RRIGS minima, respectively. The average MSEED deviation is slightly better than for met–enkephalin, although the RRIGS deviation remains essentially unchanged because of the similarities of the two RRIGS minimum energy structures. It should also be noted that the MSEED global minimum structures of the two enkephalins are very similar with a (rms) deviation of only 0.516.

Information on energy evaluations is provided in Table 18. The conclusions are qualitatively similar to those made for met–enkephalin. That is, the MSEED model produced hydration free energies which dominated the overall energy. In contrast, although the RRIGS model provided relatively more stabilization at the MSEED global solution, its contribution did not dominate the energy landscape because of the unfavorable nonbonded interactions. Similarly, favorable interactions at the unsolvated global minimum caused its overall energy

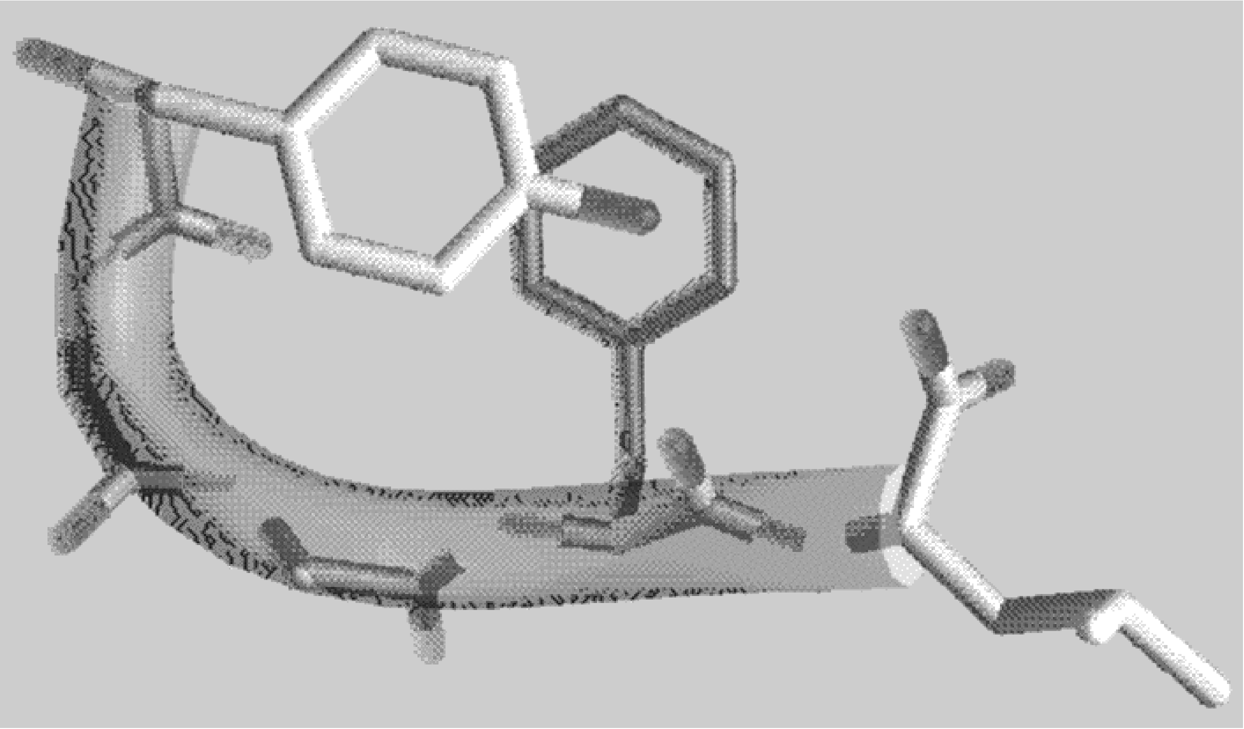

Figure 12: Plot of leu–enkephalin conformation. Global minimum energy of -46.57 kcal/mole using the RRIGS model for hydration. The structure corresponds to the conformational code of BC*H*EC (Zimmerman et al., 1977) A C^{α} worm is used to highlight the backbone structure.

Table 18: Comparison of hydration energies for leu–enkephalin. The first column refers to the hydration model used in the function evaluations, which are performed at the global solutions given in the second column. The total energy, E_{TOT}, is provided along with the contributions from hydration, E_{HYD}, nonbonded interactions (including hydrogen bonding), E_{NB}, electrostatic interactions, E_{ES}, and torsion, E_{TOR}.

	Global of	$\mathbf{E}_{TOT}$	$\mathbf{E}_{HYD}$	$\mathbf{E}_{NB}$	$\mathbf{E}_{ES}$	$\mathbf{E}_{TOR}$
MSEED	MSEED	-263.14	-268.30	-19.07	23.77	0.46
	RRIGS	-112.60	-105.03	-30.95	22.31	1.07
	UNSOL	-142.84	-133.51	-33.26	20.76	3.17
RRIGS	RRIGS	-46.57	-39.00	-30.95	22.31	1.07
	MSEED	-39.10	-44.26	-19.07	23.77	0.46
	UNSOL	-43.37	-34.04	-33.26	20.76	3.17

to be within a few kcal/mole of the global value.

4.2.4 Decaglycine

Decaglycine is a larger oligopeptide consisting of 30 dihedral angles. This peptide was first modeled using a NH_2 group at the α–amino end and a –COOH group at the α–carboxyl end. In general, energy minimizations for the unsolvated peptide indicate an enormous number of local minima which represent metastable structures corresponding to partially right–hand and left–hand α–helices. In initializing the algorithm the information in Table 3 provides 4 initial regions for each glycine residue, or a total of 10^4 initial subdomains for the entire decapeptide. This was modified to 32 initial subdomains by combining the ψ subdomains for each residue, and searching the entire ϕ space (-180,180) for every other glycine residue. A global minimum energy of -63.87 kcal/mole was located using the MSEED model . Unlike the unsolvated structure (Androulakis et al., 1997), the configuration does not exhibit a helicoidal three–dimensional pattern, but a rather extended conformation with an end–to–end C^{α} distance of 19.02 $\mathring{A}$. The global minimum energy conformation, which was located after 1331 iterations and 10,307 seconds (HP-C110), is plotted in Figure 13.

The RRIGS model was then tested on decaglycine using the acetyl terminal group (CH_3CO–) on the N–terminus and methylamide group (–$NHCH_3$) on the C–terminus. It

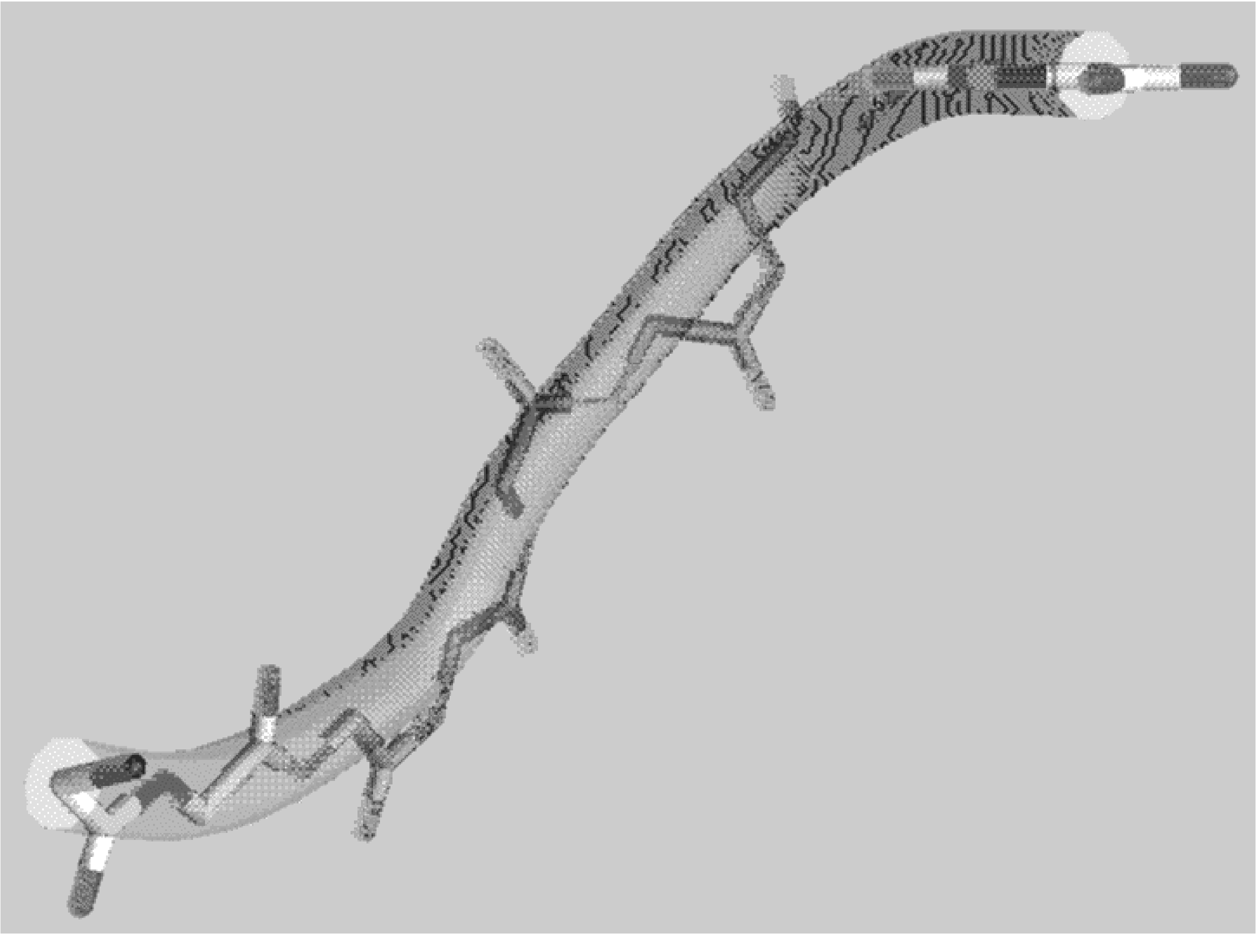

Figure 13: Plot of decaglycine conformation, using NH_2 amino and –COOH carboxyl end groups. Global minimum energy of -63.87 kcal/mole using the MSEED model for hydration. A C^α worm is used to highlight the backbone structure.

Table 19: Energy contributions at global minimum solutions of solvated decaglycine. The NH_2 and –COOH end groups were used for the MSEED example, while the CH_3CO– and –$NHCH_3$ end groups were used for the RRIGS example. The total energy, E_{TOT}, is provided along with the contributions from hydration, E_{HYD}, nonbonded interactions (including hydrogen bonding), E_{NB}, electrostatic interactions, E_{ES}, and torsion, E_{TOR}.

	$\mathbf{E}_{TOT}$	$\mathbf{E}_{HYD}$	$\mathbf{E}_{NB}$	$\mathbf{E}_{ES}$	$\mathbf{E}_{TOR}$
MSEED	-63.87	-109.90	2.32	43.49	0.22
RRIGS	-87.23	-65.71	-52.93	31.38	0.03

has been shown that these end groups stabilize the formation of α–helical structures for unsolvated decaglycine (Ripoll et al., 1991). The same modified partitioning scheme was employed, and the 3 additional end group dihedral angles were allowed to vary over the entire [-180,180] domain. The global minimum structure, which is plotted in Figure 14, was found to be fully α–helical with an energy of -87.23 kcal/mole. This structure has a 0.136 C^{α} (rms) deviation from the unsolvated global minimum structure, which was also found to be fully α–helical. The solvated α–helical structure was found after 1402 iterations and 13,003 seconds (HP-C110). A breakdown of the individual energy contributions for both decaglycine examples is given in Table 19.

5 Conclusions

In this paper, a procedure was presented for identifying the global minimum energy of solvated peptides. In general, global minimum energy structures have been identified by considering only the potential energy contributions. With one recent exception (Meirovitch and Meirovitch, 1996), considering the effects of solvation through the use of continuum models has been limited to local search techniques. The proposed global search procedure based on the αBB algorithm has been shown to be efficient for both the MSEED and RRIGS implementations. For illustrative purposes, the method was tested using the ECEPP/3 force field and two independent solvation models. The MSEED solvation model is based on solvent–accessible surface areas and, using the JRF parameter set, its hydration free energy is added at local minima only. In contrast, implementing the RRIGS solvent–accessible volume

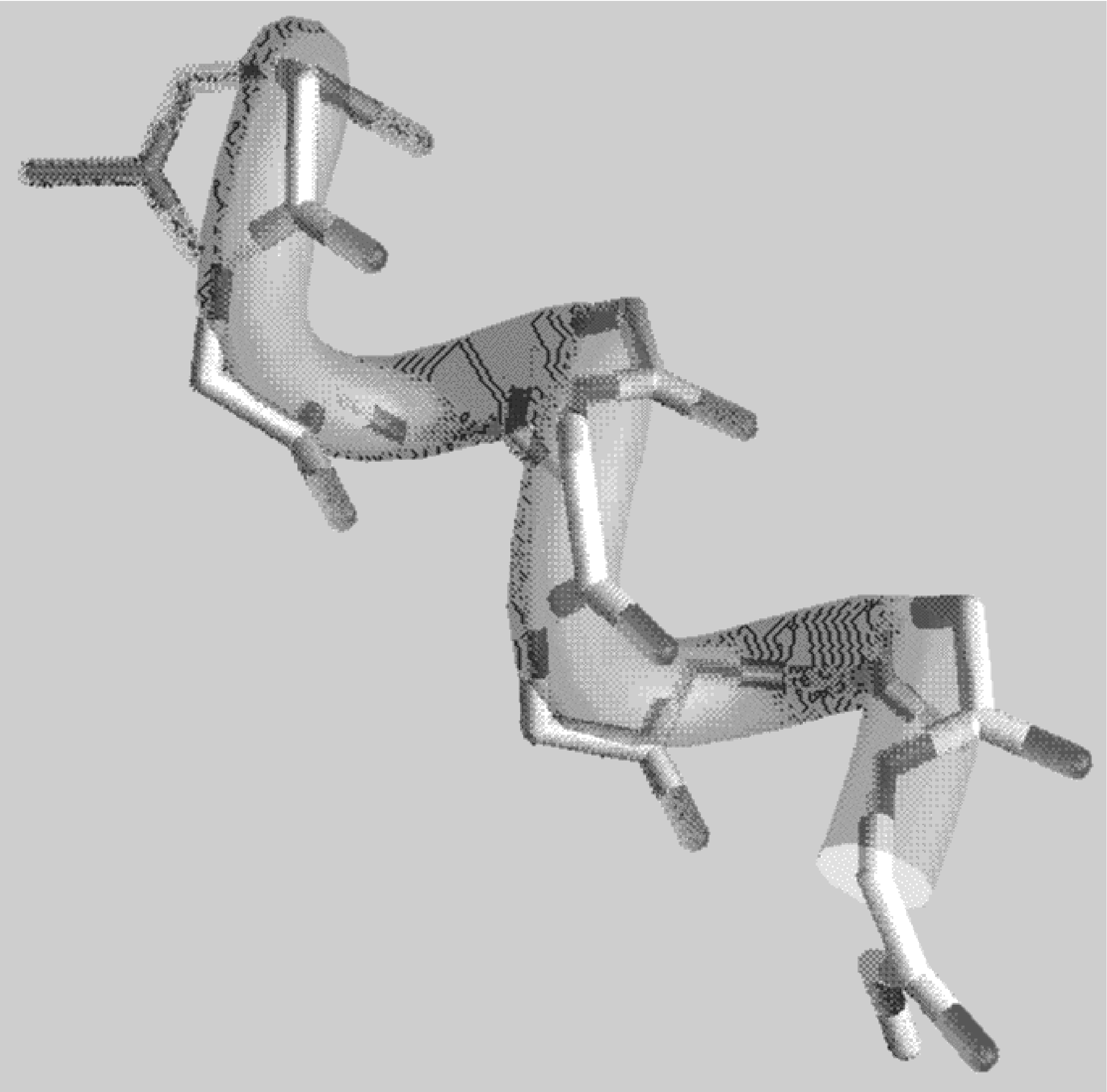

Figure 14: Plot of decaglycine conformation, using CH_3CO– amino and –$NHCH_3$ carboxyl end groups. Global minimum energy of -87.23 kcal/mole using the RRIGS model for hydration. A C^{α} worm is used to highlight the backbone structure.

of hydration shell model requires function and gradient information at each step of the local minimizations. The procedure was tested on the 20 naturally occurring amino acids, three pentapeptides and the decapeptide decaglycine. In treating the oligopeptides, distribution patterns of the dihedral angles of the naturally occurring amino acids were used in excluding parts of the domain space.

A comparison of the two solvation models was made based on identifying global minimum energy conformations. In all cases, the MSEED model was found to predict more extended conformations than those predicted by considering only potential energy terms. In contrast, both Ac–Ala$_4$–Pro–NHMe and decaglycine, whose unsolvated structures exhibit helical regions, retained those features when modeled using RRIGS. On the other hand, as with MSEED, the RRIGS predicted enkephalin structures were found to have extended conformations, with the proximity of the aromatic side chains a central difference between the two solvated conformations. A close correspondence between the met– and leu–enkephalin peptides for both models was also found, along with reasonable agreement with experimental observations. Obviously, in cases where regular secondary structure is expected to dominate, the RRIGS model proves to be a suitable hydration model. For other proteins, such as the two enkephalins, the choice of solvation model can only be validated by experimental information, which, unfortunately, has low accurracy for short linear peptides.

Other important comparisons can be made using the terminally blocked single residue analysis. In general, these results indicated that the RRIGS model was more strongly dominated by the ECEPP/3 potential model. This was evidenced by the dominance of the C_7 minima. In contrast, the low energy regions of the MSEED energy landscape were shifted towards the C_5 region. This shift was also apparent in the oligopeptide examples, where the MSEED hydration energy dominated the location of the global minimum. That is, the change in nonbonded energy between local minima was overshadowed by the change in hydration energy. This can be seen in the global minimum structures of met–enkephalin and leu–enkephalin which exhibit extended conformations and wide separation between aromatic side chains. In contrast, for the RRIGS model the nonbonded energies play an important role. The enkephalins are again extended, but favorable interactions of the aromatic side chains result in more short range interactions. The prediction of partial α–helical structure for Ac–Ala$_4$–Pro–NHMe and full α–helical structure for both forms of decaglycine using the RRIGS model also displays the importance of short range interactions in this model.

The analysis of the terminally blocked peptides also qualitatively predicted the trends of the values for hydration energies in the oligopeptide examples. Specifically, the single residue analysis showed that four residues caused the MSEED hydration energy to be more stabilizing than the RRIGS hydration energy at their corresponding global minimum. Those residues included the three aromatic residues, tryptophan, tyrosine and phenylalanine, and the ringed imidazole residue, histidine. In addition, the aliphatic residues, isoleucine, valine, leucine and alanine, were shown to be the least stabilizing in terms of the difference in hydration energies. Therefore, the Ac–Ala_4–Pro–NHMe example, which has a large aliphatic content, possesses a less stabilizing MSEED hydration energy at the corresponding global minimum. In contrast, the met–enkephalin and leu–enkephalin examples, which both contain two aromatic residues (tyrosine and phenylalanine), have much more stabilizing MSEED hydration energies at their corresponding global minima.

Acknowledgments

The authors gratefully acknowledge financial support from the National Science Foundation, Air Force Office of Scientific Research, and the National Institutes of Health.

References

Adjiman C.S., Androulakis I.P., and Floudas C.A., 1997a, A global optimization method, αBB, for general twice–differentiable NLPs – II. Implementation and computational results. submitted for publication.

Adjiman C.S., Androulakis I.P., and Floudas C.A., 1997b, Global optimization of MINLP problems in process synthesis and design. *Comput. Chem. Eng.* **21**, S445–S450.

Adjiman C.S., Androulakis I.P., Maranas C.D., and Floudas C.A., 1996, A global optimization method, αBB, for process design. *Comput. Chem. Eng.* **20**, S419–S424.

Adjiman C.S., Dallwig S., Floudas C.A., and Neumaier A., 1997c, A global optimization method, αBB, for general twice–differentiable NLPs – I. Theoretical advances. submitted for publication.

Adjiman C.S. and Floudas C.A., 1996, Rigorous convex underestimators for general twice–differentiable problems. *J. Glob. Opt.* **9**, 23–40.

Allinger N.L., Yuh Y.H., and Lii J.H., 1989, Molecular mechanics. the mm3 force field for hydrocarbons. *J. Am. Chem. Soc.* **111**, 8551–8565.

Androulakis I.P., Maranas C.D., and Floudas C.A., 1995, αbb : A global optimization method for general constrained nonconvex problems. *J. Glob. Opt.* **7**, 337–363.

Androulakis I.P., Maranas C.D., and Floudas C.A., 1997, Global minimum potential energy conformation of oligopeptides. *J. Glob. Opt.* **11**, 1–34.

Anfinsen C.B., Haber E., Sela M., and White F.H., 1961, The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *J. Proc. Nat. Acad. Sci. USA* **47**, 1309–1314.

Augspurger J.D. and Scheraga H.A., 1996, An efficient, differentiable hydration potential for peptides and proteins. *J. Comp. Chem* **17**, 1549–1558.

Brooks B., Bruccoleri R., Olafson B., States D., Swaminathan S., and Karplus M., 1983, Charmm: A program for macromolecular energy minimization and dynamics calculations. *J. Comp. Chem.* **4**, 187–217.

Burley S.K. and Petsko G.A., 1985, Aromatic–aromatic interaction: A mechanism of protein structure stabilization. *Science* **229**, 23–28.

Camerman A., Mastropaolo D., Karle I., Karle J., and Camerman N., 1983, Crystal structure of leucine–enkephalin. *Nature* **306**, 447–450.

Connolly M.L., 1983, Analytical molecular surface calculations. *J. Appl. Cryst.* **16**, 548–558.

Dauber-Osguthorpe P., Roberts V.A., Osguthorpe D.J., Wolff J., Genest M., and Hagler A.T., 1988, Structure and energetics of ligand binding to peptides: Escherichia coli dihydrofolate reductase–trimethoprim, a drug receptor system. *Proteins* **4**, 31.

Dejaegere A. and Karplus M., 1996, Analysis of coupling schemes in free energy simulations: A unified description of nonbonded contributions to solvation free energies. *J. Phys. Chem.* **100**, 11148–11164.

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

Eisenhaber F., Lijnzaad P., Argos P., Sander C., and Scharf M., 1995, 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**, 273–284.

Floudas C., 1997, Deterministic global optimization in design, control, and computational chemistry. In Biegler L., Coleman T., Conn A., and Santosa F. (eds.), *Large Scale Optimization with Applications, Part II: Optimal Design and Control*, vol. 93, (pp. 129–184), IMA Volumes in Mathematics and its Applications, Springer–Verlag.

Gill P.E., Murray W., Saunders M.A., and Wright M.H., 1986, *NPSOL 4.0 User's Guide*. Systems Optimization Laboratory, Dept. of Operations Research, Stanford University, CA.

Graham W.H., II E.S.C., and Hicks R.P., 1992, Conformational analysis of met–enkephalin in both aqueous solution and in the presence of sodium dodecyl sulfate micelles using multidimensional nmr and molecular modeling. *Biopolymers* **32**, 1755–1764.

Honig B., Sharp K., and Yang A., 1993, Macroscopic models of aqueous solutions: Biological and chemical applications. *J. Phys. Chem.* **97**, 1101–1109.

Kitao A., Hirata F., and Go N., 1993, Effects of solvent on the conformation and the collective motions of a protein. 2. structure of hydration in melittin. *J. Phys. Chem.* **97**, 10223–10230.

Kollman P., 1993, Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395–2417.

Lambert M.H. and Scheraga H.A., 1989, Pattern recognition in the prediction of protein structure. i. tripeptide conformational probabilities calculated from the amino acid sequence. *J. Comp. Chem* **10**, 770–797.

Levitt M., 1983, Protein folding by restrained energy minimization and molecular dynamics. *J. Mol. Biol.* **170**, 723–764.

Li Z. and Scheraga H.A., 1988, Structure and free energy of complex thermodynamic systems. *J. Mol. Struct. (Theochem.)* **179**, 333–352.

Madison V. and Kopple K.D., 1980, Solvent–dependent conformational distributions of some dipeptides. *J. Am. Chem. Soc.* **102**, 4855–4863.

Maranas C.D., Androulakis I.P., and Floudas C.A., 1996, A deterministic global optimization approach for the protein folding problem. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 23, (pp. 133–150), American Mathematical Society.

Maranas C.D. and Floudas C.A., 1992, A global optimization approach for lennard-jones microclusters. *J. Chem. Phys.* **97**, 7667–7677.

Maranas C.D. and Floudas C.A., 1993, Global optimization for molecular conformation problems. *Annals of Operations Research* **42**, 85–117.

Maranas C.D. and Floudas C.A., 1994a, A deterministic global optimization approach for molecular structure determination. *J. Chem. Phys.* **100**, 1247–1261.

Maranas C.D. and Floudas C.A., 1994b, Global minimum potential energy conformations of small molecules. *J. Glob. Opt.* **4**, 135–170.

Meirovitch H. and Meirovitch E., 1996, New theoretical methodology for elucidating the solution structure of peptides from nmr data. 3. solvation effects. *J. Phys. Chem.* **100**, 5123–5133.

Momany F.A., Carruthers L.M., McGuire R.F., and Scheraga H.A., 1974a, Intermolecular potential from crystal data. iii. *J. Phys. Chem.* **78**, 1595–1620.

Momany F.A., Carruthers L.M., and Scheraga H.A., 1974b, Intermolecular potential from crystal data. iv. *J. Phys. Chem.* **78**, 1621–1630.

Momany F.A., McGuire R.F., Burgess A.W., and Scheraga H.A., 1975, Energy parameters in polypeptides. vii. *J. Phys. Chem.* **79**, 2361–2381.

Némethy G., Gibson K.D., Palmer K.A., Yoon C.N., Paterlini G., Zagari A., Rumsey S., and Scheraga H.A., 1992, Energy parameters in polypeptides. 10. *J. Phys. Chem.* **96**, 6472–6484.

Némethy G., Pottle M.S., and Scheraga H.A., 1983, Energy parameters in polypeptides. 9. *J. Phys. Chem.* **87**, 1883–1887.

Neumaier A., 1997, Molecular modeling of proteins and mathematical prediction of protein structure. accepted for publication *SIAM Rev.*

Noguti T. and Go N., 1983, A method of rapid calculation of a second derivative matrix of conformational energy for large molecules. *J. Phys. Soc. Japan* **52**, 3685–3690.

Pardalos P.M., Shalloway D., and Xue G. (eds.), 1996, *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding*, vol. 23 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, Amer. Math. Soc.

Perrot G., Cheng B., Gibson K.D., J. Vila K.A.P., Nayeem A., Maigret B., and Scheraga H.A., 1992, Mseed: A program for the rapid analytical determination of accessible surface areas and their derivatives. *J. Comp. Chem* **13**, 1–11.

Ramachandran G.N. and Saisekharan V., 1968, Conformation of polypeptides and proteins. *Advances in Protein Chemistry* **23**, 283–437.

Ripoll D.R., Vásquez M.J., and Scheraga H.A., 1991, The electrostatically driven monte carlo method: Application to conformational analysis of decaglycine. *Biopolymers* **31**, 319–330.

Scheraga H., 1996, *PACK: Programs for Packing Polypeptide Chains*. online documentation.

Scheraga H.A., 1992, Predicting three–dimensional structures of oligopeptides. In Lipkowitz K.B. and Boyd D.B. (eds.), *Reviews in Computational Chemistry*, vol. 3, (pp. 73–142), VCH Publishers.

Straatsma T.P. and McCammon J.A., 1992, Computational alchemy. *Annu. Rev. Phys. Chem.* **43**, 407–435.

van Groningen W.F. and Berendsen H.J.C., 1987, *GROMOS*. Groningen Molecular Simulation, Groningen, The Netherlands.

Vásquez M., Némethy G., and Scheraga H.A., 1983, Computed conformational states of the 20 naturally occurring amino acid residues and of the prototype residue α–aminobutyric acid. *Macromolecules* **16**, 1043–1049.

Vásquez M., Némethy G., and Scheraga H.A., 1994, Conformational energy calculations on polypeptides and proteins. *Chemical Reviews* **94**, 2183–2239.

Vila J., Williams R.L., Vásquez M., and Scheraga H.A., 1991, Empirical solvation models can be used to differentiate native from non-native conformations of bovine pancreatic trypsin inhibitor. *Proteins* **10**, 199–218.

von Freyberg B. and Braun W., 1993, Minimization of empirical energy functions in proteins including hydrophobic surface area effects. *J. Comp. Chem.* **14**, 510–521.

Wawak R.J., Gibson K.D., and Scheraga H.A., 1994, Gradient discontinuities in calculations involving molecular surface area. *J. Math. Chem.* **15**, 207–232.

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

Weiner S., Kollman P., Nguyen D., and Case D., 1986, An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **7**, 230–252.

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

Zimmerman S.S., Pottle M.S., Némethy G., and Scheraga H.A., 1977, Conformational analysis of the 20 naturally occurring amino acid residues using ecepp. *Macromolecules* **10**, 1–9.