DIMACS Series in Discrete Mathematics
and Theoretical Computer Science
Volume 00, 0000

A Deterministic Global Optimization Approach for the Protein Folding Problem

C.D. MARANAS, I.P. ANDROULAKIS, AND C.A. FLOUDAS*

ABSTRACT. A deterministic global optimization algorithm is proposed for locating the global minimum potential energy conformations of oligopeptide chains. The ECEPP/3 detailed potential energy model is utilized to model the energetics of the atomic interactions. The minimization of the total potential energy is formulated on the set of peptide dihedral angles. Based on previous work on the microcluster and molecular structure determination, a procedure for deriving convex lower bounding functions for the total potential energy function is utilized which involves a number of important properties. The global optimization algorithm α**BB** which has been shown to be ϵ–convergent to the global minimum potential energy conformation through the solution of a series of nonlinear convex optimization problems is utilized. The ECEPP/3 potential model is interfaced with α**BB** in the program **GLOFOLD**, and provisions have been made to accommodate user specified partitioning of the dihedral angles into three sets. The first one (i.e., global variables), consists of dihedral angles where branching occurs. The second set (i.e., local variables) includes the dihedral variables where branching is not necessary. The third set, (i.e., fixed variables) includes the dihedral angles which are kept fixed. The proposed deterministic global optimization is applied on a number of oligopeptide folding problems.

1. Introduction

The *protein folding* problem is one of the most challenging problems in biochemistry. Predicting how a protein would fold is of paramount academic and industrial interest. Many products of the rapidly developing biotechnology industry are novel proteins. It is already possible to design genes to direct the synthesis of such proteins. Yet failure to fold properly is a common production

1991 *Mathematics Subject Classification.* Primary 90B80, 90C20, 90C35, 90C27; Secondary 65H20, 65K05.

Key words and phrases. global optimization, protein folding, convex lower bounding, α**BB**.

This research was partially funded by the NSF grants CBT–8857013, CTS–9221411, the Air Force Office of Scientific Research, Exxon Co., and Mobil Co..

* Author to whom correspondence should be addressed.

concern. It is possible nowadays to produce proteins with a given amino acid sequence and therefore, knowledge of how the protein would fold would allow one to predict and fine–tune its chemical and biological properties. This would greatly simplify the tasks of interpreting data collected by the human genome project, understanding the mechanisms of hereditary and infectious diseases, designing drugs with specific therapeutical properties, and growing biological polymers with specific material properties.

From a chemical point of view, a *protein* is essentially a polymer chain composed by a sequence of various amino acid residues connected with peptide bonds. Proteins in living cells are composed of only 20 different amino acid residues. The general form of these amino acid residues is shown in Figure 1. The form of the side chain R (e.g., methyl, butyl, benzoic, etc.) defines all different amino acid residues. The chemical structure of a protein is illustrated in Figure 2.

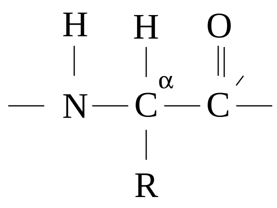

FIGURE 1. Amino acid residue with side chain R.

Note that, the side groups R_n vary from one residue to the other. Also E_{amino}, $E_{carboxyl}$ are the amino and carboxyl end groups respectively. The repeating unit -NC_αC'- connected with peptide bonds defines the *backbone* of the protein. Although, it appears linear in Figure 2, covalent bond angle requirements and interatomic forces bend and twist the chain in a way characteristic for each protein. The protein chain "curls up" into a *unique* three–dimensional geometric conformation called the *folded state* of the protein. It is exactly this configuration which defines the shape of the protein surface as well as the particular chemically active groups present on the surface which in turn determine the biological function of the protein. Predicting this energetically most favorable conformation based solely on the atomic interactions is the objective of this work.

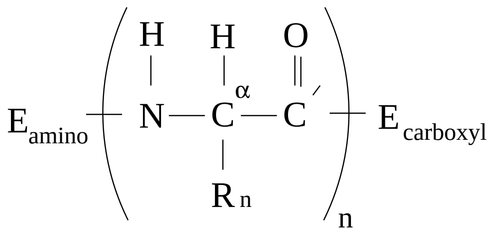

FIGURE 2. Chemical structure of proteins.

In other words, given the *primary structure* of a protein (i.e., residue sequence and type) predict its *tertiary structure* (i.e., 3–D conformation). In the next section, a mathematical description of the protein folding problem is provided.

2. Mathematical Description

The geometry of a protein can be fully described by assigning a three–dimensional *coordinate vector*,

$$r_i \;=\; \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \quad i = 1, \ldots, N$$

which specifies the position of each atom $i = 1, \ldots, N$ in the protein molecule. The *bond vector* between two atoms (i,j) connected with a covalent bond is defined as:

$$r_{ij} \;=\; \begin{pmatrix} x_j - x_i \\ y_j - y_i \\ z_j - z_i \end{pmatrix}$$

The corresponding *bond length* is then equal to the Euclidean distance between atoms i and j,

$$|r_{ij}| \;=\; \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}.$$

The *covalent bond angle* θ_{ijk} formed by the two adjacent bond vectors r_{ij} and r_{jk} can be computed by the following formulae (See Figure 3).

$$\cos(\theta_{ijk}) = \frac{r_{ij} \cdot r_{jk}}{|r_{ij}|\,|r_{jk}|}, \quad \sin(\theta_{ijk}) = \frac{r_{ij} \times r_{jk}}{|r_{ij}|\,|r_{jk}|}.$$

Here $r_{ij} \cdot r_{jk}$ is the *dot product* of the bond vectors r_{ij} and r_{jk},

$$r_{ij} \cdot r_{jk} = (x_j - x_i)(x_k - x_j) + (y_j - y_i)(y_k - y_j) + (z_j - z_i)(z_k - z_j)$$

and $r_{ij} \times r_{jk}$ is the *cross product*,

$$r_{ij} \times r_{jk} = \begin{pmatrix} (y_j - y_i)(z_k - z_j) - (z_j - z_i)(y_k - y_j) \\ (z_j - z_i)(x_k - x_j) - (x_j - x_i)(z_k - z_j) \\ (x_j - x_i)(y_k - y_j) - (y_j - y_i)(x_k - x_j) \end{pmatrix}.$$

The *dihedral angle* $\omega_{ijkl} \in [-180^o, 180^o]$ or the complementary *torsion angle* $\phi_{ijkl} = \omega_{ijkl} - 180^o$ measure the relative orientation of two adjacent covalent angles θ_{ijk} and θ_{jkl} (see Figure 3). It is defined as the angle between the normals through the planes defined by atoms i, j, k and j, k, l respectively, and can be calculated from the following relations:

$$\begin{aligned} \cos(\omega_{ijkl}) &= \frac{(r_{ij} \times r_{jk}) \cdot (r_{jk} \times r_{kl})}{|r_{ij} \times r_{jk}|\,|r_{jk} \times r_{kl}|}, \\ \sin(\omega_{ijkl}) &= \frac{(r_{kl} \times r_{ij}) \cdot r_{jk}\,|r_{jk}|}{|r_{ij} \times r_{jk}|\,|r_{jk} \times r_{kl}|} \end{aligned}$$

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. Under this assumption, the dihedral angles along the backbone fully determine the geometric shape of the folded protein.

The names of the dihedral angles of a folded protein chain follow a standard nomenclature. The dihedral angle between the normals of the planes formed by atoms $C'_{i-1}N_iC_{\alpha,i}$ and $N_iC_{\alpha,i}C'_i$ respectively is called ϕ_i where $i-1$ and i are two adjacent amino acid residues. The one defined by planes $R_iC_{\alpha,i}C'_i$ and $C_{\alpha,i}C'_iN_{i+1}$ 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 $C_{\alpha,i}C'_iN_{i+1}$ and $C'_iN_{i+1}C_{\alpha,i+1}$. The letter χ is utilized to denote the dihedral angles which are associated with the side groups R_i. Also the letter θ is used to name the dihedral angles associated with the two end groups. Figure 4 pictorially illustrates these conventions.

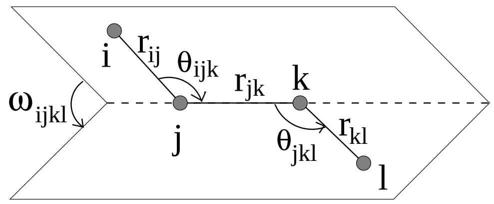

FIGURE 3. Bond vectors, covalent bond angles and dihedral angle

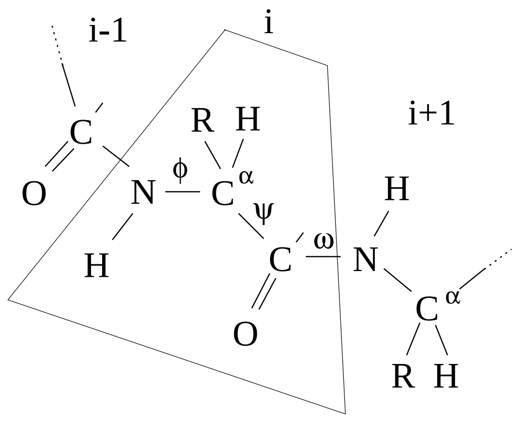

FIGURE 4. Dihedral angles in a protein

3. Potential Energy Model

Molecular mechanics calculations employ an empirically derived set of potential energy contributions for approximating these atomic interactions. This set of potential energy contributions, called the *force field*, contains adjustable parameters that are selected in a such a way as to provide the best possible agreement with experimental data. The main assumption introduced in molecular mechanics is that every parameter is associated with *a specific interaction* rather than a specific molecule *(transferability assumption).* These parameters are bond lengths; covalent bond angles; bond stretching, bending, or rotating constants; non–bonded atom interaction constants, etc. Thus, whenever a specific interaction is present, the same value for the parameter can be used even if this interaction occurs in different molecules [**4**]. Note that experimental results provide sufficient evidence that this is a reasonable assumption in most cases. Many different models have been proposed for approximating the *force field*, and some of the most popular ones are: ECEPP [**10, 11, 12**], MM2 [**1**], ECEPP/2 [**15**], CHARMM [**3**], AMBER [**18**], GROMOS87 [**17**], MM3 [**2**], and ECEPP/3 [**14**].

In this work the ECEPP/3 potential model is utilized. In this potential model, it is assumed that the covalent bond lengths and angles are fixed at their equilibrium values, and thus the protein conformation is only a function of the dihedral angles. This implies that ECEPP/3 accounts for only energy interaction terms which depend on the dihedral angles. The conformational energy is treated as the sum of electrostatic, nonbonded, hydrogen bond and torsional contributions, plus an additional loop closing potential if the polypeptide contains one or more intramolecular disulfide bonds. Also the fixed internal conformational energy of the pyrolidine ring is added for each propyl or hydroxyprolyl residue contained in the polypeptide. The first three energy contributions are computed for each atom pair (i,j) whose interatomic distance is a function of at least one dihedral angle. This set of atomic pairs is denoted as $\mathcal{P}$ and includes the atomic pairs which are separated by at least two other atoms.

The electrostatic energy U_{ES} is computed for each atomic pair $(i,j) \in \mathcal{E}S = \mathcal{P}$ as a Coulomb potential interaction between two atom–centered monopole partial charges q_i and q_j where D is the dielectric constant.

$$U_{ES} \;=\; \sum_{(i,j)\in\mathcal{E}S} \frac{q_i q_j}{D\,|r_{ij}|}$$

A modified Lennard–Jones 12–6 potential is used to approximate the nonbonded interaction energies between atomic pairs $(i,j) \in \mathcal{N}B$ [**14**]. The set $\mathcal{N}B$ contains all atomic pairs $\mathcal{P}$ except for hydrogen bonding pairs.

$$U_{NB} \;=\; \sum_{(i,j)\in\mathcal{N}B} F\frac{A_{k(i)k(j)}}{|r_{ij}|^{12}} - \frac{C_{k(i)k(j)}}{|r_{ij}|^{6}}$$

Here $k(i)$ returns the atom type of atom i in the protein chain. The coefficients $A_{k(i)k(j)}$, $C_{k(i),k(j)}$ are assigned specific values for each combination of atom types $k(i)$ and $k(j)$. F is assigned a value of 0.5 for 1–4 interactions and 1.0 for for 1–5+ interactions. An atom pair interaction is defined as 1–4 when the distance between the interacting atoms is a function of only one intervening dihedral angle. Any other interactions $(i,j) \in \mathcal{P}$ is considered to be 1–5+.

Hydrogen–bond interactions are the ones between designated donor and acceptor atoms. The donors (H) are amine, amide, hydroxyl or carboxyl acid hydrogens and the acceptors (X) are uncharged ring nitrogens, amide nitrogens, or hydroxyl ester, carbonyl, or carboxylic acid oxygens. A 12–10 potential function is used to model the hydrogen bond interactions.

$$U_{HX} = \sum_{(i,j)\in\mathcal{HX}} F \frac{A'_{k(i),k(j)}}{|r_{ij}|^{12}} - \frac{B'_{k(i),k(j)}}{|r_{ij}|^{10}}$$

Note that $\mathcal{P} = \mathcal{NB} \cup \mathcal{HX}$.

Torsional energy terms are included in the potential energy model to bring the experimental and computed rotational barrier into agreement. These terms are computed for all ω dihedral angles and for some designated side–chain χ and end group θ dihedral angles, but not for any ϕ and ψ angles. Let $\mathcal{TOR}$ be the set of dihedral angles for which a torsional term is calculated. The form of the potential function used is:

$$U_{TOR} = \sum_{k\in\mathcal{TOR}} \frac{U_{o,k}}{2}\left(1 + c_k \cos n_k t_k\right)$$

Here $U_{o,k}$ is the difference between the experimental barrier and the one calculated from the electrostatic, nonbonded, and hydrogen–bond potential functions, t_k is the value of the k^{th} dihedral angle for which a torsional term is included, n_k gives the symmetry of the barrier, and $c_k \in \{-1, 1\}$ defines the sign for the cosine term.

The cystine loop–closing energy U_{LOOP} and torsional energy U_{CYST} are computed as the sums of terms for all disulfide bonds in the peptide. The loop–closing potential penalizes any deviation of the interatomic distances S_iS_j, $C_i^\beta\ S_j$, and $C_j^\beta\ S_i$ from their experimentally observed values (see Figure 5). Let $\mathcal{SS}$ be the set of all disulfide bonds in the peptide. Then U_{LOOP} is defined as:

$$U_{LOOP} = \sum_{(i,j)\in\mathcal{SS}} B\left[\left(r_{S_iS_j} - r^o_{S_iS_j}\right)^2 + \left(r_{S_iC_j^\beta} - r^o_{S_iC_j^\beta}\right)^2 + \left(r_{S_jC_i^\beta} - r^o_{S_jC_i^\beta}\right)^2\right]$$

Note that $r^o_{S_iS_j} = 2.04A^o$ is the experimentally observed disulfide bond distance and $r^o_{S_iC_j^\beta} = r^o_{S_jC_i^\beta} = 3.052A^o$ are the distances between $S_{i/j}$ and $C^\beta_{j/i}$ so as the covalent bond angles $\theta_{C_i^\beta S_iS_j}$ and $\theta_{S_iS_jC_j^\beta}$ assume the experimentally observed value of 104^o. B is a penalty parameter assigned the value $B = 100\ kcal/mol\text{Å}^2$. Torsional contributions to U_{CYST} from the angles $C_i^\alpha C_i^\beta S_i S_j$ and $S_i S_j C_j^\beta C_j^\alpha$ are calculated as mentioned above with $U_o = 1.5 kcal/mol$. The contribution from the dihedral angle $C_i^\beta S_i S_j C_j^\beta$ is computed as a penalty term on the interatomic distance $r_{C_i^\beta C_j^\beta}$ according to the relation:

$$U_{SS} = A \sum_{(i,j)\in \mathcal{SS}} \left(r_{C_i^\beta C_j^\beta} - r^o_{C_i^\beta C_j^\beta} \right)^2$$

The penalty parameter A is set equal to 10 $kcal/mol$Å^2 and the experimentally observed interatomic distance $r^o_{C_i^\beta C_j^\beta}$ to 3.855Å.

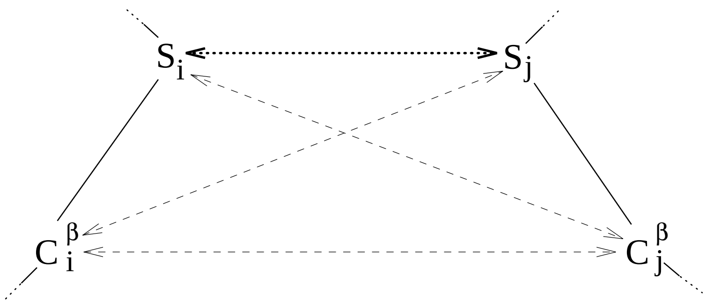

FIGURE 5. Disulfide bonding

Finally, the additional potential energy term U_{PRO} is added to account for the internal conformation energy of proline and hydroxyproline residues. This internal conformation energy depends on whether the peptide bond is on *cis* or *trans* configuration. The total potential energy of the peptide chain, in the context of the potential model ECEPP/3, can then be written as the sum of a number of different interaction and correction terms.

$$U = U_{ES} + U_{NB} + U_{HX} + U_{TOR} + U_{LOOP} + U_{CYST} + U_{PRO}$$

Note that all these terms are functions of the dihedral angles.

4. Problem Formulation

The potential 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. This implies that there are $\phi_i, \psi_i, \omega_i,\ i = 1, \ldots, N_{RES}$ dihedral angles along the backbone of the peptide chain. Also let $k = 1, \ldots, K^i$ denote the dihedral angles of the side group in the i^{th} residue and $j = 1, \ldots, J^N$ the dihedral angles of the amino end group and $j = 1, \ldots, J^C$ of the carboxyl end group respectively. This defines the side group dihedral angles $\chi_i^k,\ i = 1, \ldots, N_{RES},\ k = 1, \ldots, K^i$ and the amino $\theta_j^N,\ j = 1, \ldots, J^N$ and carboxyl $\theta_j^C,\ j = 1, \ldots, J^C$ end group dihedral angles respectively. Based on these definitions the potential model minimization energy problem can be formulated as follows:

$$\min \quad U(\phi_i, \psi_i, \omega_i, \chi_i^k, \theta_j^N, \theta_j^C)$$

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

Here U is the expression for the total potential energy as a function of the peptide dihedral angles. The specific expressions comprising U have been described in detail in the previous section. Note that U is a nonconvex function of these dihedral angles involving numerous local minima even for small peptide systems. These local minima correspond to metastable states of the polypeptide chain. A single global minimum defines the energetically most favorable peptide conformation. A plethora of different methods has been proposed for finding this conformation [**16**]. Most methods attempt to locate this point by tracing, deterministically or stochastically, single or multiple paths on the potential energy surface conjecturing that some of them will converge to the global minimum potential energy point. A review on these methods can be found in [**8**]. The key limitation of these methods is that the obtained conformations depend heavily on the supplied initial conformation expressing the bias of the researcher towards which is the most appropriate conformation. This is why, in practice, many trial geometries need to serve as initial points in an attempt to lessen the initial point dependence. However, there is no guarantee that important conformations are not overlooked. The need for a method that can guarantee convergence to the global minimum potential energy conformation motivated our initial effort to introduce such a method for microclusters [**6, 7**], and small acyclic molecules [**8, 9**] allowing for nonbonded atomic pair interactions. The approach, α**BB** has been extended to constrained optimization problems in [**5**]. In this paper, the approach is extended to peptide systems interacting with realistic potential energy models (i.e., ECEPP/3). In the next section, a brief description of α**BB** customized for the protein folding problem, is provided.

5. Global Optimization

The deterministic branch and bound type global optimization algorithm α**BB** [**8, 5**] is utilized which brackets the global minimum solution by constructing converging lower and upper bounds. These bounds are successively refined by iteratively partitioning the initial feasible region into many subregions. Upper bounds to the global minimum can be obtained by local minimizations of U. Lower bounds are obtained by minimizing a convex function L which is always less than the original nonconvex function U. This function L can be constructed by augmenting U through the addition of a convex separable quadratic term for each dihedral angle.

$$
\begin{aligned}
L \; = \; U + \alpha \{ \quad & \sum_{i=1}^{N_{RES}} \left(\phi_i^L - \phi_i\right)\left(\phi_i^U - \phi_i\right) + \\
& \sum_{i=1}^{N_{RES}} \left(\psi_i^L - \psi_i\right)\left(\psi_i^U - \psi_i\right) + \\
& \sum_{i=1}^{N_{RES}} \left(\omega_i^L - \omega_i\right)\left(\omega_i^U - \omega_i\right) + \\
& \sum_{i=1}^{N_{RES}} \sum_{k=1}^{K^i} \left(\chi_i^{k,L} - \chi_i^k\right)\left(\chi_i^{k,U} - \chi_i^k\right) + \\
& \sum_{j=1}^{J^N} \left(\theta_j^{N,L} - \theta_j^N\right)\left(\theta_j^{N,U} - \theta_j^N\right) + \\
& \sum_{j=1}^{J^C} \left(\theta_j^{C,L} - \theta_j^C\right)\left(\theta_j^{C,U} - \theta_j^C\right) \}
\end{aligned}
$$

Note that $\phi_i^L, \psi_i^L, \omega_i^L, \chi_i^{k,L}, \theta_j^{N,L}, \theta_j^{C,L}$ and $\phi_i^U, \psi_i^U, \omega_i^U, \chi_i^{k,U}, \theta_j^{N,U}, \theta_j^{C,U}$ are lower and upper bounds respectively on the dihedral angles $\phi_i, \psi_i, \omega_i, \chi_i^k, \theta_j^N, \theta_j^C$. Also α is a nonnegative parameter which must be greater or equal to the negative one half of the minimum eigenvalue of U inside the current dihedral angles rectangle. Qualitatively, the effect of adding this extra term to U is to make L convex by overpowering the nonconvexity characteristics of U with the addition of the term 2α to all of its eigenvalues. This function L, defined inside some rectangular region, involves a number of important properties which enable us to construct a global optimization algorithm for finding the global minimum of U in the space defined by the dihedral angles. These properties, whose proof is given in [**8**], demonstrate that (i) L is always a valid underestimator of U; (ii) L matches U at all corner points of the box constraints; (iii) L is convex; (iv) the maximum separation between L and V is *bounded* and proportional to α and to the square of the diagonal of the current box constraints; and (v) the underestimators L constructed over supersets of the current set are always *less tight* than the underestimator U constructed over the current box constraints for every point within the current box constraints.

Based on these properties a deterministic branch and bound type global optimization algorithm is proposed for locating the global minimum potential energy of U by constructing converging lower and upper bounds. The approach is implemented in the **GLOFOLD** package. Qualitatively, the steps of the approach are as follows:

Step 1 An upper bound on the global minimum solution of U is obtained by minimizing U with a local solver (i.e., MINOS [**13**]). The current best upper bound is updated to be the minimum over the stored ones.

Step 2 The current rectangle is partitioned in two by bisecting along the longest side.

Step 3 The convex function L is minimized inside both resulting subrectangles. If the solutions are less than the current best upper bound they are stored, otherwise they are discarded (fathoming).

Step 4 The rectangle involving the minimum solution for $\min L$ is selected for further partitioning and the corresponding solution is erased from the lower bounds stack.

Step 5 If the current best upper and lower bounds are within ϵ then terminate, otherwise continue with Step 1.

The approach is shown in [**8**] to terminate in a finite number of iterations to an ϵ–global minimum solution.

6. Implementation: GLOFOLD

The proposed approach has been interfaced with ECEPP/3 and implemented in C, in the program **GLOFOLD**. A schematic diagram of the interface between ECEPP/3 and α**BB** is shown in Figure 6.

The peptide dihedral angles are partitioned into three sets. The first one (i.e., global variables), consists of dihedral angles where branching occurs. The second set (i.e., local variables) includes the dihedral variables where branching is not performed. The third set, (i.e., fixed variables) includes the dihedral angles for which there exists sufficient (experimental) evidence for keeping them fixed.

The information required by the user, in the current implementation of **GLOFOLD**, is provided in two files. The first one, required by ECEPP/3, contains information about the sequence and number of the amino acid residues and the type of the end groups. Also dihedral angles are initialized and output file numbers are assigned. The second file contains information related with the global optimization phase. In particular, (i) number of dihedral angles, (ii) convergence tolerances, (iii) type of starting point, (iv) lower/upper bounds on dihedral vari-

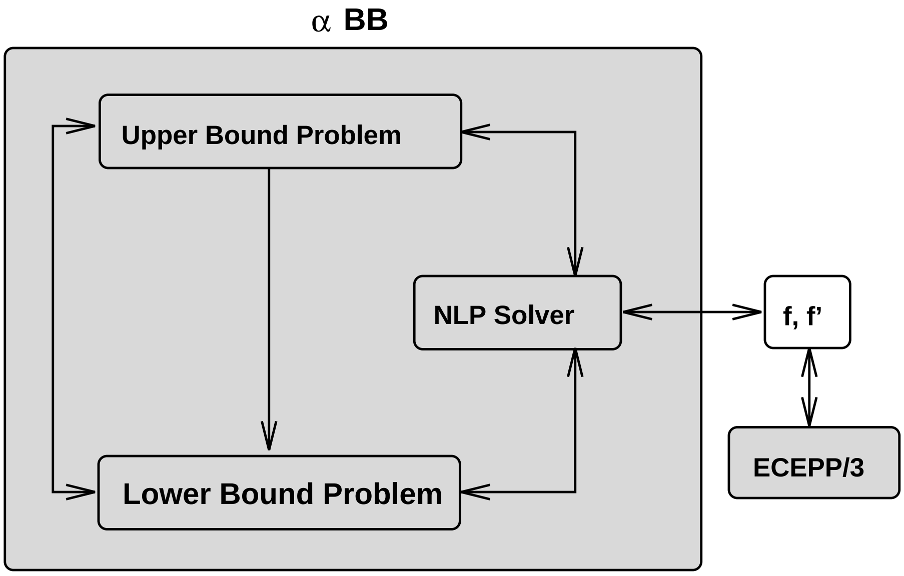

FIGURE 6. Interface between ECEPP/3 and α**BB**

ables, (v) values for α parameter, and (vi) variables where branching occurs are required. In addition to the ability of **GLOFOLD** to locate the global minimum total potential energy oligopeptide conformation, low energy protein conformations close to the global minimum one can be obtained with multi-start local optimizations initiated from the obtained solutions for the lower bounding problems. Furthermore, PDB format files are created for all solutions which can be readily interfaced with graphics programs.

7. Computational Studies

GLOFOLD has been tested on two classes of oligopeptide folding problems. The selected relative convergence tolerance is 10^{-2} and computational requirements in seconds are reported for an HP–730 workstation. First, the method was applied on all 20 naturally occurring amino acids. The amino end group used was acetyl, and the carboxyl end group was methyl. Note that all dihedral angles were treated as global variables except for the three θ angles in the end groups which were treated as local variables. The results are summarized in Table 1. The CPU time required is presented for each residue, as well as the average $< CPU >$ time needed for residues of the same size. The computational effort compares favorably with that reported in [**19**] employing a simulated annealing implementation and fewer dihedral angles. The computational effort increases, as expected, as a function of the number of global variables n, but it stays under the n^3 curve (see Figure 7). Even though in [**19**] a different potential model (i.e. AMBER) was utilized the large differences in computational requirements are quite suggestive about the relative computational efficiencies.

Amino acid	# Dihedrals	Energy	Iter	CPU	$< CPU >$
Pro	5	-19.81	28	6	**6**
Gly	6	-6.33	67	14	**14**
Ala	7	-5.18	141	55	
Cys	7	-5.84	142	45	**50**
His	8	-8.92	298	173	
Phe	8	-8.43	298	169	
Ser	8	-7.86	184	102	
Trp	8	-9.56	306	227	**167**
Asn	9	-22.95	345	220	
Asp	9	-20.05	452	239	
Thr	9	-9.59	285	208	
Tyr	9	-8.48	753	506	
Val	9	-4.19	644	387	**312**
Gln	10	-18.99	601	460	
Glu	10	-15.87	640	386	
Ile	10	-2.54	388	352	
Leu	10	-5.72	1123	613	
Met	10	-6.91	1284	641	**480**
Lys	11	-7.98	922	1070	**1070**
Arg	13	-31.84	1000	1660	**1660**

TABLE 1. Naturally occurring amino acids

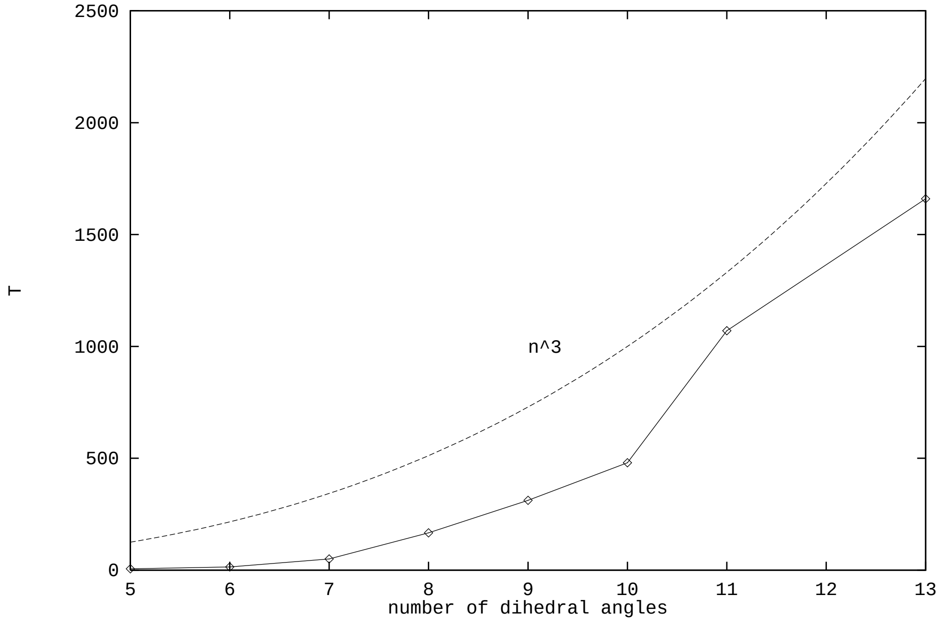

FIGURE 7. CPU times as a function of the number of global dihedral angles.

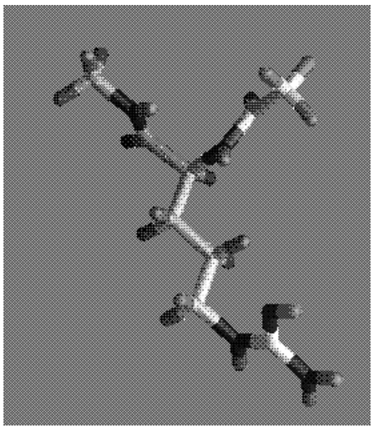

FIGURE 8. NaN'm–arginine.

No.	Name
1	NaN'm – L-Alanyl-L-proline(D)
2	NaN'm – L-Alanyl-L-proline(U)
3	NaN'm – L-Prolyl(D)-L-proline(D)
4	NaN'm – L-Prolyl(D)-L-proline(U)
5	NaN'm – L-Prolyl(U)-L-proline(D)
6	NaN'm – L-Prolyl(U)-L-proline(U)

TABLE 2. Oligopeptide names

Residue	θ_1^1	θ_1^2	ϕ_1	ψ_1	ω_1	χ_1	χ_2	ϕ_1	ψ_1	ω_1	θ_2^1	U^*
1^{known}	180.0	180	-149.0	67	177	180	–	-68.8	79	180	180	-17.39
1^{found}	60.0	-180	-149.0	67	177	178	–	-68.8	79	179	60	-17.39
2^{known}	180	180	-150.0	77	178	180	–	-53.0	-38	180	180	-17.08
2^{found}	-180	180	-150.0	77	178	-61	–	-53.0	-38	-180	-180	-17.08
3^{known}	180	180	-68.8	162	176	–	–	-68.8	81	178	180.0	-32.33
3^{found}	60	180	-68.8	162	176	–	–	-68.8	80	178	60.0	-32.34
4^{known}	180	176	-53.0	157	177	–	–	-68.8	81	178	180	-31.38
4^{found}	55	176	-53.0	157	-179	–	–	-68.8	81	178	180	-31.38
5^{known}	180	179	-68.8	169	181	–	–	-53.0	-38	179	180	-31.30
5^{found}	-63	179	-68.8	169	177	–	–	-53.0	-38	179	180	-31.30
6^{known}	180	174	-53.0	167	179	–	–	-53.0	-37	179	180	-30.22
6^{found}	-53	174	-53.0	167	-179	–	–	-53.0	-37	-179	60	-30.22

TABLE 3. Oligopeptide results

The second set of examples involves a number of oligopeptides listed in Table 2. The results are summarized in Table 3. Dihedral angle and energy values are reported at the global minimum solution. Note that in all cases the obtained solution is at least as good as the best over the ones reported in the literature. In case (3) a slightly improved solution is found, over the best known so far. The computational performance and efficiency of a simulated annealing implementation was next studied on the oligopeptide examples of Table 3. The parameters employed in the simulated annealing implementation follow the suggestions of [**19**] and are shown in Table 4. Nine different combinations of Markov chain lengths and number of successive annealing steps are considered (see Table 4). Note that MC is the Markov chain length and NT the number of successive annealing steps. Furthermore, the initial temperature was $T_0 = 5.0$, the maximum allowable step was set to $\delta = 90^o$ and the cooling schedule factor was $\beta = 0.9$. The results in Table 5 indicate that the α**BB** algorithm always locates the global minimum conformation and requires less CPU time than the simulated annealing implementation. On top of that, the failure rate (numbers in parentheses) for the simulated annealing is very high. Notably, in certain

MC	NT		
	25	50	100
100	1	2	3
250	4	5	6
500	7	8	9

TABLE 4. Parameters used in different simulated annealing runs

		αBB	Simulated Annealing								
Example	D.A.		1	2	3	4	5	6	7	8	9
		cpu	cpu (%F)	cpu (%F)	cpu (%F)	cpu (%F)	cpu (%F)	cpu (%F)	cpu (%F)	cpu (%F)	cpu (%F)
1	4	29.8 s	8.26 s	16.6 s	33.0 s	20.7 s	41.3 s	82.5 s	41.2 s	82.5 s	165. s
			(100)	(100)	(100)	(100)	(100)	(90)	(90)	(90)	(90)
2	4	40.6 s	8.20 s	16.6 s	33.0 s	20.6 s	41.5 s	82.7 s	41.3 s	84.4 s	164. s
			(90)	(100)	(90)	(100)	(90)	(90)	(90)	(70)	(90)
3	5	60.5 s	9.60 s	19.4 s	39.1 s	24.2 s	48.6 s	97.1 s	49.0 s	97.4 s	193. s
			(100)	(60)	(80)	(90)	(100)	(90)	(60)	(100)	(100)
4	5	62.2 s	9.60 s	19.2 s	39.1 s	24.2 s	48.7 s	39.1 s	48.9 s	97.1 s	195. s
			(70)	(80)	(80)	(80)	(90)	(80)	(100)	(100)	(90)
5	5	140. s	9.70 s	19.3 s	39.1 s	24.3 s	48.8 s	96.7 s	48.5 s	96.9 s	194. s
			(90)	(100)	(90)	(90)	(100)	(90)	(90)	(80)	(90)
6	5	138. s	9.70 s	19.4 s	38.7 s	24.2 s	48.9 s	96.5 s	48.6 s	96.7 s	193. s
			(100)	(100)	(100)	(100)	(90)	(80)	(90)	(90)	(70)
			88	85	83	87	81	82	84	81	79

TABLE 5. Comparison of α**BB** and simulated annealing

cases this failure rate reaches 100 %.

8. Conclusions

A deterministic global optimization method was described for locating the global minimum potential energy conformations of oligopeptide chains based on α**BB**. The ECEPP/3 detailed potential energy model was selected to model the energetics of the atomic interactions and the minimization of the total potential energy was formulated on the set of polypeptide dihedral angles. The proposed approach was implemented in C, in the program **GLOFOLD** and provisions were made to accommodate user specified partitioning of the dihedral angles into three sets. The first one (i.e., global variables), consisted of dihedral angles where branching occurs. The second set (i.e., local variables) included the dihedral variables where branching was not necessary. The third set, (i.e., fixed variables) included the dihedral angles which were kept fixed. **GLOFOLD** was applied to a number of oligopeptide folding problems. Computational performance compared favorably with a simulated annealing implementation.

References

1. N.L. Allinger, J. Am. Chem. Soc. **99** (1977), 8127.
2. N.L. Allinger, Y.H. Yuh, and J.-H. Lii, J. Am. Chem. Soc. **111** (1989), 8551.
3. B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **8** (1983), 132.
4. A.J. Hopfinger, *Conformational Properties of Macromolecules*, Academic Press, New York, NY, 1973.
5. C.D. Maranas I.P. Androulakis and C.A. Floudas, accepted in J. Global Opt. (1995).
6. C.D. Maranas and C.A. Floudas, J. Chem. Phys. **97** (1992), no. 10, 7667.
7. ———, Ann. Oper. Res. **42** (1993), 85.
8. ———, J. Global Opt. **4** (1994a), 135.
9. ———, J. Chem. Phys. **100** (1994b), no. 2, 1247.
10. F.A. Momany, L.M. Carruthers, R.F. McGuire, and H.A. Scheraga, J. Phys. Chem. **78** (1974a), 1595.
11. F.A. Momany, L.M. Carruthers, and H.A. Scheraga, J. Phys. Chem. **78** (1974b), 1621.
12. F.A. Momany, R.F. McGuire, A.W. Burgess, and H.A. Scheraga, J. Phys. Chem. **79** (1975), 2361.
13. B.A. Murtagh and M.A. Saunders, *MINOS5.0 Users Guide*, Systems Optimization Laboratory, Dept. of Operations Research, Stanford University, CA., 1983, Appendix A: MINOS5.0, Technical Report SOL 83-20.
14. G. Némethy, K.D. Gibson, K.A. Palmer, C.N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H.A. Scheraga, J. Phys. Chem. **96** (1992), 6472.
15. G. Némethy, M.S. Pottle, and H.A. Scheraga, J. Phys. Chem. **89** (1983), 1883.
16. G. Xue P.M. Pardalos and D. Shalloway, J. Global Opt. **4** (1994), no. 2, 117.
17. W.F. van Groningen and H.J.C. Berendsen, *GROMOS*, Groningen Molecular Simulation, Groningen, The Netherlands, 1987.
18. S. Weiner, P. Kollmann, D. Nguyen, and D. Case, J. Comput. Chem. **7** (1986), 230.
19. S.R. Wilson and W. Cui, Biopolymers **29** (1990), 225.

(C.D. Maranas) Department of Chemical Engineering, Princeton University, Princeton, NJ 08544–5263 USA

E-mail address, C.D. Maranas: kmaranas@titan.Princeton.EDU

(I.P. Androulakis) Department of Chemical Engineering, Princeton University, Princeton, NJ 08544–5263 USA

E-mail address, I.P. Androulakis: androula@titan.Princeton.EDU

(C.A. Floudas) Department of Chemical Engineering, Princeton University, Princeton, NJ 08544–5263 USA

E-mail address, C.A. Floudas: floudas@titan.Princeton.EDU