A Quadratic Assignment Formulation of the Molecular Conformation Problem

A.T. Phillips[1]
J.B. Rosen[2]

ABSTRACT

The molecular conformation problem is discussed, and a concave quadratic global minimization approach for solving it is described. This approach is based on a quadratic assignment formulation of a discrete approximation to the original problem.

Keywords: *Molecular Conformation, Constrained Global Minimization, Quadratic Assignment Problem*

---

1. Computer Science Department, United States Naval Academy, Annapolis, MD 21402.
2. Computer Science Department, University of Minnesota, Minneapolis, MN 55455.

1. Introduction

It is generally agreed that one of the most important, and difficult, problems in molecular biophysics and biochemistry is the protein folding problem (Gierasch and King 1990). The protein folding problem, simply stated, is as follows:

> Given a known primary sequence of amino acids, predict its native, or folded, state in 3-dimensional space.

That is, can one predict how newly made proteins -- which resemble loosely coiled strands and are typically inactive in their unfolded configurations -- will "fold" into specifically shaped balls able to perform crucial tasks in a living cell? The solution to this problem is of more than academic interest. Many major hoped-for products of the biotechnology industry are novel proteins. It is already possible to design genes to direct the synthesis of such proteins, yet failure to fold properly, which greatly determines the functionality of the protein, is a important production concern.

The value of computation of protein folding patterns is that although it is now quite simple to determine the precise amino acid sequence of a protein from DNA sequence analysis, such an analysis provides no information as to what the native (i.e. folded) 3-dimensional structure of the protein might be, and thus which amino acid residues in the protein might be next to which other amino acid residues. Knowledge of such 3-dimensional structures could be of great help in determining the nature of sites on a protein that might be involved in enzyme action or binding to other proteins, membranes, DNA, small molecules, etc. Currently the 3-dimensional structure of proteins can only be ascertained from X-ray crystallography analysis, an expensive and time-consuming process that moreover requires a protein to be pure and then to be crystallized -- not an easy process. The shortcut of direct computation has long been appealing, since it is fairly well documented that the primary structure (the sequence of amino acids) completely determines the secondary and tertiary structures (short and long range folding patterns) of the protein. Furthermore, the process of folding is spontaneous subsequent to the biosynthesis of the protein, and *should* be predicted from the sequence based on energy minimization considerations.

Unfortunately, direct computation of the native state of a protein, in the absence of any simplifying assumptions, has proven to be an intractable problem for all but the smallest of proteins. While many computational simplifications are possible (Troyer and Cohen 1991), one simple and popular approach is to model each complex amino acid residue as a single “sphere” centered on the C_α carbon position, and to model each peptide linkage between residues by a virtual bond between spheres. This C_α force field, or “string of beads”, model ignores the secondary structure of each residue, information which might well be a factor in determining the native state. Hence, it is clear that a computational solution to this simplified “molecular conformation problem” does not in itself solve the protein folding problem; however, this general approach easily allows global optimization techniques to be applied, and could be useful in a more general set of “minimum energy conformation” applications.

Hence, this paper presents a novel approach for predicting the native structures of a linear sequence of beads (residues, in the case of protein folding). This approach is based on two important assumptions:

1) for any specific molecular conformation, a corresponding potential energy function can be computed, and

2) the native state corresponds to the global (or near global) minimum of this energy function.

These assumptions appear to be valid based on results for small proteins (Troyer and Cohen 1991). Clearly, the success of such an approach will depend greatly on both the potential energy function selected and on the method used to compute the global, or near global, minimum of a function with potentially many local minima. In this paper, the molecular conformation problem is formulated so that it can be solved by a two stage approach. The problem is first modeled by a discrete approximation on a 3-dimensional lattice. This discrete lattice model can be formulated as a quadratic assignment problem and then transformed into a *continuous* concave quadratic global minimization problem. The global solution to this concave minimization problem can then be used as starting point for the second stage -- a "relaxed" continuous minimization problem. The result of this second stage should provide a global, or near global, minimum of the potential energy function, and hence a prediction of the native, or folded, state of the linear molecule. This two-stage approach has been used successfully to find the minimum energy conformation for very large problems based on a much simpler molecular model (Maier, Rosen, and Xue 1992; Xue 1992; Xue, Maier, and Rosen 1992).

2. The Molecular Conformation Problem

In the string of beads model, the molecule to be folded consists of a linear sequence of n beads $a_1,a_2,...,a_n$, where a_i denotes the i^{th} bead in the primary sequence. For every pair of consecutive beads a_i and a_{i+1}, let l_i be the bond length representing the distance between them. Also, for every three consecutive beads a_{i-1}, a_i, and a_{i+1}, let θ_i represent the bond angle corresponding to the relative position of the third bead with respect to the line containing the previous two. Likewise, for every four consecutive beads a_{i-2}, a_{i-1}, a_i, and a_{i+1}, let φ_i represent the torsion angle corresponding to the relative position of the fourth bead with respect to the plane containing the previous three. Hence, the molecular conformation problem is to determine a set of bond lengths l_i, i=1,...,n-1, bond angles θ_i, i=2,...,n-1, and torsion angles φ_i, i=3,...,n-1, which properly represent the native state of the molecule. See Figure 1 for an example.

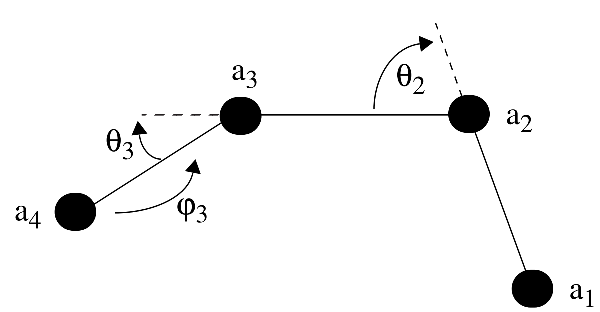

FIGURE 1.
Native Conformation for a Four "Bead" Sequence

If $y \in \mathbf{R}^{3n-3}$ is defined to be the vector of (l_i,θ_i,φ_i) triples for i=1,...,n-1 (where $\theta_1 = \varphi_1 = \varphi_2 = 0$ can be assumed) and f(y) is an appropriate potential energy function, then the *continuous* global minimization approach for solving the molecular conformation problem is simply

$$\min \quad f(y) \qquad \text{(EQ 1)}$$

Because of the large number of state variables needed to define a minimum energy conformation and the possibly exponential number of local minimizers which can occur on the energy surface (Hoare 1979), a direct computation of the global minimum in this fashion is not practical. Instead, the minimization can be carried out in two stages. In the first stage the state variables are discretized in order to form a 3-dimensional lattice. A minimization over this discrete space provides a suitable starting point for the second stage. In this second stage, the lattice restrictions are relaxed, and a possibly lower energy function value may be obtained by a continuous minimization with respect to the variables y.

3. Quadratic Assignment Formulation

In order to formulate a discrete approximation to the molecular conformation problem, the original continuous problem in 3-dimensional space is approximated by a discrete problem using a suitable 3-dimensional lattice with N sites, $N \geq n$. If s_j represents lattice site j, for j=1,...,N, then a total of nN zero-one variables x_{ij} are sufficient to completely determine the assignment of the beads a_i to the lattice sites s_j. More precisely, if $x_{ij} = 1$, then a_i is assigned to lattice site s_j. If $x_{ij} = 0$, then a_i is not assigned to lattice site s_j. Only two types of constraints are required:

1) Each bead must occupy exactly one lattice site. Hence,

$$\sum_{j=1}^{N} x_{ij} = 1, \qquad i=1,...,n. \qquad \text{(EQ 2)}$$

2) At most one bead occupies each lattice site s_j, That is,

$$\sum_{i=1}^{n} x_{ij} = 1, \qquad j=1,...,N. \qquad \text{(EQ 3)}$$

The objective function consists of both a linear and a quadratic term. The linear term

$$\sum_{i=1}^{n} \sum_{j=1}^{N} d_{ij} x_{ij}$$

represents the "direct" contribution d_{ij} to the total potential energy when the bead a_i is assigned to lattice site s_j. For example, the polarity (or lack of it) of bead a_i might affect which lattice sites are preferred (hydrophobic beads might prefer to be more in the interior of the folded chain and hence more in the center of the lattice structure, but this is by no means guaranteed). The quadratic term

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{k=1}^{n}\sum_{l=1}^{N} p_{ijkl}x_{ij}x_{kl}$$

represents the "pairwise" contribution p_{ijkl} to the total potential energy when the bead a_i is assigned to lattice site s_j and bead a_k is assigned to lattice site s_l. For example, if each bead represents an amino acid residue, then the pairwise Lennard-Jones potential energy (Fisher 1964) between amino acids is a function of the types of the two amino acids and their distances apart in the lattice structure. That is, the term p_{ijkl} depends only on a_i, a_k, and $\| s_j - s_l \|_2$. Hence, it is convenient to write the total potential energy function in the quadratic form

$$E(x) = c^T x + \frac{1}{2}x^T Q x \qquad \text{(EQ 4)}$$

where $x \in \mathbf{R}^{nN}$ denotes the zero-one vector with elements x_{ij}, $c \in \mathbf{R}^{nN}$ denotes the vector with elements d_{ij}, and $Q \in \mathbf{R}^{(nN)x(nN)}$ is a real symmetric (typically indefinite) matrix with elements p_{ijkl}. It follows that the 3-dimensional minimum energy conformation of the bead sequence restricted to the discrete lattice structure is given by the solution to the following *quadratic assignment problem*:

$$\min_{x \in P} \; E(x) \qquad \text{(EQ 5)}$$

where P consists of constraints (EQ 2), (EQ 3), and the integer restrictions $x_{ij} \in \{0,1\}$ for i=1,...,n and j = 1,...,N.

Clearly, the usefulness of this approach is very heavily dependent on the selection of the "proper" lattice structure, and on the choice of the most appropriate potential function. In the absence of any a priori knowledge about the folded structure of the molecule, the selection of the best lattice structure can be extremely difficult. Obviously, determining this "proper", i.e. perfect, selection is as hard as the conformation problem itself. Hence, it should be clear that a very fine grained, but sufficiently large, lattice must be selected -- the exact structure of which could be spherical, rectangular, or even random. That the lattice might not allow for the global minimum conformation, or even for any *stable* conformation at all, is not important. The lattice restrictions will be removed prior to the second stage of the method, and the "non-stable" conformer will be used as a starting point for a local minimization in which the conformer is allowed to relax into a stable "minimum" (albeit local) energy configuration.

The appropriate choice of the potential energy function is also a crucial factor in guiding the search for conformers toward the minimum energy configuration. As stated above, the energy term p_{ijkl} represents the pairwise contribution to the total potential energy when bead a_i is assigned to lattice site s_j and bead a_k is assigned to lattice site s_l. More specifically, the diagonal elements of Q are zero, i.e. $p_{ijij} = 0$ for i=1,...,n and j=1,...,N since the energy contribution of bead a_i when located at lattice site s_j is already provided by the term d_{ij}. Furthermore, the constraints (EQ 2) and (EQ 3) can be enforced by simply assigning an appropriate penalty value to some of the off-diagonal terms in Q. Specifically, for a sufficiently large constant $\gamma > 0$, set $p_{ijil} = \gamma$ for all i, j, l with $j \neq l$, and $p_{ijkj} = \gamma$ for all i, j, k with $i \neq k$. The first of these will not allow x_{ij} and x_{il} to both be unity (i.e., bead a_i can occupy at most one site). The second of these will not allow x_{ij} and x_{kj} to both be unity (i.e., beads a_i and a_k cannot both occupy lattice site s_j). For each bead a_i, there will usually be a number of sites s_j for which $d_{ij} < 0$, so that a_i will be assigned to some lattice site. Note that if this cannot be guaranteed, then the requirement that each bead be assigned to *exactly* one lattice site could always be satisfied by adding the single constraint

$$\sum_{i=1}^{n} \sum_{j=1}^{N} x_{ij} = n. \qquad \text{(EQ 6)}$$

Finally, in order to enforce the requirement that two consecutive beads a_i and a_{i+1} remain within an allowable distance of the required bond lengths (i.e. an approximate bond length between two consecutive beads is typically known, but a small deviation from this value may be permitted), the term $p_{ij(i+1)l}$ must be very large if the distance between lattice sites s_j and s_l is not within an allowable tolerance. For example, if l_i represents the approximate bond length between beads a_i and a_{i+1}, then one possible choice is to let

$$p_{ij(i+1)l} = \beta_i(\|s_j - s_l\|_2 - l_i)^2$$

where β_i is some constant that determines the penalty to be imposed for large deviations away from l_i. Similarly, it may be desired to force a sequence of three consecutive beads to remain within an allowable tolerance of a predetermined (i.e. fixed) bond angle. Unfortunately, incorporating this component into the *pairwise* energy term p_{ijkl} is not possible because of its dependence on a sequence of *three* beads. Hence, this potential requirement is not explored further in this model. Note, however, that such a requirement may be enforced during the second, or relaxed, stage given by (EQ 1).

For all other terms p_{ijkl} any appropriate energy potential function may be used. One such function of current interest is $p_{ijkl} = f_{ik}(r_{jl})$, where f_{ik} is called the Lennard-Jones pairwise potential between the beads a_i and a_k, and r_{jl} is the distance between the lattice sites s_j and s_l. The function $f_{ik}(r)$ has the following form:

$$f_{ik}(r) = \varepsilon_{ik}\left(\left(\frac{\sigma_{ik}}{r}\right)^{12} - 2\left(\frac{\sigma_{ik}}{r}\right)^{6}\right)$$

where ε_{ik} and σ_{ik} are constants related to the two specific beads (e.g. amino acids) involved. Notice that the minimum of $f_{ik}(r) = -\varepsilon_{ik}$ and is obtained at $r = \sigma_{ik}$. A plot of the function $f_{ik}(r)$ for $\varepsilon_{ik} = \sigma_{ik} = 1$ is given in Figure 2. Also notice that the Lennard-Jones pair potential has the property that $p_{ijkl} = p_{klij}$, since $f_{ik}(r_{jl})$ depends only on the types and relative positions of the beads a_i and a_k.

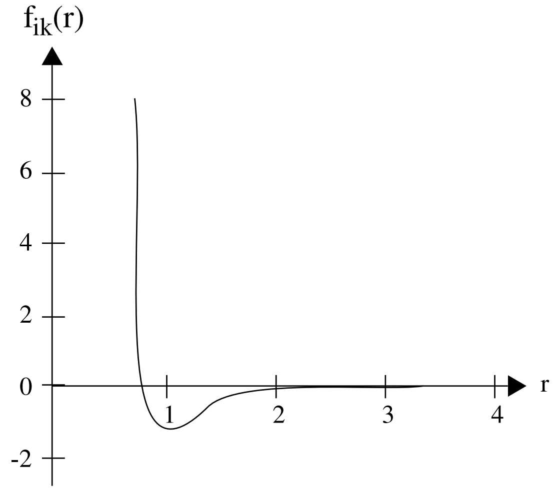

FIGURE 2.
Lennard-Jones Pair Potential Function for $\varepsilon_{ik} = \sigma_{ik} = 1$

This analysis shows that the quadratic assignment problem (EQ 5) is equivalent to an *unconstrained* quadratic zero-one program. One way to solve this is by a branch and bound method applied directly to the quadratic zero-one program. Several such methods are known, but probably the most efficient is one which uses bounds on the gradient components to fix many of the zero-one variables at their optimal values early in the solution process (Pardalos and Rodgers 1990a; Pardalos and Rodgers 1990b)).

4. A Simplified Assignment Problem

A simplified quadratic assignment problem can be formulated for the case where the linear molecule consists of a total of n *identical* elements. For this simpler situation, the problem reduces to assigning one element to each of exactly n selected lattice sites. For a lattice with N sites ($N \geq n$), this gives a total of only N zero-one variables x_i, i=1,...,N. If an element is assigned to site s_j, then $x_j = 1$, otherwise $x_j = 0$.

As in the more general case, the total potential energy consists of a linear and a quadratic term. The linear term, representing the direct energy contribution d_j of an element assigned to site s_j, will be given by

$$\sum_{j=1}^{N} d_j x_j .$$

The sum of the pairwise energy contributions p_{ij}, when two elements are assigned to sites s_i and s_j, gives the symmetric quadratic term

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}x_ix_j$$

where $p_{ii} = 0$ since the direct contributions are already accounted for by the linear term. For this simpler case, the total potential energy can be written

$$E(x) = c^Tx + \frac{1}{2}x^TQx$$

where $x \in \mathbf{R}^N$ with zero-one elements x_i, $c \in \mathbf{R}^N$ with elements d_i, and $Q \in \mathbf{R}^{NxN}$ is a real symmetric matrix with elements p_{ij}. This problem can therefore be stated in the form

$$\min_{x \in P_n} E(x)$$

where

$$P_n = \left\{x_i : \sum_{i=1}^{N} x_i = n,\ x_i \in \{0, 1\}\right\}.$$

5. Concave Quadratic Global Minimization Formulation

The (discrete) quadratic assignment problem (EQ 5) can easily be shown to be equivalent to the (continuous) minimization of a strictly concave quadratic function over a polytope. In particular, let λ_{max} be the maximum (real) eigenvalue of the matrix $Q \in \mathbf{R}^{(nN)x(nN)}$, and let $\mu = 1 + \lambda_{max}$. Then, since $(x_{ij})^2 = x_{ij}$ (recall that $x_{ij} \in \{0,1\}$), the energy function $E(x)$ can be rewritten as

$$E'(x) = c'^Tx + \frac{1}{2}x^TQ'x \tag{EQ 7}$$

where $c' = c + (\mu/2)e$, and $Q' = Q - \mu I$ is a symmetric negative definite matrix. Note that $E(x) = E'(x)$.

It is well known that the global minimum of a strictly concave quadratic function is attained at an extreme point of the feasible polytope (Rockafellar 1970). By relaxing the integer restrictions in the polytope P above to get the new polytope P' consisting of the constraints (EQ 2), (EQ 3), and the bounds $0 \le x_{ij} \le 1$ for i=1,...,n and j = 1,...,N, it is easy to see that the extreme points of P' correspond to the feasible points of P. Hence, a global minimum for the strictly concave quadratic problem

$$\min_{x \in P'} E'(x) \tag{EQ 8}$$

will also be a global minimum of the discrete quadratic assignment problem (EQ 5).

The original quadratic assignment problem (EQ 5) is therefore equivalent to the minimization of a concave quadratic function on the unit hypercube with (possibly) the one linear equality constraint (EQ 6) and nN variables, each of which is restricted to the interval [0,1]. The global, or near global, solution to this problem provides a convenient starting point for the "relaxed" *local* minimization problem (EQ 1).

Two different computational methods for this class of problem have recently been developed. The first such method finds the global minimum of a concave quadratic function on a polytope by the use of linear underestimating functions (Phillips and Rosen 1988). An important feature of this method is that for every local minimum obtained, a bound is also computed which bounds the difference between the local and (unknown) global minimum function values. Furthermore, this bound can be made as small as desired, at the cost of additional computation. Some recent improvements in this algorithm are described in Phillips and Rosen (1993).

The second method is essentially stochastic, and finds a large number of local minima by solving multiple cost row linear programs (Phillips, Rosen, and van Vliet 1992). This method is very well suited for parallel implementation because each local minimization can be performed as a completely independent calculation. A detailed computational comparison of these two methods has recently been completed (Phillips and Rosen 1992) and shows that the first method is faster in most cases, but the second method often gives additional useful information on local minima with function values close to the global minimum.

The solution of the quadratic assignment problem via global minimization (EQ 8) will give the conformation with the minimum potential energy over all possible conformations *on the chosen lattice*. Clearly the requirement that each bead must be located exactly at a lattice site is a restriction on the allowable conformations, which may prevent attainment of the minimum energy state. Therefore the lattice minimization given by the solution of (EQ 8) is considered to be the *first stage* of a two-stage process. The second stage consists of eliminating the lattice restriction and directly minimizing the potential energy function f(y) as described in section 2. The key to this second stage computation is that only a local minimization of f(y) is required starting with $y=y_0$, where y_0 represents the minimum energy configuration obtained from the first stage. This second stage minimization is therefore an unconstrained local minimization with 3(n-1) variables (recall that $\theta_1 = \varphi_1 = \varphi_2 = 0$ are fixed). Several efficient computational methods are available for this purpose as well (Liu and Nocedal 1989; Nash 1985).

6. Lattice Assignment to Satisfy a Distance Matrix

An important molecular structure problem, closely related to the molecular conformation problem, is that of finding a configuration (although not necessarily a minimum energy one) which satisfies a specified “distance matrix”. The distance matrix consists of up to n(n-1)/2 positive quantities δ_{ik} which specify the (approximate) distance between some, or all, pairs of elements a_i and a_k (e.g. amino acids, atoms, etc.) of the molecule (note that not all of the distances between pairs need to be specified). That is, the quantity δ_{ik}, usually measured experimentally, is the approximate “desired” distance between elements a_i and a_k. Hence, the problem is to find a lattice assignment of the elements a_i, for i=1,...,n, which gives the “best fit” (in the least squares sense) to the specified distance matrix.

The distance between lattice sites s_j and s_l is given by $\| s_j - s_l \|_2$, so if an assignment of a_i to site s_j and a_k to site s_l is made, then the error in satisfying the specified distance δ_{ik} (from the distance matrix) can be measured by $(\delta_{ik} - \| s_j - s_l \|_2)^2$. Therefore, an assignment is desired such that the sum of all such terms is minimized.

Let $K = \{ (i,k) : \delta_{ik} > 0 \}$ be the set of all specified distances. Then the desired assignment is that which minimizes the quadratic function

$$\frac{1}{2}\sum_{(i,\,k)\,\in\,K}\left(\sum_{j=1}^{N}\sum_{l=1}^{N} p_{ijkl}x_{ij}x_{kl}\right)$$

where $p_{ijkl} = (\delta_{ik} - \| s_j - s_l \|_2)^2$. This problem is therefore of the form (EQ 5), with E(x) given by (EQ 4) with c = 0. A significant fact about this special case is that the lower bound $E(x) \geq 0$ is known. If the lattice permits an assignment which exactly satisfies the distance matrix, then E(x) = 0 for any optimal assignment.

7. A Sample Test Problem

As initial test cases, the global optimization techniques described above can be applied to several simple, tractable molecular structures for which the conformational space has been exhaustively explored. One such simple structure is butane, a system of four hydrocarbons which satisfy the same characteristic folding models as do small peptides. The 3-dimensional *trans* conformation for butane (the conformation with minimum energy) is well known and is shown in Figure 3. In this case, since all of the molecules are actually identical hydrocarbons, the complexity of the energy potential function should be reduced. One energy potential function f(y), $y \in \mathbf{R}^9$ (with fixed $\theta_1 = \varphi_1 = \varphi_2 = 0$), for butane that has gained widespread acceptance is

$$f(y) = K_l\sum_{i=1}^{3}(l_i - l_0)^2 + K_\theta\sum_{i=2}^{3}(\theta_i - \theta_0)^2 + V_{3/2}(1 + \cos 3\varphi_3) + \varepsilon_{14}\left(\left(\frac{\sigma_{14}}{r_{14}}\right)^{12} - 2\left(\frac{\sigma_{14}}{r_{14}}\right)^{6}\right)$$

where r_{14} is the distance between the first and fourth molecules (and is therefore completely dependent on the variables l, θ, and φ), $K_l = 310.0$ kcal/ang^2, $K_\theta = 40.0$ kcal/rad^2, $V_{3/2} = 1.3$ kcal, $\varepsilon_{14} = 0.06$ kcal, $\sigma_{14} = 3.6$ ang (i.e. the well depth and location for the minimum of the lone Lennard-Jones pairwise term are 0.06 kcal and 3.6 ang, respectively), $l_0 = 1.526$ ang, and $\theta_0 = 70.5°$ (recall that in this formulation θ_i represents the bond angle corresponding to the relative position of the third bead with respect to the line containing the previous two).

Discretizing this problem using a 5x5x5 lattice with a uniform grid spacing of 1.4 ang, we get the quadratic assignment problem (EQ 5) with 500 zero-one variables and 129 linear constraints. Of course the equivalent concave quadratic global minimization formulation also requires 500 variables (although not restricted to be zero-one), but only requires one linear constraint (EQ 6) and 500 upper and lower bounds on the values of the x_{ij}. The terms p_{ijkl} of energy function E(x)

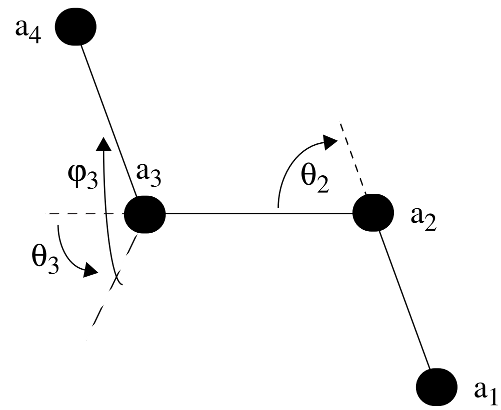

FIGURE 3.
"United Atom" Butane in its *Trans* Conformation
$l_1 = l_2 = l_3 = 1.526$ ang; $\theta_2 = \theta_3 = 70.5°$; $\varphi_3 = 180°$

defined in (EQ 4), which are used to construct the concave function E'(x) defined in (EQ 7), can be chosen as follows:

$$p_{ijkl} = \begin{cases} 0 & (i=k) \text{ and } (j=l) \\ M & (i=k) \text{ or } (j=l)\text{, but not both} \\ K_l(\|s_j - s_l\|_2 - l_0)^2 & (k=i+1) \text{ and } (j \neq l) \\ f_{ik}(r_{jl}) & \text{otherwise} \end{cases}$$

where M is a sufficiently large penalty term, and $f_{ik}(r_{jl})$ is the Lennard-Jones pair potential with ε_{ik} = 0.06 kcal, σ_{ik} = 3.6 ang for all i and k. Notice that the terms in f(y) involving θ and φ were not included in the discrete approximation E(x), since they cannot be modeled as pairwise interactions.

The global, or near global, solution (there were a total of 23 local minima found) to this *lattice restricted* problem has an energy of 319.293 kcal with a 3-dimensional configuration as shown in Figure 4. The result of this minimization provides the values l_1 = 1.4 ang, l_2 = 1.98 ang, l_3 = 2.42 ang, $\theta_2 = \theta_3 = \varphi_3 = 45°$, and r_{14} = 3.43 ang. This particular lattice restricted result is closest to the *gauche*$^-$ conformation of butane (see Figure 5), which corresponds to a local minimum of the energy function, but not the global one (*trans* is global). Because of the small size of the lattice (125 total sites) and the relatively large spacing between lattice points (1.4 ang spacing with all desired bond lengths of 1.526 ang), the lattice result cannot obtain a minimum energy conformer without violating some of the desired bond lengths. In addition, because the lattice used was only five sites wide in each dimension, and the first two beads were fixed in the center of the lattice, it was not possible to obtain a conformer close to the the global *trans* conformation (this would require a lattice with a width of at least six sites in one dimension). As stated at the end of section 5, this conformer should now be used as a starting point for the local minimization of f(y) over a continuous domain in 3-dimensional space. The time to obtain this approximate solution (the first stage result only) to a very simple example was approximately 25 minutes on a Cray X-

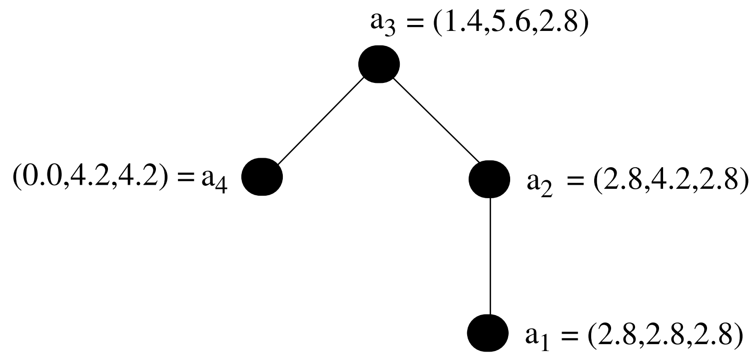

FIGURE 4.
Lattice Global Minimum Conformer for "United Atom" Butane
5x5x5 Grid with 1.4 Ang Uniform Spacing

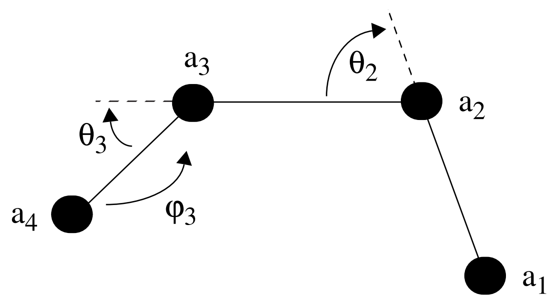

FIGURE 5.
"United Atom" Butane in its *Gauche*⁻ Conformation
$l_1 = l_2 = l_3 = 1.526$ ang; $\theta_2 = \theta_3 = 70.5°$; $\varphi_3 = -60°$

MP/464 supercomputer! Hence, we can only suggest that this approach be considered as a purely theoretical method unless/until more efficient global optimization techniques are developed.

8. Conclusions

The molecular conformation problem is a difficult and complex problem. This paper has presented an approach that models folding as a two stage process. The first stage involves a discrete lattice-type approximation that permits the original continuous model to be formulated as a "zero-one" quadratic assignment problem, and then further, as a continuous concave quadratic global minimization problem. The solution to this first stage is then used as a starting point for a relaxed continuous local minimization step which, given the correct energy function, should provide an accurate prediction of the native state of the molecule.

9. Acknowledgments

This research was supported in part by the Air Force Office of Scientific Research grant AFOSR 91-0147 and the Minnesota Supercomputer Institute.

10. References

Fisher, I.Z. 1964. Statistical Theory of Liquids. The University of Chicago Press: Chicago, IL.

Gierasch, L.M., and J. King. 1990. Protein folding: Deciphering the second half of the genetic code. *American Association for the Advancement of Science*.

Hoare, M.R. 1979. Structure and dynamics of simple microclusters. *Advances in Chemical Physics* 40:49-135.

Liu, D.C., and J. Nocedal. 1989. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* 45:503-528.

Maier, R.S., J.B. Rosen, and G.L. Xue. 1992. A discrete-continuous algorithm for molecular energy minimization. In proceedings of the *IEEE/ACM Supercomputing '92*, Minneapolis, MN, Nov. 16-20, 1992.

Nash, S.G. 1985. Preconditioning of truncated-newton methods. *SIAM Journal on Scientific and Statistical Computing* 6:599-616.

Pardalos, P.M., and G.R. Rodgers. 1990a. Computational aspects of a branch and bound algorithm for quadratic zero-one programming. *Computing* 45:131-144.

Pardalos, P.M., and G.R. Rodgers. 1990b. Parallel branch and bound algorithms for quadratic zero-one programs on the hypercube architecture. *Annals of Operations Research* 22:271-292.

Phillips, A.T., and J.B. Rosen. 1988. A parallel algorithm for constrained concave quadratic global minimization. *Mathematical Programming* 42:421-448.

Phillips, A.T., and J.B. Rosen. 1992. A computational comparison of two methods for constrained global optimization. *UMSI 92/100*. April 1992.

Phillips, A.T., and J.B. Rosen. 1993. Sufficient conditions for solving linearly constrained separable concave global minimization problems. *Journal of Global Optimization* 3(1):79-94.

Phillips, A.T., J.B. Rosen, and M. van Vliet. 1992. A parallel stochastic method for solving linearly constrained concave global minimization problems. *Journal of Global Optimization* 2(3):243-258.

Rockafellar, R.T. 1970. Convex Analysis. Princeton University Press: Princeton, NJ.

Troyer, J.M, and F.E. Cohen. 1991. Simplified models for understanding and predicting protein structure, *Reviews in Computational Chemistry* 2, K.B. Lipkowitz and D.B. Boyd (eds), VCH Publishers: New York, NY.

Xue, G.L. 1992. Parallel two-level simulated annealing: applications to molecular conformation. *AHPCRC 92-047*. April 1992.

Xue, G.L., R.S. Maier, and J.B. Rosen. 1992. Minimizing the Lennard-Jones potential function on a massively parallel computer. In proceedings of the *ACM 1992 International Conference on Supercomputing*, Washington, DC, July 19-23, 1992.