

An approach to the multiple-minima problem by relaxing dimensionality

(distance geometry/Cayley-Menger determinants/energy minimization/distance constraints)

ENRICO O. PURISIMA AND HAROLD A. SCHERAGA[†]

Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853-1301

Contributed by Harold A. Scheraga, November 26, 1985

ABSTRACT A method is presented for starting from a very-low-energy high-dimensional conformation and obtaining a low-energy three-dimensional structure by gradual contraction of the dimensionality. The contraction in dimensionality is achieved by use of Cayley-Menger determinants, of which a simplified form is derived here. Preliminary results are presented for a virtual-bond pentapeptide and for full-atom representations of several terminally blocked amino acids.

The multiple-minima problem is a severe one in the use of energy minimization in calculations of molecular conformation. Current minimization algorithms are very efficient in finding local minima but, having found a local minimum, are then trapped in the potential well of that minimum. Various approaches to the multiple-minima problem have been suggested in several recent papers (1-4). The build-up procedure is essentially an efficient and systematic search of the conformational energy space by a judicious choice of starting conformations from which to carry out minimizations (1, 2). The Monte Carlo approach avoids minimization altogether and attempts to sample conformational energy space efficiently to locate the energetically favorable regions (3). In the annealing approach, when the minimization becomes trapped in a local minimum, the temperature of the system is raised and a Monte Carlo procedure is carried out to allow the system to escape from the local potential well (4).

In this paper, we present a method for relaxing a system, not by raising the temperature but by raising the dimensionality of the space. The idea is that, in higher dimensional space, there are many more degrees of freedom in which the atoms can move about, making it easier to adjust to a low-energy conformation. Many potential barriers in three dimensions will not exist in higher dimensions. A method is presented for starting from a very-low-energy high-dimensional conformation and obtaining a low-energy three-dimensional structure from it by gradual contraction of the dimensionality. The method may also be used for escaping from a three-dimensional local potential well by starting from a three-dimensional minimum, raising the dimensionality, and then contracting it back to three dimensions.

Method

We take as the primary variables the interatomic distances themselves and use ECEPP (5, 6) (empirical conformational energy program for peptides) as the potential energy function. Like most other empirical potential functions, ECEPP is a pair-interaction potential. Since our variables are the interatomic distances themselves, the potential energy separates out into independent terms, $E = \sum f_{ij}$, where f_{ij} depends on one variable only— d_{ij} , the interatomic distance between atoms i and j . The minimum value of E is therefore attained when each f_{ij} is at its minimum value. Given the

simple nature of the f_{ij} s (Lennard-Jones, hydrogen bonding, electrostatic and torsional functions of d_{ij}), it is a simple exercise to find the value of each d_{ij} that minimizes f_{ij} . The energy corresponding to this set of values of d_{ij} then gives a lower bound on the minimum energy that the molecule can attain.

These distances, of course, will not correspond to any realizable three-dimensional structure. The distances are satisfiable by a structure only in some high-dimensional space (in general, as high as $n - 1$ dimensions, where n is the number of points in the structure). Our initial structure is then a high-dimensional one with the absolute global minimum of the energy; i.e., no other structure, whatever its dimension, can have a lower energy.

As we reduce the dimensionality of the structure, its energy will increase because the distances become more constrained. The energy will continue to increase until we reach a three-dimensional energy minimum. It is important to note that this three-dimensional energy minimum is approached from *below* rather than from above. In the usual energy minimization, we start from a high-energy structure and minimize. Because of the multiple-minima problem, we reach the nearest local minimum and, in general, there are many such local minima between the starting point and the global minimum. Since traditional energy minimization approaches the global minimum from *above*, the minimization leads only to a relatively high-energy minimum. By approaching the three-dimensional global energy minimum from below, we increase the likelihood of ending up in a low-energy three-dimensional minimum. The method is similar in spirit to Crippen's energy-embedding procedure (7, 8). The major difference lies in the technique used to achieve the gradual contraction of the dimensionality.

The key to this gradual contraction of the dimensionality is the following theorem from distance geometry which we restate here in somewhat more physical terms (9).

THEOREM. A set of n points in $n - 1$ dimensional space is embeddable in three dimensions if the following conditions hold:

(i) There exist four points (p_1, p_2, p_3, p_4) in the set that are exactly three-dimensional.

(ii) The four-dimensional volume formed by the simplex of points (p_1, p_2, p_3, p_4, p_i) is zero for all $i = 1, \dots, n$.

(iii) The five-dimensional volume formed by the simplex of points ($p_1, p_2, p_3, p_4, p_i, p_j$) is zero for all $i, j = 1, \dots, n$.

The square of the four- and five-dimensional volumes of conditions (ii) and (iii) can be calculated using Cayley-Menger (CM) determinants (9) as follows.

Four dimensional:

$$V^2 = \frac{(-1)^{4+1}}{2(4)(4!)^2} \text{CM}(R; p_i), \quad [1]$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ECEPP, empirical conformational energy program for peptides; CM, Cayley-Menger.

[†]To whom reprint requests should be addressed.

Five dimensional:

$$V^2 = \frac{(-1)^{5+1}}{2(5)(5!)^2} \text{CM}(R; p_i, p_j), \quad [2]$$

where

$$\text{CM}(R; p_i) = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & r_{12} & r_{13} & r_{14} & t_{1i} \\ 1 & r_{21} & 0 & r_{23} & r_{24} & t_{2i} \\ 1 & r_{31} & r_{32} & 0 & r_{34} & t_{3i} \\ 1 & r_{41} & r_{42} & r_{43} & 0 & t_{4i} \\ 1 & t_{1i} & t_{2i} & t_{3i} & t_{4i} & 0 \end{vmatrix} \quad [3]$$

$$\text{CM}(R; p_i, p_j) = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & r_{12} & r_{13} & r_{14} & t_{1i} & t_{1j} \\ 1 & r_{21} & 0 & r_{23} & r_{24} & t_{2i} & t_{2j} \\ 1 & r_{31} & r_{32} & 0 & r_{34} & t_{3i} & t_{3j} \\ 1 & r_{41} & r_{42} & r_{43} & 0 & t_{4i} & t_{4j} \\ 1 & t_{1i} & t_{2i} & t_{3i} & t_{4i} & 0 & s_{ij} \\ 1 & t_{1j} & t_{2j} & t_{3j} & t_{4j} & s_{ij} & 0 \end{vmatrix} \quad [4]$$

and R refers to the four reference points. The r_{kms} are the squared distances between reference points k and m . The t_{ki} s and t_{kj} s are the squared distances between reference point k and points p_i and p_j , respectively. The t_{ki} s are also referred to as CM coordinates (10). s_{ij} is the squared distance between points p_i and p_j . The condition that the volumes vanish translate into the following equations.

$$\text{CM}(R; p_i) = 0 \quad \text{for all } i = 1, \dots, n \quad [5]$$

$$\text{CM}(R; p_i, p_j) = 0 \quad \text{for all } i, j = 1, \dots, n. \quad [6]$$

Simplification of the Determinants

The evaluation of a determinant is an expensive calculation. Since the minimization will require $O(n^2)$ determinants for each evaluation of an object function, it is important to calculate the determinant constraints as efficiently as possible. A derivation of a simplified form of the determinants using Jacobi's theorem has been given by Sippl and Scheraga (10). We present here an alternative derivation, using Gaussian elimination. (The notation in the following discussion has been kept as close to theirs as possible.)

Consider $\text{CM}(R; p_i, p_j)$. Using Gaussian elimination, we can calculate this determinant by transforming the matrix involved into a partially upper triangular form. This will not alter the value of the determinant (except perhaps for a change of sign). We first exchange rows 1 and 2. Then we perform Gaussian elimination on the remaining rows to zero out all subdiagonal elements from columns 1-5. Without actually carrying out the Gaussian elimination in detail, it is clear that the resulting partially upper triangular matrix must be of the form

$$-\text{CM}(R; p_i, p_j) = \begin{vmatrix} 1 & 0 & r_{12} & r_{13} & r_{14} & t_{1i} & t_{1j} \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & R_1 & R_2 & R_3 & A_1 & B_1 \\ 0 & 0 & 0 & R_4 & R_5 & A_2 & B_2 \\ 0 & 0 & 0 & 0 & R_6 & A_3 & B_3 \\ 0 & 0 & 0 & 0 & 0 & U & s_{ij} + W \\ 0 & 0 & 0 & 0 & 0 & s_{ij} + W & V \end{vmatrix} \quad [7]$$

where the R s involve only the r_{kms} . The A s are linear combinations of the t_{ki} s and the B s are linear combinations of the t_{kj} s with coefficients involving the r_{kms} . U and V are second-order polynomial functions of the t_{ki} s and t_{kj} s, respectively. W is a function of both the t_{ki} s and t_{kj} s.

Expanding Eq. 7 gives

$$\text{CM}(R; p_i, p_j) = \text{CM}(R) [U V - (s_{ij} + W)^2], \quad [8]$$

where

$$\text{CM}(R) \equiv -R_1 R_4 R_6.$$

Except for a multiplicative factor, $\text{CM}(R)$ is in fact the squared volume of the reference simplex; i.e., it is the CM determinant of the four reference points. Furthermore, the submatrix made up of the first six rows and columns in Eq. 7 is just the upper triangular form of the matrix in Eq. 3. This implies that

$$\text{CM}(R; p_i) = \text{CM}(R)U \quad [9]$$

and, analogously,

$$\text{CM}(R; p_j) = \text{CM}(R)V. \quad [10]$$

It should be noted that the form of $\text{CM}(R; p_i, p_j)$ in Eq. 8 is equivalent to the result obtained in ref. 10; i.e.,

$$\text{CM}(R; p_i, p_j) = [\text{CM}(R; p_i)\text{CM}(R; p_j) - \text{CM}^*(R; p_i, p_j)^2]/\text{CM}(R), \quad [11]$$

where we can make the identification

$$\text{CM}^*(R; p_i, p_j) = \text{CM}(R)(s_{ij} + W). \quad [12]$$

We now show that U is really a special case of W . Consider the determinant of the submatrix formed by deleting row 7 and column 6 in the determinant in Eq. 4 and by replacing s_{ij} by zero. Using Gaussian elimination, we have

$$\begin{vmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & r_{12} & r_{13} & r_{14} & t_{1j} \\ 1 & r_{21} & 0 & r_{23} & r_{24} & t_{2j} \\ 1 & r_{31} & r_{32} & 0 & r_{34} & t_{3j} \\ 1 & r_{41} & r_{42} & r_{43} & 0 & t_{4j} \\ 1 & t_{1i} & t_{2i} & t_{3i} & t_{4i} & 0 \end{vmatrix} = - \begin{vmatrix} 1 & 0 & r_{12} & r_{13} & r_{14} & t_{1j} \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & R_1 & R_2 & R_3 & B_1 \\ 0 & 0 & 0 & R_4 & R_5 & B_2 \\ 0 & 0 & 0 & 0 & R_6 & B_3 \\ 0 & 0 & 0 & 0 & 0 & W \end{vmatrix}, \quad [13]$$

where the elements of the array on the right-hand side of the equation have the same meaning as in Eq. 7. The right-hand side of Eq. 13 is equal to $\text{CM}(R)W$. When $i = j$, the left-hand side of Eq. 13 is just $\text{CM}(R; p_i)$, which, by Eq. 9, is $\text{CM}(R)U$. We then have

$$W(i, i) = U(i), \quad [14]$$

where we have explicitly indicated the dependence of W on i and j and of U on i .

Three-dimensionality requires that Eqs. 5 and 6 be satisfied simultaneously. Eqs. 5 and 9 imply that $\text{CM}(R)U = 0$ for all $i = 1, \dots, n$. Setting $U = 0$ in Eq. 8, and setting the right-hand side to 0, gives

$$\text{CM}(R)(s_{ij} + W)^2 = 0 \quad \text{for all } i, j = 1, \dots, n, \quad [15a]$$

which, since $\text{CM}(R) \neq 0$, is equivalent to

$$\text{CM}(R)(s_{ij} + W) = 0 \quad \text{for all } i, j = 1, \dots, n. \quad [15b]$$

It should be noted that, when $i = j$, $s_{ij} = 0$ and $W(i, i) =$

$U(i)$. Eq. 15b then reduces to $CM(R)U = 0$. We then see that Eq. 15b replaces both Eqs. 5 and 6 as the condition for the vanishing of both the four- and five-dimensional volumes.

Gaussian elimination has given us the form of the simplified equation that will guarantee the vanishing of the four- and five-dimensional volumes. To obtain an explicit expression for W in terms of the t_{ki} s, we carry out an expansion by minors of the determinant on the left-hand side of Eq. 13 using row 6 and column 6. This yields the expression

$$CM(R)W(i, j) = \sum_{k=0}^4 \sum_{m=0}^4 A_{km} t_{ki} t_{mj}, \quad [16]$$

where

$$t_{0i} = t_{0j} = 1, \quad A_{km} = A_{mk}.$$

The A_{km} s are the cofactors in the expansion and depend only on the r s. In compact matrix notation

$$CM(R)W(i, j) = \mathbf{t}_i \mathbf{A} \mathbf{t}_j^T, \quad [17]$$

where

$$\mathbf{t}_i = (1 \ t_{1i} \ t_{2i} \ t_{3i} \ t_{4i}), \quad \mathbf{t}_j = (1 \ t_{1j} \ t_{2j} \ t_{3j} \ t_{4j}).$$

The condition for three-dimensionality (Eq. 15b) then becomes

$$CM(R)s_{ij} + \mathbf{t}_i \mathbf{A} \mathbf{t}_j^T = 0 \quad \text{for all } i, j = 1, \dots, n. \quad [18]$$

Once a set of distances satisfying Eq. 18 is obtained, an embedding procedure can be used to recover the cartesian coordinates from the distances (11).

t_{ki} and s_{ij} Distances

In the discussion above and in Eq. 18 in particular, we can distinguish between two kinds of distances—i.e., t_{ki} s and s_{ij} s. Sippl and Scheraga (10) have pointed out that, when the set of n points is embeddable in E^3 , three-dimensional Euclidean space, then the t_{ki} s alone are sufficient to determine the whole structure completely because the s_{ij} s may be calculated from the t_{ki} s using

$$s_{ij} = -\mathbf{t}_i \mathbf{A} \mathbf{t}_j^T / CM(R). \quad [19]$$

Stated another way, the t_{ki} s generate a coordinate system [called CM coordinates (10)] for E^3 . In the context of minimization, an objective function with distances s_{ij} and t_{ki} as primary variables could, in principle, be recast as a function involving only the t_{ki} distances. The advantage of using CM coordinates is that the number of variables is reduced to $O(4N)$ instead of $O(N^2)$ if we were to include the s_{ij} s explicitly.

In this study, however, we chose to use the s_{ij} s and t_{ki} s as variables in the minimization and not the CM coordinates alone, for two reasons. First, we have *a priori* information about suitable starting values for the s_{ij} s from energetic considerations; i.e., each s_{ij} minimizes $E(s_{ij})$, where $E(s_{ij})$ is the component of the energy involving only s_{ij} . It is not clear how to translate this information into information about starting values for the CM coordinates without losing much of the information. Second, some of the s_{ij} s calculated from the CM coordinates may be negative—i.e., negative squared interatomic distances. This makes the calculation of such quantities as the electrostatic and nonbonded energy ill-defined. This possibility is thus a further complication in the choice of starting values for the CM coordinates.

In the section below, we discuss the components of the objective function being minimized.

Components of the Objective Function

The objective function to be minimized is given by

$$F = w_E F_E + w_{4D} F_{4D} + w_{5D} F_{5D} + w_B F_B. \quad [20]$$

The w s are the weights of the various components of the objective function. F_E is the energy component. The particular energy function used here is ECEPP (5, 6), which consists of Lennard-Jones, hydrogen bonding, and electrostatic and torsional terms, all of which are easily recast as functions of the interatomic distances themselves. The torsional terms do require a slight modification, however. The two- and three-fold torsional potentials of ECEPP, expressed in terms of distances, are

$$U_2 = U_0[1 - C^2] \\ U_3 = (U_0/2)[4 C^3 - 3 C + 1], \quad [21]$$

where

$$C = (d^2 - A)/B \\ A = (d_{\min}^2 + d_{\max}^2)/2 \\ B = (d_{\min}^2 - d_{\max}^2)/2.$$

C is really $\cos \chi$ expressed in terms of distances, where χ is the dihedral angle. The distance involved is the distance between the two terminal atoms in the set of four atoms that define a torsional dihedral angle. The problem with Eq. 21 is that, since $B < 0$, the torsional potential blows up to negative infinity as d increases beyond d_{\max} . To remedy this, a repulsive term for $d > d_{\max}$ is added; i.e.,

$$20(d^2 - d_{\max}^2)^2 \quad \text{if } d > d_{\max}. \quad [22]$$

This extra term is added to the torsional potential to produce a steep barrier in the neighborhood of d_{\max} , preventing d from increasing much beyond d_{\max} .

F_{4D} and F_{5D} are the CM-determinant constraints on the four- and five-dimensional volumes. These are penalty functions of the form

$$F_{4D} = \sum (\mathbf{t}_i \mathbf{A} \mathbf{t}_i^T)^2 \quad [23]$$

$$F_{5D} = \sum_{i < j} [CM(R)d_{ij}^2 + \mathbf{t}_i \mathbf{A} \mathbf{t}_j^T]^2. \quad [24]$$

By forcing F_{4D} and F_{5D} to zero, the condition for three-dimensionality, Eq. 18, is satisfied. Eq. 22, however, tends to overweight large d_{ij} s over small ones; i.e., it magnifies small errors in large d_{ij} . To remove this bias, two alternative reformulations of Eq. 18 are used

$$F_{5D} = \sum_{i < j} [CM(R) + \mathbf{t}_i \mathbf{A} \mathbf{t}_j^T / d_{ij}^2]^2 \quad [25]$$

and

$$F_{5D} = \sum_{i < j} [CM(R)d_{ij} + \mathbf{t}_i \mathbf{A} \mathbf{t}_j^T / d_{ij}]^2. \quad [26]$$

Eqs. 25 and 26 are obtained by dividing Eq. 18 by s_{ij} and d_{ij} , respectively.

F_B is composed of upper- and lower-bound terms. F_B allows us to incorporate any *a priori* information we may have about the allowed range of values of the distances. These bounds may come from covalent geometry, e.g., upper and lower bounds on distances between atoms separated by one variable dihedral angle, or may be bounds from experimental or other theoretical considerations. F_B is a sum of penalty

terms of the form

$$\begin{aligned} &(d_{ij} - u_{ij})^2 && \text{if } d_{ij} > u_{ij} \\ &(d_{ij} - \ell_{ij})^2 && \text{if } d_{ij} < \ell_{ij} \\ &0 && \text{otherwise.} \end{aligned} \quad [27]$$

The u_{ij} and ℓ_{ij} are user-supplied bounds.

Preliminary Results

Virtual-Bond Test Case. Before applying the procedure to a real polypeptide, the method was first tested on a much simpler system. This is a test case described by Crippen (7, 8) which consists of five points simulating a (virtual-bond) linear pentapeptide. These points interact with each other through a harmonic potential with distances in angstrom units and energy in arbitrary units

$$E = \sum K_{ij}(d_{ij} - R_{ij})^2, \quad [28]$$

where the target-distance and force-constant matrices were chosen as (7, 8)

$$\mathbf{R} = \begin{pmatrix} 0.0 & 3.8 & 4.0 & 4.1 & 4.5 \\ 3.8 & 0.0 & 3.8 & 4.3 & 4.4 \\ 4.0 & 3.8 & 0.0 & 3.8 & 4.7 \\ 4.1 & 4.3 & 3.8 & 0.0 & 3.8 \\ 4.5 & 4.4 & 4.7 & 3.8 & 0.0 \end{pmatrix} \quad [29]$$

and

$$\mathbf{K} = \begin{pmatrix} 0.0 & 20.0 & 1.0 & 1.1 & 1.2 \\ 20.0 & 0.0 & 20.0 & 1.3 & 1.4 \\ 1.0 & 20.0 & 0.0 & 20.0 & 1.5 \\ 1.1 & 1.3 & 20.0 & 0.0 & 20.0 \\ 1.2 & 1.4 & 1.5 & 20.0 & 0.0 \end{pmatrix} \quad [30]$$

The objective function that was minimized was

$$F = \sum K_{ij}(d_{ij} - R_{ij})^2 + w_{4D}F_{4D}. \quad [31]$$

In this particular example, all of the distances are variable and there is no five-dimensional volume to calculate. w_{4D} was increased gradually over a few cycles of minimization to attain three-dimensionality. The resulting distance matrix was

$$\mathbf{D} = \begin{pmatrix} 0.00 & 3.77 & 4.40 & 3.61 & 4.91 \\ 3.77 & 0.00 & 3.77 & 5.16 & 3.98 \\ 4.40 & 3.77 & 0.00 & 3.77 & 5.02 \\ 3.61 & 5.16 & 3.77 & 0.00 & 3.77 \\ 4.91 & 3.98 & 5.02 & 3.77 & 0.00 \end{pmatrix}, \quad [32]$$

with an energy of 2.05 units. This matrix is identical, to within round-off errors, to the distance matrix of the global minimum reported by Crippen, which he obtained by performing an exhaustive search of the conformational space (7, 8). His energy-embedding procedure, however, was unable to locate this global minimum but instead found a local minimum with an energy of 2.58 units.

Tests with Terminally Blocked Amino Acids. The tests with the ECEPP potential were carried out on several terminally blocked amino acids, *N*-acetyl-Xaa-*N'*-methylamide, where Xaa = alanine, lysine, methionine, and serine, respectively.

Since ECEPP is a rigid-geometry potential—i.e., with fixed bond lengths and bond angles—a natural set of four reference points that satisfy condition (i) of the theorem is the set of atoms around a C^α atom—i.e., N, H^α , C^β , and C' . The distances among these points are fixed by the covalent geometry and define a perfectly three-dimensional structure. Since these distances are fixed, $CM(\mathbf{R})$ and the A_{kms} are constants. The variables will be those $t_{ki}s$, $t_{kj}s$, and $s_{ij}s$ that are not fixed by covalent geometry or some other constraint—e.g., a fixed dihedral angle about a peptide bond. [In general, one can alternatively always augment the n points with an external set of four three-dimensional points that will then serve as the reference points for condition (i) of the theorem.]

Upper bounds for the distances were initially assigned values equal to the maximum length that the molecule can adopt in the most extended conformation possible. Lower bounds were initially arbitrarily set to 1.3 Å. Where information was available about the covalent structure, these upper and lower bounds were adjusted accordingly. At this point, further adjustment could have been made if any *a priori* information about the distances were available. The upper and lower bounds were then refined by using the triangle inequalities

$$\begin{aligned} u_{ij} &\leq u_{ik} + u_{jk} \\ \ell_{ij} &\geq \ell_{ik} - u_{jk}. \end{aligned} \quad [33]$$

The starting values for each variable distance were then obtained as the calculated value (within the required upper and lower bounds) that minimized the ECEPP interaction-energy function of the pair of atoms defining that distance. The computations were carried out in two modes, with and without the imposition of a triangle inequality (12) that had to be satisfied by the starting distances.

The weight w_E of the energy component of the objective function was arbitrarily set to 1. w_{4D} and w_{5D} were initially set to values such that

$$|w_E F_E| \approx w_{4D} F_{4D} + w_{5D} F_{5D}. \quad [34]$$

For the results reported here, Eq. 25 was used for F_{5D} . w_E was set to 100. The minimization was carried out in several stages. At each stage, w_{4D} and w_{5D} were typically increased by a factor of 10. The ECEPP energy of the resulting three-dimensional structure was then minimized further using dihedral angles as the variables.

The results for these illustrative examples are shown in Table 1. For alanine, the ECEPP global minimum (13) was attained only when the starting distances were first refined

Table 1. Comparison of calculated and global-minima conformations

| Residue | Degrees | | | | | | |
|---------|---------|--------|----------|----------|----------|----------|----------|
| | ϕ | ψ | χ^1 | χ^2 | χ^3 | χ^4 | χ^5 |
| Ala*† | -80 | 76 | 61 | | | | |
| Lys* | -81 | 75 | -65 | -179 | 180 | 179 | 177 |
| Lys† | -81 | 75 | -65 | -179 | 180 | -179 | 62 |
| Met* | -77 | -30 | -67 | -179 | 180 | 60 | |
| Met† | -79 | 76 | -67 | -179 | 180 | 60 | |
| Ser* | -160 | 159 | 180 | 179 | | | |
| Ser† | -78 | 78 | 68 | 56 | | | |

All other dihedral angles are fixed at 180°.

*Calculated minimum.

†ECEPP global minimum (13).

by applying the triangle inequalities (12) prior to minimization of the objective function.

For lysine, a structure with an energy 0.01 kcal/mol (1 kcal = 4.18 J) higher than the global minimum was attained when the upper bounds were initially increased to five times their value and then gradually reduced to the required value as the minimization progressed.

Under conditions similar to that used for lysine, we obtained structures for methionine and serine that were 0.5 kcal and 3.4 kcal higher than the respective global minima.

Discussion

The major difficulty in attaining the ECEPP global minima is that our method, in a sense, simulates flexible geometry—i.e., a force-field with variable bond lengths and bond angles. This can be seen in the following example of a chain of four points in a two-dimensional space. Suppose the interaction energy between the ends of the chain is given by the curve in Fig. 1. Assume, further, that bond lengths and bond angles are fixed. Given the rigid geometry, only two conformations are possible in two-dimensional space. Let these be represented by conformations a and b in Fig. 2 with the corresponding energies indicated in Fig. 1. In three-dimensional space, conformation c of Fig. 2 has the global minimum of the energy, shown in Fig. 1. As we force the three-dimensional volume of conformation c to zero to make it two-dimensional, conformation c will evolve into conformation a. It will not evolve into conformation d because that would entail an initial *increase* in the three-dimensional volume of the structure. We see then that the algorithm leads to the higher-energy two-dimensional rigid-geometry conformation a instead of the lower-energy conformation b.

If, on the other hand, we allow flexibility in the bond geometry, then conformation d in Fig. 2 would be the two-dimensional flexible-geometry global energy minimum. In this case, conformation c would evolve into conformation d, thereby properly locating the two-dimensional global energy minimum. Of course, the existence of different flexible- and rigid-geometry global energy minima depends on the choice of flexibility parameters for the bond geometry. The point, however, is that, by allowing added dimensions, the algorithm imparts flexible-geometry characteristics to the system even though the bond lengths and bond angles are maintained fixed in the higher-dimensional space.

A similar situation may arise in the use of the rigid-geometry ECEPP potential that could prevent the attainment of some low-energy conformations as the dimensionality of the space is reduced from a high one down to three dimensions.

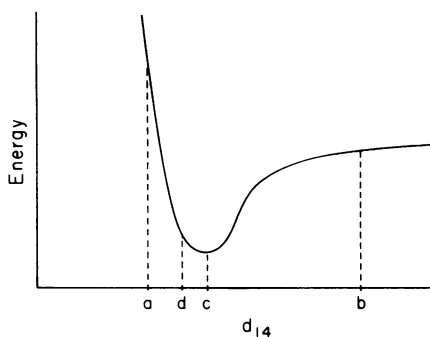


FIG. 1. Dependence of energy on the distance d_{14} . The labeled points on the d_{14} axis are the energies corresponding to the conformations shown in Fig. 2.

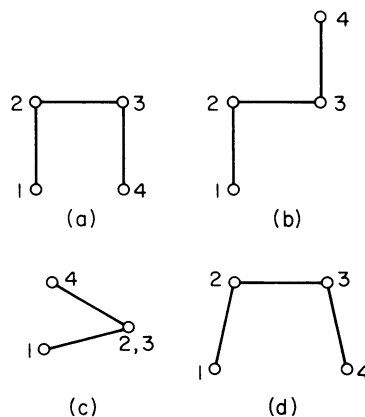


FIG. 2. Conformations: a and b, two-dimensional structures compatible with rigid geometry; c, the three-dimensional rigid-geometry structure (looking along the 2–3 bond) corresponding to the minimum energy of the curve in Fig. 1; d, the two-dimensional flexible-geometry global energy minimum.

Conclusion

The algorithm described here is an attempt to surmount the multiple-minima problem in the energy minimization of polypeptides. To assess the performance of our method on the terminally blocked amino-acid test cases, consider the positions of our computed structures in ordered lists of ECEPP minima (13) of these test cases. Alanine has 9 local minima. Lysine has 178 local minima within 3 kcal/mol of the global one. Methionine has 87 local minima within 5 kcal/mol of the global one. Serine has 52 local minima within 5 kcal/mol of the global minimum. Our method located the global minimum for alanine, the second lowest energy minimum for lysine, the second lowest for methionine, and the 14th lowest for serine. This method has accomplished the nontrivial task of locating rather low-energy minima.

Note Added in Proof. A recent variation of this method has led to the global minimum, even for serine.

We are indebted to M. J. Sippl for valuable discussions. This work was supported by research grants from the National Institute of General Medical Sciences (GM-14312) and from the National Science Foundation (DMB84-01811).

- Scheraga, H. A. (1981) *Biopolymers* **20**, 1877–1899.
- Vázquez, M. & Scheraga, H. A. (1985) *Biopolymers* **24**, 1437–1447.
- Paine, G. H. & Scheraga, H. A. (1985) *Biopolymers* **24**, 1391–1436.
- Vanderbilt, D. & Louie, S. G. (1984) *J. Comput. Phys.* **56**, 259–271.
- Momany, F. A., McGuire, R. F., Burgess, A. W. & Scheraga, H. A. (1975) *J. Phys. Chem.* **79**, 2361–2381.
- Némethy, G., Pottle, M. S. & Scheraga, H. A. (1983) *J. Phys. Chem.* **87**, 1883–1887.
- Crippen, G. M. (1982) *J. Comput. Chem.* **3**, 471–476.
- Crippen, G. M. (1984) *J. Comput. Chem.* **5**, 548–554.
- Blumenthal, L. M. (1970) *Theory and Application of Distance Geometry* (Chelsea, New York), pp. 97–105.
- Sippl, M. J. & Scheraga, H. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2283–2287.
- Sippl, M. J. & Scheraga, H. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2197–2201.
- Braun, W., Bösch, C., Brown, L. R., Gö, N. & Wüthrich, K. (1981) *Biochim. Biophys. Acta* **667**, 377–396.
- Vázquez, M., Némethy, G. & Scheraga, H. A. (1983) *Macromolecules* **16**, 1043–1049.