JOURNAL OF COMPUTATIONAL PHYSICS 24, 96–107 (1977)

A Novel Approach to Calculation of Conformation: Distance Geometry

GORDON M. CRIPPEN

Department of Pharmaceutical Chemistry, School of Pharmacy,
University of California, San Francisco, 94143

Received September 1, 1976; revised November 8, 1976

This paper considers the calculation of the conformation of a molecule by the unusual means of proposing the matrix **D** of all interatomic distances subject to a priori energetic and geometric constraints, and then calculating the corresponding atomic coordinates. The necessary and sufficient conditions on **D** from distance geometry are cited. Results are given for trials of numerical methods for carrying out such conformational calculations on cyclohexane and trypsin inhibitor.

1. INTRODUCTION

The problem of calculating the conformation of a molecule, particularly a large molecule such as a protein, has turned out to be surprisingly difficult. One of the most common methods (see for instance [1]) has been to place the atoms in some simple starting positions which satisfy a priori geometric constraints, such as fixed bond angles and bond lengths, and then alter the initial conformation by small shifts of atomic coordinates such that the geometric constraints always remain satisfied. The process continues until some energy-like function of all pairwise interatomic distances is minimal. Although the energy function may have a simple form with respect to the distance d_{ij} between any two atoms i and j, it is in general a very complicated function of the torsional angles, the natural conformational variables when bond lengths and angles are to be held fixed. The situation rapidly worsens when there are additional geometric constraints, for example, maintaining ring closure in cyclic molecules.

The approach taken in this paper is to reverse the order of calculation: first choose interatomic distances leading to a favorable energy, and then alter these distances such that the geometric constraints are fulfilled. A similar line of attack has been taken earlier by Tinoco *et al.* [2], where the formation of double-stranded segments in t-RNA molecules of known sequence was calculated using fairly simple geometric and energetic constraints on a qualitative distance matrix. The result, however, was a prediction of pairings of various parts of the chain with other parts, and not a prediction of the three-dimensional structure of the molecule, as we propose to undertake. In general, we proceed by first bounding all distances from above and below

ISSN 0021-9991

by using the a priori geometric constraints. These bounds are then made more stringent by applying certain geometric theorems. Finally, we add energetic considerations consistent with the bounds to arrive at a set of interatomic distances, which are then converted into atomic coordinates.

Section 2 considers the more elementary properties of the matrix **D** of interatomic distances d_{ij}, and how certain entries can be fixed and all others bounded from below. In Section 3 methods are given for computing an upper bound on the d_{ij}, and we cite the theorem for insuring that the proposed **D** matrix corresponds to a three-dimensional structure. In the fourth section numerical methods are discussed for converting a proposed distance matrix into atomic coordinates. The fifth section is a demonstration of the power of the approach on cyclohexane, a highly constrained molecule. Section 6 is concerned with the applicability of the distance matrix method, especially for large molecules.

2. BASIC PROPERTIES OF D

If one is to propose a distance matrix **D** of distances d_{ij} between n points (atoms, for our purposes), then **D** must have certain obvious properties regardless of how distances are calculated and how many spatial dimensions are allowed:

(1) **D** is a symmetric $n \times n$ matrix: $d_{ij} = d_{ji}$.

(2) Diagonal elements are all zero: $d_{ii} = 0$.

(3) All off-diagonal elements are strictly greater than zero: $d_{ij} > 0$, $i \neq j$ (otherwise some points i and j would be identical and a smaller order **D** would be appropriate).

Since **D** is symmetric, it is sufficient to refer to only the upper triangle in all that follows. It has been observed [3] that the local structure is specified by entries close to the diagonal of **D**, while the long range or global structure is given by elements far from the diagonal. Thus the $d_{i,i+1}$ are the bond lengths, and these together with the $d_{i,i+2}$ specify the bond angles, for a simple chain of atoms numbered sequentially along the chain. Similarly the $d_{i,i+3}$ correspond to the dihedral angles about rotatable bonds. Any desired ring closures are indicated by fixing d_{ij} at some small value, where i and j are the beginning and ending atoms of the loop. We see that the incorporation of any desired local geometric constraints, such as fixed bond lengths or cross linking, can be immediately incorporated into the proposed distance matrix.

It is also easy to calculate a matrix **S** of lower bounds on the distances. The s_{ij} are either equal to those corresponding fixed d_{ij} as explained above or are otherwise equal to the sum of the van der Waals radii of the atoms i and j. In the case of a polymer, this corresponds to a self-avoiding chain; for a self-intersecting chain, the far off-diagonal s_{ij} would be zero.

3. PROPERTIES OF D FROM DISTANCE GEOMETRY

It is possible to propose a distance matrix **D** satisfying all the conditions of the previous section, and yet not have it correspond to a realizable conformation. As a simple example, observe that there is no *two*-dimensional arrangement of four points corresponding to

$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}. \qquad (1)$$

The matrix is that of the corners of a tetrahedron, requiring three dimensions. Analogous problems arise when trying to embed an n-tuple of points in ordinary three-dimensional Euclidean space, E_3. The necessary and sufficient conditions for embedding n points in E_r for any given r, are given in a theorem due to Blumenthal [4]. We cite it here specialized to three dimensions, with appropriate changes in nomenclature.

THEOREM (Blumenthal). *A necessary and sufficient condition that a semimetric* $(n+1)$*-tuple may be irreducibly congruently embeddable in* E_3 $(3 \leqslant n)$ *is that an ordering of the points exists so that* (i) $\operatorname{sgn} \Delta(1,\ldots,k+1) = (-1)^{k+1}$ *for all* $k = 1,\ldots, m \leqslant 3$, *and* (ii) *for any other points* u *and* v, $m+1 \leqslant u$, $v \leqslant n+1$, *then* $\Delta(1,\ldots,m+1,u) = \Delta(1,\ldots,m+1,v) = \Delta(1,\ldots,m+1,u,v) = 0$. *Here* $\Delta(1,\ldots,k)$, *etc. is the Cayley–Menger determinant of the distances* d_{ij} *between the k points*

$$\Delta(1,\ldots,k) = \begin{vmatrix} d_{11}^2 & d_{12}^2 & \cdots & d_{1k}^2 & 1 \\ d_{21}^2 & d_{22}^2 & & \vdots & \vdots \\ \vdots & \vdots & & \cdot & \cdot \\ d_{k1}^2 & d_{k2}^2 & \cdots & d_{kk}^2 & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{vmatrix}. \qquad (2)$$

COROLLARY 1. *Part* (i) *is true for* $k = 1$ *as long as* $d_{12}^2 > 0$, *as it must.*

$$\Delta(1,2) = \begin{vmatrix} 0 & d_{12}^2 & 1 \\ d_{21}^2 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 2d_{12}^2.$$

COROLLARY 2. *Part* (i) *for* $k = 2$ *and* $m \geqslant 2$ *is equivalent to saying the* $n+1$ *points are not all collinear and the distances between the first three points satisfy the triangle inequality*

$$d_{12} + d_{23} \geqslant d_{13} \qquad (3)$$

or any permutation of the subscripts 1, 2, *and* 3.

Proof. Squaring Eq. (3) twice, one obtains

$$d_{12}^4 + d_{23}^4 + d_{13}^4 - 2d_{12}^2 d_{13}^2 - 2d_{23}^2 d_{13}^2 - 2d_{12}^2 d_{23}^2 \leqslant 0 \tag{4}$$

but the expression in (4) is exactly the result of evaluating $\Delta(1, 2, 3)$. Equality holds in (3) and (4) only if the three points are collinear.

COROLLARY 3. *Part* (i) *for* $k = m = 3$ *is equivalent to the* $n + 1$ *points being not all coplanar and the distances between the first four points satisfying what might be called the "quadrangle inequality." Denote the various squared distances by lower case unsubscripted letters thus*:

$$\Delta(1, 2, 3, 4) = \begin{vmatrix} 0 & a & c & d & 1 \\ a & 0 & b & e & 1 \\ c & b & 0 & f & 1 \\ d & e & f & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}. \tag{5}$$

Note that $d = d_{14}^2$. Then referring to Fig. 1, the quadrangle inequality says that d must lie between $d_{\min}$ (when the points are in the planar *cis* configuration (as illustrated) with respect to rotation about the 2-3 bond) and $d_{\max}$ (when in the planar *trans* configuration), assuming all other distances a, b, c, e, and f, to be fixed. The proof of this corollary is rather lengthy but straightforward, so it will be found in the Appendix. Even then the proof is restricted to the simpler case where $a = b = f$, corresponding to equal "bond lengths," or the "freely jointed chain," in polymer chemistry terminology.

COROLLARY 4. *Part* (ii) *of the theorem is satisfied if the bordered matrix* $\mathbf{D}(1,..., n + 1)$ *of squared distances corresponding to* $\Delta(1,..., n + 1)$ *has rank* $m + 2$. *In particular, for* $n + 1 > 4$ *points not all coplanar located in* E_3 *rank of* $\mathbf{D}$ *must be* 5.

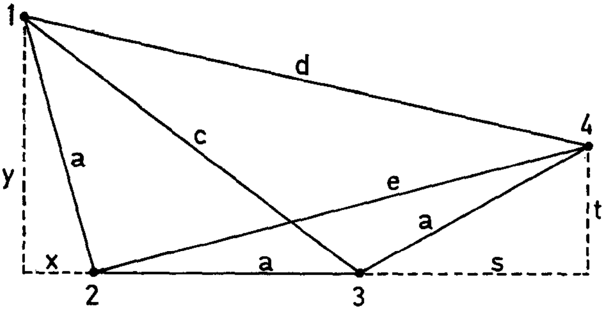

FIG. 1. Four numbered points with mutual squared distances denoted a, c, e and d corresponding to (5) when $a = b = f$. The arrangement shown is planar *cis* with respect to rotation about the 2–3 bond. See the Appendix for the use of distances x, y, s, and t.

Proof. The rank of a matrix is the order of its largest nonzero minor. If the rank of **D** is $m + 2$, then the points can be numbered so that for the first $m + 1$ of them, $\Delta(1, \ldots, m + 1) \neq 0$, and $\Delta(1, \ldots, m + 1, u) = \Delta(1, \ldots, m + 1, v) = \Delta(1, \ldots, m + 1, u, v) = 0$, since all minors of order higher than $m + 2$ must be zero. (Note that the border adds an extra row and column.)

These corollaries are useful in both gaining some insight into the implications of Blumenthal's theorem and numerically applying it. The triangle inequality of Corollary 2 is especially useful as an easy way to establish an upper bound on the distance matrix, because it obviously must hold for all triplets of points, not just the first three. The algorithm to construct an upper bound distance matrix $\mathbf{D}_{\max}$ is simply to initially set the entries equal to some a priori fixed values (as in the case of bond lengths, say), or otherwise to very large numbers; then for every ordered choice of three points, if (3) does not hold, set d_{13} to $d_{12} + d_{23}$; iterate the previous step until no alterations can be made for any element of $D_{\max}$.

As an example of the exhaustive use of the triangle inequality, consider the distances between residues 5, 14, 30, 38, 51, and 55 in the small protein, bovine pancreatic trypsin inhibitor (BPTI). From chemical studies it is known that not only is residue i connected to residue $i + 1$ for $i = 1, \ldots, 58$, but also that there are cross links between 5–55, 14–38, and 30–51. The former type of links are given unit length and then the latter type are known to be twice as long. Initially our distance information for these residues may be summarized in Table I. Applying the triangle inequality exhaustively yields the distances shown in Table II, a drastic reduction in uncertainty. For comparison, the true distances are given in Table IV.

The application of the quadrangle inequality of Corollary 3 is not so straightforward. Clearly the condition must hold for any quartet of points, but if not, the offending distance cannot be unambiguously identified since, in Fig. 1, any one of the six distances could be taken to be bounded by the values of the other five. One approach has been to start with the $\mathbf{D}_{\max}$ resulting from the triangle inequality and reduce its elements further, but as little as possible, so that $\Delta(1, 2, 3, 4) \geqslant 0$ for all (unordered) quartets of points. This is done by (i) selecting the quartet with the

TABLE I

Maximal Distances between BPTI Residues Known from Chemical Studies. 1 unit ≈3.8 Å

Residue	5	14	30	38	51	55
5	0	9	25	33	46	2
14		0	16	2	37	41
30			0	8	2	25
38				0	13	17
51					0	4
55						0

TABLE II

Maximal Distances between BPTI Residues after Application of the Triangle Inequality

Residue	5	14	30	38	51	55
5	0	9	8	11	6	2
14		0	10	2	12	11
30			0	8	2	6
38				0	10	13
51					0	4
55						0

TABLE III

Maximal Distances between BPTI Residues after Application of Quadrangle Inequality

Residue	5	14	30	38	51	55
5	0	9.00	6.03	10.12	4.82	2.00
14		0	7.41	2.00	8.59	10.18
30			0	7.23	2.00	5.71
38				0	8.78	11.02
51					0	4.00
55						0

TABLE IV

Experimentally Determined Distances between BPTI Residues by X-ray Crystallography [7]. 1 unit = 3.8 Å

Residue	5	14	30	38	51	55
5	0	5.79	2.62	4.66	2.37	1.54
14		0	5.86	1.64	5.82	6.66
30			0	5.22	1.68	2.24
38				0	4.81	5.52
51					0	1.54
55						0

algebraically smallest Δ; (ii) reducing (or holding fixed) the mutual distances according to the gradient of Δ until $\Delta \geqslant 0$ by moving along the "steepest descent" path in small steps, (iii) repeating step (i) until $\Delta \geqslant 0$ for all quartets. Typically the Δ's are initially all negative and several orders of magnitude larger in absolute value than their final values.

Using the above procedure on our example of maximal BPTI interresidue distances, we arrive at the distances given in Table III. Whereas the triangle inequality step between Tables I and II was quite an inexpensive calculation for even a 58 × 58 matrix, the exhaustive application of the quadrangle inequality is much more time consuming, and has not been tested on systems of more than eight points, where the reduction of some elements can be as much as 20%. It is important to note that the changes in distances are nontrivial, and that the resemblance to Table IV, the correct distances, is improved. The quadrangle inequality is a significant geometric constraint above and beyond the triangle inequality.

The higher order conditions of part (ii) of the theorem involve evaluation of $(n-4) + (n-4)(n-5)/2 = (n^2 - 7n + 12)/2$ determinants if there are n points altogether. Corollary 4 permits part (ii) to be checked by a single determination of rank for the $(n+1) \times (n+1)$ bordered matrix of squared distances. This is conveiniently done by Gaussian elimination with full pivoting and then noting how many nonzero rows result. Thus it is easy to check whether a given distance matrix is embeddable in E_3. However, it is difficult to alter a given matrix until the points are embeddable. Although there are only $3n - 6$ degrees of freedom to be adjusted for $n \geqslant 3$ points (namely the x, y, and z coordinates, less rigid translation and rotation) for a fully flexible chain, there are $n(n-1)/2$ variable distances in the full distance matrix, and there are $(n-4)(n-3)/2$ entries in the triangularized bordered distance matrix which must become zero. Therefore, one must solve $(n-4)(n-3)/2$ simultaneous nonlinear equations in $n(n-1)/2$ unknowns, which is quite difficult for large n. The better algorithms for solving systems of nonlinear equations generate an approximation to the Jacobian matrix involving $[n(n-1)/2]^2$ storage locations and considerable computer time. As soon as $n > 10$ the approach becomes infeasible. Nevertheless, we have successfully performed test calculations using Brown's method [5] for nonlinear equations to embed five points in E_2. Convergence was extremely rapid, but the program could not be scaled up to large numbers of points for the reasons just mentioned. Other equation solving algorithms which do not compute the Jacobian matrix, such as Schechter's SOR method [6] did not give good convergence.

4. CALCULATION OF COORDINATES

The ultimate check on such conformational calculations via distance matrix, and often the most useable final form, is a set of Cartesian coordinates. Assuming for the moment that we have successfully proposed an energetically favorable distance matrix fulfilling all the above conditions, it remains to show how to convert the distances into coordinates $\mathbf{v}_i = (x_i, y_i, z_i)$, $i = 1, \ldots, n$. The straightforward analytical approach is surprisingly unstable

$$\begin{aligned}
\mathbf{v}_1 &= (0, 0, 0), \\
\mathbf{v}_2 &= (d_{12}, 0, 0), \\
\mathbf{v}_3 &= \left(\frac{d_{13}^2 - d_{23}^2 + d_{12}^2}{2d_{12}}, [d_{13}^2 - {x_3}^2]^{1/2}, 0\right), \\
\mathbf{v}_4 &= \left(\frac{d_{14}^2 - d_{24}^2 + d_{12}^2}{2d_{12}}, \frac{d_{24}^2 - d_{34}^2 + d_{13}^2 - d_{12}^2 + 2x_4(d_{12} - x_3)}{2y_3}, [d_{14}^2 - {x_4}^2 - {y_4}^2]^{1/2}\right), \\
\mathbf{v}_i &= \left(\frac{d_{1i}^2 - d_{2i}^2 + d_{12}^2}{2d_{12}}, \frac{d_{2i}^2 - d_{3i}^2 - d_{12}^2 + d_{13}^2 + 2x_i(d_{12} - x_3)}{2y_3}, \pm[d_{1i}^2 - {x_i}^2 - {y_i}^2]^{1/2}\right), \\
&\qquad \text{for} \quad i = 5, \ldots, n
\end{aligned} \tag{6}$$

where the sign of z_i is chosen so that $| d_{4i} - \| \mathbf{v}_i - \mathbf{v}_4 \| |$ is minimal.

Note that we could have chosen z_4 to be negative, thus reversing the handedness of the whole resulting figure. The distance matrix of a collection of points is unique only up to a rigid translation, rotation, or mirror reflection. The inherent numerical instability of (6) is readily apparent, for example when $d_{13} \approx d_{23} \gg d_{12}$, and small errors in the d's result in large errors in x_3.

If the given distances satisfy all the conditions of the previous section to a good approximation, then we have developed stable iterative methods of coordinate calculation.

(i) Initially place the n points at some arbitrary positions.

(ii) For $i = 2, \ldots, n$; let $F = \sum_{j=1}^{i-1} (d_{ij\text{matrix}}^2 - d_{ij\text{calc}}^2)^2$. Minimize F with respect to the x, y, and z coordinates of point i. $d_{ij\text{matrix}}$ is the given distance, and $d_{ij\text{calc}}$ is that calculated from the coordinates set so far.

(iii) Finally refine the coordinates by minimizing $\sum_{i=1}^{n} \sum_{j=i+1}^{n} (d_{ij\text{matrix}}^2 - d_{ij\text{calc}}^2)^2$ with respect to all coordinates.

In our trials we calculated $\mathbf{D}$ from the positions of the C^α atoms in the small protein, pancreatic trypsin inhibitor, as determined by x-ray crystallography [7]. Here $n = 58$, but convergence was rapid using the Fletcher–Reeves conjugate gradient minimization algorithm [8], with coordinates being altered by only a few percent or less in step (iii). Less than ten seconds execution time was required on the CDC 7600 computer. We experienced no difficulty with converging on spurious local minima where the function value was greater than zero.

Deriving coordinates from a rough approximation to a "correct" distance matrix is a more poorly defined process, but our trials on trypsin inhibitor have been successful. Convergence is slower and weighting of terms in step (iii) is desirable [9].

5. CYCLOHEXANE

The distance geometry approach is especially well suited to conformational calculations on molecules with many strong geometric constraints. As a simple example, we have taken the six carbon atoms of cyclohexane, assuming fixed C_i–C_{i+1} bond lengths of 1.54 Å and fixed (tetrahedral) C_{i-1}–C_i–C_{i+1} bond angles of 109.47° (and therefore C_{i-1}–C_{i+1} distances of 2.51 Å). The result is only three possibly variable distances: $a = d_{14}$, $b = d_{25}$, and $c = d_{36}$. The conformations of this molecule with the above assumptions have certainly been investigated theoretically before. However, it is instructive to compare the ease of our approach to, say, the 95 equations in the paper by Gō and Scheraga [10], who solved the same problem in terms of torsional angles.

Table V shows the basic distance matrix, **D** for cyclohexane. A reasonable lower bound matrix **S**, taking the van der Waals' radii of the carbon atoms into account, would be just Table V with $a = b = c = 2$ Å. From the triangle inequality, $\mathbf{D}_{\max}$ would have $a = b = c = 4.06$ Å. Applying the quadrangle inequality to the sets of atoms 1–2–3–4, 2–3–4–5, and 3–4–5–6 gives minimal values of $a = b = c = 2.57$ and maximal distances of $a = b = c = 3.87$ Å. Embedding in E_3 eliminates two of the degrees of freedom, leaving us with one independently variable distance (when the molecule has C_2 symmetry) [10]. Having chosen a for instance, b and c were determined using Corollary 4. The proposed values of a, b, and c were inserted into the full, bordered matrix **B** of squared distances; this was converted to upper triangular form **T** by Gaussian elimination with full pivoting (for greatest numerical stability); the measure of the deviation of **B** from rank 5 was taken to be $f(a, b, c)$:

TABLE V

The Distance Matrix (Å) for the Six Carbon Atoms of Cyclohexane with Fixed Bond Lengths and Bond Angles

0	1.54	2.51	a	2.51	1.54
	0	1.54	2.51	b	2.51
		0	1.54	2.51	c
			0	1.54	2.51
				0	1.54
					0

$$f(a, b, c) = [\max_{i,j} | b_{ij} |] \left[\sum_{k=6}^{7} \sum_{l=k}^{7} t_{kl}^2 \right] \tag{7}$$

where the b's are elements of **B**, and the t's are elements of **T**. The double sum in Eq. (7) varies inversely with the largest element of **B**, so the first term in brackets is intended as a rough normalization of f for comparing the results of widely differing values of a, b, and c. The refining of **D** beyond the quadrangle inequality consisted of choosing a and then varying b and c until $f \approx 0$ and f was minimal. In this way, we were able to calculate $a = b = c = 2.949$ as our best approximation ($f = 0.0002$) to the "chair" conformation, which would be favored if uniform energetic repulsions dominated. If, on the other hand, atoms 1 and 4 were particularly attractive, their minimal distance could be $a = 2.567$, and if $b = c = 2.949$, this is actually realizable ($f = 0.0007$) and corresponds to the "boat" form. It was impossible to achieve the quadrangle inequality maximal distance of 3.87, the closest approach being 2.949 for b and c in the boat conformation.

D for the boat conformation was directly converted to cartesian coordinates, using Eq. (6), resulting in Table VI. The propagation of numerical errors in Eq. (6) did not present any difficulties in this case.

TABLE VI

Cartesian Coordinates (in Å) for the Six Cyclohexane Carbon Atoms in the Boat Conformation
($a = 2.567$, $b = c = 2.949$)

x	y	z
0	0	0
1.54	0	0
2.05	1.45	0
0.86	2.42	0
0	2.18	1.26
−0.51	0.72	1.26

6. DISCUSSION

The intent underlying this investigation has been to perform conformational calculations with the interparticle distances as the primary variables, so that one could take advantage of the direct relation between the distances and the pairwise energies, which together with the geometric constraints, presumably determine the conformation. In contrast to most physical situations, where the lower dimensionality versions of a problem are easier, the opposite is true here. In $(n - 1)$-dimensional Euclidean space, one could place n or fewer particles with almost any desired interparticle distances, such as at the distance of minimal energy for each interaction. Since we are restricted to only three dimensions, the purely geometric constraints become very important for $n \geqslant 4$, as we have seen in Section 3.

On the other hand, because a priori fixed distances can be so readily incorporated in this approach, calculations may be relatively tractable on highly constrained molecules (either physically having limited flexibility or having experimental evidence to fix or reduce the range of many distances).

Our main interest lies in the calculation of conformation of globular proteins, which typically consist of thousands of atoms grouped into a few hundred amino acid residues. Manipulation of the distance matrix for all atoms at once is quite infeasible, but in this approach the detail of the investigation can be conveniently varied. For instance, one can begin by determining the distance matrix for a single representative atom from each residue, using a crude residue–residue energy-like potential. Then rows and columns for all atoms of a pair of close residues can be added, and the local conformation could be worked out in detail using atom–atom energy functions. At the present stage of development, we are still concerned with ways of proposing residue–level distance matrices which satisfy the powerful geometric constraints outlined in this paper.

7. APPENDIX

Proof of Corollary 3. Evaluating (5), one obtains

$$\begin{aligned}\Delta(1,2,3,4) = 2\{&-a^2f - b^2d - c^2e - d^2b - e^2c - f^2a + bde + bce + cef + cde\\ &+ aef + ace + abf + abd + acf + bdf + bcd + adf - bef\\ &- ade - abc - cdf\}. \end{aligned} \tag{A1}$$

We will consider the special case where $a = b = f$; the general case is merely more laborious. Then (A1) becomes:

$$\Delta(1,2,3,4; a = b = f) = 2\{-a^3 + 2a^2d - ce^2 - d^2a - e^2c + 3ace + cde\}. \tag{A2}$$

Now referring to Fig. 1, if we take all distances but $d^{1/2}$ to be fixed, then the "quadrangle inequality" says that $d^{1/2}$ can vary between its lower bound $d_{\min}^{1/2}$ in the planar *cis* conformation, as illustrated, and its upper bound $d_{\max}^{1/2}$ in the planar *trans* conformation by rotating about the 2–3 "bond." We now calculate $d_{\max}$ and $d_{\min}$ using Fig. 1 and the pythagorean theorem

$$x + y = a, \tag{A3}$$

and

$$c = y + (a^{1/2} + W_c x^{1/2})^2, \tag{A4}$$

where

$$\begin{aligned} W_c &= +1 \qquad c > 2a,\\ &= -1 \qquad c < 2a. \end{aligned} \tag{A5}$$

Eliminating y and solving for x,

$$x^{1/2} = \frac{|\,c/2 - a\,|}{a^{1/2}} \tag{A6}$$

and similarly

$$s^{1/2} = \frac{|\,e/2 - a\,|}{a^{1/2}}. \tag{A7}$$

Hence, from (A5), (A6), and (A7)

$$\begin{aligned} d_{\min} &= (a^{1/2} + W_c x^{1/2} + W_e s^{1/2})^2 + (y^{1/2} - t^{1/2})^2 \\ &= (1/4a)(c + e - 2a)^2 + (y^{1/2} - t^{1/2})^2. \end{aligned} \tag{A8}$$

Similarly, we obtain

$$d_{\max} = (1/4a)(c + e - 2a)^2 + (y^{1/2} + t^{1/2})^2. \tag{A9}$$

Combining (A4) and its equivalent involving e and t with Eqs. (A6) and (A7) we find

$$y = c - c^2/4a \qquad \text{and} \qquad t = e - e^2/4a \tag{A10}$$

which upon substitution into (A8) and (A9) yields

$$\begin{aligned} d_{\min} &= a + (ce/2a) - 2[(c - c^2/4a)(e - e^2/4a)]^{1/2}, \\ d_{\max} &= a + (ce/2a) + 2[(c - c^2/4a)(e - e^2/4a)]^{1/2}. \end{aligned} \tag{A11}$$

From the quadrangle inequality we know that

$$-2a(d - d_{\min})(d - d_{\max}) \geqslant 0, \tag{A12}$$

but substituting Eq. (A11) and simplifying the expression results in the right-hand side of (A2) being greater than or equal to zero. Clearly equality holds only if the points are coplanar.

ACKNOWLEDGMENTS

This work was supported by a grant from the Academic Senate of the University of California. The author thanks Drs. I. D. Kuntz, Jr. and F. A. Grünbaum for their stimulating and helpful discussions, and Dr. R. Langridge for providing the x-ray crystallographic coordinates used.

REFERENCES

1. H. A. SCHERAGA, *Chem. Revs.* **71** (1971), 195.
2. I. TINOCO, O. UHLENBECK, AND M. LEVINE, *Nature (London)* **230** (1971), 362.
3. I. D. KUNTZ, *J. Amer. Chem. Soc.* **97** (1975), 4362.
4. L. M. BLUMENTHAL, "Theory and Applications of Distance Geometry," 2nd ed., pp. 98–105, Chelsea, New York, 1970.
5. K. M. BROWN, Tech. Rept. 68-23, Dept. Comp. Sci., Cornell University, Ithaca, N.Y., 1968.
6. S. SCHECHTER, *Trans. Amer. Math. Soc.* **104** (1962), 179.
7. R. HUBER *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **36** (1971), 141.
8. R. FLETCHER AND C. M. REEVES, *Comput. J.* **7** (1964), 149.
9. I. D. KUNTZ, manuscript in preparation.
10. N. GŌ AND H. A. SCHERAGA, *Macromolecules* **6** (1973), 273.