Ring Closure and Local Conformational Deformations of Chain Molecules[1,2]

Nobuhiro Gō[3] and Harold A. Scheraga[4]

Department of Chemistry, Cornell University, Ithaca, New York 14850.

Received December 4, 1969

ABSTRACT: A mathematical method is developed to provide a (computerized) solution to two problems hitherto arising in conformational energy calculations of oligomers and polymers, when bond lengths and bond angles are maintained fixed. The two problems are the calculations of the sets of dihedral angles which lead to (a) exact ring closure in cyclic molecules and (b) local conformational deformations of linear or cyclic molecules. Most of the emphasis is placed on polypeptide chain molecules. The method is applied to prove that no cyclic tri- and tetrapeptides with planar *trans*-peptide units having Pauling–Corey bond lengths and bond angles are geometrically capable of existing. The problems of calculating first and second derivatives of energy functions for cyclic molecules and for local deformations are also treated.

I. Introduction

In conformational energy calculations of alkanes and polypeptides, the bond-stretching and bond angle-bending degrees of freedom have usually been treated in two different ways, *viz.*, by considering both of these quantities as variables (I), and introducing bond-stretching and bond angle-bending potentials, or as fixed (II). In treatment I, either internal coordinates (bond lengths, bond angles, and dihedral angles)[5,6] or cartesian coordinates[7] of the atoms in the molecule may be used as variables, whereas, in treatment II, only the dihedral angles (internal coordinates) are treated as variables.[8,9] For polypeptide chains, the number of variables may be reduced further by assuming that the peptide unit is planar, although small departures from planarity[10] might be expected to occur from a consideration of the energy of rotation about the peptide bond.[11] The main advantage of treatment II is that fewer independent variables are involved than in treatment I. For example, in the case of *n*-decane there are 9 variables (dihedral angles) in treatment II, compared to 96 (when cartesian coordinates are used) or 90 (when internal coordinates are used) in treatment I. Obviously, treatment II involves much less computer time, even though treatment I is a more realistic one. Also, the computer program can have a simpler structure in treatment I, especially when cartesian coordinates are used as the variables.[12,13]

The theoretical basis of treatments I and II, together with a consideration of the procedures for calculating the conformational free energy of the molecule in each treatment, was presented elsewhere.[14] There it was pointed out that treatment II could be used as the first step of a perturbation treatment of the problem, in the sense that conformations which give minimum energy in treatment II may be used as the starting conformations of an energy minimization procedure in treatment I. It was also shown that the energy decrease in this second energy minimization process is usually small, unless there is considerable steric hindrance in the starting conformation. In other words, treatment II is generally a good one not only in providing a good approximation to the correct conformation, but also in providing a good approximation to the correct energy. In view of these facts, it is generally recommended that energy minimization be carried out first in a space of smaller dimensionality corresponding to the use of constant bond lengths and bond angles (and planar peptide units in the case of polypeptides). Then, if necessary, a further calculation can be performed in a space of larger dimensionality involving variable bond lengths and bond angles, starting with the conformation obtained in the first energy minimization procedure.

However, a difficulty arises in treatment II when one is dealing with a cyclic molecule (the first problem to which this paper is devoted), *viz.*, the calculation of a set of dihedral angles which corresponds to exact ring closure is not trivial.[15] One of the methods used to achieve ring closure is the use of a fictitious ring-closing

---

[1] This work was supported by research grants from the National Science Foundation (GB-7160 and GB-7571X) and from the National Institute of General Medical Sciences of the National Institutes of Health, U. S. Public Health Service (GM-14312).

[2] Presented, in part, before the Division of Biological Chemistry at the 158th meeting of the American Chemical Society, Sept 1969, New York, N. Y.

[3] On leave of absence from the Department of Physics, Faculty of Science, University of Tokyo, Tokyo, Japan, 1967–1970.

[4] To whom requests for reprints should be addressed.

[5] N. L. Allinger, *J. Amer. Chem. Soc.*, **81**, 5727 (1959).

[6] K. D. Gibson and H. A. Scheraga, *Biopolymers*, **4**, 709 (1966).

[7] K. B. Wiberg, *J. Amer. Chem. Soc.*, **87**, 1070 (1965).

[8] P. De Santis, E. Giglio, A. M. Liquori, and A. Ripamonti, *J. Polym. Sci., Part A*, **1**, 1383 (1963).

[9] R. A. Scott and H. A. Scheraga, *J. Chem. Phys.*, **44**, 3054 (1966).

[10] G. N. Ramachandran, *Biopolymers*, **6**, 1494 (1968).

[11] J. F. Yan, F. A. Momany, R. Hoffmann, and H. A. Scheraga, *J. Phys. Chem.*, **74**, 420 (1970).

[12] Hendrikson[13] has used an intermediate treatment in which bond lengths are kept constant and bond angles are variables. In this treatment, which has a number of variables intermediate between those of treatments I and II, only internal coordinates may be treated as variables. The complexity of the computer program in this treatment is also intermediate between those of treatments I and II.

[13] J. B. Hendrikson, *J. Amer. Chem. Soc.*, **83**, 4537 (1961).

[14] N. Gō and H. A. Scheraga, *J. Chem. Phys.*, **51**, 4751 (1969).

[15] In Hendrikson's treatment,[12,13] it is not difficult to find a set of dihedral angles and bond angles which correspond to exact ring closure.

potential.[16,17] But the inclusion of such an artificial potential also introduces some problems in the numerical computations. Therefore, the first purpose of this paper is to present a mathematical method to calculate the set of dihedral angles which correspond to exact ring closure in treatment II.

Another difficulty, having the same mathematical origin in treatment II as the ring-closure problem, arises in a different aspect of conformational energy calculations. In any conformational energy minimization procedure, one selects a certain starting conformation which is then altered, step by step, in such a way that the corresponding conformational energy is decreased monotonically. In order to carry out the energy minimization efficiently, the alteration of the conformation in each step should be determined by making maximum use of information, obtained in the previous steps, about the appearance of the conformational energy surface in the vicinity of the conformation being considered. From this point of view, it is in general desirable to avoid very drastic changes in conformation in the energy minimization procedure, since drastic changes prevent one from making maximum use of information accumulated in the previous steps. However, in treatment II, it is not easy to make only *moderate* changes in conformation for the following reason. A change of a few degrees in a dihedral angle located near the middle of a long chain molecule causes a drastic change of the overall conformation of the molecule, unless it is accompanied by simultaneous changes in nearby dihedral angles, *i.e.*, effectively by a cooperative variation of angles which confines the conformational change to a *local* section of the chain. In other words, in order to produce only a small change in conformation, a certain sequence of dihedral angles must be changed simultaneously in such a way that the positions of atoms located far from the bonds (about which rotations take place) do not change by large amounts. However, up to now, the calculation of such changes in a set of dihedral angles, which cause *local* deformations of conformation in long polymer chains, has not been carried out in treatment II. Thus, the second purpose of this paper is to solve this mathematical problem. It is possible that such local deformations, which involve only geometrical rather than energetic criteria, may provide a means of passing to successively lower minima in the multidimensional energy space.

These two problems (ring closure and local deformations) can be solved by the same mathematical method. In section II, these two problems are shown to be those of solving a set of algebraic equations for six unknowns. In section III, special cases of chain molecules are treated in which rotation can take place around all backbone bonds; in these cases, the set of algebraic equations for the six unknowns can be reduced to a single equation in a single unknown. In section IV, the special case of polypeptide chains with planar peptide units is treated; the method developed in section III is shown to be applicable also for polypeptide chains, even when the peptide units are assumed to be planar, and the method is illustrated numerically. In section V, the results of section IV are applied to cyclic tri- and tetrapeptides whose unit structures are those given by Pauling and Corey;[18] it is proved that cyclic tri- and tetrapeptides cannot exist geometrically, if the peptide unit having Pauling–Corey geometry is assumed to be planar and in the *trans* conformation. In section VI, the calculation of first and second derivatives of energy functions is treated; this is of importance in energy minimization, in calculating statistical weights, and for treating local deformations in general chain molecules.

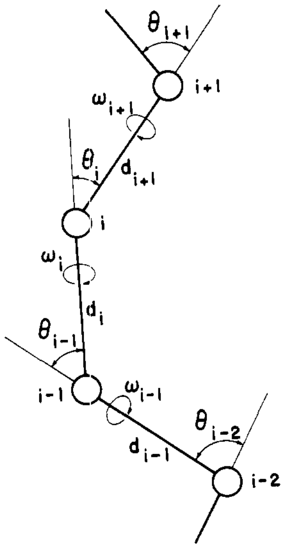

Figure 1. Definitions of bond lengths d_i, bond angles θ_i, and dihedral angles ω_i for a chain molecule.

II. Mathematical Expression of the Problems

In this section, the two problems indicated in section I are formulated as algebraic equations.

Consider a portion of a chain molecule consisting of backbone atoms linked serially (Figure 1). The definitions of bond length d_i, bond angle θ_i, and dihedral angle ω_i are given in Figure 1. A local coordinate system is set up for each backbone atom as follows. The ith local right-handed rectangular cartesian coordinate system is defined in such a way that the components of the position vectors of atoms $i - 1$, i, and $i + 1$ in this local coordinate system are (0,0,0), (d_i,0,0), and ([$d_i + d_{i+1} \cos \theta_i$],$d_{i+1} \sin \theta_i$,0), respectively. If a given point in space is expressed by position vectors $\mathbf{r}_i$ and $\mathbf{r}_{i-1}$ with respect to the ith and ($i - 1$)th coordinate system, respectively, the relation between them is given by

$$\mathbf{r}_{i-1} = \mathbf{T}_{i-1}\mathbf{R}_i\mathbf{r}_i + \mathbf{p}_{i-1} \tag{1}$$

where

$$\mathbf{T}_{i-1} = \begin{pmatrix} \cos \theta_{i-1} & -\sin \theta_{i-1} & 0 \\ \sin \theta_{i-1} & \cos \theta_{i-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

$$\mathbf{R}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \omega_i & -\sin \omega_i \\ 0 & \sin \omega_i & \cos \omega_i \end{pmatrix} \tag{3}$$

---

(16) K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci., U. S.*, **58**, 1317 (1967).

(17) R. A. Scott, G. Vanderkooi, R. W. Tuttle, P. M. Shames, and H. A. Scheraga, *ibid.*, **58**, 2204 (1967).

(18) See, for example, L. Pauling, "The Nature of the Chemical Bond," 3rd ed, Cornell University Press, Ithaca, N. Y., 1960, p 282.

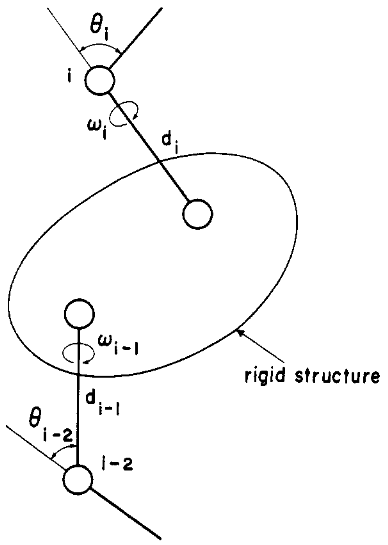

Figure 2. Definitions of bond lengths d_i, bond angles θ_i, and dihedral angles ω_i for a chain molecule with rigid portions in its backbone structure.

$$\mathbf{p}_{i-1} = \begin{pmatrix} d_{i-1} \\ 0 \\ 0 \end{pmatrix} \tag{4}$$

In accordance with treatment II, we keep bond lengths and bond angles constant, and regard only the dihedral angles as variables. In some cases, it may be useful to treat some parts of a molecule as rigid, *e.g.*, the planar amide group or the rigid benzene ring. In such cases, two consecutive bonds along the chain, about which rotation takes place, do not meet at a common atom (Figure 2). For such cases, $\mathbf{T}_{i-1}$ in eq 1 takes a more general form of a 3 × 3 orthogonal matrix, and the vector $\mathbf{p}_{i-1}$ in eq 1 may have nonvanishing y and z components. For example, an explicit form of $\mathbf{p}_{i-1}$ of such a general character will be given in section IV for a polypeptide chain. In the remainder of this paper, $\mathbf{T}_{i-1}$ and $\mathbf{p}_{i-1}$ will be assumed to be a general 3 × 3 orthogonal matrix and a general vector, respectively, unless stated otherwise. Equations 2 and 4 are special (and important) cases of $\mathbf{T}_{i-1}$ and $\mathbf{p}_{i-1}$, respectively.

We shall treat first the second problem, that of a local conformational deformation, and will treat the first problem of exact ring closure later. Let us assume that we want to change n consecutive dihedral angles $\omega_1,\ldots,\omega_n$ located in the middle of a long chain molecule without affecting the locations and orientations of the local coordinate systems other than those numbered 1 to $n - 1$. In particular, the locations and orientations of the local coordinate systems numbered 0 and n should remain fixed for any deformations. The equations for transformation of coordinates among these $(n + 1)$ local coordinate systems are given by repetitive

$$\begin{aligned} \mathbf{r}_0 &= \mathbf{T}_0\mathbf{R}_1\mathbf{r}_1 + \mathbf{p}_0 \\ \mathbf{r}_1 &= \mathbf{T}_1\mathbf{R}_2\mathbf{r}_2 + \mathbf{p}_1 \\ &\cdot \\ &\cdot \\ &\cdot \\ \mathbf{r}_{n-1} &= \mathbf{T}_{n-1}\mathbf{R}_n\mathbf{r}_n + \mathbf{p}_{n-1} \end{aligned} \tag{5}$$

use of eq 1. Since the location and orientation of the local coordinate system n relative to the local coordinate system 0 is fixed, there must be a constant orthogonal matrix $\mathbf{T}$ and a constant vector $\mathbf{p}$ such that

$$\mathbf{r}_0 = \mathbf{T}\mathbf{r}_n + \mathbf{p} \tag{6}$$

The problem is to find a set of values for n dihedral angles $\omega_1,\ldots,\omega_n$ which are compatible with eq 5 and 6. These equations cannot be satisfied, *i.e.*, a deformation cannot be confined to a local region, for an *arbitrary* set of values for these n variables. Instead, some relations must exist among these n variables in order to confine the deformation locally. In other words, the number of independent variables for local deformations must be less than n, *viz.*, $(n - 6)$ as shown below.

The location and orientation of a local coordinate system can be specified by six variables, *viz.*, three for the location of the origin and three for the Eulerian angles. Or, alternatively, they can be specified by a vector $\mathbf{s}$ (position vector of the origin), a vector $\mathbf{u}$ (unit vector in the direction of the x axis), and a vector $\mathbf{v}$ (unit vector in the direction of the y axis). There are three relations among $\mathbf{u}$ and $\mathbf{v}$, *viz.*

$$\begin{aligned} \mathbf{u}\cdot\mathbf{u} &= 1 \\ \mathbf{v}\cdot\mathbf{v} &= 1 \\ \mathbf{u}\cdot\mathbf{v} &= 0 \end{aligned} \tag{7}$$

Therefore, there are six degrees of freedom for the three vectors $\mathbf{s}$, $\mathbf{u}$, and $\mathbf{v}$. For example, the relative position of coordinate system 1 with respect to coordinate system 0 is given, according to the first of eq 5, by

$$\begin{aligned} \mathbf{s} &= \mathbf{p}_0 \\ \mathbf{u} &= \mathbf{T}_0\mathbf{R}_1\mathbf{e}_1 \\ \mathbf{v} &= \mathbf{T}_0\mathbf{R}_1\mathbf{e}_2 \end{aligned} \tag{8}$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are column vectors, whose transposed (row) vectors are $\mathbf{e}_1^+ = (1,0,0)$ and $\mathbf{e}_2^+ = (0,1,0)$, respectively. By using all of the equations in eq 5 we can obtain three vectors $\mathbf{s}_n$, $\mathbf{u}_n$, and $\mathbf{v}_n$ corresponding to the relative position and orientation of coordinate system n with respect to the local coordinate system 0

$$\mathbf{s}_n = \mathbf{p}_0 + \mathbf{T}_0\mathbf{R}_1\mathbf{p}_1 + \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{p}_2 + \ldots + \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\ldots\mathbf{T}_{n-2}\mathbf{R}_{n-1}\mathbf{p}_{n-1} \tag{9}$$

$$\mathbf{u}_n = \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\ldots\mathbf{T}_{n-2}\mathbf{R}_{n-1}\mathbf{T}_{n-1}\mathbf{R}_n\mathbf{e}_1 \tag{10}$$

$$\mathbf{v}_n = \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\ldots\mathbf{T}_{n-2}\mathbf{R}_{n-1}\mathbf{T}_{n-1}\mathbf{R}_n\mathbf{e}_2 \tag{11}$$

Here, the three vectors $\mathbf{s}_n$, $\mathbf{u}_n$, and $\mathbf{v}_n$, which are equal to $\mathbf{p}$, $\mathbf{T}\mathbf{e}_1$, and $\mathbf{T}\mathbf{e}_2$, respectively, because of eq 6, are constant. Also, $\mathbf{u}_n$ and $\mathbf{v}_n$ satisfy eq 7. Therefore, the number of independent equations among eq 9, 10, and 11 is six, which means that there are six independent relations among n variables. Hence, the number of independent variables is $(n - 6)$. This means, for example, that at least seven consecutive dihedral angles must be changed simultaneously in order to deform the conformation of a chain molecule locally; and, one of these can be varied independently, the other six being determined by the value assumed by the independently varied one.

We can now solve the problem of calculating sets of dihedral angles $\omega_1,\ldots,\omega_n$ which are compatible with eq 5 and 6, or which give a locally deformed structure, by the following steps. (1) Arbitrarily choose $(n - 6)$ dihedral angles out of the n angles $\omega_1,\ldots,\omega_n$ as the independent variables. (2) Assign an arbitrary set of values to these ($n - 6$) independent variables. (3) Solve eq 9, 10, and 11 for the other six dihedral angles for the given set of values assigned to the ($n - 6$) independent variables. Equations 9, 10, and 11 may have no solution, or one or more solutions. When there is no solution, no local deformation is possible for the assigned set of values of the ($n - 6$) independent variables, no matter what values the other six angles may assume. When there are more than one solution, there are more than one set of values that the other six angles can assume, for the assigned set of values of the ($n - 6$) independent variables, in order to deform the chain locally. As one of the arbitrary choices, we select the ($n - 6$) angles $\omega_1,\ldots,\omega_{n-6}$ as the independent variables. Then the equations which must be solved for $\omega_{n-5},\omega_{n-4},\ldots,\omega_n$ for a given set of values of $\omega_1,\ldots,\omega_{n-6}$ are

$$\begin{aligned}\mathbf{s} = {} & \mathbf{p}_{n-6} + \\ & \mathbf{T}_{n-6}\mathbf{R}_{n-5}\mathbf{p}_{n-5} + \\ & \mathbf{T}_{n-6}\mathbf{R}_{n-5}\mathbf{T}_{n-5}\mathbf{R}_{n-4}\mathbf{p}_{n-4} + \\ & \mathbf{T}_{n-6}\mathbf{R}_{n-5}\mathbf{T}_{n-5}\mathbf{R}_{n-4}\mathbf{T}_{n-4}\mathbf{R}_{n-3}\mathbf{p}_{n-3} + \\ & \mathbf{T}_{n-6}\mathbf{R}_{n-5}\mathbf{T}_{n-5}\mathbf{R}_{n-4}\mathbf{T}_{n-4}\mathbf{R}_{n-3}\mathbf{T}_{n-3}\mathbf{R}_{n-2}\mathbf{p}_{n-2} + \\ & \mathbf{T}_{n-6}\mathbf{R}_{n-5}\mathbf{T}_{n-5}\mathbf{R}_{n-4}\mathbf{T}_{n-4}\mathbf{R}_{n-3}\mathbf{T}_{n-3}\mathbf{R}_{n-2}\mathbf{T}_{n-2}\mathbf{R}_{n-1}\mathbf{p}_{n-1}\end{aligned} \tag{12}$$

$$\mathbf{u} = \mathbf{T}_{n-6}\mathbf{R}_{n-5}\mathbf{T}_{n-5}\mathbf{R}_{n-4}\mathbf{T}_{n-4}\mathbf{R}_{n-3}\mathbf{T}_{n-3}\mathbf{R}_{n-2}\mathbf{T}_{n-2}\mathbf{R}_{n-1}\mathbf{T}_{n-1}\mathbf{R}_n\mathbf{e}_1 \tag{13}$$

$$\mathbf{v} = \mathbf{T}_{n-6}\mathbf{R}_{n-5}\mathbf{T}_{n-5}\mathbf{R}_{n-4}\mathbf{T}_{n-4}\mathbf{R}_{n-3}\mathbf{T}_{n-3}\mathbf{R}_{n-2}\mathbf{T}_{n-2}\mathbf{R}_{n-1}\mathbf{T}_{n-1}\mathbf{R}_n\mathbf{e}_2 \tag{14}$$

where

$$\begin{aligned}\mathbf{s} &= \mathbf{R}_{n-6}^{-1}\mathbf{T}_{n-7}^{-1}\ldots\mathbf{R}_1^{-1}\mathbf{T}_0^{-1}[\mathbf{s}_n - \mathbf{p}_0 - \\ &\qquad \mathbf{T}_0\mathbf{R}_1\mathbf{p}_1 - \ldots - \mathbf{T}_0\mathbf{R}_1\ldots\mathbf{T}_{n-8}\mathbf{R}_{n-7}\mathbf{p}_{n-7}] \\ \mathbf{u} &= \mathbf{R}_{n-6}^{-1}\mathbf{T}_{n-7}^{-1}\ldots\mathbf{R}_1^{-1}\mathbf{T}_0^{-1}\mathbf{u}_n \\ \mathbf{v} &= \mathbf{R}_{n-6}^{-1}\mathbf{T}_{n-7}^{-1}\ldots\mathbf{R}_1^{-1}\mathbf{T}_0^{-1}\mathbf{v}_n\end{aligned} \tag{15}$$

are expressed in terms of the ($n - 6$) independent variables. The vectors **u** and **v** in eq 15 satisfy eq 7. Therefore, the number of independent equations in eq 12, 13, and 14 is 6, which is the same as the number of unknowns. The quantity **s** in the first of eq 15 is the position vector of the origin of the nth local coordinate system with respect to the ($n - 6$)th one.

We now turn to the first problem, that of closing a ring structure exactly. Let us consider a cyclic chain molecule in which there are n variable dihedral angles $\omega_1,\ldots,\omega_n$, and hence n local coordinate systems. The equations for the transformation of coordinates among these coordinate systems are given by eq 5, in which $\mathbf{r}_0,\mathbf{T}_0$, and $\mathbf{p}_0$ are replaced by $\mathbf{r}_n,\mathbf{T}_n$, and $\mathbf{p}_n$, respectively. In other words, ring structures are obtained by identifying the coordinate system 0 with the coordinate system n. This means that the problem of closing a ring structure exactly can be regarded as a special case of the problem of local deformations. Therefore, the equations for the problem of ring closure are obtained as special cases of the equations for local deformations. Thus, eq 6 should be replaced by $\mathbf{r}_0 = \mathbf{r}_n$. In eq 9, 10, 11, and 15, $\mathbf{s}_n$, $\mathbf{u}_n$, $\mathbf{v}_n$, $\mathbf{p}_0$, and $\mathbf{T}_0$ should be replaced by $\mathbf{0}$, $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{p}_n$, and $\mathbf{T}_n$, respectively. Equations 12, 13, and 14 remain unchanged.

As in the problem of local deformations, there are ($n - 6$) independent variables in a cyclic chain molecule of n variable dihedral angles. It is of interest to examine some consequences of this fact. For cyclopentaglycyl, $n = 10$ if we assume that the amide groups are planar. Hence, four of these dihedral angles can be treated as independent variables. If two glycyl residues are replaced by prolyl residues, as in

┌—Gly·Gly·Gly·Pro·Pro—┐
└————————————————┘

$n = 8$ since the ϕ angles[19] in prolyl residues may be fixed as a good approximation. Hence, there are only two independent variables for this molecule. In other words, the conformation of this molecule can be specified by assigning values to two dihedral angles which are chosen as the independent variables. Therefore, we can express the entire conformational energy surface of this molecule on several two-dimensional maps. The reason why we need more than one map is that, when there are multiple solutions to eq 12, 13, and 14 for a given set of values of the two independent variables, we have a corresponding number of possible conformations and conformational energies. The calculation of the entire conformational energy surface for this molecule is presented in the accompanying paper.[20]

Our two problems presented in section I (ring closure and local deformation) have been expressed in terms of the same mathematical expressions, *i.e.*, eq 12, 13, and 14, with **u** and **v** satisfying eq 7. The methods of solving these equations are described in sections III–VI, as indicated at the end of section I. In sections III, IV, and VI, eq 15 will not be used, *i.e.*, instead, the vectors **s**, **u**, and **v** will be treated as arbitrary vectors (which are then six independent variables since they satisfy eq 7) rather than taking $\omega_1,\ldots,\omega_{n-6}$ as the independent variables. Therefore, $\omega_{n-5},\ldots,\omega_n$ are the only variables which appear in sections III, IV, and VI. Hence, there is no point in assigning serial numbers from $n - 5$ to n to six unknowns. Therefore, in the remainder of this paper, we will replace $n - 6$, $n - 5$, $n - 4$, $\ldots$, n by 0, 1, 2, $\ldots$, 6, respectively, as suffixes in eq 12, 13, and 14. The resuffixed equations will be referred to as eq 12a, 13a, and 14a. In section V, where the method of solving eq 12a, 13a, and 14a is applied to a discussion of cyclic tri- and tetrapeptides, eq 15 will then be used.

III. Cases of Rotation about All Backbone Bonds

In this section, we focus attention on cases in which rotation can take place about all backbone bonds. Hence, in these cases, all constant matrices $\mathbf{T}_i$ and constant vectors $\mathbf{p}_i$ have the forms of eq 2 and 4, respectively. It will be shown in this section that, for these special cases, the set of simultaneous equations 12a, 13a, and 14a, for the six unknowns $\omega_1,\ldots,\omega_6$, where **u** and **v** satisfy eq 7, can be reduced to a single equation in

---

(19) J. T. Edsall, P. J. Flory, J. C. Kendrew, A. M. Liquori, G. Nemethy, G. N. Ramachandran, and H. A. Scheraga, *Biopolymers*, **4**, 121 (1966); *J. Biol. Chem.*, **241**, 1004 (1966); *J. Mol. Biol.*, **15**, 399 (1966).

(20) N. Gō and H. A. Scheraga, *Macromolecules*, **3**, 188 (1970).

a single unknown which can be solved, for example, by Newton's method.[21]

By using the explicit forms of $\mathbf{R}_i$ and $\mathbf{T}_i$ given in eq 2 and 3, eq 12a, 13a, and 14a can be transformed to

$$\mathbf{s} = \mathbf{q}_0 + \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{q}_1 + \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{T}_2\mathbf{R}_3\mathbf{T}_3\mathbf{R}_4\mathbf{q}_2 \quad (16)$$

$$\mathbf{T}_4^{-1}\mathbf{R}_4^{-1}\mathbf{T}_3^{-1}\mathbf{R}_3^{-1}\mathbf{T}_2^{-1}\mathbf{R}_2^{-1}\mathbf{T}_1^{-1}\mathbf{R}_1^{-1}\mathbf{T}_0^{-1}\mathbf{u} = \begin{pmatrix} \cos\theta_5 \\ \sin\theta_5\cos\omega_5 \\ \sin\theta_5\sin\omega_5 \end{pmatrix} \quad (17)$$

$$\mathbf{T}_5^{-1}\mathbf{R}_5^{-1}\mathbf{T}_4^{-1}\mathbf{R}_4^{-1}\mathbf{T}_3^{-1}\mathbf{R}_3^{-1}\mathbf{T}_2^{-1}\mathbf{R}_2^{-1}\mathbf{T}_1^{-1}\mathbf{R}_1^{-1}\mathbf{T}_0^{-1}\mathbf{v} = \begin{pmatrix} 0 \\ \cos\omega_6 \\ \sin\omega_6 \end{pmatrix} \quad (18)$$

where

$$\mathbf{q}_i = \begin{pmatrix} d_{2i} + d_{2i+1}\cos\theta_{2i} \\ d_{2i-1}\sin\theta_{2i} \\ 0 \end{pmatrix} \quad (19)$$

where $i = 0, 1, 2$. The merit of these expressions is that two unknowns (ω_5 and ω_6) do not appear in the three equations (*i.e.*, x, y, and z components) of eq 16 and in the x component of eq 17. Therefore, these four equations can be solved for the four unknowns ω_1, ω_2, ω_3, and ω_4. Once these four unknowns are determined, ω_5 can be obtained uniquely from the y and z components of eq 17. Then ω_6 is determinable uniquely from the y and z components of eq 18, and the already determined values of ω_1, ω_2, ω_3, ω_4, and ω_5. Therefore, the whole problem is reduced in essence to that of solving four simultaneous equations (x, y, and z components of eq 16 and x component of eq 17) for the four unknowns ω_1, ω_2, ω_3, and ω_4. The x component of eq 17 can be written as

$$\mathbf{u}^+\mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{T}_2\mathbf{R}_3\mathbf{T}_3\mathbf{R}_4\mathbf{T}_4\mathbf{e}_1 - \cos\theta_5 = 0 \quad (20)$$

where $\mathbf{u}^-$ is the transpose of $\mathbf{u}$. In the next paragraph, we will solve eq 16 for ω_2, ω_3, and ω_4 for given $\mathbf{s}$ and ω_1; *i.e.*, ω_2, ω_3, and ω_4 will each be expressed as a function of ω_1. Then these expressions will be substituted into eq 20, so that the latter may then be regarded as an algebraic equation for ω_1 only.

To obtain ω_2, ω_3, and ω_4 each as a function of ω_1, we proceed as follows. First we can express ω_2 as a function of ω_1. If the vector $\mathbf{s}$ and the angle ω_1 are regarded as given quantities in eq 16, then the vector $\mathbf{r}$ of eq 21 is also a given quantity

$$\mathbf{r} = \mathbf{T}_1^{-1}\mathbf{R}_1^{-1}\mathbf{T}_0^{-1}(\mathbf{s} - \mathbf{q}_0) \quad (21)$$

Equation 16 may then be written in terms of $\mathbf{r}$ as

$$\mathbf{r} = \mathbf{R}_2\mathbf{q}_1 + \mathbf{R}_2\mathbf{T}_2\mathbf{R}_3\mathbf{T}_3\mathbf{R}_4\mathbf{q}_2 \quad (22)$$

This is equivalent to a set of the two equations

$$\mathbf{r} = \mathbf{R}_2(\mathbf{q}_1 + \mathbf{q}_3) \quad (23)$$

and

$$\mathbf{q}_3 = \mathbf{T}_2\mathbf{R}_3\mathbf{T}_3\mathbf{R}_4\mathbf{q}_2 \quad (24)$$

From eq 24, we have

$$q_3^2 = q_2^2 \quad (25)$$

Solving eq 23 for $\mathbf{q}_3$, and substituting it into eq 25, we have

$$y\cos\omega_2 + z\sin\omega_2 = (r^2 + q_1^2 - q_2^2 - 2\rho_1 x)/2\sigma_1 \quad (26)$$

Here, x, y, and z are the components of the vector $\mathbf{r}$, which are given quantities, as mentioned above, and ρ_i and σ_i are the x and y components, respectively, of $\mathbf{q}_i$ (given by eq 19). The necessary and sufficient condition for the existence of solution(s) of eq 26 is given by

$$q_1 + q_2 \geq r \geq q_1 - q_2 \quad (27)$$

and

$$[\rho_1(r^2 + q_1^2 - q_2^2) + \sigma_1\{(q_1 + q_2)^2 - r^2\}^{1/2}\{r^2 - (q_1 - q_2)^2\}^{1/2}]/2q_1^2 \geq x \geq [\rho_1(r^2 + q_1^2 - q_2^2) - \sigma_1\{(q_1 + q_2)^2 - r^2\}^{1/2}\{r^2 - (q_1 - q_2)^2\}^{1/2}]/2q_1^2 \quad (28)$$

Here, r, q_1, and q_2 are the lengths of the vectors $\mathbf{r}$, $\mathbf{q}_1$, and $\mathbf{q}_2$, respectively. When these conditions are satisfied, eq 26 has the solution

$$\cos\omega_2 = \frac{yw \pm z(r^2 - x^2 - w^2)^{1/2}}{r^2 - x^2}$$
$$\sin\omega_2 = \frac{zw \mp y(r^2 - x^2 - w^2)^{1/2}}{r^2 - x^2} \quad (29)$$

where

$$w = (r^2 + q_1^2 - q_2^2 - 2\rho_1 x)/2\sigma_1 \quad (30)$$

We can now express ω_4 as a function of ω_1. By substituting eq 29 into eq 23, we have

$$\mathbf{q}_3 = \begin{pmatrix} x - \rho_1 \\ w - \sigma_1 \\ \pm(r^2 - x^2 - w^2)^{1/2} \end{pmatrix} \quad (31)$$

If we write eq 24 as

$$\mathbf{R}_3^{-1}\mathbf{T}_2^{-1}\mathbf{q}_3 = \mathbf{T}_3\mathbf{R}_4\mathbf{q}_2 \quad (32)$$

then the x component of this equation yields

$$\cos\omega_4 = [\rho_2\cos\theta_3 - (x - \rho_1)\cos\theta_2 - (w - \sigma_1)\sin\theta_2]/\sigma_2\sin\theta_3 \quad (33)$$

This equation determines the value of the dihedral angle ω_4 under the condition

$$1 \geq [\rho_2\cos\theta_3 - (x - \rho_1)\cos\theta_2 - (w - \sigma_1)\sin\theta_2]/\sigma_2\sin\theta_3 \geq -1 \quad (34)$$

When this condition is satisfied, we have

$$\sin\omega_4 = \pm[1 - \{\rho_2\cos\theta_3 - (x - \rho_1)\cos\theta_2 - (w - \sigma_1)\sin\theta_2\}^2/(\sigma_2\sin\theta_3)^2]^{1/2} \quad (35)$$

The ± sign in eq 35 is independent of those in eq 29 and 31. The three ± (and ∓) signs in eq 29 and 31 are interdependent in a sense that, when the upper (respectively, lower) sign is taken in one equation, the upper (respectively, lower) signs must be taken in the other two equations. Because of the ± signs in eq 29, there are in general two solutions for ω_2 for given $\mathbf{s}$ and ω_1, pro-

---

[21] See, for example, D. D. McCracken and W. S. Dorn, "Numerical Methods and Fortran Programming," John Wiley & Sons, Inc., New York, N. Y., 1964, p 133.

vided that the conditions of eq 27 and 28 are satisfied. And moreover, because of the ± signs in eq 35, there are in general two solutions for ω_4 for given **s** and ω_1, provided that the condition of eq 34 is satisfied. Because of the independence of the two groups of ± signs, one in eq 29 and 31 and the other one in eq 35, there are in general four sets of solutions of ω_2 and ω_4 for given **s** and ω_1. After ω_2 and ω_4 are determined, the value of ω_3 can be obtained from the y and z components of eq 32. The result is

$$\cos\omega_3 = \frac{(\rho_2\sin\theta_3 + \sigma_2\cos\omega_4\cos\theta_3)\times(-q_x\sin\theta_2 + q_y\cos\theta_2) + \sigma_2 q_z\sin\omega_4}{(-q_x\sin\theta_2 + q_y\cos\theta_2)^2 + q_z^2}$$

$$\sin\omega_3 = \frac{(\rho_2\sin\theta_3 + \sigma_2\cos\omega_4\cos\theta_3)q_z - (-q_x\sin\theta_2 + q_y\cos\theta_2)\sigma_2\sin\omega_4}{(-q_x\sin\theta_2 + q_y\cos\theta_2)^2 + q_z^2} \tag{36}$$

where q_x, q_y, and q_z are cartesian components of $\mathbf{q}_3$ (given in eq 31). By substituting eq 29, 33, 35, and 36 into eq 20, we obtain a single algebraic equation in the single unknown ω_1. This equation can be solved numerically, and a numerical example will be given in section IV. Equations 27, 28 and 34 are the necessary and sufficient conditions for eq 16 to be soluble for ω_2, ω_3, and ω_4 for given **s** and ω_1. When these conditions are satisfied, eq 16 can be converted to a single algebraic equation in the single unknown ω_1. But, this equation may, or may not, have solution(s) for ω_1. Therefore, eq 27, 28, and 34 constitute necessary conditions for eq 16 and 20 to be soluble for ω_1, ω_2, ω_3, and ω_4 for given **s**. The determination of ω_5 and ω_6 is then carried out as indicated in the paragraph preceding eq 20.

IV. Case of Polypeptide Chains

In this section, we consider the case of polypeptide chains having Pauling–Corey geometry.[18] It will be shown here that, even though the backbone of such a chain contains bonds about which no rotation occurs (*i.e.*, the C′–N bonds), equations of the form of eq 12a, 13a, and 14a are reducible to the form of eq 16, 17, and 18. Therefore, the method of reducing a set of algebraic equations in six unknowns to a single equation in one unknown (developed in section III) is also applicable to this case. Numerical illustrations of the method developed in section III will be given for the particular case of polypeptide chains.

The backbone of a polypeptide chain is made up of repeating structures of the form –C′–C$^\alpha$–N–, indexed according to the standard nomenclature and conventions of Edsall, *et al.*[19] In the Pauling–Corey geometry, four backbone atoms (C_i^α, C_i', N_{i+1}, and C_{i-1}^α) are in a single plane, *i.e.*, the dihedral angles around the backbone bond C_i'–N_{i-1} are kept fixed at 0°. Since there are two kinds of variable dihedral angles, ϕ and ψ, in the backbone, we define two kinds of local coordinate systems along the chain. An even-numbered coordinate system $2i$ is defined by atoms C_i^α, C_i', and N_{i+1} in such a way that these atoms have the cartesian coordinates of Table I[22] in this coordinate system. An odd-numbered coordinate system $2i - 1$ is defined by atoms N_i, C_i^α, C_i' similarly (see Table II[22] for the cartesian coordinates of these atoms in this coordinate system). Using the data of Tables I and II, the matrices and vectors which appear in the formula for the transformation of coordinates have the forms

TABLE I

CARTESIAN COORDINATES[a] OF ATOMS IN LOCAL COORDINATE SYSTEM $2i$ IN ÅNGSTRÖM UNITS

Atom	x	y	z
C_i^α	0.000	0.000	0.000
C_i'	1.530	0.000	0.000
N_{i+1}	2.067	1.206	0.000
C_{i-1}^α	3.519	1.436	0.000

[a] These coordinates correspond to a planar *trans*-peptide unit, with Pauling–Corey geometry.[18] They differ from those of T. Ooi, R. A. Scott, G. Vanderkooi, and H. A. Scheraga, *J. Chem. Phys.*, **46**, 4410 (1967).

TABLE II

CARTESIAN COORDINATES OF ATOMS in LOCAL COORDINATE SYSTEM $2i - 1$ IN ÅNGSTRÖM UNITS

Atom	x	y	z
N_i	0.000	0.000	0.000
C_i^α	1.470	0.000	0.000
C_i'	1.980	1.443	0.000

$$\mathbf{R}_{2i+1} = \mathbf{R}\phi_{i+1} \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_{i-1} & -\sin\phi_{i-1} \\ 0 & \sin\phi_{i-1} & \cos\phi_{i+1} \end{pmatrix} \tag{37}$$

$$\mathbf{T}_{2i} = \mathbf{T}_\alpha \equiv \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix},\ \alpha = 9.0° \tag{38}$$

$$\mathbf{p}_{2i} = \mathbf{p}_2 \equiv \begin{pmatrix} 2.067 \\ 1.206 \\ 0.000 \end{pmatrix} \tag{39}$$

$$\mathbf{R}_{2i} = \mathbf{R}_{\psi_i+\pi} \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\psi_i + \pi) & -\sin(\psi_i + \pi) \\ 0 & \sin(\psi_i + \pi) & \cos(\psi_i + \pi) \end{pmatrix} \tag{40}$$

$$\mathbf{T}_{2i-1} = \mathbf{T}_\beta \equiv \begin{pmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{pmatrix},\ \beta = 70.5° \tag{41}$$

$$\mathbf{p}_{2i-1} = \mathbf{p}_1 \equiv \begin{pmatrix} 1.470 \\ 0.000 \\ 0.000 \end{pmatrix} \tag{42}$$

The appearance of the nonvanishing y component in $\mathbf{p}_{2i}$ is a consequence of the planarity of the peptide unit (see Table I). However, even though the y component is nonvanishing, eq 12a, 13a, and 14a (with the quantities given by eq 37–42) can still be reduced to eq 16, 17, and 18. In this case of the polypeptide chain with planar peptide units, the vectors $\mathbf{q}_i$ defined by eq 19 have the following form, which is obtained by substituting eq 37–42 into eq 19

---

[22] For more detailed geometrical data, see H. A. Scheraga, *Advan. Phys. Org. Chem.*, **6**, 103 (1968).

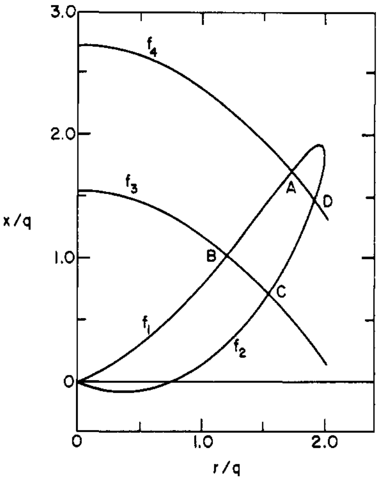

Figure 3. The region which satisfies eq 45, 46, and 50.

$$\mathbf{q}_0 = \mathbf{q}_1 = \mathbf{q}_2 = \mathbf{T}_\alpha \mathbf{p}_1{}^+\mathbf{p}_2 = \mathbf{q} \equiv \begin{pmatrix} \rho \\ \sigma \\ 0 \end{pmatrix} \equiv \begin{pmatrix} 3.519 \\ 1.436 \\ 0.000 \end{pmatrix} \quad (43)$$

Then eq 16 takes the following form in this case

$$\mathbf{s} = \mathbf{q} + \mathbf{T}_\alpha \mathbf{R}_{\phi_1} \mathbf{T}_\beta \mathbf{R}_{\psi_1+\pi}\mathbf{q} + \mathbf{T}_\alpha \mathbf{R}_{\phi_1} \mathbf{T}_\beta \mathbf{R}_{\psi_1+\pi} \mathbf{T}_\alpha \mathbf{R}_{\phi_2} \mathbf{T}_\beta \mathbf{R}_{\psi_2+\pi}\mathbf{q} \quad (44)$$

The idea developed in section III was to solve eq 44 for ψ_1, ϕ_2, and ψ_2 for given **s** and ϕ_1. The condition for the existence of solutions for eq 44 is given by

$$2q \geq r \geq 0 \quad (45)$$

and

$$f_1(r) \geq x \geq f_2(r) \quad (46)$$

where

$$\begin{aligned} f_1(r) &= [\rho r^2 + \sigma r(4q^2 - r^2)^{1/2}]/2q^2 \\ f_2(r) &= [\rho r^2 - \sigma r(4q^2 - r^2)^{1/2}]/2q^2 \end{aligned} \quad (47)$$

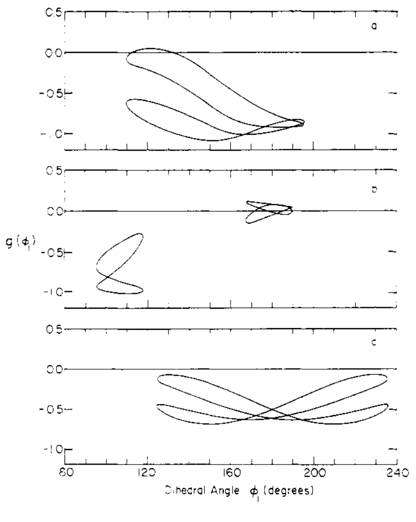

Figure 4. Graphs of $g \equiv g(\varphi_1)$ for (a) $\mathbf{s}$ = (2.913, 3.398, −2.407), $\mathbf{u}$ = (0.730, 0.280, 0.624); (b) $\mathbf{s}$ = (8.821, 1.824, −3.101), $\mathbf{u}$ = (0.059, −0.852, −0.521); (c) $\mathbf{s}$ = (9.000, 0.000, 0.000), $\mathbf{u}$ = (0.000, 1.000, 0.000).

Here q (= 3.8007) is the length of the vector **q** given by eq 43, and

$$r^2 = (\mathbf{s} - \mathbf{q})^2 \quad (48)$$

$$x = t_x \cos\beta + t_y \sin\beta \cos\phi_1 + t_z \sin\beta \sin\phi_1 \quad (49)$$

where t_x, t_y, and t_z are the x, y, and z components of the vector $\mathbf{T}_\alpha{}^{-1}(\mathbf{s} - \mathbf{q})$. Equations 45 and 46 are special cases of eq 27 and 28, respectively. The condition of eq 34 takes the following form in this case

$$f_3(r) \geq x \geq f_4(r) \quad (50)$$

where

$$\begin{aligned} f_3(r) &= \frac{2\sigma^2(\sin\alpha + \sin\beta) + 2\sigma\rho(\cos\alpha + \cos\beta) - r^2 \sin\alpha}{2(\sigma\cos\alpha - \rho\sin\alpha)} \\ f_4(r) &= \frac{2\sigma^2(\sin\alpha - \sin\beta) + 2\sigma\rho(\cos\alpha + \cos\beta) - r^2 \sin\alpha}{2(\sigma\cos\alpha - \sigma\sin\alpha)} \end{aligned} \quad (51)$$

The range in the $(r/q, x/q)$ plane, in which the point $(r/q, x/q)$ satisfies eq 45, 46, and 50, is shown in Figure 3. If a point $(r/q, x/q)$ falls into a quadrilateral-like region ABCD, there exist in general four sets of solutions of eq 44 for given **s** and ϕ_1 (see discussion following eq 35). This is a necessary condition that the set of eq 16, 17, and 18 (with its quantities interpreted as in eq 37–42) have one or more solutions for a given set of vectors **s**, **u**, and **v**. From Figure 3, we can conclude that r must lie in a narrower range than that indicated in eq 45, *viz.*

$$1.9069q \geq r \geq 1.2051q \quad (52)$$

Once the point $(r/q, x/q)$ falls into the "good" region, the left-hand side of eq 20, or

$$g(\phi_1) = \mathbf{u}^+ \mathbf{T}_\alpha \mathbf{R}_{\phi_1} \mathbf{T}_\beta \mathbf{R}_{\psi_1+\pi} \mathbf{T}_\alpha \mathbf{R}_{\phi_2} \mathbf{T}_\beta \mathbf{R}_{\psi_2+\pi} \mathbf{T}_\alpha \mathbf{e}_1 - \cos\beta \quad (53)$$

(in this case of polypeptide chains), can be calculated by solving eq 44 for ψ_1, ϕ_2, and ψ_2 (in terms of ϕ_1) and substituting the results into eq 53. By following the same procedure for different values of ϕ_1, we can obtain a graph of $g \equiv g(\phi_1)$, a few examples of which are shown in Figure 4. The curves are found to exist either in one continuous range (Figures 4a and 4c) or in two separate ranges (Figure 4b). The ranges in which the curves of $g \equiv g(\phi_1)$ exist will be discussed in section V in connection with a proof of the impossibility of the existence of cyclic tetrapeptides having the Pauling–Corey planar *trans*-peptide structure. The value of ϕ_1 at which $g(\phi_1) = 0$ is a solution of eq 20. Once ϕ_1 (or, in general, ω_1) is obtained, the other unknowns can be determined by eq 17, 18, 29, 33, 35, and 36. A computer program has been written in Fortran IV, which finds all the roots of eq 20 (for the case of polypeptide chains) for a given set of vectors **s**, **u**, and **v** by means of Newton's method.[21] It takes about 0.1 sec to solve the equation for one set of the vectors **s**, **u**, and **v** on an IBM 360/65 computer. The number of solutions of the equation $g(\phi_1) = 0$ was found empirically to vary from zero to eight. In the examples shown in Figures 4a, 4b, and 4c, the number of solutions is 2, 4, and 0, respectively.

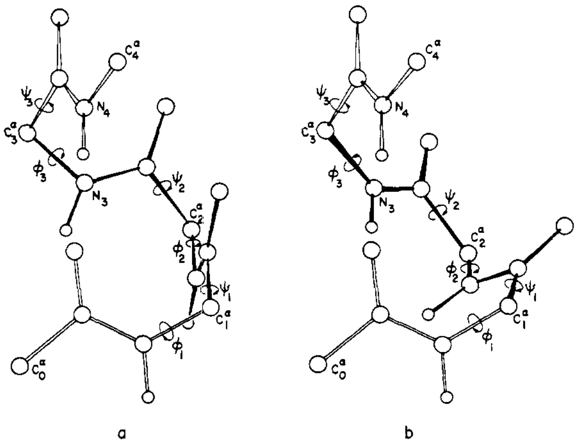

Figure 5. Conformations corresponding to the two roots for the curve in Figure 4a. The positions of the atoms which are connected by unshaded chemical bonds are the same in the two conformations. The conformation in (a) is a part of the right-handed α helix. The values of the dihderal angles ϕ_1, ψ_1, ϕ_2, ψ_2, ϕ_3 and ψ_3 are 132.0, 123.0, 132.0, 123.0, 132.0, and 123.0° for (a) and 112.7, 171.4, 88.4, 157.0, 88.4, and 131.5° for (b).

The conformations corresponding to the two roots found in Figure 4a are illustrated in Figures 5a and 5b together with their dihedral angles. In both Figures 5a and 5b, peptide unit 4 (consisting of the six atoms from C_3^α to C_4^α) is fixed in the *same position* relative to peptide unit 1 (consisting of the six atoms from C_0^α to C_1^α). The two conformations shown in Figure 5 are the only two geometrically possible ones for the given relative positions of peptide units 4 and 1. The conformation in Figure 5a is a part of the right-handed α-helix, and has a hydrogen bond between O_0 and $(H)^{N_4}$. In the conformation of Figure 5b, a bifurcated hydrogen bond[23,24] can be formed between O_0 and $(H)^{N_3}$ and $(H)^{N_4}$, but the atoms N_1 and N_2 (whose interatomic distance is 2.31 Å) violate the steric hindrance criterion[20,25] slightly. In general, it is quite possible that a conformation obtained by solving the (geometrical) equations presented here may be sterically hindered, because (repulsive) steric effects are not taken into account in eq 12a, 13a, and 14a or in eq 16, 17, and 18. Of course, energy minimization will relieve such steric hindrances.

V. Impossibility of Existence of Cyclic Tri- and Tetrapeptides with Pauling–Corey Planar *trans*-Peptide Structure

In this section, the result of section IV will be used to prove that cyclic tri- and tetrapeptides having the Pauling–Corey planar *trans*-peptide structure cannot exist geometrically, *i.e.*, the ring cannot be closed in such structures.

In the case of cyclic tripeptides, which have six variable dihedral angles, **s** of the first of eq 15 is equal to zero. Substituting **s** = **0** in eq 48, we obtain $r = q$. Since this value falls outside the range of eq 52, cyclic tripeptides with planar *trans*-peptide units having Pauling–Corey geometry cannot exist.

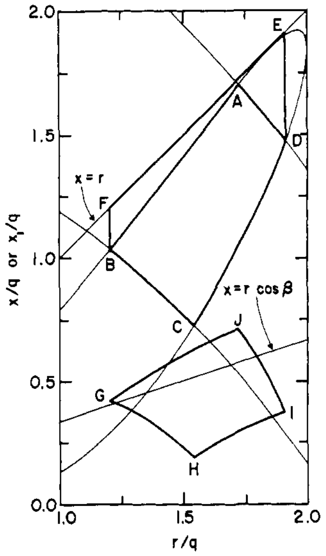

Figure 6. Diagram showing the impossibility of existence of cyclic tetrapeptides with the planar *trans*-peptide structure having Pauling–Corey geometry. See text for details.

In the case of cyclic tetrapeptides, where $n = 8$, the first of eq 15 (making use of eq 38, 39, 41, 42, 43) takes the form

$$\mathbf{s} = -\mathbf{R}_{\psi_4+\pi}{}^{-1}\mathbf{T}_\beta{}^{-1}\mathbf{R}_{\phi_4}{}^{-1}\mathbf{T}_\alpha{}^{-1}\mathbf{q} \tag{54}$$

where the dihedral angles $\omega_1, \ldots, \omega_8$ in eq 12 and 15 are interpreted as ϕ_4, $\psi_4 + \pi$, ϕ_1, $\psi_1 + \pi$, ϕ_2, $\psi_2 + \pi$, ϕ_3, and $\psi_3 + \pi$, respectively. The impossibility of the existence of cyclic tetrapeptides with planar *trans*-peptide units having Pauling–Corey geometry will be proved by showing that the points $(r/q, x/q)$, calculated from eq 48 and 49 when **s** is given by eq 54, fall outside the "good" region in Figure 3 for any values of ϕ_1, ϕ_4, and ψ_4 in eq 49 and 54; only these three angles appear in these two equations.

In order to show this, we first consider the movement of the point (r,x) as ϕ_1 is changed for fixed values of ϕ_4 and ψ_4. The value of r is independent of ϕ_1, as can be seen from eq 48. From eq 49, it follows that, as ϕ_1 changes in the range $[0,2\pi]$, x changes in the range

$$x_1 \geq x \geq x_0 \tag{55}$$

where

$$\begin{aligned} x_1 &= t_x \cos\beta + (r^2 - t_x{}^2)^{1/2} \sin\beta \\ x_0 &= t_x \cos\beta - (r^2 - t_x{}^2)^{1/2} \sin\beta \end{aligned} \tag{56}$$

For any **s**, the lower limit x_0 is always lower than points which satisfy eq 45, 46, and 50, *i.e.*, those points which lie within the region ABCD of Figure 6 (as already shown in Figure 3); this results from the inequality

$$r\cos\beta \geq t_x \cos\beta - (r^2 - t_x{}^2)^{1/2} \sin\beta \tag{57}$$

and the observation from Figure 6 that the point $(r, r\cos\beta)$ lies below the region ABCD. When r satisfies eq 52, there are three possibilities for the position of the upper limit x_1: (a) the point (r,x_1) lies below the region ABCD, (b) the point (r,x_1) lies in the region ABCD, (c) the point (r,x_1) lies above the region ABCD. As

---

[23] W. H. Baur, *Acta Cryst.*, **19,** 909 (1965).

[24] G. Nemethy, D. C. Phillips, S. J. Leach, and H. A. Scheraga, *Nature*, **214,** 363 (1967).

[25] S. J. Leach, G. Nemethy, and H. A. Scheraga, *Biopolymers*, **4,** 369 (1966).

ϕ_1 changes from zero to 2π, the point (r,x) makes a round trip between the lower limit (r,x_0) and the upper limit (r,x_1). There are solutions of eq 44 for a given vector **s** in a range or ranges of ϕ_1 where the corresponding point (r,x) lies in the region ABCD during the trip. In case a, the point never comes into the region ABCD. Therefore, there is no solution of eq 44, or there exists no curve $g \equiv g(\phi_1)$, for any ϕ_1. In case b, the point lies in the region ABCD for one continuous range of ϕ_1. Therefore, curves $g \equiv g(\phi_1)$ look like Figure 4a or 4c. In case c, the point (r,x) passes through the region ABCD twice. Therefore, curves $g \equiv g(\phi_1)$ exist in two different ranges of ϕ_1 as in Figure 4b. Summarizing the above, we can conclude the following: the necessary and sufficient condition that there exist any such sets of values for the angles ϕ_1, ψ_1, ϕ_2, and ψ_2 that satisfy eq 44 for a given vector **s** is that the point (r,x_1), where r and x_1 are defined by eq 48 and the first of eq 56, respectively, lie in a region enclosed by BCDEF in Figure 6. The upper limit of the region is given by $r = x$, as can be seen easily from the first of eq 56.

The values of r and x_1 are determined completely when a vector **s**, or, in the case of cyclic tetrapeptides, ϕ_4 and ψ_4 in eq 54, are given. As a point (ϕ_4,ψ_4) moves in the region of $0 \leq \phi_4 \leq 2\pi$, $0 \leq \psi_4 \leq 2\pi$ in the (ϕ_4,ψ_4) plane, the point (r,x_1) moves in a region of the (r,x_1) plane. The boundary of this region is determined by points which satisfy the equation

$$\begin{vmatrix} \partial r/\partial\phi_4 & \partial r/\partial\psi_4 \\ \partial x_1/\partial\phi_4 & \partial x_1/\partial\psi_4 \end{vmatrix} = 0 \tag{58}$$

From this equation, the region is found to be the one enclosed by points GHIJ in Figure 6. The region GHIJ does not overlap with the region BCDEF. Hence, there are no solutions of eq 44 when **s** is given by eq 54 for arbitrary ϕ_4 and ψ_4. Therefore, cyclic tetrapeptides with planar *trans*-peptide units having Pauling–Corey geometry cannot exist.

It should be emphasized here that the nonexistence was proved only for cyclic tri- and tetrapeptides which have *planar trans-peptide units with Pauling–Corey geometry*. The proof does not exclude the possibilities of synthesizing cyclic tri- and/or tetrapeptide molecules, in which at least one peptide unit does not have the planar *trans* conformation with Pauling–Corey geometry. In fact, cyclo-tri-L-prolyl and cyclo-tetraglycyl have been synthesized by Rothe, *et al.*[26] and by Schwyzer, *et al.*,[27] respectively. Venkatachalam[28] studied the possible sterically allowed conformations of cyclo-tri-L-prolyl by imposing a condition of threefold symmetry, but allowing for departures from Pauling–Corey geometry; he actually found the molecule to be sterically allowed if the peptide units are almost in the *cis* conformation (the deviation from planarity being approximately $\Delta\omega = 25°$). Ramakrishnan and Sarathy[29] studied the conformations of cyclo-tetraglycyl and cyclo-tetra-L-alanyl which they assumed to have either fourfold or twofold symmetry; they found that cyclo-tetraglycyl and cyclo-tetra-L-alanyl with nonplanar *trans*-peptide units (the nonplanarity being $\omega = -9$ to $-12°$) and departures from Pauling–Corey geometry are sterically possible, and they located an energetically favored conformation which turns out to have fourfold symmetry in the sterically allowed range of conformations. Schwyzer, *et al.*,[27] proposed a conformation with twofold symmetry for cyclo-tetraglycyl with *trans*-peptide units; such a model would have to have departures from Pauling–Corey geometry. In the present paper, no assumption about symmetries of molecules was imposed. Therefore, the above proof of the nonexistence of cyclic tri- and tetrapeptides with *planar trans-peptide units having Pauling–Corey geometry* excludes the possibility of existence of any conformations of such molecules whether or not they have any symmetry.

VI. Derivatives of Dependent Angles

In sections III and IV, methods were developed for solving eq 12a, 13a, and 14a for a given set of vectors **s**, **u**, and **v** for special forms of $\mathbf{T}_i$ and $\mathbf{p}_i$. In actual applications, **s**, **u**, and **v** are functions of the angles which are chosen as the independent variables as, for example, in eq 15; therefore, the solutions ω_i of eq 12a, 13a, and 14a are also functions of such independent variables. In this section, we outline a method for calculating first and second derivatives of dependent angles (*i.e.*, the solutions ω_i of eq 12a, 13a, and 14a) with respect to the independent variables. Since no assumptions will be made in this section about the forms of $\mathbf{T}_i$ and $\mathbf{p}_i$, the method to be described is a general one. Before presenting it, we devote the next two paragraphs to a discussion of the uses to which first and second derivatives of dependent angles are put in connection with conformational energy calculations.

First and second derivatives of the conformational energy surface play a very important role in studies of polypeptides. The first derivatives of the energy function are used in some energy minimization procedures.[30] The second derivatives of the energy function are used in the calculation of the statistical weights of minimum-energy conformations.[14,31–34] The calculations of the derivatives of an energy function can be reduced to those of the derivatives of the relative position vectors of atoms in the molecule with respect to the dihedral angles chosen as independent variables.[34] If some of the dihedral angles change as functions of other dihedral angles chosen as independent variables, then we need derivatives of the dependent angles with respect to the independent ones in order to calculate the derivatives of the relative position vectors of the atoms in the molecule.

Another example of the use of the first derivatives of dependent variables with respect to independent ones

---

(26) M. Rothe, K. D. Steffen, and I. Rothe, *Angew. Chem. Intern. Ed. Engl.*, **4**, 356 (1965).

(27) R. Schwyzer, B. Iselin, W. Rittel, and P. Sieber, *Helv. Chim. Acta*, **39**, 872 (1956).

(28) C. M. Venkatachalam, *Biochim. Biophys. Acta*, **168**, 397 (1968).

(29) C. Ramakrishnan and K. P. Sarathy, *Biochim. Biophys. Acta*, **168**, 402 (1968).

(30) K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci., U. S.*, **58**, 420 (1967).

(31) N. Go, M. Gō, and H. A. Scheraga, *ibid.*, **59**, 1030 (1968).

(32) K. D. Gibson and H. A. Scheraga, *Physiol. Chem. Phys.*, **1**, 109 (1969).

(33) K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci., U. S.*, **63**, 242 (1969).

(34) M. Gō, N. Gō, and H. A. Scheraga, *J. Chem. Phys.*, in press.

is the problem of local deformations of general chain molecules, which was formulated mathematically in section II as the solution of eq 12a, 13a, and 14a for a given set of vectors **s**, **u**, and **v** (whose actual values are given in terms of the independent variables in eq 15). In sections III and IV, special cases (which, however, are those of most interest) were discussed in which eq 12a, 13a, and 14a could be reduced to a single algebraic equation in one unknown, which, in turn, could be solved quite efficiently by Newton's method. In fact, all roots of eq 20 with ω_2, ω_3, and ω_4 being given as functions of ω_1, could be located very rapidly (see section IV). However, the reduction of eq 12a, 13a, and 14a to a single algebraic equation in one unknown is not possible for more general cases. It is very difficult to locate all roots of eq 12a, 13a, and 14a by the multidimensional Newton method. But, the latter is an efficient one for locating a root if a good approximate estimate of the root is known by any method; this is the situation in the problem of a *local* deformation of a chain molecule as can be seen below. Let us assume that we already have a chain molecule in a certain conformation, and we want to make a local deformation of conformation in a certain section of the molecule. Then, we change the value of the independent variable(s) in eq 15 by such a small amount that the vectors **s**, **u**, and **v** do not change by a large amount. We solve eq 12a, 13a, and 14a by the multidimensional Newton method, *starting* from the values of the unknowns corresponding to the undeformed conformation. If a large deformation (confined to a local section of a chain molecule) is to be made, this can be accomplished by a number of small stepwise deformations. The first derivatives of the dependent variables with respect to the independent ones are used in this connection. Since the method described below for obtaining the derivatives does not assume any particular form for $\mathbf{T}_i$ and $\mathbf{p}_i$, the method outlined above for making local deformations can be applied in general to any chain molecule.

Turning now to the method for computing derivatives, we note that, for any $\mathbf{T}_i$ and $\mathbf{p}_i$, eq 14a can always be transformed to eq 18. Thus, ω_6 and its derivatives are given explicitly in terms of **v** and of $\omega_1,\ldots,\omega_5$ and their derivatives. The variable ω_6 does not appear in either eq 12a or 13a, since $\mathbf{R}_6\mathbf{e}_1$ in eq 13a is equal to $\mathbf{e}_1$. Therefore, the six equations in eq 12a and 13a, of which five are independent (three in eq 12a, and three in eq 13a of which two are independent), can be considered as the ones for obtaining the five unknowns $\omega_1,\ldots,\omega_5$. Once the derivatives of $\omega_1,\ldots,\omega_5$ are calculated from eq 12a and 13a, the derivatives of ω_6 can be calculated from eq 14a. Taking the first derivatives of both sides of eq 12a and 13a with respect to an independent variable ζ (any one of the variables in eq 15), we have

$$\mathbf{s}_\zeta = \sum_{i=1}^{5} \mathbf{s}_i\omega_{i\zeta} \tag{59}$$

$$\mathbf{u}_\zeta = \sum_{i=1}^{5} \mathbf{u}_i\omega_{i\zeta} \tag{60}$$

where $\mathbf{s}_\zeta$, $\mathbf{u}_\zeta$, and $\omega_{i\zeta}$ are the first derivatives of **s**, **u**, and ω_i with respect to ζ (with $\mathbf{s}_\zeta$ and $\mathbf{u}_\zeta$ being given quantities and $\omega_{i\zeta}$ being unknowns); $\mathbf{s}_i$ and $\mathbf{u}_i$ are vectors given by the right-hand sides of eq 12a and 13a, respectively, with $\mathbf{R}_i$ being replaced by $\partial\mathbf{R}_i/\partial\omega_i$, but $\mathbf{R}_j$ $(j \neq i)$ kept unchanged. Only two of the three linear equations of eq 60 are independent. The first derivatives $\omega_{i\zeta}$ can be obtained by solving the five simultaneous linear equations in eq 59 and 60 (three of eq 59 and two independent ones of eq 60). Taking second derivatives of both sides of eq 12a and 13a with respect to the independent variables ζ and ξ, which may or may not be identical with each other, we have

$$\mathbf{s}_{\zeta\xi} = \sum_{i,j=1}^{5} \mathbf{s}_{ij}\omega_{i\zeta}\omega_{j\xi} + \sum_{i=1}^{5} \mathbf{s}_i\omega_{i\zeta\xi} \tag{61}$$

$$\mathbf{u}_{\zeta\xi} = \sum_{i,j=1}^{5} \mathbf{u}_{ij}\omega_{i\zeta}\omega_{j\xi} + \sum_{i=1}^{5} \mathbf{u}_i\omega_{i\zeta\xi} \tag{62}$$

where $\mathbf{s}_{\zeta\xi}$, $\mathbf{u}_{\zeta\xi}$, and $\omega_{i\zeta\xi}$ are the second derivatives of **s**, **u**, and ω_i with respect to ζ and ξ; $\mathbf{s}_{ij}$ and $\mathbf{u}_{ij}$ are vectors given by the right-hand sides of eq 12a and 13a, respectively, with $\mathbf{R}_i$ and $\mathbf{R}_j$ being replaced by $\partial\mathbf{R}_i/\partial\omega_i$ and $\partial\mathbf{R}_j/\partial\omega_j$, respectively, for $i \neq j$, or with $\mathbf{R}_i$ being replaced by $\partial^2\mathbf{R}_i/\partial\omega_i^2$ for $i = j$, but $\mathbf{R}_k$ $(k \neq i, j)$ kept unchanged. Substituting the values of $\omega_{i\zeta}$ and $\omega_{j\xi}$, obtained by solving eq 59 and 60, into eq 61 and 62, we have a set of linear equations for the unknowns $\omega_{i\zeta\xi}$, which can be solved easily by the usual methods.