cluding a RDC of >95% conversion, as in the present work) were independently prepared by Prof. T. Painter (private communication).
(14) Meille, S. V.; Allegra, G.; Naggi, A.; Torri, G.; Casu, B. presented at the XIIth International Carbohydrate Symposium, Utrecht, 1984; Abstract p 525.
(15) Rees, D. A.; Morris, E. R.; Thom, D.; Madden, J. K. "Shapes and Interactions of Carbohydrate Chains" in "The Polysaccharides" Aspinall, G. O., Ed.; Academic Press: 1982; Vol. 1, p 196.

Chain Closure with Bond Angle Variations

Robert E. Bruccoleri† and Martin Karplus*

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138.
Received February 13, 1985

ABSTRACT: The chain-closure algorithm of Gō and Scheraga (Gō, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 178) has been modified to allow bond angle variations. This modification greatly increases the domain of applicability of the algorithm. Of particular interest is its use for bridging deletions or introducing additions in the homology modeling of proteins. Examples from an α-helix, a β-sheet, a cyclic polypeptide, and several proteins are presented.

1. Introduction

In the modeling of a homologous or mutant protein from a known X-ray structure, a deletion or insertion in the sequence requires the construction of a peptide backbone across two fixed endpoints. Generally, this problem is solved by energy minimization,[1] by plastic or metal models (which solve part of the minimization problem mechanically),[2] or by human manipulation of a model using molecular graphics.[3] A related problem arises in the conformational search of the possible structures for a polypeptide loop or cyclic polypeptide. In such a search, some of the backbone torsions in the loop are varied freely and the remainder must be assigned to values such that the loop closes back onto the attached end.

The ring closure and local conformational deformation procedure of Gō and Scheraga[4] solves this problem algorithmically. It determines the set of dihedral angles of a polymer chain required to bridge the chain across two fixed, oriented endpoints. Their procedure assumes that the bond lengths and bond angles of the polymer chain are fixed. We have found that this restriction is very severe in that it greatly reduces the range of endpoints for which a closure can be found. It is known that bond angles do vary by a few degrees in biopolymers[5,6] and that the entropy of proteins has significant contributions from bond angle variations.[7,8] We have developed a simple modification to the Gō and Scheraga procedure that allows small bond angle variations, and we illustrate its applicability by choosing examples from known protein structures.

Section 2 describes the method, and Section 3 presents the results and discussion.

2. Method

In this section, we develop the chain-closure algorithm including both dihedral angle and bond angle variation. To make clear the nature of our modifications to the original approach of Gō and Scheraga, we must briefly review their procedure. In what follows, we refer to equation n in their paper by the notation eq GSn; details are omitted, and the interested reader should consult the original paper.[4] For the free torsion angles of a polymer chain, we define coordinate systems as shown in Figure 1. For atom i, the x axis is along the bond vector between atom $i-1$ and the succeeding atom, the y axis is the component of the succeeding bond vector perpendicular to the x axis, and the z axis is given by the cross product of the other two axes. The bond vector between atom $i-1$ and the succeeding atom is the axis of rotation for the free torsion angle, ω_i.[9] In the case of a protein, the peptide torsion angle (i.e., the torsion angle about the C–N bond) is assumed to be planar and fixed (either *cis* or *trans*) so one atom in the backbone does not need to have a coordinate set defined for it. Gō and Scheraga omit the carbonyl carbon, and thus, the α-carbon and the peptide nitrogen are used for defining the coordinate systems.

The essence of the Gō and Scheraga procedure is the transormation of the coordinate systems given by eq GS1–3

$$\mathbf{r}_{i-1} = \mathbf{T}_{i-1}\mathbf{R}_i\mathbf{r}_i + \mathbf{p}_{i-1} \qquad \text{(GS1)}$$

$$\mathbf{T}_{i-1} = \begin{bmatrix} \cos\theta_{i-1} & -\sin\theta_{i-1} & 0 \\ \sin\theta_{i-1} & \cos\theta_{i-1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{(GS2)}$$

$$\mathbf{R}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega_i & -\sin\omega_i \\ 0 & \sin\omega_i & \cos\omega_i \end{bmatrix} \qquad \text{(GS3)}$$

$\mathbf{p}_{i-1}$ is a vector that specifies the translation between two coordinate systems. In the case of peptides, the z component of $\mathbf{p}_{i-1}$ is zero, but the y component is nonzero for translations between two coordinate systems that skip the carbonyl carbon (as shown for $\mathbf{p}_{i-1}$ on the left in Figure 1). $\mathbf{T}_{i-1}$ specifies the rotation associated with the angle, θ_{i-1}, and operates in the xy plane. $\mathbf{R}_i$ specifies the rotation associated with a torsion angle, ω_i, and operates in the yz plane.[10] In these equations, the degrees of freedom are the torsion angles, ω_i. For a peptide, Gō and Scheraga define the atom numbering such that the odd-numbered torsions are the φ torsion angles and the even-numbered torsions are the ψ torsion angles.

If we know the position of two amino acid residues that we wish to connect with a peptide chain containing a known number of residues, the problem of chain closure is equivalent to finding a series of coordinate transformations that transform one endpoint coordinate system into the other. To transform one arbitrary coordinate system into another requires six degrees of freedom, three

† Present address: Cellular and Molecular Research Laboratory, Massachusetts General Hospital, Boston, MA 02114.

0024-9297/85/2218-2767$01.50/0

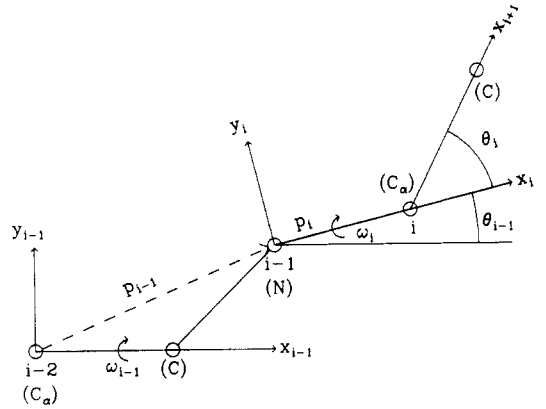


Figure 1. Local coordinate systems. An illustration of the relationship between atoms, bond lengths, bond angles, torsion angles, and local coordinate systems as defined by Gō and Scheraga for a polypeptide. Given the repetitive nature of the backbone, there are two different local coordinate systems, one having an origin on C_α (coordinate system $i-1$) and one on N (coordinate system i), and both are shown. The angle θ_{i-1} is the angle between the x axis for coordinate system $i-1$ and the x axis for coordinate system i. The matrix, $\mathbf{R}_{i-1}$, which accounts for the torsion angle rotation about the x_{i-1} axis, incorporates the isomerization of the peptide torsion angle (defined about the C–N bond), be it *cis* or *trans*.

for translation of the origin and three for the rotations necessary to superimpose the axes. Since there is one free torsion for each transformation described in eq GS1–3, six free torsions are necessary to solve the chain-closure problem. With each amino acid residue having two free torsions, three residues are necessary for chain closure.

Gō and Scheraga express the necessary transformation between the fixed endpoints using three vectors, **s**, **u**, and **v**. The vector, **s** is the translation from the first endpoint to the second, **u** is the direction of the x axis of the second endpoint coordinate system in the first endpoint coordinate system, and **v** is the direction of the y axis, defined analogously. Since **s**, **u**, and **v** can be computed directly from the coordinates of the two known endpoints, the solution of the problem can now be obtained, at least in principle. We can express **s** as a series of six coordinate transformations and **u** and **v** as a series of successive vector rotations (eq GS12–14).

$$\mathbf{s} = \mathbf{p}_0 + \mathbf{T}_0\mathbf{R}_1\mathbf{p}_1 + \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{p}_2 + \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{T}_2\mathbf{R}_3\mathbf{p}_3 + \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{T}_2\mathbf{R}_3\mathbf{T}_3\mathbf{R}_4\mathbf{p}_4 + \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{T}_2\mathbf{R}_3\mathbf{T}_3\mathbf{R}_4\mathbf{T}_4\mathbf{R}_5\mathbf{p}_5 \quad \text{(GS12)}$$

$$\mathbf{u} = \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{T}_2\mathbf{R}_3\mathbf{T}_3\mathbf{R}_4\mathbf{T}_4\mathbf{R}_5\mathbf{T}_5\mathbf{R}_6\mathbf{e}_1 \quad \text{(GS13)}$$

$$\mathbf{v} = \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{T}_2\mathbf{R}_3\mathbf{T}_3\mathbf{R}_4\mathbf{T}_4\mathbf{R}_5\mathbf{T}_5\mathbf{R}_6\mathbf{e}_2 \quad \text{(GS14)}$$

where $\mathbf{e}_1$ is (1, 0, 0), and $\mathbf{e}_2$ is (0, 1, 0). These transformations are equated to the values computed from the given endpoints, and these vector equations are solved for the free torsions. This yields six nonlinear equations in six unknowns, the torsion angles, ω_i.

In practice, Gō and Scheraga's next step was to reduce these equations into one scalar equation in one unknown, the angle ω_1; that is

$$\mathbf{u}^+\mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{T}_2\mathbf{R}_3\mathbf{T}_3\mathbf{R}_4\mathbf{T}_4\mathbf{e}_1 - \cos\theta_5 = 0 \quad \text{(GS20)}$$

where $\mathbf{u}^+$ is the transpose of **u**. We refer to the left side of this equation as $G(\omega_1)$. The values for ω_2, ω_3, and ω_4 (needed for specifying the elements of the $\mathbf{R}_i$) are all expressed in terms of ω_1. The manipulations necessary for these substitutions require the solution of two independent quadratic equations; thus two independent signs are introduced into the equation. Further, these substitutions can reduce the allowed domain of ω_1 in a variety of ways: ω_1 can cover the entire angle space, it can be limited to just one subrange, it can be limited to two disjoint subranges, or at worst, $G(\omega_1)$ may not defined for any value of ω_1. The choice of the two signs does not affect the bounds on the permissible values of ω_1, so if $G(\omega_1)$ is defined, it has four possible values.

Given these characteristics of eq GS20, its solution requires iterations over the possible subranges of ω_1 as well as iterations over the two independent signs. We solve for zeroes of $G(\omega_1)$ using Muller's method and deflation.[11,12] Muller's method is a robust and accurate algorithm that fits three points of a function to a parabola and looks for zeroes at the roots of the parabola. If in the course of applying Muller's method, we detect a change in sign for $G(\omega_1)$, then we switch to the Modified regula falsi method,[11] which is also very efficient and always successful when the root can be bracketed.

We have noticed that Muller's method finds the value of ω_1 that yields the minimum magnitude of the function if it has no roots. This is a consequence of its use of a parabolic fit. When Muller's method encounters a negative discriminant in the solution for the zeroes of the fitted parabola, it sets the discriminant to zero and continues. The resulting estimate of the root is also the point where the slope of the parabola is zero.

Our modification to the Gō and Scheraga procedure is invoked when no zeroes to $G(\omega_1)$ are found. The principle is to attempt to minimize the minimum magnitude of $G(\omega_1)$ by steepest descents minimization[13] applied to the peptide bond angles. Since θ_i and therefore $\mathbf{T}_i$ all depend on peptide bond angles, these variations do affect the value of $G(\omega_1)$. Since the assumption of peptide planarity requires the θ_{2i} be computed as the difference between two peptide bond angles, the six $\mathbf{T}_i$ matrices require nine peptide bond angles for their evaluation, and, therefore, we have nine angles to vary. The maximum variation in any bond angle is limited by some arbitrary small angle that we designate MAXDT. We normally use a value of 5° for MAXDT because the energy of such a deformation is roughly kT at room temperature, assuming the bond angle force constants used with the macromolecular mechanics program, CHARMM.[14] We proceed in two different ways depending on whether there are values of ω_1 where G is defined.

If there are values of ω_1 where G is defined, an approximation to the derivative of the minimum magnitude of G with respect to the nine peptide bond angles is computed. There are two parts to this approximation, the numerical computation of the derivative and the computation of the minimum magnitude of $G(\omega_1)$.

At the start of computing the derivative, the minimum magnitude of $G(\omega_1)$ is known from the previous attempt at finding its roots. Thus, computing the derivative requires perturbing the values of each of the bond angles by the maximum of 0.001 radians or MAXDT/10 and determining the minimum magnitude of G at the perturbed angle. The numerical derivative is then the change in minimum magnitude of G divided by the perturbation. It is not always possible to perturb a particular bond angle because the perturbation may result in $G(\omega_1)$ becoming undefined or it may change the perturbed angle by more than MAXDT degrees from its initial value. We try both positive and negative perturbations, and if both fail, we set the derivative to zero.

Because small perturbations in the bond angles do not produce large changes in the shape of $G(\omega_1)$, we approximate the minimum magnitude of G by computing just a

single value of G using the same signs as found during the root search and using a value of ω_1 that has the same relative position with respect to the limits of ω_1 as shown in the equation below

$$\omega_1^{new} = \omega_{1,low}^{new} + (\omega_1^{old} - \omega_{1,low}^{old}) \frac{\omega_{1,hi}^{new} - \omega_{1,low}^{new}}{\omega_{1,hi}^{old} - \omega_{1,low}^{old}} \quad (1)$$

where *new* refers to values after a bond angle perturbation, *old* refers to the initial value, *low* refers to the lower bound for ω_1, and *hi* refers to the upper bound for ω_1. For example, if $G(\omega_1)$ is defined for values of ω_1 from 20° to 100°, the minimum magnitude of $G(\omega_1)$ is at $\omega_1 = 40°$, and the defined range after perturbing the bond angle is 10° to 110°, we would approximate the minimum magnitude of G to be $G(35°)$. If the number of defined regions should change, *low* and *hi* refer to the limits over all defined regions. If ω_1^{new} is computed to be a value where $G(\omega_1)$ is undefined, then it is adjusted to the closest defined value for ω_1.

Once we have the derivative of the minimum magnitude of $G(\omega_1)$, we use steepest descents minimization[13] to lower the minimum magnitude of $G(\omega_1)$. The use of steepest descents generally finds the smallest angle variation necessary for closure. A normalized step of 1 is used for the initial trial. If the first step results in a higher minimum magnitude of $G(\omega_1)$, the step is decreased by a factor of 10. If a step decreases the minimum magnitude, the step size is increased by a factor of 1.5. The minimization proceeds until the minimum magnitude goes to zero (and the roots extracted), the values of the bond angles hit the limits of bond angle variation (the search fails), or $G(\omega_1)$ becomes undefined (the search also fails).

If G is not defined for any value of ω_1, we search over the bond angles to find values where it is defined. The search can be performed by two different methods, one is quick and the other is thorough. The first method is to subtract and then add MAXDT to each angle in turn stopping either when G is defined or when all 18 possibilities are exhausted. Each angle is left at the largest value before moving to the next, because this results in more successful searches than leaving it at the initial value. The second method is to try all possible combinations of the lowest and highest values for all nine angles. If G is defined for a particular combination, we sample the value of G in the defined regions and use the angles that yield the smallest magnitude for G. If either method finds a set of bond angles for which G is defined, the steepest descent search described above is applied if necessary to obtain the roots of $G(\omega_1)$. In practice, the quick search is only slightly less effective than the thorough search, but is much more efficient, so only the quick search is used.

3. Results and Discussion

In this section, we apply the method to regular α-helices and β-sheets, to a cyclic pentapeptide, and to proteins. Unless otherwise noted, the peptide torsion angle is *trans*.

3.1. α-Helix and β-Sheet. As the first example of the effect of flexibility in the bond angles, we consider an α-helix with the Pauling–Corey geometry.[4,15] We constructed a helix using these bond lengths and bond angles and with φ, ψ torsion angles of −48° and −57°. We then deleted three amino acid residues from the helix and attempted the Gō and Scheraga procedure to close the gap, which succeeded. If instead of the initial geometry of 109.5° for the N–C_α–C angle, we use a value of 112.0°, the Gō and Scheraga procedure fails to find a solution. Figure 2A shows a plot of $G(\omega_1)$. When we permit bond angle adjustments, our modified procedure changes the nine bond angles in the three amino acid residues as shown in Table I; Figure 2B shows the modified $G(\omega_1)$. The torsion angles (φ_i, ψ_i) for the new helix are (−50.7, −47.6), (−54.9, −53.3), and (−54.4, −55.1) with a root-mean-square shift of backbone positions of 0.0672 Å.

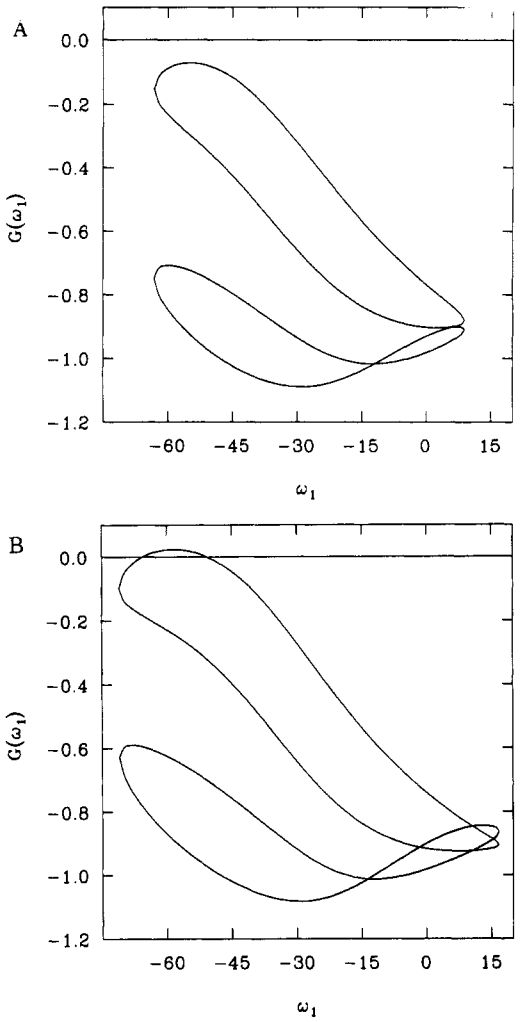


Figure 2. $G(\omega_1)$. (A) shows the values of $G(\omega_1)$ as a function of ω_1 when closing a Corey–Pauling α-helix, but where the new value for N–C_α–C bond angle has been changed from 109.5 degrees to 112.0 degrees. (B) displays how the function changes when the bond angles are modified as shown in Table I.

Table I
Modified Bond Angles for α-Helix Closure[a]

angle	C_α–C–N	C–N–C_α	N–C_α–C
Corey–Pauling standard	114.000	123.000	109.500
modified residue 1	113.416	122.812	110.304
modified residue 2	113.165	123.051	110.067
modified residue 3	113.047	123.098	110.404

[a] All angles are in degrees.

This example is useful for illustrating the CPU time requirements of the procedure. The CPU time necessary to apply the algorithm is 290 ms on a VAX 11/780 without a floating-point accelerator. By comparison, the closure using the normal peptide angles, so that no bond angle variations are necessary, requires 120 ms.

Tables II and III show the effect of the varying the three bond angles in an amino acid residue of an α-helix or β parallel pleated sheet. Here, we start with prototypical secondary structure backbones constructed with Pauling–Corey bond lengths and angles. The α-helices are constructed with torsions given above; the β-sheets have φ, ψ torsion angles of −119° and +113°. We then make a series

Table II
Backbone Torsions with Peptide Bond Angle Changes[a]

bond angles[b]				new torsion angles[d]					
C_α-C-N	C-N-C_α	N-C_α-C	best rms[c]	φ_1	ψ_1	φ_2	ψ_2	φ_3	ψ_3
				α-Helix					
109.0	118.0	104.5	0.39	-34.5	-95.9	-18.8	-94.9	0.5	-67.6
109.0	118.0	109.5	0.18	-40.3	-70.6	-37.0	-74.6	-25.6	-60.0
109.0	118.0	114.5	0.13	-46.6	-51.4	-49.8	-61.8	-45.0	-54.1
109.0	123.0	104.5	0.31	-32.9	-84.0	-28.6	-89.2	-8.6	-64.7
109.0	123.0	109.5	0.14	-41.6	-57.8	-47.2	-71.0	-34.7	-58.1
109.0	123.0	114.5	0.14	-44.4	-51.1	-49.9	-61.2	-50.4	-54.7
109.0	128.0	104.5	0.26	-32.7	-73.1	-36.8	-86.8	-16.1	-62.8
109.0	128.0	109.5	0.21	-43.8	-46.3	-56.0	-71.1	-40.7	-57.4
109.0	128.0	114.5	0.28	-48.2	-37.3	-61.5	-60.0	-59.0	-53.5
114.0	118.0	104.5	0.45	-41.2	-106.6	-7.5	-92.1	-3.5	-69.3
114.0	118.0	109.5	0.17	-46.4	-73.6	-33.4	-65.0	-34.9	-59.4
114.0	118.0	114.5	0.13	-47.5	-67.3	-33.9	-71.1	-30.3	-60.3
114.0	123.0	104.5	0.33	-38.6	-92.1	-20.9	-83.4	-13.7	-66.3
114.0	**123.0**	**109.5**	**0.00**	**-48.0**	**-57.0**	**-48.0**	**-57.0**	**-48.0**	**-57.0**
114.0	123.0	114.5	0.07	-46.8	-54.8	-43.4	-66.5	-41.7	-58.8
114.0	128.0	104.5	0.22	-37.5	-79.5	-31.8	-78.4	-22.4	-64.3
114.0	128.0	109.5	0.13	-50.0	-44.2	-59.4	-54.9	-55.2	-56.7
114.0	128.0	114.5	0.20	-44.7	-49.9	-43.3	-67.4	-51.3	-58.5
119.0	118.0	104.5	0.54	-49.3	-119.9	7.7	-89.0	-10.9	-69.5
119.0	118.0	109.5	0.19	-54.8	-68.2	-35.7	-49.8	-52.3	-55.7
119.0	118.0	114.5	0.17	-55.9	-63.2	-37.1	-67.0	-29.2	-61.9
119.0	123.0	104.5	0.40	-45.2	-101.0	-11.4	-78.2	-20.1	-66.9
119.0	123.0	109.5	0.12	-52.1	-62.9	-42.3	-51.2	-52.0	-57.5
119.0	123.0	114.5	0.13	-59.3	-36.4	-61.6	-47.8	-53.9	-57.9
119.0	128.0	104.5	0.26	-43.3	-85.4	-26.0	-70.0	-30.4	-64.7
119.0	128.0	109.5	0.09	-51.3	-54.7	-50.4	-49.7	-57.4	-58.4
119.0	128.0	114.5	0.17	-55.9	-40.3	-55.2	-50.8	-59.8	-58.8
				β-Sheet					
109.0	118.0	104.5	0.77	-102.6	-135.8	100.4	126.0	-132.0	137.3
109.0	118.0	109.5	0.13	-118.2	122.8	-129.0	123.7	-129.9	113.8
109.0	118.0	114.5	0.44	-105.6	124.1	114.1	122.7	-143.4	100.0
109.0	123.0	104.5	0.68	-90.8	133.4	93.5	123.6	-132.9	129.8
109.0	123.0	109.5	0.99	-104.0	-133.8	75.3	120.2	-111.9	146.2
109.0	123.0	114.5	0.30	-108.6	113.6	-115.4	115.5	-133.6	103.0
109.0	128.0	104.5	0.51	-71.2	-128.2	89.0	123.6	-143.8	116.3
109.0	128.0	109.5	0.43	-59.3	-116.1	81.7	113.3	-141.7	108.4
109.0	128.0	114.5	0.23	-110.5	107.7	-119.9	110.2	-126.7	104.4
114.0	118.0	104.5	0.88	-111.7	-124.8	93.7	122.7	-126.9	143.7
114.0	118.0	109.5	0.30	-111.3	126.0	-116.1	122.6	-135.1	105.5
114.0	118.0	114.5	0.57	-103.0	128.5	-107.8	117.3	-139.9	96.7
114.0	123.0	104.5	0.78	-101.9	-125.4	89.5	116.1	-122.3	137.0
114.0	**123.0**	**109.5**	**0.00**	**-119.0**	**113.0**	**-119.0**	**113.0**	**-119.0**	**113.0**
114.0	123.0	114.5	0.41	-105.8	114.3	-104.4	110.6	-131.2	99.2
114.0	128.0	104.5	0.69	-91.3	-125.0	84.4	113.4	-121.9	129.5
114.0	128.0	109.5	0.26	-124.9	109.6	-129.2	104.2	-105.5	117.9
114.0	128.0	114.5	0.27	-108.0	105.5	-106.2	103.6	-122.6	101.0
119.0	118.0	104.5	0.20	-119.9	123.7	-115.3	129.8	-131.1	113.9
119.0	118.0	109.5	0.49	-107.7	132.2	-110.3	120.4	-135.2	101.4
119.0	118.0	114.5	0.70	-101.8	134.7	-104.7	112.1	-135.0	95.3
119.0	123.0	104.5	0.87	-110.7	-113.7	84.0	114.5	-119.6	142.0
119.0	123.0	109.5	0.29	-112.3	115.2	-105.4	113.8	-125.5	105.4
119.0	123.0	114.5	0.57	-103.7	117.7	-96.9	107.2	-129.2	96.3
119.0	128.0	104.5	0.75	-100.2	-115.6	82.3	106.9	-114.6	133.9
119.0	128.0	109.5	0.12	-119.7	104.0	-109.7	103.4	-109.0	112.3
119.0	128.0	114.5	0.41	-106.0	105.0	-94.9	99.4	-119.8	98.4

[a] The boldfaced line shows the prototypical bond angles. [b] Bond angles in degrees for all three peptides being bridged. [c] Root-mean-square position deviation in Å from the standard structure for the backbone atoms constructed with the torsion angles listed. [d] New backbone torsion angles (in degrees) selected from all the closures found that yield the smallest root-mean-square torsion angle difference from the standard structure.

of calls to the modified chain-closure procedure where we have deleted three residues, and we pass different values for the bond angles to be used for closure; all other bond and torsion angles are kept fixed. The corresponding starting angles used in each of three residues are the same (e.g., the angle for C_α-C-N is the same in all three). The values for bond angles are generated systematically by adding or subtracting 5° from each angle or leaving them alone. Table II shows the torsion angles resulting from chain closure for a given combination of bond angles. Since the chain-closure procedure returns multiple solutions, only the structure that has the smallest root-mean-square (rms) torsion angle deviation from the original structure is shown. For most cases, closure was possible with the given combination of angles. For the cases when closure failed, Table III shows the modified bond angles.

There are a number of features to note in the two tables. First, Table III illustrates that closure of an α-helix is more sensitive to bond angle variations, as 11 of the possible combinations of perturbed bond angles require angle adjustment for closure, whereas only one β-sheet combination fails to close. Second, the table also shows that the amount

Table III
Peptide Angles for Which Closure Was Impossible without Modification[a]

angle[b]				new angles[d]								
C_α–C–N	C–N–C_α	N–C_α–C	best rms[c]	C_α–C–N	C–N–C_α	N–C_α–C	C_α–C–N	C–N–C_α	N–C_α–C	C_α–C–N	C–N–C_α	N–C_α–C
						α-Helix						
109.0	118.0	114.5	0.13	108.5	117.3	111.8	107.8	118.2	112.1	108.1	118.0	111.8
109.0	123.0	114.5	0.14	109.0	121.6	110.6	107.2	123.8	112.7	108.4	123.1	110.3
109.0	128.0	114.5	0.28	109.1	126.8	111.1	106.8	129.3	113.2	109.1	127.3	111.0
114.0	118.0	114.5	0.13	112.5	116.9	109.5	112.8	116.3	109.5	110.1	119.1	110.3
114.0	123.0	114.5	0.07	112.3	121.9	109.5	111.1	123.0	109.5	109.0	125.3	109.5
119.0	118.0	109.5	0.19	118.9	118.0	109.4	119.0	117.9	109.3	118.9	118.0	109.4
119.0	118.0	114.5	0.17	116.1	117.2	109.5	120.9	113.0	109.5	114.0	119.9	110.3
119.0	123.0	109.5	0.12	118.5	123.0	108.5	118.8	122.7	108.2	118.4	123.0	108.6
119.0	123.0	114.5	0.13	114.2	120.4	109.5	119.3	118.5	109.5	114.0	127.5	109.5
119.0	128.0	109.5	0.09	118.5	127.8	108.0	118.4	128.0	107.9	118.0	128.5	108.0
119.0	128.0	114.5	0.17	115.0	123.0	109.5	116.5	126.6	109.5	114.0	131.8	109.5
						β-Sheet						
109.0	128.0	104.5	0.51	108.6	128.4	104.5	109.2	128.0	104.1	109.4	127.5	105.7

[a] MAXDT was set to 5° for these calculations. [b] Input bond angles in degrees for all three peptides being bridged. [c] Root-mean-square deviation of closest matching conformation found from the standard structure for the backbone atoms. [d] Modified peptide bond angles listed in sequence for all three peptides.

of bond angle variation required to form a continuous chain is small. Finally, Table II shows that the variations in bond angles have a lesser effect on the α-helix structure than on the parallel β-sheet structure; i.e., the rms positional deviations are smaller and the torsion angles vary less. In addition, many values for the torsion angles in the β-sheets are not within the energetically allowable ranges as seen in a Ramachandran plot.[16] This indicates that there is just one class of peptide arrangements that give rise to an α-helix but there are several that form a β-sheet.

3.2. Cyclic Pentapeptide. We consider the determination of possible structures for the cyclic pentapeptide, Gly-Gly-Gly-Pro-Pro. The complete conformational space of this pentapeptide was computed by Gō and Scheraga.[17] Since the proline rings limit the φ angle in the backbone, there are eight free torsions. Gō and Scheraga varied the two proline ψ angles with a grid of 5° over the complete angle space and solved for the six glycine torsions. Figure 3A shows the range of ψ angles for which solutions were found and how many solutions were found at each such point. A total of 348 combinations of ψ_4 and ψ_5 could be closed. Figure 3B shows similar information when MAXDT is 5°, and we use the thorough method of searching for defined values of G. Here, 1565 grid points had solutions. It is evident that small bond angle variations have a striking effect on the set of possible solutions in this restricted system.

This example is useful for illustrating the effect of the search method when $G(\omega_1)$ is undefined. The thorough search took 125.5 min of CPU time on a VAX 11/780 without a floating-point accelerator. If the quick method of searching is used along with MAXDT of 5°, then 18.5 min of CPU time is necessary and the number of grid points with solutions drops slightly to 1507.

3.3. Proteins. Since the most important applications of the chain-closure procedure are likely to involve homology modeling, we examine its use for known proteins. For all three residue segments of a protein, we attempt to close the chain after deleting the backbone atoms in the three residues; side chains were not considered in this test. The bond lengths and angles used for closure are taken from the PARAM4 parameter set of CHARMM.[14] For prolines, we attempt closure using both *cis* and *trans* peptide torsion angles; only the *trans* geometry is used for the peptide torsion angle for all other amino acid residues. Three proteins are examined, flavodoxin, plastocyanin, and McPC 603. Flavodoxin is an electron-transport protein

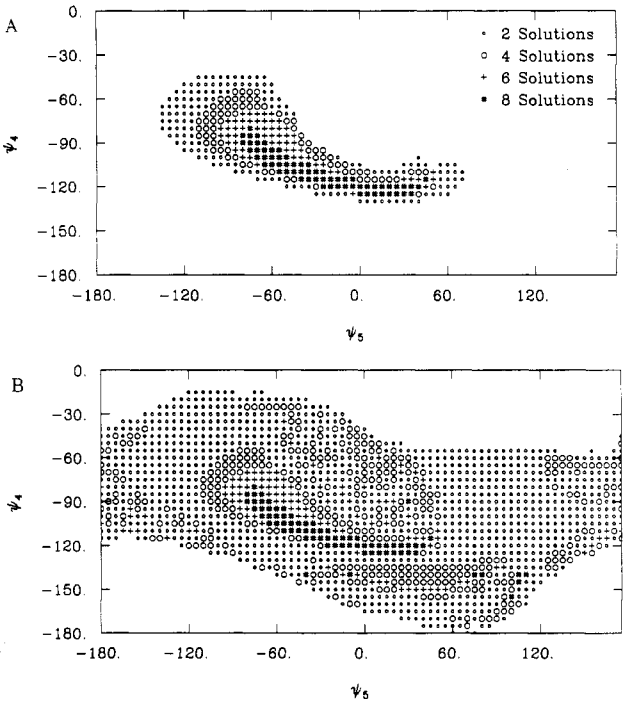


Figure 3. Pentapeptide chain closure. In these two plots we see how many conformations can be found for closing a Gly-Gly-Gly-Pro cyclic pentapeptide. The ψ torsions for the prolines were sampled every 5°. (A) shows the chain closures found when no angle variations are permitted. (B) shows the chain closures found when 5° bond angle variations are permitted and the thorough search is applied when $G(\omega_1)$ is not defined for any value of ω_1.

containing many α-helices.[18] Plastocytanin is a copper-binding plant electron-transport protein containing mostly β-sheets.[19] McPC 603 is a phosphorylcholine-binding mouse immunoglobulin.[20] It is also mostly β-sheet, but the structure we are using is an early refinement of the structure with irregular geometries.

Table IV shows how many segments were not closed by the Gō and Scheraga procedure and by our procedure allowing a 5° angle variation and the quick search method for G. In all cases, the bond angle variation led to a significant improvement in the number of closures found.

We can examine in further detail why flavodoxin and McPC 603 present more difficulties for the chain-closure algorithm than does plastocyanin. Figure 4 shows the residues in flavodoxin where chain closure was not possible

Table IV
Failure Rate of Chain Closure in Three Proteins[a]

protein	total gaps attempted	no angle variation		5° angle variation	
		no. missed	% of segments	no. missed	% of segments
Fv McPC 603	227	53	23.3	10	4.4
flavodoxin	134	51	38.1	1	0.7
plastocyanin	95	8	8.4	1	1.1

[a] Closure was attempted for all possible gaps generated by deleting backbone atoms in the native structures.

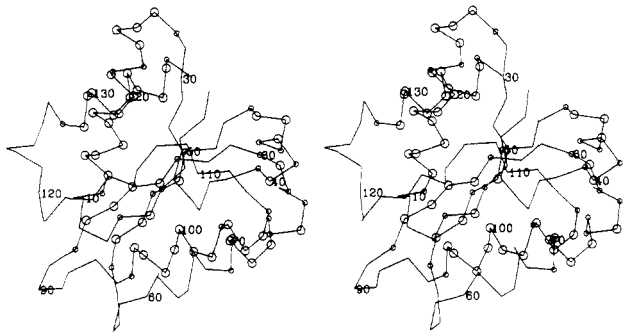


Figure 4. Flavodoxin unclosable residues. Superimposed on a C_α carbon plot of flavodoxin are the residues where chain closure failed when no bond angle variation was permitted. The large circles correspond to the central residue where closure failed. The smaller circles are for residues adjacent to the center. The absence of a circle indicates chain closure was successful.

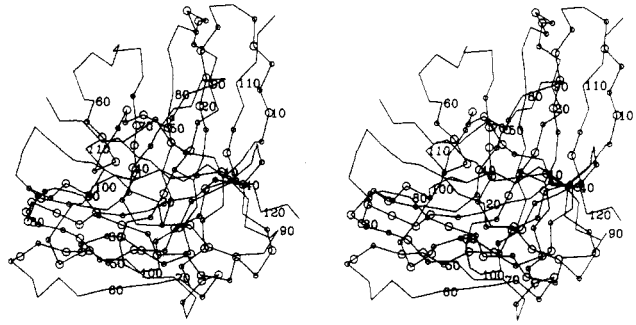


Figure 5. McPC 603 unclosable residues. Similar to Figure 4 except shown for McPC 603. The heavy chain is in the front with its sheets roughly horizontal.

when no bond angle variation was permitted. Most of the failures are localized in the α-helices, consistent with the results in section 3.1. In the case of McPC 603 (Figure 5), it is difficult to identify a structural feature that correlates with the failure of closure. A priori, one might expect the same behavior as in plastocyanin. However, a significant difference between the coordinates used for these two structures is that plastocyanin has more consistent geometries. Table V shows the energy of the backbone atoms (C_α, N, C) that result from the bonds, bond angles, and ω torsions in the peptide bond as computed by CHARMM.[14] Since the Gō and Scheraga algorithm is limited to generating conformations with ideal geometry; it is reasonable to surmise that it cannot reproduce the nonideal conformations found in McPC 603. If the distribution of backbone energy (see Figure 6) is compared with failures to close (Figure 5), some correlation is observed, but it is not striking.

4. Conclusions

A modification is introduced to the chain-closure algorithm of Gō and Scheraga so as to include limited bond angle variation in addition to the torsion angle search. It is clear that small-scale adjustments in bond angles during conformational searches are desirable, especially for α-helices and low-resolution structures. The cost of providing this enhancement is small, and when applied to conformational searches, will likely be recovered in the increased flexibility in choosing other degrees of freedom.

Although permitting bond angle variations improves the utility of the chain-closure procedure, it would be desirable to incorporate small deviations in the planarity of the peptide torsion angle along the lines developed by Madison.[21]

Table V
Backbone Strain Energy in Three Proteins

protein	E_{bond}[a]	E_{angle}[b]	E_ω[c]	E_{tot}[d]	$E_{tot}/N_{residues}$
Fv McPC 603	1477.0	1323.9	80.1	2881.0	12.260
flavodoxin	49.9	63.2	33.4	146.5	1.062
plastocyanin	49.5	40.7	30.0	120.2	1.214

[a] Sum of the bond energies (in kcal/mol) for the backbone atoms, (C_α, N, C). [b] Sum of the bond angle energies (in kcal/mol) for the backbone atoms. [c] Torsion energy (in kcal/mol) of the peptide linkage (the ω torsion angle only). [d] The sum of these three energies.

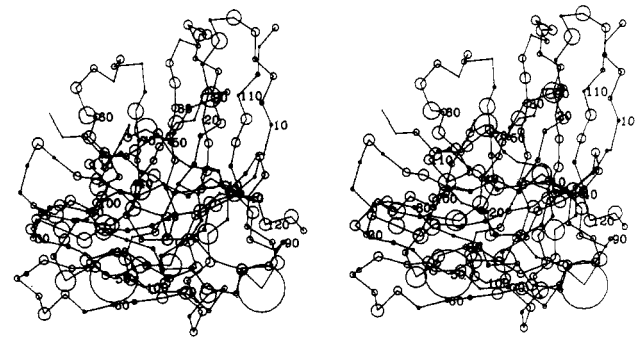


Figure 6. McPC 603 backbone energy. Superimposed on a C_α carbon plot of McPC 603 are the sums of the bond, bond angle, and peptide torsion angle energies for each residue as indicated by the size of the circles used to draw the C_α position. The sizes of the circles are scaled such that a residue energy of 50 kcal/mol yields a circle of radius 1 Å.

Acknowledgment. We thank Dr. David Davies for graciously providing coordinates before publication and Dr. Edgar Haber and John Newell for providing the resources of the Cardiac Computer Center, Massachusetts General Hospital. R.E.B. thanks Dr. Jiri Novotny for many helpful discussions. This work was supported in part by a grant from the National Institutes of Health.

Registry No. Cyclic pentapeptide, 28072-78-8.

References and Notes

(1) Warme, P. K.; Momany, F. A.; Rumball, S. V.; Tuttle, R. W.; Scheraga, H. A. *Biochemistry* **1974**, *13*, 768.
(2) Browne, W. J.; North, A. C. T.; Phillips, D. C.; Brew, K.; Vanaman, T. C.; Hill, R. L. *J. Mol. Biol.* **1969**, *42*, 65.
(3) Furie, B.; Bing, D. H.; Feldmann, R. J.; Robison, D. J.; Burnier, J. P.; Furie, B. C. *J. Biol. Chem.* **1982**, *257*, 3875.
(4) Gō, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 178.
(5) Schulz, G. E.; Schirmer, R. H. "Principles of Protein Structure"; Springer Verlag: New York, 1979.

(6) Ramachandran, G. N.; Kolaskar, A. S.; Ramakrishnan, C.; Sasisekharan, V. *Biochim. Biophys. Acta* **1974**, *359*, 7298.
(7) Kushick, J. N.; Karplus, *M. Macromolecules* **1981**, *14*, 325.
(8) van Gunsteren, W. F.; Karplus, M. *Nature (London)* **1981**, *293*, 677.
(9) Although the choice of ω_i for a general torsion angle conflicts with the more common use of ω as the peptide torsion angle, consistency is maintained with Gō and Scheraga, *ref* 4.
(10) The IUPAC convention for torsion angles (IUPAC–IUB Commission on Biochemical Nomenclature, *Biochemistry* **1970**, *9*, 3471.) shifted values for torsion angles by 180°. For example, under the old convention, a *cis* bond would have a torsion angle of 180°; now it is 0°. Gō and Scheraga used the old convention for their work, and the definition for R_i depends on this usage. However, we have used the IUPAC convention when listing any numerical values for torsion angles in this paper.
(11) Conte, S. D.; de Boor, C. "Elementary Numerical Analysis, An Algorithmic Approach"; McGraw-Hill: New York, 1972.
(12) Muller, D. E. *Mathematical Tables and Other Aids to Computation* **1956**, *10*, 208.
(13) Levitt, M.; Lifson, S. *J. Mol. Biol.* **1969**, *46*, 269.
(14) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus M. *J. Comput. Chem.* **1983**, *4*, 187.
(15) Pauling, L. "The Nature of the Chemical Bond"; Cornell University Press: Ithaca, NY, 1960.
(16) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95.
(17) Gō, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 188.
(18) Burnett, R. M.; Darling, G. D.; Kendall, D. S.; LeQuesne, M. E.; Mayhew, S. G.; Smith, W. W.; Ludwig, M. L. *J. Biol. Chem.* **1974**, *249*, 4383.
(19) Colman, P. M.; Freeman, H. C.; Guss, J. M.; Murata, M.; Norris, V. A.; Ramshaw, J. A. M.; Venkatappa, M. P. *Nature (London)* **1978**, *272*, 319.
(20) Segal, D. M.; Padlan, E. A.; Cohen, G. H.; Rudikoff, S.; Potter, M.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 4298.
(21) Madison, V. *Biopolymers* **1972**, *12*, 1837.

Orientational Order in the Nematic Phase of Low Molecular Weight Analogues of Nematic Polymers

R. Capasso, P. Iannelli, A. Roviello, and A. Sirigu*

Dipartimento di Chimica, Universitâ di Napoli, 80134 Napoli, Italy.
Received February 13, 1985

ABSTRACT: The nematic order parameter as a function of the temperature has been measured by X-ray diffraction methods for some compounds having the formula $CH_3(CH_2)_mCOOROOC(CH_2)_mCH_3$ (m = 3, 4) or $CH_3(CH_2)_4COOROOC(CH_2)_nCOOROOC(CH_2)_4CH_3$ with n = 7–10, R = $-C_6H_4-C(CH_3)=HC-C_6H_4-$. Dimeric compounds with even n afford order parameters higher than for any other of the examined compounds. Extrapolation to the isotropization temperature leads to a limiting value of ~0.58. No significant difference has been found between dimers with odd n and monomeric compounds. Order parameters extrapolated at the isotropization temperature approach the Maier–Saupe limit. The behavior of the dimeric compounds is coherent with the pattern defined for the thermodynamic data concerning their isotropization and with the behavior of structurally homologous linear polymers also showing nematic mesomorphism.

Introduction

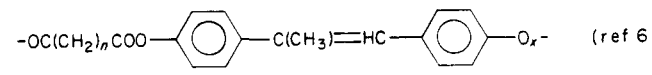
The study of mesophasic properties of semiflexible polymers has attracted interest on the problem concerning the reciprocal influence between contiguous rigid and flexible parts of the same molecule.

Recent results point out that orientational order of the rigid parts and conformational structure of the flexible portions play complementary roles. High levels of orientational order of the rigid parts in a nematic phase are more compatible with the extended conformations of the flexible parts.[1] The odd–even fluctuations of the thermodynamic quantities connected to the isotropization transition appear to be a consequence of this. Theoretical calculations[2] and some experimental investigations[3,4] are fairly convergent to the conclusion that for these semiflexible polymers the degree of order at the nematic–isotropic phase transition should be only moderately higher than for ordinary low molecular weight "monomeric" nematogens.

Low molecular weight nematogens whose molecules are structurally characterized by a "rigid–flexible–rigid" sequence have been examined as the most simple models for polymeric homologues.[5] Series of such "dimeric" model compounds having a variable number of carbon atoms in the flexible spacer reproduce the main features observed for the nematic phase of homologous series of polymers, namely, a relatively large molar isotropization enthalpy, particularly for the even-type molecules (i.e., those molecules having an even number of atoms in the backbone chain of the flexible spacer), a very much greater odd–even fluctuation of this quantity than observed for low molecular weight nematogens having a single rigid group in the molecule ("monomers"), and a greater thermal stability of the mesophase as compared with that exhibited by these monomeric homologues.

In a recent study Griffin, Sigaud, and Yoon,[3] on the basis of magnetic susceptibility measurements, have shown that the orientational order parameter at the nematic–isotropic phase transition for an even-type dimer has an intermediate value between those found for the homologous monomer and polymer.

We have measured by X-ray diffraction methods the degree of orientational order in molten fibers of semiflexible polymers of the formula



(ref 6)

In this paper the results of an investigation on the degree of order characterizing the nematic phase of dimeric and monomeric homologues are discussed.

0024-9297/85/2218-2773$01.50/0