(130) Poulos, T. L.; Price, P. A. *J. Biol. Chem.* **1971**, *246*, 4041.
(131) Robinson, G. W. *J. Biol. Chem.* **1970**, *245*, 4832.
(132) Donovan, J. W. *J. Biol. Chem.* **1969**, *244*, 1961.
(133) Herskovits, T. T.; Sorensen, M. *Biochemistry* **1968**, *7*, 2523.
(134) Herskovits, T. T.; Sorensen, M. *Biochemistry* **1968**, *7*, 2533.
(135) Wetlaufer, D. B. *Adv. Protein Chem.* **1962**, *17*, 303.
(136) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1.
(137) Auer, H. E.; Doty, P. *Biochemistry* **1966**, *5*, 1716.
(138) Peggion, E.; Verdini, A. S.; Cosani, A.; Scoffone, E. *Macromolecules* **1969**, *2*, 170.
(139) Arfmann, H. A.; Labitzke, R.; Wagner, K. G. *Biopolymers* **1975**, *14*, 1381.

Differential Geometry and Polymer Conformation. 2. Development of a Conformational Distance Function[1]

S. Rackovsky[2a] and H. A. Scheraga*[2b]

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853.
Received March 21, 1980

ABSTRACT: In paper 1, we described a differential-geometric representation of protein backbone structure which reflects the influence of both short- and medium-range interactions. In this paper, we continue our development of this representation, particularly of its application to the comparison of protein structures. Initially, two mathematical features of the representation are discussed. They are (a) the dependence of the curvature (κ_i) and torsion (τ_i) on the sign of p_ip_{i+1} (where p_i is the scalar product of two nonperpendicular unit vectors at the ith and $(i-1)$th α-carbon atoms, respectively), and the conditions under which p_ip_{i+1} changes sign, and (b) the existence (and size and shape) of a gap in the (κ, τ) plane, centered on the τ axis. Both of these features lead to discontinuities in κ and τ, analogous to those which exist in systems with periodic boundary conditions. Application of these results to the distribution of residues in a large protein sample in (κ, τ) space reveals the existence of a *continuum of bend structures*, ranging from α_R-like bends through flat structures to α_L-like bends. A corrected treatment of chain handedness is then given, thereby greatly increasing the utility of the differential-geometric method for determining chain handedness. A discussion is given of the inversion of the (κ, τ) representation to recover C^α coordinates, and results of such an inversion are presented for bovine pancreatic trypsin inhibitor. We then develop a function to represent the conformational distance between structures represented by any two points in the (κ, τ) plane. The utility of this distance function is demonstrated by comparison with a new superposition method for backbone segments composed of four C^α atoms which requires no computational optimization of superposition. The differential geometric comparison method is applied to two specific cases: (a) the comparison of reduced and oxidized cytochrome *c*, treated preliminarily in paper 1, and (b) the comparison of oxidized cytochrome *c* and glyceraldehyde phosphate dehydrogenase. The results demonstrate the validity and utility of the differential-geometric comparison method. This method is then used to illustrate the detection of conformational similarity between two portions of a protein molecule, the test case being two domains of ferredoxin. Finally, a general discussion of the differential-geometric representation is given. It is pointed out that this representation is complementary to other representations of protein structure in current use, in that it operates on a length scale (that of four C^α atoms) not conveniently treated by these others. It is therefore capable of revealing structural features that are not transparently evident in other representations. In addition, it is noted that comparison of the (κ, τ) and (ϕ, ψ) representations reveals a type of degeneracy in protein folding in which a given type of structure on one length scale can be attained in several ways through combinations of structures on smaller length scales.

I. Introduction

In the first paper in this series[3] (hereafter referred to as paper 1), a differential-geometric (DG) representation of protein backbone structure was presented. The fundamental unit of folding treated by this representation is a length of backbone consisting of four C^α atoms (three virtual bonds in the virtual-bond representation). The conformation of such a unit is described in terms of the curvature (κ_i), the torsion (τ_i) and the two virtual bond angles (θ_i, θ_{i+1}). It was demonstrated in paper 1 that the various types of characteristic backbone structure fall into regions of varying degrees of localization in the (κ, τ) plane, independent of the virtual bond angles. It was also pointed out that the DG representation provides a method for comparing the local folding of two backbone conformations at every point along the chain. The preliminary method suggested in paper 1 involved *separate* comparisons of κ and of τ of the two conformations at each residue.

In this paper, we improve the comparison procedure by constructing a function which represents the conformational distance between two points in the (κ, τ) plane and by using this function to measure the structural similarity of two four-C^α units. For this purpose, we first develop a number of mathematical properties of the DG representation. We also correct an error in the discussion of chain parity (handedness) in paper 1, thereby greatly extending the usefulness of the DG p_ip_{i+1} for determining local parity.

In section II, we summarize the DG notation briefly. In section III, we begin our treatment of mathematical questions. Two points will be covered: (1) the dependence of κ_i and τ_i on the sign (τ_i), p_ip_{i+1}, and the conditions under which p_ip_{i+1} changes sign, and (2) the existence, and the size and shape, of a discontinuity in the (κ, τ) plane, centered on the τ axis. In section IV, we give a corrected treatment of parity. In section V, we discuss the inversion of the (κ, τ) representation to recover C^α coordinates. Section VI deals with the construction of the distance function in (κ, τ) space and the comparison of conformations. Two methods are presented—a superposition method which allows the determination of actual atomic deviations between two four-C^α units without the necessity for computational optimization of superposition and the DG method. The former is used as a standard by which to judge the accuracy of the latter. The DG method is then used in section VII to discuss the comparisons of several

0024-9297/80/2213-1440$01.00/0 © 1980 American Chemical Society

proteins. We close with some general remarks about the differential-geometric approach to protein backbone structure and its relationship to other representations.

II. Summary of DG Notation

We briefly summarize the notation and concepts developed in paper 1, for convenient reference.

The local coordinate frame at residue i is defined by three unit vectors, $\hat{\mathbf{t}}_i$, $\hat{\mathbf{n}}_i$, and $\hat{\mathbf{b}}_i$, the tangent, normal, and binormal, respectively (the caret denotes a unit vector). These are defined by

$$\hat{\mathbf{t}}_i = \frac{\mathbf{v}_i + \mathbf{v}_{i+1}}{|\mathbf{v}_i + \mathbf{v}_{i+1}|} \tag{1}$$

$$\hat{\mathbf{n}}_i = p_i \hat{\mathbf{m}}_i = p_i \frac{\mathbf{v}_i - \mathbf{v}_{i+1}}{|\mathbf{v}_i - \mathbf{v}_{i+1}|} \tag{2}$$

$$p_i = \frac{\hat{\mathbf{m}}_i \cdot \hat{\mathbf{n}}_{i-1}}{|\hat{\mathbf{m}}_i \cdot \hat{\mathbf{n}}_{i+1}|} \equiv \text{sgn}\ (\hat{\mathbf{m}}_i \cdot \hat{\mathbf{n}}_{i-1}) \tag{3}$$

$$\hat{\mathbf{b}}_i = \hat{\mathbf{t}}_i \times \hat{\mathbf{n}}_i \tag{4}$$

where

$$\mathbf{v}_i = \mathbf{r}_i - \mathbf{r}_{i-1} \tag{5}$$

is a virtual-bond vector and $\mathbf{r}_i$ is the position vector of the ith C^α. The curvature (κ_i) and torsion (τ_i) at residue i are defined in terms of these unit vectors:

$$\kappa_i = \frac{q_i \chi_i}{|\mathbf{v}_{i+1}|} = \frac{q_i \cos^{-1}(\hat{\mathbf{t}}_{i+1} \cdot \hat{\mathbf{t}}_i)}{|\mathbf{v}_{i+1}|} \tag{6}$$

$$q_i = p_i\ \text{sgn}\ [(\hat{\mathbf{t}}_i \times \hat{\mathbf{t}}_{i+1}) \cdot \hat{\mathbf{b}}_{i+1}] \tag{7}$$

$$\tau_i = \frac{u_i \gamma_i}{|\mathbf{v}_{i+1}|} = \frac{u_i \cos^{-1}(\hat{\mathbf{b}}_i \cdot \hat{\mathbf{b}}_{i+1})}{|\mathbf{v}_{i+1}|} \tag{8}$$

$$u_i = \text{sgn}\ (\hat{\mathbf{b}}_{i+1} \cdot \hat{\mathbf{n}}_i) \tag{9}$$

It should be noted in connection with eq 2-4 that, by definition, the initial condition for the recursion relation involving $\hat{\mathbf{n}}_i$ is that $p_2 = +1$.

Because κ_i and τ_i are constructed from the unit vectors of the local coordinate frames at residues i and $i + 1$, they are determined entirely by the coordinates of the four C^α's, i.e., C^α_{i-1}–C^α_{i+2}.

The sign conventions embodied in q_i and u_i (which can assume the values ±1) were designed to ensure that there is a unique relationship between the (κ, τ) representation and the coordinate representation. It was shown in paper 1 that the sign q_i distinguishes between the two possible values of the virtual bond angle θ_{i+1} at C^α_{i+1}, for given χ_i, τ_i, and θ_i. The sign u_i essentially defines the sense of rotation about the central virtual bond $\mathbf{v}_{i+1}$ of the four-C^α unit. As stated in paper 1, the sign factor p_i guarantess that any planar chain segment has zero torsion.

The virtual bond angle at C^α_i is denoted by θ_i, and two auxiliary angles are defined by

$$\omega_i = \theta_i/2 \tag{10}$$

and

$$\lambda_i = \pi/2 - \omega_i \tag{11}$$

In the ensuing work, use will be made of the vectors

$$\mathbf{T}_i = \mathbf{v}_i + \mathbf{v}_{i+1} \tag{12}$$

$$\mathbf{M}_i = \mathbf{v}_i - \mathbf{v}_{i+1} \tag{13}$$

$$\mathbf{N}_i = p_i \mathbf{M}_i \tag{14}$$

$$\mathbf{B}_i = \mathbf{T}_i \times \mathbf{N}_i \tag{15}$$

It should be noted that $\mathbf{T}_i$, $\mathbf{M}_i$, $\mathbf{N}_i$, and $\mathbf{B}_i$ are parallel to $\hat{\mathbf{t}}_i$, $\hat{\mathbf{m}}_i$, $\hat{\mathbf{n}}_i$, and $\hat{\mathbf{b}}_i$, respectively.

III. Discontinuities in κ_i and τ_i

A. Discontinuity in Bend Region Due to Change in Sign of $p_i p_{i+1}$. In paper 1, it was observed that certain conformational displacements of one four-C^α unit relative to another seem to result in an inversion of the sign of κ_i and a change in both the sign and value of τ_i. We demonstrate in this section and in the relevant appendices that this is due to a change in the value of the product $p_i p_{i+1}$ (which assumes the values ±1).

The following theorems are proven in Appendix I.

Theorem 1. (a) The sign of κ_i and the sign and magnitude of τ_i depend on $p_i p_{i+1}$.

(b) As the conformation of the four-C^α unit passes through a point in conformational space where $p_i p_{i+1}$ changes sign, the pair (κ_i, τ_i) undergoes a discontinuous change of values

$$(\kappa_i, \tau_i) \rightarrow (\tilde{\kappa}_i, \tilde{\tau}_i) \tag{16}$$

where

$$\tilde{\kappa}_i = -\kappa_i \tag{17}$$

$$\tilde{\tau}_i = -(\pi/|\mathbf{v}_{i+1}| - |\tau_i|)\ \text{sgn}\ (\tau_i) \tag{18}$$

Theorem 2. The condition for a change of sign of $p_i p_{i+1}$ can be stated as follows. Let α be the angle between $\hat{\mathbf{t}}_i$ and $\mathbf{v}_{i+1}$, ϵ be the angle between $\hat{\mathbf{m}}_i$ and $\hat{\mathbf{t}}_i$, and ρ be the angle between $-\mathbf{v}_{i+2}$ and $\hat{\mathbf{m}}_i$. Then

$$p_i p_{i+1} = -1 \text{ if } \rho > \rho_0 \tag{19}$$

and

$$p_i p_{i+1} = +1 \text{ if } \rho < \rho_0 \tag{20}$$

where

$$\rho_0 = \frac{\pi}{2} - \sin^{-1}\left\{-\frac{|\mathbf{v}_{i+1}|}{|\mathbf{v}_{i+2}|} \cos(\alpha + \epsilon)\right\} \tag{21}$$

These theorems are true for arbitrary geometry (i.e., virtual bond lengths which vary from the standard value of 3.8 Å).

Theorem 1 provides a mathematical basis for the observation noted at the beginning of this section and in paper 1, and theorem 2 enables us to understand which types of conformations are affected by the discontinuity in (κ_i, τ_i) [i.e., in which regions of the (κ, τ) plane the discontinuity occurs].

The results of Appendix I suggest that these conformations should include bends and closely related structures. Figure 1 shows that this is indeed the case. In this figure we show the distribution of values of (κ_i, τ_i) of four-C^α units with $p_i p_{i+1} = +1$ for a sample of 22 proteins used in forthcoming work.[4] These values all fall in the region of low-torsion (i.e., small τ) bends identified in paper 1 and in adjacent areas. As we have just demonstrated (in theorem 1), this region is attainable from the corresponding region with $p_i p_{i+1} = -1$ by a slight deformation of the four-C^α unit. Application of eq 16, using values of (κ, τ) from Figure 1, indicates that the low-torsion bend region connects structurally at its $\tau < 0$ edge with the right-handed helical region (κ = 0.3–0.4; $\tau > 0.4$), and at its $\tau > 0$ edge with the left-handed helical region (κ = 0.3–0.4; $\tau < -0.4$). It is therefore seen that a *structural continuum* of bends exists, ranging from bends which resemble four-C^α segments of α_R helices, through essentially flat bends, to those which resemble four-C^α segments of α_L helices.

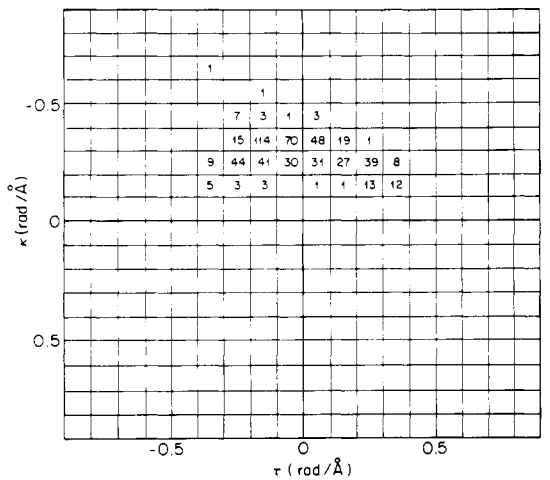

Figure 1. Distribution of residues with $p_ip_{i+1} = +1$ in a sample of 22 proteins. The numbers shown are number of occurrences per (0.1 × 0.1) rad/Å interval in the (κ, τ) plane.

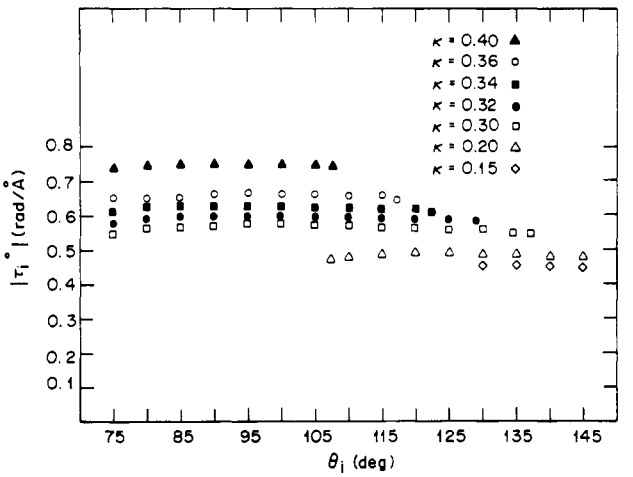

Figure 2. Dependence of $|\tau_i^0|$ (defined in text) on θ_i, for fixed positive κ_i. Corresponding curves for negative κ_i can be obtained from eq I-11, which gives the transformation from $p_ip_{i+1} = +1$ to −1.

Physically, the discontinuous jump of values of (κ_i, τ_i) discussed in theorem 1 is produced when helical four-C$^\alpha$ segments or closely related structures are flattened in the direction of U-shaped bend structures.

We are now able to understand the existence of some points in the plot of (κ, τ) for residues in the interior of helices (Figure 6 of paper 1) which fall at a distance from the main helical region. These are points where the local conformation is slightly distorted from a perfect helix, sufficient to invert p_ip_{i+1}.

Our next task is to determine quantitatively the path in (κ, τ) space that constitutes the boundary between the regions where $p_ip_{i+1} = -1$ and $p_ip_{i+1} = +1$. At any point on this boundary, three equations are satisfied simultaneously:

$$(\mathbf{v}_{i+1} - \mathbf{v}_{i+2})\cdot(\mathbf{v}_i - \mathbf{v}_{i+1}) = 0 \tag{22}$$

$$(\mathbf{v}_{i+1} + \mathbf{v}_{i+2})\cdot(\mathbf{v}_i + \mathbf{v}_{i+1}) = |\mathbf{v}_{i+1} + \mathbf{v}_{i+2}||\mathbf{v}_i + \mathbf{v}_{i+1}| \cos \chi_i \tag{23}$$

$$|(\mathbf{v}_{i+1} \times \mathbf{v}_i)\cdot(\mathbf{v}_{i+2} \times \mathbf{v}_{i+1})| = |\cos \gamma_i||\mathbf{v}_{i+1} \times \mathbf{v}_i||\mathbf{v}_{i+2} \times \mathbf{v}_{i+1}| \tag{24}$$

Equation 22 is simply the condition for inversion of p_ip_{i+1} (eq I-23). Equation 23 is the definition of χ_i (eq 6), and eq 24 arises from the definitions of γ_i, $\hat{\mathbf{b}}_i$, and $\hat{\mathbf{b}}_{i+1}$. In Appendix II we solve these equations for $|\tau_i^0|$, the value of $|\tau_i|$ at the point of inversion, as a function of χ_i (which determines κ_i) and θ_i, assuming standard geometry. As shown in Figure 2, $|\tau_i^0|$ is quite insensitive to θ_i, for fixed κ_i. In Figure 3, we show the inversion boundary (computed in Appendix II) superimposed on a map of the distribution in (κ, τ) space of all residues of the 22 proteins mentioned above. It is very clear that the distribution of residues follows the boundary; exceptions are due to deviations from standard geometry.

In summary, in this section we have (a) established the existence of the inversion between $p_ip_{i+1} = +1$ and −1 in certain regions of (κ, τ) space, (b) deduced the conformational condition that is satisfied at the point of inversion, viz., the condition on ρ with respect to ρ_0, and (c) examined the effect of the inversion on values of (κ, τ) of four-C$^\alpha$ units which undergo the inversion. These properties are required to define the distance between two points in the (κ, τ) plane and will be used in the construction of the distance function to be established in section VI. In section IIIB, we shall examine another property of (κ, τ) space.

B. Discontinuity Centered at the τ Axis Due to Change in Sign of q_i. Another discontinuity which can

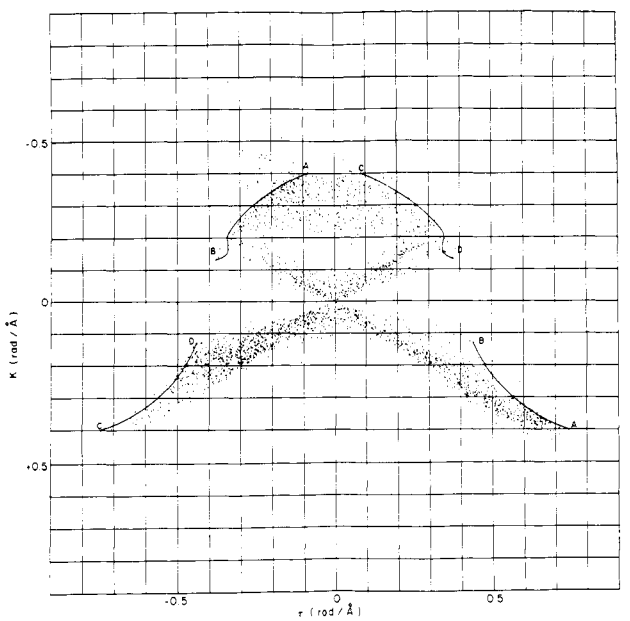

Figure 3. Inversion boundaries for p_ip_{i+1} superimposed on the actual distribution of residues (the dots) in the sample of 22 proteins of Figure 1 in the (κ, τ) plane. The two curves AB are equivalent conformationally, as are the two curves CD; these curves were computed by the method of Appendix II.

arise in (κ_i, τ_i) is due to a change in the sign of q_i unconnected with any change in the sign of p_ip_{i+1}. Mathematically, this means that the factor multiplying p_ip_{i+1} in eq I-15 can change sign independent of p_ip_{i+1}. Physically, this effect can be understood readily by remembering that q_i distinguishes between the two possible values of θ_{i+1} for given values of θ_i, τ_i, and χ_i (as pointed out in section II and, in greater detail, in paper 1). For certain values of τ_i and χ_i these two possible values of θ_{i+1} lead to conformations which are almost identical. This point is discussed in greater detail in Appendix III, in which the following theorems are proven.

Theorem 3. Let P_1 and P_2 be points (κ_1, τ_1) and (κ_2, τ_2) in the (κ, τ) plane such that (a) $p_ip_{i+1} = -1$ for both P_1 and P_2 and (b) $\kappa_1 < 0$, $\kappa_2 > 0$. Then any path in the (κ, τ) plane representing the transformation of a four-C$^\alpha$ unit characterized by (κ_1, τ_1) to a conformation characterized by (κ_2, τ_2) will involve a discontinuous jump across a gap in the (κ, τ) plane which includes the τ axis. (It should be re-

Table I
Corrected Parity Relationships[a] at Residue i

$\chi_i > \lambda_i$			$\chi_i < \lambda_i$		
	u_i			u_i	
q_i	+1	−1	q_i	+1	−1
+1	R	L	+1	R	L
−1	L	R	−1	R	L

[a] As shown in paper 1, positive and negative values of q_i correspond to the smaller and larger values, respectively, of θ_{i+1}. With this information and the sign of the torsion (i.e., of u_i), and the use of a molecular model, it is possible to deduce the direction of twist of the chain and, hence, the entries in this table.

membered that a point with $p_i p_{i+1} = +1$ and $\kappa < 0$ is equivalent to a point with $p_i p_{i+1} = -1$ and $\tilde{\kappa} > 0$, by eq 16.)

Theorem 4. The gap in the (κ, τ) plane is butterfly (or bowtie) shaped, with dimensions which are a strong function of θ_i.

Details of the θ_i dependence of the shape of the gap, and a method for approximating this shape analytically, are also discussed in Appendix III.

In this and the preceding subsection, we have established the existence and investigated the properties of two types of discontinuity in the (κ, τ) plane; one is caused by a discontinuity in $p_i p_{i+1}$, and the other is a discontinuity about the τ axis in the (κ, τ) plane. Before proceeding (in section VI) to the detailed construction of the distance function in the (κ, τ) plane, we digress slightly and discuss a number of ancillary points connected with the DG representation.

IV. Parity in (κ, τ) Space

In this section we correct an error in the division of the (κ, τ) plane into regions of definite parity (right or left handedness) as set out in Table I of paper 1. The corrected results are embodied in Table I of this paper. The corrected parity map in the (κ, τ) plane is shown in Figure 4. Application of eq 16 transforms all the points representing structures for which $p_i p_{i+1} = +1$ (whose distribution is shown in Figure 1) to new points with $\kappa_i > 0$. Comparison of Figures 1 and 3 shows that, once this transformation is carried out, relatively few points fall into the indeterminate region. Because the indeterminate region now has low population, the DG method is a *much more powerful method* for determining local parity than we erroneously concluded in paper 1.

V. Inversion of the DG Representation

As was noted briefly in footnote 16 of paper 1, it is possible (in standard geometry, where all virtual bond lengths are the same) to make the inverse transformation from $\{\kappa_i, \tau_i\}$ to C^α coordinates (using eq 13 of paper 1), given the initial virtual bond angle θ_2. We outline briefly how this is done and present representative results for bovine pancreatic trypsin inhibitor (BPTI). If the X-ray coordinates of the protein under consideration depart significantly from standard geometry, then eq 13 of paper 1 cannot be used to determine the virtual bond angles, and it is necessary to use observed values of the virtual bond lengths and bond angles. This does not complicate the inversion process significantly.

The method, which we illustrate here for standard geometry, involves the use of successive matrix transformations, familiar from earlier polymer work.[5] In this case, however, the local coordinate frame at each C^α_i is the $(\hat{\mathbf{t}}_i, \hat{\mathbf{n}}_i, \hat{\mathbf{b}}_i)$ frame rather than the traditional one of Flory.[5] A vector $\mathbf{r}^{(i)}$ which is expressed in the local coordinate frame

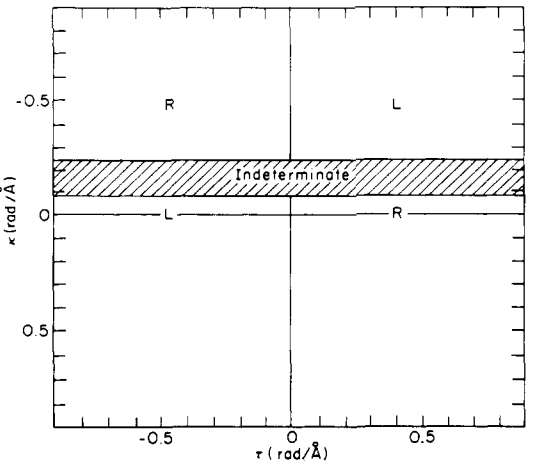

Figure 4. Corrected parity (handedness) regions in the (κ, τ) plane. In the region marked "indeterminate", parity cannot be determined solely from a knowledge of (κ, τ).

at C^α_i can be expressed in the local coordinate frame of C^α_{i-1} by the transformation

$$\mathbf{r}^{(i-1)} = \mathbf{A}^{(i-1,i)}\mathbf{r}^{(i)} + \mathbf{v}_i^{(i-1)} \tag{25}$$

where $\mathbf{v}_i^{(i-1)}$ is the virtual bond vector from C^α_{i-1} to C^α_i, expressed in the local frame at $i-1$, and the matrix is given by

$$\mathbf{A}^{(i-1,i)} = \begin{bmatrix} \hat{\mathbf{t}}_i\cdot\hat{\mathbf{t}}_{i-1} & \hat{\mathbf{n}}_i\cdot\hat{\mathbf{t}}_{i-1} & \hat{\mathbf{b}}_i\cdot\hat{\mathbf{t}}_{i-1} \\ \hat{\mathbf{t}}_i\cdot\hat{\mathbf{n}}_{i-1} & \hat{\mathbf{n}}_i\cdot\hat{\mathbf{n}}_{i-1} & \hat{\mathbf{b}}_i\cdot\hat{\mathbf{n}}_{i-1} \\ \hat{\mathbf{t}}_i\cdot\hat{\mathbf{b}}_{i-1} & \hat{\mathbf{n}}_i\cdot\hat{\mathbf{b}}_{i-1} & \hat{\mathbf{b}}_i\cdot\hat{\mathbf{b}}_{i-1} \end{bmatrix} \tag{26}$$

We define the matrix

$$\mathbf{T}^{(2,m)} = \mathbf{A}^{(2,3)}\mathbf{A}^{(3,4)}\cdots\mathbf{A}^{(m-1,m)} \tag{27}$$

Then iteration of eq 25 gives the vector $\mathbf{r}^{(i)}$ in the coordinate frame of C^α_2 (the first coordinate frame along the chain in the DG method[3]) as

$$\mathbf{r}^{(2)} = \mathbf{T}^{(2,i)}\mathbf{r}^{(i)} + \sum_{q=2}^{i-1} \mathbf{T}^{(2,q)}\mathbf{v}_{q+1}{}^{(q)} \tag{28}$$

It remains only to show how the vectors of interest can be expressed in the local coordinate frame. We express $-\mathbf{v}_i$, $\mathbf{v}_{i+1}$, and $\mathbf{v}_{i+2}$ in the coordinate frame at C^α_i. This problem is addressed for arbitrary geometry in section VIA. Here we treat only the case of standard geometry ($|\mathbf{v}_i| = l = 3.8$ Å).

It is clear by construction that the components of the virtual bond vectors along $\hat{\mathbf{t}}_i$, $\hat{\mathbf{n}}_i$, and $\hat{\mathbf{b}}_i$ are

$$(\mathbf{v}_{i+1})_t = l \sin(\theta_i/2) \tag{29}$$

$$(\mathbf{v}_{i+1})_n = -p_i l \cos(\theta_i/2) \tag{30}$$

$$(\mathbf{v}_{i+1})_b = 0 \tag{31}$$

and

$$(-\mathbf{v}_i)_t = -l \sin(\theta_i/2) \tag{32}$$

$$(-\mathbf{v}_i)_n = -p_i l \cos(\theta_i/2) \tag{33}$$

$$(-\mathbf{v}_i)_b = 0 \tag{34}$$

The three components of $\mathbf{v}_{i+2}$ can be determined by solving the following set of equations:

$$|\mathbf{v}_{i+2}| = l \tag{35}$$

$$\mathbf{v}_{i+2}\cdot\mathbf{v}_{i+1} = -l^2 \cos\theta_{i+1} \tag{36}$$

$$\hat{\mathbf{t}}_i\cdot(\mathbf{v}_{i+1} + \mathbf{v}_{i+2}) = |\mathbf{v}_{i+1} + \mathbf{v}_{i+2}| \cos\chi_i \tag{37}$$

Equation 35 is simply the statement of the virtual bond

Table II
Comparison of Observed Intramolecular Distances between C^α Atoms with Results of Inversion of $\{\kappa_i, \tau_i\}$ for BPTI

(i, j)	R_{ij}^{calcd}, Å	R_{ij}^{obsd}, Å
(1, 58)[a]	8.38	8.23
(1, 29)	12.90	12.01
(29, 58)	12.34	12.31

[a] End-to-end distance.

length. Equation 36 is the statement of the virtual bond angle, and eq 37 is the statement of the curvature κ_i (eq 1 and 6). The result is

$$(\mathbf{v}_{i+2})_t = l \sin(\theta_i/2)(2\cos\chi_i - 1) \quad (38)$$

$$(\mathbf{v}_{i+2})_n = -p_i l\{[\tfrac{1}{2}(1 - \cos\theta_i) - \cos\theta_{i+1} - 2\cos\chi_i \sin(\theta_i/2)\sin(\theta_{i+1}/2)]/\cos(\theta_i/2)\} \quad (39)$$

$$(\mathbf{v}_{i+2})_b = \alpha_b[l^2 - (\mathbf{v}_{i+2})_t - (\mathbf{v}_{i+2})_n]^{1/2} \quad (40)$$

and

$$\alpha_b = p_{i+1}u_i \quad (41)$$

Equation 41 is derived in Appendix VI. Thus, eq 38–41, together with eq 29–34, provide the components of $\mathbf{v}_i$, $\mathbf{v}_{i+1}$, and $\mathbf{v}_{i+2}$ in the $(\hat{\mathbf{t}}_i, \hat{\mathbf{n}}_i, \hat{\mathbf{b}}_i)$ frame at C^α_i, and eq 28 transforms them to the local frame at C^α_2. The Cartesian coordinates of the C^α atoms are thereby determined.

We now apply this method to the inversion of $\{\kappa_i, \tau_i\}$ for BPTI, using the experimental value of θ_2 and eq 13 of paper 1 to determine the virtual bond angles (the θ_i's) and hence the components of the $\mathbf{v}_i$'s. The set of $\{\kappa_i, \tau_i\}$ was determined directly from the X-ray coordinates of BPTI, and the inverse transformation is used here to recover the Cartesian coordinates from $\{\kappa_i, \tau_i\}$. The use of eq 13 of paper 1 is justified here by the fact that the X-ray structure of BPTI is a highly refined one,[6] which conforms closely to standard geometry. Selected intramolecular $(C^\alpha_i \cdots C^\alpha_j)$ distances are shown in Table II, as calculated from the actual X-ray coordinates and from $\{\kappa_i, \tau_i\}$, using $|\mathbf{v}_i| = 3.8$ Å. It is readily seen that the two sets of distances correspond within the accuracy of the X-ray coordinates. The small remaining deviations are due to residual divergences of the virtual bond lengths in the X-ray coordinates of BPTI from the standard value of 3.8 Å. As we remarked previously, it is to be expected that less satisfactory correspondence will result for X-ray structures which are less highly refined than the BPTI structure if the above method based on standard geometry is used. In such cases, however, actual virtual bond lengths and angles should be used, and equally good results should obtain. It is not clear, though, why one would want to perform such an inversion on structures that deviate significantly from standard geometry.

VI. Comparison of Conformations

We have now laid the groundwork for the actual construction of the distance function in (κ, τ) space which will be used for the comparison of conformations. In this section we outline two methods for performing such a comparison. One will be the comparison in (κ, τ) space. The other, which we call the standard superposition method (SSM), will provide actual deviations of individual atoms (in angstroms) between the two four-C^α units being compared when they are in a standard relative orientation (*without* computational optimization of superposition). The results of the SSM will be used as a standard by which to judge the accuracy of the differential-geometric comparison method. We show, by a comparison of the results obtained by the two methods, that the comparison method in (κ, τ) space is indeed an accurate reflection of the conformational differences between two structures. We begin with the standard superposition method.

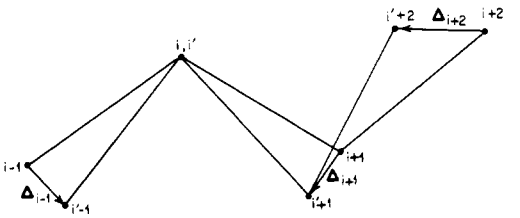

Figure 5. Relative orientation of the two four-C^α units that are compared in the standard superposition method.

A. Standard Superposition Method. The comparison of protein conformations by superposition methods is common practice.[7-9] The method in general use is to rotate one of the two structures relative to the other until optimal superposition of equivalent atoms is obtained. The resulting fit is then characterized by the root-mean-square (rms) deviation between corresponding atoms.

We now present a (computational-optimization)-free superposition method for the comparison of the individual four-C^α units making up the two structures of interest, rather than the comparison of larger blocks of backbone. We adopt this approach because we are interested in the detailed comparison of *local* conformations, rather than in a more coarse-grained, larger scale approach.

We suggest that, for the comparison of four-C^α units, it is possible to extract more information about differences in folding by using an alternative relative orientation of the two units to that obtained by optimized-superposition methods; our alternative method does not require any computational optimization. This relative orientation method is illustrated in Figure 5. The two four-C^α units to be compared (designated by primed and unprimed symbols and parameters) are so arranged that C^α_i and $C^\alpha_{i'}$ are superimposed and the planes determined by $(C^\alpha_{i-1}, C^\alpha_i, C^\alpha_{i+1})$ and $(C^\alpha_{i'-1}, C^\alpha_{i'}, C^\alpha_{i'+1})$ coincide. The orientation is then fixed by requiring that the tangent vectors at C^α_i and $C^\alpha_{i'}$ be parallel

$$\hat{\mathbf{t}}_i \cdot \hat{\mathbf{t}}_{i'} = +1 \quad (42)$$

Then the vectors

$$\Delta_j = \mathbf{r}_{j'} - \mathbf{r}_j \quad (43)$$

give the deviation of one conformation from the other on an atom-by-atom basis. It should be noted that, by construction

$$\Delta_i = 0 \quad (44)$$

The standard superposition (SS) that we have just described is essentially chosen to optimize automatically the superposition of the *three* C^α *units* $(C^\alpha_{i-1}, C^\alpha_i, C^\alpha_{i+1})$ and $(C^\alpha_{i'-1}, C^\alpha_{i'}, C^\alpha_{i'+1})$. Therefore, the vector Δ_{i+2} gives a precise measure of the difference in the manner in which the backbone continues in the two units under study. We show in Appendix IV that the Δ_j's can be calculated from $\{\kappa_i, \tau_i\}$ together with the virtual bond lengths and virtual bond angles. The derivation is carried out for *arbitrary* geometry, in which case it is necessary to use observed values of the virtual bond length and virtual bond angle. In the case of standard geometry, the virutal bond length is fixed, and it is sufficient to know the initial virtual bond angle θ_2, as demonstrated in paper 1.

We now turn to the distance function in (κ, τ) space.

B. The (κ, τ) Space Distance Function. The construction of this function is based on the mathematical

features of the (κ, τ) representation discussed in earlier sections of this paper. We consider two four-C^α units, characterized by DG parameters (κ, τ) and (κ', τ') for their respective C^α_i's. We denote the distance between the conformations of these two units [i.e., the distance between (κ, τ) and (κ', τ') in the (κ, τ) plane] by ρ_d. The procedure for calculating ρ_d follows.

1. Inversion of p_ip_{i+1}. In order to eliminate the effect of a discontinuity in p_ip_{i+1}, we must arrange that p_ip_{i+1} has the same sign for both units being compared. If either four-C^α unit has $p_ip_{i+1} = +1$, we perform the transformation to $p_ip_{i+1} = -1$ as given in eq I-17

$$(\kappa_i, \tau_i) \rightarrow (-\kappa_i, -[\pi/|\mathbf{v}_{i+1}| - |\tau_i|] \text{ sgn } \{\tau_i\}) \quad (45)$$

Henceforth, we shall assume that this transformation has been carried out where necessary and that both (κ, τ) and (κ', τ') have the same value of p_ip_{i+1}, which we choose here as -1.

2. Calculation of ρ_d. Case 1. sgn (κ) = sgn (κ'). In this case, both points are on the same side of the discontinuity centered on the τ axis. It should be noted that, once the transformation of eq 45 has been carried out, there is a periodic boundary condition at the lines

$$|\tau| = \pi/|\mathbf{v}_{i+1}| = 0.8267 \quad \text{(ideal geometry)} \quad (46)$$

corresponding to the condition that

$$|\gamma_i| = \pi \quad (47)$$

i.e., values of γ_i (which determine τ_i; eq 8) of $\pm\pi$ are equivalent. This arises precisely because the rotation of $\hat{\mathbf{b}}_{i+1}$ relative to $\hat{\mathbf{b}}_i$ by $+\pi$ is equivalent to a rotation by $-\pi$, with the same value of κ. This periodic boundary condition is easily taken into account as follows. We define ρ_d in this case by

$$\rho_d = \text{Min } [\rho_d', \rho_d''] \quad (48)$$

where

$$\rho_d' = [(\kappa - \kappa')^2 + (\tau - \tau')^2]^{1/2} \quad (49)$$

and

$$\rho_d'' = [(\kappa - \kappa')^2 + (\tau - \tau' + 2u'\pi/|\mathbf{v}_{i+1}'|)^2]^{1/2} \quad (50)$$

Here $|\mathbf{v}_{i+1}'|$ is the magnitude of $|\mathbf{v}_{i+1}|$ in the four-C^α unit characterized by (κ', τ') and

$$u' = \text{sgn } (\tau') \quad (51)$$

Case 2. sgn (κ) = −sgn (κ'). We assume, without loss of generality, that $\kappa < 0$ and $\kappa' > 0$.

In this case, the two points in the (κ, τ) plane are on opposite sides of the discontinuity around the τ axis. It should also be remembered that the shape of the discontinuity is a strong function of θ_i, the virtual bond angle at C^α_i (see eq III-7 and the discussion preceding it). In general, the two four-C^α units to be compared will have different values of θ_i, in addition to different values of (κ, τ). These factors must be taken into account in calculating ρ_d.

One can imagine many ways of transforming a four-C^α unit with DG parameters (κ, τ) and virtual bond angle θ [which we will denote by $(\kappa', \tau'; \theta)$] to the conformation represented by $(\kappa', \tau'; \theta')$. Each pathway between the two conformations is of the type shown in Figure 6; the size of the discontinuity, and therefore the length of the path, depends on the value of θ that the four-C^α unit has when κ changes sign. We shall measure the distance along two alternate paths: that on which the shape of the discontinuity corresponds to θ and that on which it corresponds to θ'. The distance between the two points will be defined

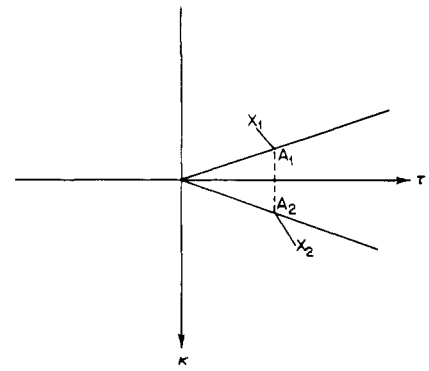

Figure 6. Pathway in the (κ, τ) plane between two points, x_1 and x_2, both with $p_ip_{i+1} = -1$ and with opposite signs for κ_i. Since points A_1 and A_2 are conformationally the same, the pathway is from x_1 to A_1 and from A_2 to x_2.

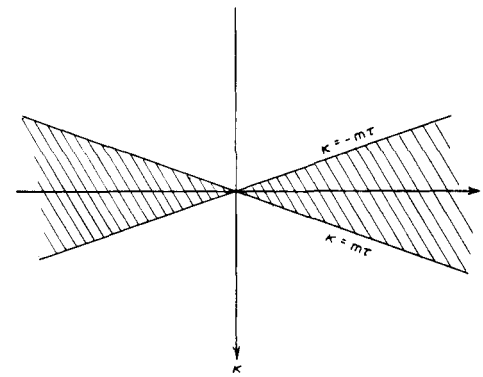

Figure 7. Discontinuity region in the (κ, τ) plane, as modeled in the (κ, τ) distance function.

as the minimum of these two. The forbidden region of the (κ, τ) plane will be modeled by a butterfly-shaped region bounded by the lines

$$\kappa = \pm m\tau \quad (52)$$

where m, for a given θ, is obtained from eq III-7 (see Figure 7). Thus, we denote the fact that the size of the gap (and hence the value of ρ_d) depends on θ' by writing

$$\rho_d = \text{Min } [\rho_d(\theta), \rho_d(\theta')] \quad (53)$$

where

$$\rho_d(\theta^k) = \min_{\tau''} \{[(\kappa + m(\theta^k)|\tau''|)^2 + (\tau - \tau'')^2]^{1/2} + [(m(\theta^k)|\tau''| - \kappa')^2 + (\tau'' - \tau')^2]^{1/2}\} \quad (54)$$

is the minimum distance between the two points, for $\theta = \theta^k$. Equation 54 denotes the fact that, with a value of θ to fix the size of the gap picked, the distance between the two points (κ, τ) and (κ', τ') must still be minimized by picking the value of τ at which the path connecting the two points crosses the gap; this optimal value of τ is denoted by τ''. In terms of Figure 6, this corresponds to picking the points A_1 and A_2 such that the length of the path connecting x_1 and x_2 is minimal.

VII. Results and Discussion

In this section we demonstrate the utility of the distance function in (κ, τ) space for the comparison of local folding in polypeptides. We do so by demonstrating a correlation between the distance between two conformations as measured in (κ, τ) space and that measured by the SS method. We then show how the DG comparison method can be used to uncover structural resemblances in protein fragments. We close this section with some general com-

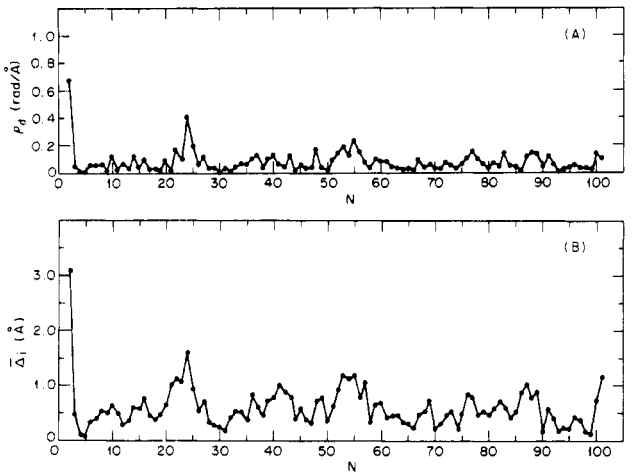

Figure 8. Comparison of CYTI and CYTR: (A) using DG comparison method; (B) using SS comparison method.

ments on the DG representation.

As in paper 1, we again consider the structures of the oxidized (inner) and reduced forms of tuna cytochrome *c*, refined independently of each other by Dickerson and colleagues.[10,11] As in paper 1, the two forms of the molecule will be abbreviated as CYTI and CYTR, respectively. It was shown by Dickerson and colleagues[10,11] that the structures of the oxidized and reduced forms are qualitatively similar. The preliminary comparison given in paper 1, which involved the *separate* comparison of κ and τ values, indicated that differences remain between the two structures. We now see, using the more sophisticated distance function formulated above, precisely where these differences lie and what their nature is.

In Figure 8A, we show a plot of ρ_d as a function of residue number. [It should be remembered that, in an N-residue chain, values of (κ_i, τ_i) exist for $i = 2, 3, ..., N - 2$.] This plot is juxtaposed (Figure 8B) with a corresponding chain plot of $\bar{\Delta}_i$, defined as

$$\bar{\Delta}_i = \tfrac{1}{3}(|\Delta_{i-1}| + |\Delta_{i+1}| + |\Delta_{i+2}|) \qquad (55)$$

The parameter $\bar{\Delta}_i$ is simply the average of the three Δ's described in section VIA, resulting from the standard superposition of the four-C^α units occurring at the ith positions of the two structures. Clearly, the magnitude of $\bar{\Delta}_i$ is proportional to the structural difference between the two four-C^α units. It therefore provides a useful standard by which to judge the success of the DG comparison method. As we shall show in greater detail below, the DG representation goes far beyond the SS comparison in the useful structural information that it conveys.

It can be seen in Figure 8A that there are substantial stretches of the two molecules which are quite similar structurally; these are characterized by values of $\rho_d \lesssim 0.1$ rad/Å. Experience suggests that this condition indicates similarity within the limits of accuracy of the X-ray coordinates. There are substantial deviations at several locations in the molecule. The large difference in folding at the N terminus between the two chains (commonly due to poor electron density at the chain ends) is manifested very clearly. There is a very substantial divergence of local folding between the two structures at residues 22–25. Smaller divergences also occur, e.g., at residues 48, 51–56, 77, 83, 87–89, and at the C terminus (residues 100, 101).

Comparison with Figure 8B shows corresponding peaks in $\bar{\Delta}_i$. The large peaks in ρ_d (>0.1 rad/Å) correspond precisely with peaks in $\bar{\Delta}_i$ characterized by $\bar{\Delta}_i > 1.0$ Å. Smaller peaks in ρ_d are associated with smaller peaks in $\bar{\Delta}_i$. We also point out the almost complete correspondence between trends in the two plots. Increases in ρ_d are generally matched by increases in $\bar{\Delta}_i$; decreases in ρ_d are generally matched by decreases in $\bar{\Delta}_i$.

Table III
Analysis of Peak in ρ_d in Comparison of CYTI and CYTR (See Text)

residue no.	(κ, τ), rad/Å	
	CYTI	CYTR
22	(0.35, −0.71)	(0.36, 0.78)
23	(−0.15, 0.16)	(−0.07, 0.11)
24	(0.25, −0.71)	(0.17, −0.31)
25	(0.08, −0.14)	(0.15, −0.31)

There are, however, exceptions to this rule. The C terminus of the molecule, at residues 99–101, should be noted, for example. The plot of ρ_d shows an increase at residue 100, followed by a decrease at residue 101; $\bar{\Delta}_i$ increases monotonically from 99–101. The basis for these differences in detail lies in the different philosophies underlying the two comparison methods. In the SS method, actual atomic deviations are calculated. This is not the case in the DG method. It should be remembered (as was emphasized in paper 1) that four-C^α units with different values of θ_i (and therefore nonzero values of $\bar{\Delta}_i$) can have the same values of (κ, τ). The DG parameters (without θ_i) describe a folding *mode*, rather than actual coordinates. Nonzero values of ρ_d correspond to separation in an abstract space in which the points represent different *types* of folding. An analogy which sheds light on this concept is provided by the classical vibrating string with fixed ends. It is well-known that such a string can vibrate in any one of an infinite set of normal modes, characterized by 0, 1, ... nodes. Specification of the mode in which the string is vibrating is not sufficient to give the coordinates of every point on the string at a given time. One additional parameter must be provided—the amplitude of the vibration. Nevertheless, two strings vibrating in the same mode, but with different amplitudes, can be said to be "folded" similarly. In the DG representation (κ, τ) plays a role analogous to that of the vibrational mode of the string, and θ_i is analogous to the amplitude. The close general correspondence between the chain plots of ρ_d and $\bar{\Delta}_i$ indicates that, as a practical matter, the differences in *type* of folding are usually accompained by relative atomic displacements. The existence of exceptions of the type noted above shows that differences can occur.

In order to indicate the power of the DG method to reveal details of local folding, it is worthwhile to analyze one of the peaks in the ρ_d chain plot. We consider the large peak at residues 22–25. In Table III, we show the (κ, τ) values of this segment for CYTI and CYTR, after the transformation of eq I-17 has been carried out. The values of (κ, τ) for residue 22 of CYTI indicate that this protein has a *left*-handed, somewhat flattened helical bend at this position [see paper 1 for discussion of the various regions of the (κ, τ) plane]. CYTR shows a *right*-handed, more flattened bend. In both proteins, the local conformation at residue 23 falls into the extended region. The magnitudes of κ and τ at residue 23 are both larger in CYTI, indicating that in CYTI the chain is less straight and less flat (more twisted) than in CYTR. At residue 24, CYTR has local structure typical of a left-handed, somewhat twisted extended region. CYTI has a much more twisted left-handed structure, more open than, but not very different from, a left-handed helical bend. At residue 25, CYTI has a fairly flat, extended local structure, while CYTR has the same type of local folding as at residue 24—a more twisted (left-handed) extended structure.

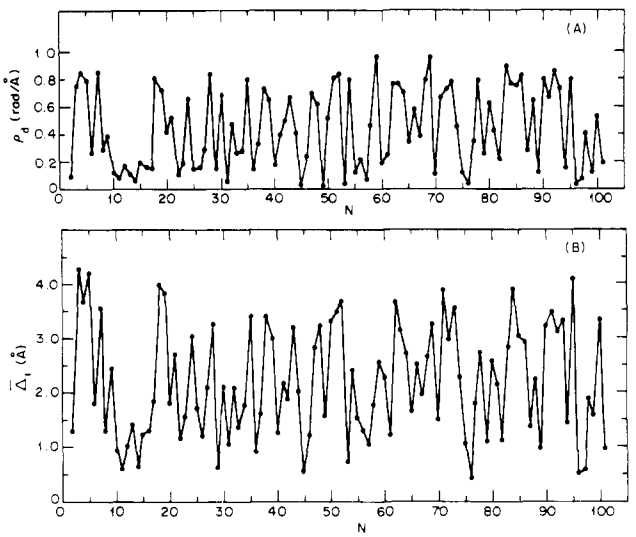

Figure 9. Comparison of CYTI and GPDR: (A) using DG comparison method; (B) using SS comparison method.

A more detailed discussion of the structural significance of the various regions of the (κ, τ) plane will be given in the next paper in this series.[4] It is clear from the foregoing, however, that the DG approach provides more than simply a convenient method for conformational comparison. It also provides detailed information about the local folding of each molecule separately. It is interesting to observe that the section of backbone represented by this peak in ρ_d is located on the outside of the molecule.[10,11] This is also the case for the peak at residues 51–56. Because these backbone segments are exposed to solvent, they have more freedom of motion than segments confined in the interior of the protein. This reflects itself experimentally in more diffuse electron density and, therefore, in somewhat different coordinate assignments and local folding. The differential geometric analysis is thus able to correlate with intuitive ideas about the folding of the two molecules (see Addendum).

It is of interest to apply the DG and SS methods to the comparison of two very different structures, in order to show that the agreement between the two methods is not an artifact arising from the high degree of similarity between CYTI and CYTR. We therefore conducted a comparison between CYTI and the first 103 residues of lobster glyceraldehyde phosphate dehydrogenase, red subunit (GPDR).[12] The results of this comparison are shown in Figure 9. It is clear that the chain plot of ρ_d is again very highly correlated with that of $\bar{\Delta}_i$, despite the very low similarity of the two structures. The correspondence between the results obtained from the two methods is thus real.

We now give another example of the power of the DG method for detecting similarity between conformations, by comparing two domains of ferredoxin whose similarity has been noted previously.[9] In Figure 10, we show the comparison of residues 1–26 of ferredoxin[13] with successive 26-residue segments of this protein, beginning at the N terminus of the molecule. We plot the average value of ρ_d, i.e., $\bar{\rho}_d$, for each 26-residue comparison, as a function of the number of the second residue in each 26-residue strip [this being the first residue in the strip for which (κ, τ) exists]. The first point, of course, represents comparison of identical structures, so that $\bar{\rho}_d = 0$. It is clear that the point at $N = 29$, with $\bar{\rho}_d = 0.124$, is much lower than the other points, indicating very good matching between residues 1–26 and 28–53. In fact, the mean value of $\bar{\rho}_d$ for the nonidentical strips, including 28–53, is 0.408 ± 0.081

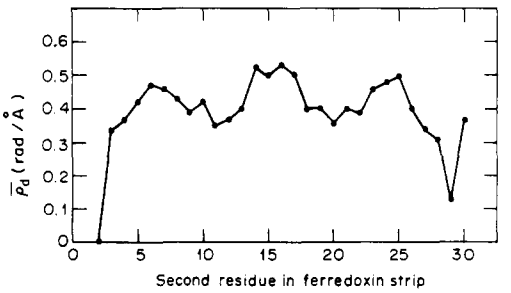

Figure 10. Comparison of residues 1–26 of ferredoxin with successive 26-residue strips of ferredoxin. The abscissa is the number of the second residue in the strip with which 1–26 is being compared. The ordinate is the average distance between coincident residues in the two strips.

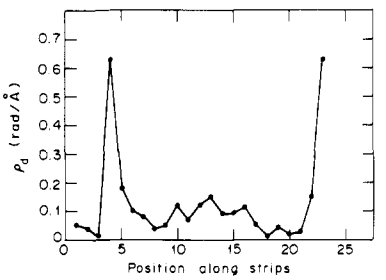

Figure 11. Chain-plot comparison of residues 1–26 and 28–53 of ferredoxin.

rad/Å, so that the match of 1–26 and 28–53 is 3.5 standard deviations better than the average. In Figure 11 we show a chain-plot comparison of 1–26 with 28–53. There is clearly excellent agreement, except at the 4th position along the two strips (residues 5 and 32) and at the end of the strip (residues 24 and 51). Inspection of the actual optimal superposition[9] reveals, however, that there is a one-residue insertion in the second domain, so located that the C^α of residue 31 is equivalent to that of residue 5 and the C^α of residue 34 is equivalent to that of residue 7. This is brought out very clearly by the large peak in ρ_d at positions 4 and 5 along the strip, corresponding to the attempted matching of local conformations at residues 5 and 32 and 6 and 33. The divergence of local folding at the end of the strip, noted above, is also clearly visible in the superposition.[9] The DG method has successfully chosen the one of the two possible registrations of the 26-residue first domain with the 27-residue second domain which is best with respect to similarity of local folding. The similarity in *local* folding at positions 1–3 along the strip is real, although residues 2–4 are out of registration with residues 29–31 in the superposition,[9] due to the one-residue insertion. We have thus demonstrated the ability of the DG comparison method to detect conformational similarity without the use of optimization.

We close this section with some general remarks on the differential-geometric approach to protein backbone structure. The principal thrust of the first two papers in this series has been the development of the fundamental equations, and application to the comparison of structures. This reflects our view that a method is needed for the comparison of *local* backbone folding, as shown by four-C^α segments of the backbone. In this paper, we have demonstrated that the DG method is capable of pinpointing divergences in local folding between two conformations in the neighborhood of each residue of the backbone, without the need for optimization of superposition of the two conformations. We have also demonstrated, however, that the DG approach provides precise and detailed information about each conformation separately. This is because the

DG approach is a *representation*, in the same sense that the set of (ϕ, ψ) gives a complete representation of the molecular structure. The (κ, τ) representation, however, operates on a different, longer length scale than the (ϕ, ψ) representation; the former describes the structure of four-C^α units, while the latter describes the structure of individual residues. The (κ, τ) representation, in turn, operates on a shorter length scale than the distance representation,[14] in which backbone structure is determined by specifying the distances between all pairs of C^α's in the molecule, including those remote from each other in space. We stress the fact that these three approaches are complementary. Although they all contain the same information, the method of presentation of each emphasizes different aspects of the backbone folding. Thus, for example, bends, which are somewhat difficult to classify in terms of (ϕ, ψ),[15] are handled in an elegant and structurally intuitive manner by the DG representation. The recently developed distance geometric methods,[14] which are based on the distance representation, are a somewhat more natural method for the study of long-range interactions. We will discuss the extension of the DG representation to longer length scales in a forthcoming publication.[4]

A corollary of this hierarchy of length scales in proteins is the existence of "degeneracy" in protein organization, a point which is clearly brought out by the DG representation. For example, as noted above,[15] there are 11 classes of bends in terms of (ϕ, ψ). It was demonstrated in paper 1 that bends of distinctly different (ϕ, ψ) types can fall quite close to each other in the (κ, τ) plane, indicating a high degree of structural similarity in terms of actual C^α arrangement. In general, the chain has several possible ways on the single-residue scale of accomplishing a given arrangement on the four-C^α length scale. This "degeneracy" undoubtedly extends to longer scales in the molecule. It is intriguing to speculate that this feature may play a role in protein self-organization.

Summary

In this paper, we have dealt with a number of mathematical aspects of the differential-geometric representation of protein backbone structure and have presented a more refined method for the comparison of protein backbone conformations than that of paper 1. This comparison method requires no optimization of superposition and reveals differences in *local* folding in the neighborhood of *each* residue along the backbone. The following mathematical points were discussed.

(1) The dependence of κ and τ on the factor $p_i p_{i+1}$, which can assume the values ± 1. The result of small changes in conformation, which cause $p_i p_{i+1}$ to change sign, was demonstrated, as was the proper method for including such changes in the distance function in the (κ, τ) plane.

(2) It was shown that there is a discontinuity in the (κ, τ) plane, centered on the τ axis, arising from the inversion of q_i, the sign of κ_i, which can assume the values ± 1. The dependence of the form of this forbidden region on θ, the virtual bond angle, was discussed, and a method for including this effect in the (κ, τ) distance function was outlined.

(3) An error in the discussion of local parity (chain handedness) in paper 1 was corrected. It was shown that, as a result of this correction, the DG method is a much more powerful tool for the discussion of local parity than suggested in paper 1.

(4) A brief discussion of the inverse process, for passing from a set of (κ_i, τ_i) and an initial θ_2 to a set of coordinates for C^α's, was given. Results of this transformation were summarized for bovine pancreatic trypsin inhibitor.

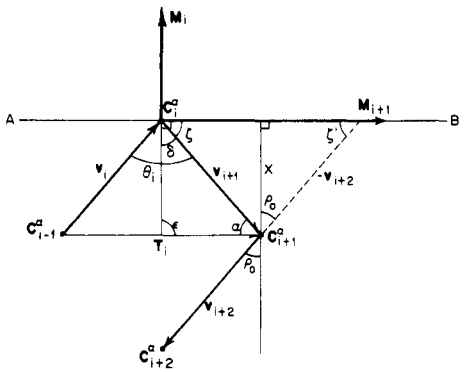

Figure 12. Construction for determining the condition for inversion of $p_i p_{i+1}$. See eq I-23.

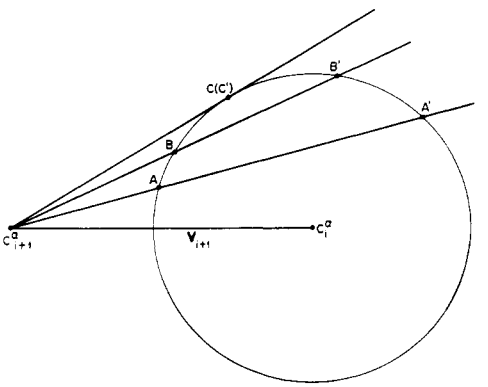

Figure 13. Effect of increasing τ_i at fixed χ_i, for $\chi_i < \lambda_i$. View, looking backward along $\hat{\mathbf{t}}_i$ (represented by the central black dot) toward C^α_i. The virtual bond $\mathbf{v}_{i+1}$ connects C^α_i and C^α_{i+1}. C^α_{i-1} lies directly behind C^α_{i+1}. By the definition of curvature (eq 6), C^α_{i+2} must lie on a cone Γ_i with apex at C^α_i, whose axis is coincident with $\hat{\mathbf{t}}_i$, and which makes an angle χ_i with $\hat{\mathbf{t}}_i$. The circle is a cross section of this cone in the plane Q which is perpendicular to $\hat{\mathbf{t}}_i$ and contains C^α_{i+1}. C^α_{i+2} must also lie in a plane O determined by the torsion τ_i. The oblique lines represent the intersections of the planes O and Q for various values of τ_i. As noted in the text and in paper 1, the two intersections (A and A') of O and Γ_i for a given τ_i (and likewise B and B' for another value of τ_i) represent the two possible locations of C^α_{i+2}, given χ_i and τ_i, and are distinguished by different values of q_i, the sign of κ_i. At the point C (identical with C'), which has the maximum possible τ_i for the given χ_i, the conformations with $q_i = +1$ and -1 become identical.

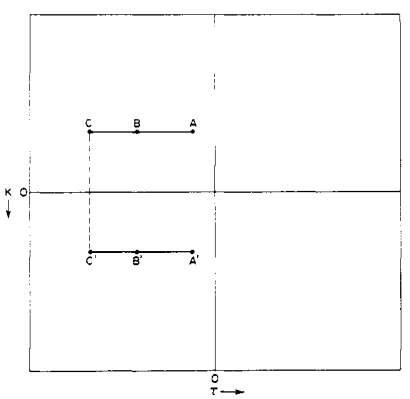

Figure 14. Path traced in the (κ, τ) plane as the deformation of Figure 13 is performed, with C and C' corresponding to the same conformation.

(5) A standard superposition method was described for determining atomic displacements of the C^α's of one four-C^α unit relative to those of another without computational optimization of superposition. It was shown that this method provides information about local folding de-

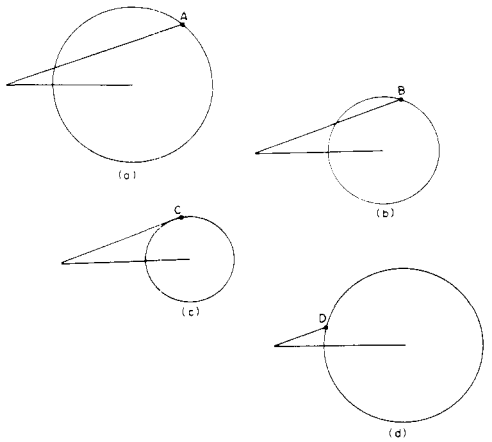

Figure 15. Effect of decreasing χ_i at fixed τ_i, for $\chi_i < \lambda_i$. View of same type as in Figure 13. In a–c, χ_i decreases; in d, χ_i has begun to increase, along the second branch of q_i.

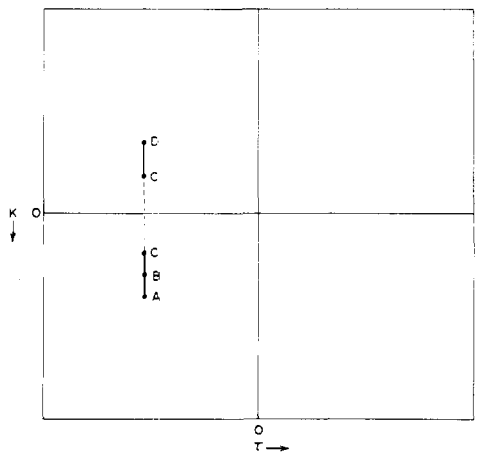

Figure 16. Path in the (κ, τ) plane traced out by the transformation of Figure 15, with both C's representing the same point.

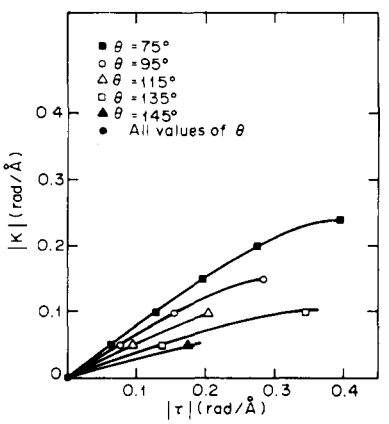

Figure 17. Plot of $|\kappa_i|$ as a function of $|\tau_i|$ for various θ_i, along the boundary of the gap centered on the τ axis.

viations and is therefore a useful standard for judging the accuracy of the DG comparison method.

(6) The distance function in the (κ, τ) plane was formulated and its accuracy for conformational comparison demonstrated by comparison with the standard superposition method outlined above. It was shown that the DG method provides detailed structural information about each conformation separately, making it possible to pinpoint the location of local structural differences on a four-C^α length scale and to describe the nature of these differences precisely. Three specific cases were treated: comparison of oxidized and reduced cytochrome *c*, comparison of oxidized cytochrome *c* with glyceraldehyde phosphate dehydrogenase, and comparison of the two domains of ferredoxin. The latter comparison is a particularly striking demonstration of the power of the method for the detection of structural similarity.

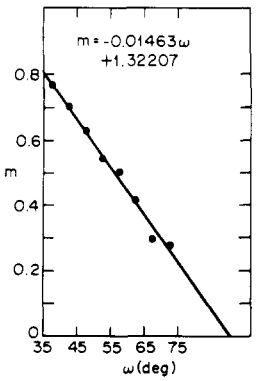

Figure 18. Plot of m (eq III-6 and III-7) vs. ω.

We closed the paper with a general discussion of the properties of the DG representation and its relationship to other representations in common use.

Acknowledgment. We thank Drs. G. Némethy and H. Wako for helpful advice and discussions, M. S. Pottle and S. Rumsey for programming assistance, and the referee for helpful suggestions on presentation.

Appendix I. Proofs of Theorems 1 and 2

A. Proof of Theorem 1. We begin by noting that, according to eq 12–15

$$\mathbf{B}_i = p_i(\mathbf{v}_i + \mathbf{v}_{i+1}) \times (\mathbf{v}_i - \mathbf{v}_{i+1}) \tag{I-1}$$

Upon rearrangement, this gives

$$\mathbf{B}_i = 2p_i(\mathbf{v}_{i+1} \times \mathbf{v}_i) \tag{I-2}$$

from which it follows that the unit vectors are

$$\hat{\mathbf{b}}_i = p_i \frac{\mathbf{v}_{i+1} \times \mathbf{v}_i}{|\mathbf{v}_{i+1} \times \mathbf{v}_i|} \tag{I-3}$$

and similarly

$$\hat{\mathbf{b}}_{i+1} = p_{i+1} \frac{\mathbf{v}_{i+2} \times \mathbf{v}_{i+1}}{|\mathbf{v}_{i+2} \times \mathbf{v}_{i+1}|} \tag{I-4}$$

From eq 8, I-3, and I-4

$$\gamma_i = \cos^{-1}(\hat{\mathbf{b}}_i \cdot \hat{\mathbf{b}}_{i+1}) \tag{I-5}$$

$$= \cos^{-1}\left\{p_i p_{i+1} \frac{\mathbf{v}_{i+1} \times \mathbf{v}_i}{|\mathbf{v}_{i+1} \times \mathbf{v}_i|} \cdot \frac{\mathbf{v}_{i+2} \times \mathbf{v}_{i+1}}{|\mathbf{v}_{i+2} \times \mathbf{v}_{i+1}|}\right\} \tag{I-6}$$

Furthermore, from eq 2, 4, and 9

$$u_i = \text{sgn}(\hat{\mathbf{b}}_{i+1} \cdot \hat{\mathbf{n}}_i) \tag{I-7}$$

$$= \text{sgn}(p_i p_{i+1}[\hat{\mathbf{t}}_{i+1} \times \hat{\mathbf{m}}_{i+1}] \cdot \hat{\mathbf{m}}_i) \tag{I-8}$$

$$= p_i p_{i+1} \text{ sgn}([\hat{\mathbf{t}}_{i+1} \times \hat{\mathbf{m}}_{i+1}] \cdot \hat{\mathbf{m}}_i) \tag{I-9}$$

Therefore, from eq 8, I-6, and I-9, inversion of the sign of $p_i p_{i+1}$ causes τ_i to assume the new value $\bar{\tau}_i$ given by

$$\bar{\tau}_i = -\left(\frac{\pi - |\mathbf{v}_{i+1}||\tau_i|}{|\mathbf{v}_{i+1}|}\right) \text{sgn}(\tau_i) \tag{I-10}$$

$$= -(\pi/|\mathbf{v}_{i+1}| - |\tau_i|) \text{ sgn}(\tau_i) \tag{I-11}$$

[In the case of ideal geometry, $|\mathbf{v}_{i+1}| = 3.8$ Å, and $\bar{\tau}_i = -[0.83 - |\tau_i|] \text{ sgn}(\tau_i)$.] The tangent vector, $\mathbf{t}_j$, is independent of the p_j's. Therefore

$$\chi_i = \cos^{-1}(\hat{\mathbf{t}}_i \cdot \hat{\mathbf{t}}_{i+1}) \tag{I-12}$$

is unaffected by a change in sign of $p_i p_{i+1}$. From eq 2, 4, and 7, however

$$q_i = p_i \operatorname{sgn}[(\hat{\mathbf{t}}_i \times \hat{\mathbf{t}}_{i+1}) \cdot \hat{\mathbf{b}}_{i+1}] \tag{I-13}$$

$$= p_i \operatorname{sgn}[p_{i+1}(\hat{\mathbf{t}}_i \times \hat{\mathbf{t}}_{i+1}) \cdot (\hat{\mathbf{t}}_{i+1} \times \hat{\mathbf{m}}_{i+1})] \tag{I-14}$$

$$= p_i p_{i+1} \operatorname{sgn}[(\hat{\mathbf{t}}_i \times \hat{\mathbf{t}}_{i+1}) \cdot (\hat{\mathbf{t}}_{i+1} \times \hat{\mathbf{m}}_{i+1})] \tag{I-15}$$

Therefore, inversion of the sign of $p_i p_{i+1}$ causes κ_i to assume the value $\tilde{\kappa}_i$ given by

$$\tilde{\kappa}_i = -\kappa_i \tag{I-16}$$

Hence, we see that, when $p_i p_{i+1}$ changes sign

$$(\kappa_i, \tau_i) \rightarrow (\tilde{\kappa}_i, \tilde{\tau}_i) \tag{I-17}$$

where $\tilde{\kappa}_i$ and $\tilde{\tau}_i$ are given by eq I-16 and I-11, respectively. This confirms mathematically the observation made in paper 1.

[It should be noted, incidentally, that the quantity in square brackets in eq I-15 can be expanded by using standard vector identities

$$(\hat{\mathbf{t}}_i \times \hat{\mathbf{t}}_{i+1}) \cdot (\hat{\mathbf{t}}_{i+1} \times \hat{\mathbf{m}}_{i+1}) = (\hat{\mathbf{t}}_i \cdot \hat{\mathbf{t}}_{i+1})(\hat{\mathbf{t}}_{i+1} \cdot \hat{\mathbf{m}}_{i+1}) - (\hat{\mathbf{t}}_i \cdot \hat{\mathbf{m}}_{i+1}) \tag{I-18}$$

$$= (\hat{\mathbf{t}}_{i+1} \cdot \hat{\mathbf{m}}_{i+1}) \cos \chi_i - (\hat{\mathbf{t}}_i \cdot \hat{\mathbf{m}}_{i+1}) \tag{I-19}$$

In ideal geometry, where all $|\mathbf{v}_i| = 3.8$ Å, the local coordinate frame is orthogonal and

$$\hat{\mathbf{t}}_{i+1} \cdot \hat{\mathbf{m}}_{i+1} = 0 \tag{I-20}$$

In this case, therefore

$$q_i = -p_i p_{i+1} \operatorname{sgn}(\hat{\mathbf{t}}_i \cdot \hat{\mathbf{m}}_{i+1}) \tag{I-21}$$

$$= -p_i \operatorname{sgn}(\hat{\mathbf{t}}_i \cdot \hat{\mathbf{n}}_{i+1}) \tag{I-22}$$

and we have a much simpler formula for q_i than eq 7].

B. Proof of Theorem 2. We now turn to a determination of the conditions under which the product $p_i p_{i+1}$ inverts sign. It follows from eq 2, 3, and 13 that $p_i p_{i+1} = \operatorname{sgn}(\mathbf{M}_{i+1} \cdot \mathbf{M}_i)$; hence, $\mathbf{M}_{i+1}$ and $\mathbf{M}_i$ are perpendicular at the point of inversion, i.e.

$$\mathbf{M}_{i+1} \cdot \mathbf{M}_i = 0 \tag{I-23}$$

As can be seen from Figure 12, eq I-23 is satisfied by any $\mathbf{v}_{i+2}$ lying on a cone forming an angle ρ_0 with $\mathbf{M}_i$. In order to identify the point of inversion, it is necessary to determine ρ_0. To do so, it is convenient to take C^{α}_{i+2} coplanar with C^{α}'s $i-1$, i, $i+1$; this can be done without loss of generality.

The angle α (required for use in eq I-32 and I-33 can be determined as follows. From eq 12

$$\mathbf{v}_i = \mathbf{T}_i - \mathbf{v}_{i+1} \tag{I-24}$$

Therefore

$$|\mathbf{v}_i|^2 = |\mathbf{T}_i|^2 + |\mathbf{v}_{i+1}|^2 - 2\mathbf{T}_i \cdot \mathbf{v}_{i+1} \tag{I-25}$$

$$= |\mathbf{T}_i|^2 + |\mathbf{v}_{i+1}|^2 - 2|\mathbf{T}_i||\mathbf{v}_{i+1}| \cos \alpha \tag{I-26a}$$

Hence

$$\cos \alpha = \frac{|\mathbf{T}_i|^2 + |\mathbf{v}_{i+1}|^2 - |\mathbf{v}_i|^2}{2|\mathbf{T}_i||\mathbf{v}_{i+1}|} \tag{I-26b}$$

Now, from eq 12

$$|\mathbf{T}_i|^2 = |\mathbf{v}_i|^2 + |\mathbf{v}_{i+1}|^2 - 2|\mathbf{v}_i||\mathbf{v}_{i+1}| \cos \theta_i \tag{I-27}$$

Therefore

$$\cos \alpha = (|\mathbf{v}_{i+1}| - |\mathbf{v}_i| \cos \theta_i)/[|\mathbf{v}_i|^2 + |\mathbf{v}_{i+1}|^2 - 2|\mathbf{v}_i||\mathbf{v}_{i+1}| \cos \theta_i]^{1/2} \tag{I-28}$$

In nonideal geometry, the local coordinate frame is not orthogonal, i.e., $\epsilon \neq 90°$ (Figure 12). The angle ϵ is determined by

$$\mathbf{M}_i \cdot \mathbf{T}_i = |\mathbf{M}_i||\mathbf{T}_i| \cos \epsilon \tag{I-29}$$

Applying the definitions of $\mathbf{M}_i$ and $\mathbf{T}_i$

$$\cos \epsilon = (|\mathbf{v}_i|^2 - |\mathbf{v}_{i+1}|^2)/|\mathbf{M}_i||\mathbf{T}_i| \tag{I-30}$$

where, according to eq 13

$$|\mathbf{M}_i|^2 = |\mathbf{v}_i|^2 + |\mathbf{v}_{i+1}|^2 + 2|\mathbf{v}_i||\mathbf{v}_{i+1}| \cos \theta_i \tag{I-31}$$

and $|\mathbf{T}_i|$ is given by eq I-27. Now

$$\delta = \pi - \alpha - \epsilon \tag{I-32}$$

and, since AB is perpendicular to $\mathbf{M}_i$ by construction

$$\zeta = \pi/2 - \delta = \alpha + \epsilon - \pi/2 \tag{I-33}$$

The quantity x defined in Figure 12 is given by

$$x = |\mathbf{v}_{i+1}| \sin \zeta \tag{I-34}$$

But

$$\sin \zeta' = x/|\mathbf{v}_{i+2}| \tag{I-35}$$

Hence

$$\sin \zeta' = \frac{|\mathbf{v}_{i+1}|}{|\mathbf{v}_{i+2}|} \sin \zeta \tag{I-36}$$

Therefore

$$\rho_0 = \pi/2 - \zeta' \tag{I-37}$$

$$= \pi/2 - \sin^{-1}\left\{\frac{|\mathbf{v}_{i+1}|}{|\mathbf{v}_{i+2}|} \sin \zeta\right\} \tag{I-38}$$

Using eq I-33 and the properties of the trigonometric functions, we obtain

$$\rho_0 = \pi/2 - \sin^{-1}\left\{-\frac{|\mathbf{v}_{i+1}|}{|\mathbf{v}_{i+2}|} \cos(\alpha + \epsilon)\right\} \tag{I-39}$$

In general, therefore, if ρ is the angle defined by

$$-\mathbf{v}_{i+2} \cdot \mathbf{M}_i = |\mathbf{v}_{i+2}||\mathbf{M}_i| \cos \rho \tag{I-40}$$

then, as can be seen from Figure 12 and eq 3, the sign of the component of $\mathbf{M}_{i+1}$ on $\mathbf{M}_i$ (i.e., $p_i p_{i+1}$) depends on the relative sizes of ρ and ρ_0; i.e.

$$p_i p_{i+1} = -1 \text{ for } \rho > \rho_0 \tag{I-41a}$$

and

$$p_i p_{i+1} = +1 \text{ for } \rho < \rho_0 \tag{I-41b}$$

Therefore, $p_i p_{i+1}$ changes sign when the conformation is such that $\rho = \rho_0$, where ρ_0 is given by eq I-39. [It should be noted that, for standard geometry, where $|\mathbf{v}_i| = 3.8$ Å for all i

$$\alpha = \cos^{-1}[\sin(\theta_i/2)] \tag{I-42}$$

and

$$\epsilon = \pi/2 \tag{I-43}$$

so that

$$\rho_0 = \theta_i/2 \tag{I-44}$$

Also, for standard geometry, $\mathbf{v}_{i+2}$ moves on a cone, one of whose edges is collinear with $\mathbf{v}_{i+1}$ when $\rho = \rho_0$. Inspection of Figure 12 shows that $\rho = \rho_0$ when $\chi_i = \lambda_i$; hence, if $\rho < \rho_0$, then $p_i p_{i+1} = +1$ (by eq I-41b), a condition that can occur only when $\chi_i > \lambda_i$.]

We have thus established that, in certain regions of (κ, τ) space, a small displacement of the conformation of a

four-C^α unit, sufficient to cause ρ to change magnitude relative to ρ_0 (eq I-41), can lead to a change in sign of $p_i p_{i+1}$ and, therefore, to a jump in (κ, τ) given by eq I-17. This jump is similar to that which occurs in systems with periodic boundary conditions—e.g., in the (ϕ, ψ) representation of a peptide, when a small increase in a dihedral angle that lies near 180° causes the point (ϕ, ψ) to reappear near the −180° line. This behavior is taken into account in a systematic fashion in the distance function described in section VI.

Appendix II. Determination of Inversion Pathway ($|\tau_i^0|$ as a Function of κ_i and θ_i) for Standard Geometry

In this appendix, we solve eq 22–24. We shall assume standard geometry, which is sufficient for the illustrative purposes of this calculation and simplifies the algebra. Therefore

$$|\mathbf{v}_j| = 3.8 \text{ Å} \tag{II-1}$$

for all j.

Let us denote the angle between $\mathbf{v}_i$ and $\mathbf{v}_{i+2}$ by $\Omega_{i,i+2}$. Then, from eq 22 and II-1, at a point of inversion

$$\cos \Omega_{i,i+2} = -(\cos \theta_i + \cos \theta_{i+1} + 1) \tag{II-2}$$

Expanding eq 23 and using eq II-1 and II-2, we find

$$-\cos \theta_i - \cos \theta_{i+1} = 2 \sin (\theta_i/2) \sin (\theta_{i+1}/2) \cos \chi_i \tag{II-3}$$

Similarly, from eq 24, II-1, and II-2 we find

$$|\cos \theta_{i+1} \cos \theta_i + \cos \theta_i + \cos \theta_{i+1} + 1| = \sin \theta_i \sin \theta_{i+1} |\cos \gamma_i| \tag{II-4}$$

Therefore, at a point of inversion, we have eq II-3 and II-4 as simultaneous equations. For convenience, let us establish the notation

$$x \equiv \cos \theta_i \tag{II-5}$$

$$y \equiv \cos \theta_{i+1} \tag{II-6}$$

Then, making use of the properties of the trigonometric functions and the experimentally observed bounds[16] on θ

$$75° \leq \theta \leq 145° \tag{II-7}$$

eq II-4 and II-3 can be rewritten as the simultaneous set of equations

$$(1 + x)^{1/2}(1 + y)^{1/2} = (1 - x)^{1/2}(1 - y)^{1/2} |\cos \gamma_i| \tag{II-8}$$

$$-(x + y) = (1 - x)^{1/2}(1 - y)^{1/2} \cos \chi_i \tag{II-9}$$

From eq II-8

$$\cos \gamma_i = \pm \left[\frac{(1 + x)(1 + y)}{(1 - x)(1 - y)} \right]^{1/2} \tag{II-10}$$

But y can be determined in terms of x by eq II-9. The solution is readily seen to be

$$2y = (-2 + \cos^2 \chi_i)x - \cos^2 \chi_i \pm (1 - x)^{1/2} \cos \chi_i [(4 - \cos^2 \chi_i)x + (4 + \cos^2 \chi_i)]^{1/2} \tag{II-11}$$

In general, only one choice of sign gives a value of y which is also a solution of eq II-9. That value of y, together with eq II-10, enables us to determine $|\tau_i^0|$ as a function of θ_i and χ_i by numerical evaluation.

Appendix III. Proofs of Theorems 3 and 4

In this appendix we prove theorems 3 and 4 of section IIIB. Recognition of the existence of a discontinuity is important in constructing a distance function since the distance between any two points separated by the discontinuity must be reduced by the size of the gap in the path connecting them.

A. Proof of Theorem 3. One way of showing that there is a discontinuity is to consider the effect of increasing τ at fixed χ. As shown in Figure 13 (and previously in paper 1), for given τ and χ, there are two possible positions of C^α_{i+2}. It was shown in paper 1 that these are characterized by opposite signs of q_i. As $|\tau|$ increases, these two points approach each other along the surface of the cone determined by the value of χ. At the point C (≡C′) the two points coalesce. The corresponding path (e.g., going from A to A′) traced out in (κ, τ) space is shown in Figure 14. It can be seen that points C and C′, which coincide in actual conformation, are separated by a gap in Figure 14. This is the discontinuity that we wish to demonstrate in the (κ, τ) plane.

Alternatively, we can follow a path of decreasing χ at fixed τ. Four stages in such a path are shown in Figure 15. The corresponding path in (κ, τ) space is shown in Figure 16. Once again we note a discontinuity in the (κ, τ) plane in going from A to D.

From the illustrations in Figures 13–16, it is seen that the existence of this gap arises as a combination of two factors. One is that, for $\chi_i < \lambda_i$, there is a maximum possible value of τ_i for a given χ_i (i.e., $|\kappa_i|$); this is a purely *geometric* factor. The other is the mathematical convention that q_i, the sign of κ_i, distinguishes between the two possible positions of C^α_{i+2} for given τ_i and χ_i.

B. Proof of Theorem 4. We now examine the size and shape of the gap. It follows from the foregoing that the point at which the discontinuity occurs is the point at which the two solutions for θ_{i+1} in terms of χ_i, γ_i, and θ_i are equal (the details of this solution are discussed in paper 1). From eq A-12 of paper 1, these two solutions become equal when the following condition (in standard geometry) is satisfied

$$\cos \gamma_i \cos \omega_i [\cos^2 \gamma_i \cos^2 \omega_i - \cos^2 \chi_i + \sin^2 \omega_i]^{1/2} = 0 \tag{III-1}$$

where ω_i is given by eq 10.

The solution

$$\cos \omega_i = 0 \tag{III-2}$$

is ruled out by the limits[16] on the angle θ

$$75° < \theta < 145° \tag{III-3}$$

The solution

$$\cos \gamma_i = 0 \tag{III-4}$$

corresponding to $\gamma_i = \pi$, is possible only for $\chi_i > \lambda_i$, for which there is clearly no discontinuity across the τ axis (as stated above, the gap arises only when $\chi_i < \lambda_i$). Therefore, the only solution of interest is given by setting the bracketed term in eq III-1 equal to zero, from which we have

$$\cos \gamma_i = \pm[(\cos^2 \chi_i - \sin^2 \omega_i)/\cos^2 \omega_i]^{1/2} \tag{III-5}$$

along the discontinuity; i.e., eq III-5 is satisfied at all points along the borders of the discontinuity gap. This equation enables us to study the discontinuity as a function of χ_i and θ_i. The result is shown in Figure 17, in which we plot $|\kappa_i|$ as a function of $|\tau_i|$ along the boundary of the gap for various values of θ_i. Each of these curves describes the size and shape of the gap around the τ axis for the given value of θ_i. It is clear that the size and shape of the gap is a strong function of θ_i. It is also clear that, for a given θ_i, $|\kappa_i|$ is very nearly linear over a substantial range in $|\tau_i|$. It is useful in the construction of the distance function in the

(κ, τ) plane to plot this initial slope m of the curves of Figure 17 as a function of θ_i, where m is defined by

$$|\kappa_i| = m|\tau_i| \tag{III-6}$$

As is shown in Figure 18 this is well fit by a straight line with the formula

$$m = -0.0146\omega + 1.3221 \tag{III-7}$$

when $\omega = \theta/2$ is measured in degrees.

Inspection of Figure 3 reveals the presence of a butterfly-shaped gap about the τ axis, of the type discussed above.

Appendix IV. The Standard Superposition Method

In Appendix V we derive the simultaneous equations for the components of Δ_{i-1} and Δ_{i+1} in the local coordinate frame $(\hat{\mathbf{t}}_i, \hat{\mathbf{m}}_i, \hat{\mathbf{b}}_i)$. It should be noted that, to convert to the $(\hat{\mathbf{t}}_i, \hat{\mathbf{n}}_i, \hat{\mathbf{b}}_i)$ frame, the $\mathbf{m}_i$ components should be multiplied by p_i (see eq 2). The equations are

$$\Delta_{i+1}\cdot\hat{\mathbf{t}}_i - \Delta_{i-1}\cdot\hat{\mathbf{t}}_i = |\mathbf{T}_i'| - |\mathbf{T}_i| \tag{IV-1}$$

$$\Delta_{i+1}\cdot\hat{\mathbf{t}}_i + \Delta_{i-1}\cdot\hat{\mathbf{t}}_i = |\mathbf{M}_i| \cos\eta_i - |\mathbf{M}_i'| \cos\eta_i' \tag{IV-2}$$

$$\Delta_{i+1}\cdot\hat{\mathbf{m}}_i - \Delta_{i-1}\cdot\hat{\mathbf{m}}_i = |\mathbf{T}_i'| \cos\eta_i - |\mathbf{T}_i| \cos\eta_i \tag{IV-3}$$

$$\Delta_{i+1}\cdot\hat{\mathbf{m}}_i + \Delta_{i-1}\cdot\hat{\mathbf{m}}_i = |\mathbf{M}_i| - |\mathbf{M}_i'| \cos H_i \tag{IV-4}$$

The angles η_i', η_i, and H_i are defined in Appendix V. Since Δ_{i+1} and Δ_{i-1} are both in the plane of $\mathbf{v}_i$, $\mathbf{v}_{i'}$, $\mathbf{v}_{i+1}$, $\mathbf{v}_{i'+1}$, and therefore perpendicular to $\hat{\mathbf{b}}_i$

$$\Delta_{i+1}\cdot\hat{\mathbf{b}}_i = \Delta_{i-1}\cdot\hat{\mathbf{b}}_i = 0 \tag{IV-5}$$

We now need the components of the position of C^α_{i+2} in each of the two four-C^α units in the $(\hat{\mathbf{t}}_i, \hat{\mathbf{m}}_i, \hat{\mathbf{b}}_i)$ frame. But it should be noted that, by eq 12, the position vectors of these two C^α_{i+2}'s in the local frame are simply $\mathbf{T}_{i+1}$ and $\mathbf{T}_{i+1}'$. We therefore calculate the components of these vectors in the local coordinate system.

The details of this calculation are relegated to Appendix VI. The results are as follows:

$$\mathbf{T}_{i+1}\cdot\hat{\mathbf{t}}_i = |\mathbf{T}_{i+1}| \cos\chi_i \tag{IV-6}$$

$$\mathbf{T}_{i+1}\cdot\hat{\mathbf{m}}_i = (|\mathbf{T}_i||\mathbf{T}_{i+1}| \cos\chi_i - 2|\mathbf{v}_{i+1}|^2 + 2|\mathbf{v}_{i+2}||\mathbf{v}_{i+1}| \cos\theta_{i+1})/|\mathbf{M}_i| \tag{IV-7}$$

$$\mathbf{T}_{i+1}\cdot\hat{\mathbf{b}}_i = \frac{2p_{i+1}u_i|\mathbf{v}_{i+2}\cdot(\mathbf{v}_{i+1} \times \mathbf{v}_i)|}{|\mathbf{T}_i||\mathbf{M}_i||\sin\eta_i|} \tag{IV-8}$$

where $|\mathbf{v}_{i+2}\cdot(\mathbf{v}_{i+1} \times \mathbf{v}_i)|$ is calculated from eq III-10. It should be noted once again that the $\hat{\mathbf{n}}_i$ component of $\mathbf{T}_{i+1}$ is obtained from eq IV-7 by multiplying by p_i. Precisely analogous results hold for the components of $\mathbf{T}_{i+1}'$ in the $(\hat{\mathbf{t}}_i', \hat{\mathbf{n}}_i', \hat{\mathbf{b}}_i')$ system.

Of course, in calculating the magnitudes of the Δ_j, we must take cognizance of the fact that neither the $(\hat{\mathbf{t}}_i', \hat{\mathbf{n}}_i', \hat{\mathbf{b}}_i')$ system nor the $(\hat{\mathbf{t}}_i, \hat{\mathbf{n}}_i, \hat{\mathbf{b}}_i)$ system is orthogonal in general and that the angles between axes in the two systems are generally different. This is straightforward.[17] The components of Δ_{i-1} and Δ_{i+1} are already expressed in the $(\hat{\mathbf{t}}_i, \hat{\mathbf{n}}_i, \hat{\mathbf{b}}_i)$ frame as solutions of eq IV-1 to IV-4. It remains only to observe that the magnitude of Δ_j $(j = i + 1, i - 1)$ in this frame is given by

$$|\Delta_j|^2 = \Delta_t^2 + \Delta_m^2 + 2\Delta_t\Delta_m \cos\eta_i \tag{IV-9}$$

where

$$\Delta_t = (\Delta_j\cdot\hat{\mathbf{t}}_i - \Delta_j\cdot\hat{\mathbf{m}}_i \cos\eta_i)/\sin^2\eta_i \tag{IV-10}$$

$$\Delta_m = (\Delta_j\cdot\hat{\mathbf{m}}_i - \Delta_j\cdot\hat{\mathbf{t}}_i \cos\eta_i)/\sin^2\eta_i \tag{IV-11}$$

The components of $\mathbf{T}_{i+1}'$ are given in the $\hat{\mathbf{t}}_i', \hat{\mathbf{n}}_i', \hat{\mathbf{b}}_i')$ frame by the analogues of eq IV-6 to IV-8. The transformation of the $(\hat{\mathbf{t}}_i', \hat{\mathbf{m}}_i')$ components of $\mathbf{T}_{i+1}'$ to the $(\hat{\mathbf{t}}_i, \hat{\mathbf{m}}_i)$ components is given by

$$(\mathbf{T}_{i+1}')_{t_i} = (\mathbf{T}_{i+1}')_{t_i'} - (\mathbf{T}_{i+1}')_{m_i'} \sin H_i/\sin\eta_i \tag{IV-12}$$

$$(\mathbf{T}_{i+1}')_{m_i} = (\mathbf{T}_{i+1}')_{m_i'} \sin\eta_i'/\sin\eta_i \tag{IV-13}$$

where

$$(\mathbf{T}_{i+1}')_{t_i'} = \frac{(\mathbf{T}_{i+1}'\cdot\hat{\mathbf{t}}_i') - (\mathbf{T}_{i+1}'\cdot\hat{\mathbf{m}}_i') \cos\eta_i'}{\sin^2\eta_i'} \tag{IV-14}$$

$$(\mathbf{T}_{i+1}')_{m_i'} = \frac{(\mathbf{T}_{i+1}'\cdot\hat{\mathbf{m}}_i') - (\mathbf{T}_{i+1}'\cdot\hat{\mathbf{t}}_i') \cos\eta_i'}{\sin^2\eta_i'} \tag{IV-15}$$

Analogues of eq IV-14 and IV-15 hold for $(\mathbf{T}_{i+1})_{t_i}$ and $(\mathbf{T}_{i+1})_{m_i}$. It is also clear that

$$(\mathbf{T}_{i+1}')_{n_i} = p_i(\mathbf{T}_{i+1}')_{m_i} \tag{IV-16}$$

and

$$(\mathbf{T}_{i+1}')_{b_i} = \mathbf{T}_{i+1}'\cdot\hat{\mathbf{b}}_i = p_i'p_i\mathbf{T}_{i+1}'\cdot\hat{\mathbf{b}}_i' \tag{IV-17}$$

Finally

$$|\Delta_{i+2}|^2 = [(\mathbf{T}_{i+1}')_{b_i} - (\mathbf{T}_{i+1})_{b_i}]^2 + [(\mathbf{T}_{i+1}')_{t_i} - (\mathbf{T}_{i+1})_{t_i}]^2 + [(\mathbf{T}_{i+1}')_{n_i} - (\mathbf{T}_{i+1})_{n_i}]^2 + 2[(\mathbf{T}_{i+1}')_{t_i} - (\mathbf{T}_{i+1})_{t_i}][(\mathbf{T}_{i+1}')_{n_i} - (\mathbf{T}_{i+1})_{n_i}] \cos\eta_i \tag{IV-18}$$

which completes the determination of the magnitudes of the Δ_j. Equation IV-9 gives the magnitudes of Δ_{i+1} and Δ_{i-1}, and eq IV-18 gives that of Δ_{i+2}; these are measures of the deviations of the two structures in the SSM. It should be noted that the determination of the components of $\mathbf{T}_{i+1}$ (which is equivalent to the determination of $\mathbf{v}_{i+2}$) in the $(\hat{\mathbf{t}}_i, \hat{\mathbf{n}}_i, \hat{\mathbf{b}}_i)$ frame essentially completes the representation of the virtual bond vectors in the local coordinate frame, for arbitrary geometry, as mentioned in the paragraph following eq 28. The representation of $-\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ in the local frame at C^α_i is straightforward.

Appendix V. Derivation of Eq IV-1 to IV-4

In this appendix, we derive eq IV-1 to IV-4, for the components of Δ_{i-1} and Δ_{i+1} in the $(\hat{\mathbf{t}}_i, \hat{\mathbf{m}}_i, \hat{o}_i)$ local coordinate frame. We begin by noting that, from eq 12, 5, and 43

$$\mathbf{T}_i' = \mathbf{T}_i + \Delta_{i+1} - \Delta_{i-1} \tag{V-1}$$

Taking the dot product of $\mathbf{T}_i$ with both sides and using eq 42, one arrives at the result

$$\Delta_{i+1}\cdot\mathbf{T}_i - \Delta_{i-1}\cdot\mathbf{T}_i = |\mathbf{T}_i'||\mathbf{T}_i| - |\mathbf{T}_i|^2 \tag{V-2}$$

Dividing through by $|\mathbf{T}_i|$ and using eq 1, we arrive at

$$\Delta_{i+1}\cdot\hat{\mathbf{t}}_i - \Delta_{i-1}\cdot\hat{\mathbf{t}}_i = |\mathbf{T}_i'| - |\mathbf{T}_i| \tag{V-3}$$

which is eq IV-1.

We now note that, for arbitrary geometry, the local coordinate frame is not necessarily orthogonal. We therefore define angles η_i and η_i' between the tangent and normal by

$$\mathbf{T}_i\cdot\mathbf{M}_i = |\mathbf{T}_i||\mathbf{M}_i| \cos\eta_i \tag{V-4}$$

and

$$\mathbf{T}_i'\cdot\mathbf{M}_i' = |\mathbf{T}_i'||\mathbf{M}_i'| \cos\eta_i' \tag{V-5}$$

and their difference by

$$H_i = \eta_i' - \eta_i \tag{V-6}$$

It is readily seen from eq V-4 and the definitions of $\mathbf{T}_i$ and $\mathbf{M}_i$ that

$$\cos \eta_i = \frac{|\mathbf{v}_i|^2 - |\mathbf{v}_{i+1}|^2}{|\mathbf{v}_{i+1} + \mathbf{v}_i||\mathbf{v}_i - \mathbf{v}_{i+1}|} \quad \text{(V-7)}$$

with an analogous equation for $\cos \eta_i'$. Clearly, if $|\mathbf{v}_i| = |\mathbf{v}_{i+1}|$ (as in standard geometry), $\eta_i = \pi/2$. It should be noted that, by construction, $\hat{\mathbf{b}}_i$ is orthogonal to $\hat{\mathbf{t}}_i$ and $\hat{\mathbf{n}}_i$ ($\hat{\mathbf{m}}_i$) in arbitrary geometry.

Now

$$\mathbf{M}_i' = \mathbf{M}_i - \Delta_{i-1} - \Delta_{i+1} \quad \text{(V-8)}$$

by reasoning analogous to that which led to eq V-1. Taking the dot product of eq V-8 with $\mathbf{M}_i$, and noting that the angle between $\mathbf{M}_i'$ and $\mathbf{M}_i$ is H_i, leads to the result

$$\Delta_{i-1}\cdot\mathbf{M}_i + \Delta_{i+1}\cdot\mathbf{M}_i = |\mathbf{M}_i|^2 - |\mathbf{M}_i||\mathbf{M}_i'| \cos H_i \quad \text{(V-9)}$$

Dividing through by $|\mathbf{M}_i|$ leads to eq IV-4.

Likewise, taking the dot product of eq V-1 with $\mathbf{M}_i$ leads to eq IV-3; the dot product of eq V-8 with $\mathbf{T}_i$ leads to eq IV-2.

Appendix VI. Derivation of Eq IV-6 to IV-8

In this appendix we derive eq IV-6 to IV-8 and the analogous equations for $\mathbf{T}_{i+1}'$.

Equation IV-6 follows directly from eq 6, 1, and 12. To derive eq IV-7 we note that

$$\mathbf{T}_{i+1}\cdot\mathbf{T}_i = |\mathbf{T}_i||\mathbf{T}_{i+1}| \cos \chi_i \quad \text{(VI-1)}$$

$$= (\mathbf{v}_i + \mathbf{v}_{i+1})\cdot(\mathbf{v}_{i+1} + \mathbf{v}_{i+2}) \quad \text{(VI-2)}$$

$$= |\mathbf{v}_{i+1}|^2 + \mathbf{v}_{i+2}\cdot\mathbf{v}_i - |\mathbf{v}_{i+1}||\mathbf{v}_i| \cos \theta_i - |\mathbf{v}_{i+2}||\mathbf{v}_{i+1}| \cos \theta_{i+1} \quad \text{(VI-3)}$$

Therefore

$$\mathbf{v}_{i+2}\cdot\mathbf{v}_i = |\mathbf{T}_i||\mathbf{T}_{i+1}| \cos \chi_i + |\mathbf{v}_{i+1}||\mathbf{v}_i| \cos \theta_i + |\mathbf{v}_{i+2}||\mathbf{v}_{i+1}| \cos \theta_{i+1} - |\mathbf{v}_{i+1}|^2 \quad \text{(VI-4)}$$

Now

$$\mathbf{T}_{i+1}\cdot\mathbf{M}_i = (\mathbf{v}_{i+1} + \mathbf{v}_{i+2})\cdot(\mathbf{v}_i - \mathbf{v}_{i+1}) = \mathbf{v}_{i+2}\cdot\mathbf{v}_i - |\mathbf{v}_{i+1}||\mathbf{v}_i| \cos \theta_i - |\mathbf{v}_{i+1}|^2 + |\mathbf{v}_{i+1}||\mathbf{v}_{i+2}| \cos \theta_{i+1} \quad \text{(VI-5)}$$

Substituting eq VI-4 and dividing by $|\mathbf{M}_i|$, we find

$$\mathbf{T}_{i+1}\cdot\hat{\mathbf{m}}_i = (|\mathbf{T}_i||\mathbf{T}_{i+1}| \cos \chi_i - 2|\mathbf{v}_{i+1}|^2 + 2|\mathbf{v}_{i+2}||\mathbf{v}_{i+1}| \cos \theta_{i+1})/|\mathbf{M}_i| \quad \text{(VI-6)}$$

which is eq IV-7.

The third component of $\mathbf{T}_{i+1}$ can be written

$$\mathbf{T}_{i+1}\cdot\mathbf{B}_i = \mathbf{T}_{i+1}\cdot(\mathbf{T}_i \times \mathbf{N}_i) \quad \text{(VI-7)}$$

Using eq 12–14, we find

$$\mathbf{T}_{i+1}\cdot\mathbf{B}_i = 2p_i(\mathbf{v}_{i+1} \times \mathbf{v}_i)\cdot\mathbf{v}_{i+2} \quad \text{(VI-8)}$$

The absolute magnitude of the vector triple product in eq VI-8 can be calculated by using the well-known relation

$$(\mathbf{A}\cdot\mathbf{B})^2 = |\mathbf{A}|^2|\mathbf{B}|^2 - |\mathbf{A} \times \mathbf{B}|^2 \quad \text{(VI-9)}$$

When applied to the product in question, eq VI-9 gives (after some algebra)

$$|\mathbf{v}_{i+2}\cdot(\mathbf{v}_{i+1} \times \mathbf{v}_i)| = |\mathbf{v}_{i+1}|\{|\mathbf{v}_{i+2}|^2|\mathbf{v}_i|^2 \sin^2 \theta_i - (\mathbf{v}_i\cdot\mathbf{v}_{i+2})^2 + 2(\mathbf{v}_i\cdot\mathbf{v}_{i+2})|\mathbf{v}_{i+2}||\mathbf{v}_i| \cos \theta_i \cos \theta_{i+1} - |\mathbf{v}_i|^2|\mathbf{v}_{i+2}|^2 \cos^2 \theta_{i+1}\}^{1/2} \quad \text{(VI-10)}$$

This can be evaluated by using eq VI-4.

It remains only to calculate the sign of $\mathbf{T}_{i+1}\cdot\hat{\mathbf{b}}_i$. From eq 9

$$u_i = \text{sgn } (\mathbf{B}_{i+1}\cdot\mathbf{N}_i) \quad \text{(VI-11)}$$

But

$$\mathbf{B}_{i+1}\cdot\mathbf{N}_i = (\mathbf{T}_{i+1} \times \mathbf{N}_{i+1})\cdot\mathbf{N}_i \quad \text{(VI-12)}$$

Using eq 12–15, we find

$$(\mathbf{T}_{i+1} \times \mathbf{N}_{i+1})\cdot\mathbf{N}_i = 2p_ip_{i+1}(\mathbf{v}_{i+2} \times \mathbf{v}_{i+1})\cdot(\mathbf{v}_i - \mathbf{v}_{i+1}) \quad \text{(VI-13)}$$

$$= 2p_ip_{i+1}(\mathbf{v}_{i+1} \times \mathbf{v}_i)\cdot\mathbf{v}_{i+2} \quad \text{(VI-14)}$$

$$= p_{i+1}(\mathbf{T}_{i+1}\cdot\mathbf{B}_i) \quad \text{(VI-15)}$$

where the last line follows from eq VI-8.

We thus have

$$u_i = p_{i+1} \text{ sgn } (\mathbf{T}_{i+1}\cdot\mathbf{B}_i) \quad \text{(VI-16)}$$

or

$$\text{sgn } (\mathbf{T}_{i+1}\cdot\mathbf{B}_i) = p_{i+1}u_i \quad \text{(VI-17)}$$

which is equivalent to 41. Equations VI-17 and VI-8 and the observation that

$$|\mathbf{B}_i| = |\mathbf{T}_i \times \mathbf{M}_i| \quad \text{(VI-18)}$$

$$= |\mathbf{T}_i||\mathbf{M}_i||\sin \eta_i| \quad \text{(VI-19)}$$

give eq IV-8.

Addendum

After submission of this paper, we received from Professor R. E. Dickerson preprints of new, high-resolution studies[18,19] on CYTI and CYTR and a comparison of these structures by superposition methods. It is interesting to note that many of the differences present in the 1977 structures have been eliminated by the new refinement, including those in the area of residue 24. The differences in the vicinity of residue 55 apparently remain and constitute a real difference between the oxidized and reduced structures.

References and Notes

(1) This work was supported by research grants from the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (Grant No. GM-14312) and from the National Science Foundation (Grant No. PCM79-20279).
(2) (a) NIH Postdoctoral Fellow, 1977–1978. (b) Author to whom requests for reprints should be addressed.
(3) Rackovsky, S.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 1168.
(4) Rackovsky, S.; Scheraga, H. A. *Macromolecules*, to be submitted.
(5) Flory, P. J. "Statistical Mechanics of Chain Molecules"; Wiley-Interscience: New York, 1969.
(6) Deisenhofer, J.; Steigemann, W. *Acta Crystallogr., Sect. B* **1975**, *31*, 238.
(7) Warme, P. K.; Tuttle, R. W.; Scheraga, H. A. *Comput. Programs Biomed.* **1972**, *2*, 248.
(8) Endres, G. F.; Swenson, M. K.; Scheraga, H. A. *Arch. Biochem. Biophys.* **1975**, *168*, 180.
(9) Rossmann, M. G.; Argos, P. *J. Mol. Biol.* **1976**, *105*, 75.
(10) Takano, T.; Kallai, O. B.; Swanson, R.; Dickerson, R. E. *J. Biol. Chem.* **1977**, *252*, 759.
(11) Takano, T.; Trus, B. L.; Mandel, N.; Mandel, G.; Kallai, O. B.; Swanson, R.; Dickerson, R. E. *J. Biol. Chem.* **1977**, *252*, 776.
(12) Moras, D.; Olsen, K. W.; Sabesan, M. N.; Beuhner, M.; Ford, G. C.; Rossman, M. G. *J. Biol. Chem.* **1975**, *250*, 9137.
(13) Adman, E. T.; Sieker, L. C.; Jensen, L. H. *J. Biol. Chem.* **1976**, *251*, 3801.
(14) Crippen, G. M. *J. Comput. Phys.* **1977**, *24*, 96.
(15) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Biochim. Biophys. Acta* **1973**, *303*, 211.
(16) Nishikawa, K.; Momany, F. A.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 797.
(17) Selby, S. M., Ed.; CRC Standard Mathematical Tables; Chemical Rubber Company: Cleveland, Ohio, 1972; p 358.
(18) Takano, T.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.*, in press.
(19) Takano, T.; Dickerson, R. E. In "Interaction between Iron and Proteins in Oxygen and Electron Transport"; Ho, C., Ed.; Elsevier/North-Holland: New York, in press.