Differential Geometry and Polymer Conformation. 1. Comparison of Protein Conformations[1a,b]

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

Department of Chemistry, Cornell University, Ithaca, New York 14853.
Received January 16, 1978

ABSTRACT: A representation of polymer backbone structure based on concepts of differential geometry is constructed. This representation reflects the influence of both short- and medium-range interactions. It is demonstrated that various types of ordered backbone structures are well characterized in terms of the differential-geometric representation. Bends, in particular, are classified in a very natural way in this picture. It is then shown that the differential-geometric representation provides a useful and objective method for comparing protein conformations. This method requires no optimization or model building and provides chain plots which indicate changes in folding in the neighborhood of *each residue* of the protein molecule.

I. Introduction

The comparison of protein conformation is of considerable current interest, both as a basis for the study of evolutionary relationships between proteins[3] and in connection with the physical chemistry of protein conformation.[4-9] Such investigations depend on the availability of an accurate criterion for judging the degree of similarity between two protein conformations. The index used in most recent studies[4-9] has been the root-mean-square deviation (RMSD) of interatomic distances between the two conformations. It is becoming increasingly clear, however, that this is an unsatisfactory criterion for comparing two conformations. For example, Hagler and Honig[10] generated conformations of the small protein Bovine Pancreatic Trypsin Inhibitor with low RMSD relative to the native conformation but with incorrect folding; as a consequence, they pointed out that an "acceptable" RMSD does not guarantee that the topology of the chain is represented correctly.

Another method which has been used[3] for comparison of conformations is that of optimal superposition of the two conformations. This shares the disadvantage that the resulting index of quality is a RMSD. It is also possible to compare distance maps,[7,8] which, in principle, contain all the necessary information about the conformation. They are, however, extremely difficult to relate to the actual folding of the protein. Evidently, a better method of comparison is required.

In considering the question of the comparison of two structures, a number of points should be made. First, it is unreasonable to expect that conformational similarity (or the lack thereof) can be represented adequately by a single number. Such a number can be only an average quantity, lacking in sufficient structural content to enable meaningful comparisons to be made.

Second, in comparing two conformations, side-chain atoms are of secondary importance. The primary requirement for conformational similarity is *backbone* similarity. Thus, the inclusion of side-chain data in the computation of parameters of similarity tends to obscure the real quality of the fit. Therefore, in seeking a method of comparison, we shall confine our attention to comparing backbone conformations. It should be noted, however, that the correct specification of backbone folding will severely limit the possibilities for side-chain conformation, and is therefore a more stringent criterion for molecular similarity than it at first seems.

Finally, it should be noted that the term "topology of the chain", as we have used it above, and as it is frequently used in the context of the present problem, is not appropriate. The presence of disulfide bridges, which introduce topological diversity to protein structures, can be regarded as a consequence of the correct juxtaposition of the appropriate cysteine residues, which is governed by the conformation of the un-crosslinked chain. Therefore, in discussing the backbone chain itself, we shall disregard disulfide bonds. With the neglect of the disulfide bonds, it follows that all single-unit protein molecules are simple unknotted curves and therefore are topologically equivalent. They can be compared, however, to space curves of different forms. Hence, an appropriate mathematical framework for the description of such curves is not topology but *differential geometry*.

In this paper, we shall develop a representation of polymer chain conformation based on concepts of differential geometry. We shall see that this description provides a straightforward method for comparing conformations. It also provides a local description of the chain which reflects both short- and medium-range order. It is therefore of interest in connection with the development of protein folding algorithms.

II. Differential Geometry of Polymer Chains

Classical differential geometry (CDG)[11,12] deals with continuous space curves. In order to describe such a curve without reference to an external coordinate system, a local reference frame is constructed at each point of the curve. The three unit vectors which comprise the local frame at a given point are the tangent, the principal normal, and the binormal. The limiting processes by which these vectors are determined at each point of the curve guarantee that the orientation of the local frame reflects the structure of the curve in the neighborhood of the point of interest.

Once the concept of the local frame is defined, it is possible to specify the curve completely in terms of two functions of the arc length s, $\kappa(s)$ and $\tau(s)$, the curvature and torsion. These functions are defined in terms of derivatives of the tangent and binormal vectors, respectively, with respect to arc length; i.e., they measure the rate of change of the direction of the tangent and binormal, respectively, when moving along a curve. The reader is referred to standard texts[11,12] on differential geometry for a full discussion of these concepts.

We wish to use similar ideas to develop a formalism for the description of polypeptide chains. We must therefore adapt methods that are used for continuous curves to make

0024-9297/78/2211-1168$01.00/0

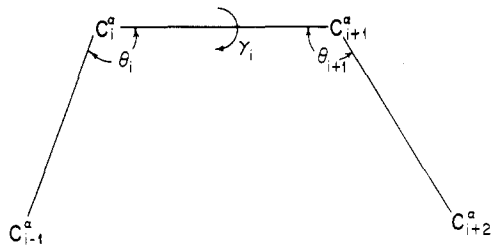

Figure 1. Virtual-bond representation of the polypeptide chain, showing the definitions of the virtual-bond angles θ_i and the virtual-bond dihedral angle for rotation γ_i.

them applicable to the treatment of an ordered set of discrete points in three-dimensional space.

As a preliminary, we define our view of the polypeptide chain. It will be most convenient to adopt the virtual-bond representation of the backbone, in which the C^{α} atoms of successive residues are considered to be connected by imaginary bonds (Figure 1). We shall assume that the peptide groups are in the planar trans conformation; hence, the virtual bonds are of uniform length (however, in using X-ray data on proteins, the observed virtual-bond distances will vary somewhat because of departures from the idealized geometry assumed here). The relation between the virtual-bond angles θ and bond dihedral angle γ and the traditional backbone dihedral angles ϕ and ψ has been given elsewhere.[13]

The first step in the formulation of a differential-geometric description of a polypeptide chain is the definition of the tangent vector at any point of the set, i.e., at any C^{α} atom of the chain. In the differential geometry of continuous curves, the tangent is defined as $d\mathbf{X}/ds$, where $\mathbf{X}$ is the position of a point on the curve and s is the arc length along the curve. (The direction of the tangent is specified by the sense of the arc length.) We therefore might have defined the tangent as a unit vector proportional to $\mathbf{r}_{i+1} - \mathbf{r}_i$, where $\mathbf{r}_i$ is the position of the ith point in the set of C^{α}'s. Indeed, this would lead to the standard virtual-bond representation based on the virtual-bond angle and virtual-bond dihedral angle. We shall choose an alternative definition of the tangent at C^{α}_i—one which includes information about the structure of the chain in the entire neighborhood of the point of interest, viz., the two neighboring points, $i - 1$ and $i + 1$. Our definition has the added advantage that the tangent line (but not its direction) remains constant if the direction of the chain is reversed. The resulting tangent also corresponds to an intuitive picture of the tangent at a point, in that it makes equal angles with the two virtual-bond vectors associated with the point i (in the assumed idealized geometry).

We therefore define the unit tangent vector at the point C^{α}_i by

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

The caret denotes a unit vector, and

$$\mathbf{v}_i = \mathbf{r}_i - \mathbf{r}_{i-1} \tag{2}$$

is the ith virtual-bond vector.

In order to define the unit principal normal at C^{α}_i, we must first define the osculating plane at that point. By analogy with CDG, we define this as the plane determined by the points $i - 1$, i, and $i + 1$. The unit principal normal at C^{α}_i is then defined, up to a sign, as the unit vector in the osculating plane which is perpendicular to the tangent. We shall determine the sign by the intuitively evident requirement that any chain segment lying in a plane have zero torsion. As will be seen below, the convention we now prescribe satisfies this requirement. We define an auxiliary unit vector $\hat{\mathbf{m}}_i$, along the principal normal, as

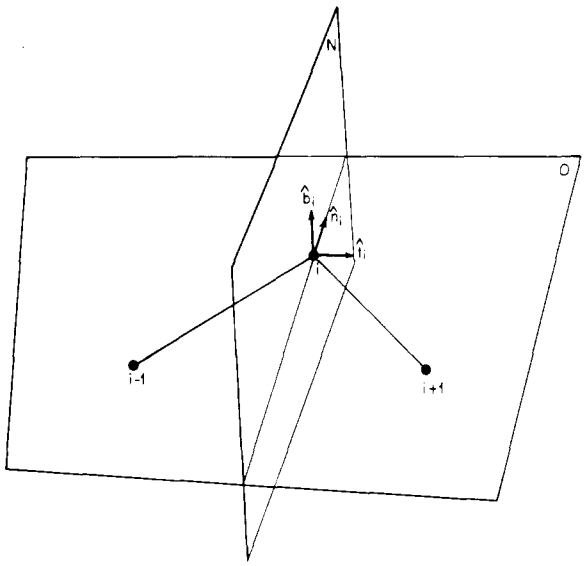

Figure 2. The local frame at C^{α}_i for the case $p_i = +1$. If $p_i = -1$, the directions of $\hat{\mathbf{n}}_i$ and $\hat{\mathbf{b}}_i$ are reversed. The osculating plane, defined by C^{α}_{i-1}, C^{α}_i, and C^{α}_{i+1}, is labeled O, and the normal plane N. It should be noted that $\hat{\mathbf{t}}_i$ and $\hat{\mathbf{n}}_i$ lie in the plane O.

$$\hat{\mathbf{m}}_i = \frac{\mathbf{v}_i - \mathbf{v}_{i+1}}{|\mathbf{v}_i - \mathbf{v}_{i+1}|} = \frac{2\mathbf{r}_i - \mathbf{r}_{i-1} - \mathbf{r}_{i+1}}{|2\mathbf{r}_i - \mathbf{r}_{i-1} - \mathbf{r}_{i+1}|} \tag{3}$$

We then take the unit normal at i as

$$\hat{\mathbf{n}}_2 = \hat{\mathbf{m}}_2 \qquad \text{for } i = 2 \tag{4}$$

(where $i = 2$ is the first C^{α} at which a local frame can be defined) and

$$\hat{\mathbf{n}}_i = p_i \hat{\mathbf{m}}_i \qquad \text{for } 2 < i < N \tag{5}$$

where N is the number of C^{α}'s in the chain, and

$$p_i = \frac{\hat{\mathbf{m}}_i \cdot \hat{\mathbf{n}}_{i-1}}{|\hat{\mathbf{m}}_i \cdot \hat{\mathbf{n}}_{i-1}|} = \text{sgn}\,(\hat{\mathbf{m}}_i \cdot \hat{\mathbf{n}}_{i-1}) \tag{6}$$

The quantity p_i is ± 1, depending on the sign of the projection, $\hat{\mathbf{m}}_i \cdot \hat{\mathbf{n}}_{i-1}$, of $\hat{\mathbf{m}}_i$ on $\hat{\mathbf{n}}_{i-1}$. For physical reasons, $\hat{\mathbf{m}}_i$ is never perpendicular to $\hat{\mathbf{n}}_{i-1}$; hence, p_i is never zero. It follows from the definitions in eq 1 and 3 that, for idealized geometry,

$$\hat{\mathbf{t}}_i \cdot \hat{\mathbf{n}}_i = 0 \tag{7}$$

It should be noted that local coordinate systems are defined at points $i = 2, 3, \ldots, N - 1$.

The unit binormal is now defined by

$$\hat{\mathbf{b}}_i = \hat{\mathbf{t}}_i \times \hat{\mathbf{n}}_i \tag{8}$$

and the construction of the local frame is complete (Figure 2).

It is now possible to define the curvature κ_i and torsion τ_i at any point. By analogy with the limiting process by which curvature is defined in CDG,[12] we write

$$\kappa_i = q_i \frac{\cos^{-1}(\hat{\mathbf{t}}_i \cdot \hat{\mathbf{t}}_{i+1})}{|\mathbf{v}_{i+1}|} = \frac{q_i \chi_i}{|\mathbf{v}_{i+1}|} \tag{9}$$

where χ_i is the angle between $\hat{\mathbf{t}}_i$ and $\hat{\mathbf{t}}_{i+1}$. There are two conventions in general use for the sign of the curvature. In one, the curvature is positive everywhere. The other allows the curvature to change sign, for example, at a point of inflection of a plane curve. We shall adopt an analogue of the second convention by setting

$$q_i = p_i \text{ sgn } [(\hat{\mathbf{t}}_i \times \hat{\mathbf{t}}_{i+1})\cdot\hat{\mathbf{b}}_{i+1}] \quad (10)$$

This definition, together with that of u_i of eq 12, will insure a unique description of the chain in terms of the set of $\{\kappa_i,\tau_i\}$, as will be seen below.

In a like manner, we define the torsion at i as[12]

$$\tau_i = u_i \frac{\cos^{-1}(\hat{\mathbf{b}}_i\cdot\hat{\mathbf{b}}_{i+1})}{|\mathbf{v}_{i+1}|} = \frac{u_i\gamma_i}{|\mathbf{v}_{i+1}|} \quad (11)$$

where γ_i is the dihedral angle defined in Figure 1. The sign of τ_i is defined by

$$u_i = \text{sgn } (\hat{\mathbf{b}}_{i+1}\cdot\hat{\mathbf{n}}_i) \quad (12)$$

It is readily seen that eq 4–6, 8, and 11 guarantee that the torsion of any planar zig-zag chain is zero, as stated previously. In section III, we shall discuss the relationship between (κ_i,τ_i) and parity (right or left handedness of the twist of the chain).

It is clear from the foregoing considerations that κ_i and τ_i depend on the coordinates of four C^α's (C^α_{i-1} – C^α_{i+2}), or alternatively on two sets of dihedral angles (ϕ_i,ψ_i; ϕ_{i+1},ψ_{i+1}). Therefore, κ_i and τ_i reflect both short-range (single-residue) and medium-range (i to $i+4$) interactions along the backbone. It should also be noted that a subunit of the chain, consisting of four successive C^α's, is the smallest length of backbone over which the chain can be said to be folded. This is because the positions of three C^α's serve only to define a plane. In order to indicate how the chain continues *out* of this plane, the coordinates of a fourth C^α are required. The (κ,τ) representation therefore relates directly to the *elementary unit* of chain folding.

The various angles involved in the virtual-bond and differential-geometric descriptions of the chain can be shown to be related by the following equation (see Appendix):

$$\cos\omega_i \cos\omega_{i+1} \cos\gamma_i = p_ip_{i+1}(\sin\omega_i \sin\omega_{i+1} - \cos\chi_i) \quad (13)$$

where p_i is defined in eq 6 and $\omega_i = \theta_i/2$ of Figure 1. Several points can be deduced from eq 13. First, it may be noted that this is an equation for ω_{i+1} in terms of ω_i, χ_i, and γ_i. It implies that, *for a given χ_i and γ_i (i.e., for a given curvature and torsion), there is no unique value for ω_i or ω_{i+1}; i.e., ω_i and ω_{i+1} can vary (within limits given by eq 13) for a given curvature and torsion.* The relation between ω_{i+1} and ω_i depends on χ_i and γ_i. It follows that, if well-defined values of χ_i and γ_i characterize given types of ordered backbone structure, then a range of ω_{i+1} and ω_i can be tolerated. In section III, we shall see that various types of ordered backbone structure do indeed have characteristic values of χ and γ.[14] We can therefore understand the observation[15] that no correlation could be found between backbone structure and virtual-bond angle.

Second, for given ω_i, χ_i, and γ_i, there are in general two solutions for ω_{i+1}, in the allowed range[13] of θ's. The question of uniqueness, therefore, is raised: is it possible to reconstruct the chain uniquely, given the complete set of $\{\kappa_i,\tau_i\}$? In fact, all ambiguity in reconstructing the chain is removed by the choices that we have made above for the signs of κ and τ. As we remark in the Appendix, the two solutions of eq 13 for given θ_i, χ_i, and γ_i can be denoted as $\theta_{i+1}^{(+)}$ and $\theta_{i+1}^{(-)}$. Then the sign convention of eq 10 leads to correspondence between $q_i = -1$ and $\theta_{i+1} = \theta_{i+1}^{(+)}$ and between $q_i = +1$ and $\theta_{i+1} = \theta_{i+1}^{(-)}$.

From the above remarks, it is clear that, if the set of $\{\kappa_i,\tau_i\}$ is given, it is necessary to specify one virtual-bond angle, say θ_2, the first along the backbone, in order to define the backbone coordinates uniquely; for, if ω_2 is given, ω_3 is determined, and so on down the chain.[16] The possibility exists that there might be classes of conformations, which we denote by $C(\{\kappa_i,\tau_i\};\theta_2)$, with identical curvature and torsion, but with different θ_2. Such similarities would not always be made apparent by conventional comparison techniques, which deal only with coordinates. In comparing two conformations of a given protein, it is improbable that pairs of this type will occur. The extremely complex and precise fit between different parts of the molecule make it unlikely that two overlap-free conformations of the same protein can exist with the same set of $\{\kappa_i,\tau_i\}$ and with different θ_2. However, such a possibility does exist when comparing conformations of two different protein molecules; in such a case, the values of θ_2 should be presented together with $\{\kappa_i,\tau_i\}$. It is intriguing to ask whether there exist sets of proteins which, in their native conformations, are not manifestly similar in coordinate space, but are nevertheless members of a class $C(\{\kappa_i,\tau_i\};\theta_2)$. We know of no such examples at present, but the idea is presented as an interesting speculation.

III. Identification of Backbone Structural Regions in (κ,τ) Space

Having outlined the differential-geometric description of the chain, we can now examine the distribution of the amino acid residues in (κ,τ) space, and attempt to correlate the values of $\{\kappa_i,\tau_i\}$ with the presence of different types of backbone structures. For this purpose, we have calculated values of $\{\kappa_i,\tau_i\}$ from the X-ray coordinates of a set of 13 proteins considered in a previous study.[17]

In Figure 3 we show the distribution of all residues in the sample plotted in intervals of 0.1 radian/Å. Before discussing this distribution, it is necessary to consider the relationship between the DG representation and parity (right or left handedness of local structure) in order to be able to specify the twist of the chain from a knowledge of (κ_i,τ_i).

We define an angle λ_i by

$$\lambda_i = \frac{\pi}{2} - \frac{\theta_i}{2} \quad (14)$$

where λ_i is the angle between $\mathbf{v}_{i+1}$ and $\hat{\mathbf{t}}_i$. Then we can distinguish two possible situations: $\lambda_i > \chi_i$ or $\lambda_i < \chi_i$. For each case, there are two possible signs for τ_i. These four possibilities are shown in Figure 4. For given values of

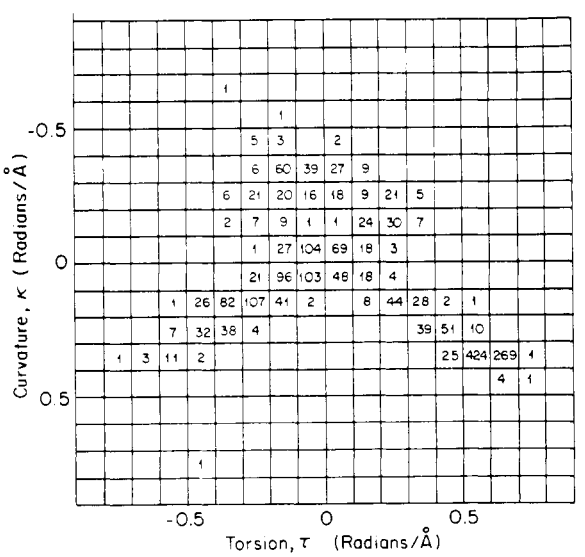

Figure 3. The distribution of residues in the sample of 13 proteins in (κ,τ) space, shown as the number of residues in each interval of (0.1,0.1).

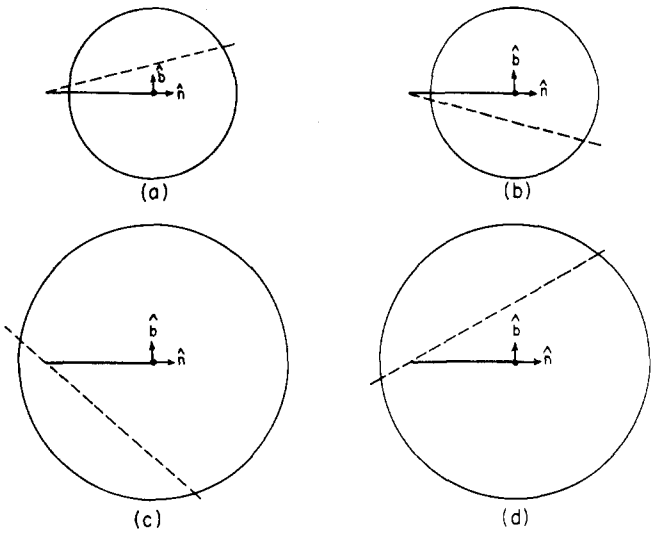

Figure 4. Views, looking backward along $\hat{\mathbf{t}}_i$ (represented by the central black dot) toward C^{α}_i. The heavy line is $\mathbf{v}_{i+1}$ (whose magnitude is 3.8 Å), terminating in C^{α}_{i+1}. C^{α}_{i-1} is directly behind C^{α}_{i+1}; the observer is therefore in O_i, the osculating plane associated with C^{α}_i. By the definition of curvature (eq 9), C^{α}_{i+2} must lie on a cone Γ_i with apex at C^{α}_i (and whose axis is $\hat{\mathbf{t}}_i$), which makes an angle χ_i with $\hat{\mathbf{t}}_i$. The circles are cross sections of this cone in a plane Q which is perpendicular to $\hat{\mathbf{t}}_i$ and includes C^{α}_{i+1}. C^{α}_{i+2} must also lie in a plane O_{i+1} determined by the torsion at C^{α}_i. The dashed lines represent the intersection of this plane with Q. These two conditions, together with the length of the virtual bond, limit the position of C^{α}_{i+2} to two possibilities. These are two points, one on each of the two straight lines which together form the intersection of Γ_i and O_{i+1}. By definition, $\hat{\mathbf{t}}_{i+1}$ lies on Γ_i. It therefore can be seen by inspection that the sign convention of eq 10 distinguishes between these two points. By imagining the various possible positions of C^{α}_{i+2} for the four cases shown, and by making reference to molecular models, the results of Table I are easily confirmed. It should be noted that, in (a) and (b), $\chi_i < \lambda_i$; in (c) and (d), $\chi_i > \lambda_i$. In (a) and (d), τ_i is negative; in (b) and (c), τ_i is positive.

χ_i and τ_i, there are two possible directions for $\mathbf{v}_{i+2}$. These are characterized by different signs for κ_i (as pointed out above), corresponding to the smaller or larger value of θ_{i+1}. Inspection of Figure 4, with the aid of a molecular model, reveals that the parity of the structure of the chain at a given residue is related to the signs of κ_i and τ_i as shown in Table I. From eq 14 and the known limits[13] on θ, it is seen that $17.5° < \lambda < 52.5°$. From this and the definition of κ_i (eq 9), together with the length of a virtual bond (3.8 Å), the following correspondences can be made:

$$\text{if } |\kappa_i| < 0.08, \text{ then } \chi_i < \lambda_i$$

and

$$\text{if } |\kappa_i| > 0.24, \text{ then } \chi_i > \lambda_i \qquad (15)$$

For the region $0.08 < |\kappa_i| < 0.24$, the relationship between χ_i and λ_i is indeterminate without additional information (i.e., the value of θ_i). From Table I, we see that the parity is independent of the relation between χ_i and λ_i for $\kappa_i < 0$ ($q_i = -1$); thus, this indeterminacy arises only for the case $\kappa_i > 0$ ($q_i = +1$) (see Figure 5). Thus, the parity at most residues can be determined by inspection of their values of (κ_i,τ_i), and this information is used in identifying the backbone structures that occur in various regions of (κ_i,τ_i) space.

Instead of using the composite distribution of Figure 3, we consider the different types of ordered backbone structure (viz., extended structures and α helices) separately. In order to locate the regions in (κ,τ) space corresponding to these ordered structures, the assignments

Table I
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	L	R
−1	L	R	−1	L	R

[a] As shown in section II, 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) (see Figure 4), 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.

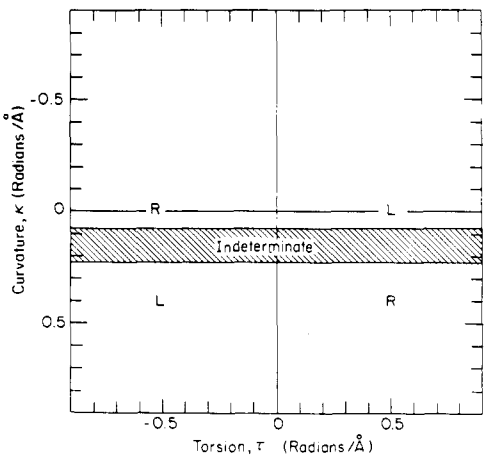

Figure 5. The regions of different parity in (κ,τ) space, determined by Table I together with eq 15. In the region marked "indeterminate", parity cannot be determined solely from a knoweledge of (κ,τ).

of conformational states by Maxfield and Scheraga[18] were used.

In Figure 6, the values of (κ_i,τ_i) for 82 interior residues from α-helical regions are shown (the interior residues comprise all but the first and last residues of each helical segment). It can be seen that the right-handed α-helical region is very localized and well defined at $0.3 < \kappa_i < 0.4$, $0.5 < \tau_i < 0.7$. The three residues with $-0.3 > \kappa_i > -0.4$ are probably mislocated because of inaccuracies in the X-ray data, which tend to cause reversals in the sign of κ_i and changes in the values of τ_i. The effect of experimental error on the signs of κ_i and the values of τ_i will be considered further in section IV.

This behavior is to be contrasted with that exhibited by the values of (κ_i,τ_i) for 173 interior residues from extended structures, shown in Figure 7. There is a very wide variation in (κ,τ) for extended structures, which permits the chain to be both left-handed and right-handed at specific residues (although there is a clear preponderance of right handedness at those residues which fall in the determinate region). The breadth and variation of (κ,τ) in extended structures undoubtedly reflects the wide limits used to define the extended region in (ϕ,ψ) space.[18] Reference to Figure 5 reveals that residues in this region are mostly right handed, which reflects the well-known[19] right-handed twist found in β-pleated sheets.

We consider next the question as to whether bends in proteins are well-characterized in (κ,τ) space. In order to investigate this question, we use the data of Isogai et al.[20] to locate 82 single bends in eight of the thirteen proteins in our sample and classify them by bend type.[21] The result of this classification is exhibited in Figure 8. A bend is considered to be composed of the C^{α}'s at positions $i - 1$,

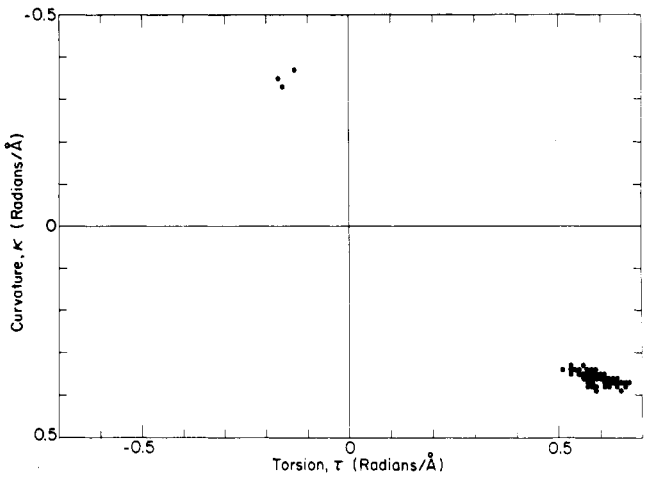

Figure 6. Values of (κ_i,τ_i) for 82 residues in the interior of α helices.

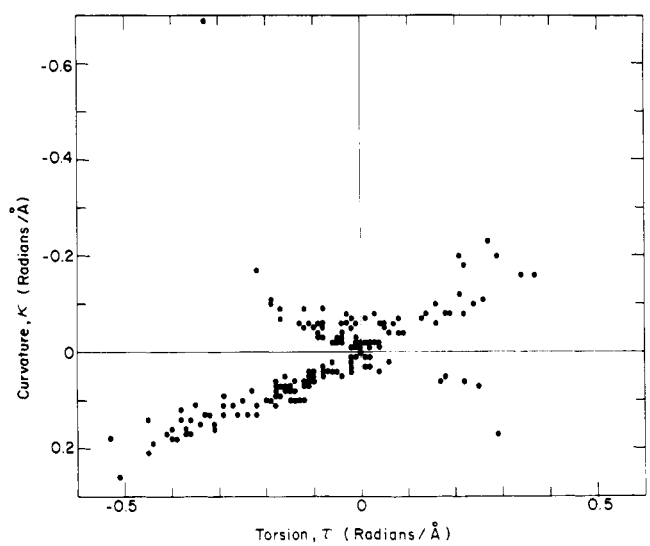

Figure 7. Values of (κ_i,τ_i) for 173 residues in the interior of extended structures.

i, $i+1$, and $i+2$, and (κ_i,τ_i) is plotted. The values very clearly fall into two principal regions, delineated roughly by $-0.4 < \kappa_i < -0.3$, $-0.2 < \tau_i < 0.1$ and $0.3 < \kappa_i < 0.4$, $0.5 < \tau_i < 0.7$. We shall refer to these as low-torsion and high-torsion bends, respectively. The high-τ-bend region corresponds essentially to the α-helix region (see Figure 6), and it is therefore understandable that most of the points in this region represent type I bends (for which residue i is in the right-handed α-helical conformation[21]). It should be noted that some type III and type IV bends are also found in this region. There is a small mirror-image region with negative torsion in the left-handed α-helical region; it is composed primarily of type III′ bends, which are expected to fall in this area since both central residues of a type III′ bend are in the left-handed α-helical conformation.[21] The low-τ region contains bends of all types. The ratio of right- to left-handed bends in this region (see Figure 5) is about 2:1. There is one type VI bend[21] at about (−0.7, −0.3), representing a cis peptide group of a proline residue at position $i+1$.

The negative curvature and low torsion of the low-τ bends indicate that they are characterized by a relatively large value of θ_i and an essentially planar structure. The high-τ bends have lower values of θ_i and are twisted out of the plane. It would appear that the differential-geometric approach provides a somewhat more natural classification of bends than the approach based on (ϕ,ψ)

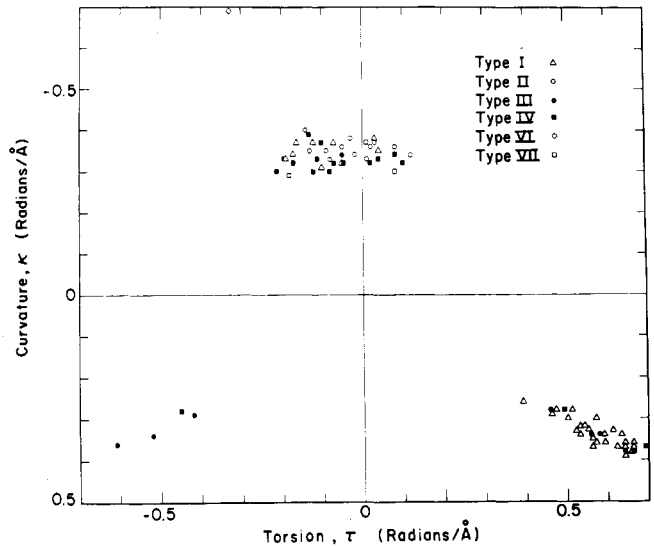

Figure 8. Values of (κ_i,τ_i) for residue i in 82 bends defined by C^{α}_{i-1}–C^{α}_{i+2}.

values. The overwhelming majority of bends fall into only two classes.

It was of interest to determine if the values of $(\kappa_{i+1},\tau_{i+1})$ of bends also fall into a natural classification. These values were therefore plotted for the same set of bends. The results, shown in Figure 9, suggest that the values of $(\kappa_{i+1},\tau_{i+1})$ do not characterize bend structures very well. Perhaps this is to be expected, since these values reflect the position of C^{α}_{i+3}, and therefore represent the relationship between the bend itself and the structure which continues the chain. Only one marked regularity emerges from the figure, viz., that $\tau_{i+1} > 0$ preponderates for type II bends.

A comparison of Figure 3 with Figures 6–8 indicates that the region with $0.1 < |\kappa_i| < 0.3$ is primarily the province of residues in nonregular structures. The relative absence of regular structures from this region (Figures 6–8) supports this conclusion. Although there is some spillover of extended structures into this region, they are undoubtedly in the minority.

IV. Comparison of Conformations

In order to investigate the utility of the differential-geometric representation for the comparison of conformations, it was decided to compare two sets of X-ray coordinates of the same protein, which had been shown independently to be folded similarly. In this way, it is possible to obtain an estimate of the limits which can be considered to represent "identity" of conformation, and of the sensitivity of the method to conformational differences. For this purpose, we chose cytochrome c, which can be crystallized in both oxidized and reduced forms.[22,23] In the oxidized form, the crystal has two molecules per unit cell; these are designated as the inner and outer molecule of the cell. It has been demonstrated[23] qualitatively that the folding of the backbone of all three forms is similar.

We therefore calculated $\{\kappa_i,\tau_i\}$ for cytochrome c oxidized inner (CYTI) and cytochrome c reduced (CYTR), and made plots of the differences between their curvatures and torsions $(\kappa_I - \kappa_R, \tau_I - \tau_R)$ as a function of position in the chain (see Figure 10). If the conformations are identical, these difference functions should be equal to zero at every point. Inspection of the chain plots reveals long stretches of good agreement, punctuated by points (values of the residue number) at which there are larger differences in κ and τ. It should be noted that the deviations in κ and

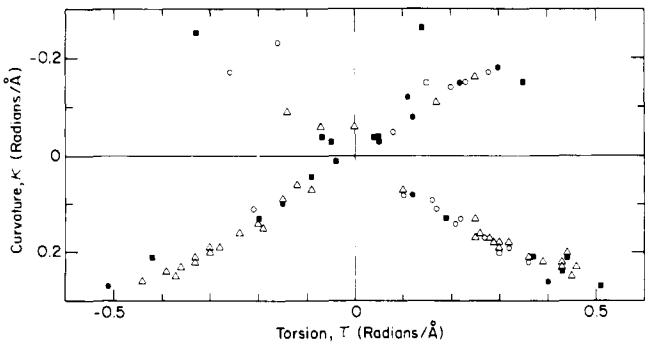

Figure 9. Values of $(\kappa_{i+1},\tau_{i+1})$ for the same set of bends as in Figure 8.

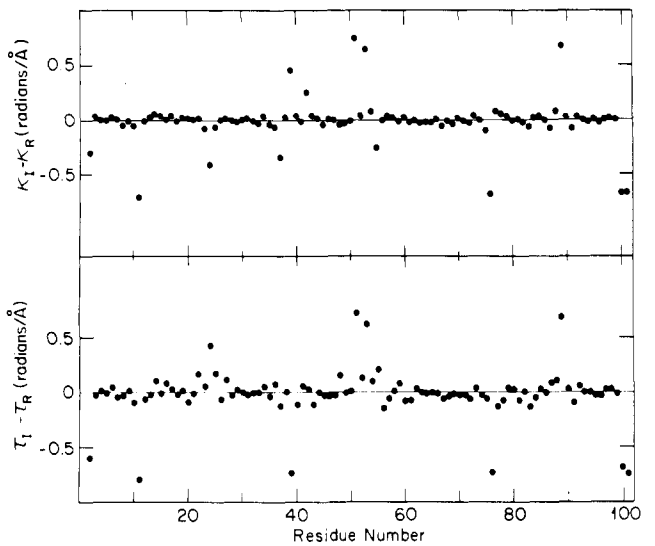

Figure 10. Comparison of CYTI and CYTR (see text). In this plot, algebraic values of κ_i and τ_i were used.

those in τ are completely correlated; i.e., they occur at the same residue. These are due to differences in local structure at these points, which are certainly to be expected, given the experimental uncertainties in X-ray crystallography. It is also possible that these may be real differences in conformation. The degree of scatter about zero in the regions of good agreement is somewhat larger for the $\tau_I - \tau_R$ plot than for the $\kappa_I - \kappa_R$ plot. This suggests that τ_i is somewhat more sensitive than κ_i to atomic displacements.

In order to understand the deviations in these chain plots more fully, we plotted the differences in absolute values of the curvature and torsion ($|\kappa_I| - |\kappa_R|$, $|\tau_I| - |\tau_R|$) in Figure 11. It can be seen that the deviations in curvature are thereby almost eliminated, whereas those in torsion remain. This indicates that differences in C^α positions tend to affect the value of τ and the sign of κ.

In future work, it is planned to calibrate the method by studying conformations with systematically introduced differences. It is readily seen from the present results, however, that the differential-geometric method is quite sensitive to changes in local folding, and is therefore useful for comparison of conformations.

The advantage of using $\{\kappa_i,\tau_i\}$ to compare protein structures, rather than the more traditional methods (e.g., optimal superposition of two structures), can be illustrated readily. Consider two conformations which are identical, except for a rotation about one bond in the middle of the chain of one. The effect of this rotation will be to produce a molecule with two domains, each of which is identical to the corresponding domain of the other conformation.

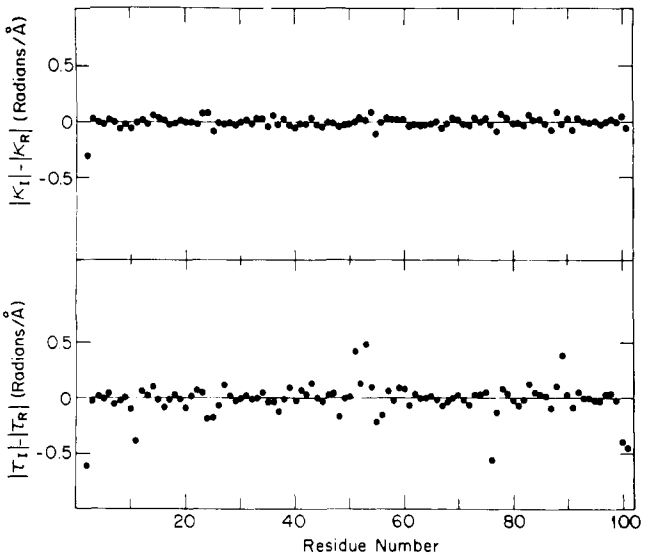

Figure 11. The same comparison as in Figure 10, but using absolute values of κ_i and τ_i.

We now imagine that we are trying to compare them by the method of superposition. We therefore rotate one rigidly relative to the other until the optimum overall fit is achieved. We then calculate the root-mean-square deviation between the two conformations. This will be quite high, despite the identity of the two domains, because of the impossibility of superimposing the two molecules. However, plots of (κ_i,τ_i) as a function of position in the chain will show complete identity except at the one residue where the rotation was performed, and perhaps in its immediate neighborhood. Furthermore, there is no need for any optimization or model building. The method is sensitive to changes in folding, and provides chain plots which illustrate these changes at *every residue* of the backbone, rather than providing only a single number. Differences between conformations therefore can be pinpointed, and their precise nature identified.

V. Conclusion

A description of polymer chains based on differential-geometric considerations has been formulated and applied to the backbones of protein molecules. The resulting representation reflects the short- and medium-range structure at each residue, and provides well-defined descriptions of the various types of ordered backbone structures. Its usefulness for comparing protein conformations has been demonstrated.

It is of interest to point out that the differential-geometric representation is likely to be useful in protein folding algorithms [applied in (κ,τ) space], because of its unique ability to incorporate medium-range interactions in a straightforward manner. In this connection, we are now studying the differential-geometric properties of the various amino acids, and the relationship between side-chain conformation and differential geometry.

Acknowledgment. We thank Dr. George Némethy for helpful discussion and for providing the data on bends on which Figures 8 and 9 are based.

Appendix. Derivation of Equation 13

We begin by noting that the vector $\mathbf{w}_i$, defined by

$$\mathbf{w}_i = p_i(\mathbf{v}_{i+1} \times \mathbf{v}_i) \tag{A1}$$

(with p_i defined in eq 6) is parallel to $\hat{\mathbf{b}}_i$, so that

$$\mathbf{w}_i \cdot \mathbf{w}_{i+1} = |\mathbf{w}_i||\mathbf{w}_{i+1}| \cos \gamma_i \quad \text{(A2)}$$

($\mathbf{v}_i$ is the virtual-bond vector defined in eq 2). We also define the vector $\mathbf{l}_i$ by

$$\mathbf{l}_i = \mathbf{v}_i + \mathbf{v}_{i+1} \quad \text{(A3)}$$

By definition, $\mathbf{l}_i$ is parallel to $\hat{\mathbf{t}}_i$, so that

$$\mathbf{l}_i \cdot \mathbf{l}_{i+1} = |\mathbf{l}_i||\mathbf{l}_{i+1}| \cos \chi_i \quad \text{(A4)}$$

From the definition in eq A1, we can write

$$\mathbf{w}_i \cdot \mathbf{w}_{i+1} = p_i p_{i+1} (\mathbf{v}_{i+1} \times \mathbf{v}_i) \cdot (\mathbf{v}_{i+2} \times \mathbf{v}_{i+1}) \quad \text{(A5)}$$

Equation A3 can be rearranged to

$$\mathbf{v}_i = \mathbf{l}_i - \mathbf{v}_{i+1} \quad \text{(A6)}$$

Using eq A2, A5, and A6, we obtain

$$p_i p_{i+1} (\mathbf{v}_{i+1} \times \mathbf{l}_i) \cdot (\mathbf{l}_{i+1} \times \mathbf{v}_{i+1}) = |\mathbf{w}_i||\mathbf{w}_{i+1}| \cos \gamma_i \quad \text{(A7)}$$

But

$$|\mathbf{w}_i| = |\mathbf{v}|^2 \sin (\pi - \theta_i) = |\mathbf{v}|^2 \sin \theta_i \quad \text{(A8)}$$

where $|\mathbf{v}|$ is the length of a virtual bond, and is a constant equal to 3.8 Å, and θ_i is the virtual bond angle defined in Figure 1.

The dot product on the left-hand side of eq A7 can be expanded, using standard vector identities. The result, together with eq A4 and A8, gives

$$|\mathbf{v}|^4 \sin \theta_i \sin \theta_{i+1} \cos \gamma_i = p_i p_{i+1} |\mathbf{v}|^2 |\mathbf{l}_i||\mathbf{l}_{i+1}| \left[\sin \frac{\theta_{i+1}}{2} \sin \frac{\theta_i}{2} - \cos \chi_i \right] \quad \text{(A9)}$$

From the law of cosines, we have

$$|\mathbf{l}_i|^2 = 2|\mathbf{v}|^2 (1 - \cos \theta_i) \quad \text{(A10)}$$

Insertion of eq A10 into eq A9, and the use of standard trigonometric identities, gives

$$\cos \omega_i \cos \omega_{i+1} \cos \gamma_i = p_i p_{i+1} [\sin \omega_i \sin \omega_{i+1} - \cos \chi_i] \quad \text{(A11)}$$

where $\omega_i = \theta_i/2$. Equation A11 is eq 13.

This equation can be solved for ω_{i+1}. The result is

$$\sin \omega_{i+1} = [\sin \omega_i \cos \chi_i \pm \cos \gamma_i \cos \omega_i (\cos^2 \gamma_i \cos^2 \omega_i - \cos^2 \chi_i + \sin^2 \omega_i)^{1/2}] / [\sin^2 \omega_i + \cos^2 \gamma_i \cos^2 \omega_i] \quad \text{(A12)}$$

The two solutions can be designated as $\omega_{i+1}^{(+)}$ and $\omega_{i+1}^{(-)}$, in correspondence with the sign chosen in eq A12. It is then the case (and numerically demonstrable) that $q_i = -1$ corresponds to $\omega_{i+1} = \omega_{i+1}^{(+)}$ and $q_i = +1$ corresponds to $\omega_{i+1} = \omega_{i+1}^{(-)}$. The virtual bond angle is then

$$\theta_{i+1} = 2\omega_{i+1}^{(\pm)} \quad \text{(A13)}$$

References and Notes

(1) (a) This work was supported by research grants from the National Science Foundation (Grant No. PCM75-08691) and from the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (Grant No. GM-14312). (b) Presented before the Division of Biological Chemistry at the 176th meeting of the American Chemical Society, Miami, Beach, Fla., September 1978.
(2) (a) NIH Postdoctoral Fellow, 1977–78; (b) author to whom requests for reprints should be addressed.
(3) M. G. Rossmann and P. Argos, *J. Mol. Biol.*, **105**, 75 (1976).
(4) P. K. Warme and H. A. Scheraga, *Biochemistry*, **13**, 757 (1974).
(5) P. K. Warme, F. A. Momany, S. V. Rumball, R. W. Tuttle, and H. A. Scheraga, *Biochemistry*, **13**, 768 (1974).
(6) M. Levitt and A. Warshel, *Nature (London)*, **253**, 694 (1975).
(7) M. Levitt, *J. Mol. Biol.*, **104**, 59 (1976).
(8) I. D. Kuntz, G. M. Crippen, P. A. Kollman, and D. Kimelman, *J. Mol. Biol.*, **106**, 983 (1976).
(9) M. K. Swenson, A. W. Burgess, and H. A. Scheraga in "Frontiers in Physico-Chemical Biology", B. Pullman, Ed., Academic Press, New York, in press [Paris (1977)].
(10) A. T. Hagler and B. Honig, *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 554 (1978).
(11) D. J. Struick, "Differential Geometry", 2nd ed., Addison-Wesley, Reading, Mass., 1961.
(12) E. P. Lane, "Metric Differential Geometry of Curves and Surfaces", University of Chicago Press, Chicago, Ill., 1940.
(13) K. Nishikawa, F. A. Momany, and H. A. Scheraga, *Macromolecules*, **7**, 797 (1974).
(14) It is appropriate to observe that previous correlations have been made between γ and various types of ordered backbone structure. See, for example, ref 13 and 15. While this work was in progress, a paper appeared which dealt with χ, among other parameters [G. M. Crippen and I. D. Kuntz, *J. Theor. Biol.*, **66**, 47 (1977)]. None of these papers deals with the construction of a complete description of the backbone, which requires not only *both* χ and γ, but also their associated signs.
(15) K. Srinivasan, R. Balasubramanian, and S. S. Rajan, *J. Theor. Biol.*, **67**, 299 (1977).
(16) It is appropriate to note at this point that the conversion from a set of $\{\kappa_i, \tau_i\}$ and a value of θ_2 to a set of C^α coordinates is a computationally rapid and straightforward procedure. In preliminary tests on Bovine Pancreatic Trypsin Inhibitor, use of the experimental values of $\{\kappa_i, \tau_i\}$ and θ_2 led to a set of C^α coordinates which were essentially indistinguishable from the experimental X-ray coordinates. This and other numerical aspects of the differential-geometric representation will be discussed in greater detail in a future publication.
(17) S. Rackovsky and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5248 (1977).
(18) F. R. Maxfield and H. A. Scheraga, *Biochemistry*, **15**, 5138 (1976).
(19) C. Chothia, *J. Mol. Biol.*, **75**, 295 (1973).
(20) Y. Isogai, G. Némethy, and H. A. Scheraga, to be published.
(21) P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Biochim. Biophys. Acta*, **303**, 211 (1973).
(22) T. Takano, O. B. Kallai, R. Swanson, and R. E. Dickerson, *J. Biol. Chem.*, **252**, 759 (1977).
(23) T. Takano, B. L. Trus, N. Mandel, G. Mandel, O. B. Kallai, R. Swanson, and R. E. Dickerson, *J. Biol. Chem.*, **252**, 776 (1977).