functional chelating agents will probably continue to be clinical use of radioactive metal ions bound to proteins, particularly monoclonal antibodies. The ability to localize metal complexes at particular sites is also likely to have other uses in biochemistry and medicine.

The applications of unattached chelates, especially in the study of fundamental physical properties of biological molecules and supramolecular complexes, are just beginning to be explored. The existence of a diverse collection of chelating agents suggests a number of experiments in which the physical properties of chelated metal ions can be used to probe biological systems.

We owe a great debt to those students and colleagues who are named among the references. This work was supported mainly by research grants CA16861 and GM25909 and by research career development award CA00462 to C.F.M. from the National Institutes of Health, and by research instrumentation grant CHE 81-14966 to the UCD Chemistry Department from the National Science Foundation. T.W. acknowledges support from NIH Molecular and Cellular Biology Training Grant GM07377 and from an Earl C. Anthony Graduate Fellowship.

Differential Geometry and Protein Folding

S. RACKOVSKY and H. A. SCHERAGA*

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853

Received June 24, 1983 (Revised Manuscript Received January 26, 1984)

Proteins are biologically produced, specific-sequence copolymers whose monomer units are the 20 naturally occurring amino acids. They are distinguished from synthetic polymers by their ability to undergo a reversible thermally or chemically induced transition from an unfolded, biologically inactive form to a native form that executes, at most, small fluctuations about a well-defined conformation and that exhibits full biological activity.[1] This "native" conformation is so sharply defined that many proteins can be crystallized and their structures elucidated by X-ray diffraction methods. Both experimental evidence and theoretical considerations suggest that this refolding is not the result of a random search by the molecule but rather is an efficient directed process that depends on a sequence of nucleation steps for the rapid attainment of the correct conformation. In this picture, the imposition of renaturing conditions leads to the formation of regions of structure whose frequency and/or amplitude of fluctuation are substantially less than those of the remainder of the chain. These nuclei either interact with one another or cause other regions of the chain to interact in a manner that leads to correct folding in a minimal number of steps.

This mechanism implies the existence of a hierarchy of time and length scales that characterize the folding process. It further suggests that these two scales are correlated, i.e., those structures that form in the shortest times are those characterized by the shortest length scale, and larger structures are formed at later times in the folding process.

Inspection of X-ray structures of proteins indeed reveals the presence of certain characteristic structures. These characteristic structural features of proteins[2] can be classified as either of undetermined length scale or defined on a particular scale. Right-handed α-helices and extended strands are not limited in their length by any structural factors internal to themselves. Bends, on the other hand, are defined by the relative placement of precisely four α-carbons, so that they can be regarded as existing on the four-C^{α} length scale along the backbone. The same may be true of one of the two structural elements that together constitute a β-bulge.[2]

The observation of characteristic structures in folded proteins and the postulated folding mechanism together raise a basic question: What can be learned about protein folding by a systematic study of known protein structures on successive length scales? The goal of the work that forms the subject of this Account is to address this question.

The Differential-Geometric Representation

The primary tool for the study of protein structure on a given length scale is a representation or mathematical description that functions on that scale. This is, in principle, only a matter of convenience, since *any* representation of molecular structure contains information about structure on the length scale of interest. It should be emphasized, however, that the extraction of relevant information from an inappropriate representation can be a tedious procedure, which makes an intuitive picture of the data difficult to obtain. Thus, the distance-matrix representation, in which the structure of the molecule is specified by giving the distances between all pairs of atoms, is best suited to

S. Rackovsky, a native of Omaha, NB, received his Ph.D. degree in 1972 from the Massachusetts Institute of Technology. Thereafter, he was a research associate at MIT for a year, a postdoctoral fellow at McGill University for 2 years, and at Cornell University for 5 years; at Cornell he held NIH and Todd postdoctoral fellowships. After 2 years with the Xerox Corp., he joined the Department of Physics and Astronomy at the University of Rochester, where he is now a Visiting Assistant Professor. His research interests are in the chemical physics of proteins and the electronic properties of condensed systems.

H. A. Scheraga was born in Brooklyn, NY, in 1921. He attended The City College of New York, where he received his B.S. degree, and went on to graduate work at Duke University, receiving the Ph.D. in 1946 and, in 1961, an ScD. (Hon). Following postdoctoral work at Harvard Medical School, he joined the faculty at Cornell University, where he is Todd Professor of Chemistry. His research interests are in the physical chemistry of proteins and other macromolecules, chemistry of blood clotting, and structure of water and dilute aqueous solutions.

(1) Anfinsen, C. B.; Scheraga, H. A. *Adv. Protein Chem.* **1975**, *29*, 205.
(2) Richardson, J. S. *Adv. Protein Chem.* **1981**, *34*, 167.

0001-4842/84/0117-0209$01.50/0

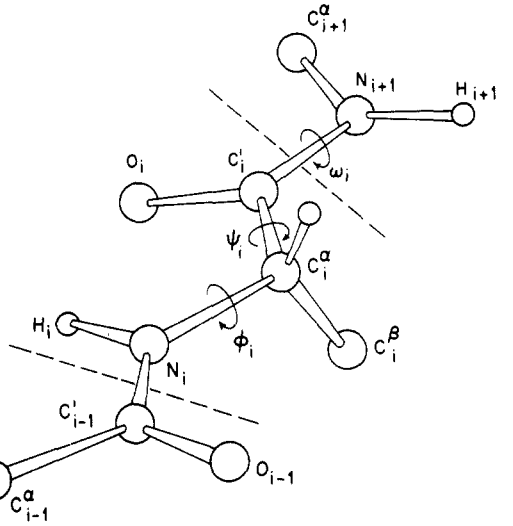

Figure 1. Perspective drawing of a section of a polypeptide chain, representing *two peptide units* and the attachment of a side chain in the L-amino acid configuration. A *residue* is shown enclosed by dashed lines. The figure indicates the standard IUPAC/IUB notation for atoms and dihedral angles.

the study of long-range interactions and other properties in which chain connectivity is not of primary interest.[3] The (ϕ,ψ) representation, in which one specifies the dihedral angles for rotation about the two backbone bonds of each residue (Figure 1), gives a detailed picture of the backbone atomic arrangement within each residue and is most suited for studying very local structure. While both representations contain the same information—the relative positions of all the atoms in the structure—it is difficult to visualize long-range structure by using (ϕ,ψ) data or single-residue conformations from a set of distances that extend beyond a single residue.

These two representations, which are those in common use, represent the two extremes of length scale of interest. The differential-geometric (DG) representation is a first step in the process of filling the gap in this spectrum of length scales. It is designed to operate on the four-C^{α} length scale, which is the next length scale above that treated by the (ϕ,ψ) representation. This is readily seen, since the positions of four successive C^{α}'s are controlled by the values of two neighboring sets of (ϕ,ψ). It is also clear that, in the virtual-bond backbone, in which the actual backbone is represented by bonds connecting successive α-carbons, a four-C^{α} unit is the smallest segment of backbone that can be said to be folded. This is because three points (i.e., the coordinates of three C^{α}'s) determine a plane; in order to describe how the backbone progresses through space, one must give the coordinates of one more point (α-carbon) (see Figure 2).

The DG representation is based on an analogy between the virtual-bond backbone and a curve in three-dimensional space. The former can be thought of as a discrete analogue of the latter, in which the continuous curve is replaced by a set of points connected by straight lines. In ideal geometry, the virtual bonds of the protein backbone are all of the same length—3.8 Å. In the same way that two parameters, $\kappa(s)$ and $\tau(s)$ (the curvature and torsion as functions of

(3) Crippen, G. M. *J. Med. Chem.* **1981**, *24*, 198.

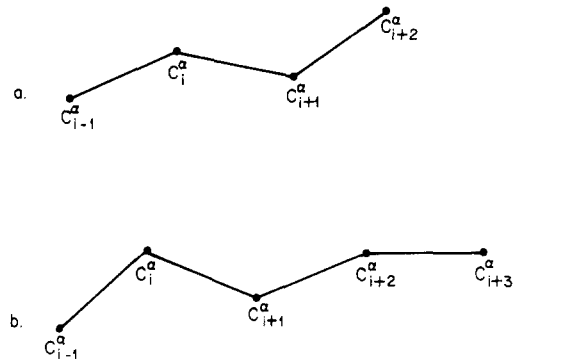

Figure 2. (a) Four-C^{α} unit whose conformation is described by (κ_i,τ_i). (b) Five-C^{α} unit whose conformation is described by $(\kappa_i,\tau_i,\kappa_{i+1},\tau_{i+1})$. These are representations of virtual-bond structures that, in general, are not planar.

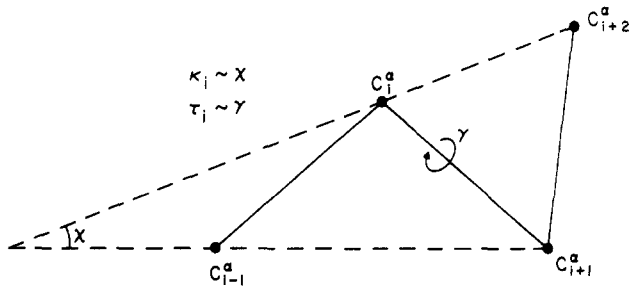

Figure 3. Definition of κ_i and τ_i. The curvature is proportional to the angle χ between the $C^{\alpha}_{i-1} \rightarrow C^{\alpha}_{i+1}$ and $C^{\alpha}_i \rightarrow C^{\alpha}_{i+2}$ vectors. The torsion is proportional to the dihedral angle γ for rotation about the $C^{\alpha}_iC^{\alpha}_{i+1}$ virtual bond. In general, the four-C^{α} unit is not planar.

arc length s), are sufficient to describe completely the course of a continuous curve in three-dimensional space, we define κ_i and τ_i (the curvature and torsion at C^{α}_i) to describe the conformation of the virtual-bond backbone. As mentioned above, the DG representation operates on the four-C^{α} length scale, in the sense that the coordinates of four α-carbons are used to define one (κ_i,τ_i) pair. By convention, (κ_i,τ_i) are determined by the coordinates of C^{α}_{i-1}, C^{α}_i, C^{α}_{i+1}, and C^{α}_{i+2}. Therefore, values of (κ_i,τ_i) exist for $i = 2, 3, \cdots, N-2$, for an N-residue protein. This is illustrated in Figure 2.

A feeling for the structural meaning of κ_i and τ_i can be obtained from Figure 3. κ_i is proportional to the angle between the vector connecting C^{α}_{i-1} and C^{α}_{i+1} and that connecting C^{α}_i and C^{α}_{i+2}. Since the former vector defines the chain direction at C^{α}_i and the latter the direction at C^{α}_{i+1}, κ_i describes the way in which the chain changes direction in the four-C^{α} unit under examination. The torsion, τ_i, is proportional to γ_i, the dihedral angle for rotation about the central virtual bond of the four-C^{α} unit. Thus, τ_i describes the way that the backbone twists in the four-C^{α} unit. The reader is referred to ref 4 for mathematical details (a discussion of which we have minimized in this general exposition).

Characteristic Structural Features

The DG representation can be used to study the structure of the above-mentioned characteristic architectural features of proteins on the four-C^{α} length scale.[5] As a first step, the regions of the (κ,τ) plane occupied

(4) Rackovsky, S.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 1168.
(5) Rackovsky, S.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 1259.

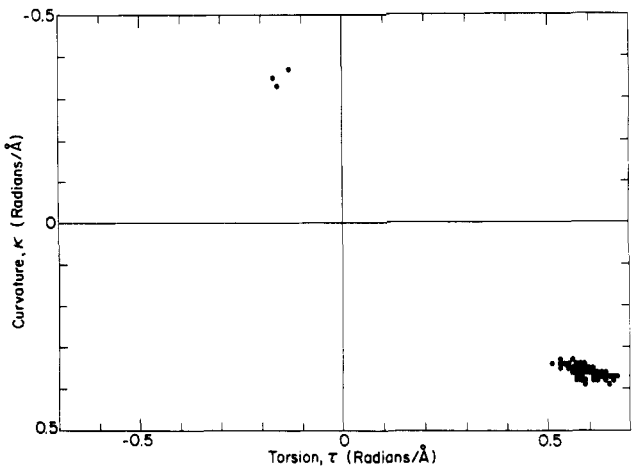

Figure 4. Values of (κ_i,τ_i) for 82 residues in the interior of α-helices. The data in this and in Figures 5 and 6 were obtained from a sample of protein molecules with known X-ray structures, in which the ordered backbone structures were identified by independent (non-DG) means.

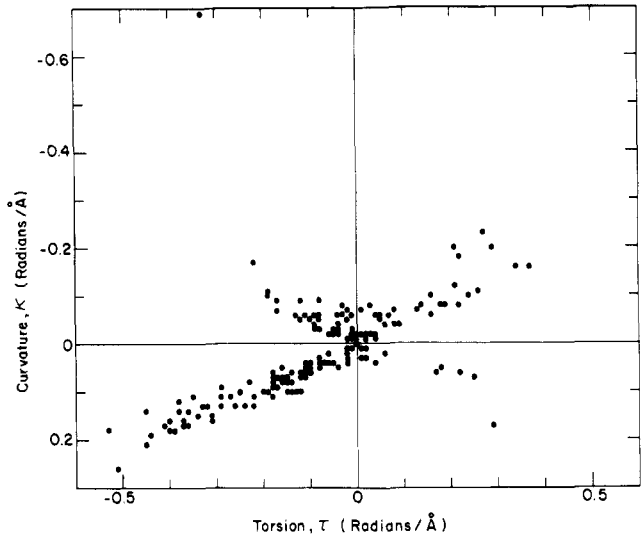

Figure 5. Values of (κ_i,τ_i) for 173 residues in the interior of extended structures.

by the various features must be delineated. Figures 4–6 (based on data for proteins of known X-ray structure) show the distribution of values of (κ,τ) for four-C^α units of helices, extended structures, and bends, respectively. Inspection reveals several interesting facts: Points corresponding to four-C^α segments in the interior of α-helices are quite tightly localized in the (κ,τ) plane; the distribution of points corresponding to four-C^α segments of extended strands is much broader than that of helical points; and the distribution of points representing bends is partially coincident with, but broader than, the distribution of helical points. [It should be remembered that bends are *defined* by the relative positions of four successive C^α's, so that the DG representation is *capable of representing bends by one* (κ,τ) *pair*.]

These points gain in significance when viewed in light of certain mathematical properties of the DG representation.[6] First, the parity (handedness) of a given four-C^α structure can be inferred from its values of (κ,τ) alone over most of the (κ,τ) plane. Second, a discon-

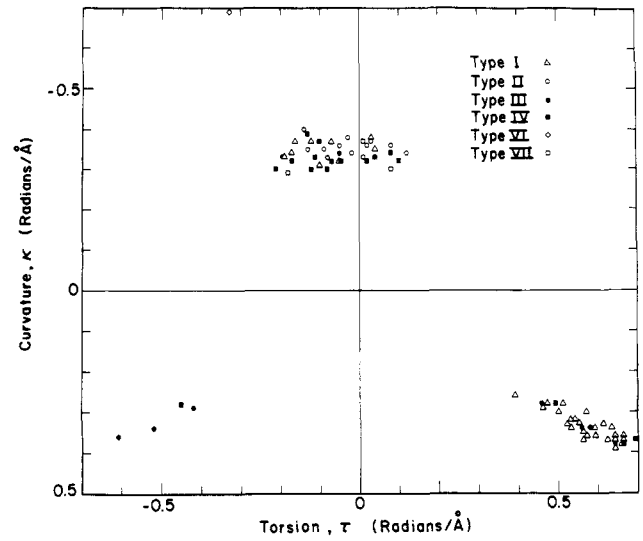

Figure 6. Values of (κ_i,τ_i) for residue i in 82 bends defined by C^α_{i-1}–C^α_{i+2}. The symbols correspond to the various types of bends[1] indicated in the inset.

tinuity in values of (κ,τ) (similar to the periodic boundary condition familiar in many applications) occurs in the helix-bend region of the (κ,τ) plane. By use of these facts, it can be shown that the three apparently isolated points in Figure 4 represent structures that are very similar to those represented by the larger group in the lower right quadrant. This point is even more relevant to Figure 6, from which it can be deduced that there exists a structural continuum of bends, ranging from *right*-handed structures that resemble four-C^α segments of α_R-helices, through flat structures, to four-C^α segments of *left*-handed α_L-helices. This fact is by no means clear from the (ϕ,ψ) representation. Indeed, Figure 6 shows that there is little correlation between the (ϕ,ψ) classification of bends (indicated by the bend types in the inset) and their structurally clear distribution in the (κ,τ) plane. The same analysis can be applied to Figure 5, from which it can be seen that four-C^α segments of extended strands are distributed over a broad region encompassing both right- and left-handed four-C^α structures. The majority of points fall into the left-handed region, reflecting the fact that, while β-sheets are right-handed in twist, the strands that make them up tend to be left-handed.[7]

It has thus been established that various structural features characteristic of proteins occur in characteristic regions of the (κ,τ) plane and that analysis of these regions can reveal structural information not otherwise readily apparent. In the next section we begin to discuss the mechanism of protein folding by attacking the problem of the distribution of structural features from a different point of view.

Nucleation Structures and Folding

Rather than investigating the distributions of given types of structures (e.g., α-helix) on the four-C^α length scale, one can ask which structures become apparent when one examines the distribution of *all* four-C^α units in the (κ,τ) plane. One would like to know, for example, about the distributions of four-C^α structures intermediate between the "classical" structural types, viz., α-helices, extended structures, and bends. A related, more

(6) Rackovsky, S.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 1440.

(7) See ref 5, footnote 15.

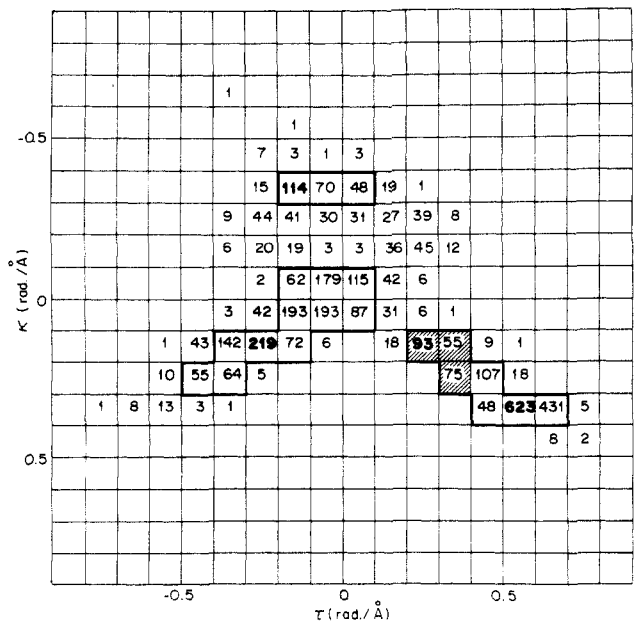

Figure 7. Distribution of (κ,τ) values for a sample of 22 proteins. Islands of greater-than-random occupation are indicated by heavy outlines. The number in each square is the number of points that fall into that square. Occupation number in boldface indicates that that square is a local maximum in the distribution.

general question of particular importance is that of the distribution of four-C^α units in the (κ,τ) plane. This is important because one can make the plausible assumption that those structures that occur with high frequency are of low energy and are therefore likely to act as nucleating structures in the unfolded molecules.

With this in mind, an analysis was made of the distribution in the (κ,τ) plane of all four-C^α units from 22 proteins of known X-ray structure.[5] The resulting histogram (Figure 7) was analyzed to reveal those regions that are occupied with greater-than-average frequency. These consist of three "islands", one of which contains a pair of peaks.

It follows from an analysis of the histogram, under the hypothesis set forth above, that there are five structural types that act as possible four-C^α nucleating structures. These are designated as E_L, E_O, E_R, A_R, and A_O. (A_L is observed with low frequency in proteins). Here E denotes extended and A denotes bend/helix structures on the four-C^α length scale. The subscripts L, O, and R denote left-handed, nearly flat, and right-handed structures, respectively. A remarkable result can be demonstrated by considering the distribution of these nucleating structures in light of the mathematical properties of the (κ,τ) plane referred to above. This is that the possible nucleating structures form a near-continuum that extends throughout the accessible regions of the (κ,τ) plane except for the left-handed helical region. Therefore, any four-C^α structure is within fluctuation range of a possible nucleation structure. Of course, there are significant differences in frequency of occurrence of the various nucleating structures. Nevertheless, this result suggests that selectivity of nucleation is rather weak on the four-C^α length scale. One may presume that specificity increases as nuclei grow to longer length scales. We will address this question in the next section.

Before doing so, we remark that the approach that we have outlined provides a quantitative basis for the concept of the existence of distinct structural features in protein backbone structures. The "classical" structural types correspond to peaks in the (κ,τ) distribution, some of which can be divided into distinct regions in a revealing fashion. However, there also exist *intermediate* structures, which can be put on an equally quantitative footing by using this methodology. The "classical" features thus correspond merely to maxima in a structural continuum. It should also be pointed out that, as was demonstrated explicitly for bends, there exists a degeneracy in backbone folding, in that a given four-C^α structure can be achieved by more than one choice of (ϕ,ψ) variables. We do not yet know whether this fact is of significance in protein folding.

Table I
Nearest-Neighbor Correlations in Five-C^α Units

positive correlation	negative correlation
E_XE_Y	E_XA_R
A_OE_X	A_RE_X
E_OA_O	
A_RA_R	
A_RA_O	
A_OA_R	
A_OA_O	

Extending the Length Scale

In order to study nucleation on the next length scale, the five-C^α scale, it is necessary to consider the distribution of nearest-neighbor pairs of DG parameters, i.e., the distribution of $(\kappa_i,\tau_i,\kappa_{i+1},\tau_{i+1})$ values.[5] An analysis analogous to that performed on the (κ,τ) distribution can be carried out, although it must be done numerically rather than graphically. In this case, two factors are important in identifying probable nucleating structures. As in the single-site (κ,τ) distribution, one must identify regions that are occupied with greater-than-random frequency. However, the nearest-neighbor pairs in the five-C^α distribution are also drawn from the single-site distribution. Therefore, one can ask which pairs occur with greater-than-random (or less-than-random) correlation, i.e., which nearest-neighbor pairs of values of (κ,τ) occur with greater frequency than would be expected on the basis of their frequency in the single-site distribution. Those five-C^α structures that exhibit *both* greater-than-random frequency and greater-than-random correlation are regarded as probable nucleation structures on the five-C^α length scale.

It is not surprising that the nucleating structures on the five-C^α scale turn out to be combinations of those identified on the four-C^α scale. It is observed, however, that there are significant variations in the tendency of these four-C^α structures to associate (see Table I). Almost all combinations of E_XE_Y (X, Y = L, R, O) structures occur as peaks with positive correlation. There is a strong preference for nearly-flat E_OE_O structures, but E_OE_L, E_LE_L, and E_LE_O structures also occur with high frequency. Combinations of E_L or E_O with E_R are less frequent, due to the relative scarcity of E_R structures.[7]

As expected, there is a very large peak due to A_RA_R structures, arising from α-helices. Smaller peaks are also observed (with positive correlation) due to A_RA_O, A_OA_R, and A_OA_O peaks.

The foregoing structures (extended strands and α-helices) are usually thought of as repeating. The DG

representation makes possible a quantitative definition of this concept, in terms of the average conformational distance between the (κ,τ) pairs making up the points of a given peak in the nearest-neighbor distribution. It can be shown that, in fact, some of the extended peaks (e.g., E_RE_L, E_LE_R, E_RE_O) are not really repeating structures. There is, however, a strong preference for more nearly repeating types of E_XE_Y structures.

Of particular interest are those nucleating structures that are not repeating. It is observed that E_XA_R and A_RE_X peaks have less-than-random correlation, i.e., *A_R and E_X structures tend to avoid each other.* On the other hand, A_O and E structures have a strong tendency to associate, with both A_OE_X and E_OA_O structures showing strong positive correlation. It therefore seems that the A_O bend structure is crucial for the nucleation of nonrepeating structures in proteins.

One can use the nearest-neighbor distribution to demonstrate the increased selectivity of nucleation on the five-C^α scale. Analysis of the single-site (four-C^α) (κ,τ) distribution shows that the peaks with above-average occupancy cover 30% of the occupied area of the distribution and include 81% of the points (four-C^α units). As noted previously, this suggests weak selectivity. In the five-C^α distribution, the peaks with above-average occupancy cover only 15.6% of the occupied area and include 59.5% of the points. The decrease in area and population imply a significant increase in selectivity.

As with the single-site distribution, the analysis of the nearest-neighbor distributions provides a quantitative basis for the concept of the classical "repeating structures", showing that they correspond to peaks in a continuous distribution of structural types, which also exhibits peaks corresponding to nonrepeating structural types.

In related work,[8] an analysis of a particular class of five-C^α structures, the double bends, has been carried out by using differential-geometric and other methods.

Roles of the Amino Acids

In the foregoing analyses of four- and five-C^α structures, no correlation was made with amino acid sequence. This is clearly an overwhelmingly important factor in protein folding, and it is to be expected that significant correlation will be found between amino acid composition and the (κ,τ) distributions of four-C^α structures. Such a study has been carried out,[9] using the same protein sample that formed the data base for the four- and five-C^α distributions.

With use of tools developed to compare (κ,τ) distributions and the assumptions noted above as to the relationship between high-frequency structures and nucleation, two conclusions can be drawn. First, the 20 naturally occurring amino acids can be divided into two groups. The larger group (group I) is responsible for nucleation of A_L, A_R, and A_O structures when located at the second position of a four-C^α unit and of A_R and E_O structures when located at the third position. The smaller group (II, composed of Pro, Gly, His, Tyr, Cys, Asn, Trp) is responsible for the nucleation of E structures when located at the second position of a four-C^α unit and of all structures except A_R and E_O when at the third position. Second, since the group II amino acids constitute only 28% of the sample, it is clear that nucleation of A_X structures must be the dominant nucleation event, although E structures are also formed.

It should be noted that this analysis, based on the *complete* distributions of (κ,τ) values for the 20 amino acids, is different in approach from studies on the amino acid compositions of particular structural features, e.g., bends.[10] By working with complete distributions, we are able to analyze *simultaneously* the compositional preference of given regions of the (κ,τ) plane and the conformational preference of given amino acids. It should also be noted that the (κ,τ) distributions can also be regarded as empirical potential energy maps on the four-C^α length scale, averaged over nearest-neighbors. This is in the same spirit as the observation that there is good agreement between the observed (ϕ,ψ) distributions of given amino acids and their (ϕ,ψ) conformational energy maps.[11] An analysis of the implications of this concept is planned.

The picture of protein folding that emerges from the foregoing work provides a unifying framework for various previous proposals. Both the hairpin bend[12] and the α-helix[13] have been proposed as primary nucleation structures. Both are, of course, examples of A structure; which one actually becomes the nucleation structure will clearly depend on the sequence of the particular protein under consideration. It should be emphasized, however, that, in contrast to previous work, we are not concerned here with the identification of the first kinetically visible nucleation structure. The structures that we have proposed may or may not fulfill this role. They may, instead, be rapidly formed structures whose presence is necessary to insure that the first kinetically visible folding steps lead to energetically favorable structural intermediates.

Comparison of Conformations

An interesting application of the DG representation is to the comparison of protein conformations. Traditionally, this problem has been treated by the use of a root-mean-square deviation (RMSD) of interatomic distances between the two conformations of interest,[14] often together with an optimal superposition procedure.[15] Although these methods have found frequent application, they suffer from certain problems that have been noted by various authors: The use of a single number as the index of quality for comparing two very large molecules is inherently unsatisfactory.[4,16] Moreover, the interatomic distance and optimal superposition methods are oversensitive to overall agreement between the two structures. Structural similarities or divergences on a more local level are not well treated by this approach.[17] Finally, the method of optimal superposition requires fairly time-consuming computational procedures.

(8) Isogai, Y.; Némethy, G.; Rackovsky, S.; Leach, S. J.; Scheraga, H. A. *Biopolymers* **1980**, *19*, 1183.
(9) Rackovsky, S.; Scheraga, H. A. *Macromolecules* **1982**, *15*, 1340.
(10) For example, see: Zimmerman, S. S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *24*, 4126.
(11) Némethy, G.; Scheraga, H. A. *Q. Rev. Biophys.* **1977**, *10*, 239.
(12) Matheson, R. R., Jr.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 819.
(13) Finkelstein, A. V.; Ptitsyn, O. B. *J. Mol. Biol.* **1976**, *103*, 15.
(14) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59.
(15) Rossman, M. G.; Argos, P. *J. Mol. Biol.* **1976**, *105*, 75.
(16) Hagler, A. T.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 554.
(17) Sippl, M. J. *J. Mol. Biol.* **1982**, *156*, 359.

A complete solution of the comparison problem demands recognition of the fact that, like other aspects of protein folding, comparison must be carried out on a hierarchy of length scales. As noted above, the overall similarity of shape of two conformations does not guarantee that there will be reasonable similarity of *local* folding. Nor does approximate similarity of local folding lead to overall similarity of structure.[18] The DG approach provides a previously unavailable solution to this problem on the four-C^{α} length scale. Since this is the elementary folding unit in the virtual bond backbone, it makes it possible to pinpoint differences and similarities in folding on a residue-by-residue basis.

A conformational distance function was constructed, making appropriate use of the mathematical properties of the DG representation noted above.[6] Its utility has been demonstrated by several examples. One of particular interest arose in connection with the structure of the cyclic decapeptide antibiotic Gramicidin S. The conformation of this molecule had been predicted[19] before the X-ray structure[20] became available. It was demonstrated[21] that the predicted and observed structures are very similar and that all divergences in backbone conformation can be accounted for by intermolecular interactions in the crystal, which arise from the formation of a very interesting intermolecular four-stranded β-sheet. It is planned in future work to extend the comparison method to longer length scales.

Recently, the importance of constructing a hierarchy of comparisons on different length scales was also recognized by Sippl,[17] who proposed such a comparison hierarchy based on the distance representation.

A modified differential-geometric comparison method was applied by Krigbaum and Lin[22] in studies of folding of a model protein on a body-centered cubic lattice. They observed no correlation in this case between comparisons based on RMSD and superposition methods and the differential-geometric comparison, which they also express in the form of a single, averaged number for the whole molecule. This may well be due to the fact, noted above, that the two methods are sensitive to backbone structure on very different length scales. One would expect this effect to be emphasized in a lattice model, in which local backbone structure is expressed in a limited number of unphysical conformations. It should also be remarked that there are pitfalls associated with the use of the differential-geometric representation on a lattice that do not arise in actual conformations.

Other Work

In two recent papers[23,24] Louie and Somorjai have proposed a differential-geometric treatment of backbone structure based on approximating the backbone by a continuous curve. In ref 23, they discussed idealized α-helices and β-sheets and proposed methods whereby these two characteristic structures may interconvert. This treatment is based entirely on the mathematical features of the two (helical and sheet) surfaces, and the authors do not address the problem of the relationship between the interconversion processes that they suggest and the physics of conformational changes in actual molecules. In ref 24, they demonstrate an algorithm for producing a continuous-curve approximation to a protein backbone by piecewise-fitted helices. This approach seems to provide a numerical basis for the type of structural diagrams used effectively by Richardson[2] to represent protein structure.

The question of parity has been treated in a very general fashion by Braun,[25] who proposed a function that defines handedness for either long- or short-range structures. He has demonstrated the connection between this function and the DG definition of handedness in the short-range case. The method provides a unified treatment of the various cases (such as crossover connections, etc.) where parity has been considered.[2,11]

Concluding Remarks

The development of the differential-geometric representation has enabled us to focus directly on the properties of protein structures on intermediate length scales. The resulting structural information, interpreted in light of reasonable hypotheses about protein folding, promises to increase our understanding of the nucleation processes that govern the remarkable efficiency of the folding process. The foregoing Account traces the first steps in this approach to the exploration of the connection between protein structure and dynamics.

The work described here was supported by research grants from the National Science Foundation (PCM79-20279) and the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (GM-14312). We thank S. Fitzwater, N. Gō, Z. I. Hodes, Y. Isogai, H. Meirovitch, R. A. Nemenoff, J. C. Owicki, Y. Paterson, W. J. Peer, M. R. Pincus, M. S. Pottle, S. Rumsey, H. Wako, S. S. Zimmerman, and, in particular, G. Némethy for many helpful discussions.

(18) Burgess, A. W.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 1221.
(19) Dygert, M.; Gō, N.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 750.
(20) Hull, S. E.; Karlsson, R.; Main, P.; Woolfson, M. M.; Dodson, E. J. *Nature (London)* **1978**, *275*, 206.
(21) Rackovsky, S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6965.
(22) Krigbaum, W. R.; Lin, S. F. *Macromolecules* **1982**, *15*, 1135.
(23) Louie, A. H.; Somorjai, R. L. *J. Theor. Biol.* **1982**, *98*, 189.
(24) Louie, A. H.; Somorjai, R. L. *J. Mol. Biol.* **1983**, *168*, 143.
(25) Braun, W. *J. Mol. Biol.* **1983**, *163*, 613.