A method for the characterization of foldings in protein ribbon models

Gustavo A. Arteca and Paul G. Mezey

Department of Chemistry and Department of Mathematics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

The ribbon model of chain macromolecules is a useful tool for analyzing some of the large-scale shape features of these complex systems. Up to now, the ribbon model has been used mostly to produce graphical displays, which are usually analyzed by visual inspection. In this work we suggest a computational method for characterizing automatically, in a concise and algebraic fashion, some of the important shape features of these ribbon models. The procedure is based on a graph-theoretical and knot-theoretical characterization of three well-defined projections of a space curve associated with the ribbon. The labeled graphs can be characterized by the handedness of the crossovers in the ribbon that are the vertices of the graph. The method can be used to provide a fully algebraic representation of the changes occurring when a molecule, such as a protein, undergoes conformational rearrangements (folding), as well as to provide a shape comparison for a pair of related molecular ribbons. This algebraic representation is well suited for easy storage, retrieval, and computer manipulation of the information on the ribbon's shape. Illustrative examples of the method are provided.

Keywords: molecular shape, protein folding, graph theory, knot theory

The structure of such large and complex molecular systems as proteins can be described by considering different levels of organization. The possibility of several scales for viewing the macromolecule's shape features allows one to employ different types of models that complement one another. Occasionally, molecular surface models (such as the hard-sphere van der Waals models) are used to represent a molecule. However, it is cumbersome to analyze the shape features of such models for molecules with a very large number of atoms. Alternative approaches are needed in these cases. The ribbon model of macromolecules is one of the more appropriate tools for these structures. This model obviously omits many details of the molecular structure, but it retains sufficient information to represent the secondary structure in proteins, the patterns of the tertiary structure, and the foldings of the macromolecular backbone.

One of the most appealing and commonly used methods for the construction of macromolecular ribbons is that of Richardson.[1] Several alternative implementations are also available.[2–4] In Richardson's model, α helices are represented by solid cylindrical helices of solid ribbons, whereas β strands occur as thick arrows, and the nonrepetitive loops connecting β strands and helices appear as thick "strings." As shown in Ref. 4, a satisfactory overall description can indeed be given as a 3D figure using ribbons for all the secondary structures. Other representations of ribbon-like models of proteins have also been given. An interesting approach[5,6] consists of representing the β strands by arrows as above, but the α helices are modeled by solid cylinders without internal structure. Both approaches will be used in this work.

These models are normally used for providing graphical displays of the macromolecular structure on a large scale. Ribbon models are undoubtedly helpful for recognizing the folding pattern and tertiary structure in proteins (antiparallel α and β structures, and parallel α/β structures, among others).[7] Usually, the description of the shape features of these molecular ribbons is based on subjective visual inspection. This procedure is somewhat unreliable when one needs precise shape comparison, as is the case when studying differences between molecules or a given molecule undergoing conformational changes, such as a continuous folding. Such studies can be placed on a firmer basis if an unbiased, systematic shape characterization is available. Such descriptions have been proposed for pairs of molecules using electronic density functions,[8–12] and extensive work has been carried out recently on the characterization of molecular surfaces.[13–27] These methods, based on results from algebraic and differential topology,[13–22] graph theory,[23–27] fractal dimensionality,[28] and two-dimensional Fourier transforms,[29] provide local and global descriptions of molecular surfaces. Their implementation for the study of large mo-

Address reprint requests to Dr. Mezey at the Department of Chemistry and Department of Mathematics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0.
Received 15 May 1989; accepted 21 February 1990

lecular systems may appear cumbersome, but the above methods are useful if one needs to pay attention only to certain local features of some part of the molecular surface. On the other hand, a study of global shape features of a ribbon model of a macromolecule can rely on alternative methods.

In this work we propose two techniques for the characterization of molecular ribbons. The procedure is simple and it provides a rigorous basis for the intuitive, observational analysis of a ribbon in a graphical display. Our proposal is to assign an oriented space curve (a one-dimensional object) to the ribbon, a topologically two-dimensional surface. The characterization of this curve can be performed by defining, in the first place, three preferential observational directions, and by projecting the curve onto planes defined by pairs of the above directions. The projections can be characterized in terms of plane graphs,[30] taking into consideration the overcrossings in the space curve. Moreover, a class of knots can be built from the projected view of the space curve. These knots can be characterized by polynomials, which provides an alternative algebraic characterization to the original ribbon.

The scope and applications of the method are described in some detail. The article is organized as follows. In the following section the theoretical basis is provided. The next section deals with some illustrative examples in terms of simple space curves. A discussion is also provided on the usual features of interest to observe in the case of a molecular ribbon, and how they appear in this formulation. The fourth section discusses some typical examples of tertiary structure in proteins, at different levels of approximation, in the context of the method developed. The fifth section presents an alternative knot-theoretical characterization of the space curves. Further comments and conclusions are found in the last section.

MOLECULAR RIBBONS AND MOLECULAR SPACE CURVES

The macromolecular models we discuss in this work provide a representation for molecules in terms of two-dimensional surfaces. This is the case of hard-sphere van der Waals surfaces, ribbon, and tubular models. In what follows, we will indicate any of these general surfaces as a set of points R in three-dimensional space. In what follows, our discussion will be presented using simple qualitative geometric ideas. For the sake of completeness, a more technical, topological discussion is presented in the appendix.

We are concerned in this work with the case of R being a ribbon-like surface. Let us consider a polyatomic molecule with N atomic nuclei; let $\mathbf{R}_\alpha = (X_{1\alpha}, X_{2\alpha}, X_{3\alpha})$ be the position vector for the αth nucleus of mass M_α, with respect to some arbitrary Cartesian laboratory frame. It must be noticed that not all atoms are taken into consideration when building a molecular ribbon. Some groups are often disregarded, and the conformational arrangements of others are usually simplified according to some conventional criteria.[2] As mentioned earlier, the aim of such constructions is to emphasize some large-scale structural features of the molecule, neglecting some small-scale, less significant characteristics. We assume that a model surface R is already available, and we shall not be concerned with the criterion used for its construction.

If the set R represents a single, noncyclic *ribbon model* (RM) that has zero thickness, then R will be equivalent (in a topological sense) to a rectangular planar domain and also to a two-dimensional disk, closed or open depending on whether the boundary of the ribbon is considered or not.[30] These models can take into account some internal structure of molecules by giving the ribbons nonzero thickness.[4] This latter case will be referred to as a *thick ribbon model* (TRM), as opposed to the more conventional ribbon model. A single, noncyclic TRM is topologically a solid ball. In this respect, a tubular representation of a molecular backbone is equivalent to a TRM.

Let us suppose first that the surface R represents a RM. (Only the strictly ribbon-like part of the model will be taken into consideration here; disulfide bridges or metal bonds[2] between branches of the ribbon will not be represented by ribbons.) The case of TRMs is discussed later.

A RM has two well-distinguished cross sections, a short one across the ribbon (*transversal*), and a longer one along it (*longitudinal*), that are parallel with the shorter and longer sides, respectively, of the rectangular representation. Let $\mathbf{r}(t)$ be a parametric space curve,[31]

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} \qquad 0 \le t \le 1 \qquad (1)$$

that is located on R, $\mathbf{r}(t) \in R$ for all t, along the longitudinal direction of the ribbon. The three unit vectors of an orthogonal Cartesian framework taken as a reference are indicated by the usual symbols $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$. The change in parameter t from 0 to 1 provides an orientation to the space curve, and hence to the ribbon. Here $\mathbf{r}(0)$ corresponds to the "starting" point of the curve (beginning of the ribbon) and $\mathbf{r}(1)$ to the "end" point.

Functions $x(t)$, $y(t)$, and $z(t)$ in equation 1 need to be only sectionally continuous. This will be the case, for example, where the ribbon is formed by several disjoint sections. The above functions, when continuous, do not even need to have continuous derivatives. Nonetheless, in the case of the usual molecular models, these functions will be differentiable within each section.

There is an infinite number of curves $\mathbf{r}(t)$ that can be traced on the surface R along the longitudinal direction. A possible choice would be to take $\mathbf{r}(t)$ as a sequence of geodesic curves, leading from $\mathbf{r}(0)$ to $\mathbf{r}(1)$.[31] Although in practice it is rather complicated to determine geodesic line segments for these molecular models, the existence of a minimum number of geodesic line segments constrained to R, as well as their curvature changes, may provide valuable information on the shape characteristics of the ribbon.

In practice, one can resort to a much simpler and convenient alternative, consisting of choosing the space curve as the median line of the ribbon. Each RM is topologically equivalent to a rectangular domain in a plane. Let us denote the chosen transformation converting the RM into the rectangular domain by Q. The median line m of the rectangular domain is the C_2 axis connecting the images of beginning and end of the ribbon. The median of the RM is defined as the inverse transform $Q^{-1}m$ of the median line m of the rectangular domain. Figure 1*a* displays schematically the relation between a ribbon and its associated space curve $\mathbf{r}(t)$, derived as the median line of the RM.

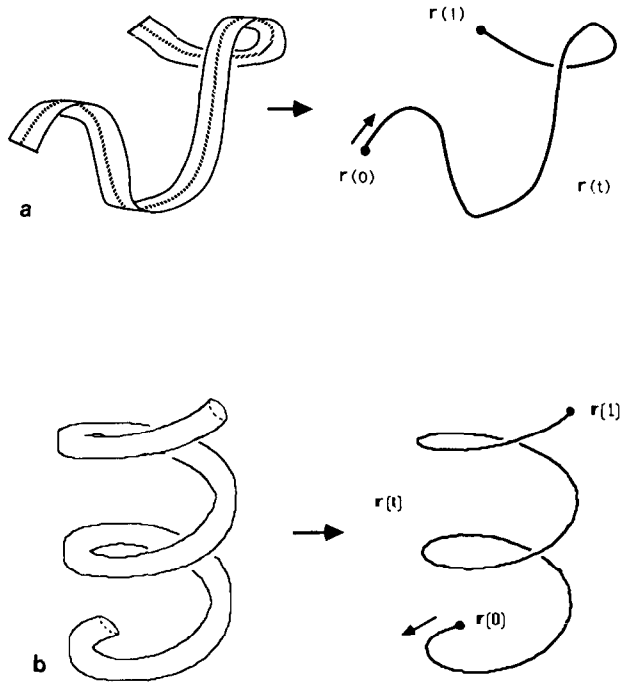


Figure 1. Relation between (a) *flat ribbon models and* (b) *thick ribbon models and their corresponding space curves*

Similarly, the TRM, topologically equivalent to a cylinder, can be transformed into a topologically equivalent rectangular solid block by a transformation Q'. The inverse transform $Q'^{-1}m'$, of the median line m' of the block (its longest C_2 axis) defines the median line of the TRM. Figure 1*b* shows schematically the relationship between this ribbon model and its corresponding space curve $\mathbf{r}(t)$, derived from the median line of a block, topologically equivalent to the solid ribbon model.

The replacement of a molecular ribbon by an associated space curve (a *molecular space curve*) eliminates many structural details. Yet, the space curve retains the essential information needed to describe folding patterns as well as some other features of the backbone structure of large molecules. Our main concern now is to develop an effective, simple algebraic characterization of a molecular space curve $\mathbf{r}(t)$ obtained from the ribbon.

The curve $\mathbf{r}(t)$ mimics the features of the molecular backbone. If the two end points are joined, then, in general, such an object is a topological knot and, accordingly, it can be studied by means of the branch of topology known as knot theory.[32] The usefulness of knot theory to several chemical applications (for instance, molecular chirality) is well documented. (See, for example, Refs. 33–43 and others quoted therein.) Its relevance to the description of the entanglement of macromolecular chains has also been recognized.[44,45] A knot-theoretical description permits one to recognize the occurrence of certain topological features, which remain invariant to conformational motions allowed to the molecule (as long as the chains are not broken and no cross-linking occurs). Among these invariants, one can mention the linking number[45] and the writhing and twisting numbers.[46–49] These three numbers may be used for the characterization of a ribbon model.

However, the above description may disregard important characteristics of the macromolecular folding. Overcrossings of side chains, changes in tertiary structure in proteins, and the opening of protein cavities all represent relevant features for understanding protein dynamics and reactivity (e.g., Refs. 50–53). Yet, these features are not represented in such a description, because they do not lead to a change in the knot type or fundamental group.[32] Moreover, in an overwhelming number of cases, the protein backbone structure, as described by molecular ribbons, tubes, or space curves, leads to simple loops that are not knotted, that is, to *unknots*.[32,43] Consequently, the standard description of the original space curve as a knot will not be very informative, in general. To derive a more informative description, a transformation could be introduced on the space curve, such that it may assign a knotted structure to an unknotted one. Such transformations are discussed in the last section.

Alternatively, it is possible to associate *plane graphs* to projections of the space curves. This provides a discrete characterization of the role of conformational changes (foldings) that lead to changes in the overcrossing pattern of the macromolecular chains. The knottedness is not explicitly considered here. The plane graph characterization of curves $\mathbf{r}(t)$ is described in the next section.

GRAPHICAL CHARACTERIZATION OF MOLECULAR SPACE CURVES

Let $\mathbf{r}(t)$, $0 \leq t \leq 1$, be our space curve, associated with the backbone of a chain molecule, as discussed in the previous section. Let us further assume that $\mathbf{r}(t)$ is confined to a finite region of the 3-space.

To characterize in a simple way some of the shape features of the curve, we introduce here a graph-theoretical characterization of three of its projections. This description provides a discrete characterization of the curve that is simple and appropriate for computational manipulation. Nevertheless, other formulations could be followed. Continuous descriptions may prove valuable in other cases. The characterization of continuous space curves by means of topological invariants such as writhing and twisting numbers has already been explored.[46–48] Also, if $\mathbf{r}(t)$ is differentiable everywhere, it is possible to define its continuous curvature and torsion, according to the Frenet–Serret formulas.[31] The changes in torsion and curvature as functions of the curve's arc length (as given by the intrinsic equations of the curve) contain valuable structural information regarding the possible molecular folding patterns. Important information is also provided by the conditions that a ribbon must satisfy to admit a continuous and differentiable geodesic curve. The study of this possibility, however, is beyond the scope of this work.

The projections of the space curve $\mathbf{r}(t)$ are chosen according to the axes of a preferential coordinate framework. This framework will be taken to be the one given by the three orthogonal axes of inertia. These axes are defined by the triplet of three orthogonal vectors that diagonalize the matrix of inertia $\mathbb{A}$, whose components are given by[54]

$$(\mathbb{A})_{ik} = \sum_{\alpha=1}^{N} \sum_{j=1}^{3} M_\alpha (X_{\alpha j}^2 \delta_{ik} - X_{\alpha i} X_{\alpha k}) \qquad (2)$$

where δ_{ik} is the Kronecker delta. Nuclear coordinates $X_{\alpha s}$ in equation 2 are measured from the center-of-mass origin. Notice that the definition of the three *viewing* directions involves all the atoms in the molecule, even though not necessarily all of them are used in the actual construction of the molecular space curve. The choice of the three axes, hereforth indicated by the triplet (Q_1, Q_2, Q_3), attached to the center-of-mass, guarantees that translations and rigid rotations will not generate changes in the derived shape descriptors. This allows one to focus the attention on the dependence of the curve's shape on the internal degrees of freedom.

Let us consider a cube that encloses the space curve (the curve is supposed to be bounded). One can always find one such a cube so that the set of three of its C_4 symmetry axes passing through the center of faces coincide with (Q_1, Q_2, Q_3). The chosen projections will be taken as views along the three axis from a viewer outside of the cube. For instance, a projection of the molecular space curve to the plane Q_1–Q_2 is equivalent to seeing a photograph of the bounded curve from far along the axis Q_3 (a similar interpretation applies for the other two directions). $P(Q_i)$ will indicate the projection of the curve $\mathbf{r}(t)$ onto a plane $Q_i = Q_i$, where Q_i is any coordinate value along the axis Q_i. The result of this operation is a plane curve, denoted as $\mathbf{q}_i(t)$:

$$\mathbf{q}_i(t) = P(Q_i)\mathbf{r}(t) \qquad 0 \leq t \leq 1 \tag{3}$$

Figure 2 shows schematically the relation between a molecular ribbon and the three viewing directions chosen. The

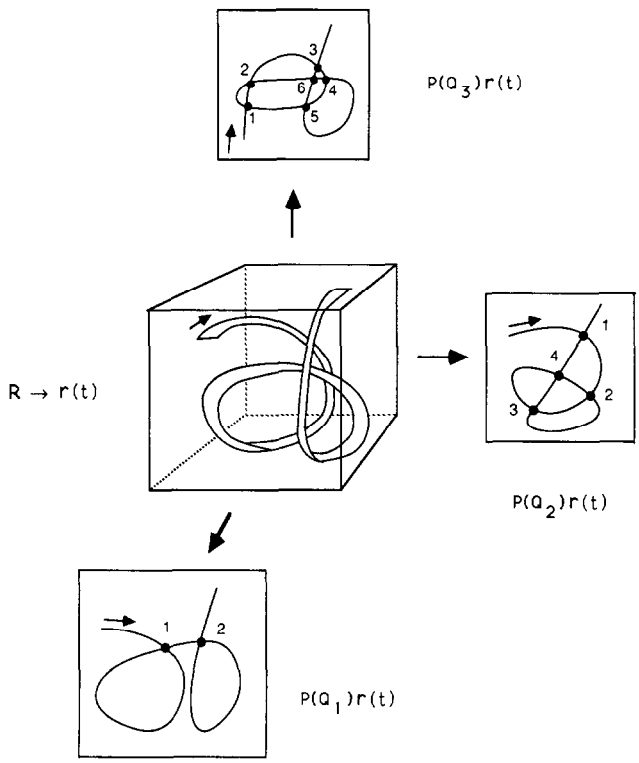


Figure 2. Relation between a flat ribbon model and its three corresponding planar curve projections.
Crossings are indicated as vertices, marked according to their occurrence along the space curve for the given orientation

results of the projections appear as plane curves exhibiting a number of crossings; these crossings are the consequence of observing the space curve curling over onto itself from a certain direction of space. The occurrence of these crossings permits us to characterize the plane curves $\mathbf{q}_i(t)$, i = 1, 2, 3, and, in turn, the space curve $\mathbf{r}(t)$.

Let us associate a graph g to each plane curve $\mathbf{q}_i(t)$, described here in terms of intuitive concepts. A more technical, set-theoretical presentation is given in the appendix.

The vertices of these graphs are the crossings of the plane curves (corresponding to overcrossings in the original space curve), and the edges are the segments of the curve connecting crossings. These segments can connect two different crossings, or a crossing with itself. Segments of the curve not providing connections between vertices will not contribute to the graph. From these intuitive notions the formal definition of the graph g_i can be given (see also the appendix).

1. *Vertices of the graph.* These are those points or connected point sets on the planar projection that represent an overlap of the original space curve, as viewed from the given preferential direction in 3-space. Notice that overlaps can occur not only as isolated points, but also entire parallel sections of the projected curve can appear to cover each other. The vertices are collected in an ordered set $V(g_i)$:

$$V(g_i) = (\nu_{i1}, \nu_{i2}, \nu_{i3}, \ldots) \tag{4}$$

The ordering is established from the orientation along the original space curve. The vertices are numbered according to their occurrence when moving along the curve. Note that when a crossing is arrived at the second time around, no new serial number is assigned to the crossing.

Any bounded molecular space curve will generate graphs with a finite number of vertices. The vertices are actually given by the actual number of crossing points on the curve only if from a given view the number of points where overlap occurs is finite. (See discussion in the appendix regarding this property.)

2. *Edges of the graph.* If two vertices of the graph are connected by a section of the planar projection of the space curve, then there will be an edge connecting them in the graph. In most cases multigraphs will be found (i.e., more than one edge connects two vertices). The set of edges of a graph g_i will be indicated by $E(g_i)$. According to our construction, a segment of the plane curve not starting from or not ending at a crossing (an overlap, in general) will not contribute an edge to g_i.

These simple intuitive notions are sufficient to follow the discussion below. Since the projections can be constructed from the space curve in an automated fashion, the determination of the graph can be performed by a computer.

The three graphs g_1, g_2, and g_3 associated with the space curve $\mathbf{r}(t)$ obtained from the ribbon G provide a concise description of some of its shape features. The basic information defining the graph is contained in its adjacency matrix, which can be easily stored, manipulated, and analyzed by a computer. Nevertheless, one can introduce additional information within the graph. For example, one can associate each vertex with a crossing index, characterizing the type of overcrossing in the space curve that defines the vertex in the graph. If ν_{ij} indicates the jth vertex of the graph g_i (numbered according to the order of occurrence along the

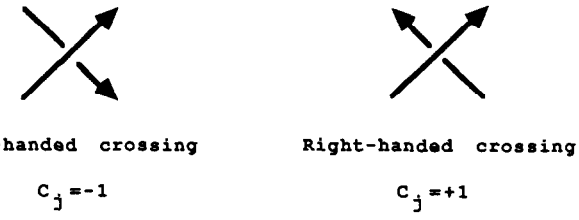


Figure 3. Convention for the handedness of crossings

curve), its corresponding crossing index C_j is defined as follows: $C_j = 1$, if the vertex represents a *right-handed crossing*, and $C_j = -1$, if it represents a *left-handed crossing*. Figure 3 illustrates the handedness of crossings.

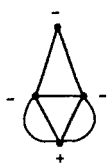
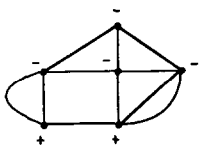
Three types of pathological crossings can occur: (a) three or more sections of the curve overcross each other at a point; (b) two or more segments meet at a point without actually overcrossing each other; (c) the vertex represents an overlap of entire segments of the curve (nonpointwise crossing). The first case will be characterized by computing C_j as the sum of crossing indices for all distinct, *pairwise* crossings occurring at the point. We will use the notation $C_j = 0$ for case b, whereas in case c the crossing index is that of a nondegenerate crossing or crossing of type a or b that can be obtained from the actual one by a continuous local deformation. This complementary information, as vertex labels for a graph with n vertices, is stored in the vector $\mathbf{C}(g_i)$:

$$\mathbf{C}(g_i) = (C_1, C_2, \ldots, C_n) \tag{5}$$

As an illustrative example, Table 1 shows the graphs obtained from the projections displayed in Figure 2. The plus and minus signs indicate the values $C_j = \pm 1$.

The set of graphs g_i and their characterization in terms of vector $\mathbf{C}(g_i)$ provide a rather comprehensive description of some of the structural features present in the original ribbon model. If one follows the change of a ribbon model along a conformational path (for example, a continuous folding of the protein backbone), both the graph and the set of crossing indices can change. However, this change will

Table 1. Crossing graphs for the three projections of the space curve associated with the ribbon model in Figure 2

	Projection: Q_1 = const
	Projection: Q_2 = const
	Projection: Q_3 = const

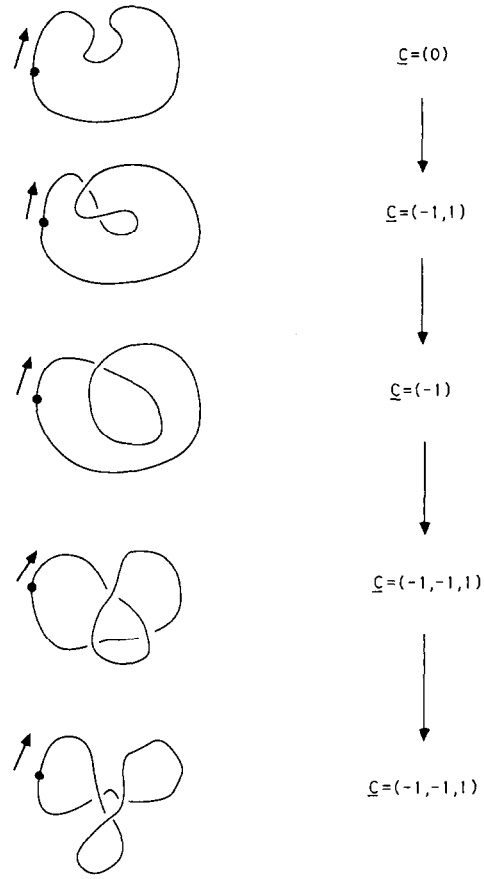


Figure 4. Characterization of foldings in a closed space curve by means of the vectors of crossing indices (equation 5)

be *discrete:* Only for certain values of the parameters defining the conformational rearrangement will the folding pattern exhibit essential changes as reflected by the projections and graphs. Consider, for example, a closed space curve, one of whose projections is followed during a folding. This case is schematically depicted in Figure 4. It may correspond to observing rearrangements in a large molecular cycle or loop. The right side of the diagram shows the changes in $\mathbf{C}(g_i)$, which occur only at specific conformations. Note, as a comparison, that the curve depicted in the figure remains always an unknot, while several changes are found in the graphs associated with its projection.

The characterization in terms of the graph g_i is more informative than that provided exclusively in terms of the vector $\mathbf{C}(g_i)$. Consider, for example, the projection of a pair of twisted curves depicted in Figure 5. Both curves A and B correspond to unknots.[55] Even though both are clearly distinguishable in terms of graphs, they possess the same set of crossing indices:

$$C(g_i) = (1, 1, 1, 1, -1, -1, 1, -1, -1, -1, -1, -1, -1) \tag{6}$$

To illustrate how degenerate vertices can occur simultaneously with vertices of pointwise crossings, consider, for example, the following space curve:

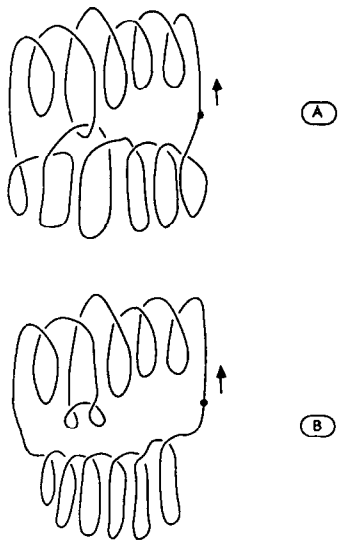


Figure 5. Example of two different closed space curves with the same vectors of crossing indices (equation 5)[55]

$$\mathbf{r}(t) = (\cos^3 t + \sin^2 t)\mathbf{i} + (\sin t + \cos^2 t)\mathbf{j} + \sin 3t\mathbf{k} \tag{7}$$

When the interval is $0 \leq t \leq 2\pi$, then $\mathbf{r}(t)$ is a twisted closed loop in three-space. This interval can be rescaled, leading to a new variable $t' \in [0, 1]$, compatible with the model described above. A representation of this curve in three dimensions is given in Figure 6 (top). For simplicity, we have considered the three Cartesian projections, also shown in Figure 6: x = const (center left), y = const (center right), and z = const (bottom). Table 2 contains the results for the corresponding graph-theoretical characterization. The following features can be observed:

1. All points in the plane curve obtained by the x = const projection are doubly degenerate. That is, the curve covers itself entirely. Accordingly, it is contractible to a single vertex as explained above.
2. The curve obtained by the y = const projection generates a graph with only pointwise left-handed and right-handed vertices. Notice that with the characterization provided by $C(g_i)$ the graph is not bipartite.
3. The z = const projection of the curve contains a single degenerate point. As the parameter t' sweeps over the interval $[0, 2\pi]$, the curve starts at, passes by, and ends at this point without actually overcrossing itself.

As an example of a curve exhibiting a simpler behavior, we can consider the following distorted helix:

$$\mathbf{r}(t) = (\cos t + \sin t)\left(\frac{t}{\pi}\right)^{1/2}\mathbf{i} + \frac{t(\sin t + 1)}{\pi}\mathbf{j} + \frac{t^{1/2}(\cos t + 1)}{\pi}\mathbf{k} \tag{8}$$

with $0 \leq t \leq 5\pi$. This interval can be rescaled, leading to a new variable $t' \in [0, 1]$, compatible with the model described above. Its graph-theoretical characterization is shown in Table 3, and it contains only pointwise vertices. As a particular feature, observe that the projections x = const

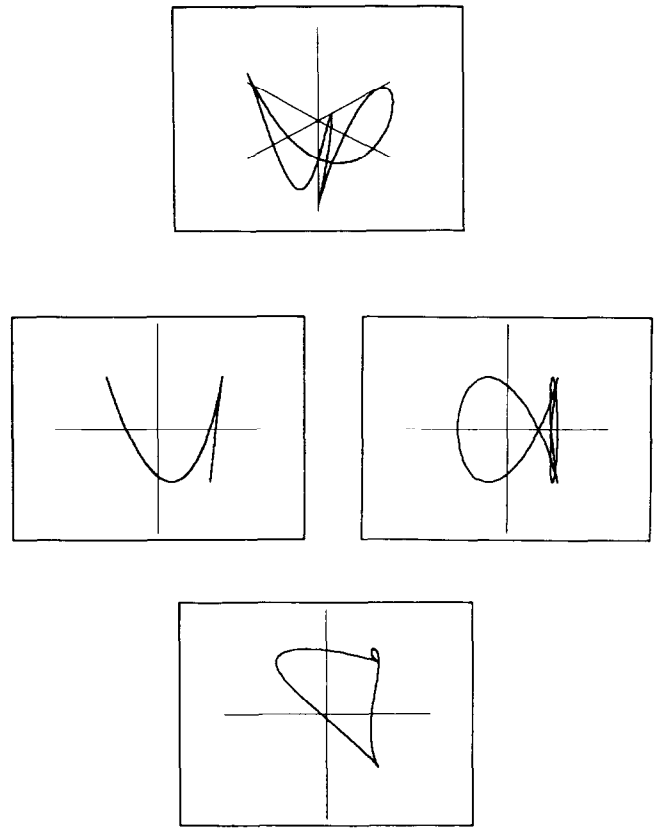


Figure 6. Space curve and Cartesian planar projections for equation 7.
Top: *Three-dimensional view of the space curve.* Center left: x = *const projection.* Center right: y = *const projection.* Bottom: z = *const projection*

Table 2. Crossing graphs for the three projections of the space curve given by equation 7

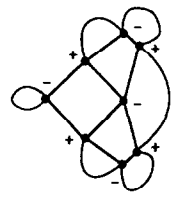
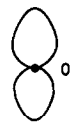
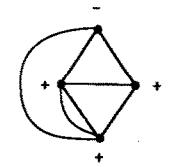
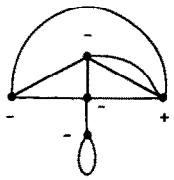
0	Projection: x = const
	Projection: y = const
	Projection: z = const

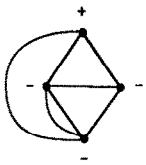
Table 3. Crossing graphs for the three projections of the space curve given by equation 8



Projection: x = const



Projection: y = const



Projection: z = const

and z = const have the same associated graphs, but they possess, under the same orientation, different sets of crossing indices.

Degenerate situations with overcrossing of sections of the curve can be handled as shown above, yet they are not very likely to occur. Nevertheless, degenerate crossovers can indeed occur for real ribbon or tubular chemical models. For example, a projection of a protein ribbon with doubly wound parallel β-sheet structure exhibits several β strands that generate a degenerate overcrossing.

The nature of the overcrossings, displayed by the protein tertiary structure from a given observational direction, deserves an additional comment. In our present analysis, we provide an entirely graphical description of the skeleton, based on overcrossing, disregarding the distances between the segments being superimposed. This description, as will be shown below, allows one to provide a quick and simple first approach to the shape characterization of the molecular backbone. However, if our goal is not pure shape description, but a description of those shape features that are dominated by short-range interactions, then a more realistic physical picture must include information of the interaction between residues. These features can eventually be included in our description. For example, one could ignore in the analysis those crossovers due to superposition of two or more segments whose distances are over a sensible cutoff value.

ANALYSIS OF SOME ILLUSTRATIVE PROTEIN MODELS

The technique described in the previous sections provides an approach for a rational, simple characterization of the structural features in a protein ribbon model. In the ribbon model, only some of the structural elements are represented; accordingly, this characterization accounts only for shape features at the scale of such elements, and disregards the conformational details of smaller molecular fragments, such as the rotation of a methyl group.

We show here some results obtained by applying the method discussed in the previous section. The algorithm for the method can be summarized as follows:

1. Construct the ribbon model corresponding to the macromolecule of interest, from the knowledge of the molecular geometry.
2. Determine (numerically) the points belonging to the space curve associated with the molecular model. The space curve for a ribbon can be built by taking the median line of the RM.
3. Determine the axes of inertia of the molecule and take them as viewing directions for projecting the space curve.
4. Describe the projections in terms of plane graphs, whose vertices are characterized by crossing indices.
5. Follow the above characterization for the foldings in the molecular backbone, and determine the domains in configurational space where the shape features, as defined by the graphs and crossing indices, remain invariant.

The existence of bridges between segments of the molecular backbone is recognized to play an essential role in determining the folding patterns in macromolecules.[39] These bridges restrict the motions allowed to a given molecule, while retaining certain shape characteristics. In our case, the problem is posed in a different way: The relevant conformational motions are determined by the condition of conserving the shape features (as described by the projected plane graphs).

Usually, the results available from the literature on ribbon models of macromolecules correspond to a single view of the model, chosen according to some subjective criterion. We use here some of these results for illustrative purposes.

Since the structure of a protein shows different levels of organization, one can display these hierarchies by introducing the secondary structural elements in steps. A single view is sufficient to illustrate the characterization of these shape features, and we provide some examples of this approach in what follows.

Suppose that each α helix is treated as a single, structureless element. If the helix is viewed as a rod, cylinder, or a strand,[6] the overcrossings *within* the helix are disregarded in first instance; only the crossovers of the helix as a whole with the rest of the molecule are retained. In this case, the corresponding molecular space curve $\mathbf{r}(t)$ can be built by taking the central axis of the helix, whenever an α helix is involved. The characterization of the foldings with and without considering the internal structure of the helices provides some simple description of some important structural features.

Figure 7 shows insulin, a small, irregular protein. The ribbon is in fact formed by two disjoint pieces (once the disulfide bridges are omitted). The drawings to the left and right represent the models with structureless and full structure helices, respectively. The corresponding graph characterizations are indicated in the lower part of the figure; the occurrence of new crossings is marked by dashed lines in the graph to the right. Notice that the internal structure of the helices does not introduce any new overcrossing be-

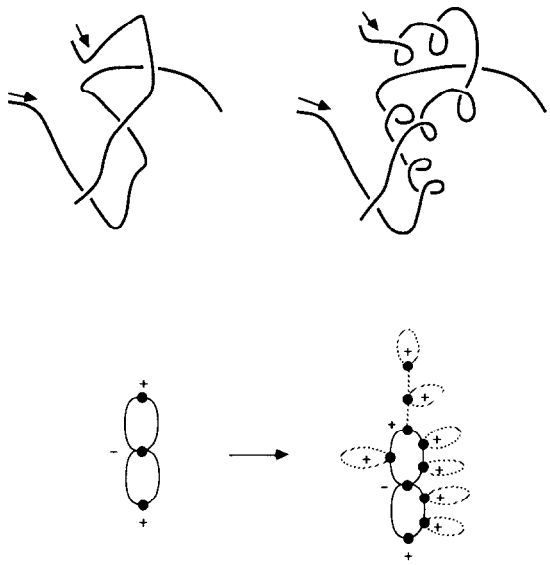


Figure 7. Front view projection of the ribbon models for the insulin protein. Top left: *Space curve obtained from the simplified ribbon model (the α helices are replaced by cylinders).* Top right: *Space curve obtained from the full ribbon model. Below each one the corresponding crossing graphs are indicated*

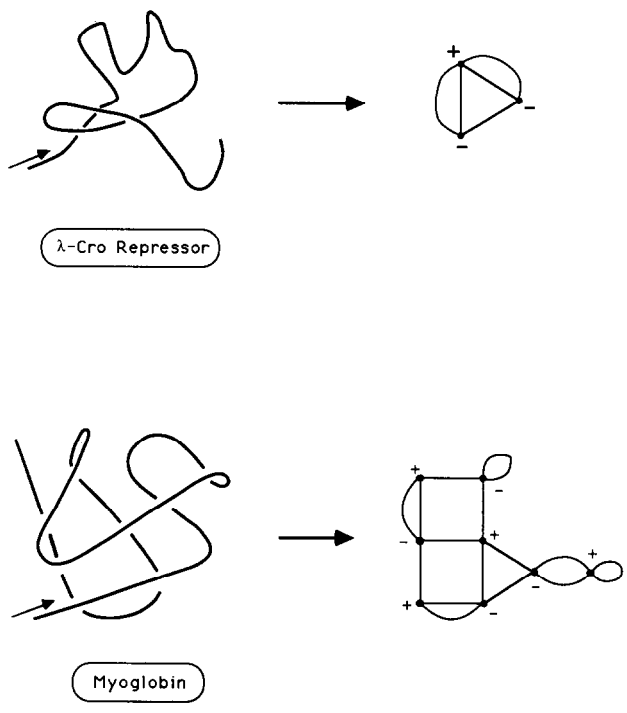


Figure 8. Front view projection of the ribbon models for the λ-Cro repressor protein and myoglobin, and their corresponding crossing graphs. Space curves were obtained from the simplified ribbon model (the α helices are replaced by their cylindrical axes)

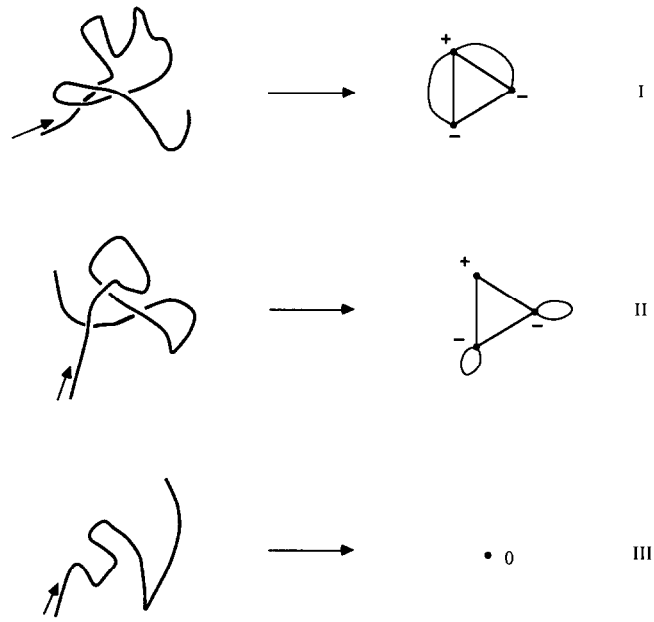


Figure 9. Graph characterization of three orthogonal projections of the λ-Cro repressor protein

tween the two sections of the space curve. That is, the α helix overcrosses only with itself. The fact that only crossings of the plus (+) type appear suggests that the axes of the α helices are close to perpendicular to the viewing direction.

Figure 8 contrasts two other examples of small proteins with skeletons represented by a single space curve. The figure displays the λ-Cro repressor protein[2] (upper) and myoglobin[56] (lower) tertiary structures, as obtained from the x-ray data. Both skeletons have been simplified by omitting the internal structure of the α helices (three in λ-Cro repressor and eight in myoglobin). Figure 9 completes the analysis of the λ-Cro repressor by providing its characterization in terms of three perpendicular projections.

The above examples illustrate how the proposed description of protein tertiary structures is actually done. In cases, when the x-ray coordinates are available, the choice of three alternative, orthogonal views for the molecular skeleton is straightforward. One can take the projections along the three Cartesian directions used for the atomic coordinates, as given, for example, in the Protein Data Bank. These directions are not arbitrary, since they are related to the axes of the protein crystal. This procedure for characterizing molecular space curves in terms of graphs, coupled with a good three-dimensional display facility (e.g., FRODO or md-TOM), provides a tool for quick analysis and labeling of macromolecular structures.[57] An illustrative example of the description provided for protein folding is given in the following section after an alternative approach based on knot theory is developed. More detailed examples of dynamical changes in other proteins will be published elsewhere.[57]

KNOT-THEORETICAL DESCRIPTION OF MOLECULAR SPACE CURVES

A mathematical knot K is a closed space curve in 3D, where the degree and type of knottedness can be characterized by

various projections of the curve onto planes.[32] A *regular projection* is one that corresponds to no degenerate crossings. If the knot is represented by a string, then the identity of the knot is not affected by changing the shape of the string, and all motions of the string are allowed as long as it is not cut or rejoined anywhere. The various allowed 3D arrangements of the string are called *placements*. It is possible to select placements so that the number of crossings of the string is minimized in a regular projection; the *crossing number* is the number of crossings obtained in such a case. If the string is given an orientation, then each nondegenerate crossing can be characterized by its handedness, as a right-handed or a left-handed crossing, as illustrated in Figure 3.

It is possible to assign polynomials to each knot, based on the handedness of crossings in any regular projection of any placement of the knot. These polynomials have a remarkable property: They are invariant to the choice of placement of the knot. That is, for a given knot, one obtains the *same* polynomial, independent of how complicated is the actual arrangement of the string, and by how much the actual number of crossings exceeds the crossing number. Hence, these polynomials are topological invariants of the knots and can be used for their characterization.[32] Among these polynomials, the Jones polynomial $V_K(t)$ of a knot K is of major interest,[34,35] since it can serve, among other roles, as a tool for detecting chirality of knots. A knot is topologically chiral if no rearrangement of the string can bring the knot into superposition with its mirror image.

Other procedures are known to characterize knots.[58,59] However, the use of polynomials presents advantages. Various polynomials can be used. Jones and Alexander polynomials are examples of one-variable polynomials, and there are also examples of multivariable polynomials.[34,35] It must be noted that more than one knot can have the same polynomial. Accordingly, a number of detailed topological descriptions have been developed to distinguish a greater number of situations. In practical use, however, one has to compromise between the number of situations that can be differentiated and the simplicity in the calculation. In our case, we will restrict the analysis to the Jones polynomials. Their relatively easy algebraic computation as well as the fact that they permit us to distinguish a large number of different knots (and their chirality properties) are important advantages.

Our interest in knot-theoretical representations stems from the fact that these polynomials provide a nonvisual shape characterization of curves in 3D space; hence, they are of relevance to the characterization of space curves representing the backbone structure of chain molecules.

Recently, there has been considerable interest in the applications of knot-theoretical techniques to polymer chains that do form actual knots.[33-43] Furthermore, the knot-theoretical polynomials have also been applied to the analysis of chirality properties of general molecules that may not form knots by themselves, leading to the concept of *chirogenicity*, and to a concise, nonvisual, computer-based analysis of molecular chirality.[40]

In the present work our problem is different. Our goal is to obtain a nonvisual, polynomial description of the shape of 2D projections of the ribbon model of chain molecules, using techniques ordinarily applied to knots. Consider a projection of the ribbon, according to the conventions described in the previous sections, and assume that all crossings are nondegenerate. The original space curve of the median line of the ribbon is not, in general, a knot, since the two end points of the median line are usually not joined. However, for the given projection we may convert the space curve of the median into a knot K_a by the following steps:

1. Attach to each end point of the space curve a straight line segment, perpendicular to the viewing plane and pointing away from the viewer. If these line segments are long enough, then they must reach a plane that is parallel with the viewing plane and lies beyond the most distant point of the original space curve.
2. Join the far ends of these line segments by another straight line segment, parallel with the viewing plane.

This procedure converts the median curve into a closed curve, that is in general a knot, denoted by K_a. (In the strict sense, the simple loop, letter O, is not knotted; hence, it is often called the *unknot* denoted by U. However, for sake of simplicity, in this study we shall refer to it as a formal member of the family of knots.) In the case of proteins, this procedure corresponds to a formal joining of the C and N termini. This operation does not have any physical interpretation; it simply provides a rule to construct a loop from an original open curve. From this loop K_a we can obtain a characterization of the original curve.

We now analyze the resulting object K_a on two levels:

(a) We describe the object itself, regarded as a knot K_a in 3D space, by taking the corresponding Jones polynomial $V_{K_a}(t)$.[34]

(b) We consider the projection of K_a to the original viewing plane, and find a new knot K_b that is compatible with the projection and preserves the most crossings. The Jones polynomial $V_{K_b}(t)$ of the knot K_b is used to characterize the projection. (Knot K_b is not necessarily unique; one may take all such K_b knots for characterization, or the first one according to some ordering, vide infra.)

In Figure 10 some examples for the simplest knots and their numerical codes[32] are shown. In Figure 11 the knots and links of five crossings or less are shown, and their Jones polynomials are also given.

The construction of the Jones polynomial $V_{K_a}(t)$ in the level (a) characterization of the ribbon model follows the standard procedure. We shall not repeat these steps here. The mathematical approach is described in detail in Refs. 34 and 35. However, for our purposes, one may use the fairly detailed, pictorial description that has been given in the chemical literature.[40]

The characterization on level (b) is a special case of a more general reconstruction and characterization problem: What is the family $\{K_b\}$ of all knots, compatible with a given projection, assuming no degenerate crossings? Of course, the original knot K_a obtained in steps 1 and 2 is sufficient to generate the given projection. However, it is common that from experimental results only a single projection is available, where the crossing information is ambiguous or partially or fully missing, and in these instances one faces a partial or the full reconstruction problem of knots from a given projection. Our problem on level (b) is related but somewhat different. Our task is to characterize the projec-

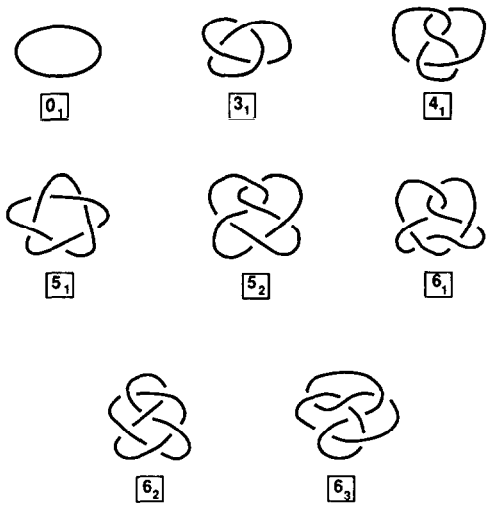


Figure 10. Knots of crossing numbers not exceeding six. For each pair of topologically chiral knots only one knot is shown. Appearances are misleading; for example, the "figure-8 knot", denoted by 4_1, appears chiral since the actual 3D arrangement of the string is chiral. However, by moving the right-hand loop of the string over the rest of the knot, and by minor "cosmetic" changes, one obtains the mirror image of the arrangement shown. Hence, this knot is topologically achiral

tion, without direct reference to the actual space curve K_a. By selecting one or several of the knots K_b that generate the same 2D projection (with crossing information suppressed), and by using their Jones polynomials $V_{K_b}(t)$, a nonvisual characterization of the projection can be obtained. The actual projection may well contain more crossings than the crossing number of the knot K_a of problem a. Since the Jones polynomial $V_{K_a}(t)$ is independent of the actual number of crossings shown by the given projection, and it depends only on the topology of the knot K_a, the characterization of the knot K_a of level (a) by the Jones polynomial $V_{K_a}(t)$ does not provide a detailed enough characterization of the projection itself. Hence, for the characterization of the projection we shall use the family of Jones polynomials $\{V_{K_b}(t)\}$ of the family $\{K_b\}$ of all knots, compatible with a given 2D projection (with crossing information suppressed), or the polynomials of selected knots K_b from the family $\{K_b\}$.

The general reconstruction problem can be treated as follows. As an illustrative example, we consider the median line of the myoglobin tertiary structure, shown in Figure 12.[56] In this example, already considered to show the application of the graph-theoretical characterization of molecular space curves, we have replaced the eight α helices by their central axes.

In general, we assume that the extension lines of steps 1 and 2 of the conversion of the median curve into the knot K_a add no new crossings and only nondegenerate new crossings, respectively, to the projection. The latter condition can always be fulfilled by an infinitesimal distortion of the ribbon model. (In the case of the example no new crossing occurs.) Hence, all n crossings of the projection can be characterized by the numbers $C_j = +1$ or -1, collected into a vector

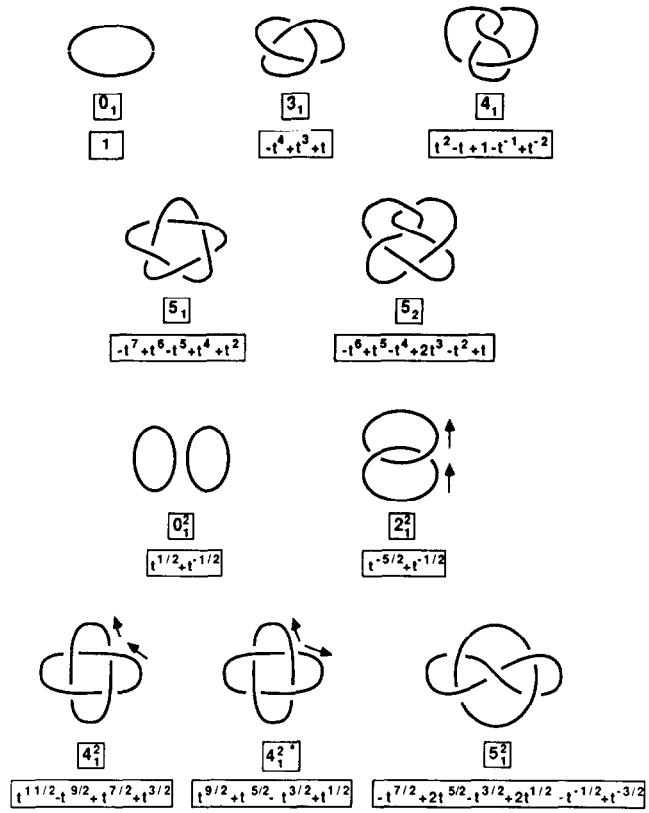


Figure 11. Knots and links of crossing number five or less and their Jones polynomials (in rectangular frame under the numerical knot symbol). These polynomials can be used for the characterization of the knots and the ribbon models they represent. The polynomial can be derived from the crossing information of any placement of the given knot. Interestingly, the same *polynomial is obtained for every placement, even for highly folded and twisted arrangements of the string, with projections showing a large number of crossings. Also note that for the mirror image* K* *of each knot* K *the polynomial can be obtained from the polynomial of the original knot by replacing the variable* t *with its reciprocal* t^{-1}. *Hence, the topological chirality of knots is implied if the Jones polynomial is different from the polynomial obtained by substituting* t *with* t^{-1}. *For example, the chirality of the trefoil knot 3_1 and the achirality of the figure-8 knot 4_1 are properly reflected in their Jones polynomials. This method can be applied for a nonvisual analysis of the chirality of the molecular ribbon model.*

$$\mathbf{C} = (C_1, C_2, \ldots, C_n) \tag{9}$$

All possible knots with the same 2D projection (with crossing information suppressed) and with arbitrarily chosen handedness for the crossings can be reconstructed by suitably modifying some or all n of the C_j numbers. By taking an n-dimensional vector

$$\mathbf{v} = (v_1, v_2, \ldots, v_n) \tag{10}$$

with elements

$$v_n = +1 \text{ or } -1 \tag{11}$$

a new vector $\mathbf{C}^v$ is generated from the reference vector $\mathbf{C}$,

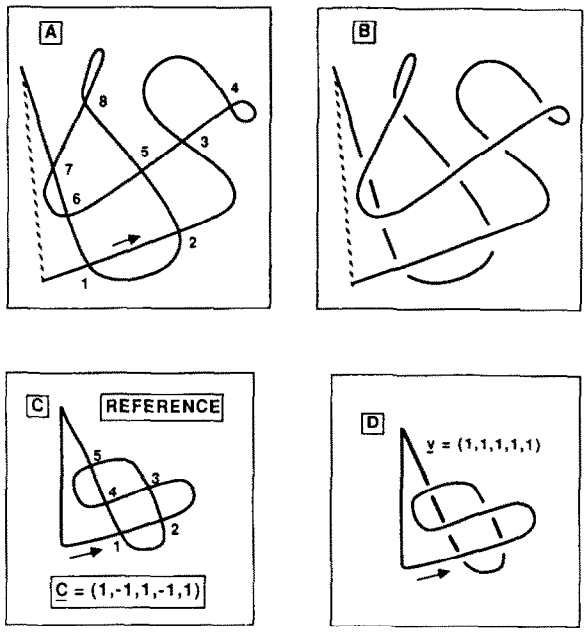


Figure 12. Reference knots for the representation of the tertiary structure of the myoglobin molecule. In part A *the oriented median line and its extension into knot is shown, following the method described in the text. The crossings of the given projection are numbered 1–8. The actual crossing information is displayed in part* B. *If our task is the reconstruction of all possible knots compatible with the planar projection of part* A, *then crossings 3, 4, and 8 are not essential and can be omitted, since in none of the possible knots compatible with the planar projection can they contribute to knottedness. This leads to the reference projection* C *where the crossings are renumbered 1–5. The actual crossing information, if available, can be specified by the vector* $\mathbf{C} = (1, -1, 1, -1, 1)$, *where the elements +1 and −1 represent right-handed and left-handed crossings, respectively. The same crossing information is also displayed in part* D, *where the elements of vector* **v** *indicate the switching of handedness relative to the reference vector* **C**, *elements 1 and −1 indicating no switch and switch, respectively. If no 3D crossing information is available, then the reference vector* **C** *may be chosen with all its elements equal to 1.*

by taking

$$\mathbf{C}^v = (C_1^v, C_2^v, \ldots, C_n^v) \tag{12}$$

of elements

$$C_j^v = v_j C_j \tag{13}$$

If crossing information for a reference projection is not available, then all elements of the reference vector **C** may be chosen as unity. By taking all the 2^n possible n-dimensional vectors **v** of the form of equation 10, the crossing vectors $\mathbf{C}^v$ of *all possible knots (and links) compatible with the given 2D projection (with crossing information suppressed) will be generated.* The family of knots obtained is denoted by $\{K_b\}$, and the corresponding family of Jones polynomials is $\{V_{K_b}(t)\}$. Note that topologically equivalent knots may be obtained by two or more different choices of **v** and $\mathbf{C}^v$ vectors, and some choices of **v** vectors may be inconsistent with the 2D projection in the sense that they cannot lead to any knot.

Since our purpose is to exploit the full characterization power of the Jones polynomials, it is of some interest to determine those knots K_b^n of family $\{K_b\}$ that cannot have simpler 2D projections than the actual 2D projection of knot K_a; that is, those knots K_b^n of family $\{K_b\}$ that have crossing numbers equal to n. If no such knot (or link) exists, then one may take a knot that has a crossing number deviating the least from the number of crossings in the projection. In some instances, certain crossings of the projection cannot contribute to knottedness; hence, they can be eliminated from the knot model (see the example in Figure 12). We shall regard n as the number of crossings obtained after eliminating those crossings that cannot contribute to knottedness. The Jones polynomials of these knots are in most instances different from, and more complicated than that of the actual knot K_a; hence, they provide more detailed information on the projection.

One may take the family of all these Jones polynomials $V_{K_b^n}(t)$ for characterization; alternatively, one may select just one of these polynomials, for example, according to the following criteria. The vectors **v** can be ordered by the lexicographic order (which would be used in a dictionary of n-letter words of an alphabet of just two letters 1 and −1) that provides an ordering of the knots K_b, and, hence, of knots K_b^n. We may choose the first K_b^n knot from the family $\{K_b\}$ for the characterization of the projection, and use its Jones polynomial $V_{K_b^n}(t)$ as a concise, nonvisual descriptor.

The myoglobin tertiary structure is represented by the projection A of knot B, shown in Figure 12,[56] and generated following steps 1 and 2. Among the eight crossings of the projection A, those of serial numbers 3, 4, and 8 can be eliminated, since for no combination of the possible choices of crossings can they contribute to knottedness. The remaining crossings generate the reference projection C and the corresponding knot D. The reference vector **C** of reference projection C is chosen as the actual crossing vector $\mathbf{C} = (1, -1, 1, -1, 1)$ of the reference knot D. However, in the general case the reference vector **C** can be chosen arbitrarily; one of the possible choices is $\mathbf{C} = (1, 1, 1, 1, 1)$. For the actual choice $\mathbf{C} = (1, -1, 1, -1, 1)$ the vector **v** of reference knot D is $\mathbf{v} = (1, 1, 1, 1, 1)$.

In Figure 13 all knot types that are compatible with the reference projection C (crossing information suppressed) and their Jones polynomials are shown. The switching vectors **v**, given with respect to reference vector $\mathbf{C} = (1, -1, 1, -1, 1)$, are also specified. The *cake knot* 5_2^*, where the asterisk indicates that this knot is the mirror image of the standard cake knot 5_2, is the first knot in the order of switching vectors **v** that has the maximum possible crossing number; in the case of the example, this number is $n = 5$. If the standard procedure for knot pairs that are mirror images is followed,[40] then the corresponding Jones polynomial

$$V(t) = -t^{-6} + t^{-5} - t^{-4} + 2t^{-3} - t^{-2} + t^{-1} \tag{14}$$

is obtained from that of the standard cake knot 5_2, by replacing the variable t with t^{-1}. This polynomial provides a

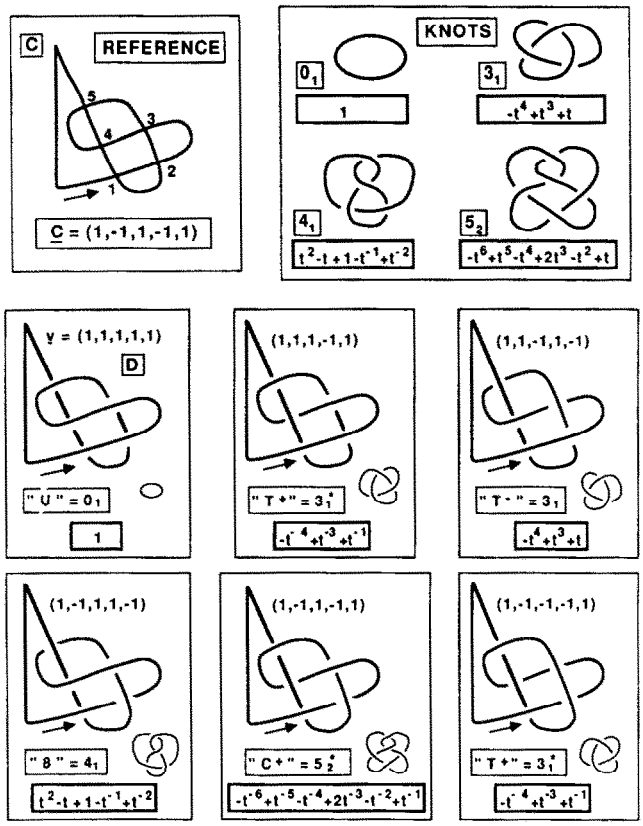


Figure 13. The collection of knots compatible with the planar reference projection of the myoglobin tertiary structure. By generating all possible assignments of switching vectors **v** *to the reference vector* **C**, *most resulting knots are equivalent to the unknot* U, *whereas in other cases both trefoil knots* 3_1 *and* 3_1^*, *the figure-8 knot, and the "cake knot"* 5_2^* *are obtained. Besides the original reference unknot* U *of vector* **v** = (*1, 1, 1, 1, 1*), *only those knots are shown which are different from* U *and precede the unknot of vector* **v** = (*1, −1, −1, −1, −1*) *in the order provided by the vectors* **v**. *The remaining knots are the mirror images of those shown. Among all these knots, the first occurrence of a knot with the highest possible crossing number, cake knot* 5_2^*, *belongs to vector* **v** = (*1, −1, 1, −1, 1*). *If no crossing information is available, then this knot and its Jones polynomial are used for the topological characterization of the projection.*

concise, nonvisual characterization of the given projection of the backbone of the myoglobin tertiary structure.

As a second example, we consider the simplified tertiary structure of the λ-Cro repressor protein, studied already in a previous section (Figure 8). The projection of the space curve (projection I in Figure 9) exhibits only three crossovers. The reference vector **C** of reference projection C is the actual crossing vector **C** = (−1, 1, −1) of the reference knot *K*. For this choice **C** the vector **v** of reference knot *K* is **v** = (1, 1, 1). In the present case the knot *K* is an unknot.[32]

In Table 4 we present the results of all knot types and their Jones polynomials that are compatible with the reference projection C. The switching vectors **v**, given with respect to reference vector **C** = (−1, 1, −1), are also specified. The most complicated knots occurring are the left- and right-handed trefoil knots (3_1 and 3_1^*, respectively). The Jones polynomials

$$V_{3_1}(t) = -t^4 + t^3 + t \tag{15a}$$

$$V_{3_1^*}(t) = V_{3_1}\left(\frac{1}{t}\right) \tag{15b}$$

are obtained, where we have used the rule of $t \rightarrow 1/t$ replacement for knot pairs that are mirror images.[40] This polynomial provides a concise characterization of the given projection of the backbone of the protein tertiary structure.

Table 4. Knot-theoretical characterization of a projection of the simplified ribbon model for the λ-Cro repressor protein[a]

Switching vector	Knot type	Jones polynomial
(1, 1, 1)	0_1	1
(−1, 1, 1)	0_1	1
(1, −1, 1)	3_1	$-t^4 + t^3 + t$
(1, 1, −1)	0_1	1
(−1, −1, 1)	0_1	1
(−1, 1, −1)	3_1^*	$-t^{-4} + t^{-3} + t^{-1}$
(1, −1, −1)	0_1	1
(−1, −1, −1)	0_1	1

[a]The projection of the molecular space curve exhibits three crossings.

The above description can be applied to characterize shape changes along a folding path. To illustrate this type of application we use again the λ-Cro repressor protein, since its simplicity allows a concise characterization.

We consider here a model of a protein folding given in terms of simple conformational rearrangements. For the sake of illustration, we restrict the analysis to the structures produced by rigid rotation of two bonds in the λ-Cro repressor protein. The two rotations, represented by angles ϕ_1 and ϕ_2, are indicated in Figure 14. These two rotations represent the change in the relative orientations among the three α helices. As starting point for the rigid rotations ($\phi_1 = 0$, $\phi_2 = 0$) we consider projection I (cf. Figures 8 and 9). The viewing direction is fixed to be the same as for the initial structure $\phi_1 = 0$, $\phi_2 = 0$, while performing the rotations.

The lower part of Figure 14 shows how the configuration subset represented by the ϕ_1 and ϕ_2 torsions can be partitioned into regions characterized by different shape descriptors. This diagram can be constructed by visual inspection of computer displays or a wire model of the molecular backbone. The shape of the boundaries between the *shape regions* is therefore qualitative. Notice the occurrence of two types of boundaries. Those parallel to the $\phi_1 = 0$ axis represent the occurrence of a change of crossing patterns between the axes of the α_1 and α_2 helices. This crossing is independent of the ϕ_2 rotations. On the other hand, curved boundaries correspond to the appearance of overcrossing of the tail loop of the protein with the protein head. This crossing depends on both angles.

The regions labeled with various numbers correspond to

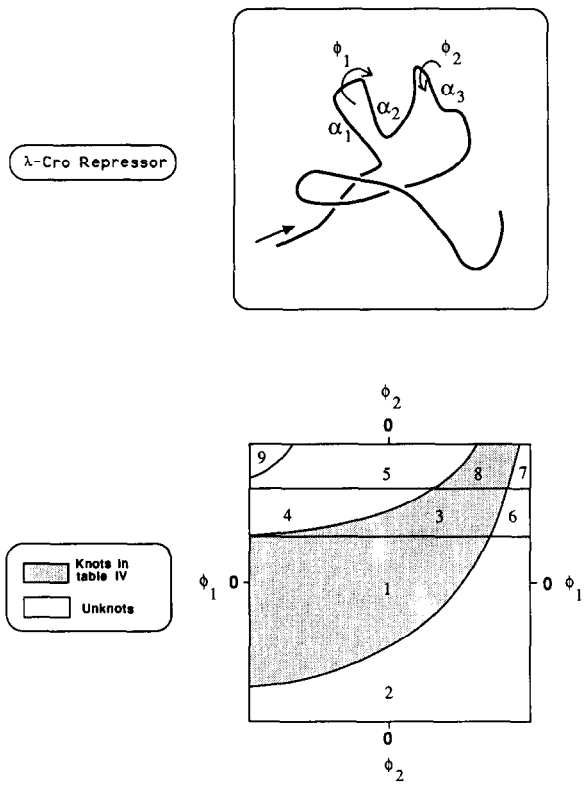


Figure 14. Description of a model process of protein folding. The system considered is the λ-Cro repressor protein undergoing rotations indicated by angles ϕ_1 and ϕ_2. These angles correspond to rotations around C–N bonds in the loops linking α helices α_1 and α_2, and α_2 and α_3, respectively, as indicated in the upper part of the figure. The viewing direction is kept constant. The rotation by ϕ_1 is defined so that only the piece of the backbone stretching from the head up to the α_1 helix moves. The rotation about ϕ_2 moves only the backbone segment from the helix α_3 up to the tail. The case of $\phi_1 = 0$ and $\phi_2 = 0$ corresponds to the x-ray structure. Each region indicated in the map by a number represents configurations with the same shape description, as given by graphs. Shaded and unshaded regions correspond to different knot-theoretical descriptions. The shape of the boundaries between regions is qualitative.

the occurrence of structures with various graphs when projected. These structures are indicated in Table 5, characterized by the crossing vectors of their graphs. By contrast, the shading represents the knot-theoretical description. Shaded regions (those characterized by the graphs 1, 3, and 8) have a knot description, as in Table 4. Unshaded regions correspond to projections that have a single knot descriptor, the unknot. The graph description happens to be more detailed, a characteristic likely to occur in most examples with relatively few crossings.

A folding path would appear as a curve in the $\phi_1 - \phi_2$ map. The diagram in Figure 14 allows one to recognize the extent in configuration space of the various shape types the molecular backbone is adopting while folding. For example, the shaded regions correspond to the most entangled structures, while unshaded regions stand for the more open structures. The diagram clearly shows the range of values for torsion angles where these closed and open structures are more likely to be found.

Table 5. Graph characterization in terms of crossing vectors for projection *I* of λ-Cro repressor[a]

Structure	Crossing vector
1	(−1, 1, −1)
2	(−1)
3	(−1, 1, 1, −1, −1)
4	(1, −1, 1, −1, −1)
5	(1, −1, 1, −1)
6	(1, −1, −1)
7	(1, −1)[b]
8	(−1, 1, 1, −1)
9	(1, −1)[b]

[a]The structures correspond to the graphs found by performing rotations in the angles ϕ_1 and ϕ_2 shown in Figure 14.
[b]Structures 7 and 9, though having the same crossing vectors, can be distinguished by their graphs.

This example provides an illustration of how to use graph- and knot-theoretical descriptors for describing the shape of molecular space curves under continuous deformation. More examples for actual dynamical changes in small proteins will be published in a forthcoming article.[57]

CONCLUSIONS

In this work we have proposed two methods for characterizing macromolecular ribbon models and their foldings. The graph-theoretical method consists of generating the overcrossing graphs of the plane curves obtained by plane projections of a space curve associated to the ribbon along three preferential directions.

A second method is to derive new space curves from the original one, by "switching" its crossing pattern in a nontrivial manner. As we have seen, this transformation may lead to curves of different knot type, thus providing a simple algebraic characterization to the projected image, hence to the placement of the original curve.

Both procedures can be implemented in an automated way and carried out without any subjective, visual inspection. Using either one of the two methods, it is possible to provide a quick, algebraic description of ribbon models, and to analyze ribbons without the need for a graphical, three-dimensional display. This feature makes the present approach potentially valuable in computer-aided drug design and in the modeling of protein–drug interactions, where it can be used as a nonvisual research tool, complementing the graphical analysis of molecular shape.

APPENDIX

The purpose of this appendix is to provide a more precise technical discussion of the graph-theoretical characterization of ribbon models. The basic ideas have been discussed in the text, and will not be repeated here.

Let $\mathbf{r}(t)$ be the space curve determined from the ordinary

or thick ribbon models, as discussed in the text (equation 1):

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} \qquad t \in I = [0, 1] \qquad (A.1)$$

The parametric equation A.1 defines the set of points belonging to the space curve.

Any section of the curve, restricted to a subinterval T of I, can thus be represented as a set:

$$c(T) = \{\mathbf{p} \in {}^3\mathbb{R}: \mathbf{p} = \mathbf{r}(t), t \in T \subset I\} \qquad (A.2)$$

The set $c(I)$ represents the whole curve.

Let (Q_1, Q_2, Q_3) be the triplet of Cartesian axes of inertia, attached to the center of mass of the space curve (see text). If $P(Q_i)$ indicates the projection to a plane $Q_i = Q_i$, with Q_i a constant coordinate value on the axis Q_i, then the result of this operation is a plane curve that will be denoted as $\mathbf{q}_i(t)$:

$$\mathbf{q}_i(t) = P(Q_i)\mathbf{r}(t) \qquad t \in I \qquad (A.3)$$

By analogy with equation 2, let us introduce the sets $c_i(T)$, $i = 1, 2, 3$, representing subsections of these projected plane curves:

$$c_i(T) = \{\mathbf{p} \in {}^2\mathbb{R}: \mathbf{p} = \mathbf{q}_i(t), t \in T \subset I\} \qquad (A.4)$$

with $c_i(I)$ the whole plane curve. The characterization of the space curve $\mathbf{r}(t)$ is transformed into the description of three planar curves, represented by the sets $c_1(I)$, $c_2(I)$, and $c_3(I)$.

As discussed in the text, we associate a graph g_i to each plane curve $\mathbf{q}_i(t)$. These graphs will have the curve's crossings as vertices, and the segments of the curve connecting crossing points as edges. These segments can connect two different crossing points, or a crossing point with itself. Segments of the curve not providing vertex–vertex connections are dropped from the graph. From these intuitive notions the formal definition of the graph $g_i = g[c_i(T)]$, associated with the set $c_i(T)$ of points in the curve $\mathbf{q}_i(t)$, can be given as follows.

Vertices. Take the ith projected curve, $\mathbf{q}_i(t)$, and let U_i be defined as the set of those points of curve $\mathbf{q}_i(t)$ that belong to more than one value of parameter t:

$$U_i = \{\mathbf{q}_i(t): \mathbf{q}_i(t) = \mathbf{q}_i(t') \qquad t \neq t'\} \qquad (A.5)$$

The maximum connected components ν_{ij} of U_i are the vertices:

$$U_i = \bigcup_j \nu_{ij} \qquad \nu_{ij} \cap \nu_{ij'} = \varnothing \qquad j \neq j' \qquad (A.6)$$

In each set ν_{ij} the infimum of parameter values t of points $\mathbf{q}_i(t) \in \nu_{ij}$ is denoted by t_j. The indices $j = 1, 2, \ldots$, follow the order of the infimum values t_j. The vertices are collected in an ordered set $V(g_i)$:

$$V(g_i) = (\nu_{i1}, \nu_{i2}, \nu_{i3}, \ldots) \qquad (A.7)$$

According to this definition, any nonpathological bounded curve will generate a graph with a finite number of vertices. A vertex may be a pair of lines overlapping along some finite length. In the particular case where each component ν_{ij} is a single point set, the vertices are actually given by a finite number of crossing points on the curve.

With this generalized definition of vertices, it is possible that an entire curve is associated with a trivial graph. For example, the view of an α helix along its rotational axis shows a circle due to the overcrossing of every helicoidal loop. Accordingly, this view will have an associated graph consisting of only one vertex.

Edges. Take all nonvertex points of curve $\mathbf{q}_i(t)$,

$$W_i = \{\mathbf{q}_i(t) \; \forall \; t\} \setminus U_i \qquad (A.8)$$

and denote the maximum connected components of W_i by ω_{ik}:

$$W_i = \bigcup_k \omega_{ik} \qquad \omega_{ik} \cap \omega_{ik'} = \varnothing \qquad k \neq k' \qquad (A.9)$$

A set

$$e_i(j, j', k) = \nu_{ij} \cup \nu_{ij'} \cup \omega_{ik} \qquad (A.10)$$

is an edge between vertices ν_{ij} and $\nu_{ij'}$ if and only if it is connected. The edge set $E(g_i)$ is the family of edges:

$$E(g_i) = \{e_i(j, j', k): e_i(j, j', k) \text{ connected}\} \qquad (A.11)$$

ACKNOWLEDGMENTS

This work was supported by operating and strategic research grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada. P.G.M. acknowledges stimulating discussions with Gerald Maggiora and Mark Johnson of the Computational Chemistry Group, The Upjohn Company. G.A.A. acknowledges inspiring discussions with Orlando Tapia, while visiting the Uppsala Biomedical Centre, Swedish University of Agricultural Sciences. We thank also Professors Whittington and van Rensburg (University of Toronto) for bringing some useful references to our attention.

REFERENCES

1. Richardson, J. S. *Adv. Protein Chem.* 1981, **34**, 167
2. Richardson, J. S. *Methods Enzymol.* 1985, **115**, 359
3. Carson, M., and Bugg, C. E. *J. Mol. Graphics* 1986, **4**, 121
4. Carson, M., *J. Mol. Graph.* 1987, **5**, 103
5. Lesk, A. M., and Hardman, K. D. *Science* 1982, **216**, 539
6. Lesk, A. M., and Hardman, K. D. *Methods Enzymol.* 1985, **115**, 381
7. Richarson, J. S. *Methods Enzymol.* 1985, **115**, 341
8. Carbó, R., Leyda, L., and Arnau, M. *Int. J. Quantum Chem.* 1980, **17**, 1185
9. Martín, M., Sanz, F., Campillo, M., Pardo, L., Pérez, J., and Turmo, J. *Int. J. Quantum Chem.* 1983, **23**, 1627
10. Bowen-Jenkins, P. E., Cooper, D. L., and Richards, W. G. *J. Phys. Chem.* 1985, **89**, 2195
11. Richard, A. M., and Rabinowitz, J. R. *Int. J. Quantum Chem.* 1987, **31**, 309
12. Richard, A. M., and Rabinowitz, J. R. *Int. J. Quantum Chem.* 1988, **34**, 207
13. Mezey, P. G. *Int. J. Quantum Chem. (Quant. Biol. Symp.)* 1986, **12**, 113
14. Mezey, P. G. *J. Comput. Chem.* 1987, **8**, 462
15. Mezey, P. G. *Int. J. Quantum Chem. (Quant. Biol. Symp.)* 1987, **14**, 127

16. Arteca, G. A., and Mezey, P. G. *Int. J. Quantum Chem. (Quant. Biol. Symp.)* 1987, **14**, 133
17. Arteca, G. A., and Mezey, P. G. *J. Comput. Chem.* 1988, **9**, 554
18. Arteca, G. A., and Mezey, P. G. *J. Mol. Struct. Theochem.* 1988, **166**, 11
19. Arteca, G. A., and Mezey, P. G. *Folia Chim. Theor. Lat.* 1987, **15**, 115
20. Mezey, P. G. *J. Math. Chem.* 1988, **2**, 325
21. Arteca, G. A., Jammal, V. B., Mezey, P. G., Yadav, J. S., Hermsmeier, M. A., and Gund, T. M. *J. Mol. Graphics* 1988, **6**, 45
22. Arteca, G. A., Jammal, V. B., and Mezey, P. G. *J. Comput. Chem.* 1988, **9**, 608
23. Connolly, M. L. *Visual Comput.* 1987, **3**, 72
24. Harary, F., and Mezey, P. G. *J. Math. Chem.* 1988, **2**, 377
25. Mezey, P. G. *J. Math. Chem.* 1988, **2**, 299
26. Arteca, G. A., and Mezey, P. G. *Int. J. Quantum Chem.* 1988, **34**, 517
27. Arteca, G. A., and Mezey, P. G. *Theor. Chim. Acta* 1989, **75**, 355
28. Åqvist, J., and Tapia, O. *J. Mol. Graphics* 1987, **5**, 30
29. Leicester, S. E., Finney, J. L., and Bywater, R. P. *J. Mol. Graphics* 1988, **6**, 104
30. See, for example, Munkres, J. R. *Elements of Algebraic Topology*. Addison-Wesley, Menlo Park, CA, 1984
31. Lass, H. *Vector and Tensor Analysis*. McGraw-Hill–Kogakusha, Tokyo, 1950
32. See, for example, Crowell, R. H., and Fox, R. H. *Introduction to Knot Theory*. Springer-Verlag, Berlin, 1977
33. Walba, D. M. Stereochemical topology. In *Chemical Applications of Topology and Graph Theory*, R. B. King, Ed. Elsevier, Amsterdam, 1983
34. Jones, V. F. R. *Bull. Am. Math. Soc. (NS)* 1985, **12**, 103
35. Freyd, P., Yetter, D., Hoste, J., Lickorish, W. B. R., Millett, K., and Ocneanu, A. *Bull. Am. Math. Soc. (NS)* 1985, **12**, 239
36. Walba, D. M. *Tetrahedron* 1985, **41**, 3161
37. Wasserman, S. A., and Cozzarelli, N. R. *Science* 1986, **240**, 110
38. Connolly, M. L., Kuntz, I. D., and Crippen, G. M. *Biopolymers* 1980, **19**, 1167
39. Kikuchi, T., Némethy, G., and Scheraga, H. A. *J. Comput. Chem.* 1986, **7**, 67
40. Mezey, P. G. *J. Am. Chem. Soc.* 1986, **108**, 3976
41. Millett, K. C. *J. Comput. Chem.* 1987, **8**, 536
42. Simon, J. *J. Comput. Chem.* 1987, **9**, 718
43. Sumners, D. W. *J. Math. Chem.* 1987, **1**, 1
44. Delbrück, M. *Proc. Symp. Appl. Math.* 1962, **14**, 55
45. Fuller, F. B. *Proc. Symp. Appl. Math.* 1962, **14**, 64
46. Fuller, F. B. *Proc. Natl. Acad. Sci. USA* 1971, **68**, 815
47. Le Bret, M. *Biopolymers* 1979, **18**, 1709
48. De Santis, P., Morosetti, S., and Palleschi, A. *Biopolymers* 1983, **22**, 37
49. Hao, M.-H., and Olson, W. K. *Biopolymers* 1989, **28**, 873
50. Karplus, M., and McCammon, J. A. *Annu. Rev. Biochem.* 1983, **53**, 263
51. Colonna-Cesari, F., Perahia, D., Karplus, M., Eklund, H., Brändén, C. I., and Tapia, O. *J. Biol. Chem.* 1986, **261**, 15273
52. Tapia, O., Eklund, H., and Brändén, C. I. Molecular, electronic, and structural aspects of the catalytic mechanism of alcohol dehydrogenase. In *Steric Aspects of Biomolecular Interactions*, G. Náray-Szabó and K. Simon, Eds. CRC Press, West Palm Beach, FL, 1987
53. Jaenicke, R. *Prog. Biophys. Molec. Biol.* 1987, **49**, 117
54. Landau, L. D., and Lifshitz, E. M. *Mechanics*. Pergamon, Oxford, 1960
55. Curve *A* actually corresponds to the space curve associated to the ribbon-like picture by M. Escher, "Bonds of Eternal Union." [Reproduced in *J. Mol. Graphics* 1988, **6**, 67]
56. Rawn, J. D. *Biochemistry*. Patterson, Burlington, NC, 1989
57. Arteca, G. A., Tapia, O., and Mezey, P. G., to be published.
58. Dowker, C. H., and Thistlethwaite, M. *C. R. Acad. Sci. (Canada)* 1982, **VI**(2), 129; *Topology and Its Applicat.* 1982, **16**, 19
59. Thistlethwaite, M. *London Math. Soc. Lecture Notes* 1985, **93**, 1