Small- and large-scale global shape features in macromolecular backbones

Gustavo A. Arteca*
Department of Chemistry, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0

A methodology to analyze and characterize small- and large-scale shape and structural features in a macromolecule is presented. The procedure involves the construction of a continuum of spherical shape maps. This continuum is built by analyzing the overcrossing pattern of a molecular space curve from every direction in 3-space as seen by a viewer with a reduced vision field. The method derives global shape descriptors (shape maps) that characterize the backbone as a whole (large-scale structure), but also allows one to focus the analysis on details of the backbone's small-scale structure. A continuous function is proposed as a simplified descriptor derived from the shape maps. This function discriminates reasonably well among elements of secondary structure and supersecondary structural motifs in proteins. The procedure is applied to α-helices, β-sheets, and models of α-helical packing.

1. Introduction

The basic notion in rational drug design is the expectation that molecules with similar shapes will exhibit comparable physical and chemical properties [1]. Consequently, current studies in computer-assisted molecular pharmacology rely heavily on techniques and criteria to make quantitative assessments of molecular shape and molecular similarity. In this work, we are concerned with the characterization of some aspects of the three-dimensional shape of large chain molecules.

The analysis of biomacromolecules requires an approach completely different from that used for small molecules. Not only do the models used to convey the three-dimensional shape of the molecule have to be different, but it is also necessary to pay attention to quite different shape features. For instance, many properties of large molecules such as proteins depend mostly on *global folding characteristics* rather than on local surface features [2–7].

The simplest model of a chain molecule that describes its essential folding features is the so-called molecular backbone. This is a space curve defined by the sequence of main chain atoms.

In our case, we are especially interested in proteins. Here, the backbone is defined by the sequence of straight-line segments joining neighboring α-carbon atoms belonging to each amino acid residue. Protein backbones usually present a

*Present address: Département de Chimie et Biochimie, Laurentian University, Ramsay Lake Road, Sudbury, Ontario, Canada P3E 2C6.

hierarchical organization provided by the occurrence of some basic substructures such as helices, strands, and turns, which are known as the *elements of secondary structure* (*or secondary-structural motifs*) in a protein. Recognizing the occurrence of such motifs is one of the aims of any characterization techniques.

There are several procedures proposed to describe quantitatively the shapes of protein backbones. Some of them, such as the well-known Ramachandran diagrams [3–7] and distance maps [8], provide a very detailed local geometrical description in terms of angles and distances.

In contrast, topological (as opposed to geometrical) techniques provide *a global description of the backbone's shape*. These approaches use the same basic input information regarding the location of nuclei, but describe the shape in terms of features which consider the backbone as a whole entity. Many of these characterizations do not essentially change for small nuclear motions and thus provide a criterion to estimate the role of configurational rearrangements in distorting the molecular shape.

Graph-theoretical and simple topological descriptors have been proposed to classify protein structures and assist in homology searches in data-bases [9–16]. Louie and Somorjai [17–19] have applied differential topology to the study of molecular space curves. In the present work, we continue the development of a methodology to characterize protein backbones based on the knowledge of the complete pattern of overcrossings [20–24]. The procedure uses geometric information (the overcrossings between segments of the backbone) to derive a global topological characterization of the fold.

The method providing shape descriptors for a backbone is discussed in refs. [20,24]. The technique is based on the analysis of the crossing patterns of a macromolecular backbone viewed from all possible directions in space. The procedure assigns a *spherical shape map* to each given molecular backbone. A spherical shape map classifies the points on the spherical surface of a ball enclosing the backbone into regions according to the crossings in plane curves derived by projecting the backbone to a plane tangent to the sphere at each point. As a result, one has a classification into equivalence classes of the 3-space surrounding the molecular curve. Each equivalence class is characterized by a distinct value of a shape descriptor. The simplest of such shape descriptors is the number of crossings in the projected plane curves. We shall refer to this approach as a *large-scale global shape characterization of the backbone*, since the complete crossing pattern is taken into account.

In this work, we generalize this method by also incorporating into the analysis the *small-scale shape features*. These features correspond to the details of the relative distances between crossings and their distribution in 3-space. Again, the procedure is global, but in this case it is possible to discriminate the occurrence of special arrangements at small geometrical scale from those at large geometrical scale in the analysis.

The method we propose involves the construction of a *continuum of spherical shape maps* built from the overcrossing pattern of a space curve viewed from every

direction in space by an observer with a reduced vision field. We introduce a continuous function derived from the reduced-view spherical shape maps that discriminates among structural motifs in proteins. The procedure is illustrated with example motifs and the packing of α-helices.

2. The reduced-view spherical shape maps of a space curve

The methodology we present in this section is an extension of the procedure developed in refs. [20,24] for the characterization of molecular space curves. In this section, we briefly discuss the original approach and the present generalization.

A molecular space curve can be represented as a parametric function $\boldsymbol{r}(t)$ in 3-space in terms of the parameter t, $t \in [0, 1]$. This curve is bounded and oriented. The starting point of the curve is $\boldsymbol{r}(0)$ and the end point is $\boldsymbol{r}(1)$. In the case of protein backbones, the curve $\boldsymbol{r}(t)$ is defined by a sequence of straight-line α-carbon segments $\boldsymbol{r}_i - \boldsymbol{r}_{i+1}$, $i \geq 1$, where $\boldsymbol{r}_i$ is the position vector of the ith α-carbon atom on the curve, where $i = 1, 2, \ldots, n$, with n being the number of amino acid residues.

The essential idea behind the characterization method is to explore *the details of the overcrossing pattern of the space curve from all directions in 3-space.* In order to take into account all possible viewing directions for the backbone, we enclose it inside the smallest possible sphere centred at the geometrical centre $\boldsymbol{r}_0$ of the curve [24]. We shall indicate with S and R the spherical surface of the ball enclosing the curve and its radius, respectively. Thus, we have:

$$S = \{s \in {}^3\mathbf{R} : \|s - \boldsymbol{r}_0\| = R\}, \tag{1}$$

where $\|\boldsymbol{r}_i - \boldsymbol{r}_0\| \leq R$ for all $i = 1, 2, \ldots, n$. Radius R is the smallest value that satisfies this condition.

When viewed from an arbitrary position in 3-space, the segments of the curve $\boldsymbol{r}(t)$ may overcross. These features have been used to characterize the curves in terms of graphs and their associated matrices [20,21]. The same information can be used to compute topological invariants such as knot polynomials [20–23].

The vectors obtained by joining the centre of the ball with each point on the spherical surface S define viewing directions used to analyze the backbone. Let $\boldsymbol{\upsilon}$ be a generic viewing direction. The situation is depicted schematically in fig. 1. Here, we show a molecular space curve enclosed in the sphere S and one viewing direction $\boldsymbol{\upsilon}$. An observer located very far from the sphere will see the backbone as a 2D-photograph obtained by projecting the curve to a plane tangent to the sphere and perpendicular to $\boldsymbol{\upsilon}$. If one indicates the projection operation to the tangent plane as $P_{\boldsymbol{\upsilon}}$, then $P_{\boldsymbol{\upsilon}}\{\boldsymbol{r}(t)\}$ is a plane curve derived from the space curve. This plane curve exhibits *actual crossings* wherever the original space curve has *overcrossings* [20–24]. These crossings allow one to construct the so-called *spherical shape map* of a space curve [24]. This map provides a global shape characterization of the backbone. The characterization makes use of large-scale shape features since all the overcrossings in the backbone are accounted for.

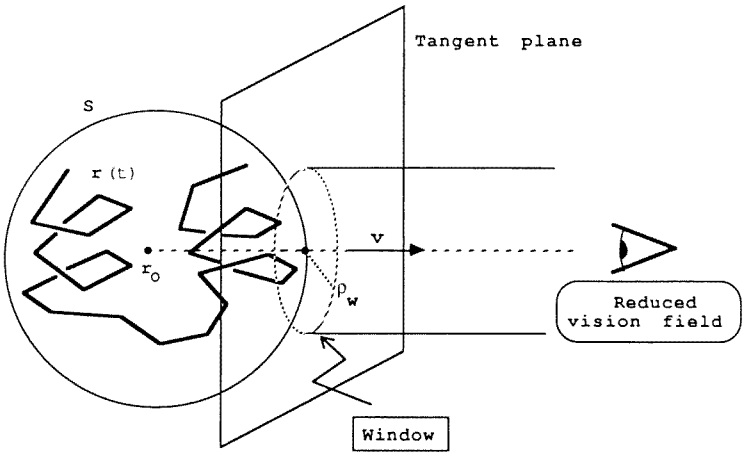

Fig. 1. Schematic representation of a protein backbone enclosed in the sphere S. The geometrical centre of the backbone coincides with the centre of the sphere. The viewing direction υ is indicated by a dashed line joining the viewer with the centre. The dashed-line circle represents the restricted viewing window (with radius ρ_w) used to monitor details in the molecular shape of the backbone.

Another, more thorough characterization can be formulated within the same approach. For instance, one can describe the changes in crossing patterns from every viewing direction by considering the distance at which the overcrossings occur and the size of the backbone. In other words, one can pay attention to changes in shape features when moving from large-scale to small-scale.

In this work, we propose an approach which accomplishes this by monitoring the view of the projected curve by means of a *circular viewing window*. This is shown in fig. 1, where a circular window of radius ρ_w is centred along the viewing direction. Figure 2 clarifies the role of the viewing window in the shape characterization. The diagram on the left-hand side of fig. 2 shows a generic projected plane curve derived from the backbone. The curve can be seen inside a circular window of radius R. Within this view (a "full" projection), we observe four crossings. The right-hand side diagram shows the view reduced by the window. (Note that the window is indicated as a full-line circle in the left-hand side diagram.) Within this restricted view, the original four crossings are reduced to two. The curve will exhibit two crossings for a range of values of the window radius. If ρ_w is small enough, the plane curve will have *no crossings* for this viewing direction. On the other hand, there is a critical value of the window radius, $\rho_w^0 \leq R$, above which the projected curve shows the same overcrossing pattern as the full projection $P_\upsilon\{r(t)\}$.

In order to have a useful but simple characterization scheme, we shall take into consideration only the number of crossings. The handedness of each of the crossings can also be incorporated by computing the *crossing index vector* characterizing

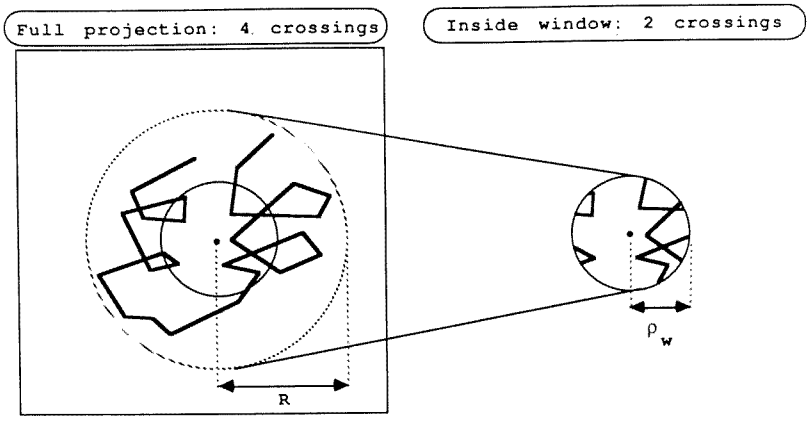

Fig. 2. Small-scale shape features revealed by the viewing window. The diagram on the left-hand side shows a full planar projection of a molecular backbone enclosed in the sphere with radius R. The full-line circle in the centre corresponds to the viewing window. The diagram to the right shows the restricted view within the window. Since the procedure ignores all details outside the window, the number of crossings is reduced from four to two in this case.

the view (cf. ref. [24]). However, we shall see that the number of crossings associated with a given view and a given window radius is sufficient to provide a quite detailed analysis.

Let $\nu(\rho_w; \upsilon)$ be the number of crossings exhibited by the projected curve $P_\upsilon\{r(t)\}$ for a window of radius ρ_w. According to the above discussion,

$$\nu(\rho_w; \upsilon) \leq \nu(\rho_w^0; \upsilon) = N^*(\upsilon), \quad 0 \leq \rho_w \leq \rho_w^0 \leq R. \tag{2}$$

The *integer number function* $\nu(\rho_w; \upsilon)$ is a *shape descriptor* of the curve viewed along the given direction. The same approach can be followed for any other viewing direction υ. If the procedure is repeated for all points on the sphere S, we can divide the sphere itself into equivalence regions on the condition that points on each region define viewing directions leading to the same number of crossings for a given window radius ρ_w.

We can introduce a single integer number N as the one feature that all points belonging to an equivalence class have in common. This integer N can take any value between 0 (no crossings) and $\nu^*(\rho_w)$, which is the maximum number of crossings observed on the sphere with a window of radius ρ_w (that is, $\nu^* = \max \nu(\rho_w; \upsilon)$, over υ). One such equivalence class is defined as the set $S_N(\rho_w)$:

$$S_N(\rho_w) = \{\upsilon \in S, \ \nu(\rho_w; \upsilon) = N\}, \quad N = 0, 1, \ldots, \nu^*(\rho_w). \tag{3}$$

Note that in practice *there is always* a window radius ρ_w for which the set $S_0(\rho_w)$ *is nonempty*. This is due to the fact that a real, macromolecular backbone cannot fill the ball completely and, accordingly, there will always exist a radius ρ_w small enough so that the curve does not exhibit crossings for some projection. The same cannot be said for other regions $S_N(\rho_w)$, $N \geq 1$. We shall show below that the study of the region on the sphere where the projected plane curve has no crossings ($N = 0$) may suffice to derive a detailed characterization of the backbone.

Each of the equivalence classes on the spherical surface can be formed by a number of disjoint regions (maximum components) $S_N^{(i)}(\rho_w)$,

$$S_N(\rho_w) = \bigcup_i S_N^{(i)}(\rho_w), \tag{4}$$

whose number and relative location depends on the type of space curve $\boldsymbol{r}(t)$. These equivalence classes provide a partitioning and a topological structure to the spherical surface S:

$$S = \bigcup_{N=0}^{\nu^*} S_N(\rho_w). \tag{5}$$

A partitioning of the 3-space around the curve $\boldsymbol{r}(t)$ given by (5) is *a reduced-view spherical shape map*. By taking $\rho_w = R$, we find the special type of spherical shape maps discussed in ref. [24].

In summary, we can encode some essential shape features of the backbone's fold by using a continuum of spherical shape maps derived from the overcrossing pattern revealed by windows of varying size. As a result, the shape characterization of a fold is transformed into the characterization of the shape regions (equivalence classes) on the sphere S. Macromolecules exhibiting comparable shape features would change their spherical shape maps as a function of ρ_w in a similar manner.

The continuum of spherical projection maps is computed in a totally automated manner by using a program developed in our laboratory [25]. In what follows, we shall restrict ourselves to the analysis of the regions on S with no crossings, instead of the whole maps. We shall refer to these as $S_0(\rho_w)$ maps.

Figure 3 shows our first illustrative example of the shape characterization provided by the $S_0(\rho_w)$ maps. The space curve considered is a model 13-atom perfectly cylindrical helix. The helix has three loops and a nearest-neighbor distance of 1.0 Å. The four diagrams in fig. 3 represent only hemispheres (the hemispheres not displayed are derived by an inversion of the map through the centre of the sphere). The dotted sections represent the directions in 3-space where the projected helix does not overcross itself, that is, the $S_0(\rho_w)$ region. For this example, one has $\rho_w^0 < 1$ Å. Accordingly, the upper left diagram corresponds to the full-projection spherical shape map [24]. The features shown are characteristic of helices, since these do not exhibit overcrossings when observed along directions close to the normal to the helical axis [24]. The reduction of the window radius ρ_w shows that, predictably, there are more directions in 3-space along which no crossings occur.

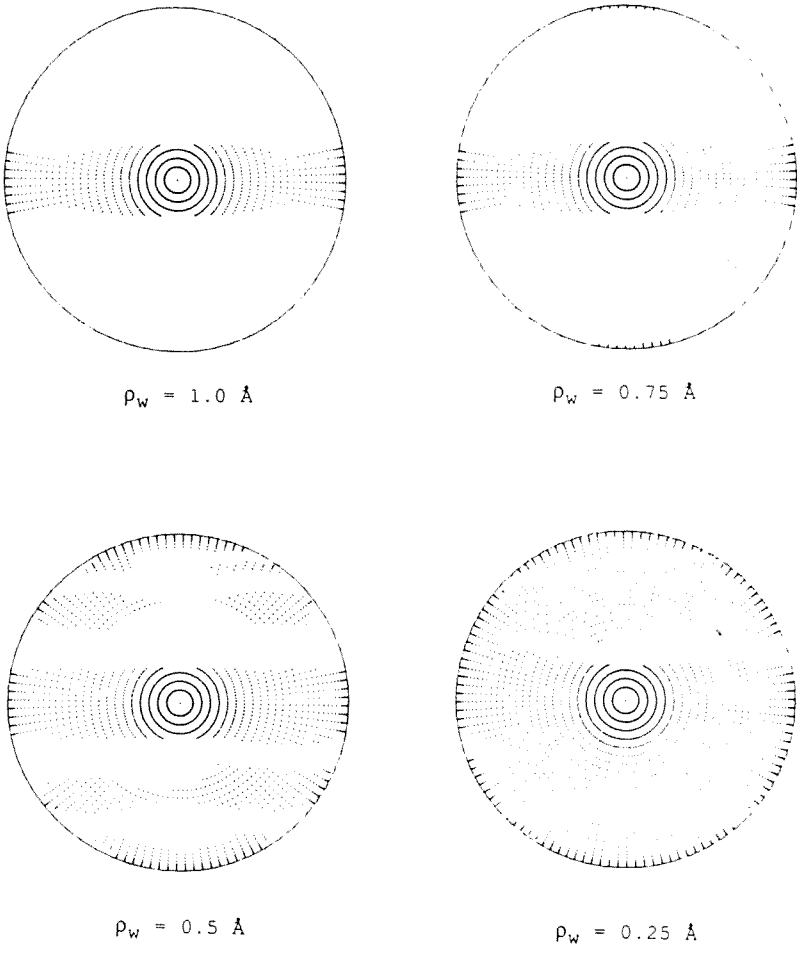

Fig. 3. Series of spherical shape maps of a 13-atom perfect, cylindrical helix (three loops) for various values of the viewing window radius. The dotted regions correspond to directions on the viewing sphere where the space curve *does not overcross* (that is, the S_0 region). The radius $\rho_w = 1$ Å is above the critical value. The area of the S_0 regions increases with a smaller ρ_w. The patterns on the sphere, i.e. the shape of $S_0(\rho_w)$, also provide information to characterize the backbone.

The location of these new directions as a function of ρ_w depends on the *length, number, and type of helices* enclosed. The information retrieved from these maps is discussed in section 3.

3. Areas of $S_0(\rho_w)$ regions as shape descriptors: applications to α-helices and β-sheets

The $S_0(\rho_w)$ maps discussed in the previous section characterize some of the essential shape features of a molecular space curve. It would be desirable to use part

of the information encoded in these diagrams in a numerical fashion. This is particularly relevant regarding the nonvisual analysis of molecular folding patterns.

A simple approach would be to use the area of the $S_0(\rho_w)$ region as a shape descriptor. Let $A_N(\rho_w)$ be the surface area of the region on S with the viewing directions along which the backbone exhibits N overcrossings (observed through the viewing window of radius ρ_w). It is evident that

$$\sum_{N=0}^{\nu^*} A_N(\rho_w) = 4\pi R^2, \tag{6}$$

where R is the radius of the sphere. If one reduces the size of the viewing window, fewer directions leading to overcrossings are found. Note that directions leading to crossings will always be found, no matter how small ρ_w is (as long as the curve does overcross somewhere). However, these directions become isolated points on the sphere S for small ρ_w, and thus they do not contribute to the total area. In other words, for infinitesimally small windows, the $S_0(\rho_w)$ region will effectively cover the whole sphere:

$$\lim_{\rho_w \to 0} A_0(\rho_w) = 4\pi R^2. \tag{7}$$

Since we are paying special attention to the region with no crossings, we shall characterize the reduced-view spherical shape maps by the fractional area $A(\rho_w)$:

$$A(\rho_w) = A_0(\rho_w)/4\pi R^2. \tag{8}$$

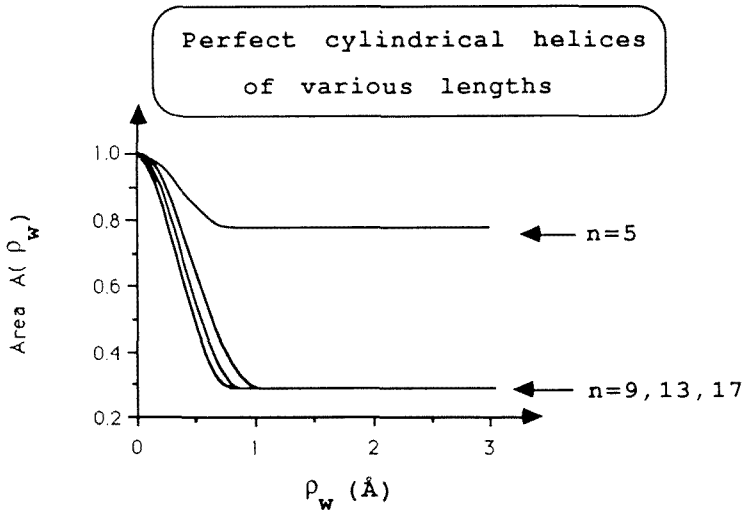

Fig. 4. Change in the area of the no-crossing region $S_0(\rho_w)$ with the window radius ρ_w for a series of perfect, cylindrical helices of various lengths. The number of atoms $n = 5$, 9, 13, and 17 defines helices with 1, 2, 3, and 4 loops, respectively.

Figure 4 shows the result of this analysis for a series of perfect cylindrical helices of various lengths. The diagram shows the fractional area $A(\rho_w)$ as a function of the window radius for helices of $n = 5, 9, 13$, and 17 atoms. These correspond to helices with 1, 2, 3, and 4 loops, respectively.

The results in fig. 4 reveal a number of interesting facts. Note that the shorter helix has a no-crossing area larger than that of the other helices. This is consistent with the fact that the fewer the loops or turns, the smaller the number of crossings in a backbone. Note that the areas decrease when the number of atoms increase. The fact that the area $A(\rho_w)$ reaches the same limit value at $\rho_w \to \infty$ for $n = 7$, 13, and 17 is a consequence of dealing with very symmetrical helices. The results suggest that the change in area becomes smaller as the helix becomes longer. One could predict a critical number of atoms n_c beyond which all perfect helices will have the same description.

Figure 5 shows the results of a similar analysis for examples of actual protein α-helices. The examples are three helices of different length in sperm-whale myoglobin [26]. We have chosen the helices labeled as 3, 1, and 7 in the Protein

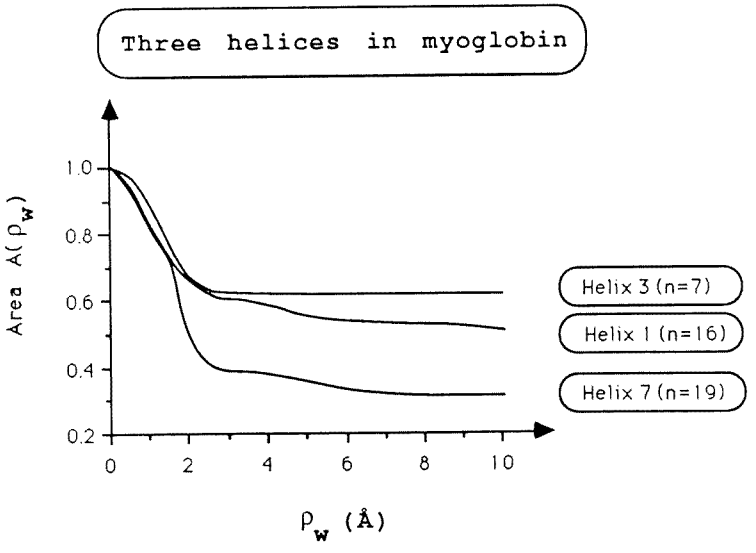

Fig. 5. Change in the area of the no-crossing region $S_0(\rho_w)$ with the window radius ρ_w for three helices of various lengths in sperm-whale myoglobin. The helices are labeled according to their numbering in the Protein Data Bank structure. The number between parentheses indicates the number of atoms for each helix.

Data Bank file corresponding to 7, 16, and 19 atoms in length, respectively. These helices are distorted cylindrical helices. The results reflect some of the trends revealed in fig. 4. The area $A(\rho_w)$ decreases with the length of the helix. Note that

the asymptotic values of $A(\rho_w)$ for large window radii are different for each helix in this case. However, the overall functional dependence of $A(\rho_w)$ is the same for all helices. These diagrams allow one to recognize the occurrence of a helix and to characterize some of its particular features such as length and deviation from perfect helicity.

We can contrast the results for helices with those for a different secondary structural motif. Figure 6 shows three $S_0(\rho_w)$ maps for a three-stranded β-sheet with $n = 17$ atoms. As in fig. 3, the dotted areas correspond to sections on the sphere S

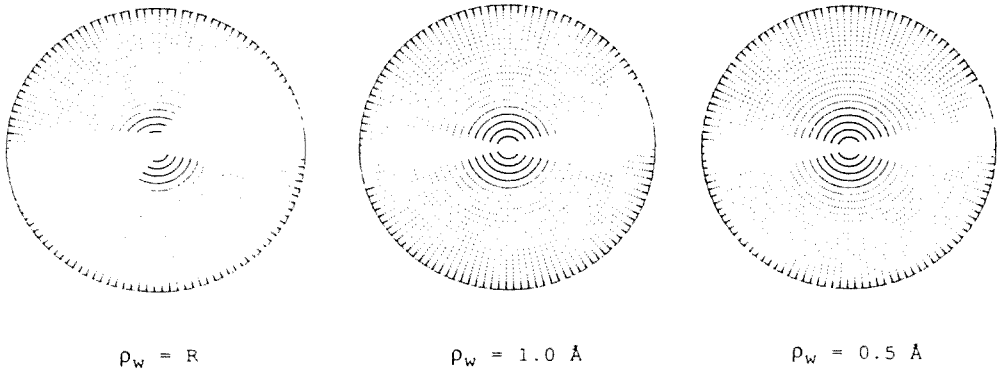

Fig. 6. Series of spherical shape maps of a 17-atom model of a three-stranded β-sheet for various values of the viewing window radius. The dotted area corresponds to directions showing no crossings.

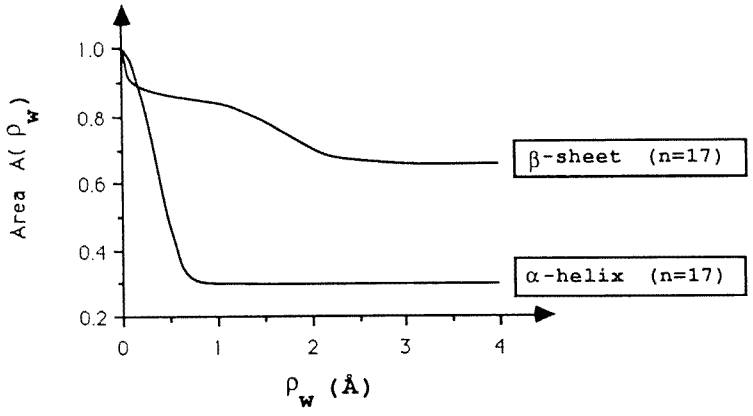

Fig. 7. Change in the area of the no-crossing region $S_0(\rho_w)$ with the window radius ρ_w for models of β-sheet and α-helices with the same number of atoms ($n = 17$).

defining viewing directions from where the β-sheet does not overcross. Note that the region where crossings *do occur* now occupies a central elongated band. This behavior is roughly the opposite to that found in helices. The no-crossing region

becomes larger when the window radius decreases but its overall shape remains the same. Figure 7 compares the areas $A(\rho_w)$ of the β-sheet and a cylindrical α-helix with the same number of atoms (in this case, 17). It is evident that the two structures are clearly distinguished by their ranges of values in $A(\rho_w)$ and by the functional dependence of the area with the window radius. The area decreases quickly with an increase in ρ_w for an α-helix, whereas it shows a rather slow change for a β-sheet.

The $A(\rho_w)$ functions discriminate between the two secondary structural elements. The same procedure can be applied to irregular structures where it is no longer possible to recognize any motif. In general, our approach provides a quantitative measure for similarity between the essential shapes of any two backbones and permits one to recognize the formation of supersecondary motifs. This is especially useful when comparing proteins with different degrees of homology and when studying computer-simulated unfolding processes [27].

4. Characterization of α-helical bundles

In this section, we apply the methodology to the characterization of the packing of α-helices (i.e. α-helical bundles).

Figure 8 shows schematically the type of models of α-helical bundles studied. The figure displays the case of a set of three connected helices. We have considered cylindrical helices with 12 atoms each (three turns). Note that the connection of the

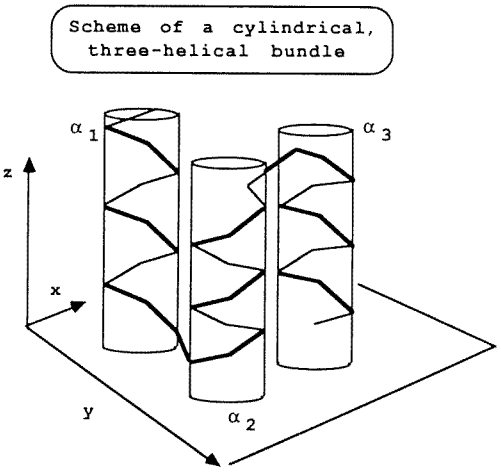

Fig. 8. Schematic representation of a 3-helical bundle. Each helix is modeled as a perfect, cylindrical α-helix with three turns (12 atoms).

helices alternates between the top and the bottom of the helices and that the helical axes are parallel.

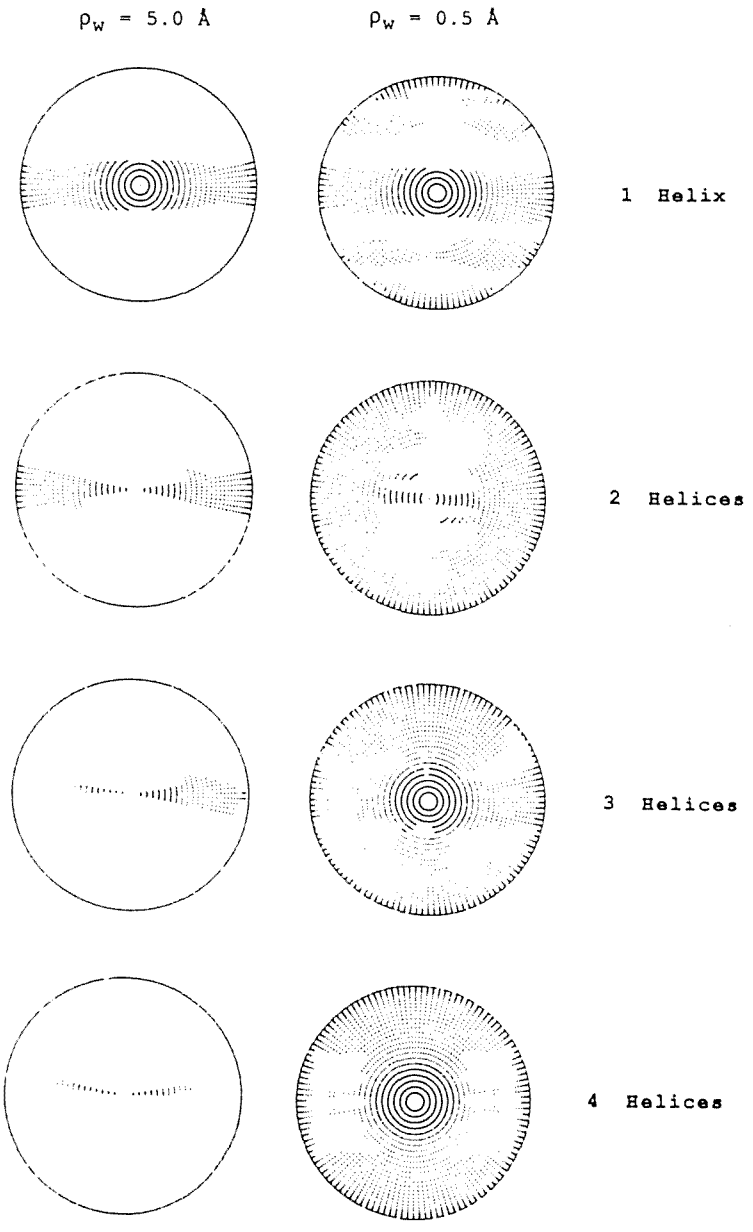

Fig. 9. $S_0(\rho_w)$ maps for α-helical bundles. The results correspond to two window radii (5.0 and 0.5 Å) for 1, 2, 3, and 4-helical bundles. The changes in the approximate symmetry of the bundle are reflected in the maps.

Figure 9 shows the results for the $S_0(\rho_w)$ maps for bundles with 1, 2, 3, and 4 helices. The results are given for two window radii $\rho_w = 5.0$ and 0.5 Å, which are larger than and smaller than ρ_w^0, respectively. It is interesting to compare the maps. Note that the $S_0(5.0$ Å) region becomes smaller as the number of helices packed

grows larger, whereas $S_0(0.5\ \text{Å})$ actually increases. Moreover, the shapes of the maps are quite altered. For a large window, we observe the occurrence of the central band which is typical for α-helices. This band identifies the direction of the axis of each helix (perpendicular to the band). The fact that this band becomes smaller and more disconnected as the bundle increases in size accounts for the fact that *the total number of crossings (a large-scale feature) increases.*

In contrast, the location of the viewing directions leading to non-crossings appears to be rotated approximately 90° for small window radii. This is particularly evident in the case of the more symmetrical 2- and 4-helical bundle $S_0(\rho_w)$ maps. The maps for $\rho_w = 0.5$ Å reveal the *small-scale shape features associated with the relative location and distribution of overcrossings in 3-space.* For example, the two blank regions shown in the $S_0(0.5\ \text{Å})$ map for the 4-helical bundle correspond to directions where the crossings appear very close to each other. This identifies the regions where one finds the actual helices, as opposed to the *interhelical regions* where one finds few crossings when using small windows. This analysis shows how the small- and large-scale shape features are discriminated in our approach.

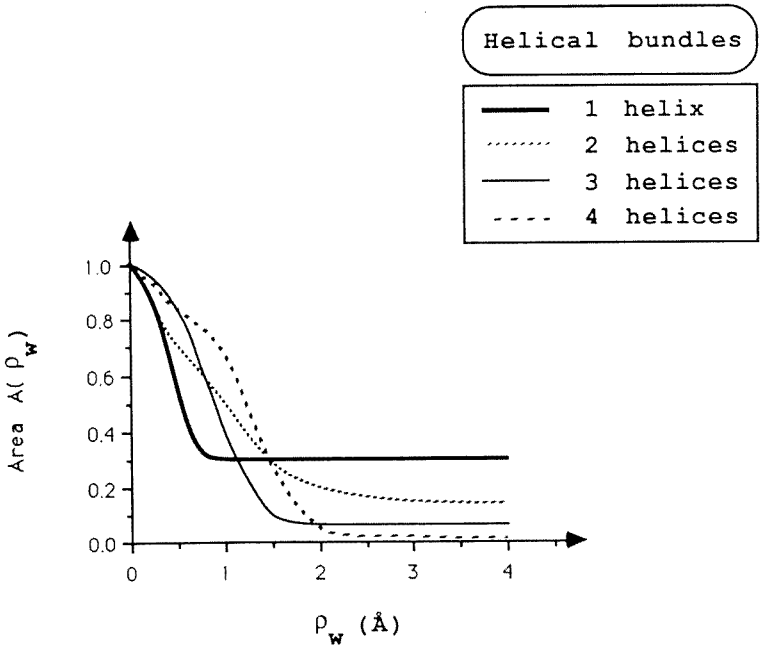

Fig. 10. Change in the area of the no-crossing region $S_0(\rho_w)$ with the window radius ρ_w for 1, 2, 3, and 4 α-helical bundles.

Figure 10 shows the surface areas $A(\rho_w)$ for the helical bundles. The large- and small-scale features are revealed by studying the large and small ρ_w values. At large ρ_w values, the area of the no-crossing region depends on the number of

helices. The larger the number of helices, the larger the total number of crossings. Accordingly, the areas decrease with an increment in the number of helices. The behavior is much more convoluted for small window radii. When ρ_w is sufficiently small, the areas become larger as the number of helices packed increases.

The results in fig. 10 show, at any rate, the *overall* functional dependence of $A(\rho_w)$ on ρ_w which is characteristic of α-helices. The *details* of this functional dependence give one additional clues to recognize the occurrence of several helices packed in bundles.

5. Further comments and conclusions

The procedure discussed above provides an alternative approach to the analysis of macromolecular structure. The method leads to a simple, algorithmical three-dimensional topological shape description in terms of a continuum of spherical maps derived from the complete overcrossing pattern of the molecular backbone.

By introducing the dependence of the spherical shape maps on the radius of a variable viewing window, one can distinguish large-scale or small-scale shape features in the backbone. As has been shown, large-scale features are mostly accountable for the type of structural motif and its size. On the other hand, small-scale structural features can serve to recognize the number and relative disposition of the basic motifs in 3-space. This procedure is suitable for studying the formation (or disappearance) of structural motifs during protein folding. The potential of this methodology to complement molecular dynamics studies of computer-simulated protein unfolding is discussed in ref. [27].

Displays of $S_0(\rho_w)$ maps give a detailed characterization of the molecular shape. However, we have shown that a good deal of this information can be retained by analyzing only the fractional surface area $A(\rho_w)$ associated with the no-crossing region S_0 as a function of the window radius. *This approach provides a single variable function which acts as a continuous, numerical global shape descriptor of space curves.*

For the sake of illustration, we have restricted the analysis to the use of the non-crossing region $S_0(\rho_w)$. Without any modification, the procedure can be applied to any other overcrossing pattern of interest. Moreover, the overcrossing pattern can be specified in more detail by taking into account the handedness of the crossings, rather than just their number [20–24].

Several further refinements are possible. For instance, in refs. [22–24] we have discussed the possibility of introducing a cut-off ε at the distance at which the overcrossings occur. A more detailed characterization can be formulated by considering both a cut-off and a reduced view. In this case, one would have a continuum of spherical shape maps defined in terms of the two parameters ε and ρ_w. This information can be rendered in a manageable fashion as contour lines of constant surface area for $S_0(\rho_w)$ in a two-dimensional ε–ρ_w diagram. This possibility will be discussed elsewhere.

Acknowledgements

This work is part of the ongoing research in molecular modeling at the Mathematical Chemistry Research Unit (Topology Program) at Saskatoon under the direction of P.G. Mezey, to whom the author is indebted for many fruitful discussions. Support of this Program from the Natural Sciences and Engineering Research Council (NSERC) of Canada is gratefully acknowledged. The author would also like to thank Drs. O. Tapia and O. Nilsson for several illuminating comments during a visit to the Department of Physical Chemistry, University of Uppsala, Sweden.

References

[1] M.A. Johnson and G.M. Maggiora (eds.), *Concepts and Applications of Molecular Similarity* (Wiley, New York, 1990).
[2] G.E. Schulz and R.H. Schirmer, *Principles of Protein Structure* (Springer, New York, 1979).
[3] C.R. Cantor and P.R. Schimmel, *The Conformation of Biological Macromolecules, Biophysical Chemistry, Part I* (Freeman, San Francisco, 1980).
[4] C. Chothia, Ann. Rev. Biochem. 53(1984)537.
[5] R. Jaenicke, Progr. Biophys. Mol. Biol. 49(1987)117.
[6] G.M. Maggiora, B. Mao, K.C. Chou and S.L. Narasimhan, in: *Protein Structure Determination. Methods of Biochemical Analysis*, Vol. 35, ed. C.H. Suelter (Wiley, New York, 1990).
[7] C. Brändén and J. Tooze, *Introduction to Protein Structure* (Garland, New York, 1991).
[8] N.S. Goel and R.L. Thompson, *Computer Simulations of Self-Organization in Biological Systems* (Macmillan, New York, 1988).
[9] M. Delbrück, Proc. Symp. Appl. Math. 14(1962)55.
[10] F.B. Fuller, Proc. Symp. Appl. Math. 14(1962)64.
[11] F.B. Fuller, Proc. Nat. Acad. Sci. US 68(1971)815.
[12] T. Kikuchi, G. Némethy and H.A. Scheraga, J. Comput. Chem. 7(1986)67.
[13] M. Le Bret, Biopolymers 18(1979)1709.
[14] P. De Santis, S. Morosetti and A. Palleschi, Biopolymers 22(1983)37.
[15] M.-H. Hao and W.K. Olson, Biopolymers 28(1989)873.
[16] E.M. Mitchell, P.J. Artymiuk, D.W. Rice and P. Willett, J. Mol. Biol. 212(1990)151.
[17] A.H. Louie and R.L. Somorjai, J. Theor. Biol. 98(1982)189.
[18] A.H. Louie and R.L. Somorjai, J. Mol. Biol. 168(1983)146.
[19] A.H. Louie and R.L. Somorjai, Bull. Math. Biol. 46(1984)745.
[20] G.A. Arteca and P.G. Mezey, J. Mol. Graph. 8(1990)66.
[21] G.A. Arteca and P.G. Mezey, in: *Theoretical and Computational Models for Organic Chemistry*, NATO ASI Series, ed. S.J. Formosinho, I.G. Csizmadia and L.G. Arnaut (Kluwer, Dordrecht, 1991).
[22] G.A. Arteca and P.G. Mezey, in: *Structure, Reactivity, and Interactions*, ed. S. Fraga (Elsevier, Amsterdam, 1992).
[23] G.A. Arteca, O. Tapia and P.G. Mezey, J. Mol. Graph. 9(1991)148.
[24] G.A. Arteca and P.G. Mezey, Biopolymers 32(1992)1609.
[25] G.A. Arteca, Program Sphere, University of Saskatchewan (1991).
[26] H.C. Watson, Progr. Stereochem. 4(1969)299.
[27] G.A. Arteca, O. Nilsson and O. Tapia, J. Mol. Graph, in press.