 CMES, vol.68, no.3, pp.221-237, 2010

Orienting a Protein Model by Crossing Number to Generate the Characteristic Views for Identification

Chikit Au[1], Yiyu Cai[2], Jianmin Zheng[3] and Tony Woo[4]

Abstract: A protein model (such as a ribbon model) can be created from the atomic coordinates in the protein data base files. These coordinates are obtained by X-ray crystallography or NMR spectroscopy with the protein arbitrarily oriented. As such, identifying or comparing a novel structure with a known item using protein model in the protein data base can be a timely process since a large number of transformations may be involved. The identification efficiency will be improved if the protein models are uniformly oriented. This paper presents an approach to orient a protein model to generate the characteristic views with minimum and maximum crossings respectively. The projection directions for these characteristic views are determined by a set of crossing maps (C-maps). Re-orientating the protein models in the protein data base to two characteristic views will facilitate the process of identification.

Keywords: Characteristic views, crossing, projection, C-map, spherical polygons

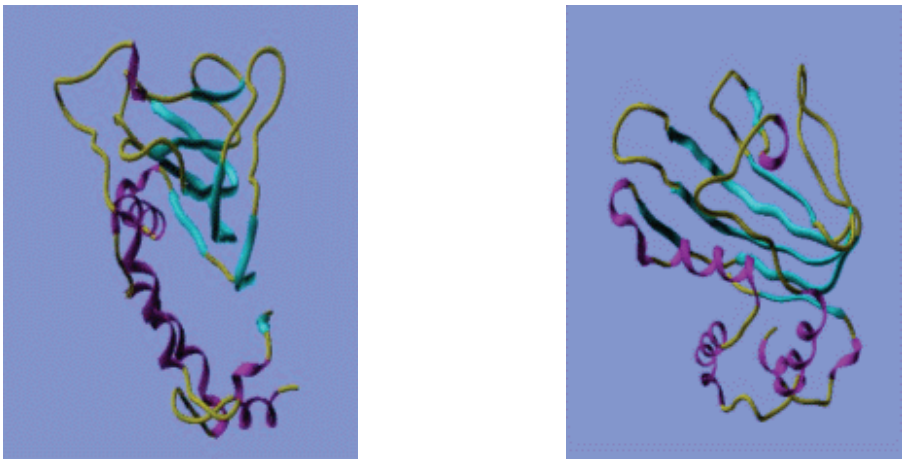
1 Introduction

There are currently a huge number of protein structures in the data base, depending on whether the site is public or private. Current protein data bases (PDB) contain about 100,000 known structural domains [am Busch, Mignon, Simonson (2009)]. The number is bound to increase owing to the automated techniques of X-ray crystallography and NMR spectroscopy. Unlike an object resting on a plane, protein suspended in liquid has no "neutral" pose. Most techniques to see the invisible nano-world are to take snapshots. Inherently, these processes are four-dimension (space and time) to three-dimension (space) projection. Since the information captured by snapshot largely depends on the projection direction, the majority of the

[1] Department of Engineering, University of Waikato, New Zealand. E-mail: ckau@waikato.ac.nz
[2] School of Mechanical Engineering, Nanyang Technological University, Singapore
[3] School of Computer Engineering, Nanyang Technological University, Singapore
[4] School of Mechanical Engineering, Nanyang Technological University, Singapore

atomic coordinates of a protein in the PDB obtained by X-ray crystallography or NMR spectroscopy is generally having an arbitrary orientation.

The protein data base files can be viewed by using a visualization program which often creates a protein model (such as a ribbon model) to highlight the way a protein chain folds. The concept that concerns protein three-dimensional structural homology is done partly by visual inspection [Rogen, Bohr (2003)] using computer graphics [Orengo, Michie, Jones, Swindells, Thornton (1997), Conte, Brenner, Hubbard, Chothia, Murzin (2002)]. Although the advanced graphic tools can rotate or translate the model freely to yield various views, the non-uniform orientation of the proteins in PDB causes difficulties in visually comparing two protein models geometrically since it could involve an unpredictable number of transformations. Furthermore, it is also not easy to identify a protein structurally with the two-dimensional images, which are three-dimension to two-dimension projection, obtained from the visualization program. Figure 1 depicts the difficulty of recognizing a protein with two-dimensional images. Two different views of a protein model with PDB code 1b2v are shown with different orientations. It can be seen that the visualization can be totally different for the same protein model. Both α-helices and β-sheets are packed in Figure 1(a) but they are well stretched in Figure 1(b).



(a) (b)

Figure 1: A protein model (PDB code 1b2v) view from different projection (helices in red, strands in blue and loops in yellow)

Hence, it will be helpful in identifying the proteins if their models are all uniformly

oriented. As such, it may also facilitate the process of comparing a novel structure with a known item in the protein data base, for instance.

2 Related Work

Objects with same orientations are the basic requirement for structural or geometric comparison. Various approaches for geometric comparison of images are proposed [Otterloo (1991), Scott, Nowak (2006)]. Basically, these approaches are for two-dimensional image matching. One of the two-dimensional objects, usually a contour extracted from the image, is used as a reference while the other object is re-positioned into the same orientation for the comparison task.

The majority of work involving orienting a protein model relates to obtaining the two-dimensional images from a three-dimensional structure with the topology preserved. TOPS cartoon [Westhead, Sidel, Flores, Thornton (1999), Flores, Moss, Thornton (1994), Levitt, Chothias (1976)] is a popular two-dimensional representation of a protein. It is a simplified representation of protein structure which emphasizes the secondary structure elements. A circle and a triangle are employed to represent an alpha helix and a beta strand respectively - both connected by straight lines. The result is somewhere between a three-dimensional and a two-dimensional view. The spatial proximity is preserved between secondary structure elements along with the chirality of the connections which is of great importance to topological studies of a protein.

[Barlow, Richards (1995)] addressed the creation of two-dimensional representation by using Sammon mapping [Borg, Groenen (1997)] which aimed to preserve all distances between the α-carbons. The limitation of these approaches was the loss of the spatial meaning of the protein structure. Thus, they did not give the three-dimensional spatial information to the viewers.

[Sverud, Maccallum (2003)] used Kohonen's self organizing map to reduce the three-dimensional α-carbon coordinate to two dimensional. The original protein model was then oriented to a view which is most closely to the two-dimensional mapping. The view quality was measured by a web-based user survey. However, it might not be practical to perform the user survey each time a technical development is made. Furthermore, the view quality is subjective. A consistent result may not be generated every time for the same protein.

Obviously, it is almost impossible to orient a protein before it is snapshot by X-ray crystallography and NMR spectroscopy. However, the model of the snapshot protein such as the ribbon model created by a visualization program can be oriented before generating the two-dimensional views. Orienting a protein model involves a sequence of rotations which requires a reference. Since reference co-ordinate

system does not actually exist, the most appropriate reference should be an intrinsic characteristic of the protein model. A protein backbone can be modeled as an open string, connecting sequential residues via their α-carbons. Consequently, a protein structure can be analyzed by the writhe and the crossing number. [Levitt (1983)] used the writhe to differentiate the chain threading while [Arteca (1993), Arteca, Tapia (1999)] characterized the shape and global structural homologies by the average crossing number. In fact, the writhe of a curve is actually defined based on the self-crossing number [Au (2008)]. As a result, the crossing number is considered as an intrinsic characteristic to orient a protein model.

This paper presents an approach, from both theoretical and the applied perspective, to the issue of orienting a protein model. A view of a protein model from a specific orientation is characterized by its crossing number. A crossing map (C-map), which is a spherical polygon on a unit sphere, is defined for each protein. By manipulating the C-maps, the orientations with optimal crossings are produced. In doing so, all the protein models can be uniformly oriented based on these characteristics to produce a two-dimensional image for identification and comparison.

3 Maximization and Minimization Principle

A protein is primarily a polypeptide backbone structure. For this investigation, a string of line segments is used to represent the protein in three-dimensional space. When a string is projected onto a view plane along a specific projection direction, two phenomena exist in the projected image due to occlusions:

i. the images of line segments overlap partially or totally and

ii. the images of line segments intersect.

Define the number of intersections between two line segment images on a view plane as the crossing number. Obviously, the crossing number differs as per projection. Equally obviously, there is a maximum crossing number and a minimum crossing number over all the projection directions. If the crossing number is to be used as measures of complexity, then the optimally complex projection giving the least and the most number of crossings should have room in view characterization.

The intent of a view is to manifest as much as possible. This is referred as maximization principle. Since a crossing gives the depth information between two spatial line segments, a characteristic view should possess as much crossings as possible – maximum crossing number. Yet, when there are "too many" details (sufficient to "confuse" a viewer), complexity reduction amendment is necessary such as conversing a solid line into a dashed or hidden is typical.

On the other hand, the overlapping of the line segment images causes information loss since two (or more) segment images are either totally or partially mapped onto the same location of the view plane. Obviously, line segment overlapping (either totally or partially) has more information loss than intersection. Hence, (total or partial) overlapping of images should be avoided and the crossing number should be minimum in a characteristic view. This is the minimization principle.

Therefore, two characteristic views with maximum and minimum crossings are employed to characterize the protein orientation.

4 Projection Direction and Crossing

The crossing of the images of two skewed line segments is determined by the visibility. A point **q** is visible to point **p** if there is projection direction **pq** connecting the points **p** and **q**. Two skewed line segment images cross each other if a point on one line segment is visible to a point on the other line segment.

4.1 Two skewed line segments

Consider two skewed line segments j and k with end points $\mathbf{p}_j$, $\mathbf{p}_{j+1}$ and $\mathbf{p}_k$, $\mathbf{p}_{k+1}$ as shown in Figure 2(a). The vectors $\mathbf{p}_j\mathbf{p}_k$, $\mathbf{p}_j\mathbf{p}_{k+1}$, $\mathbf{p}_{j+1}\mathbf{p}_k$ and $\mathbf{p}_{j+1}\mathbf{p}_{k+1}$ are four extreme projection directions to have the images of line segment j crossing that of k. Plotting these four projection directions on a unit sphere **S** with centre **o** defines a C-map (or crossing map) $\mathbf{S}_{jk}$ which is a convex spherical polygon as depicted in Figure 2(b).

The unit sphere is partitioned into three regions: the C-map $\mathbf{S}_{jk}$, its boundary $\partial\mathbf{S}_{jk}$ and their complement $\bar{\mathbf{S}}_{jk}$. For any point **v** on the unit sphere, it defines a direction **ov** projecting line segments j and k on the view plane.

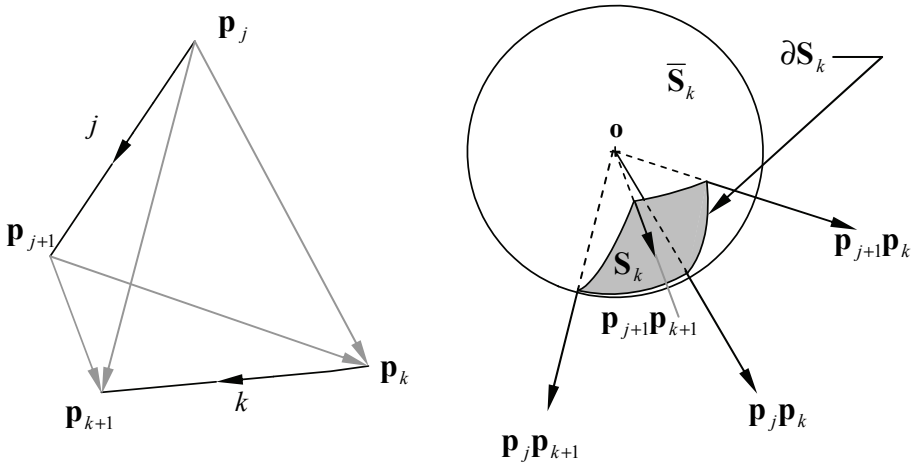
1. The images of line segments j and k intersect each other on the view plan when $\mathbf{v} \in \mathbf{S}_{jk}$.

2. The images of line segments j and k do not intersect each other on the view plane when $\mathbf{v} \in \bar{\mathbf{S}}_{jk}$.

3. The images of points $\mathbf{p}_j$, $\mathbf{p}_{j+1}$, $\mathbf{p}_k$ or $\mathbf{p}_{k+1}$ lie on the images of the line segments on the view plane when $\mathbf{v} \in \partial\mathbf{S}_{jk}$.

4.2 Two line segments are on the same spatial plane

The C-map degenerates into an arc of the great circle on the unit sphere when two line segments are on the same spatial plane. Hence, only two partitions arise: the degenerated spherical quadrilateral boundary $\partial\mathbf{S}_{jk}$ and its complement $\bar{\partial}\bar{\mathbf{S}}_{jk}$.



(a) Line segments j and k (b) Spherical quadrilateral

Figure 2: Crossing of two skewed line segment

The images of the line segments partially or totally overlap on the view plane when $\mathbf{v} \in \partial\mathbf{S}_{jk}$. By comparing with the situation of two skewed line segments, either

1. the complement $\bar{\mathbf{S}}_{jk} = \emptyset$ and $\bar{\partial}\bar{\mathbf{S}}_{jk} = \mathbf{S}_{jk}$ which implies these two line segments intersect on the spatial plane, or

2. the complement $\mathbf{S}_{jk} = \emptyset$ and $\bar{\partial}\bar{\mathbf{S}}_{jk} = \bar{\mathbf{S}}_{jk}$ which implies these two line segments do not intersect on the spatial plane.

The region of $\bar{\partial}\bar{\mathbf{S}}_{jk}$ covers the whole unit sphere in both situations.

5 C-map Re-organization

For a string of m line segments, $m(m\text{-}1)$ C-maps $\mathbf{S}_{jk}$ ($\forall k \neq j, j = 1, \cdots, m$ and $k = 1, \cdots, m$) will form on the surface of the unit sphere. A point $\mathbf{v}$ in the C-map $\mathbf{S}_{jk}$ implies a crossing between the images of line segments j and k after projecting onto the view plane along the direction $\mathbf{ov}$.

Figure 3(a) shows a string of six line segments. The C-maps of line segment L_1 ($\mathbf{S}_{1k}$, $\forall k = 2, \cdots, 6$) is shown in Figure 3(b). Hence, line segment L_1 has five C-maps. The C-map $\mathbf{S}_{12}$ degenerates into an arc (of great circle) on the unit sphere since line segments L_1 and L_2 lie on the same spatial plane. C-maps $\mathbf{S}_{16}$ and $\mathbf{S}_{15}$

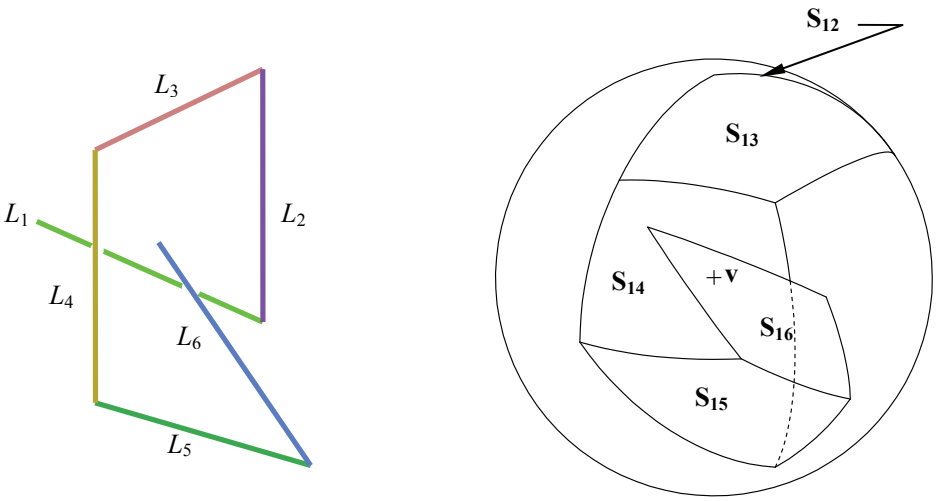


Figure 3: A string and the C-maps of line segment L_1

overlap map $\mathbf{S}_{14}$, which implies the image of line segment L_1 crosses the images of two line segments if point **v** is inside the overlapping region. For instance, if a point **v** is in the overlapping region of $\mathbf{S}_{14} \cap \mathbf{S}_{16}$, then the image of line segment L_1 crosses the images of line segment L_4 and L_6 on the view plane when the projection direction is along **ov** (where point **o** is behind point **v** in Figure 3(b)) as shown in Figure 3(a).

Since every C-map covering a region on the unit sphere refers to one crossing between a pair of line segment images, there can be some C-maps that overlap each other. A C-map with overlapping means various crossings in different region within the same map. For example, the C-map $\mathbf{S}_{16}$ shown in Figure 3(b) has two regions. One region (overlapping map $\mathbf{S}_{14}$) has two crossings while the other region has only one crossing. A way of sorting out this intransitive relation is to use clustering. A cluster consists of a set of C-maps with common intersection.

Table 1 shows, for example, that C-map $\mathbf{S}_{26}$ is in cluster $\mathbf{C}_5$ with C-map $\mathbf{S}_{14}$. It is also in cluster $\mathbf{C}_3$, with C-maps $\mathbf{S}_{14}$, $\mathbf{S}_{15}$ and $\mathbf{S}_{25}$. It can be seen that the maximum number of maps in a cluster is 4 (in cluster $\mathbf{C}_3$ or $\mathbf{C}_9$), which means the maximum crossing number is 4.

The minimum crossing number depends upon the following conditions:

1. $\mathbf{S} - \cup_{\substack{j,k \\ j \neq k}} \mathbf{S}_{jk} = \emptyset$ which implies the unit sphere is completely covered

Table 1: Clustering of C-maps

MAP	CLUSTER											
	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$	$\mathbf{C}_5$	$\mathbf{C}_6$	$\mathbf{C}_7$	$\mathbf{C}_8$	$\mathbf{C}_9$	$\mathbf{C}_{10}$	$\mathbf{C}_{11}$	$\mathbf{C}_{12}$
$\mathbf{S}_{13}$	✓											
$\mathbf{S}_{14}$			✓	✓	✓	✓						
$\mathbf{S}_{15}$			✓	✓		✓						
$\mathbf{S}_{25}$			✓	✓								
$\mathbf{S}_{26}$			✓		✓							
$\mathbf{S}_{31}$							✓					
$\mathbf{S}_{35}$		✓										
$\mathbf{S}_{41}$									✓	✓	✓	✓
$\mathbf{S}_{51}$									✓	✓		✓
$\mathbf{S}_{52}$									✓	✓		
$\mathbf{S}_{53}$								✓				
$\mathbf{S}_{62}$									✓		✓	

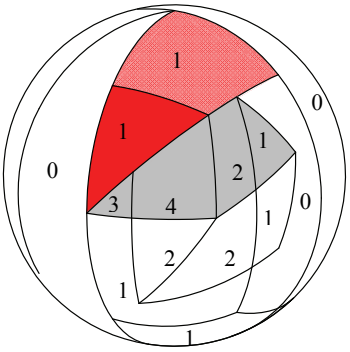
by the C-maps and the minimum crossing number can be obtained from the clustering table.

2. $\mathbf{S} - \cup_{\substack{j,k \\ j \neq k}} \mathbf{S}_{jk} \neq \emptyset$, and then the minimum crossing of the string is zero.

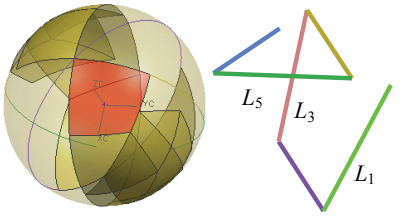
Reorganizing the C-map $\mathbf{S}_{jk}$ ($\forall j \in [1,m], \forall k = [1,m]$ and $k \neq j$) to $\mathbf{S}_i^n$ ($\forall i \in I, n \in I$) based on a crossing number n yields a new set of maps such that $\mathbf{S}_i^n = \bigcap_{\forall j,k} \mathbf{S}_{jk}$ and $n = (\text{number of } j \times \text{number of } k)$ (the subscript i is an integer to identify the reorganized maps. Its range depends upon the number of line segments in the string and its spatial structure). Each reorganized map covers a region with consistent crossing number. $\mathbf{S}_i^n$ is not necessarily an original C-map. It is the result of Boolean set operation applied to several C-maps $\mathbf{S}_{jk}(\forall j \in [1,m], \forall k = [1,m]$ and $k \neq j)$. Therefore, a C-map $\mathbf{S}_{jk}$ is a convex spherical polygon, but a re-organized C-map $\mathbf{S}_i^n$ is not necessary convex.

Figure 4(a) shows a unit sphere with reorganized crossing maps of the string depicted in Figure 3(a). Each crossing map refers to a specific crossing number. The grey C-map is $\mathbf{S}_{61}$ (crossing between L_6 and L_1), which is separated into four reorganized maps $\mathbf{S}_i^1$,$\mathbf{S}_{i+1}^2$, $\mathbf{S}_{i+2}^3$ and $\mathbf{S}_{i+3}^4$. Each has a consistent crossing number 1, 2, 3 and 4 respectively.

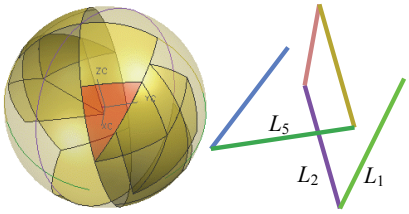
Every C-map is a convex spherical polygon. Various algorithms and their complexity for separating and intersecting the spherical polygons are well discussed in



(a) Reorganized maps



(b) Image with projection direction inside the quadrilateral map.



(c) Image with projection direction inside the triangular map.

Figure 4: A unit sphere with reorganized C-maps

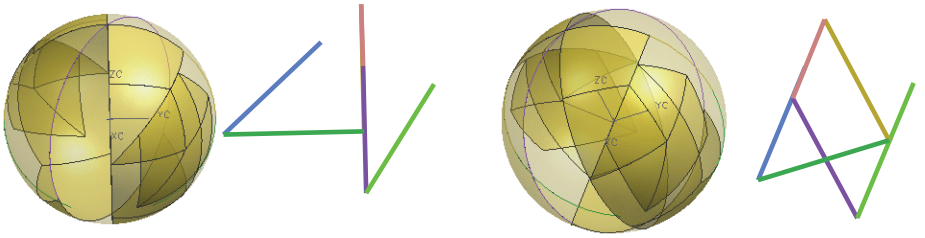
reference [Chen, Chou, Woo (1993), Gan, Woo, Tang (1994), Woo (1994)].

Figure 4(b) and 4(c) show two views of the string when the projection directions are inside the red quadrilateral and triangular map respectively. The string and the unit sphere are rotated until the projection direction is coincident with the view direction which is perpendicular to the paper. Both maps give crossing number of 1. For the projection direction inside the red quadrilateral map, the image of line segment L_3 crosses the image of line segment L_5. When the projection direction is inside the triangular map, the crossing is given by the intersection of the images of line segments L_2 and L_5.

6 Characteristic views

Since every point $\mathbf{v}$ on the unit sphere corresponds to the projection direction $\mathbf{ov}$. The unit sphere is basically categorized into two categories: the boundaries $\partial\mathbf{S}_i^n$ of the map $\mathbf{S}_i^n$ (with n crossings) and its complement. A point on the boundaries $\partial\mathbf{S}_i^n$ either gives an overlapping (totally or partially) of line segment images or has an end point projected onto the image of a line segment. Hence, the vector $\mathbf{ov}$ for$\mathbf{v} \in \partial\mathbf{S}_i^n$, $i \in I$ is a bad projection direction.

Figure 5 illustrates two situations with the projection directions falling on the bound-



(a) Partial overlapping of line segment images

(b) Total overlapping of point image with line segment images

Figure 5: Views generated by bad projection directions

aries of the C-maps. Both views show total overlapping of end point image with line segment images and partial overlapping of the line segment images.

The characteristic views of a protein have optimal crossings n_0 (either maximum value $n_0 = n_{\max}$ or minimum value $n_0 = n_{\min}$). The projection direction $\mathbf{ov}_0$ to produce a characteristic view with n_0 crossings has the point $\mathbf{v}_0$ in either map $\mathbf{S}_u^{n_0}$ (for some value u) when $n_0 \neq 0$ or map $\mathbf{S} - \cup_i \mathbf{S}_i^n$ (where $\mathbf{S}$ is the complete surface of the unit sphere) when $n_0 = 0$.

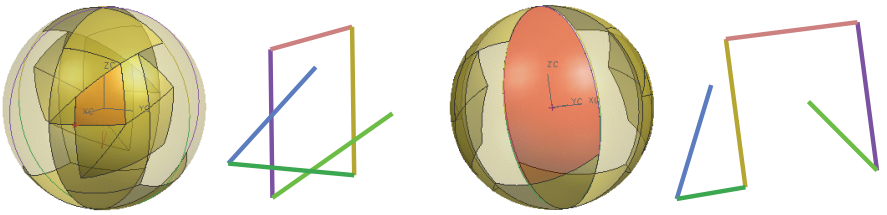
A C-map is a region on the unit sphere and there can be more than one C-map with the optimal crossing numbers. Since the boundary of each map is a bad projection direction and the point $\mathbf{v}_0$ is inside the map, it should locate as far as possible from the boundary. A heuristic approach is used to evaluate the location of point $\mathbf{v}_0$ by:

1. selecting the map (either $\mathbf{S}_u^{n_0}$ or $\mathbf{S} - \cup_i \mathbf{S}_i^n$) with maximum area; and
2. picking a location for point $\mathbf{v}_0$ which is within the map but is far away from its boundaries.

The characteristic views with maximum crossings and minimum crossings of the string shown in Figure 3(a) are listed in Figure 6(a) and 6(b) respectively. The C-maps for their projection directions (which are coincident with the paper normal) are also shown.

7 Examples

Two characteristic views of a ribbon model of a protein Apolipoprotein (a) Kringle IV Type 7 (PDB code: 1i71) is generated by this approach. Figure 7(a) shows the backbone of the protein expressed in the form of a set of line segments. Its re-organized C-maps with crossing numbers are flattened and plotted in Figure 7(b).



(a) Characteristic view with maximum crossings

(b) Characteristic view with minimum crossings

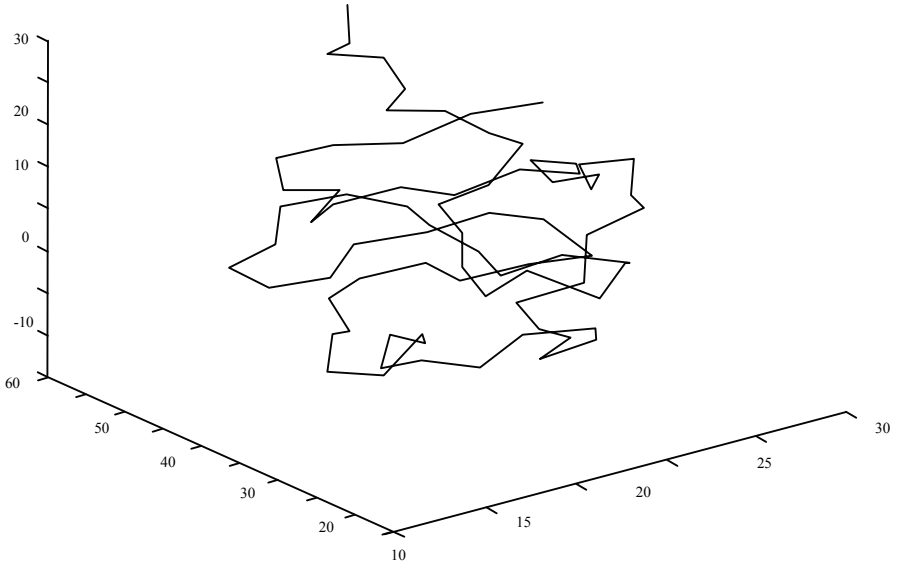
Figure 6: Characteristic views of a string

Each point on the unit sphere is expressed in terms of two angles, (θ, ψ) measured form x and z axes of the unit sphere. It can be seen that there may be more that one model orientation to produce a characteristic view (of either maximum or minimum crossing number).

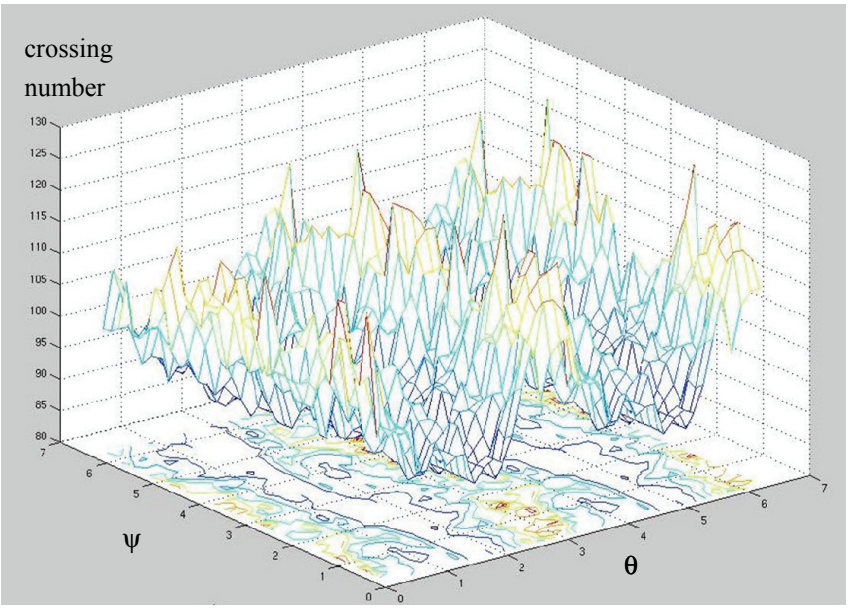
The two characteristic views of the protein are shown in Figure 8. Figure 8(a) gives the view with maximum crossing number while Figure 8(b) shows a minimum crossing. Both backbones are also depicted in the figures.

Figure 9 gives another protein model orientation example. Figure 9(a) shows the original view a Heme-Binding Protein A model (PDB code: 1b2v) extracted from the protein database. The other view of the same protein generated by SOM [Sverud, MacCallum (2003)] is also shown in Figure 9(b). It is hard to decide if these two images refer to the same protein since they are projected on the view plan with different orientation.

The characteristic views of the protein model 1b2v with maximum crossing number and minimum crossing number are listed in Figures 9(c) and 9(d) respectively. In this example, protein helical structure plays a central role in crossing number generation. If the helices are well stretched, the crossing number of the protein view is likely reduced. Otherwise, if a helix is projected along its axial direction or two helices are projected near each other, it is likely to produce a high number of crossings. Note in the crossing number computing, all protein secondary structures are represented as a string of line segments with zero cross-sectional dimension. (Note: It can be seen that the view generated by SOM is incidentally similar to the characteristic view with minimum crossing. Turning the SOM view by 90^0 counterclockwise and flipping it over produce a characteristic view as shown in Figure 9(d)).

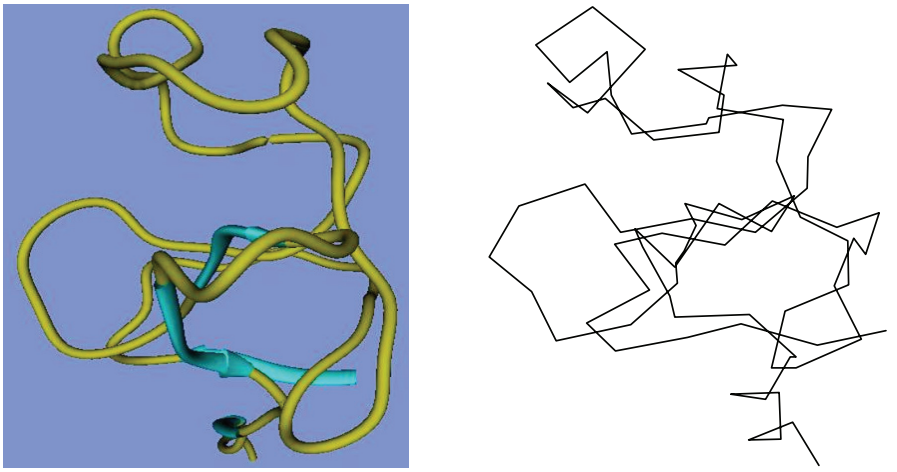


(a) The backbone represented by a string of line segments

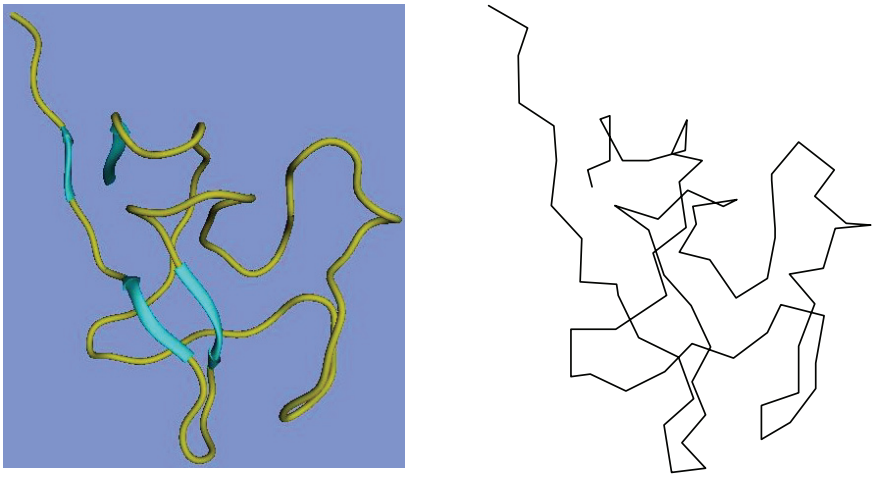


(b) The re-organized C-maps

Figure 7: An Apolipoprotein (a) Kringle IV Type 7 and its C-maps.



(a) View with maximum crossing number at $(\theta,\psi) = \left(\frac{\pi}{2},\frac{\pi}{40}\right)$

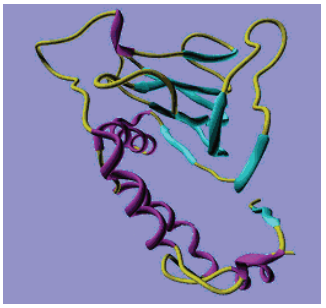


(b) View with minimum crossing number at $(\theta,\psi) = \left(\frac{\pi}{40},\frac{3\pi}{2}\right)$

Figure 8: The characteristic views of the Apolipoprotein (a) Kringle IV Type 7.

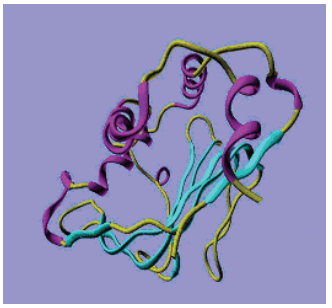
8 Discussion

In fact, the issue of generating characteristic (two-dimensional) views from three-dimensional structure is commonly found in the engineering design, computer vision and medical imaging.

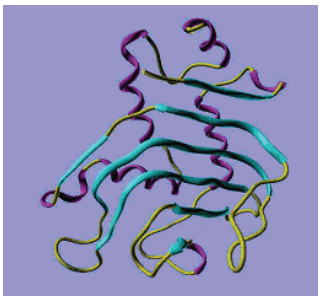


(a) The Original view of Protein 1b2v

(b) The best view of Sverud and MacCallum's SOM method



(c) A characteristic view with maximum crossings



(d) A characteristic view with minimum crossings

Figure 9: The characteristic views of the protein model 1b2v.

In engineering design, projecting a three-dimensional object onto a two-dimensional medium to obtain a line drawing, which should be able to be represented by a single edge-vertex graph, appears to be a fundamental task. A moment's reflection, however, validates the fact that there does not exist a characteristic view to reveal all the edges and vertices of the object for an arbitrary mechanical drawing neither. In human vision, the issue of whether three-dimensional object recognition should rely on internal representations that are inherently three-dimensional or on collections of two-dimensional views is still being explored, while an aspect graph [Koenderink, Van Doom (1976), (1979)] is used to represent a three-dimensional object by a set of two-dimensional views in computer vision.

Perhaps useful is the approach employed in computer tomography [Cucchiara, Lamma, Sansoni (2004)] or magnetic resonance imaging, the radiologists or imaging professionals often do image scans in a pre-defined orientation. The axial, sagittal and coronal views are typically used in diagnostic or therapeutic proce-

dures. These pre-defined orientations are inherent form of the object such as a human body.

There can be many surface curvatures defined at a point on a surface. It was Gauss who noted that, of the infinitely many curvatures, only two of them really matter – one exhibiting the maximum and the other the minimum – the product of which is the celebrated Gaussian curvature. Similarly, two views with minimum and maximum crossings are employed to characterize a protein among the infinite number of views.

One way to measure the complexity of a task is to have several procedures that accomplish the same task and to compare them by counting the number of steps in each procedure (Turing model [Turing (1952)]). The minimum number of steps is defined (by Kolmogorov [Li, Vitanyi (2008)]) as the complexity of the task. This gives an idea of using the minimum crossings in a characteristic view to measure the geometric complexity of a protein.

Despite the crossing of the line segment images is considered as information loss due to obstruction, it also gives the three dimensional spatial information of the line segments. This can easily be illustrated by comparing the top view with the isometric view of a pyramid. A top view of a pyramid consists of eight line segments without any crossing. However, it does not give any depth information while there is one crossing in the isometric view which shows more spatial information of the object. Hence, the view with maximum crossings of a protein is selected as the other characteristic view to show the spatial information.

9 Conclusion

A view of a protein reveals the most about its structure. This information facilitates the understanding of the protein structure, the classification of the structure and the formulation of hypothesis about protein functions. A geometric approach is developed for protein view characterization based on the crossing number. By simplifying a protein as a string of line segments, its crossing in a two dimensional view generated by projection is shown by computing its C-maps. Two characteristic views with optimal crossings can easily be generated. One of the major applications of these characteristic views is to facilitate the processes of protein identification, comparison and structural matching. It will be easier to identify a protein if its characteristic views of maximum and minimum crossing are included into their entry in PDB. The authors hope to report on a future date the protein structural comparison and matching by using the C-maps.

References

am Busch M. S.; Mignon, D.; Simonson, T. (2009): Computational protein design as a tool for fold recognition. *Proteins-Structure Function and Bioinformatics*, Vol. 77, no. 1, pp.413-419.

Arteca, G. A. (1993): Overcrossing spectra of protein backbones – characterization of 3-dimensional molecular shape and global structural homologies. *Biopolymers*, Vol. 33, no. 12, pp.1829-1841.

Arteca, G. A.; Tapia, O. (1999): Characterization of fold diversity among proteins with the same number of amino acid residues. *Journal of Chemical Information and Computer Sciences*, Vol. 39, no. 4, pp.642-649.

Au, C. K. (2008): The geometric interpretation of linking number, writhe and twist for a ribbon. *CMES: Computer Modeling in Engineering & Science*, Vol. 29, no. 3, pp.151-162.

Barlow, T. W.; Richards, W. G. (1995): A novel representation of protein-structure. *Journal of Molecular Graphics*, Vol. 13, no. 6, pp.373-376.

Borg, I.; Groenen, P. (1997): Modern Multidimensional Scaling. Springer-Verlag, New York.

Chen, L. L.; Chou, S. Y.; Woo, T. C. (1993): Separating and intersecting spherical polygons for computing visibility on 3-, 4-, and 5-axis machines. *ACM Transactions on Graphics*, Vol. 12, no. 4, pp.305-326.

Conte, L. L.; Brenner, S. E. ; Hubbard, T. J. P. ; Chothia, C. ; Murzin, A. G. (2002) : SCOP database in 2002 : refinements accommodate structural genomics. *Nucleic Acids Research*, Vol. 30, no. 1, pp.264-267.

Cucchiara, R.; Lamma, E.; Sansoni, T. (2004): An image analysis approach for automatically re-orienteering CT images for dental implants. *Computerized Medical Imaging and Graphics*, Vol. 28, no. 4, pp.185-201.

Flores, T. P.; Moss, D. S.; Thornton, J. M. (1994): An algorithm for automatically generating protein topology cartoons. *Protein Engineering*, Vol. 7, no. 1, pp.31–37.

Gan, J. G.; Woo, T. C.; Tang, K. (1994): Spherical maps: their construction, properties, and approximation. *ASME Transaction Journal of Mechanical Design*, Vol. 116, no. 2, pp.357-363.

Koenderink, J. J.; van Doorn, A. J. (1976): The singularilarities of the visual mapping, *Biological Cybernetics*, Vol. 24, pp.1-59.

Koenderink, J. J.; van Doorn, A. J. (1979): The internal representation of solid shape with respect to vision. *Biological Cybernetics*, Vol. 32, pp.211-216.

Levitt, M. (1983): Protein folding by restrained energy minimization and molecular dynamics. Journal of Molecular Biology, Vol. 170, pp.723-764.

Levitt, M.; Chothia, C. (1976): Structural patterns in globular proteins. *Nature*, Vol. 261, no. 5561, pp.552–557.

Li, M.; Vitanyi, P. (2008): An introduction to Kolmogorov Complexity and its Applications. Springer Verlag.

Orengo, C. A.; Michie, A. D.; Jones, S.; Swindells, M. B.; Thornton, J. M. (1997): CATH – a hierarchic classification of protein domain structures. *Structure*, Vol. 5, no. 8, pp.1093-1108.

Otterloo, P. (1991): A contour-orientated approach to shape analysis, Prentice Hall International (UK) Ltd, Englewood Cliffs, NJ, pp.90-108.

Rogne, P.; Bohr, H. (2003): A new family of global protein shape descriptors. *Mathematical Biosciences*, Vol. 182, no. 2, pp.167-181.

Scott, C.; Nowak, R. (2006): Robust contour matching via the order-preserving assignment problem. *IEEE Transactions on Image processing*, vol. 15, no. 7, pp.1831-1838.

Sverud, O.; MacCallum, R. M. (2003): Towards optimal views of proteins. *Bioinformatics*, Vol. 19, no. 7. pp.882–888.

Turing, A. M.(1952): The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London, Series B*, Vol. 237, no. 641, pp.37-72.

Westhead, D. R.; Slidel, T. W.; Flores, T. P.; Thornton, J. M. (1999): Protein structural topology: automated analysis and diagrammatic representation. *Protein Science*, Vol. 8. no. 4, pp.897–904.

Woo, T. C. (1994): Visibility maps and spherical algorithms. *Computer-aided Design*, Vol. 26, no. 1, pp.6-16.