Algorithm for ribbon models of proteins

Mike Carson and Charles E Bugg

University of Alabama at Birmingham, Comprehensive Cancer Center, 252 BHS, THT 79, University Station, Birmingham, AL 35294, USA

A method is presented to draw smooth, 3D ribbon models of proteins. The procedure calculates closely-spaced guide coordinates based on the peptide plane and passes regular, nearly parallel B-spline curves through them. This becomes a simple process with a graphics device having built-in B-spline generating capabilities such as the Evans and Sutherland PS300. Examination of ribbons such as these provides a useful tool for the crystallographer. Any irregularity in the ribbon is a strong visual cue, suggestive of potential problem areas during the refinement process.

Keywords: proteins, 3D ribbon models, B-spline curves, ubiquitin, purine nucleoside phosphorylase, calmodulin

received 25 November 1985, accepted 6 January 1986

The 'ribbon drawing' of Richardson[1] provides an elegant method enabling the vizualization folding and secondary structure of proteins. Computer programs for the generation of drawings of this type have been written by several authors. Lesk and Hardman[2] plot the ribbon approximated as a series of linked trapezoids, one segment per residue. Pique[3] produces a series of short line segments perpendicular to a smooth curve fit to the protein backbone. The method presented below models the protein ribbon in much the same manner as the ribbons one uses to wrap presents, i.e., composed of many approximately parallel smooth threads running along the length of the ribbon. The key point in relation to protein chemistry is that the peptide plane is used as the basis for the geometrical construction. The key point relating to graphics is that B-splines[4] are used to produce the smooth and regular curves that model the ribbon.

ALGORITHM:

Defining peptide planes

The minimal required input is a list of the protein's alpha carbon and carbonyl oxygen coordinates ordered by sequence.

The following procedures are carried out to define the peptide plane:

- Get vector $\boldsymbol{A} = \boldsymbol{C}_{\alpha i+1} - \boldsymbol{C}_{\alpha i}$
- Get vector $\boldsymbol{B} = \boldsymbol{O}_i - \boldsymbol{C}_{\alpha i}$
- Form vector $\boldsymbol{C} = \boldsymbol{A} \times \boldsymbol{B}$
- Form vector $\boldsymbol{D} = \boldsymbol{C} \times \boldsymbol{A}$
- Normalize $\boldsymbol{C}$ and $\boldsymbol{D}$

Vector $\boldsymbol{C}$ is normal to the peptide plane approximated by atoms $\boldsymbol{C}_{\alpha i}$, $\boldsymbol{O}_i$, and $\boldsymbol{C}_{\alpha i+1}$. The 'handedness' chosen here has $\boldsymbol{C}$ pointing away from the helix axis for right-handed α-helices. Vector $\boldsymbol{D}$ lies parallel to the peptide plane and is perpendicular to vector $\boldsymbol{A}$.

Generating guide coordinates

The required inputs are the number of threads necessary to approximate the ribbon (nine is convenient), an assignment for each residue of secondary structural class (helix, sheet, coil or turn), and the desired ribbon widths (for example, 3.0 Å for secondary structure, 1.0 Å for turns and random coils).

The following procedures are carried out to define the B-spline guide points:

- Form point P as the midpoint of $\boldsymbol{C}_{\alpha i} \rightarrow \boldsymbol{C}_{\alpha i+1}$
- If the peptide plane is part of a helix, translate in the direction of vector C, away from the helix axis (this step is elaborated on in the discussion)
- Scale vector $\boldsymbol{D}$ by one half the desired ribbon width
- Form points $P_- = P - \boldsymbol{D}$ and $P_+ = P + \boldsymbol{D}$

After these procedures have been carried out the line segment defined by $P_- \rightarrow P_+$ is of the desired ribbon width and parallel to the peptide plane. The guide coordinates are evenly spaced along this line segment. A set of guide points is then generated from the next peptide plane in the sequence, and the test below is performed.

Determining the connectivity of the guide coordinates

It is desired to have a number of threads that make up the roughly parallel long fibres of the ribbon. A complication arises when the direction of the carbonyl oxygen flips, as is always the case between adjacent residues of β sheets. This can be monitored by taking the vector product of the $\boldsymbol{D}$ vectors of successive peptide planes. If the angle between $\boldsymbol{D}_{i-1}$ and $\boldsymbol{D}_i$ is greater than 90°, a flip is noted. This information is used to determine if the line segment defined in the previous step will have its first point, P_-, or its last point, P_+, assigned to the first thread of the ribbon.

Outputing the B-spline

For the specific case of the Evans and Sutherland PS300, all that is necessary is the firmware command:

ribbon_thread_1 : = BSPLINE ORDER = 4 N = n_{pp}

 0263-7855/86/020121-02 $03.00 © 1986 Butterworth & Co (Publishers) Ltd

X_1, Y_1, Z_1
X_2, Y_2, Z_2
: : :
$X_{npp}, Y_{npp}, Z_{npp}$
CHORDS = $10*n_{pp}$;

Where 'ribbon-strand-1' is a user-defined name and 'n_{pp}' is the number of peptide planes in the protein. The coordinate triples are the guide coordinates of the entire curve approximated by '$10*n_{pp}$' line segments (i.e., sampled 10 times between adjacent control points). A similar call would be issued by those with a different type of graphics system with built-in B-spline capabilities. Without such facilities, one would be forced to write a procedure to generate the line segments of the B-spline parameterization. A PASCAL algorithm for performing the parameterization is available (p 323 of Reference 4).

DISCUSSION

The algorithm is conceptually simple, requiring only vector algebra for calculations and (hopefully) special graphics facilities for the curve-fitting. The procedure expects the user to have made a prior assignment of the secondary structure. An algorithm for the automatic classification of secondary structure that works especially well on helices has been given by Louie and Somorjai[5].

It must be noted that the curves are very smooth, and do not necessarily pass through their guide coordinates (except for the first and last points). Instead, the curves are gently guided in the direction of the control points. This is the reason why any helical stretch of the molecule must be treated specially, with the control points translated away from the helix axis. A 1.5 Å translation is suggested to produce a reasonable helix diameter. If this is not done, very slender helices are produced.

The ribbon drawing algorithm given in this paper is illustrated with structures under investigation in this laboratory. All figures were displayed employing the program Frodo[6] and photographed directly from the Evans and Sutherland PS300 screen. Colour Plate 1 emphasizes the construction of these ribbons with the protein ubiquitin[7]. A white 13-thread smooth ribbon is overlaid with a two-thread ribbon in which the flip of the carbonyl is not monitored. The red thread approximately traces the positions of the carbonyl oxygen atoms, and the blue thread the positions of the amide hydrogen atoms. Hydrogen bonding is readily seen by close contacts between the red and blue curves on adjacent segments of the ribbon.

Colour Plate 2 depicts a considerably larger protein, human purine nucleoside phosphorylase (PNP)[8], with the threads of the ribbon given the colours of the 'rainbow'. The addition of colours on our display generates white for a ribbon viewed edge-on.

Colour Plate 3 shows a different view of ubiquitin, displayed as a light yellow ribbon with a red stripe along one edge. In the four strands of the β-sheet behind the helix (viewed from left to right), the red side is found on, respectively, the right, right, left, right hand sides of the strand. The N-terminus is bottom left-centre. The direction that the peptide chain runs is down, up, up, down, again looking left to right. Thus, highlighting one fixed edge of the ribbon does not guarantee a natural assignment of the parallel/antiparallel nature of the sheet. This assignment would require a more sophisticated check than the local testing of the dipole flip between adjacent peptide planes.

It is possible to monitor the change in direction of the dipole of the peptide plane and force or suppress the flip of ribbon connectivity at convenient points. In stretches of random coil or turns there will generally be a position where the angle between successive dipoles is not quite the 90° required to affect the transition. Forcing the transition in the highlighting from one edge of the ribbon to the other, upon encountering a 60–90° change in peptide orientation, causes a negligible effect on the smoothness of the ribbon and allows for a better display of secondary structural features.

Colour Plate 4 shows the N-terminal calcium binding domains of the protein calmodulin[9], forcing a flip at three preselected sites, where successive dipoles exhibit a 60–90° change in orientation. Only half of this molecule is shown. The two additional domains of the C-terminus are nearly identical in structure to those shown. The N-terminus of the molecule lies in the bottom centre of the picture and turns up into a helix. Because of the assigned flips, the green band on the outside of the yellow ribbon effectively marks the direction of the carbonyl groups in every helix. The green band is correctly positioned on the outside edge of each of the short stretches of anti-parallel beta structure which bind the two domains. Parallel structure would have been represented by the green band on the same side of each of the ribbons.

ACKNOWLEDGEMENTS

The authors would like to thank Mrs Cathy Holcomb for help in the preparation of this manuscript and Dr Jack Sack, Rice University, USA for providing the PS300 version of Frodo. This research was supported by NIH research grants CA-13148 and DE-02670.

REFERENCES

1 **Richardson, J S** 'The anatomy and taxonomy of protein structure' *Adv. Protein Chem.* Vol 34 (1981) p 167
2 **Lesk, A M and Hardman, K D** 'Computer-generated schematic diagrams of protein structures' *Science* Vol 216 (1982) p 539
3 **Pigue, M** *What does a protein look like?* (Video tape) University of North Carolina, USA (1982)
4 **Newman, W M and Sproull, R F** *Principles of interactive computer graphics* McGraw-Hill, USA (1979)
5 **Louie, A H and Somorjai, R L** 'Differential geometry of proteins; helical approximations *J. Mol. Biol.* Vol 168 (1983) p 143
6 **Jones, T A** 'A graphics model building and refinement system for macromolecules' *J. Appl. Crystallogr* Vol 11 (1978) p 268
7 **Senadhi, V-K et al.** 'Three-dimensional structure of ubiquitin at 2.8 Å resolution' *Proc. Natl. Acad. Sci. USA* Vol 82 (1985) p 3582
8 **Ealick, S E et al.** 'The crystal structure of human erythrocytic purine nucleoside phosphorylase at 6 Å resolution' *Ann. N.Y. Acad. Sci.* Vol 451 (1985) p 311
9 **Babu, Y S et al.** 'Three-dimensional structure of calmodulin' *Nature* Vol 315 (1985) p 37