# Ribbon models of macromolecules

**Mike Carson**

University of Alabama at Birmingham, Center for Macromolecular Crystallography, 252 BHS Box 79 THT, University Station, Birmingham, AL 35294, USA

*A smooth 3D ribbon model of a protein is easily constructed by creating a set of nearly parallel B-spline curves fitted to the peptide plane. These models may be coded by residue to depict such information as secondary structure, residue type or temperature factors. The computation of the curves defining the ribbon model is explained in detail. Solid surface models of the protein backbone may be rendered, using the ribbon curve as a basis. Nucleic acid molecules may be represented as ribbon models in much the same fashion. The method is applicable to both vector and raster devices, and is illustrated with the Evans and Sutherland PS300 and the Silicon Graphics Iris workstation.*

*Keywords: proteins, 3D ribbon models, B-spline curves*

Received 17 February 1987
Accepted 24 March 1987

The Richardson[1] 'ribbon drawing' gives an excellent representation of the folding and secondary structure of a protein molecule. A previous paper by the present author[2] described a simple algorithm to create ribbon models, as illustrated in Figure 1. The peptide planes are used to generate a series of 'guide coordinates' equally spaced along the desired ribbon width. Smooth and regular cubic B-spline curves fitted to these guide coordinates form the roughly parallel threads of the ribbon. The method is almost trivial to implement using a graphics device with built-in functions to perform the curve fitting.

This paper describes in detail the steps necessary to compute the line segments approximating the curve of the B-spline parameterization. This computation is necessary to colour-code the ribbon by residue on the Evans and Sutherland PS300. The line segments can define a polygonal net suitable for rendering on a raster device such as the Silicon Graphics Iris 2400T workstation. The computation of a 'ribbon curve', which may be used for the construction of solid diagrams, is then described.

The primary interest of this laboratory is protein-structure determination. However, the techniques presented are also applicable to nucleic acids. Figure 2 illustrates a ribbon drawing of a t-RNA molecule. The guide points are based on a vector defined by the positions of the P and O5 atoms of the phosphate–sugar backbone. This vector, scaled by the desired ribbon width, is centred at the phosphate position, and guide points are spaced equally along it. A more chemically meaningful definition may be possible. The figure produced captures the helical nature and folding pattern, and could be sectioned to show such information as base type, base tilt-angle or sugar conformation.

## CUBIC B-SPLINE CURVES

The matrix algebra required to create the line segments needed to model a complicated smooth curve is given below, following the presentation of the Iris manual and Clark[3]. The mathematical basis is the parametric cubic curve. These curves are the lowest-order representation that can ensure continuity of position, slope and curvature at the point where two such curve sections meet. A parametric cubic curve is defined as a 3rd-order polynomial for a dummy variable $t$, such that $x(t) = a_x t^3 + b_x t^2 + c_x t + d_x$, with similar equations for $y(t)$ and $z(t)$. A cubic curve, $C(t)$, may then be expressed as a vector product,

$$C(t) = [t^3\ t^2\ t\ 1](a\ b\ c\ d) = \mathbf{TP}$$

In practice, a curve section is given by specifying a set of four guide coordinates and a 'curve basis' that establishes how these guide coordinates determine the shape of the curve section. That is, $\mathbf{P} = \mathbf{BG}$, where **B** is the basis matrix and **G** the guide points. There are several classes of parametric cubic curves, all differing in the constraints imposed. Each has its own basis matrix. A B-spline curve is constrained to have continuous 1st and 2nd derivatives where curve sections meet. The B-spline curves do not, in general, pass through their guide coordinates.

### Computation of line segments

A 4 × 4 matrix, **M**, is computed from guide coordinates, the curve basis and the desired sampling as follows: $\mathbf{M} = \mathbf{SBG}$.

If $N$ line segments are desired to approximate the curve section in question, the sampling matrix, **S**, is given by:

$$\mathbf{S} = \begin{bmatrix} 6/N^3 & 0 & 0 & 0 \\ 6/N^3 & 2/N^2 & 0 & 0 \\ 1/N^3 & 1/N^2 & 1/N & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The basis matrix, **B**, for the cubic B-spline is given by:

 0263-7855/87/020103-04 $03.00  

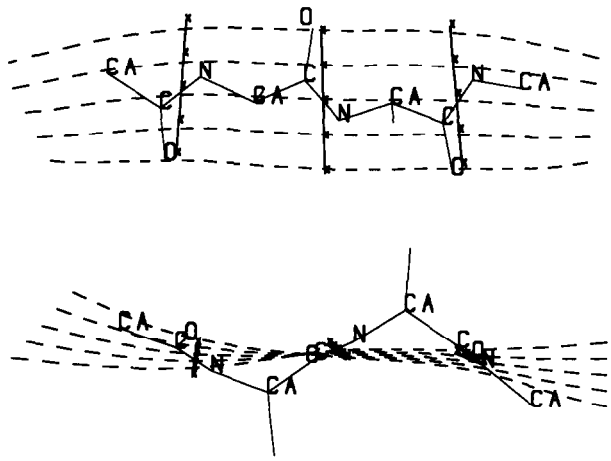

*Figure 1. Creation of ribbon models. The guide points (X) lie on a line in the peptide plane, halfway between the α-carbons. The five threads of the ribbon are shown as dashed lines. Plot from PS300 screen*

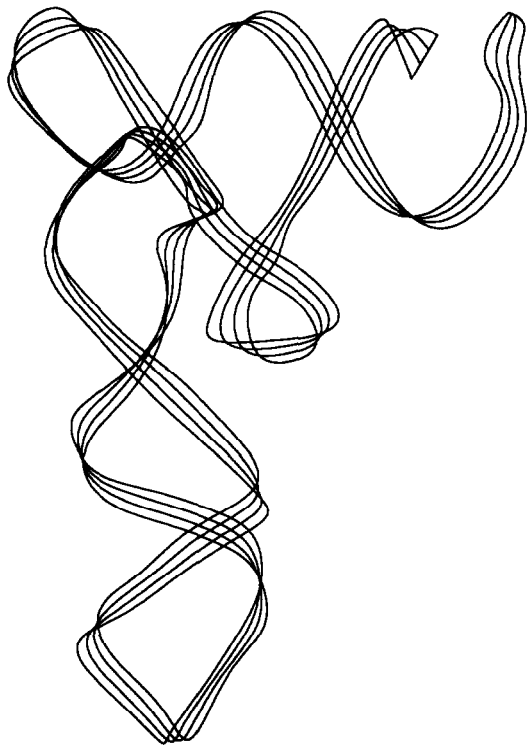

*Figure 2. B-splines of model FMet t-RNA (from unpublished results of Harvey and Prabhakaran). Plot from PS300 screen*

$$\mathbf{B} = 1/6 \begin{bmatrix} -1 & 3 & 3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}$$

The geometry matrix, **G**, is created from the four guide coordinates by:

$$\mathbf{G} = \begin{bmatrix} x1 & y1 & z1 & 1.0 \\ x2 & y2 & z2 & 1.0 \\ x3 & y3 & z3 & 1.0 \\ x4 & y4 & z4 & 1.0 \end{bmatrix}$$

The line segments are then generated using the 'forward difference algorithm' given in the C language below:

```
/*    Comments:
*
*     M[4][4] is the current matrix
*          (C arrays start with element 0)
*     N is the number of line segments to be drawn
*     Move_to() & Line_to () represent generic graphics functions
*/
Move to (M[3][0]/M[3][3], M[3][1]/M[3][3], M[3][2]/M[3][3]);
for ( k = 0; k < N; k = k + 1){
     for (i = 3; i > 0; i = i - 1){
          for (j = 0; j < 4; j = j + 1 ){
               M[i][j] = M[i][j] + M[i - 1][j];
          }
     }
     Line_to(M[3][0]/M[3][3], M[3][1]/M[3][3], M[3][2]/M[3][3]);
}
```

If another guide coordinate, $x5$, exists, the procedure is repeated with **G** calculated using $x2$ to $x5$. This will create a second section of $N$ line segments which join smoothly with the first section.

Consider the first five residues, 1 . . . 5, which define the protein's first four peptide planes. The four guide coordinates from these planes define a section of the curve corresponding to amino acid 3. The residues 2 . . . 6 would give a smoothly joining section corresponding to amino acid 4. Figure 1 shows four such sections in a β-sheet. Since the first four peptide bonds define the section for the third residue, it is necessary to construct dummy 'peptide planes' at the end of the molecule. The plane-defining vectors of the first plane are assigned to the N-terminus and to the first α-carbon to have the curve sections in register. Two additional dummy planes are set in similar fashion at the C-terminus.

In order to create a smooth ribbon coloured by residue on the PS300, one must compute the line segments of the B-spline and then sort into vector lists by residue. Colour Plate 1 gives an example of this technique. The protein ubiquitin[4] is colour-coded with respect to atomic temperature factors.

## B-splines on the PS300

The PS300 has a built-in firmware command to create B-splines. The example below gives the PS300 code needed to create a single curve approximating the backbone of a protein consisting of 50 amino acids, with ten line segments to be generated for each residue:

```
alphacurve := BSPLINE order = 4 N = 50
              x1, y1, z1        x2, y2, z2 . . .
              x50, y50, z50
              CHORDS = 500;
```

The 'ORDER = 4' specifies a cubic B-spline. The 50 coordinate triples could be α-carbon positions.

Using the methodology presented in the previous section, the vector list of line segments could be calculated and displayed. One would find this curve to be identical to that generated by the built-in command, except at the ends of the curves. This troublesome fact is a consequence of the PS300 implementation, where the B-spline is forced to begin on its first guide point and to end on its last guide point. This may be viewed as special constraints on the curve, with separate basis matrices required to impose these constraints[5]. The matrices for the beginning and end of the curve are given below. $\mathbf{B}_a$ is for the first four guide points, $\mathbf{B}_b$ for the next

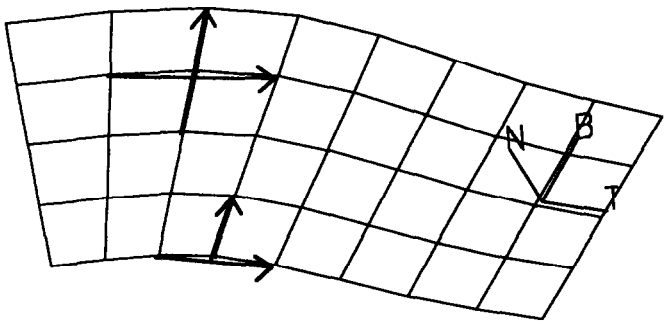


*Figure 3. Quadrilaterals constructed from adjacent B-spline curves. Vertex normals are taken as the vector cross product involving adjacent points on the grid. Arrows indicate how this is applied to a point on the interior and one on an edge. Orthonormal curve basis vectors on the central thread are indicated by T, B and N*

four, $\mathbf{B}_y$ for the next-to-last four guide points, and $\mathbf{B}_z$ for the last four.

$$\mathbf{B}_a = 1/12 \begin{bmatrix} -12 & 21 & -11 & 2 \\ 36 & -54 & 18 & 0 \\ -36 & 36 & 0 & 0 \\ 12 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{B}_b = 1/12 \begin{bmatrix} -3 & 7 & -6 & 2 \\ 9 & -15 & 6 & 0 \\ -9 & 3 & 6 & 0 \\ 9 & 7 & 2 & 0 \end{bmatrix}$$

$$\mathbf{B}_y = 1/12 \begin{bmatrix} -2 & 6 & -7 & 3 \\ 6 & -12 & 6 & 0 \\ -6 & 0 & 6 & 0 \\ 2 & 8 & 2 & 0 \end{bmatrix}$$

$$\mathbf{B}_z = 1/12 \begin{bmatrix} -2 & 11 & -21 & 12 \\ 6 & -15 & 9 & 0 \\ -6 & -3 & 9 & 0 \\ 2 & 7 & 3 & 0 \end{bmatrix}$$

It is impossible to create smoothly joined sections of the same curve with this implementation. Individual line segments must be computed.

### B-splines on the Iris

Smooth sectioned ribbon line drawings, colour-coded by residue, are easily created with the built-in functions of the Iris. However, the line quality is not as fine as with the PS300.

## RIBBONS ON RASTER

### Polygonal rendering on raster devices

Most raster systems offer facilities to create filled polygons. Figure 2 shows how a network of quadrilaterals is constructed from the endpoints of the B-spline line segments. Each quadrilateral may be split into two triangles to ensure a mesh of convex planar polygons. These polygons may then be rendered to create raster ribbon models. For a shaded image, it is necessary to compute a normal vector at each vertex of the polygon. A reasonable approximation using neighbouring points to form a vector cross product is seen in Figure 3.

The Iris is a raster machine with graphics library calls to perform the rendering of smooth shaded filled polygons with hidden surface removal. The methods of Gouraud shading, backfacing and Z-buffering are presented in standard graphics texts[6,7]. Colour Plate 2 gives an example of these procedures, showing the calcium ions bound to the protein calmodulin[8].

The ribbon surface has a distinct front, back, top and bottom. Here, the back is shaded grey and is always on the inside of the helices. The top is coloured cyan and marks the carbonyl direction of the helices. The front-to-back disposition of the small two-stranded sheet indicates its antiparallel nature. Small twists in the ribbon must be made to achieve this effect. The steps required to twist and to colour by thread to illustrate secondary structure are discussed in the original paper[2]. There is an error to report: the commentary on Colour Plate 3 should be taken as if the figure (slide) was viewed with a 180° rotation about Y.

### Ribbon space curve

Colour Plate 2 shows a 2D surface winding through 3D space. Consider the space curve that passes along the central thread of the ribbon, with its normal at every point coincident with the surface normal. This curve is then a geodesic in the ribbon surface and has interesting properties with respect to secondary structure (a paper on this is in preparation).

For present purposes, this ribbon curve is used as a basis for the construction of solid models. It may be approximated as in Figure 3, sampled at each vertex, with the vertex normal taken as the 'normal' and the vector spanning the width of the ribbon as the 'binormal'. As known from elementary differential geometry, these two vectors and the 'tangent' vector to the curve form an orthogonal basis set at every point on the curve. For example, a 'tube surface' of a space curve is constructed as the locus of points on a circle, spanned by the normal and binormal, passed along the length of a curve.

### Solid ribbon models

The solid ribbon model is developed from the ribbon curve with an ellipse. The major axis is along the binormal, the minor axis along the normal. The length of each axis is a function of the secondary structure. The turns and random coils are essentially tubes, while the ellipses broaden for helices and sheets. A quadrilateral grid as in Figure 3, created with the adjacent ellipses, is used for the actual rendering. Colour Plate 3 illustrates this, using the protein PNP[9].

Ribbon models clearly display the secondary structure of the tertiary backbone. To display more of the tertiary structure on the ribbon surface, the segments of the generating ellipses of each residue section are divided into four groups. The top of the ellipse corresponds to either the carbonyl oxygen direction or the amide NH direction of the peptide plane; the bottom is assigned to the other. The sense of this direction flips in the centre of a section of a residue belonging to an extended sheet. The front and back represent the broad faces of the ellipse. The front is designated as the direction in which the side chain is pointing. The normal vector at the centre of the section and the Cα → Cβ vector are taken, normalized and formed into a dot product to test the side-chain direction.

Colour Plate 4 illustrates the protein ubiquitin, colour-coded by residue type. Deep blue and deep red mark the tops and bottoms of the ellipsoidal surface, revealing the hydrogen bonding pattern of the polypeptide chain. Note the alternation of colours in the β-sheets and the constant carbonyl direction at the top of the helix. The front is brightly coloured with respect to the residue class. The front is always on the outside of the helix, while there is alternation in the familiar pattern of the β-sheet. The back is rendered in a drab grey, with an arrow marking the direction of the polypeptide chain. The coloured stripe bordering each grey back section indicates the type of the residue pointing away from the viewer.

## DISCUSSION

It is hoped that this paper will increase awareness of B-spline curves. Lesk and Hardman[10] used spline curves as part of their arrow-and-cylinder Artek plots, but gave no details. The work done here is in response to the need for a ribbon model for interactive 3D viewing on high-performance computer-graphics devices. It is influenced by the work at the University of North Carolina[11]. The implementation chosen is a function of the equipment in the author's laboratory.

This ribbon model is geometrically based on the peptide plane, yet is not restrained to the planes. In the colour plates, neighbouring strands of β-sheet align smoothly. An atomic model would form the standard pleated β-sheet. The Cα atoms of a sheet lie alternately above and below the surface of the ribbon. In the colour plates, the helices appear cylindrical, while an atomic model of the peptide planes would form a more polygonal outline. The algorithm is dependent on secondary stucture to compensate for the width of the helices. The adjustment places the Cα roughly at the surface of the ribbon. The ribbon surface cuts the corner a little, rounding sharp turns.

It is possible to fit curves with constraints of one's choice. Similar formulations with different constraints and/or more guide coordinates could force the ribbon surface to follow the plane more closely, or to pass through all the α-carbons. Smoother curves are also possible. The PS300 B-spline command allows construction of higher-order polynomials by adjustment of the 'order' parameter. Here the order is set to 4, forming a cubic curve with continuous 1st and 2nd derivatives. Setting the order higher produces only marginally smoother curves, while setting it lower, destroys the smoothness. The presence of a curve with a continuous 2nd derivative appears mandatory from visual inspection. The Iris offers a variety of curve formulations. Bezier and Cardinal splines were tried in various ways, and the number of B-spline guides was increased to force a pleat in the sheet, but no alternative was deemed satisfactory. A bicubic patch of B-splines can provide a more analytical surface representation and may provide a smoother raster image.

The algorithm is an effective artifice. The representations are aesthetically pleasing, having the feel of Richardson's original ribbon drawings. The ribbon curve presented gives a practical basis for ribbon modelling it also allows fanciful computer recreation, as the viewer may look from afar or from the cockpit of a 'peptide plane' that flies through a line drawing of the protein, hugging the ribbon surface. Viewing of the actual raster rendering of the surface, which proceeds a polygon at a time from the N- to the C-terminus of the molecule, is highly recommended to anyone speculating on the folding pathway of the molecule.

## COMPUTER PROGRAMS

The software developed to produce these ribbons is available in two packages. The first set runs in a VAX/VMS environment and is written in the C language. The program runs interactively, reading atomic coordinates and processing options, then writing PS300 firmware. The firmware is loaded into the PS300 for display using the PSFrodo firmware. The line drawings of ribbons or of individual threads or sections are assigned as 'Mol' objects, the display of which is toggled with buttons. These may be viewed interactively in real time. The Frodo[12] source code could be modified to call these routines directly.

The second set runs in a Unix environment on the Iris and is written in the C language. Line drawings are displayed and may be transformed in real time. Once a suitable view is established, a variety of rendering options may be chosen. Readers interested in obtaining the programs should contact the author.

## ACKNOWLEDGEMENTS

The author would like to thank his crystallographic colleagues for their advice and their atomic coordinates, Charlie Bugg for his enthusiasm, encouragement and support, Scott Rowland for the photography, and Mary Ann Lynch for help in the preparation of the manuscript.

## REFERENCES

1 **Richardson, J S** 'The anatomy and taxonomy of protein structure' *Adv. Protein Chem.* Vol 34 (1981) p 167
2 **Carson, M and Bugg, C E** 'Algorithm for ribbon models of proteins' *J. Mol. Graph.* Vol 4 No 2 (June 1986) pp 121–122
3 **Clark, J H** *Parametric curves, surfaces and volumes in computer graphics and computer-aided geometric design* Technical Report No 221, Computer Systems Laboratory, Stanford University, USA (1981)
4 **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
5 **Harrington, S** *Computer graphics: a programming approach* McGraw-Hill, USA (1983)
6 **Foley, J D and Van Dan, A** *Fundamentals of interactive computer graphics* Addison-Wesley, USA (1984)
7 **Newman, W M and Sproull, R F** *Principles of interactive computer graphics* McGraw-Hill, USA (1979)
8 **Babu, Y S et al.** 'Three-dimensional structure of calmodulin' *Nature* Vol 315 (1985) p 37
9 **Ealick, S E et al.** 'The crystal structure of human erythrocytic purine nucleoside phosphorylase at 6 Å resolution' *Ann. NY Acad. Sci.* Vol 451 (1985) p 311
10 **Lesk, A M and Hardman, K D** 'Computer-generated schematic diagrams of protein structures' *Science* Vol 216 (1982) p 539
11 **Pique, M** *What does a protein look like?* (Video tape) University of North Carolina, USA (1982)
12 **Jones, T A** 'A graphics model building and refinement system for macromolecules' *J. Appl. Crystallog.* Vol 11 (1978) p 268