NEW PROGRAMS

SETOR: Hardware-lighted three-dimensional solid model representations of macromolecules

Stephen V. Evans

Department of Biochemistry, University of British Columbia, Vancouver, British Columbia, Canada

SETOR is designed to exploit the hardware lighting capabilities of the IRIS-4D series graphics workstations to render high-quality raster images of macromolecules that can undergo rotation and translation interactively. SETOR can render standard all-atom and backbone models of proteins or nucleic acids, but focuses on displaying protein molecules by highlighting elements of secondary structure. The program has a very friendly user interface that minimizes the number of input files by allowing the user to interactively edit parameters, such as colors, lighting coefficients, and descriptions of secondary structure via mouse activated dialogue boxes. The choice of polymer chain representation can be varied from standard vector models and van der Waal models, to a B-spline fit of polymer backbones that yields a smooth ribbon that approximates the polymer chain, to strict Cardinal splines that interpolate the smoothest curve possible that will precisely follow the polymer chain. The program provides a photograph mode, save/restore facilities, and efficient generation of symmetry-related molecules and packing diagrams. Additionally, SETOR is designed to accept commands and model coordinates from the standard input stream, and to control standard output. Ancillary programs provide a method to interactively edit hardcopy plots of all vector and many solid models generated by SETOR, and to produce standard HPGL or PostScript files. Examples of figures rendered by SETOR of a number of macromolecules of various classes are presented.

Keywords: molecular modeling, interactive computer graphics, solid models, secondary structure, hardware lighting, IRIS-4D workstations

Color Plates for this article are on pages 127–128.

Address reprint requests to S.V. Evans, National Research Council of Canada, Institute for Biological Sciences M54, Ottawa, Canada, K1A 0R6.

Received 6 March 1992; accepted 1 June 1992

INTRODUCTION

As the number and complexity of determined macromolecular structures increases, it has become more necessary to create simplified and visually understandable representations of these molecules. Some of the most useful and effective simplifications are the diagrams developed by Richardson,[1] where the course of the backbone of a protein is approximated by a smooth ribbon. These initial diagrams were usually sketched by hand, a process that required several hours of careful work to produce an esthetically pleasing model. Programs have since been written to produce very effective line drawings of protein ribbons,[2] but which still offer only a static view of the molecule. Programs to render depth-cued ribbon models of macromolecules rendered in three dimensions have also been written by Carson,[3,4] who used B splines as the basis for interpolating a smooth curve along polymer backbones to produce visually pleasing vector and solid models.

With the advent of computer hardware that is dedicated to rendering lighted surfaces, the capabilities of computer graphics systems have now increased to the extent that it has become possible to render lighted solid models of macromolecules in three dimensions that can be manipulated in real time. SETOR has been designed to combine the advantages of hardware lighting and high-speed transformations available on the Silicon Graphics IRIS-4D series computers with a facility to render hardcopy black and white outlines of most images on a conventional plotter, to provide a full spectrum of rendering capabilities.

PROGRAM DESCRIPTION

Overview

SETOR is written in C utilizing the Silicon Graphics Inc. Graphics Library. The basic structure of the program has keyboard input kept to a minimum, with as many variables as possible available for user modification via mouse activated

dialogue boxes. Consequently, SETOR has a very user friendly interface.

SETOR will accept orthonormal coordinates in either Wayne Hendrickson's PROLSQ format,[5] Brookhaven PDB format,[6] or formatted CHARMm coordinate files. As the coordinates are read they are separated into polymer chains based on the chain labels or *segid*, and, in the case of PDB files, by *TER* records as well. For efficiency in memory management, space is allocated dynamically as each coordinate file is read. The maximum number of polymer chains is currently set to eight, but can be easily redimensioned at compilation time. The maximum number of independently displayable residues is only limited by the computer hardware (to about 65000). Each polymer chain and residue can be independently colored, rendered, and displayed.

In interactive mode, the user is presented with a hierarchy of pop-up menus from which program variables can be accessed. As SETOR does not, in general, use input files to determine program function, but relies on interactive control, many different representations can be explored in a very short time. As well, no matter what method the user chooses to represent the molecule, all models in SETOR are subject to three-dimensional manipulation. This means that the user can choose to view and record the molecule while it is in motion, or to fine tune the most advantageous single orientation for viewing.

Cubic splines

Although SETOR can display standard wire-frame vector or solid models, or van der Waal's spheres, most representations of molecules within it make use of a series of connected cubic splines interpolated along the polymer backbone. Two sorts of cubic splines are employed. The use of B splines in a manner similar to that suggested by Carson[3,4] produces a very smooth interpolation of the polymer backbone; however, a B spline does not in general pass though any of the control points used to define it. The Cardinal spline will not produce as smooth a ribbon as the B spline (Color Plate 1a), but it will curve to exactly fit the spline through each control point in the backbone, and so will be a truer representation of the course of the polymer chain. Additionally, a Cardinal spline has more flexibility in the route chosen between the control points, where instead of interpolating a smooth curve between control points the spline can curve sharply about each control point to yield an elegant α-carbon type skeleton (Color Plate 1b).

While the spline is used to define the course of the ribbon, the plane of the ribbon can be set within SETOR either to follow the main chain carbonyl oxygen bond vectors, the amide bond planes, or to be perpendicular to the α–β bond of the side chains.

SETOR can render the backbone as a two-dimensional "flat" ribbon, or as a three-dimensional ribbon with triangular, rectangular, circular, or elliptical cross section.

Representation of secondary structure

Even ribbon diagrams of molecules can be quite complex if the molecules being rendered contain a large number of atoms, and so SETOR provides a facility for inserting simplified representations of elements of secondary structure by assigning each residue along the polymer chain a secondary structure descriptor. The descriptors provided are based on the schematic diagrams of Richardson,[1] and have very specific definitions within SETOR.

The first two secondary structure descriptors set the cross section without changing the course of the ribbon spline:

- **ribbon.** The default secondary structure descriptor for all residues read from a coordinate file, the ribbon descriptor has user-definable breadth and width. Ribbons are the most common method of representing the spline, and can be used to illustrate regions of well-defined secondary structure such as helices or β-sheets; however, as will be seen below, a much more pleasing image is obtained by specifically restraining those elements of secondary structure.
- **rope.** This descriptor has an isometric cross section, regardless of what sort of three-dimensional ribbon spline is rendered. The rope is usually chosen to represent regions of less well-defined secondary structure.

The remaining descriptors restrain the course of the ribbon spline, or replace it with a simplified representation:

- **arrow.** An arrow usually is placed at the C-terminal end of β-sheets to show the direction of the polypeptide chain.
- **sheet.** The course of the ribbon spline through each set of residues designated as sheets is restrained to lie in a smooth curve.
- **helix.** Two sorts of operations are possible with helix descriptor. SETOR can either take the residues defined as a helix and perform some idealization to present a more visually pleasing ribbon, or can perform a least-squares fit of the tube axis to the atomic positions, and replace that section of the polypeptide chain with a smooth straight tube.
- **loop.** The course of the ribbon spline is set to the shortest smooth path between the endpoints of a loop segment, and is used to "clean up" loops with poorly defined secondary structure, and thus render a more stylized diagram.

Secondary structure descriptors are either input via the secondary structure editor (Color Plate 2), or by picking the residues off the screen with the cursor, or by reading a file of descriptors. The secondary structure editor provides a highly suggestive means of positioning elements of secondary structure within the molecule by using icons to represent each residue and type of secondary structure element. The mouse-activated editor makes it possible to explore different representations of protein structure in just a few minutes. A selection of molecules of various classes illustrating the use of secondary structure representations is given in Color Plate 3.

Lighted surfaces

Each lighting environment consists of the properties of the light, surfaces, and surroundings. These parameters are described fully in standard graphics textbooks, and the *Silicon Graphics Programming Guides*. The following section is an

outline of the manner in which the available lighting tools have been implemented in SETOR.

To display lighted surfaces, the direction, color, and intensity of the light used must be specified. The IRIS computers support at least 8 independent light sources in hardware. SETOR implements these lights in two ways. First, SETOR defines 3 lights as "external" light sources that are always located outside the viewing frustum. Each external light source can be interactively positioned so as to provide overall illumination of the macromolecule from any direction. The remaining 5 lights are implemented as "bound" light sources, where each can be positioned at a user-specified residue. This provides a light source that is visible in the rendering, and that serves to illuminate the interior of the molecule and give depth to the image. In addition to the position of each light, its other properties (such as color and intensity) can be individually and interactively adjusted. Color Plate 4 illustrates the use of internal light sources for the calcium binding protein oncomodulin.[7]

As all objects are illuminated and viewed by light, the surface properties of an object refer to those that define the ability of a surface to scatter and reflect light. As with the light source characteristics, surface properties can be manipulated interactively. To maintain a consistent image, all surfaces rendered in SETOR share the same qualities of ambient and diffuse reflectance, as well as specular highlights and shininess. Surfaces of different objects are distinguished by their colors, which are set independently.

Finally, the global parameters of the scene can be adjusted from within SETOR. The most important quality is the degree to which light is attenuated with distance. A strong depth-cued effect can be created using lights positioned in front of the objects with the lighting attenuation turned on. Color Plate 5 shows the van der Waal's surface of myoglobin[8,9] illuminated by an attenuated light source. (Note that for IRIX 3.3.2, light attenuation is implemented differently on IRIS VGX and GT series computers.)

Selection of colors

With 8 bits of color available on each of the red, green, and blue portions of the image, there are 16.8 million different colors available on the fully configured IRIS computers. The SETOR color editor (Color Plate 6) provides a convenient method of interactively selecting practically any of these colors. As colors are selected they can be individually bound to the spline backbones, to any and all side chains and residues, to elements of secondary structure, to the background color, or to the lights that illuminate the scene.

Streaming input

As an alternative to interactive sessions, SETOR can also accept commands and coordinates from an input stream. This enables preprogrammed sequences to be produced, with control over most of the aspects of image rendering that are available in interactive mode. As well, the user can customize SETOR's output stream so as to direct software controlling recording apparatus.

Polygon antialiasing and the accumulation buffer

The IRIS-4D VGX platforms provide an elegant method of polygon antialiasing in the form of an accumulation buffer. This can be implemented in real time if the accumulation buffer exists in hardware, but can also be accomplished in those systems with the software accumulation buffer, although the latter option takes several seconds to render a single frame.

Symmetry-related molecules

SETOR generates symmetry-related molecules by reading and parsing standard crystallographic symmetry operators as in the *International Tables for X-ray Crystallography*,[10] and calculating the corresponding rotation and translation matrices to use in orthonormal space. Symmetry operations can be specified either by manually typing in the desired relationships, or by specifying the space group (SETOR will read its own database of the symmetry elements of the most common chiral space groups). The input symmetry relationships apply to all polymer chains. Once the symmetry is specified, the corresponding molecules can be toggled to display using either the keyboard or the mouse. SETOR can display up to 24 symmetry-related molecules.

Hardcopy plots

Although SETOR is designed primarily to render high-quality color raster images, often a black-and-white hardcopy vector plot is more convenient, and so SETOR is equipped with facilities for reproducing most images on laser printers. The SETOR package contains several ancillary programs, one of which is a vector image editor (called *setorplot*, Color Plate 7) that allows the interactive manipulation of vector images output from SETOR. Among these features are the ability to alter individual line thicknesses of objects that had different colors in SETOR, stereo versus mono plots, stereo depth, font, label text, label position, and label size. Figure 1a shows a ribbon plot of flavodoxin[11] using SETOR and setorplot.

Setorplot will also reproduce vector outlines of many of SETOR's solid models, and implements its own z-buffer to remove hidden surfaces in just a few CPU seconds. For example, the outline of the solid model of Taka amylase[12] in Figure 1b, produced using setorplot, took only 12 CPU seconds to remove all hidden surfaces and plot the image on an IRIS 4D/440 VGX.

Setorplot interfaces with additional programs that can produce output that is suitable for PostScript and some HPGL printers.

Hardware requirements and availability

Currently SETOR only runs on Silicon Graphics IRIS-4D GT or higher machines, including the Personal IRIS, although there are plans to port the program to other platforms. The program requires 24 bitplanes RGB with full z-buffer capability and at least 16-Mb memory. SETOR is available on a nonprofit basis to any academic institution.

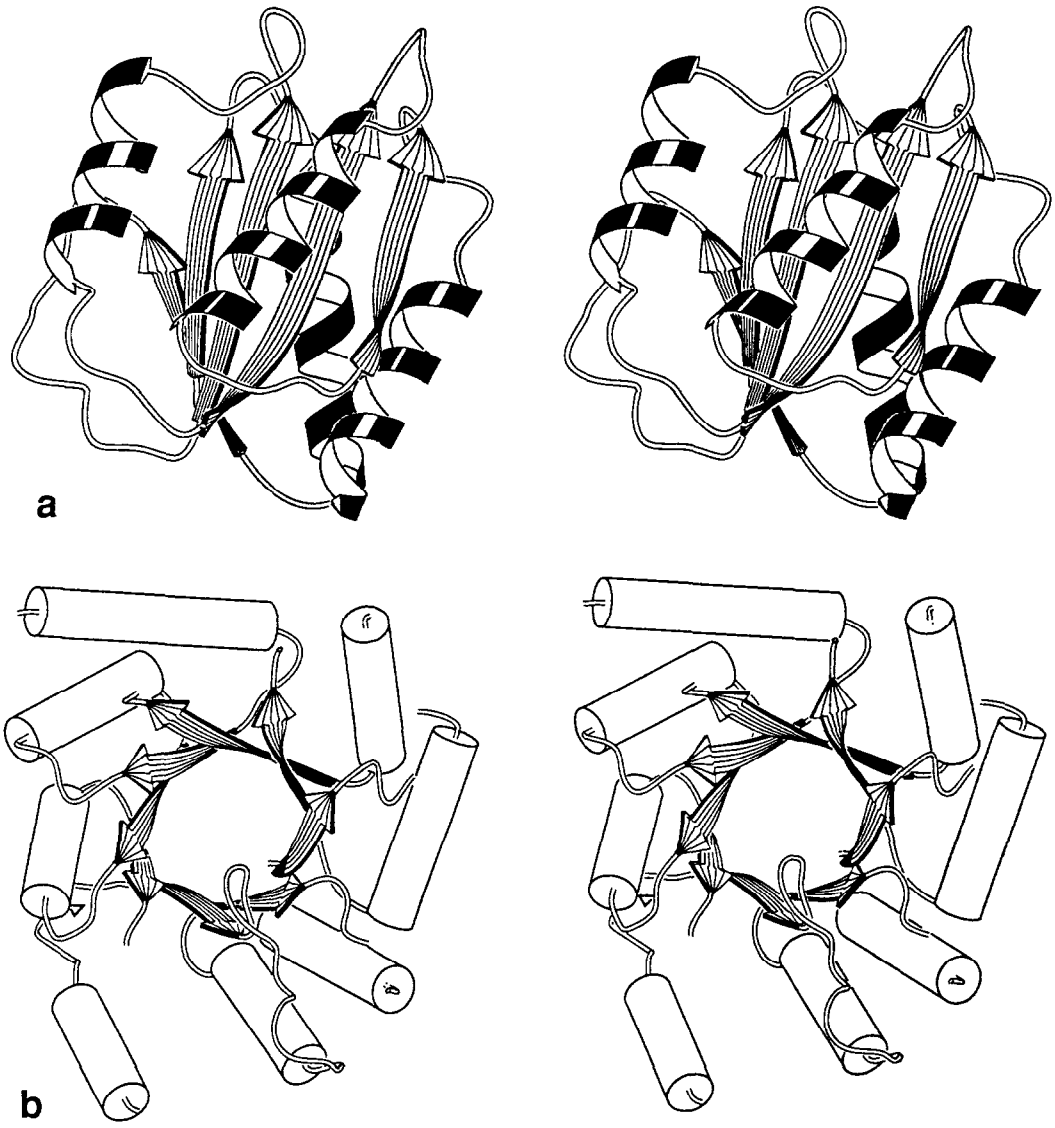

Figure 1. Stereoviews of vector outlines of solid models of (a) flavodoxin and (b) The $(\alpha - \beta)_8$ barrel of Taka amylase, produced using SETOR and setorplot. Setorplot took 7 and 12 CPU seconds on an IRIS 4D/440 VGX, respectively, to render the two images with all of the hidden lines removed.

ACKNOWLEDGEMENTS

I would like to thank Albert M. Berghuis, Dr. Gary D. Brayer, Dr. David R. Rose, and Dr. Alastair K. Muir for helpful suggestions, and Benny Cifelli of Silicon Graphics Inc. for generously allowing me extended access to several 4D demo workstations.

REFERENCES

1 Richardson, J.S. The anatomy and taxonomy of protein structure. *Ad. Protein Chem.* 1981, **34** (2) 167–339

2 Priestle, J.P. RIBBON: a stereo cartoon drawing program for proteins. *J. Appl. Crystallogr.* 1988, **21**, 572–576

3 Carson, M., and Bugg, C.E. Algorithm for ribbon models of proteins. *J. Mol. Graphics* 1986, **4** (2) 121–122

4 Carson, M. Ribbon models of macromolecules. *J. Mol. Graphics* 1987, **5** (7) 103–106

5 Hendrickson, W.A., and Konnert, J. PROLSQ. in *Biomolecular Structure, Function, Conformation and Evolution.* (R. Srinivasan, Ed.) Pergamon Press, Oxford, 1981, 43–57

6 Bernstein, F.C. The protein data bank: a computer archive. *J. Mol. Biol.* 1977, **112,** 535–542

7 Ahmed, F.R., Przybylska, M., Rose, D.R., Birnbaum, G.I., Pippy, M.E., and MacManus, J.P. Structure of Oncomodulin refined at 1.85-Å resolution. *J. Mol. Biol.* 1990, **216**, 127–140

8 Evans, S.V., and Brayer, G.B. The structure of

horse heart metmyoglobin to 2.8-Å resolution. *J. Biol. Chem.* 1988, **263,** 4263–4268
9 Evans, S.V., and Brayer, G.B. High resolution study of the three dimensional structure of horse heart metmyoglobin. *J. Mol. Biol.* 1990, **213**, 885–897
10 *International Tables for X-ray Crystallography*. Vol. I. (N.F.M. Henry and K. Lonsdale, Eds.) Kynoch Press, Birmingham, England
11 Smith, W.W., Burnett, R.M., Darling, G.D., and Ludwig, M.L. Structure of the semiquinone form of flavodoxin from *Clostridium MP*. *J. Mol. Biol.* 1977, **117**, 195–225
12 Matsuura, Y., Kusunoki, M., Harada, W., and Kakudo, M. Structure and possible catalytic residues of Taka-amylase A. *J. Biochem.* 1984, **95**, 697–702