VENUS — a program to display protein structure using raster colour graphics

Yuichi Iga and Noritake Yasuoka

Crystallographic Research Center, Institute for Protein Research, Osaka University, Yamadaoka 3-2, Suita, Osaka 565, Japan

A Fortran program is described which retrieves the coordinates and connectivities of macromolecular structures and displays them in a variety of styles. These include the wire model, the ball-and-stick model and the space-filling model. The program picks up the coordinates of the α carbon atoms of a protein molecule. They are connected with either lines or platelets to give a wire or pleat model. The coordinates may also be connected by a smooth line, a cubic spline, to give a ribbon model. These models are very helpful in understanding the folding pattern, as well as the secondary structure of a given protein molecule. A user can display a limited sequence of a chain, eg to inspect the active site. Seven colouring schemes are available to differentiate backbone atoms from side-chain atoms, or hydrophilic residues from hydrophobic residues, and so on. Alphanumerics can be displayed to label atoms and residues.

Keywords: raster scan graphics, Fortran, macromolecular structures, modelling

received 2 April 1984, revised 29 May 1984

Rapid advances in LSI technology have made it possible to produce raster scan graphics displays at a moderate price. Although they are still inferior to random scan graphics displays in resolution, they are almost equal in line-drawing capacity. The increasing application of raster graphics in fields such as CAD/CAM (computer-aided design/computer-aided manufacture) will lead to mass production and therefore to lower prices.

It is necessary to use graphical representations in order to understand the 3D structure of a molecule. Some computer programs have been developed and widely used for this purpose in chemical crystallography. Among them, ORTEP by C K Johnson[1] and PLUTO by S Motherwell[2] are widely used to prepare figures on a digital plotter. The construction of a graphics program is strongly dependent on a given hardware, and consequently graphics programs are less portable than plotter programs. BILDER by R Diamond[3] and FRODO by T A Jones[4] are implemented in some protein crystallography laboratories.

We have been developing an online information retrieval system for protein structure data[5]. The coordinates and the connectivities of protein molecules are stored on a mainframe computer and are easily retrieved to produce graphical representations. We have modified PLUTO to prepare pictures on a storage-type graphics display.

Recently a new model of a colour graphics display was installed in our institute. In the present paper we describe a Fortran program which was developed as part of our system in order to display macromolecular structure. The program is named VENUS, after the planet in the solar system, with respect to PLUTO.

OUTLINE OF HARDWARE

The graphics device used is an N6960 colour graphics display supplied by Nippon Electric Co, an OEM version of GR2403 produced by Seiko Electronic Industries. The display is connected to an ACOS 850 computer via an RS232C interface. The size of the 1024 × 1024 CRT (cathode ray tube) is 275 mm × 275 mm. Seven colours are available: white, red, blue, green, yellow, magenta and cyan. The display is equipped with four frames of digital video memories (1024 × 1024 × 4) and a 760 kbyte segment buffer.

FUNCTIONS OF VENUS

The program is written in ACOS Fortran and uses the graphics package GDSP/SPLOT; it is approximately 8000 statements in length. A summary of the characteristics of the program follows.

Command architecture

User interaction with the program is accomplished by the selection of a command from the command menu. A cross-hair cursor is shown on the CRT and can be moved by a pen and tablet. When the cursor reaches a desired command, the user presses the pen stylus, and the selected command is recognized by the program. There are eight sections in the command menu: FILE, FORM, SELECT, MODE, ROTATE, SETUP, ZOOM and EXIT (see Figure 1).

Each section has its own commands and these are displayed in place of the section menu. A command is selected by operating the stylus. Some commands are accompanied by subcommands, and these are displayed

under the command menu if selected. The instructions in VENUS have a tree structure with three levels. A user has to pay attention to the structure to obtain the desired representations.

Description of a dataset

The coordinates and the connectivities are supplied to the program in formatted form. The name of a dataset is input after the FILE section is selected. A sample of data is given in Figure 2. The format of the coordinate data is the same as that used in the Brookhaven Protein Data Bank[6] and that of the connectivity follows PLUTO. These data are easily retrieved by our system. A user can prepare the data in this form from his own structure using a program developed by the present authors (unpublished work).

A variety of display modes

Several representations are used to display the molecular structure. They include the wire model, the pleat model, the smooth-line model, the ribbon model, the

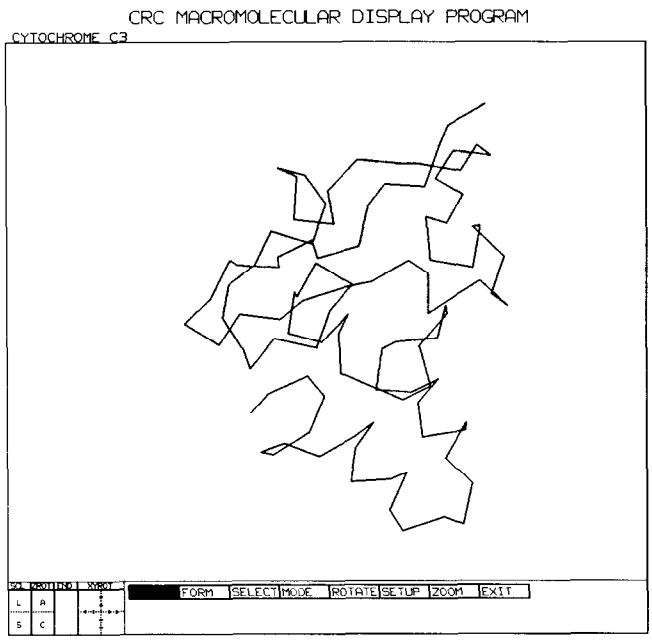

Figure 1. The first menu of the Crystallographic Research Center display program, VENUS. The name of the protein is shown over the frame, and the command menu is shown below

```
DATA 2PAR       ALLATOM
CELL   1.0   1.0   1.0   90.0   90.0   90.0
COMPND     PREALBUMIN (HUMAN PLASMA)
ATOM     58  N    CYS A  10       -5.617  10.210  23.270  1.00  0.   000
ATOM     59  CA   CYS A  10       -5.491  11.476  22.525  1.00  0.   000
ATOM     60  C    CYS A  10       -6.472  11.349  21.338  1.00  0.   000
ATOM     61  O    CYS A  10       -6.648  10.235  20.850  1.00  0.   000
ATOM     62  CB   CYS A  10       -4.104  11.563  21.847  1.00  0.   000
ATOM     63  SG   CYS A  10       -2.761  10.449  22.310  1.00  0.   000
ATOM     64  N    PRO A  11       -6.982  12.502  20.920  1.00  0.   000
ATOM     65  CA   PRO A  11       -7.959  12.526  19.826  1.00  0.   000
ATOM     66  C    PRO A  11       -7.381  12.305  18.419  1.00  0.   000
JOIN      58        59        60        61
JOIN      59        62        63
JOIN      60        64
JOIN      64        65        66        67
JOIN      64        70
```

Figure 2. A part of a dataset for the coordinates and connectivities of prealbumin is shown. The former is in the format of the Protein Data Bank and the latter in the format in PLUTO

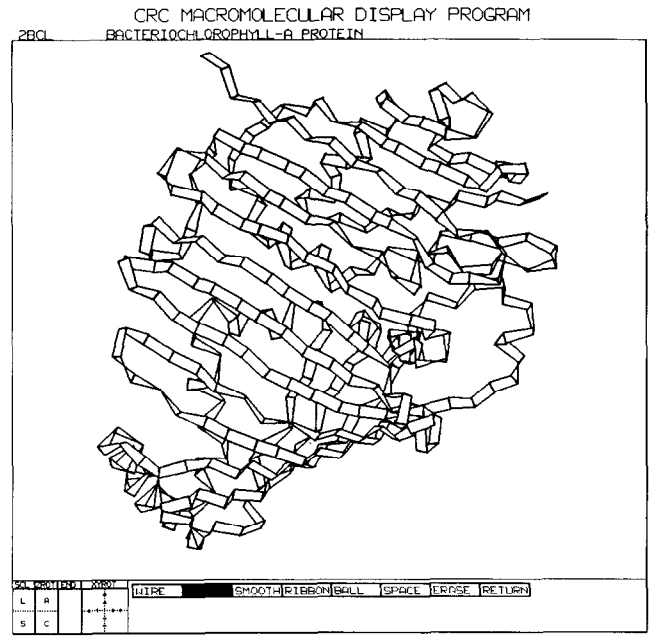

Figure 3. An example of the pleat model is shown for Bacteriochlorophyll A protein

ball-and-stick model and the space-filling model. The hidden-line algorithm is applied to all the representations.

It is not necessary to explain the wire model. The following three representations are less well known and are suitable for visualizing the characteristic features of the folding pattern and the secondary structure of a given protein. Let the ith α carbon atom be C_i. A line, 0.5 Å in length, is defined from each C_i, which is perpendicular to the plane passing through C_{i-1}, C_i and C_{i+1}. The ends of each line are connected in sequence to give a pleat model. An example is shown in Figure 3. The coordinates of α carbon atoms can be connected by a cubic spline[7]. This gives a smooth-line model.

The smooth-line model is widened along the tangent to give a ribbon model. The colour of the front side is different from that of the reverse side. An example is shown in Colour plate 1 (see p C). The ball-and-stick model is well known. The stick bond may be painted or unpainted. The space-filling model represents the van der Waals surface of the molecule. Each atom is assigned an accepted radius, but this may be altered by the user. Several modes of space-filling model are available: a drawing of the envelope is the simplest of these and is not time-consuming. The 3D impression of the structure may be increased by distributing a number of points on the van der Waals surface. We use the following equation:

$$D = a\cos\theta + b\cos^2\theta + c$$

where D is the density of dots on the surface, θ is the angle made by the eye line with a given radius and a, b and c are constants (the default values are 1.2, −0.2 and 0). Similar equations to represent the shaded surface are discussed by Brickman[8]. The density of the dots is modulated according to the z coordinate of each atom:

$$D' = D\{1 - 0.8(z - z_{\min})/(z_{\max} - z_{\min})\}$$

where z_{max} and z_{min} denote the maximum and minimum value for the z coordinate of the atoms. This modulation gives the displayed picture a 3D appearance. An example is shown in Colour plate 2. A combination of the described display modes is available, eg a ball-and-stick model may be superimposed on a space-filling model.

Display range, atom labels and colours

It is not always best to display the whole molecule. We have already described the suitability of the α carbon model for understanding the secondary structure. In order to examine the structure of the polypeptide in detail, it is adequate to display the backbone model.

In proton NMR (nuclear magnetic resonance) studies of a protein, the behaviour of particular residues can be detected. The signals corresponding to aromatic residues, especially histidine residues, are often observed separately in the spectra. In these studies it may be helpful to show the molecular model as backbone atoms with histidine residues. It is often important to examine the local structure of a protein in detail. VENUS may be used to display the active sites of an enzyme (catalytic sites or substrate binding sites) in detail. Such a representation is widely used in the field of drug design[9,10]. For these purposes commands have been included which enable the desired regions of a given enzyme to be extracted. Another command enables the user to select atoms within a given radius, say 10 Å. It is also important to be able to display a molecular model with atom labels. The name of a residue or an atom, with or without sequence number, can be appended to the model.

Seven colours are available to differentiate the types of residues or atoms and these are very helpful in understanding the structure. An example is shown in Colour plate 3.

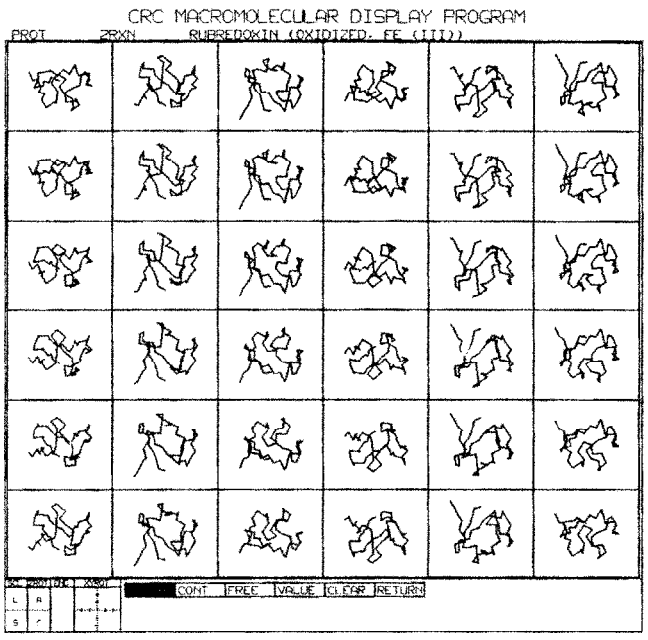

Figure 4. The connected α carbon atoms of Rubredoxin are rotated successively by 10 degrees around the x *axis. The resulting 36 figures are shown simultaneously*

Pseudo-dynamic manipulation of a model

An N6960 graphics display is 2D and it is difficult to rotate and translate the model in real-time. VENUS, however, has some interactive modes. The greater the number of vectors displayed, the more difficult is the dynamic operation. Therefore we deal with α carbon atoms only and prepare 36 models which are rotated in sequence by 10° around the x-axis. These figures are simultaneously shown on a display (Figure 4). When a user selects CONT subcommand, the figures are displayed successively. The molecule in the picture then looks as if it is rotating. Enlargement and reduction of the displayed picture and clipping of the model are also possible. This function also enables a user to define the desired volume of the molecule. The mode of the display is specified by instruction.

CONCLUSIONS

A program to display the structure of a protein molecule in various modes, including dynamic interactions, using raster graphics is described. This type of display is quite inexpensive (about 5 million yen (U.S. $20 000) at the beginning of 1984) and may be widely used in CAD/CAM applications. A 3D raster graphics display is also being developed. We are planning to develop a program system for this display in the field of protein crystallography and molecular design.

ACKNOWLEDGMENT

The authors would like to express their sincere thanks to Dr Masami Kusunoki for helpful discussions.

REFERENCES

1 **Johnson, C K** 'ORTEP-II: a fortran thermal ellipsoid plot program for crystal structure illustrations' *ORNL-5138* Oak Ridge National Laboratory, TN, USA (March 1976)

2 **Motherwell, S** 'Program for plotting molecular structure illustrations — PLUTO78' Crystallographic Data Centre, University Chemical Laboratory, Cambridge, UK (April 1978)

3 **Diamond, R** 'BILDER: A computer graphics program for biopolymers and its application to the interpretation of the structure of tobacco mosaic virus protein discs at 2.8 Å resolution' in **Srinvasan, R et al (eds)** *Biomolecular structure, function, conformation, and evolution* Pergamon Press (1981) pp 567–588

4 **Jones, T A** 'FRODO: a graphics model building and refinement system for macromolecules' *J. Appl. Crystallogr.* Vol 11 (1978) pp 268–272

5 **Taketani, M, Iga, Y, Matsuura, Y, Yasuoka, N, Kakudo, M and Isomoto, Y** 'On-line information retrieval system on protein structure data and interactive graphics display in protein crystallography' in **Glaeser, P S (ed)** *Data for science and technology* Pergamon Press (1981) pp 84–87

6 **Bernstein, F C, Koetzle, T F, Williams, G T B, Meyer, E F, Brice, M D, Rodgers, J R, Kennard, O, Shimanouchi, T and Tasumi, T** 'The Protein Data Bank: a computer-based archival file for macro-

molecular structures' *J. Mol. Biol.* Vol 112 (1977) pp 535–542

7 **Ahlberg, J H, Nilson, E N and Walsh, J L** 'The theory of splines and their applications' Academic Press (1967)

8 **Brickman, J** 'Shaded surface in molecular raster graphics' *J. Mol. Graph.* Vol 1 (1983) pp 62–67

9 **Blaney, J M, Jorgensen, E C, Connoly, M L, Ferrin, T E, Langridge, R, Oatley, S J, Burridge, J M and Blake, C C F** 'Computer graphics in drug design: molecular modeling of thyroid hormone–prealbumin interactions' *J. Med. Chem.* Vol 25 (1982) pp 785–790

10 **Bash, P A, Pattabiraman, N, Huang, C, Ferrin, T E and Langridge, R** 'Van der Waals surfaces in molecular modeling: implementation with real-time computer graphics' *Science* Vol 222 (1983) pp 1325–1327