Graphic display of nucleic acid structure by a microcomputer

Kazutada Watanabe, Kiyoshi Yasukawa and Koujiro Iso

Department of Chemistry, College of Arts and Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153, Japan

Received 16 August 1983

ABSTRACT

A microcomputer program for the display of a three dimensional structure of a nucleic acid and a protein has been developed. This program can generate a wire model and a ball and stick model on a color CRT display. The phosphate atoms on the backbone, backbone, bases and all atoms can be respectively displayed. Rotation and enlargement of a partial structure is easily accomplished.

INTRODUCTION

One of the most important subjects in molecular biology is the structure of proteins and nucleic acids. Exponential increase in knowledge of this structure is greatly dependent upon the contribution of x-ray crystallography. The structural display of proteins and nucleic acids is essential not only to study and compare with them but also to obtain a direct image of their functions. The topological relationship among atoms in proteins and nucleic acids can be provided by a schematic diagram generated by a computer. Thus, many programs have been developed for the main frame computer (1,2,3,4,5,) and are used as powerful tools in biochemical research. Recent advance in microcomputers allows accommodation of a generous memory space and it is possible to draw high quality diagrams which are partially compatible with a main frame computer. An inexpensive micro-computer has advantages over a main frame computer in the expense of CPU time and file storage, local data processing, and user ability even without knowledge of a central operating system.

The programs described here have been developed for a microcomputer to display the schematic diagrams of nucleic acids and proteins based on data of atomic coordinates.

HARDWARE

The diagram was drawn in either a PC-8801 or a PC-9801 microcomputer (Nippon Electric Co.Ltd.) equipped with a 5-inch dual drive double density floppy disk system (PC-8031) and a printer (PC-8821). The microcomputer in our laboratory is connected with a micromodem and can be used as a remote terminal for a Hitachi M-280H main frame computer.

SOFTWARE

This program is written in a BASIC language and is so prepared that neither knowledge of a microcomputer or a computer language is necessary for drawing the diagrams. Nucleic acids and proteins consisting of less than 2000 atoms, about 100 bases of nucleic acids (t-RNA) and 200 amino acid residues (Fab-L fragment of IgG) can be drawn using a 5-inch mini-floppy disk.

Preparation of atomic coordinates

The data, including atomic coordinates, can be prepared either from a keyboard or a host computer. Since it is difficult to key in all the data on atomic coordinates of a protein, a

Fig. 1 Flow diagrams of collecting the data of atomic coordinates of nucleic acids and proteins.

program (TSSb) has been prepared to transfer the atomic coordinate data of the Brookhaven Protein Data Bank from the host computer to a floppy disk of microcomputer. Using this program, only the data on nucleic acids and proteins which is necessary for drawing the schematic diagrams need to be extracted and stored. Therefore, the atomic coordinates can be collected and processed not only from a keyboard but also from a disk file which is copied from the Protein Data Bank in the host computer and modified by the microcomputer (Fig. 1).

When the atomic coordinates are keyed in from a keyboard, the data of atom name, residue name, residue number, x-, y-, z-axis, linkage number and linkage among atoms must be input.

Schematic diagram

The structure can be displayed showing only the phosphate (α -carbon) atoms of the main chain, the main chain atoms, the base (side chain) atoms or all the atoms in a specified range. The structural model is chosen from between a wire model and a ball and stick model (Fig. 2). After completion of the drawing on the CRT screen, the diagram can be stored in the floppy disk for later rapid display on the screen.

Running this program (MODEL), the name of the nucleic acid

"structural display of t-RNA"

starting residue & last residue No. 1,76

```

command  window      : h-range 30,40          l-range 10,20
command  section     : p-range 0,40
command  rotation    : axis X                angle 30
command  rotation    : axis Y                angle -45
command  scale       : 1 angstrom=0.5 cm
command  display     .

```

1. wire model	2. ball & stick model	<u>2</u>
0. P	1. main chain	2. bases
		3. all atoms
		<u>3</u>
1.XY	2.YZ	3.ZX
		4.YX
		5.ZY
		6.XZ
		<u>1</u>

Fig. 2 A hard copy of the CRT screen indicating the functional choices. The underlined characters are specified from keyboard.

Fig. 3 Drawings of the structural models of t-RNA. a) a wire model b) a ball and stick model. Each model was drawn from X-Y direction.

(protein) is displayed on the screen and the range of residues to be displayed will be asked. From a single residue to an entire molecule can be drawn. The second command level will be input from programmable function keys. The function of each key is shown on the CRT.

After conditioning of the display by these commands, hitting the 5th function key of "display" will activate the third option of model kinds to be displayed. A wire model and a ball and stick model are accommodated in this program. The fourth is the choice of phosphate (P) (α -carbon), backbone, base (side chain) or all atoms. The fifth is the atomic coordinate choice. An "xy" indicates that the x-axis is the horizontal and y-axis is the longitudinal axis. Since six choice (xy,yz,zx,yx,zx,xz) are possible for the coordinate, the schematic diagrams can be drawn from six directions. Moreover, when the atomic coordinates are rotated by 3° around an arbitrary axis, a pair of stereoscopic diagrams of the molecule can be drawn. Selection of the last option will cause the microcomputer to start calculations and drawings. About 15 minutes is required to draw the total 1630 atoms of t-RNA either a wire model or a ball and stick model. Figure 3 shows an example of both for t-RNA. This diagram is drawn from the xy direction of the atomic coordinates given by Sussman et al. (6). However, a rotation of about -43 degrees of the atomic coordinates around the y-axis and on the Z-X plane makes it easy to see the L-shape of t-RNA (Fig. 4). For more detailed analysis of the structure, enlargement of a part of the structure is necessary.

"WINDOW" and "SECTION" commands at the 2nd option makes this possible. For example, the range between 30 and 40 of x-axis and 10 and 20 of y-axis are enlarged by the "WINDOW" command; this range is drawn in Fig. 5. However, this diagram is still complicated. Therefore, the range of z-axis is specified as between 0 and 40. Residues in this range can then be clearly seen in the diagram (Fig. 6). It is also possible to remove backbone atoms or side chain atoms for the drawing (Fig. 7). When the size of the drawing is specified by a "SIZE" command, figures can be compared with each other in the same scale.

The diagram on a screen can be stored on the floppy disk.

Fig. 4 Ball and stick models of t-RNA rotated around y-axis about -43 degrees on the Z-X plane.
a) phosphates. b) all atoms.

Fig. 5 The range of t-RNA between 30 and 40 of x-axis, and 10 and 20 of y-axis displayed by a ball and stick model.

Once stored, it is easy and rapid to put it on the screen. It is recommended that a diagram be stored when an entire molecule is drawn since it takes a long time (15 min) to draw all the atoms of an entire molecule.

The data of atomic coordinates is easily retrievable. It can be printed out by selecting the "COORDINATE" command at the 2nd option level. The atomic coordinate data after rotation can be obtained. Calculation of the distance between two atoms, maximum and minimum values of atomic coordinates and the identities of the atoms for each axis is easily made.

Fig. 6 A section between 0 and 40 of Z-axis in the "WINDOW" range.

Fig. 7 Extracted drawing of bases in the "WINDOW" range.

To study the structure of biomolecules, construction of molecular models is essential. Several kinds of molecular models are available. However, it is not easy to make an entire molecule of nucleic acid and protein without any drawing of the structure. Our program supplies diagrams of the same size with a molecular model for any part of a molecule from every direction. Without these drawings, it is extremely difficult to construct a molecular model even in the presence of atomic coordinate data. It is not only a hard work to draw these schematic diagrams by hand, but is also difficult to judge the most suitable direction for the model construction.

This is the first microcomputer program for a schematic drawing of t-RNA. Many microcomputer programs have been developed for storage and retrieval of nucleic acid sequences and to search for the sequence homology and restriction sites. These are for the **primary** and the **secondary** structure of the nucleic acids. The next step is to know the three dimensional structure of nucleic acids and proteins. Recently, a great deal of attention has been paid to the display of a three dimensional structure of nucleic acids and proteins generated by main frame computer (1,2,3,4,5). However, not all researchers can utilize either main frame computer or an expensive graphic time sharing system. By contrast, it is easy to get an inexpensive and convenient microcomputer. Once the data on atomic coordinates are stored in the floppy disk of a microcomputer, the graphic display of three dimensional structure of biomolecules can be

easily generated in the absence of a main frame computer. This program is equipped to draw both proteins and nucleic acids. Therefore, this kind of microcomputer program will certainly become useful and important in the study of the structure of nucleic acids and proteins.

REFERENCES

1. Sasaki, K., Dockerill, S., Adamiak, D.A., Ticle, I.J., and Blundell, T. (1975) *Nature* 257, 751-757.
2. Johnson, C.K., (1976) ORNL-5138, Oak Ridge National Laboratory, Tennessee.
3. Robinson, A.L. (1978) *Science* 200, 749-752.
4. Langridge, R., Ferrin, T.E., Kuntz, I.D. and Connolly, M.L. (1981) *Science* 211, 661-666.
5. Lesk, A.M., and Hardman, K.D. (1982) *Science* 216, 539-540.
6. Sussman, J.L., Holbrook, S.R., Warrant, R.W., Church, G.M., and Kim, S.-H. (1978) *J. Mol. Biol.* 123, 607-630.