PASS: Simple molecular graphics system for personal computers

G. Chelvanayagam

Department of Computer Science, University of Western Australia, Perth, Australia

An interactive tool for displaying and manipulating molecular structures is presented. The system has a user friendly, menu driven interface and provides good quality graphics for viewing proteins. Full screen stereo viewing and a high degree of flexibility in the investigation of specific sites are among its key attributes. The low cost of the system allows it a diverse range of applicability.

Keywords: personal computer, molecular graphics, proteins

INTRODUCTION

Many computer systems have been reported for viewing and manipulating molecular structures. Some are designed to run on microcomputers or personal computers[1–5] and in general have a range of basic functions. Functions include drawing modes of skeletal models, ball-and-stick models and some form of surface models based on van der Waal's radii. Rotations and translations can also be performed at varying speeds depending on the particular hardware. Overall these systems are useful but tend to be limited to quite small molecules.

Powerful systems have been written for larger computers using dedicated graphics terminals like the ES350 or ES330.[6–8] In addition to the standard basic functions, these systems allow a high degree of flexible site inspection in larger molecules, real-time manipulation, full screen stereo viewing and the ability to interface to molecular modeling, molecular dynamics and energy minimization packages. More recently, the development of new systems has turned to the stand alone architecture of SUN and IRIS-type workstations.[7,8]

The tradeoff between versatility and cost of a system is well recognized. Consequently, the objective in designing PASS was to produce a low cost, user friendly system that incorporated features found in expensive systems. Desirable features include good quality graphics for larger molecules (including proteins), reasonably fast speeds, full screen stereo viewing and flexibility in the investigation of specific sites (siteseeing).

Address reprint requests to Dr. G. Chelvanayagam at the Department of Biocomputing, European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, West Germany.
Received 12 June 1990; accepted 17 July 1990

SYSTEM OVERVIEW

Configuration

To ensure that the monetary cost of the system would be low, PASS was designed for a PC/AT running under DOS 3.2 or its equivalent. A graphics card capable of supporting 2 pages at a resolution of 640 × 350 and 16 colors is required. An EGA card with 256K of graphics RAM is preferred. Output from the graphics card should be to a high resolution color monitor. If hardcopies of on-screen images are required an EP-1201A or compatible dot-matrix printer must be connected via a parallel port. Full screen stereo requires a set of SEGA 3D glasses and a modified second parallel port. It is also convenient to have a hard disk for the storage of data.

Implementation

TURBO C was chosen[9] as the language to implement PASS because of its modularity, portability, dynamic data structures and built-in graphics primitives, as well as its association with the PC family. Modularity allows easy modifications to the code. Extensions can be linked in as they are developed. Dynamic memory allocation is particularly suitable when dealing with molecules because they tend to vary in size. The code was designed to be as portable as possible so that enhanced versions could be ported onto a SUN or similar workstation.

Data format

Data input to the system is required in the standard Brookhaven format.[10] Data files must reside on the same disk or, if a hard disk is installed, in the same directory as the PASS application program for the system to locate them. Additionally, they must be assigned the extension PDB (e.g., 1MBN.PDB for Myoglobin).

Internally, linked lists are used to store atomic, secondary structure, residue and informative data. Templates are used to define the connectivity of atoms. Novel molecules must be defined in terms of a template and appended to the template file.

Interface

The user interface is menu driven and simple to use, both by novices and experienced users.[11] Menus pull down on selection to allow one to choose the desired display options or information windows for the currently loaded data file. This method of interface allows maximum use of the screen for the actual display of the molecular structure. User input to the system is via the keyboard. Arrow keys are helpful to the novice to traverse the menus. One-alpha character input, however, can be faster and more convenient for the more experienced user.

Pop-up windows containing general information (including remark records from the data file) about the molecular structure loaded can be selected from the menu.

Representing structures

PASS provides five different representations of molecular structures: skeletal models, schematic models, ball-and-stick models, van der Waal's models and ribbon models. In each of these models, there is flexibility in view definition, scaling, translating and rotating.

Skeletal models (wire frame models) provide a very basic but useful representation of molecular structures, particularly on a level concerning atomic or residue interactions. Color coding each atom makes it easy to identify residues even when they are not labeled. This model is easy and fast to generate and hence is the best to use while manipulating a protein. The display of the side-chain atoms in proteins is optional. Ball-and-stick models are drawn identically to skeletal models with the exception that circles are drawn at each atomic position.

If correctly marked in the Brookhaven data file, the secondary structure of proteins can be viewed. (see Color Plate 1.) Helices are drawn as tubes, and sheet strands as twisted arrows whose direction points from the N to the C terminus of the strand. This follows from the work of Lesk and Hardman.[12] All turns are highlighted by using a different color from the random coil. Secondary structure types can be viewed individually (for example viewing only the helices in a class-1 protein[13]) or in any combination of type, including no secondary structure.

Atomic van der Waal's surfaces are generated using the dot sphere approach of Bash.[14] Spheres are drawn about each atom with a radius corresponding to the atom's van der Waal's radius. The spheres are generated by plotting points along circular paths on equally spaced, consecutive planes normal to a vector in the direction of one of the atom's covalent bonds. Before plotting, each point is tested to ensure that it does not lie within the van der Waal's radius of any covalently bonded atom. This method was chosen over that of plotting a set of uniformly distributed points over the spherical surface because it provides more depth information in a two-dimensional view, even though it is more expensive to calculate.

Ribbon models are constructed using the elegant method of Carson.[15] (See Color Plate 2.) Cubic β-spline curves are drawn through sets of four grid points using a curve basis to dictate how the guide coordinates affect the curve. The need for four guide points imposes the restriction that at least four residues be viewed while in this model. Ribbon models follow from Richardson's[16] description of a protein that provide the viewer with an excellent feel for both the secondary structure and the folding of a protein.

Highlighting

By default, atom types are color coded for easy identification. Individual residue types, however, can be highlighted by displaying them in one of fifteen different colors. Alternatively, residues can be displayed in a color coded fashion based on their hydropathic character. This form of highlighting can use either the three-state hydropathic character model of Zubay (internal, neutral or external)[17] or the seven-state model of Dickerson and Geis.[18]

Labeling can be done by atom name, residue name or residue number. Labels can be opted in or out.

The user can save particular views as static images. When redrawn, these images may not be manipulated but can be used to build composite pictures drawn in different modes to highlight different regions of importance. (See Color Plate 3.)

Siteseeing

To investigate a site of interest a view option is used. With this option the user simply enters by residue number the residue or regions of interest. This causes the system to create a new temporary data structure containing only the atoms in the selected residues. The new view can then be scaled (zoomed in or out) and manipulated faster than a view of the whole structure. (see Color Plate 4.) The increase in speed depends on the total number of atoms in the selected view. Up to 19 different regions of the structure can be selected to appear in any view. Each region may consist of one or more residues. Each time a view is selected, it is automatically recentered on the screen.

The visibility of any water molecules present is optional: Water molecules can be viewed or hidden through a single selection from the menu.

Large molecules

Turbo C has a reachable address space of 1M, about 360K of which is used to address DOS and the PC BIOS. This imposes a limit of 640K on any Turbo C application (without the use of extended memory). Given that the executable PASS system requires about 240K and that 50K must be reserved for dynamic buffers for local images occurring under the pull-down menus, 350K is available for loading in from data files. Of this 350K, further space must be reserved for individual views. The actual amount of space depends on the size, in residues, of the view. Typically this value would be small and certainly no more than 50K.

With the data structures chosen and the restriction that the system be able to load and manipulate structures of at least 5000 atoms, atomic coordinates are stored as integers

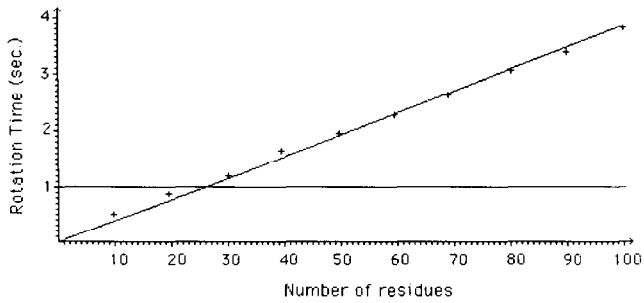

Figure 1. A plot of the time of rotation versus varying numbers of residues in Myoglobin using PASS on an AT clone (16 Mhz cpu clock without a co-processor). A skeletal model, with the side chains included, was used to represent the structure. The figure shows that a structure of about 26 residues can be rotated in less than a second. These results would be slower for a PC but faster for a 386 machine. The degree of rotation has little effect on the results

rather than as floating point (real) numbers. In Turbo C, integers require only 16 bits for storage, while floating point numbers require 32 bits, doubling the storage needed for each x,y and z coordinate for each atom. Storing the coordinates as integers has the added advantage that no mapping function is necessary from internal to screen coordinates. The only drawback with this approach is the small round-off error associated with each scaling and rotation action. Thus, it is best to use views while manipulating rather than the original structure. A simplified, but slower, version of PASS is also available, which maintains a composite transformation matrix to circumvent this restriction.

Response times

In the system, molecular structures can be manipulated in one of three ways: translations, rotations and scaling. Each of these actions requires recalculation of the (x,y and z) coordinates for each atom. Consequently, the time necessary to complete an action on a structure depends upon the number of atoms in the structure. Figure 1 shows the time taken to perform a single rotation with varying numbers of residues. The figure illustrates that the time taken for a rotation has an approximate linear dependence on n, where n is the number of residues. Less than 30 residues can easily be rotated in real time using a skeletal model.

Stereo viewing

Stereo viewing has proven to be a valuable tool in viewing molecular structures. This system can provide full screen stereo viewing if SEGA 3D glasses are connected as described by Chelvanayagam.[19] The principle of tacheostoscopy is used so that left and right eye views are rapidly interchanged on the screen. Side-by-side stereo views are also available both in stationary and rotating modes. Although side by side images are comparatively small, they do not require use of the 3D glasses. Hardcopies of side-by-side stereo can be generated using one of the print options.

Future developments

Extensions planned for the system include interfacing it to molecular modeling, molecular dynamics and energy minimization systems written for personal computers. Immediate modifications will allow the system to accept input through a mouse as well as store individual views in a Brookhaven format.

CONCLUSIONS

A useful, low cost tool has been presented for displaying and manipulating molecular structures on a personal computer. The tool is user friendly and provides good quality graphics while allowing flexibility in the investigation of specific sites on structures. The ability to view structures with full screen stereo greatly enhances the system's performance.

ACKNOWLEDGEMENTS

The author is indebted to John Sadleir (Dept. of Biochemistry, Univ. of W.A.) and Du Huynh (Dept. of Computer Science, Univ. of W.A.) for their help and advice in the design of this system.

REFERENCES

1 Hubbard, R.E. Color molecular graphics on a microcomputer. *J. Mol. Graph.* 1983, **1**, 13–16
2 Clarke, B. A molecular graphics suite of programs for a microcomputer to display molecules from Cambridge crystallographic data files and the alpha-carbon backbone of proteins from the protein data bank crystal files. *Comput. Chem.* 1988, **12**, 65–82
3 Rowlett, R. A low cost system for the graphical display of space-filling models of proteins. *Comp. & Chem.* 1985, **9**, 301–315
4 Watanabe, K. *et al.* Graphic display of nucleic acid structure by a microcomputer. *Nucleic Acids Res.* 1984, **12**, 801–817
5 Cozzini, P. *et al.* Computer graphic tool for color structure display on personal computers. *J. Mol. Graph.* 1985, **3**, 90–92
6 Schomburg, D. and Reichelt, J. BRAGGI: A comprehensive protein modeling program system. *J. Mol. Graph.* 1988, **6**, 161–165
7 Dayringer, H.E. *et al.* Interactive program for visualization and modeling of proteins, nucleic acids and small molecules. *J. Mol. Graph.* 1986, **4**, 82–87
8 Ferrin, T.E. *et al.* The MIDAS display system. *J. Mol. Graph.* 1988, **6**, 13–27
9 *Borland Turbo C 2.0 User's Guide* (1989)
10 Bernstein, F.C. The protein data bank: A computer archive. *J. Mol. Biol.* 1977, **112**, 535–542
11 Chelvanayagam, G. *MDI-TC: A simple menu driven interface for Turbo C applications.* Univ. of Western Australia technical report 90/6
12 Lesk, A.M. and Hardman, K.D. Computer generated

schematic diagrams of protein structures. *Science* 1982, **216**, 539–540
13 Schulz, G.E. and Schirmer, R.H. *Principles of Protein Structure* Springer-Verlag, New York (1979) 87–88
14 Bash, P.A. *et al.* van der Waal's surfaces in molecular modelling: Implementation with real time computer graphics. *Science* 1983, **222**, 1325
15 Carson, M. Ribbon models of macromolecules. *J. Mol. Graph.* 1987, **5**, 103–106
16 Richardson, J.S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 1981, **34**, 167
17 Zubay, G. *Biochemistry* Addison-Wesley, Reading (1983) 12
18 Dickerson, R.E. and Geis, I. *The Structure and Actions of Proteins.* Harper & Row, New York (1969) 16–17
19 Chelvanayagam, G. and McKeaig, L. Stereo viewing on the PC/AT with EGA graphics. *J. Mol. Graphics,* in press
20 Watson, H.C. The stereochemistry of the protein myoglobin. *Prog. Stereochem.* 1969, **4**, 299
21 Marquart, M. *et al.* Trypsin Inhibitor. *Acta Crystallogr. Sect. B* 1983, **39**, 480