Molecular modeling on the Commodore Amiga

John S. Garavelli

National Biomedical Research Foundation, Washington, DC, USA

The Amiga 3000 is nearly ideal for desktop molecular modeling. It uses the Motorola 68030 32-bit processor, running at 16 or 25 MHz, with either the 68881 or 68882 math coprocessor and eight custom processing chips. The video system is NTSC compatible 15.75- or 31.5-KHz RGB analog or digital. The system video interface offers user selectable 320 × 200, 640 × 480, 768 × 480 or 1280 × 400 resolution with a displayable palette of 4 096 colors and interlaced, noninterlaced and overscan modes. NTSC compatibility permits simple and inexpensive video signal synchronization, so that composed monitor images can be displayed and recorded with standard video equipment. The processor, in conjunction with the custom video chips, can manipulate models mathematically at a sufficient rate to present moving or animated images in real time. Users can interact with the molecular image with mouse or joystick controls. A third-party manufacturer, Haitex Resources, has produced an inexpensive stereo image display interface for the Amiga. This stereo image display system consists of liquid crystal shutter goggles attached to a control module on one mouse/joystick port and synchronized at 60 Hz with alternating left/right screen displays.

Several simple molecular graphics programs for the Amiga have been available in the public domain since 1988. The ease of video interfacing and the availability of stereo image display have stimulated several developers to begin assembling molecular modeling packages. At the National Biomedical Research Foundation the Amiga is being tested as an inexpensive stereo-image color graphics workstation for molecular modeling, biopolymer sequence analysis and medical imaging. We hope to use it as a prototyping system for program development and testing, with a large cost savings over graphics display processors. We are seeking to incorporate secondary and tertiary structural considerations into the process of aligning distantly related proteins. This would enable us to make alignments that have a much higher probability of representing accurately the evolutionary relationship of two proteins, the structural relationship of antigenic sites, or the functional relationship of related domains in different proteins. The stereo-image display capabilities of the Amiga allow us to align sequences to known tertiary structures of related sequences. The Amiga's software portability and networking capabilities allow us to use the tested software on other, more expensive computers. Its stereo-image graphics display capabilities provide an essential tool for examining the tertiary structure and functional domains of proteins and for investigating the relationships between genetic expression, protein sequence, protein structure and protein function.

Keywords: desktop molecular modeling, computer graphics, video graphics, stereo graphics

INTRODUCTION

The Amiga 2000 and 3000 series computers are made by Commodore Business Machines and are based on the Motorola 68000 processor. The 2000 series has a 68000 processor; the 3000 series has a 68030 processor. The National Biomedical Research Foundation (NBRF) recently acquired an Amiga 2000 upgraded with a 68030 processor and a 68882 floating-point coprocessor. These processors run at 33 MHz rather than the standard 16 MHz; processors running at 50 MHz are available. The standard memory for a 2000 is 1 Mbyte, to which we have added 4 Mbyte. For a 3000 the standard memory is 2 Mbyte, expandable to 18 Mbyte on the motherboard. One Gbyte of memory is addressable. There are a set of custom chips designed to provide fast memory movement, video graphics and animation, 4-channel sound generation and English text-to-speech generation.

There are a large number of peripheral devices available for the Amiga. Parallel, RS 232 serial and two 9-pin ports are standard. The standard disk drive is the 3.5-inch 880-Kbyte floppy drive. Internal or external hard disk drives are available, and SCSI is supported. The NBRF Amiga has an internal 100-Mbyte 19-msec hard drive with an SCSI controller and connections for additional ports.

The video system is NTSC compatible with 15.75- or 31.5-KHz RGB analog or digital signals; PAL compatibility is also provided. The system video interface offers user selectable 320 × 200 or 640 × 400 resolution with a displayable palette of 4 096 colors and interlaced or noninterlaced modes; VGA is supported. In overscan mode, higher resolution is possible. NTSC compatibility is very important; it permits simple and inexpensive video signal synchronization so that composed monitor images can be

Address reprint requests to Dr. Garavelli at the National Biomedical Research Foundation, 3900 Reservoir Road, NW, Washington, DC 20007, USA.
Received 10 October 1990; accepted 15 October 1990

displayed and recorded with standard video equipment. A third-party manufacturer, Haitex Resources, has produced an inexpensive stereo-image display interface for the Amiga. This stereo-image display system consists of liquid crystal shutter goggles attached to a control module on one mouse/joystick port and synchronized at 60 Hz with alternating left/right screen displays. Stereo images can also be recorded on video and, with another attachment on the VCR, viewed in stereo on regular television monitors.

The Amiga can be configured to emulate other standard microprocessors and network with laboratory computer systems. With an optional Bridgeboard the Amiga becomes a multiprocessor with IBM-PC hardware running MS-DOS or OS/2 software concurrently with the Amiga's own multitasking system. With third-party software, Macintosh proprietary ROM chips and a variable-speed floppy drive, the Amiga can run Macintosh software, although not in multitasking operations. With third-party software, the standard Amiga 3.5-inch floppy drive can read IBM formatted disks; with both software and the variable-speed drive, it can read Macintosh formatted disks.

Networking hardware and software is available to support Ethernet, Novell NetWare, TCP/IP, NFS, Arcnet and X-Windows. The NBRF Amiga is running Ethernet with DECNET support software.

The AmigaDOS operating system has offered true multitasking since it was introduced in 1985. It provides users and programmers with keyboard command, mouse/menu and icon interfaces. The latest release supports a REXX-like interprocess communications protocol. For the monosyllabically inclined, a UNIX operating system was released in Fall 1990. The AmigaDOS operating system is coded in C, and programmers may access the operating system's own processing functions. For example, the handling of queues, or doubly linked lists, is a standard C programming problem easily accomplished with function calls to the operating system's own linked list handling routines. BASIC, FORTRAN, Modula and other compilers are available.

RATIONALE FOR HARDWARE SELECTION

One recent objective of the NBRF has been to acquire molecular graphics and modeling capabilities. We hope that these capabilities will simplify and improve our work in collecting and comparing protein sequences for the Protein Identification Resource.[1] We wanted to supplement our large (and somewhat more expensive) graphics display processor with an inexpensive microprocessor, which would function as a graphics workstation. We decided on the Amiga for the following reasons:

First, the graphics capabilities are outstanding: High resolution graphics are available as standard equipment. They were not an afterthought, added after the computer was introduced, and they do not require one to purchase expensive graphics processor boards or separate monitors. The simple video interface was extremely important. Having done molecular modeling with other graphics display units, it has been my experience that, invariably, after I have worked to produce a beautiful model of a molecular interaction or a dynamic protein-folding simulation, coworkers will say, "That's great! But how do I get a color picture of it for the journal or a motion picture for the presentation?" Here you see the simple and very inexpensive answer. Its stereo interface was a plus. The NBRF already has two Tektronix stereo display units costing $12 000 each, and the most inexpensive graphics display processor available with a stereo interface is more than $30 000. The Amiga's stereo interface cost $100.

Second, the multifunctionality is unsurpassed. The Protein Identification Resource receives protein sequences submitted by researchers in an unbelievable variety of formats. For the last few years we have been receiving submissions on floppy disks but have had no way to process them. The IBM PCs were not networked with the VAX, we did not have a Macintosh, and the VAX floppy drive cannot read either format. The Amiga allows us to read both IBM and Macintosh disks and to transfer them over the network to the VAX. As an added attractive feature, a computer virus borne on an IBM or Macintosh disk we received would not be able to work in the foreign operating system of the Amiga.

Third, the multitasking operating system and the ease of programming made it attractive as a developmental system. The processor, in conjunction with the custom video chips, can manipulate models at a sufficient rate to present moving or animated images in real time. Users can interact with the molecular image by means of mouse or joystick controls.

Finally, and probably most important, it was an inexpensive alternative. The NBRF's Amiga system cost $6 000.

SOFTWARE CONSIDERATIONS

The major unattractive feature was the lack of good molecular modeling programs for the Amiga. Except for local television stations, which use Amigas to produce video graphics for weather reports and commercials, it is not a popular machine. It is the same problem that the Macintosh faced when it was first introduced: People would not buy one because no one was writing commercial programs for it, and no one was writing commercial programs for it because not enough people were buying it.

There are two public domain molecular graphics programs for the Amiga, each demonstrating in its own way the machine's strengths.

One program is CAMM, written in 1987 by Paul T. Miller, then a high school student. It uses a menu interface, which unfortunately is attached to a different screen from the graphics display. The program can display single color stick models, multicolor circles and a computationally inexpensive, but reasonable, representation of space-filling models. The models rotate interactively under joystick control. It has no chemical information for molecular modeling and the input format uses graphics units. This program demonstrates the ease of programming the Amiga and the feasibility of interactive graphics.

The other program is Molecule3D, written by Dwight Blubaugh, a chemistry graduate student, and released as part of the Haitex stereo interface. The monitor has a double image, which can be shuttered to produce a stereo image by attaching the liquid-crystal goggles to the computer or the VCR. This program displays only colored stick models

and the flight path is not under interactive control. It is not an animation sequence, however. The input format used is chemically meaningful but nonstandard.

SOFTWARE DEVELOPMENT OBJECTIVES

I have used an Amiga 2000 for one year; NBRF received its Amiga 3000 in early August 1990. In May 1990 major parts of a VAX FORTRAN program for molecular modeling file format conversion[2] were translated into VAX C. That C code was ported to the Amiga and the user interface was constructed. Most of the graphics display routines were written after the 3000 arrived and the program was demonstrated three weeks later at the Molecular Graphics Society workshop on desktop molecular modeling on August 23–25. The program, BIOMOL, can be started from either the command line keyboard interface or GUI, and it currently operates almost entirely through the mouse/menu interface. BIOMOL is designed to read and write to a number of molecular modeling and sequence database file formats, but only the Brookhaven Protein Data Bank[3] (PDB) format is presently available.

Once a model in PDB format is read, the bonding is determined by distance parameters, and hydrogens are not added. It is intended later that bonding and hydrogens will be added by the usual residue library method.

The display presently consists of colored stick models. The graphics handling of the program is not very sophisticated at this point. The display is not yet double-buffered, which is necessary to use the stereo interface. The model is moved under mouse control; the mouse pointer is essentially being converted into a selection pointer, a slide device or a rotation device. As a rotation device, the direction of the mouse determines the rotation axis, which is in the z-plane and perpendicular to the mouse direction, and the distance traveled determines the amount of rotation. The user should be able to select some portion of the model and "drag" it around. The meter for mouse pointer motion can be turned on and off. We plan to add controls for zoom and z-plane clipping.

At this writing, the only depth-cueing is perspective. The atoms are not yet depth-sorted. With the initial selection of graphics resolution and bit-planes used here, depth-fading is not possible.

The intended use of this program is to fit protein sequences with unknown conformations to proteins with similar sequences and known conformations. Such fitting would be useful in assessing the reliability of sequence alignment algorithms and the conformational significance of protein sequence motifs and features.

ACKNOWLEDGMENTS

This work was supported in part by grant GM45529 from the National Institutes of Health. The Protein Identification Resource is supported by grant LM05206 from the National Library of Medicine of NIH. Mr. David Orloff is acknowledged for his assistance in preparing photographs.

REFERENCES

1 George, D.G., Barker, W.C. and Hunt, L.T. The protein identification resource (PIR). *Nucleic Acid Res.* 1986, **14**, 11–15

2 Garavelli, J.S. The effort to define a standard molecular description file format. In *Chemical Structure Information: Interfaces, Communication, and Standards* (W.A. Warr, Ed.). Amer. Chem. Soc., Washington, DC (1989) 337–44

3 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, **112**, 535–42