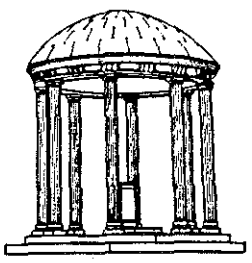
Twelfth Annual Report
Interactive Graphics for Molecular Studies

TR86-007

March 1986

Frederick P. Brooks, Jr.
Michael Pique
Helga Thorvaldsdottir



The University of North Carolina at Chapel Hill
Department of Computer Science
CB#3175, Sitterson Hall
Chapel Hill, NC 27599-3175

This research was sponsored in part by the Division of Research Resources, National Institutes of Health, Grant #RR-02170-03.

IV. Narrative Description

Frederick P. Brooks Jr.
Michael Pique
Helga Thorvaldsdottir

A. Summary of Research Progress

1. Objectives and Operation

We have built, and operate as a service resource, an effective interactive computer resource for seeing, manually manipulating, and computationally modifying mathematical models of complex molecules. We believe that our present resource, called GRIP, has been shown to be as complete and useful as any in existence. One measure of the utility of the GRIP Resource system is that at least seven of our scientific collaborators have obtained their own graphics systems as a direct result of their successful work here.

Our resource has dual objectives. We are a resource providing powerful computer graphics facilities and expert assistance to chemists studying macromolecular structure, both in our laboratory and by building software for export. We are also computer scientists dedicated to advancing the art of interactive computation and interactive three-dimensional graphics. The objectives are complementary. Our chemist collaborators provide the essential focus and a real, complex, and interesting driving problem for our computer science research; our computer science research in turn provides our clients with more powerful tools to improve their insight into very complex structures.

Fundamental to our approach are the following principles:

- The GRIP Resource is designed to help chemists get results from their research, and its success is measured only by theirs.
- Our systems are designed to help the chemists visualize their molecules, their density maps, etc., so that they can use their knowledge to guide computational proceses. That is, they are an aid to, not a surrogate for, human thinking and manipulation. Hence a strong emphasis is placed on human factors research and on human engineering of the system.
- Our system is designed to serve many users, not one or two, so it includes an armory of alternative tools and techniques.

A corollary of these objectives is that we are heavily dependent on observation of and feedback from real users attempting to solve real problems. Our users are almost exclusively working on the structures of molecules of considerable biochemical interest: proteins and nucleic acids. We advance health-oriented biochemical research by enhancing the productivity of individual researchers through better tools.

Facilities used by the project include:

- VAX-11/780
- the UNIX operating system (Berkeley 4.2 bsd version for the VAX)
- 4 MB memory
- 600 MB disk storage

- 1600/6250 bpi dual density tape drive
- Vector General 3303 vector graphics display
- Evans & Sutherland Picture System-330 color vector graphics display on Ethernet
- Two Adage Ikonas RDS-3000 image processing and display systems, each with 1024 by 1024 pixels at 6 bits/pixel (or 512 by 512 at 24 bits/pixel), color map, two internal high speed processors, cross-bar switch for remapping pixel values, video digitizer, and write mask.
- several data tablets, mice, joysticks, stereo devices
- Polhemus position sensor
- Ethernet link to department VAX, which has hard copy plotters, printer, and a dial-up connection to the national
- Masscomp 500 color graphics workstation, with two 68010 processors, floating point boards, Ethernet array processor, and A-D, D-A subsystem.
- Sun 2/50 Ethernet monochrome workstations with file server, Laserwriter,
- Videotape recording, editing, playback facilities.

2. Project Progress and (Computer Science) Results

2.1 GRINCH

GRINCH is a system for the ab initio interpretation of density maps, the only such molecular graphics system of which we are aware. It uses a ridge-line representation of electron density, instead of the conventional contour representation. This uses about two orders of magnitude fewer line segments, which so simplifies the picture that a biochemist can study a whole map at once and hope to see structure in it.

GRINCH Exports

We have found each such port to require customization or alteration of the basic system to fit a different hardware/operating system environment, or especially, different communication between host and display. GRINCH was originally written in C under Unix for a VAX11/780, and the first version used an Adage Ikonas 3000 display system.

In previous work the system was revised to run on the popular Evans & Sutherland PS-300 color vector display. Last year Pique ported the system to a VAX VMS system running the Eunice emulator for Unix, at the University of Chicago. He then successfully completed a port to an IBM 3081-PS 300 configuration under VM/CMS at the University of Connecticut. The user list for our Resource enumerates all the systems that have been installed to date.

This year Dr. Michael Carson of the University of Alabama at Birmingham completed a port of the system to the native VMS operating system on the VAX, a more efficient mode, but a harder port, than the previous one to the Eunice emulator under VMS. Carson is in turn exporting his version to Harvard and to Eli Lilly. So for the first time we saw a spread of the system that we did not have to make happen with our own resources.

Pique completed a port to a Scripps Institute system that uses an E&S Multi-picture System, quite different from the PS-300.

Pique also installed GRINCH on a VAX-Eunice configuration at Prof. Donald Voet's laboratory at the University of Pennsylvania.

Support of Exported GRINCH Systems

To no one's surprise, supporting multiple diverse installations of a software system is a lot more work than supporting only the one in the originating laboratory. This work includes bug fixes, distribution of enhancements, maintaining the master code in such a way that the several versions are all contained within it or generated from it, and answering telephone queries. The amount of such work is a measure of how extensively the remote systems are being used. By that measure we have only a few remote sites that are making extensive use .

Thomas Palmer this year updated and extended Douglas Schiff's GRINCH User Manual to describe the many different configurations and the aspects that vary with the host, display, or operating system.

GRINCH Speedup

Lee Westover reorganized part of the GRINCH image redrawing implementation which shortened the response time by an estimated 25

Douglas Schiff revised the clipping routine so that only parts of the picture that have been changed are re-subjected to clipping before being redrawn.

Neela Srinivasan installed and made operable at considerable effort graphics subroutines supporting the use of Ethernet rather than the E&S. 9600 baud (!) serial connection between the host and the display. This adds another substantial improvement to GRINCH performance, as well as to every other PS-300 program.

New Capabilities

Background Objects. Douglas Schiff designed and Richard Holloway built a capability to add to a GRINCH picture an arbitrary line-graph background object, with special color to make it distinct from the map data and the fitted molecular model. Such background objects might, for example, be contour representations of density, useful when viewing local parts of a map, or pre-derived α-carbon locations, or cofactors crystallized with the molecule.

Split Links. Douglas Schiff designed and Ray Van Dyke implemented a capability to split lines in the ridge-line graph into sub-elements, assigning different interpretations to each part if one chooses. This is useful when low resolution or noise has created a ridge-line that should be fitted with several bonds.

2.2 Molix

Our Molix subproject aims to provide state-of-the-art molecular graphics on workstations, as opposed to traditional minicomputers with vector-graphics displays. The five-year objective is to support, on a system costing less than $50,000, the functions available in 1984 on VAX-PS-300 systems, or equivalent.

Last year we selected the Masscomp 500 as our initial delivery vehicle. Although this machine is clearly insufficient to achieve our final goal, it had the best graphics and floating-point performance of the workstations we evaluated.

PS-300/Masscomp 500 Configuration.

One approach to our goal is to hold to full function, and attempt to realize it on intermediate-cost configurations. This strategy has the advantage that while we are developing such software, system costs continue to decline. Last year we attached our PS-300 to the Masscomp and ported the GRINCH system to that configuration, which costs about $90,000. This combination proves to be quite attractive. It offers the display speed, color, and image complexity of the PS-300, with ample compute power and file capacity for a single project's use. Profs. David and Jane Richardson of Duke University acquired and now operate such a system, using Grinch as well as Duke University acquired and now operate such a system, using GRINCH as well as other tools.

The UNC Department of Biochemistry decided this year to acquire such a configuration for their molecular graphics work. We helped them configure it. **Frodo.** James Boughton and Neela Srinivasan are currently porting Frodo, Alwyn Jones's popular system for detailed fitting of maps, to the PS-300/Masscomp 500 and Unix configuration. We are using the Unix version ported to Fortran 77 by Dr. John Sack of Rice University.

GRINCH on Unaugmented Masscomp. The other approach is to start with the cost goal and see how much function and speed can be realized. Helga Thorvaldsdottir and Sam Black have ported GRINCH to a color Masscomp 500, a machine whose configuration cost is well under our $50,000 target.

Since the Masscomp has a raster display rather than a vector one, both the concept and the details of molecular display have to be revised. A straight vector-raster conversion is so slow that real-time motion of protein and nucleic-acid molecules cannot even be approximated.

First a fairly straightforward port was done and made to work. Since then the work has been aimed at simplifying the displays and speeding up the drawing. The most recent work has been on revising the user interface to make it smoother. The speed is now good enough to make the system usable. We plan to test it on real users shortly.

2.3 Enzyme-Substrate Docking

A major effort this year has been devoted to devising techniques for the display of enzyme-substrate docking. This is important both to biochemists attempting to understand how proteins and nucleic acids work, and the precise relationship of structure to function, and to chemists doing analytic drug design, attempting to find or synthesize drugs with specified active properties and without undesirable side effects.

We are building a system in which the scientist can fit one molecule up against another, perceiving and experiencing the forces predicted by various energy models. Such a system should prove useful both for the analysis and perfection of energy models and for their use in drug design. This requires work both on the displays and on the real-time evaluation of complex energy models.

Force Display – the Argonne Remote Manipulators (ARMS). If one asks how one would like to visualize the docking of two complicated 3-D surfaces against each other, it is hard even to imagine a satisfactory visualization. If one imagines what is wanted, it is a display in which one can move models of the molecules while both seeing them and *feeling* all the forces and torques. Ivan Sutherland first put forth this vision in 1965.

In 1971 we acquired a pair of Argonne Remote Manipulators, which are electrically coupled in 7 degrees of freedom (including pinch) and which allow the operator of the master manipulator to feel at his station the forces exerted on the remote slave manipulator. Joseph Capowski made this device into a computer display system. Jerome Kilpatrick prepared the software to synthesize

forces experienced in a virtual-world model and to display those forces to the operator, together with a visual display of the action of the virtual slave in the virtual world.

The computer power available at that time would not allow the real-time modeling of a very complex virtual world. We settled for a tabletop, seven child's blocks, and the slave tongs. Kilpatrick showed that the augmentation of visual display by force display significantly improved the user's perception of the model world.

Although our objective at that time was molecular modeling, we mothballed the system until available computer power should come up to the task, about two more orders of magnitude of computer speed.

At last the time has come, and we have taken the arms out of mothballs. John Hughes, Michael Pique, Michael Kotliar, and Ming Ouh-Young have been finding and restoring old electronics, and substituting new programs for some of the old electronics. The arms are now interfaced to a Masscomp 500 which has the special subsystem for high-speed, high-precision D-A and A-D conversion. As of the time of this report, the position of all seven degrees of freedom can be sensed. We have not yet attempted to drive the force motors with our new circuitry and software.

Energy Models–Hydrophobic Forces. David Blow, F.R.S., Professor of Biophysics at Imperial College, London, spent three months of his sabbatical with us during this year. He believes the hydrophobic/hydrophilic force component of docking macromolecules in solution may contribute as much as one-third of the total free energy. Most current energy models ignore this component because it is inherently difficult to calculate, and because the form of the model for this component is itself debatable. One of our formal all-evening project-team discussions was spent with the computer scientists listening wide-eyed to four different chemists explaining their thoughts on water, and their results in modeling and measuring its effects.

During his stay here, Prof. Blow worked with Douglas Schiff in reformulating his hydrophobic force program, translating it to Unix Fortran, getting it operating here, and attempting some specialized displays to help visualize the situations modeled.

We are designing all our force-evaluation software so that the precise energy model used is a pluggable component. We will be using the conventional Scheraga-type model, the Blow model, and any others that our collaborators want to try.

Energy Models–Grid Evaluation. Dr N. Pattabiraman and Professor Robert Langridge of UCSF have developed a method of pre-evaluating energy models at points on a grid near the active site of a substrate or enzyme. Then as the cofactor is approached to the active site, the total free energy can be calculated rapidly looking up for each atom the precalculated values of interactions with all the atoms of the other molecule.

This method seems made-to-order for our needs. We had Dr. Pattabiraman visit here and explain more. Ray Van Dyke and Hala Abdalla, guided by Douglas Schiff, are now programming the grid evaluator for our system.

Energy Models–Fast Calculation by Array Processor. The Masscomp 500 has an optional array processor, which must, however, be programmed in a very awkward subroutine language. A software class team consisting of Robert Weir, Allan Chang, Eric Dashman, Candee Ellis and Gang Yang, working together with Michael Pique, prepared a pre-processor that accepts a much more natural arithmetic language. This is being taken over as a standard offering by Masscomp and will be offered to all their users.

Experiments by Michael Pique indicate that the array processor, used to the hilt in native mode, can evaluate energy models as much as 20 times as fast as the VAX11-780. The student

team showed that with their pre-processor and an easy programming language, speed-ups as large as 8-10 can be achieved.

Energy Models–Fast Evaluation by Custom VLSI. Last year we entered into a collaboration with the Department of Informatics at Technische Hogeschool Twente in the Netherlands, whereby they are designing a custom chip to evaluate Scheraga-type models. Two workers from that group have each spent a month in our laboratory learning our problem and VLSI design techniques. Ir. Ben van den Dolder was here in 1984, Drs. Corrie Huijs in 1985.

Drs. Huijs, together with Prof. G.A. Blaauw of THT, have been studying the appropriate algorithms to use in such a custom chip. Drs. Huijs especially studied whether such a chip held the potential of being at least 10x faster than the Masscomp array processor. Her results are positive, so work is proceeding. Actual chip design has been totally stopped by delays in getting Dept. of Commerce export licenses for even the very old Caesar chip design tools.

2.4 Advanced Computer Graphics Technologies

One of our regular on-going efforts is to explore advances in computer graphics technology to see how those can be harnessed to molecular problems. This is one way computer scientists can be especially useful to the community of molecular-structure scientists.

Head-Mounted Display. For some years we have, in collaboration with Prof. Henry Fuchs, been working on a head-mounted display which would allow the user to walk around in a room, seeing virtual objects superimposed on the real scene via half-silvered mirrors. Our project's objective in all this is to create a room-filling molecule. The chemist could walk about *in* the molecule, and manipulate its twistable bonds with a close-up view of local structure, density, charge, etc.

Head-mounted displays have two problems: how to design the necessary light-weight display and optics for a realistic wide-angle view; and how to track the position and orientation of the head in the model space.

Progress this year has been on the tracking. Our Graphics Laboratory acquired a standard small Polhemus magnetic sensor. This device sets up a magnetic field and then reports x,y,z,roll, pitch, yaw for any of several sensors in the field. The field is about one cubic meter.

We installed the Polhemus. Andrew Glassner, working on another project, did initial calibration and experiments with it. Richard Holloway, Michael Pique, John Hughes, and Sam Black have interfaced it to the Masscomp and are working on getting very rapid updates of the readings.

We have been working with the Polhemus manufacturers to get a large-volume system built, which would be half paid for by us and half by another project in the department. Our goal is a volume at least 3 meters on each side.

Using Casio 2-inch TV's, we are making a restricted-volume prototype system with the present Polhemus. We will generate the appropriate perspective views on the PS-300, capture them with a video camera, and display them on the little TV's in a prototype head-mounting. Ming Ouh-Young has built an RF video modulator to interface the camera signals to the Casios.

Tektronix Stereo Shutters. Tektronix has developed a different concept for stereo display, for which we have served as the field test site. We find it to be the most satisfactory stereo device we have seen in fifteen years of trying them all.

A liquid-crystal shutter circularly polarizes the light passing through. The shutter can be switched from clockwise to counter-clockwise polarization in (almost) the retrace time of a standard video raster image. The display is then programmed to alternate left and right eye views, and to trigger the shutter to switch at the proper time.

In contrast to electronically-switched eyeglasses, the Tektronix shutter is made large size, 13 inches, and placed over the display. The user wears ordinary inexpensive polarizing eyeglasses. No wires; no voltages at the temples. Multiple users can view the same display for the cost of extra glasses, under $5 each. One viewer can look at multiple displays, each switching independently.

This year James Lipscomb has coordinated our continued exploitation of this device, using first laboratory prototypes from Tektronix, then factory-made units, which now costs about $3,100, with controller and glasses. We are outfitting most of the displays in the UNC Graphics Laboratory with such screens, using a grant from the NSF. Our project is expecting to receive a prototype 19" screen during the coming year. These will sell for $6,000 with controller and glasses.

Advances this year included interfacing the stereo shutter to the Masscomp, which was straightforward (originally we interfaced it only to the Adage Ikonas). We also interfaced it to our video recorders, so we can now readily make and show stereo videotapes. This substantially helps us demonstrate our work at remote places, for the shutter is readily carried in a suitcase. This was shown at the Drug Information Association meeting in Chicago; at the Molecular Graphics Society meeting in Oxford, England; at SIGGRAPH in San Francisco; at the University of California at San Diego; and at the Research Institute of Scripps Clinic in La Jolla, California. Preliminary tests by James Lipscomb indicate that the stereo shutter can be used with a video projector for larger audiences.

A most challenging task here has been the interfacing of the shutter to the E&S PS-300, by Phil Stancil and John Hughes. Since it is a vector display, not a raster one, the length of time between refreshes is not constant but depends upon the complexity of the picture drawn. It is now working in our laboratory. Our collaborators at Burroughs-Wellcome have acquired a shutter for their PS-300, which is hosted on a VMS system. We have been working with them to install it and modify our stereo software for that environment.

Smooth Motion of Spherical Models–Pixel-Planes. The state of the computer graphics art is that one can easily achieve smooth motion of Kendrew (stick-figure) models of molecules. Our experiments of earlier years have shown arbitrary, user-directed smooth motion to radically help the kinetic depth effect, an extremely powerful monocular depth cue.

One of our objectives for the five-year project is to achieve similar smooth-motion displays of space-filling (CPK) models of molecules. Space-filling models appear to be especially important for docking studies. Earlier work by Michael Pique has achieved smooth motion for intersecting-sphere models of up to a few score of atoms, and near-smooth motion for a few hundered. This was achieved using the very fast bit-sliced microprocessor of the Adage Ikonas, the fastest general-purpose display commercially available. Moreover, Pique used all the tricks. We believe his Fast Spheres program, now commercially distributed by Intermetrics, to define the upper bound of what can be done with today's off-the-shelf technology.

Our next-generation effort towards this objective involves the Pixel-Planes graphic display engine invented by UNC professor Henry Fuchs. This general-purpose polygon display engine consists of custom VLSI frame-buffer chips that incorporate a very rudimentary processor for each pixel. The machine, funded jointly by NSF and DARPA, is currently under construction in UNC's Microelectronics Systems Laboratory, with an objective of being showable at SIGGRAPH '86 in August.

Dr. Richard Potter of our project is working on Pixel-Planes programs to show spherical models of macromolecules.

Algorithms. We held a special evening discussion of project members with all the local applied mathematicians, computational geometers, and algorithm designers in January, 1985. The object was to identify potential new algorithms, and areas where new algorithms would be of especial interest to the project. The resulting list is contained in Appendix B.

2.5 Trailblazer Molecular Graphics Facility

It is our objective to maintain in our laboratory a facility whose hardware capabilities serve as a forecast of what will be available in a few years at a cost suitable for every chemistry laboratory. This enables us to work now at developing long lead-time software, so it will be ready when the machines are affordable.

Project Director Michael Pique was an invited participant at the 1985 Cold Spring Harbor Computer Graphics and Molecular Modelling Conference. Appendix A contains an extended abstract of the talk he gave there. It articulates a view of the future of molecular graphics as seen from a computer scientist's viewpoint.

Ethernet. As reported above, Neela Srinivasan has made Ethernet run between the PS-300 and other hosts. We tried for two years to make the PS-300 parallel interface run under Unix, but without success. Evans and Sutherland, who had donated the parallel interface to our project, graciously swapped it for an Ethernet attachment. Although this is in principle not so fast as the parallel interface, in practice it seems easier to use, it is much more flexible, and it is very fast.

All the machines in the Graphics Laboratory, and indeed all the department's machines, are now Ethernetted together. The use of Ethernet on the PS-300 means in particular that it can be switched from the VAX host to the Masscomp host by software, without any wire plugging and unplugging. Masscomp-PS-300 transfers experience the same speedup as VAX-PS-300 transfers.

Masscomp Upgrade. As part of our continued collaboration with Masscomp, they have upgraded the machine in the Graphics Laboratory to the configuration described in (A. 1), above.

To this we have added the DACP-500 A-D, D-A Data Acquisition Unit for control of the ARMs. We also served as a Beta test site for the Version 3.0 operating system.

Document System. More than half of the work of building programs is writing documents. We installed and began operating a network of Sun equipment for document and program preparation. The initial installation includes a Sun file server, a Laserwriter, and two diskless Sun 2/50 workstations. Over the next quarter we will enlarge the system to provide another disk dedicated to molecular graphics and more workstations so that more of our team use them at once.

A software-lab student team consisting of Erika Baily, Jamie Bird, Cloyd Goodrum, and Andrew Smoak built software to enable the writing tools available for the troff formatter under Unix to be used with the more powerful TEX formatter on the Sun. These tools include a spelling checker, a diction checker, and a style checker. Thomas Palmer of our project worked with the class project output to produce product-quality, checked-out tools which are now installed. His major effort has been rewriting the very nice Occam editor originally built by James Sneeringer, adapting it to the Sun and its window system.

Video and Film Facilities. NSF funds were used to upgrade the Graphics Laboratory's facilities for making visual records of our work. A computer-controlled Dunn camera, capable of high quality images on Polaroid or regular color film, in a variety of sizes has been installed.

The video facilities have been enhanced by provision of a second Sony Umatic Recorder and a Sony editor.

Blox. NSF funds were also used for the purchase of a major software tool, the Blox system made by Rubel Software for the generation of interactive, Macintosh-style interfaces. It supports the Suns, the Masscomps, the Vaxes, and a variety of displays, not including any of our high-performance ones. We expect it to make experimentation with user interfaces an order of magnitude easier, and to lead to some standardization of the interfaces coming out of our laboratory.

2.6 Visualizations

Computer graphics makes possible many more ways of visualizing the abstractions we call molecules and density maps than did the physical models of brass or plastic. One of the thrusts of this five-year project is the devising and borrowing of visualizations, and their testing for usefulness in producing insight.

Density Clouds. Lee Westover has implemented a new way of visualizing electron density, in effect showing it as a cloud of transparency that models the density function. Illumination effects on the cloud as it is rotated before the viewer help perception.

Connolly Surfaces. Michael Pique has translated Michael Connolly's "ms" program for molecular surface calculation from Fortran into C. This enabled a variety of other improvements in speed and flexibility to be undertaken, yielding a 100% speedup. Planned is a "bigcore" version that keeps major workfiles in main memory, thus avoiding many file accesses. Appendix C describes the changes already incorporated.

Surfaces by Space-Sweep Algorithm. New UNC professor and chairman Jurg Nievergelt is a well-known algorithm designer whose most recent work has been space-sweep algorithms for computational geometry. Douglas Schiff is writing his dissertation under Prof. Nievergelt's direction, working first on a convex-hull algorithm for molecule shape description and then on a space-sweep algorithm for solvent-accessible surfaces. We expect this to lead to improvements in "ms" and in its derivatives, Connolly's "ams" and "rams", standard tools of great utility. The dream is an incremental-calculation version of "ms" that would allow the surface to be locally recalculated as fast as the biochemist interactively manipulates the underlying atoms.

2.7 Industrial Collaborations

Our scientific work depends primarily upon collaborations with the biochemists with whom and for whom we build molecular graphics systems. We are fortunate to have also a variety of industrial collaborators who help us to stay on the leading edge of technology.

Burroughs-Wellcome is both an industrial collaborator and employer of a group of scientists who are our scientific collaborators as well. Once again this year B-W made a gift to UNC for the support of one graduate student on the project. This year Douglas Schiff has been our Burroughs-Wellcome Fellow.

Massachusetts Computer Corporation (Masscomp) is our collaborator in the Molix project, furnishing us a fully-configured machine for us to use in porting existing tools and developing new ones. They renewed our agreement this year.

Part of our collaboration is that we have a chance to serve as Beta test sites for new operating systems and products in the graphics area.

Tektronix is our collaborator in investigating stereo techniques. They originally developed the shutter for the purpose of providing few-color screens using monochromatic tubes. We have worked with them to define and demonstrate its uses as a stereo device, and potential markets for it as such.

Evans and Sutherland Computer Corporation originally gave our project the E&S PS-300 for molecular graphics studies. This year they gave us the Ethernet connection apparatus and its software support. This has made a major difference in response time to users and in program loading speed.

3. Objectives for Next Project Year

We shall be pursuing the same projects and objectives as outlined in Section 2 above, where they are discussed in context and in more detail. The priority order is shown below:

3.1 Docking Studies.

Our major push will be to complete the ARM's force-feedback subsystem, including software, and test it with real molecule force models. We have as a special objective to get this running by October, when we will host a SIGGRAPH-sponsored workshop on interactive 3-D computer graphics.

3.2 Molix.

We plan to user-test and field the Masscomp-only version of GRINCH, in a configuration costing less than $50,000.

3.3 Space-Filling Models on Pixel-Planes.

We plan to demonstrate real-time user-directed motion of protein-sized molecules on the Pixel-Planes machine, once again in time for the October workshop.

3.4 Advanced Graphics Technology.

We plan to demonstrate a restricted-volume head-mounted display working with real molecular data.

3.5 GRINCH Field Support.

We plan to support GRINCH users, whether visitors to our laboratory or users of GRINCH systems installed elsewhere, with consultation, bug fixes, feature enhancements, and documentation enhancements.

3.6 Building Move.

We will move our laboratory into the new Computer Science building. This will slow other work.

4. Collaborative Research and (Biochemical) Results

Margaret Eastman

University of North Carolina at Chapel Hill
Bovine prothrombin

Margaret Eastman used the facility frequently to examine energy-minimized conformations of a cyclic hexapeptide from bovine prothrombin. She made stick-figure and shaded-surface CPK views of the model for study and publication. This dissertation research is under the direction of UNC Professor L. G. Pedersen.

Robert Egan

University of Miami
Bacterial chlorophyll A

Robert Egan used the facility to produce views in connection with earlier work on the chlorophyll A protein, which comes from a green photosynthetic bacteria.

Margaret C. Etter **

University of Minnesota
Cyclohexadione-benzene

We prepared computer graphics visualizations of a model structure of an unusual organic cryptand structure with six cyclohexanedione molecule intermolecularly hydrogen bonded into a macrocyclic hexameric ring, termed a cyclahexamer, with a benzene molecule trapped in the center. Etter suggests to us that this might be a useful model for neutral molecule guest binding at receptor sites. We hope to use this as a test case for the force-feedback docking project.

Francis Jurnak

University of California, Riverside
Elongation Factor Tu

Francis Jurnak visited our facilities a number of times in 1985. She used both GRINCH and GRIP-75 to interpret the electron density map of the Elongation Factor - Tu. There are three domains in the protein, but Jurnak worked mainly on the DGP domain, which is the largest and functionally the most important.

Judith Kelly *

University of Connecticut
Beta-lactamase

GRINCH running on the IBM 3804/PS300 configuration at the University of Connecticut was used for the preliminary fitting of beta-lactamase. The resulting backbone tracing was then taken and compared to a penicillin target. Dr. Kelly's collaborators from Belgium did some of the GRINCH work.

Susan Lord

University of North Carolina at Chapel Hill
Fibrinogen

Lord superimposed the amino acid sequence of a peptide sequence known to be important for recognition and cleavage by thrombin, onto the backbone structures of several known beta turns. Only one of these yielded a model in which a certain pair of residues were near each other. She made photographs of the stick-figure and CPK space-filling visualizations to discuss with Herald Sheraga's NMR research group at Cornell.

Patrick Mize

Becton-Dickenson
c1-s

The amino acid sequence is know for c1-s, a component of complement, but the structure has not been solved. The structure of the acitive site is of special interest. The facility was used to study the shape of the active site, in the hope of then being able to predict the structure of the substrates.

Duncan E. McRee *
Jane Richardson
David Richardson

Duke University
Sulfate Reductase

GRINCH running on the Masscomp 500/PS300 configuration at Duke University was used to work on Sulfate Reductase. Work was mainly done on the active site and some of the sequence on it was traced. The sequence from a related species is known.

Dave Richardson used a display program based on Douglas Schiff's DOCK to teach a graduate biochemistry seminar that designed the four-helix synthetic protein "Felix".

Jan Richardson **

Burgess Publishing
Aspirin

We prepared illustrations of the molecular structure of aspirin for a chemistry textbook.

R. Sarma

State University of NY at Stony Brook
Protein S

GRINCH was used to fit a model to an electron density map of Protein - S, a bacterial protein from myxococcus xanthus. The map was calculated to 2.8A resolution.

Paul Sigler *
Richard Schevitz

University of Chicago
TRP repressor

GRINCH running on the VAX 750/PS300 configuration at the University of Chicago was used to begin interpretation of the electron density map.

Craig Smith *

University of Alabama at Birmingham
ax1 sea anemone toxin

GRINCH running on the VAX 750/PS300 configuration at the University of Alabama at Birmingham is being used for the main-chain tracing of 65-residue disulphide-rich sea anemone toxin. Smith has not had much success because the map connectivity is unclear but he continues to use GRINCH.

Hope Taylor

Duke University
Ribonuclease - S'

Hope Taylor came on many short visits to use GRIP to refine her model of semi-synthetic ribonuclease - S'. She hopes to finish that work by the middle of 1986.

Ramalingam Veerappapillai
Robert Egan

University of Miami
Human alpha-lactalbumin

The amino acid sequence of the human alpha-lactalbumin protein has been know for a number of years. Both GRINCH and GRIP-75 were used to build a model of this protein.

Ed Westbrook *

Argonne National Labs
Delta5-3 ketosteroid isomerase

GRINCH running on the VAX 750/PS300 configuration at the University of Chicago was used to fit a model to an electron density map of Delta5-3 ketosteroid isomerase. All 125 residues were traced and built into the map using GRINCH. Work on detailed interpretation is proceeding.

* Work done in investigator's own lab using UNC-developed software.

** Investigator did not visit UNC; we made pictures to order.

4.1 Advisory Council

Our Advisory Council during 1985 has consisted of the following persons, with their affiliations and principal interests:

Prof. David Blow, Imperial College, energy modeling for molecular interaction

Dr. Michael Cory, Burroughs-Wellcome, analytic drug design

Prof. Jan Hermans, UNC Dept. of Biochemistry, molecular dynamics and modeling

Prof. David Richardson, Duke Dept. of Biochemistry, protein structure and the design of new proteins

Prof. Jane Richardson, Duke Dept. of Biochemistry, protein structure and the design of new proteins

Dr. William V. Wright, IBM Corporation, molecular graphics

The group met with project members in all-evening formal discussions on about four occasions during 1985, with subgroups meeting more often.

Prof. Robert Langridge, University of California at San Francisco, Principal Investigator of the other NIH Research Resource for Molecular Graphics, has joined the Council for 1986, but there has not yet been a meeting with him as a member. We are keeping Prof. Blow on the Council in 1986, but we expect him for at most one meeting.

5. Publications

Publications by Users

* Cory, M., McKee, D.D., Kagan, J., Henry, D.W., Miller, A.J. Design Synthesis and DNA Binding Properties of Bifunctional Intercalators - Comparison of Polymethylene and Diphenyl Ether Chains Connecting Phenanthridine, *Journal of the American Chemical Society, 107*, 2528, (1985)

* Jurnak, F. The structure of the DGP domain at EF-Tu in the location of the amino homologous to ras oncogene proteins. *Science, 230*, 32-36, (Oct 1985)

* Kelly, J.A. et al. On the origin of bacterial resistance to penicillin : a comparison of a beta-lactamase and a penicillin target. *Science*, (in press of March 1986)

Richardson, D.C, Richardson, J.S. *Interpretation of Electron Density Maps*, Chapter 14 of Diffraction Methods for Biological Macro-Molecules, vol.115b in Methods in Entymology, 189-206. Editors Wyckoff, Hirs and Timasheff. Academic Press. NY (1985)

*Schevitz, R. et al. TRP Repressor. Nature, 317, 782-786, (Oct 85)

* Taylor, H.C., Komoriya, A., Chaiken, I.M. Crystallographic structure of an active, sequence-engineered ribonuclease. *Proc. Natl. Acad. Sci. USA, Biochemistry, 82*, 6423-6426, (Oct 1985)

Tulinsky, A., Park, C., Rydel, T.R. The structure of prothrombin fragment 1 at 3.5A resolution. *Journal of Biological Chemistry, 260*, 10771-10778, (1985).

Wright, C. The crystal structure of wheatgerm agglutinin isolectin 2 refined at a nominal resolution 1.8A. *Journal of Molecular Biology* (submitted)

Conference Presentations by Builders

* Lipscomb, J.S., Brooks, F.P., Jr., Pique, M.E., and Smith, D. *UNC 1985 Computer Graphics Sampler*. Video tape, color, sound, 4 minutes. Shown at ACM SIGGRAPH'85 Conf., San Francisco, California, (July 1985). Published in *ACM SIGGRAPH Video Review,20*, (8 November 1985).

* Pique, M.E., Lipscomb, J.S., and Andersen, A.C. Trip Through Molecule of Superoxide Dismutase. Slide of one frame from the UNC segment of the movie, *The Magic Egg*. Produced by Garrickfilms and ACM SIGGRAPH. Published in *ACM SIGGRAPH'85 Technical Slide Set*, slide no.60 (July 1985).

* Pique, M.E. and Lipscomb, J.S. *What to Look for in Color Raster Graphics*. Video tape, color, sound, 14 minutes. Shown at Atlantic City conference on Scientific Computing and Automation, (March 1985).

* Pique, M.E. and Lipscomb, J.S. *How the UNC Segment of the SIGGRAPH'84 Omnimax Movie was Made*. Evans and Sutherland Users Group Meeting, ACM SIGGRAPH'84 Conf., Minneapolis, Minnesota, (July 1984).

* Lipscomb, J.S. Comparison of stereo display devices. Fourth International Meeting of the Molecular Graphics Society, Oxford, England (15-17 April 1985). Abstract published in: *Journal of Molecular Graphics, 3*, 122, (September 1985)

* Pique, M.E. Research resource: computer graphics for molecular studies. Fourth International Meeting of the Molecular Graphics Society, Oxford, England (15-17 April 1985). Abstract published in: *Journal of Molecular Graphics, 3*, 122-123, (September 1985)

* Pique, M.E. Drug Information Association, February 1985 Chicago (invited speaker)

* Pique, M.E. American Crystallographic Association, August 1985 Palo Alto (invited speaker)

* Acknowledges facility

Research Highlights - Completed Research

1. First Under-$50,000 Version of GRINCH Developed

Existing molecular graphics systems run on hardware configurations costing $150-$200 K. Our efforts at moving full-function molecular graphics software to cheaper systems were begun by porting our GRINCH system for ab initio interpretation of electron density maps of proteins to a raster-graphics workstation costing well less than $50 K. This should enable most molecular-structures chemists to have such a system in their own offices or laboratories. We expect that in turn to change habits of working in much the same way that cheap word processors has changed the way people construct documents.

2. Superior Stereo Display Device Demonstrated

Our project worked with Tektronix in the product and market definition of their new Stereo Shutter display device. The 19" shutter, which costs about $6000 with controller and glasses, can be used with most raster display devices and relatively cheap polarizing glasses to give a stereo display systems superior to any predecessors.

Molecular structures exploit their three-dimensional freedom in a way that few man-made structures do. Consequently, 3-D perception is one of the most important problems in molecular graphics. Stereo vision alone does not solve it, but it is a powerful component of the solution.

Research Highlights–Research in Progress

1. Force-Feedback Molecule Display Under Construction

We are constructing a force-display system which will allow the molecular scientist to feel and to see the goodness of geometric, electrical, and other fit of two molecules. There is reason to expect force display to be much superior to visual display alone for this application. The perception of allowable and forbidden dockings is crucial to analytic drug design and very important for understanding the action of toxins and carcinogens.

2. Space-Filling Visual Models of Molecules Movable in Real Time

Docking studies require that the visual models of molecules show the spatial extent of each molecule, not merely the stick-figure representation of its bonds. Heretofore it has been difficult to rotate and manipulate such models in computer graphics because of the immense amount of computation required. Work in progress in a collaborating project in our department is leading to a very fast display engine for polygons. Our project has shown how to use it for circles and is working on programs to harness it for the display of spherical models of the large protein and nucleic acid molecules important for life processes.

Appendix A

Technical Trends in Molecular Graphics

Michael E. Pique
University of North Carolina at Chapel Hill
Department of Computer Science
Chapel Hill, North Carolina 27514 USA
(919) 966-5053

Extended abstract of talk presented 12 December 1985
at the Cold Spring Harbor Workshop on Molecular Modelling
and Computer Graphics

Revised 27 December 1985

Introduction

We review trends in molecular graphics and make some tame predictions for the period 1986-1990.

The Human Interface

- Ideas spreading from Xerox PARC and Atari, through the Apple MacIntosh and the Commodore Amiga will reach molecular graphics during 1986: pop-up windows, pull-down menus, more than one thing going on at a time. During the next five years, users and builders will make molecular systems more like video-games, with mice and trackballs, some joysticks that are specialized by function, the working system easier to use and more fun. Joel Birnbaum writes, "The video games people have stumbled onto the right experimental model– domesticated computers and instruments should be self-evident and self-documenting, and they should offer strong positive reinforcement almost immediately. Most users should never need to write a program in the way that we now think of that job. Professional computer scientists and programmers will survive, I hope, but increasingly their job will be to evolve high-performance architectures and interfaces to shield users from the internal complexity, and to develop tools so that users can build and maintain applications without becoming programmers". We must learn what makes programs both fun and effective, why people find "Adventure" games challenging but text editors merely baffling.

- Spreadsheets with their automatic constraint-satisfaction will expand to 3-D geometric uses, simple ones at first, but a "visi-Connert/Hendrickson" refinement program in a few years. This first round of games-style interactive programs will be built by adapting existing batch programs for full-screen use: for example, working with Duncan McRee from Duke University, we at UNC converted a one-analysis-per-run protein structure predictor into a Visicalc-like screen-oriented program, where the prediction changes immediately as the user changes the protein sequence.

- We will see more tolerant systems: with *undo / redo* to reduce the consequences of mistakes (even games give you three tries for 25 cents), and *remember / replay* to ease the tedium. Molecular systems are lagging behind commercial drafting and mechanical design tools here.

- Will we have voice input? Sound or speech output? Possibly... this will require personal workstation spaces with privacy. Acceptable voice input is available now, but speech output is still poor and of unconvincing usefulness. If users can admit that sound effects are both fun and effective we will see them in wide use; the Leeds University team has been using a screeching-siren cue for energy strains for ten years.

- Current graphics hardware gives us our window into otherwise-unseeable worlds, but we do not yet have Sutherland's virtual environment, where we can see, touch, and push the objects floating around us. The kinesthetic display hardware is not ready for production yet but I expect to see lab prototype force

feedback applied to molecular modelling in three or four labs over the next five years, and I expect the Sony "Watchman" market will give us head-mounted television sets good enough to do interactive graphics with within five years.

Modelling and evaluation capabilities

- More users will demand computational completeness: measurement of lengths, angles, areas, volumes, smoothness; boolean union & intersection operations on areas and volumes. I hope we will see more effective visual encoding of results and fewer displays with every color on a peacock proudly strutting around the screen.

- Analysis of the many kinds of physical and chemical information about biochemically-interesting molecules will continue to drive the multiple visual representations, including skeletal, surface, volume, and non-geometric views like diagonal plots. Some labs will try updating these simultaneously in multiple windows.

  Display and publication of moving (internally changing) pictures [e.g., molecular dynamics, parametric plots] will remain a hard problem, but more labs will address it seriously.

- What kind of questions will the emerging expert systems for macromolecular structure be able to handle in five years: "Do any protein sequences have two prolines separated by an arginine?" - Yes. "Do any proteins crystalize with corresponding phenylalanines stacked across a dimer contact?" - ("Jane Richardson in a can") - I doubt it, though if you're willing to wait five or six hours for the inference engine to grind away, possibly. The folding problem will continue to drive this work, with the 1-D pattern-matchers and the 3-D energy-modellers running separate horse races.

The Software Environment and Software Tools

- Networking

Everything will be ethernetted together. Larger labs will have shared, distributed, filesystems for team projects and multiprocessor worksharing, like Sun Microsystem's NFS and Apollo's DOMAIN.

Local network data servers, CD-ROM players with ethernet plugs, will be supplemented by telephone dialup-on-demand, as in the NIH Prophet II scheme.

Making this work will require sharable external machine-independent data representations, e.g. the IEEE standard floating-point formats, the TCP standard for byte-ordering, Sun's remote (inter-machine) procedure calls. Some data transformations will be packaged into self-contained black-box stream modules with one input and one output: their users (whether human or other programs) will be unconcerned about where the processing happens.

Will we see (and be able to identify?) the triumph or the demise of satellite graphics: will we see looser or closer coupling between computer and display? The immediacy of the games-style picture to the driving computer is responsible for much of the effectiveness and fun, but to achieve the highest possible picture-drawing speed, engineers put special-purpose chips and processors between the CPU and the display– even the low-cost Amiga has special fast pixel-filling circuitry. Unfortunately, unless the design has good balance between the processor and bus speeds, performance suffers, like in the Picture System 300 which has a slow 68000 microprocessor between the host computer and the fast line-drawing hardware. The wheel of incarnation rolls on: as these processors "offload" the host, they become host computers that need to be programmed and then, in turn, offloaded. After the painful experiences of the molecular graphics community with the Picture System 300's "9600-baud megabyte blues" (it taking 16 minutes to load the graphics memory from the host) we will see more attention paid to system performance balance during the coming years.

- It will be easier to build, or even *buy*, the non–3-D graphics parts of molecular graphics systems.

Tripos Associates built and use successfully an integrated user interface management system ["UIMS"] in their SYBYL product: it presents menus and prompts to the user, handles and tokenizes user picks and keystrokes, thereby making the application program easier to understand and more portable. Rubel Software's BLOX package is similar, is commercially available, and runs on a wide range of host computers and graphics displays. These promise portability and uniformity of good human interfaces.

Data base management systems are supposed to insulate application programs from changes in the content or representation of the data they use or make. At the IBM United Kingdom Scientific Centre, Stephen Todd and Andrew Morffew used an existing DBMS to hold different kinds of molecular data and produce input information for different applications programs. Their experience was positive, but DBMS software will have to become faster and more reliable to catch on in most biochemistry labs.

Scanner and parser generators will aid builders in the nitty-gritty low-level jobs of breaking command strings into tokens and building tokens into command sequences. For example, the Bell Labs' Lex and Yacc programs that accept high-level specifications of an input language to be accepted and produce compilable programs that process statements in the language, eliminating much error-prone hand-coding.

The next five years will be an era of program portability, as C, Fortran-77, Fortran-8X, and some Pascal variants running under UNIX and UNIX-look-alike operating systems will allow easy migration between computers. The UNC experience porting the Grinch map interpretation system from VAX UNIX to IBM CMS; to VAX VMS; to Masscomp, Sun, Celerity, and Tektronix UNIX has been encouraging: 15,000 source lines of C needing only 5 to 50 lines changed, most of these being changes for the non-UNIX CMS environment.

- It will be easier to build or buy the 3-D graphics parts of molecular graphics systems.

Layered languages will surround the tough 3-D graphics level: a good example is how Connolly's GRANNY program maintains application data, passes text commands to device-dependent (but application-independent) Olson/O'Donnell GRAMPS which maintains and displays the 3-D picture. The next steps on this path are menu-oriented human interface managers feeding the application-specific level, and the generalization of that level to considering the molecular modelling system as a "graphics shell", e.g. Hubbard's HYDRA, mediating data formats between independent application programs.

One technical approach is internal intermediate 3-D pseudo-display lists, suitable for generating both vector and raster images.

Hardware

- The great vector-raster (non?) battle may fizzle out if (as seems to be the case) only the molecular modelling community really *wants* vector displays. There is some chance we will see transitional, hybrid high-end hardware that mixes vector and raster on the same screen. Raster has been seen as a poor-man's vector display (fewer lines, lower quality, slower) but we will see a few molecular graphics systems taking advantage of raster's special advantages. Don't be fooled on the speed issue: do not compare vectors in 1/40 second with pixels in one second; remember: 1 microsecond/inch is 25 nanoseconds/pixel at 512-by-512 and 12 at 1024-by-1024. Beware misinterpreting block-fill pixel rates as arbitrary X,Y line-fill rates.

Resolution and anti-aliasing will continue controversial: 1000-by-1000 or anti-aliased 512-by-512? The 512-by-512 gives crucial compatibility with television: allowing not only direct video output, but the possibility of interactively overlaying and mixing images from video disc with the interactive graphics. The price to master and produce 20 5-minute laserdiscs is $1000 now, additional copies $15. (By the way, there's room for about 10 megabytes a minute of digital information too. The first five minutes of that hour-long disc can hold the whole Brookhaven protein data bank along with the 9000 video frames.)

- Stereoscopic display will be common, and in many labs taken for granted, with the advent of production-quality liquid-crystal shutters from Tektronix and Leeds. Stereo needs double buffering (left eye, right eye) and moving stereo images will need yet-unheard-of quadruple buffering.

- Vector and array processors may turn out to be usable for graphics as well as modelling, but I suspect most of what they can do is already being built into the 3-D coordinate transformation parts of current displays. Using an array processor to make raster images is a difficult technical job, and will seldom be worth the trouble and software lag. Most system builders will be content to use whatever pixel-based operations are built into the display hardware.

Economic and social issues

- Computation centers have traditionally charged for help by billing for CPU cycles. Molecular modellers have traditionally used more CPU cycles per unit-of-help than average users, so have felt overcharged. (The wheel of incarnation here has the lab and personal computers offloading the computation centers.) But you can be sure that at any time, some molecular modellers will have problems too big for personal computers, just like some trips are too far to drive. Travellers collect together in airliners to go long distances, but it is much harder to find a super-computer than driving to the airport and buying a plane ticket. We need to move programs and data between machines of widely varying space and speed capabilities without taking a year to do the transportation. Perhaps UNIX will help enough: Sun's Bill Joy points out that from the IBM PC to the Cray-2, UNIX runs on a ten-thousand–to–one range of CPU speeds.

Continuing problem areas

- Some institutions have a long standing fear of buying software, have you heard "We won't spend more than $100 for any piece of software"? This attitude will pass, I hope, as people get used to buying software for their home computers. Unfortunately some are getting used to stealing it too.

- Software lag and obsolescence will get worse, what's that saying, "any program that runs is obsolete"? This is true especially for machine-specific display code, and with major display hardware upheavals coming every four years the developers are struggling to learn how to use the displays fast enough that anyone's interested when they finally have a polished product. Compilers are now so easy to write that assembler language has died out for CPU's, but there now must be tens of thousands of lines of "assembler" (i.e., low-level, non-portable, hardly-debuggable, unmaintainable) written for the PS 300. One possibility is that for the next five years the new displays will largely emulate the PS 300 instruction set.

Will graphics standards win out? GKS no, PHIGS maybe. The tension between machine-independent code and machine-dependent performance (added to natural vendor proprietary approaches) will be resolved in favor of performance: molecular modelling will continue to absorb whatever the engineers can build. But we can't afford to throw out entire working systems just to make use of new displays, we need to find the equivalent of "coding the tight loops in assembler" for graphics displays.

- Market size thoughts may push developers away from high-scientific payoff, negligible volume systems toward low-powered but high volume ones. The problem here is that hardware catches up sooner than most people expect: in 1975 the UNC GRIP system was being built on a giant mainframe computer, an IBM 360 Model 75, and some crystallographers thought it was unfortunate that the builders were basing the software on a computer that users couldn't conceivably afford, one that had an entire megabyte of memory. The VAX/780's, which held *two* megabytes, appeared in crystallographers' labs by 1978. Who is/should be working on software for the next hardware generation?

Acknowledgements

I am grateful to my colleagues on the UNC GRIP Project, especially James S. Lipscomb and Frederick P. Brooks, for helpful discussions over many years about trends and trendiness in molecular graphics. The GRIP Project work is supported by the Division of Research Resources, U.S. National Institutes of Health.

Interesting Reading

Joel S. Birnbaum. 1985. Toward the domestication of microelectronics. *Computer* **18**:11 (November 1985). Adapted from a chapter in *Cohabiting with Computers*, Joseph F. Traub, Ed., 1985, William Kaufman, Inc., Los Altos, CA.

T.J. O'Donnell and Arthur Olson. 1981. GRAMPS: a graphics language for the MPS. display. *Computer Graphics* **15**:3 (Proc. 1981 ACM SIGGRAPH Conference).

Michael Connolly and Arthur Olson. 1984. GRANNY, a companion to GRAMPS for the real-time manipulation of macromolecular models. *Computers and Chemistry* **9**:1.

William Leler. 1980. Human vision, anti-aliasing, and the cheap 4000-line display. *Computer Graphics* **14**:3 (Proc. 1980 ACM SIGGRAPH Conference).

Vernor Vinge. 1984. *True Names*. Bluejay Books, New York.

Doc 8501310
1/31/85

Algorithms of Special Interest to the GRIP Molecular Graphics Project

F.P. Brooks

Structures of special interest

- 3-D continuous (electron) density function, d(x,y,z).
- Its Fourier transform: intensity spots and phases.
- Near-tree graphs of line segments, especially those consisting of a long main chain and many short side-chains of few kinds, occasionally cross-linked. The main chain is in fact composed of planar segments joined at points.
- Clumps of spheres whose radii are different but bounded by a small range and whose centers have very constrained separations.

Nice properties to exploit

- Plenty of continuity in functions.
- Repeated calculation of slowly-changing scene – lots of frame-to-frame coherence.
- All algorithms need produce only results on points of a grid – no continuous results required.

Algorithm classes newly of interest because of hardware

- Vector or array algorithms in which all elements are processed alike.
- Pipelined algorithms (same, except also allow element-to-element relations within a vector).
- Algorithms for Pixel-Planes machine: f=Ax+By+C calculated simultaneously for all pixels p(x,y).

Geometric Algorithms

- Contouring
- Ridge-line finding
- Fusion of nearly-collinear joined line segments
- Hiding – clumps of spheres
- Intersections among spheres within clumps
- Finding all neighbors of a point, or point set, within radius r
- Bump-detection between clumps of spheres
- Surface dotting – clumps of spheres
- Surface tiling – clumps of spheres
- Shading – clumps of spheres
- Connally surfaces: defined by probing a clump of spheres with a probe sphere of finite radius.

  Corollary: Surface characterization should cover surfaces made up of both convex and concave spherical elements, but all surface elements are of these two types only.
- Structure-finding given all inter-point spacings (NMR), other constraints
- Spline-fitting developable ribbons to linear graphs, to chains of planar segments.
- Fitting constrained 3-D structures (α helices, β sheets, heme groups) to similar line graphs.

Algorithms on physics functions

- Idealization of a geometric configuration towards canonical edge lengths, angles, with weights on constraints
- Constrained/restrained structure refinement
  - Against density
  - Against x-ray intensities
  - With/without thermal parameters as result
  - (Wish to view immediate results)
- Calculation of forces, torques, energies
  - Rigid clumps – precomputed grid methods
  - Limited degrees of freedom – few side chains wiggle
  - Loose assemblies with weighted constraints
- Energy minimization/force-torque response
  - Same cases as above
  - Simplex and other methods
  - Arbitrary force models: note water, hydrophobic force
- 3-D Fourier transform algorithms, both ways
  - Difference maps a special case
  - Calculating gradients of real-space functions directly from Fourier data
  - Interactive phase determination – wish to view intermediate results

Graph Algorithms

- Transversal, matching, bridging of graphs that are nearly trees
- Character-string matching, with unequal-length interpolations, and a C_i-C_j distance table for the character set.
- Finding main-chain segments in ridge-line collections.

Appendix C

Recent changes to Connolly's "MS" molecular surface program

Michael Pique
February 1986

A program that is run infrequently or runs in negligible time is not worth much work to make run faster. A program that does a job whose usefulness is unproven is not worth much work to extend or speed up. But a program like Michael Connolly's "ms" that not only promises to be a basic tool for biochemistry for years to come, but that takes more than an hour to run on typical problems on available VAX computers, is an excellent candidate for improvements.

Our changes to "ms" so far are machine independent, in that they do not depend on special vector hardware. Some add capabilities to "ms", some open new opportunities for future development, some simply make it run faster. The net of the changes is better capabilities with more flexibility and approximately double the execution speed.

These changes include:

Translation from Fortran to C. The original "ms" was in VAX/VMS Fortran, in 1981 we made minor modifications to convert it to standard Fortran-77. We translated it to C in about 10 hours of work to ease development of improvements, and permit enhancements impossible in Fortran-77, particularly dynamic memory allocation. Since this work was done under SUN, Vax, and Masscomp UNIX we could use C's superior development tools, including the "lint" program checker and the "tcov" run analyzer. Now that "ms" is available in C it will be more easily ported to other computers.

Redesign of an internal temporary workfile after an analysis showed 85% of the data was used in only one of the three passes over the workfile. We split the workfile into one part that is read three times and one part (about 6 times larger) that is read once only. In the future "bigcore" version we intend to keep one or both of these workfiles in main memory.

Not storing atoms in the input data that are marked to be disregarded for a particular run. This required keeping a record of which input line number each atom appeared on so that output information about atomic surface areas can be properly reported by input atom number. For files in which all atoms are considered this will of course have no effect, but in cases when surfaces must be calculated only for some fraction of the atoms (those in hydrophobic sidechains, for example) it will increase contiguity of data storage, reduce memory requirements and page faults, and improve cache performance.

Reduced the number of distance comparisons by sorting the atoms along one dimension (Z, for example). This eliminated about 80% of the distance calculations involved in finding the neighbors of each atom, and then the mutual neighbors of each neighbor atom pair. This will pay off more and more as larger molecules are investigated because it changes the order of the algorithm from N-squared in the number of atoms to N times log(N). In the future "bigcore" version we may try keeping neighbor information as an N-squared bit map of neighbors, turning neighborhood searches into bit-vector compressions.

Substituted dynamic memory allocation for many Fortran-style fixed-size arrays. This allows the program's memory requirements to be flexible

with the number of atoms, probe radius, and surface dot density needed for each run. The eventual goal will be to have all internal tables dynamically allocated to eliminate all compiled-in size limits and thus to reduce the number of runs that fail after hours or days of computation simply because a table exceeded its expected size. While this makes the program run slower we feel it is worth while.

Some speedups we have made by trading space for time: the original "ms" was designed to be economical of memory to be able to run on smaller computers. With 1986's larger virtual and physical memories we've been able to speed it up by, for example, pre-computing spheres of surface points for all specified input radii instead of (as the Fortran "ms" did) recomputing the points afresh for each input atom.

We have used two tools to watch "ms" in operation: a subroutine profiler "gprof" and a statement profiler "tcov". We ran the subroutine profiler on a VAX-780 with floating point accelerator. The statement profiler was only available on SUNs so we ran it on a SUN-2 workstation. For all tests, we compute the 1.4-angstrom-probe solvent accessible surface of the superoxide dismutase (SOD) molecule.

The subroutine profiler shows how many times each routine is called and how many seconds each took up. A recent run shows 12 routines that were each using at least 1 percent of the time:

%time	cumsecs	seconds	calls	name	
29.1	1227.75	1227.75	1	main	(main program)
17.7	1972.64	744.89	702028	collid3	(a 3-atom collision check)
12.7	2507.62	534.98	519866	collid2	(a 2-atom collision check)
3.9	3080.14	163.43	2308233	cross	(vector cross product)
3.3	3219.17	139.03	150350	collid1	(a 1-atom collision check)
3.3	3358.05	138.88	dsqrt_r5	(double-precision sqrt routine)	
3.2	3491.60	133.55	465478	cat	(matrix-times-matrix)
3.1	3623.72	132.12	29128	_doprnt	(formatted output library routine)
2.8	3743.04	119.32	578623	cci	(circle-circle intersection)
2.8	3858.99	115.95	1293777	multv	(matrix-times-vector)
2.2	3953.43	94.44	976088	vnorm	(vector normalization)
1.4	4013.92	60.49	sqrt	(square root library routine)	

where "name" is the name of the routine and

"%time" is the percentage of the total running time of the program used by this routine.

"cumsecs" is a running sum of the number of seconds accounted for by this routine and those listed above it.

"seconds" is the number of seconds accounted for by this routine alone.

"calls" is the number of times this routine was invoked, if this routine is profiled, else blank.

The subroutine profiler suggests which routines would be candidates for speeding up by more careful coding, moving into in-line code, or (in some cases) replacement by vector operations. It does not, however, show how time is taken up INSIDE individual subroutines. For that information, we ran the "tcov" statement profiler: the SUN C compiler inserts counting instructions before each statement in the program; after a successful run the counts are written out and a utility program merges the count information back into a source listing. Here are samples of that listing, the statement counts are at the left margin; "#####" means the statement was never executed at all.

```
  68515 ->       /* gather mutual neighbors of iatom and jatom */
4892338 ->       for(knbr=0;knbr<nnbr;knbr++)
4823823 ->         mnbr[knbr] = knbr!=jnbr
4823823 ->            && closerthan(cj, cnbr[knbr], 2*rw+rj+rnbr[knbr]);
4823823 ->
4823823 ->       /* small loop for each mutual neighbor of iatom and jatom */
4892338 ->       for(knbr=0; knbr<nnbr; knbr++) { /* 400 */
4823823, 2425772 ->            if (! mnbr[knbr]) continue;
2398051 ->         katom = inbr[knbr];
1598674 ->         if (katom <= jatom) continue;
 799377 ->         sk = snbr[knbr];
 799377 ->         if (! (si || sj || sk))
 ##### ->              continue;


 799377 -> /* radius of circle of intersection of expanded katom with plane */
 799377 ->          rck = sqr(rk+rw) - sqr(pbk);
 289269 ->          if (rck <= 0.0) continue;
 510108 ->          rck = sqrt(rck);
2040432, 1530324 ->            do3(k) cck[k] = ck[k] - pbk * uij[k];
1530324 ->
1530324 ->          /*  call circle-circle intersection subroutine,
1530324 ->              skip to bottom of inner loop if no intersection */
 510108,  159094 -> if(! cci(cck,base,rck,height,uij,base2,alt2[0])) continue;
 159094 ->          /* probe placement at ends of altitude vectors */
1404056, 1053042 -> do3(k) alt2[1][k] = - alt2[0][k];
1053042 ->         for(iw=0;iw<2;iw++) {
2808112, 2106084 ->      do3(k) cw[iw][k] = base2[k] + alt2[iw][k];
2106084 ->             /* collision check with mutual neighbors */
 702028 ->             pair[iw] = ! collid(cw[iw],rw,cnbr,rnbr,mnbr,nnbr,
 702028 ->             3,jnbr,knbr);
 702028 ->             }
 702028 ->        /* if neither probe position is allowed, skip inner loop */
 351014,  348663 -> if ( (! pair[0]) && (!pair[1])) continue;
  2351 ->         both = pair[0] && pair[1];


  2162 ->    /* contact probe placement loop */
  2162 ->    ncon = (4 * 3.14159 * sqr(ri)) * den;
153591 ->     for(i=0;i<ncon;i++)  /*650 */ {
605716,  454287 -> do3(k) cw[0][k] = ci[k] + (ri + rw) * con[i][k];
454287 ->       /* check for collision with neighboring atoms */
151429 -> if (collid(cw[0],rw,cnbr,rnbr,mnbr,nnbr,1, jnbr,knbr))
146813 ->   continue;
 4616 ->   else fprintf(contact, ...
```

"tcov" originally reported that the most often executed lines were executed 40 million times --- this on an input file of 2162 atoms that produced 14320 surface points as output.

Although "ms" spends its time doing geometric operations it is not immediately vectorizable. Much of the work already completed should make it easier to vectorize, but nothing concrete has been done yet. We have two target machines for vectorization: the Convex C-1 mini-super-computer and the Masscomp 5000 super-micro-computer with integral single-card AP-501 array processor. UNC has a Masscomp/AP in-house and has access to two Convex C-1's at the nearby Microelectronics Center of North Carolina; Scripps Clinic has an in-house Convex C-1. The Masscomp AP-501 does about five million single-precision floating point operations per second; the Convex about twenty million. (By comparison, a Sun-3 with 68881 floating point chip does 100 to 200 thousand and with Weitek 1164/5 attached processor 600 thousand to 1.2 million per second). The Masscomp AP is programmed using an extended dialect of C known as "CAP" (C for Array Processors) written at UNC as a student project in spring 1985. The CAP program accepts a program written with APL-like vector and array operations and produces a standard C program with these operations replaced by calls to array-processor subroutines. The Convex C-1 is programmed either using a automatically vectorizing compiler (Fortran available now, C promised shortly) or in-line assembler statements to trigger hardware vector operations.

How much of this code is vectorizable, and with what limitations, remains to be seen.

Although the Masscomp AP can do pairwise distance calculations --- the heart of "ms" --- at ten times the speed of a VAX-780/FPA, it has only a small (64K byte) working storage so must continually transfer point lists from the main CPU memory, or worse, from disk backing store. Vectorization and virtual memory are natural adversaries so the Convex's 128 megabytes of physical memory may be the key to a really fast molecular surface calculations.

Research that we expect to lead to improvements in "ms" and its derivatives, Connolly's "ams" and "rams", is being done at UNC by doctoral student Doug Schiff, under the direction of noted algorithm designer and author Professor Jurg Nievergelt. Doug is presently working on a convex hull algorithm for molecular shape description, and expects to soon move on to the efficient calculation of solvent accessible surfaces using a space sweep. A goal is an incremental-calculation equivalent of "ms" that would allow the surface to be locally recalculated as fast as the biochemist interactively manipulated the underlying atoms.