Eighth Annual Report, February 1982
Interactive Graphics for
Molecular Graphics System

TR88-038

July 1988

Lynn F. TenEyck

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

Eighth Annual Report
NIH Grant RR 00898-08

INTERACTIVE GRAPHICS FOR MOLECULAR GRAPHICS SYSTEM

Lynn F. TenEyck

Department of Computer Science
University of North Carolina
Chapel Hill, North Carolina

February 1982

1. Summary of Research Progress

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-75, has been shown to be as complete and useful as any in existence. One impressive measure of the power and utility of GRIP-75 is that at least seven of our clients have obtained their own graphics systems as a direct result of their successful work here.

Our resource has dual objectives. We are a service center providing powerful computer graphics facilities and expert computer assistance to chemists studying macromolecular structure. We are also computer scientists dedicated to advancing the art of interactive computation and interactive, three-dimensional graphics. The objectives are complementary. Our chemist clients 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 the accuracy of their results and, most importantly, improve their insight into very complex structures. Our overall objectives are:

1. To help our client biochemists obtain the most accurate structural information their data will support, with the most efficient use of their time, and
2. To advance the art of interactive graphics to the point that our clients are able easily to perceive the significant *chemical* information in their structures, at all levels of detail.

We believe that the use of our system substantially improves a chemist's understanding of the molecules with which he is working. Our most valuable product is insight.

Five years of experience with GRIP-75 have shown us a number of inadequacies. We are now building a second-generation version of this system, called GRIP-X, designed to be more comprehensive in the biochemical problems that can be attacked, more powerful in the mathematical tools available, more varied in the visualizations available for molecules and maps, smoother in the user interface, and constructed as product-quality, documented, exportable software.

Fundamental to our approach are the following objectives:

- The GRIP system is designed to help chemists get results from their research, and its success is measured only by theirs.
- GRIP is designed to help the chemist visualize his molecules, his density maps, etc., so that he can use his knowledge to guide computational processes. That is, it is 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.
- GRIP is designed to serve many users, not one or two, so it must include an armory of alternative tools and techniques.
- GRIP is designed to interface smoothly with any batch computations its users must do, and to incorporate on-line facilities for all computations that can reasonably be done "while you wait."
- We as computer scientists are interested in GRIP as a test-bed for research in man-machine systems design, in man-machine interaction, and in the design of distributed computing systems.

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. GRIP-75 was the trailblazing system in manipulating a molecular model to fit complex experimental data. Our clients have published over thirty papers in the biochemical literature containing results derived from the GRIP system. Features pioneered in GRIP are finding their way into other molecular graphics systems. The presence of the facility at UNC has been instrumental in maintaining a close collaboration with biochemists at Duke, who come over regularly to use our system. At least one of our new faculty members, Dr. R. Snodgrass, based his decision to come to UNC on the GRIP system, which demonstrated to him our ability to build large innovative systems.

Our facility consists of a VAX 11/780 computer with a Vector General VG3303 vector display unit and a full complement of interactive input devices; and a PDP-11/45 computer with a Vector General VG3 vector display unit and a high-speed connection to the University of North Carolina Computation Center 360/75. Our present production system runs on the 360/75 and the PDP-11/45. The VAX 11/780 system is for the development of our next system.

The Computer Science Department obtained money from NSF for another VAX for departmental use. By pooling resources we have come up with a situation in which everybody wins. In exchange for some of the currently surplus capacity of our VAX we have obtained, through the Computer Science Department and the Microelectronics Center of North Carolina, the use of 1.25 MB of memory (in addition to the 750 KB we bought) and a magnetic tape drive. We share the GRIP VAX with other graphics users in the department; in return we get the use of the Ikonas raster graphics system and the full resources of the department VAX — for example, the Versatec printer/plotter and a Hewlett-Packard four-pen plotter.

Our effective configuration on the VAX system is

- VAX 11/780 with UNIBUS but no MASSBUS;
- the UNIX† operating system (Fourth Berkeley version for the VAX)
- 2 MB memory
- 600 MB disk storage
- 800/1600 bpi dual density tape drive
- Vector General 3303 vector graphics display
- Ikonas RDS-3000 image processing and display system, 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
- Summagraphics data tablet
- 16-channel (expandable to 32) analog to digital converter
- high speed parallel link to department VAX, which has hard copy plotters, printer, and a dial-up connection to the national "USENET" UNIX network*

†UNIX is a Trademark of Bell Laboratories.

*The usefulness of the "USENET" network of UNIX systems can hardly be overstated. News and mail can be exchanged with over 150 computer systems across the country, including systems at Columbia University, the University of California at San Francisco, and the University of California at La Jolla. The systems mentioned specifically are all doing work in interactive computer graphics, and I believe that most of them receive research support from the

Another large gain from pooling resources with the department VAX has been reduction of maintenance costs. We can use each machine to verify faults in the other machine. Rapid identification of defective circuit boards translates directly to shorter and cheaper service calls.

Division of Research Resources.

1.1. Summary of GRIP-75 Usage by year, 1975-1981

Table I summarizes the use of the GRIP-75 system for all purposes since we began demonstrations and productive operation on July 15, 1975. We have not tried to estimate the system time spent on development before the beginning of productive operation but know it to be many hundreds of hours. Because we changed from manual to machine logging of GRIP sessions in mid-1976, we believe the true buildup of system usage to be substantially greater than suggested by these data. We have observed that users tend to overestimate the time they spend using GRIP.

Table I. GRIP-75 Usage by years

	1975	1976	1977	1978	1979	1980	1981	Totals
Production	329	581	781	1034	1108	556	349	4738
Demonstrations	12	50	161	137	43	57	76	536
Development	297	186	198	213	345	109	51	1399
Totals	638	817	1140	1384	1496	722	476	6673

Table II gives the use of the GRIP-75 system by year for each team of biochemists. These teams are identified by their principal investigators.

Table II. GRIP Production Time (Hours)

User	1975	1976	1977	1978	1979	1980	1981	Total
Hermans	7	29	42	13	11	-	-	102
Kim	200	321	105	24	49	42	-	741
Richardson	83	79	91	188	384	297	169	1291
Lipscomb	12	-	-	102	-	-	-	114
Carter	27	21	-	4	5	16	-	73
Jensen		46	62	-	-	-	-	108
Tsernoglou		85	11	-	-	-	-	96
James			19	-	-	-	-	19
Low			55	191	65	-	153	464
Davies			85	92	-	-	-	177
Schiffer			35	28	26	28	-	89
Amma			163	35	-	-	-	198
Wright			41	-	96	-	-	137
Hendrickson			72	80	34	-	-	186
Schevitz				74	-	-	-	74
Love				94	-	-	-	94
Kartha				109	-	-	-	109
Taylor					89	45	17	134
Rich					78	-	-	78
Sarma					188	-	-	188
Olson					20	-	-	20
Bugg					39	-	-	39
Premilat					24	-	-	24
Amzel						98	-	98
Hardman						27	-	27
Eggleston						3	3	6
Roth							7	7
Total Hours	329	581	781	1034	1108	556	349	4738

Table III is a list of the biochemists whose research teams have used the GRIP system and their institutions. These are listed in the order of their first use of GRIP. For groups sending more than one biochemist to use our system, the principal investigator is given first and the names of his colleagues follow indented. The institutions of these colleagues are given only where they differ from the principal investigator's. In five cases — Sarma, Rich, Love, Sigler, and Amzel — the principal investigator has made little or no direct use of GRIP himself.

Table III. GRIP Users (1975-1981)

J. Hermans	University of North Carolina
D. R. Ferro	Istituto di Chimica delle Macromolecole
J. E. McQueen	
I. D. Kuntz	University of California School of Pharmacy, San Francisco
M. Vacatello	Institute of Chemistry, University of Naples, Italy
S. H. Kim	Duke University
J. L. Sussman	
R. W. Warrent	
S. R. Holbrook	
G. M. Church	
W. Shin	
C. J. Alden	
D. C. Richardson	Duke University
J. S. Richardson	
E. D. Getzoff	
J. A. Tainer	
D. McRee	
W. N. Lipscomb	Harvard University
J. L. Crawford	
C. W. Carter	University of North Carolina
R. A. Jones	
L. H. Jensen	University of Washington
K. Watenpaugh	
R. E. Stenkamp	Yale University
D. Tsernoglou	Wayne State University
G. A. Petsko	
M. James	University of Alberta
L. T. J. Delbaere	
G. Brayer	
B. W. Low	Columbia University
A. Sato	
M. Kimball	
S. Ginnell	

D. R. Davies M. C. Liu E. A. Padlan	National Institutes of Health
M. Schiffer	Argonne National Laboratory
E. L. Amma R. L. Girling R. C. Paslay	University of South Carolina
H. T. Wright	Princeton University
W. Hendrickson	Naval Research Laboratory
P. B. Sigler A. Podjarny R. W. Schevitz	University of Chicago
W. E. Love W. E. Royer	Johns Hopkins University
G. Kartha	Roswell Memorial Institute
H. C. Taylor	Berkeley Springs Research Consortium
A. Rich N. Woo	Massachusetts Institute of Technology
R. Sarma A. Laudin	SUNY, Stony Brook
A. Olson	National Resource for Computational Chemistry
C. Bugg R. Almassy J. Fontecilla	University of Alabama
S. Premilat	Université de Nancy, France
M. Amzel	Johns Hopkins University
K. Hardman	IBM Yorktown Heights Research Center
D. Eggleston	University of North Carolina
B. Roth L. Kuyper	Burroughs-Wellcome
M. Connolly	Yale University

1.2. Changes in Resource Direction and Problems

It is obvious from the tables that the usage of GRIP-75 is rapidly declining. That is not unexpected. Computer graphics systems which do the same job as GRIP-75 are becoming fairly common. We believe that none of the other systems are quite as powerful as GRIP-75, but they are not that far behind. Our past experience demonstrates that chemists *will* travel long distances to work on a system that offers them unique facilities. We are developing such facilities for GRIP-X.

In the past we have devoted much of our effort to client service and to research in how our clients made effective use of our system. (This is shown dramatically by the figures in Table I comparing production use and development use of GRIP-75 during the period 1977-1979.) We are now devoting nearly all of our efforts towards exploiting the things we have learned about human factors in the design and use of graphics systems, and to exploring advanced computer graphics techniques to be incorporated into our new system. These efforts are described much more fully in the next section. Once we have developed our new system we will again enter a client service phase to study the utility and power of our new methods. We expect that the unique combination we will offer of powerful graphics and good fitting and refinement aids will attract many clients.

1.3. Core Research and Development Summaries

This section describes the various projects in graphics and crystallography which we have worked on during the past year. Some of the projects involve people not connected with the facility, working under the direction of members of the facility. This is one of the great advantages of placing a facility of this type in a computer science department. The software engineering class, COMP 145, needs a steady stream of projects for the student teams of programmers. These teams work for a client, producing software to his specifications. We have served as "clients" for a number of teams, and have obtained useful prototype software as a result. (COMP 145 teams rarely produce production quality software, but are very useful for producing prototypes and test versions.) Two of the projects — advanced graphics and semiautomatic interpretation of electron density maps — are reported in substantial detail because of their importance to our facility.

1.3.1. Advanced Graphics for Molecular Modelling

During the past year we have devoted a large part of our effort to graphics research in order to develop useful tools for GRIP-X. It has been obvious to us for some time that an advanced system is going to have to deal in more than just atomic coordinates and connectivity; other molecular properties will also have to be treated. All of these properties are different abstractions of the molecule and often require different representations. The following paragraphs describe some of the areas in which we have been working.

1.3.1.1. High Performance Raster Graphics

There are aspects of molecular graphics for which raster graphics are much better suited than line-drawing graphics. Examples are space filling representations of molecules and full color displays. The Ikonas color graphics image processing system attached to our VAX has a frame buffer that can be used either for a picture 512 by 512 pixels with 24 bits per pixel, or for pictures 1024 by 1024 pixels with six bits per pixel; a color map, which can map any pixel value into 30 bits of color information; a cross-bar switch, which can arbitrarily permute the bits in each pixel before they go through the color map; an internal 32

bit processor with a 200 nsec cycle time; a multiply-accumulator module, which can multiply a pair of 16-bit numbers and accumulate a 35-bit sum in 200 nsec; a video digitizer for storing video signals into the frame buffer; and a high resolution monitor. We are continuing experiments begun last year to determine how a device of this nature might best be used for interactive molecular graphics.

One of the best possible depth cues, for those who can see it, is stereo. The Ikonas monitor runs at 60 Hz non-interlaced. By using the cross-bar switch to exchange images every frame and having the user view the display through an appropriate shuttering device we expect to be able to show very stable stereo images with full color and 512 line resolution, or eight colors and 1024 line resolution. (We have requisitioned the special glasses required for this.) In a system of this nature the limit on the complexity of the picture is determined by the fineness of the raster, rather than by the number of vectors to be drawn. Any picture that can be stored in the frame buffer can be viewed in full stereo by this method. This raises the very interesting question of a suitable representation for an electron density map on a stereo raster graphics system.

The other major depth cue commonly used in interactive three-dimensional graphics is the kinetic depth effect, by which a moving image on the screen appears to be moving in space. This is a very good illusion, particularly when the motion is under the direct and immediate control of the user. Raster graphics systems usually cannot present the kinetic depth effect because of the very large number of pixels which must be rewritten to produce a new picture. The Ikonas can rewrite a picture very quickly using the internal bit-sliced microprocessor. A team of students in COMP 145, the Software Engineering class taught here by Professor Brooks, tackled the problem of generating moving molecular pictures on the Ikonas. The team included Mr. Larry Lifshitz of the GRIP project, and was supervised by Mr. Michael Pique, also of the GRIP project. The project was quite successful for structures up to about 50 atoms. The illusion of a model of solid spheres rotating in space is remarkable. Two papers on the project have been submitted to SIGGRAPH 82, a major computer graphics conference.

1.3.1.2. Representations of Molecular Properties

The conventional stick figure (Kendrew) representation of a molecule is a good representation of two properties — the positions in space of the atomic centers, and the bonding relationships between atoms. It is not a good representation of the spatial extent of the molecule, and carries no information concerning such properties as chemical reactivity or electrical charge. Dr. Michael Connolly has developed a method of representing the solvent-accessible surface of a molecule by placing dots on the surface. This permits study of molecular surfaces. We obtained Dr. Connolly's programs, and arranged for him to visit our laboratory for a week to work with us and to discuss the problem. We have been studying these pictures for several months now. Some of our previous results concerning three-dimensional illusions apply to these pictures; smooth motion and intensity depth cueing are good methods for enhancing the spatial perception of the images. We also believe that we have determined some of the important visual properties of the surfaces for perception as surfaces rather than as collections of dots. One of the important properties, we believe, is that the dots are placed on a surface which has continuous first derivatives. This aids interpolation of the surface in the human perceptual system. Another important variable is the density of the dots over the molecular surface. We expect these results to be important as we develop faster ways of generating these images, and develop ways to manipulate the surfaces interactively.

An obvious technique for encoding chemical and physical information about molecular properties is color. We have been investigating this intensively. At present we are using the Ikonas raster graphics system with special purpose microcode routines to get color pictures which can be rotated interactively. This is not completely satisfactory, as the 512 by 512 resolution is visible to the user. It does, however, let us experiment with the medium and determine good ways of using color. The work by Tom Williams described later in this report depends heavily on color coding of information. Our conclusion is that in an interactive system a large number of colors is desirable, but for simple color coding applications perhaps as few as seven will be acceptable. A line drawing display should have just as good resolution in color as in monochrome if smooth motion and stable images are to be achieved. We are presently negotiating for an evaluation unit of a new color graphics display tube for our Vector General 3300 display unit. We will be able to use this unit in color with no hardware changes to get a range of seven colors.

1.3.1.3. Automatic Generation of Abstractions of a Structure

Much of the structural information about a molecule is best communicated in terms of high level abstractions. Features such as β-sheets and α-helices are best shown as ribbons and cylinders, or similar simplifications. Outstanding examples of this level of abstraction are the elegant drawings prepared by J. Richardson of Duke University in her *Protein Structure Coloring Book*. M. Pique has been experimenting with methods to automatically generate such drawings given only a coordinate set. A procedure he has developed for running a ribbon along a molecular backbone (with shading and highlighting) shows very clearly how any molecule is folded up.

1.3.1.4. Ridge Line Representations of Electron Density Data

The most common method for representing the electron density function on a graphics system is to contour the function in three dimensions at a constant level. The resulting level curves form a sort of wire cage enclosing the molecule. This representation has much to recommend it. It clearly shows the volume occupied by the molecule. It also gives a good idea of the shape of the molecule in any relatively small volume of space. This representation contains a great many lines and rapidly becomes confusing as one tries to view larger and larger volumes. Also, information concerning the positions of the maxima of the electron density function within the cage is not shown.

As part of his work on algorithmic fitting of molecular models to electron density maps Tom Williams has been working with ridge line representations of the electron density map. Ridge lines run through the tubes that would be defined by the cage of the contoured representation connecting the maxima. The ridge lines are much closer to the stick figure representation of the molecule. This representation is very useful for looking at large volumes and for tracing molecular boundaries.

1.3.2. Semiautomatic Interpretation of Electron Density Maps

1.3.2.1. Introduction

Tom Williams has been developing a method for computer aided initial fitting of electron density maps. By *initial fitting* we mean estimating from an electron density map the positions of the mainchain atoms and perhaps the carbonyl and C-beta atoms. This method produces atomic coordinates for all of the atoms in each of the residues that have been located using the amino acid

sequence if it is available. The major part of his effort has been concentrated on designing a method that attempts to make best use of the special skills of both the machine and the human user. A fitting system is being developed to show the practicality of the method. Further work would be needed to produce a production quality system, or, as seems more useful, to incorporate his method into GRIP-X.

1.3.2.2. Representation of the Electron Density Map

The most commonly used representation for electron density is contours drawn in sets of parallel planes. The contours show surfaces in the electron density function either at one or a small number of density values. The difficulty with using contours for initial fitting is that many lines are needed to accurately show the surface and a global overview is often necessary to resolve ambiguities. There is a limit to the number of lines that can be comprehended and interpreted at once.

Ridge line representations of the electron density map were first suggested by Dr. C. K. Johnson of Oak Ridge National Laboratory. A ridge line representation shows the essential information about the location of high density and how it is connected, using many fewer lines than does a contour representation. Ridge lines, named by analogy to geographic terrain, form a network or graph connecting peaks(local maxima) through passes. If there was one and only one peak for each atom in the molecule then producing the ridge lines would completely solve the problem - each graph vertex would correspond to an atom and each edge to a bond. For molecules the size of proteins this never occurs because of insufficient map resolution, phase errors, noise in the data collection process, atom mobility in the crystal, etc. The problem has been reduced to determining the correspondence between the ridge lines and the model. Some edges of the ridge line graph will lie along bonds, some edges are spurious and do not correspond to any part of the model, and some edges are missing, leaving bonds unaccounted for. The task then is to match the two graphs, the molecular model and the ridge lines. The ridge lines are approximated by an algorithm similar to that of Swanson [SWANS79].

1.3.2.3. The Division of Labor

Humans are very good at pattern recognition. They can bring together diverse sources of specialized information to determine whether some feature matches a pattern. They can balance a feature's match to a local pattern with the feature's global context to make an overall decision. These are things that computers do with much difficultly. Attempts to interpret a map by a computer alone have not been very successful. Computers however are capable of accurate storage and recall of large amounts of information , pattern recognition for local, well-defined patterns, and numerical calculations.

The computer's task then is to remember the current state of interpretation and display it at all times. The ridge line graph and model are shown on a graphics screen with the edges color coded by type. The different types of edges are molecular model, uninterpreted map and map interpreted as each of the features mainchain, sidechain, carbonyl, and bridge. The bridge type designates disulfide bonds and hydrogen bonds. This encoding helps the user to see at a glance the portion of the mainchain traced so far, which edges attached to the mainchain have not been accounted for yet, and many other useful pieces of information about the current interpretation. Having this information always clearly visible should make the task of interpretation easier. The current display implementation does color line drawing on an IKONAS raster graphics

system.

The user of the system is given the task of identifying features in the ridge lines and adding edges to match the ridge lines with the molecular model. The computer, of course, needs to be told about this feature. Having this communication simple and fast makes it easier for the user to experiment with various interpretations rather than being locked into one interpretation because it is so difficult to try others or spending inordinate amounts of time trying the envision what the consequences would be.

The communication takes place through relatively simple local patterns. These patterns must be simple enough to be predictable in their application to the ridge lines yet powerful enough to convey the shape of a feature like a sidechain or a residue. The user selects a pattern and an edge in the ridge line graph that must be part of the match. The computer applies this local pattern and changes the current interpretation of the map to reflect the match. The computer changes only edges that are visible so that the user can immediately see the exact effects of the match. There are two types of patterns for each feature: one that matches the whole feature and one that unconditionally matches a single edge. There is also a pattern that will cause the type of a feature to revert back to unknown; all other patterns match only ridge line edges of type unknown.

The mainchain pattern, for example, matches a path that connects the given edge to the closest existing mainchain edge (that is, one that has already been interpreted as corresponding to the mainchain of the model) if it exists or else to the farthest edge that can be reached. Further restrictions on the pattern are that it may not form cycles or branches in the mainchain. Thus to designate a path in the ridge lines as mainchain, the user merely selects the mainchain pattern and an edge on that path.

1.3.2.4. Commands

There are three main types of commands available: viewing commands that affect which edges will be seen, flow of control commands that select the map or the state of its interpretation, and feature specification commands.

The viewing commands are themselves of two types. One set changes the scale and centering of the portion of the map that is visible on the display device. The other set selects any combination of model, map that has not been associated with a model residue, and map that has been associated. This selective display capability allows one, for example, to show only the model that has been fit so far, or that model plus the unassociated portions of the map. This last combination effectively shows the molecular model replacing the now invisible portions of the map that correspond to the model.

In addition to these interactive commands the viewing angle can be changed dynamically. The lowest level of map that is visible can be changed dynamically also. This is analogous to changing the contour level for a contour representation of a map.

The flow of control commands include the LOAD and SAVE commands. The current centering and scale are saved in addition to the state of interpretation so that one can easily switch between alternate interpretations for comparisons. The derivation of the saved versions is maintained so that one can tell if, for example, two saved versions are alternative interpretations derived from a common previous interpretation. UNDO and REDO commands allow the user to back up after issuing a command or to re-execute the command if it has been UNDOne. The current implementation allow the last 20 commands to be

UNDOne. The UNDO command will operate on any command with the restriction that only the last LOAD can be UNDOne and SAVEs cannot be UNDOne. REDO, of course, must be used to undo an UNDO command. There is a batch program to extract the fit model coordinates from a saved map interpretation.

The feature commands deal with three levels of features. First, features of the type mainchain, sidechain, carbonyl oxygen, and bridge. Next are residues built of mainchain, sidechain, and carbonyl. Finally chains of residues are matched with the amino acid sequence. The patterns that specify the lowest level features have already been discussed.

Specifying a residue requires that the lower level features be already identified. The user selects an amino acid type, a sidechain, and a carbonyl, the mainchain being implied from these. The computer matches the edges of the map features to the model for that residue type and does least squares fits of individual rigid body pieces to the map. Rigid bodies may extend to include bonds from adjacent residues that have been fit previously. For example, if the succeeding residue has been fit then the rigid body CA-C'-O is extended to include the N and CA atoms of that succeeding residue. The individual fits are combined to given an overall fit of the residue to the map. Within an individual rigid body fit, ideal bond lengths and bond angles are preserved since the rigid body is only rotated and translated for the best fit. Combining the individual fits creates a compromise between ideal geometry and a best fit to the electron density.

To register residues with the known sequence, the two ends of a chain of residues are specified. The computer rates how well the residue types that have been assigned to these residues, match the sequence starting at each possible position in the sequence and extending in each direction. A list of the best matches is presented to the user for selection rather than just taking the best match. This allows the user to apply special information to the decision process that may be very difficult to supply directly to the computer.

1.3.2.5. Results

Ridge lines have been computed for a 1.9A rubredoxin map, a sea snake neurotoxin map, a 2.0A but poor quality SOD map, and a 2.5A staph nuclease map. Only the rubredoxin map has been interpreted completely to the mainchain and sidechain level. The ridge lines are superb and interpreting it with the method is very easy. Only six edges had to be added to bridge gaps in the mainchain. In one 45 minute session Williams counted and identified sidechains without using the sequence. Subsequent comparison with a well fit model and the sequence showed that 40 percent of the residue types were correctly identified. If residue types with side chains of similar size and shape are considered equivalent (for example, THR and VAL), another 40 percent of the residues were chosen "correctly." The position of the iron atom was clear as well as the four CYS residues bonded to it.

Only a few residues have been fit to the ridge lines for the purpose of testing the fitting portion of the system. The other maps have not been examined in detail. Experience with these and poorer resolution maps may require modification of patterns or addition of new types of patterns but extensive modification of the system is not expected to be required.

1.3.3. Design of the New GRIP

The field of protein crystallography has moved rapidly in the last five years. GRIP-X will have to be far more flexible and powerful than GRIP-75 if we are to lead in the field of molecular graphics. We identify three categories of users: *fitters*, *viewers*, and *movers*. *Fitters* are protein crystallographers adjusting or creating a molecular model to fit an electron density map. *Viewers* are those studying a series of structures looking for homologies or attempting to characterize macromolecular structures. *Movers* are biophysical chemists studying molecular interactions or molecular dynamics. All three categories of users place special demands on a molecular graphics system.

Fitters typically work in two stages. The first stage is obtaining a molecular model, by hook or by crook. The second stage, at least in recent years, is to refine the molecular model against the crystallographic data. Both steps require the ability to edit the molecular model conveniently. Lack of such facility is one of the deficiencies in GRIP-75. We are designing the new GRIP to include editing of the structure at several levels — the atomic level, the residue level, and perhaps the secondary and tertiary structure levels.

There is no general agreement concerning the best methods to be used during the refinement stage. Thus GRIP must not only support refinement, it must support research on refinement methods. We are proposing to make available to the user a variety of automatic and semi-automatic aids to fitting, such as bump checkers, on-line geometry idealizers, and perhaps a utility to evaluate measures of the current goodness of fit to the experimental electron density map. We are also planning to include a convenient interface to batch programs for such tasks as structure factor calculation, partial structure difference map calculation, and structure factor least squares refinement. Finally, we will support generation and manipulation of structures with non-standard geometry. (Such structures can easily arise during refinement.)

Viewers generally have very elaborate graphics requirements. They need to be able to manipulate several structures at once, with both automatic and manual control of relative orientation. They need to be able to edit the picture (as distinguished from editing the molecule) very extensively. Viewers need the ability to identify and manipulate as units various substructures within the molecule — helices, sheets, and domains, for example; to selectively display chemically significant features of the molecule; and to show various abstractions of the molecular structure. (Common structural abstractions are the α-carbon skeleton of proteins, ribbon diagrams, and representation of helices by cylinders or β-sheets by broad arrows.) Present plans for GRIP-X do not include immediate full support of all of these features, but do include the hooks necessary to install them. We will provide the ability to identify and selectively display various subsets of a structure.

We do not yet have a satisfactory analysis of the requirements of *movers*. Obviously they need some of the features required by viewers. Just as obviously, they need a wide variety of automatic aids for evaluating different conformations and interactions. We presently believe that the same features of the system design which will provide automatic aids for fitters will also be adequate to incorporate the automatic aids required by movers. We are starting a collaboration with researchers at the Burroughs-Wellcome Company in Research Triangle Park who are working on problems of drug design. They are concerned with fitting model compounds into active sites, estimating binding properties, and the like. We hope that they will become the clients we need to focus our efforts in this field as the crystallographers have focussed our efforts in the fitting problem in the past.

The design and implementation of GRIP-X is in progress cooperatively with the IBM United Kingdom Scientific Centre at Winchester, England. We are striving for a common architecture for our systems (that is, they should appear the same to a user) even though we will necessarily have different implementations. Dr. William V. Wright, who directed work on GRIP-75, is now at the IBM U. K. Scientific Centre. The new design incorporates some novel human interaction concepts developed by Dr. Wright. It will also provide, in addition to the manipulation of structures by knobs and joysticks as in the present system, some very interesting ways of specifying structural manipulations from a data tablet. (We are continuing our popular tradition of providing our clients with a wide selection of techniques and strategies.)

We had severe communications problems during the previous year. We now have a weekly conference telephone call, and provision for exchanging written documents electronically over the IBM VMNET. We also exchange visits several times a year. This has substantially improved our collaboration.

The present division of labor has the U. K. group doing most of the architectural design of the new system (in particular, Dr. Wright) and work on the design of a molecular database system. We are developing a language for implementing graphics systems which we hope will be of general utility rather than just useful for this one system. Both groups are considering implementation questions and data structures.

1.3.4. Crystallographic Software

Dr. TenEyck has been developing a method of least squares refinement of macromolecular structures which promises to be faster, cheaper, and more robust than existing methods. A COMP 145 team has taken on the task of fitting all of the separate programs used in the method into one system. The system is being designed as a test bed for refinement methods in general, with the various steps of the refinement process all given well defined interfaces to the system. We hope that the availability of a system like this, coupled with good graphics for monitoring the progress of the refinement, will lead directly to more powerful methods of refining crystal structures.

1.4. Collaborative Research and Service

1.4.1. Calculations on Protein Structure and Interactions

Prof. Jan Hermans, of the biochemistry department of the University of North Carolina, has been a collaborator of ours since the earliest days of the GRIP system. The first GRIP system was written to aid his research in 1971. Lately his research interests have moved towards molecular dynamics. We have provided Prof. Hermans access to our VAX on a time-available basis for as long as we have spare capacity, with no commitment to maintain the service. The projects described below have been greatly aided by our facilities, but their scientific content is due entirely to Prof. Hermans.

An existing program that performs molecular dynamics calculations for a protein crystal (and which can perform energy minimization calculations) has been made to run satisfactorily on the VAX under the UNIX operating system. This is a copy of a program that also has run on a CDC Cyber and a VAX 11/780 under the VMS operating system. No "production" runs have been done so far on the VAX here. Other programs that need to be brought up on the VAX are a program that performs a Monte Carlo simulation of the protein-solvent crystal, and a program that calculates simulated electron density maps from the molecular dynamics and/or Monte Carlo results.

A second project is the development of MACREF, a macromolecule refinement procedure with energy minimization and least-squares crystallographic refinement. This is a collaboration with Dr. K. D. Watenpaugh of the University of Washington in Seattle. The program contains the first program mentioned above, and also PREDATOR, a procedure which prepares the data structures used in the energy refinement calculation. PREDATOR has been extensively edited on the VAX, and is now in Ratfor form. The ultimate objective is to have MACREF and PREDATOR both in Ratmac, a highly portable programming system developed by Prof. J. Stewart of the University of Maryland.

A third project concerns an excluded-volume theory of polymer-protein interactions based on polymer chain statistics. An algebraic theory has been developed and has been thoroughly compared with results of Monte Carlo simulations of exclusion of spheres by random segmented chains (and vice versa). The agreement is essentially perfect. Results of the theory also agree well with available experimental data. A manuscript describing the algebra and the comparison of the two sets of theory and experimental data has been submitted to *J. Chem. Phys.* A 15 minute talk on this project has been presented at the February 1982 meeting of the Biophysical Society in Boston.

Prof. Hermans work is supported by grants from the National Science Foundation (PCM76-22723 and PCM81-12234) and the National Institutes of Health (HL-20319 and HL-26309).

1.4.2. Interactions of Drugs with Proteins

We are beginning a collaboration with scientists at the Burroughs-Wellcome Company in Research Triangle Park, North Carolina, to develop improved methods for studying drug interactions with proteins and the evaluation of compounds for possible biological activity. This is a very complicated field with many opportunities for advanced graphics and automated aids for evaluating molecular interactions. We have been looking for clients working in this field to give us the expert chemical guidance we will need and the feedback as to which of our techniques are successful. The development of GRIP-75 owes a great deal to the heavy early use of the system by Dr. Sung-Ho Kim. We need someone to play a similar role in our development of molecular interaction software.

Initially we have aided in studies by Dr. Barbara Roth of the binding of trimethoprim to the enzyme dihydrofolate reductase. So far this work has been a rather basic docking exercise with a small number of degrees of conformational freedom in the trimethoprim. We are using this simple initial study as a prototype to determine what tools will be appropriate for this task.

Dr. Mike Corey of Burroughs-Wellcome uses the PROPHET system to generate and maintain files of chemically related compounds as part of the drug design process. PROPHET does not provide much aid in determining three-dimensional relationships within a series of structures, nor is the viewing and plotting hardware adequate to Dr. Corey's needs. (This is in part due to the fact that PROPHET is running on a computer 1,000 miles away, communicating over a 1200 baud telephone connection.) We have been discussing Dr. Corey's research problems with him to determine just which parts of his work are most suitable for automation and graphical aids. As an interim measure, a COMP 145 team is developing an interface between our VAX and PROPHET which will enable a PROPHET user to use our high performance three-dimensional graphics viewing system instead of Tektronix storage tube graphics.

1.4.3. Sea Snake Neurotoxin

Prof. Barbara Low of Columbia University and Dr. Steve Ginnell, a research associate of hers, spent about 150 hours fitting the structure of a neurotoxin to a greatly improved 1.4 A resolution electron density map phased by the density modification procedures of Dr. Douglas Collins, Texas A&M University. The two visits this year are a continuation of work done here beginning in 1977. Knowledge of the structure of this neurotoxin is important to understanding the nature of the neural connections it blocks.

1.4.4. Studies on Interpretation of Density Maps

Drs. D. and J. Richardson of Duke University are beginning a research project to study the effect of resolution on the accuracy of the interpretation of electron density maps. They believe that in maps with poorer than 3 A resolution certain types of misinterpretations become common. They are attempting to confirm this in controlled studies using maps calculated to different resolutions of known structures. There is surprisingly little systematic knowledge of the effect of resolution on the accuracy of a structure, partly because it is very difficult to separate the effects of resolution and phase errors. This study should give a lot of insight into the nature of low resolution maps, and could act as a valuable guideline for crystallographers deciding how much work to invest in data collection before the attempt to interpret their maps.

1.5. Training

Three new graduate research assistants joined the project this year. The training of new research assistants is a natural consequence of the number of students in this department seeking professional Master's degrees instead of doctorates. We give our clients sufficient training to use the system when they arrive, but do not train them in the operation of the underlying computer systems. Instead we provide them with trained assistants as "shepherds" to take care of that aspect of their use of the system. We find that it takes one to two days for a client to become comfortable with the graphics and structure manipulation features of GRIP-75, and the remainder of his visit can be devoted to productive work.

2. Administrative Changes

The major administrative change is that Dr. TenEyck is leaving in May to return to Oregon. His position has been advertised in *Science* and *Nature*, and resumes are coming in. Interviewing has started.

3. Highlights

Advanced Graphics

We have devoted a great deal of effort this year to new visual representations of molecules and density maps, and to making old representations more useful. An example is incorporation of dynamic motion into a full color raster graphics display, under the continuous viewpoint control of the user. Another example is variations of M. Connolly's surface representation, in which the surface points are replaced by small vectors normal to the surface. We also experimented with composite representations and with a dual workstation which has both line-drawing and raster graphics systems displaying the same molecular data in two different ways. We are learning a lot about the relative strengths and weaknesses of the two types of display running them together in this fashion. One measure of the success of our efforts in molecular graphics is that at the 1981 SIGGRAPH computer graphics conference, our images were used by Ikonas, Vector General, and Ramtek in their hardware exhibits.

Semiautomatic Density Map Interpretation

Tom Williams' work on computer aided interpretation of electron density maps has been extremely fruitful. Using a ridge line representation of the map, in which line segments connect local maxima, he has developed a method which saves a great deal of labor in the problem of going from the first electron density map to rough molecular coordinates. This reduces a step which normally takes weeks to a few days. Others have tried ridge line representations before, but not in quite this fashion. As a bonus, it appears that the ridge line representation may be the long-sought representation of electron density which permits computer graphics users to see large volumes of map without exceeding the capabilities of their displays.

Software for Protein Crystallography

One of the GRIP-X goals is to serve as a general workbench for crystallographers. We have started implementing standard crystallographic programs on our computers to support this goal. The system being developed is sufficiently flexible to support research in crystallographic computing methods as well as simply supporting crystallographic computing. Crystallography has long needed a suitable test bench for comparison of different methods in the same environment. We believe our system will provide this function.

4. Resource Advisory Committee and Allocation of Resources

Table IV lists the members of our Advisory Committee. We currently offer the facility, and such help as we can give, free of charge to any chemist:

- who has a scientifically interesting problem, as assessed by the Committee,
- whose work is at a stage where our facility might be useful,
- who is willing to commit his time, travel money, and effort to a serious use of the facility, and
- who is willing to give us written and oral feedback from his experience.

Table IV: Advisory Committee Members

Name	Degree	Title	Department	Institution
F. P. Brooks	Ph.D.	Kenan Professor & Chairman	Computer Science	UNC-CH
J. Hermans	Ph.D.	Professor	Biochemistry	UNC-CH
D. Richardson	Ph.D.	Professor	Biochemistry	Duke U.
J. Richardson	M.S.	Assistant Professor	Anatomy	Duke U.
W. V. Wright	Ph.D.	Senior Systems Architect	UK Scientific Centre	IBM England

5. Dissemination of Information

The availability of GRIP-75 is widely known among crystallographers. We publicise the facility by announcements and notices at scientific meetings, by demonstrations to all interested parties, and by word of mouth.

5.1. Announcements of Availability

During the 1981-1982 grant period the availability of GRIP-75 was announced at the meeting of the American Crystallographic Association in College Station, Texas, in March 1981; at the Computers and Chemistry meeting in Tallahassee, Florida, in March 1981; at the meeting of the International Union of Crystallography in Ottawa, Canada, in August, 1981; and will be announced at the Washington, D. C. meeting of the American Crystallographic Association in March, 1982. In addition GRIP-75 was included in the list of molecular graphics facilities published by the Brookhaven Data Bank, and of course is in the list of Biotechnology Research Resources published by DRR. These announcements should reach all of the scientists who might be interested in our facility. Our experience indicates that most of our clients hear about our facility from other crystallographers, then look up our telephone number in the DRR Biotechnology Research Resources list.

5.2. Demonstrations

Some 76 hours of 1981 resource time were used in demonstrations. During 1981 we demonstrated GRIP to the following biochemists and computer scientists.

Dr. T. Kehl	University of Washington
Dr. J. Mudge	Digital Equipment Corp
Dr. D. Borgwardt	Univ. of Washington
Dr. R. Dobler	Siemans AG Munich
Dr. M. Schmidt	Siemans AG Munich
Dr. M. Theofanos	
Dr. R. Kieburtz	Stony Brook - SUNY
Dr. T. Whitted	Bell Labs
Dr. J. Champness	Wellcome Foundation
Dr. M. Mc Kendry	Univ. of Illinois
Dr. B. Baxter	University of Utah
Mr. J. Harlow III	Harris Semiconductors
Dr. E. Luks	Bucknell University
Mr. E. Johnson	Paramount Pictures
Dr. K. Knowlton	Bell Labs
Mr. D. Turner	University of Kent
Dr. R. Churchouse	British Computer Board
Mr. H. Norton	
Dr. T. Leighton	
Dr. G. Petsko	M.I.T.
Dr. T. Kunii	Univ. of Tokyo
Dr. H. Blum	N.I.H.
Dr. R. Salemme	University of Arizona
Mr. J. Tolbert	Southern Bell
Dr. W. Franta	University of Minnesota
Dr. J. Ferrill	University of Minnesota
Dr. K. Culik	Wayne State University

Dr. C. Molnar	Washington Univ.
Dr. K. Schwams	Carnegie Mellon Univ.
Dr. G. Lindstrom	Univ. of Utah