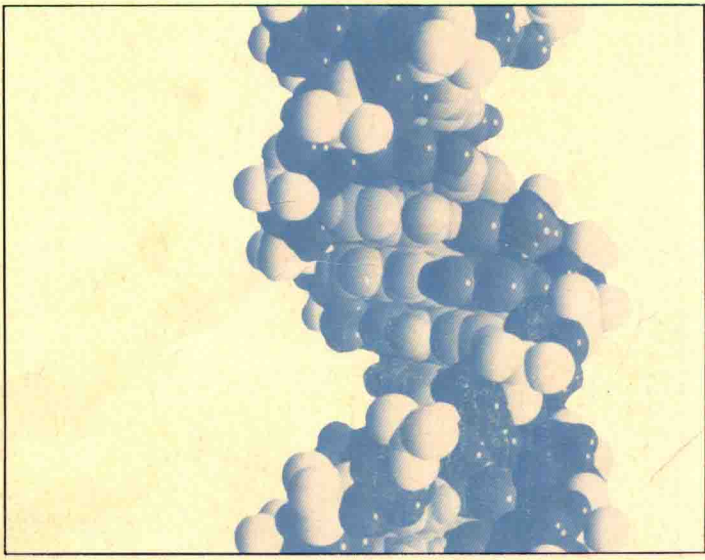
Current Communications IN MOLECULAR BIOLOGY



Computer Graphics and Molecular Modeling

Current Communications
IN MOLECULAR BIOLOGY
Cold Spring Harbor Laboratory / 1986

Computer Graphics and Molecular Modeling

Edited by
Robert Fletterick
University of California School of Medicine, San Francisco
Mark Zoller
Cold Spring Harbor Laboratory

Titles in

Current Communications in Molecular Biology
PLANT INFECTIOUS AGENTS
ENHANCERS AND EUKARYOTIC GENE EXPRESSION
PROTEIN TRANSPORT AND SECRETION
IMMUNE RECOGNITION OF PROTEIN ANTIGENS
EUKARYOTIC TRANSCRIPTION
PLANT CELL/CELL INTERACTIONS
TRANSLATIONAL CONTROL
COMPUTER GRAPHICS AND MOLECULAR MODELING

COMPUTER GRAPHICS AND MOLECULAR MODELING

International Standard Book Number 0-87969-193-X
Book design by Emily Harste
Printed in the United States of America

Cover: Random DNA sequence. (Computer graphics by Richard J. Feldmann, National Institutes of Health.)

The individual summaries contained herein should not be treated as publications or listed in bibliographies. Information contained herein can be cited as personal communication contingent upon obtaining the consent of the author. The collected work may, however, be cited as a general source of information on this topic.

All Cold Spring Harbor Laboratory publications may be ordered directly from Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, New York 11724. (Phone: Continental U.S. except New York State 1-800-843-4388. All other locations [516] 367-8425.)

Conference Participants

John E. Anderson, Dept. of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts

Patrick Argos, European Molecular Biology Laboratory, Heidelberg, Federal Republic of Germany

Carl-Ivar Brändén, Dept. of Molecular Biology, University of Uppsala Biomedical Center, Sweden

Cyrus Chothia, MRC Laboratory of Molecular Biology, Cambridge, England

Charles Cook, Hoffmann-La Roche Inc., Nutley, New Jersey

Henry Dayringer, Monsanto Co., St. Louis, Missouri

William F. DeGrado, Central Research and Development Dept., E.I. du Pont de Nemours & Co., Wilmington, Delaware

Beyrol Dominy, Pfizer Central Research, Groton, Connecticut

David Eisenberg, Dept. of Chemistry and Biochemistry, University of California, Los Angeles

Bruce W. Erickson, Dept. of Biochemistry, Rockefeller University, New York, New York

Gary Evans, Becton Dickinson Research Center, Research Triangle Park, North Carolina

Richard J. Feldmann, Division of Computer Research Technology, National Institutes of Health, Bethesda, Maryland

Thomas Ferrin, Computer Graphics Laboratory, University of California, San Francisco

Robert Fletterick, Dept. of Biochemistry and Biophysics, University of California, San Francisco

Manuel Glynias, Dept. of Pharmacology, Case Western Reserve University Medical School, Cleveland, Ohio

Jonathan Greer, Abbott Laboratories, Abbott Park, Illinois

Arnold Hagler, Agouron Institute, La Jolla, California

Roderick E. Hubbard, Dept. of Chemistry and Computing Service, University of York, England

T. Alwyn Jones, Dept. of Molecular Biology, University of Uppsala Biomedical Center, Sweden

Martin Karplus, Dept. of Chemistry, Harvard University, Cambridge, Massachusetts

Arthur Lesk, MRC Laboratory of Molecular Biology, Cambridge, England

Robert Manfre, American Cyanamid Corp., Princeton, New Jersey

Patrick Mize, Becton Dickinson Research Center, Research Triangle Park, North Carolina

Michael Pique, Dept. of Computer Science, University of North Carolina, Chapel Hill

Allen Richon, Ciba-Geigy Corp., Summit, New Jersey

Peter Stern, Dept. of Chemical Physics, Weizmann Institute of Science, Rehovot, Israel
William Taylor, Dept. of Crystallography, Birkbeck College, London, England
Dale Tronrud, Institute of Molecular Biology, University of Oregon, Eugene
Mark Zoller, Cold Spring Harbor Laboratory, New York

Preface

Many of the current problems in molecular biology and biochemistry only answer one- or two-dimensional questions. Examples of these are the primary sequence of a protein or the determination of the specific nucleotides recognized for the regulation of gene transcription. Three-dimensional problems are more formidable. For example, how does the primary structure of a protein determine tertiary structure and function? How do three-dimensional objects interact and relate? Beyond this, are questions of the fourth dimension, such as, what are the dynamics of a three-dimensional macromolecule?

The meeting on Computer Graphics and Molecular Modeling held December 10–13, 1985, at the Banbury Center of Cold Spring Harbor Laboratory brought together a diverse group of researchers to address these issues. The problems discussed were: protein sequence homology, three-dimensional structure homology in proteins, molecular design of functional macromolecules, X-ray crystallography, and software for the analysis of the energetics and dynamics of proteins. This book summarizes these presentations in the form of extended abstracts.

The Banbury Center was an ideal location for our meeting and we thank James Watson for making it available and Mike Shodell, Director of Banbury, and Bea Toliver for handling details of organization. Herb Parsons dealt effectively with the unusual audiovisual demands of such a meeting. The publication of this book in a short time is due to the cooperation of all speakers in providing their abstracts promptly and to the Publications department (Nancy Ford, Director; Judy Cuddihy).

R.F.
M.Z.

The meeting on Computer Graphics and Molecular Modeling was funded entirely by proceeds from the Laboratory's Corporate Sponsor Program, whose members provide core support for Cold Spring Harbor and Banbury meetings:

Agrigenetics Corporation
American Cyanamid Company
Amersham International plc
Becton Dickinson and Company
Biogen S.A.
Cetus Corporation
Ciba-Geigy Corporation
CPC International Inc.
E.I. du Pont de Nemours & Company
Genentech, Inc.
Genetics Institute
Hoffmann-La Roche Inc.
Johnson & Johnson
Eli Lilly and Company
Mitsui Toatsu Chemicals, Inc.
Monsanto Company
Pall Corporation
Pfizer Inc.
Schering-Plough Corporation
Smith Kline & French Laboratories
Upjohn Company

Contents

Introduction

R. Fletterick and M. Zoller*

Department of Biochemistry and Biophysics
University of California, San Francisco, School of Medicine
San Francisco, California 94143

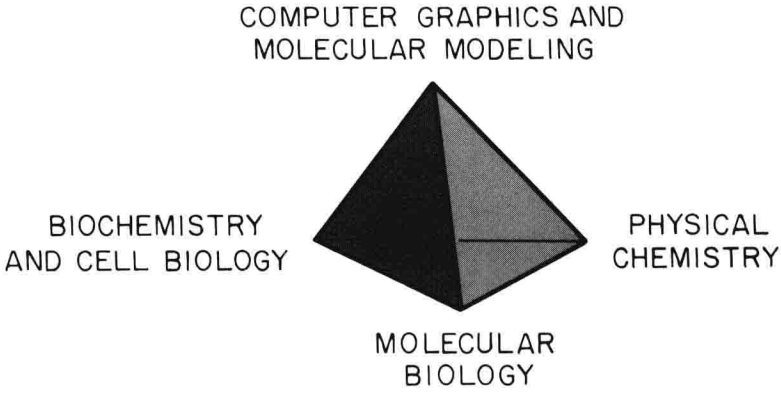
*Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

Although three-dimensional structural analysis of macromolecules has been possible for 25 years, a complete analytical description was brought about only in the last 5–10 years by the development of powerful hardware and software for computer graphics. The problems of displaying and analyzing three-dimensional macromolecules are formidable. The tools for fitting structural models to X-ray maps are now available, and three-dimensional models of proteins can now be accurately described, displayed, and compared (see Figs. 1–3 following). A second phase has now begun in which a few of the goals are to predict structure and function and to design useful proteins. Several accomplishments in the area of protein design are described in this volume. In addition, the results from energy calculations and protein dynamics suggest that structural and temporal simulations are achievable goals.

This was an appropriate time to bring together the eclectic group represented in this volume. Current methodologies of molecular biology and protein sequence analysis have accelerated the identification and determination of new protein structures and have produced access to new proteins of desired sequence. To ask meaningful questions about protein structure and function, it is important to understand the basis of protein structure and dynamics. Concomitant with the advances in molecular biology, the fields of X-ray crystallography, peptide synthesis, computer science, and protein dynamics have also been expanding rapidly. This volume is an attempt to present the most recent advances in each field and to draw the fields together through the use of the computer.

A goal for the future is the useful integration of the power of computers and the many software packages available. The computer systems in general need to be accessible to all biologists so that the different approaches toward the study of protein structure

and function can be successfully integrated (see below). (Also see Appendix: Molecular Computer Graphics Installations, compiled by the University of North Carolina.) This meeting was a modest start toward these goals.



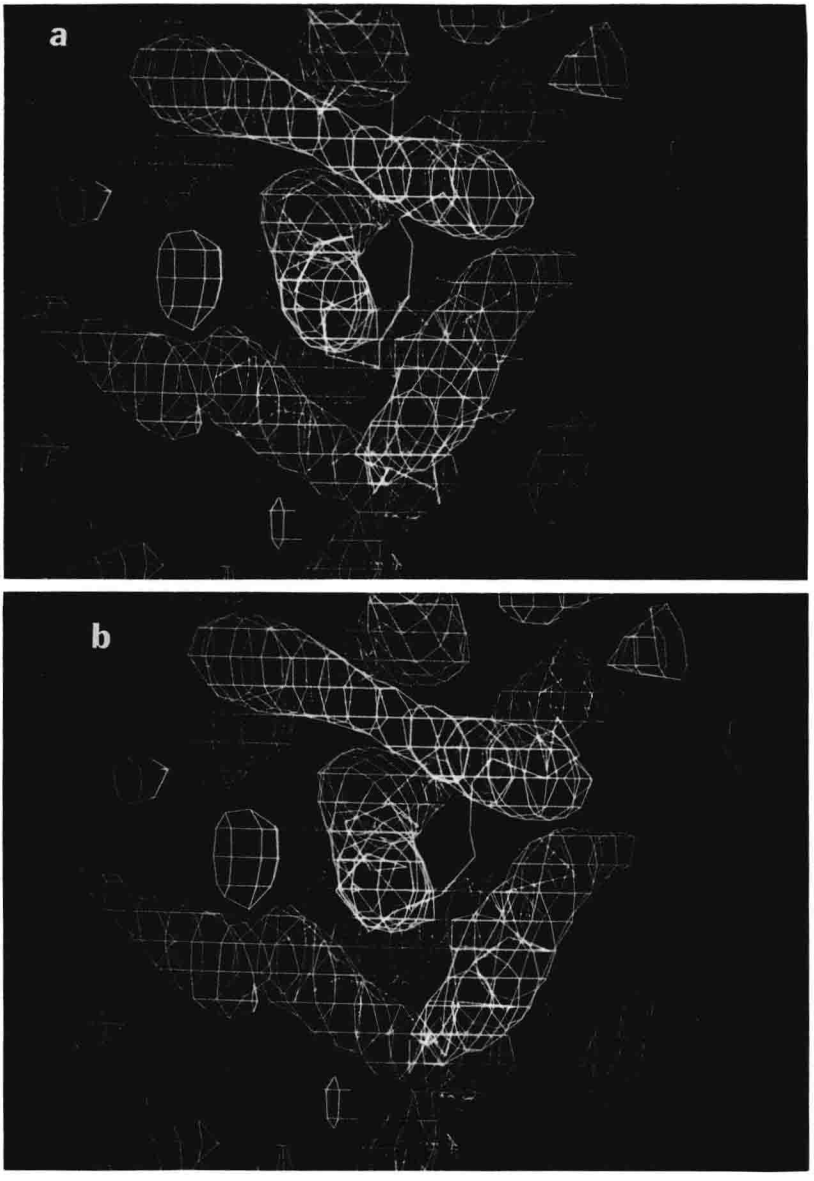


Figure 1 A monomer of R1-69 bound to DNA. (*a*) Model B-DNA (red), tilted and displaced as described in the text, is superimposed on the 7 Å electron density (blue) of the cocrystal. Also superimposed on the electron is the α-carbon backbone of λR17-85 (gold). (*b*) The same view as in *a*, but with the α-carbon backbone of the R1-69 model, constructed as described in the text, superimposed instead of λR17-85. Note the better correspondence of model α-helices with rods of density. (Fig. 1 for Anderson and Harrison, p. 95.)

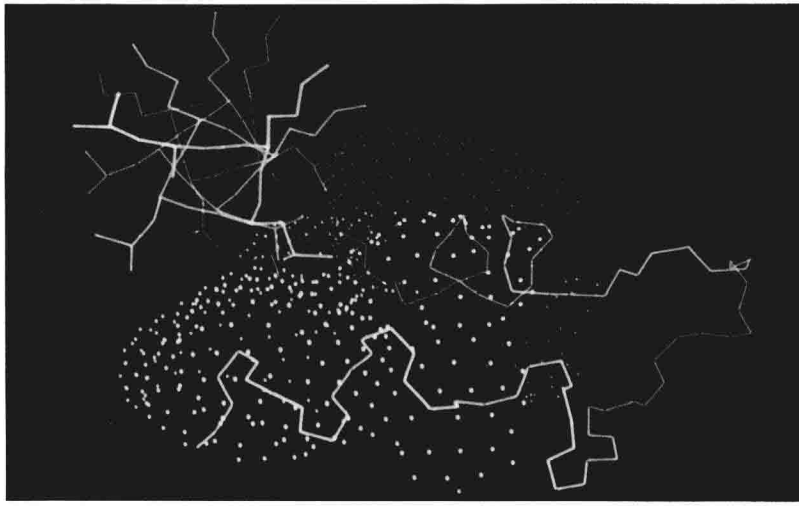


Figure 2 Proposed docking orientation for the basic peptide (Leu-Lys-Lys-Leu-Leu-Lys-Leu-Leu-Lys-Lys-Leu-Leu-Lys-Leu) to the carboxyterminal domain of the ICB-based model for calmodulin. Hydrophobic residues are colored green, basic residues blue, and acidic residues are red. The surface shown is the solvent-accessible surface generated for those residues on calmodulin that are believed to form the peptide-binding site. The white backbone is that of the protein, and the yellow backbone belongs to the peptide. (Fig. 2 for DeGrado et al., p. 85.)

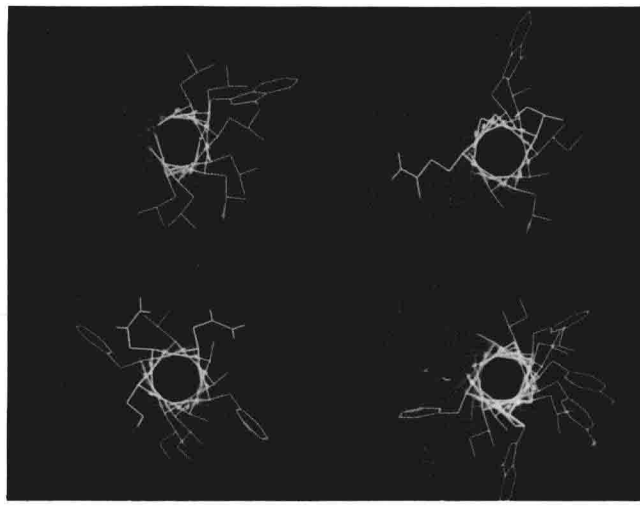


Figure 3 The amino acid sequences of calmodulin-binding peptides drawn in helical conformations looking axially down the helix. (*Upper left*) Residues 342–362 of phosphorylase B kinase; (*upper right*) residues 347–360 of skeletal muscle MLCK; (*lower left*) residues 3–17 of peptide RS20 (Lukas et al. 1986) representing a partial sequence of smooth muscle MLCK; (*lower right*) peptide 2 of Table 1. (Fig. 3 for DeGrado et al., p. 85.)

PROTEUS: Graphics Software for Proteins

H. Dayringer,* A. Tramantano, † and R. Fletterick †

*Monsanto Company, Chesterfield, Missouri 63198

†Department of Biochemistry/Biophysics
University of California, San Francisco, Medical School
San Francisco, California 94143

In the fall of 1983 Monsanto began to expand its efforts in biotechnology into areas that required significant capabilities for visualizaton of protein structures and the interaction of biomacromolecules and substrates. After exploring a number of alternative methods for obtaining computer graphics software and hardware, it was determined that developing our own capability in-house was the best long-term solution to our needs. To speed the process of development we undertook a joint venture with Professor Robert Fletterick's lab (University of California, San Francisco, Medical School). By combining their experience in the solution of protein structures and investigation of structure-function relationships and Monsanto's ability to commit money and people to development, we felt that we could achieve the best results in the minimum time. The graphics software PROTEUS was the result.

The software, written in C language for the VAX computer and the Evans and Sutherland PS300 graphics computer, provides a broad range of capabilities required for the visualization and manipulation of macromolecular structures. These capabilities include the standard abilities to rotate, translate, and scale objects and the more advanced abilities to mutate residues, insert and remove residues, and change the distance, angle, and dihedral angles between atoms and molecules. The range of functions provided by the program will be described in further detail.

The PROTEUS software processes input from a variety of sources. Molecules defined in the Brookhaven Protein Data Bank, Cambridge Crystal Data Base, FRODO DNS2, or DISCOVER (BIOSYM Inc.) formats can be read by the program. In addition, the program has its own internal format of saved files which can be read very quickly and restore the program to exactly the same state as when the save was done. When reading molecular definition files from

external sources, the connectivity of the molecule is determined if necessary using a template followed by distance search approach. This provides a very fast method for proteins but is sufficiently general that the Cambridge file can be handled. The program can also take input from a "user" format file of vectors, dots, or text and can read electron density contour maps.

Because it is written in C language, there are essentially no limits on the number of molecules or the complexity of the molecules that can be handled by this program. All data structures required to hold the information that defines the molecule – its coloring, surfaces, labels, distance monitors, and other properties – are allocated dynamically and stored as linked lists. User reference to objects, residues, and atoms are via user-assigned names for the objects and standard naming conventions for the residues and atoms. The user has complete freedom over the assignment of object alias names, and any number of such unique names can be assigned to any object.

The user interacts with the system via either a menu on the PS300, activated using the data tablet, or via a command interface through either the PS300 keyboard or the VAX log-in terminal. The menu interface is easier to use for the infrequent or novice user. The command interface is more powerful and generally faster than the menu but is somewhat harder to learn, at least for infrequent computer users.

The user can perform operations such as translation, rotation, and scaling of objects using either the PS300 dials or commands to the program (via either the command or menu interface). By default these operations are performed with respect to the world coordinate system defined by the PS300. The user is allowed to specify that the operations take place about an internal coordinate system for the object that is, by default, the moment of inertia axis system. This internal system can be redefined by the user.

In addition to performing operations on single objects, objects can be associated into global objects which can then be treated as a single object. Thus, it is possible to place two or more objects into a desired relative orientation and then rotate this entire ensemble of objects about its common center of mass or any other user-designated point. This feature is particularly useful when doing docking experiments.

Objects can be colored and labeled at the object, residue, and atom level under user control. In addition, a user-defined label string can be associated with any object. The color and size of each label type can be set for each object. The object color can be con-

trolled for each atom, residue, or object. In addition, the color of the surface can be set for each individual atom, independent of the color of the atom itself or its label. Objects can be surfaced using either a van der Waals or solvent accessible surface (Connolly 1983).

A variety of additional features are available, including addition and removal of bonds, merging and separating objects, changing torsional angles, and outputting the coordinates in a variety of formats. The user can view an object in stereo mode (Terabit viewers) and gently rock the object. Parameters such as distance, angle, and dihedral angles can be measured between any atoms.

A special feature allows the program to generate and display the symmetry-related replicates of a molecule using the space group and unit cell parameters. The user can control which replicates are displayed and can measure parameters between atoms of the replicate and the original object or another replicate.

The modeling features of the software include the ability to mutate a residue, to set distance, angle, and dihedral angle parameters, and to insert and remove residues from the protein. Mutation to a new residue preserves the backbone atom placements and provides best possible alignment of the side-chain. The parameter-setting command performs the operations either inter- or intraobject, provided that there are no conflicts such as those caused by rings.

Insertion or removal of residues is done using a modeling assist feature. The user indicates the location in the sequence where the modification is to occur. PROTEUS searches a prebuilt data base based on selection entries from the Brookhaven Protein Data Bank (PDB) and finds the 10 best models for the final structure after insertion or removal of the desired residue. The user views these models and chooses one of them. The program then uses the Hermans/McQueen (1974) algorithm to put the original protein into as nearly this final conformation as possible within the constraint that the loop must be closed.

The data base search is conducted on the basis of α-carbon distances. The PDB entries have been preprocessed to extract an α-carbon distance matrix for each residue. Beginning at each possible residue and extending for the next 18 residues, calculate the inter-α-carbon distance. Store each such "line" of distances as a matrix row in a disk file. When the user indicates the location of a change to be made to a protein, calculate a similar distance matrix for the protein, beginning back 9 residues from the site of the proposed modification. Now remove from the matrix the rows and columns in a region about the modification. The size of this region, which we call the "flexible region," is determined by the user (defaults to