

## A Computer Graphics System for the Building of Macromolecular Models into Electron Density Maps

BY JAMES R. MILLER,\* SHERIN S. ABDEL-MEGUID, MICHAEL G. ROSSMANN† AND DAVID C. ANDERSON

*Department of Biological Sciences and Department of Computer Sciences, Purdue University, West Lafayette Indiana 47907, USA*

(Received 18 August 1980; accepted 9 October 1980)

### Abstract

A brief description of the Molecular Modeling System-X graphics system hardware is followed by an explanation of the language which has been developed to realize an 'electronic Richards box'. A variety of commands permits the construction and manipulation of a protein model within an electron density distribution. Usually about ten amino-acid residues can be displayed at any one time within a box of 20 grid points on an edge. The density is changed automatically as the viewer translates the model off the edge of the screen. He can then add, subtract or modify residues as appropriate.

### Introduction

The Richards Optical Comparator (Richards, 1968) is the usual instrument employed in the interpretation of protein and nucleic acid electron density maps. Its invention by Richards represented an enormous advance over previous crude model building techniques. Nevertheless, there are significant problems related to electron density display, model construction and model inaccuracies. Many of these disadvantages can be overcome by use of a computer graphics display, although the latter introduces difficulties of its own (*e.g.* limitations on the amount of data that can be displayed and the need for rotation to obtain three-dimensional perception). Furthermore, a graphics display system bridges the gap between computation of electron density and refinement of molecular models without the intermediate construction of a physical model.

The use of computer graphics for the interpretation of electron density is not new. Several systems have been implemented in the past. Particularly noteworthy are the University of North Carolina (Britton, 1977; Wright, 1972), the University of California at San Francisco (O. Jones, 1978; O. Jones, Hack & Beutel, 1978), the MRC Laboratory of Molecular Biology

(Diamond, 1978) and the München (T. A. Jones, 1978) systems. Although some general principles are common to each installation, adaptation is invariably required to use whatever hardware is locally available. At this time various technologies are being explored and improved. We present here another method, 'BUILD', with its own advantages and disadvantages. It was adapted for use on the Washington University, St. Louis, Molecular Modeling System-X (MMS-X) graphics display system. The latter, available in about 16 crystallographic laboratories in North America, combines a minicomputer with specialized hardware.

The philosophy of the BUILD system is to minimize problems of file manipulation and other aspects of housekeeping essential to the performance of the software. To a great extent the operator carries out the manipulations he would do on a Richards optical comparator without regard to computing detail. For instance, the system infers from user manipulation of the display when it is appropriate to obtain new portions of the electron density map.

### Hardware

The MMS-X was developed, at the Washington University Computer Systems Laboratory (CSL), St. Louis, from the experience gained with four earlier MMS machines (Barry, Bosshard, Ellis & Marshall, 1975). The current device incorporates a Texas Instruments 980B minicomputer with a memory of 48 K words of 16 bits each, a Pertec dual disc drive with a total capacity of  $2.4 \times 10^6$  words, a Hewlett-Packard 1321A cathode ray tube, and a Beehive terminal. In addition, special purpose hardware, developed at CSL, is connected to the Texas Instruments minicomputer via a direct memory access channel. It consists of a  $3 \times 4$  matrix multiplier, controller, several knobs, switches, buttons and three-degree-of-freedom joysticks (Perry, 1978). The complete system is connected through a telephone line (960 characters per second) to the university's CDC 6000 computing facilities. This permits the exchange of electron densities and coordinate files.

A view of the system console is shown in Fig. 1. The Beehive cathode ray tube and keyboard are used for communication between the BUILD system and the

\* Present address: CAD/CAM Division, Control Data Corporation, 4201 North Lexington Avenue, Arden Hills, Minnesota 55112, USA.

† Correspondence should be addressed to Michael G. Rossmann at the Department of Biological Sciences, Purdue University, W. Lafayette, Indiana 47907, USA.

user. One joystick is normally used to manipulate the entire display while the other is only attached to the molecule. Each can be used to perform three-dimensional translation or (on throwing a switch) three-dimensional rotation of the display. There are options for continuous zooming and stereoviewing. There are also six potentiometers (labelled A through F) on a separate control panel which can be used to adjust specified dihedral angles.

### User interface

A language was developed which uses two-letter mnemonics for individual commands. Commands may require additional information in terms of directives or prompt the user for input in regard to the current display (Table 1). The syntax of a command can be written as:

$$CC[QQ][m-n-]\{, [QQ][m-n|-]\},$$

where CC is the mnemonic, square brackets denote optional entries, braces enclose entries repeated zero or more times, | denotes alternatives,  $m$  and  $n$  are residue numbers representing the beginning and end (inclusive) for the specific molecular portion to be affected, and - implies the negation of the command. The term 'directive' describes the  $[QQ][m-n-]$  portions of the command. Not all commands allow directives; however, directives are executed in the order of their appearance in a command. Examples are as follows.

(i) The entry DI would cause the entire molecule to be displayed.

(ii) The entry DI MC 10-15 would cause the main-chain atoms in residues 10 through 15 to be displayed.

(iii) The entry DI 18-20 would cause the main- and side-chain atoms in residues 18 through 20 to be displayed.

(iv) The entry DI 18-20, MC 10-15 would cause both (ii) and (iii) to occur.

It is noteworthy that the order of application of

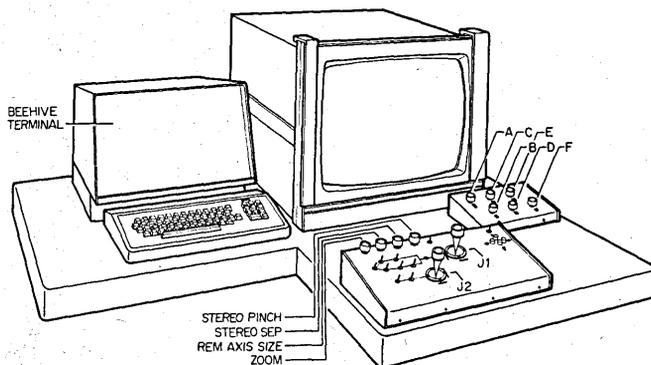
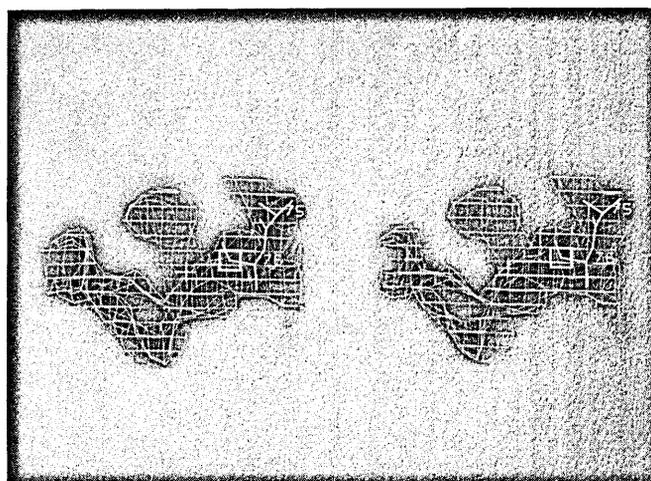


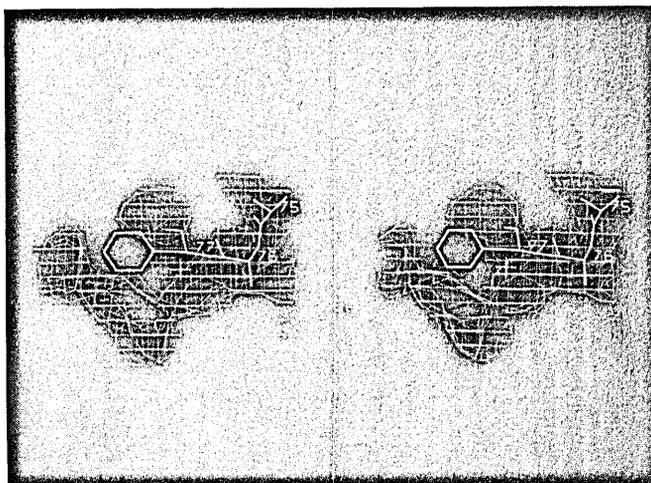
Fig. 1. Schematic drawing of MMS-X system showing: Beehive terminal and keyboard, Hewlett-Packard cathode ray tube, joysticks J1 and J2, other switches, and dihedral angle potentiometers A, B, ..., F.

Table 1. List of commands in the BUILD system

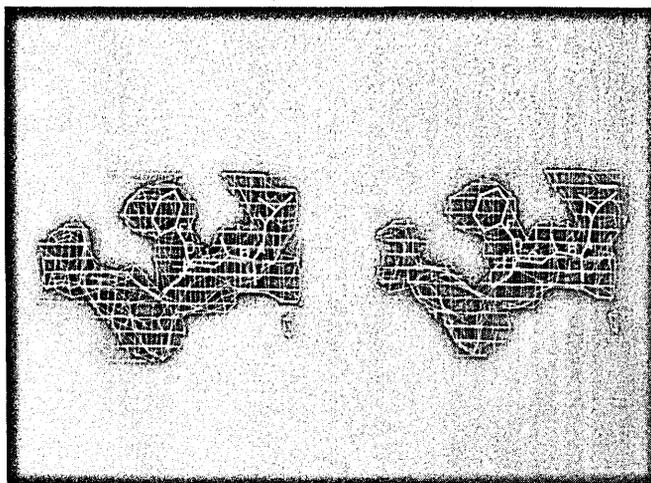
Command	Directive legal	Number of possible qualifiers	Prompt	Description
AD	No		Yes	Add residue
AN	Yes	9	Yes	Animate molecule
CK	No		No	Checkpoint system
CL	No		Yes	Change contour level
DI	Yes	1	No	Display molecule
DM	Yes	0	No	Delete molecule
DS	No		Yes	Distance between atoms
HM	Yes	0	No	Highlight main chain
LA	Yes	1	No	Label atoms (e.g. C, N, O, ...)
LI	Yes	9	No	Label intensity (adjusts brightness of labels)
LR	Yes	2	No	Label residues (e.g. Ala, Asp, ... and/or 1, 2, ...)
MO	No		No	Map on/off
NA	Yes	1	Yes	New area of map
NM	No		Yes	New map specifications
QU	No		No	Quit
RC	No		Yes	Restore conformation
RE	Yes	1	Yes	Re-number residues
RM	Yes	2	Yes	Read molecule
RR	Yes	0	Yes	Replace residues
SC	No		No	Save conformation
WM	Yes	2	Yes	Write molecule



(a)



(b)



(c)

Fig. 2. Sequence of operations in fitting a part of the SBMV structure. (a) Cursor showing where new residue (Phe) is to be added. (b) Added Phe residue in initial standard orientation. (c) Phe residue rotated into electron density by means of rotations about bonds labelled *A*, *B*, *C* and *D*.

directives is not in general commutative. For example, DI-, 13-14 would display the molecule from residues 13 to 14, whereas DI 13-14, - would cause no display at all.

#### A. Building the molecule

The AD (add residue) mnemonic is the basic model building command. It prompts for a specific residue (e.g. Ala, Lys, etc.). Unless the residue is to be 'unattached', one end of the available polypeptide must be selected for 'docking' the new residue. The desired end is selected by a cursor (Fig. 2a) which can be moved to successive polypeptide ends by depressing a key. The new residue is then added according to a standard geometry (Fig. 2b) at the chosen end of a chain.

The AN (animate) mnemonic assigns the rotary knobs (A through F) to selected bonds of the molecule. This command is 'manual' if entered without directives or 'automatic' if directives are included. In the latter case the knobs are assigned by BUILD to the first available single bonds in the range specified by the directives. These are then labelled appropriately on the cathode ray screen (Fig. 2c). In the former 'manual' mode a cursor is displayed which can be rapidly passed over the rotatable bonds in the molecule. A knob can be assigned to a bond by depressing the key (A through F). Whether automatic or manual bond selection is performed, the user is prompted to determine the 'sense' of all selected bonds, that is whether the portion of the molecule at one end of the bond or the other is to move when rotations are performed. This is accomplished by specifying one atom whose position is to remain invariant during all bond rotations. This atom is termed the 'center atom' and is selected with the cursor. The second joystick is also available for manipulation of the molecule. Unless it has been disabled with a qualifier, it is attached to the polypeptide fragment containing the center atom once the latter has been selected.

The DS command computes the distance between any two atoms specified by the cursor and thus helps to assess the quality of a given conformation. This feature might well be expanded to search automatically for unacceptable van der Waals' distances, unlikely dihedral angles and possible hydrogen bonds.

There are a number of commands which help to visualize the molecule more easily. These are HM which highlights the main chain, LA which labels atoms with their chemical types (e.g. C, N, O, S, etc.), and LR which labels residues with their consecutive numbers and/or three-letter amino-acid codes. In addition, LI can alter the intensities of the display labels and RE permits the renumbering of amino-acid residues. Finally, should the amino-acid sequence need changing (e.g. Phe to Tyr), then the alteration can be readily accomplished by means of the replace residue (RR) command.

#### B. Saving the molecule

Owing to current memory limitations there is a maximum of ten residues and 50 atoms storage capacity in core. This is backed up by the discs. Some or all of the

molecule in core can be written onto disc by use of the WM (write molecule) command and can be regained with the RM (read molecule) command. These commands edit or select from a file containing the best current molecular conformation. While the DI (display molecule) command allows for the selective display of the current conformation in core, the DM (delete molecule) command deletes all or some of the residues presently stored in core. There is also a facility (SC, save conformation) for rapidly saving one or more conformations. These can then be regained with the RC (restore conformation) command for comparison.

The CK (checkpoint) and QU (quit) commands allow the user to suspend or stop building as he desires. The checkpoint operation stores the current status of a building session which can then be regained exactly at a future time.

#### C. Electron density manipulation

The original electron density map, evaluated at points on a suitably sized grid, is maintained in the host CDC 6000 system. Portions of it (usually about  $20 \times 20 \times 20$  points) can be extracted and sent to the MMS-X graphics device using the block mode file transfer protocol (Croft, 1979). The extracted portions are contoured by the Texas Instruments 980 and stored on disc ready for display. Usually only about 30 portions can be stored on any given disc simultaneously. Each extracted portion is suitably labelled and recorded in a directory by BUILD.

When the user initiates a new investigation or changes to a completely different segment of the map, he will be prompted by the NA (new area) command for the limits in x, y and z of the first portion to be extracted as well as the contour level. From then on he can manually designate any other portion for extraction. Alternatively, manipulation of the display initiates automatic selection of other portions. As the user translates the currently selected portion of the map, centering particular regions of interest on the screen, BUILD determines whether another portion of the map would better cover the current region of interest. Neighboring portions are automatically selected by translation along the grid allowing for a predetermined overlap. Each portion will have the same size and contour level. If it is determined that another portion would be more suitable, a search is automatically initiated for this portion of the map on disc. Should it be found, the response is relatively fast (about 5 s); if it is missing, it must be requested from the CDC 6000 and the response will be slow (3 to 4 min at least).

At this time the extraction procedure can accommodate unit-cell translations but as yet no attempt has been made to implement crystallographic symmetry operators. Indeed, it is more straightforward either to compute the original map over an area somewhat larger than the asymmetric unit, or to generate the additional symmetry-related portions of the map in the host. The university CDC 6000 system is thus simply a device for

computing and storing the necessary electron density, but does not enter into the actual operation of the graphics system. The BUILD program is structured in such a manner that the host computer can be replaced by a magnetic tape drive or any other input device attached to the MMS-X.

There are four commands associated with map handling. The directives, which normally specify sections of the molecule, are either irrelevant for map control or permit only a qualifier. The NA (new area) command was described above. This is complemented by the NM (new map specification) command which permits the use of two or more maps simultaneously. This feature allows for easy comparison of similar molecules in their respective electron densities (*e.g.* isozymes, quasi-equivalent subunits in viruses, *etc.*). The contour level can be changed with the CL command, while the displayed map can be turned off (or on) completely with the MO command, leaving only the molecule on the screen.

#### Experience with the system

The BUILD system was used to construct the C subunit of southern bean mosaic virus (SBMV; Abad-Zapatero, Abdel-Meguid, Johnson, Leslie, Rayment, Rossmann, Suck & Tsukihara, 1980) to a 2.8 Å resolution map. A model was simultaneously constructed into a Richards optical comparator with respect to the same electron density map. An alanine residue was constructed at each residue position in the absence of a known amino-acid sequence (Fig. 3). However, in many places it was possible to make a reasonable guess as to the character of the residue (*cf.* Fig. 2). The computer graphics-built model and the 'Richards-box' model were then systematically compared.

The Richards-box backbone coordinates of the  $\beta$ -barrel (excluding the amino terminal arm) were fitted to the MMS-X results by a least-squares procedure which

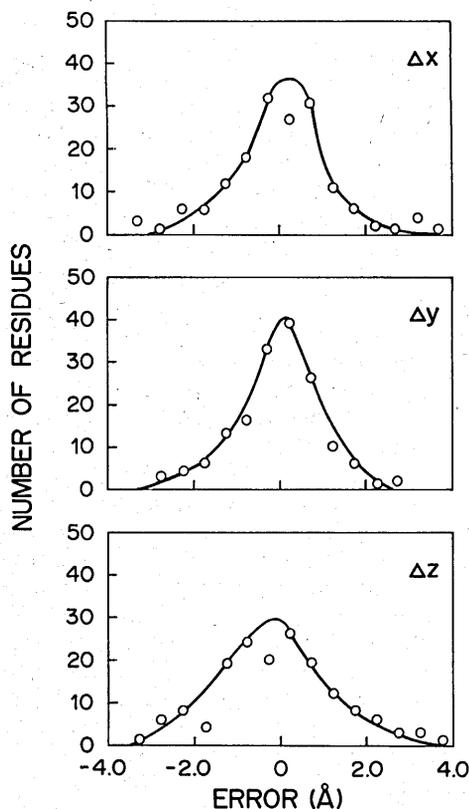


Fig. 4. Error distribution in the differences (a)  $\Delta x$ , (b)  $\Delta y$ , and (c)  $\Delta z$  between the Richards-box and computer-graphics-system coordinates of equivalenced residues.

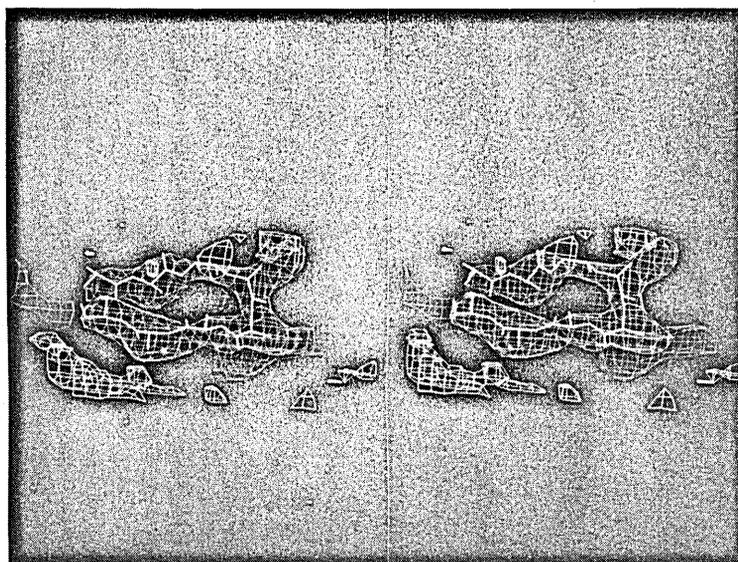


Fig. 3. Stereoview of a turn between two strands showing the density and superimposed model seen in the MMS-X.

is also able to find 'insertions' or 'deletions' in either model (Rao & Rossmann, 1973; Rossmann & Argos, 1975). There were 160 residues which could be equivalenced, while 20 residues of the Richards-box model and 5 residues of the MMS-X model had no apparent homology. Frequency distributions of the differences between homologous  $C_\alpha$  coordinates are shown in Fig. 4. These are seen to be approximately Gaussian with a standard error of around 1.1 Å in each case. All deviations greater than 2 Å were then examined more closely.

In Fig. 5 is shown a hydrogen-bonding scheme for the  $\beta$ -barrel of SBMV based on the Richards-box model. On it are marked those residues which deviate by more than  $\pm 2$  Å in  $x$ ,  $y$  or  $z$ , as well as the position of insertions and 'deletions'. This figure demonstrates that there are no systematic differences of any one coordinate relative to various parts of the molecule. However, it is seen that the major differences between the two methods of model building occurred in helices and turns. Re-examination showed that the single contour level required by the MMS-X was too high in the region of helices; therefore the multiple contour

representation in a Richards box is of significant help. In the vicinity of turns, however, it is clear that the MMS-X model fits the electron density better than the Richards-box model.

One of the difficulties in the graphics system is the problem of inspecting the quality of conformations and hydrogen-bonding arrangements. This is because of the limitation imposed on the number of residues which can be viewed simultaneously. These difficulties are likely to diminish when extra core memory becomes available in the near future.

We are grateful for help and encouragement from Dr David Barry and his colleagues at the Washington University Computer Systems Laboratory. We would like to thank Dr Celerino Abad-Zapatero for the Richards-box coordinates of SBMV and Sharon Wilder for help in the preparation of the manuscript. The work was supported by the National Institutes of Health (grants no. GM 10704 and AI 11219) and by the National Science Foundation (grant no. PCM78-16584). SSA was supported by a National Institutes of Health Postdoctoral Fellowship (no. 5 F32 GM 07103).

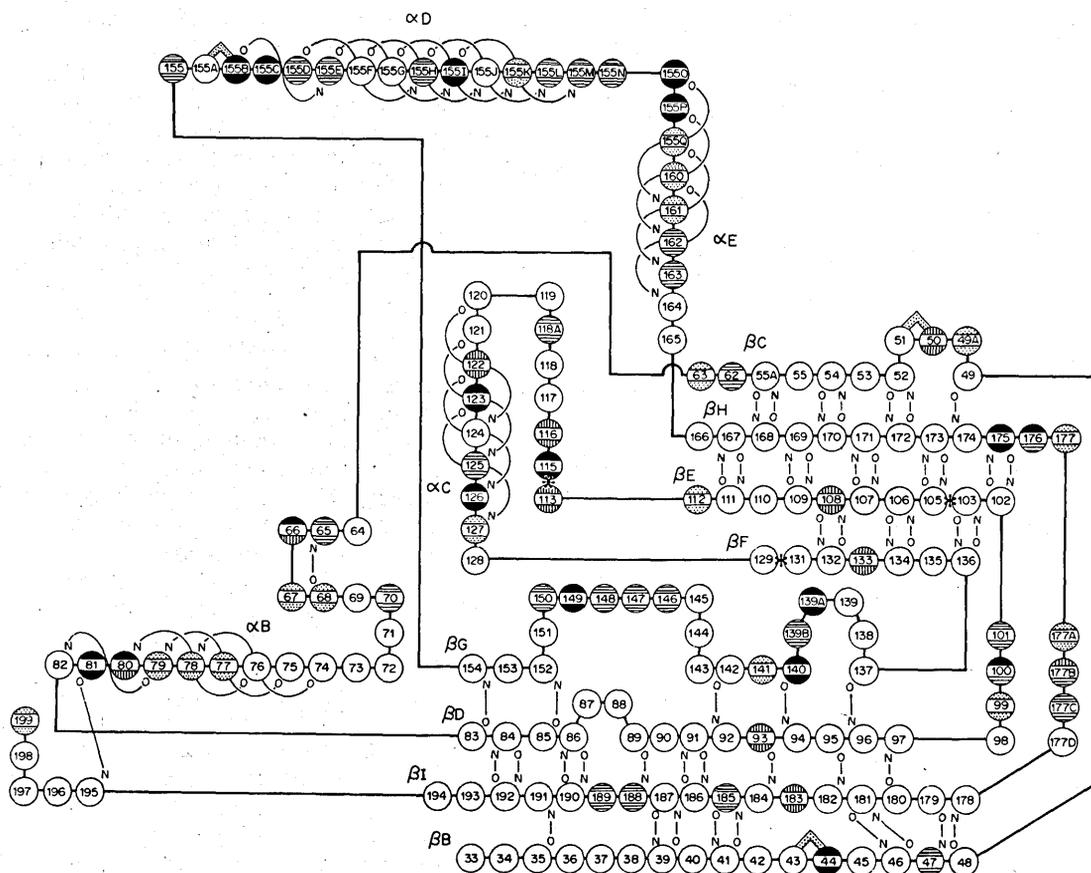


Fig. 5. Hydrogen-bonding scheme for the  $\beta$ -barrel of the SBMV C subunit based on the Richards-box model. Marked are those residues differing by more than  $\pm 2$  Å in their  $C_\alpha$  coordinates in  $x$  (●),  $y$  (⊙) or  $z$  (⊖) between the Richards-box model and the MMS-X model, as well as the positions of insertions (⊕) and deletions (⊖).

## References

- ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1980). *Nature (London)*, **286**, 33-39.
- BARRY, C. D., BOSSHARD, H. E., ELLIS, R. A. & MARSHALL, G. R. (1975). *Computers in Life Science Research*, edited by W. SILER & D. A. B. LINDBERG, pp. 137-147. New York: Plenum.
- BRITTON, E. G. (1977). *A Methodology for Ergonomic Design of Interactive Computer Graphic Systems, and its Application to Crystallography*. PhD Thesis, Univ. of North Carolina.
- CROFT, W. (1979). *Block Mode Terminal Protocol - September 1979*. Document K5-BLOCKP, Computing Center, Purdue Univ., West Lafayette, Indiana.
- DIAMOND, R. (1978). *BILDER: A Computer Graphics Program for Biopolymers and its Application to the Interpretation of the Structure of Tobacco Mosaic Virus Protein Disks at 2.8 Å Resolution*. MRC Laboratory of Molecular Biology, Univ. Medical School, Cambridge, England.
- JONES, O. (1978). *Protein Interactive Graphics System: A Computerized Model Room - July 1978*. Computer Graphics Laboratory, School of Pharmacy, Univ. of California, San Francisco, California.
- JONES, O., HACK, P. & BEUTEL, T. (1978). *BILD - A Tutorial Introduction: Three-Dimensional Picture Drawing by Computer - July 1978*. Computer Graphics Laboratory, School of Pharmacy, Univ. of California, San Francisco, California.
- JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268-272.
- PERRY, T. C. (1978). *MMS-X Architecture -- An Overview - February 1978*. Tech. Memo. no. 252. Computer Systems Laboratory, Washington Univ., St Louis, Missouri.
- RAO, S. T. & ROSSMANN, M. G. (1973). *J. Mol. Biol.* **76**, 241-256.
- RICHARDS, F. M. (1968). *J. Mol. Biol.* **37**, 225-230.
- ROSSMANN, M. G. & ARGOS, P. (1975). *J. Biol. Chem.* **250**, 7525-7532.
- WRIGHT, W. V. (1972). *An Interactive Computer Graphic System for Molecular Studies*. PhD Thesis, Univ. of North Carolina.