R. Stroud. A wholly digital version of the device, which allowed coordinates to be both read out and entered through a control console (e.g., for model building from coordinates), was constructed at the University of California at San Diego, and subsequently at the University of Arizona. The motorized versions of this superposition device produce coordinates of comparable precision to the optical and acoustic devices described previously (at least ±1 mm), and similarly allow for very rapid measurements (one atomic position per 30 sec). However, construction of the former device is more complicated than either the AIMS (which can simply be rented) or acoustic coordinate measuring devices.

With the advent of computer graphics and associated software for both map fitting and coordinate readout (this volume [12]), it appears only a matter of time before graphics obviates either building physical models of proteins in optical comparators or devising means for measuring their coordinates. Nevertheless, the techniques described here will undoubtedly continue to prove effective aids in determining crystal structures for some time to come. In addition, many modeling studies of protein conformation and interaction continue to be most readily investigated in their initial stages by the construction of physical models. Subsequent model development and analysis by computer necessitates having accurate coordinates, so that the technical feasibility of such studies depends on having an accurate and easy way of measuring model coordinates.

4 F. R. Salemme and D. G. Fehr, J. Mol. Biol. 70, 697 (1972).


By T. Alwyn Jones

Introduction

Computer graphics provides an elegant method of controlling the protein crystallographer’s interaction with his model. The graphics display allows the model to “show” the crystallographer a part of its electron density such that he can decide how a molecular fragment best fits. Once the crystallographer has made his decision, the computer merely does the bookkeeping and minor improvements. The aim of the molecular fitting program, therefore, is to create the necessary environment to allow the crystallographer to decide what atoms he wants in what piece of density.
A great deal of effort has been spent to develop the necessary control software and, in some cases, to build the hardware. At present there are approximately 10 density-fitting program systems in active use. Most laboratories are not able to build the necessary hardware, but high-performance equipment is available from a number of vendors. The program FRODO\textsuperscript{1,2,3,4} is presently implemented on the Vector General 3400 (with DEC, VAX, or PDP-11 computers), MMS-X, Evans & Sutherland PS 2 (PDP-11), MPS (VAX or PDP-11), and PS300.

The man–machine interface in FRODO varies with the available equipment. All systems use a data tablet to pick menu items and to identify atoms shown on the screen. The VG3400 and Evans & Sutherland PS300 systems use analog-to-digital converters to define the view direction, and other picture-related functions such as clipping, zooming, and intensities. Commands that invoke dihedral angle rotations, single atom, and fragment shifts are also coupled to the A-Ds. On the Evans & Sutherland PS2 and MPS equipment, pseudo A-Ds are drawn on the screen and can be activated by the data tablet pen. The MMS-X version uses the standard user control panel.

Program Flow

The following is strictly applicable to only the VG3400 version of FRODO. Other versions have some minor differences.

After starting the program the user must specify a control data set. This data set contains all of the important parameters needed by FRODO. It includes, for example, the data set name for the user's coordinate file and the regularization zone. It is a source file and can be edited (at one’s own risk). The user then gets to the CHAT interface, which has a large number of menu items activated from a terminal keyboard. Every exit from CHAT causes an update of the control file. Control normally then passes to the display loop.

To make the system easier to use an effort has been made to keep the display menu options to a minimum, and to a single "page." The data tablet pen position is marked on the screen by a cursor, and a menu item is activated by moving the pen so that the cursor is positioned over the


item and then pushing the pen into contact with the tablet. In principle, any number of commands can be activated and each is polled in turn. This means, for example, that an atom can be moved and have its contacts updated at the same time. In practice some care has to be taken, since there are a number of mixed options which may be disagreeable to the user. For example, if a group of atoms is being moved and the user activates the SAVE command, then the current fragment coordinates will get written to disk. The active commands can be seen at a glance because each has a star drawn next to it. To exit from the display loop, the user must pick a suitable menu item to either terminate the session or enter one of the utilities. Reentering the display loop causes a new loading of coordinate information from the disk.

The vectors drawn on the display are constructed from three data sets and normally show some of the atoms in the coordinate data set superimposed on a "chicken wire" representation of some sort of electron density.

The coordinate data set contains more than just atomic coordinates. The atoms are grouped together to form a residue. There are residue records to describe the type of residue (e.g., PHE, MPD), the name of the residue (e.g., A2, 10G), the position in the data set of the atom records for this residue, the center of gravity, and the radius of the residue. The residues are grouped together to form a sequence. As far as the display loop is concerned, the sequence is only important when defining viewing zones, i.e., it does not necessarily force any chemical connectivity between residues (although it may exist). The data set can also contain extra information such as lattice type (P, I, R, F, A, B, C), unit cell constants, and crystal symmetry information. This information is optional but may be required for certain commands. The user must decide at the CHAT interface how he wishes to access this data set. There are three possibilities: (1) Define the start and end residues of a zone. The program then displays all the atoms in the residues within the zone as defined by the sequence. (2) Define a point in space and a radius, and then display all of the atoms in the data set which are within the volume. (3) Define a mixture of 10 display zones plus a sphere.

In the sphere mode the user can choose an option to display any symmetry-related atoms which may fall within the volume. Both the sphere and symmetry options make use of the residue center of gravity information to decide what appears in the volume. Another option allows one to define by name which atoms are to be displayed; e.g., one can define just Cα to see the fold of a protein.

After picking which atoms are to be displayed, one must decide on a connectivity, i.e., which atoms should have a line drawn between them.
The usual connectivity scheme in FRODO is based on distance criteria so that if atoms are closer than a certain distance they are joined. In sphere and mix modes all atoms are tested together. In zone mode the connectivity is built up a residue at a time, and a specific link is made between residues. An atom with no connections appears on the display as a three-dimensional cross. If one is displaying just Cα atoms, for example, there is an option to connect the first atom to the second, the second to the third, etc. The initial connectivity does not necessarily represent chemically correct bonds. It is simply there as an initial framework for the crystallographer to decide how he is to change his structure to fit the density.

The second data set consists of linked vectors. It may represent density contoured at a number of different levels, or skeletonized electron density, or guide points, or a vectorized library of molecular data sets. The vectors are arranged in three-dimensional volume elements and in what are called contour commands (C-COMs). If the data set is a vectorized map, each C-COM corresponds to a contour level. In the vectorized molecular library each molecule is equivalent to a C-COM. The user can decide in the CHAT interface which (if any) C-COMs are to be chosen.

The third data set is an electron density map. It is also arranged in three-dimensional volume elements with each density value packed into one byte. This data set is much smaller than an equivalent vectorized map, and has the advantage that the contour levels can be changed at any time. It is, however, slower to work with than the vectorized data set. The map can also be used to automatically fit molecular fragments to the density. One often uses both a map and vector data set where the contour level has been chosen after a brief inspection of the map.

Building an Initial Model

The crystallographer usually knows the rough fold of his molecule before starting work on the display. This is best determined by extensive study of minimaps plotted on stacked plastic sheets. The structure solution of retinol binding protein by Newcomer et al. is one of the few exceptions to this rule. There are then four different ways of building the model on the display.

The first method is a relic of working with Kendrew wire models in a Richards box. FRODO has extensive model-making features to produce coordinates from a given sequence which have standard bond lengths and angles and preferred torsion angles. A zone of residues can be made and

moved to the place in the map where one wishes to begin building (not necessarily the N terminus). The fragment can be translated and rotated by the display menu command FBRT so that the first residue sits close to its density. Up to six consecutive dihedral angles can be varied at a time, and by repeated FBRTs and TORs the fragment may be made to fit the density. These coordinates are written to disk when the user is satisfied by using the SAVE command. The model-making option can be used to extend the zone of residues and an attempt made to fit these. However, it rapidly gets more difficult to do this while maintaining the constrained, rigid geometry.

In method two, by judicious choice of display zones, the user fits a few residues as described above but introduces a discontinuity in a peptide linkage by separately fitting the next zone of residues.

In method three the user introduces discontinuities directly on the screen. The screen connectivity is used to decide which atoms are affected by dihedral rotations and by fragment rotation/translation. Suppose a zone of residues fits the density but the user sees that a side chain in the center of the zone would fit much better if he could change $\phi$ (around the N-$C_\alpha$ bond) for the residue. If one changes this angle directly, all the other atoms to the end of the zone would be moved out of density. This is prevented by breaking the $C_\alpha$–C bond of the residue using the menu BOBR command and then rotating around the N–$C_\alpha$ bond. This, of course, distorts the bond angles around the $C_\alpha$ atom. More commonly, the user disconnects the side chain from the main chain and moves the small fragment straight into the desired density. This is illustrated in Fig. 1a (which is drawn on a plotter directly from the picture on our display), where a growing chain has clear density for the phenylalanine side chain. The $C_\alpha$–$C_\beta$ bond is broken and the ring moved into the density using FBRT (Fig. 1b). The same coordinates are shown from a different view in Fig. 1c, where one can clearly see density for the carbonyl oxygen. In Fig. 1d the oxygen has been moved into this density and we now have a very distorted residue.

It should be clear that to simplify the fitting process we must introduce errors in bond lengths, angles, and fixed torsion angles. These can be removed by model regularization. To prevent the buildup of errors in particular variables, the regularization should have no built-in rigid constraints (such as fixed bond lengths, for example). FRODO uses the method described by Hermans and McQueen\(^6\), which they call the method of local change. In this method each atom is shifted to minimize a weighted sum of terms representing the shift from its starting positions.

Fig. 1. Fitting a growing protein chain using method 3. (a) The growing chain with the Phe residue out of density. (b) The Phe ring is moved as a rigid group to fit the density. (c) A different view showing density for the carbonyl oxygen. (d) The carbonyl oxygen is moved into density. (e) The result of regularization with certain atoms fixed. (f) The ring stays in density after regularization.
errors in bond lengths, bond angles, and fixed torsion angles. Extra terms can easily be added to maintain preferred values for normally variable torsion angles (e.g., for an α helix) and for distances between atoms (e.g., to maintain hydrogen bond distances).

To keep the atoms in density FRODO has a menu option FIX. Any atom hit with the FIX option will not move during regularization. Before
regularizing the atoms shown in Fig. 1d, the peptide oxygens of residues 230 and 231 were fixed together with C\textsubscript{\(\beta\)} of residues 231. The result of regularization is shown in Figs. 1e and 1f. Sometimes one carries out a second regularization with nothing fixed to remove some buildup of errors (normally in the CO bond). The skills to be learned in using FRODO are to decide how to distort the model to fit the density, and then to decide which atoms to fix so that the main group of atoms stay in density and the rest are pulled into it. Often the fit is a combination of actions such as a bond break to delimit a dihedral rotation, the dihedral rotation to line up some chain with some density, another bond break to separate the fragment, and finally a rotation/translation shift into the density.

The fourth method of fitting attempts to place atoms near their density before actually sitting in front of the display. Guide coordinates are read directly from the minimap. At most one picks (for a protein) a guess for the C\textsubscript{\(\alpha\)}, O, and one side-chain atom. For regions of the molecule with secondary structure, FRODO’s model-making options are used to generate the desired conformation. Another option is then used to find the best rigid body fit to the guide coordinates. The gaps between secondary structure are filled by first inserting guide atom coordinates into the data set and making these atoms fixed. The coordinates for the remaining atoms get added (rather roughly) by the regularizer option, provided coordinates already exist for the first residue in the regularization zone. Because the guide atoms are fixed, the newly added atoms are forced to move close to their density. A second cycle without fixes is normally required to produce good stereochemistry. The last residue of the regularization zone is used as the first of the next zone. Once the complete protein has been made, the map can be inspected and improvements made as described earlier.

After making the first pass through the molecule in zone mode, the user must make a second pass in sphere mode. This is not just to ensure that chains do not interpenetrate. Since one rarely works at atomic resolution, the interpretation must be guided by contacts with neighboring residues. At this stage one can also decide how well the atoms fit the density. Atoms not fitting density are not used in later phase calculations.

The advantages of building a model with FRODO rather than in a Richards box are as follows:

1. **Speed.** One can normally build a model more quickly in a display. This gain can be lost by spending more time on details such as making better hydrogen bonds and trivial changes to improve the fit to the density.

2. **Accuracy.** All parts of the map are equally accessible in the display, but not in the box. Their crystallographic R factors are noticeably better,
often below 40%. The chicken wire representation of the density in the
display is usually clearer than stacked sections in a box, and if the crystal-
lographer can correctly interpret the density, he can build the model more
accurately.

3. **Control.** Once a newly positioned conformation has been SAVEed
with FRODO, the coordinates are safely on disk. Backup and restore
features are available on the display menu, so that accidental overwriting
of coordinate information is not a disaster. The wire model is easily de-
graded by gravitational forces prior to measuring coordinates. These co-
ordinates must be regularized and their positions checked by plotting on
density sections.

4. **Volume.** For large enzymes and viruses the actual space occupied
by a physical model becomes a problem, especially in mature laborato-
ries. A Kendrew model of the plant virus satellite tobacco necrosis virus
(STNV) would have a volume of 18 m³.

5. If the crystallographic asymmetric unit consists of more than one
molecule, the model of the first molecule built in the display can be used
as a good starting model for the second. This technique was used on
southern bean mosaic virus, in which there are three different protein
chains. One can also use coordinates from related conformations as was
done by Eklund *et al.*⁸ in their study of holo-alcohol dehydrogenase,
ADH. In this example the coordinates for the two domains of the apo-
ADH were rotated into the map, and then rebuilt to fit the density. The
second chain of the molecule was then built using the transformed first
chain as the initial model.

The main disadvantage of building a model in the display is that there
are usually so many people wanting to use it that one cannot just go and
check out an idea on the spur of the moment.

**A Tool in Crystallographic Refinement**

FRODO originated in the groups of Drs. Robert Huber and John Gass-
man in 1976. By that time it was known that protein molecules could be
refined, provided the crystallographer had the necessary patience and
computer power. At that time Huber’s group used Diamond’s constrained
real space refinement program⁹,¹⁰ to improve the model fit in maps calcu-

---


lated with $2|F_o| - |F_c|$ amplitudes and model phases. After some cycles the refinement would stop and it was then necessary to find out why. By studying various sorts of maps (with model or model/isomorphous combined phases) it was usually possible to locate errors in the model but very difficult to correct them. Since then, refinement methods have improved, and mostly gone into reciprocal space.\(^{11,12,13}\) These programs are very good at removing small random errors from the model, but they cannot correct major nonrandom errors.\(^2\) During refinement one hopes that the improvement in the phase angles is such that the nonrandom errors can be identified and corrected on the display.

When building the first model in the MIR map, I find it useful to divide the atoms into three groups: the good, the bad, and the ugly. During refinement one hopes that the distribution is pushed in the direction of the good atoms. In Munich and Uppsala an atom flag in the FRODO coordinate data set is used to decide if the atom is used in phase calculation. Normally during the course of the refinement more atoms are used, although one frequently removes atoms for a few cycles if one distrusts their behavior. We often use chopped Fouriers, in which a portion of the structure is removed from the phase calculation and a Fourier calculation (normally $2|F_o| - |F_c|$ amplitudes) made around the region of interest.

The bookkeeping qualities inherent in the computer system are an important aid in refinement. With FRODO the crystallographer can maintain a library of past models in the vector data set and check his present model against any past version. It is also helpful to keep a map copy of the MIR map since it is the only density unbiased by the model.

The main disadvantage of working with FRODO is that one is often faced with one’s own inadequacy, in that one can see everything perfectly and still not know what to do.

**FRODO Crystallographic Extensions**

An important development in computer graphics in the last few years has been the use of 32-bit computers to control the display. The increased computer power means that crystallographically meaningful calculations can be made on part of the structure when the user sits at the display. The first to be introduced is a real space refinement option.\(^{14}\) This option is an

---


extension of the fragment rotate/translate menu option and is called finger­tip refinement. The fragment moves as a rigid body to maximize the grid sum convolution $\sum P_{\text{calc}}P_{\text{obs}}$, where $P_{\text{calc}}$ is the calculated density obtained from a Gaussian function and $P_{\text{obs}}$ the observed grid density. The user controls the number of grid points over which to sum and the rotation/translation search range. The technique is illustrated in Fig. 2. The side chain $C_\beta-C_\gamma$ bond has been severed and the two fragments have been arbitrarily moved out of the density, as shown in Figs. 2a and b. The results of volume fitting each fragment are shown with the same views in Figs. 2c and d. A noninteractive version can refine a zone of residues. Each residue is split up into smaller fragments, each of which is refined as above. The result can be shown on the display so that the user can decide if there should be any fixes before regularization. These options are not seen as replacing reciprocal space refinement methods but as an aid to the FRODO user. They have, however, been used by Jones and Liljas$^{15}$ to carry out the first crystallographic refinement of a virus (satellite tobacco necrosis virus) to a resolution of 2.5 Å.

**FRODO as a General Molecular Modeling System**

FRODO was designed as a tool for the practicing protein crystallographer. However some of its features make it of more general use:

1. The flexible coordinate data set and the absence of any connectivity dictionary mean that any sort of molecular fragment can be displayed.

2. There are a number of different options available for choosing which part of the coordinate data set should be displayed. However, on a 16-bit computer there are space limitations as to how many atoms can be manipulated (200 atoms on a PDP 11/VG3400 system).

3. The MOL option can be used to create quite complicated display lists. These can be made up of any combination of atoms, zones, spheres, residue types (i.e., all Phes, say) and molecular surfaces. Display lists are sets of vectors and as such take the place of the electron density normally displayed with FRODO.

4. The fragment rotation/translation with neighbor calculations can be used to dock substrates into active sites. This has recently been enhanced to include local energy minimization (Cambillau et al. $^{16}$).

5. The edit features in the SAM option allow one to change residue types and delete or insert new residues. If there are sequence similarities between two molecules, and if the X-ray structure has been solved for one

---


FIG. 2. Fitting a fragment automatically using real-space volume fitting. (a) and (b) are different views of the starting coordinates. The dashed line represents the Cβ–Cγ bond, which is not drawn on the display because the atoms are too far apart. (c) and (d) are the results after refinement of each fragment.
of them, then its coordinates can be used as a skeleton to hang on the other sequence.

The program is cumbersome when one tries to compare two related molecules, because only one coordinate file can be active at a time. How-
ever, the main drawback of the program is that there is no entry to a complete protein data base system except via a single coordinate data set.

Acknowledgments

Drs. J. Gassmann and R. Huber encouraged me to develop a density-fitting program. Professor T. Blundell invited me to get FRODO running on an Evans & Sutherland PS 2, and Dr. I. Tickle improved my efforts. Dr. L. Bush made the MPS and Drs. J. W. Pflugrath & M. A. Saper made the PS300 conversions. Drs. C. Cambillau and E. Horjales are responsible for the energy-minimization options. Dr. J. Hermans developed the regularization method.

[13] Rationalization of Molecular Models

By Jan Hermans

Introduction

It is normal to represent the "solution" of a crystal structure, which itself approaches the electron density as closely as the accuracy of the structure factors permits, in terms of a molecular model, i.e., a collection of atoms. There are two principal reasons to do this: chemists and biochemists would pay little or no attention to a structure that was presented in any other form, and the model is an indispensable part of the principal crystallographic refinement technique applicable to proteins. Available chemical information is invariably incorporated in the model; of a protein it is preferable to have the entire amino acid sequence. If the amino acid sequence is not known, a good deal of it may be inferred from the shape of the calculated electron density; the larger side chains are easily recognized. However, individual C, N, and O atoms are not resolved in the first maps. Atomic resolution may be reached at the later stages of refinement, if the experimental data are of sufficiently high resolution.

The fact that the map does not clearly indicate individual positions for the majority of atoms causes considerable uncertainty, which can be much reduced by making the model conform to standard stereochemistry, according to measured bond lengths and bond angles of small molecules such as amino acids and peptides. These were first established in the now classic work of Pauling and co-workers. Bond lengths and bond angles have been found to be very similar for each type of chemical group, even in somewhat different physical (i.e., nonbonded) or chemical environments. In addition, such groups as aromatic rings and the peptide group...