

# XtalView/Xfit—A Versatile Program for Manipulating Atomic Coordinates and Electron Density

Duncan E. McRee

*Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037*

Received November 30, 1998, and in revised form January 29, 1999

**Xfit is a model-building and map viewing program in XtalView that is used by the structural biology community including researchers in the fields of crystallography, molecular modeling, and electron microscopy. Among its distinguishing features are built-in fast Fourier transforms that allow users flexibility in map calculations including the creation of OMIT maps and the updating of structure factors to reflect model changes from within the program. Written in C and using the freely available XView toolkit, it is highly portable to almost any X-windows based workstation including Intel-based LINUX systems. Its user interface is designed to aid in facile model-building and contains a semiautomated fitting system that allows the user to interactively and rapidly build chain *de novo* into an electron density map. The program is highly optimized to allow such features as interactive contour levels and map calculations to be completed within a few seconds. Features in the latest version including phase-combination, solvent-flattening, automated water addition, and small-probe dot contact surfaces, as well as basic design features, are discussed.**

© 1999 Academic Press

**Key Words:** computer program; electron density fitting; molecular modeling; protein crystallography.

## INTRODUCTION

The number of structures in the Protein Data Bank (Abola *et al.*, 1997) has grown exponentially and is rapidly approaching the 10 000 mark at the time of this writing. In order to solve these structures, an enormous investment in manpower is being made. Given the increasing power of computers, it is logical to harness this computing power to assist the structural biologists' efforts to determine structures and to relieve much of the tedium in solving structures and building models. Furthermore, robust computer algorithms can go a long way toward increasing the accuracy of structure determi-

nations. A new generation of software is needed and is being actively developed by a number of groups. Our effort, XtalView, is predicated on the belief that visualization of the data is key to more rapid structure determination. By providing for easy visualization, XtalView aids structural biologists to see patterns and alternatives that may be key to solving a structure and provides instant feedback on the effect of their choices.

Although XtalView can also be used for working with reflection files to phase data, this article will discuss XtalView's model building and map viewing module, Xfit. Xfit has become popular over the years with 1000 copies being distributed in 1998 alone. General design features of Xfit will be discussed and several key features of the program will be covered.

*History.* XtalView was started in 1991 at The Scripps Research Institute (TSRI) with the goal of using the emerging X11-windows system and Internet more effectively for crystallography. At that time, the use of windows-based computers was largely limited to having several VT100 terminal emulators open at once. Programs used on-line imitations of card decks for input, and output was designed for 120-column line-printers that were rapidly becoming extinct. Indeed, programs referred to the input as cards and the files were treated as tapes. The idea of a graphical user interface (GUI) had caught on in the rest of computing but protein crystallography, despite having embraced computing from the earliest days, needed some prompting to catch up. Accessible, inexpensive workstations with X11 graphics were readily available but could not be used for fitting due to the lack of software. To remedy this situation, XtalView was written to provide a graphical interface for solving protein structures and to allow map fitting on standard X11 workstations. Using this system, we completely solved *Chromatium vinosum* cytochrome *c'* by MIR and fitted the electron density maps using only a black-and-white SUN Sparcstation I. The structure was finished in

late 1991 and published in 1993 (Ren *et al.*, 1993). Although a Convex supercomputer was used for refinement with XPLOR (Brünger *et al.*, 1987), today, a PC running LINUX provides more than adequate computing power for the largest and most sophisticated fitting requirements.

From the very beginning, XtalView has been used by many groups for uses other than the intended one of protein crystallography. We used it to solve the structures of a cyclic peptide nanotube by a combination of 2-dimensional electron diffraction, molecular modeling, and back-transformation of these models (Ghadiri *et al.*, 1993). A number of other groups have used it for viewing electron microscopy-derived maps and for molecular modeling as well as for solving a large number of protein and DNA structures.

### THE XTALMGR

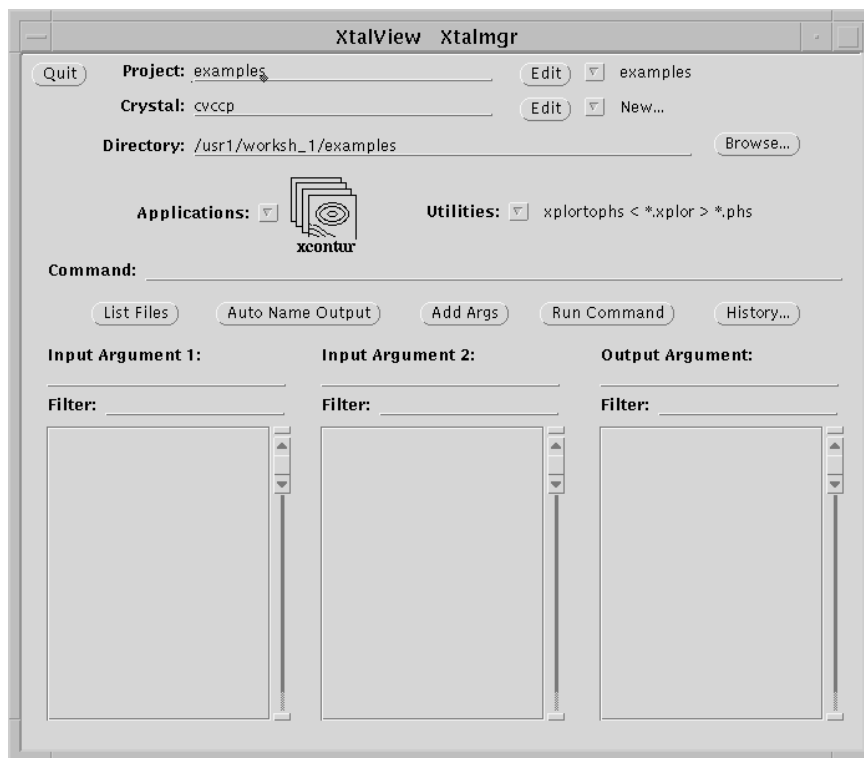
The XtalMgr program (Fig. 1) is the central control point for XtalView. It organizes the files and provides a launching point for the rest of the applications. It can be used to set up projects and edit and create crystal files. The Crystal Editor option is used to

edit/create the parameters of a crystal. Normally, Xfit is started with XtalMgr after the files to be used are selected.

### XFIT USER INTERFACE

*Atom stack.* Xfit uses an atom stack for many operations (Fig. 2A). All operations use a prefix notation. First, the operands are selected and put on the stack, and then the command is given. For example, to calculate a distance, one first clicks on two atoms to put them on the stack and then clicks on Distance on the main window. Atoms are pushed onto the stack in two ways: by picking the atom on the canvas (Fig. 2B) by left-clicking or by picking from the atom list in the Model window.

*Mouse.* To maximize portability, the primary interface is the mouse. Most UNIX systems come with a three-button mouse and XFIT was designed to work with three buttons. There are two fundamental mouse operations available in XView, click and drag. The right mouse button is reserved for the MENU button and is used to bring up the Canvas Menu. The left mouse button is used for picking atoms when it is



**FIG. 1.** XtalMgr. The XtalView XtalMgr program is used to launch other XtalView applications and to organize the user files. When an application is selected, the files that can be used with that application are listed. The user then selects the files to be used and can either name the output file by typing in a name or use the Auto Name function to have the program suggest a name. The user then selects the files to be used with the application and builds up the command line with Add Args. The application is then launched with Run Command. A history of the commands launched is kept. The list of available applications is kept in a file in the XtalView data directory so that the list of applications can be customized at each site.

clicked and for rotations when it is dragged. Rotations are always treated as orthogonal. Rotating about the axis perpendicular to the screen ( $z$ ), is done by dragging near the top of the screen or by holding down the Ctrl key while dragging. The cursor changes shape to indicate when the user is in the  $z$  rotation area.

The middle mouse button changes function for different fitting modes. In the default mode, the mouse pans the screen position in  $x$  and  $y$ . Again, the top of the screen is used for translations in  $z$ . When there are current atoms to be fit (that is, moved relative to the rest of the model), the mouse can be put into one of three modes that operate on the current atoms, rotation, translation, and torsion (bond twisting). The modes are switched by using the right mouse button to access the Canvas menu. Since the Canvas menu appears anywhere on the canvas, it is not necessary to move the mouse to access the Canvas menu, and this makes fitting more rapid.

*Orthogonal movements.* On the canvas menu are operations that rotate the viewpoint precisely  $90^\circ$  about the  $x$ ,  $y$ , and  $z$  screen axes. These useful operations are missing from many modeling programs such as "O" and Insight. They allow completely orthogonal movements that do not cross-couple in a manner analogous to the orthogonal slides and axes on an X-ray goniometer. Without these operations, when users rotate the view to move the atoms in the third direction, they inadvertently change the two directions they thought were already fixed because they do not move orthogonally with respect to the old position. With the orthogonal commands, only two drags are necessary to get the atoms in the correct position as the cross-coupling effect between the axes is eliminated.

*Activating atoms to be fit.* First one or more atoms are placed on the stack by picking and then the user selects one of seven types of groups to be fit: one atom, all picked atoms, one residue, one residue isolated by cutting the peptide bonds, all picked residues, a group defined in the model window, and the entire molecule. The specified part of the model is then colored green and made current; that is, they can then be moved relative to the rest of the model. A number of operations are performed each time the current atoms are moved, including redrawing, surface calculations, geometry calculations, and goodness-of-fit measurement. XFit can also call an external program when the atoms move and receive back geometry for display.

*Undo.* Changes to the coordinates of the model can be undone at any time by swapping them with saved coordinate states. Deleted residues are still visible in the residue list but are flagged as deleted

and not drawn or saved to output files. By toggling the delete flag the residue can be brought back at any time. The program can also be run in an autosave mode where it saves the coordinates before any change to the model.

*Hardware stereo.* Hardware stereo using the CrystalEyes system has been available on the SGI versions of the program for some time. Recently, the CrystalEyes hardware has become available for LINUX and DEC workstations and the latest version of the program supports hardware stereo on these platforms. Refer to the XtalView Web site at <http://www.scripps.edu/pub/dem-web/> for details.

## XFIT FEATURES

*Coordinate refinement.* Xfit goes beyond simple coordinate refinement to allow the user to interactively push and pull the model as if it were a mechanical model. For example, a single atom can be fit while a larger section including the atom is refined. This function can also take into account the map density so that the model is real-space refined relative to the map in a local region. This has the effect of making the model move into the density once it starts to overlap with it.

*Maps and phases.* Xfit has a built-in fast Fourier transform (FFT) capable of going in both directions from phases to maps and from models to phases. This allows for flexibility in the use of the program. For instance, the resolution of the map can be changed at any time and/or the type of map can be changed. Except for experimental phases, such as MAD and MIR, Xfit can calculate crystallographic phases from the model if given a reflection list and the PDB file. This allows updating the phases at any time during the fitting. A partial structure factor calculator also allows making OMIT maps on the fly.

*Spline maps.* Normally maps are calculated in a grid that is convenient for the FFT algorithm. The map is sampled discretely and if points in between the grid points need to be calculated, they are interpolated. This interpolation gives errors of about 20%. Spline maps avoid this by using a spectral B-spline description of the map (Grosse, 1981) that is accurate to 0.1% at any arbitrary point. This gives better looking contours and more accurate refinement and averaging.

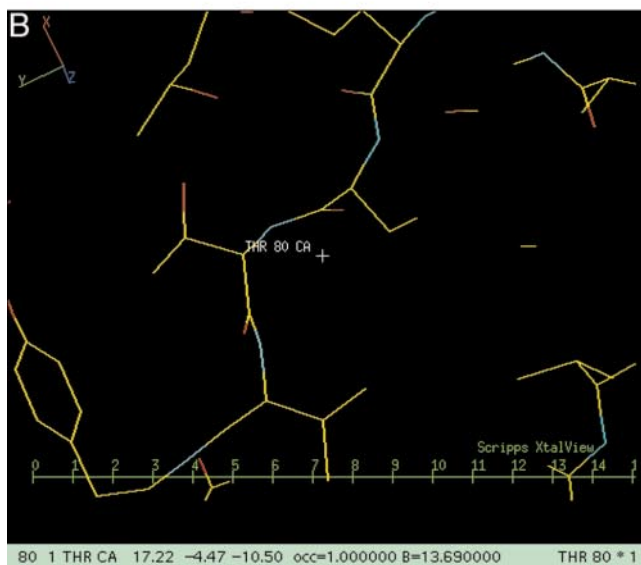
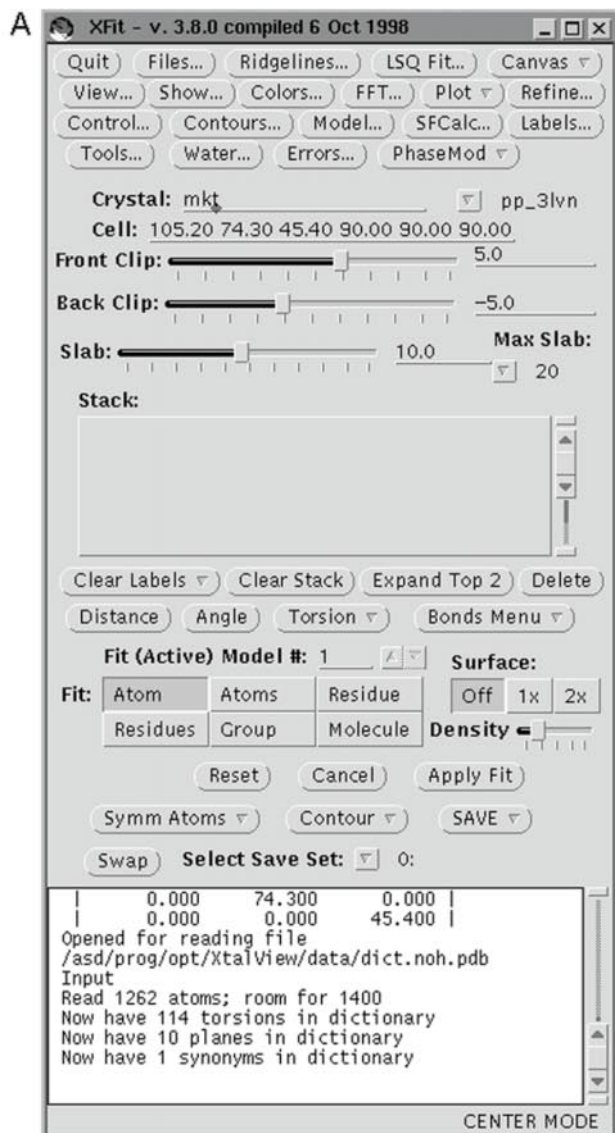
*Fixing main chain.* One of the trickiest parts of fitting a protein is getting the main chain peptide planes correct with good  $\phi$ - $\psi$  torsion angles. In Xfit, one strategy is to use the positions of the  $C_\alpha$  and  $C_\beta$  atoms to define the main-chain geometry rather than trying to move the peptide plane itself. Pentameric poly-Ala fragments (Jones *et al.*, 1991) are least-squares fit over the part to be fixed and the

geometry of the main chain is replaced with the new geometry in this local region.

**Combining phases.** There are times when it is desirable to combine two maps together by combining the phases, for example, combining MIR phases with the latest refined model phases. In general, these maps can be very powerful for overcoming phase bias in the model phases and filtering noise in the MIR map, so that the combined map is better than either alone. This can be done in Xfit by using the PhaseMod/Phase Combination command. The relative weights of the two data sets can be altered with the sliders. It is often useful to down-weight the model phases to about 50% to allow more MIR information to come through. The program uses  $\sigma_A$  weighting (Read, 1986) to calculate figure-of-merits for phase sets with  $F_o$  and  $F_c$  and no figure-of-merit.

**Improving phases.** Phases can often be improved by solvent flattening and/or histogram matching. Within Xfit, phases can be modified with the PhaseMod/Phase Improvement command. The first step is the calculation of a solvent mask. After the mask is calculated, Xfit will load and display the mask over the protein density. The percentage solvent can be adjusted if the mask seems too tight or loose to the protein density. In the second step, the mask is applied to the map to filter the phases by flattening the solvent and adjusting the histogram of the density. Xfit displays the results as each step proceeds. The methods used in the solvent flattening are too extensive for this paper but are similar to the methods used by the program DM (Cowtan and Main, 1998). Since the calculation is done outside the program, the solvent mask calculation and/or phase cycling can be performed by other programs by providing Xfit with a suitable command script. Over the years we have found that users can make more intelligent decisions of the phase modification parameters if they are easily visualized. In particular the choice of percentage solvent and integration radius can make a large difference in the quality of the output phases.

**FIG. 2.** Xfit windows. (A) The Xfit main window. At the top are three rows of buttons for bringing up the program's main pop-up windows. Below that is the crystal information. The sliders are used to control the clipping planes. In the center is a list with the atom stack, below which are two rows of buttons with a number of common atom stack operations. The basic fitting operations fill the rest of the space on the panel. Finally, at the bottom is the message window where information is echoed to the user. In the footer at the bottom the middle mouse mode is indicated. (B) The Xfit canvas. In the upper right is the gnomon indicating the direction of the model axes. A white cross indicates the center of the screen and the ruler at the bottom is used to measure distances. In the footer at the bottom, information about the last picked atom is displayed.



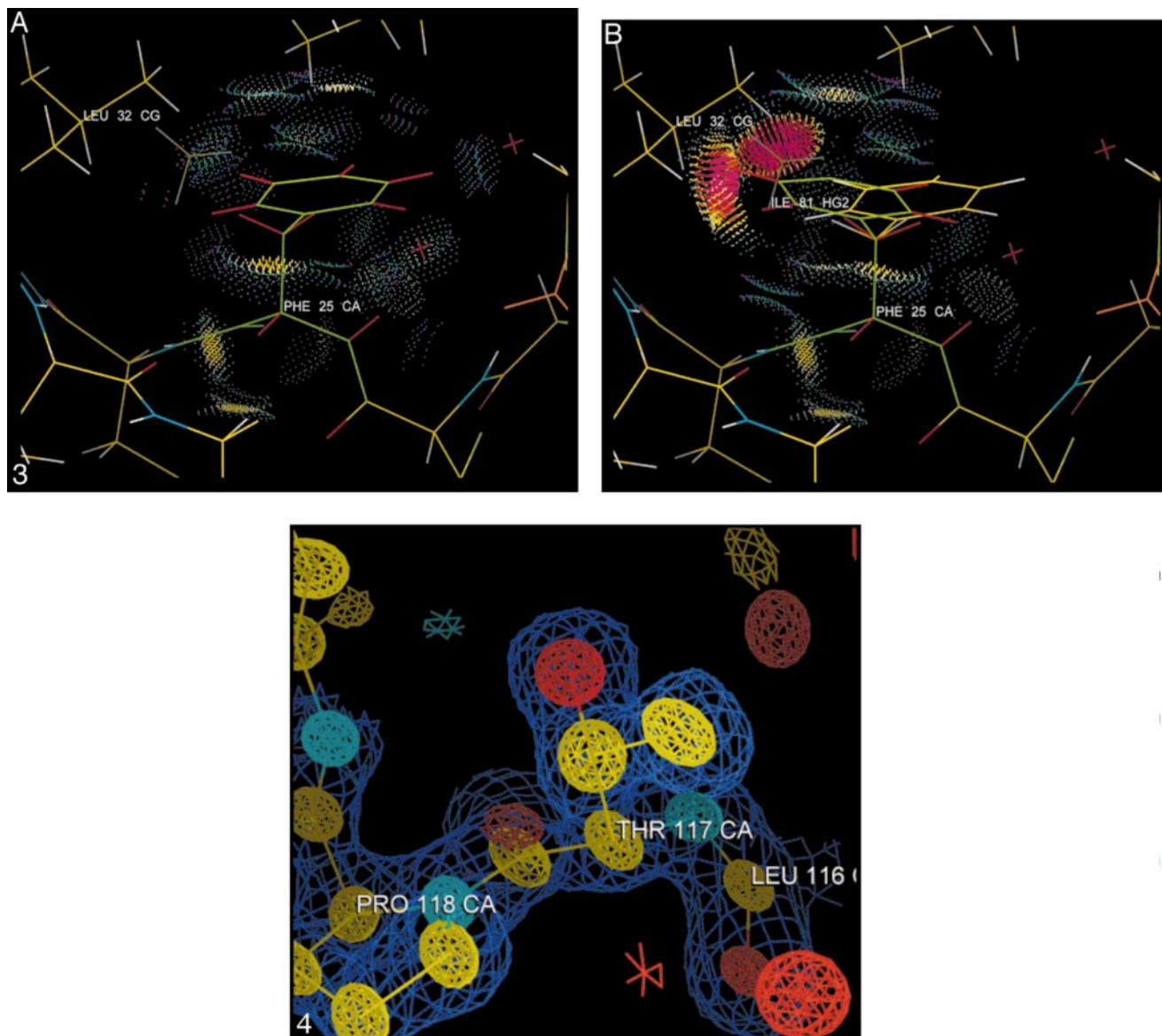
*Cut and paste.* One particularly powerful feature of the program borrowed from word processing is the ability to cut and paste or copy residues between and within a model.

*Adding a prosthetic group or ligand.* Xfit has a very general method of adding ligands and prosthetic groups. Anything that can be specified in a PDB format file, which could, for example, be downloaded from the PDB Web site or generated in a

chemistry sketching program, can be inserted into the model. The new group can be fit just like any residue within Xfit including twisting of bonds with the torsion functions.

#### SEMIAUTOMATED FITTING

Semiautomated fitting allows for rapidly building a protein or DNA model with the computer doing the work and the user making the critical decisions. We



**FIG. 3.** The small-probe contact dot surface being displayed as residue Phe 25 is fit and is indicated by the active atoms displayed in green. The rest of the model is yellow for carbon atoms, cyan for nitrogen atoms, red for oxygens, and white for hydrogens. (A) The residue is shown in its well-packed starting position as indicated by the blue and green dots with some minor interpenetration indicated by the yellow dots. (B) As the residue is rotated about the  $\chi_1$  torsion, the model comes into a severe steric clash with Leu 32 as indicated by the long red and magenta lines showing the included volume between the surface of the overlapping atoms.

**FIG. 4.** A postscript plot from Xfit showing ellipsoids calculated from anisotropic thermal parameters of a high-resolution, 1.35-Å model of 7-Fe ferredoxin, superimposed over the model (sticks) and the electron density (blue lines). The labels have been rendered with the drop-shadow option to improve legibility.

have used this method for several years and find that it greatly decreases the amount of time spent in building an initial model. This algorithm is done in three steps. (1) The main chain is traced by placing a series of  $C_\alpha$  markers using the density as a guide. (2) The  $C_\alpha$  trace is turned into a polyaniline backbone by fitting overlapping fragments. (3) The alanines are replaced with sequence and the side chain positions are found by a rotamer search.

**Backbone trace.** The program finds the position of the next residue in the backbone by considering the map density and the direction between the previous two residues. The scoring function maximizes the continuity and height of the density between the two points. The program tries to eliminate strong side-chain paths by considering a second move and adding the two probabilities together to find the final score. Thus side chains are eliminated because they will not provide a good position for the second move. Paths are sorted by the scoring function and then chosen such that they are at least  $30^\circ$  apart. Paths that form an angle less than  $100^\circ$  with the previous two residue markers are also eliminated. The top five choices are kept and the top choice is displayed. The user can easily switch among the five choices for viewing. The algorithm is fast enough that the original density can be used without any simplification and loss of information, due to skeletonization of the density.

**Fragments.** In the best scenario, the entire main chain can be found with just one fragment. However, this is rarely the case. Therefore, the program needs to work with fragments—connected segments of amino or nucleic acids. The program finds and numbers the fragments by distance criteria. Commands can then work at the fragment level by having the user click on just one atom in the fragment. Since the fragments can be added arbitrarily, they will appear in the model window in no particular order. In general, this is no problem and is taken care of later when the sequence is assigned. However, the fragments can be put into order by renumbering and then sorting them. If a chain is backwards in the map, it is a simple matter to reverse it.

**Poly-Ala backbone.** After the trace is complete, the fragment is then turned into a backbone of polyanilines using the method of overlapping pentamer fragments from a database of previously solved structures (Jones *et al.*, 1991) starting from the N-terminal end of the fragment and moving three residues at a time to the C-terminal end. It is best to get the trace fairly complete before this step, as things that are easy to fix at the  $C_\alpha$  link stage become more difficult once a complete peptide backbone is in place.

**Sequence.** Once the model has been turned into a polyaniline backbone, the program needs the user to identify one residue in the backbone that is in the sequence. The program can then build the rest of the connected chain. If it is difficult to find a match, it is a simple matter just try some guesses and see how they work. It is very easy in Xfit to change the sequence around so that different possibilities can be examined against the density.

**Results.** The first protein traced using this method was ADP ribosyl cyclase (Prasad *et al.*, 1996). The initial chain trace was finished in 3–4 h. Many protein structures have been worked on since then using the method at TSRI and elsewhere and have been the basis for numerous small improvements. Recently, we added the ability to use the backbone tracing method for nucleic acid structures for the solution of an 82-nucleotide RNA–DNA complex, the largest nucleic acid structure solved to high resolution (Nowakowski *et al.*, 1999).

#### REAL-SPACE REFINEMENT

Xfit has a number of options for fitting models with real-space refinement as either a rigid-group or about torsion angles. The unique B-spline density map description of Xfit can greatly improve the fit by providing accurate interpolations between grid points. The refinement is implemented with a complete search of a small local area while evaluating the  $R$ -density function to avoid getting stuck in local minima. This also makes the refinement very general as the search can be iterated over any model geometry parameter. The Rotomer search is particularly useful for placing side chains. It takes the current atoms and systematically rotates them through all of the  $\chi$  torsion angles and puts the side chain in the position with the best fit to the density. The  $R$ -density function options control the trade-off between accuracy and search time. Of the three options, the atom centers option is the fastest, least accurate mode and works by summing the density at the atom positions. The other two options, map correlation and map difference, build the radial electron density around each atom using form factors, atom  $B$ -values, and map resolution. Since this is essentially the same as the first step of the calculation from model to structure factors, the program can take unique advantage of the built-in structure factor calculating routines. The density is then recalculated at each search position and compared to the map after the density for atoms that are fixed to prevent steric overlaps is subtracted. Since the geometry of the model may change in each search position, Xfit does the calculation on the same grid as the map and completely recalculates the density at each step in the search. The correlation function

then maximizes  $\rho_{\text{obs}} * \rho_{\text{calc}}$  while the difference function minimizes the difference  $|\rho_{\text{obs}} - \rho_{\text{calc}}|$ . We have found that for large fragments, where only the atom centers option can be calculated in a reasonable amount of time, the large number of atom centers has a compensating effect and the resulting fit is very good. For small fragments where the computational time is small for the correlation and difference options, it makes sense to use these options as they can make a significant difference in the quality of the fit.

### GEOMETRY REFINEMENT

Xfit refines model geometry by minimizing the error in bond distances, bond angles, planes, and torsion angles against a dictionary of ideal model geometry. Bond distances are refined by minimizing the standard least-squares function,  $(d - d_{\text{ideal}})^2 / \sigma_{\text{bond}}^2$ , where  $d$  is the bond length,  $d_{\text{ideal}}$  is the ideal value of the bond length, and  $\sigma_{\text{bond}}$  is the weighting factor for bond lengths. Ideal bond lengths are extracted from the dictionary, which consists of an ideal example of a residue in PDB file format. (It is easy to add to the dictionary by reading in new PDB coordinates.) The program finds the bond with the same atom names and residue type in the dictionary and sets the ideal length to this value. The  $\sigma$ , or weighting factor, defaults to a value that can be changed by the user by moving a slider. Angles are treated as 1,3 distances and handled in the exact same manner as bonds except that they have their own  $\sigma$ . Three dictionaries are supplied with the program, one with no hydrogens, one with polar hydrogens only, and one with full hydrogens. Each dictionary consists of idealized models of each residue using the Engh and Huber (1991) parameters for both protein and DNA and for some prosthetic groups.

The main chain torsion angles,  $\phi$ ,  $\psi$ , and  $\omega$ , are refined as a special case.  $\phi$  and  $\psi$  are refined simultaneously against a 2-dimensional  $\phi$ - $\psi$  table with the  $\phi$ - $\psi$  landscape specified in arbitrary units. The current  $\phi$ - $\psi$  is looked up in the table and the direction in which it should move to increase the  $\phi$ - $\psi$  goodness is determined.  $\phi$  and  $\psi$  are then adjusted in that direction a few degrees. In this way residues that are close to  $\alpha$ -helical tend to stay  $\alpha$ -helical, and so forth.  $\omega$  is used to keep the peptide plane planar. A weak restraint is also applied to the protein side-chain  $\chi$ 1 angles in the preferred positions.

Since the parameters to be refined are interacting, and each atom can have several geometry parameters acting on it simultaneously, the program minimizes the values in several iterations. The standard diagonal least-squares method is used whereby each shift of the atom positions is calculated and a

weighted sum is kept of all of the bond, angle, plane, and torsion shifts as they are looped through. Large shifts above 0.2 Å are damped out. The atom is then moved according to the summed shift and the next cycle is calculated.

The user can add arbitrary restraints to the geometry. For instance, the ends of a fragment can be restrained to stay in the same place; the  $C_{\alpha}$  atoms of a fragment can be restrained to a small volume to prevent large movements of a fragment with initial bad geometry, and an atom can be restrained to a point in space. Thus, one strategy for moving a model into density is to add a few restraints from the model to points in the density and then running the refinement until the model is pulled into the density.

### CONFORMERS

Each protein residue with a rotatable side chain is present in several conformers in the Xfit dictionaries. The conformers of a residue can be quickly cycled through by clicking on the residue and then pressing the "c" key repeatedly, thus allowing the user to quickly evaluate each one. The lowest energy conformer is the first one listed and is the default for inserted residues. It is possible for a user to alter the conformers in the dictionary or to supply new ones by loading the dictionary as a model and fitting it. In fact, the current dictionaries were supplied by a user, Jean-Luc Pellequer, who was dissatisfied with the earlier, more limited dictionaries. With these new dictionaries, XFit provides a more comprehensive and correct distribution of conformers than in most other molecular modeling programs.

### DOT SURFACES

Dot surfaces of three types can be displayed around the current atoms: van der Waals, twice van der Waals, and small-probe contact dot surfaces. As the model is moved, the surfaces are updated in real time. The density of the dot surface is on a slider and can be increased for a more completely filled surface or reduced to speed the surface recalculation. The small-probe contact dot surface is especially useful for finding collisions between the current atoms and other atoms in the model (Word *et al.*, 1999a). The surface is calculated by a plug-in, called PROBE, written by J. Michael Word in Dave and Jane Richardson's group at Duke University (<ftp://kinemage.biochem.duke.edu>). PROBE rolls a very small probe of 0–0.25 Å on the surface of the atoms and then displays dots where the surface of the current atoms and the rest of the model join (Fig. 3A). If the surfaces intersect, indicating a collision, lines are drawn at the intersection and colored to indicate the severity of the contact (Fig. 3B). Since the surface is calculated in real time, the collisions are shown as

the user fits and provide an excellent indicator of proper packing. The Richardson group's work indicates that hydrogens are essential for a complete packing analysis (Word *et al.*, 1999b). For this reason, J. Michael Word has also supplied a program for adding hydrogens to the model called REDUCE (<ftp://kinemage.biochem.duke.edu>).

#### VIEWING THERMAL PARAMETERS

The thermal parameters of a model can provide important information about the level of disorder in a model. We have added the capability of viewing anisotropic thermal parameters as probability surfaces to Xfit, similar to the manner popularized by ORTEP (Burnett and Johnson, 1996). The anisotropic thermal parameters of a residue with the concomitant density is shown in Fig. 4. The user can set the probability with a slider in the View window and the ellipses generated can be rotated in real time, which gives a real advantage over the static pictures generated by ORTEP.

#### FIGURES AND PLOTS

The contents of the screen can be sent to a PostScript printer from the Postscript Plot window or saved into a file. With the help of Ethan Merritt at the University of Washington, we have developed a Raster3D interface (Merritt and Bacon, 1997). With this option, pictures can be made that are composed of lighted cylinders and spheres of cover quality. In addition,  $\phi$ - $\psi$  and  $B$ -value plots can be made with the View window. Xfit can also be used to find the viewpoint for the popular graphics program Molscript (Kraulis, 1991). This viewpoint can be cut and pasted into a Molscript command file after the sign of the translation is changed, which is opposite from the one Xfit uses.

#### SPLIT RESIDUES

With the advent of freezing as a very common technique in protein crystallography, it is becoming increasingly common to find residues "frozen out" into two or more conformations in the crystal. We have enhanced Xfit to make it simpler to split a residue into multiple conformers and to handle these according to the PDB guidelines. In fitting a split residue, either the whole residue can be fit as a unit or either split half can be fit separately, depending upon whether the unsplit portion of the residue is clicked on or one side of the split is selected.

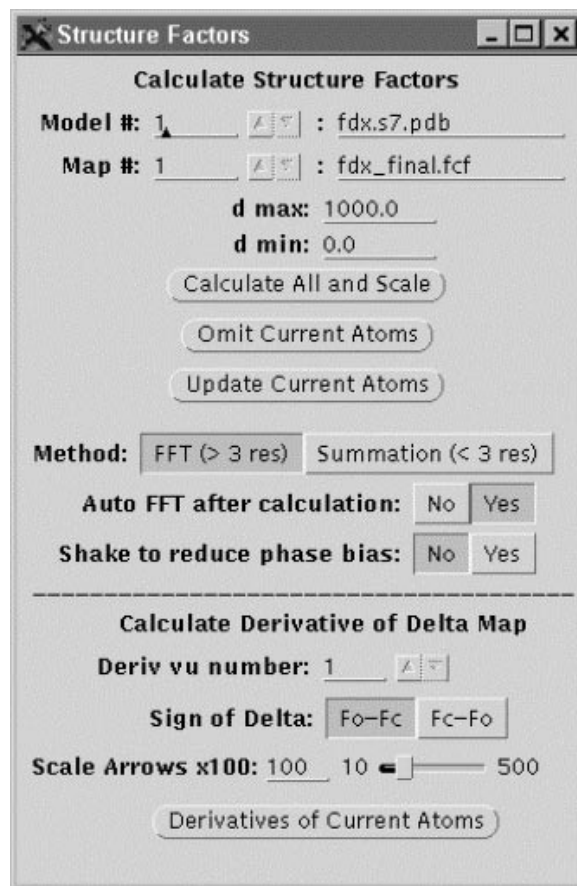
#### STRUCTURE FACTOR CALCULATIONS

A unique feature of Xfit as a graphics program is the ability to interactively calculate structure factors with a fast Fourier transform given a reflection list

and a model (Fig. 5). Advantages of calculating the structure factors internally are that they can be correctly scaled, they have the ability to calculate OMIT maps, and they have the ability to update the structure factors to reflect changes in the model. The Shake option in Xfit can be used to reduce phase bias by adding a small number to each atom coordinate ( $\pm 1/6 d_{\min}$ ) before the FFT (but does not change the model) in order to remove the correlation in atom positions forced by the least-squares algorithm in refinement programs.

#### FINDING GEOMETRY ERRORS

The Error window is used to find geometry errors in the model. The geometry is analyzed and the results put into the error list so the user can click through it. The  $\phi$ - $\psi$  plot is also calculated and displayed. Residues with bad  $\phi$ - $\psi$  values are marked on the  $\phi$ - $\psi$  plot (and put in the error list). The user



**FIG. 5.** A unique feature of Xfit is the SFCalc window, which is used to calculate structure factors from the model (top portion of the window). In the bottom portion of the window are functions used to calculate lines indicating the direction and magnitude of shift at each atom position according to the first derivative of the electron density.



can click on each error to move through the model and repair the geometry.

### WORKING WITH WATERS

Adding waters to a protein map can be a time-consuming step. Simple algorithms for adding waters by placing a water into peaks in a difference map are unsatisfactory in that they build water models with poor geometry and do not take into account knowledge of how waters are usually found in proteins.

*Automated water addition.* In Xfit, waters are added in shells around the protein, taking into account the protein, the electron density, and the crystallographic symmetry. The program finds all the peaks above a threshold in the map, sorts them from high to low, and then goes through the list. If the peak is near enough to an oxygen or nitrogen to hydrogen bond, and it is not too far away from the protein, then a water is added and its coordinates are refined. All the symmetry positions are considered and the water is added at the position that gives the shortest distance to the protein. Repeating this step adds a second shell of water hydrogen bonded to the first. The new waters are added to the Error list so that the user can quickly navigate through the list of added waters to check them.

*Manual water placement.* Manually adding waters in Xfit is extremely fast using shortcut keys. Waters can be added at any position by moving the cursor and typing "W." To delete a water all that is needed is to pick it and type "D."

### SCRIPTS AND REMOTE CONTROL OF Xfit

Another unique feature of Xfit is the ability to use it as a crystallographically intelligent viewer for other programs. Xfit periodically checks a script file to see if it has any new commands and executes them. Thus commands can be inserted into the script file that can be used, for example, to reload a model that has been changed by a molecular dynamics program in order to display the progress of a long dynamics run. In this example, after a certain number of steps, the dynamics program could write the coordinates to a PDB file, and send a loadmodel command to Xfit. A phasing program can append a command to the script file to display the map for alternate phase sets. Coupling this scripting capability with a Web page will allow time-consuming and memory-hungry crystallographic operations to be calculated locally by a Web browser running from a remote site. In collaboration with the PDB, we plan to build a Web-based electron density server using Xfit so that users can visualize the structure factor information currently being deposited for about half

of the structures at the PDB. Similarly, we are collaborating with the CCP4 software group in England (CCP4, 1994) to provide viewing options for their new graphical user interface.

### USING Xfit TO PLACE A MODEL INTO EM DENSITY

Xfit has been used by a number of electron microscopy groups for viewing their EM maps and for placing models into the EM density. The chicken-wire contours are very fast to compute and allow for real-time rotation and viewing of very large slabs of electron density. The chief modifications needed to adapt the program for EM use were to greatly increase the range available on the sliders to accommodate the larger scale of EM maps.

### USING XtalVIEW TO LOAD A PDB FILE AND MAP FROM THE WEB

For this operation both the PDB file and the corresponding CIF file from the PDB Web site must be downloaded. Not all entries have structure factors but, increasingly, many do. The CIF file contains the space group and unit cell information so that it is not necessary to enter a crystal file. (With older versions of Xfit prior to version 4, it is necessary to create a crystal file with the information on the CRYST card of the PDB file.)

### OBTAINING XtalVIEW

Xfit is a module of XtalView and is distributed by the Computational Center for Macromolecular Structures (CCMS) located at the San Diego Supercomputer Center and is extremely easy to install. CCMS, funded by the National Science Foundation, is dedicated to providing quality software with professional help to the structural biology community. To obtain XtalView, send an e-mail to the CCMS listserver to get the current download instructions:

E-mail: [ccms-request@sdsc.edu](mailto:ccms-request@sdsc.edu)

Subject:

Message: get xtalview

Help for XtalView can be reached via e-mail by sending a message to [ccms-help@sdsc.edu](mailto:ccms-help@sdsc.edu). You may also look at the Web sites <http://www.sdsc.edu/CCMS> and <http://www.scripps.edu/pub/dem-web> for late-breaking information and FAQs. Commercial users should contact the author at [dem@scripps.edu](mailto:dem@scripps.edu).

Distribution and improvements to XtalView are supported by NSF Grant BIR-9631436. Angela Loh at Digital generously donated equipment and software to aid in the DEC conversion and maintenance of the programs. Lynn Ten Eyck deserves credit for providing the FFT routines and for many of the ideas in the program. I thank Pamela Williams, Mark Israel, Ken Steube, and Jane Kwon for a tremendous amount of help to users, Mike Pique for much of the early computer graphics work, Alex Shah for many additions to the graphics code, Jeff Mandell for work on the

distribution, Erik Nelson for the B-spline code, Ethan Merritt for the Raster3D interface, Jean-Luc Pellequer for building the dictionaries, Marieke Thayer, Rashmi Nunn, John Tainer, Elizabeth Getzoff, Dave Stout, Chris Bruns, Enrico Stura, Zhong Ren, Gary Gippert, Paul Swepston, John Irwin, and many others for numerous suggestions for improvements to the programs. I also thank J. Michael Word and Dave and Jane Richardson for use of their PROBE software as an Xfit plug-in.

## REFERENCES

- Abola, E. E., Sussman, J. L., Prilusky, J., and Manning, N. O. (1997) Protein data bank archives of three-dimensional macromolecular structures, *Methods Enzymol.* **277**, 566–571.
- Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) Crystallographic R-factor refinement by molecular dynamics, *Science* **235**, 458.
- Burnett, M. N., and Johnson, C. K. (1996) ORTEP-III: A Thermal Ellipsoid Plot Program for Crystal Structure Illustrations, Oak Ridge National Laboratory Publication 6895.
- CCP4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D* **50**, 760–763.
- Cowtan, K. D., and Main, P. (1998) Miscellaneous algorithms for density modification, *Acta Crystallogr. Sect. D* **54**, 487–493.
- Engh, R. A., and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement, *Acta Crystallogr. Sect. D* **47**, 392–400.
- Ghadiri, M. R., Granja, J. R., Milligan, R. A., McRee, D. E., and Khazonovich, N. (1993) Self-assembling organic nanotubes, *Nature* **366**, 324–27.
- Grosse, E. (1981) Approximation and Optimization of Electron Density Maps, Ph.D. thesis, Computer Science Department, Stanford University.
- Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building models in electron density maps and the location of errors in these models, *Acta Crystallogr. Sect. A* **47**, 110–119.
- Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr.* **24**, 946–950.
- McRee, D. E. (1992) A visual protein crystallographic software system for X11/Xview, *J. Mol. Graphics* **10**, 44–46.
- McRee, D. E. (1993) Practical Protein Crystallography, Academic Press, San Diego.
- Merritt, E. A., and Bacon, D. J. (1997) Raster 3D—Photorealistic molecular graphics, *Methods Enzymol.* **277**, 505–524.
- Nowakowski, J., Shim, P. J., Prasad, G. S., Stout, C. D., and Joyce, G. F. Crystal structure of an 82-nucleotide RNA–DNA complex formed by the 10-23 DNA enzyme, *Nat. Struct. Biol.* **6**, 151–156.
- Prasad, S., McRee, D. E., Stura, E. A., Levitt, D. G., Lee, H. C., and Stout, C. D. (1996) Crystal structure of *Aplysia* ADP ribosyl cyclase, a homologue of the bifunctional ectozyme CD38, *Nat. Struct. Biol.* **3**, 957–964.
- Read, R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors, *Acta Crystallogr. Sect. A* **42**, 140–149.
- Ren, Z., Meyer, T., and McRee, D. E. (1993) Atomic structure of a cytochrome c' with an unusual ligand-controlled dimer dissociation at 1.8Å resolution, *J. Mol. Biol.* **234**, 433–45.
- Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999a) Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogens, *J. Mol. Biol.* **285**, 1709–1731.
- Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999b) Asparagine and glutamine: Using hydrogen atom contacts to choose side chain amide orientation and obtain distinct rotamer populations, *J. Mol. Biol.* **285**, 1733–1745.