XtalView, SHELX-97 and LINUX

SHELX Workshop, ACA Meeting, 2000

Duncan McRee, The Scripps Research Institute, La Jolla, California
dem@scripps.edu
XtalView help and questions: ccms-help@sdsc.edu

Contents

XtalView is a GUI software package for protein crystallography. It uses an X-windows interface on a number of UNIX platforms - SGI, SUN, DEC and LINUX are all supported from a single source tree. It is supported and distributed by CCMS at the San Diego Supercomputer Center (http://www.sdsc.edu/CCMS). Version 4 should be available by the time you read this.

XtalView is used for solving structures by MIR or MAD, for de novo model building, fitting between refinement cycles and modeling. Much of XtalView is covered in the book "Practical Protein Crystallography"[1] , McRee and Israel[2], McRee99[3]. The online help is at http://www.sdsc.edu/CCMS/Packages/XTALVIEW/XV1TOC.html.

In this paper I will discuss the use of the XtalView program Xfit for fitting between refinement cycles. Most of what is covered is applicable to all refinement programs but here we will specifically discuss using SHELX-97 with Xfit running on LINUX. However XtalView also runs on SGI's, SUN and DEC.

Obtaining XtalView

XtalView is free to academics and non-profits through CCMS. Send an email with the message to ccms-request@sdsc.edu with the message 'get xtalview'. Complete instructions will be return emailed to you. Others may purchase XtalView by contacting the author at dem@scripps.edu.

Fitting Features

- Point and click interface
- Rapid mouse-based fitting
- Built-in FFT to calculate maps internally
- Structure factor calculator
- On-the-fly omit maps
- Shake (jiggle) anti-bias option
- Extensive model handling facilities
- Automated waters
- Fit while refining mode
- Raster3D interface
- Up to 50 models and 20 maps
- Dynamic memory allocation

SHELX Fitting Features

Atomic-resolution structures are becoming more the norm these days. Several features have been added to XtalView to support SHELX-97.

- Reads SHELX Reflections files (LIST 6 and LIST 5)
- Calculates Sigma-A coefficient maps from Fo, Fc
- Improved split side-chains
- Displays Anisotropic B's (i.e. ORTEP style)

[1] D.E. McRee (1999), Practical Protein Crystallography, 2nd Edition, Academic Press, San Diego.

[2] Israel, M. and McRee, D.E. (2000) "XtalView" in Crystallographic Computing 7, P.E Bourne and K.D. Watenpaugh, Eds., Oxford University Press, http://www.sdsc.edu/projects/Xtal/IUCr/CC/School96/IUCr.html

[3] McRee, D.E. (1999) XtalView/Xfit - A Versatile Program for Manipulating Atomic Coordinates and Electron Density. *J. Structural Biology*, **125**, 156-165.

LINUX Installation

XtalView has been tested with a number of LINUX configurations. If you are installing a new system we recommend a Pentium 100+ and >= 16 Mb memory, preferably 32 Mb and a 3-button mouse. You need to install the xview package option and the 75-dpi font package. If you want to recompile, install the kernel building package for the needed include files as well as the gcc compiler. The X-server should be Xfree-86 3.2+. We've used Red Hat 4.0-6.1 successfully but other flavors can be used if the kernel is 2.0 or higher. In a future release we will support the Crystal Eyes PC which works with PC monitors and uses the same glasses as Crystal Eyes for the SGI. There are no dials for LINUX. We know of a number of LINUX issues that occasionally cause problems. Email ccms-help@sdsc.edu with your problem and we can probably help.

A typical manual fitting session with Xfit

Note: Items that appear on the Xfit windows appear in **bold**.

To make a PDB file for Xfit, use the WPDB 2 command in SHELXL. (Xfit reads and understands the ANISOU card.) To make the map either use the LIST 6 with the newest version of Xfit to create a reflection file or you can use LIST 5 with versions 3.1 or less. You can also use LIST 6 and SHELXPRO with the [X] XtalView option to generate an XtalView format phase file. SHELXPRO can calculate sigma-A coefficients and you then make an Fo map after reading the file in Xfit. The new version of Xfit can calculate sigma-A coefficients on the fly, anisotropically scale the data and add a bulk solvent correction.

Start Xfit.

The first thing to do is to make sure you have the correct crystal chosen. Crystals can be entered using Xtalmgr (before starting xfit).

Crystal: fdx-frozen   ▽ sirhp
Cell: 54.84 54.84 92.64 90.00 90.00 90.00

The crystal contains the unit cell and symmetry information.

To load a model open the **Files** window. Click on the **Windows** button with the right mouse button and select **Files...**

Xfit Tools

Windows ▽   Canvas Tools ▽   Expert Tools ▽   Auto Fit ▽

Files...   Control...
Ridgelines...   Contours...
LSQFit...   Model...
View...   SFCalc...
Show...   Labels...
Colors...   Dials...
FFT...   Water...
Plot   ▷   Canvas   ▷
Refine...
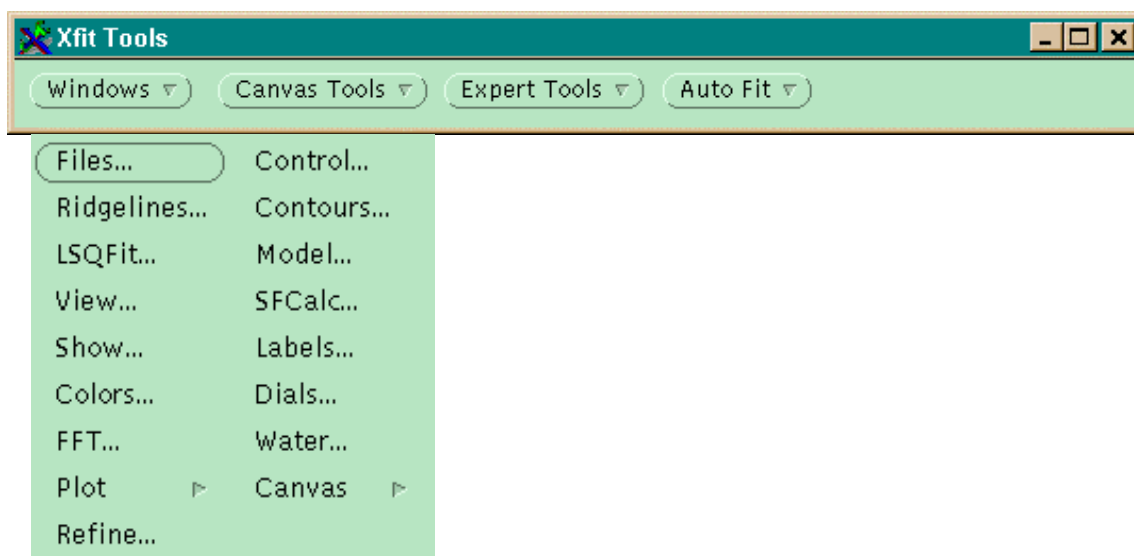
In the Files window you will see a list of all the files in the directory. If you did not start in the correct directory, you can use the directory list, below the file list, to change directories or enter the directory on the **Directory** line and press enter.
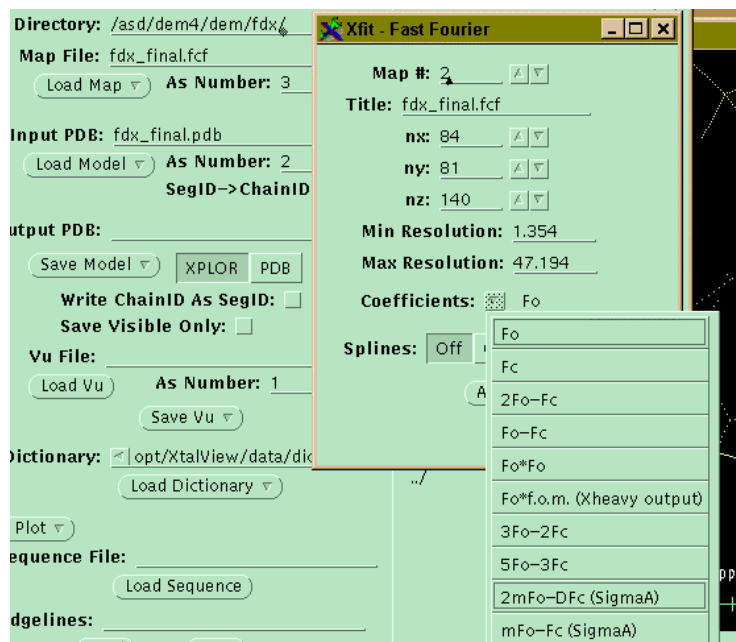
Click on your model in the window and then click on the **Load Model** button. You can load several models at once. The slot the model is loaded into is indicated by the **As Number** field. This number is automatically incremented after you load each model. If you want to overwrite a model then click this button back to the model number you want to load and then click **Load Model**. The program will ask you to confirm the overwrite.

Load the Map

Although there is a binary map format that X-fit reads (the FSFOUR format[4]) it is much more advantageous to take advantage of Xfit's FFT capabilities and read in a phase file and FFT it to a map. It is actually takes less time to do this than to load a binary map from a file. There are also other advantages:

- You can change the resolution at will
- You can change the map coefficients at will
- You can make omit maps at will*
- You can update the phases after changing the model*

[4] There are two map converters available – map2fs by Kevin Cowtan for CCP4 to FSFOUR map conversion and CNS2FSFOUR by Chris Putnam for CNS to FSFOUR map conversions for those times when the coefficients for the map are unavailable or hard to duplicate. See the XtalView web site: http://www.scripps.edu/pub/dem-web.
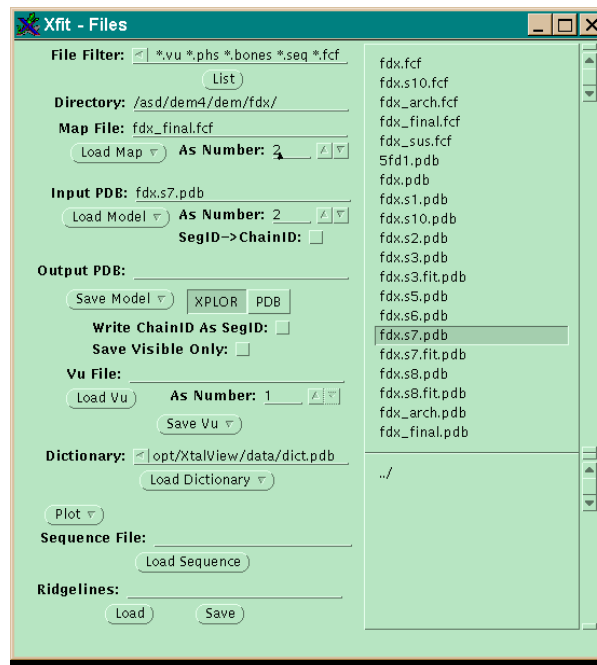
The starred items can only be done if you are using isotropic B's and the resolution is below about 1.5 Å. This is because X-fits uses an FFT that cannot handle anisotropic U's and like all FFT's it becomes inaccurate at very high resolutions. However, at these higher resolutions sigma-A maps usually are all that is needed - if you are still making decisions about where to place side-chains you are almost certainly not at very-high resolution.

Load the map by clicking on its name. (If your Xfit isn't listing .fcf files, add *.fcf to the **File Filter** field and click on the **List** button. Click on the **Load Map** button. After the phases are read in, the FFT window will pop up. Use the right mouse button on the Coefficients menu button and choose the coefficients you want. If the file contains Fo and Fc, I highly recommend using one of the sigma-A map types. At this time a map will appear in the canvas. If you want to change resolution ranges or coefficients later you can do so by bringing up the FFT window again.

Note: If you are one of those people who likes to use both a difference map and a 2Fo-Fc map at the same time, just load the map twice, changing the coefficients each time.

Contouring the Map
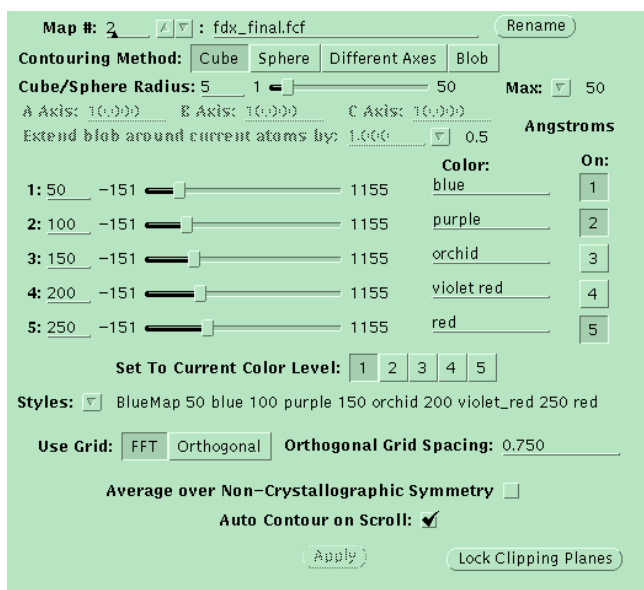
You can change the contour levels and box size in the contouring window. If you are going to inspect your map at the residue level leave the box size at 5 and **Auto Contour On Scroll** on, and as you scroll through the map the program will automatically re-contour.

The first level is set to 1 sigma of the map and the second 2 sigma and so forth. When the map is FFT'd the sigma was scaled to 50. In the messages window the scale was printed so that later you can calculate the equivalent unscaled map values[5].

---
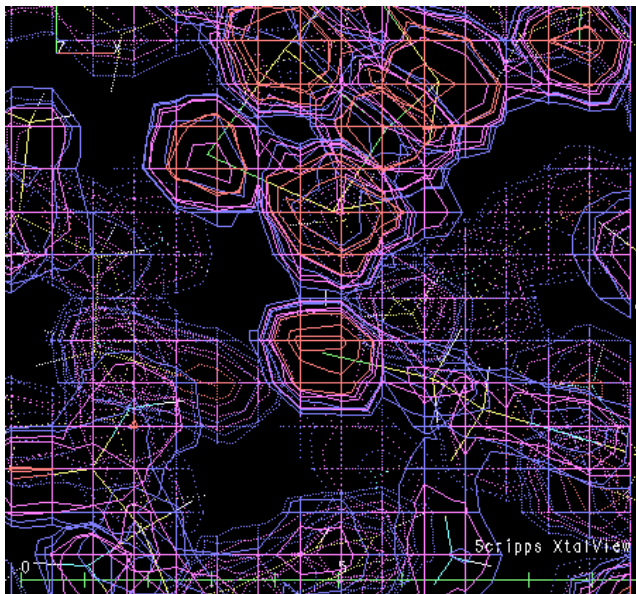
[5] If you use the Xfit structure factor calculator then your data will be placed on an absolute scale. You can get electrons/Åcubed by dividing the contour level by the scale value. Most refinement programs scale Fc to fit Fo, which are usually on an arbitrary scale. Thus the scale for imported Fc data is also arbitrary. To get around this either put the Fo data on an absolute scale, or use the Sfcalc/Calculate All and Scale option to recalculate the scale.

Use the Styles menu button to quickly choose some preset contouring options. The default is the Blue map shown above. Another useful one is Diff Map.

The canvas has a gnomon in the upper left corner that shows the direction of the Cartesian axes. At the bottom is a scale bar. At the lower right is a shameless advertisement. In the center is a white cross showing the rotation center. To rotate the picture use the **Left** mouse button. To re-center use the **Middle** mouse button. To bring up a shortcut menu use the **Right** mouse button. (Note: if you like the Insight mouse conventions you can choose this option in the Control window.) The **Middle** mouse button changes function as you choose different fitting options. The default model color is yellow for carbons with other atoms colored by atom type. You can change the coloring with the Color window. Near the top of the screen the cursor changes to a Z to indicate the z-rotation area where the rotation operations rotate in the plane of the screen and translations are in and out of the screen.
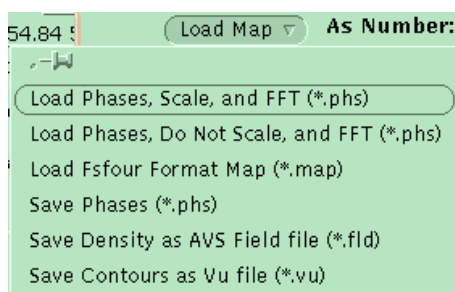
Saving Phases

If you calculate the phases from within Xfit and want to save them for the future or another program, you can use the map file menu button on the Files window. Click the **As Number** field to the number of the map to be saved and then pull down the menu with the right mouse button and choose **Save Phases** (*.phs).

XtalView Tips and Techniques

The shortcut menu

The shortcut or canvas menu is reached by using the right mouse button. It comes up wherever the mouse is positioned in the canvas window. This saves a lot of back-and-forth, back-and-forth mouse movement. The same menu can be reached from the toolbar as the Canvas Menu. The items on the menu (moving down the columns from the left):
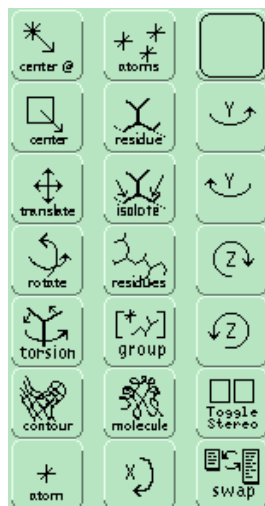
- **Center at** - centers at the last pick
- **Center** middle mouse mode
- **Translate** middle mouse mode
- **Rotate** middle mouse mode
- **Torsion** middle mouse mode
- **Contour** maps
- **Atom** - Make last atom picked fittable
- **Atoms** - Make all atoms on stack fittable
- **Residue** - make the last residue picked fittable
- **Residues** - make all the residues on the stack fittable
- **Group** - make the group containing the last pick fittable
- **Molecule** - make the entire molecule fittable
- **X +90** - rotate plus 90 degrees about the x (horizontal axis)
- **X -90** - rotate minus 90 degrees about the x (horizontal axis)
- **Y +90** - rotate plus 90 degrees about the x (horizontal axis)
- **Y -90** - rotate minus 90 degrees about the x (horizontal axis)
- **Z +90** - rotate plus 90 degrees about the x (horizontal axis)
- **Z -90** - rotate minus 90 degrees about the x (horizontal axis)
- **Toggle stereo** mode
- **Swap** regular menu and expert menu (useful in hardware stereo mode)

Dictionary

Dictionary: ◁ opt/XtalView/data/dict.pdb

(Load Dictionary ▽)
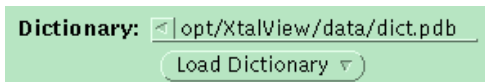
There are two dictionaries supplied with XtalView: $XTALVIEWHOME/data/dict.polar.pdb which has hydrogens on the polar residues (i.e. XPLOR), $XTALVIEWHOME/data/dict.noh.pdb which has no hydrogens and $XTALVIEWHOME/data/dict.full.pdb. The dictionary is a list of residues in PDB format with some extra cards that define planes and torsions. For alternate conformers each residue has several consecutive entries. As you cycle through conformers the next example is used to replace the current residue. If you have a prosthetic group that is not in the dictionary then simply append a PDB file with the prosthetic group you want with the Load Dictionary menu:

Load Dict _ □ ✕
(Append)
Replace
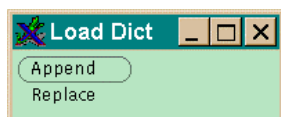
There is also a full hydrogen dictionary supplied. Remember, though, that SHELXL adds hydrogens for you, and this method is the better way to go as SHELXL will generate all the correct entries in the .ins file.

Turning the Map on and Off

Xfit - Show and Hide Objects _ □ ✕

Models:

1 fdx_final.pdb 1810 atoms
–
–
–
–

Maps:

1 fdx_final.fcf 0 lines
2 fdx_final.fcf 10756 lines
–
–
–
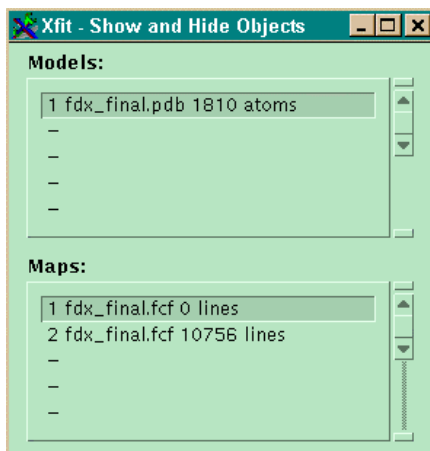
Use the Show window to turn objects on and off. Here I have two maps which I can either alternate between or show at the same time by clicking on the entry.

Label The Residues

Open the Label Window, set it up as shown here and click Label Every. This will facilitate identification during fitting.

Atom Pick Label Style: ▽ 100

(Label Every) **nth Residue:** 1

Go to the N-terminus

To get to the N-terminus click on the chain whose N-terminus want to go to and then press the "[" key. The C-terminus is reached by using the "]" key. (You can also use the Auto-Fit menu if you forget the key.)

Moving along the Chain

To move to the next residue hit the space bar - this pops you to the next residue and the map will re-contour (unless the Auto Contour on Scroll option is off on the Contour window). By hitting the space bar you can quickly inspect every residue in the model.

Fitting A Residue

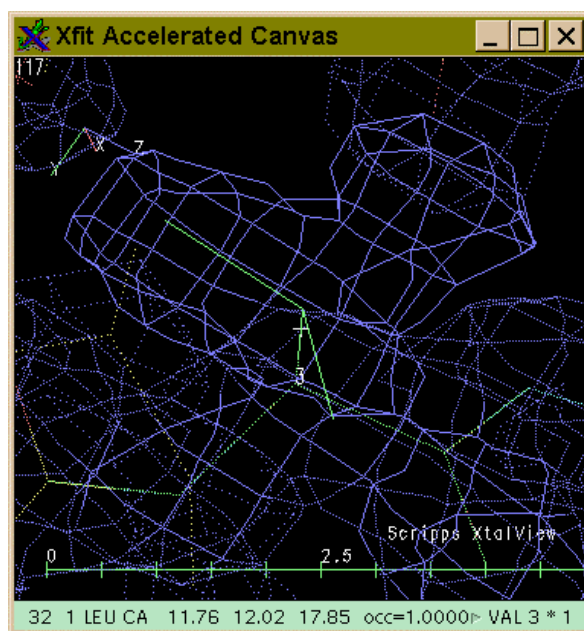In the figure below we see that Val 3 is in the wrong conformer and doesn't match the density. We'll discuss manually moving the residue. This will seem complicated as it is described but after you do it a few times it will be simple.

Practice fitting a residue - don't worry it is easy to undo any changes you make - even later after fitting several other residues.

First click on any atom in Val 3 to place it on the top of the stack. The use the Residue option either on the shortcut menu or on the main window to make the residue fittable. It will turn green to indicate it is in fitting mode. Now you can move it with the mouse. If you bring up the shortcut menu with the right button you can click on either rotate or translate. These connect the middle mouse button to the selected operation. The left mouse will continue to rotate the viewpoint. Select translate and drag with the middle mouse button. The green residue will move - the old, yellow version stays in place so that you know where it was. Move the residue a lot and you will notice that the main-chain bonds will become red to let you know they are ridiculously long. Move it back and they become green again. Move it a long ways for now. Now choose rotate and then rotate the residue with the middle mouse again. Note that the residue rotates around the atom originally picked. (If you need to move the viewpoint center you can choose center on the shortcut menu to restore the middle mouse to its default mode.) To put the residue back at the starting position, choose Reset on the main window.

Torsions

To torsion a bond, pick the atoms at the ends, picking the end that will rotate second.   Then choose torsion from the shortcut menu. Now as you drag the middle mouse horizontally, the bond will twist. For the chi angles of protein side-chains there is a quicker way - just type 1 on the keyboard to set up chi1, 2 to set up chi2 and so forth. Remember that the input focus needs to be in the canvas window or else the keystroke is sent to the window with input focus. By using the number keys to set up the chi angles and dragging with the mouse you can quickly position a side-chain.

When you are finished fitting use either Apply, to keep the changes, or Cancel to return the residue to its starting position, on the main window. The semi-colon ";" key can be used to apply the fitting. The semi-colon key means stop everything and it will also halt the refinement option we will talk about later.

Save

Save your model by opening the files window, and entering a name in the Output PDB field and pressing save. Save often to avoid losing your work. You can also start Xfit with the -auto flag to automatically save the file every time you do an operation that changes it.

Real-space Refinement

While the manual stuff is fun, after a while you just want the computer to do it. With a residue set up in fitting mode, select **RealSpace Search** on the canvas menu. The computer tries all the possible torsions for that residue to get the best fit to the density. If you have several residues being fit at once, the one last clicked on is searched.
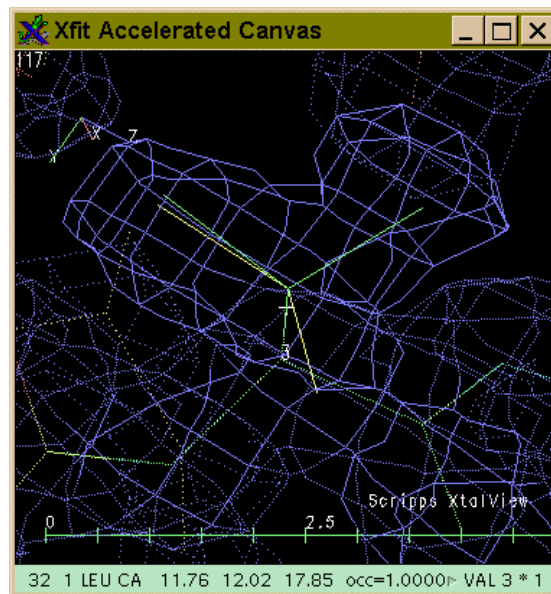
If you have more than one map, set the map to search against on the Refine window:

Geometry refinement

With a residue selected by clicking on it, choose one of the Refinement options on the canvas menu. The residues being refined against the dictionary will be highlighted as a cyan color. **Refine Range** will refine

all the residues between the last two picked **and Refine Region** will do the last picked plus its two neighbors. Pressing the space bar will do another round of refinement.

Ending Refinement

Press the semi-colon ";" key to stop all refinement and fitting. You can undo a refinement with **Refine/Undo Last Refi**.

Refine While Fitting

This is a very powerful way to fit. In this mode the geometry is refined while you drag atoms around, giving the model the feel of being mechanically linked. To do this easily you need xfit version 4 and above (it works in the older xfit versions but it requires more setup). Pick a residue with the mouse. Open the canvas menu and under **Refine** pick **RealSpaceRefi Region**. This sets up the residue clicked on plus the two neighbors to be geometry refined against the map (if you have several maps loaded, the map refined against is set in the Refine window). The refineable residues turn cyan.   Click an atom – it turns green. Drag it with the middle mouse and watch what happens. Click another atom and repeat. Try just wiggling an atom back and forth to do some local annealing. Pressing the space bar will do another round of refinement. Press the spacebar until you are done and then press the semi-colon ";" key to end fitting and refinement.

Viewing thermal parameters

A lot of information can be gleaned from the thermal parameters of the model. You can use the new options on the **View** window to view either isotropic B's or anisotropic U's as 50% probability surfaces. Click on a range of residues and choose the **Pop** button next to **First Res** and **Last Res** (or type in the residue numbers) to set up a range (you can do the whole model but it may be slow). Then choose **Thermal Ellipses** and click on **Make Vu**.
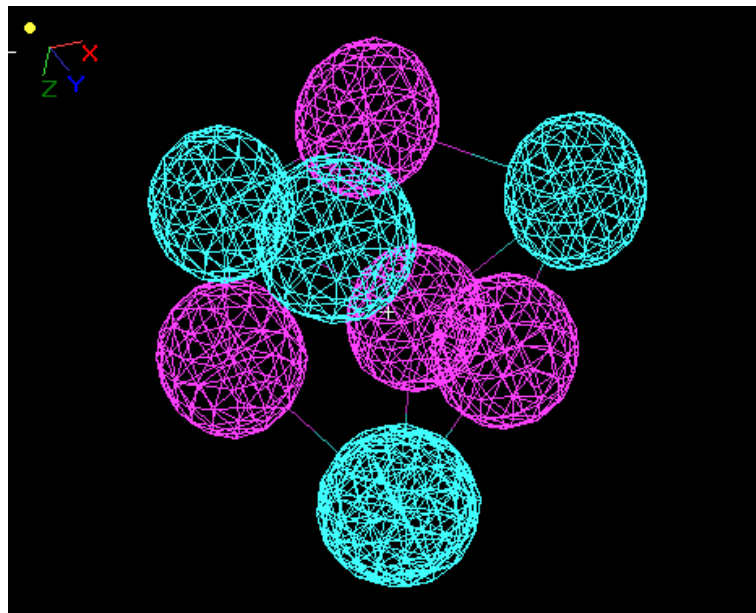
Control Window
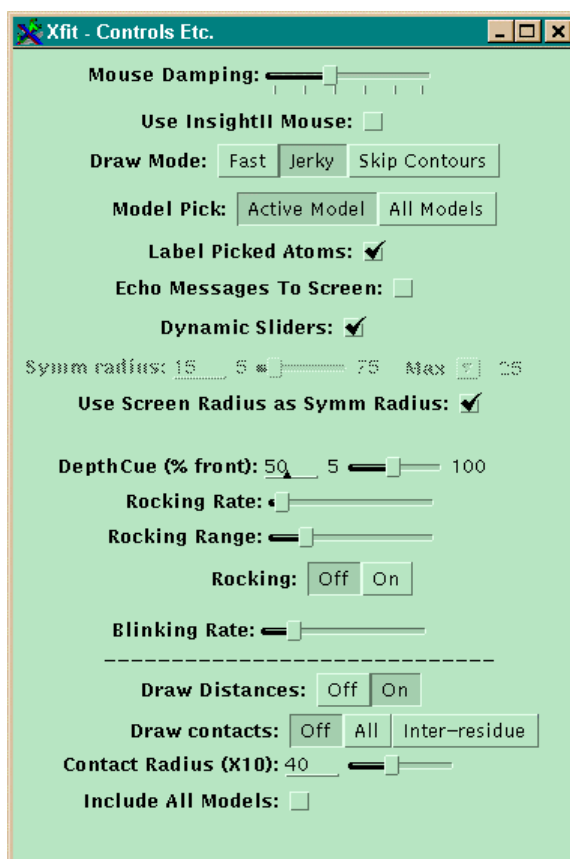
A number of miscellaneous options have ended up on the control window. On slower machines the dynamic sliders may be too slow. In this case switch **Dynamic Sliders** off. Also useful for model building is the **Draw Contact**s option which will show all contacts within the distance specified. These contacts are dynamic and update as you fit residues.

PostScript Plots

The contents of the screen can be sent to a PostScript printer from this window. The PostScript can also be saved into a file. To make a stereo plot first turn on side-by-side stereo from the View window.
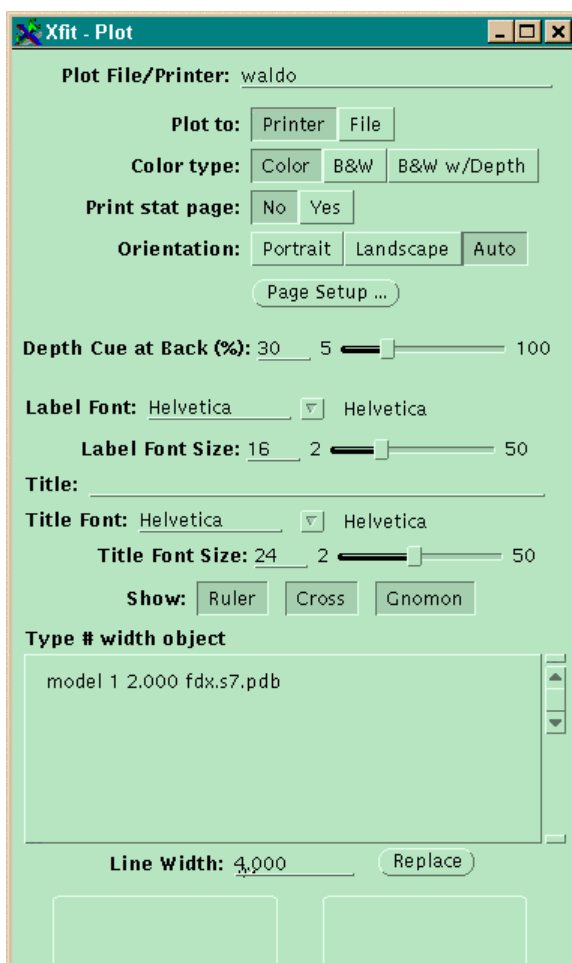
Raster 3D
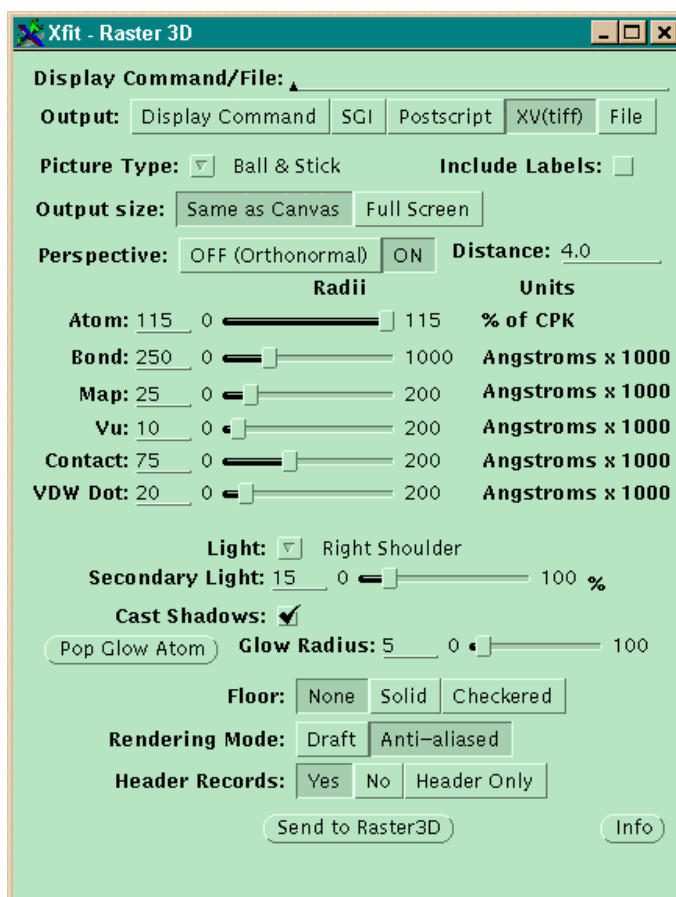
With the help of Ethan Merritt at the University of Washington, we have developed a Raster3D interface. It will be necessary to obtain Raster3D from Ethan via his Web page. With this option you can make pictures composed of lighted cylinders and spheres of cover quality.

Model Window
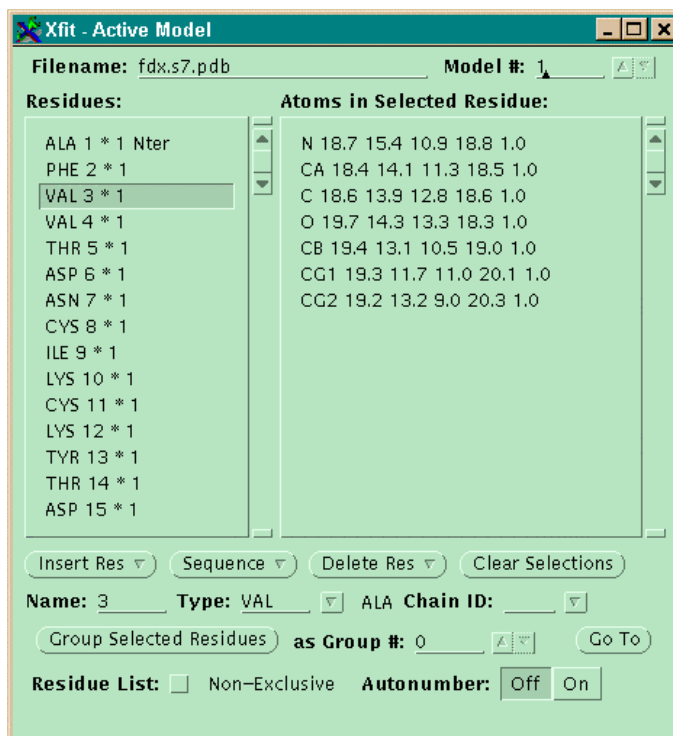
There are a number of options on the model window but the one that is especially important to point out to SHELX users is the **Split Residue** menu item on the **Insert Res** menu button. To use this, first set up a residue to be fit (click on it and choose **residue** on the shortcut menu) and then activate the torsion you want to split the side chain about.

Now choose **Split Residue** and the residue will be divided about the torsion and the two halves numbered A and B and given occupancies of 0.5. One half will become active and you can move it into the density with the middle mouse.

These are the other menus available on the Model window.
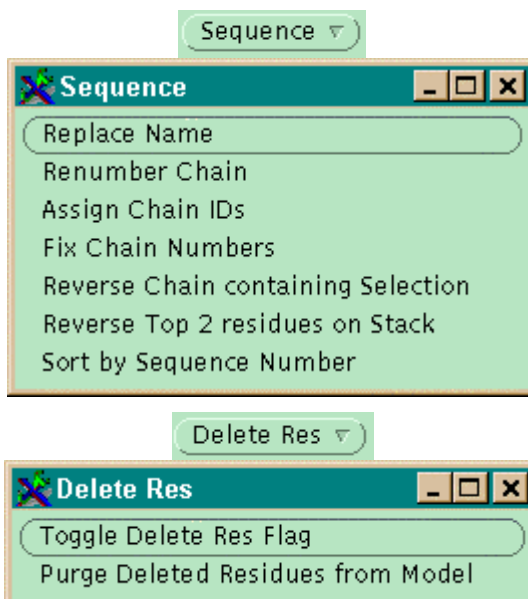
SfCalc window
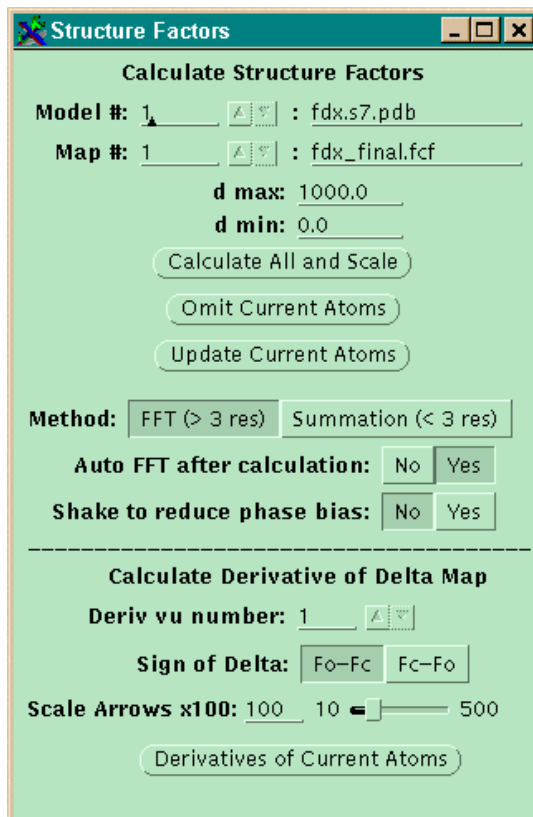
The SfCalc window is used to calculate structure factors from the model (i.e. Fc, phi). At resolutions below 1.5 and with isotropic B models, this window can be used to calculate the structure factors for a model. For other refinement programs we usually keep an empty phase file around and calculate the structure factors from the model and then FFT the map. The advantage of this is that the structure factors will be correctly scaled so that the omit map and update options can be used. To make an omit map you need to do **Calculate All and Scale** once after reading in the phase data. You may want to turn the **Shake** option on as this can help substantially in reducing phase bias. This works by adding a small number to each atom coordinate ($+/-$ 1/6 d_{min}) before the FFT (but does not change the model). The effect of this is to jiggle atoms which have been moved a small amount in order to satisfy the error residual of wrong piece of structure.

To make an omit map you choose one or more residue and make them current (green). Now choose **Omit Current Atoms**. The partial structure factors for the current atoms are calculated and then subtracted from the structure factors of the whole model and the omit map is FFT'd and contoured.

At higher resolutions, above 1.5, and especially if the B's are anisotropic, the program cannot do an accurate structure factor calculation as it uses an isotropic reverse FFT algorithm. However, if you are at the point of making your B's anisotropic and are still wondering where side-chains are located you may want to rethink your fitting strategy!

Semi-automated fitting

Tracing the main chain
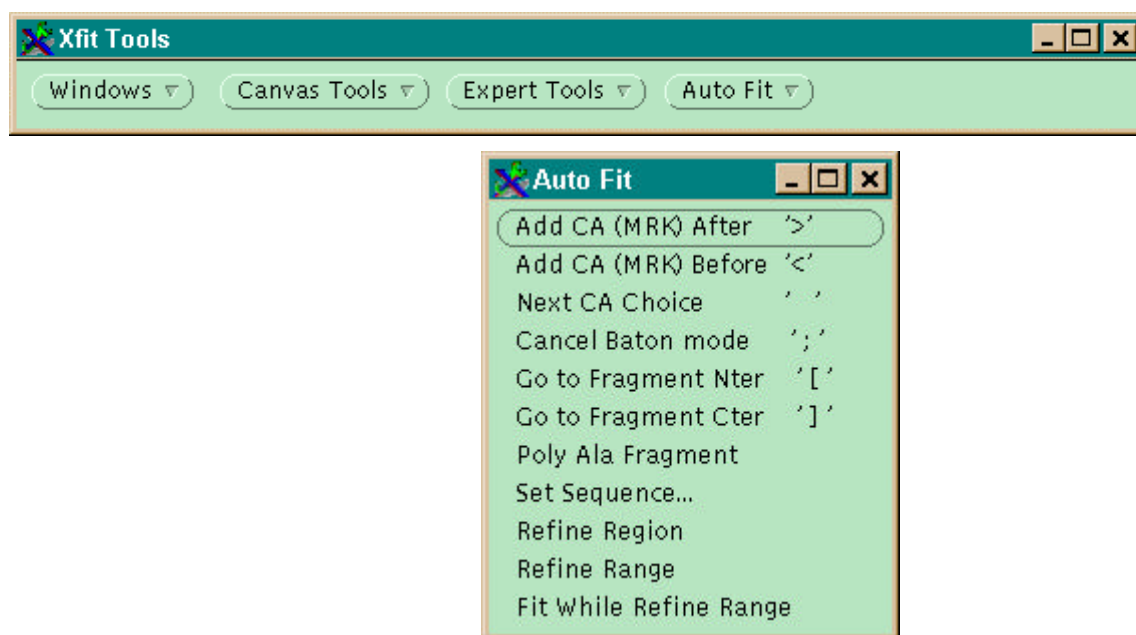
Xfit v3.5+ has a new menu added "Auto Fit" which incorporates features for automated fitting procedures that used to be done one step at a time. The new menu allows very fast fitting in the parts of the map that are clear allowing you to concentrate on the difficult bits without having to doing the repetitive fitting tasks on most of the map. By clear we mean that the main-chain density should be mostly connected at positive contour levels and the side-chains are visible.

Note that some menu items have a character in single quotes that indicate a keyboard shortcut for the operation. After you become accustomed to the keys you will find this much quicker than using the menu. Before we explain how to use the menu we need to discuss some notation.

MRK -. A MRK is a residue with a single CA atom used to mark the positions of the Ca atom of a residue. MRKs are part of the standard xfit dictionary found in $XTALVIEWHOME/data/dict.pdb[6]. If you don't have a dictionary being automatically loaded read the beginning of the xfit man page before continuing. *You must have a dictionary before automated fitting can proceed.*

Fragment/Chain - the basic unit of fitting is a string of residues connected together. The criteria for connectedness for MRK residues (i.e. CA's) is two successive ones be within 4.5 Ångstroms of each other. The **Model** window automatically numbers each fragment and marks the chain termini. The **Sequence** menu button on the Model window has a command **Fix Chain Numbers** that can be used to update the

---

[6] There are three dictionaries available in the distribution. They differ in the treatment of hydrogen atoms. One has no hydrogens, one has polar hydrogens only (i.e. hydrogens involved in hydrogen bonding) and the third has all hydrogens. The default is the polar H dictionary. You can change the dictionary type in the **Model** window or you can have a different dictionary load by inserting this line in .cshrc after you source XtalView.env:

```
setenv XFITDICT $XTALVIEWHOME/data/dict_noh.pdb
```

to load the no hydrogen dictionary. To load the all hydrogen dictionary change dict.noh.pdb to dict.allh.pdb. The polar hydrogen dictionary is dict.polarh.pdb.

chain numbers.

Focus Residue - This is the last picked residue either on the canvas or on the Model Window. It shows up as highlighted in Model window residue list and on the lower right footer of the canvas. All of the commands discussed here start with the focus residue and use it to identify the chain to be acted upon.

Current map - the map used in the auto fit commands is the one enumerated in the Refine window in the map number field. The default is map 1. If another is desired it must be changed here before using the auto fit commands.

Auto Fit Menu Commands

Add CA (MRK) After - Adds a MRK residue to the C-terminal end of the current focus residue (see above). The ">" key is a shortcut for this operation. If there is already a residue here the user is warned and given the opportunity to cancel. If there is more than one residue in the fragment the vector between the last two residues and the density of the current map (see above) is used to find the position of the residue. The 5 most likely positions are saved and can be cycled through by pressing the space bar or by the "Cycle CA" option on the menu. For small adjustments or if none of the positions looks correct, the middle mouse button moves the CA atom in refine while fitting mode keeping the distance 3.8 Ångstroms to the focus residue. The MRK residue is automatically numbered one less than the focus residue. Lastly the command makes the added residue the focus residue so that the next invocation of this command will add another residue to the end of the chain before the new one. In maps with clear main-chain density the command finds the next CA position and no further fitting is needed. Occasionally the side chain density is stronger than the main chain and in this case pressing the space-bar will find the correct choice.

Add CA (MRK) Before - Similar to the Add CA (MRK) After except that it adds to the N-terminal end of the chain/fragment starting at the focus residue. The "<" key is the shortcut.

Next CA Choice - Cycles through the CA choices for the Add MRK commands. The spacebar is a shortcut to this command.

Cancel Baton mode - Accepts the CA position and ends the refinement mode. The ";" key is the keyboard shortcut. It is good practice to always issue this command when finished adding to the ends of a fragment otherwise a later press of the spacebar may have surprising results. Note this command and the ";" key will accept any fitting operation and any refinement mode so it can be used anytime after a fitting or refinement operation t accept the results.

Go to Fragment Nter - moves to the N-terminal end of the fragment/chain containing the focus residue and makes it the new focus residue. Use it as a precursor to the Add MRK Before command.

Go to Fragment Cter - moves to the C-terminal end of the fragment/chain containing the focus residue and makes it the new focus residue. Use it as a precursor to the Add MRK After command.

Poly-Ala Fragment - Turns all the MRK or VEC residues in the fragment/chain containing the focus residue into alanines using overlapping pentamer commands. If the best pentamer contains either a glycine or proline it is kept as such instead of an alinine to prevent inserting alanines in bad positions.

Auto Fit Strategy

Starting a new model - After locating a stretch of density center the position of a CA atom on the cross in the center of the screen and open the **Model** window. Set the active model to an unused number, the **Type** to MRK, enter a likely residue number (e.g. 100) and select **New model** command on the **Insert Res** menu. A new reidue will appear in the center of the screen. Select this residue to make it the focus

residue by clicking on it with the mouse and then use the ">" key to add a second MRK residue. Since there is an ambiguity as to which direction the chain should go, the program leaves it up to you to move it. If you have some indication of the chain direction (see Fig. 3.40 and 3.43) then you can place the residue in that direction. Otherwise, just guess and reverse the chain direction later if needed. To place the CA atom use the middle mouse button and drag the residue until is in the desired position. Use the ">" key again to place the third and subsequent residues automatically.

The program finds the position of the next residue by considering the map density and the direction between the previous two residues. In considering the map density the program tries to eliminate strong side-chain paths by considering a second move and adds the two probabilities together to find the final probability. Thus side-chains are eliminated because they will not provide a good position for the second move. If the program still finds a side-chain press the spacebar to cycle through the other 4 choices. Pressing the spacebar 5 times brings you back to the original choice. You can also move the CA position with the middle mouse button as you did for the second position. In a medium resolution map it is not possible to precisely determine the CA position so instead concentrate on getting the path right and use the pentamer fitting followed by real-space refinement to better position the fragment.

Save Frequently! You never know when there is going to be an accident. Open the **File** window, and enter a filename in the **Output PDB** field and then click on **Save** just below the field. You can also use the **Save** button on the main window to write out the file.

To add residues to the N-terminal end of the fragment hit the "[" key to move to the N-terminus and then type "<" to add residues at that end. When you get to a spot where the map is not clear it is usually best to move on to another spot and fill in another chain fragment. You'll also want to hit the **Symm Atoms** on the main window occasionally to look for symmetry related residues to make suer you don't trace the same density twice. As you add you will find that a fragment has the wrong chain direction. It can be easily reversed by selecting a residue in the fragment and then in the **Model** window issue the **Reverse Chain containing Selection** on the **Sequence** menu button.

Putting fragments into order - In the best case you can find the entire main chain with one fragment. However, this is rarely the case. Since you added the fragments arbitrarily they will be appear in the model window in no particular order. In general this is no problem unless you want two fragments to join and they are won't because they are in the wrong order. To fix this you can renumber one of the fragments. Click on the first residue in the fragment to be renumbered in the Model window, enter the new number in the **Name** field for the start of the chain, and then issue a **Replace Name** command from the **Sequence** menu button followed by a **Renumber Chain**. After you have your fragments numbered as you like, you can issue a **Sort by Sequence Number** command and resort the fragments into ascending order by number. You can also reorder fragments by cut and pasting if for some reason you don't want to change the numbering system you are using. To do this, first select the **Non-Exclusive** check-box. This lets you select more than one residue at a time. Highlight all the residues in the fragment in the list by clicking on them, and then use **Cut selected into buffer** on the **Insert Res** menu button to put the fragment into the residue buffer. Then select the residue at the end of fragment you want to paste in front of and use **Paste buffer after** to put the residues back in the list. You can also **Paste buffer before** by selecting the appropriate residue. Note: the cut and paste commands can be very useful during model building to shuttle residues back and forth between models. By changing the **Model #** field between the cut/copy and the paste you can cut or copy residues between models as well as within models. For instance, you can cut a fragment out of a related protein and save a lot of fitting effort. You can then rename the first residue to the correct sequence number and renumber the chain to put it in the correct sequence order.

Reversing the chain- If you get a chain in reverse order its simple to reverse it. Open the **Model** window, select a residue in the fragment and then issue a **Reverse Chain containing Selection** command from the

Sequence menu button. This command works very well at the MRK stage but later it will all your residues with the N and C terminii facing the wrong way.

After the trace is complete or you get stuck and can't go farther, the fragment is poly-Ala'ed to form a backbone. The poly-ala command builds overlapping pentamer fragments from the N-terminal end of the fragment to the C-terminal end.   It is best to get the trace fairly complete before you do this step. You can't undo it, so save the MRK trace first. Now select a residue and from the **Auto Fit** menu button on the **Toolbar**, issue a Poly Ala Fragment and the program takes over. After it is finished the MRKs will be replaced with a poly-Ala backbone. You are now ready for the next step - assigning sequence.

Assigning the Sequence

Sequence file. Before you can assign the sequence you need a sequence. You can almost always find your sequence on the Web using one of the sequence databases and then cut and paste this into a file using your browser. Give the file the extension .seq and edit the file to check the format. The format xfit wants is as follows. The first line is the title. Subsequent lines should contain the sequence in single-letter in upper case. Numbers and punctuation will be ignored. That's it. Read this file into xfit using the File window. Put the name in the **Sequence File** field and push **Load Sequence**. In later sessions, you can put the sequence file on the xfit command line and as long as it has the .seq extension it will be loaded as a sequence file. Hint: since there isn't anyway to enter in sequence numbers in the sequence file, xfit numbers the sequence starting at the first residue. If you want to change the numbering you can insert some dummy residues (e.g. A) at the beginning of the sequence or in a gap.

Identify sequence. Now you have to find at least one residue that you can identify in the sequence. Sometimes you can identify one by a special position such as binding a prosthetic group. For instance, the histidine ligand to a heme group. Otherwise you need to look for a pattern of residues that you can identify by shape and size. Usually, the easiest to identify are the large residues such as a tryptophan. The core is the most reliable place to assign sequence as surface residues are often disordered or given a haircut by solvent flattening. Look for a repeating pattern and then try to identify that pattern in your sequence. remember the chain might be reversed so try the match in both directions. If you are still stumped or aren't sure about what a residue should look like anyway you can always just try some guesses. Its very easy in xfit to change the sequence around so you can try different possibilities.
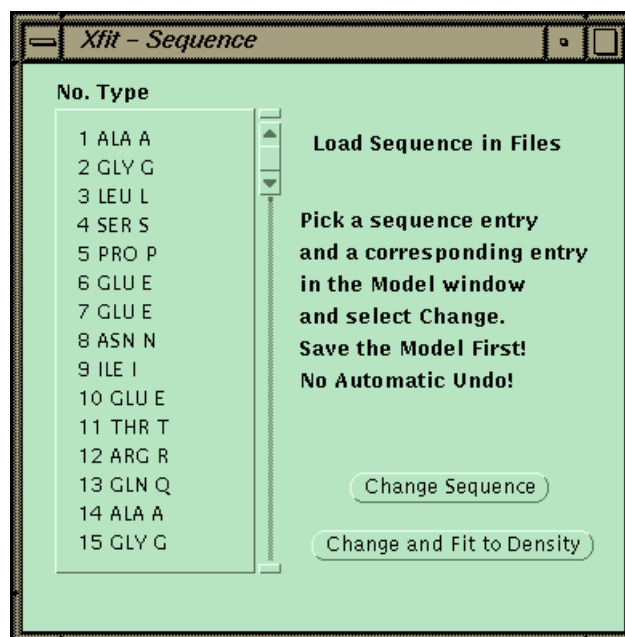
Assign sequence. Having a candidate residue click on it to select it. Open the sequence window using the sequence **Toolbar/Auto Fit/Set Sequence...** and scroll down the sequence to find the corresponding amino acid in the list and click it to select it. Now click on **Change and Fit to Density** and stand back. All the residues in the fragment will be converted according to the sequence and then all the rotamers tried to find the best match to density. You can do this as many times as you like. Just remember to save any models you want to keep. This way you could, for example, try out all the tryptophans in the sequence and then look along the model to see how it matches the sequence.

Editing waters

Two shortcut keys have been added specifically to make editing the water list faster to encourage users to actually look at the waters. Note that in the majority of cases editing the water list and removing bad waters will cause your R-free to drop even if there is a very small increase or no change in the overall R-factor. A bad water is one in which there is no density, the density is not shaped like a water, or the B-value is higher than say 80.0. For high B-value waters with sensible density you can try halving the occupancy.

- **Shift+W** – **Adds** a water where the cursor is and at the center of the slab. The new water is in fitting mode and its position can be adjusted using the middle mouse button. To quickly position the water leave the refine window up in a corner and after the water is added click on the Translation button to real-space refine the water into the center of the peak. Use a thin slab (5 Å) to get the out-of-screen direction close enough for the translation to pull it in. Use the semi-colon key to end the water fitting or go on to the next one and use the semi-colon key at the end of the water adding sequence.

- **Shift+D** – **Deletes** the last picked water/residue. Click on a water and then issue this command. If it's a water the program deletes it. If it's a residue the program asks for confirmation of the delete. You can undelete a water by going to the model window, finding the deleted water and deleting it again to toggle the DEL flag.

MAD Phasing with XtalView

With the proper treatment, xheavy can be used to phase MAD experimental data. No modifications are needed to the program and the method used is similar to that used by others to phase MAD data by

treating it as a special case of MIR phasing. The anomalous differences, between Bijvoet pairs, are treated as native anomalous, and the dispersive differences, between wavelengths, are treated as isomorphous pairs. Care must be taken to treat the member of the dispersive pair with the smaller value of f' as the native while the other is the derivative.

Nomenclature

First some nomenclature. MAD experiments are done by measuring 3 or more wavelengths. The 3 wavelengths are identified by an XAS scan at the absorption edge of the anomalous scatterer in the crystal:

- PI - point of inflection of the edge, minimum of f'

- PK - peak of absorption above the edge in energy, maximum of f'', intermediate f'

- RH - remote edge above the edge in energy, maximum of f'

- RL - remote edge below the edge in energy, maximum of f'
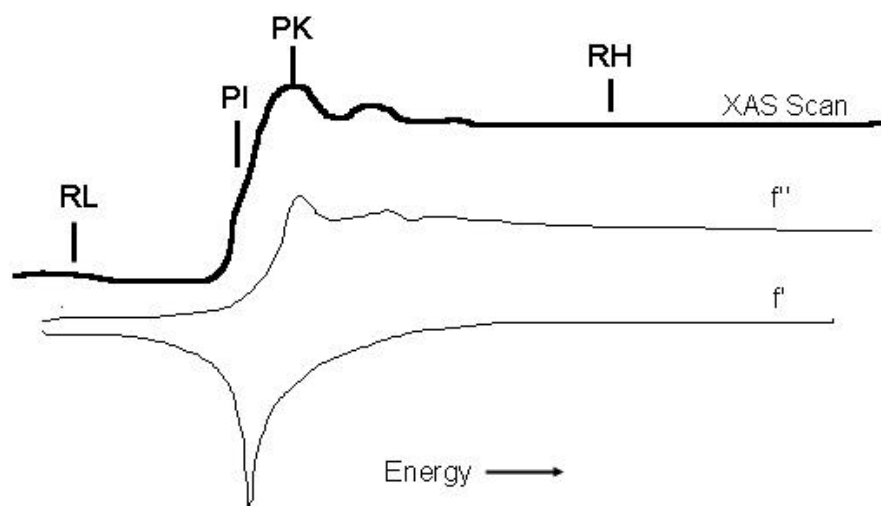
The 4th wavelength, RL, is sometimes measured below the edge. Its contribution to phasing is minimal, because it has no anomalous signal and its dispersive differences are redundant with RH, although it may be used to increase redundancy. It would be handled identically to RH except that it will have no anomalous signal. More useful may be another remote edge above the edge in energy, which would have both a dispersive difference and an anomalous signal and increase data redundancy.

If the MAD signal is strong, a decent MAD experiment can be done with just two wavelengths, PK and RH. Both wavelengths have a good anomalous signal and there is a dispersive difference in f' between the two.

Always collect PK first. It has the best signal for solving the heavy atom positions and in the best of cases you can solve a structure with just this one wavelength. The second wavelength to collect is then RH as these two wavelengths suffice to make a minimal MAD experiment. Next, collect PI, as it has the largest change in f', and collect RL only if there is plenty of time. It is probably better to get more redundancy at the other wavelengths than to collect RL, though.

Data Collection

The best way to collect the data seems to be to simply collect 3 data sets on a frozen crystal with exactly

the same strategy. Use inverse beam strategy to collect the Bijvoet pairs in lower symmetry spacegroups. In higher symmetry spacegroups it is usually trivial to get the Bijvoet pairs on a large detector face. Make sure you collect enough data to get about 95% completeness on the Bijvoet pairs. Offset the crystal that it rotates about 20⁰ off an axis to up the completeness by avoiding the blind zone overlapping an axis. We monitor the data by integrating it continuously as it comes off the machine to find the point at which the data is complete and to make sure we are getting good R-symms. The quality of the data is the ultimate key to MAD phasing. Additional redundancy is very helpful as the signal is small and averaging over a large number of reflections is helpful. We have tried strategies to collect the data in blocks and find this unnecessarily complicates data collection, reduction and phasing. Our best results have come from the simplest strategy with the highest redundancy. We have seen large improvements in map quality due to increased redundancy and high redundancy is needed for the best results in finding large number of selenium positions in larger proteins by allowing the selection of the best Bijvoet differences. R-symms should range from 2-3% for the low resolution data to 10% for the high resolution with an average of about 6% (without merging Bijvoets). Scale the Bijvoets together, but do not merge them. Output the data as I's.

Making the Fin Files

Always run the data through xprepfin, even if you think you have a fin format file to avoid bad reflections and problems with centric vs. acentric, etc. If you are using scalepack, xprepfin can read the output file and convert it to a fin file. If you made I's as suggested above, then select the **Convert I to F** option. Also select the **Data are: Bijvoet Pairs** option:
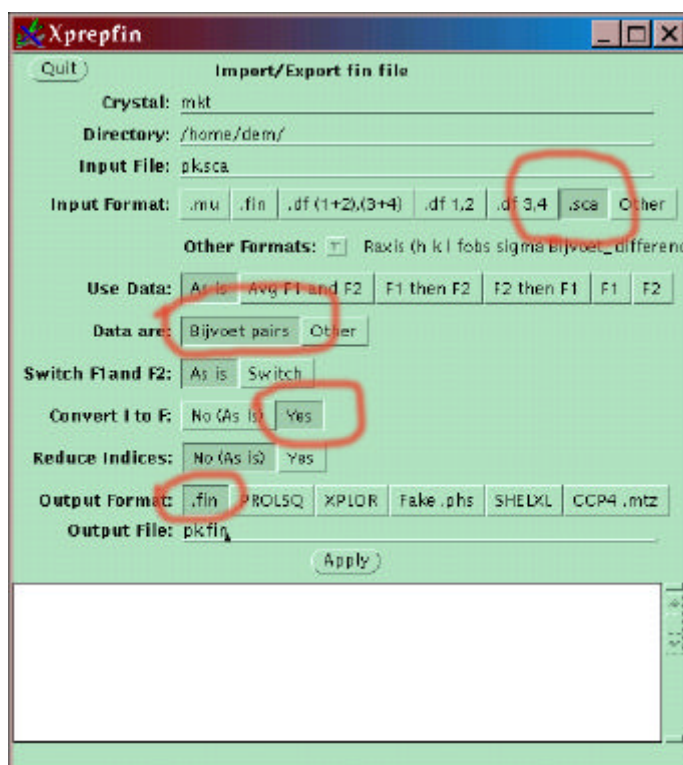
You should have one fin file for each wavelength collected. For the standard 3 wavelength MAD experiment these will be pk.fin, pi.fin and rh.fin

Difference Patterson Maps

The Bijvoet differences are treated as would any anomalous scatterer. Make the Patterson maps as usual with xfft. In xtalmgr, slect xfft as the application. Click on pk.fin as **Input Argument 1**, click **Auto Name**

Output to set the output file to pk.map, and then use **Add Args** followed by **Run Command**. This should start xfft with pk.fin as the input and pk.map as the output. Input the **Resolution Filter** – I usually choose 5-3.5 Å and set the **Outlier Filter** to 30%. If you have the newer XtalView 4 you can also choose Remove Patterson Origin which helps if there happens to be a peak near the origin. It also lets you see the size of the largest peak printed in the message window. You can change the resolution and outlier filters to maximize this number to get the best Patterson map. Generally, the highest peak will be somewhere above 400 (the map is scaled such that 1 sigma or 1 rms is equal to 50), although this is also dependent on how many sites you have; the more sites the smaller the size of each one. View the Patterson map with xcontur. In the xtalmgr, pick xcontur as the application, click on pk.map as the input argument 1, and click **Add Args**, **Run Command**.

The dispersive differences need to be merged with xmerge. Merge the pairs in this order using xmerge:

Wavelengths	Input file 1	Input File 2	Output File
PI-PK	Pi.fin	Pk.fin	Pi.pk.fin
PI-RH	Pi.fin	Rh.fin	Pi.rh.fin
PK-RH	Pk.fin	Rh.fin	Pk.rh.fin

The largest signal will be in the PI-RH pair for the dispersive differences and in PK for the anomalous differences. All the Pattersons should look basically the same. All 3 wavelengths, PI, PK and RH will have an anomalous signal with the largest signal at PK. If you can't see peaks in the Patterson maps at low resolution - say 5 Å, then there probably isn't any need to proceed. However, with well measured data, this should not be a problem. We have seen sulfur atoms, which have a very small signal, in Bijvoet difference Pattersons.

Solve the Patterson with SHELX and Shake-n-Bake if the Patterson is complicated. Actually, you can skip making all the Pattersons if you know how to run these programs and interpret the output, as they can tell you about the signal. I like to take the solution from these programs, put them into xpatpred, write out a predictions file and then overlay this label file onto the Patterson with xcontur (use the File… button in xcontur and load the prediction file you saved from xpatpred as a Labels file). The labels should fall onto the peaks in the Patterson. This gives you the confidence to go on with the next steps. If you have large peaks, and they are all accounted for by the solution, the MAD phasing cannot fail. If you have some peaks not accounted for, but the majority are, the best course is probably to shove on and with a partial solution and use these phases to make Bijvoet difference Fouriers with pk.fin and xmergephs to pick up the additional sites later.

In general, if there isn't anything in the Patterson map for a particular wavelength or wavelength pair, then it is not worth including this particular file in the phasing.

Phasing

Now you are ready to phase. Start xheavy and use the editor to add each pair. You will have 6 "derivatives": PI_ano, PK_ano, RH_ano, PI-PK, PI-RH and PK-RH. In order to avoid overestimating the figure merit each derivative should be weighted correctly. Weight each derivative at 33%. This way, the problem reduces to 1 isomorphous derivative plus one anomalous derivative. If you have an RL dataset, lower the weights to 25% since you now have 4 identical solutions. Failure to downweight will hinder the solvent flattening step by overestimating the figure of merit and not giving any room for improving the phases. All the derivatives share the same sites - call them all atom type PT (the default - in practice the subtle difference between different atom types structure factors are far below your noise levels) and set the B's to 10.0. Refine the derivatives as usual. Although, they all share the same sites, letting them all

refine individually gives the best phases.   Sometimes, one of the derivatives refines poorly – usually the Patterson also looked rather poor. Delete it from the list. Poor derivatives do more harm than good. Start at 5 Ångstroms and then raise the resolution limit to whatever limit you feel the signal continues to. Be honest here- adding lots of high resolution but poorly phased data will not help the maps. One problem is what to call the native F. If you have a data set you plan to use as the native for refinement then include this on the master fin file field, or just leave this line blank.

Once refined you can calculate the protein phases. During phasing you get several useful statistics. One is the R-anomalous of the 25% largest Bijvoet differences for the anomalous pairs. This should be 40% or less. 25-30% is ideal. The R-centric can be used for the isomorphous/dispersive pairs. These should be 60% or less, although this can be in the 70's and still give useful information. In the 50's is ideal. Another thing to check is whether the acentric and centric scale factors agree to within about 20%. The differences in the values of f' and f'' are all automatically scaled by xheavy. The phasing power for the dispersive pairs is usually surprisingly high if you are used to heavy atom derivative phasing. This seems to be because the errors are quite low since the same crystal is used for all measurements. The final figure-of-merit in our cases is usually around 0.6. This would seem low for a normal MIR experiment but the equivalent MAD maps are much better.

Absolute Configuration or "Hand"

At this point you MUST try BOTH alternate heavy atom configurations, also called "hands". The correct configuration is ambiguous in the Patterson function and cannot be found a priori. To try the alternate invert each atom around the center - i.e. change the sign on x,y,z to -x,-y,-z. (In XtalView4 you can do this automatically by opening the **Options**… window in xheavy and changing the **Heavy Atom Configuration** to –1. Also change the name of the output phase file or you will overwrite it ☺ I call the files h1.phs and h2.phs). Make both maps and examine them in xcontur. In one you will see clear solvent channels and protein features and the other map will be junk. If you can't see a difference backtrack and check all the steps.   Before closing heavy, don't forget to save the solution with **Save Derivative**. All the output goes into a file named xheavy_####.log, where #### is the process id. You might want to rename this to something you can remember for later use in writing the phasing table in the paper.

Improving the Model

At this point its worth looking for minor sites (unless the maps are so good you don't need to). What minor sites you say? Why the sulfurs of course. The sulfurs in the protein will have no dispersive signal as they have no absorption edge in the range useful for MAD. However, they do have a small but significant anomalous signal, especially in Fe MAD experiments, and sometimes at the higher energy selenium experiments. It is also very common to pick up a zinc, copper or magnesium ion that you were unaware of perviously at this stage. By making a Bijvoet difference Fourier map using xmergephs (check both **Swap F1 and F2** and **Add 90 to Phase** check boxes), the PK.fin dataset and the current protein phases you can look for peaks by making a 5-4 Ångstrom Fo-Fc map in xfft and then using xcontur and add these to the anomalous solutions. Then recalculate the phases. This Bijvoet difference Fourier map is also very useful for finding the sulfurs in cysteines and methionines to verify your sequence assignment and finding sequence markers. We have been using this technique with great success for a number of years in all our structure solutions. Even though the resolution at which the sulfurs can be seen is usually quite low, 5-4 Angstroms, this is more than sufficient for locating the sulfur containing side-chains for sequencing purposes.

One note. The site of the scatterers will be the worst spot in the map. The reason for this is technical but has to do with all the errors piling up at the spot where the heavy atom is located since it effectively defines the origin of the phasing. They usually come back with solvent flattening and in the first map after refinement.

Now you are ready to fit the maps! Look at the map by loading the phase file from xheavy into xfit and choose Fo*fom as the map coefficients in the FFT window. E-mail us with difficulties and/or questions. We have solved numerous proteins using MAD phasing with Fe, Cu, selenium and bromine edges. It seems complicated the way its described because I've put in all the details an XtalView novice might need, but once you learn it, its very quick to do. We often knock the structure off at the beam line – at least enough to verify that the maps are good. If you have diffraction data better than 2 and phases to about 2.5 Å, you can often build much of the model using wARP, or you can use the **Autotrace** functions in xfit to knock out a preliminary model in an afternoon.

Solvent Flattening

We use DM to solvent flatten our maps if needed. Many MAD maps are good enough to use unflattened and while solvent flattening increases the contrast in the protein region, it often flattens sidechains on the surface. In any case, look at both maps.

```
f2mtz hklin cua.phs hklout cua.mtz <<eofcell 34.9 70.60 53.50 90 98.12
90symmetry 4labout H K L FP FOM PHIBCTYPOUT H H H F W PFORMAT
'(3F4.0,2F9.2,F7.2)'ENDeof
```

We ran DM with the input file:

```
#This part adds the high resolution F's to the mtz file created
# before
mtzutils hklin1 hires_F.mtz \   hklin2 cua.mtz \  hklout hires_cua.mtz <<eofsymm
p21INCLUDE 1 FINCLUDE 2 FOM PHIBRUNeof#
#Now we run DM with solvent masking, histogram matching
# and symmetry averaging.
dm  HKLIN hires_cua.mtz hklout dm_hires_aver.mtz \histlib
/nfs/image/usr/local/ccp4_3.3/lib/data/hist.lib <<eofsolc 0.45mode solv hist
averncsmask nmer 2wang 8 0aver refi
# the identity matrix for the first moleculerota matrix 1.00 0.00 0.00 0.00 1.00
0.00 0.00 0.00  1.00
trans 0.0 0.0 0.0
aver refi
#the matrix and vector from xfit's LSQ Fit function
#for the second molecule
rota matrix 0.051060 -0.364196 -0.929970 0.164497 -0.915385  0.367570 -0.985096
-0.171713  0.013137trans 10.930595  11.009571  39.851776
ncycle 20
combine free
LABIN FP=F FOMO=FOM PHIO=PHIB
LABOUT PHIDM=PHIDM FOMDM=FOMDM
eof
```

 DM has the nice feature that it calculates all of the needed masks automatically including the non-crystallographic symmetry mask. At this point we put the phases back into Xfit using the script:

```
mtz2various hklin dm_hires_aver.mtz hklout dm_hires_aver.phs << eoflabin FP=FP
FOM=FOMDM PHIB=PHIDM OUTPUT USER '(3I4,x,F7.2,3x,F7.2,3x,F7.2)'END eof
```

Appendix

Quick Start Guide to XtalView

Before starting XtalView the environment must be set up. This is best done by placing the environment variables and path in your .cshrc file as described at the start of the XtalView User's Guide. Alternatively you can source the file XtalView.env included with the standard XtalView distribution. However, each user should set up their own crystal database directory as described in the XtalView User's Guide. On some machines the font path needs to be amended to include the XtalView fonts. This is done with the fonts command. Now start Xtalmgr by typing `xtalmgr&`. If you have trouble starting Xtalmgr, send email to ccms-help@sdsc.edu with a description of the problem and the error message (cut and paste the error into your email).

Xfit environment variables

- **XFITPENTDIR** - the location of the pentamer database - normally $XTALVIEWHOME/data/pbdvec

- **XFITDICT** - the location of the dictionary to use. Two choices are in the standard release: $XTALVIEWHOME/data/dict.pdb which has polar hydrogens a la XPLOR or $XTALVIEWHOME/data/dict.noh.pdb.

- **PRINTERCOMMAND** - use to specify a different printer command from lpr such as lp.-dprinter.

Menu Buttons

A menu button looks like this: . Any time a button has a downward-pointing triangle like this, pressing the right-button on the mouse over this button will bring up a menu of options. Missing this fact will hide much of XtalView's capabilities!

Xfit Canvas Window Keys

- Canvas window must have focus
- Keys are case sensitive

ESC	Toggle hardware stereo or exit photo mode
1	Activate Chi-1 torsion of side-chain if fitting
2	" Chi-2
3	" Chi-3
4	" Chi-4
5	" Chi-5
6	" Chi-6
f	Activate phi torsion between focusres and previous residue
s	Activate psi torsion between focusres and next residue
9	Decrease stereo blank interval (use with 0 to adjust stereo height alignment)
0	Increase stereo blank interval
(	Decrase stereo y stretching (use with ')' to adjust stereo aspect ratio)
)	Increase stereo y stretching
b	Go back one residue in chain
c	Next side-chain rotamer
D	Delete residue (asks to confirm if not water)
h	Rotate about vertical -1 degree
H	Rotate about vertical -10 degrees
i	Zoom in 5% (makes model smaller)
I	Zoom in 9%
j	Rotate about screen perpendicular -1 degree
J	Rotate about screen perpendicular -10 degrees
k	Rotate about screen perpendicular +1 degree
K	Rotate about screen perpendicular +10 degrees
l	Rotate about vertical +1 degree
L	Rotate about vertical +10 degrees
n	Rotate about horizontal +1 degree
N	Rotate about horizontal +10 degrees
o	Zoom out 11% (makes model larger)
O	Zoom out 15%
Q	Calculate least-squares plane of atoms on stack
q	calculate distance of top atom on stack to plane calculated with Q
p	Toggle photo mode (fullscreen)
r	Rocking off
R	Rocking on
u	Rotate about horizontal 11 degree
U	Rotate about horizontal -10 degrees
W	Add water at cursor position
x	Look down x axis
y	Look down y axis
z	Look down z axis
+	Measuring cross at screen center (shows 1.54, 2.7, 3.8 Å distances)
.	Small (default) cross
>	Add MRK residue (single CA atom) after focus
<	Add MRK residue (single CA atom) before focus
]	Go to C-terminus of focus chain
[	Go to N-terminus of focus chain
;	Apply fit and/or Refinement (use **Cancel** button on main window to cancel)

SPACE   Go to next residue

Notes:

The focusres (focus residue) is the last one clicked on and is listed in the lower left of the canvas window footer.

The focus chain is the fragment with the focus residue. Chains terminate if the distance to the next residue is too long to form a peptide bond.