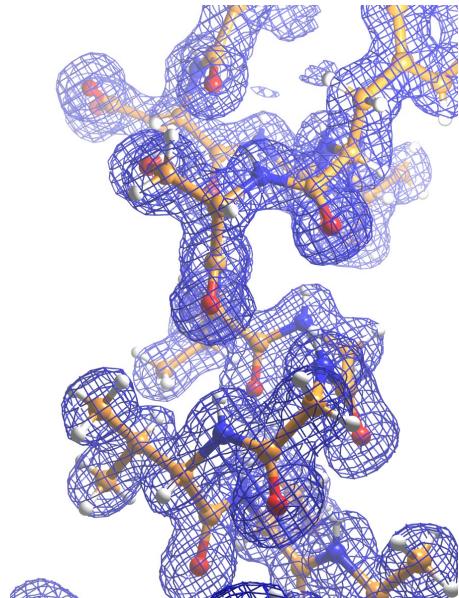


# **MIFit Manual**

Version 9

Copyright © 2008, Rigaku Americas Corporation.  
All rights reserved.



# Table of Contents

<b>MIFit Manual.....</b>	<b>1</b>
<b>Table of Contents.....</b>	<b>2</b>
<u>List of Figures.....</u>	<u>5</u>
<u>List of Tables.....</u>	<u>7</u>
<b>Acknowledgements.....</b>	<b>8</b>
<b>1 Getting Started.....</b>	<b>9</b>
<u>1.1 The MIFit software.....</u>	<u>9</u>
<u>1.2 Launching MIFit.....</u>	<u>9</u>
<u>1.3 Storage of start-up parameters.....</u>	<u>10</u>
<u>1.4 Introduction to session files.....</u>	<u>10</u>
<u>1.5 Opening model, diffraction data and map files.....</u>	<u>11</u>
<u>1.6 Creating a new document.....</u>	<u>11</u>
<u>1.7 Crystal information.....</u>	<u>12</u>
<u>1.8 Default atom colors.....</u>	<u>14</u>
<u>1.9 Stereochemical dictionaries.....</u>	<u>16</u>
<u>1.10 Saving MIFit session files.....</u>	<u>16</u>
<u>1.11 Closing a file and closing MIFit.....</u>	<u>17</u>
<u>1.12 Preferences for MIFit.....</u>	<u>17</u>
<b>2 Coordinate and map files.....</b>	<b>21</b>
<u>2.1 Loading coordinate data.....</u>	<u>21</u>
<u>2.2 Loading electron density data.....</u>	<u>23</u>
<u>2.3 Controlling the map display.....</u>	<u>26</u>
<u>2.4 Moving around the electron density map.....</u>	<u>28</u>
<b>3 Menus and toolbars.....</b>	<b>29</b>
<u>3.1 Startup menu / toolbar.....</u>	<u>29</u>
<u>3.2 General menu / tool bar.....</u>	<u>29</u>
<u>3.3 Modeling menu / toolbar.....</u>	<u>30</u>
<u>3.4 Display menu / toolbar.....</u>	<u>31</u>
<b>4 Keyboard shortcuts.....</b>	<b>32</b>
<u>4.1 Keyboard shortcuts for document handling.....</u>	<u>32</u>

<u>4.2 Keyboard shortcuts for display</u> .....	32
<u>4.3 Keyboard shortcuts for side-by-side stereo viewing</u> .....	33
<u>4.4 Keyboard shortcuts for model fitting</u> .....	33
<b><u>5 Model Building and Refitting</u></b> .....	<b>35</b>
<u>5.1 Chain tracing</u> .....	35
<u>5.2 Refitting backbone atoms</u> .....	35
<u>5.3 Building and fitting side chain atoms</u> .....	36
<u>5.4 Refitting and adding individual residues</u> .....	37
<u>5.5 Modeling discrete disorder</u> .....	39
<u>5.6 Local structure refinement</u> .....	39
<u>5.7 Adding water molecules</u> .....	41
<u>5.8 Interactive ligand fitting</u> .....	42
<u>5.9 Structure refinement with CCP4/REFMAC5</u> .....	42
<u>5.10 Using the Ramachandran plot</u> .....	43
<b><u>6 The Ligand Dictionary Editor</u></b> .....	<b>45</b>
<u>6.1 Entering ligand data</u> .....	45
<u>6.2 Conformation generation</u> .....	46
<u>6.3 The Ligand Dictionary Editor</u> .....	47
<u>6.4 Exporting ligand information</u> .....	48
<u>6.5 Restraint dictionary files for SHELX and CNS/CNX</u> .....	48
<b><u>7 Displaying structure data</u></b> .....	<b>50</b>
<u>7.1 Setting model display modes</u> .....	50
<u>7.2 Displaying model surfaces</u> .....	52
<u>7.3 Backbone and ribbon displays</u> .....	55
<u>7.4 Choosing a background color</u> .....	57
<u>7.5 Printing directly from the MIFit main canvas</u> .....	58
<u>7.6 Schematic protein:ligand interaction plots</u> .....	59
<u>7.7 Data capture for structure reports and PDB deposition</u> .....	60
<b><u>8 Running external crystallographic software</u></b> .....	<b>65</b>
<u>8.1 Accessing the CCP4 suite</u> .....	65
<u>8.2 Accessing SHELX and OpenEye software</u> .....	65
<u>8.3 The Job List menu</u> .....	66
<u>8.4 Running customized Scripts</u> .....	67

<b>9 SAD phasing.....</b>	<b>69</b>
9.1 Prerequisites.....	69
9.2 SAD phasing interface.....	69
9.3 SAD phasing output files.....	71
<b>10 Automated Structure Solution.....</b>	<b>74</b>
10.1 Overview of automated co-crystal structure determination.....	74
10.2 The automated co-crystal structure determination interface.....	75
10.3 The Job/Integrate with d*TREK interface.....	79
10.4 The Job/Molecular Replacement interface.....	80
10.5 The Job/Refinement interface.....	81
10.6 Checking requirement for restraint dictionaries.....	83
10.7 The project history file.....	84
10.8 Automated structure validation.....	84
10.9 Automated ligand fitting.....	86
10.10 Semi-automated ligand fitting.....	86
10.11 Rebuilding models using non-crystallographic symmetry.....	88
10.12 Automated multiple protein:ligand structure superposition.....	89
<b>11 Tutorial Lessons.....</b>	<b>91</b>
11.1 Lesson 1: Basic manipulation of the model display.....	91
11.2 Lesson 2: Loading and changing electron density map displays.....	94
11.3 Lesson 3: Refitting individual residues.....	96
11.4 Lesson 4: SAD phasing.....	99
11.5 Lesson 5: Chain tracing.....	102
11.6 Lesson 6: Establishing restraints for structure refinement.....	104
11.7 Lesson 7: Displaying molecular surfaces.....	106
11.8 Lesson 8: Creating schematic plots of protein:ligand interactions.....	108
11.9 Lesson 9: Creating reports for PDB deposition and publication.....	110
11.10 Lesson 10: Automated refinement.....	113
11.11 Lesson 11: Automated co-crystal structure determination.....	115
11.12 Lesson 12: Automated Molecular Replacement.....	118
11.13 Lesson 13: Automated ligand density docking.....	121
11.14 Lesson 14: Automated ligand density docking with OpenEye software.....	123
11.15 Lesson 15: Superimposing ligands from multiple related structures.....	125

<u>11.16 Lesson 16: Model and map display using files from the wwPDB.....</u>	127
<b>Appendix A: Scripting Language.....</b>	<b>129</b>
A.1 Model Commands.....	129
A.2 View Commands.....	129
A.3 Map Commands.....	130
A.4 Miscellaneous Commands.....	133
<b>Appendix B: Default Colors.....</b>	<b>134</b>

## List of Figures

<b>Figure 1.1 MIFit startup screen.....</b>	<b>10</b>
<b>Figure 1.2 MIFit with a new empty document loaded.....</b>	<b>12</b>
<b>Figure 1.3 File/Manage Crystal window.....</b>	<b>13</b>
<b>Figure 1.4 Crystal selection dialog.....</b>	<b>14</b>
<b>Figure 1.5 Dialog defining the atom coloring scheme.....</b>	<b>15</b>
<b>Figure 1.6 Dialog defining atom colors according to B-factor values.....</b>	<b>15</b>
<b>Figure 1.7 Prompt for saving changed documents.....</b>	<b>16</b>
<b>Figure 1.8 General Preferences.....</b>	<b>17</b>
<b>Figure 1.9 Environment Preferences.....</b>	<b>19</b>
<b>Figure 2.10 File/Open models, data, maps, etc... dialog.....</b>	<b>21</b>
<b>Figure 2.11 MIFit after loading a coordinate file.....</b>	<b>22</b>
<b>Figure 2.12 Dialog for selecting MTZ data labels.....</b>	<b>24</b>
<b>Figure 2.13 Model and map display.....</b>	<b>25</b>
<b>Figure 2.14 Using Direct FFT.....</b>	<b>26</b>
<b>Figure 2.15 Contour Options dialog.....</b>	<b>27</b>
<b>Figure 5.16 Pentamer match for model structure following A 759 Met.....</b>	<b>36</b>
<b>Figure 5.17 Dialog for mutating amino acids.....</b>	<b>37</b>
<b>Figure 5.18 Model/Add residue dialog.....</b>	<b>38</b>
<b>Figure 5.19 Options for structure optimization.....</b>	<b>40</b>
<b>Figure 5.20 Dialog to set parameters for water picking.....</b>	<b>41</b>
<b>Figure 5.21 Example of Ramachandran Plot.....</b>	<b>44</b>
<b>Figure 6.22 SMILES string import options.....</b>	<b>46</b>
<b>Figure 6.23 Dictionary Editor window.....</b>	<b>47</b>

<b>Figure 7.24 Render/Ball/Cylinder Size dialog.....</b>	<b>50</b>
<b>Figure 7.25 Collage of basic display methods.....</b>	<b>51</b>
<b>Figure 7.26 van der Waal surface selection dialog.....</b>	<b>52</b>
<b>Figure 7.27 Collage of surface displays.....</b>	<b>53</b>
<b>Figure 7.28 Example of solid surface showing a fragment binding pocket in hsp90.54</b>	
<b>Figure 7.29 Show/Secondary Structure/Ribbon Colors... dialog.....</b>	<b>55</b>
<b>Figure 7.30 C<math>\alpha</math> trace and ribbon diagrams.....</b>	<b>56</b>
<b>Figure 7.31 Tube and Schematic representations of a protein.....</b>	<b>56</b>
<b>Figure 7.32 Color palette for changing the MIFit canvas background color.....</b>	<b>57</b>
<b>Figure 7.33 File/Print Preview window.....</b>	<b>58</b>
<b>Figure 7.34 Example ASP (adenine binding to hsp90).....</b>	<b>60</b>
<b>Figure 7.35 Job/Report interface.....</b>	<b>61</b>
<b>Figure 7.36 Example of an HTML structure report.....</b>	<b>63</b>
<b>Figure 8.37 Example of a running job under Windows.....</b>	<b>66</b>
<b>Figure 8.38 Open Job Results dialog.....</b>	<b>67</b>
<b>Figure 8.39 Job/Run Custom Job interface.....</b>	<b>68</b>
<b>Figure 9.40 Job/SAD Phasing interface.....</b>	<b>69</b>
<b>Figure 9.41 Summary of site searches.....</b>	<b>71</b>
<b>Figure 9.42 Portion of phasing summary.....</b>	<b>72</b>
<b>Figure 10.43 Job/Cocrystal Solution dialog.....</b>	<b>76</b>
<b>Figure 10.44 Example of HTML summary containing just one job.....</b>	<b>78</b>
<b>Figure 10.45 Job/Integrate with d*TREK dialog.....</b>	<b>79</b>
<b>Figure 10.46 Job/Molecular Replacement dialog.....</b>	<b>80</b>
<b>Figure 10.47 Job/Refinement dialog.....</b>	<b>82</b>
<b>Figure 10.48 Section of project history file recording molecular replacement and refinement jobs.....</b>	<b>84</b>
<b>Figure 10.49 Example of an ‘Error list file’ with density errors.....</b>	<b>85</b>
<b>Figure 10.50 Job/NCS Modeling interface.....</b>	<b>88</b>
<b>Figure 10.51 Cocrystal Superposition interface.....</b>	<b>89</b>
<b>Figure 10.52 Six superimposed HIV protease inhibitors.....</b>	<b>90</b>
<b>Figure 11.53 Lesson 1.....</b>	<b>93</b>
<b>Figure 11.54 Lesson 2.....</b>	<b>95</b>
<b>Figure 11.55 Lesson 3.....</b>	<b>98</b>

<b>Figure 11.56 Lesson 4, SAD results.....</b>	<b>100</b>
<b>Figure 11.57 Lesson 4, main view.....</b>	<b>101</b>
<b>Figure 11.58 Lesson 5.....</b>	<b>103</b>
<b>Figure 11.59 Lesson 6.....</b>	<b>105</b>
<b>Figure 11.60 Lesson 7.....</b>	<b>107</b>
<b>Figure 11.61 Lesson 8.....</b>	<b>109</b>
<b>Figure 11.62 Lesson 9, RCSB PDB ADIT.....</b>	<b>111</b>
<b>Figure 11.63 Lesson 9, report.....</b>	<b>112</b>
<b>Figure 11.64 Lesson 10.....</b>	<b>114</b>
<b>Figure 11.65 Lesson 11, HTML summary.....</b>	<b>116</b>
<b>Figure 11.66 Lesson 11, main view.....</b>	<b>117</b>
<b>Figure 11.67 Lesson 12.....</b>	<b>120</b>
<b>Figure 11.68 Lesson 13.....</b>	<b>122</b>
<b>Figure 11.69 Lesson 14.....</b>	<b>124</b>
<b>Figure 11.70 Lesson 15.....</b>	<b>126</b>
<b>Figure 11.71 Lesson 16.....</b>	<b>128</b>
<b>Figure 11.72 Default colors.....</b>	<b>134</b>

## **List of Tables**

<b>Table 3.1 Menu and tool bar options available at startup.....</b>	<b>29</b>
<b>Table 3.2 General toolbar items and their menu equivalents.....</b>	<b>30</b>
<b>Table 3.3 Modeling toolbar items.....</b>	<b>31</b>
<b>Table 3.4 Display toolbar items.....</b>	<b>31</b>
<b>Table 4.5 Keyboard shortcuts associated with menu and toolbar items.....</b>	<b>32</b>
<b>Table 4.6 Keyboard shortcuts controlling model display.....</b>	<b>32</b>
<b>Table 4.7 Keyboard shortcuts to alter software stereo settings.....</b>	<b>33</b>
<b>Table 4.8 Keyboard shortcuts active during model fitting.....</b>	<b>34</b>

# Acknowledgements

The automated error checking capabilities used by MIFit's automated refinement script employs probability data on phi-psi distributions provided by the Richardson laboratory (Lovell et al., Proteins 50:437-450, 2003).

The data for the SAD phasing tutorial was supplied by Aiping Dong (Structural Genomics Consortium, University of Toronto) and corresponds to PDB entry 2AZP.

Many of the structures and data sets used in the tutorials were obtained from the Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)).

MIFit uses a library from The FreeType Project ([www.freetype.org](http://www.freetype.org)).

# 1 Getting Started

## 1.1 The MIFit software

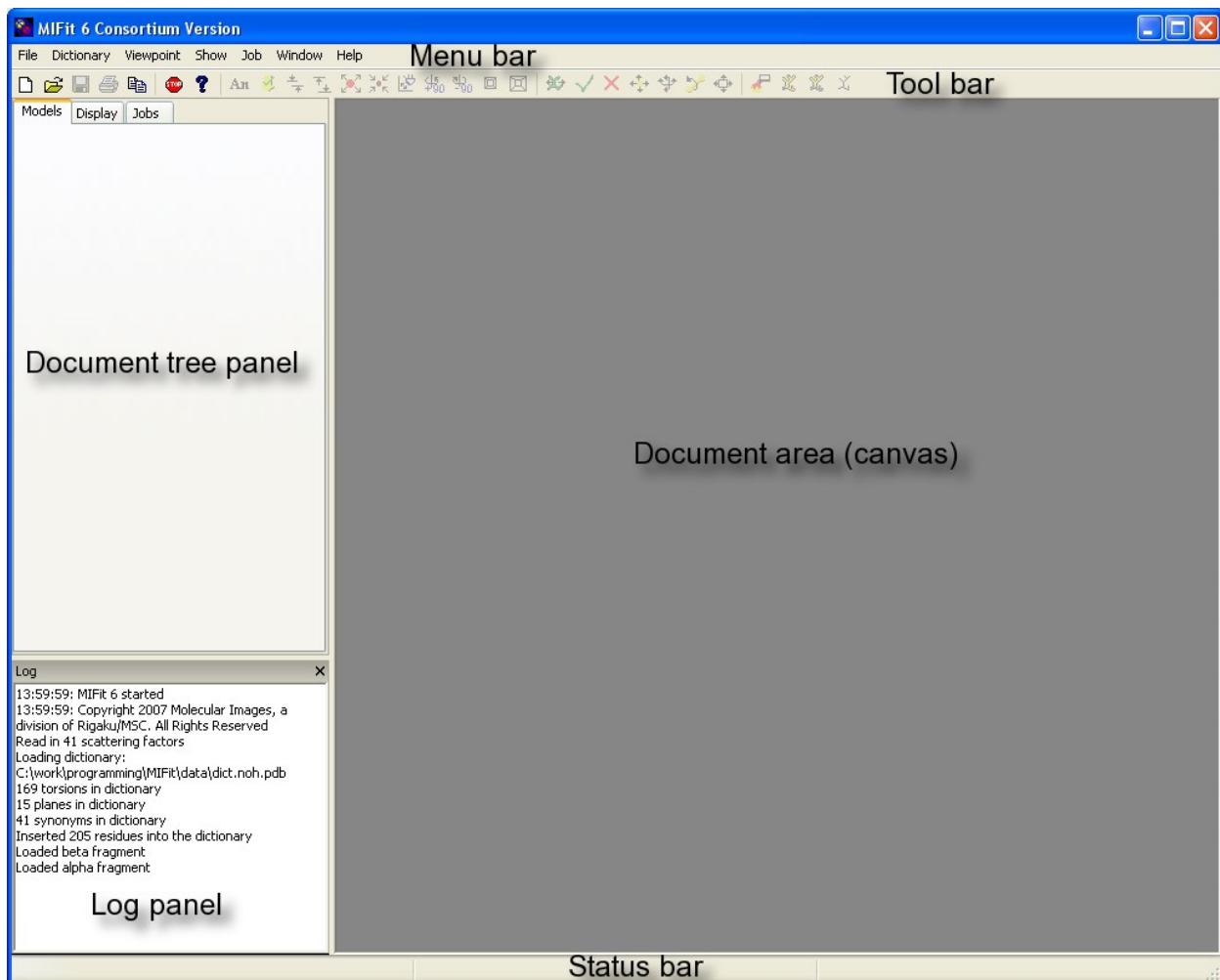
MIFit (pronounced em-ī-fit) is a molecular graphics program for protein crystal structure determination that runs on Windows and Linux operating systems. Besides capabilities for model-building and protein structure display, MIFit also provides interfaces that apply third-party crystallographic software (d\*TREK, the CCP4 suite and SHELX) for image data processing, SAD phasing, molecular replacement and refinement. These crystallographic applications are managed by a program called MIExpert. A particular focus of the MIFit/MIExpert software is the efficient solution and analysis of protein:ligand complexes.

## 1.2 Launching MIFit

On computers with a Linux operating system, MIFit may be started from the command line by simply executing the MIFit program that is found in the top level of the MIFit installation directory. It is not necessary to set any environment variables to run MIFit. However if you intend to run CCP4 applications from MIFit, the environment settings for CCP4 should be established prior to launch.

On computers with a Windows operating system, MIFit may be launched either by double-clicking on the MIFit desktop icon or by selecting **All Programs/MIFit/MIFit** from the Start menu. If CCP4 is properly registered, the CCP4 environment is automatically registered and available to MIFit.

Starting MIFit launches the interface shown in Figure 1.1. The startup window shows the most minimal menu and tool bars and contains two panes on the left side. The top left pane contains a tree representation of the document and the bottom left pane displays log information output by the program. The window decoration is provided by the window manager.



**Figure 1.1** MIFit startup screen

### 1.3 Storage of start-up parameters

On the Linux operating system, MIFit stores user and environment specific parameters in a file called `.MIFit` in the user's home directory. If this file does not exist it is created the first time that MIFit is run. Once established, the `.MIFit` file is read whenever MIFit is started. This file is not intended to be edited by users. All the values stored can be changed through the use of MIFit.

### 1.4 Introduction to session files

MIFit may be launched with or without command line parameters. The command line parameter most commonly supplied is the name of a MIFit 'session file', which is indicated by file extension `.mlw`. Session files contain information about individual projects and may be saved from the MIFit interface. A session file may be given to another person in your organization to review structure results. When a user opens the session file the exact view of the structure that was saved will be displayed, including any annotations that were provided. Since session files support access to diffraction data it is also possible to use them to provide non-crystallographers convenient access to density map information.

## 1.5 Opening model, diffraction data and map files

A document that already exists on a user's file system may be opened using the **File/Open models, data, maps, etc...** command (keyboard shortcut **Ctrl+O**). This command is used to load structure data from session (*.mlw*) or coordinate (*.pdb*) files. Files associated with atomic models that have been the subject of recent work are also kept in a history list displayed in the **File** menu. Double-clicking on file names in this list provides convenient access to these structures.

A group of files can be opened at once using the open command. Holding down 'Ctrl' or 'Shift' allows multiple selections. When a group of coordinate files are loaded together they are automatically color-coded if the **File/Preferences...** setting **Incrementally color additional models** is checked. An additional setting **Render/Dim Non-active Models** may be deselected in order to view all models with equal visibility.

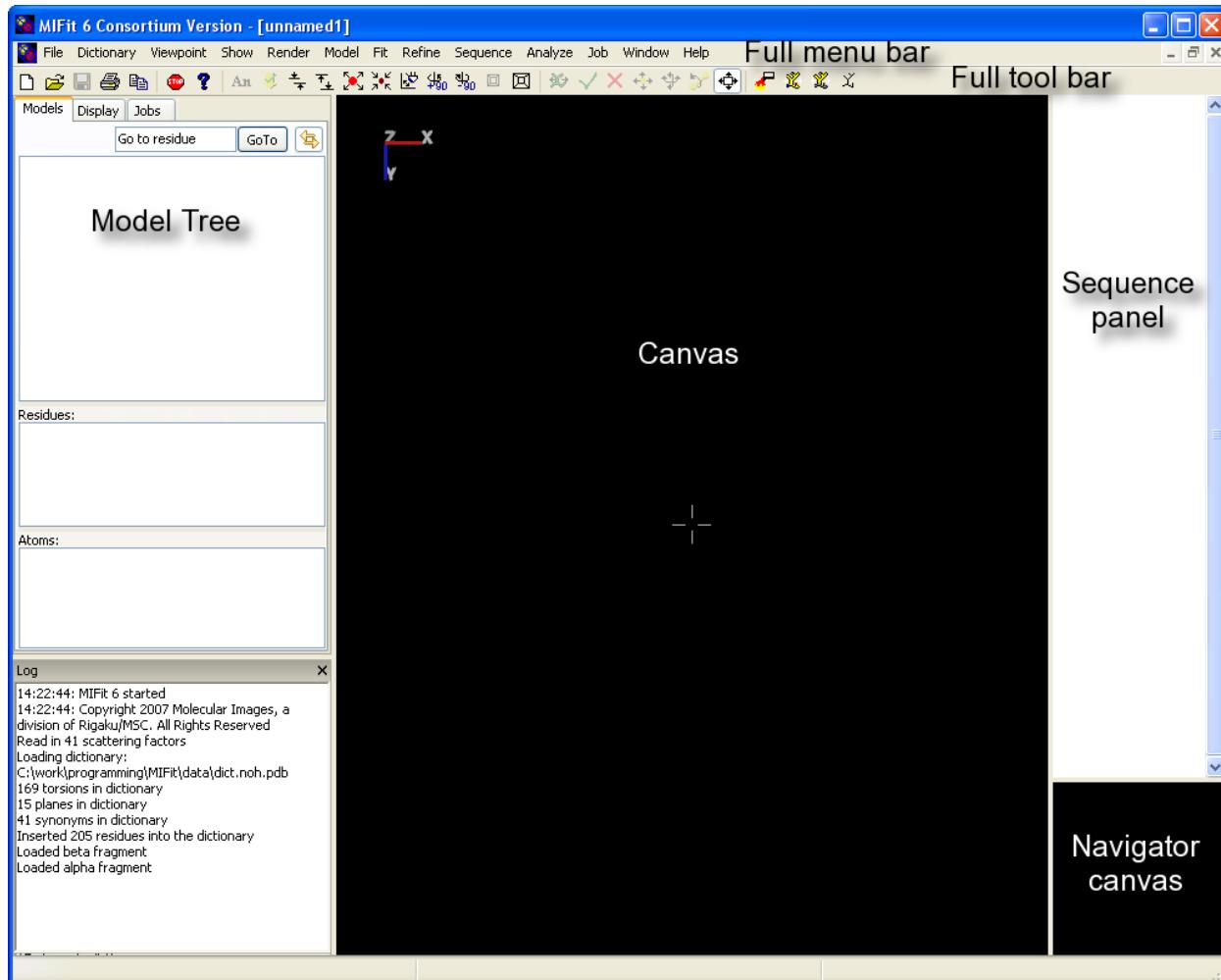
Diffraction data files and pre-computed electron density maps in the CCP4 or XtalView formats are also loaded into MIFit with the **File/Open models, data, maps, etc...** command. Diffraction data and maps are discussed in Chapter 2.

## 1.6 Creating a new document

A new MIFit document may be created with the **File/New...** command (keyboard shortcut **Ctrl+N**). The **File/Open models, data, maps, etc...** may then be used to add content to the current document. To open files in a different document, use **File/Open into new document...** command.

Under the Windows operating system, individual document windows may be cascaded, tiled vertically or tiled horizontally. Under the Linux operating system, each document window always covers the whole document area and has a tab attached at the top. Clicking on the tab of a document brings its window to the top. The Windows version of MIFit contains an additional **Window** menu with which the arrangement of the document window may be controlled.

Creating a new document with the **File/New...** command brings up an empty document window (Figure 1.2). This action changes the appearance of both the menu bar and the toolbar from that shown in Figure 1.1 to provide access to commands for displaying and manipulating structure data. Opening an existing document or importing structure data into this new document will immediately fill the model, canvas, sequence and navigator windows with document-related content.



**Figure 1.2** MIKit with a new empty document loaded

Figure 1.2 shows the document area divided into three new distinct areas: the large canvas area where the model will be displayed, a sequence window which will show the model sequence and a small square navigator canvas. This navigator canvas will show a scaled drawing of the C $\alpha$  trace of the whole model. The default name for the new empty session file is unnamed1.mlw. If further new documents were added they would be named unnamed2.mlw, unnamed3.mlw etc. The document tree pane on the left is now occupied by windows for the the model navigation tree. The model navigation tree pane is divided into separate panels for chains/segments, residues and atoms.

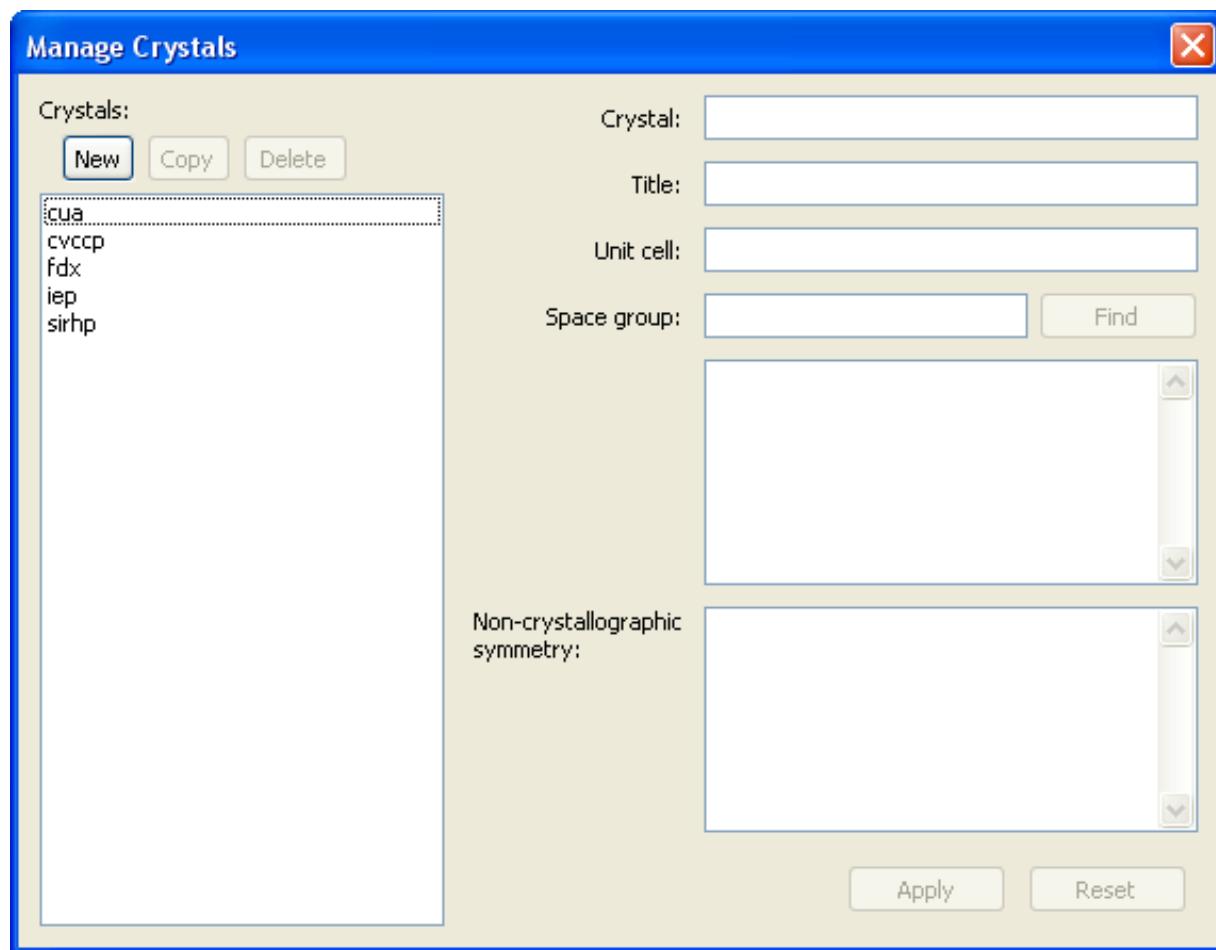
## 1.7 Crystal information

Crystal data (cell dimensions and space group information) for each structure model is automatically read from the input coordinate file, i.e. the information is taken from the CRYST1 record in a PDB format file. This information is displayed next to a small crystal icon beneath the file name for the model. Dummy values (space group P1 and cell dimensions 100.0 100.0 100.0 90.0 90.0 90.0) are entered if the input coordinate file lacks this information. When a structure factor or density map file is loaded into MIKit the crystal cell and symmetry associated with that data is

independently read and displayed. The CCP4 MTZ and map formats within their header data and using these as standard working formats facilitates straightforward handling of crystallographic calculations in MIFit.

The model symmetry controls the display of symmetry related molecules and the data symmetry controls the calculation of density maps.

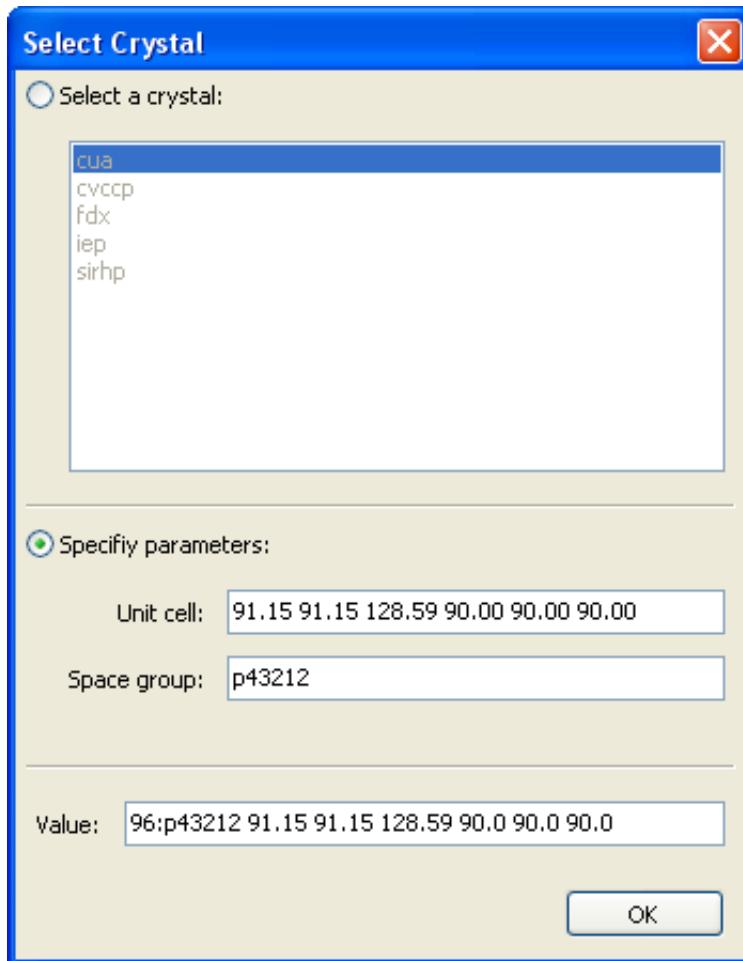
When crystal data is absent a special crystal file may be established using the **File/Manage Crystals...** command (Figure 1.3).



**Figure 1.3** File/Manage Crystal window

The **New** button under the **Crystals:** section may be used to establish a new crystal file named according to the **Crystal** field. (Use of the **Copy** or **Delete** commands may be used in conjunction with the selectable crystal list to copy or delete crystal information from the selected file.) Information for a new crystal is entered through the **Title**, **Unit cell** fields together with a **Space group** name or number. (Note that it is not necessary to enter the symmetry operations – these are determined using the **Find** button based on the **Space group** parameter). Clicking **Apply** button adds this new crystal to the set of known crystals stored in the directory defined by the **File/Preferences.../Environment/Crystal Data Directory** parameter.

Alternatively, when it is only necessary to select crystal data from a pre-existing file that is different from the current crystal data, a right-click on the crystal icon in the navigation tree provides an **Edit** command (Figure 1.4).

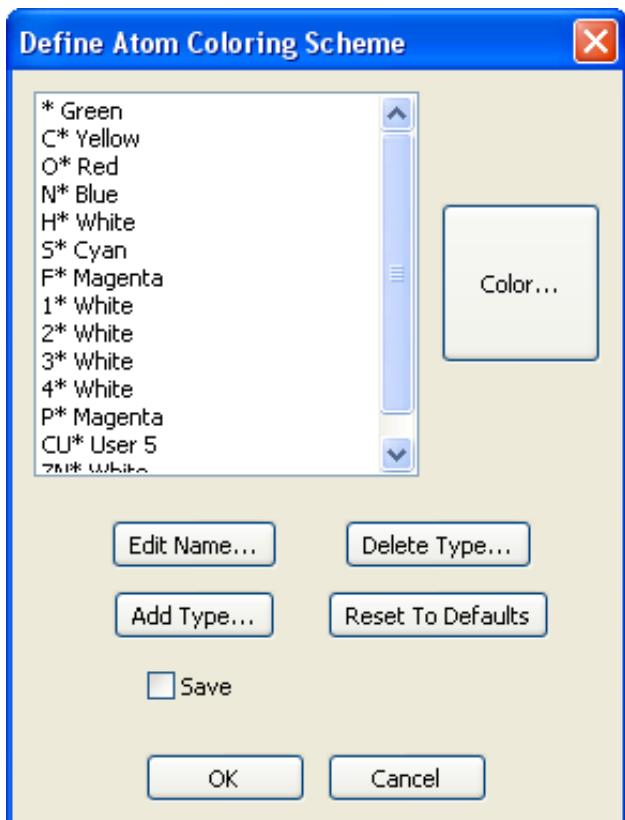


**Figure 1.4**      Crystal selection dialog

With the **Select a crystal** radio button set on, it is possible to select a crystal file from the scrollable list. If required, setting the **Specify parameters** radio button on allows the user to edit the crystal for the selected crystal. The **OK** button sets and saves the current parameters.

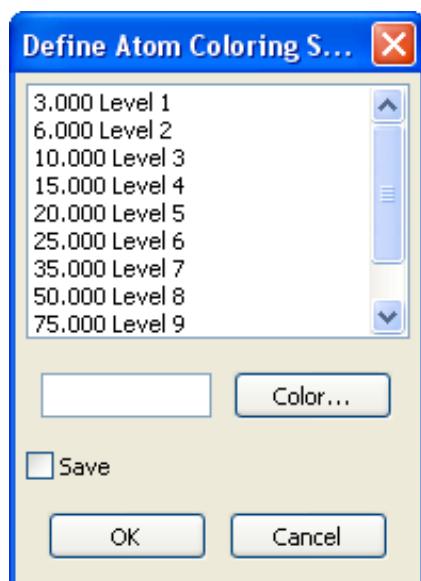
## 1.8 Default atom colors

The **File/Define Atom Colors...** command (Figure 1.5) may be used to set default atom colors according to atom names. It will not normally be necessary to change these settings from the default values.



**Figure 1.5** Dialog defining the atom coloring scheme

Atoms may also be colored according to the values provided in the B-factor column in the input coordinate file. The color definitions may be set using the **File/B-Value Colors Ranges...** command, which provides a dialog window for setting the color values (Figure 1.6).



**Figure 1.6** Dialog defining atom colors according to B-factor values

## 1.9 Stereochemical dictionaries

When MIFit is launched a dictionary file that contains stereochemical information on amino acids and other entities commonly found in protein crystal structures is loaded. This dictionary is used for model fitting and internal refinement.

Unless a custom dictionary file is specified using the **File/Preferences...** menu for environmental preferences (Figure 1.9), a default dictionary is used from the file *MIFIT/data/dict.noh.pdb*. When a new dictionary is loaded into MIFit the log window in the lower left will report statistical information on the dictionary contents.

The current dictionary may also be changed (perhaps just for the purposes of a single session) using the **Dictionary/Load New Dictionary...** command. The dictionary may also be extended using the **Dictionary/Load Append Dictionary...** command. Executing either of these items will pop up a file selection dialog window for specifying the additional dictionary files. Files with extensions *.pdb* or *.ent* (PDB format) or *.cif* (mmCIF format) may be read by MIFit for this purpose.

The task of adding a new ligand to a dictionary, including the ability to check and edit the refinement restraints is handled through the **Dictionary/Import Ligand** command. The use of this command is discussed in Chapter 6.

## 1.10 Saving MIFit session files

The **File/Save Session** command is disabled (grayed out) until any changes are made to the working document. Any file may be saved with a new name by using the **File/Save Session As...** command. MIFit also contains a safety feature so that if an attempt is made to close a document that has been modified but was not saved a warning prompt appears (Figure 1.7).

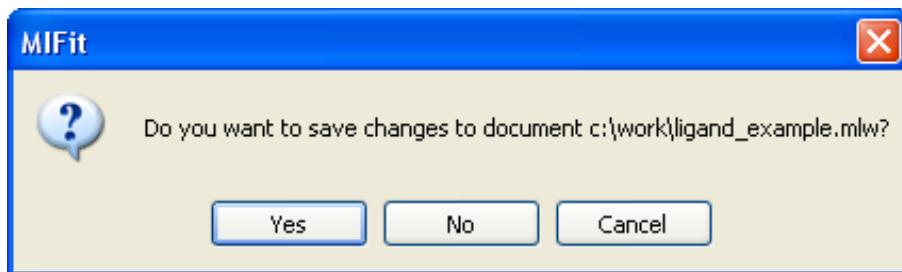


Figure 1.7      Prompt for saving changed documents

If several models are open and you switch between models, the **File/Save Session** command menu could either be active or inactive, depending on whether any changes were made to the selected structure.

A separate command (**File/Export PDB...**) is used if, rather than a session file (which also contains MIFit display information) you wish to save the current active model coordinates to the PDB file format.

## 1.11 Closing a file and closing MIFit

A selected structure document may be closed using the **File/Close** command. Note that MIFit saves structure information in session files so that even if you began your MIFit session by importing a coordinate file in PDB format, this information will be saved as a session file.

The MIFit program may be closed using the **File/Exit** command or by using the keyboard shortcut **Alt+X**.

## 1.12 Preferences for MIFit

The **File/Preferences...** menu is used to establish information that will be used in every MIFit session and which may depend on the operating environment of the computer on which MIFit is installed.

The preference controls are separated into **General**, **Environment**, and **Map contours** preferences and the tree on the left (Figure 1.1) allows toggling between the sets of preferences.

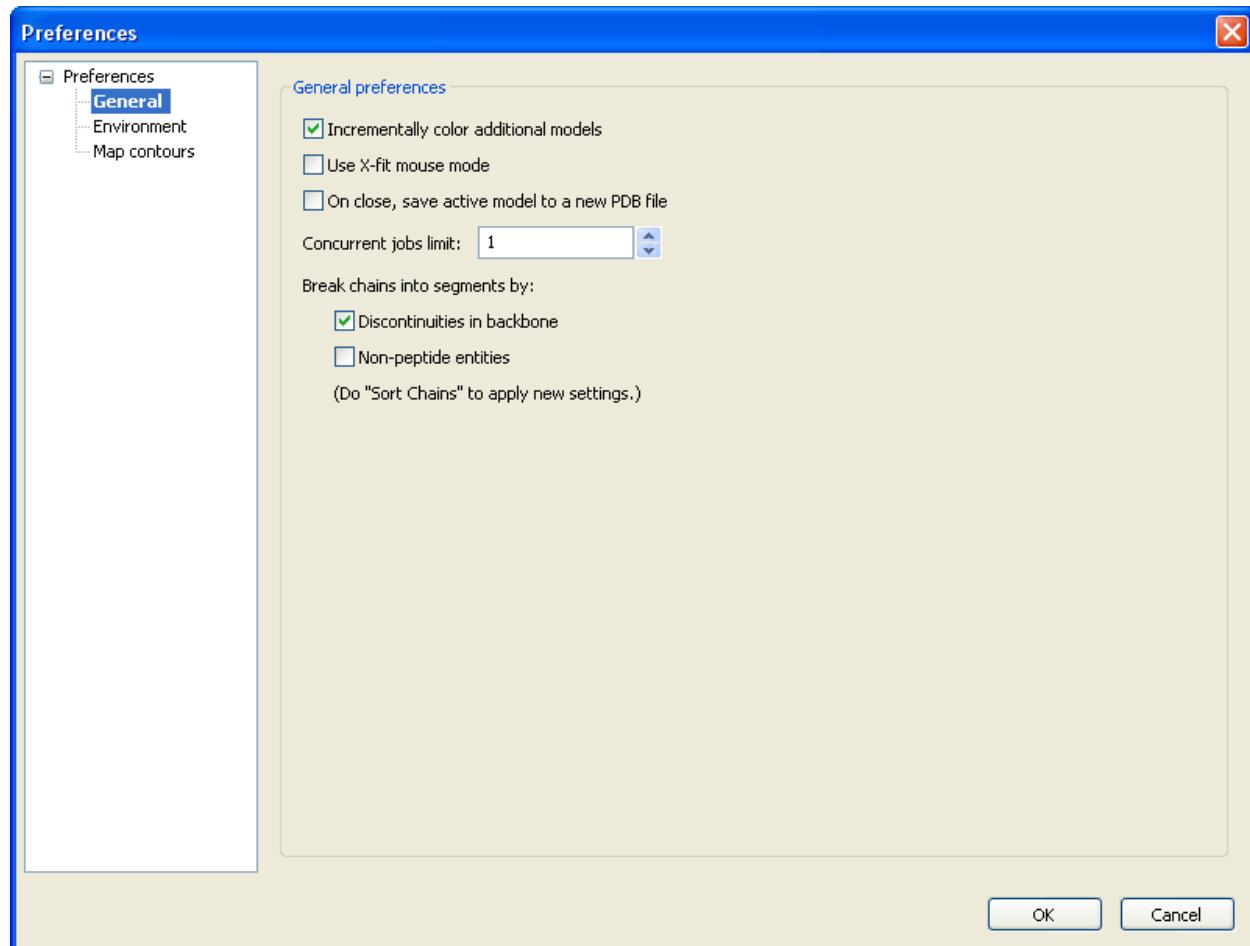


Figure 1.8 General Preferences

Checking the **Incrementally color additional models** option sets an automatic color-coding scheme for when loading multiple models, i.e. each additional model is distinguished by coloring with a different default color.

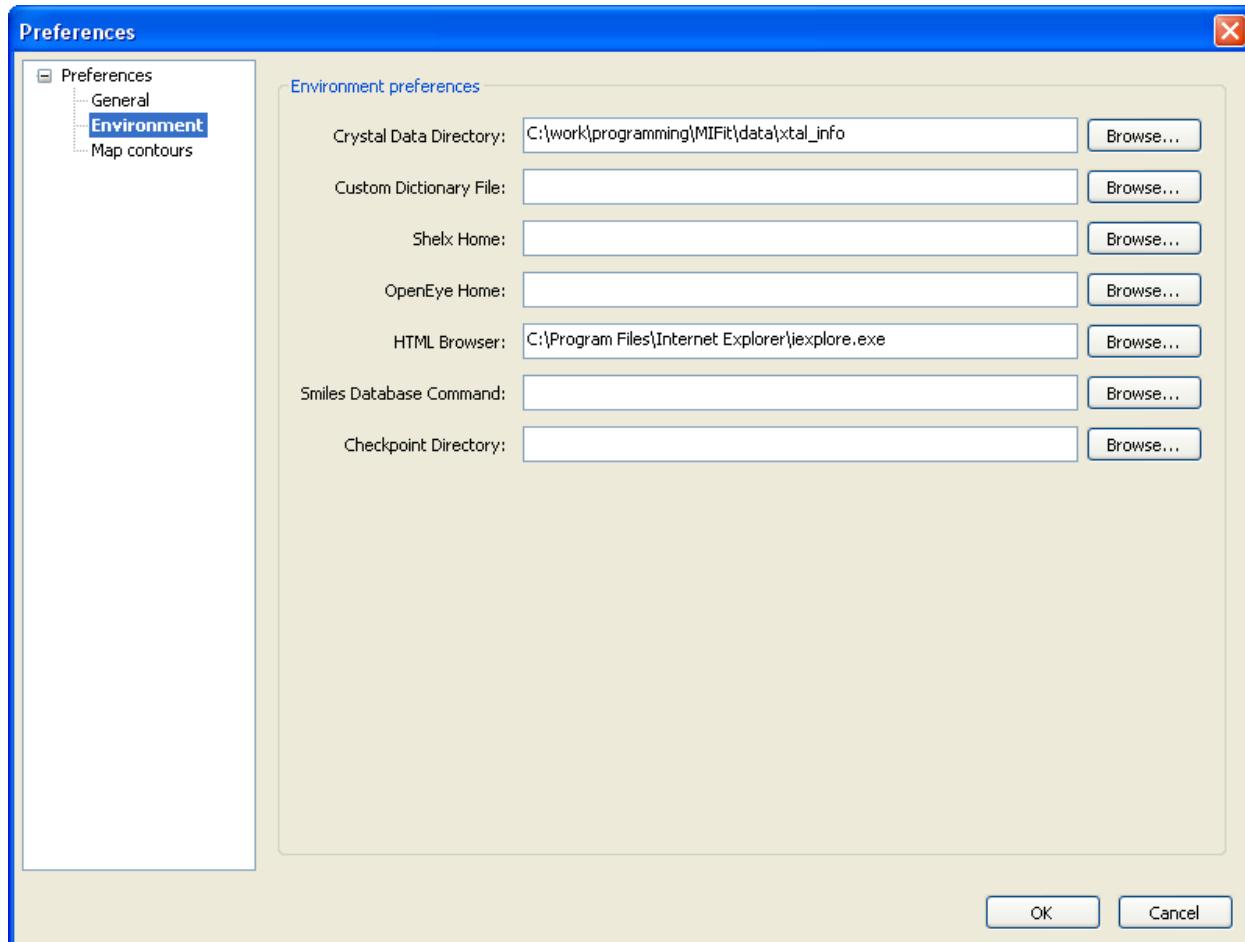
The **Use X-fit mouse mode** option may be used to set X-fit-like functions to a 3-button mouse and otherwise has no effect.

By default, MIFit will rewrite a session file containing current coordinates and information on the current session. Some users may prefer to also be prompted to write the current model to a PDB file and the **On close, save active model to new PDB file** option enables that preference.

MIFit provides interfaces to external crystallographic processes. However, on a single processor computer a MIFit session and a single background job (perhaps a refinement) will usually consume most of the computer memory and slow down interactive graphics. To avoid problems that would occur if inadvertently launching multiple background jobs the **Concurrent jobs limit** parameter may be set to limit the number of running jobs.

The MIFit structure navigation tree displays the structure broken into chain identifiers. This tree and may be configured via the checkboxes so as to further segment the individual chains upon meeting sequence number discontinuities (**Discontinuities in backbone**) and/or meeting entities outside the set of amino acids (**Non-peptide entities**). For most purposes it is useful to segment just according to discontinuities in the backbone sequence numbering.

Setting related to external software and paths are contained in the **Environment** preferences (Figure 1.9).



**Figure 1.9** Environment Preferences

The **Crystal Data Directory** contains a set of files that may be used to define the crystal cell and space group information for structure determination projects. However, in the current version of MIFit crystal information will usually be read directly from the input coordinate or data files and it will rarely be necessary for a user to setup and apply any independent crystal information. (If needed, the **File/Manage Crystals...** command may be used to establish new crystal data).

The **Custom Dictionary File** defines stereochemical information for the standard amino acids, cofactors and any ligands that you wish to add to the model. The default setting for this parameter is the dictionary file in the MIFit installation. To work on new projects, you may find it necessary to make a personal, customized dictionary file containing information on novel ligands to which you have write access.

The **Shelx Home** and **OpenEye Home** parameters are optional parameters that are needed to run components of these software systems through MIFit interfaces.

The **Shelx Home** parameter defines the path to the directory that contains the programs from the SHELX system. This path is passed directly to the applications in MIExpert by commands (**Job/SAD Phasing** or **Job/Refinement**) that have options to use components from the SHELX software.

The **HTML Browser** parameter should contain the path to a web browser executable. Some of the MIFit applications, including the SAD phasing interface and some of the automation options, return information to the user on job completion by opening a browser window.

The **SMILES Database Command** is used to specify a command that would execute a script which, if provided with a ligand registration number on the command-line, would return a SMILES string as standard-output. Typically the script would perform a look-up in a corporate small molecule database for that purpose. For example this parameter might be set to:

```
C:\Python24\python.exe "C:\my work\smilesdb.py"
```

in order to use Python to execute a script called smilesdb.py. The purpose of this command is to allow the MIFit Ligand Dictionary Editor access to SMILES data for construction ligand molecules and refinement restraints.

MIFit maintains a series of checkpoint coordinate files so that the user may recover work in the case of failure. These files are automatically deleted on a normal exit. It may be convenient to keep these files in a separate defined directory (rather than wherever the current working directory happens to be) and this file space may be defined with the **Checkpoint Directory** parameter.

Clicking on **OK** sets and saves the user parameters displayed within this interface.

# 2 Coordinate and map files

This chapter describes the process of loading coordinate and diffraction or density map data into MIFit.

## 2.1 Loading coordinate data

Atomic coordinate data may be loaded into MIFit using the **File/Open models, data, maps, etc...** command. This command opens a window (Figure 2.10) from which you may specify the type of coordinate file that you wish to load.

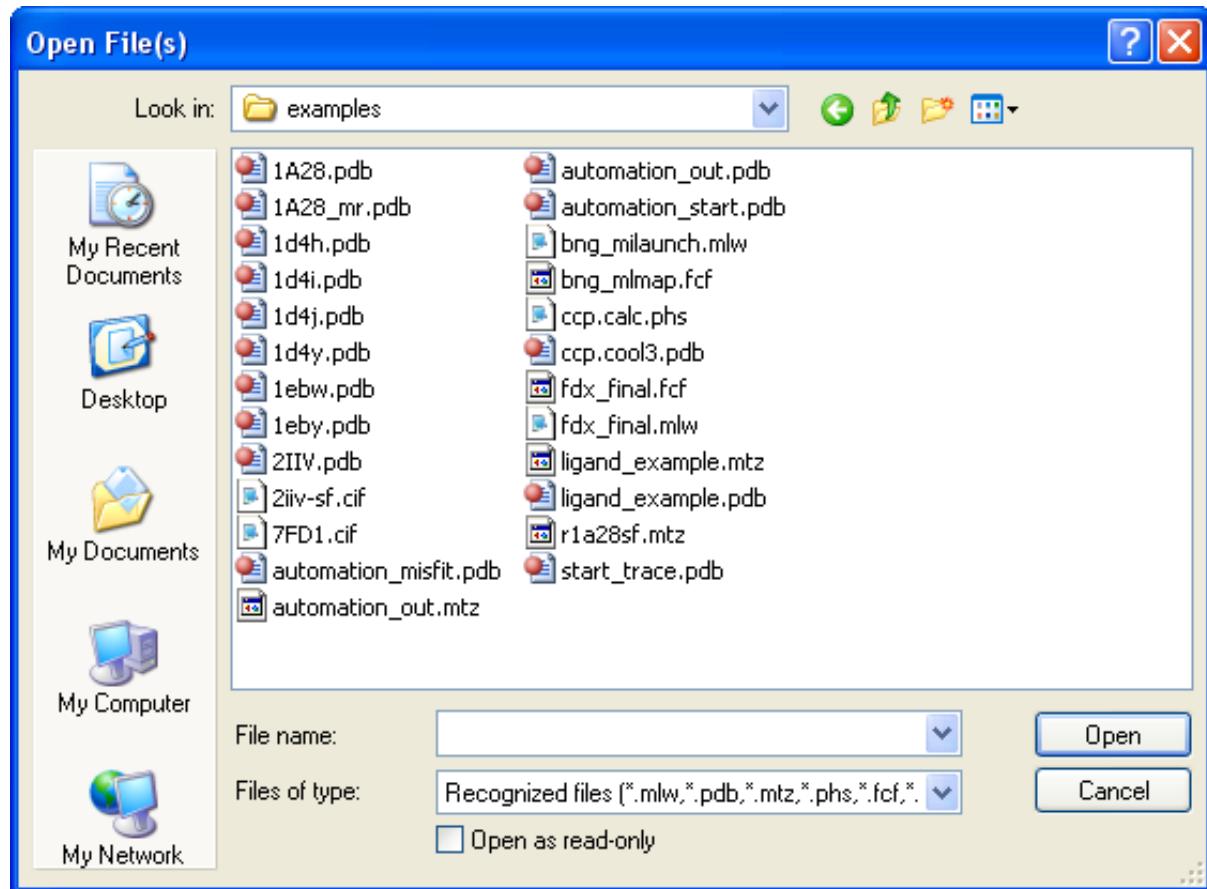
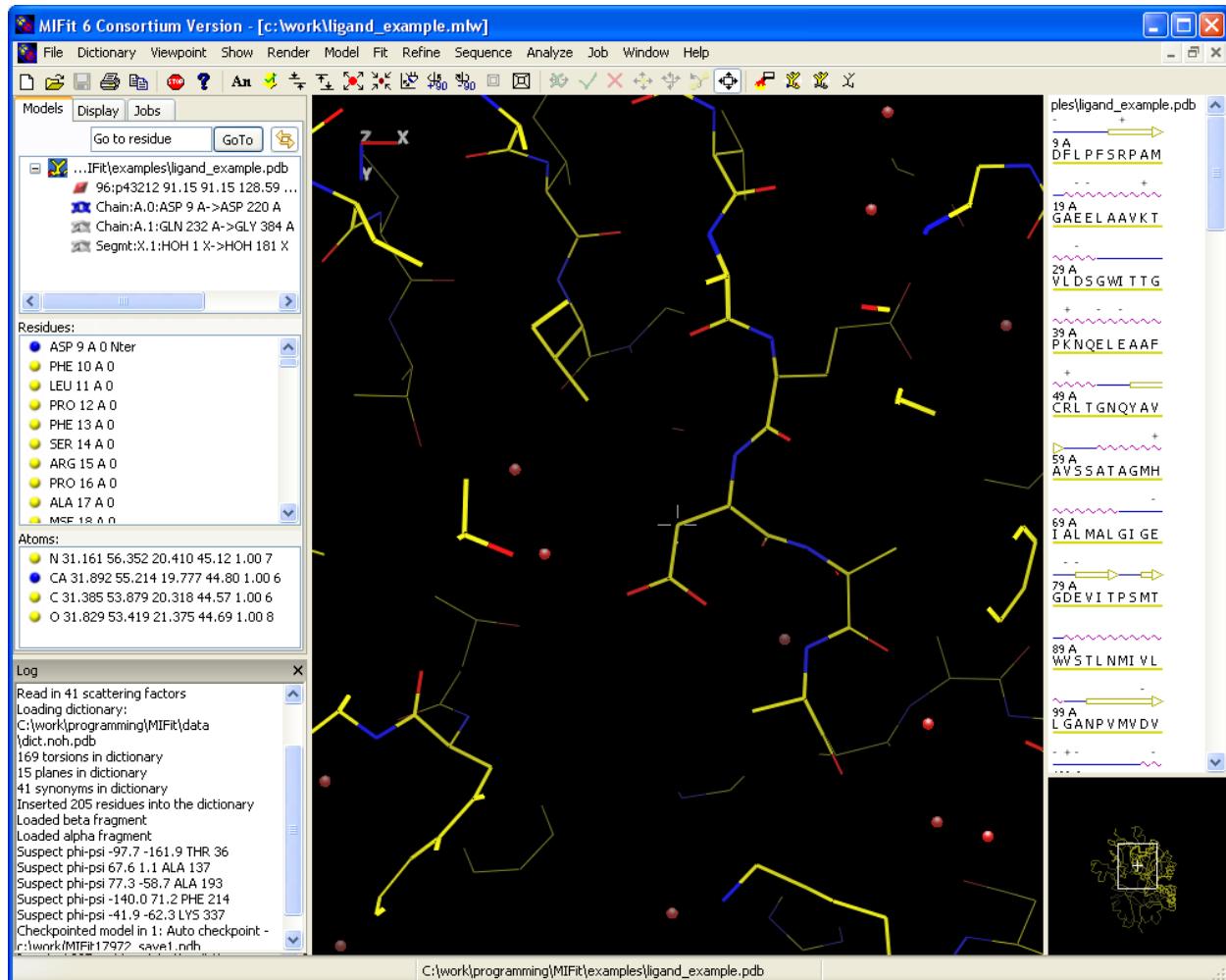


Figure 2.10 File/Open models, data, maps, etc... dialog

The options available from the **Files of type** selection allow selecting all recognized file types, just specific file types, or all files.

Once you have selected a coordinate file this structure data is immediately loaded and displayed (Figure 2.11).



**Figure 2.11** MIFit after loading a coordinate file

The three panes in the upper left contain a tree-control display of the loaded structure. Within each document, there may be one or more leaves for individual structures in the top pane. Each structure is further subdivided into chains and segments, and may also contain associated labels and annotations under the Display tab. The chains/segments for the active model are further expanded to display individual residues and the atoms within the target residue. Double-clicking on a residue in the tree control will center the view in the main canvas on that residue. Double-clicking on an atom in the tree control will center the view in the main canvas on that atom.

Alternatively, a **Go to residue** text box may be used to enter a chain-id residue number pair for centering the canvas view on the selected residue. If a chain-id is omitted the structure will center on the first residue number in the structure that matches the content of the text box. The icon next to the **GoTo** button is normally toggled on to synchronize the canvas view with selection.

The upper right window contains a sequence view with a schematic drawing of the calculated secondary structure. If multiple models are loaded the sequence displayed in the sequence view follows the active model.

Note that clicking on the calculated secondary structure toggles through helix, sheet and random coil secondary structure types. Editing the secondary structure description in this way is sometimes useful to adjust the appearance of secondary structure images rendered with the **Show/Secondary Structure/Make Ribbon** and **Show/Secondary Structure>Show Schematic Secondary Structure** commands.

Double-clicking on a residue in the sequence window will center the residue in the canvas window. With the mouse pointer resting over a residue, the residue's identity will be shown in the right half of the status bar at the bottom of the canvas, for example, as 'VAL 123 A'.

Similarly, when the mouse pointer rests over an atom in the canvas pane, information about this atom is displayed in the right half of the status bar, for example, 'CYS 123 CA (12.345 23.456 34.567) Occ=1.00 B=12.3'.

When an atom is clicked with the left mouse button in the canvas area this atom is put on a stack. A label of the form 'GLU 123 A CB' is shown to the right of the atom and the message area in the lower left corner of the canvas pane displays (in this case) '1: GLU 123 CB'. The stack may contain many atoms but the message area explicitly lists up to four of them with an additional line providing the number of any additional atoms, for example as, '+ 12 more...'.

The top of the atom stack may be cleared with the menu item sequence **Show/Stack/Clear top item** and the entire stack may be cleared with menu item sequence **Show/Stack/Clear stack**. Atomic labels and information in the message area may be erased with commands **Show/Labels/Clear Labels** or **Show/Canvas/Clear Message** respectively.

## 2.2 Loading electron density data

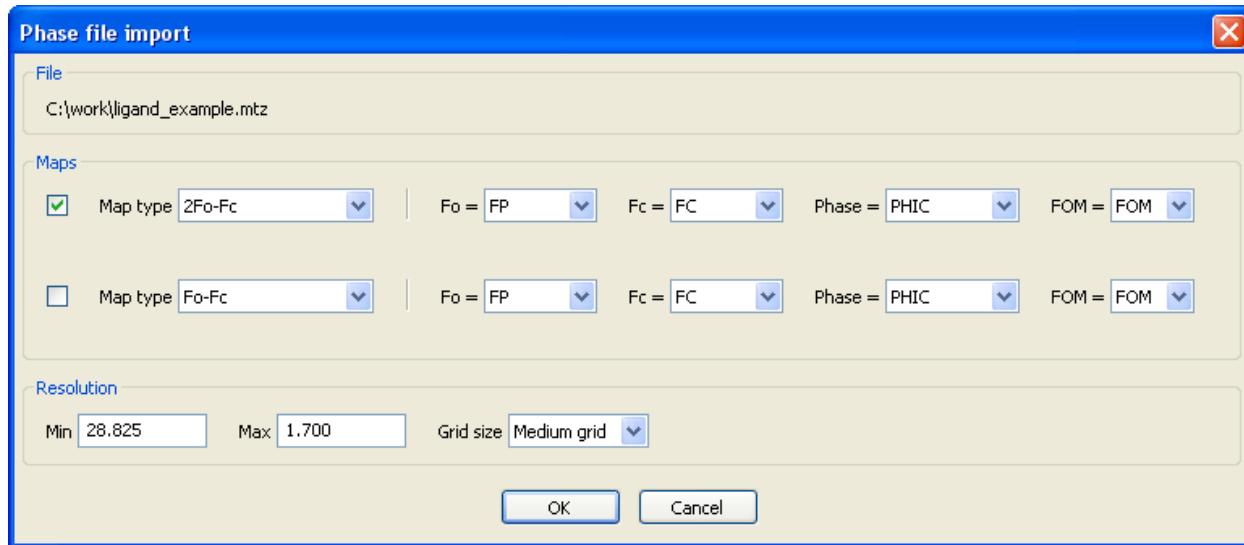
Once a coordinate data set is loaded MIFit may be used as a molecular structure viewer. However, for crystallographic model fitting we also need to load and display an electron density map. Electron density maps may be loaded into the document using the **File/Open models, data, maps, etc...** command.

When loading a data file, MIFit will read a set of structure factors and calculate a map on the fly via a Fast Fourier Transform. One reason that this mode of operation is often preferred to reading a pre-computed map is that diffraction data files require much less disk space than map files. Furthermore, calculating electron density maps within MIFit is a more flexible approach because the user may easily change the map resolution or alter the grid spacing of the map display. When reading a map file, MIFit requires a pre-calculated map be in the CCP4 or XtalView (FSFOUR) map format. There are a few situations where using pre-calculated maps is more appropriate than importing diffraction data. For example, a map may have been modified in real space by operations that are not easily reflected in reciprocal space and have no corresponding equivalent within the framework of MIFit.

The currently supported diffraction data formats are the XtalView format (*.phs*), the SHELX format (*.fcf*), the CCP4 format (*.mtz*), and the mmCIF format (*.cif*). It should be noted that many different types of data are stored in mmCIF files – obviously, the data should correspond to structure factors for this particular application.

The diffraction data format that is most commonly used, and the main input data format for automated structure solution and refinement processes (see Chapter 10), is the MTZ format from the CCP4. When data is loaded in this format the cell and space group information encoded in the

file header are automatically assigned. A dialog window appears that shows the mapping of MTZ data labels to established MIFit data types (Figure 2.12).

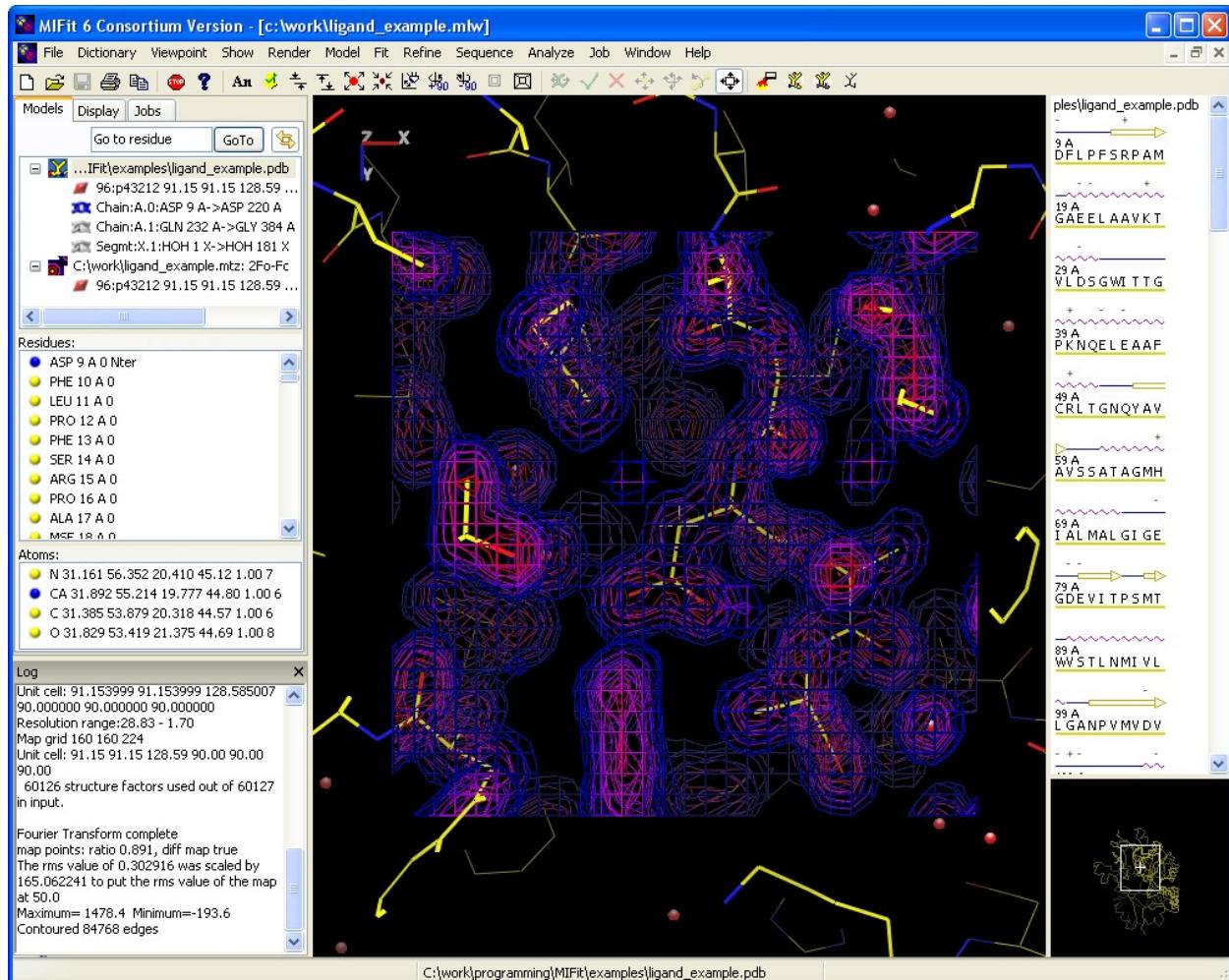


**Figure 2.12** Dialog for selecting MTZ data labels

The pull down menu for the **Map Type** parameter allows the user to select for the calculation of various standard types of electron density maps. The available map types include various types of difference and SigmaA weighted maps. MIFit attempts to infer the correct labels for different categories of MTZ file data. In the example shown in Figure 2.12 all of the data types were correctly identified. The pull down menus to the right of each data field may be used to reassign the MIFit data types to other MTZ column labels.

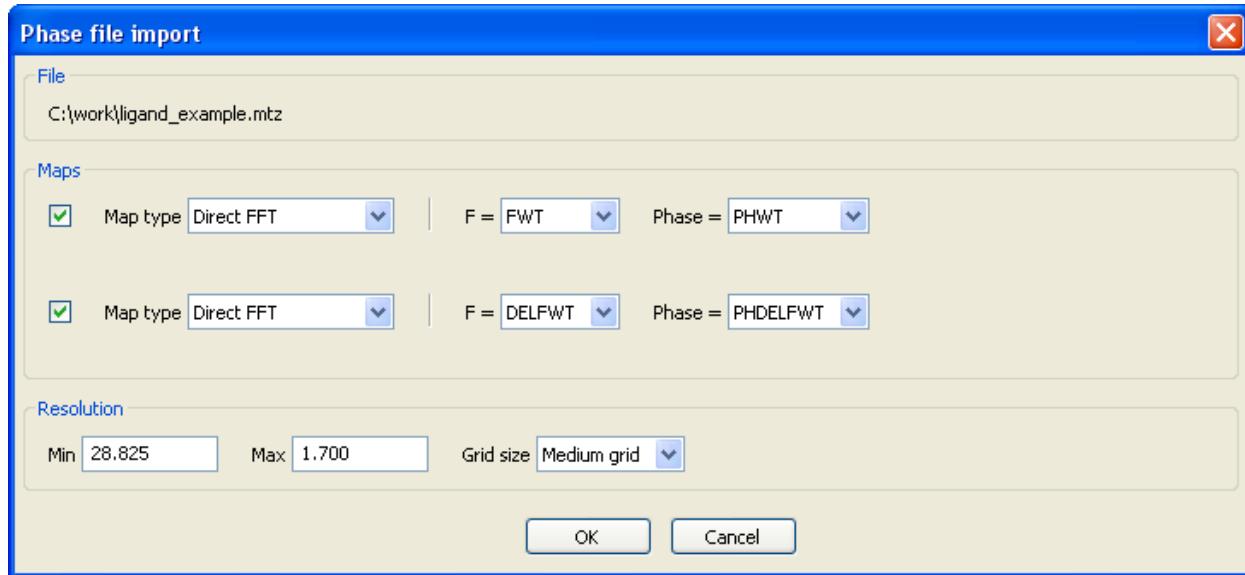
The resolution defaults (**Max Resolution** and **Min Resolution** parameters) are taken from the data limits of the data and are normally the most appropriate to apply. Setting the **Grid size** parameter to 'Medium grid' is appropriate for most model fitting work; a finer grid gives a smoother appearance and may give better images for presentation graphics, particularly with low resolution data.

After clicking on **OK**, the specified map is calculated and displayed in the canvas window (Figure 2.13).



**Figure 2.13** Model and map display

It is worth noting that there is a common situation where it might be useful to use pre-computed coefficients from a file rather than the true observed structure factor amplitude and associated phase values. Maps computed from likelihood weighted coefficients appear to have less bias and show more detail than other map types based on model phases. The CCP4/REFMAC5 program writes MTZ files containing pre-computed likelihood-weighted Fourier coefficients (FWT, PHWT and DELFWT, PHDELFWT) for the calculation of normal and difference maps. MIFit will identify this special case and pre-populate the **Map Type** with **Direct FFT**. When a Direct FFT map type is selected, MIFit will not be able to display any internally computed maps. To do so, reopen the data file and select the true observed structure factor.



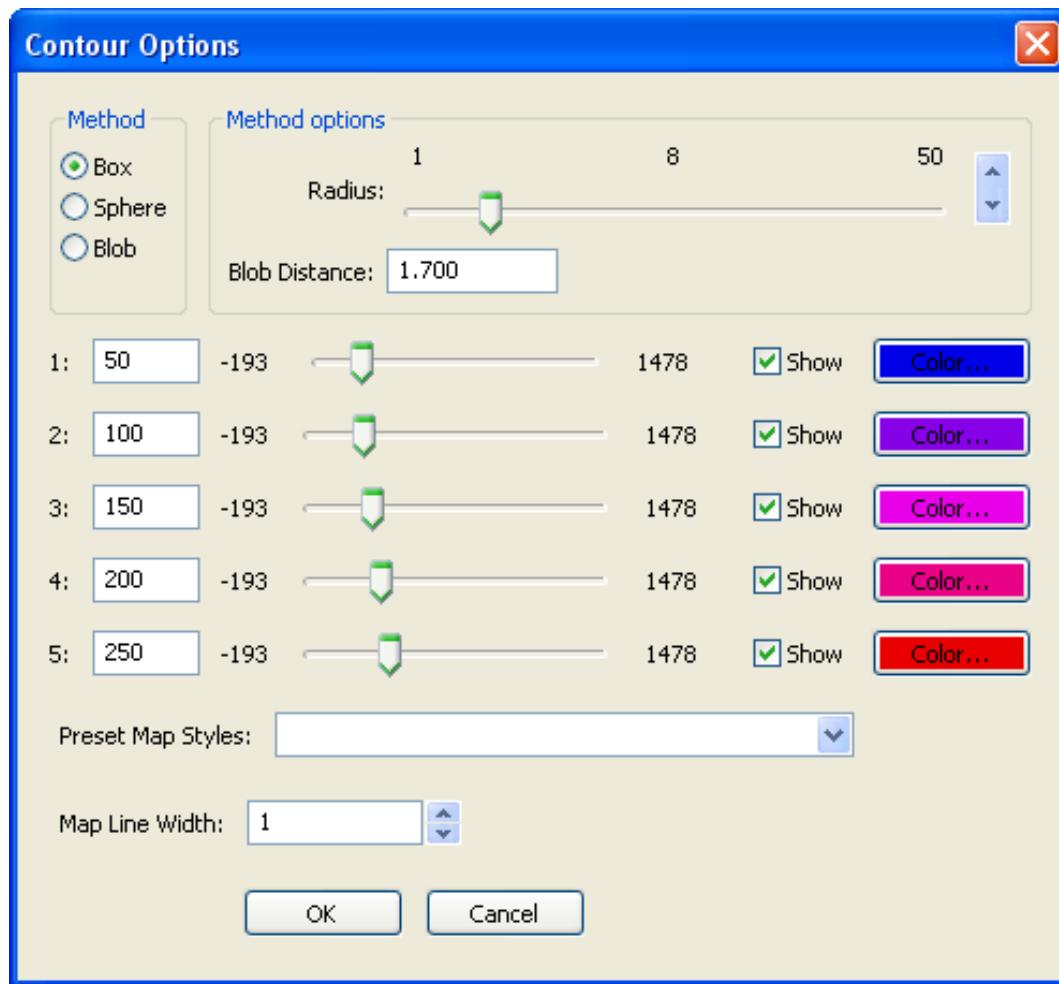
**Figure 2.14** Using Direct FFT

It may not be ideal to use the  $F_{\text{calc}}$  values from CCP4/REFMAC5 and have MIFit calculate a difference map (**Map Type** Fo-Fc) because of scaling issues with the low resolution data. It may be better to use either the pre-computed difference map coefficients (usually DELFWT and PHDELFWT) or re-compute the  $F_{\text{calc}}$  and Phase values from a model by selecting it for the Fc parameter.

There may be situations where the cell and space group data encoded in the data file header are incorrect (possibly the data were only processed to the correct point group) or absent (the .phs file format does not contain this information). In this situation the information may be loaded from a previously created crystal file. Right-clicking on the crystal icon beneath the icon representing the diffraction data provides access to an **Edit** option which contains a menu for loading new crystal files (Figure 1.4).

## 2.3 Controlling the map display

Once loaded, the electron density map display may be modified by right-clicking on the map in the Models tree (i.e. the map icon at the bottom of the tree in Figure 2.13) and selecting the **Contour options** command which will bring up the dialog shown in Figure 2.15.



**Figure 2.15** Contour Options dialog

The **Method** and **Radius** parameters control the volume of electron density that is displayed. The **Method** parameter may be used to select whether the map will be displayed as a box or sphere around a selected point. Depending on the number of contour lines displayed, the fitting operation that is to be performed and the speed of the computer, it is usual to set a map display radius in the range 5-10Å.

The **Blob Method** is used to display electron density only within the **Blob Distance** of selected residues. The selection of target residues is by the same methods that are used for model fitting (for example, selection options under the **Fit** pull down menu) and the blob parameters are inactive unless such a selection has been made. The blob option is used for creating presentation graphics and should be applied with caution; there is a fine distinction between omitting density for clarity and eliminating surrounding map noise so as to mislead an audience into thinking that the map quality is better than it really is.

The **Preset Map Styles** pull down menu allows selection of a map display style from a set of standard contour level, colors and map types. Maps calculated within MIFit are scaled so that one ‘sigma’ (the root-mean-square density fluctuation in the map) equals 50 internal map units.

Therefore the contour levels shown in Figure 2.15 (50, 100, 150 ...) correspond to density levels at 1,2,3,4,5 sigma.

The **Show** checkboxes may be used to eliminate some of the contour levels from the display. It will often be the case that a map appears cluttered by too many contour surfaces and some of them can be switched off with these options. The **Color...** options may be used to change the color of an individual contour surface from the default colors specified by the **Preset Map Styles** choices.

The **Map Line Width** parameter may be used to change the line width used in the map display. For most interactive work a line width of '1' is most appropriate but it may be useful to display thicker lines when exporting images for presentation.

## 2.4 Moving around the electron density map

MIFit automatically re-computes the display of the electron density map when the display is re-centered on a new atom or translated by more than a few angstroms. Full crystal symmetry information is imposed on the map so there are no problems with moving outside a pre-computed map volume. The **Viewpoint/Center Visible Density** command may be useful for re-centering the displayed density to lie in the center of the main canvas.

In order to locate significant density features that are not accounted for by the current model, the **Find Ligand Density** from the map context menu is available. This tool is parameterized to search electron density difference maps for relatively large density blobs (i.e. not features that could normally be accounted for by ordered water molecules). Putative ligand densities found by this tool are parameterized by a closely packed set of pseudo-atoms and may be picked as CLUST objects from the MIFit document tree. Larger clusters in this list correspond to more extensive density features. If no CLUST objects appear in the document tree after executing this command then no density blobs were sufficiently extended to qualify as ligand densities.

# 3 Menus and toolbars

There are several ways in which a user may interact with MIFit:

- via keyboard shortcuts
- via menu items
- via toolbar items

Keyboard shortcuts are discussed in Chapter 4. In this chapter we discuss and compare menu and toolbar items. Not all menu items have equivalent toolbar items but essentially all toolbar items have equivalents from among the menu items.

In addition, a right-click in the main or navigator canvas areas makes a context menu available. This menu contains frequently used menu items that are otherwise distributed among several main menus.

## 3.1 Startup menu / toolbar

 <b>File/New</b>	Create a new structure document
 <b>File/Open models, data, maps, etc...</b>	Open an existing structure document
 <b>File/Save Session,</b> <b>File/Save Session As...</b>	Save a structure document
 <b>File/Print...</b>	Print the canvas
 <b>File/Copy Canvas</b>	Copy the canvas to the clipboard (Windows only)
 <b>File/Manage Crystals...</b>	Used to abort a long operation
 <b>Help/Help...</b>	Create or edit a crystal
	Invoke help (this manual)

**Table 3.1** Menu and tool bar options available at startup

Table 3.1 shows the toolbar items and their menu item equivalents that are available when MIFit is started up. The **File/Save Session**, **File/Save Session As...** and **File/Print...** menu items are initially inactive.

## 3.2 General menu / tool bar

Once either an existing structure document has been opened or a new structure document been created, the menu and the toolbar expand to contain more items. Table 3.2 shows the additional toolbar items and their menu equivalents. These additional toolbar items are appended to the

items listed in Table 3.1. The additional menu bar items are inserted between the File and Help pulldown menu bar items on Linux systems and between the File and Window pulldown menu bar items on Windows systems.

	<b>Show/Annotation/Add annotation to model</b>	Add annotation at canvas center
	<b>Model/Checkpoint Model</b>	Save model state in PDB format temporary file
	<b>Viewpoint/Slab In</b>	Decrease display depth in z-direction
	<b>Viewpoint/Slab Out</b>	Increase display depth in z-direction
	<b>Viewpoint/Zoom Out</b>	Zoom out to expand view of model
	<b>Viewpoint/Zoom In</b>	Zoom in to shrink view of model
	<b>Viewpoint /Top View</b>	Display view down z-axis
	<b>Viewpoint /Rotate View +90</b>	Rotate about y by +90°
	<b>Viewpoint /Rotate View -90</b>	Rotate about y by -90°
	<b>Viewpoint /Decrease Perspective</b>	Decrease canvas perspective
	<b>Viewpoint /Increase Perspective</b>	Increase canvas perspective

**Table 3.2** General toolbar items and their menu equivalents

### 3.3 Modeling menu / toolbar

The items in the modeling portion of the toolbar are initially active. When an atom is selected in the canvas by clicking on it, the **Fit Residue** option becomes active. When this item is activated or when the **Fit/Fit Residue** command is selected all the **Fit** items except for **Fit/Torsion** are activated. The **Fit/Torsion** option becomes active when a bond defining a torsion angle has been selected.

	<b>Fit/Fit Residue</b>	Fit residue containing top-of-stack atom
	<b>Fit/Apply</b>	Apply modifications to selected residue(s)
	<b>Fit/Cancel</b>	Cancel all modifications
	<b>Fit/Translate</b>	Switch to translating selected residue(s)
	<b>Fit/Rotate</b>	Switch to rotating selected residue(s)
	<b>Fit/Torsion</b>	Apply torsion motion to selection

	<b>Fit/Center</b>	Right-mouse pans viewpoint
---	-------------------	----------------------------

**Table 3.3** Modeling toolbar items

When the model is modified and the **Fit/Accept Fit** command is selected the **Save file** toolbar item and the analogous menu items **File/Save Session** and **File/Save Session As...** become active.

### 3.4 Display menu / toolbar

	<b>Render/Color</b>	Color Tool
	<b>Show/Atoms or Show/Residues</b>	Show Tool
	<b>Show/Sidechains</b>	Show Sidechain Tool
	<b>Show/Atoms or Show/Residues</b>	Hide Tool

**Table 3.4** Display toolbar items

Color selection capabilities may be accessed with the **Render/Color** menu options or with the **Color Tool** toolbar item.

# 4 Keyboard shortcuts

Numerous keyboard shortcuts are available within the MIFit software in order to provide efficient control over document handling, model display and model-fitting. Once the shortcuts have been learned, many common operations can be accessed much more rapidly than through the MIFit menus or toolbar.

The tables in this chapter list the keyboard shortcuts that are currently available in MIFit. The notation ‘/’ is used to indicate two shortcuts that usually have opposite actions. The notation ‘Shift+’ indicates simultaneous selection of the Shift key with the following key. The notation ‘Ctrl+’ indicates simultaneous selection of the Ctrl key with the following key. Similarly, the notation ‘Alt+’ indicates simultaneous selection of the Alt key with the following key.

## 4.1 Keyboard shortcuts for document handling

The actions of the three keyboard shortcuts associated with menu and toolbar commands were discussed in Chapter 1.

 <b>Ctrl+N</b>	<b>File/New</b>	Create a new structure document
 <b>Ctrl+O</b>	<b>File/Open models, data, maps, etc...</b>	Open an existing structure document
<b>Alt+X</b>	<b>Exit</b>	Terminates program

**Table 4.5** Keyboard shortcuts associated with menu and toolbar items

## 4.2 Keyboard shortcuts for display

<b>X</b>	View along x-axis (from origin)
<b>Y</b>	View along y-axis (from origin)
<b>Z</b>	View along z-axis (from origin)
<b>Left / Right</b>	Rotate left / right about y-axis
<b>Up / Down</b>	Rotate up /down about x-axis
<b>Shift+I / Shift+O</b>	Slab in / Slab out (along z-axis)
<b>I / O</b>	Zoom in / Zoom out

**Table 4.6** Keyboard shortcuts controlling model display

The notations **Left/Right/Up/Down** refer to the arrow cursor keys.

## 4.3 Keyboard shortcuts for side-by-side stereo viewing

MIFit currently supports side-by-side stereo and also contains a full screen mode (**ESC** keyboard shortcut) that allows the canvas to be expanded to fill the entire monitor screen.

The side-by-side stereo display splits the canvas into a left and right half and displays the left/right eye image in the left/right half respectively. The default stereo display is straight-eye rather than cross-eye stereo. Side-by-side stereo display may also be controlled with the menu command **Viewpoint/Stereo**.

	Toggle stereo
<b>A / S</b>	Decrease / increase stereo angle
<b>Shift+A</b>	/ Decrease / increase stereo separation
<b>Shift+S</b>	
<b>Ctrl+Shift+S</b>	Toggle cross-eyed to wall-eyed stereo
<b>Q / W</b>	Decrease /increase stereo separation for Ligand Dictionary Editor
<b>X</b>	Toggle cross-eyed to wall-eyed stereo for Ligand Dictionary Editor

**Table 4.7**      Keyboard shortcuts to alter software stereo settings

Use of hardware stereo requires a quad-buffered video card, CRT monitor, and LCD shutter glasses. If the hardware is available, it may be activated using the **Viewpoint/Use Hardware Stereo** command.

## 4.4 Keyboard shortcuts for model fitting

Most of the interactive model-building shortcuts are only active while fitting the model, i.e. when one or more residues are selected for modification.

[   PageUp	Go to N-terminus of chain
]   PageDown	Go to C-terminus of chain
< / >	Insert MRK residue before / after present residue
Space	Go to next residue
<b>B</b>	Go to previous residue
;	Apply fit / Accept refinement
<b>C</b>	Select next conformer (side-chain)
<b>D</b>	Delete residue

<b>F</b>	Fit residue
<b>R</b>	Replace and fit residue (side chain refitting)
<b>Ctrl+R</b>	Refine residue
<b>Shift+W</b>	Add water molecule at cursor center
<b>Delete</b>	Delete atom
<b>1 / 2 / ... / 5</b>	Select $\chi_1$ (chi-1), $\chi_2$ , ..., $\chi_5$ side-chain torsion angle

**Table 4.8**      Keyboard shortcuts active during model fitting

# 5 Model Building and Refitting

MIFit contains a full set of tools for building protein models into density maps. MIFit's capabilities include baton style C $\alpha$  chain-tracing, converting C $\alpha$ -traces to poly-alanine chain segments, the application of pentamer libraries to refit main chain atoms, convenient side chain mutation and density refitting, automated and manual water-picking, ligand placement and adjustment.

## 5.1 Chain tracing

Nowadays, the major portions of structure models are initially built into experimentally phased electron density maps using automated procedures (for example, with the arp/wARP, RESOLVE or MAID software) or were derived from molecular replacement searches. Nevertheless, it may be necessary to manually trace portions of the map that were either not fit by these methods or were misfit.

'Smart' baton-style placement of C $\alpha$  markers (residue type MRK) is available via the **Model/Add MRK Before** and the **Model/Add MRK After** commands. Convenient keyboard short cuts equivalent to these commands are '<' and '>'. The function of these commands is to predict (based on density and geometric constraints) and add a C $\alpha$  atom either before or after the current C $\alpha$  atom. If the predicted position appears to be incorrect then the right mouse button may be used to adjust the C $\alpha$  site at fixed distance of 3.8Å from the previous C $\alpha$ .

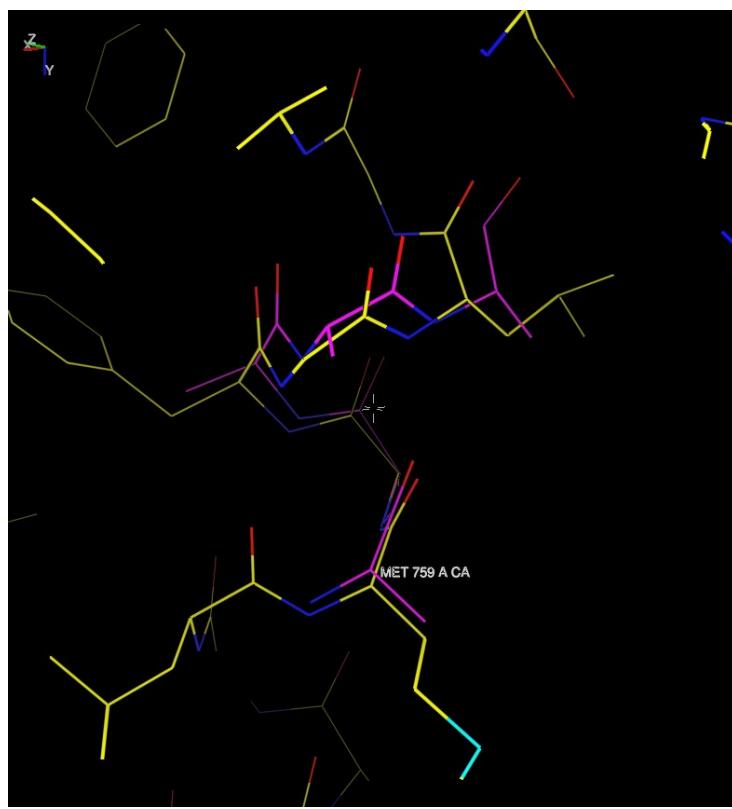
The **Fit/Accept Fit** command may be used to accept the current set of positions and cancel the chain tracing.

Selecting the end points of the sequence of MRK points and then selecting the **Model/Poly-ALA Range** will convert the C $\alpha$  trace to a poly-alanine trace. Alternatively, if the entire chain consists of MRK points then the **Model/Poly-ALA Chain** command may be used to carry out the conversion. The accuracy of the poly-alanine trace depends on the accuracy of the C $\alpha$  placement. The endpoints of the trace tend to be the least reliable because the atomic positions are less constrained than those in the central region. In the absence of a following residue, the positions of atoms in the C-terminal amino acid are not fully defined so this amino acid remains a MRK residue.

## 5.2 Refitting backbone atoms

The MIFit software contains commands to check and refit protein backbone based on the use of a pentamer fragment library. A tool is also available for 'flipping' individual amino acids so that the carbonyl oxygen is rotated about the peptide plane.

To employ the pentamer library the user must first select an atom from the model displayed in the main canvas. The first C $\alpha$  in the pentamer will belong to the amino acid associated with the selected atom. Next, selecting the **Fit/Fix Backbone/Suggest Backbone Match** command will display in purple the pentamer that best fits the current structure (Figure 5.16)



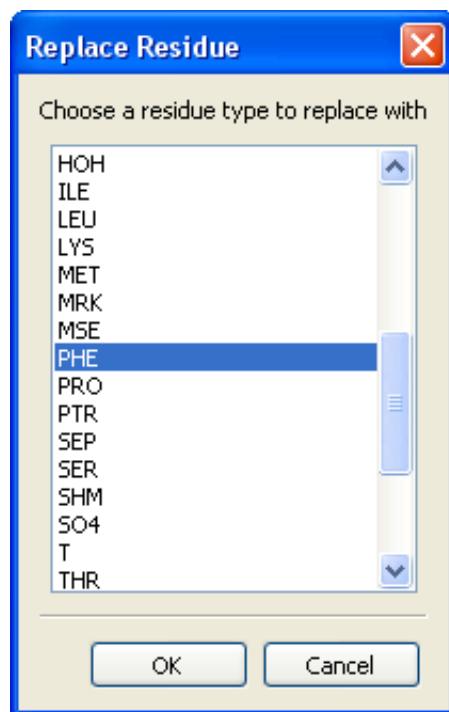
**Figure 5.16** Pentamer match for model structure following A 759 Met

Portions of the associated protein backbone may be shifted to match the displayed pentamer using the **Fit/Fix Backbone** commands **Replace Middle 3**, **Replace First 4**, **Replace Last 4** or **Replace All 5**. If none of these commands appears to be useful then the **Fit/Fix Backbone/Clear Backbone Match** command may be used to eliminate the pentamer from the display and the tree control.

To simply flip an individual peptide the user should click on an atom in the peptide plane and then select the **Fit/Fix Backbone/Flip Peptide** command. The peptide will then be observed to rotate so that the carbonyl atom appears on the opposite side of the peptide plane.

### 5.3 Building and fitting side chain atoms

Side chain groups may be changed from one type to another by selecting an atom in the main canvas and then using the **Model/Replace residue** command. The action of this command is simply to mutate one amino acid type to another without reference to any electron density map. After selecting this command a dialog window (Figure 5.17) appears from which the new amino acid type may be selected.



**Figure 5.17** Dialog for mutating amino acids

An extremely useful command for both model-building and model correction (i.e. adjusting the position of a side chain to improve the fit to the electron density) is the **Model/Replace and Fit** command. This command also spawns the dialog selection for mutating an amino acid, with default value set to the original amino acid type. The action of this command is to replace the amino acid side chain but also to make the best fit to the electron density. An alternative method for selecting this command is the keyboard shortcut ‘R’. The proposed fit (with the amino acid remaining colored green to indicate that it is still ‘live’) may be accepted or rejected using the toolbar icons and .

It is also possible to cycle through possible conformers for a particular amino acid type by first selecting an atom in the main canvas and then using either the **Model/Next Conformer** command or the keyboard shortcut ‘C’.

A fitting method that allows arbitrary interactive adjustment of side chain torsion angles is to select the amino acid and then choose either the **Fit/Fit residue** command or the keyboard shortcut ‘F’. The numeric keyboard keys (1, 2 ...) may then be used to select a torsion angle ( $\chi_1, \chi_2 \dots$ ) for rotation. A grey arrow appears on the screen to indicate the selected torsion and the right-mouse button controls the torsion about the indicated bond. Toolbar icons and may be used to accept or reject the fit.

## 5.4 Refitting and adding individual residues

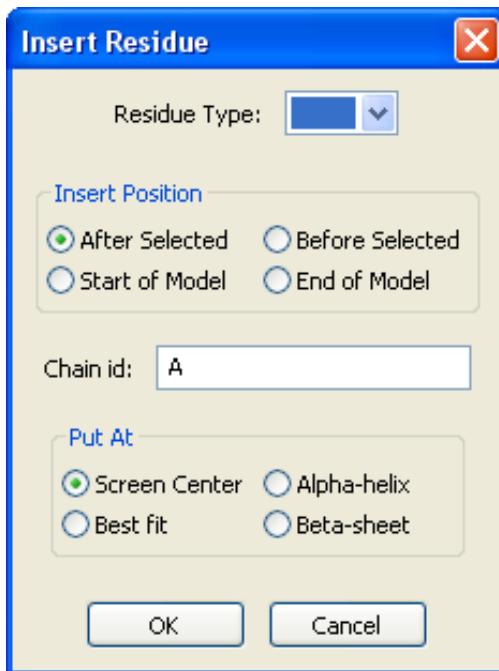
An entity (amino acid, water etc) in the current model may be activated for fitting by using either the **Fit/Fit residue** command or the keyboard shortcut ‘F’. Active entities are colored green in the main MIFit canvas. Entities may be translated by selecting the translate icon (+) and adjusted using the right-mouse button (the left-mouse button retains control of the view orientation).

The rotation icon ( ) may be selected to change the right-mouse button to control the rotation of the active entity. Clicking on a bond in the active entity sets an arbitrary torsion with the rotating atoms at the arrow head marking the bond. The atom closest to selected position is marked with the arrow head. The torsion icon ( ) may be selected to set the right-mouse button control to the torsion.

Similarly, atomic groups other than individual amino acids may be activated for fitting. The **Fit** commands **Fit Atom**, **Fit Residues**, **Fit Residue Range**, **Fit Atoms** and **Fit Molecule** may be used in conjunction with the selection of atoms from the main canvas to activate portions of the structure for fitting. Note that for multiple selections the list in the bottom left hand corner of the main canvas records the entities currently in the stack. The **Show/Stack** commands may be used to manipulate the contents of the stack.

To accept a fit the **Fit/Accept Fit** command or the icon may be used. To cancel a fit the **Fit/Cancel Fit** command or the icon may be used. If you do not wish to deactivate the model but do wish to return to the original position of the active entity then the **Fit/Reset Fit** command may be selected.

The **Model/Add residue** command may be used to add new entities to the model.



**Figure 5.18**    Model/Add residue dialog

The pull down menu corresponding to the **Residue Type** parameter in the resulting dialog box (Figure 5.18) is used to specify entity that is to be added to the model. The **Insert Position** options describe where in coordinate set the new entity will be placed. The **Chain id** parameter box may be filled if you wish to set a new chain id for the inserted residue. This option is not normally used for protein model-building but may be useful if the entity added is a ligand.

The **Put At** options may be used to simply add the new residue at the screen center, extend an existing chain in **Alpha-helix** or **Beta-sheet** conformation or extend the existing chain with **Best fit** of the new residue to the electron density.

The **Model/Delete Residue** (keyboard shortcut **D**) may be used to remove a selected entity from the structure, i.e. after using this command the entity disappears from both the graphics display and the tree control and is no longer a part of the model.

The **Model/Rename Residue** may be used to change the residue number for an individual entity.

The **Model/Delete atom** command (keyboard shortcut **Delete**) may be used to delete an individual atom from the structure i.e. the last atom selected from the canvas disappears from the canvas display and the tree control and is no longer part of the model. This command is mostly useful for pruning disordered side chains.

## 5.5 Modeling discrete disorder

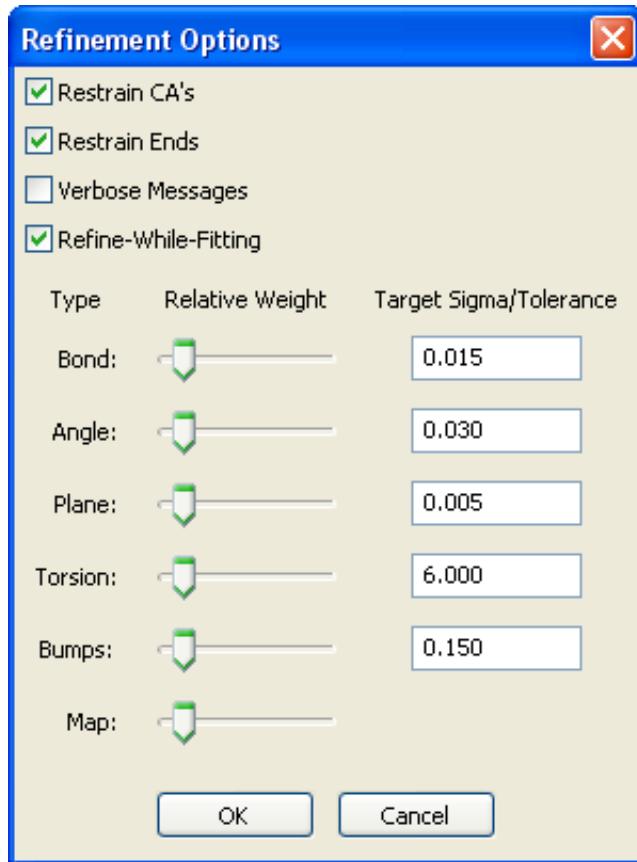
Discrete disorder (the presence of multiple well defined conformations) is very simple to model with MIFit and is handled transparently by the refinement program CCP4/REFMAC5. Well refined structures at resolutions better than 2Å will usually show some abnormal side chain densities that may be reasonably interpreted in terms of two major side chain conformations.

To setup an amino acid for modeling discrete disorder, select an atom within the amino acid in which the disorder needs to be modeled and activate it with **Fit/Fit residue** (the keyboard shortcut **F**). The **Fit/Disorder/Split Fit** command may then be used to duplicate the entire amino acid, with the duplicate copy slightly offset from the original copy for clarity. The duplicated copy may be manipulated with the interactive fitting tools i.e. the translation () , rotation () and torsion adjustment () options available from the tool bar.

More commonly, the required modeling requires split just a side chain or part of a side chain into multiple conformations. To do this, specify a side chain torsion angle using one of the methods described in Section 5.3. The **Fit/Disorder/Split Fit** command may then be used to split the side chain at the specified torsion and the new copy may be rotated into position. As usual, the tool-bar icons  and  may be used to accept or reject the fit.

## 5.6 Local structure refinement

The entire molecule or portions of the molecule may be optimized within MIFit using commands beneath the **Refine** menu. The **Refi/Refi Options** command (Figure 5.19) shows the options for the optimization. It will be noted that the refinement target includes a term for matching the electron density as well as tethers for the C $\alpha$  positions and the ends of the refinement zone.



**Figure 5.19** Options for structure optimization

An active region for structure optimization is colored pale blue. Following optimization, the changed structure may be accepted with the **Refine/Accept Refine** command, canceled with the **Refine/Cancel Refine** command or allowed to remain active but reset to the unrefined position with the **Refine/Reset Refine** commands.

Once a region is active, holding down the keyboard space bar results in a continuous refinement of the selected region to convergence. This is often a powerful optimization technique, resulting in large shifts in the atomic coordinates to fit the electron density.

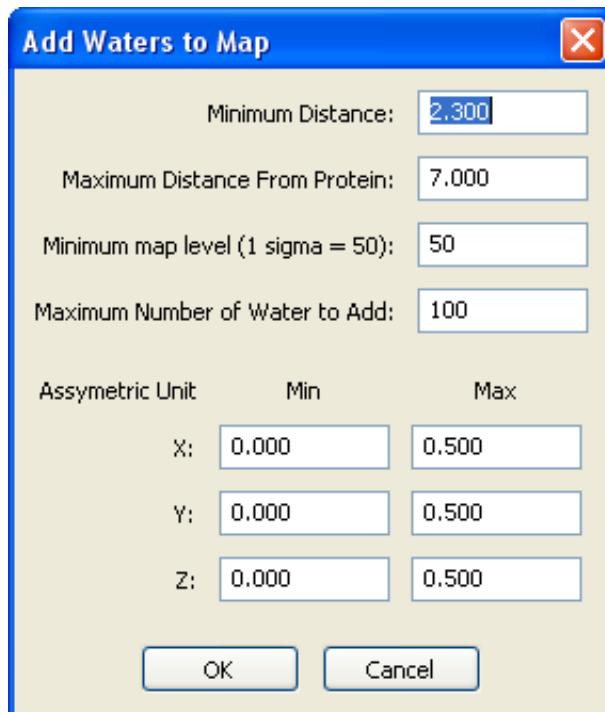
The entire molecule may be optimized using the **Refine/Refine Molecule** option. More usually, an amino acid is selected for optimization. The individual selected amino acid may be optimized using the **Refine/Refine Residue** command or the amino acid the amino acids on either side of it may be included in the optimization by using the **Refine/Refine Local Region** command.

Selecting two atoms by clicking on them in the main canvas and then applying the **Refine/Refine Range** command will result in all amino acids between the first and the last selected atom being included in the optimization.

## 5.7 Adding water molecules

The MIFit system contains several mechanisms for adding water molecules to the atomic model. One method is to fold water-picking into automated refinement cycles through the **Job/Refinement** interface (Chapter 10).

MIFit also provides an independent method for water-picking using the **Model/Add Waters...** command (Figure 5.20).



**Figure 5.20** Dialog to set parameters for water picking

This command is intended to pick waters from a map loaded into the main canvas. This will usually be a difference map, in which densities for ordered water molecules are visible as peaks around the protein. The resulting dialog box from this command contains criteria for selecting waters within a border around the protein (via the **Minimum distance from protein** and **Maximum distance from protein** parameters). The **Minimum map level** parameter sets the threshold level for density to be considered as a possible water site. For a difference map this would ordinarily be set to a value of 3-4 sigma; for a conventional map a value of ~1 sigma might be appropriate. The **Maximum number of waters to add** parameter is useful to avoid overloading the map with water molecules. Since the optimal density map level parameter is somewhat problem dependent this is useful to prevent adding too many waters if that value is initially set too low. The **Asymmetric unit** parameter need not ordinarily be changed – it will normally correspond to a unique region of the crystal cell.

Upon selecting **OK** the water fitting process run and the new waters will be displayed in the main canvas. The number of waters added will be reported in the lower left of the status bar.

These methods serve to fit the majority of ordered water molecules in a protein structure. To delete an individual water molecule the most efficient method is to select it by clicking on it in the main canvas and then right-click to obtain a menu containing **Delete** residue command. Selecting this command will delete that water molecule.

Individual waters may either be added using the **Model/Add residue** command, setting the **Residue Type** pulldown menu to ‘HOH’ (the standard code for a water molecule). A potentially faster method is to use the keyboard shortcut **Shift+W**. This option caused the new water to appear at the mouse position in the main canvas (i.e. not at the canvas center).

## 5.8 Interactive ligand fitting

After loading a protein model into the main canvas, the **Model/Add residue** command may be used to include the ligand entity in the model (Figure 5.17).

The **Residue Type** parameter is a pulldown menu that allows selection of any entity in the MIFit dictionary. The **Insert Position** parameter controls where in the coordinate file the ligand atom record will appear. The **Put At** parameter controls position of the ligand in the structure – for simply adding a ligand to the model the ligand density should be moved to the screen center and the **Screen Center** option should be selected.

Unlike many model building programs, MIFit does not require any dictionary data in order to manipulate (rotate and translate) a ligand molecule or rotate a part of the ligand about an arbitrary torsion angle.

In a scenario where the atomic model (including ligand) and electron density map are loaded in the main canvas and you wish to fit the ligand to a density feature: activate a ligand for fitting with (i) a single-click on any atom in the ligand (ii) a click on the **Fit/Fit Residue** menu item or by using the keyboard shortcut ‘f’. These actions will highlight the ligand molecule in green. The left mouse button now controls rotation of the entire canvas view and the right mouse button may be set to rotate or translate the ligand molecule. As already described for fitting amino acids, the two icons to the right of the ‘Cancel Fit’ icon are used to toggle between translational (+) and rotational (+) fitting modes.

Clicking on a bond in an active residue (ligand) establishes an arbitrary torsion angle within the ligand. This torsion angle is highlighted as a grey arrow along the rotatable bond. The atomic group closest to the point that was clicked will be the part of the ligand that will move. If the icon to the right of the icons that control the translation and rotation operations is selected (), a portion of the ligand may be rotated about this bond using the right mouse button. It will usually be fairly obvious to a human with minimal chemistry training which torsion angles within a ligand are allowed relatively free rotation.

Taken together, these options allow the crystallographer to not only quickly orient a ligand molecule in density but also to fit groups of atoms within the ligand by twisting them about rotatable bonds.

## 5.9 Structure refinement with CCP4/REFMAC5

MIFit contains two interfaces for running refinement jobs with the CCP4/REFMAC5 software. The **Job/Refinement** command (available in the MIFit automation system and described in Chapter 10) is intended for relatively long refinements. This command contains built-in options

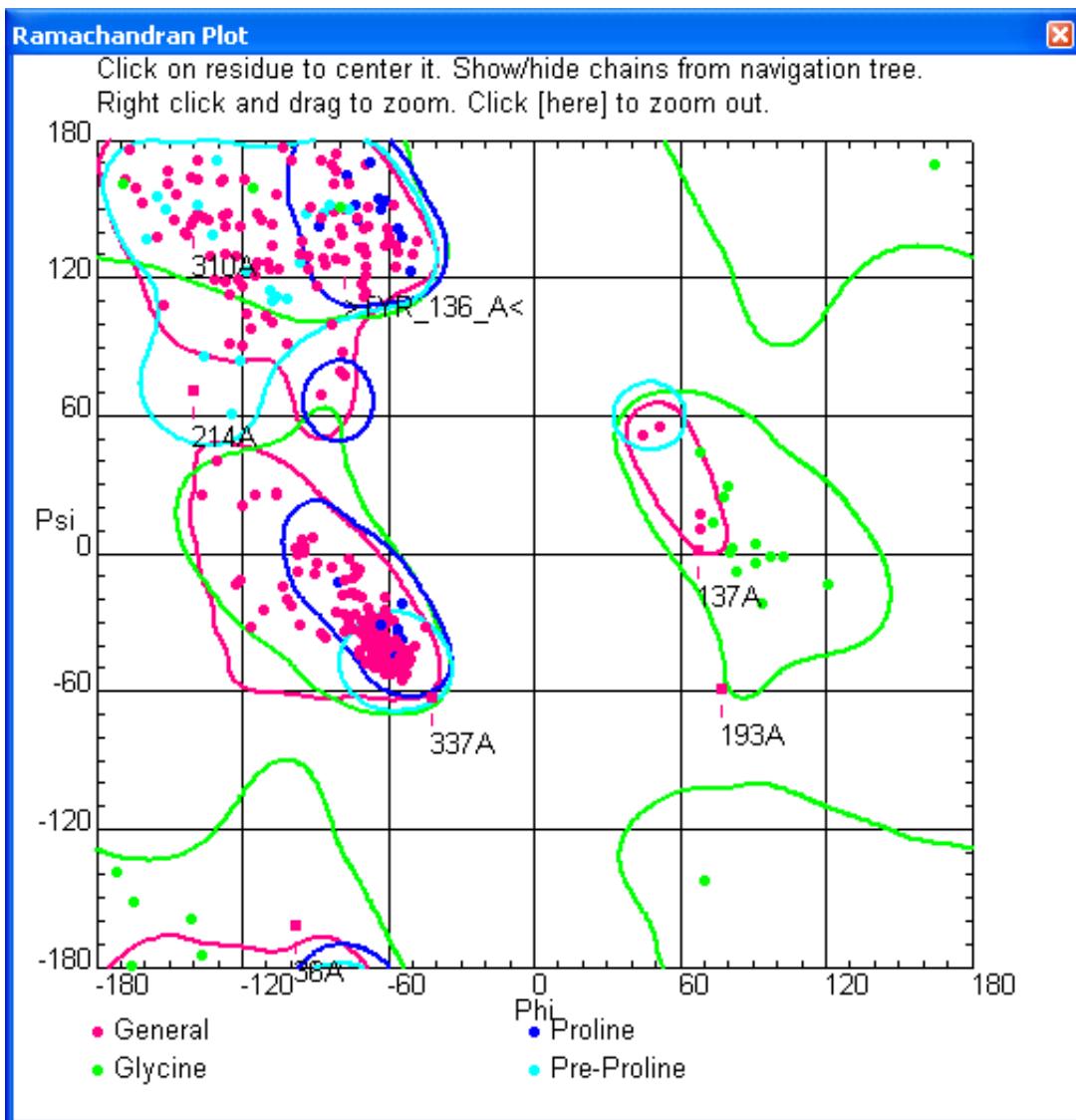
for water-picking, generating arp/wARP maps and provides a system for logging refinement progress and error detection.

## 5.10 Using the Ramachandran plot

The **Analyze/Ramachandran Plot** command displays the phi-psi positions of the visible amino acids in the current model (Figure 5.21). Shown on the plot are allowed regions based on the Richardson phi-psi data. Note that the Richardson's analysis created different allowed regions for 'general', 'glycine', 'proline' and 'pre-proline' residues and these contours are displayed with different colors. Text on the plot identifies outlying residues. By right clicking and dragging on the plot you can zoom in on a crowded region; a click in the specified zone 'Click [here]' zooms back out. By clicking on a residue in the plot the user can center the model on that residue in the MIFit canvas.

The display of phi-psi points is connected to the show/hide status in the navigation tree. i.e. it is, for example, possible to display just the amino acids for a single chain by hiding other chains in the tree control.

The plot is interactive in that a change in the phi-psi angles for an amino acid in the model (for example during model-fitting) is reflected by the associated changes in the phi-psi marker position in the plot. This feature is useful for model-building since the user can adjust the model into allowed regions on phi-psi space, which will usually be maintained through refinement.



**Figure 5.21** Example of Ramachandran Plot

# 6 The Ligand Dictionary Editor

The MIFit software contains specialized tools to simplify the task of generating refinement restraints for novel ligands. Ligand restraint files may be exported for use with the CCP4/REFMAC5 refinement software or converted for use with SHELX or CNS/CNX.

## 6.1 Entering ligand data

The **Dictionary/Import Ligand** command provides several paths for entering small molecule information into the MIFit dictionary. This command branches into **Cif**, **Mol**, **Pdb** and **Smiles** options, which may be used to enter small molecule information in any of these formats.

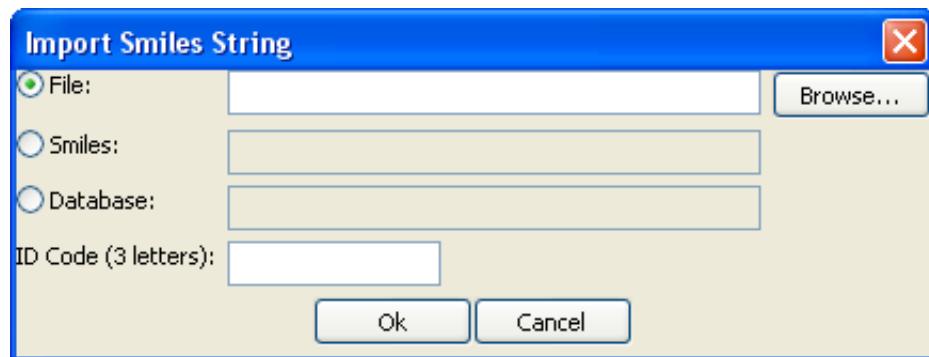
In order to develop information for a novel ligand an effective approach will often be to draw the ligand with a molecular sketcher (which will efficiently manage valence and hydrogen location information) and then load the resulting model into the dictionary editor as a MOL or SMILES file. Alternatively, a set of template coordinates may already be available from the Protein Data Bank web sites. Useful resources for finding those small molecules that have already been found associated with protein structures are (<http://www.ebi.ac.uk/msd-srv/chempdb/cgi-bin/cgi.pl> and <http://ligand-depot.rcsb.org/>). Both the PDB and the mmCIF data formats are available from these sites. Computational chemistry software for working with macromolecules is usually capable of writing energy-minimized structures in PDB format and these provide good input parameters for the Ligand Dictionary Editor.

The **Dictionary/Import Ligand/Cif**, **Mol** and **Pdb** options provide browsers for loading ligand coordinate information in mmCIF (\*.cif), MOL (\*.mol, \*.sdf, \*.sd) or PDB (\*.pdb) formats.

From the **Mol** option, the user must also provide a 3 character entity code after entering the ligand data file since, unlike the PDB and mmCIF formats, this information is not a part of that format. To avoid confusion and operational conflicts, it is best not to use three character codes that have already been used by the Protein Data Bank for structures in the public domain. These entry codes may be searched from the PDB web site; currently about 5% of the available name space appears to have been taken. In particular, the CCP4/REFMAC5 software includes mmCIF dictionary entries for the majority of these ligands and it is convenient to take advantage of these files.

Both 2D and 3D mol file types are supported, depending on the 2D/3D parameter embedded in the mol file format.

The **Smiles** option leads to an interface that provides three different methods for importing the SMILES string (Figure 6.22).



**Figure 6.22** SMILES string import options

The **File** option may be used with the associated **Browse...** button to enter a file (with extension, *\*.smi*) that contains a SMILES string. Alternatively, the **Smiles** option provides a text field in which to cut-and-paste a SMILES string.

The **Database** option may be used if it is possible to provide a script that is capable of returning a SMILES string when a ligand identification number ('regno') is provided through interface. To utilize this capability the script and the command to execute it should be provided as the **File/Preferences.../Environment/Smiles Database Command** parameter. This capability is mainly intended for users with access to a corporate small molecule data base since these databases typically store SMILES data.

The ID Code (3 letters) field is required to provide the Ligand Dictionary Editor with a 3 character entity code for the ligand since these codes are not part of the SMILES description.

## 6.2 Conformation generation

Although normally used to provide coordinate data for single conformers, the Ligand Dictionary Editor also contains the capability for generating multiple conformers for the ligand. Both the **Dictionary/Generate Conformers** command and the **Conformers/Generate Conformers** command within the Ligand Dictionary Editor may be used for this purpose. The ligand conformations are generated in torsion angle space and are subsequently stored in the MIFit dictionary file as explicit coordinates, in a way analogous to the storage of amino acid side chains conformers.

The automated ligand fitting option, **Refine/Find Ligand Fit and Conformer**, (available only in the enhanced versions of MIFit) will automatically generate conformers if only a single conformer is present in the dictionary. If the ligand contains hydrogen atoms then these should be stripped from the coordinates before generating conformations for ligand fitting. The **Atoms/Remove Hydrogens** option may be used to eliminate hydrogen atoms from the entity.

It should be noted that when ligand data is entered in mmCIF format the restraint information (i.e. the definitions of chiral centers, planar groups etc) are read from the mmCIF file rather than generated by the MIFit code. This behavior is mostly useful for checking the mmCIF dictionary files prepared for use with the CCP4/REFMAC5 program. However, the torsion angles are re-derived by MIFit for the purpose of generating the conformational ensemble.

## 6.3 The Ligand Dictionary Editor

The Dictionary Editor interface (Figure 6.23) appears after completing the entry of ligand data.

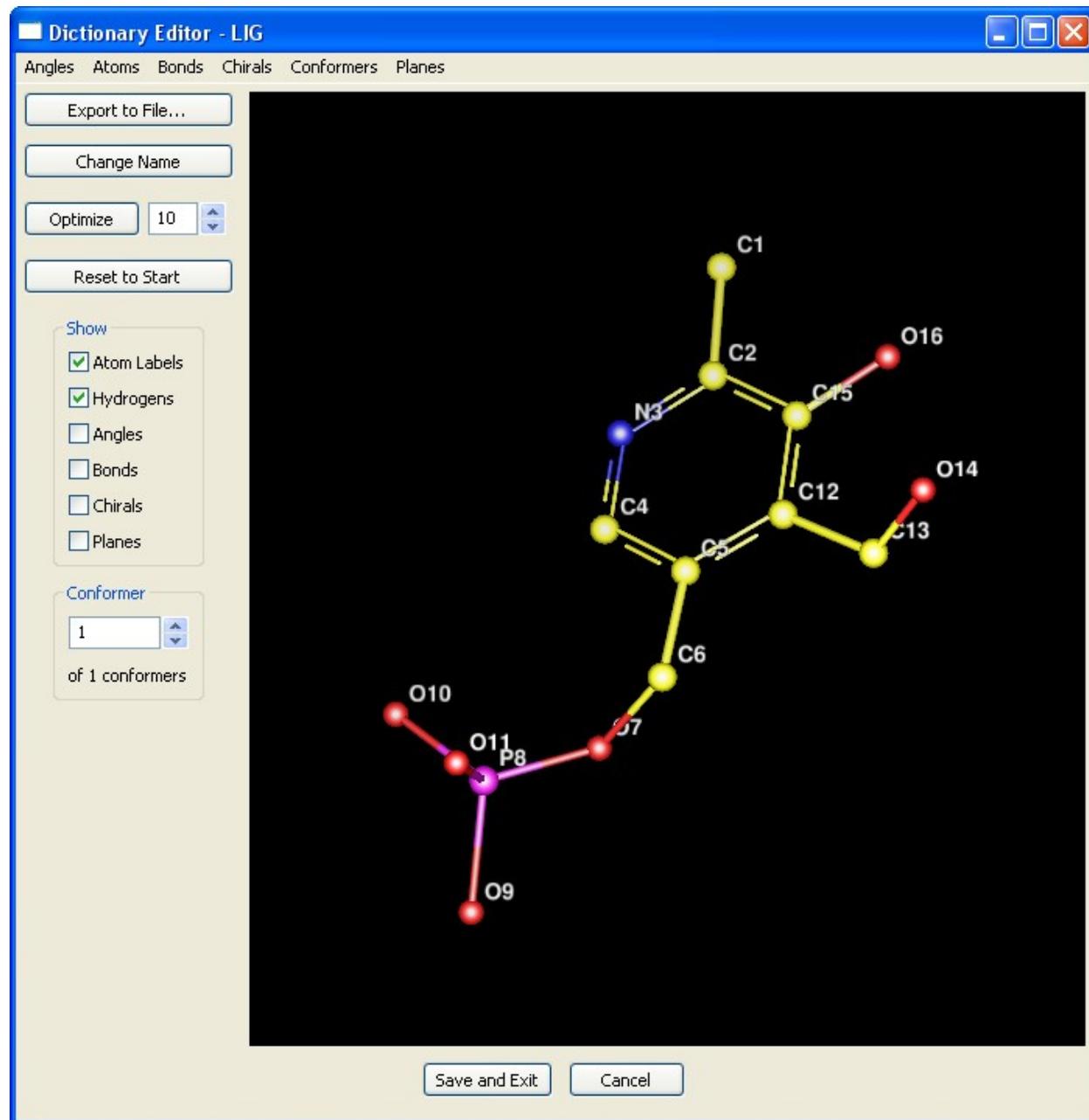


Figure 6.23 Dictionary Editor window

This application contains a window for viewing the molecule and displaying the current restraints. The left and right mouse buttons control the rotation and translation of the molecule; by holding Shift while pressing the left or right mouse button the size of the molecule may be adjusted. The keyboard short cut ‘|’ controls the display of split screen stereo, as it does for the main canvas.

A right-mouse click may be used to provide access to a menu that allows changing the drawing style. Currently supported styles are **Line**, **Ball and Line**, **Stick** and **Ball and Stick**.

The **Show** check boxes on the left hand side of the Ligand Dictionary Editor provide options for displaying the various refinement restraints established for this molecule and for voiding the display of hydrogens and atom labels. For example, checking the **Planes** box will display green meshes across those sets of atoms that a refinement program will restrain to lie in the same planes. The **Refine** button optimizes the ligand atom positions towards consistency with the current set of restraints. This option is most often used when it is necessary to change a plane definition.

The **Angles**, **Atoms**, **Bonds**, **Chirals** and **Planes** pull down menus at the top of the Ligand Dictionary Editor window provide the tools for changing the restraint definitions. These tools work in conjunction with interactive atom picking from the molecular image. The **Atoms** pull down menu also contains options for removing all hydrogen atoms or individual selected atoms from the ligand.

The most common problem with automatically generated restraints is that the plane group definitions are only partially correct. The Ligand Dictionary Editor software attempts to deduce the correct subsets of atoms to combine into planes from the input coordinate data but this is a relatively difficult problem. In order to add or remove atoms from a plane first click on the plane identifier (for example, a '1' marking plane-1). The plane mesh will then change color to red, indicating that the selected plane is active. The relevant options under the **Planes** pull down menu are now active. Initially, only the **Remove Plane** option is available. After clicking on an atom in the model, options to add or remove atoms also become available and may be used to change the atomic composition of the plane.

It is important to note that the reliability of the automatically derived restraints depends on the quality of the input coordinates. Energy minimized ligand models will usually give correct restraints whereas inaccurate coordinate sets are likely to miss some restraints. For mmCIF files, which encode restraint information, the embedded restraint information is used.

## 6.4 Exporting ligand information

Once the molecule is parameterized by a correct set of refinement restraints the **Export to File...** button may be used to write ligand dictionary information in the mmCIF format for refinement with CCP4/REFMAC5 or as a coordinate file in PDB or MOL format. The **Save as type** selection in the file dialog is used to specify the format.

After selecting the **Save and Exit** button in the main Dictionary Editor window the ligand coordinates and associated restraints are added to the MIFit dictionary.

## 6.5 Restraint dictionary files for SHELX and CNS/CNX

The mmCIF (REFMAC) dictionary files output by the Ligand Dictionary Editor may also be converted for refinement with either SHELX or the CNS/CNX family of programs. Although MIFit maintains its own dictionary format, the mmCIF format is used as the most standardized vehicle for information exchange. In addition, the mmCIF dictionary files available within the

CCP4 installation provide a valuable resource and the conversion tools allow the utilization of these data with the other refinement programs. The twin commands **Job/Convert Cif to SHELX** and **Job/Convert Cif to CNS/CNX** may be used perform this format conversion.

Once the application for SHELX dictionary conversion has executed a dictionary file for the SHELX program will appear with default name *mi\_shelx.lib* and in the same directory as the input mmCIF data file. If the command is subsequently repeated with a different mmCIF dictionary file then the additional restraints will be appended to the file *mi\_shelx.lib*.

The output files from this application that are intended for use with the CNS/CNX program have names based on the input ligand entity name. For example, if a ligand is described with entity code ‘MMM’ within the input mmCIF file then output is a parameter file, *MMM.par*, and a topology file, *MMM.top*. As with the files produced for SHELX, these files will appear in the same directory as the input mmCIF file.

# 7 Displaying structure data

## 7.1 Setting model display modes

MIFit provides many options for displaying protein structures. These options may be selected from the **Render** menu. The basic choices for displaying the atomic model in the main canvas are listed at the top of the menu as:

- Sticks
- Knob and Stick
- Ball and Cylinder
- CPK

The **Render/Sticks** option represents the atomic model by drawing lines between the atomic centers. The **Render/Knob and Stick** option adds to the stick representation by marking the atomic positions with small spheres. The **Render/Ball and Cylinder** option is a version of the **Render/Knob and Stick** option that presents a more three-dimensional appearance. The **Render/CPK** option provides a fully space-filling representation of the molecule.

For the **Render/Ball and Cylinder** display mode the user is able to specify the diameter of the balls and cylinders representing the atoms in the atomic model. The **Render/Set Ball/Cylinder Size...** command provides access to a dialog window (Figure 7.24) in which the **Ball Diameter** field allow specification of the ball size as a percentage of the CPK diameter and the **Cylinder Diameter** field allows control of the cylinder size as a percentage diameter of the ball diameter. Setting the **Cylinder Diameter** parameter to ~90% of the **Ball Diameter** parameter provides a way to make thick sticks with a full three-dimensional appearance.

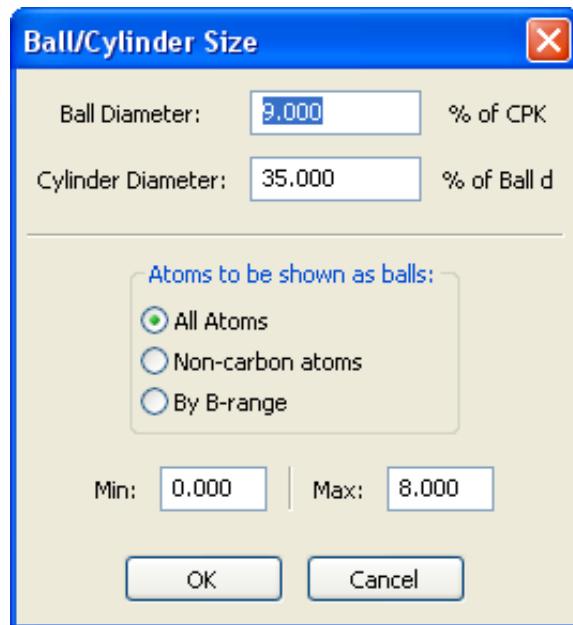


Figure 7.24 Render/Ball/Cylinder Size dialog

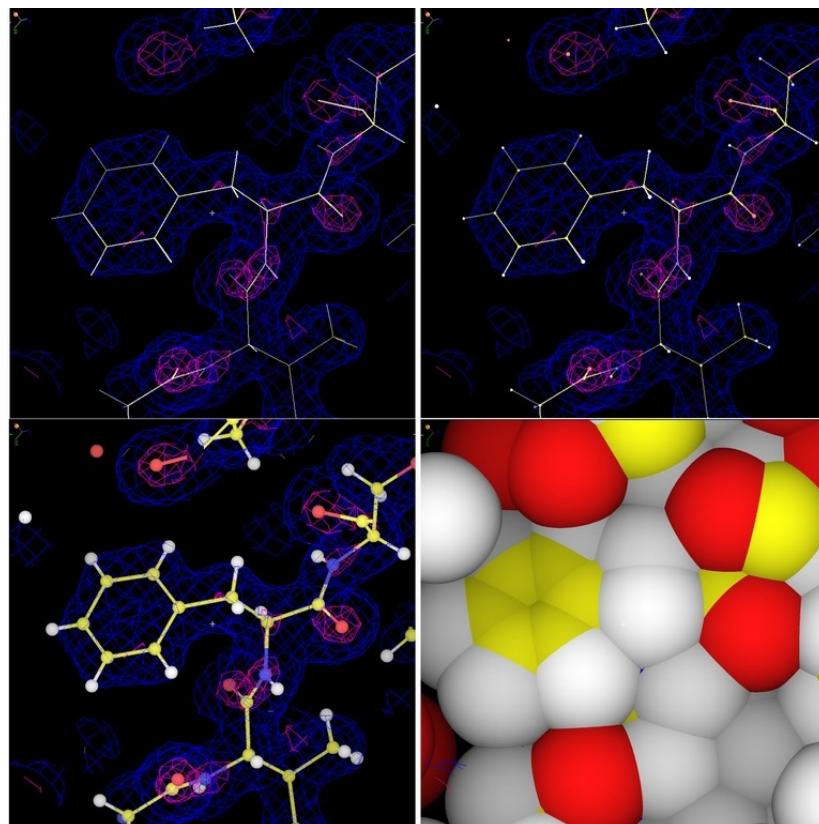
The **Render/Smooth Lines** option is usually selected as this improves the appearance of lines by anti-aliasing. However, this option also slows down the interactive aspects of the model display and, depending on the speed of the host computer, it may be necessary to set **Render/No Line Smoothing** for general model fitting work.

The **Render/Depthcue Lines** and **Render/Depthcue Colors** options both provide a sense of three-dimensionality in the z-direction in the absence of stereo display and these options are normally toggled on. Sometimes it may be more useful to maintain a brighter image by turning off the depth cueing of colors. Depth cueing of only lines still provides some sense of a three dimensional image.

The **Render/Line Thickness** option is used to change the line thickness (in pixels) for the model display. This option is usually set so that lines have a thickness of one pixel but it may occasionally be useful to brighten the image by making the lines slightly thicker.

All model display styles allow the concurrent display of electron density contours, although the electron density surfaces may not be very clearly visible when the **Render/CPK** option is applied. However the CPK representation combined with the display of difference density contours is a good way to visualize both the crystallographic aspects of ligand binding sites and surface cavities.

All of the model display options are available stereo displays as well as for mono viewing. If several model documents are open, each open model document may be set to different model display parameters and/or stereo mode.



**Figure 7.25** Collage of basic display methods

Shown from top-left in clockwise direction in Figure 7.25 are stick, knobs-and-stick, CPK and balls-and-stick displays.

When displaying multiple models during model-fitting operation it may be helpful for the non-active (reference) models to be less displayed with less intensity than the working model. Conversely, if displaying two models in order to create a figure to illustrate their similarities, the models would be displayed with equal visibility. For this reason a **Dim Non-active models** option is available and the extent of the dimming may be controlled with the **Set Amount to Dim Non-active Models** parameter.

## 7.2 Displaying model surfaces

The basic model display options available via the **Render** menu were discussed in the previous paragraphs. A different type of model display involves the representation of molecular surfaces. These representations are somewhat akin to the display of electron maps in that they may appear concurrently with a normal model display option. However, unlike electron density maps, which require structure factor data for their calculation, surface calculations do not require any additional information beyond the atomic model.

### Dot surfaces

Options related to dot surface displays are selected from either the **Show/Dot Surface** menu (Error: Reference source not found) or from inside the canvas window by right-clicking the mouse to pop up the QuickMenu. Selecting the **Show/Dot Surface/van der Waal Surface** option creates a dialog window (Figure 7.26) that contains options for controlling the calculation of the surface.

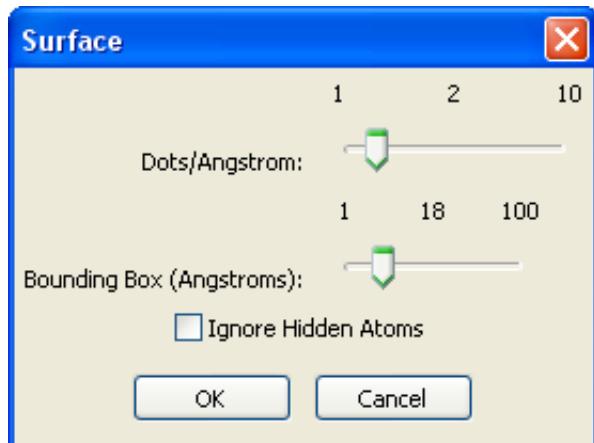


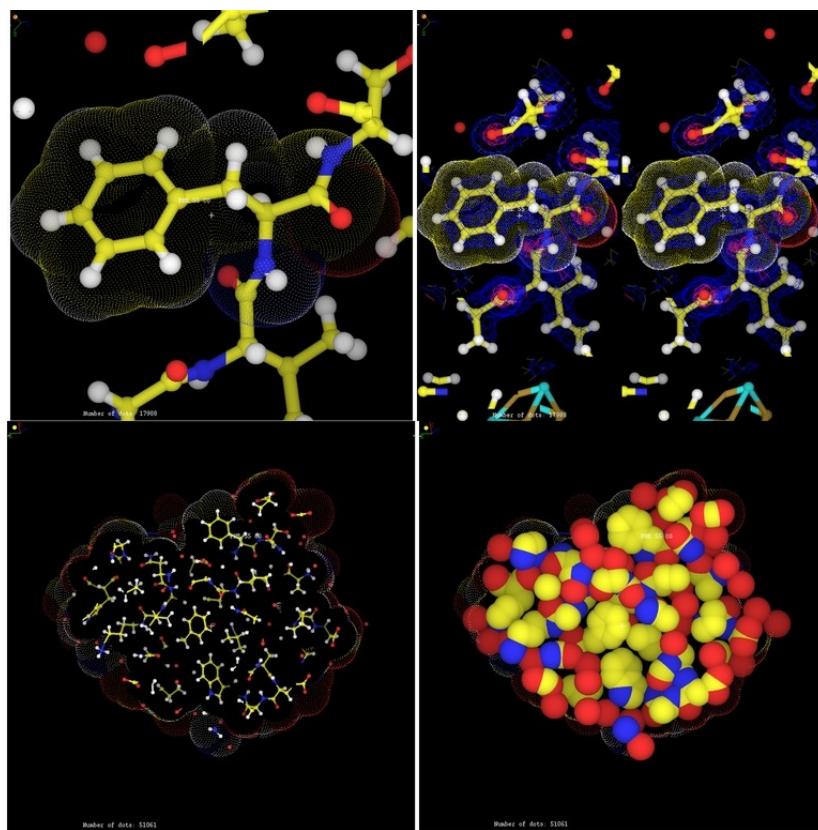
Figure 7.26 van der Waal surface selection dialog

Only one dot surface can exist in MIKit at any given time. The maximum number of dots/Angstrom is 10 and for a high dot density the display somewhat resembles a CPK display. When the surface extends over large parts of a protein structure the response of the interactive display may be slowed down.

The majority of the **Show/Dot Surface** options result in a dialog window similar to the one shown in Figure 7.26 except that in some cases the bounding box is an unnecessary parameter. A particularly useful surface is the solvent exposed surface that may be generated using the **Show/Dot Surface/Solvent Exposed Surface** command. This is a relatively smooth surface that provides a means of visualizing niches and cavities in the protein that are capable of binding small molecules.

The **Show/Dot Surface/Clear Surface** command may be used to eliminate an existing dot surface.

A collage of some representative surface displays is shown in Figure 7.27. As illustrated by the top right image in this collage, different types of model display may be combined. The illustration in the top right of the collage demonstrated that surface displays may also be viewed in stereo.



**Figure 7.27** Collage of surface displays

## Solid Surfaces

For some purposes it is an advantage to create opaque, solid surfaces rather than semi-transparent dot surfaces. The solid surfaces hide the ‘clutter’ of the atomic model representation.

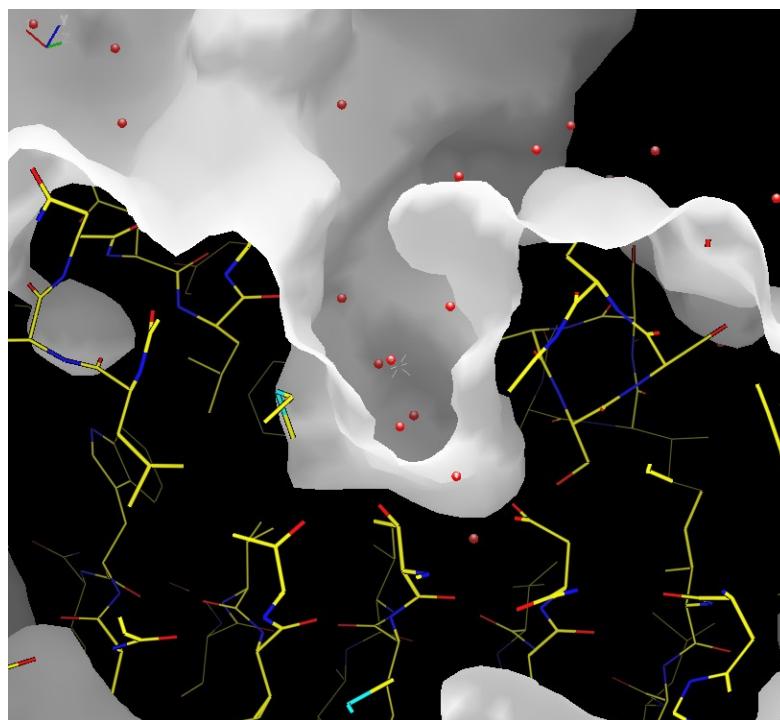
Commands related to displaying solid surfaces may be accessed from the **Show/Solid Surface** menu. The top set of commands **Build Surface**, **Color Surface**, **Color Surface by atom type** and **Clear Surface** are greyed out (inaccessible) until an atom is selected.

**Build Surface** renders the surface in either **Molecular Surface Mode** or **Accessible Surface Mode**. The surface produced by **Molecular Surface Mode** is similar in extent to a Van der Waals surface while the surface produced in the **Accessible Surface Mode** is expanded by the radius of a nominal ‘water’ probe molecule.

Surfaces can be built based on an **Atom Selection**, a **Single Residue Selection**, a **Multiple Residue Selection**, **Peptide Selection** or a **Molecule Selection**. The **Peptide Selection** option is the most commonly used command since this will create a surface based on all the protein (amino acid) component of a model, leaving cavities containing ligands unfilled. Note that it is possible to build multiple surfaces so, for example, it is possible to render surfaces for two different models showing component of a molecular assembly in the same document.

Solid surfaces are often used to provide a quick visual assessment of whether a designed ligand molecule is able to bind to a particular site without clashing with the protein. Surface depictions are also frequently seen in journal articles, providing an overview of ligand binding cavities and protein surface topography. Software requirements of fast interactive rendering versus creating an accurate and attractive appearance are somewhat contradictory so several commands have been provided to allow the user fine-tune the appearance of the surface display. The **Set Smooth Level** command allows the user to provide increasing numbers of surface smoothing iterations (on a scale of 0-5). Note the smoothing changes the exact dimensions of the surface. The **Standard Quality**, **High Quality** and **Ultra High Quality** settings change the size of the grid on which the surface is initially computed. These settings improve the accuracy of the surface at cost of some computer time in the initial surface computation and slowing down graphics speed for interactive displays.

Figure 7.28 shows a molecular surface with standard quality setting.



**Figure 7.28** Example of solid surface showing a fragment binding pocket in hsp90

## 7.3 Backbone and ribbon displays

Displaying all of the atoms in a model shows the greatest amount of detail but may also make it difficult to visualize the overall features of a structure. An understanding of the fold and major topological features of the entire structure may be facilitated by displaying just the Ca-carbon trace or the chain trace as a ‘ribbon diagram’, a display mode popularized by Jane Richardson in the 1980s.

The **Show/Backbone** command may be set to **Show backbone as atoms** (i.e. the default all atom representation), **Show backbone as CA trace** or **Hide Backbone** in order to control the representation of the backbone. Similarly, the **Show/Sidechains** command contains options **Show sidechain atoms** and **Hide sidechain atoms** to manage the display of all side chains.

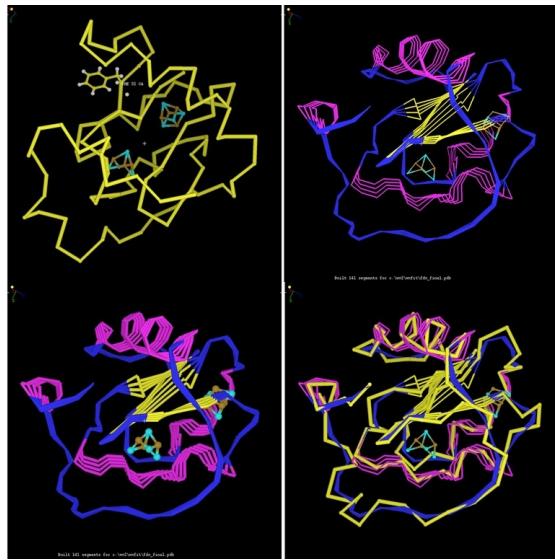
In order to create an alpha-carbon trace you can select **Show/Backbone/Show backbone as CA trace** and then select **Show/Backbone/Hide sidechain atoms** to hide the side chains.

Similarly, the **Show/Secondary Structure/Make Ribbon** command creates a ribbon image of the current model. The **Show/Secondary Structure/Clear Ribbon** command removes the ribbon image and the **Show/Secondary Structure/Ribbon Colors...** command spawns a dialog window (Figure 7.29) that allows the user to change the default colors for the helix, sheet and random coil secondary structure elements.



**Figure 7.29** **Show/Secondary Structure/Ribbon Colors...** dialog

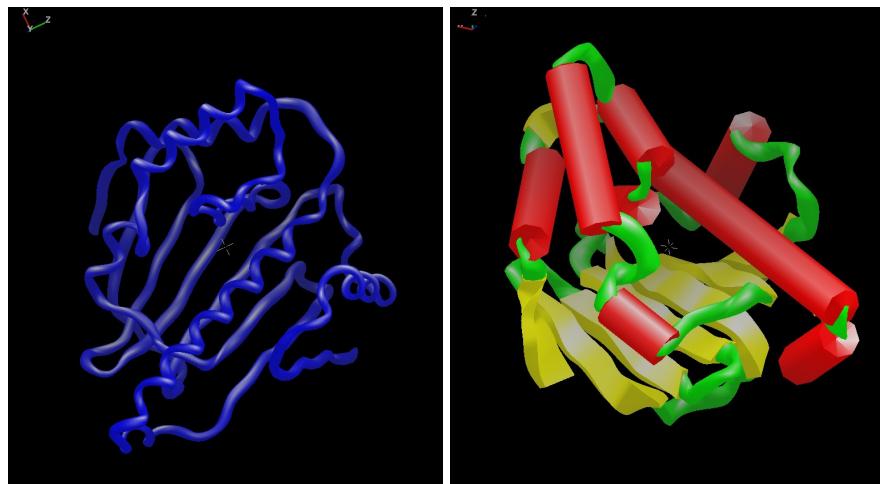
A collage of Ca-carbon and ribbon visualizations is shown in Figure 7.30.



**Figure 7.30** C $\alpha$  trace and ribbon diagrams

The top-left image in Figure 7.30 shows how an alpha-carbon trace may be combined with an all atom representation of a prosthetic group (here, a Fe-S cluster). The atomic model may be displayed as a stick, knob-and-stick or a ball-and-stick (shown here) representation. The bottom-right image shows that alpha-carbon traces may be combined with ribbon diagrams. All of these images could also be rendered in stereo and could also include electron density or dot surfaces.

MIFit also contain options for creating ‘worm’ and schematic displays of the protein chain trace with the **Show/Secondary Structure/Show Tube Secondary Structure** and the **Show/Secondary Structure/Show Schematic Secondary Structure** command (Figure 7.31). The **Show/Secondary Structure/Hide Secondary Structure** command may be used to remove these displays.



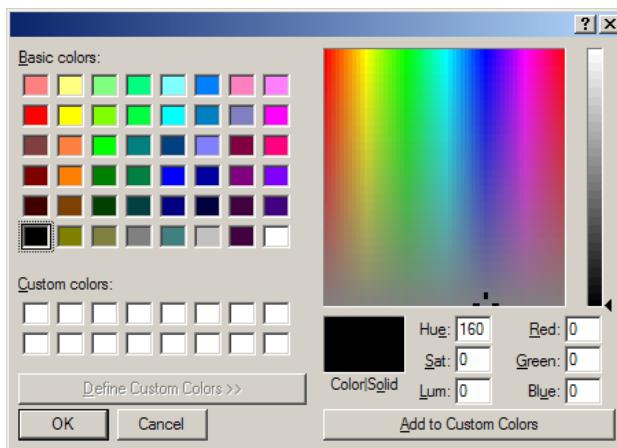
**Figure 7.31** Tube and Schematic representations of a protein

## 7.4 Choosing a background color

A black background is the standard (default) choice for interactive work with atomic models. However, if you wish to print the display canvas on overhead film or capture the image for a PowerPoint presentation, black is usually a poor choice. An image that appears attractive and displays with good contrast on a computer screen will often appear dark and hard to see when viewed as an overhead or on a PowerPoint slide. For example, Figure 7.28 and Figure 7.30 look perfectly reasonable when viewed on a computer screen contain too much black when printed. Although these issues are to some extent a matter of personal preference, light (often white) backgrounds often look best for these types of presentation.

Printing on paper those images that contain black backgrounds also has several negative consequences. One practical issue, especially when using plain paper in ink-jet printers, is that what emerges from the printer is a soggy welled piece of paper as this soaks up a lot of black ink. The paper takes quite a while to dry and become useable. Another side effect is on the lifetime of the ink cartridge. Manufacturer-specified ink cartridge life times are based on the assumption that roughly 5% of the paper is actually covered with ink, which is a reasonable estimate when printing text. Should you choose to print pictures with very dark or black backgrounds the use of ink will be much greater and you will need to replace the ink cartridge much more frequently than usual. Since paper is usually white this may also seem like the logical choice for the background color. This is not entirely true and you may have to experiment with various background color selections, depending on what images you print, the type of paper and the type of printer. Often a light color such as cyan, blue or brown will work well.

The background color in the main MIFit canvas may be changed using the **Render/Set Background Color** command (Figure 7.32).



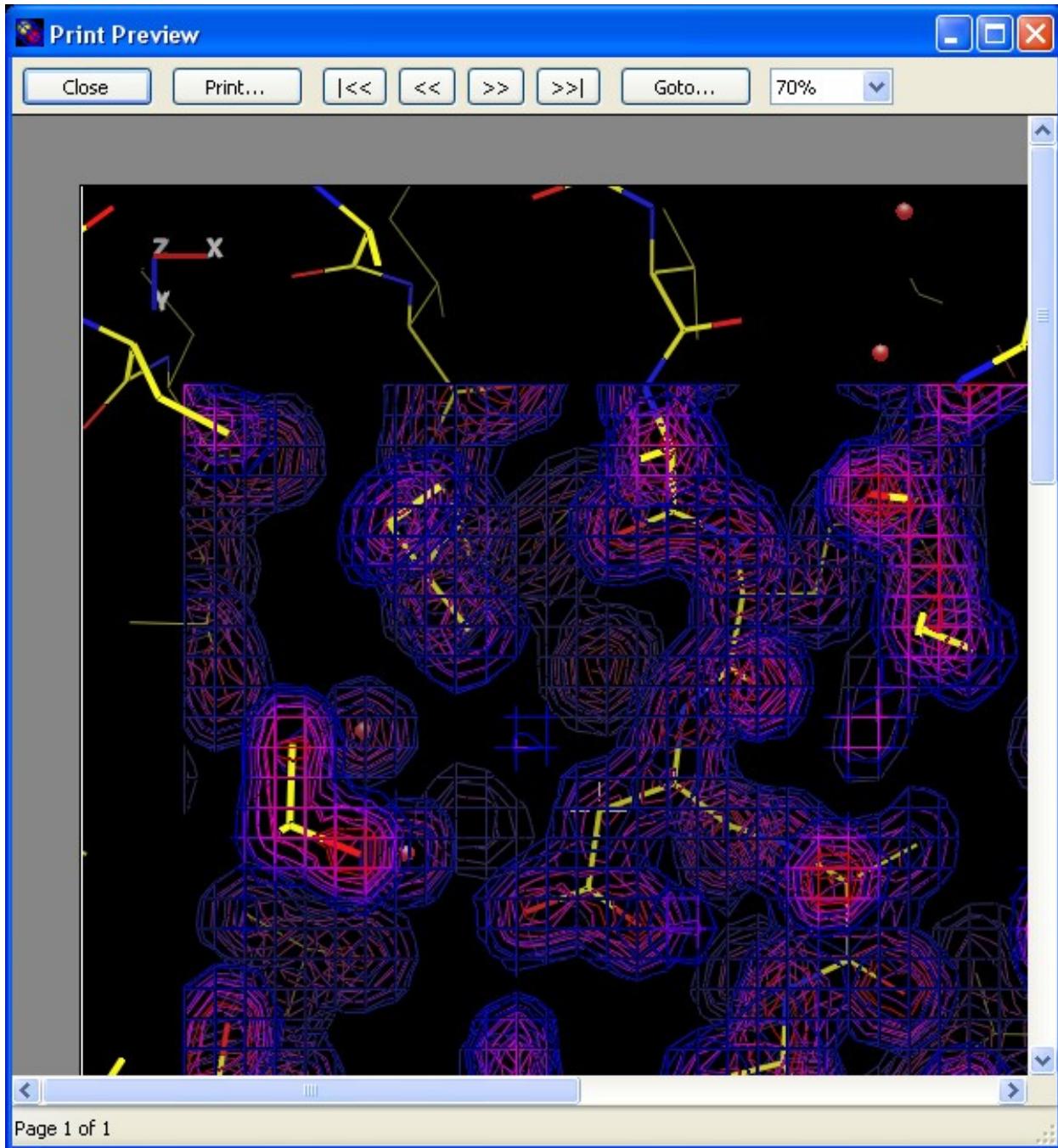
**Figure 7.32** Color palette for changing the MIFit canvas background color

This command provides a color palette from which a color may be selected by clicking on it. After selecting **OK** the background color for the main MIFit canvas will change to the selected color.

## 7.5 Printing directly from the MIFit main canvas

While displaying a structure with MIFit on the computer screen one might want to produce a hardcopy of the MIFit canvas contents. The available options for printing from the **File** menu are **File/Print...** and **File/Print Preview**.

The **File/Print Preview** command shows what will appear when the canvas is printed (Figure 7.33).



**Figure 7.33** File/Print Preview window

Buttons other than **Close**, **Print...** and the magnification parameter box (on the right) are defaults of the windowing system and have no effect in this particular application. The **Print...** button in the **File/Print Preview** window will bring up a print dialog window in the same way as the **File/Print...** command. Changing the value of the magnification parameter allows the user to alter the magnification displayed on the screen but this selection has no influence on the printed picture.

Note that it is also possible to capture the screen view directly for pasting into an open document (perhaps a Word or PowerPoint file) using the **File/Copy Canvas** command. As an alternative the **File/Export Image As...** option may be used to write the canvas image to one of a variety of standard image formats.

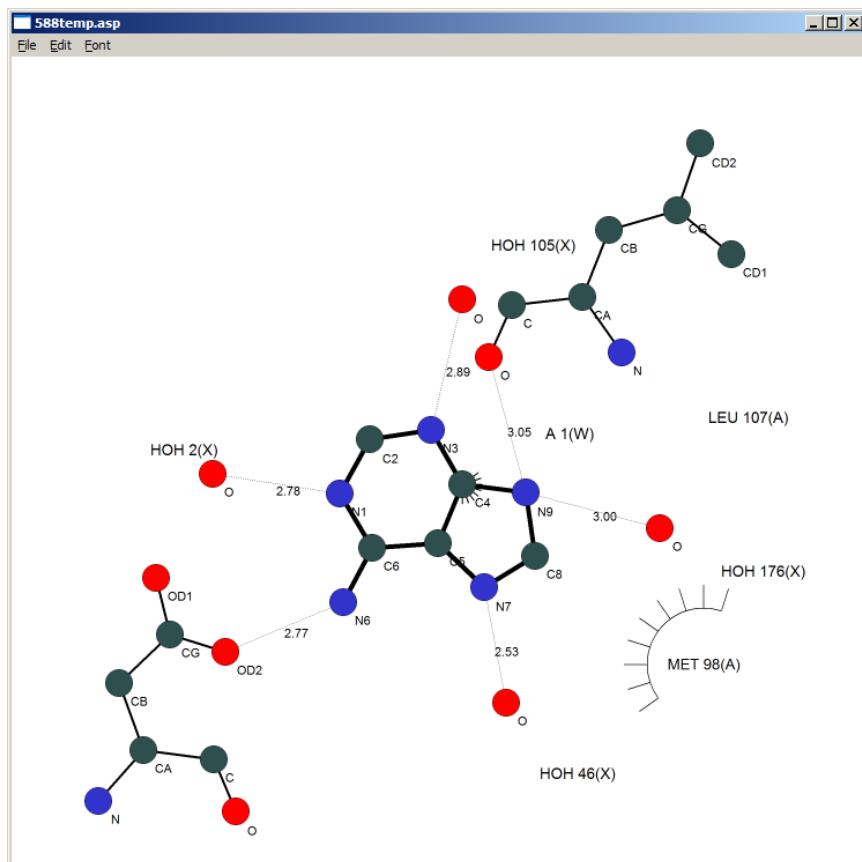
## 7.6 Schematic protein:ligand interaction plots

MIFit contains an application for generating schematic active site plots (ASP's) that represent protein:ligand interactions in a way that is similar to the images created by the 'ligplot' program (A.C. Wallace, R.A. Laskowski, J.M. Thornton, Protein Engineering, 8, 127-34, 1995).

To create an ASP: (i) open a session or PDB file of the protein:ligand complex, (ii) single click on any atom in the ligand, (iii) select the **File/Export Active Site Plot** option.

Following these actions a new window will appear that shows a schematic representation of the protein:ligand interactions (Figure 7.34). Only those amino acids and solvent entities in contact with the ligand are contained within this view. Hydrophobic and hydrogen bond contacts are defined by the atomic distances between protein and ligand atoms so this plot gives more 'correct' results with more accurately refined models. The orientation of the plot and selection of entities within it are carried out by the application i.e. it does not matter how the protein is oriented in the MIFit canvas or which ligand atom was selected to generate the plot.

The output image may be adjusted dragging the lower right hand corner with the mouse to remove white space from the bottom and left of the plot. The fonts used for text elements in the image may be changed using the **Edit** and **Font** pulldown menus. The **Edit/Add Text** option may be used to add title information to the figure.



**Figure 7.34** Example ASP (adenine binding to hsp90)

The **File** pulldown menu contains options of opening, closing, printing and saving the ASP images and data. In particular, the **File/Export Image** option provides a menu for saving the plot in png, tiff or bmp formats.

## 7.7 Data capture for structure reports and PDB deposition

MIFit must have access to a current version of the CCP4 software for this application to work (see Chapter 8).

MIFit contains an application for the generation of structure annotation data in mmCIF format for submission to the RCSB PDB and for automatically creating ‘boilerplate’ structure reports in HTML or text format. This application may be accessed through the **Job/Report** interface (Figure 7.35) and executes an automated structure reporting application within MIExpert.

The philosophy behind this application is that as much of the annotation and validation information as possible should be calculated directly from the input model and diffraction data. Data processing statistics may be parsed from data merging log files from SCALA, SCALEPACK or D\*TREK.

If necessary, more elaborate user annotation (typically relating to gene names and other ‘non-electronic’ information) may be parsed from a special template file. This approach to structure annotation is discussed in J.Badger et al, Acta Cryst F61:818-820, 2005.

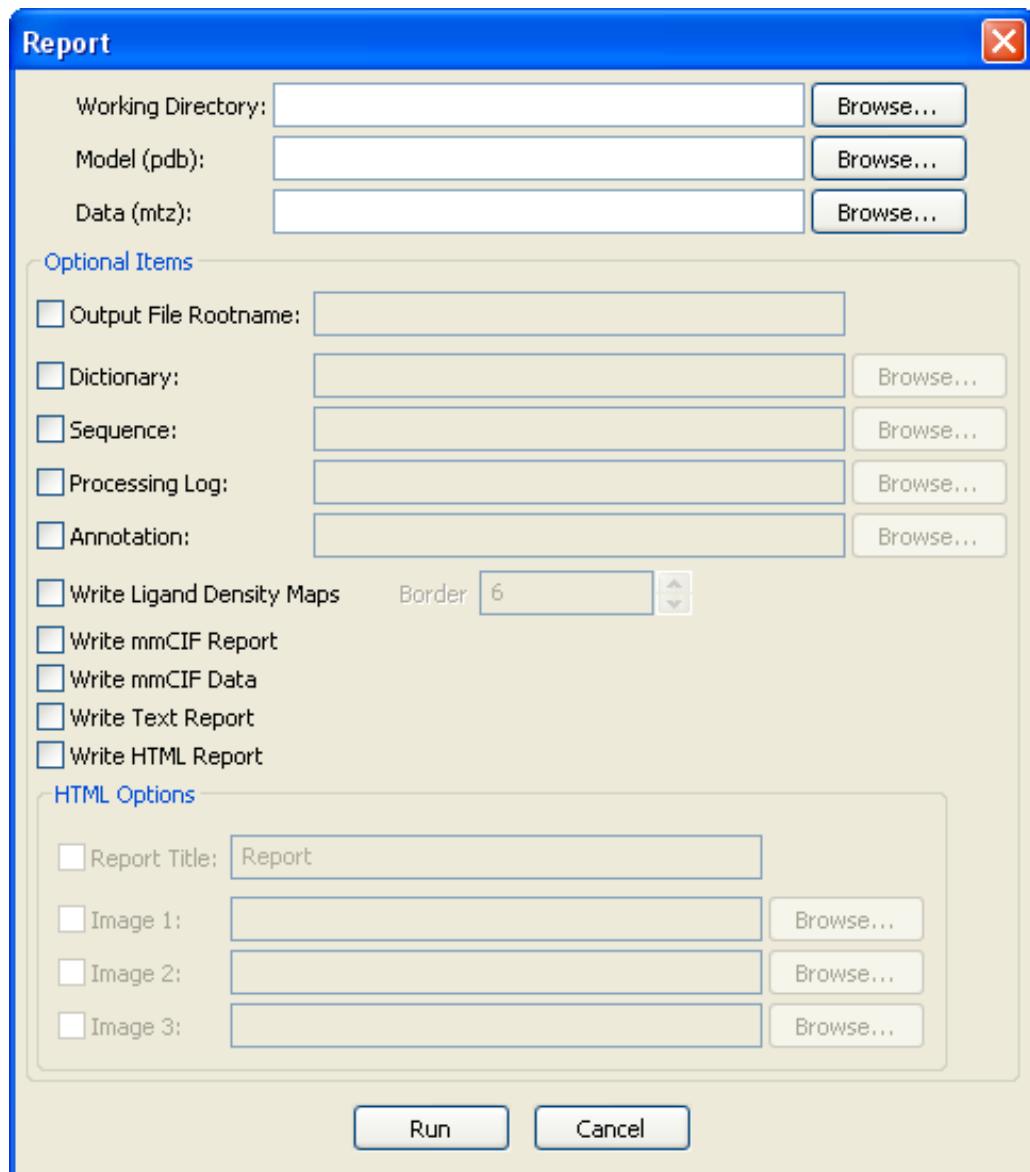


Figure 7.35 Job/Report interface

The **Working Directory** parameter specifies the directory where calculations will be performed and outputs will be written. The **Model (pdb)** parameter specifies the file name for the model that is to be reported and the **Data (mtz)** parameter specifies the file name for the associated diffraction data.

An optional input is the **Output File Rootname** parameter for specifying of root name of the output files. This parameter allows multiple reports to be written into the same directory since it avoids the problem of all report files having the same names. If the parameter is not set then the default file root name *pdbdeposit* is supplied.

The **Job/Report** application will automatically identify ligand molecules from the input coordinate file but if specific ligands need to be excluded from this identification they may be identified

by an independent chain-id in the input coordinate file and the **Ligand Chain ID** parameter may be used to identify them.

The **Dictionary** parameter is available to include any refinement dictionaries (in the mmCIF format, as used in refinement with CCP4/REFMAC5) needed for stereochemical assessment of the structure.

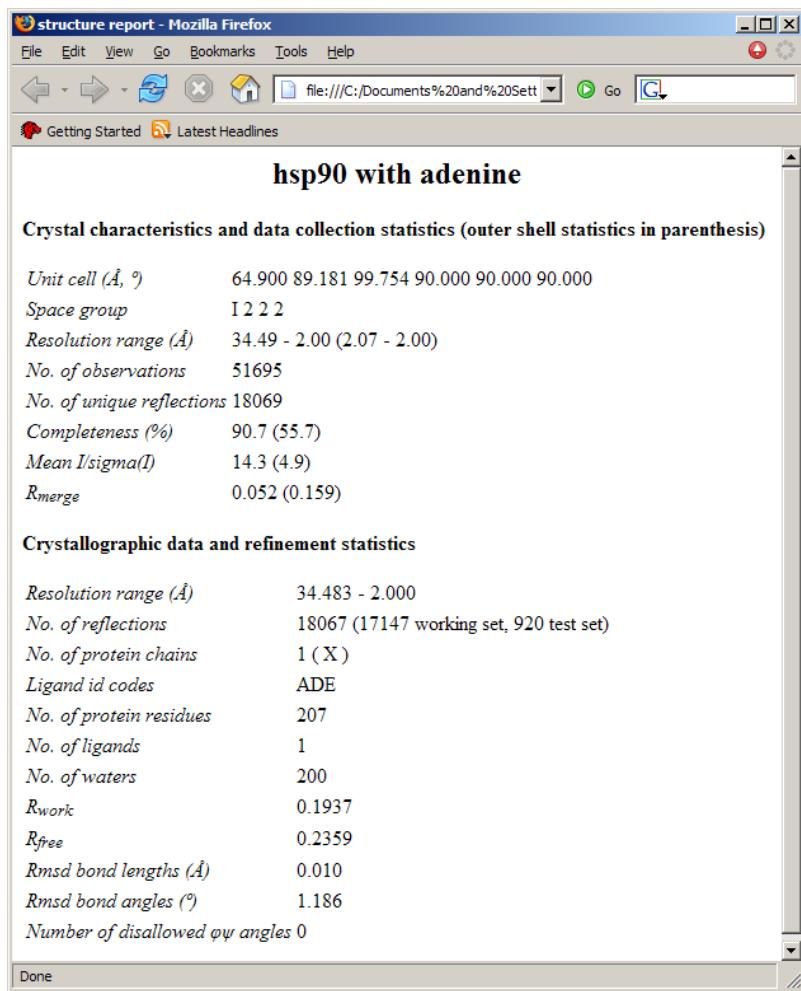
The **Sequence** parameter provides an entry point for passing the protein sequence in FASTA format into the output mmCIF annotation file. The required sequence is the protein that was in the crystal, regardless of whether any portions were disordered in the electron density map.

The **Processing Log** parameter may be used to enter the data merging log files from any of the SCALA, SCALEPACK or D\*TREK programs. These log files are parsed for statistical information (for example,  $R_{\text{merge}}$ ) relating to the data processing and this information is contained in the resulting mmCIF and HTML reports.

The **Annotation** parameter is an option for reading a file of user annotation (for example depositor contact information, literature and gene annotation). This option is mainly useful for avoiding repeated entry of the same information in a PDB deposition. An annotation file template is *deposit3d.template* and is located in the *../data/Scripts* directory of the MIFit installation.

The **Write mmCIF Report**, **Write mmCIF Data**, **Write Text Report** and **Write HTML Report** checkboxes may be used to specify which types of report to create. The mmCIF report is a formal and potentially very complete description of the structure and the structure determination process. This report also contains the structure coordinates. The mmCIF report may be uploaded through the RCSB PDB ADIT interface. If the required annotation items are filled out this file contains the complete information needed for a structure deposition.

The HTML report creates a ‘journal style’ list of structure data (Figure 7.36). The text report contains a small number of key items and might be more useful for incorporating into (for example) a PowerPoint slide.



**Figure 7.36** Example of an HTML structure report

If the **Write HTML Report** option is selected then further options are available for adding a title and molecular images to the report. At present the molecular images must be square in order to avoid distortion when these are resized to fit the report.

The **Write Ligand Density Maps** option is used to write out files corresponding to the portions of the likelihood-weighted electron density map that surround ligands in the structure. These files may be useful for import into other display programs that are enabled to read the CCP4 map format but may not be well suited to handling maps that cover the entire protein. The ligand identity is encoded in the map file names. For example, for a ligand molecule called ‘LIG’ with chain id ‘W’ and residue number ‘1’ the file *pdbdeposit\_W1\_LIG.map* would be created. The **Border** parameter may be used to specify the extent of the density map in angstroms around the each ligand. In addition to these partial maps, the command also writes a complete likelihood-weighted map of the unit cell as file *pdbdeposit\_fullcell.map*.

The **Run** command launches the process. The run time is typically similar to a single CCP4/REFMAC5 cycle since REFMAC5 is used to calculate structure factors and stereochemical agreement. If the **Write HTML Report** option was selected the resulting report will appear in a browser upon completion of the job. Note that by expanding the **Job List** menu at the top of the

document tree pane, and then right-clicking and selecting the option to view the log file, you can view the run diagnostics of the reporting run. Access to this log file is only enabled after the job has completed.

# 8 Running external crystallographic software

MIFit runs external crystallographic programs (CCP4 and SHELX) through a command-line library of crystallographic methods called MIExpert. Applications managed in this way include SAD phasing (Chapter 9) and the automated structure solution applications (Chapter 10), which include molecular replacement and refinement. All of these applications may be launched from the GUIs within the MIFit **Job** menu by using the MIExpert application as a standalone command-line ‘server’ for common refinement tasks. In command-line mode, executing the MIExpert.exe program from the MIFit installation provides help options for the various commands.

MIExpert protects the user from common problems with using the CCP4 suite (for example, handling paths containing spaces on Windows systems, some atomname justification issues in PDB files). MIExpert also manages various data management tasks by daisy-chaining different applications into tasks. For example, molecular replacement calculations may be initiated from merged intensity data since the conversion to amplitude data in mtz format is handled internally.

## 8.1 Accessing the CCP4 suite

MIExpert uses the CCP4 software suite as a basis for SAD phase determination, molecular replacement and structure refinement. To execute these scripts the user’s operating environment should contain the CCP4 environment variables and the CCP4 programs should be in the user’s path.

On the Windows operating system the relevant environment variables should always be automatically available once the CCP4 suite is installed. User environment variables may be checked by (i) right-clicking on the **My Computer** icon on the desktop, (ii) selecting **Properties**, (iii) selecting **Advanced**, (iv) selecting **Environment Variables**.

On Linux systems the window in which MIFit is launched will need to ‘know’ the CCP4 environment. This will not be the case unless execution of the CCP4 setup process (i.e. by sourcing the CCP4 setup script) is incorporated into the user’s login process. You can check to see which environment variables are established in a particular window by using the Linux/UNIX *printenv* command.

Before attempting to run any CCP4 applications (applications controlled by menus under **Jobs**) a check is performed to make sure that the CCP4 environment is established. The current set of applications were initially developed and tested using CCP4 6.0.1 and appear to be fully compatible with the current version of CCP4 (6.0.2).

## 8.2 Accessing SHELX and OpenEye software

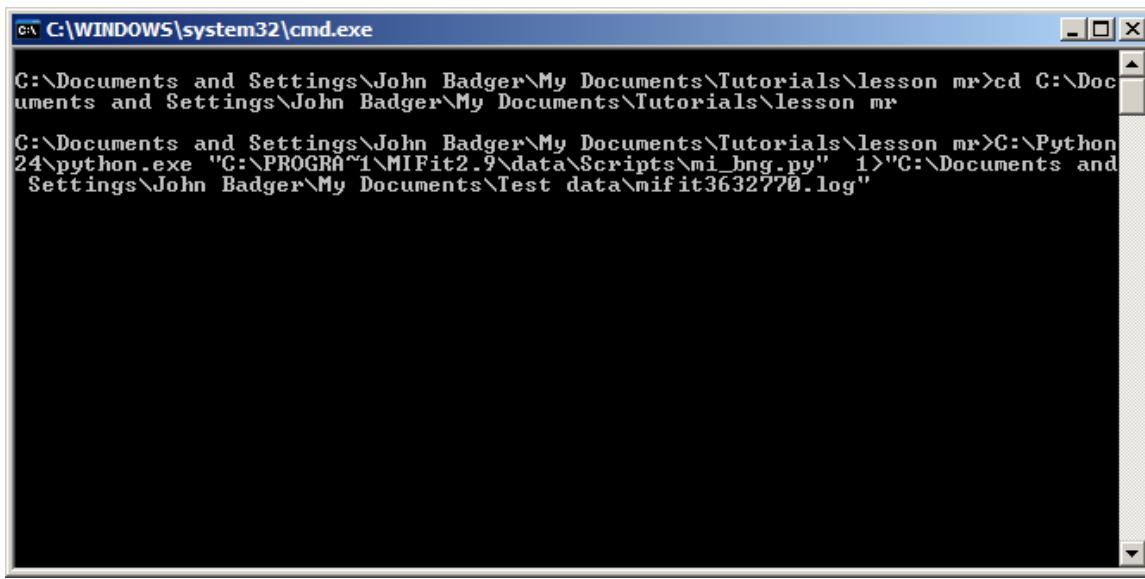
Applications within MIExpert use the SHELX package to find anomalous scatterer sites for SAD phasing and to provide an alternative refinement option to CCP4/REFMAC5.

SHELX does not employ any environment variables or an install process that would always place the installation at a standard installation on the host computer. To access the SHELX programs the user needs to set the path to the SHELX directory containing the executables to the

**Shelex Home** parameter in the **File/Preferences...** menu. This information is passed via a command-line argument to the MIExpert application for SAD phasing and refinement.

### 8.3 The Job List menu

When an external application is executed on a Windows computer you will see a small text port appear for the duration of the job (Figure 8.37).

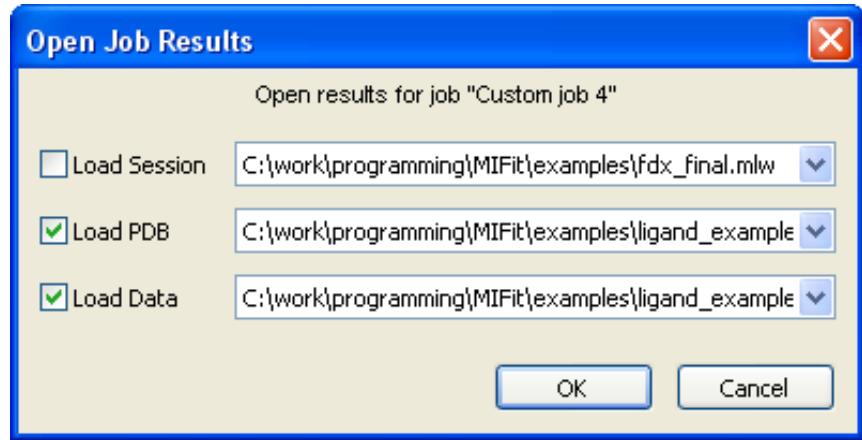


```
C:\WINDOWS\system32\cmd.exe
C:\Documents and Settings\John Badger\My Documents\Tutorials\lesson mr>cd C:\Documents and Settings\John Badger\My Documents\Tutorials\lesson mr
C:\Documents and Settings\John Badger\My Documents\Tutorials\lesson mr>C:\Python24\python.exe "C:\PROGRA~1\MIFit2.9\data\Scripts\mi_bng.py" 1>"C:\Documents and Settings\John Badger\My Documents\Test data\mifit3632770.log"
```

**Figure 8.37** Example of a running job under Windows

In order to manage external jobs the **Jobs** tab in the document tree pane on the left of the MIFit interface contains a **Job List**. This list contains all the jobs that have been run in the current MIFit session. Each job is identified by a serial number. A right mouse-click on a job provide a set of options, **Delete Job**, **Job Properties**, **Show Log File**, **Clean Successful**, **Clean All** and **Detach Job**. Jobs are color coded yellow (running), green (successfully completed) or red (failed).

The most useful command is **Open Results...** which provides an easy way to open results from a completed job. The Open Job Results dialog (Figure 8.38) lists session, PDB, and MTZ files in reverse chronological order, since in most cases the most recent files are the results of the job. Toggle on or off the types of files to be loaded with the checkboxes on the left.



**Figure 8.38** Open Job Results dialog

Another useful command is **Show Log File**, which will show all of the diagnostic information from the job that would normally be printed to a terminal. Note that this command is only accessible after the job is completed in order to avoid interfering with a running job. If a job should unexpectedly fail it is often possible to find the cause (or at least, the point of failure) by checking these logs.

Sometimes a job is killed or otherwise fails but MIFit continues to list the job as active. This is potentially a problem as the number of simultaneous jobs is limited through the **File/PREFERENCES.../General** menu. The **Detach Job** command may be used to eliminate this job.

## 8.4 Running customized Scripts

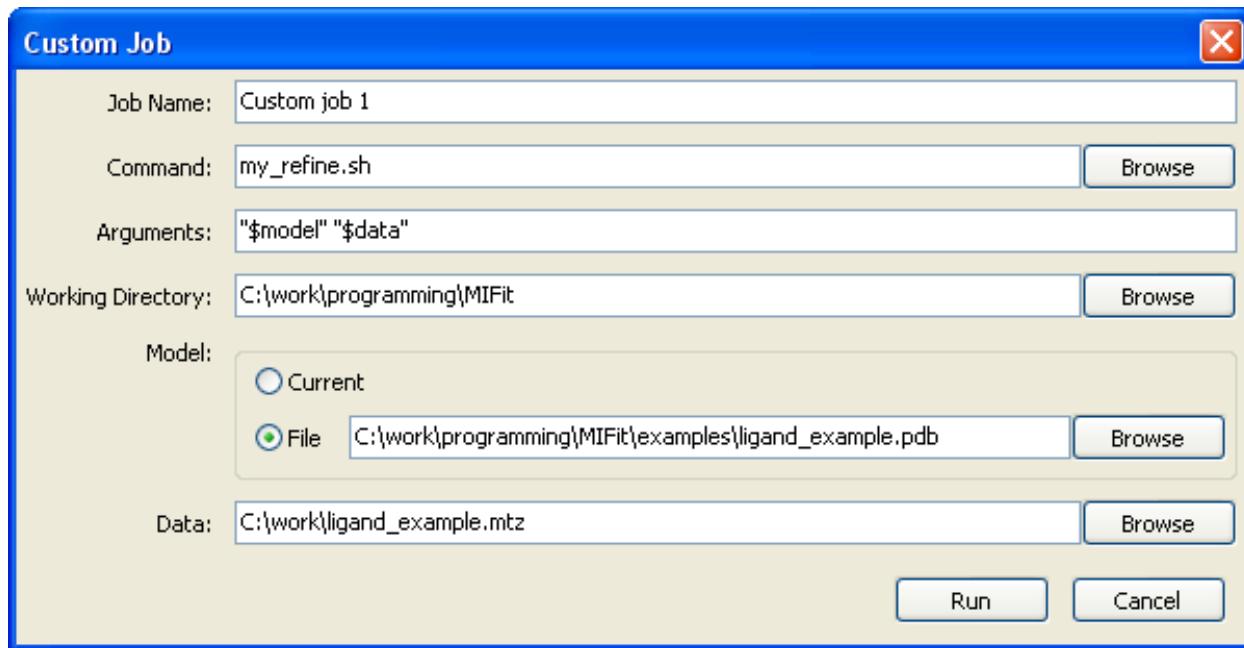
The MIFit **Job/Run Custom Job** interface (Figure 8.39) has been developed to provide a very simple but general mechanism for running external crystallographic applications and optionally reloading resulting coordinate (pdb) or diffraction data (mtz) in to a new MIFit session. This interface allows the user to apply a customized script-driven processing procedures to a model without leaving the MIFit interface.

The **Job Name** parameter is used to provide a temporary identification of the job.

The **Command** parameter specifies the command-line command that the user wishes to execute. The exact form of the command will depend on the operating system and type of script. For example, for a python script on Windows or a C-shell script on Linux the parameter would just be the path to the script. For a python script on Linux it might also be necessary to include the path to the python interpreter in the command.

The **Arguments** parameter provide the input to the script. By supplying the arguments “\$model” and “\$data” paths to the model (supplied by the **Model** parameter) and the diffraction data file (supplied by the **Data** parameter) will be passed as the first two command-line arguments to the script. (The quotes are included here to handle paths containing spaces and might be unnecessary on a Linux file system.). For the **Model** parameter either the current active model loaded into MIFit or a model from a file may be passed to the script. Additional arguments could also be included as part of the **Arguments** parameter, as needed by the script.

The **Working Directory** parameter specifies the directory in which calculations will be performed and the output files are expected to appear.



**Figure 8.39** Job/Run Custom Job interface

The parameters used by the **Job/Run Custom Job** interface are preserved across MIFit sessions. Jobs run through this interface may be managed by through MIFit **Jobs** tab at the top of the navigation tree control.

# 9 SAD phasing

## 9.1 Prerequisites

The SAD phasing application within MIFit uses SHELXD as the anomalous scattering site finder and the CCP4 software to prepare data from intensities (TRUNCATE), for site refinement (MLPHARE) and for phase refinement (DM). Chapter 8 describes the steps that need to be taken to access CCP4 and SHELX installations from MIFit.

## 9.2 SAD phasing interface

The SAD phasing interface (Figure 9.40) is available via the **Job/SAD Phasing** command.

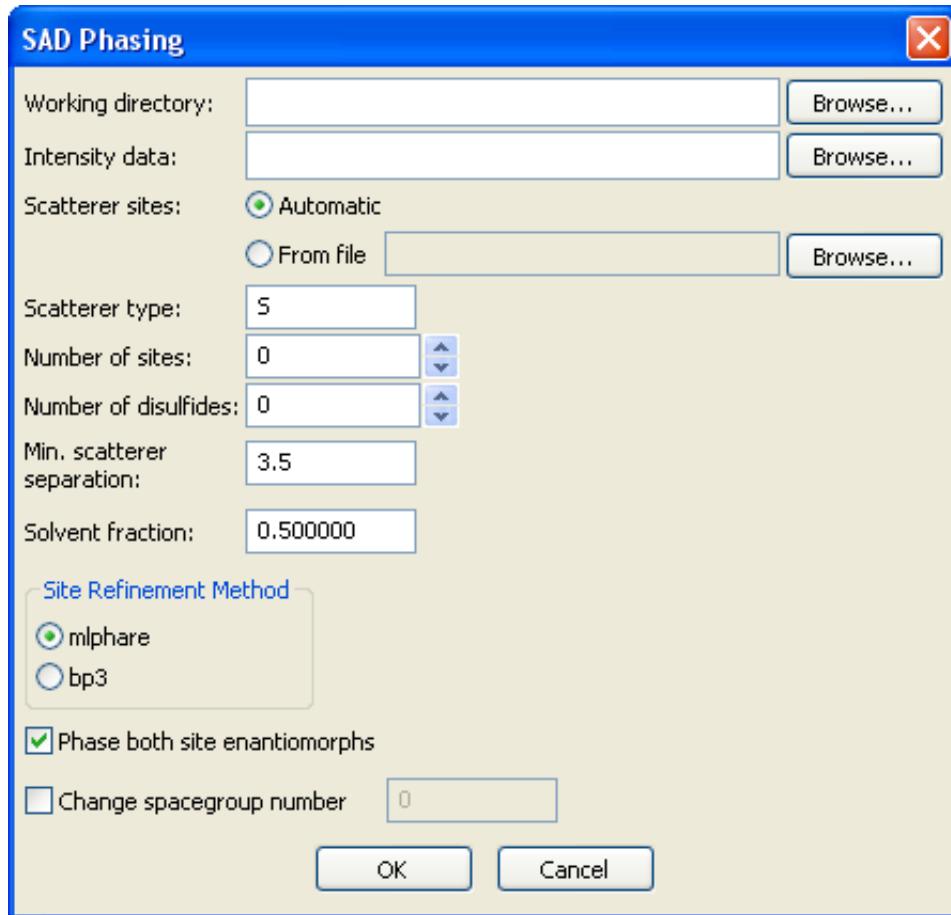


Figure 9.40 Job/SAD Phasing interface

The **Working directory** parameter is the directory in which you wish the SAD phasing outputs to appear. This directory may be located or created via the the associated **Browse...** button.

The **Intensity data** parameter should contain the path to a file containing merged intensity data, in which the Bijvoet reflection mates have been kept separate. Input data from CCP4/SCALA,

SCALEPACK or d\*TREK should be automatically recognized. After performing an internal conversion to the CCP4 MTZ file format, these data are processed through the CCP4/TRUNCATE program prior to carrying out phasing calculations.

The **Scatterer sites** parameter may be set to **Automatic** or **From file**, depending on whether you are starting a new structure determination and need to locate the sites ab initio or wish to enter sites from a previous calculation. The **Browse...** button associated with the **From file...** option provides a mechanism for loading files containing anomalous scattering sites. The **Files of type** filter in the browser contains a pulldown for PDB (.pdb) or RES (.res) input files, which correspond to PDB and SHELX formats.

The **Scatterer type** parameter may be used to set the element code for the anomalous scattering centers. The default value is ‘S’ (sulfur) since this phasing system has been successfully tested on examples of sulfur-SAD phasing from a chromium source. ‘SE’ (selenium) would also be a common setting.

The **Number of sites** parameter is the number of scattering sites that will be determined. This is the number of scattering centers to be found and in the case S-SAD phasing of a protein that contains disulfides it is smaller than the number of individual sulfur atoms. (In these cases site location proceeds more robust for detection of the disulfide ‘superatoms’ with a subsequent resolution of these atoms into individual sulfur atom pairs.) The **Number of disulfides** parameter is used to set the number of the located sites to split into sulfur pairs and is only applicable to S-SAD phasing. This step is important for achieving optimal phasing power in cases where the initial site identification used disulfide ‘superatoms’.

The **Min scatterer separation** is a parameter that determines the shortest allowed distance between two sites (in angstroms). The default value of 3.5 is satisfactory for most purposes since atoms will not normally approach more closely than that. Note that this setting will force the site finder to locate disulfides as superatoms rather than individual sulfur atoms.

The **Solvent fraction** parameter is the expected solvent content of the crystal and is used for the density modification (phase refinement) step.

The **Site Refinement Method** parameter may be set to **mlphare** or **bp3** to select for the CCP4/MLPHARE or the CCP4/BP3 program for site refinement and phasing.

The **Phase both site enantiomorphs** option is used to generate phases for both the initial anomalous scatter constellation ('hand 1') and the inverted constellation ('hand 2'). This option would be set on for an initial calculation but would be turned off once a set of sites is known to lead to a map with protein-like characteristics. This would be the case if a set of sites was being adjusted (perhaps removing erroneous sites and adding sites on unfilled features) in order to improve the phasing.

The **Change spacegroup number** parameter may be used if it is necessary to change the space group from the value given in the input file of intensity data. Other than mistakes in processing, it may be necessary to use this parameter to test phasing in enantiomorphic space groups.

The **OK** button launches the phasing procedure. It will typically take a few minutes to phase a protein, with the majority of the time taken in running multiple site finding trials. Upon completion of the phasing process, the text window will disappear and a browser will appear in which phasing diagnostics are reported. The **Job List** control may also be used to examine log data from the phasing run.

### 9.3 SAD phasing output files

A SAD phasing run will create a number of files that will appear in the directory defined by the **Working directory** parameter.

Results from automated site location are listed in a file called *mi\_phase\_summary.html* (Figure 9.41). This file lists the output files from searches at multiple resolution cutoffs (2.13Å to 3.33Å in 0.2Å increments in the example) together with correlation coefficients between these sites and the data. The solution that is automatically passed forward from refinement is the one with the highest correlations.

The screenshot shows a Mozilla Firefox browser window with the title "SHELXD site finder summary - Mozilla Firefox". The address bar shows "file:///C:/Documents%20and%20Set...". The main content area is titled "SHELXD site finder summary". It contains three table rows for parameters:

Job time	Thu May 25 15:37:42 2006
Working directory	C:\Documents and Settings\John Badger\My Documents\Tutorials\lesson 4
Data file	mi_sad_sort_f.mtz

Below these is a table of search results:

File	Res. (Å)	CC	CC(weak)
<a href="#">mi_sad_2.131.res</a>	2.131	46.06	27.03
<a href="#">mi_sad_2.331.res</a>	2.331	48.15	28.50
<a href="#">mi_sad_2.531.res</a>	2.531	49.74	28.60
<a href="#">mi_sad_2.731.res</a>	2.731	50.73	28.70
<a href="#">mi_sad_2.931.res</a>	2.931	50.93	28.74
<a href="#">mi_sad_3.131.res</a>	3.131	51.38	29.13
<a href="#">mi_sad_3.331.res</a>	3.331	50.40	27.09

At the bottom is a "Done" button.

**Figure 9.41** Summary of site searches

Overall results from the phasing process are reported in the HTML file *mi\_phase\_summary.html*. Figure 9.42 shows a portion of that file that reports the figures of merit resulting from the site refinement and the final occupancies and temperature factors for the ten sulfur atoms. In this case the occupancies are almost equal across these atoms and the B-factors are relatively low, indicating correct and well-behaved sites. Although relatively low (< 0.40) the figures of merit are greater than 0.3 across several resolution shells, indicating that there is relatively weak but still significant phasing power.

Site refinement and phasing summary - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Getting Started Latest Headlines

**FOM as a function of resolution after site refinement**

Res. (Å)	No. phased data	FOM
9.13	163	0.3140
6.22	457	0.3539
4.71	904	0.3521
3.79	1511	0.3092
3.17	2237	0.2601
2.73	3081	0.2397
2.39	4074	0.1634
2.13	2866	0.0492

**Occupancies and B-factors after site refinement**

Site	Atom	Occ	B
1	S	1.198	32.494
2	S	1.152	33.481
3	S	1.130	30.638
4	S	1.060	37.011
5	S	0.919	29.255
6	S	1.006	41.059
7	S	1.125	35.347
8	S	0.951	42.595
9	S	0.981	29.757
10	S	0.975	27.181

Done

Figure 9.42 Portion of phasing summary

Experimentally phased map data are most quickly accessed by loading the resulting session files *mi\_sad.mlw* and *mi\_sad\_i.mlw*, corresponding to phasing in the original and inverted hand of the anomalous scatterer constellation. The ‘protein residual’ values in the phasing summary give some indication as to which is more likely to be the correct solution (a lower value is better) but it is sensible to inspect the maps for both solutions. When judging these maps it is important to be able to identify distinctly protein-like features (for example, alpha helices and beta sheet) in order to be sure that a phasing solution has really been obtained.

Also present in the working directory are a crystal file for MIFit (*sad\_crystal*), MTZ files and logs from the phasing procedure and .phs files that contain pre-computed coefficients for anomalous difference maps. Details of these file names are provided in the phasing summary.

In some cases it will be possible and necessary to adjust the anomalous scatter sites (one of files *mi\_sad\_phased.pdb*/*mi\_sad\_phased\_i.pdb*) and repeat the phasing calculations to obtain a more optimal map. This could be the case if there are some incorrect sites in the initial solution or if the procedure to convert superatoms to individual sulfurs was only partially successful. In the latter case the anomalous difference Fourier map will show oblate features for the disulfides and atoms may be fitted to these features using MIKit.

# 10 Automated Structure Solution

## 10.1 Overview of automated co-crystal structure determination

An automated structure determination process has been developed within MIExpert to provide ‘pre-refined’ structures and maps for co-crystal structure analysis in as efficient and convenient a manner as possible. The MIFit GUI for running this process is described in section 10.2.

### Summary

This process prepares refinement data from a file of integrated intensities (D\*TREK, SCALEPACK and SCALA (MTZ) formats are supported) or images, runs molecular replacement with CCP4/MOLREP and performs preliminary refinement calculations (including water-picking) with CCP4/REFMAC5 to obtain a pre-refined model of the co-crystal structure. A MIFit session file is created at the completion of the refinement process with pre-computed map data and with a view centered on the expected ligand site. An option is available to provide an HTML report summarizing results from a set of cocrystal solution jobs. Summary information on the structure solution process is provided in the file project\_history.txt in the working directory. A convenient list of probable structure errors is also provided for each structure. The entire automated process will normally complete within a few minutes. For example, on a 2GHz laptop, it takes 3.5 minutes to complete this process on a 2Å resolution data for hsp90.

If the host computer contains a D\*TREK installation this application may start with the unprocessed image data and will perform the preliminary image data processing and merging. Obviously, image data processing is a relatively long operation and will typically add ~15 minutes to the run time.

The interface and associated script allow for the entry of multiple data sets.

### Reference data and structure transformations

If a reference data set is provided the process will re-index the new data (for applicable space groups) to find the indexing that is most consistent with the reference. Following molecular replacement, the script will also apply symmetry operations to place the resulting model as close as possible to the input model. The aim of these operations is to ensure that all structures within a series of co-crystals are conveniently placed at the same position in the crystal cell. The cross-validation flags in the reference data set are preserved in the setup of refinement data for the new structure.

### Molecular Replacement and Refinement

If the structure is known to exist in multiple conformational states, a set of possible starting models may be automatically input to the molecular replacement process. In the case of searches involving multiple trial models, the model that gives the lowest R-factor following molecular replacement is carried forward for further refinement. The refinement is run in two stages, initially with the model directly from molecular replacement and then by four short runs that are interspersed with water-picking calculations using CCP4/ARP-WATERS. All data are used for re-

inement calculations together with the 'mask' bulk solvent correction and isotropic B-factor refinement. Should a special restraint library be needed for the refinement (i.e. if the MR model contains non-standard components) it may be supplied as part of the input.

For high volume protein:cocrystal structure determination projects the diffraction data will have usually been processed in the (known) space group for the crystal. If only the point group is determined the molecular replacement procedure may be setup to try all possible space groups within the point group.

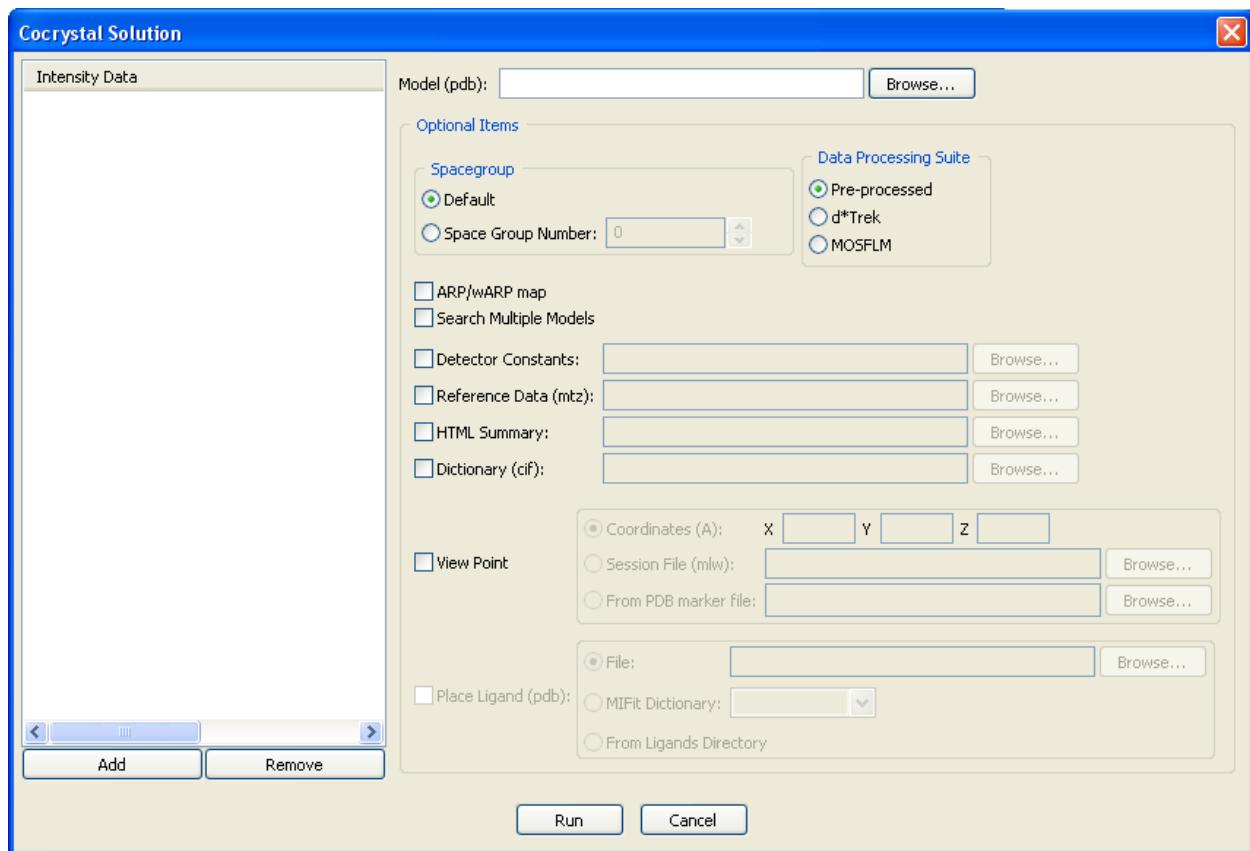
## Ligand Placement

A ligand may be specified from a PDB file or from the MIFit dictionary. If the OpenEye ligand-fitting software is available, this software is used to provide a flexible fit to the difference density at the user-defined target point. In the absence of OpenEye software a 6D search is performed in order to place an input ligand in the expected binding site. This capability may be useful for fitting small and relatively rigid ligands (for example, from fragment screening experiments) or in cases where the crystallographer has a good prior expectation for the conformation of the ligand (for example, when working on a series of related molecules).

An additional option is available to carry out accurate 6D searches on a set of ligands if the PDB files are placed in the data directory. The resulting file presents the best fits for all of the ligands. This option is relatively time-consuming (several minutes per ligand) and is mainly intended to fit the small sets of molecules used in mixture experiments with shape diverse fragments.

## 10.2 The automated co-crystal structure determination interface

The **Job/Cocrystal Solution** interface (Figure 10.43) automates co-crystal structure solution.



**Figure 10.43** Job/Cocrystal Solution dialog

The minimum requirements for running this application are a molecular replacement search model and a file of intensity data.

The molecular replacement search model may be entered as the **Model (pdb)** parameter.

The filenames for the data set(s) to be analyzed may be entered under the **Intensity Data** parameter using the **Add** button. An important and useful feature of this interface is that a series of related data sets may be added to the **Intensity Data** list and the structure solution operations will work on each of them in turn. If a data set is incorrectly included in the processing list it may be removed with the **Remove** button. For this application the working directory is assumed to correspond to the location of the data set. i.e. it is best to place each set of intensity data within a separate directory. If the extension for the data set is .img or .osc then it is assumed that this corresponds to an image data file and the application will attempt to process all image data in that directory. A BNG subdirectory is created as the working directory for all subsequent steps.

A set of optional items are available for these structure solution jobs.

The **Data Processing Suite** option may be set to **None** if the input file is merged integrated data or **d\*TREK / MOSFLM** if the user has one of these data processing programs installed and wishes to process the image data as part of this process. For the MOSFLM option the executable is assumed to be called *ipmosflm* in the *bin* directory of a CCP4 installation. MOSFLM is not currently generally available on Windows although it is anticipated that this will eventually

change. Automated image data processing is somewhat less robust than the rest of the structure solution process although the d\*TREK option has been found on good quality data.

When image data processing is performed, the results from the structure determination steps will appear in a *BNG* subdirectory of the directory containing the image data.

The **Detector Constants** option is only applicable if the structure solution job includes automated image data processing and information (usually the beam center or crystal to detector distance) is incorrect. The browse button associated with this option may be used to select a text file (*.txt*) that may contain values for the beam center (*beam\_center*) in mm and the crystal to detector distance (*xtal\_detector\_distance*) in pixels.

For example,

```
beam_center 1013.0000 1009.0000  
xtal_detector_distance 50.0000
```

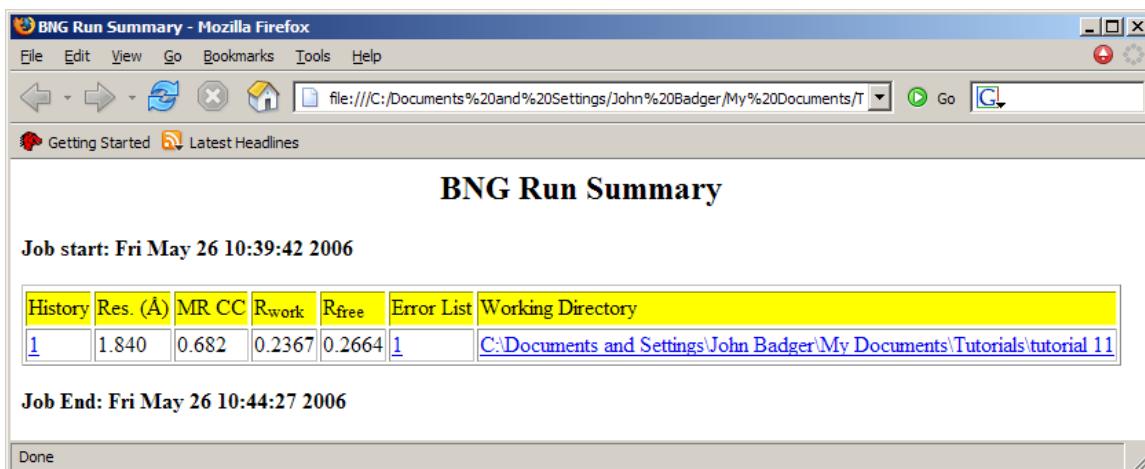
The **Spacegroup** option may be set at **Default** to use the space group encoded in the header of the diffraction data file. If this value is incorrect but the data was processed in the same point group the correct **Space Group Number** may be entered. The space group number must be provided if this application is to run image data processing with d\*TREK or MOSFLM.

The **ARP/wARP map** checkbox may be used to compute phases for an ‘arp-wARP map’ i.e. a map computed from a model that has been subject to atom deletion and pseudo-water insertion into unaccounted data. In cases where reasonably high resolution data is available this procedure sometimes leads to definite improvements in the interpretation of putative ligand density. Note that although the displayed map is phased on an atomic model adjusted by the ARP-wARP procedure the displayed model was not subject to the atom deletion/addition procedure. This is because most users prefer to work with a complete model rather than one ‘chewed up’ by ARP-wARP.

Although not strictly necessary, it will often be desirable to provide a **Reference Data (mtz)** parameter to apply the indexing conventions and R-free flags used in this reference data to the new data set. In some space groups the provision of this data set is necessary to ensure that the final model is placed in the crystal at approximately the same position as the input search model.

A potentially useful option is the provision of a **Search Multiple Models** checkbox which, when toggled on, will perform MR calculations using all PDB files in the same directory as the model specified by the **Model (pdb)** parameter. The model that gave the best MR solution (lowest R-factor) is carried forward for subsequent refinement. This capability is useful when a co-crystal is known to exist in multiple crystal forms (say, an ‘open’ and a ‘closed’ conformation) as the MR process will select the most appropriate model for subsequent refinement and model-fitting.

The **HTML Summary** option is a path to a directory in which a summary of the results from a set of structure solution jobs will appear. If this option is applied the summary will appear in browser when the job completes.



**Figure 10.44** Example of HTML summary containing just one job

The **Dictionary (cif)** parameter may be used to provide any CCP4/REFMAC5 dictionary that might be required to run refinement on the input model.

The **View Point** option with the **Points** parameter toggled on requires **X**, **Y**, **Z** as parameters corresponding to the view point in angstroms. This option provides an input for centering the view in the resulting MIFit session file. Alternatively, with the **Session** parameter toggled on, the required input is a previous session file. In that case both the view center and other display attributes (view direction, slab thickness, map colors etc) are applied to the new session file. Except for the case where the **ARP/wARP map** option is specified, these options will also eliminate water molecules from the vicinity of the target point in order to facilitate the generation of clearer difference density maps.

The **Place Ligand (pdb)** parameter may be used to provide ligand placement near the **View Point** target as a final step in the structure solution process. Unless the OpenEye ligand docking software (*afitt* or *flynn*) is installed this operation is purely a density docking process by 6D search i.e. it adds the molecule to the protein but does not attempt to refine it. For this reason, no dictionary is required for the ligand. The rigid-body docking is mainly useful if the ligand conformation can be anticipated in advance or the ligand has few degrees of freedom (i.e. for fragment screening).

The dictionary coordinates may be taken from the ligand specified by the **File** parameter or from the **MIFit Dictionary**. The **From Ligands Directory** provides a mechanism in which ligand pdb files in each subdirectory ‘ligands’ of the data directory will be used for ligand-fitting. i.e. this provides an implicit mechanism for associating a different ligand for each data set. If several ligand (up to 4) are found in the ligands subdirectory then all of them will be fit. This capability is intended for making trial fits of fragments from mixture experiments.

Once the **Run** button is clicked the automated structure solution begins.

If the procedure had never been run before, the model resulting from this process is *refine\_3.pdb* with associated data file *refine\_3.mtz*. If the 6D ligand placement option was applied a coordinate file *refine\_3\_ligfit.pdb*, which contains the fitted ligand, is also created. Crystal symmetry operations are applied to superimpose the final refined model at the same point in the crystal as the input model.

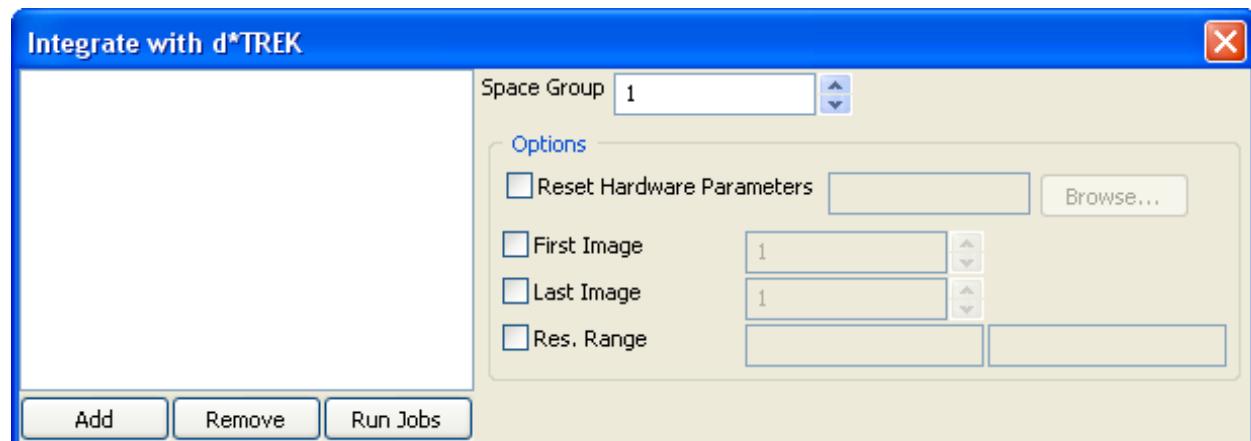
A MIFit session file (*bng\_milaunch.mlw*) that displays the final model in the context of the final electron density map and the final difference map is available for facile review of results. For technical reasons this file loads phased reflection files in .fcf format (*bng\_mlmap.fcf*) rather than the final mtz file. (Note that the difference map is not produced when the arp-wARP map option is selected.)

Local abnormalities in the final structure are listed in file *refine\_3\_errors.txt* and a MIFit crystal file (*bng\_crystal*) corresponding to the new data set is created in the working directory. Note that the list of local abnormalities may also be imported into a MIFit graphics session as an annotation.

As co-crystal solution jobs complete the emerging session files may be viewed using MIFit. At least on a relatively powerful laptop computer, session files may be reviewed concurrent with ongoing structure solution processes.

## 10.3 The Job/Integrate with d\*TREK interface

Automated processing of image data may be performed using the **Job/Integrate with d\*TREK interface** (Figure 10.45). This interface is not intended to replace traditional image data processing interfaces for analyzing data on new or difficult-to-process crystals; it is intended to automate processing of routine ‘reasonable’ data sets obtained in high throughput co-crystallography and fragment screening.



**Figure 10.45** Job/Integrate with d\*TREK dialog

The **Add** and **Remove** buttons on the left of the GUI are used to create a list of folders containing image data to process.

Since this GUI is intended to relieve the crystallographer from the task of processing data from previously characterized crystals an expected space group is available and should be entered as the **Space Group** parameter. No other information is needed to auto-process image data.

In some cases the hardware parameters encoded in the headers of the image files (including the beam center position and crystal to detector distance) may not be correct. These values may be rectified by providing a parameter file using the **Reset Hardware Parameters** option. The format of this file is described in section 10.2.

The d\*TREK software contains the concept of a **Beam Mask** which may be used to eliminate portions of the detector surface from use in data integration. If a beam mask is available it may be entered as the **File** parameter.

In some cases it may be necessary to process only a subset of the available images (for example, if the crystal became severely radiation damaged during data collection) or to a predetermined resolution limit. For these purposes the **First Image**, **Last Image** and **Res.Range** parameters are available. Image numbers may be entered in the **First Image** and **Last Image** parameters. The **Res.Range** parameters are low and high resolution limits in angstroms.

Data processing usually takes many minutes. After completion the standard d\*TREK files *ScalAverage\_1.ref* and *dtscaleaverage\_1.log* contain the integrated merged intensity data and merging statistics.

## 10.4 The Job/Molecular Replacement interface

The automated molecular replacement application may be run using the **Job/Molecular Replacement** command (Figure 10.46). This application employs the CCP4/MOLREP and CCP4/PHASER programs to carry out molecular replacement calculations involving complete rotation-translation searches.

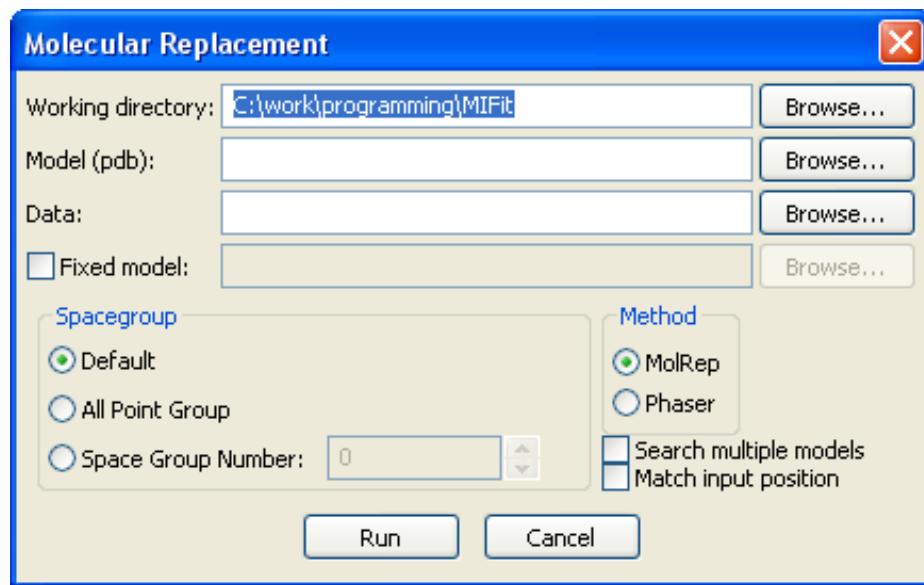


Figure 10.46 Job/Molecular Replacement dialog

Required parameters for these calculations are the **Working directory** parameter, which defines where the molecular replacement calculation outputs will be created, **Model (pdb)**, which defines the input search model and **Data (mtz)**, which defines the input structure factor data. Note that data may be entered as structure factor amplitudes in mtz format or as merged intensities in d\*TREK (.ref), SCALEPACK (.sca) or mtz (.mtz) formats.

The **Fixed model** checkbox allows the user to enter a partial model to hold fixed while carrying out a search with the model defined by the **Model (pdb)** parameter. This might be useful for (example) solving a structure with two different molecules in the crystal asymmetric unit where it is

necessary to locate the two molecules independently. However, it should be noted that the molecular replacement process is often capable of automatically determining when to search for multiple models in a crystal. For example, if your input model is a monomer but the crystal asymmetric unit contains a dimer then the process will usually place two molecules. Also, when multiple molecules are present the process will try to place them in the crystal so that they are adjoining rather than being separated by empty space.

The **Spacegroup** option may be set at **Default** to use the spacegroup encoded in the header of the diffraction data file, **All Point Group** in order to assess all possibilities within the point group (assuming data was processed in the most general space group), or a **Space Group Number** may be entered.

The **Method** option may be set to **MolPep** or **Phaser** to select between CCP4/MOLREP and CCP4/PHASER as the molecular replacement engine.

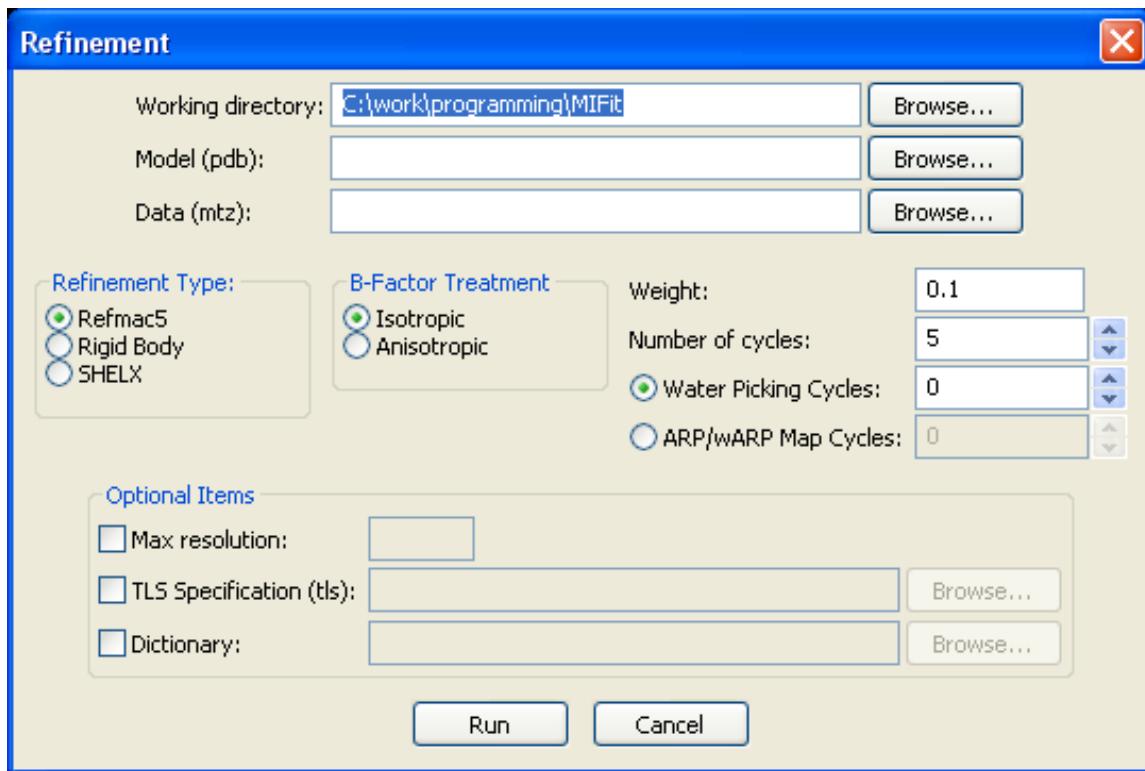
If the **Search multiple models** checkbox is selected the molecular replacement process will be run using all PDB files that are found in the same directory as the model defined by the **Model (pdb)** parameter. This option may be useful when a new structure is being solved and it is unclear which of several homologous structures would be the best candidate for solving the structure by MR. The results from multiple searches can be assessed by looking at the project history file.

The **Match input position** checkbox may be used to apply crystal symmetry operations to place the model resulting from the MR process as closely as possible to the input model. This option is useful when working on co-crystal or other projects where there are precedent structures in the same space group since it is usually convenient to have all structures placed at the same position in the crystal cell.

The **Run** button is used to launch the molecular replacement application. Output files resulting from this application are generated in the working directory and identified by file root '*molrep\_#*' where the '#' is the serial number corresponding to this molecular replacement run. The project history file is consulted in order to identify the next available sequence number and records the input, output and summary of each molecular replacement run.

## 10.5 The Job/Refinement interface

Automated refinement jobs may be run using the **Job/Refinement** command (Figure 10.47).



**Figure 10.47 Job/Refinement dialog**

The **Working directory** parameter specifies the location where the output files resulting from the refinement will be generated.

The **Model (pdb)** parameter specifies the input model and the **Data (mtz)** parameter specifies the refinement data file.

The **Refinement Type** options include **Rigid Body**, which performs rigid-body refinement of the independent chains within the coordinate set, and options to perform refinement on all atomic parameters with either **Refmac5** or **SHELX**.

Individual atomic temperature factors may be refined with either **Isotropic** or **Anisotropic** restraints.

The refinement weight applied to the x-ray term in restrained refinement may be specified using the **Weight** parameter. In practice, the optimal weight tends to larger values with higher resolution data and also tends to diminish as the refinement approaches convergence.

The **Number of cycles** parameter specifies the number of refinement cycles within a refinement run. A value of 5-15 cycles is often sufficient for preliminary refinement of a co-crystal structure.

The **Water Picking Cycles** parameter intersperses water-picking with refinement cycles. For example, if this parameter is set to a value of '3' then electron density maps will be scanned for water molecules following each of three refinement runs and the process will then be completed by one further refinement run.

The **ARP/wARP Build Cycles** parameter is similar to water picking parameter but includes deletion of protein atoms and the inclusion of dummy protein atoms at positions incompatible with real water molecules. This procedure is quite effective for improving map quality with high resolution data, typically converging in about five cycles. Note that protein molecules subject to this procedure often become somewhat ‘chewed up’ and it may be useful to use the map obtained by this procedure but base model refitting on the model prior to this procedure.

The optional items that may be supplied through this interface include **Max resolution**, which allows the user to truncate the upper resolution of data used in the refinement.

The **TLS Specification** parameter may be used to input specifications of bodies to use for TLS refinement.

The **Dictionary** parameter may be used to provide a restraint dictionary for an input structure that contains a novel small molecule ligand.

The **Run** button is used to launch the refinement application. At the conclusion of the refinement process a file containing putative errors appears in a browser window.

Output files resulting from this application are generated in the working directory and identified by file root ‘*refine\_#*’ where the ‘#’ is the serial number corresponding to this refinement run. The project history file is consulted in order to identify the next available sequence number and records the input, output and summary of each refinement run. Output files retained from the refinement process with CCP4/REFMAC5 include coordinates, phased reflection data, standard CCP4/REFMAC5 summary files (free text and mmCIF formats) and the ‘error list’ file. In order to simplify loading the resulting structure and phased data into a MIFit a session file called *refine\_.mlw* is created.

## 10.6 Checking requirement for restraint dictionaries

Protein structures that contain cofactors or other small molecule ligands may require the input of restraint information for their refinement. The CCP4/REFMAC5 installation contains dictionaries in mmCIF format for most of the small molecule entities found in public domain structures. Many of these dictionaries are in a form such that the ligand stereochemistry is automatically known to CCP4/REFMAC5 (i.e. in the same way that the standard amino acids are known) but others are in a ‘minimal’ form that and require preprocessing to a complete description before being used. In addition, some structures contain several novel ligands, or a combination of known and novel ligands. It will not always be obvious before attempting to run a refinement whether all of the entities in a coordinate file are accounted for and that the refinement will be able to proceed.

In order to check a coordinate file and, if necessary, combine various sources of restraint information, the **Job/Set Refmac5 restraints** command is available. The input for this command is the coordinate file that you wish to use for refinement and, in the same directory as that file, any mmCIF dictionary files that are considered necessary for the refinement. The dictionary files should be named according to the three character residue code in the coordinate file with file extension *.cif*. For example, if the structure contains a novel ligand ‘MMM’ there should be dictionary file available called *MMM.cif*.

The results from running this pre-refinement check are reported in the project history file. If a restraint dictionary was created that contains the information from several inputs it is named *re-*

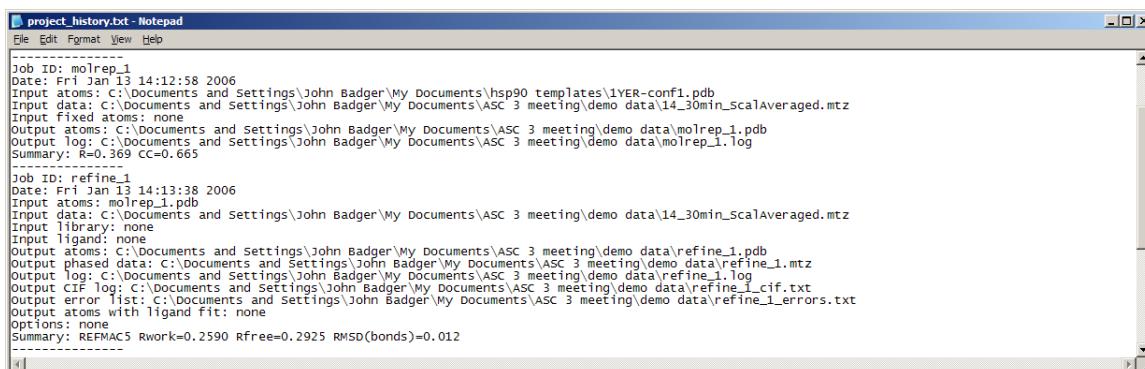
*straints #.cif*, where # is a serial number reflecting the number of times this command was executed in the current working directory.

Outputs from this process are most conveniently viewed via the **Job List** control in the document tree pane (see Section 8.3).

## 10.7 The project history file

Applicable scripted applications employ the concept of a history file (called *project\_history.txt*) that is created in and accessed from the directory in which calculations are performed. This file logs the key input and output files from each application run together with a summary of the result (Figure 10.48). This file may make a useful record, for example, of a series of molecular replacement calculations starting from different models. To maintain accurate and consistent bookkeeping standard file root names and serial numbers are used to name the output files resulting from these scripted applications. Besides providing an indexed record of the script runs, the project history file ensures that the outputs from program runs are named sequentially and that files are not overwritten.

The main automation scripts log results to the project history file upon completion i.e. failed jobs do not result in an entry.



```
project_history.txt - Notepad
File Edit Format View Help
-----
Job ID: molrep_1
Date: Fri Jan 13 14:12:58 2006
Input atoms: C:\Documents and Settings\John Badger\My Documents\hsp90 templates\1YER-conf1.pdb
Input data: C:\Documents and Settings\John Badger\My Documents\ASC 3 meeting\demo data\14_30min_Scalaveraged.mtz
Input ligand atoms: none
Output atoms: C:\Documents and Settings\John Badger\My Documents\ASC 3 meeting\demo data\molrep_1.pdb
Output log: C:\Documents and Settings\John Badger\My Documents\ASC 3 meeting\demo data\molrep_1.log
Summary: R=0.369 CC=0.665
-----
Job ID: refine_1
Date: Fri Jan 13 14:13:38 2006
Input atoms: molrep_1.pdb
Input data: C:\Documents and Settings\John Badger\My Documents\ASC 3 meeting\demo data\14_30min_Scalaveraged.mtz
Input library: none
Input ligand: none
Output atoms: C:\Documents and Settings\John Badger\My Documents\ASC 3 meeting\demo data\refine_1.pdb
Output phased data: C:\Documents and Settings\John Badger\My Documents\ASC 3 meeting\demo data\refine_1.mtz
Output log: C:\Documents and Settings\John Badger\My Documents\ASC 3 meeting\demo data\refine_1.log
Output cif: C:\Documents and Settings\John Badger\My Documents\ASC 3 meeting\demo data\refine_1.cif
Output error list: C:\Documents and Settings\John Badger\My Documents\ASC 3 meeting\demo data\refine_1_errors.txt
Output atoms with ligand fit: none
options: none
Summary: REFMAC5 Rwork=0.2590 Rfree=0.2925 RMSD(bonds)=0.012
-----
```

**Figure 10.48** Section of project history file recording molecular replacement and refinement jobs

## 10.8 Automated structure validation

Structure refinements run via the **Job/Refinement** and the **Job/Cocrystal Solution** interfaces automatically create a text file that reports certain global structure quality values and a list of amino acids in which a severe abnormality was detected (Figure 10.49). For structure refinements carried out through the **Job/Refinement** interface, this file appears in a browser upon completion of the refinement. This file may also be loaded as an annotation for a given structure. This provides a display on the main canvas of the error type and a selectable list for navigating through the error indications. Select the **Display** tab in the navigation tree, right-click on the **Annotations** icon and select **Import from error list** to load this list.

Global quality metrics include the standard crystallographic indices  $R_{\text{work}}$  and  $R_{\text{free}}$  as calculated by CCP4/REFMAC5, the percentage of residues outside the core area of the Richardson's Ramachandran plots (treating general, Pro and Gly amino acids as separate cases), the percent-

age of residues with severely abnormal  $\chi_1$  angles (using the Richardson's side chain torsion angle data as implemented in CCP4/ROTAMER program) and the total number of amino acids that appear to contain local errors. These metrics give a good indication of the quality of a structure model at a particular point in the refinement. The text file is concise enough to be printed and included in a laboratory notebook.

More useful for achieving model improvement is the information from the list of local errors, which may be used as targets for model examination and rebuilding efforts. Although proteins do contain some genuine structural anomalies (with interesting energetic rationales) and there is a 'grey zone' in assessing what degree of misfit between data and map is acceptable, the calculation routes and thresholds used to identify most of these errors types have been tested on very large numbers of structures and will usually indicate portions of the model that need to be checked and corrected. The statistical data used for these tests and calculation routes are more sophisticated and modern than standards embodied in the PROCHECK (1993) program and the density checks supply a means of detecting geometrically correct features that misfit the data.

```

refine_4_errors.txt - Notepad
File Edit Format View Help
#
# Coordinates: refine_4.pdb
# Data: refine_4.mtz
#
# Rwork: 0.2033
# Rfree: 0.2450
# Percentage of residues outside Richardson phi-psi core: 0.0
# Percentage of residues with abnormal rotamers: 0.0
# Percentage of residues flagged: 5.3
#
# Residue list codes for severe abnormality types:
# (G)eometry, (V)an der waals, (O)mega, (P)hi-psi, (C)is peptide,
# (R)otamer chi-1, (D)ensity
#
A 28 GLN . . . . . D
A 31 SER . . . . . D
A 60 ARG . . . . . D
A 75 GLU . . . . . D
A 85 GLN . . . . . D
A 158 GLU . . . . . D
A 176 THR . . . . . D
A 192 GLU . . . . . D
A 208 LYS . . . . . D
A 209 LYS . . . . . D
A 212 GLN . . . . . D
#

```

**Figure 10.49** Example of an 'Error list file' with density errors.

More specifically, the criteria for flagging structure errors are:

Geometry errors – bonds length deviations greater than 6x the refinement sigma or bond angles greater than 8x the refinement sigma used by REFMAC5

Van der Waals errors – close contact deviations greater than 4.25x the refinement sigma used by REFMAC5

Omega errors – deviations of 4x the true deviation ( $5.6^\circ$ ) from the peak ( $178.9^\circ$ )

Phi-Psi errors – lies outside a prescribed area containing 99.95% of correct amino acids in the general area, 99.8% of correct amino acids for GLY and 99.8% of data for PRO in the Richardson tabulations

Cis peptide errors – non-PRO amino acid flagged as cis by REFMAC5

Rotamer errors (chi-1) – deviations of more than 45° in chi-1 from the nearest conformer as calculated by ROTAMER, using rotamer data from the Richardson lab

Density errors – identification of a 5 sigma peak/hole within 1.5 Å of any model atom or a N/CA/C/O atoms outside a 0.85 sigma envelope in the likelihood map or a significant number (depending on residue type) number of atoms outside a 0.85 sigma envelope in the likelihood map

In addition to reporting these abnormalities via the ‘error list file’, they are also documented as ‘REMARK 500’ records in the resulting .pdb file.

## 10.9 Automated ligand fitting

A fully automated application is available via the **Job/Cocrystal Solution** commands for placing a copy of a ligand using a 6D search procedure. The required input is a PDB file of the ligand. If multiple conformers are present then they should be separated by TER records. This automated fitting option is mainly useful for placing rigid molecules and for projects where a specific ligand conformation is anticipated (perhaps by analogy with similar compounds for which structures have already been determined). A tool for fitting ligands in the context of an interactive MIFit session is discussed in section 10.9.

For the 6D ligand search option to be applied a ligand target point must be specified. The knowledge of this target point is used to eliminate water molecules from the ligand site (which would otherwise overlap the ligand density and interfere with placement) and to ensure that the placed ligand is put in a ‘standard’ site and not in density related by crystallographic symmetry. This option may be applied by entering coordinates for the ligand to be fit as the **Place Ligand (pdb)** parameter. The associated X,Y,Z parameters define the position of the target site in angstroms.

The search procedure uses CCP4/FFEAR program to carry out a complete rotation-translation search using the ligand difference density as the target. In order to speed-up this process an initial search is run on a relatively coarse angular grid (15-20° increments) followed by a local search around the best solution on a finer grid (3° increments). The search speed depends on the space group and the number of data included in the calculation (data to 2.8-2.4 Å resolution are used). The placement of very small ligands takes longer than larger ones because the determination of the correct orientation requires higher resolution data and a finer search grid.

Ligand placement does not require a ligand dictionary since the operating processes only involve a 6D search and no refinement is performed. (If examination of the resulting model shows that the ligand placement was successful then a ligand dictionary may need to be supplied for subsequent refinement cycles).

## 10.10 Semi-automated ligand fitting

MIFit contains a command (**Refi/Find Ligand Fit and Conformer**) for docking ligands to difference density during model-building sessions.

This command performs a rotation-translation search for a ligand arbitrarily overlapped onto the ligand density in a difference map and selects the best fitting conformer from the conformers present in the MIFit dictionary (see Chapter 6 for information on loading the MIFit dictionary via the Ligand Dictionary Editor). This command applies a genetic algorithm (D.E.McRee, Acta Cryst. D60, 2276-2279, 2004) to search positional parameters (rotation and translation space) and to select the ligand conformer for which the atom centers overlap best with the ligand difference electron density.

The ligand fitting method works best on difference maps that show strong, well-defined ligand densities. Although there is no formal limitation on the quality of the target density, the interface to this application currently limits the map data to be better than 2.5Å resolution. Trials suggest that this is the resolution at which the shape of the density will usually be sufficiently distinctive to correctly fit a ligand. The fitting process will usually work better if stray electron density features (usually corresponding to ordered water molecules near the ligand) are modeled so as to remove them from the difference map.

Large and flexible ligands (more often biologically relevant cofactors than molecules that synthesized in industrial lead discovery and optimization programs) tend to be the more difficult to fit than small rigid examples; molecules that contain more than 60 non-hydrogen atoms will probably be too large to fit and are prevented by the interface. On the other hand, compact molecules (say consisting of 3-4 planar groups separated by torsion angles) are often fit quickly and correctly.

To use this ligand fitting tool, a set of ligand conformers must be present in the MIFit ligand dictionary (Chapter 6). When the ligand density has been identified in a difference density map, the ligand molecule may be added to the protein coordinate data by selecting the required entity using the **Model/Add residue** command. This selection creates the dialog window shown in Figure 5.18.

The **Residue Type** pulldown may be used to select the ligand from the list of entities in the MIFit dictionary. For this application the **Insert Position** parameter should usually be ‘End of Model’ and the **Put At** parameter will be ‘Screen Center’. The **Chain id** field may be used to override that default chain selection by entering another single character chain-id – usually when a user wishes to enter a new, unique chain-id for the ligand. Clicking on **OK** will place the ligand in an arbitrary orientation at the screen center, which should have been adjusted to overlap the ligand difference density.

The ligand may be selected using the **Fit/Fit Residue** command (or keyboard shortcut f) and the fitting process may be launched using the **Refine/Find Ligand Fit and Conformer** command. The process will terminate when either the maximum allowed number of generations is reached in the Genetic Algorithm. Scores and progress for a running process are displayed in the log window pane and the right segment of the status bar. It should be noted that the position of the ligand that is displayed in the graphics during a running ligand docking process is a current position and not the best fit to that point. The best fit is loaded when the process completes. This procedure typically takes ~1 minute and is directly proportional to the number of conformations (often 200-300).

In some cases the results in a ligand that is only approximately fit to the ligand density. When this occurs, or when an approximate fit for the ligand was obtained via interactive model-build-

ing. Another useful possibility is to use the **Refine/Refine Residue** option to apply real-space refinement to optimize the fit of ligand to map.

## 10.11 Rebuilding models using non-crystallographic symmetry

When building models of protein crystals in which there is more than one subunit per asymmetric unit it is usually the case that model-fitting efforts are concentrated on one copy (usually defined by a single chain id in the PDB coordinate file) and it would be desirable to replicate this rebuilt model across all copies of the protein.

The **Job/NCS Modeling** interface (Figure 10.50) is designed to provide a convenient mechanism for this task.

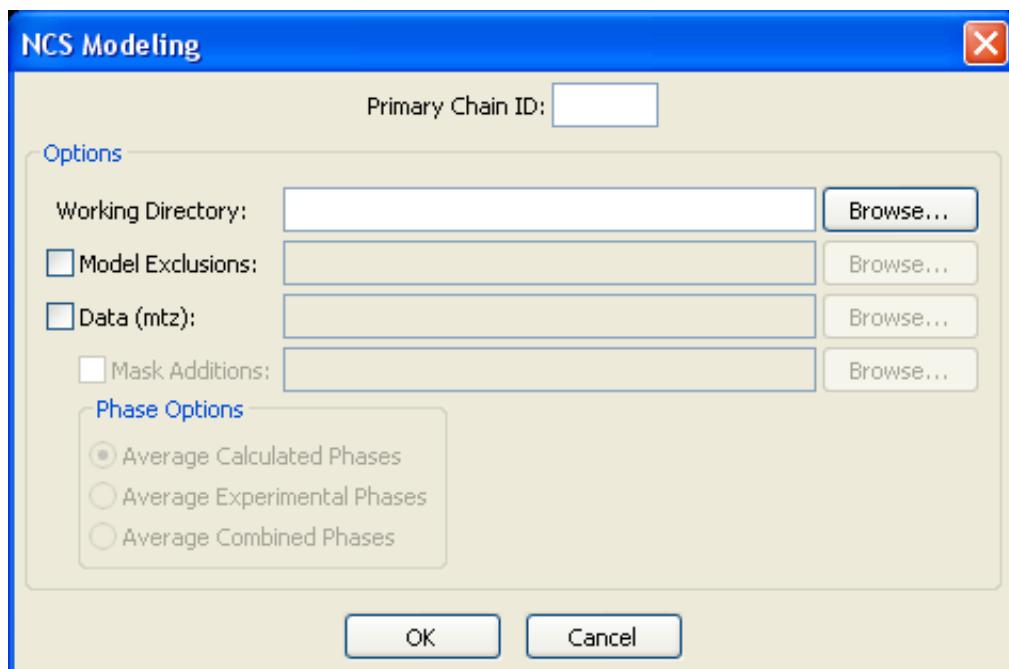


Figure 10.50 Job/NCS Modeling interface

This interface is available when a protein model is loaded into the MIFit canvas. Required parameters are the **Primary Chain ID**, which defines the template chain (the model that the user wishes to replicate over less completely modeled chains in the structure), and the **Working Directory** where the rebuilt model will be placed.

The optional **Model Exclusions** parameter may be used to provide the path to a file that contains data lines for chain-id, start-amino-acid-number, end-amino-acid-number over which the NCS symmetry will not be applied. For example, 'B 120 130', would exclude chain B, amino acids 120-130 from rebuilding.

When launched by selecting **OK** the rebuilt model is returned to the MIFit canvas.

Optionally, a diffraction data file may be entered with the **Data (mtz)** parameter and the density map is averaged within the model envelope to provide phases according one of the **Average Cal-**

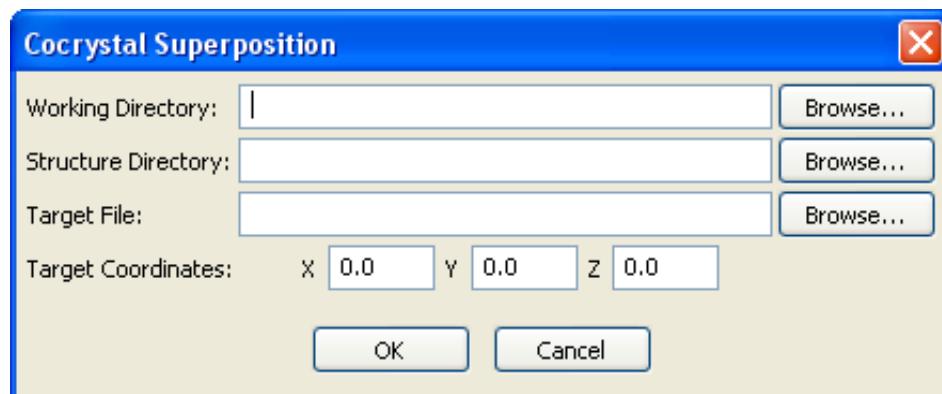
**culated Phases**, **Average Experimental Phases** and **Average Combined Phases** selections. The process will automatically calculate phases but expects the input file to contain Hendrickson-Lattman ABCD coefficients to use experimental phases.

The **Mask Additions** option provides a specification (a file containing lines with chain-id, amino acid-number) for including atomic groups outside the primary chain in the subunit mask. The output mtz file is named after the input pdb file but prefixed with *ncs*\_.

## 10.12 Automated multiple protein:ligand structure superposition

A very common scenario in protein crystallography is that several closely related structures have been solved (i.e. the proteins have identical sequences and are frequently crystallized in the same space group). The structure analyst wants to compare the binding modes of many ligands within the active site. To perform an accurate comparison, only atoms around the ligand binding sites should be used for the structure superposition. The method used by MIFit is to base the structure superposition on Ca atoms within 15Å of an active site coordinate within a reference structure.

The **Job/Cocrystal Superposition** command is used to run the structure comparison.



**Figure 10.51** Cocrystal Superposition interface

This command is only accessible when a model is loaded into the MIFit main canvas. It will often be convenient to load the reference structure that will be used as a basis for the structure superposition.

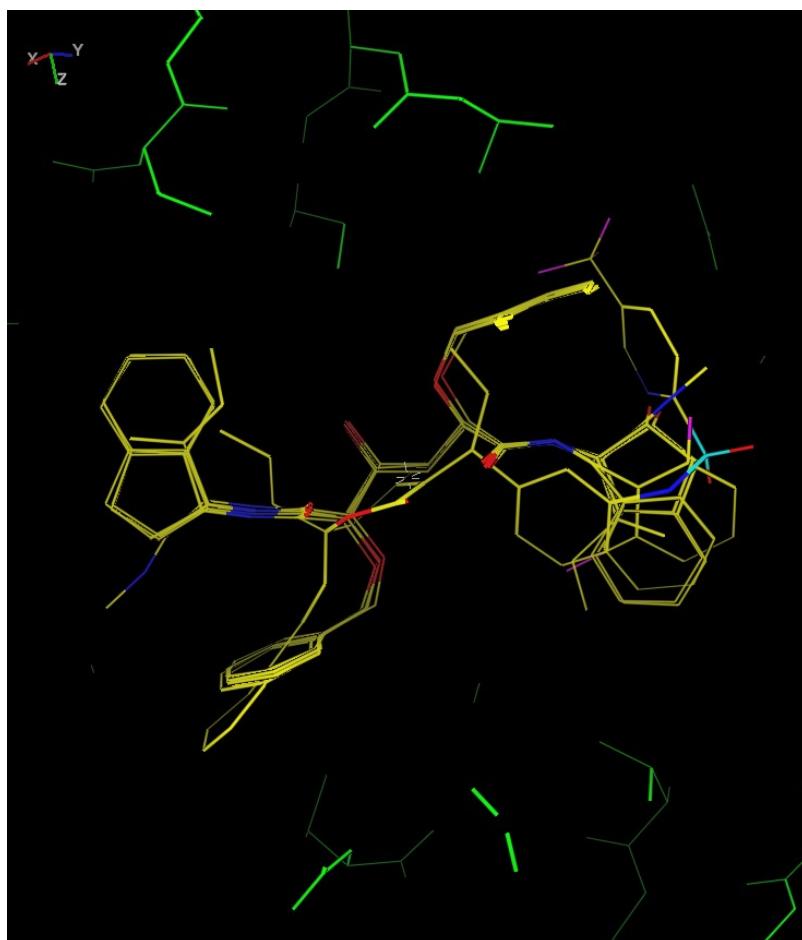
The **Working Directory** parameter defines the path to the directory in which the results from the structure comparison will appear.

The **Structure Directory** should contain all the coordinate files (PDB format) for the protein:ligand structures that will be compared. These structures must have consistent sequence naming conventions. i.e. residue A 22 in one structure corresponds to residue A 22 in all the others.

The **Target File** parameter defines the particular coordinate file that will be used as the reference structure for comparisons. This file is usually one of the structures in the **Structure Directory** but need not be.

The **Target Coordinates** are the **X**, **Y** and **Z** positions in angstroms in the **Target File** for the center of the ligand binding site. This position is used to select the surrounding Ca atoms for structure superposition.

After clicking on **Run** the superposition process will return a coordinate file called *allligands.pdb* to the **Working Directory** and load this file into the MIFit main canvas. This file contains all of the ligands with the structure set superimposed together and distinguished by unique chain ids.



**Figure 10.52** Six superimposed HIV protease inhibitors

Figure 10.52 shows a typical result of the structure superposition application). In this example the reference protein structure was loaded into MIFit and has been colored green to distinguish it from the model containing the six ligands. It is evident from this image that parts of some of the bound molecules superimpose extremely closely but the part of the binding site on the right of the image is filled by molecules in a wide range of conformations.

# 11 Tutorial Lessons

## 11.1 Lesson 1: Basic manipulation of the model display

This lesson introduces some of the mechanisms for controlling the view and representation of an atomic model.

1. Copy the file *automation\_out.pdb* from the *examples* directory of your MIFit installation into a new directory.
2. Select the **File/Open models, data, maps, etc...** command to load a coordinate file in PDB format. Use the **Look in** pulldown menu in the resulting browser window to locate the file in the directory used for (1). Load the file (by clicking on it) and select **Open**. You should see the protein model appear in the main canvas window.
3. Try rotating the model by dragging the left mouse button in the canvas window.
4. Try zooming in and out by dragging the left mouse button while holding down the Shift key. Dragging the left mouse button up zooms in and dragging the mouse button down zooms out.
5. Try panning the model by dragging with right mouse button in the canvas window.
6. Notice that if the cursor is positioned in the extreme top of the main canvas the function of the left mouse button changes so that dragging it left or right rotates the view about the Z-axis. The function of the right mouse button also changes so that it controls the slab.
7. Try clicking on the slab icons ( and ) in the tool bar to slab the canvas view in and out.
8. Try re-centering the display by double-clicking on an atom in the main canvas.
9. Try centering on a residue by double-clicking on an amino acid in the **Residues:** section of the navigation tree on the left.
10. Try centering on a residue by entering a chain-id, residue name pair (say, ‘X 35’) in the **Go to residue** parameter box at the top of the navigation tree on the left.
11. Try centering on a residue by double-clicking on an amino acid in the sequence window on the right.
12. Try single-clicking on an atom in the main canvas. Notice that this atom is placed in the stack in the lower left of the canvas. The stack is used to identify targets for commands that operate on specific atoms or residues.
13. Try coloring a residue by clicking on a residue and then clicking on the  button and selecting the **Color Last Picked Residue** option. In the resulting dialog box select a color by clicking on it and then use the **Method** pulldown menu to select to ‘All atoms’. When you select **OK** you should see the color of all atoms in the selected residue change. Try some of the other color options. You can also perform most of the same commands from the **Render/Color** options.

14. Try changing the rendering style by selecting the different options from the **Render** pull down menu. Many people prefer to use the **Sticks** option for speed and simplicity when model fitting.
15. Try the command **Viewpoint/Center model on screen** to zoom and slab the model automatically to fill the screen.
16. Try changing the model to a backbone only representation by selecting the **Show/Backbone>Show backbone as CA trace** and the **Show/Sidechains/Hide side-chain atoms** commands. You should now see only connected CA atoms of each residue (and any ligands or prosthetic groups, if they were present). It is simpler to see the overall fold of the protein in this representation
17. Try the command **Show/Secondary Structure/Make Ribbon** to make a ribbon representation. A popup dialog box asks whether you wish to retain the original representation of the model (the CA trace) in addition to building the ribbon diagram. You may select ‘Yes’ to remove the CA trace. The resulting model will be represented in different colors for helix, sheet and coil portions of the molecule.
18. Try changing the ribbon colors by selecting the **Show/Secondary Structure/Ribbon Colors...** command. Clicking on each of the **Helix Color...**, **Sheet Color...** and **Random Coil Color...** buttons will provide a color palette for selecting new colors for each of these secondary structure types. Select red for the helix, yellow for the sheet and green for random coil. Click **OK** to see the model re-colored with these choices. You may also note that the secondary structure definitions displayed in the sequence view may be toggled through alpha, beta and coil secondary structure types by clicking on the secondary structure symbol in that view (then **Show/Secondary Structure/Make Ribbon** will reload the ribbon diagram with the new definition).
19. Select **Show/Secondary Structure/Clear Ribbon** to remove the ribbon display.
20. Select **Show/Secondary Structure>Show Tube Secondary Structure** to draw the molecule as a glossy tube. You may answer either ‘Yes’ or ‘No’ to the dialog box that asks ‘Do you want to hide residue atoms’. Select **Show/Hide Secondary Structure** to remove the tube representation.
21. Select **Show/Secondary Structure>Show Schematic Secondary Structure** to draw the molecule with schematic secondary structure rendering. You may answer either ‘Yes’ or ‘No’ to the dialog box that asks ‘Do you want to hide residue atoms’.
22. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



**Figure 11.53** Lesson 1

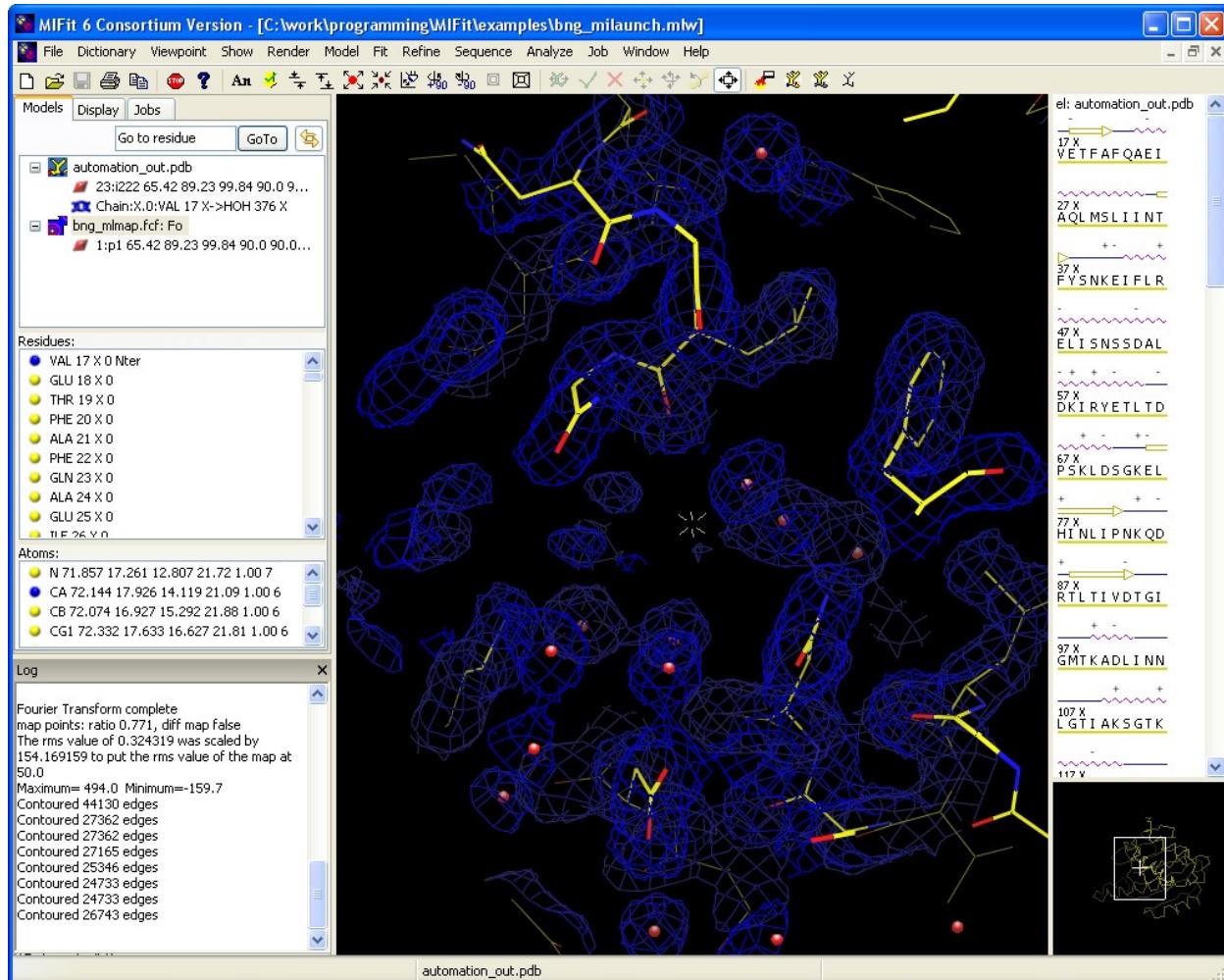
## 11.2 Lesson 2: Loading and changing electron density map displays

This lesson introduces some of the mechanisms for loading and adjusting the display of electron density maps.

1. Copy files *automation\_out.pdb*, *bng\_mlmap.fcf* and *bng\_milaunch.mlw* from the *examples* directory of your MIFit installation into a new directory. These files are a coordinate file, a file of phased x-ray data and a MIFit session file for loading the coordinates and data.
2. Use the **File/Open models, data, maps, etc...** command to load the file *bng\_milaunch.mlw* from the directory (1) in which it was placed. You will need to change the file filter to select for files with extension ‘.mlw’. You should see the electron density map appear around the protein model in the main canvas. Note that space group and cell information (beside the crystal icon) are read from input coordinate file for the model. Similarly, this information is read from the .fcf data file although this format encodes symmetry operators directly rather than the space group.
3. Try using the right mouse button to pan across the map. You should see that the map is automatically re-contoured as you scroll across the main canvas and you move to the edge of the current box of map density.
4. The map size and contour levels can be changed using the **Contour Options...** command in the map’s menu which is accessed by right-clicking on the map in the Models tree. Move the **Radius** scrollbar to change the size of the display density to a 9Å box. From the pull down menu in the **Preset Map Styles** list box select the first option - ‘Blue Map 1,2,3,4 5 sigma’ – and click on **OK**. The lowest contour level for displaying this map is set to one sigma (Crystallographers refer to the root-mean-square density fluctuation of the map in terms of ‘sigma’, although this may be confusing as it does not relate to the error in the electron density for this type of map). MIFit internally scales maps so that one sigma is set to 50 units. The other density levels are initially set to 2 sigma, 3 sigma etc for up to 5 contour levels.
5. Again select the **Contour Options...** in the map’s menu Try turning off contour levels 2 and 4 by removing the checks from the associated **Show** parameters. Click on **OK** and you should see these two intermediate contour levels disappear from the display, leaving a less cluttered image.
6. Try changing the resolution of the map by selecting the **FFT Phases...** command in the map’s menu, setting the **Min Resolution** parameter to 3.0 and selecting **OK**. The operating default for the FFT had been set to create the map using the full resolution limits of the data but you should now see the map at 3Å resolution showing much less detail.
7. Try changing the grid spacing for the map contours by selecting the **FFT Phases...** option again and changing the **Grid** option from the default (‘Medium Grid’) to ‘Fine Grid’. Click on **OK** and you should see that the map is returned at the same resolution with a much smoother appearance. The ‘Medium Grid’ setting is used for most model building activities since extra contour lines impede graphics performance and may

make the model hard to see but presentation images sometimes look better when over-contoured on a fine grid.

8. Zoom out the model and select the **Show/Symmetry Atoms/Show symmetry atoms as atoms** command. You should see some symmetry related atoms (purple) appear around the molecule. Alternatively, you can use the **Show/Symmetry Atoms/Show symmetry atoms as CA trace** to see just a CA trace of the symmetry related model.
9. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



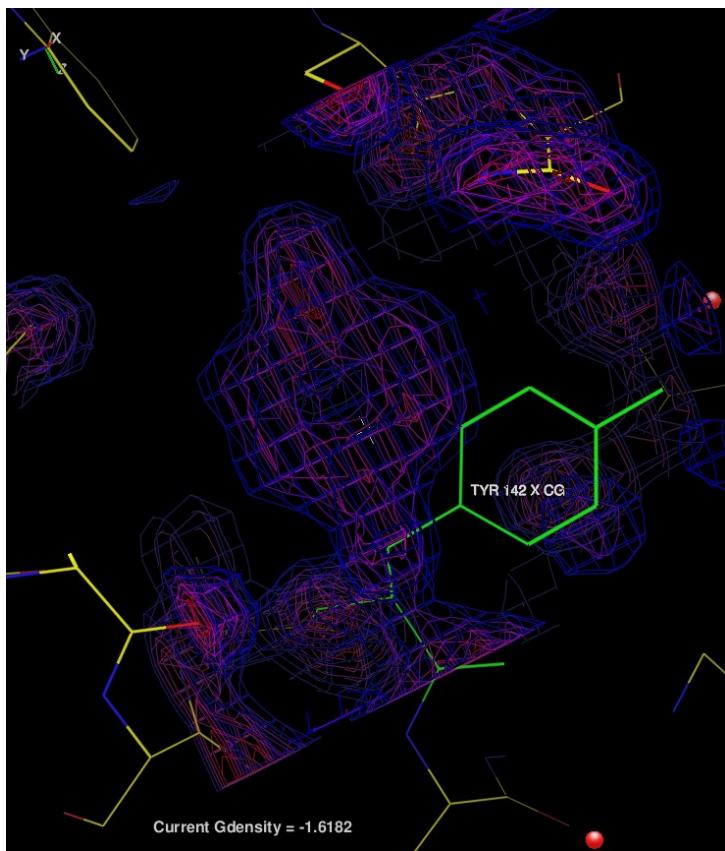
**Figure 11.54** Lesson 2

## 11.3 Lesson 3: Refitting individual residues

*Scenario: You have been refining a protein structure which is correct except for the positions of a few residues which still misfit the electron density map after the last cycles of refinement. You wish to refit these amino acids. The lesson illustrates methods for refitting individual amino acids.*

1. Copy files *automation\_misfit.pdb* and *automation\_out.mtz* from the *examples* directory of your MIFit installation into a new directory. These files are a coordinate file that contains a few model-fitting errors and file of phased x-ray data in mtz format from a refinement with CCP4/REFMAC5.
2. Select the **MIFit File/Open models, data, maps, etc...** command to load file *automation\_misfit.pdb* and *automation\_out.mtz* from the directory established in (1). (Ctrl+click to select both files.)
3. A dialog will appear to allow selection of MTZ columns labels. In this example we will create the electron density map from pre-computed likelihood-weighted structure factor coefficients and phases that were calculated by the REFMAC5 program. i.e. we will not use the true values for  $F_{\text{obs}}$  and phase but will simplify the calculation route by utilizing pre-computed map coefficients. Since MIFit automatically recognizes MTZ files with standard REFMAC5 column labels, the dialog will automatically be set to load these values. Simply click on **OK**.
4. Scroll down to residue X 142 in the **Residues:** section of the navigation tree and double click on it. This action will re-center the model and map on the C $\alpha$  atom in residue X 142 Tyr. You will notice that the side chain of this amino acid does not fit the density.
5. Select the residue to be fit (X 142 Tyr) by clicking on any atom in it in the main canvas. Now select either the **Fit/Fit Residue** command or the  icon in the toolbar or the keyboard shortcut 'f'. The residue should turn green to indicate that it is active for fitting.
6. Try translating the residue by either clicking on the **Fit/Translate** command or the  icon in the toolbar. Using the right mouse button you can now move the selected residue around the canvas.
7. Try rotating the residue by either clicking on the **Fit/Rotate** command or the  rotate icon in the toolbar. Using the right mouse button you can now rotate the selected residue around the canvas.
8. Reset the position of the amino acid back to the starting position but leaving it in fitting mode by selecting the **Fit/Reset Fit** command.
9. To activate rotation of the side chain about the  $\chi_1$  angle (i.e. along the bond between the C $\alpha$  and C $\beta$  atoms) enter '1' on the keyboard. You should see a grey arrow that points from the C $\alpha$  to the C $\beta$  atom appear on the model in the main canvas window. Use the right mouse button to rotate the side chain about  $\chi_1$  into the electron density.

10. The side chain also needs to be adjusted about the  $\chi_2$  angle (i.e. the bond between the C $\beta$  and C $\gamma$  atoms). Enter '2' on the keyboard and you should see the grey arrow shift to point from C $\beta$  to C $\gamma$ . Use the right mouse button to rotate the side chain about  $\chi_2$  into the density.
11. In a real fitting session you would have accepted this fit either by selecting the **Fit/Accept Fit** command, by clicking on the  icon on the toolbar or by using the keyboard shortcut ';' . Here, select **Fit/Reset Fit** so that the lesson can continue to show other fitting methods. Select the  (cancel fit) icon to deactivate this residue.
12. A convenient way to refit side chains is to use the **Model/Replace and fit** command (or the keyboard shortcut 'r'). To try this click on any atom in X 142 Tyr in the main canvas. After selecting the replace-and-fit command a dialog box will popup containing a menu with all possible entity types. Since we want to retain the default amino acid type (Tyr) select **OK**. You will see that the Tyr side chain move to fit the electron density. To save this fit you could select the  icon from the toolbar. Alternatively, to cancel this fit click, select the  icon from the toolbar.
13. We will now correct a common type of main chain fitting error, a 'peptide flip', in which the carbonyl oxygen is rotated approximately 180 degrees from the correct orientation. Make sure that Chain:X in the navigation tree is still selected (icon is colored blue) and scroll the **Residues:** list to residue X 37 and double click on it. This action will re-center the model and map on the C $\alpha$  atom in X 37 Phe. You will notice that the main chain of this amino acid does not fit the density.
14. Click on the carbonyl (O atom) in X 37 Phe. Select **Fit/Fix Backbone/Flip Peptide**. You will see the peptide plane flip over so that the carbonyl now properly fits the electron density.
15. You will now use a pentamer library to check a part of the protein structure. Make sure that Chain:X in the navigation tree is still selected (icon is colored blue), scroll down to residue X 192 and double click on it. This action will re-center the model and map on the C $\alpha$  atom in Glu X 192 Glu.
16. Click on any atom in X 192 Glu in the main canvas. Select **Fit/Fix Backbone/Suggest Backbone Match**. You will see a pentamer match to the backbone, starting at X 192 Glu, appear on the screen. In this case the protein conformation is similar to the pentamer and there is no reason to change it. If the atomic model dims you may turn off the dimming by unchecking the **Render/Dim Non-active models** command. The image of the pentamer match can be removed with the **Fit/Fix Backbone/Clear Backbone Match** command. In cases where the backbone should be refit to match the pentamer the four **Fit/Fix Backbone/Replace** commands could be used to replace parts of the structure.
17. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



**Figure 11.55** Lesson 3

## 11.4 Lesson 4: SAD phasing

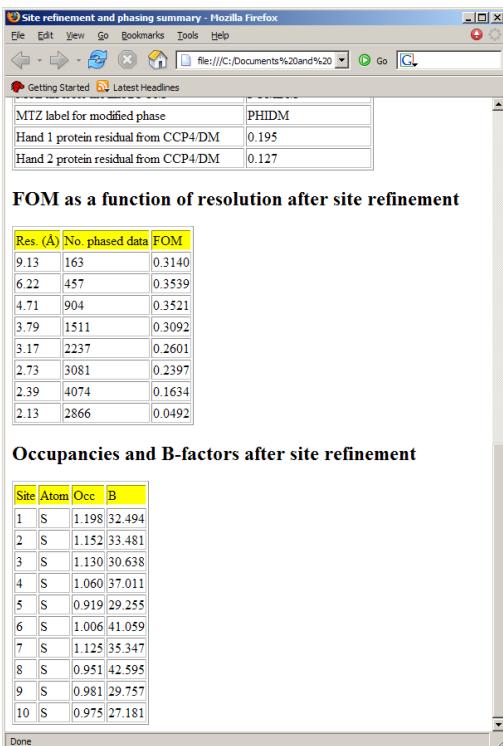
(This tutorial requires that MIFit has access to SHELXD and CCP4.)

*Scenario: You have collected full data (i.e including anomalous differences) to 2.1Å on a Cr wavelength source from a new 318 amino acid protein that contains 10 sulfur atoms. You wish to compute an electron density map from this SAD data for model-building. This data was supplied by Aiping Dong (Structural Genomics Consortium, University of Toronto) and corresponds to PDB entry 2AZP.*

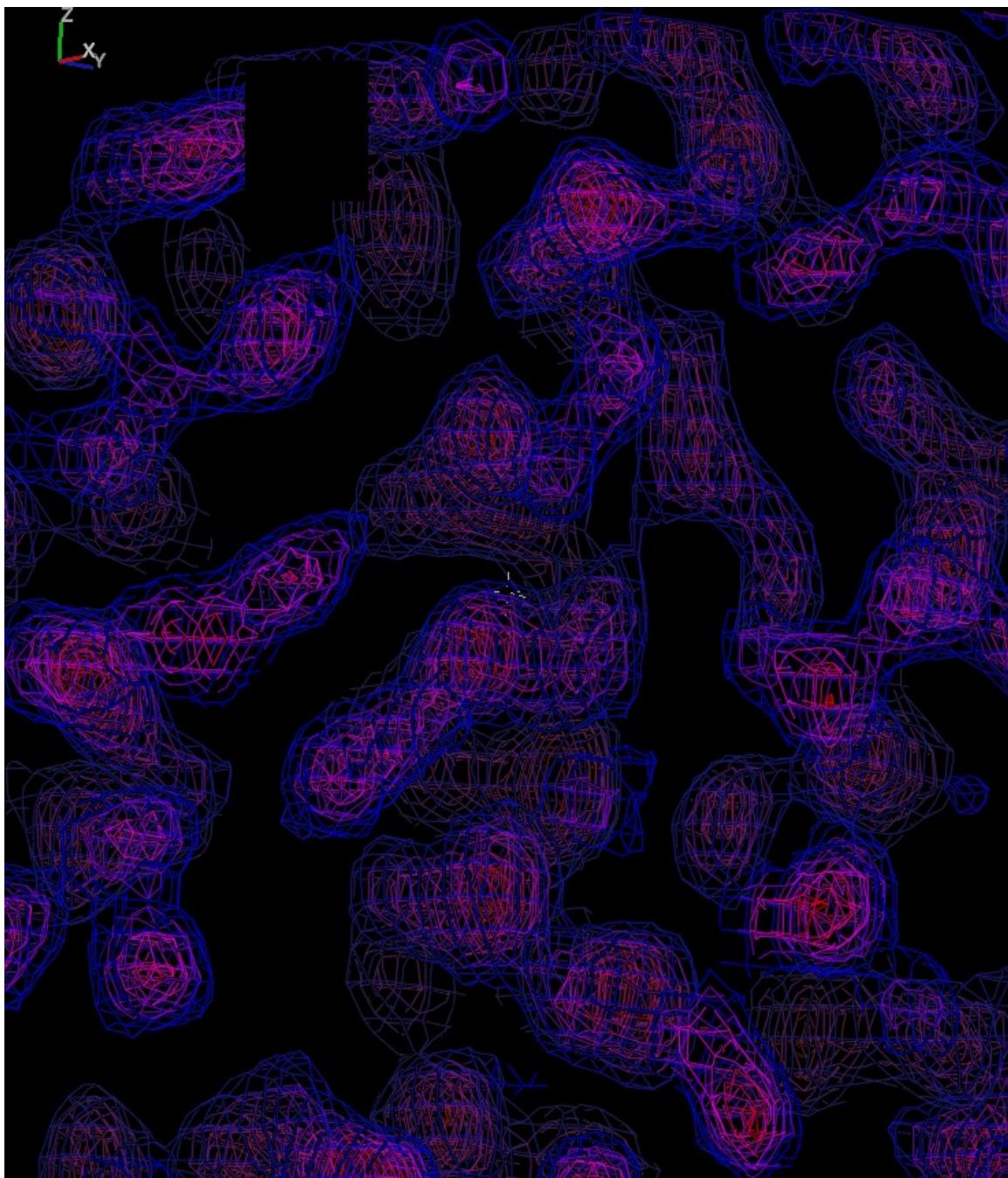
1. Copy file *ano.sca* from the *examples* directory of your MIFit installation into a new directory. This file contains the integrated and merged intensity data for the SAD phasing process.
2. Select the **Job/SAD Phasing** application from the MIFit interface.
3. Use the **Browse...** button to set the **Working directory** parameter to the directory created in (1).
4. Use the **Browse...** button to set the **Intensity data** parameter to the file *ano.sca*.
5. Since this is an entirely ab initio structure determination the method for determining **Scatter sites** should remain set to automatic. The **Scatter type** should remain set to S (sulfur), the **Number of sites** should be set to 10, the **Number of disulfides** should be set to 0, the **Min scatterer separation** parameter should remain set at 3.5Å.
6. Set the **Solvent fraction** parameter to 0.49 (the expected solvent volume for this protein size/crystal cell). The option to **Phase from both site enantiomorphs** is needed so that checkbox should remain selected and the option to **Change spacegroup number** should not be invoked. The **Site Refinement Method** option may remain set to **mlphare** to use the CCP4/MLPHARE program for site refinement.
7. Click on **OK** and the SAD phasing job should launch. The job will typically take 5-10 minutes (with most of the time spent running multiple site finding trials). Upon completion, a browser window will appear providing diagnostic information on the phase determination. For potentially successful phase determination the figures of merit should be greater than 0.30 in several resolution shells and most of the sites should have consistent occupancies and relatively low B-factors. In this example all sites have occupancies of ~1.0 with B-factors in the range 27-42Å<sup>2</sup>. The hand with the lowest density modification residual is more likely to be the correct solution (in a genuinely new example it is wise to inspect both maps).
8. Click on the **Jobs** tab near the top of the navigation tree. In the **Jobs List** you will see an icon representing the job (colored yellow for running, green for completed and red for failed). Right-click on the bottommost entry in the job list and select **Show Log File**. This action will provide a view of summary information from the structure superposition process. Note that this command is inaccessible until the job completes.
9. The simplest way to load maps for visual inspection is to make use of the pre-computed session files. Use the **Open Results...** command in the job's menu which is accessed by right-clicking on the job in the **Jobs List**. In the Open Job Results dialog,

the **Load Session** selection will already show the *mi\_sad.mlw* file because it is the most recent *.mlw* file. Uncheck **Load PDB** and **Load Data**; they are not needed for this example. Click **OK**. The loaded session will display two maps: one map for phasing with hand 1, the original site constellation, and the other map for phasing from the inverted site constellation. The map contours are colored with different styles to distinguish them. They can be inspected separately by toggling each map on and off with the map's **Show/Hide** command (accessed by right-clicking on the map in the Models tree). Here, the map corresponding to the hand with the lowest residual is correct.

10. You should search the map display in order to visually identify protein secondary structure features, indicative of an interpretable map.
11. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



**Figure 11.56** Lesson 4, SAD results



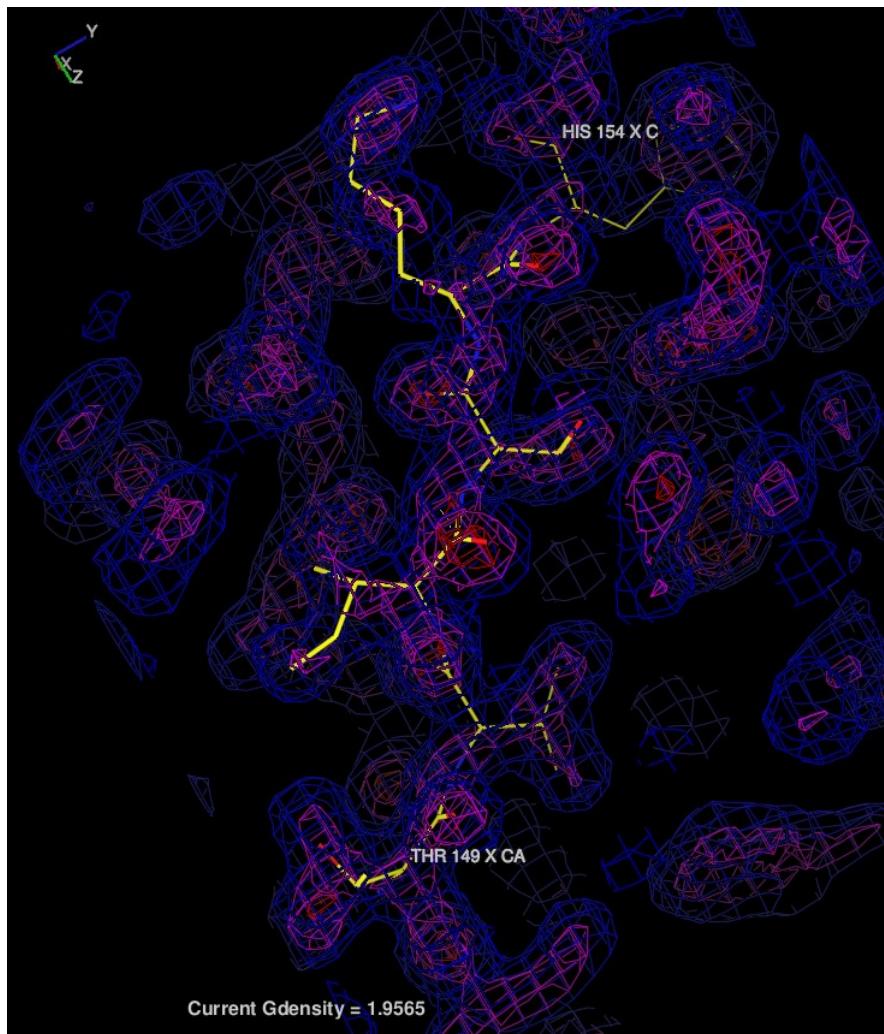
**Figure 11.57** Lesson 4, main view

## 11.5 Lesson 5: Chain tracing

*Scenario: You have just obtained an electron density map and wish to build a model into the electron density. (Note: in many cases automated model-fitting programs are now used to build initial models into experimentally phased maps or molecular replacement methods are used to obtain an approximate model. The manual building process is often limited to tracing loops and model correction.)*

1. Copy files *start\_trace.pdb*, *start\_sequence.seq* and *automation\_out.mtz* from the *examples* directory of your MIFit installation into a new directory. The file *start\_trace.pdb* just contains two C $\alpha$  marker atoms ('MRK' residues) and is used to initiate the model-building process and guide the tutorial. The file *start\_sequence.seq* contains a sequence fragment for the piece of the protein chain that we will build. The file *automation\_out.mtz* contains phased structure factor data that will be used to compute an electron density map into which the model will be built.
2. Use the command **File/Open models, data, maps, etc...** to load coordinate file *start\_trace.pdb* and data file *automation\_out.mtz* from the directory in which you placed it. In the Phase file import dialog, disable loading of the difference map by clicking on the checkbox to the left of the second group of maps settings. Click **OK**. The two-marker structure and map should appear in the main canvas.
3. Make sure that the region of the map has appropriate volume and contour levels for model building. Select the map's **Contour Options** command (accessed by right-clicking on the map in the Models tree). In the resulting dialog box, move the **Radius** slider to about 10 in order to set a 10Å box for the map volume. You may wish to set the **Preset Map Styles** option to 'Blue Map, 1,2,3,4,5 sigma' and then uncheck the display of contour levels **2** and **4** to remove some of the contour lines from the map display. Select **OK** and you will see the map view change to these settings.
4. In the **Go to residue** parameter box at the top of the navigation tree enter 'X 150'. This action will center the display on C $\alpha$  marker X 150.
5. Click on the (centered) MRK 150 entity in MIFit main canvas and then select the **Model/Add MRK After** command. You should see a green line appearing to connect that marker to the expected position for the next C $\alpha$  atom.
6. By continuing to select **Model/Add MRK After** or by using the keyboard short cut '**SHFT>**' trace out four more C $\alpha$  positions.
7. Select **Refine/Accept Refine** and **Fit/Accept Fit** to accept this trace.
8. Select **Model/Poly-Ala Chain** to convert this chain trace to poly-alanine.
9. The C-terminal residue is not fully converted to ALA (it remains as a MRK residue) and may be deleted by clicking on the MRK atom in the main canvas. Then right-click in the main canvas and select **Delete Residue** from the pop-up menu.
10. Since this is a high resolution map showing atomic detail it may be useful to optimize the poly-ALA trace before adding side chains. Click on an atom at the N- and C-terminal ends of the chain. Select **Refine/Refine Range** to activate this chain for refine-

- ment (it will change color to blue) and then hold down the space bar to drive the refinement to convergence. Accept the result by selecting **Refine/Accept Refine**.
11. Select **Sequence/Read Lower Sequence** and use the dialog box to enter the file *start\_sequence.seq*. You will notice that the sequence code TVITKHN has appeared in the sequence window.
  12. Select atoms in residues X 149 and X 154 (i.e. both ends of the poly-alanine fragment) and then select the **Model/Replace with Sequence** option. After selecting **Yes** in the popup dialog you will see the poly-alanine fragment replaced by the correct amino acid sequence.
  13. The fit is imperfect – as an optional exercise you could use some of the methods learned from tutorial 3 to refit individual amino acids. You could also repeat the model refinement (step 13 above) with this model.
  14. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



**Figure 11.58** Lesson 5

## 11.6 Lesson 6: Establishing restraints for structure refinement

*Scenario: You have completed the preliminary refinement of a new protein structure and now wish to create a dictionary for a small molecule ligand molecule within the structure for use with CCP4/REFMAC5. You have obtained a SMILES string describing the molecular structure of the ligand.*

1. Copy the file *ligand\_example.smi* from the *examples* directory of your MIFit installation into a new directory. This file contains a SMILES string for the ligand. Note that the Dictionary Editor also supports entry of ligand coordinate information in the form of MOL, PDB and mmCIF formats.
2. Select the MIFit command **Dictionary/Import Ligand/Smiles**. Make sure that the **File** option is toggled on and use the **Browse...** button to enter the name of the file containing the SMILES string. Enter a three character entity code (for example, 'LIG') for the ligand in the **ID Code** field. Select **Ok**.
3. You may click on the **Chirals** and **Planes** checkboxes to confirm that the interpretation of the ligand in terms of restraints was correct. You will find that there will be no chiral center restraints and one plane covering the aromatic ring and atoms immediately connected to it. Incorrect restraint assignments may be changed using the pull down menus at the top of the Dictionary Editor window.
4. To write out the refinement dictionary, Select the **Export to File...** button. Enter a file name, select **Refmac mmcif Dictionary File (\*.cif)** in the **Save as type** field, and click on **OK**. Many crystallographers prefer to ignore the relatively weak and often poorly defined torsion restraints from ligand refinement and select **No** in the resulting 'Write Torsions?' dialog box.
5. Select the **Save and Exit** button to exit from the Ligand Dictionary Editor. If you expand the Dictionary icon in the tree on the left of the MIFit interface you will see that entity LIG has been added to the list of entities (double-clicking on these entities re-opens the dictionary editor).
6. You may select **File/Exit** if you wish to shut down MIFit. When you exit from MIFit you will be asked if you wish to write the new ligand entry into the MIFit dictionary file for future use.

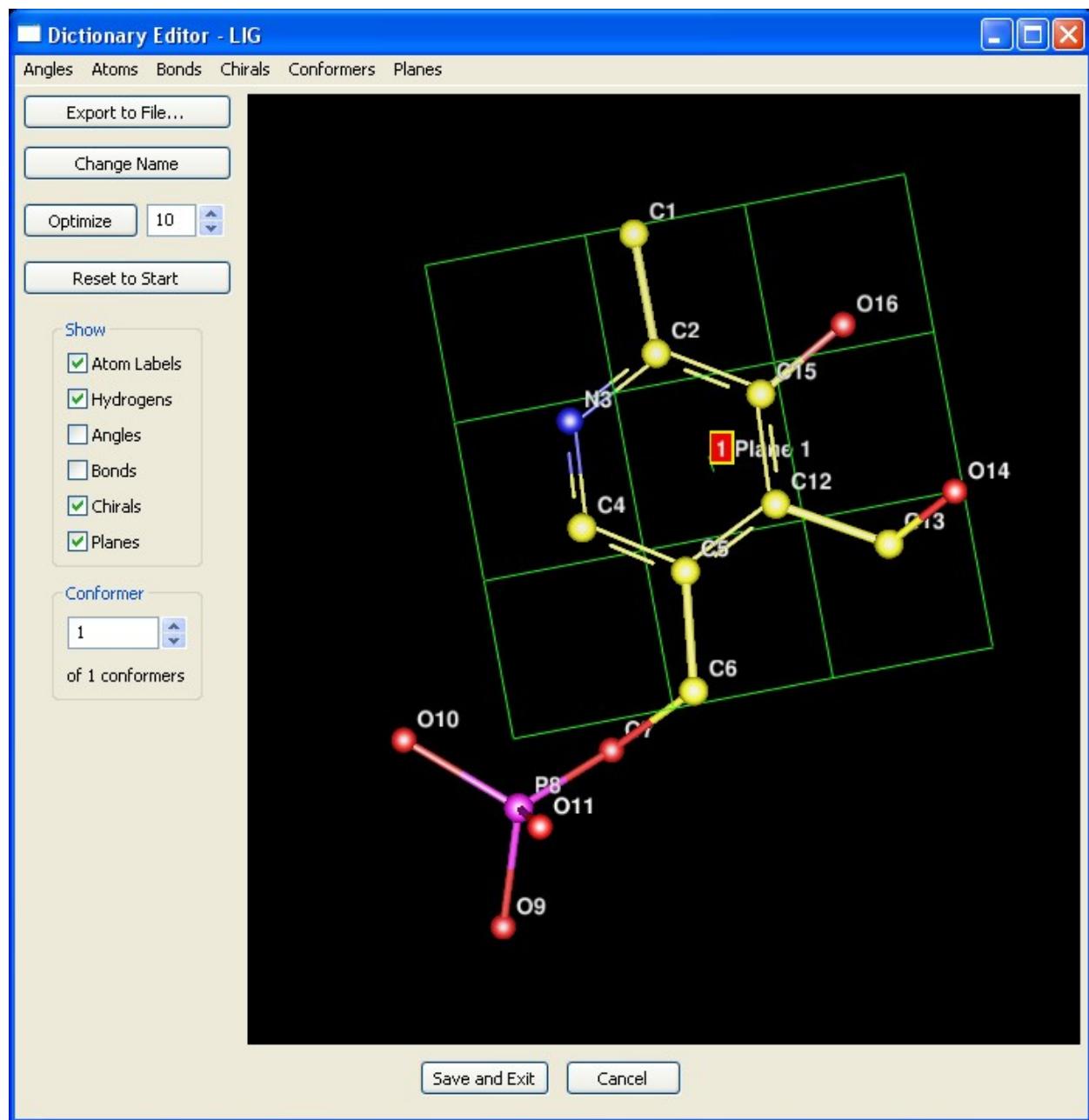
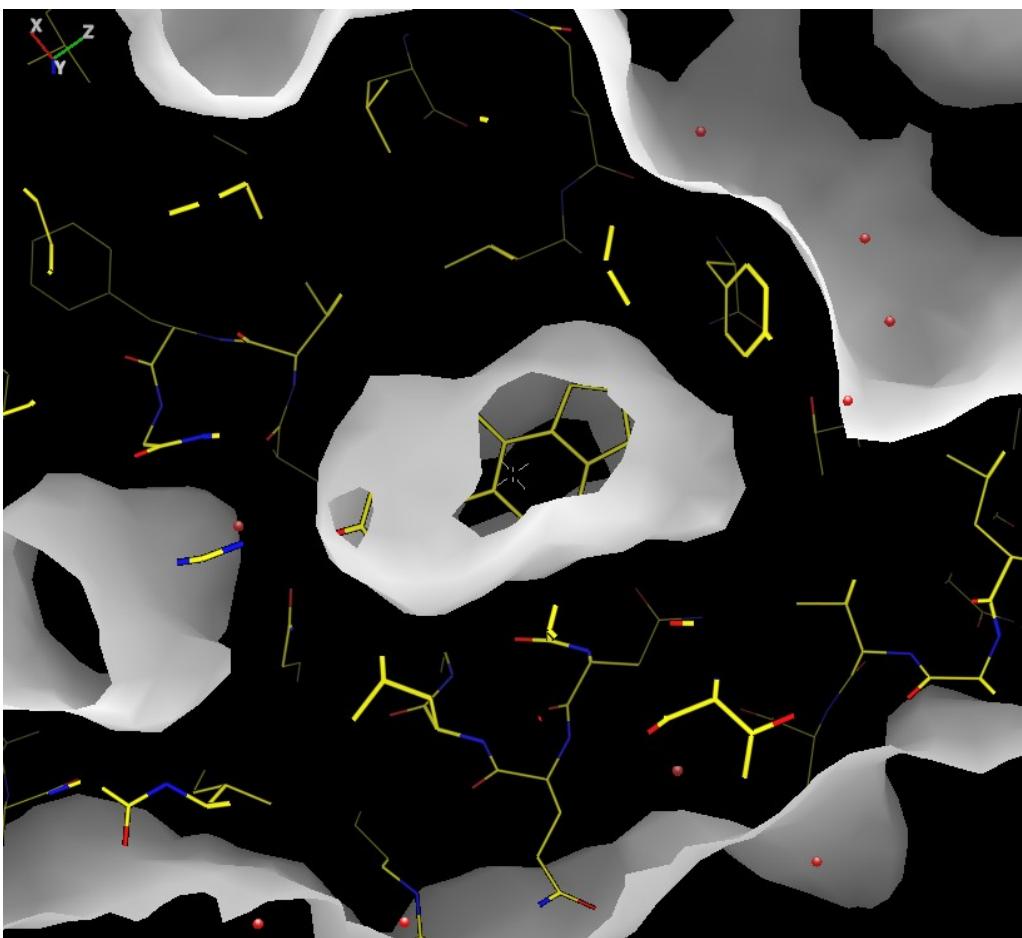


Figure 11.59 Lesson 6

## 11.7 Lesson 7: Displaying molecular surfaces

*Scenario: You have completed the refinement of a protein:ligand structure and you wish to examine the ligand binding volume and the protein surfaces.*

1. Copy file *1A28.pdb* from the *examples* directory of your MIFit installation into a new directory. This file contains a protein structure with a bound ligand (progesterone).
2. Use the MIFit command **File/Open models, data, maps, etc...** to load coordinate file *1A28.pdb* from the directory in which you placed it. The structure will appear in the main canvas.
3. Click on Segmt B.4 in the top pane of the navigation tree. Then double-click on entity STR 2 in the **Residues:** section of the tree. These actions will center the canvas display on the bound ligand STR 2.
4. Click on any atom in the main canvas.
5. Make sure that the **Show/Solid Surface/Molecular surface** mode option is checked and select the **Show/Solid Surface/Build Surface** command. After a few seconds you should see the solvent exposed surface including a completely enclosed cavity that contained the STR molecule, i.e. in this protein the bound ligand is not accessible from the outside.
6. You may wish to click on the zoom out icon  and the slab out icon  to obtain a more global view of the protein surface.
7. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.

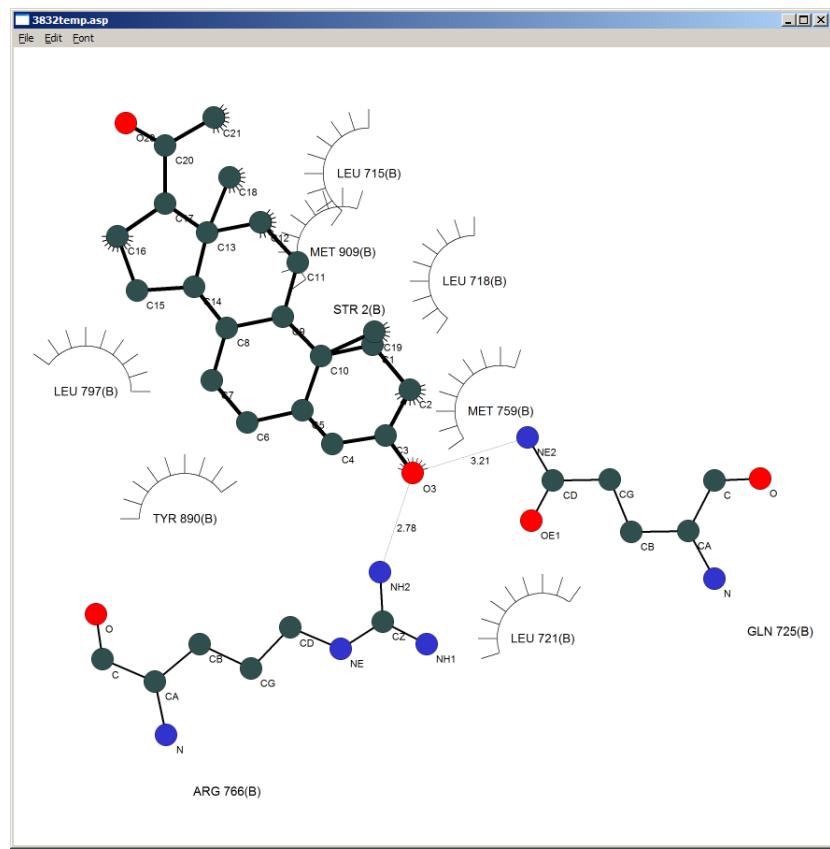


**Figure 11.60** Lesson 7

## 11.8 Lesson 8: Creating schematic plots of protein:ligand interactions

*Scenario: You have just completed the refinement of a protein structure containing a ligand molecule (progesterone) and wish to create a schematic (2D) plot of the protein-ligand interactions for a publication.*

1. Copy file *1A28.pdb* from the *examples* directory of your MIFit installation into a new directory. This file contains a protein structure with a bound progesterone molecule.
2. Use the MIFit command **File/Open models, data, maps, etc...** to load coordinate file *1A28.pdb* from the directory in which you placed it. The structure will appear in the main canvas.
3. Select the chain including STR 2 B from the navigation tree. Enter B2 in the **Go to residue** parameter box at the top of the navigation tree and enter. These actions will center the canvas display on the bound ligand STR 2. These actions will center the display on the bound ligand STR. Note that the orientation of the ligand in the main MIFit canvas will not affect the view point used for the schematic plot.
4. Click on any atom in the ligand in the main MIFit canvas.
5. Select the **File/Export Active Site Plot** command. The active site plot should appear. Note that by holding on the bottom-left corner of the plot with the mouse you can shrink the image display size to remove white space on the left.
6. You may select **File/Export Image** in the active site plot window if you wish to write out the image in one of the *png*, *tif* or *bmp* formats.
7. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



**Figure 11.61** Lesson 8

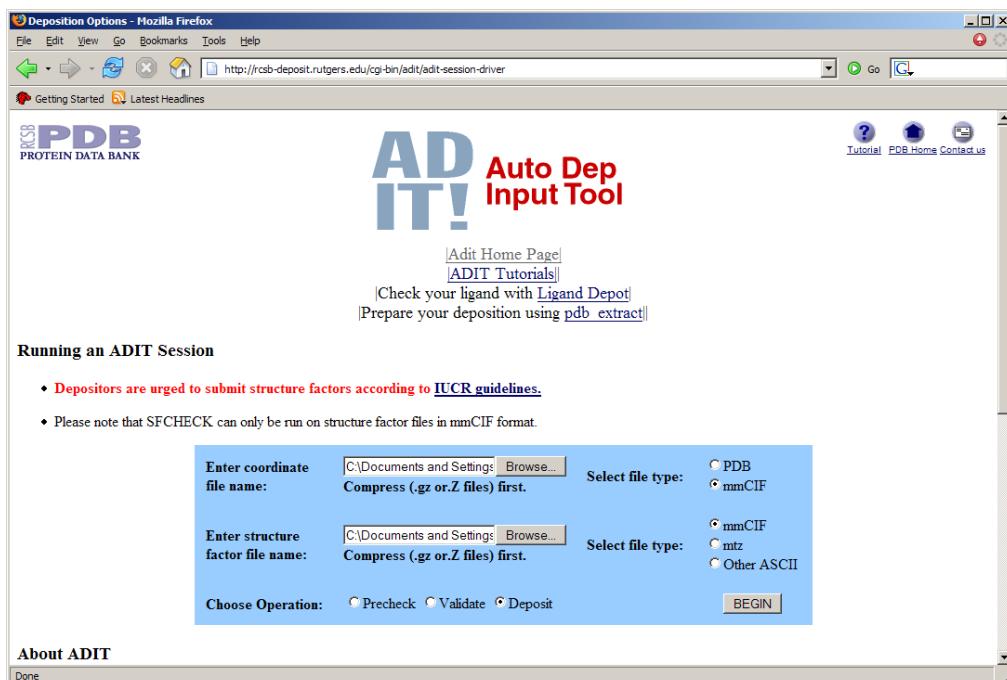
## 11.9 Lesson 9: Creating reports for PDB deposition and publication

(This tutorial requires that MIFit has access to CCP4.)

*Scenario: You have just completed the refinement of a new protein structure. You wish to assemble the structure determination statistics into a table for a structure report and deposit the structure to the Protein Data Bank.*

1. Copy the files *automation\_out.pdb*, *automation\_out.mtz*, *automation\_start.log* and *automation\_sequence.txt* from the *examples* directory of your MIFit installation into a new directory. Files *automation\_out.pdb* and *automation\_out.mtz* are the refined coordinates and refinement data respectively. File *automation\_start.log* is a data processing log containing data merging statistics from the d\*TREK program. File *automation\_sequence.seq* is the full protein sequence encoded by single character amino acids codes (i.e. the sequence of the protein in the crystal, regardless of whether the structure was sufficiently ordered to model).
2. Select the MIFit **Job/Report** command. Use the **Browse...** buttons to set the **Working Directory** parameter to the directory containing the structure data established in (1), the **Model (pdb)** parameter to the *automation\_out.pdb* file and the **Data (mtz)** parameter to the *automation\_out.mtz* file.
3. Select the **Sequence** checkbox and use the **Browse...** button to locate and add the file *automation\_sequence.seq*. To do this you will need to change the **File of type** parameter to ‘Sequence Files (\*.seq)’. Select the **Processing Log** checkbox and use the **Browse...** button to locate and add the file *automation\_start.log*. (If you had wished to include ‘non-electronic’ information at this point to the PDB deposition then the **Annotation** checkbox could have been selected and a completed annotation template file of the type found in *../examples/deposit3d.template* could have been included. Otherwise this information can be introduced into the RCSB PDB deposition interface. This option is mainly useful for avoiding repetitive input on related depositions.)
4. Select the **Write mmCIF Report** and the **Write HTML Report** options in order to generate a mmCIF report suitable for PDB deposition and an HTML report containing structure determination statistics. Select the **Write mmCIF Data** checkbox to generate a set of structure factor data in mmCIF format.
5. Select the **Run** command. On the Windows OS you will see a small text port appear while the reporting calculations execute. You will then see an HTML report appear in a browser.
6. Select **Jobs** tab near the top of the navigation tree. Then right-click on the bottom-most entry in the **Jobs List** and select **Show Log File**. This action will provide a view of summary information from the report generation process. Note that this command is grayed-out (inaccessible) until the job completes.
7. Look in the working directory. You will see file *pdbdeposit.htm*, containing the HTML report, file *pdbdeposit.cif*, a mmCIF file suitable for deposition to the PDB and an mmCIF x-ray data file *pdbdeposit\_hkl.cif* to provide to the PDB.

- The file *pdbdeposit.cif* potentially contains all information necessary for a structure deposition using the RCSB PDB ADIT interface at <http://rcsb-deposit.rutgers.edu/adit/>. From the ADIT interface you can pre-check the deposition file using the ‘Precheck’ option. Previewing the deposited entry using the ‘Preview Entry’ option after applying the ‘Deposit’ option allows for correction and inclusion of any missing data items.
- You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



**Figure 11.62** Lesson 9, RCSB PDB ADIT

**Structure report - Mozilla Firefox**

File Edit View Go Bookmarks Tools Help

file:///C:/Documents%20and%20Settings/John

Getting Started Latest Headlines

## Structure determination and model refinement statistics

**Crystal characteristics and data collection statistics (outer shell statistics in parenthesis)**

Unit cell ( $\text{\AA}$ , $^\circ$ )	65.419 89.227 99.842 90.000 90.000 90.000
Space group	I222
Resolution range ( $\text{\AA}$ )	34.58 - 1.84 (1.91 - 1.84)
No. of observations	137717
No. of unique reflections	22847
Redundancy	6.03 (2.36)
Completeness (%)	88.8 (43.1)
Mean $I/\sigma(I)$	27.8 (4.7)
$R_{\text{merge}}$	0.038 (0.174)

**Crystallographic data and refinement statistics**

Resolution range ( $\text{\AA}$ )	32.71 - 1.84
No. of reflections	22844 (21674 working set, 1170 test set)
No. of protein chains	1 (X)
Ligand id codes	-
No. of protein residues	207
No. of ligands	0
No. of waters	153
$R_{\text{work}}$	0.2105
$R_{\text{free}}$	0.2452
Rmsd bond lengths ( $\text{\AA}$ )	0.008
Rmsd bond angles ( $^\circ$ )	1.063
Number of disallowed $\phi\psi$ angles	0

Done

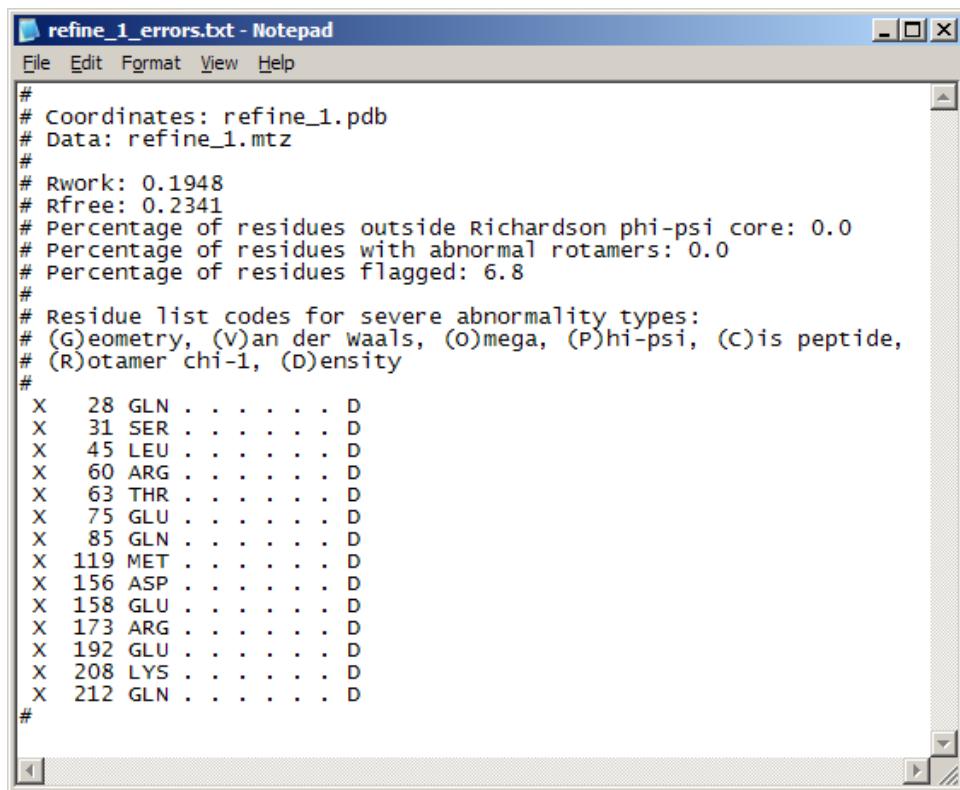
**Figure 11.63** Lesson 9, report

## 11.10 Lesson 10: Automated refinement

(This tutorial requires that MIFit has access to CCP4.)

*Scenario: You have a reasonably good protein model and wish to continue to finalize the structure with further refinement including water-picking and automated error detection.*

1. Copy the files *automation\_out.pdb* and *automation\_out.mtz* from the *examples* directory of your MIFit installation into a new directory. File *automation\_out.pdb* is the model that will be refined. File *automation\_out.mtz* is the refinement data in MTZ format.
2. Select the **Job/Refinement** command from the MIFit interface. In the resulting interface use the **Browse...** button to select the directory defined in (1) as the **Working directory**. Use the associated **Browse...** buttons to load *automation\_out.pdb* as **Model (pdb)** and *automation\_out.mtz* as the **Data (mtz)** parameter.
3. Leave the **Refinement Type** parameter as Refmac5 and the **B-Factor Treatment** as Isotropic.
4. The **Weight** parameter controls the relative contribution of the data and the stereochemical restraints to the refinement process, with a lower value indicating a lower contribution from the data. This is a relatively high resolution structure (which allows larger weights) but is also near the end of the refinement (for which smaller weights are used) so leave this as 0.1.
5. We are going to include water picking in this refinement so leave the **Number of cycles** parameter as 5. Select **Water Picking Cycles** and set the value to 2. With these settings, three refinement runs will be employed with water picking between runs (i.e. refinement/water-pick/refinement/water-pick/refinement).
6. Leave the **Optional Items** alone and click on **Run**. On the Windows OS you should see the text port appear to indicate that the structure solution job is running. Upon completion a browser will pop-up containing a summary of possible errors in the structure.
7. In a real-life application the next step would be to load the output coordinates (*refine\_1.pdb*) and phased structure factor data (*refine\_1.mtz*). A MIFit session file (*refine\_1.mlw*, in the working directory) is available to accelerate that step and may be accessed with the **File/Open models, data, maps, etc...** command. See lesson 2). Although metrics for identifying structure misfits are somewhat fuzzy it is interesting to look through the amino acids in the associated error list. To import the error list as an annotation select the **Display** tab, right click on **Annotations** and apply the **Import from error list** command. This action can be used to load the error list into the **Annotations** tree and the main canvas.
8. If you like, select **Jobs** tab near the top of the navigation tree. Then right-click on the bottommost entry in the **Jobs List** and select **Show Log File**. This action will provide a view of summary information from the refinement process. Note that this command is inaccessible until the job completes.
9. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



The screenshot shows a Windows Notepad window titled "refine\_1\_errors.txt - Notepad". The menu bar includes File, Edit, Format, View, and Help. The content of the text file is as follows:

```
#  
# Coordinates: refine_1.pdb  
# Data: refine_1.mtz  
#  
# Rwork: 0.1948  
# Rfree: 0.2341  
# Percentage of residues outside Richardson phi-psi core: 0.0  
# Percentage of residues with abnormal rotamers: 0.0  
# Percentage of residues flagged: 6.8  
#  
# Residue list codes for severe abnormality types:  
# (G)eometry, (V)an der waals, (O)mega, (P)hi-psi, (C)is peptide,  
# (R)otamer chi-1, (D)ensity  
#  
X 28 GLN . . . . . D  
X 31 SER . . . . . D  
X 45 LEU . . . . . D  
X 60 ARG . . . . . D  
X 63 THR . . . . . D  
X 75 GLU . . . . . D  
X 85 GLN . . . . . D  
X 119 MET . . . . . D  
X 156 ASP . . . . . D  
X 158 GLU . . . . . D  
X 173 ARG . . . . . D  
X 192 GLU . . . . . D  
X 208 LYS . . . . . D  
X 212 GLN . . . . . D
```

Figure 11.64 Lesson 10

## 11.11 Lesson 11: Automated co-crystal structure determination

(This tutorial requires that MIFit has access to CCP4.)

*Scenario: You are collecting many data sets on related protein:ligand complexes as part of ligand discovery and optimization project. You wish to obtain pre-refined structures of these as rapidly and as in convenient way as possible.*

1. Copy the files *automation\_start.pdb* and *automation\_start.ref* from the *examples* directory of your MIFit installation into a new directory. File *automation\_start.pdb* is the initial search model that will be used for the structure determination – it is a well refined example of this protein structure. File *automation\_start.ref* is a file of intensity data (in this case, processed using the d\*TREK program).
2. Select **Job/Cocrystal Solution** from the MIFit interface. You will see the ‘Cocrystal Solution’ menu appear.
3. Click on the **Add** button in the lower left of the menu. Use the **Look in** browser to locate and load the file of intensity data, *automation\_start.ref*. Note that data merged using any of d\*TREK, SCALEPACK or CCP4/SCALA may be used by this process. If you had multiple data sets corresponding to the same type of crystal then multiple co-crystal solution jobs could be run by adding them all into this selection.
4. Use the **Browse...** button to load the model *automation\_start.pdb* as the **Model (pdb)** parameter.
5. An option that is useful when multiple data sets are loaded into the cocrystal solution menu is **HTML Summary**. This provides a summary table of all the structure determination jobs. To see what this looks like, select the **HTML Summary** checkbox and use the associated **Browse...** button to determine the directory in which this file will be deposited - select the directory established in (1) for the data.
6. A useful option is to set the position on which the display will be centered in the resulting MIFit session file. To do this, check the **View Point** selection, make sure that **Points** is selected and enter values 61, 33 and 26 as **X**, **Y** and **Z** parameters. These numbers are the position of the target site on which to center the display in angstroms. (If a previous session file were available that specified this view it could have been entered through the **Session** parameter). Besides setting the viewpoint, this option will also remove waters from the target site area (potentially occupied by a ligand) in order to provide more useable difference map for ligand fitting.
7. Select **Run**. On the Windows OS you should see the text port appear to indicate that the structure solution job is running. The job runs molecular replacement and several cycles of refinement and water-picking. On a 2GHz laptop the job takes about 3.5 minutes. Upon completion a browser will pop-up containing an HTML summary of the job.
8. Select the **Jobs** tab near the top of the navigation tree. Then right-click on the bottom-most entry in the **Jobs List** and select **Show Log File**. This action will provide a view

of summary information from the structure solution process. Note that this command is grayed-out (inaccessible) until the job completes.

9. Use the **File/Open models, data, maps, etc...** command to load the session file *bng\_milaunch.mlw*. (The refinement process used in this job also creates a session file on each run, so you will also see other session files). The refined model from this run will be loaded as well as the ‘standard’ likelihood-weighted electron density map and a likelihood weighted difference map. Default MIFit contour levels and colors are used for both. You may wish to right-click on the map icons in the tree on the left and select **Show/Hide** to hide one or other of these maps. Looking at these maps you will conclude that there are several ordered water molecules but no bound ligand. This process allowed you to evaluate this data set with minimum of work.
10. Look at the directory specified in (1). Files specified by root *refine\_3* correspond to the last refinement run. File *refine\_3\_errors.txt* lists amino acids that may be in error (as indicated by abnormal geometries or a significant degree of mismatch to the electron density). Note that the job history and this error list are linked to the HTML summary file.
11. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.

BNG Run Summary

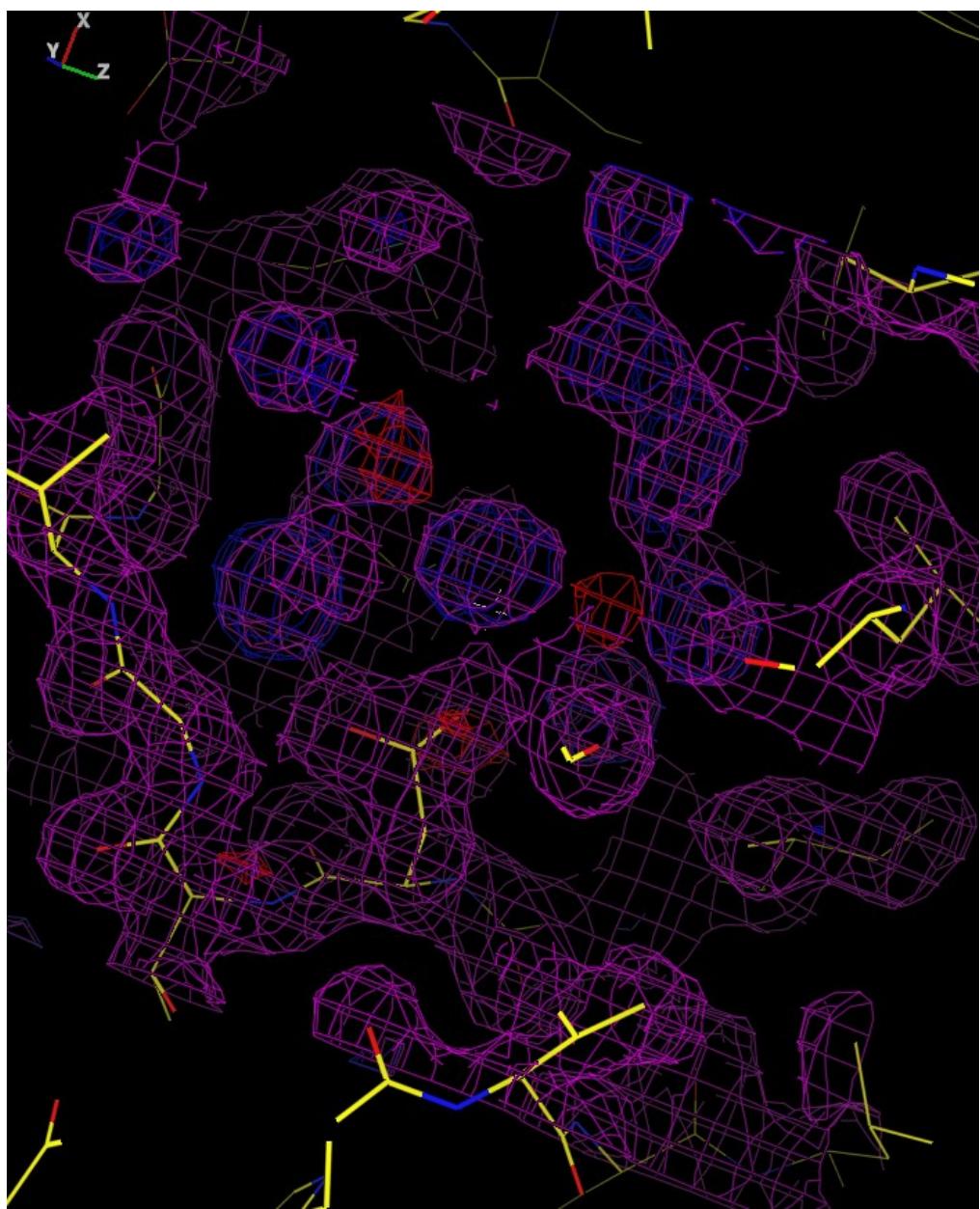
Job start: Fri May 26 10:39:42 2006

History	Res. (Å)	MR CC	Rwork	Rfree	Error List	Working Directory
1	1.840	0.682	0.2367	0.2664	1	C:\Documents and Settings\John Badger\My Documents\Tutorials\tutorial 11

Job End: Fri May 26 10:44:27 2006

Done

**Figure 11.65** Lesson 11, HTML summary



**Figure 11.66** Lesson 11, main view

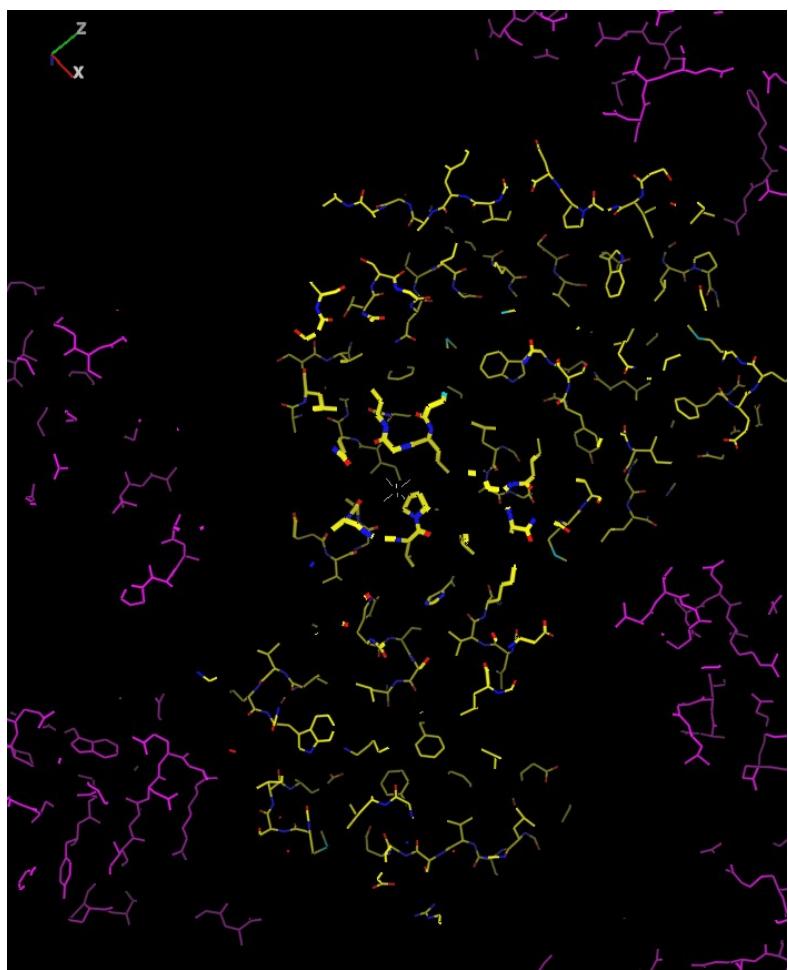
## 11.12 Lesson 12: Automated Molecular Replacement

(This tutorial requires that MIFit has access to CCP4.)

*Scenario: You have just obtained data on a new crystal form of a previously solved structure. In contrast to the previously solved structure, which contained a monomer in the crystal asymmetric unit, the new crystal form may contain a dimer in the crystal asymmetric unit.*

1. Copy the files *1A28\_mr.pdb* and *r1a28sf.mtz* from the *examples* directory of your MIFit installation into a new directory. File *1A28\_mr.pdb* is the initial search model (a monomer) that will be used for the structure determination. File *r1a28sf.mtz* contains structure factor data for the new crystal form.
2. Select **Job/Molecular Replacement** from the MIFit interface. You will see the ‘Molecular Replacement’ menu appear.
3. Use the **Browse...** button associated with the **Working directory** parameter to find and load the directory defined in (1).
4. Use the **Browse...** button associated with the **Model (pdb)** parameter to load file *1A28\_mr.pdb*. Use the **Browse...** button associated with the **Data (mtz)** parameter to load file *r1a28sf.mtz*.
5. The options to input a **Fixed model**, **Search multiple models** and **Match input position** should remain unchecked. The **Fixed model** parameter would be used if a molecular replacement search had been able to establish one component of the crystal structure (perhaps a single molecule from a complex or a domain of a large flexible structure). The **Search multiple models** parameter (a directory containing many MR search models) can be used in difficult molecular replacement cases to conveniently test a number of different models. The **Match input position** is useful when MR is applied to structures with same space group and cell as previously solved structures (i.e. co-crystals). It will adjust the MR solution by symmetry operations to match the initial search model.
6. The **Spacegroup** option may remain set to **Default** to use the space group specified by the diffraction data file header. The **Method** option may be set to **MolRep** to use the CCP4/MOLREP program to perform the MR.
7. Select **Run**. On the Windows OS you should see the text port appear to indicate that the molecular replacement job is running. The text port disappears when the job completes. This process takes about 1 minute on a 2GHz laptop.
8. Select the **Jobs** tab near the top of the navigation tree. Then right-click on the bottom-most entry in the **Jobs List** and select **Show Log File**. This action will provide a view of summary information from the molecular replacement process. Note that this command is grayed-out (inaccessible) until the job completes.
9. Use the **File/Open models, data, maps, etc...** command to load the resulting model *molrep\_1.pdb* from the directory defined in (1) into MIFit. You will notice that two molecules were found by the molecular replacement process (indicated by and A and

- a B chain in the top pane of the navigation tree) to fit the diffraction data and that these were automatically packed together.
10. Look at the files in the directory defined in (1). The *project\_history.txt* file logs the automation process. File *molrep\_1.log* is the log file from the molecular replacement run.
  11. Although the default behavior of most molecular replacement programs is to exclude solutions in which molecules grossly overlap in the crystal it is always useful to check the packing of the solution.
  12. Select **Show/Symmetry Atoms/Show symmetry atoms as atoms** to generate all symmetry related atoms in the volume displayed in the main MIFit canvas. You may wish to zoom out or move around the model and regenerate symmetry related positions to get a full sense of the crystal arrangement.
  13. Since this appears to be a correct solution ( $R < 0.35$  with good packing) you may also wish to view this model in the context of an electron density map. A convenient way to do this is to perform a quick refinement of the current structure and then load the resulting refined likelihood weighted electron density map. The **Job/Refinement** command is described in the next tutorial lesson.
  14. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



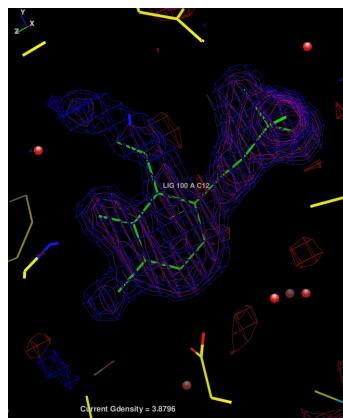
**Figure 11.67** Lesson 12

## 11.13 Lesson 13: Automated ligand density docking

*Scenario:* You have an almost completely refined structure and now wish to add and fit a ligand to residual density within the target site.

1. Copy the files *ligand\_example.pdb*, *ligand\_example.mtz* and *ligand\_example.smi* from the *examples* directory of your MIFit installation into a new directory. File *ligand.pdb* is the parent protein model, file *ligand\_example.mtz* contains data from the last cycles of refinement in MTZ format and file *ligand\_example.smi* contains a SMILES string corresponding to the ligand that is to be fit.
2. To establish the ligand parameters select the **Dictionary/Import Ligand/Smiles** command. With the **File** option toggled on, use the **Browse...** button to select file *ligand\_example.smi* from the dictionary defined in (1). Enter a three character code (for example, LIG) as the **ID Code (3 letters)** parameter. Click on **OK**. You should see the Ligand Dictionary Editor appear.
3. If you like, you can confirm that the ligand restraints were correctly defined (see Chapter 6). Click on **Save and Exit** to load the ligand into the dictionary.
4. Next, load the protein model and map using the **File/Open models, data, maps, etc...** dialog box to locate file *ligand\_example.pdb* and *ligand\_example.mtz* from (1). In the Phase file import dialog, uncheck the first map settings since this example only requires the difference map using the DELFWT and PHDELFWT columns.
5. Use the left (rotate) and right (translate) mouse buttons to make sure that the large density feature near residue A 163 is centered on the cross hairs.
6. To add the ligand to the map select the **Model/Add residue** command. Use the **Residue Type** pulldown to select the LIG entity. Select ‘End of Model’ for the **Insert Position** parameter. Make sure that ‘Screen Center’ is toggled on for the **Put At** parameter. You may enter a new chain-id ‘W’ in the Chain-id field. Select **OK** to add the ligand to the model.
7. Click on any atom in the ligand and select **Fit/Fit Residue**. Alternatively you may use the keyboard shortcut **f**. You should see the ligand turn green, indicating that it is active.
8. Select the **Refine/Find Ligand Fit and Conformer** command to automatically fit the ligand. The ligand docking search will continue until the maximum number of trials is reached or a sufficiently good fit is established. At the completion of this process a good fit of the ligand will usually be obtained. Since this search starts from a randomized starting point, fully reproducible results are not obtained on every run and occasionally the search will result in a plausible misfit.
9. You may sometimes be able to optimize the fit by selecting the **Refine/Rigid-Body Refine Current Atoms** option. If just one or two torsion groups are incorrect then they may be fixed by clicking on the both defining the torsion group near the moveable atoms and rotating around that bond with the right mouse button.

10. Once you have achieved a satisfactory fit you may accept the fit by clicking on the  icon from the toolbar.
11. Ligand fits may also be optimized using real-space refinement. Click an atom on the ligand and then select **Refine/Refine Residue**. The ligand color will change to blue to indicate that it is active for refinement. Hold down the space bar to refine the ligand to convergence. Select **Refine/Accept Refine** to accept the optimized solution.
12. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



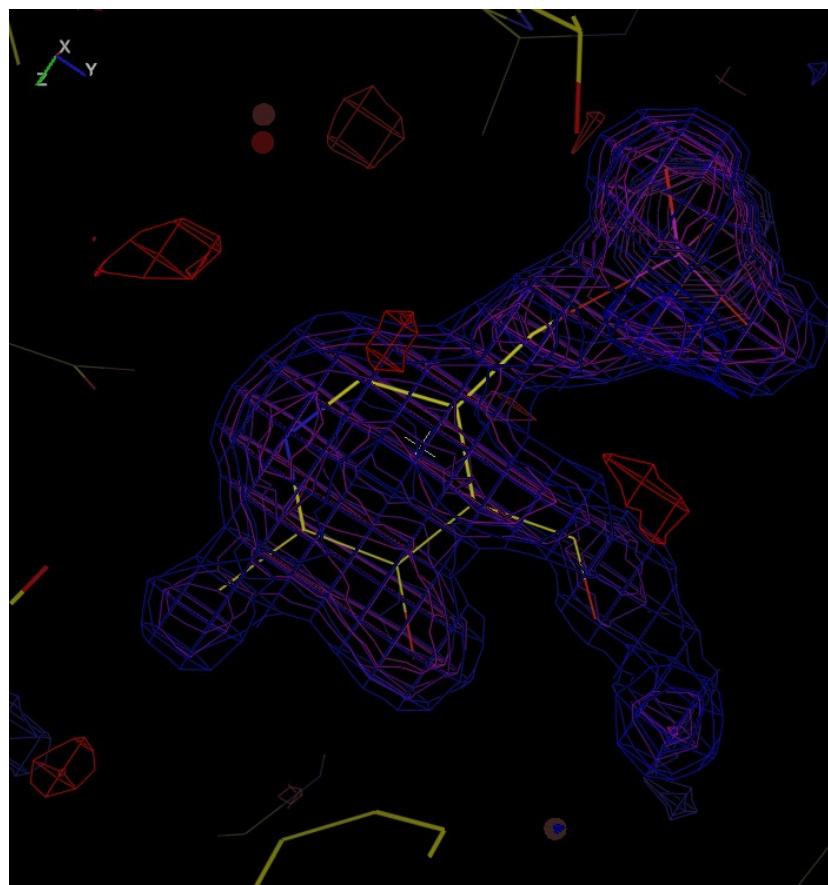
**Figure 11.68** Lesson 13

## 11.14 Lesson 14: Automated ligand density docking with OpenEye software

(This tutorial requires that MIFit has access to CCP4 and the OpenEye ligand density fitting software.)

*Scenario: You have an almost completely refined structure and now wish to add and fit a ligand to residual density within the target site.*

1. Copy the files *ligand\_example.pdb*, *ligand\_example.mtz* and *ligand\_example.smi* from the *examples* directory of your MIFit installation into a new directory. File *ligand.pdb* is the parent protein model, file *ligand\_example.mtz* contains data from the last cycles of refinement in MTZ format and file *ligand\_example.smi* contains a SMILES string corresponding to the ligand that is to be fit.
2. Next, load the protein model using the **File/Open models, data, maps, etc...** dialog box to locate file *ligand\_example.pdb* and *ligand\_example.mtz* from (1). In the Phase file import dialog, uncheck the first map settings since this example only requires the difference map using the DELFWT and PHDELFWT columns.
3. Use the left (rotate) and right (translate) mouse buttons to make sure that the large density feature near residue A 163 is centered on the cross hairs.
4. Select the **Job/Run OpenEye Ligand Fitting** command. In the resulting dialog box, select the **Browse...** button associated with the **Ligand** field. Change the filter from files of type '.pdb' to files of type '.smi' (for a SMILES string) and load the file *ligand\_example.smi*. Select **OK** in the OpenEye interface.
5. The OpenEye ligand fitting software should run for several seconds and then a model containing the correctly fitted ligand should appear in the main canvas. A dialog that controls options for re-centering the view should appear, select **No** in this dialog.
6. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



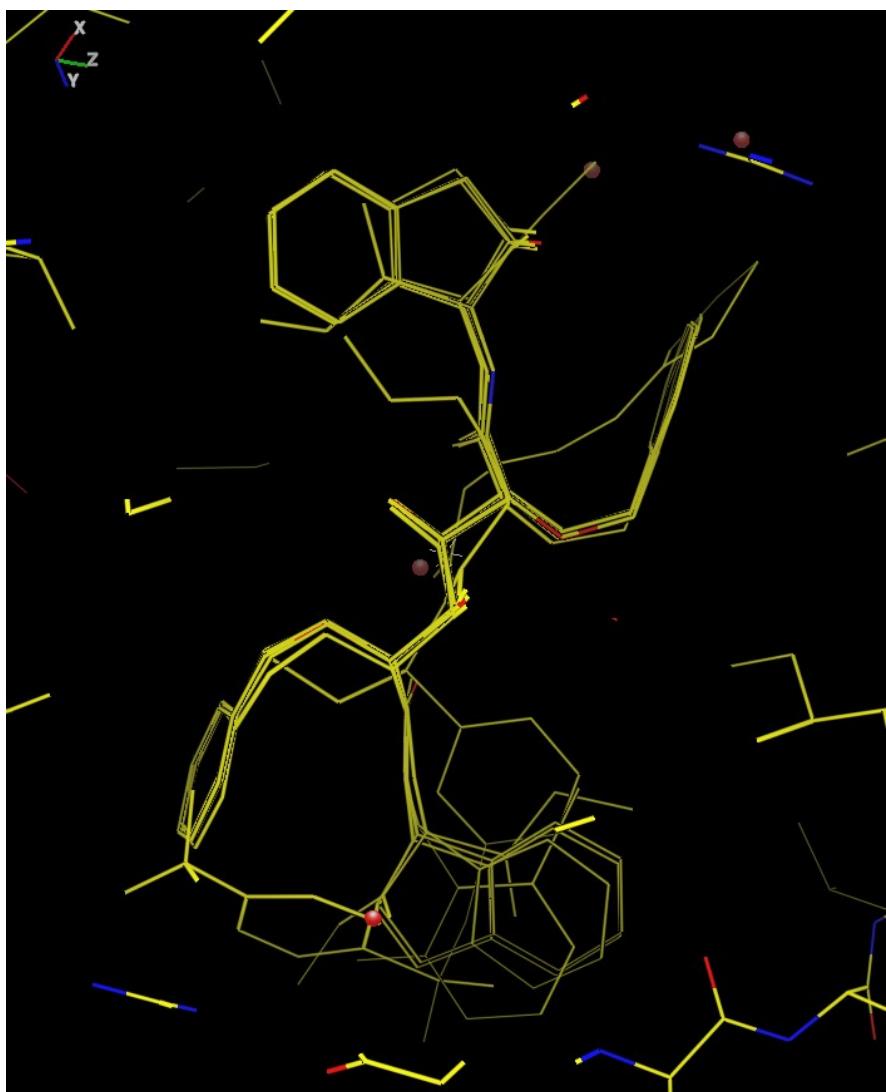
**Figure 11.69** Lesson 14

## 11.15 Lesson 15: Superimposing ligands from multiple related structures

(This tutorial requires that MIFit has access to CCP4.)

*Scenario: You have solved the structures of six HIV protease:ligand complexes and wish to display and make an accurate comparison of the binding positions of the ligands in each of them.*

1. Copy files *1d4h.pdb*, *1d4i.pdb*, *1d4y.pdb*, *1eby.pdb*, *1ebw.pdb*, *1d4j.pdb* from the *examples* directory of your MIFit installation into a new directory. Each of these files contains the HIV protein with a different ligand. This step is required because the superposition application will superimpose all PDB files from within a specified directory.
2. Create a working directory (different from the directory established in (1)) into which files containing the superimposed ligands will be placed.
3. Use the MIFit command **File/Open models, data, maps, etc...** to load coordinate file *1d4h.pdb* from the directory in which you placed it. The structure will appear in the main canvas.
4. In the top pane of the navigation tree select segment *\_3*. Double-click on entity BEH 501 in the residues list. These actions will center the display on the bound ligand BEH in this reference structure.
5. Select the Job/Cocrystal Superposition menu.
6. Set **Working Directory** to the name of the directory specified in (2) above. Set **Structure Directory** to the name of the directory specified in (1), in which you stored the set of six coordinate files. Set **Target File** to structure *1d4h.pdb*. All of the above selections may be made using the **Browse...** buttons. Set the **X**, **Y** and **Z** parameters in the **Target Coordinates** field to a point 14, 23, 5. These are coordinates in angstroms that are close to atom C01, a central atom in the ligand that is used to define the binding site volume for the structure superposition.
7. Select **OK**. You will see a small text port appear for a few seconds while the superposition process executes. You will then see the set of superimposed ligands appear in the main canvas. A dialog box for recentering the view will appear; select **Yes** in this box. The coordinate file containing these ligands (*allligands.pdb*, in the working directory) will show as a second model in the navigation tree on the left. If you click on the model icon for this file you will see that the superimposed ligands in this file are identified by separate chains.
8. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.

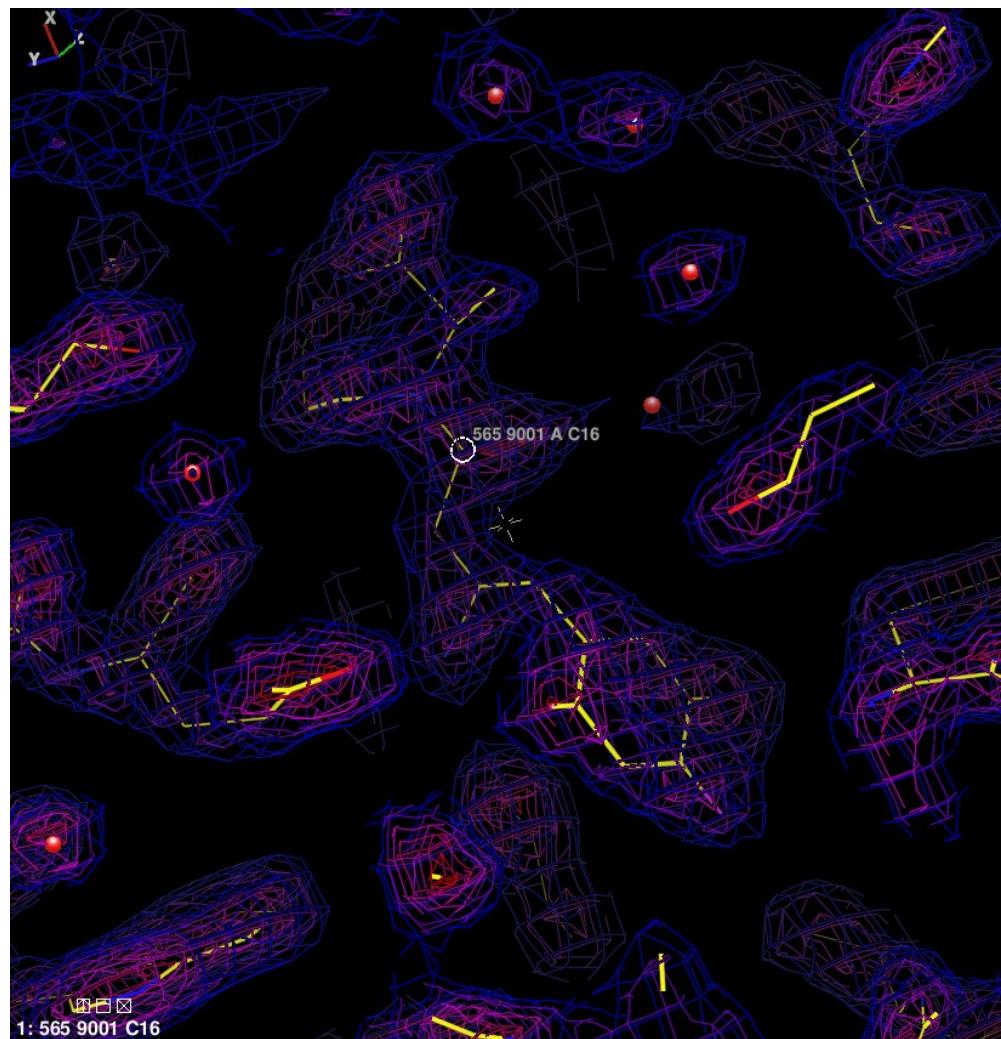


**Figure 11.70** Lesson 15

## 11.16 Lesson 16: Model and map display using files from the wwPDB

*Scenario: You are working with computational chemists analyzing a structure from the PDB and wish to assess the reliability of the model against the associated electron density map.*

1. Copy files *2IIV.pdb* and *2iiv-sf.cif* from the *examples* directory of your MIFit installation into a new directory. These are the coordinate and diffraction data files that can be downloaded from the PDB web sites. The diffraction data file is in mmCIF format.
2. Use the MIFit command **File/Open models, data, maps, etc...** to load coordinate file *2IIV.pdb* from the directory in which you placed it. The structure will appear in the main canvas. You will notice that the crystal data (displayed next to the crystal icon beneath the coordinate file, at the top of the navigation tree) is correct.
3. Use the command **File/Open models, data, maps, etc...** to load data file *2iiv-sf.cif*. Although this file did not contain any space group or cell information you will notice that the crystal data displayed next to the corresponding crystal icon is correctly copied from the coordinate file. In the Phase file import dialog, the Fc and Phase will show Model 1 values indicating that the structure factors and phases will be computed from the model. After clicking OK, some diagnostic information on this calculation will appear in the **Log** pane in the lower left. You will see that the R-factor in this calculation is 0.264, which is a reasonable value. Note that reported R-factors are a function of a number of factors including scaling methods, bulk solvent correction and data limits and rarely correspond exactly to the results of a recalculation.
4. The **Fast Fourier Transform Map** dialog box will appear. Change the **Map type** option to '2Fo-Fc' and select **OK**. You should see a map appear in the main canvas.
5. Using the map's **Contour Options** (accessed by right-clicking on the map in the Models tree), move the **Radius** slider to about 10 in order to set a 10Å box for the map volume. You may wish to set the **Preset Map Styles** option to 'Blue Map, 1,2,3,4,5 sigma' and then uncheck the display of contour levels 2 and 4 to remove some of the contour lines from the map display. Select **OK** and you will see the map view change to these settings. You may wish to use the slider to slightly increase the lowest density value (say, from 50 to 60).
6. In the top pane of the navigation enter 'A 9001' in the **Go to residue** text box at the top of the tree control and click on the **GoTo** button. The canvas view should recenter on the ligand, entity 565. In this case the ligand is very well defined and makes a good fit to the map.
7. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.



**Figure 11.71** Lesson 16

# Appendix A: Scripting Language

MIFit may be started with a session or script file as a command line parameter. These files are recognized by the extension .mlw. This capability is mostly used to allow other applications control MIFit. For example, the automated structure solution application in MIExpert has an option to launch MIFit after structure solution calculations are completed. However, a user may also easily create a script using a text editor.

The MIFit script commands are case insensitive and each command should be given on a separate line. Command key words are in **bold** and arguments are in *italics*. All commands are of the form. Some commands require a prerequisite, which will be noted. The character # marks the remainder of a line as comments.

**command** *val1 val2*

# - comment

## A.1 Model Commands

**loadpdb** *n pathname*

Load the PDB coordinate file in pathname and number it *n*. Subsequent commands can switch between multiple maps by specifying *n*.

Prerequisites: none

**atomlabel** *m resid atomid chained*

Label an atom. The atom is labeled with the default label as if it was picked in interactive mode.

*m* — model number starting from 1. Use 1 if there is just one model.

*resid* — the name of the residue, e.g. 101.

*atomid* — the name of the atom, e.g. CA.

*chainid* — the chain identifier, e.g. A. If *chainid* is a space use a “\*”

Example: atomlabel 1 100 CA \*

Prerequisites: at least one model loaded.

## A.2 View Commands

**rotx** *val*

Rotate about the x world axis (horizontal) *val* degrees.

Prerequisites: an open window

**roty** *val*

Rotate about the y world axis (vertical) *val* degrees

Prerequisites: an open window

**rotz** *val*

Rotate about the z world axis (out of screen) *val* degrees

Prerequisites: an open window

**backclip** *val*

Set the back clipping plane to *val*.

Prerequisites: an open window

**frontclip** *val*

Set the front clipping plane to *val*.

Prerequisites: an open window

**zoom** *val*

Set the zoom level to *val*.

Prerequisites: an open window

**perspective** *val*

Set the perspective level to *val*. 0 is no perspective, or orthonormal display.

Prerequisites: an open window

**stereo** *on|off* *v1* *v2*

Set the side-by-side stereo to *on* or *off*. *v1* is the separation in pixels of the two image centers, and *v2* is the stereo\_angle is the angle in degrees (between 3 and 8 degrees are useful values. Use a positive value for wall-eye and a negative value for cross-eye stereo). *v1* and *v2* are not needed for the *off* command.

Prerequisites: an open window

**translation** *x* *y* *z*

Set the center to the coordinate (*x,y,z*) in Ångstroms.

Prerequisites: an open window

**rotation** *v11* *v12* *v13* *v21* *v22* *v23* *v31* *v32* *v33*

Set the rotation to the 3x3 matrix specified in *v11*...*v33*.

Example: `rotation 1 0 0 0 1 0 0 0 1`

Prerequisites: an open window

## A.3 Map Commands

**loadmap** *n* *pathname*

Load the map in *pathname* and number it *n*. Subsequent commands can switch between multiple maps by specifying *n*, specifically with the command `maptocont n`.

Prerequisites: none

**loadmapphase** *n* *pathname*

Load the phases in the *pathname* and number it *n*. Subsequent commands can switch between multiple maps by specifying *n*, specifically with the command **maptocont** *n*. A density map will be FFT'd from the phases directly after loading. Note that, at present, data in the .phs file format are not properly loaded from session files.

Prerequisites: none

**color** *c*

Set the current color to the *c*. Subsequent coloring commands will use this color. For color numbers see Appendix 2: Colors

Prerequisites: none

**maptocont** *n*

Set the map to be contoured. Subsequent contouring commands will refer to the map loaded with this number (see **loadmap** and **loadmapphase**).

Prerequisites: a loaded map with the same number

**contourmap** *n*

Contour the map number *n* with the current values.

Prerequisites: other contouring commands and a loaded map at *n*

**maplinewidth** *v*

The line width of map lines in pixels. Changes only the current map.

Prerequisites: loaded map

**contourcolor** *v*

Set the color of contour level *v* to **color** *c* (see **color**).

Prerequisites: a loaded map

**contourleveldefault** *l1 l2 l3 l4 l5*

Set the map contour level values. Must specify 5 levels. Note: **contourlevels** command turns individual levels on and off.

Example: **contourleveldefault** 50 100 150 200 250

Prerequisites: a loaded map

**contourlevels** *d1 d2 d3 d4 d5*

Turn on and off individual contour levels. A 0 value turns off a level and a non-zero turns them on.

Example: **contourlevels** 1 0 0 0 1

Turns on only the first and fifth levels.

Prerequisites: loaded map

**contourradius** *v*

Set the radius (half-width) of the area to be contoured to *v*.

Prerequisites: a loaded map

### **fftapply**

Forces an FFT call to calculate density values. Used after changing the map dimensions with the commands **coefficients**, **fftnx**, **fftny** and/or **fftnz**.

Prerequisites: a loaded map with phases

### **coefficient s**

Set the map coefficients for the FFT to one of "Fo","Fc","2Fo-Fc","Fo-Fc","Fo\*fom","3Fo-2Fc","5Fo-3Fc","2mFo-DFc" or "Fo-DFc". Case is not important.

Prerequisites: a loaded map with phases

### **resmin dmin**

Set the minimum resolution (outer resolution) of the map to *dmin*.

Prerequisites: a loaded map with phases

### **resmax dmax**

Set the maximum resolution (outer resolution) of the map to *dmax*.

Prerequisites: a loaded map with phases

### **unitcell a b c alpha beta gamma**

Set the unit cell to *a b c alpha beta gamma*.

Prerequisites: a loaded map

### **spacegroupno n**

Set the spacegroup number to *n*. (See International Tables for spacegroup numbering. This same information can also be found in the file \$MIFITHOME/data/symlib.)

Example: **spacegroupno 19**

Sets the spacegroup to P212121.

Prerequisites: a loaded map with phases

### **crystal name**

Load the crystal name.

Prerequisites: Crystal must be previously specified with File/Crystal with MIFit (or, for experts, be located in the *crystal\_info* directory or the current working directory).

### **name name**

Name the map *name* identification to the user.

Prerequisites: a loaded map

### **fftnx n**

Set the number of intervals in the x direction to *n*.

Note if the number is not a valid value (i.e. factorable by 2,3,5,7, it will be set to the next highest valid value).

Prerequisites: a loaded map with phases

**fftny** *n*

Set the number of intervals in the y direction to *n*.

Note if the number is not a valid value (i.e. factorable by 2,3,5,7, it will be set to the next highest valid value).

Prerequisites: a loaded map with phases

**fftnz** *n*

Set the number of intervals in the z direction to *n*.

Note if the number is not a valid value (i.e. factorable by 2,3,5,7, it will be set to the next highest valid value).

Prerequisites: a loaded map with phases

## A.4 Miscellaneous Commands

**silentmode**

This prevents dialog boxes from popping up during command processing.

Prerequisite: None

**clear**

Clears out all models, maps and vu objects.

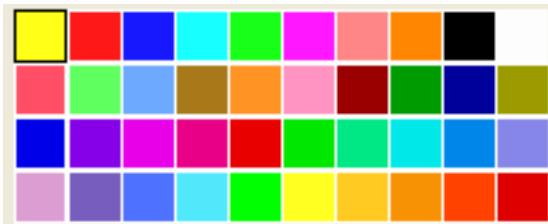
Prerequisite: None

**directory** *dir*

Sets the current working directory to *dir*. Handy to simplify file names and to specify where temp files will be written.

Prerequisite: None

# Appendix B: Default Colors



**Figure 11.72** Default colors

First row: colors 1-10

Second row: colors 11-20

N	RGB			Internal name
0	0	0	0	Black
1	255	255	20	Yellow
2	255	20	20	Red
3	20	20	255	Blue
4	20	255	255	Cyan
5	20	255	20	Green
6	255	20	255	Magenta
7	255	128	128	Pink
8	255	128	0	Orange
9	0	0	0	Brown
10	253	253	253	White
11	255	72	96	User 1
12	90	255	90	User 2
13	102	164	255	User 3
14	162	115	20	User 4
15	255	143	32	User 5
16	255	143	190	User 6
17	150	0	0	User 7
18	0	150	0	User 8
19	0	0	150	User 9
20	150	150	0	User 10

Third row: colors 21-30

Fourth row: colors 31-40

N	RGB			Internal name
21	0	0	230	Map 1
22	128	0	230	Map 2
23	230	0	230	Map 3
24	230	0	128	Map 4
25	230	0	0	Map 5
26	0	230	0	Map 6
27	0	230	128	Map 7
28	0	230	230	Map 8
29	0	128	230	Map 9
30	128	128	230	Map 10
31	218	152	207	Level 1
32	113	87	185	Level 2
33	72	107	254	Level 3
34	75	230	251	Level 4
35	0	255	0	Level 5
36	255	255	30	Level 6
37	255	200	30	Level 7
38	245	141	3	Level 8
39	255	60	0	Level 9
40	220	0	0	Level 10