MIFit Manual
Version 3.0 for Windows and LINUX

Duncan E. McRee and John Badger
© Rigaku 2003-6
All Rights Reserved
# Table of Contents

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>8</td>
</tr>
<tr>
<td>1  Getting Started</td>
<td>9</td>
</tr>
<tr>
<td>1.1 The MIFit Software</td>
<td>9</td>
</tr>
<tr>
<td>1.2 Launching MIFit</td>
<td>9</td>
</tr>
<tr>
<td>1.3 Establishing the working environment for MIFit</td>
<td>10</td>
</tr>
<tr>
<td>1.4 Storage of start-up parameters</td>
<td>12</td>
</tr>
<tr>
<td>1.5 Introduction to session files</td>
<td>13</td>
</tr>
<tr>
<td>1.6 Opening a document</td>
<td>13</td>
</tr>
<tr>
<td>1.7  Creating a new document</td>
<td>14</td>
</tr>
<tr>
<td>1.8 Crystal information</td>
<td>16</td>
</tr>
<tr>
<td>1.9 Default atom colors</td>
<td>16</td>
</tr>
<tr>
<td>1.10  Stereochemical dictionaries</td>
<td>18</td>
</tr>
<tr>
<td>1.11 Saving files</td>
<td>18</td>
</tr>
<tr>
<td>1.12 Closing a file and closing MIFit</td>
<td>18</td>
</tr>
<tr>
<td>2  Coordinate and map files</td>
<td>20</td>
</tr>
<tr>
<td>2.1 Loading coordinate data</td>
<td>20</td>
</tr>
<tr>
<td>2.2 Loading electron density data</td>
<td>22</td>
</tr>
<tr>
<td>2.3 Controlling the map display</td>
<td>26</td>
</tr>
<tr>
<td>2.4  Moving around the electron density map</td>
<td>28</td>
</tr>
<tr>
<td>3  Menus and toolbars</td>
<td>29</td>
</tr>
<tr>
<td>3.1 Startup menu / toolbar</td>
<td>29</td>
</tr>
<tr>
<td>3.2 General menu / tool bar</td>
<td>30</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.3 Modeling menu / toolbar</td>
<td>31</td>
</tr>
<tr>
<td>3.4 Display menu / toolbar</td>
<td>32</td>
</tr>
<tr>
<td>3.5 Ray-trace render menu / toolbar</td>
<td>32</td>
</tr>
<tr>
<td>4  Keyboard shortcuts</td>
<td>34</td>
</tr>
<tr>
<td>4.1 Keyboard shortcuts for document handling</td>
<td>34</td>
</tr>
<tr>
<td>4.2 Keyboard shortcuts for display</td>
<td>34</td>
</tr>
<tr>
<td>4.3 Keyboard shortcuts for stereo viewing</td>
<td>35</td>
</tr>
<tr>
<td>4.4 Keyboard shortcuts for model fitting</td>
<td>36</td>
</tr>
<tr>
<td>5 Model Building and Refitting</td>
<td>37</td>
</tr>
<tr>
<td>5.1 Chain tracing</td>
<td>37</td>
</tr>
<tr>
<td>5.2 Refitting backbone atoms</td>
<td>37</td>
</tr>
<tr>
<td>5.3 Building and fitting side chain atoms</td>
<td>38</td>
</tr>
<tr>
<td>5.4 Refitting and adding individual residues</td>
<td>40</td>
</tr>
<tr>
<td>5.5 Modeling discrete disorder</td>
<td>41</td>
</tr>
<tr>
<td>5.6 Local structure refinement</td>
<td>42</td>
</tr>
<tr>
<td>5.7 Adding water molecules</td>
<td>43</td>
</tr>
<tr>
<td>5.8 Interactive ligand fitting</td>
<td>44</td>
</tr>
<tr>
<td>5.9 Interactive structure refinement with CCP4/REFMAC5</td>
<td>45</td>
</tr>
<tr>
<td>6 The Ligand Dictionary Editor</td>
<td>47</td>
</tr>
<tr>
<td>6.1 Entering ligand data</td>
<td>47</td>
</tr>
<tr>
<td>6.2 Conformation generation</td>
<td>48</td>
</tr>
<tr>
<td>6.3 The Ligand Dictionary Editor</td>
<td>49</td>
</tr>
<tr>
<td>6.4 Exporting ligand information</td>
<td>50</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6.5 Restraint dictionary files for SHELX and CNS/CNX</td>
<td>51</td>
</tr>
<tr>
<td>7 Displaying structure data</td>
<td>53</td>
</tr>
<tr>
<td>7.1 Setting model display modes</td>
<td>53</td>
</tr>
<tr>
<td>7.2 Displaying model surfaces</td>
<td>55</td>
</tr>
<tr>
<td>7.3 Ribbon and backbone displays</td>
<td>57</td>
</tr>
<tr>
<td>7.4 Choosing a background color</td>
<td>59</td>
</tr>
<tr>
<td>7.5 Printing directly from the MIFit main canvas</td>
<td>60</td>
</tr>
<tr>
<td>7.6 Ray tracing</td>
<td>61</td>
</tr>
<tr>
<td>7.7 Schematic protein:ligand interaction plots</td>
<td>64</td>
</tr>
<tr>
<td>7.8 Data capture for structure reports and PDB deposition</td>
<td>65</td>
</tr>
<tr>
<td>8 Running external crystallographic software</td>
<td>70</td>
</tr>
<tr>
<td>8.1 The Python automation scripts</td>
<td>71</td>
</tr>
<tr>
<td>8.2 MIFit access to a Python interpreter</td>
<td>71</td>
</tr>
<tr>
<td>8.3 Accessing the CCP4 suite</td>
<td>72</td>
</tr>
<tr>
<td>8.4 Accessing SHELX</td>
<td>73</td>
</tr>
<tr>
<td>8.5 The Job List menu</td>
<td>73</td>
</tr>
<tr>
<td>8.6 Script catalogue</td>
<td>75</td>
</tr>
<tr>
<td>8.7 Automation script keywords</td>
<td>78</td>
</tr>
<tr>
<td>9 SAD phasing</td>
<td>87</td>
</tr>
<tr>
<td>9.1 Prerequisites</td>
<td>87</td>
</tr>
<tr>
<td>9.2 SAD phasing interface</td>
<td>87</td>
</tr>
<tr>
<td>9.3 SAD phasing output files</td>
<td>89</td>
</tr>
<tr>
<td>10 Automated Structure Solution</td>
<td>93</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>10.1 Overview of automated co-crystal structure determination</td>
<td>93</td>
</tr>
<tr>
<td>10.2 The automated co-crystal structure determination interface</td>
<td>94</td>
</tr>
<tr>
<td>10.3 The Job/Molecular Replacement interface</td>
<td>98</td>
</tr>
<tr>
<td>10.4 The Job/Refinement interface</td>
<td>99</td>
</tr>
<tr>
<td>10.5 Checking requirement for restraint dictionaries</td>
<td>101</td>
</tr>
<tr>
<td>10.6 The project history file</td>
<td>102</td>
</tr>
<tr>
<td>10.7 Automated structure validation</td>
<td>103</td>
</tr>
<tr>
<td>10.8 Automated ligand fitting</td>
<td>104</td>
</tr>
<tr>
<td>10.9 Semi-automated ligand fitting</td>
<td>105</td>
</tr>
<tr>
<td>10.10 Automated multiple protein:ligand structure superposition</td>
<td>107</td>
</tr>
<tr>
<td>11 Tutorial Lessons</td>
<td>110</td>
</tr>
<tr>
<td>Lesson 1: Basic manipulation of the model display</td>
<td>110</td>
</tr>
<tr>
<td>Lesson 2: Loading and changing electron density map displays</td>
<td>113</td>
</tr>
<tr>
<td>Lesson 3: Loading crystal data and displaying symmetry related molecules</td>
<td>115</td>
</tr>
<tr>
<td>Lesson 4: Refitting individual residues</td>
<td>117</td>
</tr>
<tr>
<td>Lesson 5: SAD phasing</td>
<td>120</td>
</tr>
<tr>
<td>Lesson 6: Chain tracing</td>
<td>123</td>
</tr>
<tr>
<td>Lesson 7: Establishing restraints for structure refinement</td>
<td>126</td>
</tr>
<tr>
<td>Lesson 8: Displaying molecular surfaces</td>
<td>128</td>
</tr>
<tr>
<td>Lesson 9: Creating schematic plots of protein:ligand interactions</td>
<td>130</td>
</tr>
<tr>
<td>Lesson 10: Creating reports for PDB deposition and publication</td>
<td>132</td>
</tr>
<tr>
<td>Lesson 11: Automated co-crystal structure determination</td>
<td>135</td>
</tr>
<tr>
<td>Lesson 12: Automated Molecular Replacement</td>
<td>138</td>
</tr>
<tr>
<td>Lesson 13: Automated refinement</td>
<td>141</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>Lesson 14: Automated ligand density docking</td>
<td>143</td>
</tr>
<tr>
<td>Lesson 15: Superimposing ligands from multiple related structures</td>
<td>145</td>
</tr>
<tr>
<td>Appendix 1: Scripting Language</td>
<td>148</td>
</tr>
<tr>
<td>Model Commands</td>
<td>148</td>
</tr>
<tr>
<td>View Commands</td>
<td>149</td>
</tr>
<tr>
<td>Map Commands</td>
<td>150</td>
</tr>
<tr>
<td>Miscellaneous Commands</td>
<td>154</td>
</tr>
<tr>
<td>Appendix 2: Default Colors</td>
<td>155</td>
</tr>
</tbody>
</table>
Acknowledgements

The recent development of the MIFit software was largely funded by the members of the Automatic Structure Consortium (ASC) (http://www.molecularimages.com/asc.html).

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).

MIFit uses a library from The FreeType Project (www.freetype.org).
1 Getting Started

1.1 The MIFit software

MIFit 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 provides interfaces for automated structure solution and to allow interaction with external crystallographic software (the CCP4 suite and SHELX). A particular focus of the MIFit software is the efficient solution and analysis of protein:ligand complexes.

MIFit was originally created by Duncan McRee as a successor to the XtalView/Xfit program. MIFit is currently being developed by Molecular Images (a division of Rigaku) and is usually pronounced ‘m. i. fit’.

1.2 Launching MIFit

On computers that run the LINUX operating system MIFit may be started from the command line by simply executing MIfit.exe. On computers that operate under Windows, MIFit may be launched either by double-clicking on the MIFit desktop icon or by selecting Programs/ MIFit/ MIFit from the Windows Start menu.

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 Establishing the working environment for MIFit

The **File/Properties**... 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 **Molimage Home** and **XtalView Home** parameters should normally be identical and be set as paths to your MIFit installation.

![Set Properties](image)

Figure 1.2 **File/Properties** menu

If you are running MIFit from an installation on a shared system (usually a LINUX server), file protections will typically be set by the system administrator so that individual users do have the required permission to change the files specified by the **Crystal Data Directory** and the **Dictionary File** parameters. The **Crystal Data Directory** contains a set of files that define the crystal cell and spacegroup information for structure determination projects. The **Dictionary File** defines stereochemical information for the standard amino acids, cofactors and any ligands that you wish to add to the model. To work on new projects you may find it necessary to make personal copies of these files and set the **Crystal Data Directory** and the **Dictionary File** paths to them.
The Default Crystal corresponds to the name of a crystal file in the Crystal Data Directory and may be updated to your current project using the File/Crystal.. command.

If you wish to use the SAD phasing capability (Chapter 9), automated reporting capabilities (Section 7.8) or the structure solution automation (Chapter 10) then a Python interpreter must be installed on your computer and the Python Interpreter field should be directed to that interpreter. Most installations of the LINUX operating system already include a Python interpreter but this may need to be acquired and installed on computers that run Windows.

The CCP4 Home parameter is used by the Job/Run Refmac command in order to locate the CCP4/REFMAC5 program and its associated files. The path defined by this parameter must be directed to a CCP4 installation to enable this command.

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 Left Panel Width parameter controls the width of document tree pane and the log window pane. The default value of 260 need not normally be changed.

The Xfit Mouse Mode checkbox may be used to enable a three-button mouse.

Clicking on OK sets and saves revised parameters within this interface.

1.4 Storage of start-up parameters

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. The simplest possible $MIFit$ file looks like:

```plaintext
file1=/home/paehler/xvc2/example1.mlw
MOLIMAGEHOME=/xtal/software/MolecularImages
XTALVIEWHOME=/xtal/software/xtalview-syr
CRYSTALDATA=/home/paehler/xtal_info
CRYSTAL=
XFITDICT=/xtal/software/xtalview-syr/data/dict.noh.pdb
[Window\ Position]
Width=1281
Height=942
XPos=-4
YPos=0
```
1.5 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.

MIFit session files are ASCII files that contain a mixture of plain text and XML formatting. A basic, 'empty' session file looks like:

```xml
<?xml version = "1.0" encoding="UTF-8"?><!
<wxFitDocument CreationDate="15:42:21 03/05/03" version="1.0">
  <AtomStack></AtomStack>
  <ViewPoint>
    translation 0.00 0.00 0.00
    rotation 1.0000 0.0000 0.0000 0.0000 1.0000 0.0000 0.0000 0.0000 1.0000
    zoom 20.000
    perspective 1.000
    frontclip 5.00
    backclip -5.00
    transform
    stereo off
  </ViewPoint>
</wxFitDocument>
```

1.6 Opening a document

A document that already exists on a user’s file system may be opened using the File/Open... command (keyboard shortcut Ctrl+X). 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 (which is stored in the
.MIfit file) and the paths to them are displayed in the File menu. Double-clicking on file names in this list provides convenient access to these structures.

Diffraction data files are loaded into MIFit with the Map/Load Phase file command; pre-computed electron density maps are be loaded with the Map/Load Map file command. Diffraction data and maps are discussed in chapter 2.

1.7 Creating a new document

A new MIFit document may be created with the File/New... command or the keyboard shortcut Ctrl+N. The File/Import into New Document may then be used to add content (coordinate data) to that document.

After selecting the File/New... command, MIFit prompts the user for a document template (Figure 1.3). For new model fitting applications you should apply the default option, Model Fitting Session, so just click OK.

![Figure 1.3 Available document templates for the File/ New... command](image)

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 top level menu item (Window) 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 1.4). 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 canvas, sequence and navigator windows with document-related content.

Figure 1.4 MIFit with a new empty document loaded

Figure 1.4 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 small canvas will show a scaled drawing of the Cα trace of the whole model. The tree windowpane now shows an additional item, the new empty session file, *unnamed1.mlw*. If further new documents were added they would be named *unnamed2.mlw*, *unnamed3.mlw* etc.

After **File/New** is selected the **File/Add Model to Current Document** command, which had been previously inactive, becomes available for importing model structure data. The types of file that are available through this filter are coordinate files in PDB (*.pdb, *.ent), mmCIF (*.cif) and SHELX (*.res, *.ins) formats.
1.8 Crystal information

Information on a crystal (i.e. cell dimensions, space group and associated symmetry operators) may be entered, saved and selected using the File/Crystal... command.

This command presents the ‘Set Crystal’ popup window (Figure 1.5). In this window, the pulldown menu associated with the Crystal parameter (the name of a specific crystal file) may be used to load information from other available crystal files. In addition, a new crystal may be established by editing the Crystal file name and entering the Title, Unitcell and a Spacegroup name or number. (Note that it is not necessary to enter the symmetry operations – these are determined using the Find button based on the Spacegroup parameter). Clicking OK adds this new crystal to the set of known crystals stored in the directory defined by the File/Properties.../Crystal Data Directory parameter.

![Figure 1.5 Set Crystal dialog box](image)

1.9 Default atom colors

The File/Define Atom Colors command (Figure 1.6) 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.
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 1.7).

![Figure 1.6 Dialog window defining the atom coloring scheme](image)

![Figure 1.7 Dialog window for defining atom colors according to B-factor values](image)
1.10 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, internal refinement and is a basis for ‘interactive’ reciprocal-space refinement with the Job/Run Refmac command.

The default dictionary is specified in the dialog window that appears when the File/Properties… command is issued (Figure 1.2). When a new dictionary is loaded into MIFit the log window in the lower left will report statistical information on the dictionary contents. If no dictionary is loaded, you should set the path to a dictionary by using the File/Properties command. The default dictionary for a new MIFit installation will usually be the file /MIFIT/data/dict.noh.pdb. If no dictionary is loaded then the commands to replace or refine protein entities will not work.

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 To Dictionary command. The use of this command is discussed in Chapter 6.

1.11 Saving files

The File/Save 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 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 (Figure 1.8) appears.

![Figure 1.8 Prompt for saving changed documents](image-url)
If several models are open and you switch between models, the **File/Save** command menu could either be active or inactive, depending on whether any changes were made to the selected structure.

### 1.12 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.
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… command. This command opens a window (Figure 2.1) from which you may specify the type of coordinate file that you wish to load.

The two appropriate options available from the Files of type pulldown menu are Model Fitting Session (*.mlw) and Model Fitting Session (*.pdb). These files correspond to a MIFit session file and a PDB coordinate file.

Once you have selected a coordinate file this structure data is immediately loaded and displayed (Figure 2.2).
Figure 2.2 MIFit window after loading a coordinate file

The upper left pane contains a tree-control display of the loaded structure. The tree may be expanded to the level of individual amino acid residues by double-clicking on the associated entry. Within each document, there may be one or more leaves for individual structures. Each structure is further subdivided into chains and segments, and may also contain labels and annotations. These chains/segments may be further expanded to display individual residues (or labels or annotations). Double-clicking on a residue in the tree control will center that residue in the canvas pane.

The upper right window contains a sequence view with a schematic drawing of the calculated secondary structure. 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 **Object/ Stack/ Clear top item** and the entire stack may be cleared with menu item sequence **Object/ Stack/ Clear stack**. Atomic labels and information in the message area may be erased with commands **View/ Clear Labels** or **View/ 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 either the **Map/ Load Phase file** command or the **Map/ Load Map file** command.

The **Map/ Load Phase file** command 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. The **Map/ Load Map file** command requires a pre-calculated map in the CCP4 or XtalView/xfit map format. There may be a few situations in which pre-calculated maps are more appropriate. 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.

Selecting the **Map/ Load Phase file** command creates a browser window from which the user may select and load diffraction data files in any of the currently supported formats. The menu choices currently available include the XtalView format (.phs), the SHELX format (.fcf), the SCALEPACK format (.sca), the CCP4 format (.mtz), the mmCIF format (.cif) and the D*TREK format (.ref). It should be noted that many different types of data are stored in mmCIF files – obviously, in 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 the
Map/Load Phase file command is used to select a MTZ file for data that corresponds to a new crystal. A dialog box (Figure 2.3) will appear in order to allow the user to create a new crystal file. If a crystal with corresponding cell dimensions is already present in the MIFit system, this step is skipped. This dialog box will prompt the user for the name of the new crystal file.

![Figure 2.3 Dialog box to establish the name for a new crystal](image)

After entering a name for the crystal file ('unknown' is not a good name!) and clicking on the OK button, the cell and crystal symmetry information that were extracted from the MTZ file are presented in a dialog window for confirmation (Figure 2.4). This dialog window allows the user to check the crystal file information before accepting it by clicking on the OK button.

![Figure 2.4 Dialog window listing crystal data](image)

If a map is accidentally created with crystal data that does not match the structure, the map may be deleted by right-clicking on the map entry in the tree control and selecting Delete.
When the crystal information associated with the MTZ data file is defined, a dialog window appears that shows the mapping of MTZ data labels to established MIFit data types (Figure 2.5).

![Figure 2.5 Dialog window for setting MTZ data labels](image)

In this process MIFit attempts to infer the correct labels for different categories of MTZ file data. In the example shown in Figure 2.5 all of the data types were correctly identified. The pulldown menus to the right of each data field may be used to reassign the MIFit data types to other MTZ column labels.

Once the correct data mappings have been made, pressing the **OK** button creates a dialog window for calculating the electron density map (Figure 2.6).

![Figure 2.6 Dialog window for density map calculation](image)
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. The pulldown 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. Setting the Grid 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 in the FFT menu the specified map is calculated and displayed in the canvas window (Figure 2.7).

![Figure 2.7 Model and map display](image)

It is worth noting that there is one common situation where it might be useful to set the Fo column and Phi column shown in Figure 2.5 to Fourier coefficients other 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 (usually FWT, PHWT and DELFWT, PHDELFWT) for the calculation of ‘normal’ and difference maps. It may efficient to enter these sets of values as dummy Fo column and Phi column values and then per-
form the calculation as if for the calculation of an Fobs map. This calculation will result in the likelihood-weighted map appearing in the canvas window. Obviously, following this calculation, one should not subsequently perform any operation that would require the true observed structure factor.

It appears not to be ideal to use the Fc values calculated using CCP4/REFMAC5 in the direct calculation of a difference maps (Map Type Fo-Fc in the Fast Fourier Transform Map menu) due to scaling issues with the low resolution data. To obtain this type of map it is better to use either the pre-computed difference map coefficients (usually DELFWT and PHDELFWT) or re-compute the Fc values within MIFit by selecting the Map/Calculate Structure Factors... option.

2.3 Controlling the map display

Once loaded, the electron density map display may be modified in several ways:

- by right-clicking the tree control map leaf (i.e. the map icon at the bottom of the tree in Figure 2.7) and selecting the Color/Contour option
- by selecting the Map/Contour Options menu item
- by right-clicking anywhere inside the canvas pane and selecting Map/Contour Options from the popup dialog menu

All three of these actions will bring up the dialog window shown in Figure 2.8.
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 \(5-10\text{Å} \).

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** pulldown 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** pulldown 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 2.8 (50, 100, 150,..) correspond to density levels at \(1,2,3,4,5\) sigma.

![Figure 2.8 Map/Contour Options menu](image_url)
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.

### 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 **Map/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 **Map/Find Ligand Density** tool 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 Quick Access Menu available. This menu contains a convenient agglomeration of frequently used menu items that are otherwise distributed over several menus on the menu bar.

3.1 Startup menu / toolbar

<table>
<thead>
<tr>
<th>Menu/Toolbar</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>File/New</td>
<td>Create a new structure document</td>
</tr>
<tr>
<td>File/Open...</td>
<td>Open an existing structure document</td>
</tr>
<tr>
<td>File/Save, File/Save As...</td>
<td>Save a structure document</td>
</tr>
<tr>
<td>File/Print...</td>
<td>Print the canvas</td>
</tr>
<tr>
<td>Object/Copy Canvas</td>
<td>Copy the canvas to the clipboard (Windows only)</td>
</tr>
<tr>
<td>Abort a long operation</td>
<td>Used to abort an operation</td>
</tr>
<tr>
<td>File/Crystal...</td>
<td>Create or edit a crystal</td>
</tr>
<tr>
<td>Help/Help...</td>
<td>Invoke help (this manual)</td>
</tr>
</tbody>
</table>

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**,** File/Save 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.

<table>
<thead>
<tr>
<th>Toolbar Item</th>
<th>Menu Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object/Add Annotation to Model</strong></td>
<td></td>
<td>Add annotation at canvas center</td>
</tr>
<tr>
<td><strong>Checkpoint Current Model</strong></td>
<td></td>
<td>Save model state in PDB format temporary file</td>
</tr>
<tr>
<td><strong>View/Slab In</strong></td>
<td></td>
<td>Decrease display depth in z-direction</td>
</tr>
<tr>
<td><strong>View/Slab Out</strong></td>
<td></td>
<td>Increase display depth in z-direction</td>
</tr>
<tr>
<td><strong>View/Zoom Out</strong></td>
<td></td>
<td>Zoom out to expand view of model</td>
</tr>
<tr>
<td><strong>View/Zoom In</strong></td>
<td></td>
<td>Zoom in to shrink view of model</td>
</tr>
<tr>
<td><strong>View/Top View</strong></td>
<td></td>
<td>Display view down z-axis</td>
</tr>
<tr>
<td><strong>View/Rotate View +90</strong></td>
<td></td>
<td>Rotate about y by +90°</td>
</tr>
<tr>
<td><strong>View/Rotate View -90</strong></td>
<td></td>
<td>Rotate about y by -90°</td>
</tr>
<tr>
<td><strong>View/Decrease Perspective</strong></td>
<td></td>
<td>Decrease canvas perspective</td>
</tr>
<tr>
<td><strong>View/Increase Perspective</strong></td>
<td></td>
<td>Increase canvas perspective</td>
</tr>
</tbody>
</table>

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 picked residue** option becomes active. When this item is activated or when the **Fit Fit Residue** command is selected all the **Fit** items except for **Twist Bond** are activated. The **Twist Bond** option becomes active when a bond defining a torsion angle has been selected.

- **Fit/ Fit Residue**
  - Fit residue containing top-of-stack atom
- **Fit/ Apply**
  - Apply modifications to selected residue(s)
- **Fit/ Cancel**
  - Cancel all modifications
- **Fit/ Translate**
  - Switch to translating selected residue(s)
- **Fit/ Rotate**
  - Switch to rotating selected residue(s)
- **Fit/ Torsion**
  - Apply torsion motion to selection
- **Fit/ Center**
  - Right mouse pans viewpoint

Table 3.3 Modeling toolbar items

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

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show/ Radius</td>
<td>Set Radius to Cylinder</td>
</tr>
<tr>
<td>Show/ Radius</td>
<td>Set Radius to Ball</td>
</tr>
<tr>
<td>Show/ Radius</td>
<td>Set Radius to CPK</td>
</tr>
<tr>
<td>Show/ Radius</td>
<td>Radius Tool</td>
</tr>
<tr>
<td>Show/ Color</td>
<td>Color Tool</td>
</tr>
<tr>
<td>Object/ Surface</td>
<td>Surface Tool</td>
</tr>
<tr>
<td>Show/ Show Residues</td>
<td>Show Tool</td>
</tr>
<tr>
<td>Show/ Show Sidechains</td>
<td>Show Sidechain Tool</td>
</tr>
<tr>
<td>Show/ Hide</td>
<td>Hide Tool</td>
</tr>
</tbody>
</table>

Table 3.4 Display toolbar items

The radius selection toolbar items do not operate on their own but work in conjunction with the Show/Radius menu options or with the Radius Tool toolbar item. Color selection capabilities may be accessed with the Show/Color menu options or with the Color Tool toolbar item.

3.5 Ray-trace render menu / toolbar

Scenes can be sent to a ray-tracer document using the Render/Render with Ray-Tracer... command to make a presentation quality graphic. This action creates a new document with file extension .shw and loads this document into the ray-tracer. The ray-tracer contains options that provide control over colors, backgrounds, pixel size and depth cueing. The ray-tracer also includes a quick preview feature (the Render/Preview Picture command) that allows the user to preview the picture prior to a full rendering. The Render/Raytrace Picture command creates the fully detailed image with lighting, shadows and anti-aliasing. These pictures can then be saved in any of a variety of commonly used graphics formats using File/Export Image As command from the ray-tracer canvas. More information on the ray-tracer may be found in Chapter 7.
- **Render/Preview**  
  Quick preview of picture

- **Render/Raytrace**  
  Ray trace picture (may take some time)

- **Render/Cancel Tracing**  
  Stop the ray tracing – usually to fix something and start again
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 ‘|’ indicates ‘or’ (i.e. when two shortcuts result in the same action). The notation ‘/’ is used to indicate two alternative shortcuts that usually have opposite actions. The notation ‘Ctrl+’ indicates simultaneous selection of the Ctrl key with a keyboard shortcut. Similarly, the notation ‘Alt+’ indicates simultaneous selection of the Alt key with a keyboard shortcut.

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.

| Ctrl+N | File/ New | Create a new structure document |
| Ctrl+X | File/ Open... | Open an existing structure document |
| Alt+X | Exit | Terminates program |

Table 4.1 Keyboard shortcuts associated with menu and toolbar items

4.2 Keyboard shortcuts for display

| x | X | View along x-axis (from origin) |
| y | Y | View along y-axis (from origin) |
| z | Z | View along z-axis (from origin) |
| Left / Right | Rotate left / right about y-axis |
| Up / Down | Rotate up /down about x-axis |
The notations **Left/Right/Up/Down** refer to the arrow cursor keys.

### 4.3 Keyboard shortcuts for stereo viewing

MIFit currently supports side-by-side stereo and also contains a full screen mode 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 **View/ Side-by-side Stereo**.

With top-down (hardware) stereo display the canvas is split into a top and a bottom half, displaying the left/right eye image in the top/bottom half respectively. This stereo mode is only used in conjunction with stereo hardware inserted between video card and monitor and with appropriate stereo glasses (*CrystalEyes* and *NuVision*). To take advantage of hardware stereo a Nu-Vision 60GX-NSR stereo emitter and glasses are needed. Quad-buffered stereo, which is supported by certain video cards and allows stereo in a window is not supported by MIFit at this time.

<table>
<thead>
<tr>
<th>Key</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>I</code></td>
<td>Toggle side-by-side (sw) stereo</td>
</tr>
<tr>
<td><code>O</code></td>
<td></td>
</tr>
<tr>
<td><code>ESC</code></td>
<td>Toggle full-screen display</td>
</tr>
<tr>
<td><code>F2</code></td>
<td>Toggle top-down (hw) stereo</td>
</tr>
<tr>
<td><code>F3</code></td>
<td>Decrease / increase blank interval (hw stereo)</td>
</tr>
<tr>
<td><code>F4</code></td>
<td></td>
</tr>
<tr>
<td><code>A</code></td>
<td>Decrease / increase stereo angle (sw stereo)</td>
</tr>
<tr>
<td><code>S</code></td>
<td></td>
</tr>
<tr>
<td><code>Q</code></td>
<td>Decrease / increase stereo angle (hw stereo)</td>
</tr>
<tr>
<td><code>W</code></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3 Keyboard shortcuts to manipulate stereo display
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.

[ | PgUp] Go to N-terminus of chain

] | PgDown Go to C-terminus of chain

< / > Insert MRK residue before / after present residue

Space Go to next residue

; Apply fit / Accept refinement

c | C Select next conformer (side-chain)

d | D Delete residue

f | F Fit residue

r | R Replace and fit residue (side chain refitting)

Ctrl+r | Ctrl+R Refine residue

Shft+w | Shft+W Add water molecule at view center

1 / 2 / ... / 5 Select $\chi_1$ (chi-1), $\chi_2$, ..., $\chi_5$ side-chain torsion angle

Table 4.4 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 Cα chain-tracing, converting Cα-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 majority of structure models were 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 Cα 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 Cα atom either before or after the current Cα atom. If the predicted position appears to be incorrect then the right mouse button may be used to adjust the Cα site at fixed distance of 3.8Å from the previous Cα.

The Fit/ Apply 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 Cα 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 Cα 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.

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 Cα 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 5.1).

![Figure 5.1 Pentamer match for model structure following A 759 Met](image)

Portions of the associated protein backbone may be shifted to match the displayed pentamer using the **Refi/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 **Refi/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 **Refi/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 5.2) appears from which the new amino acid type may be selected.

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 short cut ‘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 short cut ‘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 short cut ‘f’. The numeric keyboard keys (1, 2, ..) may then be used to select a torsion angle (χ1, χ2,..) 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 (ʼ) and adjusted using the right-mouse button (the left-mouse button retains control of the view orientation). The rotation icon (ʼ) 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 (ʼ) 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 Atoms**, **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 **Object/ Stack** commands may be used to manipulate the contents of the stack.

To accept a fit the **Fit/ Apply** command or the ✓ icon may be used. To cancel a fit the **Fit/ Cancel** 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** command may be selected.

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

![Model/ Add residue dialog box](Image)

**Figure 5.3 Model/ Add residue** dialog box

The pulldown menu corresponding to the **Residue Type** parameter in the resulting dialog box (Figure 5.3) is used to specify entity that is to be
added to the model. The **Insert Position** options describe where in co-
ordinate set the new entity will be placed.

The **Put At** options are not currently functional and the default value may
remain as ‘Screen Center’.

The **Model/ Delete Residue** (or equivalently, the 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 num-
ber for an individual entity.

![Model/ Rename Residue dialog box](image)

**Figure 5.4 Model/ Rename Residue** dialog box

In the example (Figure 5.4) the original amino acid number (761) may be arbitrarily changed using the associated field.

### 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 re-
finement 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 conforma-
tions.

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 acti-
vate it with **Fit/ Fit residue** command or the keyboard shortcut ‘f’. The **Fit/ Disorder/ Split Fit** command may then be used to duplicate the en-
tire amino acid, with the duplicate copy slightly offset from the original copy for clarity. The duplicated copy may be manipulated with the interac-
tive fitting tools *i.e.* the translation ( ), rotation ( ) and torsion adjust-
ment ( ) 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 toolbar icons ✔ and ❌ may be used to accept or re-
ject the fit.
5.6 Local structure refinement

The entire molecule or portions of the molecule may be optimized within MIFit using commands beneath the Refi pulldown. The Refi/Refi Options command (Figure 5.5) 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 Cα positions and the ends of the refinement zone.

![Refinement Options](image)

Figure 5.5 Options for structure optimization

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

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

Selecting two atoms by clicking on them in the main canvas and then applying the Refi/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 5.6).

![Add Waters to Map](image)

Fig 5.6 Dialog window 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 5.2).

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 (X) are used to toggle between translational (直辖) and rotational (穸) 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 Interactive 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 MI Fit 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. The Job/Run Refmac command is a convenient tool for quick refinements within the context of a model-fitting session and will be described in this section.

To use the Job/Run Refmac command both a model and a map must be loaded. If the input data was from an MTZ data file a menu (Figure 5.7) will appear to prompt the user for correct MTZ column assignments.

![Figure 5.7 Selection of MTZ labels for refinement](image)

This step is necessary to ensure that true Fo values are passed to the CCP4/REFMAC5 program since the previous map calculation may have used dummy parameters in place of the measured structure factor amplitudes.

The refinement run is controlled by a single parameter, the number of refinement cycles (Figure 5.8). Typically 5-10 cycles are sufficient for a quick refinement. The weight for balancing the contribution of data versus stereochemical terms to the refinement is determined automatically.
Figure 5.8 Menu to set the number of refinement cycles

After selecting **OK** you will see a text window appear to indicate the launch of the refinement job. You may also check the **Job List** in the document tree pane. After right-clicking on the job id, the **Show Log File** option will become accessible for viewing the job log once the job is completed. Upon job completion the refined model will appear in the MIFit main canvas and the MTZ column selection and Fast Fourier Transform menus will appear to allow the user to immediately recompute the refined map.

If the protein structure contains ligand molecules the CCP4/REFMAC5 refinement dictionaries will be created automatically provided that these molecules were entered into the MIFit dictionary. Note that the CCP4/REFMAC5 installation contains a large number of cofactors and ligands and, assuming they were modeled with consistent entity and atom names, these will be handled transparently by the refinement process. Program failures are usually due to problems with the ligand nomenclature and may be diagnosed from the **Job List/ Show Log File** menu.
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) 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 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.

The Smiles option leads to an interface that provides three different methods for importing the SMILES string (Figure 6.1).
Figure 6.1 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 small Python script that is capable of returning a SMILES string when a ligand identification number ('regno') is provided through interface. To utilize this capability the Python script must be called *smilesdb.py* and should be placed in the ..data/Scripts directory within the MIFit installation. An illustrative (dummy) script is supplied to return a SMILES string on input of the regno ‘12345’. This capability is mainly intended for users with access to a corporate small molecule data base since these databases typically store SMILES data. To use the **Database** option a path to a Python interpreter must be provided as the **File/Properties/Python Interpreter** parameter (see Chapter 1).

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

After entering small molecule data via one of the **Dictionary/Import Ligand** commands users are asked whether they wish to generate multiple conformers for the ligand. Users of the enhanced (automation) version of MIFit that intend to use the automated ligand density fitting capabilities within MIFit should select ‘yes’; in other cases ‘no’ may be selected. 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. If no ligand conformers were generated but it is decided later to use the auto-
mated ligand fitting options, both the **Dictionary/ Generate Conformers** command and the **Dictionary Editor** provide options for generating the additional conformations at a later time.

If the ligand contains hydrogen atoms then these should be stripped from the coordinates before generating conformations for ligand fitting.

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 rederived by MIFit for the purpose of generating the conformational ensemble.

### 6.3 The Ligand Dictionary Editor

The **Dictionary Editor** interface (Figure 6.2) appears after completing the entry of ligand data.

![Dictionary Editor interface](image)

**Figure 6.2 Dictionary Editor interface**
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 Dictionary Editor provide options for displaying the various refinement restraints established for this molecule. 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 pulldown menus at the top of the 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 most common problem with automatically generated restraints is that the plane group definitions are only partially correct. The 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 pulldown 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. Selecting this button creates a dialog window (Figure 6.3)
from which the user may elect to write the ligand structure information in the mmCIF format for refinement with CCP4/REFMAC5, as a coordinate file in PDB or MOL format or as a SMILES string.

![Output format](image)

**Figure 6.3 Output format options for ligand data**

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. On execution of these commands an interface appears to provide for entry of the dictionary file in mmCIF format (Figure 6.4).
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. This is the representation that is also used by the XtalView/xfit program. 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 7.1) 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.
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 7.2 Collage of basic display methods

Shown from top-left in clockwise direction in Figure 7.2 are stick, knobs-and-stick, CPK and balls-and-stick displays.

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.

Options related to surface displays are selected from either the Object/Surface menu (Figure 7.3) or from inside the canvas window by right-clicking the mouse to pop up the QuickMenu.
Selecting the **Object/Surface/van der Waal Surface** option creates a dialog window (Figure 7.4) that contains options for controlling the calculation of the surface.

Only one surface can exist in MIFit 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 **Object/Surface** options result in a dialog window similar to the one shown in Figure 7.4 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 **Object/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 **Object/Clear Surface** command may be used to eliminate an existing surface.
A collage of some representative surface displays is shown in Figure 7.5. 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 7.5 Collage of surface displays](image)

### 7.3 Ribbon and backbone 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.

An alpha-carbon representation may be generated by selecting the **Show/CA Backbone** option. Similarly, the **Show/Make Ribbon** command creates a ribbon image of the current model. The **Show/Clear Ribbon** command removes the ribbon image and the **Show/Ribbon Colors...** command spawns a dialog window (Figure 7.6) that allows the
user to change the default colors for the helix, sheet and random coil secondary structure elements.

Figure 7.6 **Show/ Ribbon Colors...** dialog window

A collage of Ca-carbon and ribbon visualizations is shown in Figure 7.7.

Figure 7.7 Ca trace and ribbon diagrams
The top-left image in Figure 7.7 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.

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, Figures 7.5 and 7.7 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. 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 6.8).

![File/Print Preview window](image)

Figure 7.8 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 parame-
ter 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 Object/Copy Canvas command.

7.6 Ray tracing

Although images may be printed directly from the MIFit canvas or captured with the Object/Copy Canvas command, MIFit also contains a ray-tracer for creating and exporting molecular images. The ray-tracer provides a superior rendering capability that is particularly useful for creating presentation graphics.

The Render/Render with Ray-tracer... command is used to access the ray-tracing application within MIFit. After selecting this command a dialog window for setting the ray-tracer parameters appears (Figure 7.9).

![Figure 7.9 Ray-tracer dialog box](image)
After clicking on the **OK** button the entire MIFit window changes to the appearance shown in Figure 7.10.

![Figure 7.10 MIFit ray-tracer window](image)

From the ray-tracer window the **View** commands may be used to adjust the size and position of the molecular image. The **Object** commands allow control of lighting, color and backgrounds for the image. Noteworthy is the **Object/Import Background...** option that allows import of an image file to use as background to the molecular display. The **Render** pulldown menu contains the command that is used to render the picture (**Render/Raytrace Picture**) as well as a **Render/Preview Picture** command to quickly preview the picture. Other commands control the display of shadows and the amount of ray-tracing that is to be performed.

Since ray-tracing is a relatively slow process it will often be best to first create the picture using quick display modes and, once it has been verified that the image parameters give the picture that is desired, improve the rendering. A useful trick is to use the **View/Picture Size...** command to initially create an image at relatively low resolution and, once the desired image is obtained, increase the resolution to generate a publication quality image.
Besides creating color illustrations, the ray-tracer has a built-in option (Render/Graylevels) to create black and white illustrations. Figure 7.12 shows an example of a black-and-white image that combines a user-selected background (a diffraction pattern), a ribbon representation of the model and a small molecule within its associated electron density.

Figure 7.11 Example of small molecule rendered as a metal object and placed on a colored background

Figure 7.12 A black-and-white image that combines a special background, a ribbon display and a ligand in electron density.
Ray-traced images may be exported in standard image file formats using the **File/Export Image As...** command. You may also save the image parameters in a showcase (file extension .shw) command file for reuse and possible modification at some later time. For example, if you were asked to create a new version of an existing image but at higher resolution (say, for a journal cover) the showcase file could be used to regenerate the image with a greater number of pixels.

The ray-tracer window may be closed using the **File/Close** command; this action returns the molecular display in the MIFit canvas.

### 7.7 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 **Model/Export Active Site Plot** option.

Following these actions a new window will appear that shows a schematic representation of the protein:ligand interactions (Figure 7.13). 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.
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.

### 7.8 Data capture for structure reports and PDB deposition

*MIFit must have access to a Python interpreter and a current version of the CCP4 software for this application to work (cf the **File/ Properties..** interface described in chapter 1 and 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 7.14) and executes an automated structure reporting script (*mi_deposit3d.py*).

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 a publication on an earlier version of the *mi_deposit3d* script (J. Badger *et al.*, Acta Cryst **F**61:818-820, 2005).

![Figure 7.14 Job/Report interface](image)

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 7.15). The text report contains a small number of key items and might be more useful for incorporating into (for example) a PowerPoint slide.
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 **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

This chapter describes the method used by MIFit to interact with external crystallographic programs (CCP4 and SHELX). 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 are launched from the MIFit Job menu. Please note that the automated structure solution applications are only available in the enhanced (automation) version of MIFit.

8.1 The Python automation scripts

Most if the external crystallographic applications that are interfaced to the MIFit system are controlled via a set of Python scripts. These scripts completely encapsulate and automate the setup and execution of third party software (predominantly from the CCP4 distribution). The required input for a scripted application is a small keyword (parameter) file that may be generated through the MIFit GUI or by editing a previous example.

By placing all of the program-specific control over the third party application into Python scripts

- It is simple to repair the execution of the MIFit application (for example, if the underlying CCP4 application evolves) without requiring changes in the compiled code in the MIFit program
- Customization and enhancements to the execution of these applications may be rapidly prototyped and implemented at individual MIFit sites
- Tasks for which the execution time is excessive within the context of an interactive model-building program may be run externally from MIFit
- The automated co-crystal structure solution process may be launched by triggering a script (mi_bng.py) in which a MIFit session for examining the results is launched by the script rather than vice versa

The scripted applications may be run in a variety of ways (Figure 8.1) including the traditional MIFit GUI, in standalone modes and within the framework of another automated structure solution system.
Although readily moved or duplicated, the Python scripts in the MIFit installation are subject to the same licensing regulations as the MIFit program. The scripts may be used freely at sites that hold valid MIFit licenses but may not be redistributed or used elsewhere.

### 8.2 MIFit access to a Python interpreter

In order to run a Python script, MIFit must be able to access a Python interpreter. Computers that run the LINUX operating system generally contain a Python installation as part of their operating software but it is necessary to obtain and install Python on computers that use the Windows operating system.

To set or check that your MIFit installation is connected to a Python interpreter look at the MIFit **File/ Properties...** menu (Figure 8.2).
This menu contains a **Python Interpreter** parameter which should be directed to the Python executable in the Python installation. The automation scripts within MIFit were developed using Python 2.0 and should be compatible with subsequent versions of Python.

### 8.3 Accessing the CCP4 suite

MIFit uses the CCP4 software suite as a basis for SAD phase determination, molecular replacement and structure refinement. The Python scripts have been written to circumvent troublesome issues with the CCP4 software that arise when path names on the Windows operating system contain spaces.

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 the scripts test one of the CCP4 environment variables to see if the CCP4 environment is established. In standalone operating modes a warning message will be printed if the environment does not appear to be established. If a script is launched via the MIFit GUI and terminates almost immediately it is a good idea to check the log file accessible from the Job List control in the top left of the document tree pane for messages that might indicate the cause of failure (see section 8.5 for more information on the Job List control).

The current set of scripted applications were developed and tested using CCP4 5.0.2. There is some potential that future releases of the CCP4 software might break the current applications or contain new features and modified protocols that it would be desirable to include. As mentioned before, an advantage of using Python scripts rather than compiled MIFit code to run the CCP4 software is that it is easier to modify the scripts to handle changes in emerging new versions of the CCP4 codes.

8.4 Accessing SHELX

MIFit uses the SHELX package to find anomalous scatterer sites for SAD phasing and to provide an alternative refinement option.

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 may create an environment variable, $SHELXBIN, that is the path to the directory that contains the SHELX executables. This may be a convenient method for LINUX users. Alternatively, the Python script variables shelx_dir_windows and shelx_dir_linux in the mi_sad.py (SAD phasing) and the mi_refine.py (structure refinement) scripts may simply be edited to point to that directory for Windows and LINUX respectively. This method should be used for a personal computing environment using Windows.

8.5 The Job List menu

When an external application is executed you will see a small text port appear for the duration of the job (Figure 8.3).
In order to manage external jobs the **MIFit Objects** list in the document tree pane on the left of the MIFit interface contains a **Job List**. Selecting the ‘+’ on the left of the **Job List** expands the list to contain (in run order) 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** and **Clean All**.

The most 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.
8.6 Script catalogue

Table 8.1 lists the Python scripts available in the `/data/Scripts` directory. Most of these scripts were developed for the automated (enhanced) version of MIFit; those scripts that are available in the basic version of MIFit are highlighted in blue.

<table>
<thead>
<tr>
<th>Script file</th>
<th>Parameter file</th>
<th>MIFIT GUI</th>
</tr>
</thead>
<tbody>
<tr>
<td>mi_integrate.py</td>
<td>mi_runintegrate.txt</td>
<td>No</td>
</tr>
<tr>
<td>mi_dataprep.py</td>
<td>mi_rundataprep.txt</td>
<td>No</td>
</tr>
<tr>
<td>mi_sadphase.py</td>
<td>mi_sadphase.txt</td>
<td>Yes</td>
</tr>
<tr>
<td>mi_molrep.py</td>
<td>mi_runmr.txt</td>
<td>Yes</td>
</tr>
<tr>
<td>mi_refine.py</td>
<td>mi_runrefine.txt</td>
<td>Yes</td>
</tr>
<tr>
<td>mi_restraints.py</td>
<td>mi_runrestraints.txt</td>
<td>Yes</td>
</tr>
<tr>
<td>mi_convertlib.py</td>
<td>mi_runconvertlib.txt</td>
<td>Yes</td>
</tr>
<tr>
<td>mi_bn.py</td>
<td>mi_runbng.txt</td>
<td>Yes</td>
</tr>
<tr>
<td>mi_deposit3d.py</td>
<td>mi_rundeposit3d.txt</td>
<td>Yes</td>
</tr>
<tr>
<td>mi_ligandoverlap.py</td>
<td>mi_ligandoverlap.txt</td>
<td>Yes</td>
</tr>
<tr>
<td>mi_lsqligands.py</td>
<td>mi_lsqligands.txt</td>
<td>No</td>
</tr>
<tr>
<td>smilesdb.py</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td>mi_imager.py</td>
<td>mi_runimager.txt</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 8.1 Python scripts for automated crystallography
The purposes of these scripts are summarized as follows:

**mi_integrate.py**
Runs D*TREK for automated (i.e. offline and non-graphical) image integration and data merging.

**mi_dataprep.py**
Runs CCP4 programs to reduce integrated intensity data (from CCP4/SCALA, D*TREK or SCALEPACK) to amplitude data and assign $R_{\text{free}}$ flags with options to reassign the space group, re-index data to a standard reference and retain $R_{\text{free}}$ flags.

**mi_sadphase.py**
Runs SHELXD to find the anomalous scatterer sites, CCP4/MLPHARE to refine the site locations and CCP4/DM to refine SAD phases.

**mi_molrep.py**
Runs CCP4/MOLREP with options to fix a protein fragment, for running a set of MR trials from multiple starting models and for repositioning the resulting model via symmetry operations back onto the input model.

**mi_refine.py**
Runs SHELX or CCP4/REFMAC5 with options to allow rigid-body refinement, perform water-picking, change the number of refinement cycles, set the x-ray weight, select the type of temperature factor refinement and perform a 6D search to place a ligand.

**mi_restraints.py**
Checks a coordinate file and aggregates all required ligand dictionaries, makes any special covalent restraints and outputs a single restraint dictionary for CCP4/REFMAC5.

**mi_convertlib.py**
Converts a CCP4/REFMAC5 mmCIF restraint dictionary to a format suitable for use with SHELX or CNS/CNX.

**mi_bng.py**
A wrapper script for automating cocrystal structure solution by running data preparation, molecular replacement and refinement processes and subsequently preparing files for MIFit,(optionally) launching the MIFit session.

**mi_deposit3d.py**
Calculates and parses structure data to produce a mmCIF annotation file that may be used to deposit structure data to the Protein Data Bank (up-
loading the file through the RCSB ADIT interface) or as a basis for creating internal structure reports in HTML or text formats.

**mi_ligandoverlap.py**

Runs CCP4/LSQKAB to automatically superimpose the target sites of multiple related coordinate files (*i.e.* coordinates with identical sequence numbering) and provide a single file that contains superimposed ligands from all structures.

**mi_lsqligands.py**

Runs CCP4/LSQKAB to overlap ligand conformers and write all conformers within a given rms deviation of the target.

**smilesdb.py**

This file is a dummy script with the function of accepting a small molecule registry number from the Ligand Dictionary Editor interface and returning the SMILES string for that ligand to the Ligand Dictionary Editor. Replacing this script with a script of the same name with code for querying a corporate database allows MIFit to interact with this information.

**mi_imagery.py**

Launches MIFit with view and style parameters taken from a previous session (mlw) file. This application may be used to generate a series of related sessions and session files, for example, to build a gallery of similar images of a ligand in density.

Most of the applications listed in Figure 8.4 may be run through the MIFit GUI as applications under the **Job** pulldown menu. In a few cases scripts do not need an individual interface since they serve as part of a larger process or an interface is not appropriate. The application **mi_bng.py** is a special case of a script that combines several scripts into a single process for automated co-crystal structure determination (see Chapter 10).

As indicated in Figure 8.1 and Table 8.1, the role of the MIFit GUI in running scripted applications is simply to write the parameter file needed to run the script and to execute the scripted application. Parameter files are typically just a few lines long and contain basic information such as the paths to input data. Examples of parameter files for all of the Python scripts are located in the ../data/Scripts directory of your MIFit installation.

To execute a script in standalone mode the parameter file should be present at the point of execution and the operating system must have access to a Python interpreter. On the Windows operating system double-clicking on the Python script icon with a parameter file in the same folder will exe-
execute the script. On LINUX/UNIX operating systems the command `python 'script-name'` with the parameter file in working directory will execute the script.

### 8.7 Automation script keywords

Table 8.2 list of all of the keywords used in the Python scripted automation system. Note that all keywords are parsed as upper case text and start and terminate with an `*`.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Description</th>
<th>Scripts applied</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>BOTHHANDS</em></td>
<td>Option with ‘yes’ or ‘no’ value to phase with input sites and inverted site constellation</td>
<td>mi_sadphase.py</td>
</tr>
<tr>
<td><em>BUILD CYCLES</em></td>
<td>The number of atom addition and deletion cycles. Note that each build cycle is placed between two refinement cycles and that this keyword is incompatible with water picking.</td>
<td>mi_refine.py</td>
</tr>
<tr>
<td><em>BREF TYPE</em></td>
<td>A flag (‘isotropic’ or ‘anisotropic’) to set the type of temperature factor refinement</td>
<td>mi_refine.py</td>
</tr>
<tr>
<td><em>BEAM MASK</em></td>
<td>Full path to a beam mask file for the detector. May be set to ‘rigakuccd’ for a Rigaku CCD detector.</td>
<td>mi_integrate.py</td>
</tr>
<tr>
<td><em>CIF IN</em></td>
<td>Full path to CIF library file to convert to SHELX or CNS/CNX format</td>
<td>mi_convertlib.py</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Module</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td><em>CYCLES</em></td>
<td>Number of refinement cycles for a single run (typically 5-20)</td>
<td>mi_refine.py</td>
</tr>
<tr>
<td><em>DATALOGFILE</em></td>
<td>Full path to log file from data merging (SCALA, SCALEPACK or D*TREK)</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>DICTIONARY</em></td>
<td>Full path to MIFit dictionary</td>
<td>mi_lsqligands.py</td>
</tr>
<tr>
<td><em>FRAGCENTER</em></td>
<td>Approximate XYZ coordinates in angstroms for the expected ligand binding site (set to ‘none’ for no ligand placement)</td>
<td>mi_refine.py, mi_bng.py</td>
</tr>
<tr>
<td><em>FRAGFITFILE</em></td>
<td>Full path to a coordinate file containing a ligand for a 6D search (set to ‘none’ if not applicable)</td>
<td>mi_refine.py, mi_bng.py</td>
</tr>
<tr>
<td><em>FIRSTIMAGE</em></td>
<td>Image number for the first image to process (may be omitted or set to ‘none’ for automatic determination of the first image)</td>
<td>mi_integrate.py</td>
</tr>
<tr>
<td><em>GROUPNO</em></td>
<td>Option to set a group number id for data processing output files (omit or set to ‘none’ for automated processing)</td>
<td>mi_integrate.py</td>
</tr>
<tr>
<td><em>HEADERIMAGE</em></td>
<td>Option to supply full path to a special header image to use</td>
<td>mi_integrate.py</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Module(s)</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td><em>HKLIN</em></td>
<td>Full path to integrate intensity file from SCALA, D*TREK or SCALEPACK</td>
<td>mi_dataprep.py, mi_bng.py, mi_imager.py</td>
</tr>
<tr>
<td><em>IMAGE1</em></td>
<td>Option to full path for an image file to include in an HTML report (set to ‘none’ to disable)</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>IMAGE2</em></td>
<td>Option to full path for an image file to include in an HTML report (set to ‘none’ to disable)</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>IMAGE3</em></td>
<td>Option to full path for an image file to include in an HTML report (set to ‘none’ to disable)</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>INTEGRATE_RES_RANGE</em></td>
<td>Resolution range (two numbers) for data processing (omit or set to ‘none’ for automated determination)</td>
<td>mi_integrate.py</td>
</tr>
<tr>
<td><em>MATCHPDBIN</em></td>
<td>Option with ‘yes’ or ‘no’ values to match the MR solution by symmetry operators to the input model location</td>
<td>mi_molrep.py</td>
</tr>
<tr>
<td><em>LASTIMAGE</em></td>
<td>Image number for last image to process (may be omitted)</td>
<td>mi_integrate.py</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Command</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td><em>LIBFILE</em></td>
<td>Full path to a ligand library file for REFMAC5 or SHELX (set to ‘none’ if not needed)</td>
<td>mi_refine.py, mi_bng.py, mi_deposit3d.py</td>
</tr>
<tr>
<td><em>LIGANDCHAIN</em></td>
<td>May be set to a ligand chain id to discount consideration of the entity as a ligand</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>MAPTYPE</em></td>
<td>Set to ‘normal’, ‘difference’ or ‘omit’. The ‘normal’ and ‘difference’ modes display ML-weighted maps and assume data from a previous REFMAC5 run. The ‘omit’ mode eliminates ligands from the map</td>
<td>mi_imager.py</td>
</tr>
<tr>
<td><em>MATCHPDBIN</em></td>
<td>Option with ‘yes’ and ‘no’ settings to apply symmetry operations to match the MR result to the position of the input model</td>
<td>mi_molrep.py</td>
</tr>
<tr>
<td><em>MAXRES</em></td>
<td>Maximum resolution for refinement data (omit or set to ‘none’ to use all data)</td>
<td>mi_refine.py</td>
</tr>
<tr>
<td><em>MERGE_RES_RANGE</em></td>
<td>Resolution range (two values) to restrict the resolution of data for merging (omit or set to ‘none’)</td>
<td>mi_integrate.py</td>
</tr>
<tr>
<td>Option</td>
<td>Description</td>
<td>File(s)</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td><em>MLWREFERENCE</em></td>
<td>Full path to reference session file to establish viewpoint, drawing styles, density levels and colors</td>
<td>mi_imager.py</td>
</tr>
<tr>
<td><em>MULTIPDBIN</em></td>
<td>Option with 'yes' or 'no' values to run MR on all models in the <em>PDBIN</em> directory</td>
<td>mi_molrep.py</td>
</tr>
<tr>
<td><em>MIFIT</em></td>
<td>Option that may be set to 'yes' to launch MIFit at the end of the process</td>
<td>mi_bng.py</td>
</tr>
<tr>
<td><em>MTZIN</em></td>
<td>Full path to input data in MTZ format</td>
<td>mi_refine.py, mi_molrep.py, my_deposit3d.py</td>
</tr>
<tr>
<td><em>ONLYMERGE</em></td>
<td>Option with 'yes' or 'no' values to restrict data processing to merging of integrated data (omit or set to 'no' for automated processing)</td>
<td>mi_integrate.py</td>
</tr>
<tr>
<td><em>OUTROOTNAME</em></td>
<td>Option to supply a root file name for outputs from this script (root <em>pbdde-posit</em> is applied if omitted)</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>PDBDIR</em></td>
<td>Full path to directory containing pdb files to analyze</td>
<td>mi_ligandoverlap.py</td>
</tr>
<tr>
<td><em>PDBFIXED</em></td>
<td>Full path to a fixed protein molecule (set to 'none' for no fixed</td>
<td>mi_molrep.py</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Relevant Scripts</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td><em>PDBIN</em></td>
<td>Full path to input coordinates in pdb format</td>
<td>mi_refine.py, mi_molrep.py, mi_restraints.py, mi_bng.py, mi_imager.py, mi_deposit3d.py</td>
</tr>
<tr>
<td><em>REFERENCEMTZ</em></td>
<td>Full path to a reference mtz file for assigning R-free data and reindexing (set to 'none' if no reference)</td>
<td>mi_dataprep.py, mi_bng.py</td>
</tr>
<tr>
<td><em>REFERENCEPDB</em></td>
<td>Full path to pdb file of reference ligand</td>
<td>mi_lsqligands.py</td>
</tr>
<tr>
<td><em>RMSD</em></td>
<td>Maximum allowed deviation of ligand conformer from reference for coordinate capture (typically 1.0)</td>
<td>mi_lsqligands.py</td>
</tr>
<tr>
<td><em>REFTYPE</em></td>
<td>Set to 'shelx', 'repmac5' or 'rigid-body' to select the type of refinement</td>
<td>mi_refine.py</td>
</tr>
<tr>
<td><em>REFPROGRAM</em></td>
<td>Set to 'shelx', 'cns' or 'cnx' to define format for output library file</td>
<td>mi_convertlib.py</td>
</tr>
<tr>
<td><em>SEPARATION</em></td>
<td>Minimum allowed separation for anomalous scatterers for site finder</td>
<td>mi_sadphase.py</td>
</tr>
<tr>
<td><em>SITEFILE</em></td>
<td>Full path to file containing positions of anomalous scatterers in PDB (pdb) or</td>
<td>mi_sadphase.py</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Module</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td><em>SITENUMBER</em></td>
<td>Maximum number of anomalous scatterer sites to find and process</td>
<td>mi_sadphase.py</td>
</tr>
<tr>
<td><em>SOLVENTFRACTION</em></td>
<td>Solvent fraction</td>
<td>mi_sadphase.py</td>
</tr>
<tr>
<td><em>SSNUMBER</em></td>
<td>Number of disulfides amongst anomalous scatterer sites</td>
<td>mi_sadphase.py</td>
</tr>
<tr>
<td><em>SEQFILE</em></td>
<td>Full path to FASTA sequence file for the protein (set to ‘none’ if not used)</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>SPACEGROUP_NO</em></td>
<td>Space group number</td>
<td>mi_integrate.py, mi_dataprep.py, mi_bng.py</td>
</tr>
<tr>
<td><em>TARGETCHAIN</em></td>
<td>Chain id for ligand in <em>TARGETPDB</em> file</td>
<td>mi_lsqligands.py</td>
</tr>
<tr>
<td><em>TARGETENTITY</em></td>
<td>Entity id for ligand to compare</td>
<td>mi_lsqligands.py</td>
</tr>
<tr>
<td><em>TARGETRESNO</em></td>
<td>Residue id for ligand in <em>TARGETPDB</em> file</td>
<td>mi_lsqligands.py</td>
</tr>
<tr>
<td><em>TARGETPDB</em></td>
<td>Full path to the reference pdb file</td>
<td>mi_ligandoverlap.py</td>
</tr>
<tr>
<td><em>TARGETSITE</em></td>
<td>The XYZ coordinates in angstroms defining the target point in <em>TARGETPDB</em></td>
<td>mi_ligandoverlap.py</td>
</tr>
<tr>
<td><em>TEMPLATEFILE</em></td>
<td>Full path to template file (see deposit3d.template) for adding non-electronic annotation (set to ‘none’ if not)</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td>Key</td>
<td>Description</td>
<td>Script</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td><em>TEMPLATEIMAGE</em></td>
<td>Full path to either the directory containing images or a template image from the directory (usually the first image)</td>
<td>mi_integrate.py</td>
</tr>
<tr>
<td><em>TITLE</em></td>
<td>Title text for HTML report</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>WATERPICKCycles</em></td>
<td>The number of water picking cycles. Note that each water cycle is placed between refinement cycles</td>
<td>mi_refine.py</td>
</tr>
<tr>
<td><em>WEIGHT</em></td>
<td>Weight of x-ray term in refinement (usually 0.1-0.5)</td>
<td>mi_refine.py</td>
</tr>
<tr>
<td><em>WORKDIR</em></td>
<td>Full path to directory where calculations will be performed</td>
<td>all scripts</td>
</tr>
<tr>
<td><em>WRITECIF</em></td>
<td>A 'yes' or 'no' option to write mmCIF data for RCSB PDB submission</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>WRITEHTML</em></td>
<td>A 'yes' or 'no' option to write selected information in HTML format</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>WRITEHKL</em></td>
<td>A 'yes' or 'no' option to write reflection data in mmCIF format</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td><em>WRITETEXT</em></td>
<td>A 'yes' or 'no' option to write selected information in text format</td>
<td>mi_deposit3d.py</td>
</tr>
<tr>
<td>mat</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8.2 Keywords used by automation scripts
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). The SAD phasing interface runs a Python script as an intermediary for controlling the phasing applications. Chapter 8 describes the steps that need to be taken to access Python, CCP4 and SHELX installations from MIFit.

9.2 SAD phasing interface

The SAD phasing interface (Figure 9.1) is available via the Job/ SAD Phasing command.

![Figure 9.1 Job/ SAD Phasing interface](image)

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 **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 9.2). 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.
Overall results from the phasing process are reported in the HTML file `mi_phase_summary.html`. Figure 9.3 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 are merit are greater than 0.3 across several resolution shells, indicating that there is relatively weak but still significant phasing power.
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 MIFit.
10 Automated Structure Solution

The applications described in this chapter are only available in the enhanced (automation) version of MIFit.

10.1 Overview of automated co-crystal structure determination

An automated structure determination process has been developed 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 (script mi_bng.py) 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), 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 co-crystal 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.

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 refinement 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.

**6D Ligand Search**

Optionally, the process may carry out an accurate 6D search 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 expectation for the conformation of the ligand (for example, when working on a series of related molecules). If a small number ( <10 ) of ligand conformations are supplied with the input coordinate file, separated by TER records, then all of these conformations will be evaluated versus the density map. The best scoring conformation will be selected for addition to the model.

**10.2 The automated co-crystal structure determination interface**

The **Job/Cocrystal Solution** interface (Figure 10.1) automates co-crystal structure solution.
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.

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

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.

The Spacegroup Number parameter is rarely used (co-crystal data are usually processed in the correct, known space group) but may be used to change the space group number within the limits of the point group.

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.

![BNG Run Summary](image)

Figure 10.2 Example of HTML summary containing just one job

The Dictionary parameter may be used to provide any CCP4/REFMAC5 dictionary that might be required to run refinement on the input model.
The **Place Ligand (pdb)** parameter and associated X, Y, Z parameters may be used to perform a rigorous and complete 6D search to fit the ligand coordinates to density near the specified point in angstrom units. This operation is purely a density docking process (it adds the molecule to the protein but does not attempt to refine it) so no dictionary is required for the ligand. The input coordinates may correspond to a single molecules or contain a small number of conformers (< 10) separated by TER records in the coordinate file.

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 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.

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

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 crystal file (*bng_crystal*) corresponding to the new data set is created in the working directory.

In addition, phased reflection files *bng_mlmap.fcf* and *bng_mlmap.phs*, containing likelihood-weighted coefficients in 'fcf' (SHELX) and 'phs' (XtalView) formats are always created for subsequent use. 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. (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*.

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/Molecular Replacement interface

The molecular replacement application (*mi_molrep.py*) may be run using the **Job/Molecular Replacement** command (Figure 10.3). This application employs the CCP4/MOLREP program to carry out molecular replacement calculations involving complete rotation-translation searches.

![Molecular Replacement interface](image)

**Figure 10.3 Job/Molecular Replacement interface**

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.

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.

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.

### 10.4 The Job/Refinement interface

Refinement jobs utilize the script `mi_refine.py` and may be run using the **Job/Refinement** command (Figure 10.4).

![Figure 10.4 Job/Refinement interface](image)

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

The **Ligand (pdb)** parameter may be used to provide a coordinate file for a 6D ligand search directed at the site specified by the **X, Y, Z** parameters. At present this option is mainly useful for rigid ligands are those in which the likely ligand conformation is known prior to the refinement (for example, when working with a series of very similar ligands or when fitting a known ligand into sites related by non-crystallographic symmetry).

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 ‘\texttt{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 \texttt{refine\_#.mlw} is created.

10.5 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 (\textit{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 \texttt{Job/Set Refmac5 restraints} command (which runs \texttt{mi_restraints.py}) is available (Figure 10.5). 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 \texttt{.cif}. For example, if the structure contains a novel ligand ‘MMM’ there should be dictionary file available called \texttt{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 \texttt{restraints\_#.cif}, where \# is a serial number reflecting the number of times this command was executed in the current working directory.
10.6 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 10.6). 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 book-keeping 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.
10.7 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 10.7). For structure refinements carried out through the Job/Refinement interface, this file appears in a browser upon completion of the refinement.

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 percentage 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.
10.8 Automated ligand fitting

A fully automated application is available via the **Job/Refinement** and the **Job/Cocrystal Solution** commands for placing a copy of a ligand (or selecting from an ensemble with fewer than 10 conformers) 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, working processes in which an external program is available to generate a small number of major conformations 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/FFFEAR program to carry out a complete rotation-translation search using the ligand difference density as the tar-
get. 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 3.0-2.4Å 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).

10.9 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 map data should normally be better than 2.5Å resolution for the shape of the density to 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 more difficult to fit than small rigid examples; molecules that contain more than 40 non-hydrogen atoms will probably be too large to fit. 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 10.8.

![Figure 10.8 Model/Add residue dialog window.](image)

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’. 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 Refi/Find Ligand Fit and Conformer command. The process will terminate when either the maximum allowed number of generations is reached in the Genetic Algorithm or the fit score reached a ‘good enough’. For ‘good’ densities in ~2Å resolution maps final scores will often be in the range of 7-8. 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.

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-building, the commands
Refi/Rigid-Body Refine Current Atoms (performs a rotation-translation refinement of the entire ligand) and Refi/Find Best Conformer (checks conformers after superimposing upon the current ligand model) may be used to improve the fit to the density map. Sometimes a succession of these commands is needed to ‘toggle in’ the ligand to correctly fit the electron density.

10.10 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 Cα atoms within 15Å of an active site coordinate within a reference structure.

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

![Cocrystal Superposition interface](image)

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 10.10 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

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.. command to load a coordinate file in PDB format. Set the Files of type pulldown filter to ‘Model Fitting Session (*.pdb)’. 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 using the left mouse button while holding down the Shift key.

5. Try panning the model by dragging with right mouse button in the canvas window.

6. Notice that if the cursor is in the extreme top of the main canvas that the function of the left mouse button changes to rotate the view about the Z-axis and the function of the right mouse button controls the slab.

7. Try clicking on the slab icons in the tool bar to slab 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 sequence window on the right.

10. Try single-clicking on an atom in the main canvas. Notice that this atom is placed in the stack on the left. The stack is used to identify targets for commands that operate on specific atoms or residues.

11. 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 using the Method pulldown selection 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 **Show/ Color** options.

12. Try changing the rendering style by selecting the different options from the **Render** pulldown menu. Many people prefer to use the **Sticks** option for speed and simplicity when model fitting.

13. Try the command **View/ Center model on screen** to zoom and slab the model automatically to fill the screen.

14. Try changing the model to a backbone only representation with **Show/ CA Backbone**. 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.

15. Try the command **Show/ 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.

16. Try changing the ribbon colors by selecting the **Show/ 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 colored this way.

17. You may select **File/ Close** or **File/ Exit** to close this session or shut down MIFit.
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 **MIFit File/Open..** command to load the file `bng_milaunch.mlw` from the directory (1) in which it was placed. If this was the first time that you have ever loaded this structure you will be prompted to add crystal parameters to the MIFit repository of known crystals. Select **OK** to accept the defaults for these prompts. You should see the electron density map appear around the protein model in the main canvas.

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. Try changing the size and contour levels by selecting the **Map/Contour Options...** command. Move the **Radius** scrollbar to increase the size of the display density from a 7Å box to a 9Å box. From the pulldown menu in the **Preset Map Styles** list box select the first option - 'Blue map 1, 2, 3, 4 5 sigma' – and click on **OK**. You should see the map volume expand slightly but no change in the display colors and contour levels as this is the default choice for standard electron density maps. 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 **Map/Contour Options...** command. Try turning off contour levels 2 and 4 by unchecking 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 **Map/FFT Phases...** command, 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 **Map/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. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.
Lesson 3: Loading crystal data and displaying symmetry related molecules

This lesson demonstrates how to supply MIFit with cell and spacegroup information for a structure and how to display crystallographic symmetry.

1. Copy the file automation_out.pdb from the examples directory of your MIFit installation into a new directory. This file is a coordinate file.

2. In order to display crystallographic symmetry (for example, to check whether parts of a structure make contacts with other molecules in the crystal cell) and to compute electron density maps for some input data formats (for example, the XtalView .phs format) it is necessary to create and load a file that contain crystal cell, space group and symmetry information. Note that some data formats (for example, the .mtz and the .fcf formats) contain cell and space group data and if these data are already loaded the molecular symmetry will be correctly assigned without needing to refer to a correct crystal file. To see how to set up a crystal file for a new model, click on File/Crystal….

3. The Crystal parameter contains a pulldown menu that allows you to select from previously created crystal files. You may also edit this parameter to generate a new crystal file. To see how to establish a new crystal file, enter a new crystal file name in this field (for example, sym_tutorial).

4. Enter a crystal title (for example, test) in the Title field. In the Unit-cell field enter the cell dimensions ‘65.419 89.227 99.842 90.00 90.00 90.00’. In the Spacegroup field enter either the space group symbol (I222) or the space group number (23). Click on Find and you should see that the symmetry operators for this space group are loaded.

5. Select OK to write this crystal file into the MIFit directory of known crystals and to load this crystal as the default in the MIFit objects in the tree on the left of the interface.

6. Select File/ Open.. to load the coordinate file automation_out.pdb. i.e. use the Look in parameter to browse for the directory defined in (1), set Files of type to ‘Model Fitting Session (*.pdb)’ and use the File name parameter to load this file.

7. Click on the zoom out icon (▲) a few times so that most of the molecule is visible.

8. Select Model/ Generate Symmetry. When you do this the ‘Set Crystal’ menu will appear to confirm that the correct crystal data is loaded – select OK in this menu. You will see symmetry related atoms appear in the main canvas in purple. Note that only symmetry related atoms that are close to the view of the main canvas are generated. If you
zoom out further or change the center point then the symmetry related atoms will be generated for the new view. Try using **Model/Clear Symmetry** to remove the symmetry related molecules from the display.

9. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.
Lesson 4: 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… command. Change the Files of type pulldown menu to the ‘Model Fitting session (*.pdb)’ selection and use the Look in browser parameter to enter file automation_misfit.pdb from the directory established in (1). Click Open to load the coordinate file.

3. Select the MIFit Map/Load Phase file command. In the popup dialog box the Files of type parameter should be set to ‘CCP4 MTZ (*.mtz)’. Use the Look in browser option to enter file automation_out.mtz from the directory established in (1). Click Open to load the phased data file.

4. A pop-up dialog box 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. To do this, select FWT for the $F_O$ column and PHWT for the $\Phi$ column parameters and click on OK.

5. In the popup dialog box ‘Fast Fourier Transform Map’ click on OK. This action will compute and display the electron density map, which will appear in the main canvas.

6. Expand the model data to chains and segments by clicking on the model icon in the tree of the left. Click on the ‘Chain X’ icon to expand it. Scroll down to residue X 142 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.

7. 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.
8. 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.

9. Try rotating the residue by either clicking on the **Fit/ Rotate** command or the icon in the toolbar. Using the right mouse button you can now rotate the selected residue around the canvas.

10. Reset the position of the amino acid back to the starting position but leaving it in fitting mode by selecting the **Fit/ Reset** command.

11. 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.

12. 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.

13. In a real fitting session you would have accepted this fit either by selecting the **Fit/ Apply** command, by clicking on the icon on the toolbar or by using the keyboard shortcut ';'. Here, select **Fit/ Reset** so that the lesson can continue to show other fitting methods. Select the (cancel fit) icon to deactivate this residue.

14. 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.

15. 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. Go to the 'Chain X' icon in the tree on the left of the MIFit interface. Scroll down 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 electron density.

16. 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.
17. You will now use a pentamer library to check a part of the protein structure. Go to the ‘Chain X’ icon in the tree on the left of the MIFit interface. Scroll down to residue X 192 and double click on it. This action will re-center the model and map on the Cα atom in Glu X 192 Glu.

18. 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. 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.

19. You may select File/ Close or File/ Exit to close this session or shut down MIFit.
Lesson 5: SAD phasing

(This tutorial requires that MIFit has access to a Python interpreter, 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 the checkbox should remain selected and the option to Change spacegroup number should not be invoked.

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 27-42Å². In this case the density modification protein residual from phasing in hand 2 is significantly lower than in hand 1 so this is more likely to be the correct solution (in a genuinely new example it is wise to inspect both maps).
8. Click on the **Job List** icon near the top of the navigation tree to expand it. Then 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.

9. The simplest way to load maps for visual inspection is to make use of the pre-computed session files. Use the **File/Open..** command to load session file `mi_sad_i.mlw` into MIFit (file `mi_sad.mlw` corresponds to phasing with hand 1, the original site constellation). You may be prompted to load establish a new crystal by a ‘Set Crystal’ dialog box if this is the first time you have worked with this data. You may **Cancel** this request.

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.

<table>
<thead>
<tr>
<th>Res. (Å)</th>
<th>No. phased data</th>
<th>FOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.13</td>
<td>163</td>
<td>0.3140</td>
</tr>
<tr>
<td>6.22</td>
<td>457</td>
<td>0.3539</td>
</tr>
<tr>
<td>4.71</td>
<td>904</td>
<td>0.3521</td>
</tr>
<tr>
<td>3.79</td>
<td>1511</td>
<td>0.3092</td>
</tr>
<tr>
<td>3.17</td>
<td>2227</td>
<td>0.2601</td>
</tr>
<tr>
<td>2.73</td>
<td>3081</td>
<td>0.2397</td>
</tr>
<tr>
<td>2.39</td>
<td>4074</td>
<td>0.1634</td>
</tr>
<tr>
<td>2.13</td>
<td>2886</td>
<td>0.0492</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Site</th>
<th>Atom</th>
<th>Occ.</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S</td>
<td>1.198</td>
<td>32.494</td>
</tr>
<tr>
<td>2</td>
<td>S</td>
<td>1.152</td>
<td>33.481</td>
</tr>
<tr>
<td>3</td>
<td>S</td>
<td>1.130</td>
<td>30.638</td>
</tr>
<tr>
<td>4</td>
<td>S</td>
<td>1.060</td>
<td>37.011</td>
</tr>
<tr>
<td>5</td>
<td>S</td>
<td>0.919</td>
<td>26.255</td>
</tr>
<tr>
<td>6</td>
<td>S</td>
<td>1.006</td>
<td>41.659</td>
</tr>
<tr>
<td>7</td>
<td>S</td>
<td>1.125</td>
<td>35.347</td>
</tr>
<tr>
<td>8</td>
<td>S</td>
<td>0.955</td>
<td>32.594</td>
</tr>
<tr>
<td>9</td>
<td>S</td>
<td>0.981</td>
<td>26.757</td>
</tr>
<tr>
<td>10</td>
<td>S</td>
<td>0.975</td>
<td>27.181</td>
</tr>
</tbody>
</table>

**Occupancies and B-factors after site refinement**
Lesson 6: 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 Cα 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 MIFit command File/Open.. to load coordinate file start_trace.pdb from the directory in which you placed it. Remember that you can use the Look in: pulldown selection to locate the directory you established in step (1) and set Files of type.. to ‘Model Fitting Session (*.pdb)’ to find PDB coordinate files. The two-marker structure should appear in the main canvas.

3. Select the MIFit Map/Load Phase file command. In the popup dialog box the Files of type parameter should be set to ‘CCP4 MTZ (*.mtz)’. Use the Look in browser option to locate file automation_out.mtz from the directory established in (1). Select Open to load the phased data file.

4. In the resulting dialog box select FWT for the Fo column setting and PHWT for the Phi column setting. Select OK. You will be fitting to a map computed from these previously prepared Fourier coefficients.

5. Select OK in the resulting Fast Fourier Transform Map dialog box. You should now see in the main canvas a small region of map around a two atom start point for model-building.

6. The region of the map that is displayed by default is rather small for model building. Select the Map/Contour Options command. In the resulting dialog move the Radius slider to about 10 in order to expand the view of the map. You may also wish to 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 expand.
7. Select the molecule icon in the tree in the left of the canvas to expand the description into molecular segments. Click on segment \( X.1 \) to expand it and double-click on entity \( MRK \, 150 \). These actions will center the display on \( \text{Ca} \) marker \( X \, 150 \).

8. Click on the (centered) MRK 150 entity in MIFit main canvas and then select the \textbf{Model/ Add MRK After} command. You should see a green line appearing to connect that marker to the expected position for the next \( \text{Ca} \) atom.

9. By continuing to select \textbf{Model/ ADD MRK After} or by using the keyboard short cut ‘SHFT>’ trace out four more \( \text{Ca} \) positions.

10. Select \textbf{Refi/ Accept Refine} and \textbf{Fit/ Apply} to accept this trace.

11. Select \textbf{Model/ Poly-Ala Chain} to convert this chain trace to poly-alanine.

12. The C-terminal residue is not fully converted to ALA and may be deleted by clicking on one of its atoms in the main canvas. Then right-click in the main canvas and select \textbf{Delete Residue} from the pop-up menu.

13. Select \textbf{Sequ/ Read Lower Sequence} and use the dialog box to enter the file \textit{start_sequence.seq}. You will notice that the sequence code TVITKHN has appeared in the sequence window.

14. Select atoms in residues X 149 and X 154 (\textit{i.e.} both ends of the poly-alanine fragment) and then select the \textbf{Model/ Replace with Sequence} option. After selecting \textbf{Yes} in the popup dialog you will see the poly-alanine fragment replaced by the correct amino acid sequence.

15. The fit is imperfect – as an optional exercise you could use some of the methods learned from tutorial 3 to refit individual amino acids.

16. You may select \textbf{File/ Close} or \textbf{File/ Exit} to close this session or shut down MIFit.
Lesson 7: 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. Select No from the resulting ‘Multiple conformers?’ dialog box. The generation of multiple conformers would be useful for automated ligand density fitting within MIFit but this step is not needed to generate ligand restraints for refinement.

4. 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 centers and one plane covering the aromatic ring and atoms immediately connected to it. Incorrect restraint assignments may be changed using the pulldown menus at the top of the Dictionary Editor window.

5. To write out the refinement dictionary, Select the Export to File.. button. Select Refmac mmCIF Dictionary File (*.cif) from the possible output options and click on OK.

6. In the resulting menu use the Save in pulldown selection to choose a directory in which the mmCIF restraint file will be saved and the File name field to enter a file name for the file. Click on Save to write out the file. 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.

7. 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 reopens the dictionary editor).
8. 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.
Lesson 8: 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.. to load coordinate file 1A28.pdb from the directory in which you placed it. Remember that you can use the Look in: pulldown selection to locate the directory you established in step (1) and set the Files of type.. parameter to 'Model Fitting Session (*.pdb)' to find PDB coordinate files. After completing these operations the structure will appear in the main canvas.

3. Click on the molecule icon in the tree in the left of the canvas to expand the description into molecular segments. Click on segment B.4 to expand it and double-click on entity STR 2. These actions will center the canvas display on the bound ligand STR 2.

4. Click on any atom in the STR entity in the main canvas.

5. Select either the Show/Dot Surface/ Surface Residue command or the Object/Surface/ Surface Residue command. Click on the OK button in the resulting dialog box. You should see a van der Waals dot surface appear around the ligand.

6. Remove this display by selecting either the Object/Surface/ Clear Surface command or by right-clicking on the Surface Dots entry in the tree on the left and selecting Delete.

7. We will now investigate the accessibility of the ligand binding cavity by constructing a solvent accessible surface. First, delete the STR molecule by clicking on any atom in it, right-clicking anywhere in the main MIFit canvas and selecting Delete Residue.

8. Either right-click on the MIFit main canvas, selecting Surface/ Solvent Exposed Surface or select Object/ Surface/ Solvent Exposed Surface. Click on the OK button in the popup dialog box. 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.

9. 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.
10. You may select **File/ Close** or **File/ Exit** to close this session or shut down MIFit.
Lesson 9: 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.. to load coordinate file 1A28.pdb from the directory in which you placed it. Remember that you can use the Look in: pulldown selection to locate the directory you established in step (1) and set Files of type.. to 'Model Fitting Session (*.pdb)' to find PDB coordinate files. The structure will appear in the main canvas.

3. Click on the molecule icon in the tree in the left of the canvas to expand the description into molecular segments. Click on segment B.4 to expand it and double-click on entity 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 viewpoint used for the schematic plot.

4. Click on any atom in the ligand in the main MIFit canvas.

5. Select the Model/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.
Lesson 10: Creating reports for PDB deposition and publication

(This tutorial requires that MIFit has access to a Python interpreter and 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 ..data/Scripts/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. Se-
lect the **Write mmCIF** data checkbox to generate a set of structure factor data in mmCIF format.

5. Select the **Run** command. 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. Click on the **Job List** icon near the top of the navigation tree. Then 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 report generation process.

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.

8. 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/](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.

9. You may select **File/ Close** or **File/ Exit** to close this session or shut down MIFit.
Structure determination and model refinement statistics

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

- Unit cell (Å, °): 65.419 89.227 99.842 90.000 90.000 90.000
- Space group: 1222
- Resolution range (Å): 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/σ(I): 27.8 (4.7)
- Rmerge: 0.038 (0.174)

Crystallographic data and refinement statistics

- Resolution range (Å): 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
- Rwork: 0.2105
- Rpkee: 0.2452
- Rmsd bond lengths (Å): 0.008
- Rmsd bond angles (°): 1.063
- Number of disallowed ψψ angles: 0
The following tutorials relate to applications that are only available in enhanced (automation) version of MiFit

Lesson 11: Automated co-crystal structure determination

(This tutorial requires that MiFit has access to a Python interpreter and 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 co-crystal 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**. 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. Click on the **Job List** icon near the top of the navigation tree. Then 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 solution process.

9. Use the **File/Open..** command to load the session file `bng_milaunch.mlw` into MIFit. (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. *i.e.* 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.
Lesson 12: Automated Molecular Replacement

(This tutorial requires that MIFit has access to a Python interpreter and 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. Select Run. 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 running Windows.
7. Click on the **Job List** icon near the top of the navigation tree. Then 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 molecular replacement process.

8. Use the **File/Open..** 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 to fit the diffraction data and that these were automatically packed together.

9. 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.

10. 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. To do this you need to establish and load the correct crystal for this structure by selecting **File/Crystal..**

11. In the **Crystal** parameter enter `mr_demo` as the crystal file name. For **Title** enter the text, ‘mr test’. For **Unitcell** enter the cell parameters ‘58.123 64.444 69.954 90.0 95.741 90.0’. For **Spacegroup** enter the space group name ‘p21’ and click on the **Find** button to load the correct symmetry operations. Then click on **OK** to create this crystal file.

12. You may confirm that this crystal is loaded by right-clicking on the crystal icon associated with this model and selecting **Edit**. If it were not just loaded you would use the pulldown menu associated with the **Crystal** parameter to select crystal file `mr_demo`.

13. Select **Model/ Generate Symmetry** 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.

14. 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.

15. You may select **File/Close** or **File/Exit** to close this session or shut down MIFit.
Lesson 13: Automated refinement

(This tutorial requires that MIFit has access to a Python interpreter and 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 use) 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. 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 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.. command. See lessons 2 and 3). Although metrics for identifying structure misfits are somewhat fuzzy it is interesting to look through the amino acids in the associated error list.

141
8. If you like, you can click on the **Job List** icon near the top of the navigation tree. Then 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 refinement process.

9. You may select **File/ Close** or **File/ Exit** to close this session or shut down MIFit.
Lesson 14: 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. Since we are doing ligand fitting (rather than just establishing a ligand parameters), respond to the ‘Multiple Conformers?’ dialog box with Yes. You should see the Ligand Dictionary Editor appear.

3. If you like, you can confirm that the ligand restraints were correctly defined (cf chapter 6) and click on the Conformer option a few times to confirm that multiple conformers were established. Click on Save and Exit to load the ligand into the dictionary.

4. Next, load the protein model using the File/Open.. dialog box to locate file ligand_example.pdb from (1). Do not forget to set Files of type parameter to ‘Model Fitting Session (*.pdb)’.

5. Now load the pre-computed difference map coefficients in the MTZ file output from a refinement run with REFMAC5. Select Map/Load Phase File and use the dialog box to load file ligand_example.mtz. Set Fo column to DELFWT and Phi column to PHDELFWT and select OK. (If this tutorial was never run before you will be prompted by the ‘Set Crystal’ dialog box. You may select Cancel in this box if you do not wish to preserve this crystal file.) In the resulting ‘Fast Fourier Transform Map’ dialog box select OK.

6. Change the map volume and contours to more appropriate values for displaying a difference map by selecting Map/Contour Options.. Increase the Radius to 10Å and the Preset Map Styles option to ‘Difference Map -4,-3,3,4,5’. Select OK to display the map with these settings.
7. Use the left (rotate) and right (translate) mouse buttons to make sure that the large density feature near A 163 is centered on the cross hairs.

8. 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. Select **OK** to add the ligand to the model.

9. 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.

10. Select the **Refi/ 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 uses an genetic algorithm fully reproducible results are not obtained and occasionally the search will result in a plausible misfit.

11. You may sometimes be able to optimize the fit by (in turn) selecting the **Refi/ Rigid-Body Refine Current Atoms** option and the **Refi/ Find Best Conformer** 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.

12. Once you have achieved a satisfactory fit you may accept the fit by clicking on the ✔ icon from the toolbar.

13. You may select **File/ Close** or **File/ Exit** to close this session or shut down MIFit.
Lesson 15: Superimposing ligands from multiple related structures

(This tutorial requires that MIFit has access to a Python interpreter and 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.. to load coordinate file 1d4h.pdb from the directory in which you placed it. Remember that you can use the Look in: pulldown selection to locate the directory you established in step (1) and set the Files of type.. parameter to 'Model Fitting Session (*.pdb)' to find PDB coordinate files. The structure will appear in the main canvas.

4. Click on the molecule icon in the tree in the left of the canvas to expand the description into molecular segments. Click on segment _.3 to expand it and double-click on entity BEH 501. 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 CO1, 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. 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. Click on the **Job List** icon near the top of the navigation tree. Then 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.

9. You may select **File/ Close** or **File/ Exit** to close this session or shut down MIFit.
Appendix 1: 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 script mi_bng.py 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. For MIFit users who were also XtalView users it will be noticed that the script language is essentially the same except that a number of commands are no longer supported in MIFit (they are just ignored if they are present).

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.

Command \[val1\] [val2]

# - comment

Model Commands

**LoadPDB** \(n\) pathname

Load the PDB/mmCIF 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.
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

**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.

**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`. For example `[1 0 0 0 1 0 0 0 1]`.

Prerequisites: an open window

**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 command).

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.

Note: the old xfit script method of a contourlevels with a single number is still supported for backwards compatibility.
Prerequisites: loaded map

**Contourradius** \(\nu\)

Set the radius (half-width) of the area to be contoured to \(\nu\).
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** \( d_{max} \)

Set the minimum resolution (outer resolution) of the map to \( d_{max} \).

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 \( P2_12_12_1 \).

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

**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 2: Default Colors

Colors 1-10 (first row, Basic Colors)
Colors 11-20 (second row, User Colors)
Colors 21-30 (third row, Map Colors)
Colors 31-40 (fourth row, Contour Level Colors)

<table>
<thead>
<tr>
<th>n</th>
<th>rgb</th>
<th>Internal Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 0 0</td>
<td>Black</td>
</tr>
<tr>
<td>1</td>
<td>255 255 20</td>
<td>Yellow</td>
</tr>
<tr>
<td>2</td>
<td>255 20 20</td>
<td>Red</td>
</tr>
<tr>
<td>3</td>
<td>20 20 255</td>
<td>Blue</td>
</tr>
<tr>
<td>4</td>
<td>20 255 255</td>
<td>Cyan</td>
</tr>
<tr>
<td>5</td>
<td>20 255 20</td>
<td>Green</td>
</tr>
<tr>
<td>6</td>
<td>255 20 255</td>
<td>Magenta</td>
</tr>
<tr>
<td>7</td>
<td>255 128 128</td>
<td>Pink</td>
</tr>
<tr>
<td>8</td>
<td>255 128 0</td>
<td>Orange</td>
</tr>
<tr>
<td>9</td>
<td>0 0 0</td>
<td>Brown</td>
</tr>
<tr>
<td>10</td>
<td>253 253 253</td>
<td>White</td>
</tr>
<tr>
<td>11</td>
<td>255 72 96</td>
<td>User 1</td>
</tr>
<tr>
<td>12</td>
<td>90 255 90</td>
<td>User 2</td>
</tr>
<tr>
<td>13</td>
<td>102 164 255</td>
<td>User 3</td>
</tr>
</tbody>
</table>
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 
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   0  Level 6 
37  255 200   0  Level 7 
38  245 141   3  Level 8 
39  255   60   0  Level 9 
40  220   0   0  Level 10