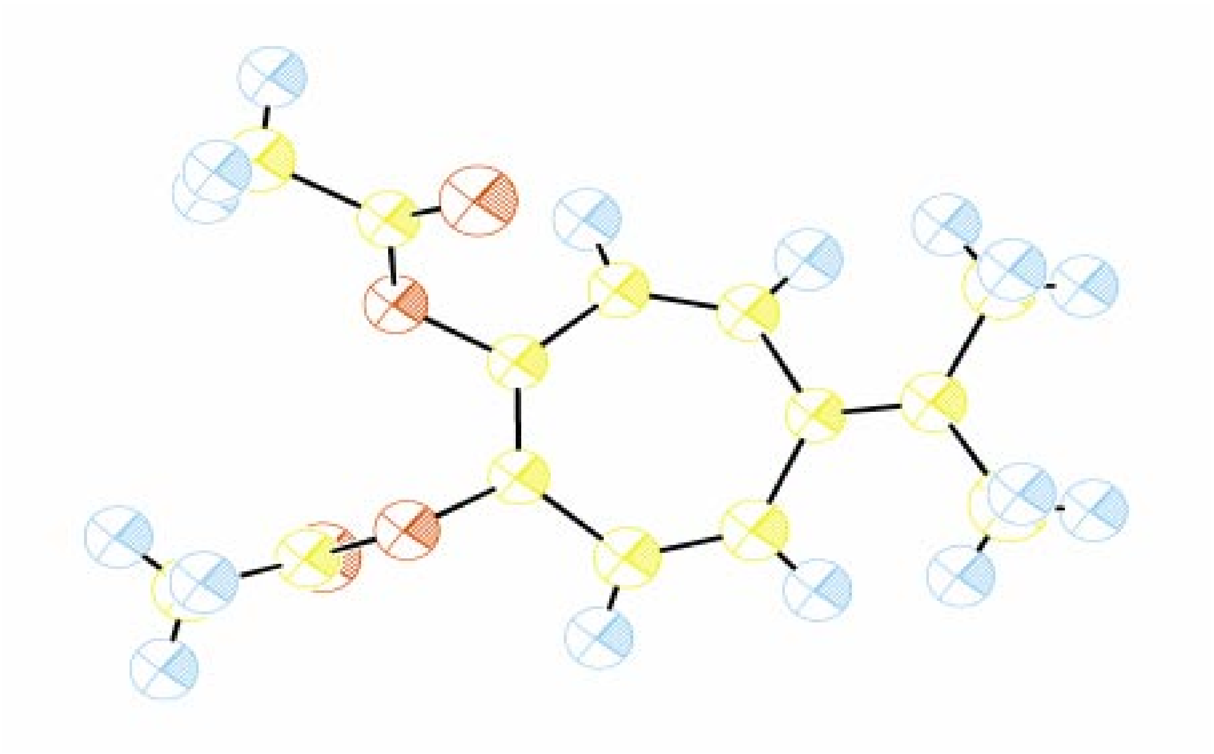
Material
Analysis and
Characterization

Nonius

maXus

Comprehensive Crystallography Software



18 February 1999

Table of Contents

1 Introduction to maXus 1-1

2 The File Menu 2-1

3 Importing Data 3-1

1 Introduction to maXus

1.1 Credits

maXus is a product principally of

Stuart Mackay, Wei Dong, Chris Edwards, Allan Henderson, Chris Gilmore,
Neil Stewart, Kenneth Shankland and Arlene Donald,
of
Chemistry Department,
The University of Glasgow,
Glasgow, Scotland.

for

MAC Science Co. Ltd.
Nakamachidai 3-12-1,
Tsuzuki-ku
Yokohama-shi 224, Japan

and

Nonius B.V.
Röntgenweg 1
P.O. Box 811
2600 AV Delft
The Netherlands

Version 3.2.1 Manual updated by
Gordon Barr

MITHRIL is Based on the Mithril90 Direct Methods Package by
Chris Gilmore, Stephen Brown and Allan Henderson

SIR92 Principal authors**:**
A. ALTOMARE, G. CASCARANO, C. GIACOVAZZO & A. GUAGLIARDI
Ist. di Ric. per lo Sviluppo di Metodologie Cristallografiche, CNR, c/o Dip. Geomineralogico - Univ. of Bari
BURLA M.C. & POLIDORI G.
Dip. Scienze della Terra - Univ. of Perugia
CAMALLI M.
Ist. Strutt. Chimica CNR Monterotondo stazione - Roma
(Thanks are also due to M. Luic, A. Nunzi, R. Spagna, I. Vickovic & D. Viterbo for their valuable contribution to the previous releases.)

DIRDIF by
P. T. Beurskens, G. Admiraal, G. Beurskens, W.P. Bosman, S. Garcia- Granda, R.O. Gould, J.M.M. Smits, and C. Smykalla.
Crystallography Laboratory, University of Nijmegen,
Toernooiveld, 6525 ED Nijmegen, The Netherlands.

RasMol was written by:
Roger Sayle
Biomolecular Structure Department
Glaxo Research and Development
Greenford, Middlesex, UK.
Copyright © 1992,1993,1994 by Roger Sayle
(rasmol@ggr.co.uk)
RasMol is a free product which is shipped with maXus without charge. Please consult the online RasMol documentation for more details. License details are also in rasmol.LICENSE in the manuals directory.

ORTEP was written by
Johnson, C.K. (1965). Report ORNL-3794.
Oak Ridge National Laboratory, Oak Ridge,
Tennessee, USA.

GX Interactive Ortep (1985) written by
C.J. Gilmore, P.R. Mallinson and K.W. Muir
Chemistry Department, Glasgow University,
Glasgow, Scotland

Portions of **ORTEP XGraphics and PostScript** based on code by Norimasa Yamazaki,
University of Electro-Communications,
Tokyo, Japan.

Original **PLOTQ** code by Keith Henderson and Colin Bannister, Chemistry Dept., The University, Glasgow

The **nedit** text editor used for editing and displaying text files is a free editor distributed under the terms detailed in the file “nedit.LICENSE” in the maXus manuals directory or the extras/nedit directory.

“This work was performed at Fermi National Accelerator Laboratory, operated by Universities Research Association, Inc., under contract DE-AC02-76CH03000 with the U.S. Department of Energy.” It was written by:
Mark Edel, Joy Kyriakopulos, Arnulfo Zepeda-Navratil,
Suresh Ravoor, Donna Reid, Jeff Kallenbach
Fermi National Accelerator Laboratory, m/s 234
P.O. Box 500
Batavia, IL 60510
edel@fnal.gov

Regular expression code by Henry Spencer,
University of Toronto

The license code in maXus makes use of the following, in the form of freely distributable implementations:
The **DES** code is Copyright © 1995-1997 Eric Young (eay@mincom.oz.au)
All rights reserved. Please see the file COPYRIGHT.DES in the manuals directory for details.
The RSA Data Security, Inc. **MD5 Message-Digest Algorithm** code is Copyright © 1991-2, RSA Data Security, Inc. Created 1991. All rights reserved. Please see the file COPYRIGHT.MD5 in the manuals directory for details.

ShelXS97 and ShelXL97 are produced by Professor George Sheldrick and not supplied with maXus. Details on obtaining and installing these programs for use in maXus are in the corresponding on-line manuals.

The Crystallographic Information File is described in the paper “The Crystallographic Information File (CIF): a New Standard Archive File for Crystallography” by S. R. Hall, F. H. Allen & I. D. Brown [Acta Cryst. (1991),
A45, 655-685], copies of which may be obtained from the Technical Editor,
IUCr, 5 Abbey Square, Chester CH1 2HU, England.

Copyright in the CIF specifications and Dictionary resides with the IUCr.

Simulated precession photographs are generated using the ’**pattern**’ program written by:
Guoguang Lu, PhD
Research Scientist
Molecular Structural Biology
Dept of Medical Biochemistry and Biophysics
Karolinska Institue
17177 Stockholm, Sweden

maXus logo designed by Gordon Barr

1.2 Support

We offer maXus users the highest levels of support.

In case of problems, or suggestions for improvements contact one of the following:

email: maXus@chem.gla.ac.uk

fax: (+141) 330 4419

In *Japan* you can contact:

MAC Science Co. Ltd.
1-5-1 Shin-Yokohama
Kohoku-ku
Yokohama-shi 222, Japan
Telephone: 81-45-473-6781
Fax: 81-45-473-1234

In *Europe* contact:

Nonius B.V
Rontgenweg 1
2624 BD Delft
The Netherlands
Telephone: 15 269 8300
Fax: 15 262 7401

In *North America* contact:

Enraf-Nonius Co.
1650 Sycamore Ave Suite 8
Bohemia, NY 11716
Phone 800-645-8379
Fax: 516 589 2068

or contact:

Prof.C.J.Gilmore
Department of Chemistry
University of Glasgow
Glasgow G12 8QQ, Scotland
Fax: +44 141 330 4419

1.3 Introduction

maXus is a powerful, state-of-the art computer program for solving, refining and publishing crystal structures automatically from X-ray diffraction data. It works in a windows environment with SUN, Silicon Graphics and Linux systems, and offers the following facilities:

- Spacegroup: A computer program to assign space groups, merge intensity data and correct for crystal decomposition. Non-standard space group settings are readily accommodated. Intensity histograms and simulated precession photograph graphics options also available.
- Test i: A program to test for centrosymmetric / non-centrosymmetric structures.
- Five programs to solve crystal structures automatically:
  - SIR
  - DIRDIF
  - MITHRIL
  - Monte-Carlo MULTAN
  - SHELXS 97 (interface only; program available from Prof. Sheldrick).
- An AUTO option to solve and refine crystal structures automatically, including H atom addition.
- A program, MODEL, to manipulate structures during the process of solution and refinement.
- A real-time, graphical least squares program, RTG-LSQ, that interactively refines crystal structures.
- A full graphical interface for the SHELXL97 least squares program
  (not supplied, but available from Prof. Sheldrick).
- Structure factor software.
- A multi-optional Fourier program, FOURIER.
- Six different absorption correction methods:
  - Numerical - with new OGL interface
  - Psi-Scan
  - Spherical
  - Sortav
  - Difabs
  - Cylindrical
- PUBLISH options for generating tables, CIF files and a report for publishing the final structure.
- Five different colour graphics options:
  - Interactive ORTEP III
  - PLUTO
  - RasMol
  - XXMOL for 3D colour graphics.

  Plus, a powerful new version of PLOTQ for viewing and output-ing 2 and 3-dimensional contoured Fourier maps.

All the options are fully described in the manual which follows. Using them, it is quite possible to completely refine a structure in less than one hour.

1.4 Current limits in maXus

1.4.1 Atoms

- Maximum number of atoms: 500.

1.4.2 Parameters

- Maximum number of refined parameters in full matrix least squares (before blocking): 2,000.
- Maximum number of refined parameters in block-diagonal least squares: 2,000.

1.4.3 Reflections

- Maximum number of reflections in least squares: No limit.
- Maximum number of reflections in Original file, before merging: 400,000.

1.4.4 Fourier

- Maximum number of reflections symmetry expanded in the Fourier program: 150,000.
- Maximum number of grid points along x axis: 200.
- Minimum Dynamic range of the map: 0.001. (The peak search will terminate if the difference between the maximum and minimum peak heights is less than this.)

1.5 Preparing to run maXus for the first time

maXus can deal with data from the following formats:

MacScience diffractometer
ASCII input
CAD4 diffractometer
KappaCCD (as a CIF file)
CIF files

To begin, use the **mkdir** command to set up a new, empty directory on your computer. e.g.

```
mkdir <directory name>
```

Whichever type of data file you have, copy it into the new directory. e.g.

```
cp <current file position> <new file position>
```

If this is a MacScience diffractometer data file, it should be called *Original*, and should not be edited.

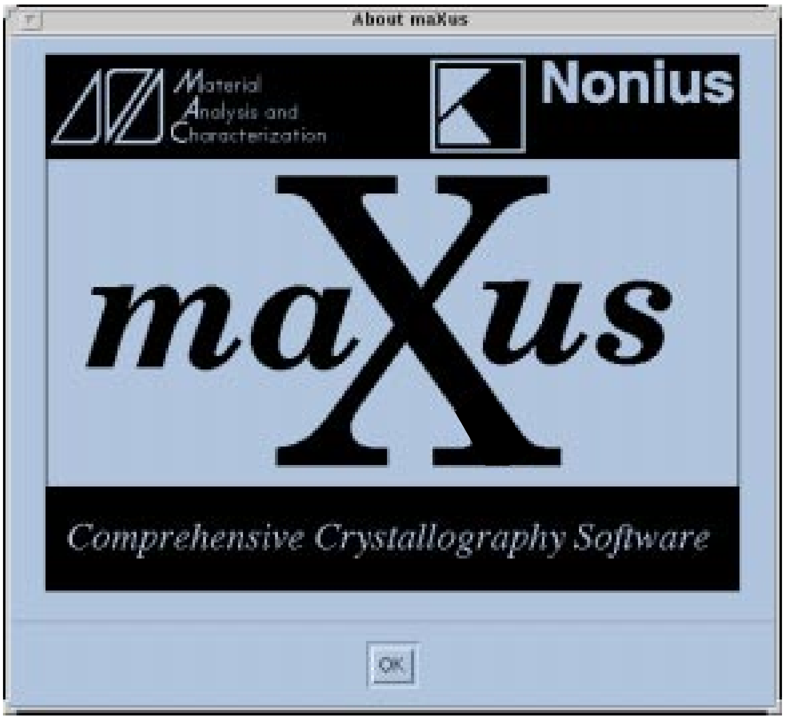
If you have one of the other types of data file, you will need to import it to a maXus format. This process is described in detail in Chapter 3. Imported files can have any name.

1.6 Running maXus

To start the maXus program, type **maxus** at the Unix command prompt.

It is strongly recommended that maXus is only called from the directory in which your data files reside.

The maXus logo will appear for a few seconds, before being automatically dismissed:



The core window will appear with your current directory displayed:

Current directory: /home/tcrystal/gbam/testdata/demo

The directory from which you launched the program is shown in the lower half of the window.

You are now ready to begin working with maXus.

2 The File Menu

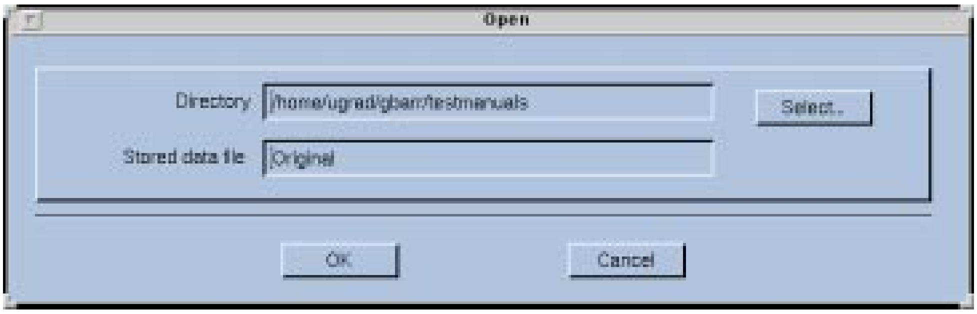
2.1 File Menu Options

Pull down the File menu (the left hand entry on the window). There are eight entries:

Each of these will now be discussed in detail.

2.1.1 File Menu: Open

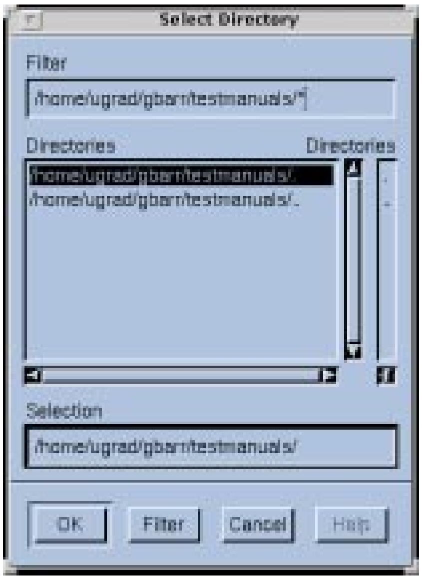
This selects the directory in which maXus runs. Click on this entry, and the Motif file browser appears:



Directory: /home/ugrad/gbarr/testmanuals

Stored data file: Original

Either enter the directory directly in the top box, or click *Select...* Another window will appear:



Filter: /home/ugrad/gbarr/testmanuals/*

/home/ugrad/gbarr/testmanuals/.
/home/ugrad/gbarr/testmanuals/..

Selection: /home/ugrad/gbarr/testmanuals/

Use this to search through your file store. Once you have selected the required directory, then click OK. You are now ready to use maXus.

If the directory chosen does not exist, another box will appear with the message:

Can't open directory

Click OK and try again.

2.1.2 File Menu: Import

This invokes the programs which import intensity data from the following sources:

- CAD4 diffractometer.
- ASCII from other sources.
- CIF from a suitable CIF file.
- Archive from a maXus cif-archive file that loads structures solved by maXus and stored in a suitable archive directory.

These options are described in full in Chapter 3 of this manual.

2.1.3 File Menu: History

This has two options:

- Display: This brings up a text pane, showing the history file (see Section 2.2.4) . This keeps a log of all of the modules and programs you run for a particular compound, and the date and time you run them. This can be very useful to keep track of and summarise the work done so far. The file is also used by the CIF generation program (see the Publishing chapter, Section 17.7)

  For example:

```
  History records:

            Date                    Program

Mon      Jun      21 09:57:40 BST 1999 Space group determination:  P 21/c
Mon      Jun      21 09:57:42 BST 1999 Sort-merge data:  FULL     R(merge) =    0.0306
Mon      Jun      21 09:59:55 BST 1999 Structure solution using  sir92
Mon      Jun      21 10:00:41 BST 1999 Least squares, maXus  on data uncorrected for absorption
                                       R =   0.0925 weighted R =   0.1482 No. of parameters =    29
Mon      Jun      21 10:08:30 BST 1999 Least squares, maXus  on data uncorrected for absorption
                                       R =   0.0774 weighted R =   0.1371 No. of parameters =    29
Mon      Jun      21 10:11:32 BST 1999 Fourier program, Max=    1.069 min =  -1.343 rms =    0.150
Mon      Jun      21 10:13:26 BST 1999 Geometry program.
```

A dialog box will then appear:

History output complete

Clicking *OK* will dismiss the text pane.

If the file is empty (because you have not worked with this compound before, or because the Reset option has been invoked (see below), you will be informed:

History file empty

- Reset

This option deletes all entries in the maXus.history file. A dialog box checks that you really mean to do this:

Delete all history info ?

Use this option with extreme caution; this file is vital to the process of generating CIF files for publishing your work (see Section 17.7).

2.1.4 File Menu: List HKL

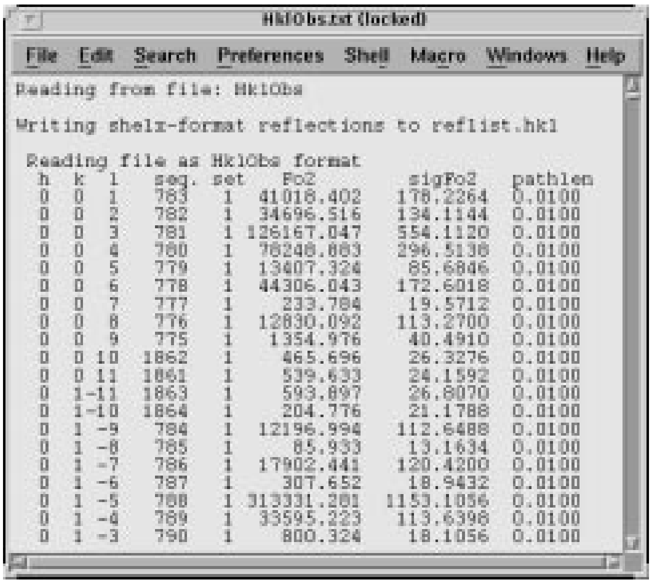
There are three submenus:

(1) Observed
(2) Calculated
(3) Corrected. This has further options:

Numerical
Difabs
Psi-scan
Spherical
Cylindrical

When running, maXus stores the processed observed and calculated structure factors in two files called HklObs and HklCal respectively. These are binary files and hard to read. Generally, the user does not need access to them, but on some occasions it can be useful to examine them.

The option *ListHKL* lists these files in a text window, for example:

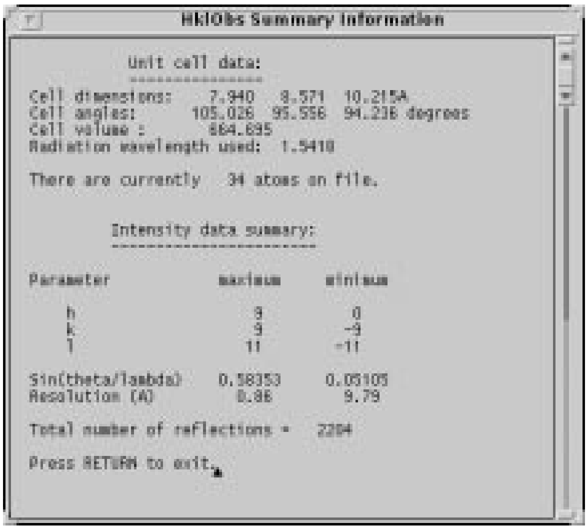


```
Reading from file: HklObs

Writing shelx-format reflections to reflist.hkl

 Reading file as HklObs format
  h   k   l    seq. set    Fo2          sigFo2    pathlen
  0   0   1    783   1   41018.402    178.2264   0.0100
  0   0   2    782   1   34696.516    134.1144   0.0100
  0   0   3    781   1  126167.047    554.1120   0.0100
  0   0   4    780   1   78248.883    296.5138   0.0100
  0   0   5    779   1   13407.324     85.6846   0.0100
  0   0   6    778   1   44306.043    172.6018   0.0100
  0   0   7    777   1     233.784     19.5712   0.0100
  0   0   8    776   1   12830.092    113.2700   0.0100
  0   0   9    775   1    1354.976     40.4910   0.0100
  0   0  10   1862   1     465.696     26.3276   0.0100
  0   0  11   1861   1     539.633     24.1592   0.0100
  0   1 -11   1863   1     593.897     26.8070   0.0100
  0   1 -10   1864   1     204.776     21.1788   0.0100
  0   1  -9    784   1   12196.994    112.6488   0.0100
  0   1  -8    785   1      85.933     13.1634   0.0100
  0   1  -7    786   1   17902.441    120.4200   0.0100
  0   1  -6    787   1     307.652     18.9432   0.0100
  0   1  -5    788   1  313331.281   1153.1056   0.0100
  0   1  -4    789   1   33595.223    113.6398   0.0100
  0   1  -3    790   1     800.324     18.1056   0.0100
```

The Numerical, Difabs, Psi-scan, Spherical, Cylindrical options list the binary files generated by the absorption correction programs (See Chapter 6 for details)

2.1.5 File Menu: HKL Info

This option produces a text pane with brief summary of the unit cell and intensity data:



```
          Unit cell data:
          ---------------
Cell dimensions:    7.940   8.571  10.215A
Cell angles:      105.026  95.556  94.236 degrees
Cell volume :       664.695
Radiation wavelength used:  1.5418

There are currently   34 atoms on file.


         Intensity data summary:
         -----------------------

Parameter            maximum    minimum

    h                    9          0
    k                    9         -9
    l                   11        -11

Sin(theta/lambda)    0.58353    0.05105
Resolution (A)         0.86       9.79

Total number of reflections =   2204

Press RETURN to exit.
```

Press the return key to dismiss the window.

2.1.6 File Menu: Unlock

When any of the maXus processes are running remaining programs in the suite are locked so that they cannot be run. This is to prevent problems caused by more than one program trying to write to the same file, and the chaos that can result. The only exceptions to this are the *Weight Analysis* and *Info* programs under the *Refine* menu - these bring up histograms that can be useful in designing a suitable least squares weighting scheme.

Sometimes the more experienced user may wish to unlock a current process so that something else can be run. Selecting *Unlock* will bring up a dialogue box:

Are you sure

Make your choice - but be careful!

2.1.7 File Menu: Tidy

The maXus program generates a large number of files. When the structure analysis is complete you may wish to reduce this to a bare minimum, and the *Tidy* facility offers this option to the user. It should be used with caution, and is fully documented in Chapter 19

2.1.8 File Menu: About

Select *About*. The following box, with support addresses appears:

maXus is a computer program for the solution and refinement
of crystal structures from X-ray diffraction data

MAC Science Co., Ltd.
Nakamachidai 3-12-1, Tsuzuki-ku
Yokohama 224
Japan

Fax: +81 45 4731234
Tel: +81 45 4736781

Nonius B.V.
Rontgenweg 1
P.O. Box 811, 2600 AV Delft
The Netherlands

Fax: +31 15 262 74 01
Tel: +31 15 269 83 00

Nonius
390 Central Ave
Bohemia, NY
11716, USA

Fax: +1-516-589-7053
Tel: +1-516-589-2885

For support contact your local representative or
E-Mail: maXus@chem.gla.ac.uk

```
License: using file /local/maxdev/lib/maXus.license
Current date is 31/Jul/1998
Hostid in file is: 0(0x0, 0)
maXus is running in time-limited DEMO mode
Expiry date in File is 31/Dec/1998
```

2.1.9 Exit

This exits the maXus program. It is the preferred way of leaving the system. You will be asked if you really meant to select this option:

Exit maXus ?

Exit quits the program, *Cancel* lets you change your mind.

2.2 Important Files

In general, the user does not need to be aware of most of the files that maXus uses, but some are mentioned in the menus from time to time and so need to be understood:

2.2.1 Original

This is your file from the diffractometer. See Chapter 3 for details of importing files.

The format of this file has changed slightly between maXus 2.0 and maXus 3.0 to allow an increase in the maximum number of reflections allowed to 400,000 (i.e. the format for the number of reflections is now I7, changed from I5). In theory, Original files from previous versions of maXus (in the old format) should still be able to be read in.

2.2.2 DrawFile

This is first written by the Spacegroup program and is continually updated. It contains the space group information, the cell dimensions, the atomic coordinates and the connectivity.

If you have used the expanded Model window option, another version of this file, (DrawFile.expand) will be present.

Do not edit these files.

2.2.3 Model

This is also first written by the Spacegroup program and is continually updated. It contains the space group information, the cell dimensions, the atomic coordinates, and is used by several programs as an alternative to the DrawFile.

If you have used the expanded Model window option, another version of this file, (Model.expand) will be present.

Do not edit these files.

2.2.4 maXus.history

This history file is saved in the working directory, and keeps track of all of the programs you have run and the order you have run them in. It is added to every time you run part of the maXus suite of programs. Its contents may be viewed using the *File Menu / History / Display* option (Section 2.1.3).

It is used in particular by the CIF generation programs to construct an accurate map of which programs in maXus have been used in which order. This information is required to produce accurate CIF files for publication.

Do not edit this file.

3 Importing Data

from
KappaCCD,
CAD4 DIFFRACTOMETER,
CIF FILES,
ASCII TEXT &
ARCHIVE FILES

3.1 Introduction to Importing Files to maXus

maXus is designed to import and process data from a wide variety of sources. This chapter details the different routes of importing this data, which are as follows:

1. Data sets that are written in the standard CIF format, including data from the KappaCCD. This will allow the importing of reflection information, and thus allow the structure solution using maXus. This provides a route for reading data from any diffractometer *via* the Flack program. It will also allow the importing of atomic information and the subsequent analysis of molecular geometry.
2. CAD4 datasets. These are produced by Nonius diffractometers.
3. Datasets that are in ASCII format. This allows formatted reflection data to be input from any source. We envisage this being used to input X-ray powder data or electron diffraction data into the maXus suite of programmes.
4. Archive data sets stored using the *Publish/Archive* option.

3.2 Importing CIF Files (including KappaCCD)

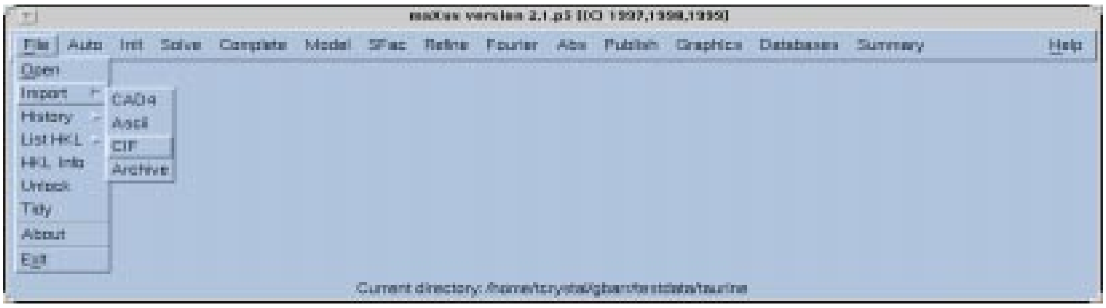
If you have either atomic or reflection data in a CIF format then by using the Import CIF option, it is possible to import the data into maXus. Atomic data can be imported and then treated as a refined structure. It can be displayed and manipulated in the Model window, and geometric calculations can be performed on the structure.

Note: It is not possible to import both Atomic and Reflection data simultaneously, as refinement or structure solution would destroy the atomic co-ordinates that were imported. This limitation will be removed in a future release of maXus.

The Import CIF makes an intelligent attempt to calculate a value for Z if neither the crystal density or a value for Z are included in the CIF file (this can be altered when the SpaceGroup program is run if incorrect).

3.3 How to Use CIF - Import

From the main menu:

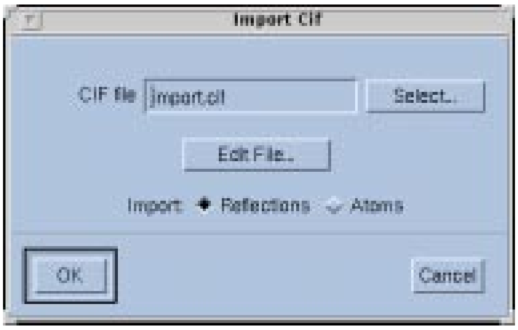


Choose the *File* option, followed by *Import*, followed by *CIF.*

This will call up the CIF Import window.

3.3.1 The CIF Window

The following window will appear:

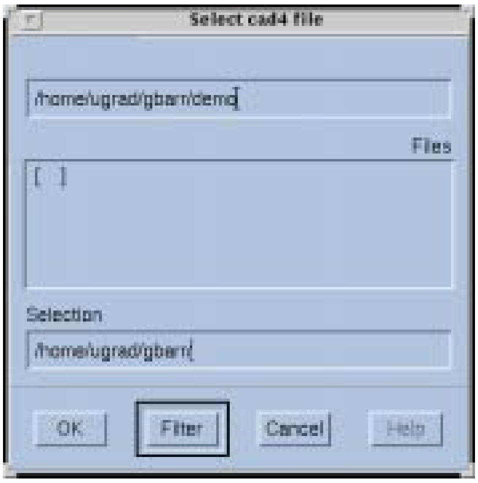


CIF file: import.cif

Import: ◆ Reflections ◇ Atoms

The field labelled CIF File is a text box and should the name of the CIF file that your are going to import should be entered here. **NOTE: This box is case sensitive.**

To change directory, or to browse through your files, click *Select.* to get a file browser:



Select cad4 file

/home/ugrad/gbarr/demo

Files

[ ]

Selection

/home/ugrad/gbarr/

Once you have chosen the file you want, click *OK*.

If the CIF file you are importing is in an older CIF format than the expected one (V. 2.1 of the CIF dictionary), you will be given the chance to convert it:

maXus

Old style cif found. Convert it to new format?

Clicking *OK* will save a copy of your original CIF file in a file called *import.cif.oldformat*. Note that this is **not** listed in the KeepFile, and so would be deleted by a *Tidy* operation.

Clicking on *Edit File* opens the new version data file named, and allows you to view or edit it as required..

The buttons labelled *REFLECTIONS* and *ATOMS* are mutually exclusive and represent whether you are intending to import atomic or reflection information. The future behaviour of maXus is dependent on which selection you make at this stage.

Once the file name and import type are selected click on the *OK* button and the CIF-Import will begin.

If you wish to exit this window without further processing use the *CANCEL* button in preference to exiting the window using the X-window control bar.

3.4 Import CIF - Reflections Option

3.4.1 Reflection - Data Fields Required

To import the reflection data, certain information is required within the CIF file. While some fields are required, sensible defaults are supplied for those which are not present in your CIF file.

A list of all the reflection field names, a description of what they mean, and their default settings can be found in Appendix C.

3.4.2 Entering the crystal morphology

Data describing the indexed crystal faces can be included in the CIF file in the form:

```
_loop
_exptl_crystal_face_index_h
_exptl_crystal_face_index_k
_exptl_crystal_face_index_l
_exptl_crystal_face_per_dist
0 0 1 0.12
1 -1 2 0.23
```

The importer will saved this data in a file called FaceDat, which will be automatically read by the NUMABS absorption correction program. See Chapter 6.

3.5 CIF Import - Atoms Option

3.5.1 Atoms - Data Fields Required

To import the atom data certain information is required within the CIF file. While some fields are required, sensible defaults are supplied for those which are not present in your CIF file.

A list of all the Atoms field names, a description of what they mean, and their default settings can be found in Appendix C.

3.5.2 Atoms - Further Processing

After you have chosen the atoms option from the main CIF - Import window and clicked on the OK button, the following window will appear warning you that the functionality of maXus will be limited from this point on.

After the CIF file containing the atomic information has been imported, only a limited number of functions within maXus will work. These functions are:

All modules accessing reflection data are now inactive

1. The model window will allow the molecule to be displayed and also allow interactive molecular geometry queries to be made.
2. The molecule can also be displayed using the following, all of which are under the Graphics option on the main menu bar:

- ORTEP
- XXMol
- RasMol
- Pluto

3. Geometrical calculations can be performed on the molecule using the following tools, all of which are under the Publish option on the main menu bar:

- Geometry
- Platon

4. You can also export the atomic information you have imported from the CIF file as CSSR format using the CSSR option, which is under the Publish option on the main menu bar.

No other functionality of maXus will be available

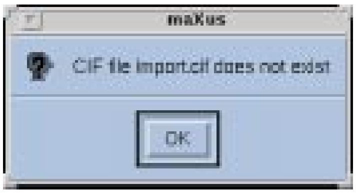
3.6 Other CIF Import Windows

3.6.1 The CIF Import Performance Meter

While the CIF Import is running a performance meter will appear at the top of the screen. When CIF Import has completed, the performance meter will be closed. It is possible to close this window while CIF Import is running. The performance meter window is shown below:

3.6.2 Error Dialogs Associated with CIF - Import

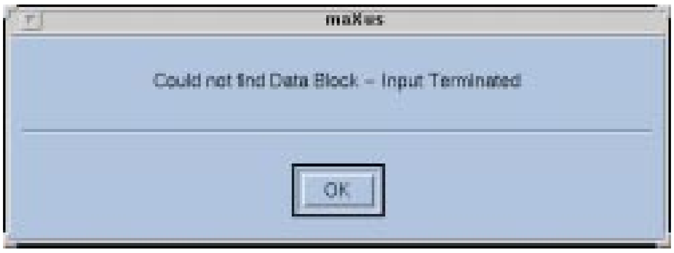
If the file entered in the CIF - Import main window does not exist, then the following error window will appear:



CIF file import.cif does not exist

Click on OK, processing will stop and the CIF - Import window can be called up again in the normal manner.

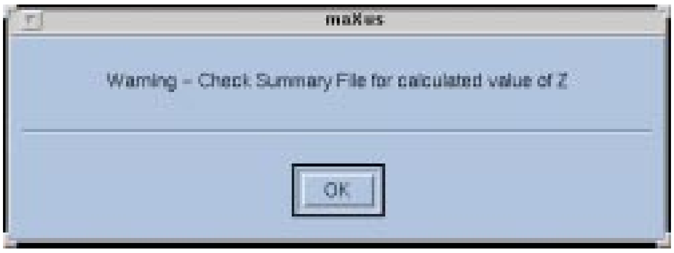
If your CIF file does not contain a valid data block then the following window will appear:



Could not find Data Block – Input Terminated

Click on OK, processing will stop and you must examine and correct the CIF file before attempting to import the file again. Check the CIF Import summary file as this will supply information regarding the file name that it has found the fault in.

If the value for the number of molecules in the unit cell that is imported and the value calculated based on unit cell parameters and cell contents differ, then the following window will appear:



Warning – Check Summary File for calculated value of Z

Click on the OK box. Processing will continue normally and the value of Z imported from the CIF file will be used in subsequent processing. Check the CIF - Import summary file at the end of processing where more information regarding this problem will be available.

If any problems are detected when trying to calculate the Laue group then the following window will appear:

Problem Determining Laue Group – Check Summary File

Click on the OK box. Processing will continue. Check the CIF - Import summary file at the end of processing where more information regarding this problem will be available. In general, problems with Z can be ignored at this stage since the Spacegroup program will compute a good value based on atomic volumes (see Chapter 5).

If a problem is encountered while decoding an imported symmetry position in xyz format then the following window will appear:

Error in Decoding Symmetry position – Import Terminated

Click on the OK box and processing will terminate. Check the CIF - Import summary file (Section 3.7) at the end of processing where more information regarding this problem will be available. Correct the symmetry operation in the CIF file prior to attempting to re-import the file.

If the CIF dictionary is not installed in the correctly then the following error window will appear:

Could not find CIF Dictionary – Import Terminated

Click on the OK box and processing will terminate. Contact the person that installed maXus at your site and inform them of this problem.

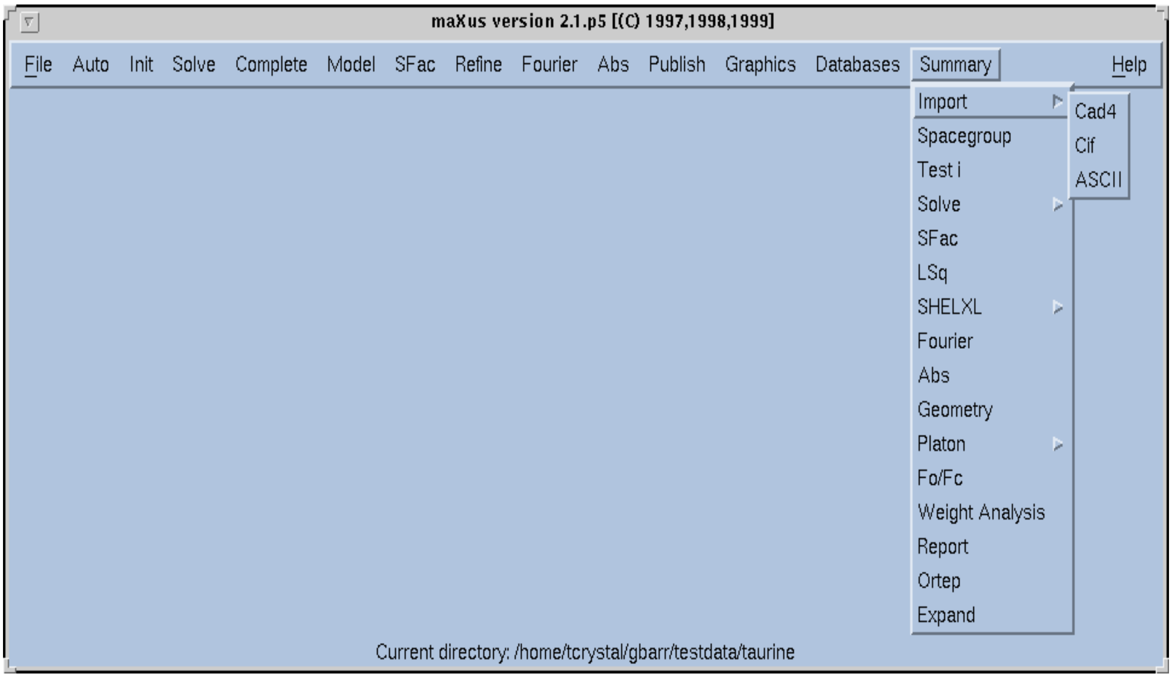
If an internal problem to maXus has occurred and the user interface cannot interact with the main maXus CIF - Import routines the following window will appear:

Could not identify import type - Import Terminated

If this occurs please inform your system manager of the problem.

3.7 The CIF Import Summary File

To access the summary file return to the main maXus window.



Current directory: /home/tcrystal/gbarr/testdata/taurine

Choose the *Summary* option, followed by *Import / CIF.*

The first section of the summary file will detail all data names in the CIF file that do not conform to the CIF dictionary that is installed along with your version of maXus.

The next line of text informs you of the name of the data block that the CIF importer will take its data from. It is important than no more than one data block is included per file, and the second and subsequent data blocks in the CIF file will be ignored.

The information that follows lists what defaults have been applied to which parameters. Much of this will be redundant information e.g. setting an isotropic temperature factor to an atom that has anisotropic values. Most of the messages are simply warnings.

See Appendix B for an example of a typical CIF - Import Summary File.

3.8 Importing CAD4 Datasets

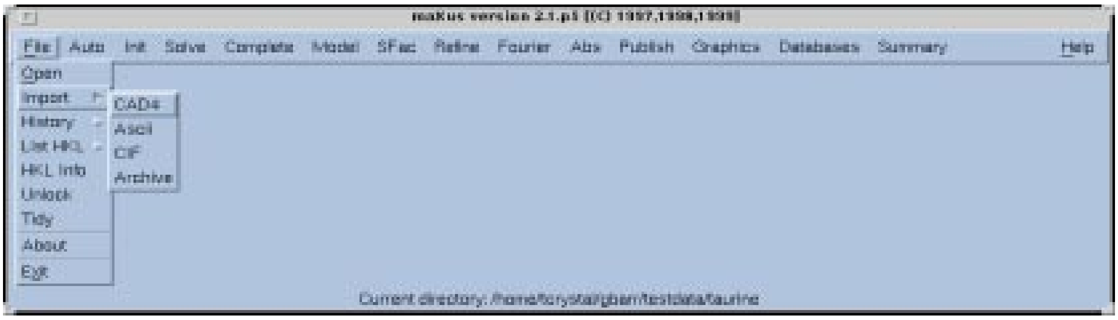
If you are using a CAD4 diffractometer you must import the file to maXus using the Import-CAD4 program. This program performs the following functions:

- Creates a maXus Original file for further processing. An Original file is a core maXus datafile that contains basic information and all reflection data;
- Captures information concerning the experimental set up for the CIF file;
- Performs all data reduction of CAD4 profiled and non profiled reflections.
- Creates a DCosines file containing the direction cosines for the absorption correction programs - see Chapter 6.
- Makes an intelligent attempt to calculate a value for Z if neither the crystal density or a value for Z are input in the GUI window(this can be altered when the SpaceGroup program is run if incorrect).

It is not possible to proceed with maXus using CAD4 data without running this program.

3.9 How to Use CAD4 - Import

From the main menu:

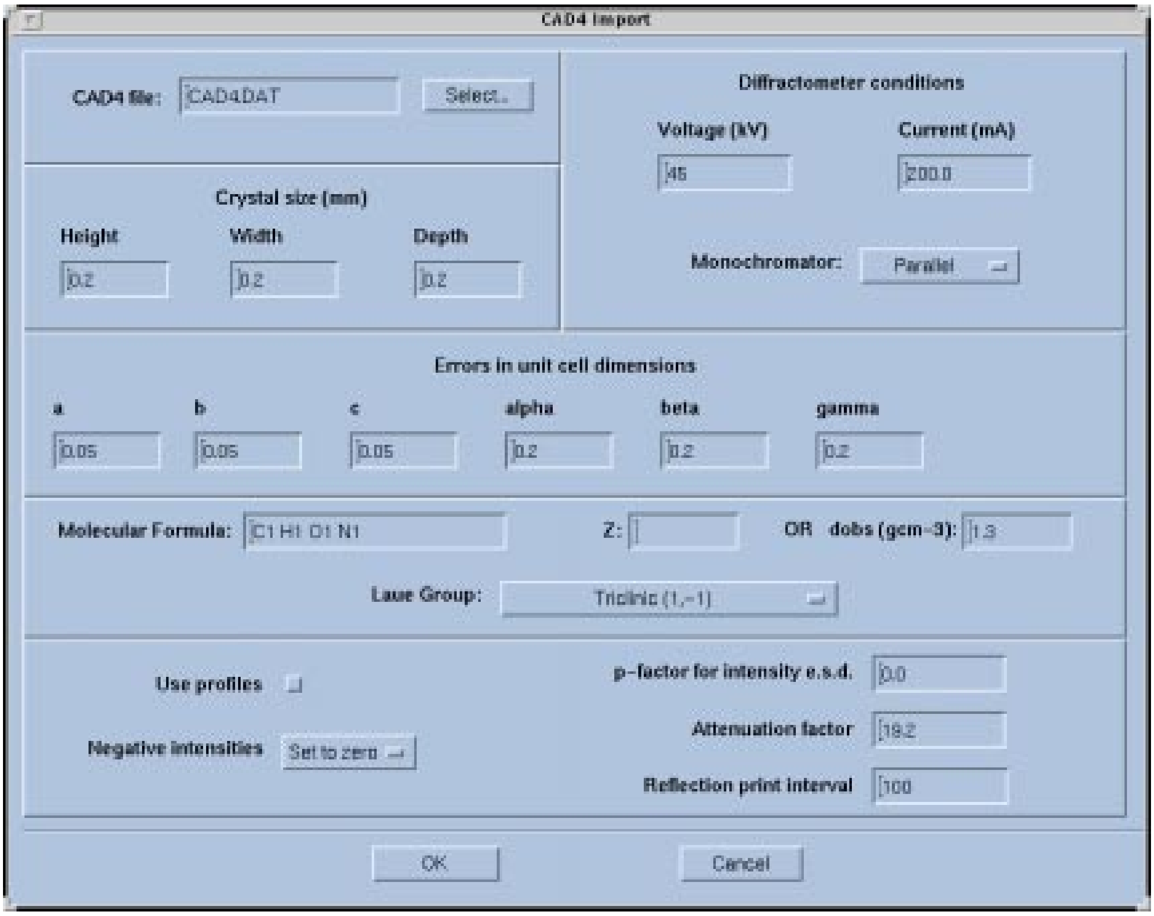


Choose the *File* option, followed by *Import*, followed by *CAD4*.

This will bring up the main CAD4 Import window.

3.9.1 The Main CAD4 Window

The following window will appear:



CAD4 file: CAD4.DAT

Voltage (kV): 45
Current (mA): 200.0
Monochromator: Parallel

Height: 0.2
Width: 0.2
Depth: 0.2

a: 0.05
b: 0.05
c: 0.05
alpha: 0.2
beta: 0.2
gamma: 0.2

Molecular Formula: C1 H1 O1 N1
Z:
OR dobs (gcm-3): 1.3
Laue Group: Triclinic (1,-1)

Use profiles ☐
Negative intensities: Set to zero
p-factor for intensity e.s.d.: 0.0
Attenuation factor: 19.2
Reflection print interval: 100

When all parameters that require to be changed from defaults have been altered, click on the OK button at the bottom of the CAD4 window. This will initiate the program.

If you wish to exit this window without processing use the CANCEL button in preference to exiting the window using the window control bar.

Each of the text fields are described below.

3.9.2 CAD4 File - Text Field

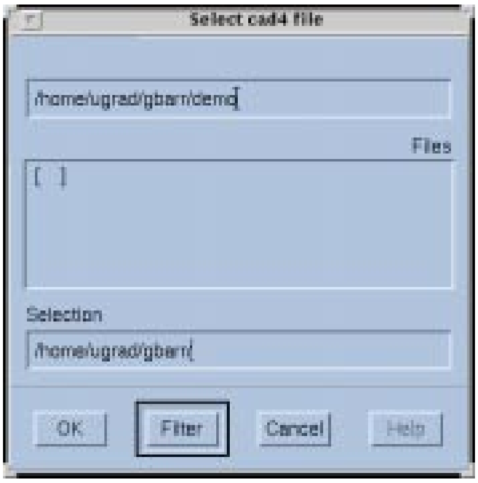
The user must supply the name of the CAD4 file containing the full dataset. The default is CAD4.DAT. ***This text box is case sensitive***.

If the file given in this box does not exist, then the CAD4 program will display the following dialogue box and then exit.

CAD4 file does not exist

It will give information in the summary file (see Section 2.4) as to where the problem occurred. To correct this problem, check both the name of your data file and the capitalisation of the file name. Also check that maXus is processing in the correct directory.

To change directory, or to browse through your files, click *Select...* to get a file browser:



/home/ugrad/gbarn/demo

Selection: /home/ugrad/gbarn/

Once you have chosen the file you want, click *OK*.

3.9.3 Crystal Size (mm): Height, Width, Depth Fields

These are the physical dimensions of the crystal used in the experiment.

The figures are in mm and the defaults are 0.2, 0.2, 0.2. These figures are only used in the output CIF file for descriptive purposes, and are not needed anywhere else.

3.9.4 Diffractometer Conditions: Voltage (kV) Field

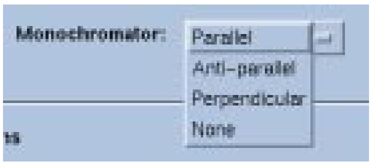
This is the voltage at which the diffractometer X-ray beam generator was operating during the experiment. The value is in kV and the default is 45. This figure is only used in the output CIF file as part of the experimental description.

3.9.5 Diffractometer Conditions: Current (mA) Field

This is the current at which the diffractometer X-ray beam generator was operating during the experiment. The value is in mA and the default is 200.0. Again, this figure is only used in the output CIF file as part of the experimental description.

3.9.6 Monochromator - Pop-Up Menu

Clicking on the menu area of the Monochromator field will cause the following pop-up menu to appear:



Monochromator: Parallel

The default is Parallel. To change to another selection click on the required selection.

This is the geometry of the CAD4 diffractometer, where Parallel/ Anti-parallel are the settings on a CAD4 diffractometer with a rotating anode and Perpendicular is the setting on a normal CAD4. The setting chosen here will determine which formula is used to calculate the Lorentz-polarisation factor.

3.9.7 Errors in Unit Cell Dimensions: a, b, c Fields

These are the estimated standard deviations on the unit cell dimensions.

The figures are in Å and the defaults are 0.05, 0.05, 0.05. These figures are used in the calculation of the errors in the volume of the unit cell. They are also used in the geometry programs as a contributor to the estimated standard deviations of bond lengths, bond angles, torsion angles and non-bonded contacts.

3.9.8 Errors in Unit Cell Dimensions: alpha, beta, gamma Fields

These are the estimated standard deviations on the unit cell angles. The figures are in degrees and the defaults are 0.2, 0.2, 0.2 Å on the edges, and 0.2° for the angles. These figures are used in the calculation of the errors in the volume of the unit cell. They are also used in the geometry programs as a contributor to the estimated standard deviations of bond lengths, bond angles, torsion angles and intramolecular contacts.

If an angle standard deviation input is zero and the associated angle is within 0.05 degrees of 90^0 or 120^0 then the angle is set from the experimental observed value to the appropriate exact value.

3.9.9 Molecular Formula Field

This is the molecular formula of the asymmetric unit.

In the formula, the atomic symbol and the number of atoms are grouped without a space or other separator. *The atomic symbol is case independent.* The list of atomic types can be separated by spaces or commas. The default for this field is C1 H1 O1 N1. This field is used extensively throughout maXus and in the CAD4 converter for verification of the value of the number of molecules in the unit cell.

Examples of acceptable and unacceptable molecular formulae:

ACCEPTABLE:

```
C11 H19 O4 N
```

```
c11 h19,O4,n
```

UNACCEPTABLE:

```
c11/h19,O4,N6
```

- since ‘/’ is not an acceptable separator.

```
C 11,H 19, O4 N6
```

- since a space between the element type and number is not permitted.

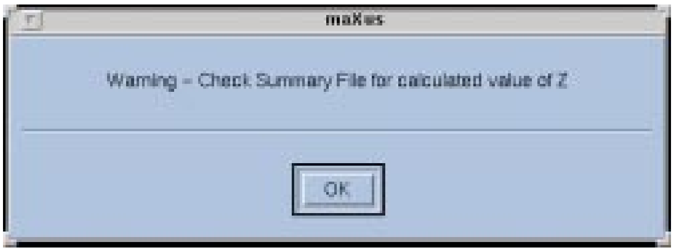
```
CN116 HO194
```

- since CN and HO are unrecognised element types.

3.9.10 Z (Number of Molecules in Unit Cell) Field

This field contains the number of molecules (Z), with the formula given in the Molecular Formula dialogue box that are contained within the unit cell.

The program will verify the user input value during processing by calculating the volume of the unit cell and comparing this with the number of non hydrogen atoms in the unit cell, each of which is assumed to require a volume of 18Å^3. If a significantly different value for Z is obtained the following warning window will appear:



Warning - Check Summary File for calculated value of Z

The user supplied and calculated values of Z will appear in the summary file. Note that this is just a warning and the program will continue with the user supplied value. Information on how to access the summary file is in Section 2.11.

The Import CAD4 program makes an intelligent attempt to calculate a value for Z if neither the crystal density or a value for Z are innput at this stage. Setting the correct value of Z is not important at this stage, since you will be given the opportunity to alter it when running the Spacegroup program later.

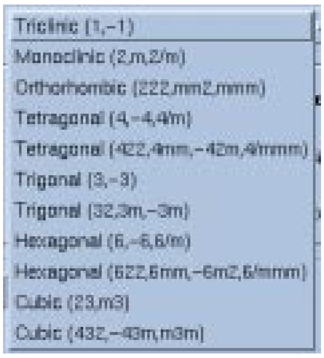
3.9.11 dobs (gcm^{-3}) (Observed Crystal Density) Field

This is the observed density of the crystal used in the experiment.

The figure is in gcm^{-3} and the default is 1.3. This value is only used in the output CIF file for descriptive purposes.

3.9.12 Laué Group - Pop-Up Menu

Clicking on the menu area of the Laué group field will cause the following pop-up menu to appear:



The default for this menu is Triclinic (1,-1). To select another Laue group click on the group you require. Rhombohedral systems should be treated as hexagonal.

The data reduction program will verify your choice from unit cell dimensions and if a discrepancy is discovered the following warning window will appear:


maXus

Problem Determining Laue Group - Check Summary File

The user supplied and the calculated Laué groups will appear in the summary file. Note that this is just a warning and the program will continue with the user supplied choice.

3.9.13 Use Profiles - Button

The button is used to determine whether or not profiles have been recorded in the CAD4 data file and hence if these should be used in the calculation of reflection intensities. If the button is *off* then the dataset will be treated as though no profiles have been recorded, even if they are present in the dataset. If the button is *on* and no profile information is present in the dataset then the button is ignored.

If the button is *on* and profiling information is present then the peak will be profiled and revised intensity and background count figures will be calculated for use in the determination of F_{obs}.

If profiles are present and the profile option has been requested then at the end of processing four windows will appear that graphically display typical reflections.

Each window will show two reflections, chosen at random, one balanced and one unbalanced with respect to background. If an unbalanced reflection could not be found it will not appear. A window for each of the following types of reflections will be produced: very weak, weak, medium, and strong.

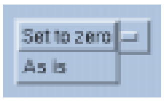
An example of such a graph is given below:

On the graph the Intensity is shown on the Y axis and the Profile Bin number is on the X axis. The h,k,l values of the reflections being displayed are given on the graph key.

To remove this window click in the CLOSE button.

3.9.14 Negative Intensities - Pop-Up Menu

Clicking on the menu area of the Negative Intensities field will cause the following pop-up menu to appear:



The default is ***Set to Zero***. To change to another selection click on the required selection.

Set to Zero will take all zero and negative reflections and set $F_{obs} = 0$ and $\sigma(F_{obs}) = 1$

As Is will allow negative intensities to be added to the Original file with their measured values for further processing.

3.9.15 p-factor for Intensity e.s.d Field

The p-factor is used in the calculation of $\sigma(F_{obs}^2)$ and prevents excessive weight being given to large structure factors. The default for this is 0.0. A similar but more flexible option is supplied in the weighting schemes of the least squares refinement programs.

In most cases, unless you know exactly what you are doing, leave this value as zero.

3.9.16 Attenuation Factor Field

Define the attenuation factor for filters appropriate to the diffractometer and radiation. The default for this is 19.2.

3.9.17 Reflection Print Interval Field

This figure determines at what interval a reflection in the dataset will be printed to the summary file. The default for this is 100 i.e. every 100th reflection will be printed.

3.10 Other CAD4 Import Windows

3.10.1 The CAD4 Import Performance Meter

While the CAD4 convertor is running a performance meter will appear at the top of the screen. When CAD4 has completed the performance meter will be closed. It is possible to close this window while CAD4 is running. The performance meter window is shown below:

Cad4 Process Indicator

3.10.2 Conversion Complete

Once the program has successfully converted a CAD4 file to maXus format the following window will appear:

Conversion Complete

Click on OK and maXus will return to the control of the main menu bar.

3.10.3 Conversion Errors

In the event that a fatal error occurs or 30 warnings about the data are issued then the program will terminate and the following dialogue box will appear:

Error - CAD4 Terminated

If this occurs click on OK and examine the summary output information carefully as this will detail the reason for the fatal error.

If the first record in your CAD4 input dataset is not a Type 21 record, the conversion program will display the following dialogue box and then terminate.

First Record in CAD4 File is not TYPE 21 - Conversion Terminating

If this occurs, Click on the OK box and then carefully examine the CAD4 dataset that you are trying to import checking for errors of any kind in the header area.

If the second record in your CAD4 input dataset is not a Type 22 record, the conversion program will display the following dialogue box and then terminate.

Second Record in CAD4 file is not TYPE 22 - Conversion Terminating

If this occurs. Click on the OK box and then carefully examine the CAD4 dataset that you are trying to import checking for errors of any kind in the header area.

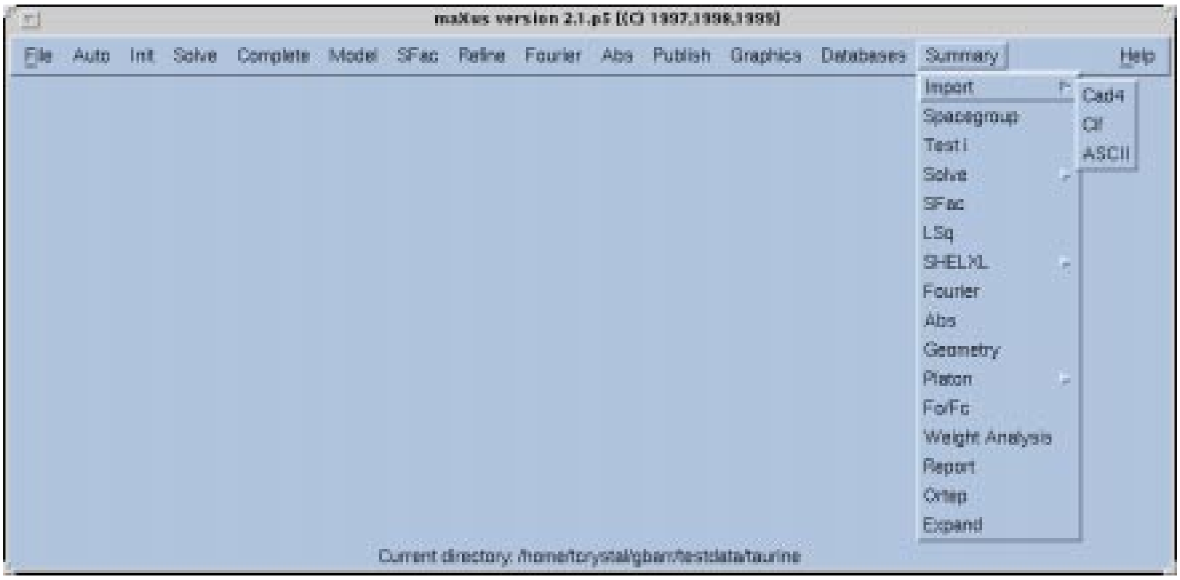
If the lines of data in your CAD4 input file are misaligned relative to each other the following dialogue box will appear and the program will terminate.

CAD4 data problem - Records are displaced - Conversion Terminating

If this occurs. Click on the OK box and then carefully examine the CAD4 dataset that you are trying to import, for any lines of data that are not lined up in the dataset.

3.11 Information in The CAD4 Import Summary File

To access the summary file go back to the main maXus window.



Choose the *Summary* option, followed by *Import / CAD4*.

The first section of the summary file will detail some of the options entered on the CAD4 screen. It will inform of the scale factor being used, the p-factor selected, the attenuation factor, and if profiles are being used. This section also details the monochromator setting and what action is to be taken with negative intensities. See Appendix A for an example.

Each time a CAD4 record is read that contains the information of the unit cell parameters then the wavelength of radiation used in the experiment is printed along with the new orientation matrix and also the metric matrix. The new unit cell and reciprocal unit cell are also output at this time. See Appendix A for an example.

Every reflection that comes within the Reflection Print Interval set on the main CAD4 screen, it is printed to the summary file for checking. See Appendix A for an example.

The final section of the summary report details the final statistics on the reflections. See Appendix A for an example.

3.12 Introduction To Importing ASCII Files

If you are trying to import a dataset in an ASCII format to maXus then you will require the Import-ASCII program. This uses an interface that will prompt you for information which maXus requires for data processing. This program performs the following functions:

- Captures information about the experiment, the unit cell and Laue group;
- Creates a maXus Original file for further processing. An Original file is a core maXus data-file that contains basic information and all reflection data;
- Makes an intelligent attempt to calculate a value for Z if neither the crystal density or a value for Z are input (this can be altered when the SpaceGroup program is run if incorrect).

It is not possible to proceed with maXus using ASCII data without running this program.

3.13 How to Use ASCII - Import

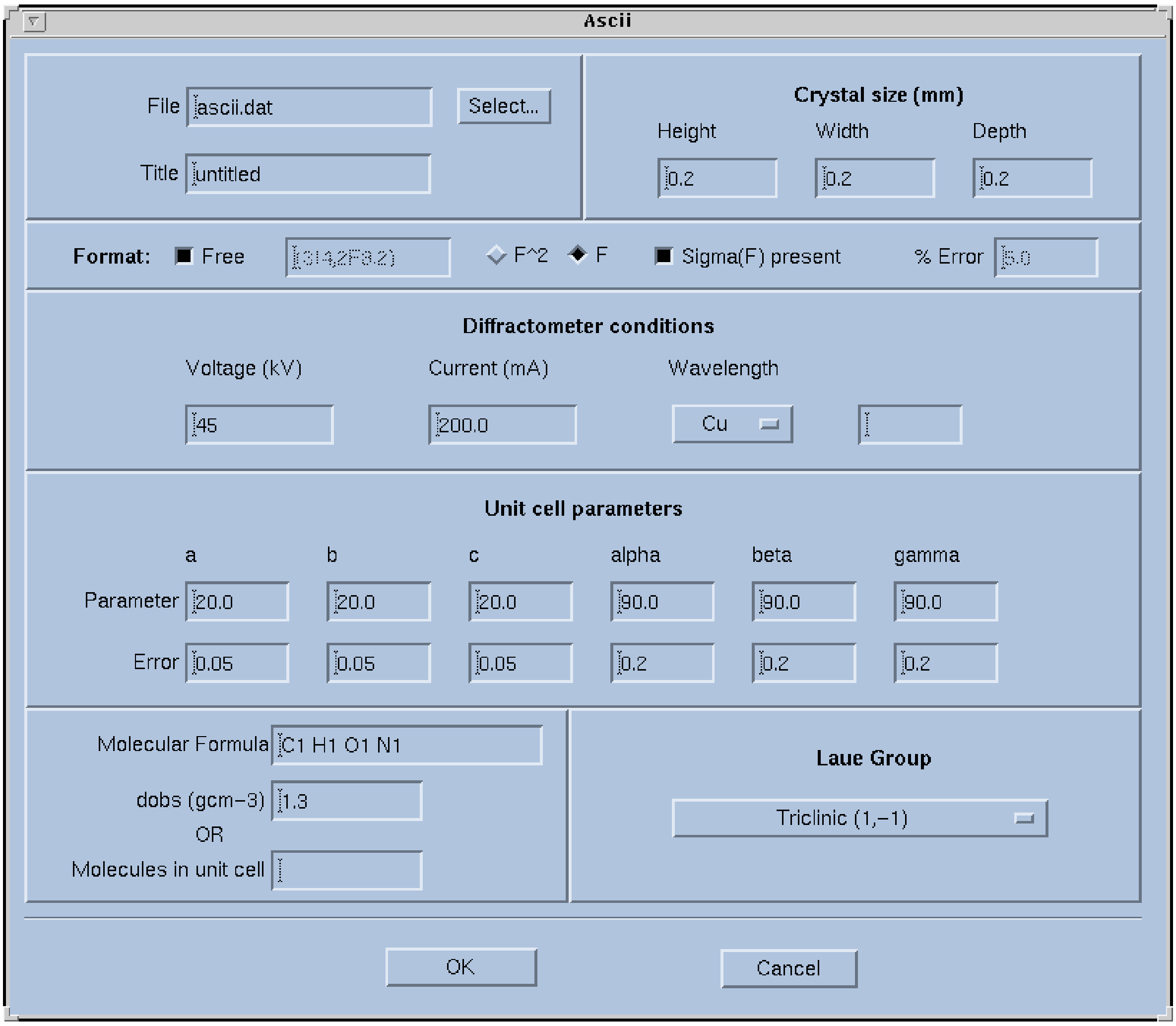
From the main menu:

Choose the FILE menu, followed by IMPORT, followed by ASCII.

This will call up the main ASCII Import window.

3.13.1 The Main ASCII Window

The following window will appear:



File ascii.dat

Title untitled

Height	Width	Depth
0.2	0.2	0.2

Format: ■ Free (3I4,2F8.2) ◇ F^2 ◆ F ■ Sigma(F) present % Error 5.0

Voltage (kV)	Current (mA)	Wavelength	
45	200.0	Cu	

	a	b	c	alpha	beta	gamma
Parameter	20.0	20.0	20.0	90.0	90.0	90.0
Error	0.05	0.05	0.05	0.2	0.2	0.2

Molecular Formula C1 H1 O1 N1

dobs (gcm-3) 1.3

OR

Molecules in unit cell

Laue Group Triclinic (1,-1)

Once all parameters that require to be changed from default have been altered click on the OK button at the bottom of the ASCII window. This will initiate the ASCII processing programs.

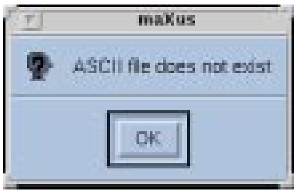
If you wish to exit this window without processing, use the CANCEL button in preference to exiting the window using the window control bar.

Each of the text fields are described below.

3.13.2 ASCII File - Text Field

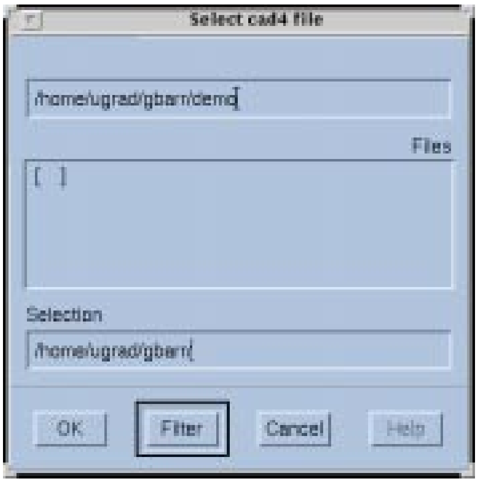
The user must supply the name of the ASCII file containing the full dataset. The default is ascii.dat. ***This text box is case sensitive***.

If the file given in this box does not exist, then the ASCII program will display the following dialogue box and then exit.



ASCII file does not exist

To correct this check both the name of your data file and the capitalisation of the file name. Also check that maXus is processing in the correct directory. Try locating your file using the file browser by clicking *Select* to get:



/home/ugrad/gbarn/demo

[ ]

/home/ugrad/gbarn/

Once you have chosen the file you want, click *OK*.

3.13.3 Title - Text Field

This is the title of the structure that will be used by all subsequent maXus processes for this structure.

3.13.4 Free Format Button

The default for this is *on*. Free format has data columns separated by at least one space or comma. Alternatively, maXus can interpret data in a fixed Fortran 77 style format. If the Free button is off, a fixed data string can be entered in the text field. The default for this is (3I4,2F8.2)

See Section 3.15 for more details on ASCII file formats.

3.13.5 F / F^2 - Buttons

The button is used to determine if the structure factor magnitude given within the dataset is F, rather than F squared. These buttons are mutually exclusive. The default is F.

3.13.6 Sigma(F) Present - Button

This button is used to determine whether data on the error in F (F^2) information is present within the data set. The default for this button is that on, i.e. sigma F information is available.

3.13.7 % Error Field

If no sigma(F) information is available then it will be simulated as a percentage of the F or (F^2) value. This value can be entered in this text field. This field has a default value of 5. This is often used with electron diffraction data sets.

3.13.8 Crystal Size (mm): Height, Width, Depth Fields

These are the physical dimensions of the crystal used in the experiment.

The figures are in mm and the defaults are 0.2, 0.2, 0.2. These figures are only used in the output CIF file.

3.13.9 Diffractometer Conditions: Voltage (kV) Field

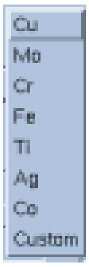
This is the voltage at which the diffractometer X-ray beam generator was operating during the experiment. The value is in kV and the default is 45. This figure is only used in the output CIF file as part of the experimental description.

3.13.10 Diffractometer Conditions: Current (mA) Field

This is the current at which the diffractometer X-ray beam generator was operating during the experiment. The value is in mA and the default is 200.0. This figure is only used in the output CIF file as part of the experimental description.

3.13.11 Diffractometer Conditions: Wavelength - Pop-Up Menu

Clicking on the menu area of the wavelength field will cause the following pop-up menu to appear:



The default is Cu. To change to another selection, click on that selection.

This is the wavelength of the radiation used during the measurement of the experimental data. Selecting *Custom* allows a Custom Wavelength (in Å) to be entered in the text field to the right of the pop-up menu.

If you are inputing electron or neutron data, these choices (Sections 3.13.9 - 3.13.11) are irrelevant and will be superceded by options chosen in the Spacegroup program.

3.13.12 Unit Cell Parameters: a, b, c, alpha, beta, gamma Fields

These are the unit cell parameters. The dimensions, a, b, c, are in Å and have defaults of 20.0Å. The angles α, β, γ are in degrees and have defaults of $90.0^{o.}$

3.13.13 Unit Cell Parameter Errors: a, b, c, alpha, beta, gamma Fields

These are the estimated standard deviations on the unit cell parameters.

The figures for the dimension parameters are in Å and the defaults are 0.05Å. The figures for the cell angles are in degrees and the defaults are $0.2^{0.}$

These figures are used in the calculation of the errors in the volume of the unit cell. They are also used in the geometry programs as a contributor to the estimated standard deviations of bond lengths, bond angles, torsion angles and non-bonded contacts.

3.13.14 Molecular Formula - Text Field

This is the molecular formula of the atoms in the asymmetric unit.

In the formula, the atomic symbol and the number of atoms are grouped without a space or other separator. *The atomic symbol is case dependent.* The list of atomic types can be separated by spaces. The default for this field is C1 H1 O1 N1. This field is used extensively throughout maXus and care must be taken when entering it.

Examples of acceptable and unacceptable molecular formulae:

ACCEPTABLE:

```
C11 H19 O4 N
```

```
c11 h19,O4,n
```

UNACCEPTABLE:

```
c11/h19,O4,N6
```

- since ‘/’ is not an acceptable separator.

```
C 11,H 19, O4 N6
```

- since a space between the element type and number is not permitted.

```
CN116 HO194
```

- since CN and HO are unrecognised element types.

3.13.15 dobs (gcm^{-3}) (Observed Crystal Density) Field

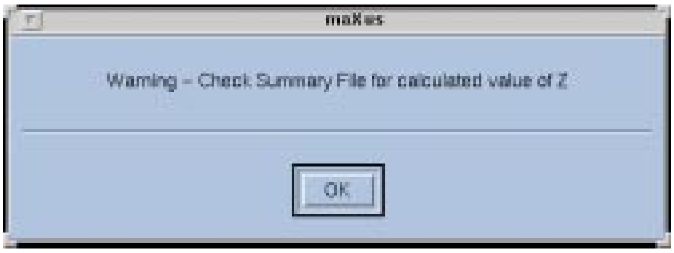
This is the observed density of the crystal used in the experiment.

The figure is in gcm^{-3} .

3.13.16 Molecules in Unit Cell (Z) Field

This field contains the number of molecules (Z), with the formula given in the Molecular Formula dialogue box, that are contained within the unit cell.

The program will verify the user supplied value during processing by calculating the volume of the unit cell and comparing this with the number of non hydrogen atoms in the unit cell, each of which is assumed to require a volume of 18Å^3. If a significantly different value for Z is obtained the following warning window will appear:



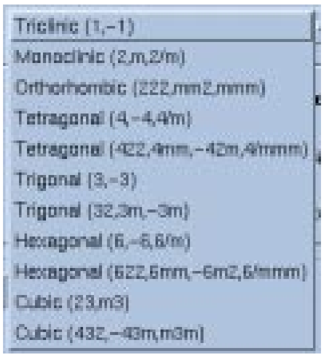
Warning - Check Summary File for calculated value of Z

Note that this is just a warning and the program will continue with the user supplied value.

If no value of Z or dobs is input at this stage, the program will make an intelligent attempt ot calculate a value for Z. Setting the correct value of Z is not important at this stage, since you will be given the opportunity to alter it when running the Spacegroup program later, where it is also computed more accurately using atomic volumes (Chapter 5).

3.13.17 Laué Group - Pop-Up Menu

Clicking on the menu area of the Laué group field will cause the following pop-up menu to appear:

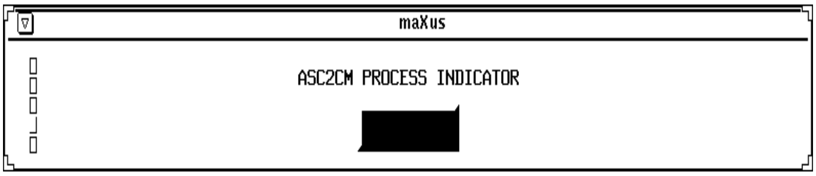


The default for this menu is Triclinic (1,-1). To select another Laue group click on the group you require. Rhombohedral systems should be treated as hexagonal.

3.14 Other ASCII Import Windows

3.14.1 The ASCII Import Performance Meter

While the ASCII convertor is running a performance meter will appear at the top of the screen. When ASCII has completed, the performance meter will be closed. It is possible to close this window while the ASCII import is running.



3.14.2 Conversion Complete

Once the program has successfully converted an ASCII file to maXus format the following dialogue box will appear.

Click on OK and maXus will return to the control of the main menu bar.

ASCII import complete

3.15 File formats for the ASCII File

3.15.1 Free Format

If the data file for input to the ASCII conversion program is a free format file, it must contain the following information separated by spaces or tabs.

h,k,l - The reflection indices. These must be integers, and can be negative.

F or F^2 - The magnitude of the structure factor or the signed square of the magnitude of the structure factor. This must be a real value.

Sigma F - This is the error on the structure factor given in the previous field. This is an optional figure and unless the program is informed of its presence, *via* the checkbox (Section 3.13.6) the data will be ignored. If this figure is not present then the ASCII converter will simulate a value using the % Error field.

Example of acceptable and unacceptable data lines from the input file:

ACCEPTABLE:

```
1 0 13 45.0
```

- Valid h,k,l, F format.

```
1 0 13 45.0 3.2
```

- Valid h,k,l, F, sigma(F) format.

UNACCEPTABLE:

```
1,0,13,45.0
```

- Value must be separated by tabs or spaces.

```
1 0 13A
```

- Value for F is missing.

```
1.0 0.0 13.0 45.0 3.2
```

- The indices of the reflection must be integer format.

3.15.2 Fixed Format

This follows the FORTRAN convention. The default type of fixed format is (3I4,2F8.2)

In other words, in each line of data, the program expects to find 3 integers, each of 4 characters, followed by 2 floating point numbers, each of which total width 8 characters (including the decimal point, with 2 places after it)

For example (if `.` represents a space):

```
               ..2..-1....6..227.92..3.95
               ..3...1...12...13.01.13.05
               ..1...5....3..222.22..1.25
Column Number: 12312341234512345678123456
Descriptor:     I3  I4   I5    F8.2  F6.2
```

This format would be labelled: (I3,I4,I5,F8.2,F6.2)

The program expects to find the data in the order h, k, l, F (or F^2), and σ F or σ F^2.

For example, the data h,k,l = (1,0,13), F = 45, and σ F= 3.2 would be represented as:

```
...1......13...45.00....3.20
```

(Where "`.`" is a space.)

Note that if you have deselected the ‘Sigma(F) present’ button in the main window, the program will crash if you are using Fixed Format, unless you change the data type accordingly - to (3I4,F8.2).

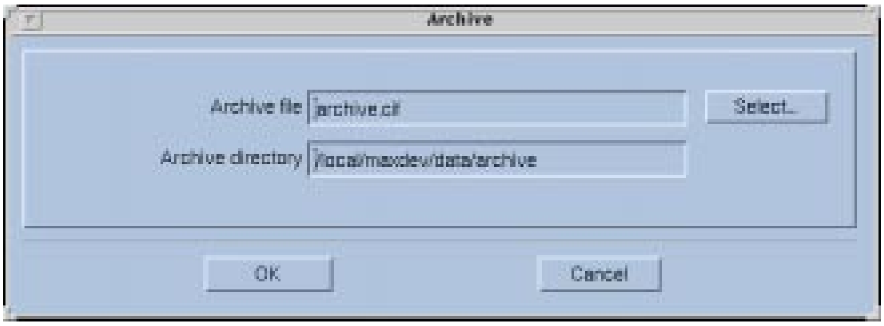
For example, the data h,k,l = (1,0,13) and F = 45.0 would be represented as:

```
...1......13...45.00
```

(Where "`.`" represents a space.)

3.16 Importing Archive Files

If you have saved a previous structure using the *Publish/Archive* option, then you can restore either the coordinates or the intensity data using the *Import Archive* option. Selecting this brings up a dialogue box:

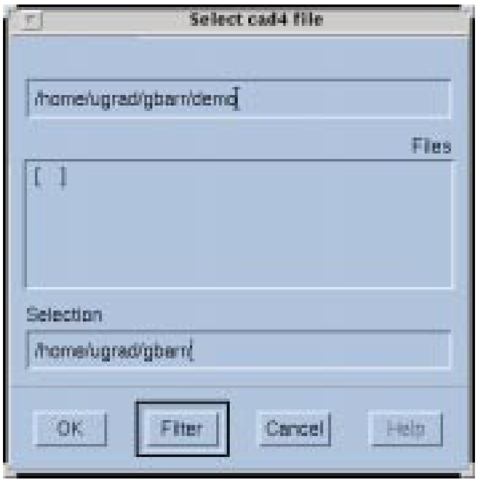


Archive file: archive.cif
Archive directory: /local/maxdev/data/archive

Enter the name of the file and the directory and click *OK*. If the file does not exist another dialogue box appears:

Archive file does not exist

To change directory, or to browse through your files, click *Select...* to get a file browser:



/home/ugrad/gbarn/demo

[ ]

Selection: /home/ugrad/gbarn/

Once you have chosen the file you want, click *OK*.

Otherwise the input proceeds and the following box appears:

CIF Import Complete

3.17 Important Details Concerning Archiving

The archive system uses standard CIF files in which data have been stored. When you invoke the *Publish/Archive* option a CIF file is generated in which the reflections are stored and this file is kept in your nominated archive directory. The *Import/Archive* option reads in the data only - it does not read in the atomic coordinates. If you want to read in coordinates, then use the *Import/CIF* option.

APPENDIX A

Example of a Typical CAD4 Import Summary File

```
SCALE FACTOR NUMBER =   1 SCALE FACTOR =   1.00000

P-FACTOR =0.0000 ATTENUATOR FACTOR = 19.2000

EXPECT PROFILES   ----- NO

MONOCHROMATOR SETTING PERPENDICULAR

NEGATIVE REFLECTIONS WILL BE SET TO ZERO

WAVELENGTH OF RADIATION USED = 0.709300

ORIENTATION   0.078982    -0.015091   0.052301
MATRIX        -0.083365   0.015251    0.049671
              -0.026798   -0.091845   -0.000387

METRIC        0.013906    -0.000002   0.000000
MATRIX        -0.000002   0.008896    0.000004
              0.000000    0.000004    0.005203

CELL    8.4801    10.6021   13.8638   90.000   90.000   90.000
        0.11792   0.09431   0.07213   90.000   90.000   90.000

 -------------------------------------------------------------------------------------------------------

     H    K    L    FREL   SIGF   NREF BL   INT BR  THETA    T    1/LP   DEC    XRAYT   JCODE   NSET

IR*  -3   -3   3    111.58 0.18   1    826  102279 444 10.27 3    0.3745 1.0000 0.0139  1       1
IR*  -1   3    5    46.37  0.08   2    623  100371 614 9.69  16   0.3514 1.0000 0.0333  2       1
IR*  -2   1    -2   43.55  0.07   3    649  101625 732 5.96  11   0.2111 1.0000 0.0461  3       1
     -7   -3   -2   20.40  0.21   4    272  4949204  18.32   7    0.7287 1.0000 0.0592  4       1
     -7   -3   -1    93.29 0.26   5    606  38027 261 18.13  3    0.7193 1.0000 0.0631  5       1
U    -7   -3   0    0.00   0.00   6    252  906  225  18.06  7    0.7162 1.0000 0.0758  6       1
     -7   -3   1    93.21  0.26   7    605  38008 282 18.13  3    0.7193 1.0000 0.0800  7       1
     -7   -2   1    49.25  0.23   8    337  15089 20217.58   4    0.6925 1.0000 0.0889  8       1
U    -7   -2   0    0.00   0.00   9    231  858  223  17.51  7    0.6893 1.0000 0.1019  9       1
     -7   -2   -1   49.08  0.23   10   339  14963 18617.58   4    0.6925 1.0000 0.1111  10      1
     -7   -2   -2   34.62  0.18   11   378  13189 241 17.77  7    0.7020 1.0000 0.1242  11      1
     -7   -2   -3   32.63  0.19   12   356  11592 246 18.09  7    0.7177 1.0000 0.1369  12      1
     -7   -2   -4   32.98  0.19   13   332  11450246 18.54   7    0.7395 1.0000 0.1500  13      1
     -7   -1   -4   46.66  0.24   14   272  12989 205 18.21  4    0.7235 1.0000 0.1589  14      1
     -7   -1   -3   28.90  0.19   15   348  9477 222 17.76   7    0.7016 1.0000 0.1719  15      1
     -4   -4   -8   20.14  0.18   100  147  4702109  17.27   7    0.6778 1.0000 1.0678  100     1
     -1   0    -12  4.94   0.44   200  115  760  146  18.09  7    0.7172 1.0000 2.1544  200     1

 ------------------------------------------------------------------
WAVELENGTH OF RADIATION USED = 0.709300

ORIENTATION   0.078982    -0.015091   0.052301
MATRIX        -0.083365   0.015251    0.049671
              -0.026798   -0.091845   -0.000387
```

```
METRIC         0.013906   -0.000002   0.000000
MATRIX        -0.000002    0.008896   0.000004
               0.000000    0.000004   0.005203

CELL   8.4801  10.6025  13.8638  90.000   90.000   90.000
       0.11792 0.09431  0.07213  90.000   90.000   90.000

 ----------------------------------------------------------------------------------------------------------

   H   K   L    FREL  SIGF  NREF BL   INT BR  THETA  T   1/LP   DEC    XRAYT  JCODE  NSET

   -5  -1  -5   38.35 0.20  300  195  1151115614.35  4   0.5442 1.0000 3.0875 300    1
W  -2  -7  -1   0.43  1.13  400  67   259  62  14.50  3   0.5508 1.0000 4.0572 400    1
   1   -2  -10  3.27  0.46  500  103  499  84  15.56  7   0.5981 1.0000 5.0289 500    1

 ----------------------------------------------------------------------------------------------------------

WAVELENGTH OF RADIATION USED = 0.709300

ORIENTATION    0.078982   -0.015091   0.052301
MATRIX        -0.083365    0.015251   0.049671
              -0.026798   -0.091845  -0.000387

METRIC         0.013906   -0.000002   0.000000
MATRIX        -0.000002    0.008896   0.000004
               0.000000    0.000004   0.005203

CELL   8.4801  10.6025  13.8638  90.000   90.000   90.000
       0.11792 0.09431  0.07213  90.000   90.000   90.000

 ----------------------------------------------------------------------------------------------------------

   H   K   L    FREL  SIGF  NREF BL   INT BR  THETA  T   1/LP   DEC    XRAYT  JCODE  NSET

   -3  -2  1    4.74  0.27  600  268  1507222 8.32   7   0.2989 1.0000 5.9967 600    1
A  -2  0   0    197.97 0.45 639  15   48275 28 4.81  -4  0.1694 1.0000 6.3628 639    1
   -1  -5  -4   7.51  0.18  700  97   129488  11.59  7   0.4273 1.0000 6.9822 700    1
D  0   -1  -2   2.53  0.29  781  642  2555454 3.51   7   0.1232 1.0000 7.7444 781    1
D  0   -3  -2   2.31  0.37  799  241  1012184 6.48   7   0.2302 1.0000 7.8825 799    1
   0   -3  -1   1.43  0.64  800  334  1276270 5.96   7   0.2110 1.0000 7.8953 800    1
D  1   -3  0    2.37  0.44  847  374  1561318 6.25   7   0.2219 1.0000 8.4194 847    1
D  1   -2  0    2.25  0.39  863  509   2110435 4.53  7   0.1596 1.0000 8.5808 863    1
------------------------------------------------------------------------------------------------------------

WAVELENGTH OF RADIATION USED = 0.709300

ORIENTATION    0.078982   -0.015091   0.052301
MATRIX        -0.083365    0.015251   0.049671
              -0.026798   -0.091845  -0.000387

METRIC         0.013906   -0.000002   0.000000
MATRIX        -0.000002    0.008896   0.000004
               0.000000    0.000004   0.005203

CELL   8.4801  10.6025  13.8638  90.000   90.000   90.000

       0.11792 0.09431  0.07213  90.000   90.000   90.000
```

```
 -----------------------------------------------------------------------------------------------------

     H    K    L     FREL  SIGF   NREF  BL   INT BR   THETA   T    1/LP    DEC    XRAYT    JCODE   NSET

     -8   -2   -1    4.78   0.69   900    199  900  167  20.05    6    0.8163 1.0000 8.9189    900       1
     -4   -8   -3    27.66  0.19   1000   240  7884189  18.99    7    0.7623 1.0000 9.9631    1000      1
      0   -6   -10   12.92  0.23   1100   110  1982116  19.02    7     0.76361.0000 11.0706  1100      1
------------------------------------------------------------------------------------------------------------------

WAVELENGTH OF RADIATION USED = 0.709300

ORIENTATION       0.078982    -0.015091    0.052301
MATRIX            -0.083365    0.015251     0.049671
                  -0.026798    -0.091845    -0.000387

METRIC            0.013906    -0.000002    0.000000
MATRIX            -0.000002    0.008896     0.000004
                  0.000000     0.000004     0.005203

CELL      8.4801    10.6025   13.8638   90.000    90.000    90.000
          0.11792   0.09431   0.07213   90.000    90.000    90.000

 -----------------------------------------------------------------------------------------------------

     H    K    L     FREL  SIGF   NREF  BL   INT BR   THETA   T    1/LP    DEC    XRAYT    JCODE   NSET
     -6   -6   -6    13.74  0.29   1200   147  1873132  20.91    6     0.86141.0000 12.0969  1200      1
     -1   -10  -5     2.36  0.78   1300   42   234  56   21.18    6     0.87621.0000 13.1864  1300      1
-----------------------------------------------------------------------------------------------------------------

WAVELENGTH OF RADIATION USED = 0.709300

ORIENTATION       0.078982    -0.015091    0.052301
MATRIX            -0.083365    0.015251     0.049671
                  -0.026798    -0.091845    -0.000387

METRIC            0.013906    -0.000002    0.000000
MATRIX            -0.000002    0.008896     0.000004
                  0.000000     0.000004     0.005203

CELL        8.4801  10.6025   13.8638   90.000    90.000    90.000
          0.11792   0.09431   0.07213   90.000    90.000    90.000

 -----------------------------------------------------------------------------------------------------

     H    K    L     FREL  SIGF   NREF  BL   INT BR   THETA   T    1/LP    DEC    XRAYT    JCODE   NSET
     -9   1    -2    22.13  0.26   1400   208  3882180  22.46    6    0.9462 1.0000 14.1481  1400      1
     -4   -4   -13   54.40  0.27   1500   220  12581 12923.36    4    0.9963 1.0000 15.2858  1500      1
     1    -7   -12   34.30  0.23   1600   187  7902128  22.91    6    0.9709 1.0000 16.2942  1600      1
-----------------------------------------------------------------------------------------------------------------

WAVELENGTH OF RADIATION USED = 0.709300

ORIENTATION       0.078982    -0.015091    0.052301
MATRIX            -0.083365    0.015251     0.049671
                  -0.026798    -0.091845    -0.000387

METRIC            0.013906    -0.000002    0.000000
```

```
MATRIX          -0.000002   0.008896    0.000004
                 0.000000   0.000004    0.005203

CELL      8.4801  10.6025  13.8638  90.000    90.000    90.000

         0.11792  0.09431  0.07213  90.000    90.000    90.000

 -----------------------------------------------------------------------------------------------------------

     H    K    L     FREL  SIGF   NREF  BL    INT BR   THETA   T     1/LP    DEC     XRAYT    JCODE   NSET
D   -10   0    -1    4.99   0.78   1616  132   712  155  24.82   6    1.0815 1.0000 16.4789  1616      1
     -6   -2   -12   10.40   0.35  1700  96    1041106  23.76   6    1.0196 1.0000 17.3786  1700      1
     -1   -12  -2    10.21  0.36   1800  101   1017106  24.06   6    1.0366 1.0000 18.4928  1800      1

 ---------------------------------------------------------------------------
-----------------------------------

I      51  INTENSITY STANDARDS
W     264  MEASURED ON PRESCAN ONLY
U     144  OF THESE HAD I LESS THAN OR EQUAL TO ZERO
A       1  USED THE ATTENUATOR
R      51  REFLECTIONS, INCLUDING INTENSITY STANDARDS, WERE REJECTED
C       0  OPTIMUM PSI SETTING NOT ATTAINABLE
D       6  PEAK DISPLACED BY MORE THAN DANG
S       0  S - COUNT LOSS OF MORE THAN 1%
T       0  T - COUNT LOSS OF MORE THAN 1% WITH ATTENUATOR
F       0  FAILED NON-EQUALITY TEST
X       0  C OR X POSITIONING PROBLEM

1846  REFLECTIONS WRITTEN TO THE ORIGINAL FILE

H   K   LFSQTHETASCANAPERTURE

MINIMUM -10 -12 -16-22.402.40    0.73    2.20

MAXIMUM 1 1  2   39192.33 24.970.902.27
```

APPENDIX B

Example of a Typical CIF Import Summary File

```
Warning: data name _refine_diff_density_rms not in dictionary!

Warning: data name _shelx_titlenot in dictionary!

Warning: data name _shelx_refln_list_code not in dictionary!

Warning: data name _shelx_f_calc_maximum not in dictionary!

Data will be accessed from data block shelxl

No name could be found. Defaulting to COMPOUND 1

No value for the observed density could be found. Defaulting to 1.25 (gcm-3)

No device type could be identified. Defaulting to CAD4

Could not find orientation matrix element _UB_13 in the CIF file - Unitary
Matrix Written

Could not find orientation matrix element _UB_23 in the CIF file - Unitary
Matrix Written

Could not find orientation matrix element _UB_31 in the CIF file - Unitary
Matrix Written

Could not find orientation matrix element _UB_32 in the CIF file - Unitary
Matrix Written

Could not find orientation matrix element _UB_33 in the CIF file - Unitary
Matrix Written

Cannot find Isotropic Temperature factor for atom: O1         Defaulting to
0.05 Angstrom squared

Cannot find Isotropic Temperature factor for atom: O2         Defaulting to
0.05 Angstrom squared

Cannot find Isotropic Temperature factor for atom: O3         Defaulting to
0.05 Angstrom squared

Cannot find Isotropic Temperature factor for atom: O4         Defaulting to
0.05 Angstrom squared

Cannot find Isotropic Temperature factor for atom: O5         Defaulting to
0.05 Angstrom squared
```

APPENDIX C

Reflection Data Fields - CIF Import

The following is a list of CIF standard names, followed by the defaults. Some information that is required will be obtained from other CIF fields as the default.

TABLE 1. CIF Fields Required For Reflection Input

Name	Default	Description
_chemical_name_common	_chemical_name_systematic	The name that will be used within maXus.
_chemical_name_systematic	_chemical_name_mineral	The name that will be used within maXus.
_chemical_name_mineral	COMPOUND 1	The name that will be used within maXus.
_chemical_formula_sum	C1 H4	The molecular formula of the compound under examination.
_cell_formula_units_Z	Calculated based on number of non hydrogen atoms each occupying 18Å^3	The number of molecules in the unit cell.
_cell_length_a	10 Å with an error of 0.1Å	The length of the unit cell along the a-axis.
_cell_length_b	10 Å with an error of 0.1Å	The length of the unit cell along the b-axis.
_cell_length_c	10 Å with an error of 0.1Å	The length of the unit cell along the c-axis.
_cell_angle_alpha	90^o with an error of 0.1^o	The cell angle alpha.
_cell_angle_beta	90^o with an error of 0.1^o	The cell angle beta.
_cell_angle_gamma	90^o with an error of 0.1^o	The cell angle gamma.
_exptl_crystal_density_meas	1.25 gcm^{-3}	The observed crystal density.
_exptl_crystal_density_diffrn	Calculated from the unit cell contents and the volume of the unit cell.	The calculated crystal density
_diffrn_radiation_wavelength	0.7093 Å (Mo K-α Radiation)	The wavelength of the radiation used in the data collection. (in a loop)
_diffrn_radiation_id_all		Which parts of the data were collected with the wavelength above; normally ALL.
_diffrn_measurement_device	CAD4	The maXus program recognises CAD4, MAC, Kappa CCD and DIP collection devices. All others will be set to the default.
_diffrn_radiation_type	X-Rays	The maXus program recognises electron and neutron radiation.
_diffrn_orient_matrix_UB_11	Entire orientation matrix is set to unitary matrix.	Element 1-1 of the orientation matrix
_diffrn_orient_matrix_UB_12	Entire orientation matrix is set to unitary matrix.	Element 1-2 of the orientation matrix
_diffrn_orient_matrix_UB_13	Entire orientation matrix is set to unitary matrix.	Element 1-3 of the orientation matrix

TABLE 1. CIF Fields Required For Reflection Input

Name	Default	Description
_diffrn_orient_matrix_UB_21	Entire orientation matrix is set to unitary matrix.	Element 2-1 of the orientation matrix
_diffrn_orient_matrix_UB_22	Entire orientation matrix is set to unitary matrix.	Element 2-2 of the orientation matrix
_diffrn_orient_matrix_UB_23	Entire orientation matrix is set to unitary matrix.	Element 2-3 of the orientation matrix
_diffrn_orient_matrix_UB_31	Entire orientation matrix is set to unitary matrix.	Element 3-1 of the orientation matrix
_diffrn_orient_matrix_UB_32	Entire orientation matrix is set to unitary matrix.	Element 3-2 of the orientation matrix
_diffrn_orient_matrix_UB_33	Entire orientation matrix is set to unitary matrix.	Element 3-3 of the orientation matrix
_symmetry_cell_setting	Best guess is calculated from the unit cell parameters	The symmetry of the unit cell.
_refln_index_h	reflection is ignored	The h index of reflection in CIF _loop
_refln_index_k	reflection is ignored	The k index of reflection in CIF _loop
_refln_index_l	reflection is ignored	The l index of reflection in CIF _loop
_refln_F_meas	_refln_F_squared_meas	The observed F magnitude.
_refln_F_squared_meas	Reflection is ignored	The observed signed F^2.
_refln_F_sigma	_refln_F_squared_sigma	The error on the observed F magnitude.
_refln_F_squared_sigma	Reflection is ignored	The error on the observed F^2.

Atom Data Fields - CIF Import

The following is a list of CIF standard names, followed by the defaults. Other information that is required will be obtained from other CIF fields as the default.

TABLE 2. CIF Fields Required For Atoms Input

Name	Default	Description
_chemical_name_common	_chemical_name_systematic	The name that will be used within maXus.
_chemical_name_systematic	_chemical_name_mineral	The name that will be used within maXus.
_chemical_name_mineral	COMPOUND 1	The name that will be used within maXus.
_chemical_formula_sum	C1 H4	The molecular formula of the compound under examination.
_cell_formula_units_Z	Calculated based on number of non hydrogen atoms each occupying 18Å^3.	The number of molecules in the unit cell.
_cell_length_a	10 Å with an error of 0.1Å	The length of the unit cell along the a-axis.
_cell_length_b	10 Å with an error of 0.1Å	The length of the unit cell along the b-axis.
_cell_length_c	10 Å with an error of 0.1Å	The length of the unit cell along the c-axis.

TABLE 2. CIF Fields Required For Atoms Input

Name	Default	Description
_cell_angle_alpha	90^{o} with an error of 0.1^{o}	The cell angle α.
_cell_angle_beta	90^{o} with an error of 0.1^{o}	The cell angle β.
_cell_angle_gamma	90^{o} with an error of 0.1^{o}	The cell angle γ.
_diffrn_radiation_wavelength	0.7093 Å (Mo K-α Radiation)	The wavelength of the radiation used in the data collection.
_diffrn_radiation_type	X-Rays	The maXus program recognises electron and neutron radiation.
_symmetry_cell_setting	Best guess is calculated from the unit cell parameters	The symmetry of the unit cell.
_atom_site_label	Atom ignored if not present.	Atom Label Identifier
_atom_site_fract_x	Atom ignored if not present.	Fractional atomic co-ordinate along x.
_atom_site_fract_y	Atom ignored if not present.	Fractional atomic co-ordinate along y.
_atom_site_fract_z	Atom ignored if not present.	Fractional atomic co-ordinate along z.
_atom_site_U_iso_or _equiv	0.05 Å^2	Isotropic temperature factor, or equivalent calculated from anisotropic factors.
_atom_site_occupancy	1.0	Standard site occupancy factor.
_atom_site_type_symbol	‘C’	Atomic element symbol
_atom_site_aniso_label	Atom ignored if not present.	This must match the corresponding Atom Label Identifier. It is used to identify the set of anisotropic temperature factors.
_atom_site_aniso_U_11	0 Å^2 with an error of 0Å^2 Isotropic assumed	The 1-1 Anisotropic temperature factor
_atom_site_aniso_U_12	0 Å^2 with an error of 0Å^2 Isotropic assumed	The 1-2 Anisotropic temperature factor
_atom_site_aniso_U_13	0 Å^2 with an error of 0Å^2 Isotropic assumed	The 1-3 Anisotropic temperature factor
_atom_site_aniso_U_22	0 Å^2 with an error of 0Å^2 Isotropic assumed	The 2-2 Anisotropic temperature factor
_atom_site_aniso_U_23	0 Å^2 with an error of 0Å^2 Isotropic assumed	The 2-3 Anisotropic temperature factor
_atom_site_aniso_U_33	0 Å^2 with an error of 0Å^2 Isotropic assumed	The 3-3 Anisotropic temperature factor
_symmetry_equiv_pos_as _xyz	_symmetry_space_group _name_H-M	A loop of the symmetry equivalent positions as presented in International Tables Vol. A.
_symmetry_space_group _name_H-M	_symmetry_Int_Tables _number	The space group name as presented in International Tables Vol. A.
_symmetry_Int_Tables _number	1	The space group number as presented in International Tables Vol. A.

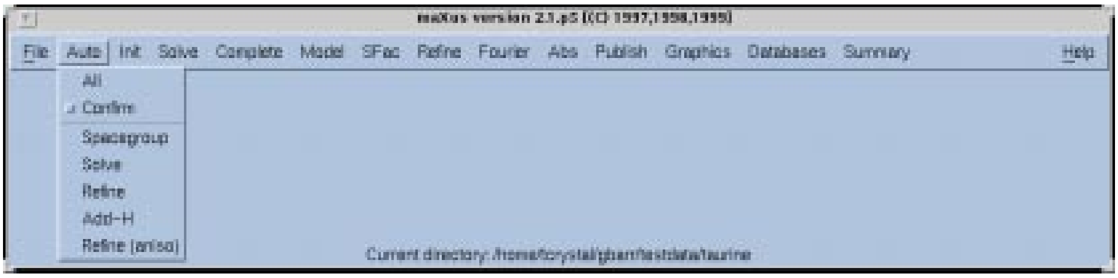
4 maXus in Automatic Mode

4.1 Solving and Refining Crystal Structures Automatically

maXus provides a procedure for automatically solving and refining crystal structures.

This first assignes the space group, attempts to solve the stucture, refines isotropically, automatically adds hydrogens, and finally refines anisotropically. To begin the process, either an Original file, or a converted data file, must be present in the directory you are working in.

From the main menu select Auto.



Current directory: /home/crystal/gban/testdata/taurine

You will find six options presented to you:

All
Confirm
Spacegroup
Solve
Refine
Add H
Refine (aniso)

These options will now be described in turn.

Selecting *All* runs each of the programs in turn. Individual programs can be run directly by selecting the relevant option from the Auto menu.

4.2 Auto Menu: All

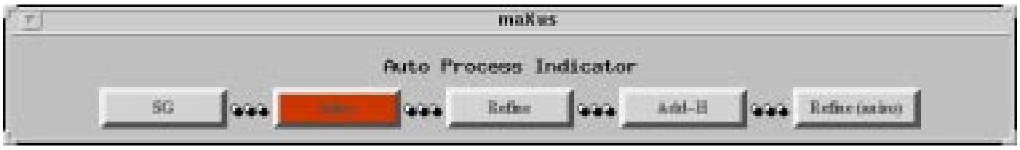
The program can be run in one of two modes. To run in Confirmation mode, select *Confirm* from the *Auto* menu, before selecting *All*. With the *Confirm* option turned on, the program will stop after each individual stage of the refining process, and produce the following dialog box:

Continue?

Click *OK* to continue, or *Cancel* to stop the process.

If *Confirm* is not selected before *All*, the program will run in its default mode of No Confirmation - it runs each of the programs in turn, without pausing between them.

While maXus is processing the data, a progress indicator is displayed to show which program is currently running:



Automatic processing using the *All* option treats the data in the following way:

(1) The automatic space group determination program, Spacegroup, is run. The program will select, if there is any ambiguity, the most common space group. (If you have already run the Spacegroup program separately, the results of this will be used; Spacegroup will not be run again.) The only time Auto Spacegroup requires user input is if a non-standard wavelength has been used. In this case, you will need to enter constants (df’, df’’ and μ) for each element type, as well as the type of data-collection device.

(2) The data are checked and normalised.

(3) Direct methods are used to attempt a structure solution. The best solution is refined first by Fourier methods, and then by least squares. An attempt is made to assign peaks to atom types and to find any missing atoms.

(4) If the R-factor is >25% after refinement, a second set of options for direct methods is selected, and the above procedure is repeated. Each set of options is slower than the previous one.

(5) Up to four sets of direct methods options are tried. If one set gives an R-factor of <25% then that solution is passed to step (6).

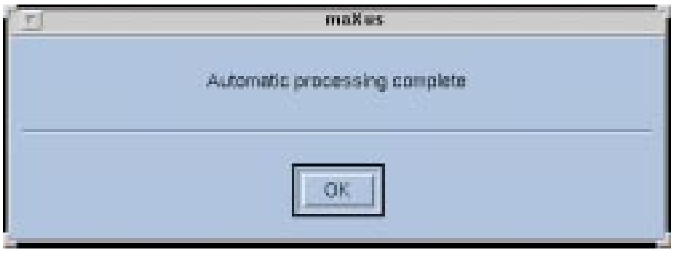
(6) The structure is then refined for three cycles with full matrix least squares, with each atom given individual isotropic thermal parameters.

(7) The H atoms are automatically added in idealised positions where possible. Note that only C atoms can have H atoms attached automatically. It is not possible to automatically add H atoms to terminal O-H groups, for example.

(8) The structure is refined anisotropically using full matrix least squares. All non-H atoms are given anisotropic thermal parameters; the H atoms ride on the attached C atoms, and are given a fixed thermal parameter.

(9) The program then stops.

If the solution was successful, the program terminates normally:



Automatic processing complete

If the program tries all possible strategies and still cannot solve the structure, the following error message appears:

Autosolve has failed

4.3 Automatic Processing: Warning!

There are some important points to considered when using the Automatic Processing feature:

This procedure is very good at assigning atom types, but it can make mistakes. ALL AUTO PACKAGES MAKE THESE MISTAKES, and many of them are appearing in the literature. YOU MUST CHECK THE ATOM TYPES before proceeding (See the chapter on the Model window; Section 12.4.1). We are continually modifying the program to improve this feature.

- H atoms: some terminal CH_3 groups may not be optimal. You can use the manual addition of H atoms in the MODEL window to add H atoms with more control (Chapter 12).
- It is usually better to run the space group determination program manually using the *Spacegroup* option under the *Init* menu. Answer the questions and then select from the space groups on offer - you have knowledge that is not available to *Spacegroup*.

The steps (1)-(7) are, in general, also accessible as individual steps as now described.

4.4 Auto Menu: Space group

This runs step (1) of *Auto/All* without user intervention. The same proviso applies as before:

- It may be better to run the manually using the option under the *Init* menu, answer the questions, and then select from the space groups on offer - you have knowledge that is not available to *Spacegroup*.

4.5 Auto Menu: Solve

This carries out the following steps:

(1) The data are checked for duplicates and systematic absences, and then normalised

(2) Direct methods are used to attempt a structure solution. The best solution is refined first by Fourier methods, then by block diagonal least squares. An attempt is made to assign peaks to atom types and to find any missing atoms

(3) If the R-factor is >25% after refinement, a second set of options for direct methods is selected, and the above procedure is repeated. Each set of options is slower than the previous one.

(4) Four sets of direct methods options are tried. (If the procedure still fails, use the direct options described in Chapter 7. Alternatively, use the other direct methods programs available in the maXus package.)

4.6 Auto Menu: Refine

The structure is refined for three cycles with full matrix least squares, with each atom given an individual isotropic thermal parameters. Any necessary constraints are applied automatically.

4.7 Auto Menu: Add-H

The H atoms are automatically added in idealised positions where possible. Note that only C atoms can have H atoms attached automatically (N atoms have ambiguities). It is not possible to add H atoms to terminal O-H groups, for example.

In order to alter these, or to add H atoms to other types of atom you have to use the MODEL program. See the Model chapter for a full description.

This option is often useful even when you are solving a structure manually.

4.8 Auto Menu: Refine (aniso)

The structure is refined anisotropically using full matrix least squares. All non-H atoms are given anisotropic thermal parameters; the H atoms ride on the attached atoms, and are given a fixed thermal parameter. Refinement proceeds until either:

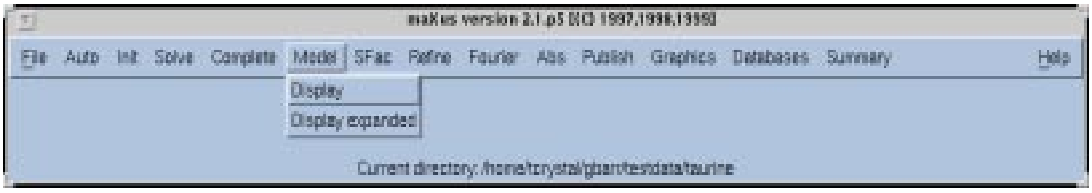
- Shift errors are less than 0.5 , or

- 10 cycles have been carried out.

Again, this option is often useful even when you are refining a structure manually.

4.9 Checking the AUTO results

Open the MODEL display window:



Click the *DISPLAY* button (bottom right hand corner), and then select *U (iso)* from the options presented for display properties.

Atoms which have negative, or non-positive definite temperature factors will have ‘?’ in them. Atoms with unusually large thermal parameters will appear with large spheres.

These are the atoms which are probably wrongly assigned:

- The ? means that the atom should have an *increased* atomic number.
- Large spheres mean that the atom should have an *decreased* atomic number.

Return to the model window and change the atom types by clicking on correct element in the periodic table, and then on the atom whose type is to be changed. Also, check the geometry around suspect atoms using the distance/angle options.

Check that atomic valencies are correct.

MODEL functions are fully described in Chapter 12.

This completes a description of the AUTO options, but there are a number of other automatic ways of carrying out steps in crystal structure analysis, and these will now be discussed.

4.10 Solving Structures Automatically

If the AUTO mode does not solve the structure for you then the following steps are recommended:

(1) Use the SOLVE options on the menu. You are presented with the following choices:

- MCMultan (see Chapter 11)
- Mithril94 (see Chapter 9)
- Dirdiff (see Chapter 8)

- SIR with 4 options numbered 1-4 (see Chapter 7).

Try the first three of these in turn, and use the MODEL option to investigate the proposed structure. Try DIRDIF if there are one or more heavy atoms present. The SIR options are identical to those already tried in *Auto/Solve*. This latter option can, however, be useful in backtracking your way through a difficult structure.

4.11 Completing Structures Automatically

You may wish to complete the structure automatically. To do this work as follows:

(1) Assign atom types, even if you are not sure, if they are not already assigned. The MODEL option enables you to do this. Now you can complete it in automatic mode.

(2) Make sure the model you propose to complete does not have more atoms of a specified type than you have indicated in the Model file e.g. if there are supposed to be 40 C atoms in your molecule, check you do not have more than this.

(3) Select *Complete/SIR* to complete using SIR (see Chapter 7) or *Complete/Dirdiff* to complete using Dirdiff (see Chapter 8).

5 Spacegroup and Test i

Programs to Determine the Space Group from Diffraction Data, Search for Alternative Symmetry, Merge the Data, Correct for Crystal Decomposition and Test for a Centre of Symmetry.

5.1 Introduction

The space group program is state of the art software designed to determine space group symmetry. It works in the following way:

- Pseudo precession photos are generated for inspection.
- It optionally corrects for crystal decay in the X-ray beam.
- The data are also optionally corrected for λ/2 contamination that occurs with all CCD devices.
- The systematic absences are scanned to determine the lattice type (P, A, I, F *etc.*).
- Internal R(int) values are computed for the 11 Laue classes to examine and quantify the diffraction symmetry. Each Laue class is associated with a set of space groups.
- The metric symmetry is examined to see if the unit cell dimensions indicate a unit cell with higher or lower symmetry than that of the input cell. Two methods are used:
    1. The construction and interpretation of the Niggli matrix.
    2. The LePage method.
- Normalised structure factor statistics, based on E-magnitudes, are used to decided if the space group is centrosymmetric or noncentrosymmetric.
- The systematic absences are examined for translational symmetry elements (glide planes, screw axes *etc.*)
- The space group is now chosen on the basis of all this information.
- The number of asymmetric units/molecules in the unit cell are computed using a database of atomic radii for various classes of compound.
- Histograms displaying data quality are produced.
- The data are optionally merged.

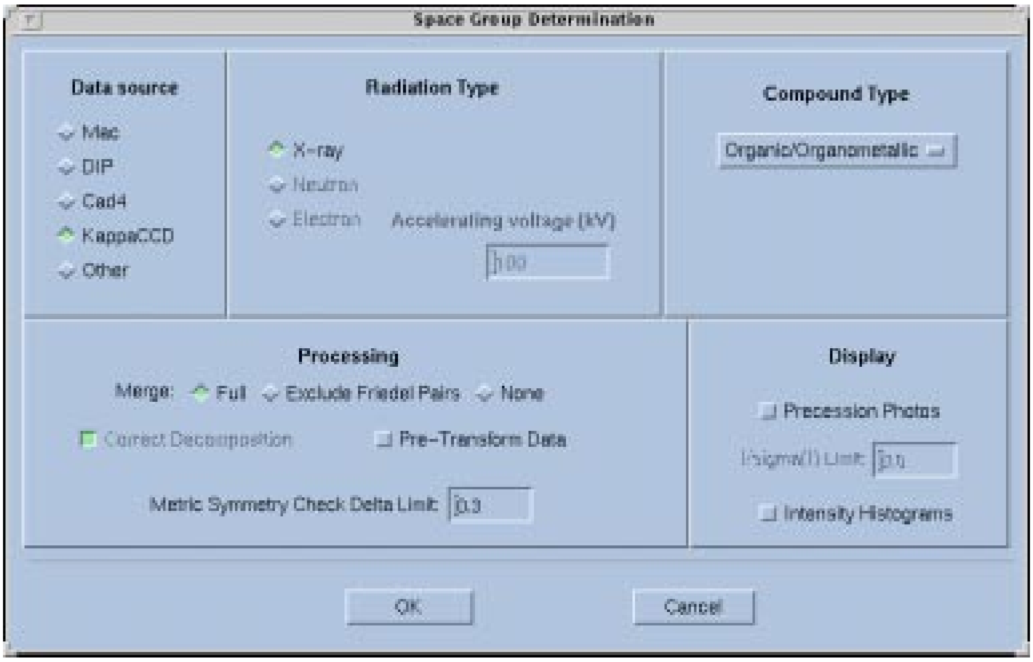
Spacegroup is the first program you must run when performing a crystal structure analysis.

It is not possible to proceed with maXus without running this program.

5.2 How to Use Spacegroup

From the main maXus window select the *Spacegroup* option from the *Init* menu.

The following dialogue box appears:



Space Group Determination

Data source
○ Mac
○ DIP
○ Cad4
● KappaCCD
○ Other

Radiation Type
● X-ray
○ Neutron
○ Electron
Accelerating voltage (kV) 1.00

Compound Type
Organic/Organometallic

Processing
Merge: ● Full ○ Exclude Friedel Pairs ○ None
☑ Correct Decomposition
☐ Pre-Transform Data
Metric Symmetry Check Delta Limit 0.3

Display
☐ Precession Photos
I/sigma(I) Limit 0.1
☐ Intensity Histograms

The choices are as follows:

5.2.1 Data Source

There are five possible source options:

- *KappaCCD*: The Nonius KappaCCD detector
- *CAD4*: The Nonius CAD-4 diffractometer
- *Mac*: A MacScience diffractometer
- *DIP*: The MacScience image plate
- *Other:* Data from synchrotrons, for example.

In general, these sources should be correctly identified for you, using information from the Original file. If, however, you have used the ASCII import then you may need to set the correct source here. Your choice is important since all of the absorption correction programs, as well as the CIF and report generators, use this information.

5.2.2 Radiation Type

Normally, this is set to X-Ray. However, if you have chosen *Other* as your data source, you can set the radiation type to *Neutron* or *Electron*. If you choose *Electron*, the Accelerating voltage (kV) should be entered into the text field; it is required for scaling purposes.

The radiation type selected is used to choose the correct scattering factors, anomalous dispersion corrections, absorption coefficients etc. **It must be correctly set at this point.**

5.2.3 Compound Type Pop-Up Menu

This gives the following options:

Intermetallic - Select this if your compound contains no C, H, N or O atoms.

Inorganic - Select this if your compound contains no C atoms, except as carbonates.

Organic/*Organometallic* - Everything else (this is the default).

If it is wrongly set, the program could calculate an incorrect value for Z (See Section 5.3.13).

5.2.4 Processing: Merge

Three options are available from the initial dialogue box. The default is Full.

- Full: This merges all equivalent reflections + Friedel pairs.
- Exclude Friedel Pairs: as above but excluding the Friedel pairs. Use this if you are doing work involving absolute configuration, and with heavy atoms present.
- None: No merging is carried out. You will receive warning messages from direct methods programs concerning duplicate reflections, but these can be ignored.

Make a suitable choice. *If absorption is going to be important, it is strongly recommended that you choose* ***None*** *and merge your data at absorption correction time.*

5.2.5 Processing: Correct Decomposition

If this option is available and on, the Original file will be searched for monitor reflections (serial diffractometer data only). They must appear first in the list, and subsequent appearances of these reflections will only be accepted if the same group of reflections are found each time. They do not need to be evenly spaced.

This option is only available for MacScience and CAD-4 data.

5.2.6 Processing: Pre-Transform Data

With this option selected, when the Spacegroup program is started, the user will be prompted to enter a 3x3 transformation matrix that transforms the unit cell and intensity data indices before the rest of the program is begun. The default is no transformation. See Section 5.3.2.

5.2.7 Processing: Metric Symmetry Check Delta Limit

The program searches for alternative metric symmetry using the LePage method (Lepage, *J. Appl. Cryst.* (1982). **15**, 255-259). In doing this, the program computes angles between planes of low order and the primitive cell. This parameter defines the maximum value of the angle which will be treated as significant. The default is 0.3 degrees.

5.2.8 Display: Precession Photos

This option produces six simulated precession photographs of the data - one for each of the 0kl, 1kl, h0l, h1l, hk0 and hk1 data.The default setting is on, and the photos appear in a postscript viewer.

Only those reflections with measured intensities, plus their Friedel pairs are displayed. For more information see Section 18.3.

In general all the reflections measured are displayed on the pseudo precession photographs, but sometimes (in cases of twinning, for example) it is useful to be able to suppress the weaker reflections. The I/sigma(I) box defines a I/σ(I) limit. If I/σ(I) is less than the set limit then the reflection is not displayed. The default value is zero.

5.2.9 Display Intensity Histograms

This option produces two histograms to help inform you about the quality of the data. The default setting is on. The graphs display the number of reflections as a function of sin θ/λ, and the number of reflections as a function of I/σ(I).

5.3 Running the Spacegroup Program

Click the OK button, and the program will begin.

(See Appendix C for a flow chart of the operations it performs.)

If the Spacegroup program has already been run, a warning box will appear:

Spacegroup has already been run

This warns you that you will be overwriting existing files and you will have to start the entire solving and refining procedure again - so be careful here. Click *Cancel* if you have changed your mind. If you click *OK* then the program begins, and a scrollable text pane appears in which any output will be displayed.

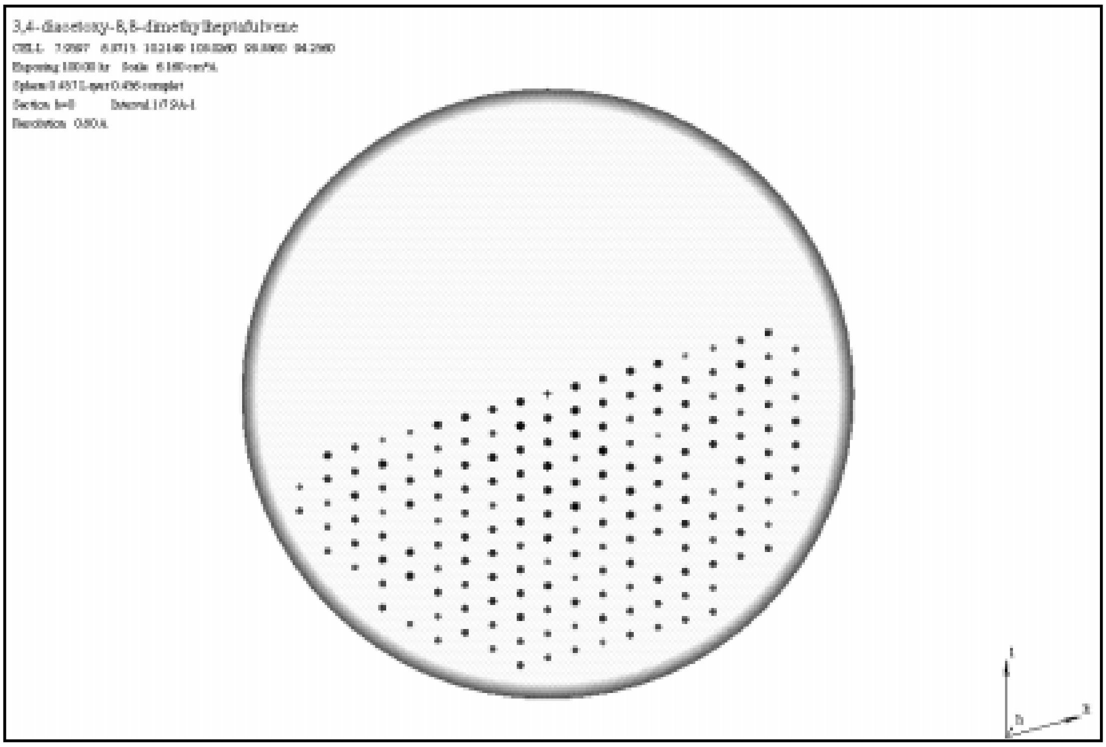
5.3.1 Pseudo Precession Photographs

If you asked for precession photographs they will appear similar to the one shown below.

Remember: Only those reflections with measured intensities, plus their Friedel pairs are displayed. Also If I/σ(I) is less than the set limit then the reflection is not displayed (where the limit was entered in the original dialogue box).

Examine these photos with care if you are having difficulties in solving the structure, *e.g*.

- Are there systematically weak reflections?
- Is there one huge reflection?
- Is there non-crystallographic symmetry?



A copy of the generated photos is saved in the PostScript file *pattern.ps* in the working directory, which can be viewed and or printed at a later date, using a PostScript viewer such as GhostScript. For more information on controlling this program see the chapter on Databases (Section 18.3).

5.3.2 Transformation of Input Data

If you selected the option to transform the unit cell and the Miller indices, you will now be prompted for the first row of the transformation matrix:

Enter Value

Enter 1st row of matrix:

1.0,0.0,0.0

Enter the required figures and press *OK*. Three number separated by spaces or commas are expected here. If you make a mistake the following dialogue box will appear

maXus

Error in matrix entry. Continue?

No cancels the transformation process. If you click *Yes* then the following box appears:

Enter Value

Re-enter 1st row of matrix:

1.0,0.0,0.0

And the process continues until all three rows have been entered. When the matrix has been completely input, you will be shown the orginal and transformed cell dimensions in the text pane, and then will be prompted:

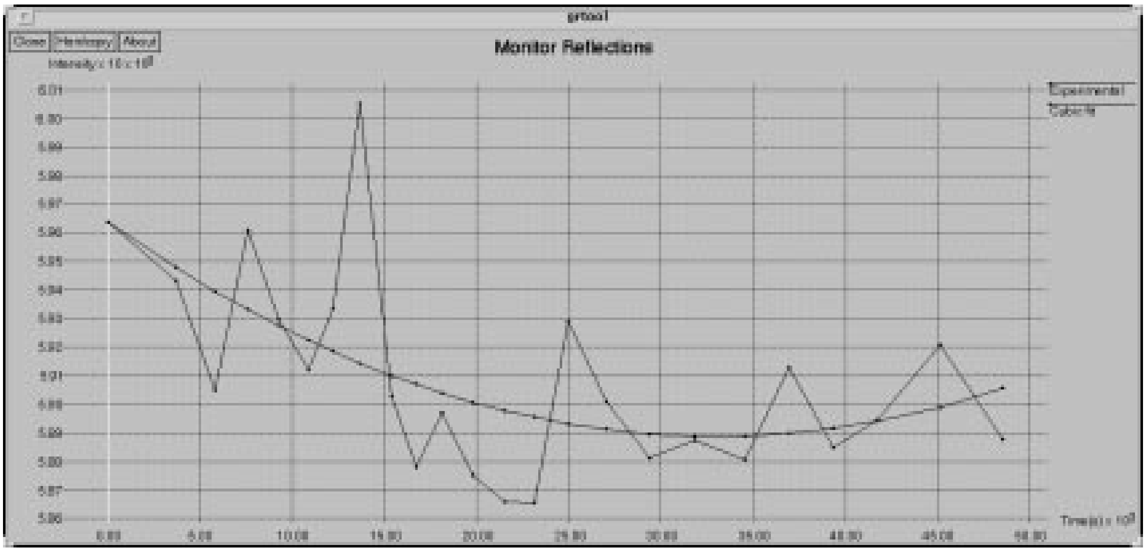
maXus

Is this OK?

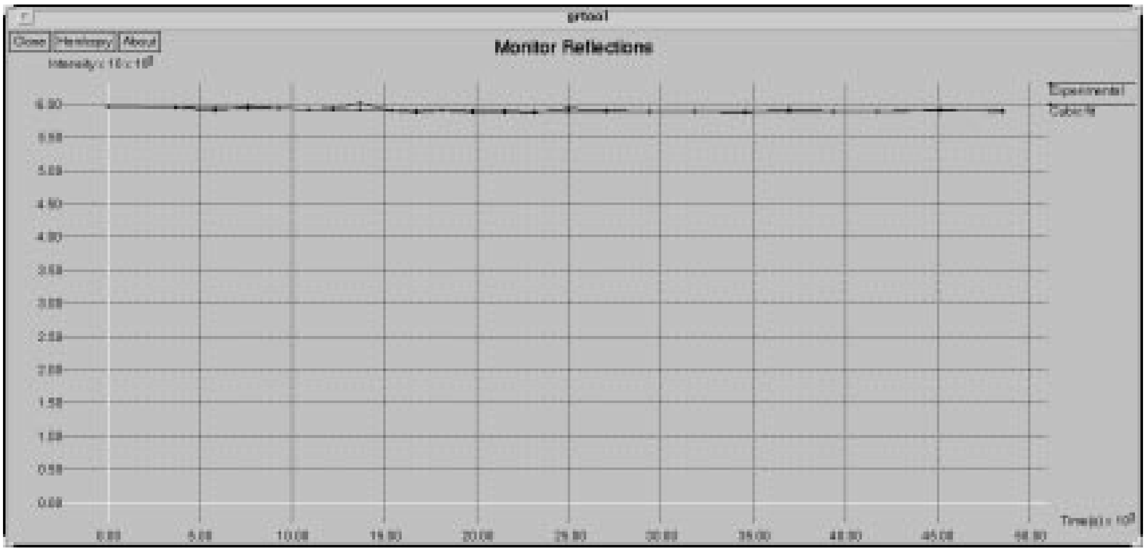
If *Yes* is selected, the program will proceed with the transformed data. Clicking *No* gives you the option to try again with a new transform, or to cancel the whole transform procedure.

5.3.3 Correction for Decay in the X-ray Beam (CAD4, MacScience only)

Following this the data are corrected for decay in the X-ray beam (assuming you have a serial diffractometer, and the decomposition option was selected). The correction process produces the following typical graphical output:



This window uses the full range in the y-axis. For a more realistic view a second window is also generated, which has a zero base line:



The solid line is a quadratic fit of the form:

$$I(n) = I(0) + an + bn^2$$

where a and b are the constants determined by non-linear least-squares, n is the reflection measurement number, $I(0)$ is the average intensity of the check reflections at the start of the data collection, and $I(n)$ is the intensity at reflection number n.

A printout may be obtained of a graph by clicking the *Hardcopy* button in the graph window, and the graphs can be dismissed by clicking *Close*.

A dialogue box will now appear:

Continue with correction?

If you choose *Yes*, the decomposition correction is now run. Clicking *No* cancels the decay correction.

Several possible errors are detected at this stage, as monitor reflections can cause problems:

- If they are not first:

Monitor reflections must be first

- or if there are too many reflections, or too many or too few sets:

>5000 Monitor reflections

<5 sets of monitor reflections

>10 standards

A faulty, or missing CAD4 decay file can cause problems:

Faulty CAD-4 decay file

No CAD-4 decay file

If there is a problem with the Original file, the following dialog box appears:

Input error on Original file

In all cases pressing *Quit* brings up the following box:

Decay correction abandoned

Click *OK* to continue. Once the decomposition correction is complete, the following box appears:

Moving Original --> Original.nodecay

If you click *OK* then the Original file is copied to Original.nodecay, and the new Original file now contains the corrected data. You can always copy back your original file if you wish using the UNIX command:

```
cp Original.nodecay Original
```

If a file called Original.nodecay already exists, the program will ask if it is OK to overwrite it. When the decay correction is complete the following dialogue box appears:

Decay Correction complete - Close Plot

5.3.4 λ/2 Contamination Correction

If your input is from a CCD device, you will now be prompted for correction for λ/2 contamination (see, Kirschbaum, Martin & Pinkerton, *J.Appl.Cryst.* (1997), **30**, 514-516). λ/2 contamination is inherent in **all** CCD detectors. The correction here is applied to the data on a *temporary basis* to help with systematic absences in marginal cases. A dialogue box appears:

Perform lambda/2 stripping?

In general, the *Yes* option is only required if you are having problems with systematic absences. The stripping procedure can take up to 30 seconds on slower computers. The stripping is only used for systematic absences and nowhere else.

5.3.5 Assignment of Lattice Type

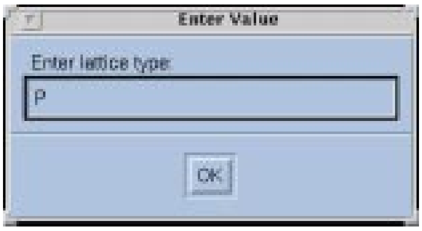
The systematic absences are examined for lattice type. Typicical output looks like this:

```
Lattice exceptions              P      A      B      C      I      F    Obv    Rev    All
 Total no of reflections        0      0      0      0   1930      0   2567   2575   3844
 Total  with I>3*sig(I)         0      0      0      0   1930      0   2567   2575   3844
 Mean intensity for all         0      0      0      01538137      0 5894411 558391 24091
 Mean intensity I>3*sig(I)      0      0      0      01538137      0 5894411 558391 24091
(Obv and Rev refer to R type cells)
Preferred lattice type is F
```

In the first row the lattice types are listed (the *Obv* and *Rev* headings refer to R type cells. *Obv* should be used to conform to IUCr standards.) In the 2nd row the total number of reflections found that violate the systematic absence for this lattice are given. The next row repeats this but only for those reflections for which I is greater than 3σ(I). The mean intensity is then given for all these reflections, followed by the mean when the weaker reflections (with I less than 3σ(I)) are removed. The preferred lattice type is determined, and checked by the user *via* a dialogue box *e.g.*

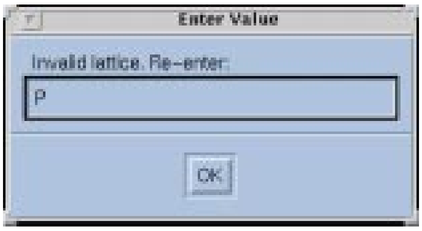
Is lattice type C OK?

If this is incorrect, and you wish to change it, click *No* and the following dialogue appears:



Enter lattice type:
P

Type your choice. Input is case insensitive; if you make a mistake, or enter an invalid lattice, the following dialogue appears:



Invalid lattice. Re-enter:
P

Enter a correct lattice type, and click OK.

5.3.6 The Niggli Matrix

The program now constructs the Niggli matrix from the lattice type and the unit cell dimensions. Typical output looks like this:

```
Unit cell:
 a =   7.9397(0.0016)          alpha= 105.03(0.014)
 b =   8.5713(0.0014)          beta =  95.56(0.017)
 c =  10.2149(0.0023)          gamma=  94.24(0.015)
 V =    664.69( 0.23) A^3
Niggli reduced cell:   7.940  8.571 10.215 105.03  95.56  94.24
Niggli matrix:        63.0388     73.4672    104.3442
                     -22.6993     -7.8523     -5.0268
Transformation matrix:     1.00    0.00    0.00
                           0.00    1.00    0.00
                           0.00    0.00    1.00
```

Niggli matrices are described in Giacovazzo, *Fundamentals of Crystallography*, OUP, 1992 pp77-80. They can be useful for detecting alternative symmetry based on metric (distance) arguments. The program uses the Niggli tables in Azaroff and Buerger, *The Powder Diffraction Method*, McGraw-Hill, 1958 pp 146-147 (also listed in Appendix B of this chapter) to see if there is an alternative lattice description. If one is found a message is produced, for example:

```
Niggli matrix indicates triclinic cell of type P
```

The main problem with Niggli matrices is defining zero. Errors are not fully propagated so decisions as to what sort of alternative cell can be found can be misleading. The LePage method described in Section 5.3.8 is much more reliable.

5.3.7 Laue Classes

The program now looks at possible equivalent reflections for the 11 Laue classes (See Burns and Glazer, ‘*Space Groups for Solid State Scientists*,’ Academic Press, 1990, pp 70-72).

A table appears in the text pane:

```
Checking Laue class from equivalent reflections:
 Laue class    R(int)     n
    -1          0.009   212
    2/m         0.313   859
    mmm         0.399   716
    4/m         0.396   629
    4/mmm       0.480   436
    -3          0.406   646
    -3m         0.480   520
    6/m         0.465   485
    6/mmm       0.520   380
    m-3         0.513   319
    m-3m        0.522   217
```

From this example you can see that -1 is the likely Laue class and, on this basis, the space group is triclinic. The correspondence is as follows:

Laue class	**Crystal class**
-1	Triclinic
2/m	Monoclinic
mmm	Orthorhombic
4/m	Tetragonal
4/mmm	Tetragonal
-3	Trigonal
-3m	Trigonal
6/m	Hexagonal
6/mmm	Hexagonal
m-3	Cubic
m-3m	Cubic

In the following example there are two choices: -1 or 2/m:

```
Checking Laue class from equivalent reflections:
 Laue class    R(int)     n
    -1          0.002    12
    2/m         0.025    99
    mmm         0.224   421
    4/m         0.389   392
    4/mmm       0.388   347
    -3          0.337    91
    -3m         0.404   260
```

```
    6/m         0.421   392
    6/mmm       0.466   347
    m-3         0.466   273
    m-3m        0.516   190
```

In this case 2/m is the obvious choice, since -1 is just Friedel’s law in action.

5.3.8 The LePage Method of Searching for Alternative Metric Symmetry

A more reliable method of searching for alternative metric symmetry comes from LePage (*J. Appl. Cryst.* (1982) 15, 255-259) and, in this implementation, is adapted from Platon (Spek, A.L., *Acta Cryst*. (1990) A46, C34).

Typical output looks like this:

```
Possible 2-Fold Axes - 2-Axis Crit = 0.50, Exp. Error = 0.30 Deg., LATT = F:

                Rows      Products       Angle Between Two Direct Axes
Nr   D    N Direct  Recip Dot Delta   1    2    3    4    5    6    7    8    9

 1  6.890 2  0 0 1  1 1 2  2 0.011    0. 120. 120. 120. 120.  90.  90. 135. 135.
 2  6.890 2  0 1-1  0 1-1  2 0.000  120.   0. 120.  90. 120. 120. 135. 135.  90.
 3  6.890 2  0 1 0  1 2 1  2 0.000  120. 120.   0. 120.  90. 120. 135.  90. 135.
 4  6.890 2  1 0 0  2 1 1  2 0.000  120.  90. 120.   0. 120. 120. 135. 135.  90.
 5  6.890 2  1 0-1  1 0-1  2 0.007  120. 120.  90. 120.   0. 120. 135.  90. 135.
 6  6.890 2  1-1 0  1-1 0  2 0.010   90. 120. 120. 120. 120.   0.  90. 135. 135.
 7  9.744 4  1 1-1  1 1 0  2 0.005   90. 135. 135. 135. 135.  90.   0.  90.  90.
 8  9.744 4  1-1 1  1 0 1  2 0.010  135. 135.  90. 135.  90. 135.  90.   0.  90.
 9  9.744 4 -1 1 1  0 1 1  2 0.010  135.  90. 135.  90. 135. 135.  90.  90.   0.

Transformation Matrix: Input (a,b,c) to Conventional Cell(a’, b’, c’):
(a’)   ( 1.00 0.00 0.00) (a)   (x’)   ( 1.00 0.00 0.00) (x)   Metrically
(b’) = ( 0.00 0.00 1.00) (b).  (y’) = ( 0.00 0.00 1.00) (y).  cubic
(c’)   ( 0.00-1.00 0.00) (c)   (z’)   ( 0.00-1.00 0.00) (z)   FOM: 0.011

                Latt     a       b       c      Alpha   Beta  Gamma      Volume
Input Cell        F     9.744   9.744   9.744   90.00  90.00  90.00      925.15
Reduc Cell        P     6.890   6.890   6.890   60.00  60.00  60.00      231.27
Conv. Cell       cF     9.744   9.744   9.744   90.00  90.00  90.00      925.15

No alternative metric symmetry detected.
```

The example input unit cell is cubic and F centred. It is possible, however, to generate other cells, and nine of these are listed in the table, but the final analysis produces the original cell. The delta parameter in the above table is an angle, and it is possible to investigate entries in this table by changing the *Metric Symmetry Check Delta Limit* option in the main dialog box (Section 5.2.7). For example setting this limit to 0.007 and re-running the program would produce the following output:

```
Checking for higher metric symmetry via LePage method:
Possible 2-Fold Axes - 2-Axis Crit = 0.50, Exp. Error = 0.01 Deg., LATT = F:
                Rows      Products       Angle Between Two Direct Axes
Nr   D    N Direct  Recip Dot Delta   1    2    3    4    5    6    7    8    9
```

```
 1  6.890 2  0 0 1  1 1 2  2 0.011    0. 120. 120. 120. 120.  90.  90. 135. 135.
 2  6.890 2  0 1-1  0 1-1  2 0.000  120.   0. 120.  90. 120. 120. 135. 135.  90.
 3  6.890 2  0 1 0  1 2 1  2 0.000  120. 120.   0. 120.  90. 120. 135.  90. 135.
 4  6.890 2  1 0 0  2 1 1  2 0.000  120.  90. 120.   0. 120. 120. 135. 135.  90.
 5  6.890 2  1 0-1  1 0-1  2 0.007  120. 120.  90. 120.   0. 120. 135.  90. 135.
 6  6.890 2  1-1 0  1-1 0  2 0.010   90. 120. 120. 120. 120.   0.  90. 135. 135.
 7  9.744 4  1 1-1  1 1 0  2 0.005   90. 135. 135. 135. 135.  90.   0.  90.  90.
 8  9.744 4  1-1 1  1 0 1  2 0.010  135. 135.  90. 135.  90. 135.  90.   0.  90.
 9  9.744 4 -1 1 1  0 1 1  2 0.010  135.  90. 135.  90. 135. 135.  90.  90.   0.

Transformation Matrix: Input (a,b,c) to Conventional Cell(a’, b’, c’):
(a’)   ( 1.00 0.00 0.00) (a)   (x’)   ( 1.00 0.00 0.00) (x)   Pseudo
(b’) = ( 0.00 0.00 1.00) (b).  (y’) = ( 0.00 0.00 1.00) (y).  cubic
(c’)   ( 0.00-1.00 0.00) (c)   (z’)   ( 0.00-1.00 0.00) (z)   FOM: 0.011

                Latt     a       b       c     Alpha   Beta  Gamma      Volume
Input Cell        F     9.744   9.744   9.744  90.00  90.00  90.00      925.15
Reduc Cell        P     6.890   6.890   6.890  60.00  60.00  60.00      231.27
Conv. Cell       cF     9.744   9.744   9.744  90.00  90.00  90.00      925.15

Checking for higher metric symmetry via LePage method:

Possible 2-Fold Axes - 2-Axis Crit = 0.50, Exp. Error = 0.01 Deg., LATT = F:
                 Rows      Products       Angle Between Two Direct Axes
Nr   D    N Direct  Recip Dot Delta   1    2    3    4    5    6    7    8    9
 1  6.890 1  0 1-1  0 1-1  2 0.000    0. 120.  90. 120. 120. 135. 135.
 2  6.890 2  0 1 0  1 2 1  2 0.000  120.   0. 120.  90. 120. 135.  90.
 3  6.890 1  1 0 0  2 1 1  2 0.000   90. 120.   0. 120. 120. 135. 135.
 4  6.890 2  1 0-1  1 0-1  2 0.007  120.  90. 120.   0. 120. 135.  90.
 5  6.890 1  1-1 0  1-1 0  2 0.010  120. 120. 120. 120.   0.  90. 135.
 6  9.744 2  1 1-1  1 1 0  2 0.005  135. 135. 135. 135.  90.   0.  90.
 7  9.744 3  1-1 1  1 0 1  2 0.010  135.  90. 135.  90. 135.  90.   0.

Checking for higher metric symmetry via LePage method:
Possible 2-Fold Axes - 2-Axis Crit = 0.50, Exp. Error = 0.01 Deg., LATT = F:
                 Rows      Products       Angle Between Two Direct Axes
Nr   D    N Direct  Recip Dot Delta   1    2    3    4    5    6    7    8    9
 1  6.890 1  0 1-1  0 1-1  2 0.000    0. 120.  90. 120. 135.
 2  6.890 1  0 1 0  1 2 1  2 0.000  120.   0. 120.  90. 135.
 3  6.890 1  1 0 0  2 1 1  2 0.000   90. 120.   0. 120. 135.
 4  6.890 1  1 0-1  1 0-1  2 0.007  120.  90. 120.   0. 135.
 5  9.744 0  1 1-1  1 1 0  2 0.005  135. 135. 135. 135.   0.
Checking for higher metric symmetry via LePage method:

Possible 2-Fold Axes - 2-Axis Crit = 0.50, Exp. Error = 0.01 Deg., LATT = F:
                 Rows      Products       Angle Between Two Direct Axes
Nr   D    N Direct  Recip Dot Delta   1    2    3    4    5    6    7    8    9
 1  6.890 1  0 1-1  0 1-1  2 0.000    0. 120.  90.
 2  6.890 0  0 1 0  1 2 1  2 0.000  120.   0. 120.
 3  6.890 1  1 0 0  2 1 1  2 0.000   90. 120.   0.
Checking for higher metric symmetry via LePage method:
Possible 2-Fold Axes - 2-Axis Crit = 0.50, Exp. Error = 0.01 Deg., LATT = F:
                 Rows      Products       Angle Between Two Direct Axes
Nr   D    N Direct  Recip Dot Delta   1    2    3    4    5    6    7    8    9
 1  6.890 0  0 1-1  0 1-1  2 0.000    0. 120.
```

```
 2  6.890 0  0 1 0  1 2 1  2 0.000  120.   0.
Transformation Matrix: Input (a,b,c) to Conventional Cell(a’, b’, c’):
(a’)   (-0.50 0.50 1.00) (a)   (x’)   ( 0.00 0.00 1.00) (x)   Metrically
(b’) = (-0.50-0.50 0.00) (b).  (y’) = (-1.00-1.00 0.00) (y).  monoclinic
(c’)   ( 0.50-0.50 0.00) (c)   (z’)   ( 1.00-1.00 1.00) (z)   FOM: 0.000

                Latt     a       b       c     Alpha   Beta  Gamma      Volume
Input Cell        F     9.744   9.744   9.744  90.00  90.00  90.00      925.15
Reduc Cell        P     6.890   6.890   6.890  60.00  60.00  60.00      231.27
Conv. Cell       mC    11.934   6.890   6.890  90.00 125.26  90.00      462.57

Possible alternative metric symmetry detected.
```

A possible alternative C centred monoclinic unit cell has been found. A dialogue box appears:

Examine alternative symmetry?

If you click *Yes*, then the whole procedure of lattice identification, Niggli matrix generation, and Laue class determination is repeated for the transformed data. After the procedure has run, you can choose whether to continue with the transformed cell, or use the original one *via* a dialogue box:

Continue with transformed data?

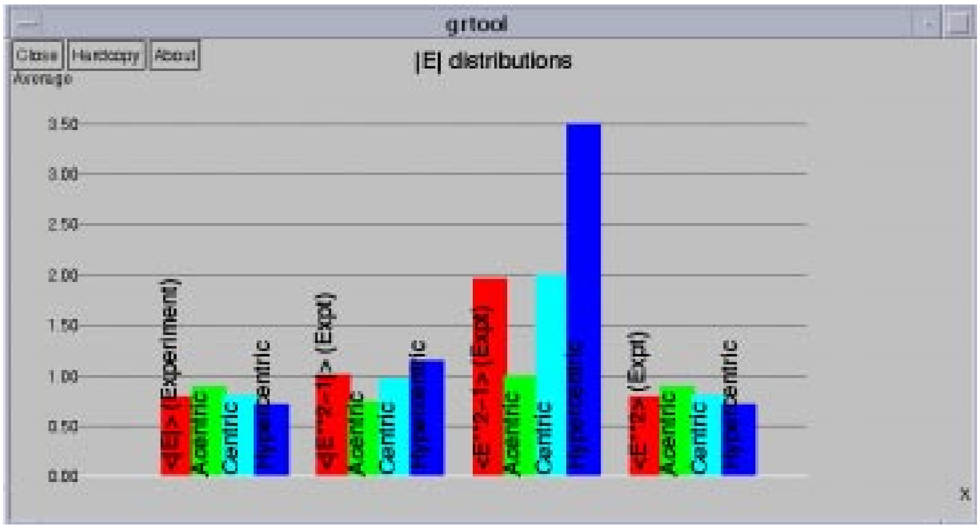
The *No* option takes you back to the original cell, the *Yes* option uses the LePage suggested cell.

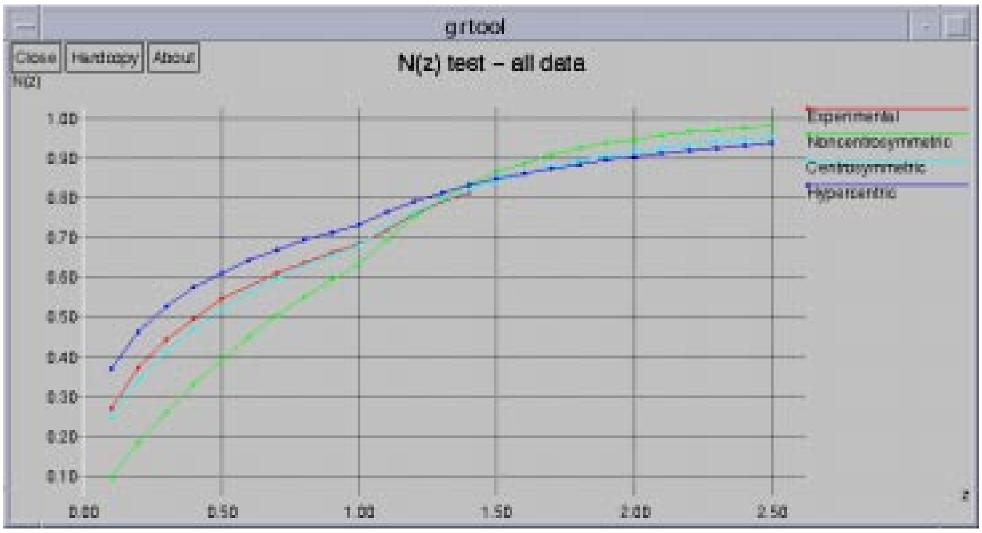
In general only use new cells from the Lepage method if you cannot solve and/or refine the structure with your chosen cell. If you are having difficulties then this method can be extremely useful.

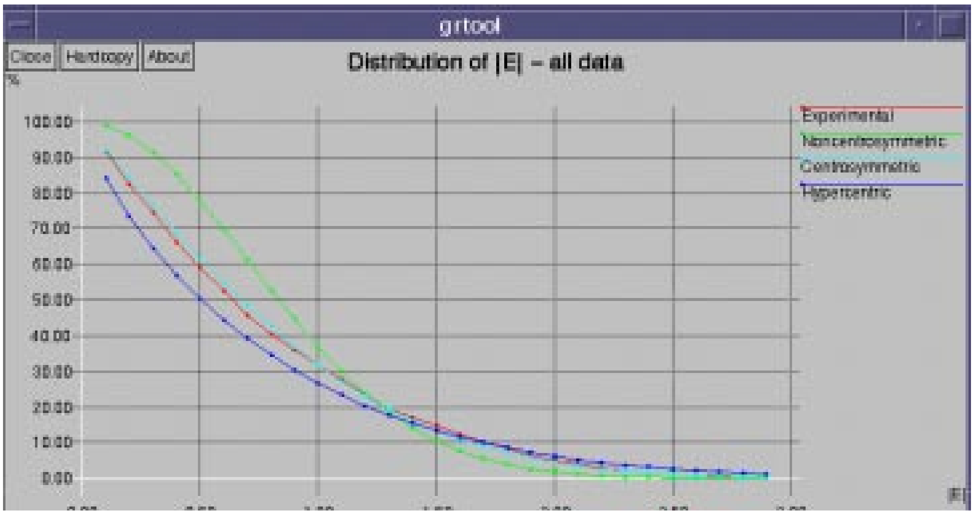
5.3.9 Testing for Centrosymmetric or Non-centrosymmetric Space Groups

Using statistics from normalised structure factors, the program now normalises the intensity data to give you statistics based on moments of E. Example output graphs are:

::


grtool
Close
Hardcopy
About
|E| distributions
Average
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
<|E|> (Experiment)
Acentric
Centric
Hypercentric
<|E**2-1|> (Expt)
Acentric
Centric
Hypercentric
<E**2-1> (Expt)
Acentric
Centric
Hypercentric
<E**2> (Expt)
Acentric
Centric
Hypercentric
X


grtool
Close
Hardcopy
About
N(z) test – all data
N(z)
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.50
1.00
1.50
2.00
2.50
z
Experimental
Noncentrosymmetric
Centrosymmetric
Hypercentric


grtool
Close
Hardcopy
About
Distribution of |E| – all data
%
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
|E|
Experimental
Noncentrosymmetric
Centrosymmetric
Hypercentric

It tries to decide if the structure is centrosymmetric or noncentrosymmetric using a chi-squared statistics based on the N(Z) test. For large data sets, a random sample of data is used; for smaller data sets, the entire set is used. In this case the output to the text pane looks like this.

```
Chisquared statistics of agreement:
Centrosymmetric: 0.0055
Noncentrosymmetric: 0.2159
Space group is probably centrosymmetric
```

Based on these statistics, the program decides if the space group is likely centrosymmetric, non-centrosymmetric, or hypersymmetric. A dialog box appears with the result:

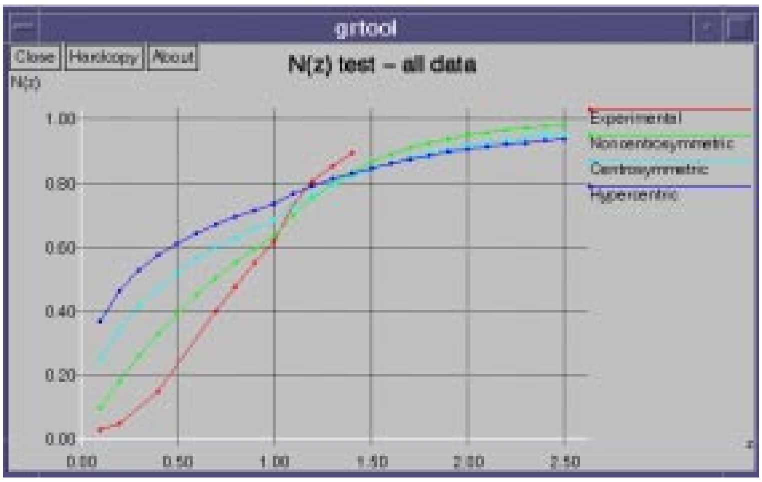
Space group probably centrosymmetric

Remember that these tests are based on Wilson statistics and are only strictly valid if:

- There are no heavy atoms present.
- Atoms are not in special positions.
- There is no pseudosymmetry
- The structure is not too small (say less than 10 non-H atoms)
- There is no excessive structural regularity.

For a full discussion see Shmeuli and Weiss, *Introduction to Crystallographic Statistics*, Oxford University Press, 1996.

For example, here is the plot of a Pt-Cl complex where the Pt and Cl atoms all lies in special positions:



(The true spacegroup is centrosymmetic, but the program auggests it is non-centrosymmetric.) If there are subtle differences, then re-running this section of the Spacegroup program separately using the *Test i* menu option (see Section 5.5) may be of more help.

5.3.10 Systematic Absences

The systematic absences are now explored to look for translational symmetry elements. Typical output looks like this:

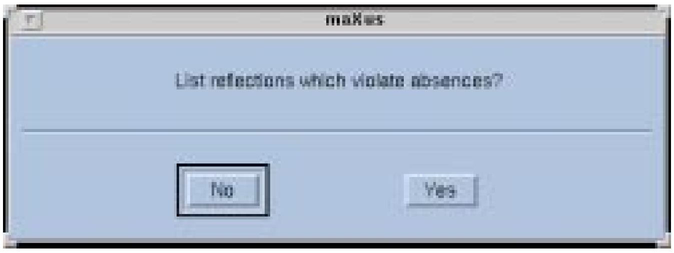
```
Checking systematic absences:
   Reflection class      For this condition  For the reverse
                         <F^2>  No of reflns <F^2>  No of reflns

    hkl:  k+l=2n         1420.7283  10328    1515.8818  10242
    hkl:  h+l=2n         1420.7283  10328    1515.8818  10242
    hkl:  h+k=2n         1468.2540  20568       1.9881      2   *
    hkl: h+k+l=2n        1373.0370   9736    1553.5410  10834
    hkl: h+k,k+l=2n      1420.8657  10327    1515.7340  10243
    0kl:   k =2n         2607.9910    795       1.9881      1   *
    0kl:   l =2n         3875.3730    377    1461.4303    419
    0kl:  k+l=2n         3865.1260    378    1464.9218    418
    h0l:   h =2n         2446.5842    740       1.9881      1   *
    h0l:   l =2n         5125.7417    353       2.8020    388   *
    h0l:  h+l=2n         5111.2681    354       2.8041    387   *
    hk0:   h =2n         1071.5219    619    1953.7457    695
    hk0:   k =2n         1071.5219    619    1953.7457    695
    hk0:  h+k=2n         1538.1475   1314       0.0000      0   *
    h00:   h =2n         6038.9531     36       0.0000      0   *
    0k0:   k =2n         7069.1064     35       0.0000      0   *
    00l:   l =2n        11790.4688     22       5.9104     36   *
 In the above list, an ’*’ signifies a possible absence.
 List of suggested systematic absences:
    hkl:  h+k=2n
    0kl:   k =2n
    h0l:   h =2n
    h0l:   l =2n
    h0l:  h+l=2n
    hk0:  h+k=2n
    h00:   h =2n
    0k0:   k =2n
    00l:   l =2n
```

If any of the main classes of reflections (h00, 0k0, 00l, hkl) have less than three entries, or are missing completely, a warning is given:

Warning! <3 hk0 reflections

The program continues to run after you click *OK*, although you should be aware that this is a potential problem, and may affect the correct determination of the space group.

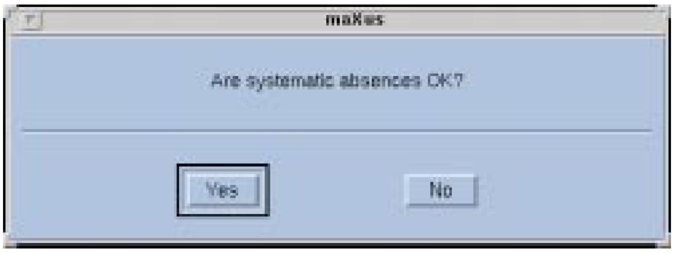
A dialogue box will now appear, asking if you wish to be given a list of reflections which violate the systematic absences:



List reflections which violate absences?

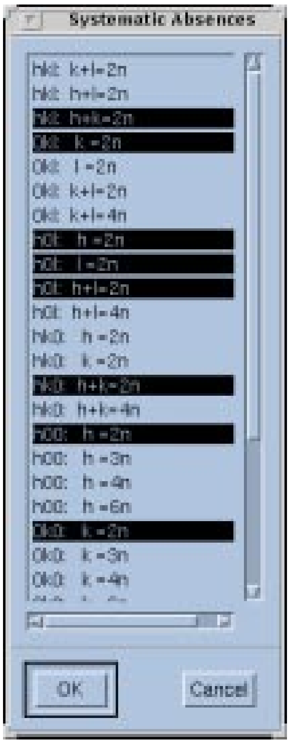
If you select *Yes*, a list of the violating reflections appears in the text window.

A dialogue box will appear to ask if the absences displayed are correct:



Are systematic absences OK?

If they are acceptable click *Yes*; the program then moves on to determine the space group. If not then *No* brings up the systematic absences browser:



```
hkl: k+l=2n
hkl: h+l=2n
hkl: h+k=2n
0kl: k =2n
0kl: l =2n
0kl: k+l=2n
0kl: k+l=4n
h0l: h =2n
h0l: l =2n
h0l: h+l=2n
h0l: h+l=4n
hk0: h =2n
hk0: k =2n
hk0: h+k=2n
hk0: h+k=4n
h00: h =2n
h00: h =3n
h00: h =4n
h00: h =6n
0k0: k =2n
0k0: k =3n
0k0: k =4n
```

The highlighted lines are those suggested by the program. Click on them once to deselect; click once on the absences you require to highlight them. Once you have selected all the absences needed, click *OK*.

5.3.11 Selecting the Space Group

A dialogue box now lets you choose possible space groups, based on the information gathered so far. For example:

```
Num  Name        %  C   NZ Rint
 1   P 1        1%[*]   1 0.009
 2   P -1      13%      2 0.009
```

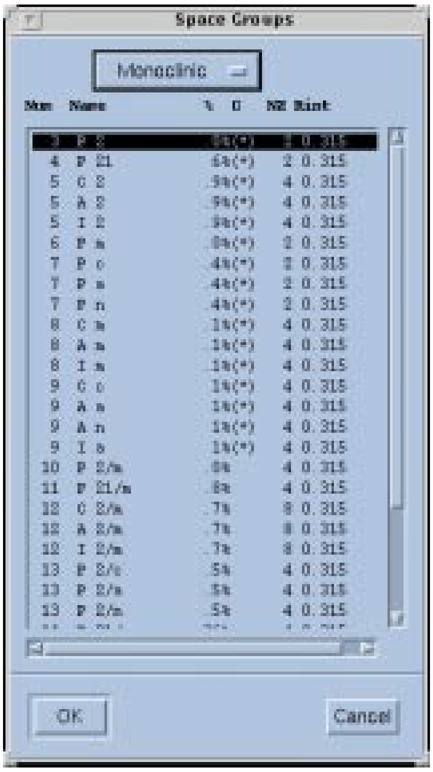
The various entries in the dialogue box for each group are (from left to right):

- The space group number.
- The spacegroup name
- The % occurrence in the Cambridge Structural Database.
- A (*) signifies a non-centrosymmetric space group
- NZ is the number of equivalent positions.
- Rint is the R(int) for the Laue class corresponding to that particular space group. It should be less than 10% for most cases.

Click on your group of your choice, and then *OK*.

Alternatively, if the group required is not shown, or the program was unable to suggest any space groups base upon the information available to it, use the space group browser by clicking *Browse...*

The space group browser appears set for the suggested crystal class:



```
Num  Name          %  C     NZ Rint
  3  P 2        .0%(*)      2 0.315
  4  P 21       .6%(*)      2 0.315
  5  C 2        .9%(*)      4 0.315
  5  A 2        .9%(*)      4 0.315
  5  I 2        .9%(*)      4 0.315
  6  P m        .0%(*)      2 0.315
  7  P c        .4%(*)      2 0.315
  7  P a        .4%(*)      2 0.315
  7  P n        .4%(*)      2 0.315
  8  C m        .1%(*)      4 0.315
  8  A m        .1%(*)      4 0.315
  8  I m        .1%(*)      4 0.315
  9  C c        .1%(*)      4 0.315
  9  A a        .1%(*)      4 0.315
  9  A n        .1%(*)      4 0.315
  9  I a        .1%(*)      4 0.315
 10  P 2/m      .0%         4 0.315
 11  P 21/m     .8%         4 0.315
 12  C 2/m      .7%         8 0.315
 12  A 2/m      .7%         8 0.315
 12  I 2/m      .7%         8 0.315
 13  P 2/c      .5%         4 0.315
 13  P 2/a      .5%         4 0.315
 13  P 2/n      .5%         4 0.315
```

You can use the pop-up menu at the top to select from alternative crystal classes. Click once on the group of your choice, and then click *OK*.

5.3.12 Non Standard Settings

For monoclinic and orthorhombic crystal systems, once a space group has been selected, the program will check to see if the chosen space group is a non-standard setting of another, conventional, space group.

For example, if the space group *I2* is selected the following dialogue box will appear:

Non Standard setting of C 2 spacegroup. Transform?

Selecting *No* abandons the transformation. If you choose *Yes* the unit cell parameters and the Miller indices will be transformed to the new setting for *C2* and this will be used in subsequent structure calculations. In the case of monoclinic systems, the space group transformation can effect the value of the beta angle and may result in a new unit cell setting in which the value of β is not in the standard range, giving a unit cell that looks distorted.

In such cases, a message will appear in the text pane:

```
The value of beta in the transformed cell is 131.29
This is not in the standard range (90-120 deg)
```

A dialog box appears asking if you wish to continue:

Beta <90 or >120. Continue?

If not, the transformation is abandoned.

Note that this is only a warning given by the program about the value of β, and transformation of the space group setting can be continued without any problems. After transformation, β is always obtuse.

If the space group transformation is carried out the following message is displayed e.g:

```
Space Group transformed from I2 to C2
Unit Cell transformed from
7.940 8.571 10.215 90.0 95.56 90.0
to 13.531 8.571 7.940 90.0 131.29 90.0
```

5.3.13 The Number of Formula Units in the Unit Cell

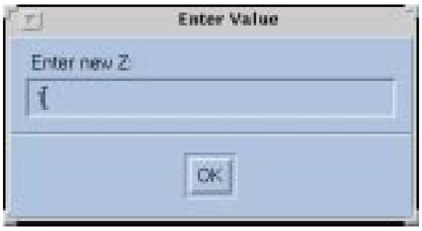
The program uses an empirical method to calculate the value of Z, the number of formula units in the unit cell, by using the molecular formula, and the atomic valences (see Appendix A).

Typical output from Spacegroup to the text pane at this stage is:

```
       Original input unit cell contents information:
 C14 H16 O4           Z= 2 fw= 248.30 dobs=1.200 dcal=1.240
 The compound is of type organic or organometallic
 The molar mass is    248.278 au
 The calculated value of Z is   2.07
 The selected value of Z is  2
 The value of F(000) based on this value of Z is    264 e
 The calculated crystal density is   1.241 g/cm^3
 The (observed volume)/(predicted volume) is  1.034
```

Z is chosen to be an integral mutliple of NZ. You will be asked if you wish to alter the calculated value of Z (number of formula units per unit cell):

If you choose *Yes*, you can enter a new value:



The text pane will show results of calculation using the new value (2 in this example):

```
Based on new Z value input by user:
 The calculated value of Z is   2.07
 The selected value of Z is  2
 The value of F(000) based on this value of Z is    264 e
 The calculated crystal density is   1.241 g/cm^3
 The (observed volume)/(predicted volume) is  1.034
 No. of formula units in the asymmetric unit =  2.000
```

You will be asked again ito confirm that you are happy with the value of Z.

5.3.14 Non-standard Wavelengths

If the value of the wavelength recorded in the data file is not recognised as one of the standard settings, the program requires you to enter values of df', df'' and mu (μ) for each element type.

df' and df'' are scattering factors, and μ is an absorption coefficient. These values are stored and used later in maXus.

The following type of dialog appears for each element:

Enter df',df'',mu for C

0. 0. 0.

The three variables are df', df'', and μ respectively.

Once the constants have been entered for each element, you are then asked for the type of data collection device used:

Enter data collection device:

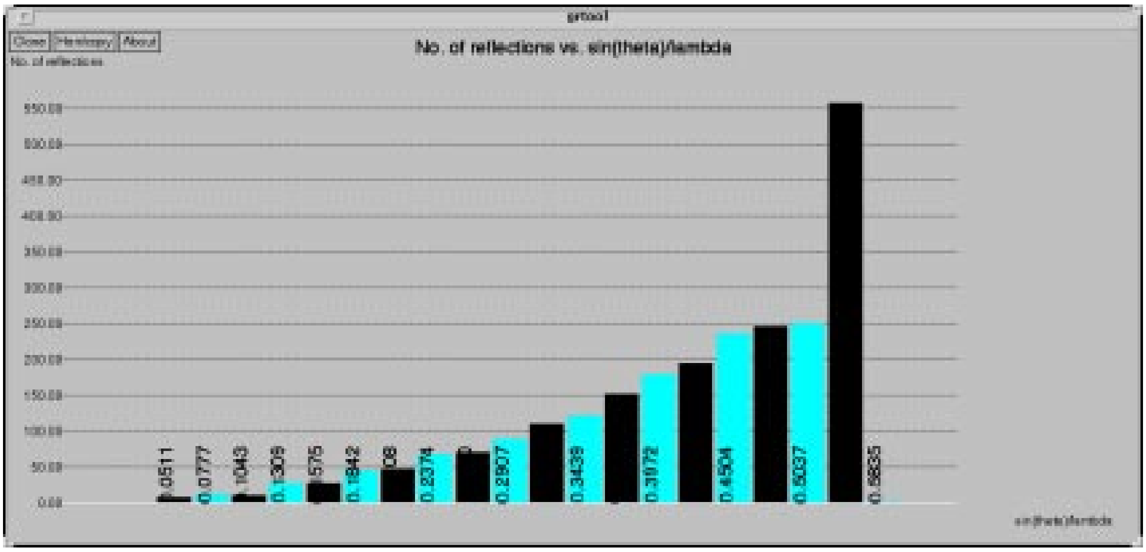
Synchrotron

The default is Synchrotron.

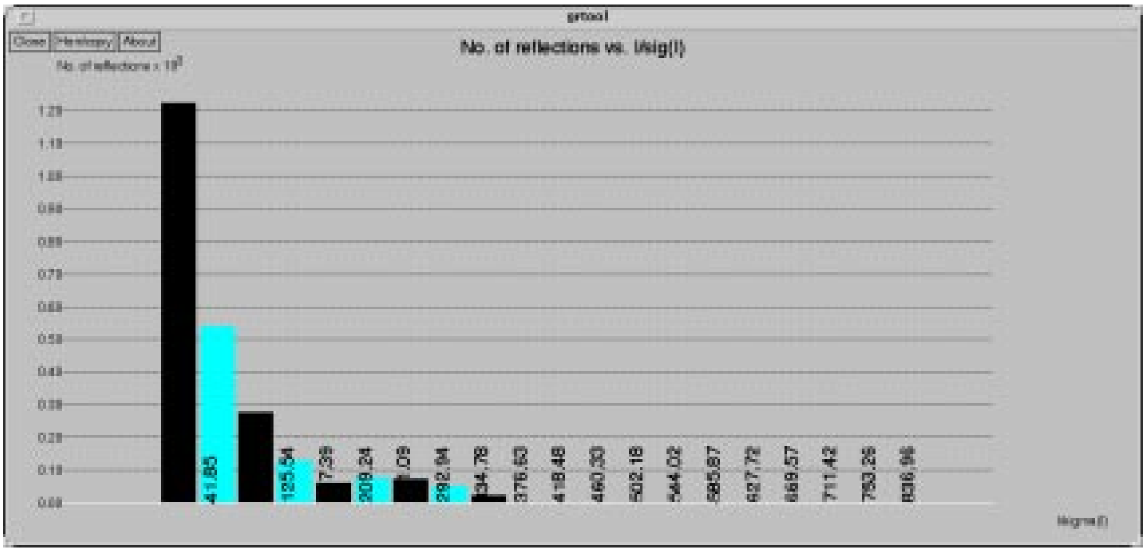
If you have used one of the standard wavelengths, these dialogs do not appear.

5.3.15 Output Histograms.

If you have selected this option in the original window, two histograms then appear which demonstrate graphically the quality of the input data. The first shows the number of reflections as a function of sin θ/λ. It should increase exponentially as in the following example:



The second shows the number of reflections as a function of I/σ(I) and is also a measure of data quality:



5.3.16 Completion

When the program has finished successfully, the message

All Done

will appear. Be sure to examine any graphs of interest before clicking *OK*, as doing this will dimiss them all.

5.3.17 Error Windows for Spacegroup software

Several error windows are associated with the program. Various checks are made of the intensity data. The messages are as follows:

Input error on Original file

This is a general error condition. *Quit* aborts from the program.

Laue Group Error

The Laue group is stored in the header of the Original file. There is a faulty entry here.

Too many reflections

There are more than 400,000 reflections on the Original file.

File faulty – rdfl line not found

This means the rdfile header in the Orginal file is missing. Replace or edit the file.

Read error on Original File

Unexpected EOF on Original

These messages mean the Original file itself is faulty. Try replacing it with a fresh copy, or re-import your diffractometer data, and run Spacegroup again.

5.4 Space Group Settings not in the Space Group Database.

maXus employs a database of space groups stored in an appropriate library directory which cannot be directly edited by the user. It contains a large number of standard and non-standard space groups.

You may wish to use a non-standard setting that is not in this list. If so, it is possible to edit the database to add your own non-standard settings. See Section 18.2.1 in Chapter 18.

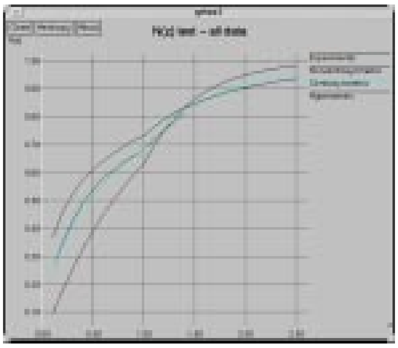
5.5 Test i Program

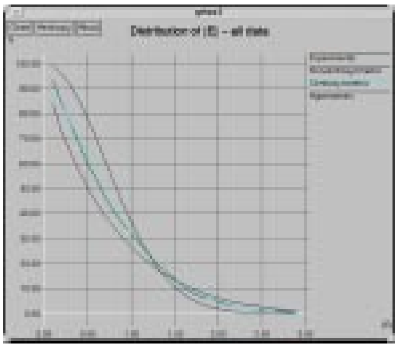
5.5.1 Running Test i

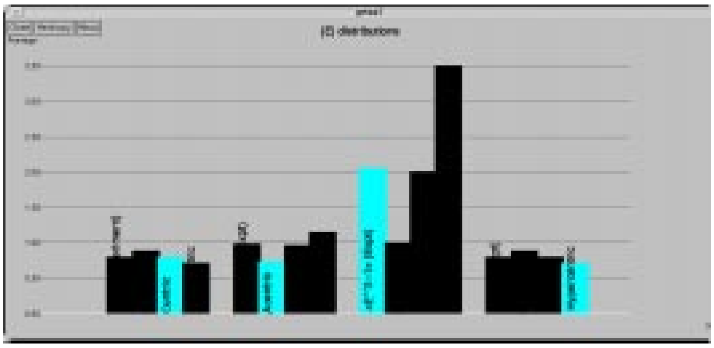
If, even after running Spacegroup you are still unsure whether you have a centrosymmetric or non-centrosymmetric crystal, this option lets you re-examine the evidence. Invoke the *Test i* option from the *Init* menu. Running the program at this point is slightly more accurate than using the Spacegroup results since duplicate reflections are removed, there is more output and correct statistcal weights are given to the E values. Often the differences are slight, however. The Test i progress indicator will appear:

A text pane appears which produces a set of intensity information based on Wilson statistics. Note that for large data sets, only a random sample of data is used; for smaller data sets, the entire set is used. There is a copy of this information in the file Testi which can be viewed with the *Summary/ Test i* option from the main maXus menu bar.

The program also brings up several graphics windows to display the data, and allow examination. These are of N(Z) tests, and |E| distributions, and examples of them are shown below:







When the program is complete the following dialogue box appears:

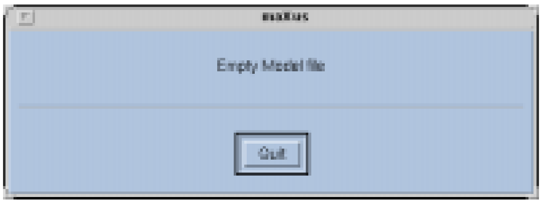
Test i complete

If you wish to print or further examine the graphical output, you should do so before clicking *OK* here, as this will cause the text pane, progress indicator and graphics windows to disappear.

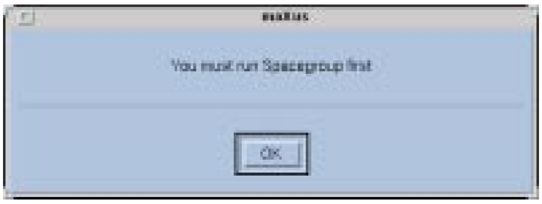
You may wish to return to the Spacegroup option at this point, and if necessary select a different space group.

5.5.2 Test i errors

If you have not run Spacegroup first then one of two possible error dialogue boxes will appear:



Empty Model file



You must run Spacegroup first

5.5.3 Limits of Test i

Remember that Test i is based on Wilson statistics. It has only strict validity if:

- There are no heavy atoms present
- There is no pseudosymmetry
- The structure is not too small (say less than 10 non H atoms)
- There is no excessive structural regularity.

For a full discussion see Shmeuli and Weiss, *Introduction to Crystallographic Statistics*, Oxford University Press, 1996.

Appendix A

Data used in calculation of Z

Source: NBS Technical Note 1190

Empirical Method for the Approximation of Densities - J K Stalick

Element Data

Element No.	Symbol	Organics & organometallics (9493)	Intermetallic compounds (5875)	Inorganics (2693)
1	H	5.40	5.5	6.7
2	He			
3	Li	18.0	14.5	6.5
4	Be	24.5	8	11
5	B	13.0.	7	10.5
6	C	13.9	6.5	22
7	N	11.0	5.5	12
8	O	10.0	10.0	10
9	F	11.5	11.5	11.5
10	He	22.0	22	22
11	Na	18.0	32.5	16.5
12	Mg	38.0	23	5
13	Al	34.5	16	13
14	Si	39.0	16	20.5
15	P	30.0	16.5	19.5
16	S	26.0	20.5	24.5
17	Cl	25.0	23	28.5
18	Ar	30.0	30	30
19	K	31.0	45	27
20	Ca	41.0	30	18
21	Sc	21.0	20.5	12
22	Ti	25.5	14.5	11
23	V	39.0	12	17.5
24	Cr	34.0	11	17
25	Mn	31.5	12	11
26	Fe	29.0	11	11.5
27	Co	27.0	10	8
28	Ni	26.5	10	5.5
29	Cu	24.0	10.5	9

Element Data

Element No.	Symbol	Organics & organometallics (9493)	Intermetallic compounds (5875)	Inorganics (2693)
30	Zn	35.0	15	11.5
31	Ga	39.0	17	11.5
32	Ge	39.0	18.5	16
33	As	36.5	23	21.5
34	Se	28.5	25.5	31
35	Br	30.5	27	34
36	Kr	40.0	40	40
37	Rb	39.0	47	36
38	Sr	39.0	44	23.5
39	Y	37.0	28	14.5
40	Zr	40.0	22	19
41	Nb	35.5	15	23.5
42	Mo	37.0	13	25
43	Tc	37.0	16	34
44	Ru	41.0	14	13
45	Rh	39.0	13	16.5
46	Pd	34.5	13	13.5
47	Ag	29.5	19	20
48	Cd	34.0	22.5	15.5
49	In	64.0	24	24.5
50	Sn	49.0	30	21.5
51	Sb	46.0	29	24.5
52	Te	48.0	34.5	25.5
53	I	44.0	40	51
54	Xe	53.5	53.5	53.5
55	Cs	53.0	58	47
56	Ba	41.0	49	29.5
57	La	65.0	34.5	23
58	Ce	60.0	32	17.5
59	Pr	60.0	31	19
60	Nd	60.0	30	19.5
61	Pm	55.0	40	25
62	Sm	55.0	29	18.5
63	Eu	55.0	32.5	20.5
64	Gd	50.0	29	18.5
65	Tb	50.0	28.5	17
66	Dy	50.0	27	16
67	Ho	50.0	26.5	16.5

Element Data

Element No.	Symbol	Organics & organometallics (9493)	Intermetallic compounds (5875)	Inorganics (2693)
68	Er	50.0	26	15
69	Tm	45.0	25.5	15.5
70	Yb	45.0	27.5	15.5
71	Lu	45.0	25	14
72	Hf	45.0	22.5	18.5
73	Ta	43.0	15.5	18.5
74	W	46.0	15	25.5
75	Re	51.0	11	10
76	Os	43.0	13	26
77	Ir	29.5	14	10
78	Pt	42.0	15	12.5
79	Au	31.0	15	22
80	Hg	38.0	21.5	23
81	Tl	58.0	32.5	32
82	Pb	43.0	31	25.5
83	Bi	49.0	34	20.5
84	Po	70.0	30	35
85	At	70.0	30	35
86	Rn	70.0	30	35
87	Fr	70.0	30	35
88	Ra	70.0	54	39
89	Ac	70.0	30	35
90	Th	85.0	30.5	24
91	Pa	70.0	26	22
92	U	61.0	24	28.5
93	Np	70.0	30	19
94	Pu	70.0	27	16
95	Am	70.0	26.5	20
96	Cm	70.0	24	35
97	Bk	70.0	30	35
98	Cf	70.0	30	35
99	Es	70.0	30	35
100	Fm	70.0	30	35
101	Md	70.0	30	35
102	No	70.0	30	35
103	Lr	70.0	30	35

Appendix B

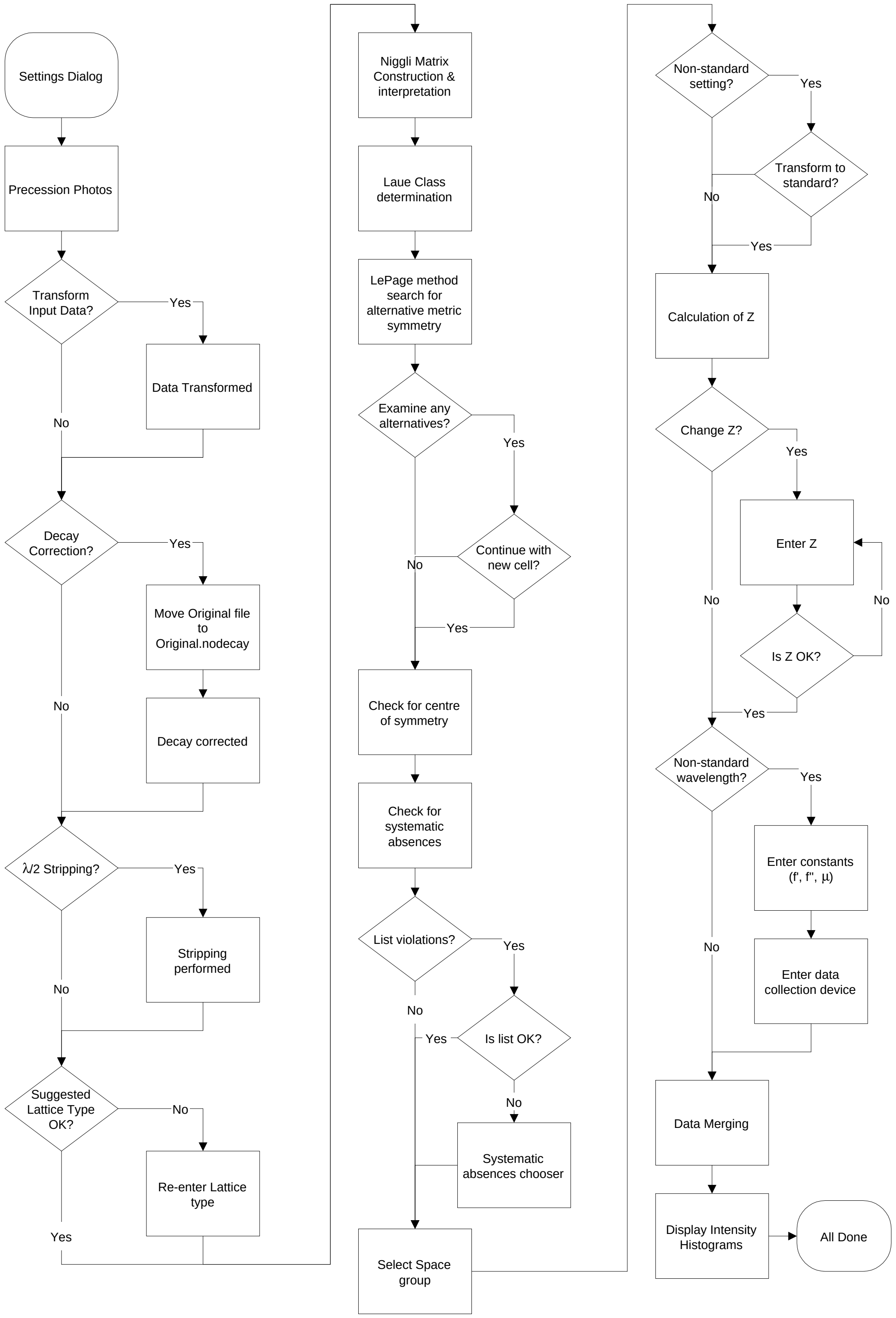
Niggli Matrices

	P	I	F	C	R
Cubic	$\begin{bmatrix} s_{11} & s_{11} & s_{11} \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} s_{11} & s_{11} & s_{11} \\ \frac{1}{3}\bar{s}_{11} & \frac{1}{3}\bar{s}_{11} & \frac{1}{3}\bar{s}_{11} \end{bmatrix}$	$\begin{bmatrix} s_{11} & s_{11} & s_{11} \\ \frac{1}{2}s_{11} & \frac{1}{2}s_{11} & \frac{1}{2}s_{11} \end{bmatrix}$		
Tetragonal	$\begin{bmatrix} s_{11} & s_{11} & s_{33} \\ 0 & 0 & 0 \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{22} \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} s_{11} & s_{11} & s_{33} \\ \frac{1}{2}\bar{s}_{11} & \frac{1}{2}\bar{s}_{11} & 0 \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{11} & s_{11} \\ \frac{1}{2}(\bar{s}_{11} - \bar{s}_{12}) & \frac{1}{2}(\bar{s}_{11} - \bar{s}_{12}) & s_{12} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{22} \\ \frac{1}{4}s_{11} & \frac{1}{2}s_{11} & \frac{1}{2}s_{11} \end{bmatrix}$			
Hexagonal	$\begin{bmatrix} s_{11} & s_{11} & s_{33} \\ 0 & 0 & \frac{1}{2}\bar{s}_{11} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{22} \\ \frac{1}{2}\bar{s}_{22} & 0 & 0 \end{bmatrix}$				$\begin{bmatrix} s_{11} & s_{11} & s_{33} \\ \frac{1}{2}s_{11} & \frac{1}{2}s_{11} & \frac{1}{2}s_{11} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{11} & s_{11} \\ s_{23} & s_{23} & s_{23} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{11} & s_{11} \\ \bar{s}_{23} & \bar{s}_{23} & \bar{s}_{23} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{22} \\ \frac{1}{2}\left(\bar{s}_{22} - \frac{1}{3}\bar{s}_{11}\right) & \frac{1}{3}\bar{s}_{11} & \frac{1}{3}\bar{s}_{11} \end{bmatrix}$

	P	I	F	C	R
Orthorhombic	$\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} s_{11} & s_{11} & s_{11} \\ \bar{s}_{23} & \bar{s}_{31} & (\bar{s}_{11}-\bar{s}_{23}-\bar{s}_{31}) \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{22} \\ s_{23} & \frac{1}{2}s_{11} & \frac{1}{2}s_{11} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \frac{1}{2}\bar{s}_{22} & \frac{1}{2}\bar{s}_{11} & 0 \end{bmatrix}$	$\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \bar{s}_{23} & \bar{s}_{23} & (\bar{s}_{11}-2\bar{s}_{23}) \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ s_{11} & s_{11} & s_{11} \end{bmatrix}$	$\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ 0 & \frac{1}{2}\bar{s}_{11} & 0 \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ 0 & 0 & \frac{1}{2}\bar{s}_{11} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \frac{1}{2}\bar{s}_{22} & 0 & 0 \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{11} & s_{33} \\ 0 & 0 & \bar{s}_{12} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{22} \\ \bar{s}_{23} & 0 & 0 \end{bmatrix}$	
Monoclinic	$\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ 0 & \bar{s}_{31} & 0 \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \bar{s}_{23} & 0 & 0 \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ 0 & 0 & \bar{s}_{12} \end{bmatrix}$	$\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \frac{1}{2}\bar{s}_{22} & \frac{1}{2}\bar{s}_{11} & \bar{s}_{12} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{11} & s_{33} \\ \bar{s}_{23} & \bar{s}_{31} & (\bar{s}_{11}-\bar{s}_{23}-\bar{s}_{31}) \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \bar{s}_{23} & \frac{1}{2}s_{11} & s_{11} \end{bmatrix}$	$\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \bar{s}_{23} & 0 & \frac{1}{2}\bar{s}_{11} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \frac{1}{2}\bar{s}_{22} & \bar{s}_{31} & 0 \end{bmatrix}$ $\begin{bmatrix} \bar{s}_{11} & s_{22} & s_{33} \\ \bar{s}_{23} & \frac{1}{2}\bar{s}_{11} & 0 \end{bmatrix}$	$\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \bar{s}_{23} & \bar{s}_{23} & \bar{s}_{12} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{22} \\ \bar{s}_{23} & \bar{s}_{21} & \bar{s}_{21} \end{bmatrix}$	$\begin{bmatrix} s_{11} & s_{11} & s_{33} \\ \frac{1}{2}s_{12} & \frac{1}{2}s_{11} & s_{12} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \frac{1}{2}s_{22} & s_{12} & s_{12} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \frac{1}{2}s_{31} & s_{31} & \frac{1}{2}s_{11} \end{bmatrix}$
Triclinic	$\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ s_{23} & s_{31} & s_{12} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{22} & s_{33} \\ \bar{s}_{23} & \bar{s}_{31} & \bar{s}_{12} \end{bmatrix}$				

Appendix C

Flowchart showing Spacegroup operation



6 Absorption Corrections

Six Absorption Correction Programs:
Numerical, Sortav, Psi-scan,
Difabs, Spherical and
Cylindrical

6.1 Introduction

There are six independent absorption correction programs within the maXus package, each with differing levels of mathematical complexity. Pull down the *Abs* menu. There are five entries:

- Numerical
- Sortav
- Difabs
- Psiscan
- Spherical
- Cylindrical

For all these methods, except Sortav, there are two further choices in a sub-menu:

- Processed
- Original

Processed is the usual option.This corrects the data files after the Spacegroup program has run, and allows the user to select either corrected or non-corrected data sets in least-squares.

Original corrects the Original file itself i.e. the raw diffractometer data. This enables the user to correct the data at the start of the analysis once and for all, but it is not possible for the user to select either corrected or non-corrected data sets in least-squares. This can be useful in cases of strong absorption where the lack of a correction can interfere with the data merging, the space group determination and the structure solution *via* direct methods or Patterson techniques.

NOTE: It is quite permissible to try all five methods (excluding Sortav): each program writes a unique corrected data file which can be selected at least-squares time so that all possible methods can be tried, and the one giving the best results selected for the final analysis. With Sortav you can try a complete analysis with and without the absorption correction.

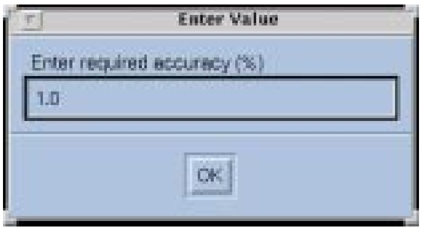
6.2 NUMABS Numerical Correction

6.2.1 Running the Program

This program performs the absorption correction by Gaussian quadrature over the crystal volume, with a variable grid selected to give 1% accuracy as a default. For each reflection it requires the direction cosines of the incident and diffracted beams and also a precise knowledge of the crystal morphology. The mean X-ray path length is calculated and can be used in the refinement of the extinction parameter in least squares. The code is an implementation of the procedure described by Coppens, 1970.

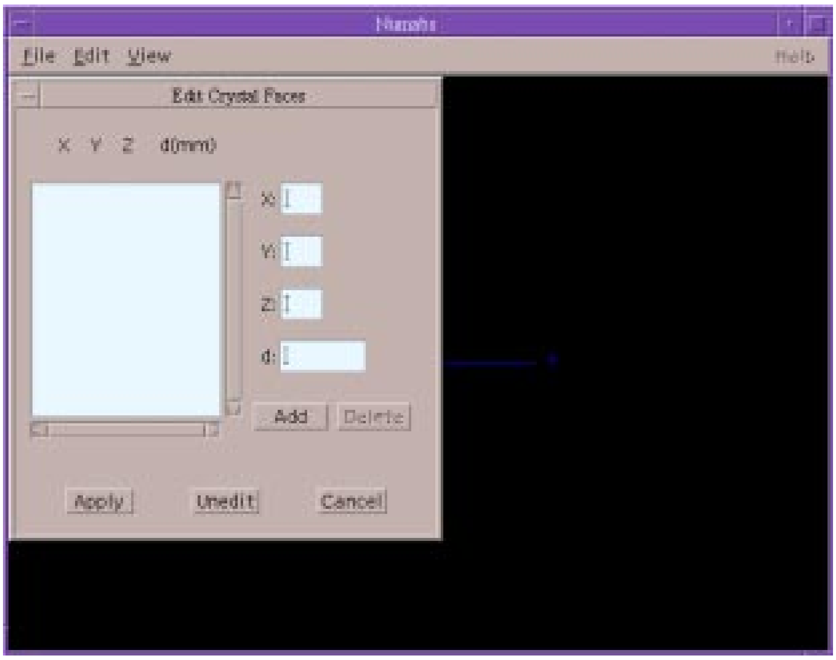
Choose one of the *Numerical* options from the *Abs* menu.

You will be given the opportunity to change the default accuracy level via a dialog box:

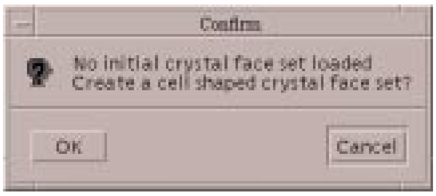


Enter required accuracy (%)
1.0

The following windows will then appear:



Unless you have imported the data from a CIF file which contained the crystal morphology (see below), the following dialog will appear:



No initial crystal face set loaded
Create a cell shaped crystal face set?

Selecting *Cancel* terminates the absorption correction.

Click *OK* allows the user to create a new crystal shape information set.

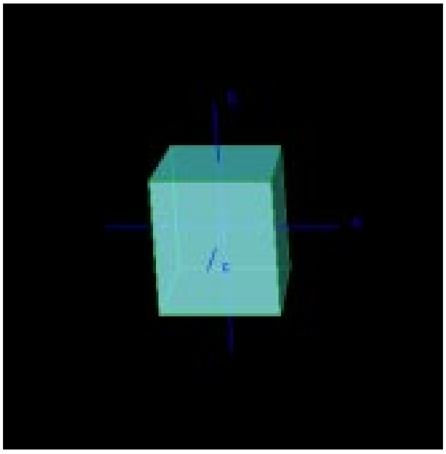
The following default values appear in the *Edit Crystal Faces* window:

X	Y	Z	d(mm)
1	0	0	0.3
0	1	0	0.4
0	0	1	0.5
-1	0	0	0.3
0	-1	0	0.4
0	0	-1	0.5

X: 0
Y: 0
Z: -1
d: 0.5107

6.2.2 Main Numabs window

The main Numabs display window at this stage will be showing only the axes. Clicking *Apply* in the *Edit Crystal Faces* window makes Numabs draw the crystal shape. As can be seen from below, the default values correspond to a cube:



This can be rotated in three-dimensions by dragging with the left mouse button, and translated by dragging with the middle mouse button.

The right mouse button in this display gives a pop-up menu allowing the *Opacity* of the cystal display to be controlled; the user may select from:

- 100%
- 75%
- 50%
- 25%

Clicking on the crystal itself brings up the *Edit Crystal Faces* window.

From the main Numabs window several menus are available:

File menu: **Save and Run**

This option should be selected once you are happy with the crystal shape. It starst the actual absorption correction process itself - see Section 6.2.4 .

Exit

This option terminates the Numabs program without performing the absorption correction. The program will check that you really do mean to select this option.

Edit menu: **Crystal Faces**

This option brings the *Edit Crystal Faces* window to the front.

View menu: **Zoom in** (Keyboard equivalent: z)

This option zooms in on the crystal image.

Zoom out (Keyboard equivalent: shift-z)

This option zooms out of the image.

Pick HKL (Keyboard equivalent: p)

Selecting this option displays the hkl values on the axes passing through the crystal. (Selecting a low value of *Opacity* may prove helpful for clarity here).

Clicking on one of the ‘dots’ next to the hkl numbers on a particular axis makes that number appear in the *Edit Crystal Faces* window.

For example, clicking on a -1 on the x-axis makes a -1 appear for the x-value.

Reset View (Keyboard equivalent: r)

This option simply resets to the default view of the image

6.2.3 Entering the crystal morphology

You can enter the indexed crystal faces in any of three ways:

The easiest way is to enter the data directly into NUMABS. It creates the FaceDat file to save the information you enter.

Editing and Adding Current Entries:

This is done by selecting the entry you wish to edit in the *Edit Crystal Faces* window by clicking once on it - it should become highlighted.

The numbers in the X, Y, Z and d text boxes down the right-hand side should change to reflect the current setting of your selected plane.

Make any required changes to the values, and click *ADD* when you are happy. If you have just changed d and not X, Y or Z, the program will check that you do wish to replace the current entry with your modified one.

If your alterations involved changes to X, Y or Z, clicking *ADD* will result in a new plane being added to the list.

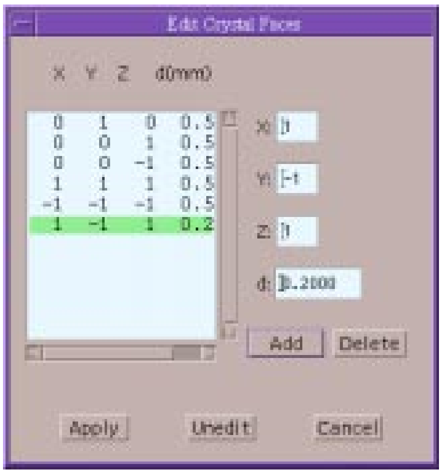
Clicking *DELETE* removes the selected plane.

UNEDIT reverts to the last saved condition (ie the last time you clicked *Apply)*.

APPLY redraws the crystal in the main NUMABS window according to the planes you have defined (in order to define a solid bounded the object must have a minimum of 6 faces!).

Example:

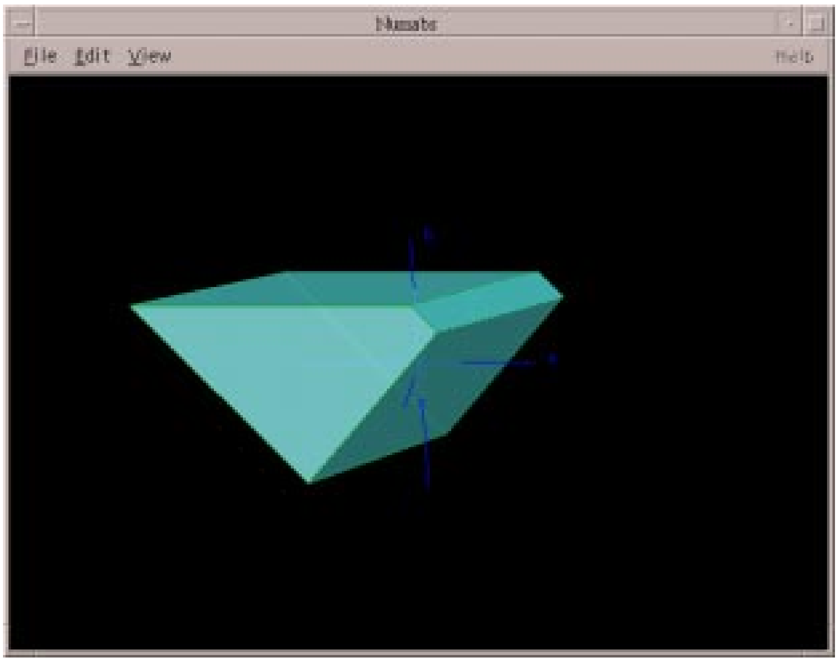
With the following planes entered:



X	Y	Z	d(mm)
0	1	0	0.5
0	0	1	0.5
0	0	-1	0.5
1	1	1	0.5
-1	-1	-1	0.5
1	-1	1	0.2

X: 1
Y: -1
Z: 1
d: 0.2000

Clicking *APPLY* gives the following crystal morphology:



- Alternatively, if you have imported data from a CIF file, which contained the morphology information in the form:

```
_loop
_exptl_crystal_face_index_h
_exptl_crystal_face_index_k
_exptl_crystal_face_index_l
_exptl_crystal_face_per_dist
0 0 1 0.12
1 -1 2 0.23
```

etc....

then the importer will have saved this data in a file called FaceDat, which will be automatically read by NUMABS, and the information, and crystal, displayed. You are now ready to run the correction. (See Section 6.2.4).

- The final option is to import the data directly by creating and/or editing a file named FaceDat in your working directory. The input must follow the format:

First line: Number of lines of data

Number of planes followed by h k l d(mm) - for example:

```
6
1 1 1 0.5
```

```
1 -1 1 0.5
0 1 0 0.2
0 0 1 0.5
-1 -1 -1 0.5
0 0 0 -1 0.4
```

Within the lines, the format is free.

NUMABS checks to see if this file is present when it is run; if so, it reads in the data from it. Any additions or alterations you make to the data are then saved to this file. A maximum of 50 faces is permitted.

6.2.4 Execution

When you are satisfied with the crystal morphology, select *SAVE AND RUN* from the *File* menu. This will invoke the absorption correction routines. Initially, the morphology is checked, and if the crystal vertices which are input produce an impossible solid, then the program will prompt the user to examine the input and edit where necessary. The problem may not be immediately obvious and so it is often advisable to keep the crystal model as simple as possible, when the dimensions are subject to reasonable errors.

Selecting *EXIT* will stop further execution and will leave the data file in the state that existed before the program was called.

Once the correction is complete, you will be prompted about merging the data. The following dialogue box appears:

Do you want to merge the data?

If you choose *No* then no merging is carried out. If however, you select *Yes* a second box appears:

Full merge or exclude Friedel pairs?

Make your selection from the two options. Data merging is then performed. A dialogue box will tell you when it is completed:

Sort-Merge of type EFP completed

The output file, HklAbsN, contains the corrected $|F_o|^2$. The corrected data will generally effect the overall scaling parameter in the refinement. However, the least squares refinement procedures will automatically perform a scale adjustment prior to an overall refinement. It is possible to examine the full output from this program by choosing 'Abs' from the 'Summary' menu. A file HklabsN is written calculated which includes X-ray beam path lengths for extinction refinement in least squares. In order that the correction be effective, great care must be taken in calculating the crystal dimensions.

6.2.5 Error Messages

Numerical absorption correction programs require the direction cosines of the incident and diffracted beams. For MacScience diffractometers these are calculated in the program; for CAD4 users a file should be present called DCosines (generated by the CAD4 import program) containing the direction cosines in formatted form. If this file cannot be found a dialogue box appears:

No direction cosines

Click *Stop* for the program to terminate. If the direction cosines file is faulty in any way, the following message appears:

Faulty direction cosines

>9500 grid points

The crystal is too complex for the calculation to proceed. Consult support.

Not a usable 3-D solid

The faces you have defined do not form a closed three-dimensional solid. Re-edit the faces list by clicking *Redo*, or quit with the *Stop* button.

Numabs terminating

Program is terminating with an error.

Numabs complete

This is normal program termination.

6.2.6 Correcting the Original File

The Original file is copied to Original.bak before any corrections are made. You can use this if things go wrong by copying it back to Original using the UNIX command

```
cp Original.bak Original
```

If you get the dialogue box:

Original.bak already exists

Warning: Original.bak file exists

This means that there is already an Original.bak file in existence which will be overwritten. This is just a precautionary warning. The calculations proceed in the usual way.

Faulty Original File

This indicates a faulty Original file. Try replacing it with another copy, and running again.

You may also then need to rerun Spacegroup.

6.3 SORTAV

SORTAV uses a heavily redundant data set to model an anisotropic absorption surface from a crystal whose isotropic dimensions (sphere or plate, for example) are known approximately. It was written by Bob Blessing at the Haupman-Woodward Institute in Buffalo, USA. You can only use this program under the following circumstances:

- KappaCCD data only.
- Direction cosines must be on the CIF file containing the intensity data.
- A large body of redundant data is present.

As well as correcting for absorbtion, SORTAV can merge data, reject statistical outliers and merge data sets or CCD frames.

IMPORTANT: You can only use Sortav with SINGLE DATA SETS, not those from multiple crystals.

6.3.1 Using SORTAV

To use it to correct for absorbtion proceed as follows:

- Import a CIF file in the usual way. This generates an Original file.
- Run Spacegroup. Do NOT merge the data.
- Run SORTAV. A backup copy of the Original file is copied into Original.bak. If this file already exists then be careful - you may have already run SORTAV on this data. If this is the case, a dialogue box appears:

Original.bak already exists

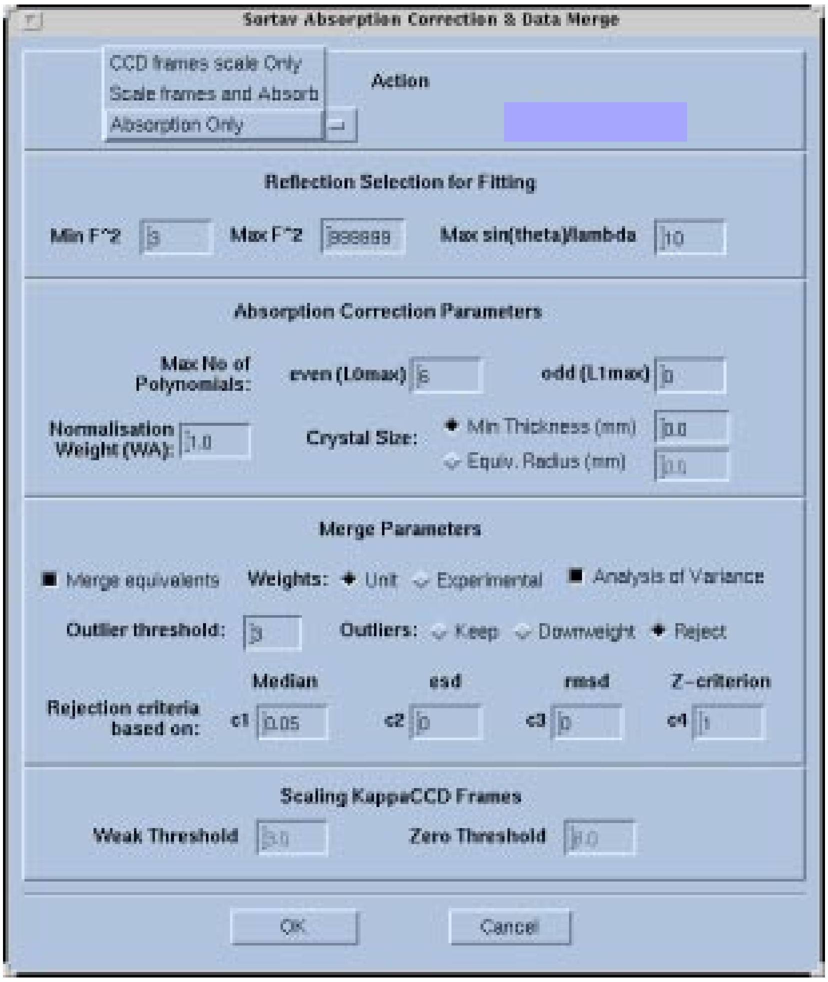
If you continue, then you get reminded with a second box:

Be very careful here

- Run Spacegroup again. You can merge if you wish at this stage, or within SORTAV itself.
- Now proceed with your structure analysis.

6.3.2 SORTAV Main Window

When you invoke SORTAV, the following window appears:



Min F^2 3
Max F^2 999999
Max sin(theta)/lambda 10
Max No of Polynomials: even (L0max) 6
odd (L1max) 0
Normalisation Weight (WA): 1.0
Crystal Size: ● Min Thickness (mm) 0.0
○ Equiv. Radius (mm) 0.0
■ Merge equivalents
Weights: ● Unit ○ Experimental
■ Analysis of Variance
Outlier threshold: 3
Outliers: ○ Keep ○ Downweight ● Reject
Rejection criteria based on: Median c1 0.05 esd c2 0 rmsd c3 0 Z-criterion c4 1
Weak Threshold 3.0
Zero Threshold 8.0

The options for presented SORTAV are as follows:

6.3.3 Action Pop-Up Menu

The options are:

- Absorption only
- Scale frames & Absorb
- KappaCCD frames scale only

Usually, you want absorption only. Be careful here as your choice may invalidate any absorption correction. Often there are too few reflections for the other options to work. If you select Scale frames or CCD frames, you also need to set KappaCCD scaling factors - see Section 6.3.13.

6.3.4 Reflection Selection for Fitting:

Reflections to be used in computing the absorption correction surface are selected here:

- **Min F^2** - i.e. the minimum F^2 value.
- **Max F^2** - i.e. the maximum F^2 value.
- **Max sin(theta)/lambda** - maximum value of sin θ/λ.

The selected data must be within the limits you have just defined.

- **Max sin(theta)/lambda** can be used to exclude from the fitting high-angle data for which absorption effects are relatively small;
- **Min F^2** can be used to exclude data too weak to carry much information about the absorption anisotropy;
- **Max F^2** can be used to exclude data too strong to be free of anisotropic extinction.

Default values are:

- **Min F^2** - 3
- **Max F^2** - 999999
- **Max sin(theta)/lambda** - 10

6.3.5 Absorption Correction Parameters: Max. No. of Polynomials

SORTAV uses spherical harmonics to map the absorption surface.

L0MAX and L1MAX are the even and odd order limits of the real spherical harmonic Y(l,m) expansion for fitting the absorption anisotropy.

For a crystal bathed in a homogeneous incident beam, the transmission surface is, in principle, centrosymmetric, since reversal of the beam direction gives the same transmission. This implies that the Y(l,m) fit should be limited to the even order, $l = 2n$, functions. Odd order, $l = 2n + 1$, functions can be used if there is also a problem with an inhomogeneous, quasiparallel beam incident from a crystal monochromator, or with a crystal that is too large or not well enough centered to be uniformly illuminated in all orientations.

Default values for a first run are L0MAX = 6 and L1MAX = 0. Increasing L0MAX may increase the accuracy of your correction, but can lead to negative transmission factors.

6.3.6 Absorption Correction Parameters: Normalisation Weight (WA)

SORTAV carries out a weighted constrained least squares optimisation. WA is a proportionality constant for weighting an absorption anisotropy restraint residual toward *A(aniso)* = 1.

The choice of an appropriate WA value is a matter of the user’s judgement of how tightly the restraint $A_{hi} = 1$ should apply in the case of the particular data set at hand, but, in general, a value of unity is a good choice.

6.3.7 Absorption Correction Parameters: Crystal Size

You can choose to define either the minimum thickness (mm) for a regular solid, or the equivalent radius (mm) for a spherical crystal.

This value is used to calculate a spherical crystal part of the overall transmission factor, which is defined to be $A = (A(sphere))/(A(aniso))$ where *A(aniso)* is the fitted absorption anisotropy correction factor (i.e. the reciprocal transmission anisotropy factor). The spherical crystal part introduces a θ dependence. If the radius is supplied as zero but non-zero values are supplied for the minimum thickness, the program will estimate radius from $A(sphere) = (A(limit))/(A(max))$ where $A(limit) = \exp(-\mu Tmin)$, and *A(max)* is the maximum transmission anisotropy factor, i.e. the reciprocal of the minimum absorption anisotropy correction factor *A(aniso)*, calculated during the Y(l,m) fitting. The estimate of the radius from minimum thickness, and *A(max)* is obtained by interpolation in the table of $A(sphere) = A(\mu R, \theta)$ from the International Tables for X-ray Crystallography, Vol. B.

If the radius is estimated from minimum thickness, and *A(max)*, the user should be careful to verify that a reasonable radius is obtained. An unreasonable radius can be obtained if the equivalent data do not thoroughly sample the transmission paths through the crystal, and the fitted parameters produce an unreasonably large *A(max)*.

If both radius and minimum thickness are supplied as zero, the program computes only an absorption anisotropy correction - essentially an anisotropic scaling - and any scattering angle dependence of the absorption correction is neglected.

You must input either the min thickness or the radius value in the text field, otherwise the absorption correction is unduly simplified.

In general, the results are not very sensitive to the crystal size.

6.3.8 Merge Parameters: Merge Equivalents Button

SORTAV can merge equivalent reflections. If this option is selected, then the new Original file contains merged data.

6.3.9 Merge Parameters: Weights Buttons

The user can select from one of two options:

- Unit weights

- Experimental weights

The default is unit weights. In most circumstances, especially for a first run of the program to identify gross outliers, one should use unit weights.

Although the relative error, $(\sigma(F^2))/F^2$, is generally larger for small F^2 than for large F^2, the absolute error, $\sigma(F^2)$, is generally smaller for small F^2 than for large F^2. Thus, if experimental weights, $W = 1/(\sigma(F^2))^2$, are used for averaging, y_{mean} is biased toward the small F^2, i.e. toward the measurements with negative errors.

For a “good” sample of n multiple equivalent measurements, one expects approximately the same F^2 and $\sigma(F^2)$ for all n measurements. Thus, approximately constant weights, i.e., unit weights, are generally appropriate for averaging.

Experimental weights are useful when one is merging data from two or more different experiments - different crystals, wavelengths, scan speeds, etc. - with significantly different average levels of random error; however, experimental weights should be used only after the gross outlier measurements have been identified, and rejected from the data set, based on a preliminary run using unit weights. The input file of control data for the program allows for a list of measurements to be rejected.

6.3.10 Analysis of Variance Button

With this button turned on, as in the default setting, the variation of the ratios $q = (rmsd)/(esd)$ is analysed as a function of $y = Fo^2$ and $s = (\sin\theta)/\lambda$ in two ways:

1. The unique data are classified in intervals of *y* and *s*, and a two-way table function of $Q(Y, S)$ compiled. the table entries are the values $(rmsd)/(esd)$ averaged within the (Y, S) blocks defined by the *Y* and *S* intervals. As the table is being compiled, the mean value of Q, Q_{mean}, and the root-mean-square deviation of the Q values from their mean, *rmsdq*, are evaluated.

2. A quadratic surface,

$$Q(Y, S) = (YS1)(A_{11}A_{12}A_{13})(Y)(A_{21}A_{22}A_{23})(S)(A_{31}A_{32}A_{33})$$

$$Q(Y, S) = (A_{11}Y)^2 + (A_{22}Ys)^2 + A_{33} + 2(A_{12}YS) + 2(A_{13}Y) + 2(A_{23}S)$$

is fitted to the (YI, SI, QI) data by least-squares to minimize $\chi^2 = \sum(WI(QI - Q(YI, SI))^2)$ with $WI = NI - 1$, where NI is the number of multiple equivalent measurements of the *i*-th unique reflection.

The goodness-of-fit, $Z = \sqrt{\left[\frac{\chi^2}{\sum(WI)}\right]\left[\frac{Nobs}{Nobs - Npar}\right]}$, where $Npar = 6$ coefficients (A_{11}, A_{22}, A_{33}, $A_{12} = A_{21}$, $A_{13} = A_{31}$, and $A_{23} = A_{32}$), and an R-factor, $R = \sqrt{\left[\frac{\chi^2}{\sum(WI)(QI)^2}\right]}$, are evaluated.

The esd's are then revised according to revised $esd = MAX(esd, Q(Y, S)esd, RMSD)$ where $Q(Y, S)$ is calculated from the fitted surface, if $Z < RMSDQ$, or looked-up in the table (without interpolation), if $Z \geq RMSDQ$.

The analysis of variance should be performed when data from a fair sampling of the Y and S ranges of the data set have been multiply measured.

6.3.11 Merge Parameters: Outlier Buttons

Three options are available:

- Keep
- Downweight
- Reject

Selecting *Keep* results in outlying data being included.

Downweight uses estimated normal relative probabilities to downweigh outlying data.

Reject uses the various fields and coefficients defined in Section Figure 6.3.12 to disregard outlying data beyond particular values.

6.3.12 Merge Parameters: Outlier threshold & Rejection Criteria

If you are merging data, the parameters c1 - c4 define the parameters used for rejecting outliers.

The program permits rejection of measurements F^2 with $ABS[F^2 - MEDIAN(F^2)] > T$, where

$$\begin{aligned} T = MAX\Big(& c1 \times YMEDIAN, c2 \times MEDIAN[\sigma(F^2)], \\ & c3 \times MEDIAN[ABS(F^2 - YMEDIAN)] \times \sqrt{\frac{N}{N-1}}, \\ & c4 \times ZCRIT(N) \times MAX(MEDIAN[\sigma(F^2)]), MEDIAN[ABS(F^2 - YMEDIAN)] \times \sqrt{\frac{N}{N-1}}\Big) \end{aligned}$$

Reasonable trial values are provided as a default:

C1 = 0.05 C2 = 0.0 C3 = 0.0 C4 = 1.0

The rationale for these values is as follows:

C1 = 0.05 because if values of $\sigma(F^2)$ are based only or mainly on counting statistics they might seriously underestimate the population standard deviation for the strong reflections. Thus even F^2 values within C2x100% of *YMEDIAN* might be rejected by the test against $\sigma(F^2)$. For weak reflections, the $\sigma(F^2)$ are essentially determined by counting statistics, but for very strong reflections, the $\sigma(F^2)$ can be essentially independent of counting statistics.

C2 = C3 = 0 and C4 = 1 to default to *ZCRIT(N)*σ.

ZCRIT(N) is the value of $Z = (ABS(\Delta))/\sigma$ corresponding to a normal probability $P = 1/(2N)$ that *Z>ZCRIT.*

N = 2, 3, 4, 5, 10, 20, 50, 100, 300, 1000

ZCRIT = 1.15, 1.38, 1.54, 1.65, 1.96, 2.24, 2.57, 2.81, 3.14, 3.48

Chauvenet’s criterion (see References, Section 6.3.16) holds that data with $Z > ZCRIT$ are sufficiently improbable in a sample of n data to be rejected. The user can override the Chauvenet criterion by supplying a negative value for C4. Indeed, any or all of the tests can be suppressed by supplying negative values for the test coefficients.

For each sample of N equivalent measurements, the test for rejection is performed only once, using the initial estimate *YMEAN* = *YMEDIAN*. Repetition of the test after recalculation of *YMEAN* could lead to eventual rejection of all n measurements. for problem cases in which the scatter of the measurements is so great that all n measurements, or all but one measurement, are rejected by the test, the initial estimates *YMEAN* = *YMEDIAN*, $ESD = MEDIAN[\sigma(YI)]$, and $RMSD = MEDIAN[ABS(YI - YMEDIAN)] \times \sqrt{N/(N-1)}$ are retained.

The program permits outlier down-weighting based on estimated relative normal probabilities.

After optional rejection of abnormal outliers from *YMAX* and/or *YMEDIAN*, the median of the remaining sample is taken to be an initial estimate of the sample mean, and the larger of either the median experimental error estimate or the median absolute deviation from the median is taken as an estimate of the sample standard deviation.

Then, with $\mu = MEDIAN(F^2)$ and, depending on the user’s choice, either $\sigma = \sigma(F^2)$ or $\sigma = MAX[MEDIAN(\sigma[YI])], MEDIAN(ABS(YI - MEDIAN(YI)) \times \sqrt{N/(N-1)})$, the relative normal probability of each measurement is estimated as $WI = \exp[-0.5 \times ((YI - \mu)/\sigma)^2]$, and used as a weight for calculating *YMEAN, ESD*, and *RMSD*.

6.3.13 Scaling KappaCCD Frames: Weak threshold, Zero threshold

This is only invoked if merging is required - i.e. if ‘Scale Frames & Absorb,’ or ‘CCD Frames Scale Only’ are selected from the pop-up menu (Section 5.3.3).

- Weak Threshold - threshold for excluding weak reflections from the fit of inter-subset scale factors. A measurement is excluded if: $F^2/(\sigma(F^2))$ < Weak Threshold
- Zero Threshold - threshold for assigning zero weight to extreme outliers in the scale factor fitting. A measurement is assigned zero weight if

$$ABS\left(F^2 - \frac{F^2}{\chi}\right)/(\sigma(F^2)) > (ZeroThreshold) \times max(Z, 1.0)$$

where Z is the standardized rms error of fit from the proceeding cycle.

Defaults are 3.0 and 6.0 respectively for these two variables. In general, the merging of frames is not recommended, because:

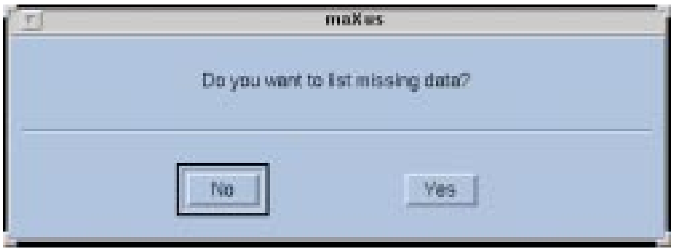
1. The process is slow and can be unstable unless there are numerous data in common between frames.
2. It can remove the effects of absorption and hence make the correction less effective.

6.3.14 SORTAV Completion

While SORTAV is running, a process indicator is displayed:

Sortav Process Indicator

After the program has searched for missing reflections out to the max sin θ/λ of the data, it asks:



Do you want to list missing data?

Please note that if you choose *Yes* and are using a large data set, this option can take several hours.

Once processing is complete, you are asked if you wish to save a copy of the output in a SHELX format:

DO you want to write a shelx data file?

If so, click *Yes*. The file is saved as sortav.shelx.

The output from SORTAV can then be viewed by selecting *Abs* from the *Summary* menu.

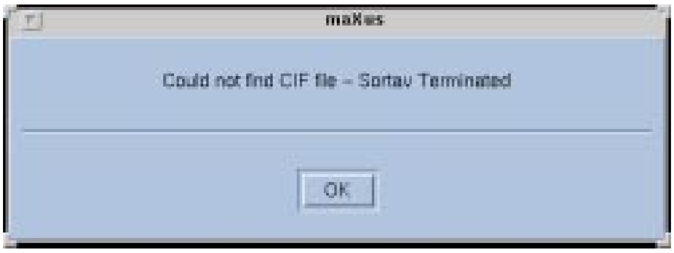
If the program terminates after a successful run, the following appears:

Sortav complete – run Spacegroup again

If the program terminates abnormally, the following dialogue box appears:

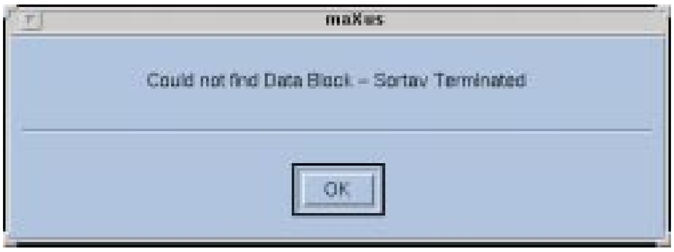
Sortav Terminating

6.3.15 Error messages



Could not find CIF file – Sortav Terminated

SORTAV reads the cif file containing the raw data. This file could not be found.



Could not find Data Block - Sortav Terminated

The CIF file does not have the correct header.

Error - no reflections found

No reflections were found in the CIF file.

More than 2500 frames

More than 2,500 kappaCCD frames detected in merge mode of SORTAV. Switch off the merge option on the main GUI.

Problem determining Laue Group - check Errors file

There is a problem with the Laue group. Read the file cif.err or the Abs summary file to check it. It may not be a problem.

Couldnot find CIF dictionary - Sortav Terminated

This is a system problem, contact support.

Insufficient data duplicates

You need a high level of data redundancy to run SORTAV; in this case the redundancy is insufficient.

Not enough equivalents

As above, but less serious so you can continue, but be careful and check the results carefully!

Singular transmission matrix

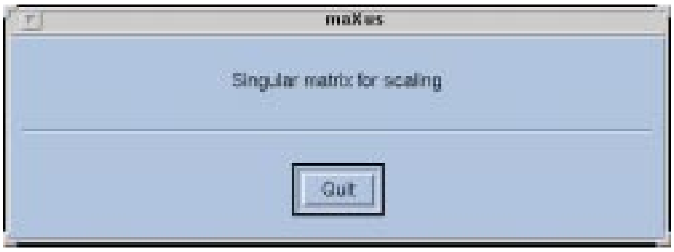
Because of insufficient duplicate data, or intensity data problems, the absorption surface cannot be modelled.

Negative transmission factors: Reduce LOMAX?

Because of insufficient data, the absorption surface cannot be modelled. This time the error gives physically impossible transmission factors.

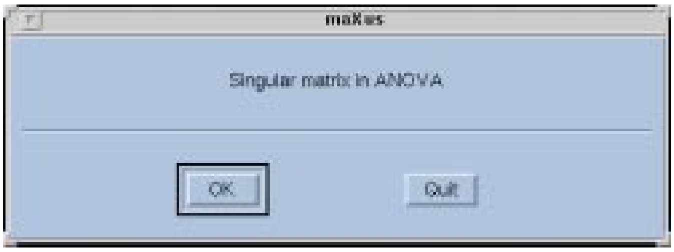
Singular normal matrix

In scaling datasets, a singular matrix was obtained. This is because of an insufficiency of reflections in common between layers. Switch off the merge option and try again.



Singular matrix for scaling

As above.



Singular matrix in ANOVA

The analysis of variance produced a singular matrix. You can still continue, but read the output carefully.

No estimate of crystal size (Tmin or radius)

You have not entered any values for crystal size in the main SORTAV window.

Read error on Original file

The Original file is faulty. Have you remembered to run Spacegroup?

Try the UNIX command `head -50 Original` to see the beginning of the file.

Original file is write-protected

Just as it says! Try the UNIX command:

```
ls -l Original
```

you should see read/write permissions for the owner of the file.

File not found

One of the files needed is not present. The output file will tell you which one.

6.3.16 References

R.H. Blessing (1995) *Acta Cryst.* **A51**, 33-37.

R.H.Blessing (1997) *J.Appl. Cryst.* **30**, 421-426.

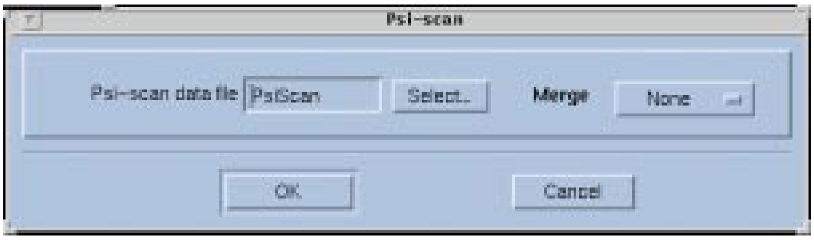
Hugh D. Young (1969). Statistical Treatment of Experimental Data, PP. 78, 162. New York: Mc Graw-Hill Co.

6.4 PSI-SCAN

This program represents a basic, but effective solution to the problem of crystal X-ray absorption. However, the program requires the psi-scan profile data file, PsiScan, which needs to be measured experimentally on a diffractometer. If this file has not been provided, then the process will terminate without further action. *The program cannot be used for CCD data*.

6.4.1 Psi-scan data input

To carry out the psi-scan correction, it is necessary to perform a psi-scan on your diffractometer. Your diffractometer manual will tell you how to do this. The program will ask you for the name of the file holding the PsiScan data, using a browser:



Psi-scan data file: PsiScan
Merge: None

Clicking *Select* brings up a window to allow you to browse through your files and directories to find the one you are looking for.

Under the *Merge* pop-up menu there are three choices:

- None: there is no post correction merging.
- Exclude FP: The data are merged, but Friedel pairs are excluded.
- Full: All the data are merged.

Of course, if you have already carried out a merge (using Spacegroup), there will be no data to merge.

Two types of psi-scan file can be used:

CAD4: A standard Nonius psi-scan file.

This can contain as many scans or as many reflections as you wish with differing step sizes, but you must scan the range 0 - 360. The data are interpolated and averaged.

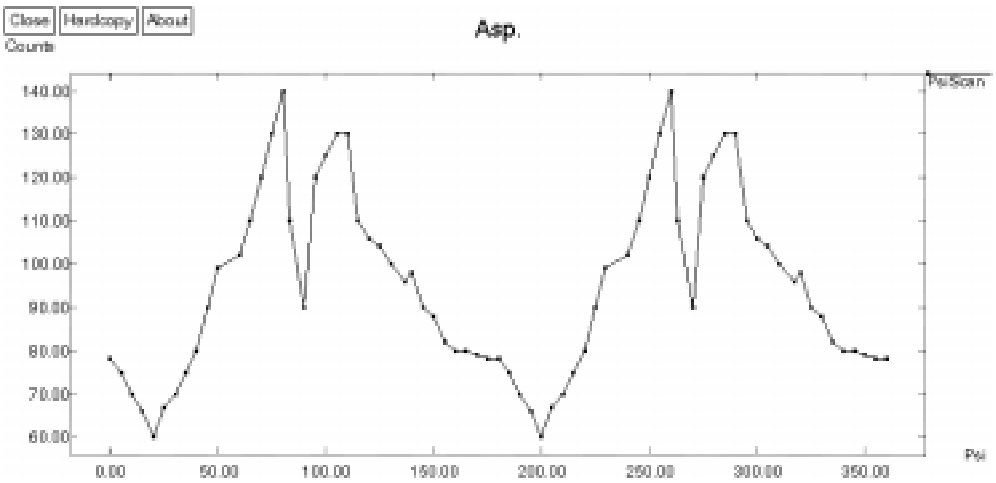
MacScience: An ASCII file of the following format is required:

- ASCII plain text file, one entry per line.
- Each entry has two fields: the angle and the counts.
- The data must span the range 0-360^o
- The angle should increase monotonically.

If the Psi-scan file does not exist, then the following error box appears:

PsiScan file does not exist

The program will now display the psi-scan profile in a graphics window. This can be printed if required using the buttons on the top right-hand side:



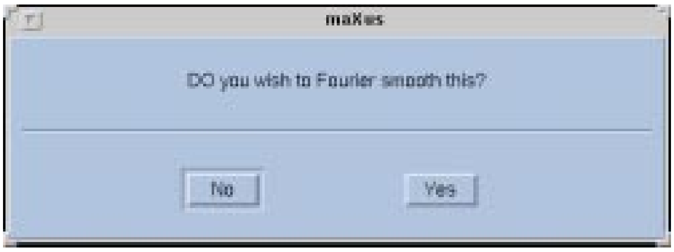
No user input is required and the progress of the execution is indicated by a limited output of corrected reflections.

Assuming the data file is in existence and everything else is OK, you will now be asked if you wish to make the scan symmetric such that $\phi = \phi + 180$ degrees:

Do you wish to make this symmetric?

Select *Yes* or *No*.

A dialogue box will ask if you want the data Fourier smoothed:



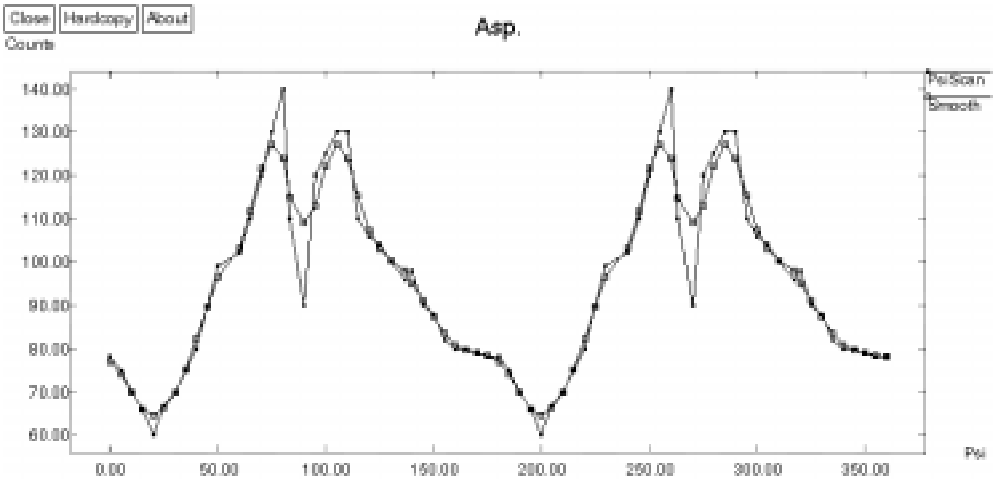
DO you wish to Fourier smooth this?

Use this option with care: it will smooth the data according to a moving cursor enclosing a specified number of points that must be input by the user:

Enter smoothing window size:

5

The larger this number the greater the smoothing. If you select this option the window is re-drawn with the smoothed curve also displayed:



Decide if you wish to keep this. Once the decisions have been made, the calculation proceeds automatically.

IMPORTANT: If you decide to repeat the calculations a new graphics window is opened on top of the existing one. In the current Silicon Graphics version these need to be dismissed at the end of the calculations by pressing the Close button at the top LH of the screen. Do NOT dismiss the window from the window bar.

It is possible to examine the full output from this program by choosing *Abs* from the *Summary*' menu.

Psiabs complete

6.4.2 Messages, Warnings and Errors

If the orientation matrix in the Original file is faulty an error message appears:

Faulty orientation matrix

The psi-scan program requires the CAD4 diffractometer to run in bisect mode; if it is not, the following message appears:

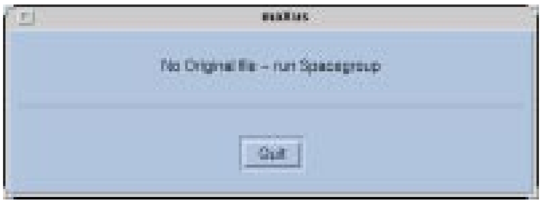
CAD4 not run in bisect mode

If you choose to continue with this data, check the program results carefully.

The program cannot use Neutron or Electron data; if you try to do so, an error will be generated:

No Neutron or Electron data

If the required data files (either Original or Processed) could not be found, one of the following is produced:



No Original file – run Spacegroup

NO processed data files were found

No RefData file found

If there is too little data present in the input file, or too much data:

Too many Psiscan data

Too fewPsiscan data

Errors will also be generated if the data spans too large a range:

Dynamic range Psiscan data too large

Or if it does not span 359 degrees:

Psiscan data must span 359 degrees

If the data file is corrupted, or the data is not in the expected order, an error is generated:

Psiscan error – records displaced

Clicking *Quit* in any of these error dialogs results in an error termination message:

Psiscan Terminated

6.4.3 Correcting the Original File

See Section 6.2.6. The procedure is identical.

6.4.4 Limitations

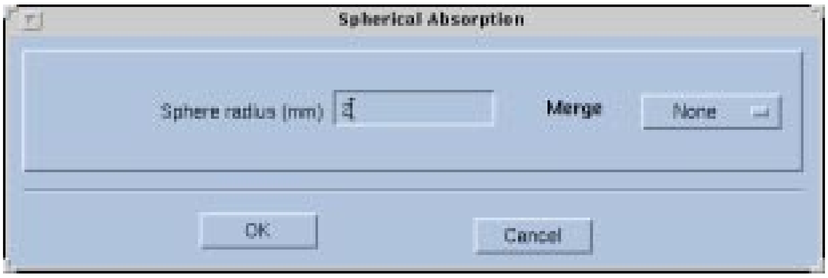
Psi-Scan cannot be used with CCD data.

6.5 SPHERICAL

For crystals which approximate to spherical in shape, this is a good option for an absorption correction. All that needs to be entered is the crystal radius in mm.

6.5.1 Running the Spherical Absorption Correction Program

Choose *Spherical* from the *Abs* menu. The execution of this program will require the input of the crystal radius as follows:



Sphere radius (mm) 3
Merge None

Values of μR>12.0 will cause major inaccuracies and will not be accepted. No further input from the user is needed.

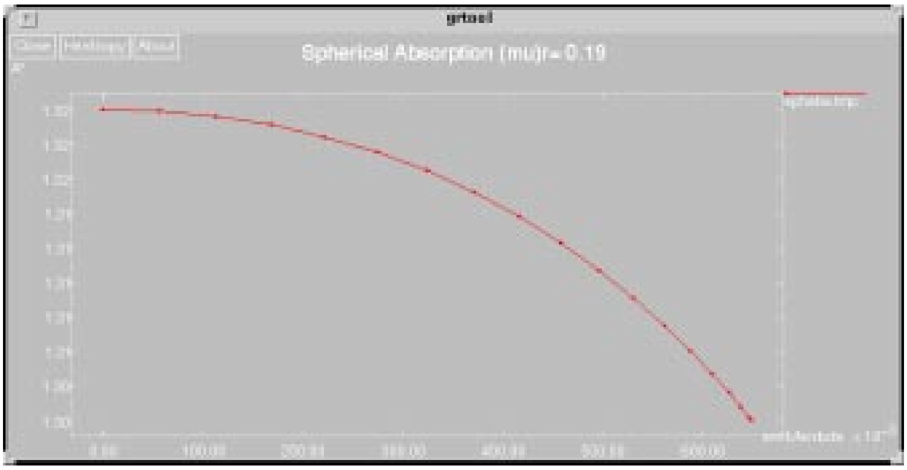
Under the merge button there are three choices:

- None: there is no post correction merging.
- Exclude FP: The data are merged, but Friedel pairs are excluded.
- Full: All the data are merged.

Of course, if you have already carried out a merge at Spacegroup time, there will be no data to merge.

It is possible to examine the full output from this program by choosing *Abs* from the *Summary* menu.

The program also produces graphical output, for example:



Upon successful completion of the program, the following is displayed:

Absorption correction complete

6.5.2 Messages and Warnings

The value you have used for μR is not in a usable range:

R must be 0-12

You cannot use electron or neutron data here:

No Neutron or Electron data

If the orientation matrix in the Original file is faulty:

Faulty orientation matrix

Error termination:

FAULT: Program terminating

Consult user support.

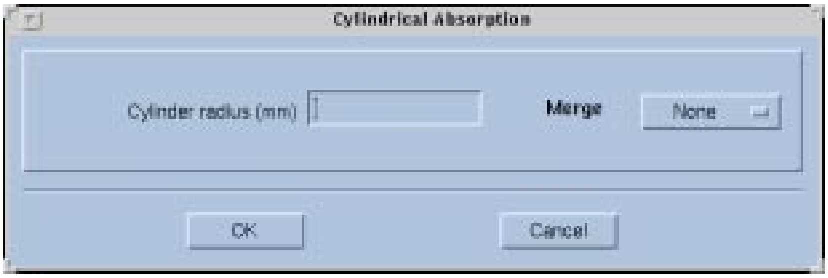
6.5.3 Correcting the Original File

See Section 6.2.6. The procedures are identical.

6.6 CYLINDRICAL

6.6.1 Running the Cylindrical Absorption Correction Program

This option can be useful for needle-shaped crystals. Choose *Cylindrical* from the *Abs* menu. The execution of this program will require the input of the crystal cylinder radius:



Values of µR>12.0 will cause major inaccuracies and will not be accepted. No further input from the user is needed.

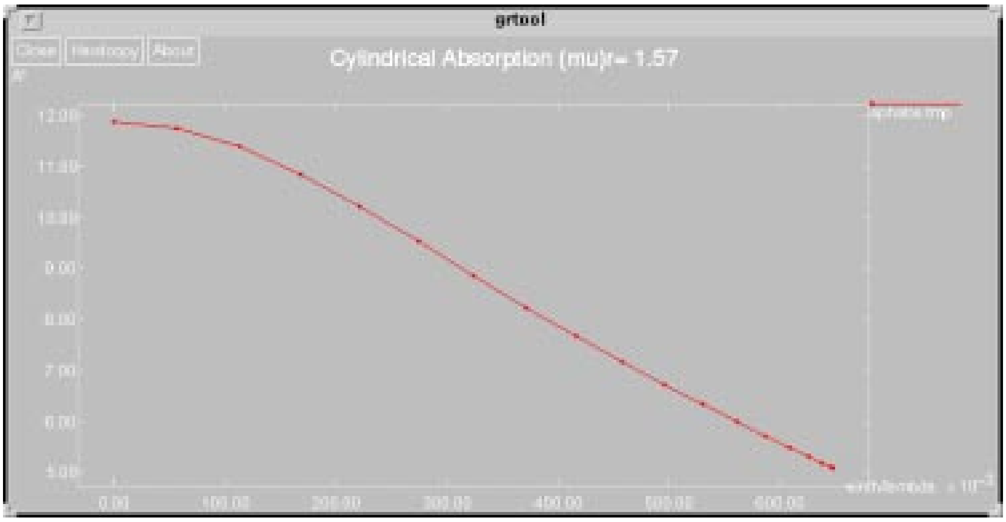
Under the merge button there are three choices:

- None: there is no post correction merging.
- Exclude FP: The data are merged, but Friedel pairs are excluded.
- Full: All the data are merged.

Of course, if you carried out a merge at Spacegroup time, there will be no data to merge.

It is possible to examine the full output from this program by choosing *Abs* from the *Summary* menu.

The program also produces graphical output, for example:



6.6.2 Messages and Warnings

All error messages and warnings are identical to those for the SPHERICAL absorption program; see Section 6.5.2.

6.6.3 Correcting the Original File

See Section 6.2.6. The procedures are identical.

6.7 DIFABS

DIFABS applies a correction for absorption/extinction by fitting a transmission surface to the differences between $|F_o|$ and $|F_c|$ at the end of **isotropic refinement**.The calculation is based on the method used by Walker and Stuart.

6.7.1 Running Difabs

Choose *Difabs / Processed* from the *Abs* menu. The user will be asked:

Include theta dependence?

Usually, you run this with theta dependence, but if there are problems, you can switch this off.

The execution of this program will require no further input from the user. On completion of the process, the workspace will clear, bringing control back to the main menu. The input reflection file is the calculated structure factor file, HklCal. It is presumed that the order of the hkl indices are consistent with the observed file, HklObs. The HklCal data is presumed to represent the best possible isotropic refinement, where the data is averaged and free from systematic absences. In principle the hydrogen scattering should be included, but omitting the hydrogen atoms does not usually give rise to problems. When the program terminates correctly, the following dialogue box appears:

Difabs Complete

Once the correction is complete, you will be prompted about merging the data. The following dialogue box appears:

Do you want to merge the data?

If you choose *No* then no merging is carried out. If however, you select *Yes* a second box appears:

Full merge or exclude Friedel pairs?

Make you selection from the two options. Data merging is then performed. A dialogue box will tell you when it has been completed:

Sort-Merge of type EFP completed

The output file contains the corrected $|F_o|^2$.The corrected data will generally effect the overall scaling parameter in the refinement. However, the least squares refinement procedures will automatically perform a scale adjustment prior to an overall refinement. It is possible to examine the full output from this program by choosing *Abs* from the *Summary* menu.

6.7.2 Warnings and Errors

Warning: Anisotropic atoms present

Atoms have been found with anisotropic temperature factors. Normally this is not permitted, but you are allowed to choose.

If there is a problem with the orientation matrix in the data:

Faulty orientation matrix

Clicking *Quit* in either gives the following dialogue box:

Difabs Terminated

6.7.3 Correcting the Original File

See Section 6.2.6. The procedures are identical.

6.7.4 Banning DIFABS?

A discussion on whether the absorption correction software, DIFABS, should be banned, was started by Tom Spek on the newsgroup *sci.techniques.xtallography* and seems to have reached some conclusions on the DIFABS issue. There is concern over the use and misuse of absorption correction software such as DIFABS to determine crystal structures. The discussion has resulted in the conclusions that DIFABS (and similar software) is a legitimate package in the crystallographer's ‘bag of tricks’ - but that users should be made aware of its possible abuses and effects (Software correction, poorly installed and aligned equipment, or being at an inappropriate stage of the refinement for using the correction software, for example).

It must be used with honesty and care.

6.7.5 Limitations

DIFABS cannot currently be used with CCD data; SORTAV generally gives better results in any case.

6.8 Data Merging: Some Advice

If absorption corrections are going to be important the following procedure is recommended:

- Do not sort-merge at Spacegroup time.
- Sort-merge at the time of the absorption correction.

7 SIR

Direct Method structure solution Using Semi-Invariants Representation

7.1 The SIR Program

The SIR (Semi-Invariants Representation) package has been developed for solving crystal structures by direct methods. The representation theory, proposed by Giacovazzo (1977,1980) allowed the derivation of powerful methods for estimating structure invariants (s.i.) and structure seminvariants (s.s.). The mathematical approach makes full use of the space group symmetry. SIR uses symmetry in a quite general way allowing the estimation and use of s.i. and s.s. in all the space groups. The range of options available to experienced crystallographers for choosing their own way of solving crystal structures is rather wide. However scientists untrained in direct methods or experienced people trustful in the SIR default mode often can solve crystal structures without personal intervention.

7.2 Solving Structures

SIR can be started as part of the Automatic Structure Solution routine (Chapter 4) which is documented separately, or directly from the *Solve* menu which gives access to the program, with four default options:

(1) Default 1
(2) Default 2
(3) Default 3
(4) Default 4

Each of these defaults invokes an ever-increasing level of complexity of commands. Try each one in turn if you are having difficulty solving a structure.

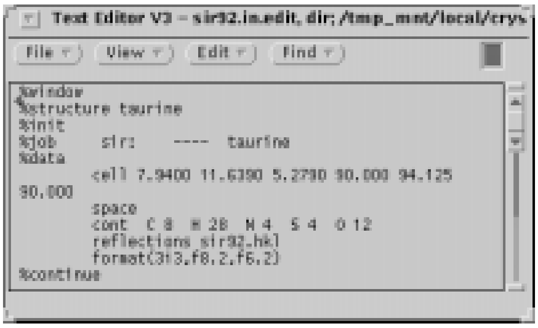
The other two options are:

(5) Edit input file
(6) Use edited

These allow the user to compose their own command sequences to SIR for difficult situations.

7.3 Edit Input File

When this option is selected, a text editor appears:



```
%window
%structure taurine
%init
%job     sir:    ----  taurine
%data
         cell 7.9400 11.6390 5.2790 90.000 94.125
90.000
         space
         cont  C 8  H 28  N 4  S 4  O 12
         reflections sir92.hkl
         format(3i3,f8.2,f6.2)
%continue
```

Use this editor to add your own commands as detailed later this manual. Save the file in the usual way.

7.4 Use Edited

The file that was edited is now used as input to SIR.

7.5 Completing a Structure

This option is run by selecting *SIR* from the *Complete* menu in the main maXus window:

Current directory: /home/crystal/gbam/testdata/demo

When SIR is used to *Complete* a structure which has already been partially solved it carries out the same default options with additional prior information provided by the existing peak list.

7.6 How SIR Works

The sequence of operations carried out in SIR isis as follows:

- **Normal**

Normalisation of diffraction intensities by the Wilson method.

- **Invariants**

Calculation of triplets (using the p10 formula) and negative quartets plus seminvariants as well

- **Phase**

Phase calculation, expansion and refinement.

- **Fourier/Least-Squares**

Density map calculation, peak search, atom assignment and least-squares refinement.

All of these procedures are explained in more detail below.

7.7 SIR System Overview

The main modules of the program are: SIR, LIST, DATA, NORMAL, SEMINVARIANTS, INVARIANTS, PHASE, FOURIER/LEAST-SQUARES, EXPORT/RESTART and PATTERSON.

- **SIR module**

It interprets commands and calls desired routines.

- **LIST module**

It is the software interface between SIR and the direct access file on which data and results are stored.

- **DATA module**

This routine reads the basic crystallographic information like cell parameters, space group symbol, unit cell content and reflections. It includes a modified version of the subroutine SYMM by Burzlaff & Hountas (1982). Symmetry operators and information necessary to identify structure seminvariants (estimated in SEMINVARIANTS routine) are directly derived from the space group symbol. Diffraction data are checked in order to find out equivalent reflections or systematically absent reflections (which are then excluded from the data set) and, eventually, reflections (weak) not included in the data set (Cascarano, Giacovazzo & Guagliardi, 1991).

- **NORMAL module**

In this module diffraction intensities are normalized using the Wilson method. Statistical analysis of intensities is made in order to suggest the presence or absence of the inversion centre and to identify the possible presence and type of pseudotranslational symmetry (Cascarano, Giacovazzo and Luic, 1988 a,b; Fan, Yao & Qian, 1988). Possible deviations (of displacive type) from ideal pseudotranslational symmetry are also detected. All the above information is not used as prior information in the next steps of SIR, unless the directive PSEUDO is given to the program. When some additional prior information, besides positivity and atomicity of electron density, is available, then a suitable renormalization of structure factors is made. SIR is able to deal with the following types of prior information:

1. pseudotranslational symmetry, identified by normalization routine or other source;
2. a well oriented and well positioned molecular fragment (Camalli, Giacovazzo & Spagna,1985; Burla,Cascarano,Fares, Giacovazzo, Polidori & Spagna,1989).

- **SEMINVARIANTS module**

This routine is not used in the default run of SIR. One-phase s.s. are estimated by means of their first and second representation as described by Giacovazzo (1978), Cascarano et al. (1984). As a default the second representation is calculated. Two-phase s.s. are estimated by means of their first representation as described by Giacovazzo et al. (1979), Burla, Giacovazzo and Polidori(1988). The estimated s.s. are stored in the direct access file; those evaluated with highest reliability will be actively used in the phasing process while the others will contribute to compute, with other phase relationships, the figure of merit CPHASE.

- **INVARIANTS module**

Up to 20,000 triplets relating reflections with normalized E values greater than a given threshold (strong triplets) are stored for active use in the phasing process. Also triplets (psizero triplets) relating two reflections with large E and one with E close to zero are generated: they are actively used in the phasing process (Giacovazzo, 1993) and define a special figure of merit (PSCOMB). Special types of triplets (psi-E triplets) based on two strong and one intermediate reflections (just below the threshold of strong reflections) are calculated and used in the FOURIER/LEAST-SQUARES module in order to extend phase information (Altomare, Cascarano, Giacovazzo & Viterbo, 1991). Negative quartets are generated by combining the psizero triplets in pairs, and those with cross-magnitudes smaller than a given threshold are estimated by means of their first representation, as described by Giacovazzo (1976). These quartets are be actively used in the phasing process (Giacovazzo, Burla & Cascarano, 1992) and will provide an important contribution to the FOM CPHASE. Active triplets may be estimated according to Cochran's (1955) p-3 distribution: the concentration parameter of the von Mises distribution is then C = 2 * E(h) * E(k) * E(h-k) / sqrt(N) (1)

Triplets can also be estimated according to their second representation (i.e. p-10 formula, as described by Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi and Polidori,1984). The concentration parameter of the new von Mises (i.e. of the same form of Cochran's) distribution is given by

G = C (1 + q) (2)

where q is a function (positive or negative) of all the magnitudes in the second representation of the triplet. The G values are rescaled on the C values and the triplets are ranked in decreasing order of G. The top relationships represent a better selection of triplets with phase value close to zero than that obtained when ranking according to C. These triplets will be actively used in the phase determination process. Triplets estimated with a negative G represent a sufficiently good selection of relationships close to 180 degrees to be used both for active use in the phasing process (Giacovazzo, Burla & Cascarano, 1992) and for the calculation of a powerful FOM (CPHASE). Triplets with G close to zero are expected to have values widely dispersed around 90 or 270 degrees and are used to compute an enantiomorph sensitive FOM. A similar FOM is also computed using quartets estimated with a very small concentration parameter. As a default, triplets are estimated according to p-10 formula. The parameter C of the Cochran distribution (say p-3) is suitably modified when prior information, such as that described above under NORMAL module, is available. Then triplet phases are no longer expected to be around zero (see quoted references) and may lie anywhere between zero and two pi.

- **PHASE module**

In the SIR program the most reliable one-phase s.s. are treated as known phases. Besides triplets, also the most reliable negative quartets and two-phase s.s. may be actively used. Each relationship is used with its proper weight: the concentration parameter of the first representation for quartets and two-phase s.s., and C or G for triplets.

- convergence/divergence procedure -

The convergence procedure (Germain, Main & Woolfson, 1970) is a convenient way of defining an optimum starting set of phases to be expanded by the tangent formula or by any other algorithm. When the p-10 formula is used, as a default, a special convergence process is devised which chooses the starting set according to

$$\alpha = (\sum G D_1(G) D_1 \alpha(k)) \times D_1 \alpha(h-k)$$

as suggested by Giacovazzo (1979) and by Burla, Cascarano, Giacovazzo, Nunzi & Polidori (1987), with D1(G) = I1(G) / I0(G) I1 and I0 represent modified Bessel functions of order one and zero respectively. The summation in (3) is over all relationships defining the reflection h. If p-3 formula is used the default choice is <alpha> = Sum (C * D1(C)) (4) Once the starting set has been defined, a good pathway for phase expansion is determined by a divergence procedure. In the divergence map, starting from the reflections in the starting set, each new reflection is linked to the preceding ones with the highest value of <α>.

- phase extension and refinement -

The starting set defined by the preceding step is usually formed by the origin (and enantiomorph) fixing reflections, a few one phase s.s. and a number of other phases which may be obtained: a) by magic integer permutation (White & Woolfson, 1975; Main, 1978), b) by a random approach (Baggio, Woolfson, Declercq & Germain, 1978; Burla, Cascarano & Giacovazzo, 1992). The option a) is the default, b) runs if the directive RANDOM is used. In this last case a large number (depending on the available computer time) of trials can be requested. If a partial structure is available (the directive PARTIAL should have been used in the normalization routine) the PHASE routine is automatically able to take that information into account. No further directives are strictly necessary. However directives SYMBOLS, SPECIALS, MAXTRIAL may be used to change default values. Phase expansion and refinement are carried out by means of a tangent formula using triplets, negative quartets, psi-zero triplets and the most reliable two-phase structure seminvariants. In the weighting scheme the experimental distributions of the alpha parameters are forced to match with the theoretical ones (Burla, Cascarano, Giacovazzo, Nunzi and Polidori, 1987). For each phase set several FOM's are computed using all invariants and seminvariants estimated by means of the representation method. Their meaning and an optimized way of combining all the computed FOM's to give a highly selective combined figure of merit (CFOM) is described in the papers by Cascarano, Giacovazzo and Viterbo (1987) and by Cascarano, Giacovazzo & Guagliardi (1992). All FOM's, as well as the combined CFOM, are expected to be equal to 1.0 for correct solutions. CFOM larger than 0.5 can be considered encouraging. If pseudotranslational symmetry is present then CFOM > 0.3 may characterize the correct solution.

- **FOURIER/LEAST-SQUARES module**

The sets of phases generated by the tangent routine are first expanded through psi-E relationships and then passed to the fast fourier transform routine written by L.F.Ten Eick (1977) and subsequently modified by the MULTAN team (Main et al. 1980). Several additional features have been introduced in the present version.

1. special positions are handled, peaks very close to symmetry elements are moved onto symmetry elements, the site symmetry is defined and the atomic occupancy factor is calculated; key numbers for designating free, coupled or fixed positional parameters for least-squares subroutines are also calculated together with symmetry conditions on the thermal ellipsoid.

2. The set of peaks provided by the peak search routine is automatically analysed in order to provide sound molecular fragments. If (as usually) the atomic species present in the unit cell are known, their atomic radii and (eventually) their chemical coordination are used in order to automatically identify fragments and relate peaks to the atomic species (automatic labelling of peaks).

The above information is automatically processed *via* least-squares Fourier cycles in order to complete the crystal structure, reject false peaks and refine structural parameters (Altomare, Cascarano, Giacovazzo & Guagliardi 1993). An isotropic diagonal matrix refinement is used which does not involve H atoms. The final R's usually vary from 0.08 to 0.15.

- **EXPORT/RESTART module**

Atomic parameters produced by the preceding module, stored in the direct access file, can be exported in ASCII file in a format suitable for other programs such as CRYSTALS, SHELX92, MOLPLOT, MOLDRAW, SCHAKAL, etc. If a graphic interface is available it is possible to delete or relabel some atoms and restart the FOURIER LEAST-SQUARES procedure. If the graphic interface is not available, it is possible to re-run the program after having modified the atom list through the command RESTART.

- **PATTERSON module**

In SIR it is possible to compute a PATTERSON map using various coefficients.

7.8 Text Command Reference

The input consists of a sequence of comments, commands and directives. The commands are headed by '%' character and directives must follow the related command. SIR recognizes the following commands:

%INITIALIZE	initialize the direct access file (to override previous results and data)
%DATA	Data input routine
%NORMAL	normalization routine
%SEMINV	seminvariants routine
%INVARIANTS	invariants routine
%PHASE	converge-tangent routine
%FOURIER	Fourier/Least-Squares routine
%EXPORT	
%END	end of the input file
%PATTERSON	
%JOB	a caption is printed in the output
%CONTINUE	the program runs in default conditions from the last given command up to the end

%STRUCTURE string	this command is used to specify the name of the structure to investigate. The program creates the name of the files needed by adding the appropriate extension to the structure name. The file names are: string.bin -> direct access file string.ins -> final coordinates file string.plt -> file for graphics If this command is not used the default string “STRUCT” (instead of the name of the structure) is used to create file names.
%WINDOW [x y]	graphic window is required. Optionally it is possible to increase the dimensions using x and y. Default values are 720,500.
%NOWINDOW	graphic window is suppressed %RESTART Fourier/Least-Squares restart routine

Directives are described below, in the sections dedicated to the various routines. All commands and directives are in free format (between columns 1-80) and are case independent. Only the first four characters are significant. The keywords can start in any position.If the first non-blank character is “>”, then the record is interpreted as a comment. SIR preserves intermediate results. For example, if invariant estimates have been already obtained during a previous run of SIR, in a new run the commands %INVARIANTS can be omitted. Commands can be given in any order, under the following conditions: - first routine used must be DATA, if it has not be used in a previous run; - Invariants and Seminvariants routines have no meaning if diffraction moduli are not normalized; - PHASE routine has no meaning if no triplets have been calculated; - FOURIER routine cannot run before PHASE routine. - PSEUDO and PARTIAL directives require their own normalization. If one renounces to use PSEUDO or PARTIAL information the SIR must be restarted from %DATA command.

The minimal information needed by SIR consists of (see Example 1):

- cell parameters
- cell content
- space group symbol
- reflections

The following limitations hold for commands and directives:

- PTEN cannot be used if the directive PSEUDO is on.
- seminvariants cannot be requested if PSEUDO is on.
- if PARTIAL is on, PSEUDO and RANDOM cannot be used.
- the directive SYMBOLS is of no use if RANDOM is on.
- PARTIAL and GENER are not compatible
- PATTERSON followed by continue is not permitted.

7.8.1 Preparation of data for DATA routine

CELL a b c alpha beta gamma

Cell dimensions. a,b and c are in angstroms, alpha, beta and gamma in degrees.

SPACEGROUP string

String is the symbol of the space group (i.e. P 21 21 21 or P 21/C or R -3), according to International Tables (1952). Blanks are necessary among the items constituting the space group symbol. SHIFT sx sy sz Origin shift components. They must be integral multiples of 1/24

CONTENTS Eli n1 El2 n2 El3 n3 .... Unit cell contents. Eli is the chemical symbol of atomic type i, ni is the corresponding number of atoms in the unit cell (to a maximum of 8 atomic types). For each chemical element up to Cf (Z=98) scattering factor constants are stored, together with information on the atomic number and weight, covalent and Van-der-Waals radii., etc. in a file.

RHOMAX x

Maximum value of (sin(theta)/lambda)**2 accepted for reflections to be used. The default is that all the data accepted.

RECORD n

Specifies the number of reflections per record when n>1.

FORMAT string

String is the run time format to read reflections. Default value for string is (3I4,2F8.2).

GENER

Reflections not in data will be generated having (sin(theta)/lambda)**2 value less than the maximum value of (sin(theta)/lambda)**2 in input.

NOSIGMA

To be used when sigma(F) values are meaningless.

REFLECTIONS string

If string is equal to FOLLOW then the program expects reflections on card reader after the%END or %CONTINUE command. Otherwise string is the name of the reflections file. Records have n reflections, each with h,k,l,F(obs), sigma (F) where h,k,l are integer (less than 63). The end of reflections is detected using one of the following: - blank record; - end of file; Negative values of F(obs) are allowed; negative values of sigma(F) are forbidden.

FOSQUARE

Program expects h,k,l, F**2, sigma(F**2).

7.8.2 Preparation of data for NORMAL routine

NREF n

Number of reflections with largest E-values subject to a minimum value of E(obs) = 1.2. Default = 4 * number of independent atoms + 50 if centrosymmetric + 50 if triclinic + 150 to a maximum of 499. If PSEUDO directive is on, the program tries to use all the E’s down to 1.2.

NZRO n

Number of smallest E-values to calculate psi(zero) triplets. Default is 1/3 of strong reflections.

BFAC x

Temperature factor if the user wants to supply it.

PSEUDO n(1,1) n(2,1) n(3,1) n(4,1) n(1,2) . . . n(4,3)

If parameters are equal to zero, or omitted, the normalization routine performs statistics for the pseudotranslation effects and renormalizes the reflections on assuming the most probable pseudotranslation as prior information. If user wants to supply a specific pseudotranslational symmetry, and this corresponds to a class given in Table 1 the n(1,1) can be set to the class number and the other values must be omitted. If the pseudotranslation is more complex, it should be specified by up to 3 sets of 4 values n(i,j) so that (n(1,j)*h + n(2,j)*k +n(3,j)*l =n(4,j)*m) j=1,3 where m is an integer number.

PARTIAL

To perform the procedure which aims at recovering the complete structure from a partial one. This instruction is followed by the list of atomic information (one per card) consisting of atomic type, x, y, z and optionally occupancy factor (default 1.0). See Example 3 - 4. Please consider the directive FRAGMENT in module FOURIER/ LEAST-SQUARES for a different approach.

7.8.3 Preparation of data for SEMINVARIANTS routine

FIRST

To perform the estimation of one-phase seminvariants *via* the first representation.

NRS1 n

The number of the strongest reflections to use for one-phase seminvariants search. Default value = number of strongest reflections chosen by normalization routine (max 350).

LIST

Print the list of the strongest reflections to use for one-phase seminvariants search.

NUMK n

The number of free-vectors used by the second representation formula for the one-phase seminvariants. The default value is computed according to the structural complexity:
n = min (natom/msym + 40, 100) where natom is the number of atoms in the cell and msym is the number of symmetry operators of the space group.

NRS2 n

n is the number of the strongest reflections to use for two-phase seminvariants search. Default value = number of strongest reflections chosen by normalization routine (max 350).

7.8.4 Preparation of data for INVARIANTS routine

NRTRIPLETS n

The number of strongest reflections for the sigma2 search. Default = number computed by normalization routine.

GMIN x

Triplets with G < x are not actively used. Default value x=0.6 (in any case x > 0.2).

EMIN x

Minimum value for the normalized structure factors in the sigma2 search. Default = minimum value of E for the strongest reflections selected by normalization routine.

NRPSIZERO n

The number of weakest reflections for the psi-zero search. Default = number computed by normalization routine.

EMAX x

Maximum value of E for the weak reflections for the psi-zero search (up to a maximum of NRPSIZERO reflections).

COCHRAN

To use the p-3 formula. (By default p-10 formula is used).

NUMK n

The number of free-vectors used by p-10. Default value is provided by SIR according to the structural complexity: n = min (natom/msym + 30, 70) where natom is the number of atoms in the cell and msym is the number of symmetry operators of the space group.

CORRECTION x

If a pseudotranslation procedure is on, a correction factor depending on x is applied to the probabilistic argument of weak-weak-weak triplets when they are present. The default value is calculated by program. x = -1 eliminates weak-weak-weak triplets, x = 1 corresponds to the largest confidence in them.

BIG n

This keyword is used to modify the number of triplets to estimate. If n is omitted, up to 20000 triplets will be estimated. (Default 8000).

7.8.5 Preparation of data for PHASE routine

LIST n

The number of reflections at the top of the divergence map for which a list of the map is obtained. Default = no printout of the map.

TWOPHASE x

Two phase seminvariants are actively used if their concentration parameter is larger than x. If the directive is omitted x is assumed as the minimum value of G (or C) among the positive triplets.Very large values of x (say 1000 or so) hinder the active use of two-phase seminvariants.

ONEPHASE n

Number of strongest one-phase seminvariants to use in the starting set.Default value for n is chosen by seminvariants routine (up to a maximum of 3). If n is equal to zero, no seminvariant will be accepted.

ORIGIN n(i) phi(i)

Codes and phases of the reflections chosen by the user to fix the origin. The total number of reflections needed to fix the origin must be given (by default the origin is chosen by the program).

ENANTIOMORPH n

Code of the reflection chosen by the user to fix the enantiomorph (by default the enantiomorph is chosen by the program).

SYMBOLS n

The number of permuted phases (max 12). Default = 5.

PERMUTE n(i)
Codes of the permuted reflections chosen by the user (included in the number of required symbols, max 12).

SPECIALS n
The number of reflections with restricted phase to be permuted (included in the number of required symbols, max 12).

PHASE n(i) phi(i) wt(i) Codes, phases and weights of reflections with known phase (max 200). If one card is not sufficient, the directive PHASE must be repeated in the other card(s).

TABLE
To print the table of alpha values.

TRIALS n(i)
To print the final phases for the specified trials.

MINFOM x n
x: the program automatically stops when a solution is found with CFOM > x . Default value = 0.95.
n: is the maximum number of phase sets (with the largest combined figure of merit - CFOM) to retain; (default value = 10)

RANDOM n
To use random phases starting sets. n: is the number of random phases. If n is omitted, it is equal to one half of the strong reflections.

SEED n
Random generation seed. n must be an integer large value and must be odd. (Default 6753).

NOREJECT
Used to retain equivalent sets of phases.

TEST
Used to restart the random number generator using the default seed.

MAXTRIALS
Maximum number of trials when random approach is used. (Default 100).

7.8.6 Preparation of data for FOURIER/LEAST-SQUARES routine

SET n
The serial number of the set from which to calculate the E-map. If ‘SET n’ is not specified, the set with the highest combined figure of merit will be used, and then if FOURIER routine is rerun the set with the next highest combined figure of merit will be used.

MAP
If MAP is specified, the Fourier-map will be printed on the lineprinter. The printing of the Fourier-map is rarely used. The output is in sections of constant y with x across the page and z running down it on a grid of about three points per angstrom. The maximum number of grid points across the page is 36 (i.e. about twelve angstroms in length). If more are needed, a new page is started.

LEVEL n
All numerical values in the Fourier-map greater than n will be underlined with **** in the lineprinter output to facilitate contouring. The default is 100. The map is automatically

scaled between +&- 999 (approximately) at the grid points at which it is calculated. Note, however, that it is printed between +&- 99.

GRID x

The grid on which the Fourier-map is evaluated will have a spacing of approximately x angstroms. The default spacing is computed by the program.

LIMITS l1 l2 l3

The E-map will be printed from 0 to l1 along x, from 0 to l2 along y, from 0 to l3 along z. (This directive can be used only if associated to MAP directive).

PEAKS n

The number of peaks to be searched for in the map. The default is the number of peaks for which sum (occ(i)*m) = 1.3 * n where occ is the crystallographic site occupancy factor, m is the number of symmetry operators, n is the number of non-hydrogen atoms in the unit cell.

LAYX

LAYY

LAYZ

This directive can be used if MAP directive is given. The Fourier-map will be printed in section of constant x (or y, or z). (The default is y).

RADIUS El x

The value x of the radius used to compute the connectivity is assigned to element El.

COORDINATION El dmin dmax [n]

The minimum and maximum value of the bond distanced are assigned to element El. n is the coordination type; Allowed values for bond angles are: n=1 85-145 Degrees n=2 55-180 Degrees

FOMIN x

Minimum value of F(obs) (on absolute scale) to be used in the recycling procedure.

SIGMA x

Reflections with F(obs)/sigma(F) greater than x will be used in the recycling procedure.

DMAX x

Distances up to x between atoms will be printed.

RECYCLE n

Used to stop the recycling procedure after cycle number n. If n is equal to zero only the E-map will be calculated.

FRAGMENT string

Used to supply a known fragment to be completed. String is the name of the file in which, for each atom, are stored Element X Y Z

7.8.7 Preparation of data for EXPORT routine

In the following directives string is used to supply the output file name. If string is omitted the default file name will be used.

CRYSTALS string
A file containing peak coordinates is created in CRYSTALS format. About default see notes on the implementation.

COMPLETE
To produce a complete QUICKSTART file for CRYSTALS.

SHELX string
A file containing peak coordinates is created in SHELX-92 format. About default see notes on the implementation.

MOLDRAW string
A file containing peak coordinates is created in MOLDRAW format.

SCHAKAL string
A file containing peak coordinates is created in SCHAKAL format.

MOLPLOT string
A file containing peak coordinates is created in MOLPLOT format.

XYZ string
A file containing peak coordinates is created in XYZ format (Cartesian coordinates).

7.8.8 Preparation of data for RESTART routine

COMPLETE
The recycling procedure restarts and continues until the R value increases. By default only one cycle of the procedure is performed.

RELABEL string species (or RENAME string species)
String is formed by atomic species and serial number (no blanks between these two items are allowed); species is the new atomic species.

DELETE string
String is formed by atomic species and serial number (no blanks between these two items are allowed).

7.8.9 Preparation of data for PATTERSON routine

LAYX, LAYY, LAYZ, MAP, PEAKS, LIMIT, GRID: See Fourier directives.

E**2 (or E*E)
Coefficients to be used in the Patterson synthesis.

F**2 (or F*F)
Coefficients to be used in the Patterson synthesis.

E*F (or F*E)
Coefficients to be used in the Patterson synthesis.

7.9 When Default SIR fails: Strategies

Very often crystal structures are solved by SIR using only default settings. In cases of failure, the user can modify the default pathway in order to improve the success rate. The number of options is quite large and some of them are suggested here. The CFOM figure of merit is in practice an estimate of the probability that a trial solution is correct. If the highest CFOM does not correspond to the correct solution, subsequent CFOM's may be successful. A special problem arises when several nearly equivalent CFOM's are available for different trials. Then the individual FOM's have to be carefully considered by the user. In particular the additional R(x,y) figure of merit (Cascarano, Giacovazzo & Guagliardi, 1992) can also provide useful suggestions. The magic integer procedure is an effective and simple way for starting a multisolution process. Second representation formula for triplet invariants proved much more powerful than first representation formula therefore it is used as a default for estimating triplets. In case of failure, the user can increase the number of symbols, or use the RANDOM procedure. The use of 1- and 2- phase seminvariants are not default choices (they require additional computing time and often they are not necessary for the success of SIR). If symmetry is rather high, several seminvariants can be reliably estimated: some of them will actively be used in phase expansion and refinement process, others will be used as figures of merit. Sometimes the additional estimation of one or few one-phase seminvariants is enough for solving crystal structures. A check for possible pseudotranslational symmetry is always made by SIR: if PSEUDO keyword is not activated then information is given but not used in the phasing process. If a pseudotranslational symmetry has been found by the normalization routine a new run of SIR may be started including the keyword PSEUDO. Sometime a molecular fragment is located in the E-map, or it is available from other sources (i.e. Patterson), which is too small for recovering the complete crystal structure *via* the usual Least-Squares - Fourier methods; then the keyword PARTIAL may be activated in the normalization routine.

Perhaps because of a faulty data collection strategy, weak reflections may not be included in diffraction data.This lack of information influences both the normalization process (scale and overall thermal factors are affected by systematic errors; the experimental E-distribution is often non-centric even when the crystal structure is centrosymmetric) and the estimation of 1- and 2-phase seminvariants and of triplet and quartet invariants: in particular, a reduced number of negative triplets (*via* p-10 formula) and of negative quartets is calculated.

Success in the structure solution may be obtained if weak reflections are introduced in the data set. Even if there are still a large number of options available four of them are quoted for special cases: a) the value of NREF (number of reflections actively used in the phasing process) is fixed by SIR. For some special structures the ratio “number of active triplets/ NREF” is too small (less than 10). Larger values of NREF may improve the phasing procedure. b) high (or low) resolution reflections may occasionally play a too important role in the first steps of convergence and divergence process. Fixing a thermal factor lower (or larger) than that provided by the normalization routine may successful change convergence and divergence process. c) an alternative space group should be carefully considered.

7.10 Examples of input for SIR

1. The following example shows the maximum default use of SIR. Most of the structures can be solved in this way. Diffraction data are in the file BOBBY.HKL in format (3I4,2F8.2), one reflection per record.

```
%DATA
CELL 9.626 9.626 9.626 90 90 90
SPACEGROUP P 21 3
CONT H 24 C 24 N 4 O 24 NA 4 CA 4
REFLECTIONS BOBBY.HKL
%CONTINUE
```

2. User wants to use in the phasing process the pseudotranslational symmetry (eventually identified in a previous run). Diffraction data follow commands with format (2(3I4,F10.4,F1.0)), two per record. Sigma(F) values are not in input file.

```
%WINDOW
%STRUCTURE FEGAS
%INIT %JOB FE2GA2S5 2H P63/MMC
%DATA
CELL 3.651 3.651 29.745 90. 90. 120.
SPACE P 63/M M C
CONT FE 4 GA 4 S 10
REFLECTIONS FOLLOW
NOSIGMA
RECORD 2
FORMAT (2(3I4,F10.4,F1.0))
%NORMAL
PSEUDO
%CONTINUE
 0 0 4 21.10900 0 0 6 82.16260
 0 0 8 139.48620 0 0 10 31.70140
 0 0 12 63.81540 0 0 14 48.66780
..............................................
 5 0 6 19.55970 5 0 7 .00000
 5 0 8 6.31500 5 0 9 28.76280
```

3. The following example shows how to run partial procedure after that a complete run of SIR has been performed.

```
%STRUCTURE CEPHAL
%JOB PARTIAL PROCEDURE RUN
%NORMAL
PARTIAL C .3824 .1494 .2856
C .4344 .1415 .1688
C .3675 .1102 .2789
C .3866 .1546 .5060
C .4153 .0957 .1638
C .4088 .1654 .1820
> Note that invariants and eventually seminvariants
> need not to be recomputed
%PHASE
%CONTINUE
```

4. The following example shows how to use the partial procedure when a molecular fragment is located from other sources (i.e. Patterson method).

```
%WINDOW
%STRUCTURE
CEPHAL
%INIT
%JOB - CEPHALOTAXIN C 2 -
%DATA
CELL 22.840 8.150 19.540 90. 117.7 90.
SPACEGROUP C 2
CONT C 144 H 168 O 32 N 8
REFLECTIONS CEPHAL.HKL
RECORD 3
FORMAT (3(3I3,F7.2,1.0))
NOSIGMA
%NORMAL
PARTIAL O .06760 -.76129 .88547 1.0
O .10620 -.53535 .97020 1.0
C .05235 -.61801 .84287 1.0
C .07655 -.48284 .89316 1.0
C .09666 -.71719 .96485 1.0
%CONTINUE
```

5. In the following example the use of one and two-phase seminvariants is requested together with 7 symbols to permute in PHASE routine. It is assumed that SIR was already run before (thus normalization and invariants are not needed).

```
%STRUCTURE BOBBY
%SEMINVARIANTS
%PHASE
SYMBOLS 7
%CONTINUE
```

6. In the following example only the E-map of the set number 7 and the complete FOURIER/LEAST-SQUARES procedure for set number 9 are requested by the user.

```
%WINDOW 850 800
%STRUCTURE BOBBY
%FOURIER
SET 7
RECYCLE 0
%FOURIER
SET 9
%CONTINUE
```

7. The user wants only view and eventually modify the structure previously produced by FOURIER routine and create a complete input file for CRYSTALS named CEPHAL.QCK.

```
%STRUCTURE CEPHAL
%WINDOW
%EXPORT
CRYSTALS CEPHAL.QCK
COMPLETE
%END
```

8. In this example random approach is used and the best 200 sets, over 2000 trials, of phases are retained in the direct access file. No Fourier is required.

```
%NOWINDOW
%STRUCTURE CEPHAL
%PHASE
RANDOM
MAXTRIALS 2000
MINFOM 1.0 200
%END
```

9. In the following example the user just knows a fragment provided by a previous Fourier procedure and wants to delete some peaks and modify the atomic specie of other.

```
%STRUCTURE AZET
%WINDOW
%RESTART
DELETE Q56
DELETE Q55
DELETE Q54
RELABEL Q48 C
RELABEL N6 C
%CONTINUE
```

10. In the following example the user knows a fragment and wants to complete it using FOURIER/LEAST-SQUARES procedure.The binary file “azet.bin” must exist.

```
%Window
%Structure azet
%Fourier
fragment azet.frg
%Continue
```

Coordinates are in the file “azet.frg” which contains

```
Cl .02944 .72012 .08865
Cl .23727 .78692 .30869
```

Appendix A

Classes of reflections corresponding to the low index pseudo- translational symmetry.

1) 1h+1k+1l= 1n
2) 1h+0k+0l= 2n
3) 0h+1k+0l= 2n
4) 0h+0k+1l= 2n
5) 1h+1k+1l= 2n
6) 1h+1k+0l= 2n
7) 1h+0k+1l= 2n
8) 0h+1k+1l= 2n
9) 1h+0k+0l= 3n
10) 0h+1k+0l= 3n
11) 0h+0k+1l= 3n
12) 1h+1k+0l= 3n
13) 1h+0k+1l= 3n
14) 0h+1k+1l= 3n
15) 1h+1k+1l= 3n
16) 1h+1k+2l= 3n
17) 1h+2k+1l= 3n
18) 2h+1k+1l= 3n
19) 1h+2k+0l= 3n
20) 1h+0k+2l= 3n
21) 0h+1k+2l= 3n
22) 0h+0k+1l= 4n
23) 0h+1k+0l= 4n
24) 1h+0k+0l= 4n
25) 1h+1k+0l= 4n
26) 1h+0k+1l= 4n
27) 0h+1k+1l= 4n
28) 1h+1k+1l= 4n
29) 2h+2k+1l= 4n
30) 2h+1k+2l= 4n
31) 1h+2k+2l= 4n
32) 2h+1k+1l= 4n
33) 1h+2k+1l= 4n
34) 1h+1k+2l= 4n
35) 1h+2k+0l= 4n
36) 1h+0k+2l= 4n
37) 0h+1k+2l= 4n
38) 2h+1k+0l= 4n
39) 2h+0k+1l= 4n

40) 0h+2k+1l= 4n
41) 3h+3k+1l= 4n
42) 3h+1k+3l= 4n
43) 1h+3k+3l= 4n
44) 1h+2k+3l= 4n
45) 1h+3k+2l= 4n
46) 3h+1k+2l= 4n
47) 1h+3k+0l= 4n
48) 1h+0k+3l= 4n
49) 0h+1k+3l= 4n
50) 0h+3k+2l= 6n
51) 0h+2k+3l= 6n
52) 2h+3k+0l= 6n
53) 3h+2k+0l= 6n
54) 3h+0k+2l= 6n
55) 2h+0k+3l= 6n
56) 2h+2k+3l= 6n
57) 3h+2k+3l= 6n
58) 3h+3k+2l= 6n
59) 0h+4k+3l=12n
60) 4h+0k+3l=12n
61) 4h+3k+0l=12n
62) 0h+3k+4l=12n
63) 3h+4k+0l=12n
64) 3h+0k+4l=12n
65) 1h+0k+0l= 2n & 0h+1k+0l= 2n
66) 1h+0k+0l= 2n & 0h+0k+1l= 2n
67) 0h+1k+0l= 2n & 0h+0k+1l= 2n
68) 1h+0k+0l= 2n & 0h+1k+1l= 2n
69) 0h+1k+0l= 2n & 1h+0k+1l= 2n
70) 0h+0k+1l= 2n & 1h+1k+0l= 2n
71) 1h+1k+0l= 2n & 1h+0k+1l= 2n
72) 1h+0k+0l= 2n & 0h+1k+0l= 2n & 0h+0k+1l= 2n

7.11 References

1. Altomare A., Cascarano G., Giacovazzo C. & Guagliardi A. (1993). *J. Appl. Cryst.* **26**,000-000. In Press.
2. Altomare A., Cascarano G., Giacovazzo C. & Viterbo D.(1991). *Acta Cryst.* **A47**, 744-748.
3. Baggio R., Woolfson M.M., Declerq J.P. & Germain G.(1978). *Acta Cryst.* **A34**,883-892.
4. Burla M.C.,Cascarano G.,Fares E.,Giacovazzo C.,Polidori G. & Spagna R. (1989).*Acta Cryst.* **A45**, 000-000
5. Burla M.C., Cascarano G. & Giacovazzo C. (1992).*Acta Cryst.* **A48**,906-912.
6. Burla M.C.,Giacovazzo C. & Polidori G.(1987).*Acta Cryst.* **A43**,797-802.
7. Burla M.C.,Giacovazzo C. & Polidori G.(1988).*Acta Crysta.* **A44**, 000-000.
8. Burla M.C., Nunzi A., Giacovazzo C. & Polidori G. (1981), *Acta Cryst.* **A37**, 677-684.
9. Burla M.C., Cascarano G., Giacovazzo C., Nunzi A. & Polidori G. (1987), *Acta Cryst.* **A43**, 370-374.
10. Burzlaff H. & Hountas A. (1982), *J. Appl. Cryst.* **15**, 464-467.
11. Camalli M.,Giacovazzo C. & Spagna R. (1985).*Acta Cryst.* **A41**, 605-613.
12. Cascarano G., Giacovazzo C. & Guagliardi A. (1991). *Acta Cryst.* **A47**, 698-702.
13. Cascarano G., Giacovazzo C. & Guagliardi A. (1992). *Acta Cryst.* **A48**, 859-865.
14. Cascarano G., Giacovazzo C. & Luic' M.(1988a).*Acta Cryst.* **A44**, 176-183.
15. Cascarano G., Giacovazzo C. & Luic' M.(1988) . *Acta Cryst.* **A44**, 183-188
16. Cascarano G., Giacovazzo C. & Viterbo D.(1987).*Acta Cryst.* **A43**, 22-29.
17. Cascarano G., Giacovazzo C., Calabrese G., Burla M.C., Nunzi A., Polidori G. & Viterbo D. (1984), *Z. Kristallogr.*, **167**, 34-47.
18. Cascarano G., Giacovazzo C., Camalli M., Spagna R., Burla M.C., Nunzi A. & Polidori G. (1984), *Acta Cryst.* **A40**, 278-283.
19. Cascarano G., Giacovazzo C., Burla M.C., Nunzi A. & Polidori G. (1984), *Acta Cryst.* **A40**, 389-394.
20. Cochran W.(1955). *Acta Cryst.***8**.473-478.
21. Fan Haifu, Yao Jia-Xing & Qian Jin-zi (1988). *Acta Cryst.* **A44**, 688-691.
22. Germain G., Main P. & Woolfson M.M. (1970), *Acta Cryst.* **B26**, 274-285.
23. Giacovazzo C. (1976), *Acta Cryst.* **A32**, 958-966.
24. Giacovazzo C. (1977), *Acta Cryst.* **A33**, 933-944.
25. Giacovazzo C. (1978), *Acta Cryst.* **A34**, 562-574.
26. Giacovazzo C. (1979), *Acta Cryst.* **A35**, 757-764.
27. Giacovazzo C. (1980), *Acta Cryst.* **A36**, 362-372.

28.Giacovazzo C. (1983), *Acta Cryst*. **A39**, 685-692.

29.Giacovazzo C. (1993). *Z. f. Kristall.*, in press.

30.Giacovazzo C., Burla M.C. & Cascarano G. (1992). *Acta Cryst.* **A48**, 901-906.

31.Giacovazzo C., Spagna R., Vickovic I. & Viterbo D. (1979), *Acta Cryst.* **A35**, 401-412.

32.Main P. (1977), *Acta Cryst.* **A33**, 750-757.

33.Main P. (1978), *Acta Cryst.* **A34**, 31-38.

34.Main P., Fiske S.J., Hull S.E., Lessinger L., Germain G., Declercq J.P. & Woolfson M.M. (1980) - MULTAN80, a system of computer programs for the automatic solution of crystal structures from x-ray diffraction data - Univ. of York, England.

35.White P.S. & Woolfson M.M. (1975), *Acta Cryst.* **A31**,53-56

36.Ten Eick,L.F. (1977). *Acta Cryst*. **A33**,486-492.

8 DIRDIF

Solving Heavy Atom Structures Automatically

8.1 Introduction

DIRDIF can be used as follows:

- To solve and complete all structures with heavy atoms, including P and S.
- To solve all structures of molecules with (partly) known geometry, whether they have heavy atoms or not.
- To deal with situations where *ab-initio* direct methods gave a misplaced fragment.
- To expand a small fragment to the complete structure.
- To solve a structure with an enantiomorph problem.
- To solve a structure with a supersymmetry or pseudo-symmetry problem.

Most of the program options are fully automated according to the black box principle, but on the other hand they are also, as circumstances require and subject to the user’s decision, characterized by their flexibility and adaptability to the current problem. The user has full control over the system using an interface program which takes care of all communication and linkage; however, the system is smart enough to allow for ‘local democracy’, resulting in an alteration of the computational flow as intermediate results indicate or require.

NOTE: DIRDIF does not solve structures in which there is no heavy atom, unless the geometry of part of the molecule is already known.

DIRDIF consists of the following modules:

- The program PATTY calculates and interprets a sharpened Patterson function.
- The program PHASEX expands and refines the phases by direct methods applied to the difference structure. The partial structure (known heavy atoms or positioned fragment) is expanded by the application of direct methods to the difference structure factors. The method is particularly powerful when:
    1. The known part of the structure is only marginally sufficient to solve the structure,
    2. When the known atoms comprise a (pseudo)symmetry problem,
    3. When the structure is a superstructure,
    4. When there is an enantiomorph problem.

The supervisor program transfers the information to continue with the calculation of a Fourier synthesis and automatic recycling until completion of the structure.

- The program ORIENT orientates a given fragment (model) by a vector search procedure in vector space (Nordman) and writes the rotated fragment to an intermediate file. The supervisor program transfers the information to continue with:
- TRACOR which translates the fragment to the right position by translation functions based on correlation in reciprocal space. The supervisor program then transfers the information to continue with PHASEX. The system solves a structure using a known fragment (in correct orientation and position) or one or more known heavy atoms. The ATOMS file with the atomic parameters of the fragment must be present.

- DIRP1 is used in difficult cases by reducing the symmetry (temporarily) to P1 and completing the structure in this space group.

The ATOMS file is overwritten with the final atomic parameters of the structure. DIRDIF uses in addition its own ATMOD file containing coordinates.It is possible to view or edit this with the usual UNIX editors.

8.2 Running DIRDIF

Choose *Dirdif* from the *Solve* menu in the main window. The following options are available from the sub-menu:-

- Default. This is used to solve structures automatically using Patterson-heavy atom methods; the program calls most of the other modules automatically, and no further user intervention is required.
- PATTY is used to solve and complete a heavy atom structure automatically. PATTY is a program for the interpretation of a sharpened Patterson. It uses Buerger's implication theory (i.e. the so called symmetry map) and checks all possible cross vectors using the minimum-function value as a selection criterion. The program is used for heavy atom structures with unknown heavy atom positions (including not-so-heavy-atoms like S or P). When the user runs PATTY, the system will automatically run the following sequence:
    1. First calculate the Patterson function (program FOUR),
    2. Then find the heavy atom(s) (program PATTY), and
    3. Expand the partial structure (programs PHASEX, followed by FOUR), and
    4. Finally recycle several times (programs PHASEX and FOUR) for further elucidation of the structure.
- ORIENT orients a fragment of known geometry in the unit cell. ORIENT is a program to find the orientation of a molecular fragment (model) by means of vector search methods (Nordman). The input model is used for the calculation of interatomic vectors. The shape function of a single Patterson peak is approximated from the shape of the origin peak, and it is used for the calculation of the amount of overlap between neighbouring vectors. Vectors are selected on weight (including overlap), length, and mutual separation. A fast cyclic search system, employing increasing resolution per cycle, leads to the best fitting orientation of the model in angular space, i.e. the final result of the program is a set of Eulerian angles (denoted A, B, C) which is applied on the original model. The selection criterion is the Nordman minimum average function value. The program is used for structures with known geometry for a small part of the molecule. Note that such a molecular fragment often is available from the user's own collection of structure determinations. When the user calls for ORIENT, the system will automatically run the following sequence:
    1. First calculate the Patterson function (program FOUR), then
    2. Find the orientation of the model (program ORIENT), and
    3. Continue with translation functions (program TRACOR), and

    4 Finally further structure elucidation (programs PHASEX and FOUR).

- HELP provides help facilities.
- DIRP1 reduces the symmetry (temporarily) to P1 and completes a partial structure in this space group. DIRP1 is special option that can be useful for the solution of a structure in case the user is very uncertain about the space group, the composition of the compound, and/or the position of some heavy atoms. The option DIRP1 causes:

    1. The reflection data to be expanded to space group P1 (or centered equivalent e.g. C1), and
    2. The program PHASEX to be executed for elucidation of the structure in P1.

  The input atom may be, for instance, one atom in the origin! After inspection of the results (made asymmetric by the enantiomorph-fixing procedure) the user decides how to continue. When the user calls DIRP1, and ‘solves’ the structure in P1, he/she must recognize and locate the symmetry elements.
- PHASEX completes a partial structure. PHASEX is a program to refine phases and expand phases of the difference structure factors (calculated for a correct set of atomic parameters). The difference structure factors are normalized (Wilson- Parthasarathy) giving *E1* values, and weights are calculated (Woolfson or Sim). Reliable phases of the *E1* values are input to a modified tangent formula to refine all input phases and to find phases for unphased reflections. PHASEX is particularly known for its power to solve special symmetry problems which arise from an input model which is centrosymmetric in a non-centrosymmetric space group (enantiomorph problem) or a model which has higher translation symmetry (superstructure effects). The program recognizes the problem, and uses a special symbolic addition procedure to solve the enantiomorph and/or origin ambiguity problem. The program is used when some atoms are known (on correct positions). The program is automatically called after the execution of PATTY, ORIENT and TRACOR. However, the you should call for PHASEX when you have your own suggestions for atomic positions: for instance you may have corrected the atoms in the ATOMS file available from a previous run (which, of course, can only be useful if something went wrong.). When the user calls for PHASEX, the system will:

    1. Arrange for structure factor calculation and normalization, then
    2. Execute PHASEX followed by a Fourier synthesis (program FOUR), and
    3. Finally organizes recycling several times (programs PHASEX and FOUR) for further elucidation of the structure/

- TRACOR: This positions the fragment. TRACOR is a program to find the position of a molecular fragment (with correct orientation) by means of reciprocal space correlation functions. The input fragment is used for the calculation of partial structure factors for all reflections (for the entire expanded data set). The partial structure factor (sum of partial structure factors of symmetry related fragments) depends on the shift vector t. The correlation between calculated and observed structure factors (intensities) determines the best value for the shift vector t. The actual calculations are done by the Fast Fourier Transform method, and employs all symmetry elements simultaneously. The program is used for expanding structural fragments with correct orientation but unknown position. The program is automatically called after the execution of ORIENT. If the ‘best’ solution from ORIENT fails to solve the structure, however, the user may call for TRACOR manually, and supply the fragment rotated by the angular settings of the second ORIENT solution. Note that a correctly oriented fragment sometimes is available as the result of a failure of ab-initio direct methods; when a recognizable fragment does not allow expansion or refinement, then the fragment may be misplaced, however the orientation is certainly correct, and the user may call for TRACOR. The program is also a powerful tool for the elucidation of heavy atom structures. For instance, the origin and the next largest non-Harker Patterson peak define a pair of heavy atoms which can be used as a well oriented model to be positioned by the program TRACOR. When the user calls for TRACOR, the system will automatically run the following sequence:

  1. First expand the reflection data and calculate structure factors, then
  2. Find the position of the fragment (program TRACOR), and
  3. Finally find the remainder of the structure (programs PHASEX and FOUR, recycling).

- INTERACTIVE runs all of DIRDIF interactively - answer the questions as they are posed to you. This is the best way to run ORIENT if you wish to use the supplied database of structural fragments. In this case an orient calculation need not have the coordinates in the ATMOD file.
- FOUR runs the DIRDIF Fourier package - this is not normally used directly by the user. FOUR is a program for the calculation of a Fourier or Patterson map. It carries out a calculation of distances and angles, assignment and shuffling of new peaks into connected atoms, plotting of the asymmetric part of the structure, and finally arranging for recycling of a part of the structure. Note that the program’s decision of which peaks to assign and which atoms to use is only preliminary, and the user must apply his or her chemical knowledge to his local graphical system to make the final decisions. The program uses input files generated by other programs. It is automatically called after the execution of other programs.

If your structure does not come out as you wish or expect, and you have detected where the solution of the structure (probably) went wrong (the output listing file might give you an indication) you can rerun a part of DIRDIF either with non default parameters, or with a changed model, or using the second solution of ORIENT or TRACOR, etc. Sometimes DIRP1 is an interesting option (e.g. to solve a space group uncertainty).

8.2.1 Completing Structures with DIRDIF

In the main maXus window, the *Complete* menu offers the choices of DIRDIF or SIR. They work differently. Try this option in difficult cases especially when you have pseudo-symmetry problems, but be sure that you have a partial structure before you begin (For more details on SIR, see Chapter 7).

8.2.2 Acknowledgements

DIRDIF is written by:

P. T. Beurskens, G. Admiraal, G. Beurskens, W.P. Bosman, S. Garcia- Granda, R.O. Gould, J.M.M. Smits, and C. Smykalla.

Appendix A

Known Fragments Stored in the Fragment Database

Each fragment is referenced by two-digit numbers XX.YY. In the list below these are recognised as in the following example:

```
ATMOD 20.08 Cyclopentane (twist-envelope) C5 H10
```

The ATMOD keyword is followed by XX.YY. These are the two numbers requested when you run ORIENT in interactive mode.

```
ATMOD 01.01 D-Tartaric acid C4 H6 O6
ATMOD 01.02 Ethylacetate C4 H8 O2
ATMOD 01.03 Valeric acid C5 H10 O2
ATMOD 01.04 Valeric acid amide C5 H11 O N
ATMOD 03.01 Butylamine C4 H11 N

ATMOD 05.01 trans-Butadiene C4 H6
ATMOD 05.02 cis-Butadiene C4 H6
ATMOD 05.03 Butanole C4 H10 O
ATMOD 05.04 Hexatriene C6 H8
ATMOD 05.05 Undecane C11 H24
ATMOD 10.01 Isoxazole C3 H3 O N
ATMOD 19.01 Benzene C6 H6
ATMOD 20.01 Cyclopropene C3 H4
ATMOD 20.02 Cyclopropane C3 H6
ATMOD 20.03 Cyclobutene C4 H6
ATMOD 20.04 Cyclobutane (flat) C4 H8
ATMOD 20.05 Cyclobutane (twist) C4 H8
ATMOD 20.06 Cyclopentadiene C5 H6
ATMOD 20.07 Cyclopentene C5 H8
ATMOD 20.08 Cyclopentane (twist-envelope) C5 H10
ATMOD 21.01 1,4-Cyclohexadiene (flat) C6 H8
ATMOD 21.02 1,3-Cyclohexadiene C6 H8
ATMOD 21.03 Cyclohexene C6 H10
ATMOD 21.04 Cyclohexane (chair) C6 H12
ATMOD 21.05 Cyclohexane (twist-boat) C6 H12
ATMOD 22.01 Cycloheptatriene C7 H8
ATMOD 22.02 Cycloheptatriene C7 H8
ATMOD 22.03 Cyclooctatetraene C8 H8
ATMOD 22.04 1,3,5-Cyclooctatriene C8 H10
ATMOD 22.05 1,3,5-Cyclooctatriene C8 H10
ATMOD 23.01 (18)Annulene C18 H18
ATMOD 24.01 Naphthalene C10 H8
ATMOD 26.01 Anthracene C14 H10
ATMOD 27.01 Bicyclo[2.1.0]pentane C5 H8
ATMOD 27.02 Bicyclo[2.2.0]hexane C6 H10
ATMOD 27.03 Bicyclo[3.1.0]hexane C6 H10
ATMOD 27.04 Bicyclo[3.2.0]heptane C7 H12
ATMOD 27.05 Bicyclo[3.3.0]octane C8 H14
ATMOD 27.06 Indene C9 H8
```

```
ATMOD 27.07 Indane = Benzocyclopentane C9 H10
ATMOD 27.08 cis-Hydrindane C9 H16
ATMOD 27.09 trans-Hydrindane C9 H16
ATMOD 27.10 Benzocyclohexane C10 H12
ATMOD 28.01 Phenanthrene C14 H10
ATMOD 28.02 cis-anti-cis-Perhydroanthracene C14 H24
ATMOD 28.03 cis-syn-cis-Perhydroanthracene C14 H24
ATMOD 28.04 cis-trans-Perhydroanthracene C14 H24
ATMOD 28.05 trans-anti-trans-Perhydroanthracene C14 H24
ATMOD 28.06 trans-syn-trans-Perhydroanthracene C14 H24
ATMOD 31.01 Bicyclo[1.1.1]pentane C5 H8
ATMOD 31.02 Bicyclo[2.1.1]hexane C6 H10
ATMOD 31.03 Bicyclo[2.2.1]heptane = Norbornane C7 H12
ATMOD 31.04 Bicyclo[3.1.1]heptane C7 H12
ATMOD 31.05 Cubane C8 H8
ATMOD 31.06 Bicyclo[2.2.2]octane C8 H14
ATMOD 31.07 Bicyclo[3.2.1]octane C8 H14
ATMOD 31.08 Bicyclo[3.2.2]nonane C9 H16
ATMOD 31.09 Bicyclo[3.3.1]nonane C9 H16
ATMOD 31.10 Adamantane C10 H16
ATMOD 31.11 Bicyclo[3.3.2]decane C10 H18
ATMOD 31.12 Tricyclo[3.3.3]undecane C11 H20
ATMOD 31.13 Diamantane C14 H20
ATMOD 31.14 Dodecahedrane C20 H20
ATMOD 31.15 Buckminster-fulleren (C60) C60
ATMOD 32.01 1,2,4-Triazole C2 H3 N3
ATMOD 32.02 Imidazole C3 H4 N2
ATMOD 32.03 Pyrazole C3 H4 N2
ATMOD 32.04 Imidazolidine C3 H8 N2
ATMOD 32.05 Pyrrole C4 H5 N
ATMOD 32.06 Pyrrolidine C4 H9 N
ATMOD 33.01 1,3,5-Triazine (flat) C3 H3 N3
ATMOD 35.01 Benzimidazole C7 H6 N2
ATMOD 35.02 Indole C8 H7 N
ATMOD 35.03 2h-Isoindole C8 H7 N
ATMOD 38.01 Oxirane C2 H4 O
ATMOD 38.02 Oxetane C3 H6 O
ATMOD 38.03 Furan C4 H4 O
ATMOD 38.04 Tetrahydrofuran C4 H8 O
ATMOD 38.05 1,4-Dioxane (chair) C4 H8 O2
ATMOD 39.01 Thiirane C2 H4 S
ATMOD 39.02 Thietane C3 H6 S
ATMOD 39.03 Thiophen C4 H4 S
ATMOD 41.01 Thiazole C3 H3 S N
ATMOD 44.01 Uracil C4 H4 N2 O2
ATMOD 44.02 Cytosine protonated C4 H6 N3 O
ATMOD 44.03 Thymine C5 H6 N2 O2
ATMOD 44.04 Adenine protonated C5 H6 N5
ATMOD 44.05 Guanine protonated C5 H6 N5 O
ATMOD 48.01 Glycine (pi = 5.97) C2 H5 N O2
ATMOD 48.02 g-Aminobutyric acid zwitterion C4 H9 O2 N
ATMOD 50.01 Cephalosporin-C skeleton C8 H4 N2 S O
ATMOD 51.01 Estradiol skeleton C18 H20 O2
ATMOD 51.02 Testosterone skeleton C19 H21 O2
```

```
ATMOD 51.03 5a-Steroid skeleton C19 H26
ATMOD 51.04 Cortisone skeleton C20 H19 O3
ATMOD 51.05 Progesterone skeleton C20 H21 O
ATMOD 51.06 Cholestero skeleton C20 H22 O
ATMOD 58.01 Morphine skeleton C17 H19 O3 N
ATMOD 59.01 (+) Biotin skeleton C6 H7 N2 S O
ATMOD 59.02 Gliotoxin skeleton C13 H7 N2 S2 O3
```

9 MITHRIL

9.1 Using MITHRIL with maXus

Mithril is a direct methods program that allows full control of the phasing process, or automatic operation.

From the maXus window:

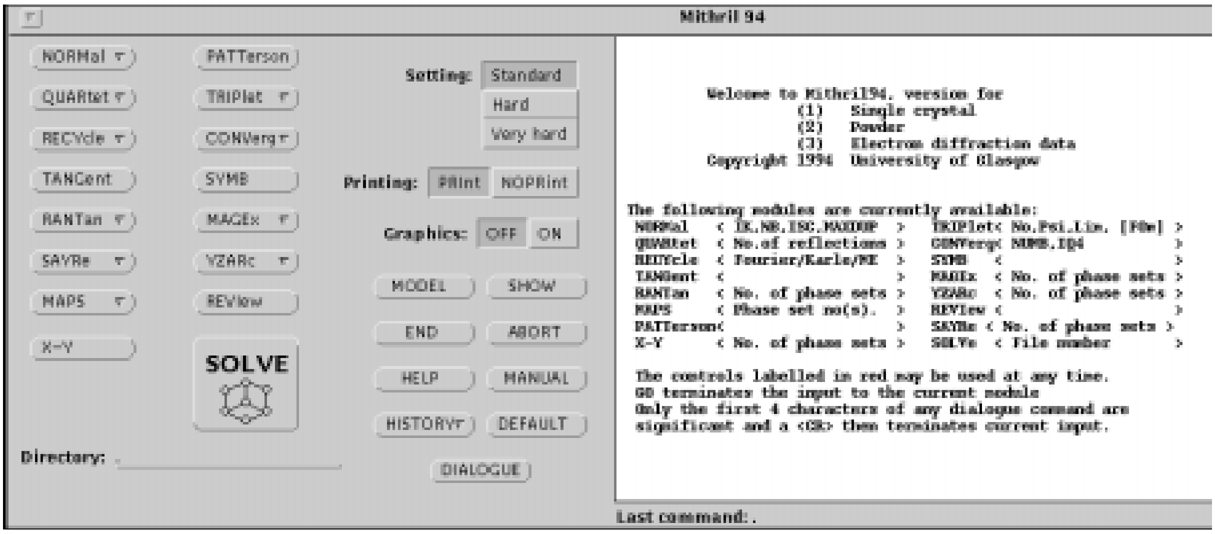
Current directory: /home/tcrystal/gbarr/testdata/bucky

From the *Solve* menu select the option *Mithril*. There are two options in the sub-menu:

1. Mithril94
2. Auto assign.

The button next to Auto assign is usually depressed. This means that once Mithril has run, it will enter SIR to assign atom types and carry out initial refinement, as long as the Fourier program in Mithril has been run. If you do not want to the program to do this automatically, then select this menu item - it will change to non-depressed. You will need to do this before running *Mithril*.

Once *Mithril94* is selected, the Mithril interface will appear:



```
        Welcome to Mithril94. version for
                   (1)   Single crystal
                   (2)   Powder
                   (3)   Electron diffraction data
        Copyright 1994  University of Glasgow


The following modules are currently available:
 NORMal   < IK,NB,ISC,MAXDUP   >   TRIPlet< No.Psi.Lim. [FOm] >
 QUARtet  < No.of reflections  >   CONVerg< NUMB,ID4           >
 RECYcle  < Fourier/Karle/ME   >   SYMB   <                    >
 TANGent  <                    >   MAGEx  < No. of phase sets >
 RANTan   < No. of phase sets  >   YZARc  < No. of phase sets >
 MAPS     < Phase set no(s).   >   REView <                    >
 PATTerson<                    >   SAYRe < No. of phase sets >
 X-Y      < No. of phase sets  >   SOLVe  < File number        >

 The controls labelled in red may be used at any time.
 GO terminates the input to the current module
 Only the first 4 characters of any dialogue command are
 significant and a <CR> then terminates current input.
```

Last command: .

Mithril94 is loaded and reads the reflection and model information from the files created by maXus. Whilst loading, the cursor will change to a small clock to indicate that Mithril is busy. When it reverts to an arrow, Mithril is ready to interact with the user.

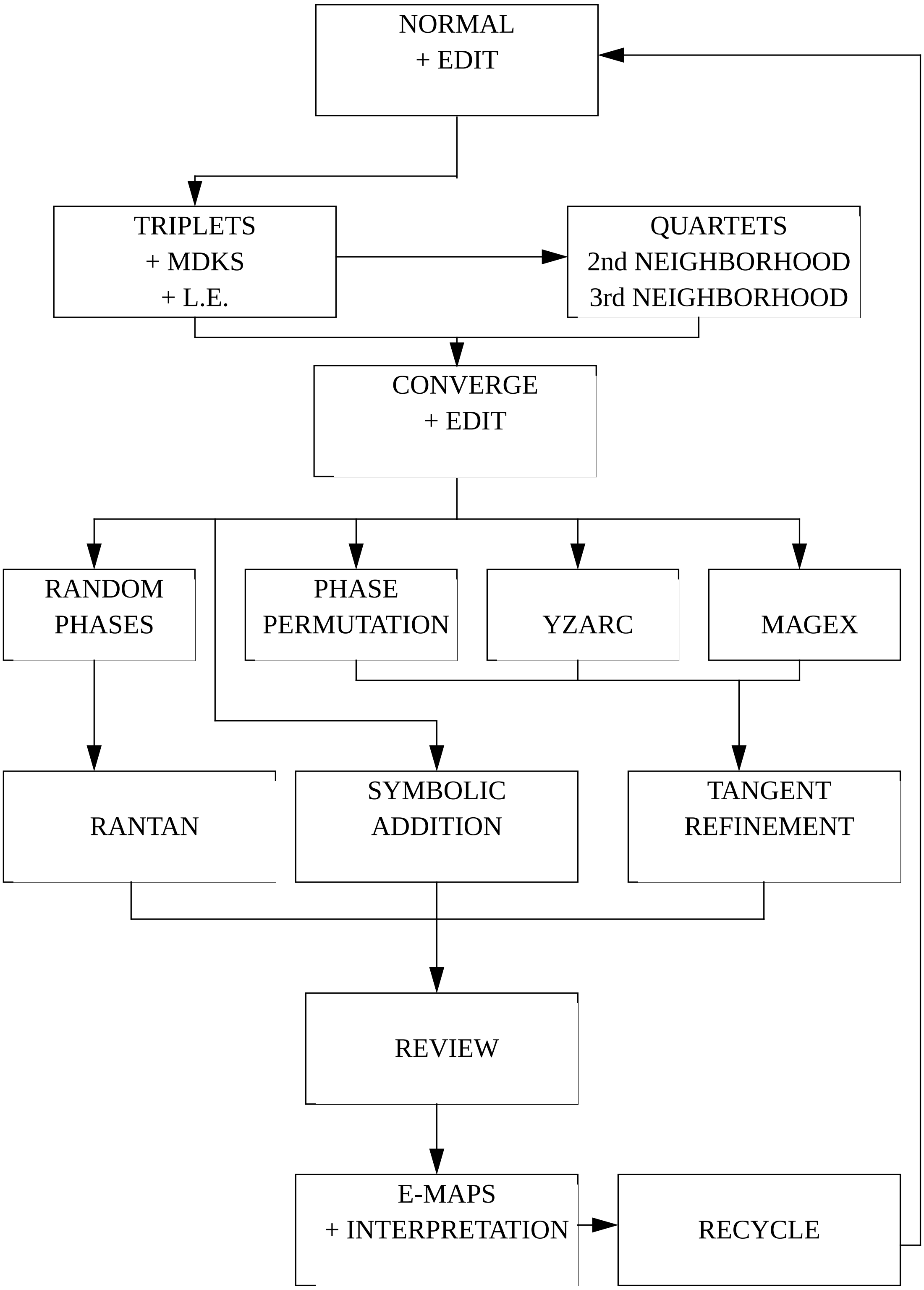


Figure 1: Mithril program modules flowchart

If you have not run Spacegroup a dialog box will appear:

Draw file does not exist

Click *OK* and the program will abort; otherwise proceed to use the commands as outlined in the following manual.

9.1.1 User Interface notes

Levels of user interaction are set dynamically and can be changed at any time, so that some modules may be run interactively and others under complete program control. The interactive modes are menu-driven with a separate menu for each module outlining available commands. Each instruction consists of a 4-character key word and a free-format list of input fields. Zero rather than blank fields generate default options. Each command has a sensible default value so that the user need only enter non-default options. Furthermore, there are four levels of default complexity, designed to cover all situations between the simplest and most difficult problems. These default levels are also dynamic and may be altered from module to module. Finally, the user need only specify those modules for which non-default options are to be used, the program will automatically run any modules that are needed but that have not been explicitly called by the user.

9.2 The SOLVE option in Mithril94

Mithril94 provides a host of facilities enabling users to tackle a very wide range of structures. The SOLVE option is intended to shield most users from the complexities of a structure solution by providing a series of automated paths to the solution. If the structure is not solved by one run with SOLVE, the next run will use a different set of program options. This philosophy is continued for up to 20 different sets of options, each set becoming progressively more computationally demanding than the last. By use of the Hard and Very hard settings, this can be extended to 60 sets.

9.2.1 Using the SOLVE option

We assume for the moment that this is the user's first attempt at solving the structure under investigation. Once Mithril has loaded, simply click the large SOLVE button. The program will read the first set of SOLVE options from a file and commence its attempt at a structure solution. As this is the first attempt at a solution, Graphics are turned on automatically to display the Plot of Wilson and Debye curves. Clicking left over this graph will dismiss it, and the Distribution of E^2 with $\sin\theta/\lambda$ is displayed. Click again to dismiss this and the Cumulative Probability Distribution will be displayed. Click again to dismiss this. The program continues to run, culminating in the generation and searching of the E-map corresponding to the "best" solution found. The program then terminates, and the user can return to the main maXus menu by clicking QUIT. Mithril output can be examined by selecting *Summary* then *Solve* from the SUMMARY FILES menu. Returning to the main maXus menu, clicking on MODEL will take you to a display of the peaks found in the “best” map, ready for atom assignment, tidying and refinement.

If you find that a particular set of SOLVE options works well for the structures that you study, then you can use the Dialogue box and type SOLVE N, where N is the number of the set of solved options that you prefer e.g. solve 4. See Section 9.2.4 for a list of the options.

9.2.2 What if SOLVE does not produce a promising solution?

Simply restart Mithril and click SOLVE again. The program will realise that the first set of options has already been tried, and will therefore use the next set and proceed as before. This process can be continued until all 20 of the SOLVE strategies have been exhausted. Of course, each of the strategies can be modified by selecting Hard or Very hard before clicking on SOLVE.

9.2.3 How do I know how many times SOLVE has been run?

Clicking on the HISTORY will show how many attempts have been made to solve the structure using SOLVE. If no attempts have been made, the program will respond with “The history list is empty”. Selecting CLEAR HISTORY from the pull down menu on the HISTORY button will clear the history list.

9.2.4 What is the program doing in a SOLVE run?

The commands executed by Mithril for each SOLVE run are shown below:

Option 1	NORMAL, MODEL, TANG, NOSTOP, DEFAULT, END
Option 2	NORMAL, MODEL, TANG, SWTR, NOSTOP, DEFAULT, END
Option 3	NORMAL, HARD, MODEL, QUART, TANG, NOSTOP, MAPS, END
Option 4	NORMAL, HARD, MODEL, QUART, RANT 250, NOSTOP, SWTR, MAPS, END
Option 5	NORMAL, MODEL, TRIP FO, QUART, SAYRE 250, START 55 37, NOSTOP, MAPS, END
Option 6	NORMAL, HARD, MODEL, X-Y 500, NOSTOP, MAPS, END
Option 7	NORMAL, HARD, MODEL, TRIP 800 FO, QUART 250, SAYRE 1000, NOSTOP, MAPS, END
Option 8	NORMAL, MODEL, MAGEX 100, TANG, SWTR, NOSTOP, END
Option 9	NORMAL, MODEL, YZARC 100, TANG, SWTR, NOSTOP, END
Option 10	NORMAL 1, SAYRE 250, NOSTOP, MAPS, DEFAULT, END
Option 11	NORMAL, BSCL 1.0, RANTAN 250, NOSTOP, MAPS, END

Option 12	NORMAL, BSCL 8.0, RANTAN 250, NOSTOP, MAPS, END
Option 13	NORM 1, MODEL, TRIP 0,1, SAYRE 500, START 77 99, NOSTOP, MAPS , END
Option 14	NORMAL, MODEL, TRIP 0,1 FO, QUART, X-Y 500, NOSTOP, MAPS , END
Option 15	NORM, MODEL, TRIP 800, QUART 250, NEIGH, RANTAN 500, SWTR, NOSTOP, MAPS, END
Option 16	NORMAL, MODEL, TRIP 0,1, QUART, NEIGH, SAYRE 400, START 123 457, NOSTOP, MAPS, END
Option 17	NORMAL 1, MODEL, TRIP 800,1 FO, QUART 250, SAYRE 1000, NOSTOP, MAPS, END
Option 18	NORMAL 1, MODEL, TRIP 600,1 FO, QUART 200, X-Y 500, NOSTOP, MAPS, END
Option 19	NORMAL 1, MODEL, TRIP 800,1 FO, QUART 250, X-Y 1000, NOSTOP, MAPS, END
Option 20	NORMAL 1, MODEL, TRIP 800,1 FO, QUART 250, POSI, NEIGH, RANTAN 1000, SWTR, NOSTOP, MAPS, END

9.2.5 What do *Hard* and *Very hard* do?

Essentially, they improve the chances of solving difficult structures and are well worth using if you are having problems. You should be aware of the fact they increase the computational time dramatically.

9.2.6 Alternatives to consider when SOLVE fails at any stage

- If, upon examination of the Mithril output file, all the figures of merit for the phase sets are very poor, then there is little point in investigating this set of solutions. In this case, the user should try the next SOLVE option.
- If, upon examination of the Mithril output file, the figures of merit look promising, then it may well be worthwhile doing one of two things:-

  1. Generating the map for the solution again, but requesting more peaks, as often it only takes a few missing peaks to turn a correct solution into one that is very difficult to understand.

  2. Examine some of the other solutions that were produced. Even though they may have inferior figures of merit, they may still be correct. The MAPS -N, where N is an integer that specifies the number of best solutions you wish to examine is useful in this respect.

  3. Try the simulated annealing option in tangent refinement.

Obviously, these two options require a higher degree of user interaction than SOLVE alone. Fuller details of the interactive use of Mithril are given in subsequent sections.

9.3 Data Collection

It may seem strange to include a discussion of data collection, but all direct methods calculation depend critically on the quality and resolution of the intensity data. These factors become more and more important as structural complexity increases or in cases of pseudo-symmetry, structural regularity (such as fused six-membered rings) or high thermal motion in the crystal. It is therefore important to plan your data collection with care especially if any of these features are expected. In particular:

1. Get your intensities as accurately as possible with at least some equivalent reflection which are subsequently merged to give a unique data set. Be suspicious of any discrepancies.
2. Measure weak reflections with a similar accuracy to the strong ones. Weak reflections generate small E-magnitudes and these play a critical part in the quartet and ψ_0 relationships which have a vital role in the Mithril94 package. Some diffractometer software places great reliance on pre-scans of peaks before measuring the intensity accurately. Weak peaks are then ignored. This increases efficiency but if the small E-magnitudes are only measured in a pre-scan, then they will have very large standard deviations and the relationships which use them will be unreliable. So make sure you input data collection parameters which will override the pre-scan options.
3. The higher the resolution, the better. The Cu sphere should be considered the minimum although for some crystals this is just not possible because of poor diffraction quality. The program does its best in these circumstances, but the chances of success are reduced. Sheldrick’s rule quantifies this as follows - *“If less than 50% of the reflections in the resolution range 1.1-1.2 Å are observed (i.e.* $|F| > 2.0\ \sigma(|F|)$*), then the structure will be difficult to solve by conventional direct methods”*. Two exceptions to this rule are very small structures and heavy atom structures. Be careful with very high angle, observed, weak reflections which become very large E-magnitudes. This is especially likely with high intensity tubes. This problem can be dealt with by the use of triplet and quartet weighting and cut-off techniques. These are discussed in more detail in the relevant sections. Another problem of unusually large E’s can be caused by solvent, which may produce very large E’s at very low resolution. It is found that it is best to remove these reflections from a data set.

9.4 Mithril94 Commands

In the remainder of this manual and in the programme itself the following nomenclature is used:

- Any text or part of a command text in lower case is optional e.g. **NORMal** means that only NORM is required but that NORMA or NORMAL is accepted.
- All commands appear in **LARGE, BOLD, CAPITALS**
- An essential blank is written as an underscore e.g. **TOP_**

Mithril94 commands consist of four or more characters which begin in the first column of each line, they may be in upper or lower case. Only the first four characters are significant, the rest are ignored. The parameters, if any, then follow on the same line in free format. Only columns

1-72 are scanned. The parameters are separated by blanks or commas. To get a default insert a parameter of zero - not a blank. This is important when entering parameters for those commands which allow more than one.

For example, in the module NORMAL the command LIMIt can be used. It has three parameters - sin θ maximum, sin θ minimum, and maximum permitted E-magnitude. If it is desired to use the first two parameters with their default values, but to specify a maximum E of 3.5 the command:

```
LIMIt 0,0,3.5
```

must be issued. The θ limits are given their default values and E-max is set to 3.5. Note that the command:

LIMIt ,,3.5 or LIMIt , ,3.5 is not acceptable.

If the fields fill the line, continuations are possible by using '='. The rest of the line is then ignored and another line is read. This should be blank in columns 1-4. As many continuations as needed may be used, but there is a limit of 200 parameters in total for any command. Example:

```
PHASe 1 3 4 5 6 7 8 9 11 23 34 45 46 56 102 =
104 119
```

Some commands require keywords instead of, or in addition to the numerical parameters. These keywords can be placed anywhere on the line. Only the first two characters are significant. E.g.

SIGMa ALL where 'ALL' is a keyword.

Comments can be inserted on a command by using '!'. Everything that follows this symbol is ignored.

There are three different types of command:-

(1) The general commands which alter the mode in which the program is run:

TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY_HARD, MODEL, SHOW, X.

These can be entered as often and whenever desired.

(2) Those which call modules -

NORMAL, TRIPLETS, QUARTETS, CONVERGE, YZARC, MAGEX, SYMB, TANGENT, RANTAN, REVIEW, MAPS, RECYCLE, PATTERSON

These commands cause the relevant module to be entered.

(3) Those which are particular commands for a given module only.

9.4.1 The General Commands

TITLe Enters a title which is printed at the top of each page. The default is a blank line. Continuations (‘=’ sign) are not allowed with this command.

END Tells the program that no further input will follow. The current module will be run to completion, and the package will then stop.

MENU In the interactive modes (LEVEL = 1,2,3) the screen is cleared and the current menu is displayed. If LEVEL = 0 the command is ignored.

NOPRint Switches off output to the print file (but not the secondary file used under LEVEL 0). This is useful if you are re-running a job for which the output has already been printed. The **PRINt** command turns the printing on again. You can use these commands as often as necessary.

LEVEl N The parameter N is an integer which can have the values 0, 1, 2 or 3. This signifies the degree of user interaction required. The levels operate as follows:

> N=0 This is used for batch jobs. No real-time user interaction is expected. No menus are displayed. A secondary print file is created which contains a list of the input commands as they are executed, the error and warning messages, and a summary of progress through the program.
>
> N=1 This is the lowest level of real-time operation. The user is expected to be sitting at a terminal. Menus are displayed on the screen, modules are run in real-time as selected by the user. A limited summary of progress and results appears on the screen. There is no secondary print file.
>
> N=2 This gives the same as N=1 but the screen output is more detailed and interrogation of the user may occur in some modules.
>
> N=3 This expects a high degree of user-machine interaction e.g. the screen is never cleared without user permission, tangent refinement stops after each solution with a request to continue and MAPS becomes more interactive. All N=2 facilities are also included.

The default will depend on your installation (at Glasgow we use N=2). Remember that the LEVEL parameter is quite flexible and can be changed as you proceed through a task.

DEFAult This removes further control of the package from the user, and runs the program to completion from the point at which the command is issued. Default parameters are used in all the subsequent modules, and the HARD and VERY_HARD options are still relevant. The default flow through the package is as follows:

(1) For structures with translational symmetry (non-symmorphic):

NORMAL - TRIPLETS - CONVERGE - TANGENT - MAPS

(2) For structures without translational symmetry (symmorphic) or for which the commands HARD or VERY HARD have been issued: _

NORMAL - TRIPLETS - QUARTETS - CONVERGE - TANGENT - MAPS

(3) For weighted Fourier recycling:

NORMAL - MAPS

(4) For Karle recycling of the non-symmorphic structures:

NORMAL - TRIPLETS - TANGENT - MAPS

(5) For Karle recycling of the structures of type (2) above:

NORMAL - TRIPLETS - QUARTETS - TANGENT - MAPS

(6) For situations where MAGEX or YZARC is being run:

NORMAL - TRIPLETS - QUARTETS - CONVERGE - MAGEX (or YZARC) - TANGENT - MAPS

or, in non-symmorphic situations with standard defaults:

NORMAL - TRIPLETS - CONVERGE - MAGEX (or YZARC) - TANGENT - MAPS

(7) RANTAN behaves just like TANGENT for default paths.

So, for example, if you have a symmorphic space group, and during or after running the triplets module you issue the command DEFAULT the modules QUARTET, CONVERGE, TANGENT and MAPS will be run under appropriate defaults, and the program will then halt. The modules MAGEX and YZARC are never called by DEFAULT, they are always user called.

HARD This gives a new level of defaults for structures which are proving difficult. It has the effect of generating more triplets; it automatically calls the quartets module with the third neighbourhood option; no reflections are rejected by converge; both MAGEX and YZARC phase larger starting sets; there is no automatic stopping during tangent refinement; more than one E-map is produced. It is possible to return to the standard level of defaults by entering the command HARD a second time.

VERY_hard This extends the options of the HARD command further by including the generation and use of positive quartets. Both commands make considerable demands on computer time and need to be used with care. It is possible in times of great need to enter both commands. This results in the cumulative effect of both options. It is possible to remove the VERY HARD option in the same way as for HARD.

MODEl In maXus, each structure has associated with it a ‘MODEL’ file in which the unit cell parameters, symmetry, lattice type, cell contents etc. are stored as Mithril94 compatible commands. The command MODEL is usually issued in NORMAL and it causes the MODEL file to be scanned, and any relevant information extracted. This saves entering the information directly from the keyboard. It is also possible to store Mithril94 commands on this file. Under these circumstances, the instruction MODEL will cause the commands to be executed before returning control to the user.

SHOW When operating interactively, it is possible to forget which options have been selected *via* PRINT and NOPRINT, or whether HARD and/or VERY HARD options are currently in operation. The command SHOW displays these parameters on the screen for an interactive job.

X Gives an immediate abort from the package.

9.4.2 Commands That Call Modules - Overview

The following commands call modules:

NORMal	**TRIPlets**	**QUARtets**	**CONVerge**	**SYMB**	**MAGEx**
YZARc	**TANGent**	**RANTan**	**REVIew**	**MAPS**	**RECYcle.**

In the interactive mode, calling a module will cause the terminal screen to be cleared, and a menu appears outlining the available commands which may be entered in any order. If you make a mistake with any command just re-enter it. The modules do, however, need to be run in a fixed order, and this is outlined in the table below. (Figure 1 may be useful here).

MODULE	MODULES WHICH MUST BE RUN BEFORE THIS MODULE (An * signifies an optional module call)
NORMAL	NONE (This is the package entry point)
TRIPLETS	NORMAL
QUARTETS	NORMAL, TRIPLETS
CONVERGE	NORMAL, TRIPLETS, QUARTETS *
MAGEX	NORMAL, TRIPLETS, QUARTETS *, CONVERGE
YZARC	NORMAL, TRIPLETS, QUARTETS *, CONVERGE
SYMB	NORMAL, TRIPLETS, QUARTETS *, CONVERGE *
TANGENT	NORMAL, TRIPLETS, QUARTETS *, CONVERGE
RANTAN	NORMAL, TRIPLETS, QUARTETS *, CONVERGE
REVIEW	NORMAL, TRIPLETS, QUARTETS *, CONVERGE, TANGENT
MAPS	NORMAL, TRIPLETS, QUARTETS *, CONVERGE, TANGENT

The consequence of this requirement, is that modules are called automatically with suitable defaults whenever the user enters a sequence of commands in which one or more of the necessary modules are missing. For example, the sequence:

NORMAL - TANGENT - MAPS

has TRIPLETS and CONVERGE missing. (QUARTETS is also missing if the structure is symmorphic, hard or very hard), so these modules are run with the appropriate default options in between NORMAL and TANGENT.

The sequence:

NORMAL - QUARTET - MAGEX - DEFAULT

will run NORMAL under user control, the TRIPLETS under default, CONVERGE under default, and MAGEX under user control. The command DEFAULT then invokes the flow already outlined on page 20, and the modules TANGENT and MAPS are then run under default. Note that although all these examples enter the package *via* NORMAL, the program can be entered at any point provided that the relevant modules have been run on a previous occasion and the necessary files have been kept. It is possible to go backwards as well as forwards in the program. When this is done the program goes directly to the module which has been called, it does not go through the remaining modules first. All the modules can be re-entered except TRIPLETS in which the MDKS option is invoked. E.g.

NORMAL - MAGEX - TRIPLETS - MAPS - YZARC - DEFAULT

will run the modules NORMAL, TRIPLETS, QUARTETS (if appropriate), CONVERGE, TANGENT, MAPS, YZARC, TANGENT, MAPS.

It is possible to override the default flow, if you really know what you are doing. Any module, except NORMAL, can be called with a “-1” as the first parameter. The module is entered, the program is informed of this, but the module is not actually run. One of the main uses of this is with the QUARTET module where it is possible to prevent quartets being generated in symmorphic space groups by using the command:

QUARTET -1

After all the commands for a module have been entered, one of the following operations will cause the module to be run:

(1) A call to another module. The current module is run and the new module entered. Calling the same module as the current one has two possible effects. In the interactive modes (LEVEL 1-3) the module is abandoned without running it, and re-entered. All the options must be re-typed. .

(2) The commands END or DEFAULT.

(3) If the LEVEL parameter is 1, 2 or 3, a carriage return or blank line. (If LEVEL is zero, blank lines are ignored on input)

9.4.3 Normal Module

For detailed discussions of the normalisation process see Rodgers (1965, 1980) and Giacovazzo (1980a). The code used here is an extensive modification of the MULTAN80 program. The normalisation procedure is often neglected by crystallographers since it is so automatic, but decisions made at this point have drastic implications for subsequent steps in the analysis, and full control over the process may be needed. Accordingly, the package offers the following features:

- Equivalent reflections and systematic absences are removed. The former distort the phasing procedure, while the latter can render relationships which rely on small E-magnitudes incorrect. In this latter category are the Q triplets and the negative quartets. Missing reflections may be added to complete a data set using Wilson statistics to estimate the magnitudes of the missing E values.
- Allowance is made for a lack of knowledge of the detailed contents of the unit cell.
- Should the traditional K-curve or Wilson plot techniques prove inadequate, a new Bayesian method of normalisation has been introduced. In addition direct input of E-magnitudes and/or phases is permitted and the normalisation procedure is bypassed. This allows interfacing to other normalisation programs.
- Editing facilities are provided to remove or modify structure factors before normalisation, or to remove or modify the subsequent E-magnitudes. The former facility alters the normalisation process, whereas the latter modifies only some of the normalised structure factors. Allied to these provisions is a set of optional θ limits and a maximum permitted E-magnitude. Experience has shown that E-magnitudes greater than approximately 3.5 can prevent the weighting schemes in a weighted multi-solution phasing environment from working with optimum efficiency as they tend to drive most weights to unity very rapidly. The relatively crude device of setting an upper limit can often remove this problem. The θ limits can be useful to exclude low-angle data which are subject to large systematic error, or the high angle reflections which can be very sensitive to small changes in overall temperature factor.
- Full control is provided over scaling and the temperature factor.
- The ability to input groups of known stereochemistry whose position and orientation in the unit cell can be either fixed or random (Main, 1976).
- The ability to normalise X-ray, electron, and neutron diffraction data.
- Normalise powders including overlapped reflections. The latter are used for normalisation but are not included in the list of reflections to be phased.

The normalisation module provides the entry point into the Mithril94 package. It must always be the first module run in an analysis. The menu looks like this:

The following commands are available; they can appear in any order except the DATA instruction which must come last.

NORMal IK, NB, ISC, MAXDUP [PHase] [NOsigma]

IK = 0 for use Wilson plot.

1 for use K-curve.

NB is the number of points to use in the Wilson plot. The default is

8*$\log_{10}$ [0.05 * Max(No. of reflections, 100)]. 10

ISC = 0 for one scale factor per parity group.

1 for use one overall scale factor.

MAXDUP. The NORMAL module checks for duplicates and systematic absences. Only the first MAXDUP found are listed. The default is 50.

[PHase] A keyword to indicate that phases are to input along with h, k, l, F, σ(F).

[NOsigma] A keyword to prevent the calculation of errors on the temperature and scale factors, and to signal that σ(F) is omitted.

LIST

Causes a full list of E’s (and F’s on an absolute scale) to be printed. Only use this if you really need it.

SFAC ATOM TYPE , a 1 , a 2 , b 1 , b 2 , c, COVALENT RADIUS, VAN DER WAALS RADIUS.

Inputs a scattering factor curve of the form:

$$f = a_1 e^{(-a_2 \rho^2)} + b_1 e^{(-b_2 \rho^2)} + c \qquad \text{(EQ 1)}$$

where $\rho = \sin\theta / \lambda$. (Moore, 1963).

Note the use of the five-parameter form here as opposed to the more usual nine-parameter form. The program has stored scattering curves for all the elements. You only need to enter an SFAC command if you need to use special scattering factors for any reason. The covalent and van der Waals radii are used by the MAPS module, and they can be omitted if desired. They are input in Angstroms.

CONTents ATOM TYPE(S), NUMBER IN THE UNIT CELL

This enters the unit cell contents (not the asymmetric unit). A list of symbols and the number of this type of atom in the unit cell must be included. It does not matter if the symbols come first or the types, the module associates the first symbol with the first numeric field, the second symbol with the second numeric field etc., so the following are all permitted:

CONTENTS 120 44 180 C O H

CONTENTS C 120 O H 44 180 etc.

If the CONTENTS command is missing, a cell in which only CH_2 groups are present, and for which the crystal density is 1.25 g/ml is assumed. Certain default options will be slightly modified in a pessimistic way if this is done e.g. the MAPS module will give you an extra 10 peaks. If an atom type is named which does not have a corresponding atomic scattering curve stored in the MITHRIL.DAT file (see SFAC) then an SFAC command will be needed for this atom.

LIMIts $\sin\theta_{max} / \lambda$, $\sin\theta_{min} / \lambda$, MAXIMUM PERMITTED E.

Any reflection with Bragg angles outside these limits is rejected. The defaults for the first two parameters are 0.0 and 1.0. The default for the third is 8.2, which represents a programming induced limit. Very large E- magnitudes can prevent a successful direct methods attempt, and this parameter merely sets any E greater than the maximum to the maximum itself. Values around 3.5 are usually successful.

NEWE H, K, L, NEW E-MAGNITUDE

This command enables the user to alter the magnitude of the specified normalised structure factor. If the new magnitude is given as zero, then the E is removed. Note that this takes place after the normalisation procedure, so the structure factor is used in the usual way in the calculation of the scaling parameters. Note also that the h, k, l indices must refer to the final E as output by NORMAL. This may have been symmetry transformed from its input form. A maximum of 100 NEWE commands are accepted.

MISSing

Find missing reflections, and, using Wilson statistics, calculate an expected F-magnitude, insert this reflection into the data, and re-normalise.

BSCL TEMPERATURE FACTOR, SCALE FACTOR(S)

This command allows the user to input his own temperature factor or scale factors. If only the temperature factor appears, then NORMAL calculates its own scale factor or factors depending in the ISC parameter on the NORMAL call. If the scale factor(s) are included on this command, then the temperature factor must be supplied as well. If it is not then a default of 4.0 is used. It is not possible to specify scale factor(s) alone and ask NORMAL to compute the temperature factor. If individual scale factors have been requested for each parity group and a single scale factor is input, this scale is applied to all the parity groups.

EDIT H, K, L, NEW F

This works like the NEWE command, except that it operates on the structure factors as input before normalisation, so the indices must refer to the input indices. A zero structure factor magnitude will cause the input reflection to be deleted. A maximum of 100 EDIT commands are accepted.

TRANs TRANSFORMATION MATRICES

This command allows the input reflections to have their indices transformed from their old values of h, k, l to new ones h', k' and l' as follows:

$$\begin{bmatrix} h' \\ k' \\ l' \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_5 & x_6 & x_7 \\ x_9 & x_{10} & x_{11} \end{bmatrix} \begin{bmatrix} h \\ k \\ l \end{bmatrix} + \begin{bmatrix} x_4 \\ x_8 \\ x_{12} \end{bmatrix} \qquad \text{(EQ 2)}$$

The elements x_1 to x_{12} are input in that order. If less than 12 elements are input, the remainder will be assumed zero.

If the TRANS command is being used in conjunction with the EDIT command, note that the reflections are checked *via* EDIT before the transformation matrices are applied.

ATOM ATOM LABEL X, Y, Z

This defines the coordinates of the group atoms. The label consists of an atom type as written on SFAC or CONTENTS commands with an optional label inside parentheses e.g. C(12) or C(12´) or Li(2A) or Li. Labels such as C12 or C2A are invalid. If the atom type and label is missing or unrecognised, then it is assumed to be carbon. A maximum of 200 atoms is permitted when totalled from all the groups with a limit of 100 atoms for all the type 3 or 4 groups.

GROUp TYPE(2/3/4/5/6), NO. IN CELL, CELL PARAMETERS

This command inputs the groups of known stereochemistry into the package. Each group (up to a limit of 10 groups) is followed by a set of ATOM commands (see below). The group type and the default cell associated with the group works as shown in the table below.

TABLE 3.

TYPE	GROUP		
	POSTION	ORIENTATION	CELL DEFAULT
2	Random	Random	Orthogonal A
3	Random	Correct	Orthogonal A
-3	Random	Correct	Crystal Cell
4	Correct	Correct	Orthogonal A
-4	Correct	Correct	Crystal Cell
5	Karle Recycling		Crystal Cell
6	Weighted Fourier Recycling		Crystal Cell

Groups of type 5 and 6 imply a recycling procedure is required. The cell parameters must refer to edges and angles not their cosines.

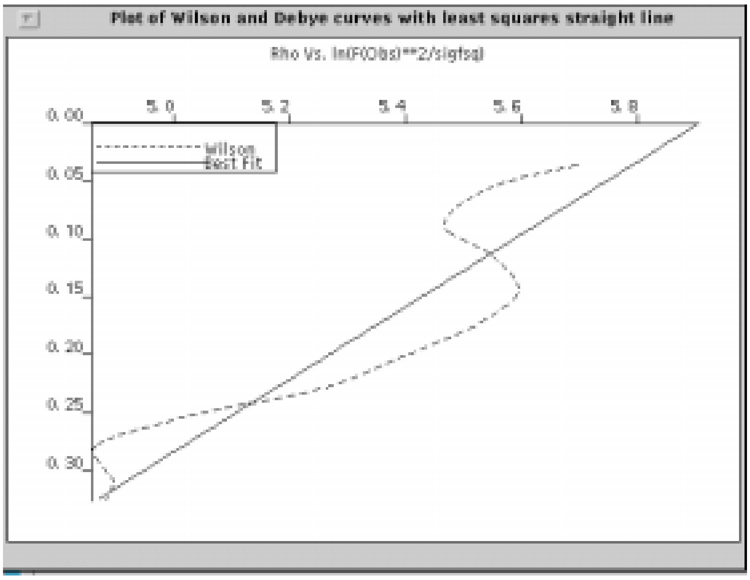
A missing or zero TYPE parameter is assumed to be 2. If the number of groups in the unit cell is missing or zero then it is assumed to be the number of equivalent positions including centres of symmetry and lattice centering**.**

NOCHeck

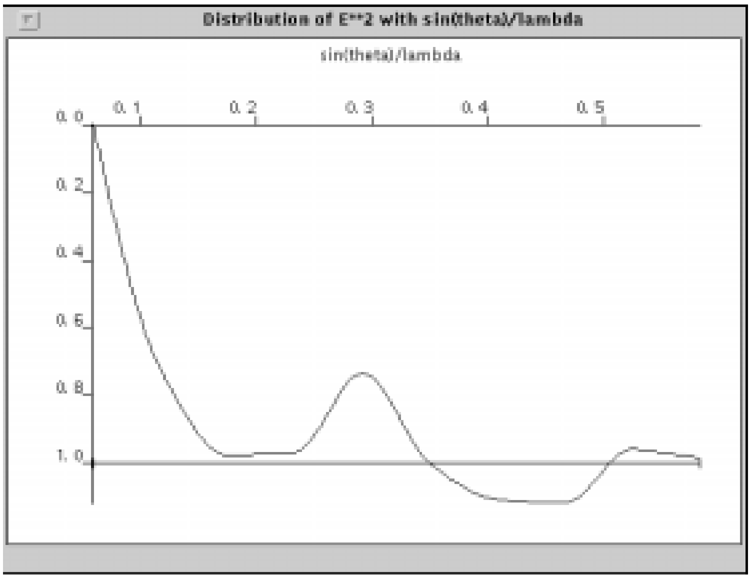
This switches off the checks for duplicate reflections and systematic absences in the intensity data. This can be useful if you already use a program to remove these reflections at data reduction time, since it prevents duplicate calculations and reduces the running time of NORMAL. However, only use it if you are certain that the duplicates and absences are missing otherwise you will cause havoc.

The commands TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY HARD, MODEL, SHOW and X are always available.

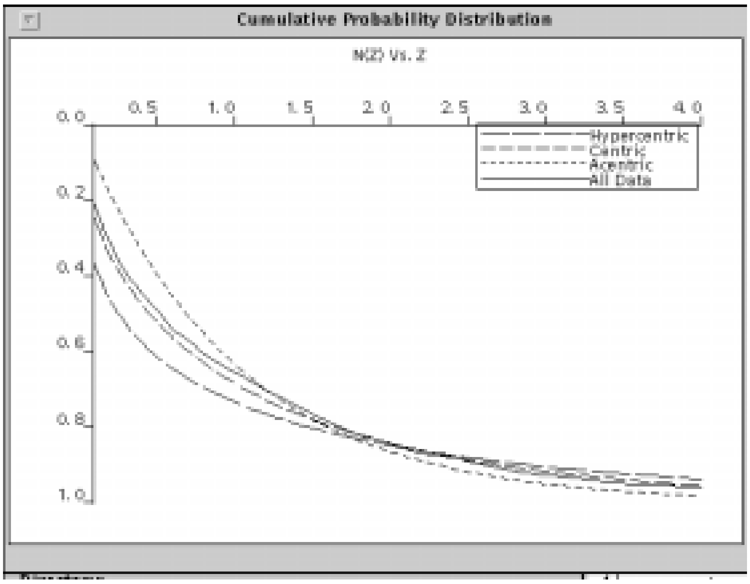
Three window appear if the graphics option is switched on. The first is the Wilson plot:



Click the box to clear it. The next is the variation of E^2 with Bragg angle:



and finally the E-statistics:



9.4.4 Triplets Module

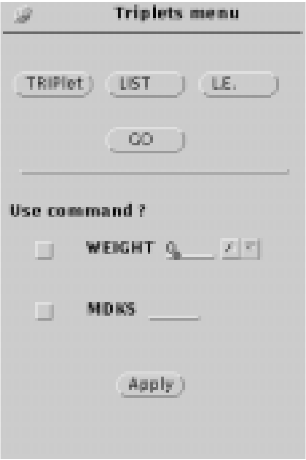
The Cochran distribution (Cochran & Woolfson, 1955) is used with the addition of two formulae for independently checking the triplet cosine. These are the MDKS formula (Hauptman, 1972) and a related technique (LE) in which a quintet extension of a triplet is used to derive a joint conditional probability distribution involving six E-magnitudes (Gilmore & Hauptman, 1985). The distribution is manipulated to give a system of 10 simultaneous linear equations in which the triple-phase invariant is one of the undetermined variables, and can be calculated in two ways. One is from the 10 simultaneous equations. Another method is *via* linear least-squares since three of the variables in the least-squares system are in fact the E-magnitudes involved in the triplet itself, and are thus known. The module requires that both estimates should agree within reasonable limits, otherwise both are ignored. Both MDKS and LE methods give only approximate estimates for the cosines, but they can be used to alter the relative weights of the triplets and for indicating which relationships may be troublesome.

The triplets are searched for those which give indications of the phases of one-phase seminvariants. If MDKS or LE has been used, an analysis of these triplets in terms of the estimated cosine is also given. This can be a useful adjunct in the decisions concerning the reliability of the Σ_1 phases.

A cut-off, input as a fraction of the largest value of $\sin^2\theta/\lambda^2$ can be used to remove triplets with two or more reflections having a resolution greater than this cut-off. A weighting scheme based on $\sin^2\theta/\lambda^2$ may be used to down weight unreliable triplets that involve any high resolution reflections. (See Gilmore & Brown, 1988). This can be very useful for very high resolution data sets.

This module must always be run, and will be executed with defaults if not called explicitly. The TRIPLETS module initialises the invariants file so that any invariants previously stored on it are lost.

The menu looks like this:



The available commands, which can be entered in any order, are as follows:

TRIPlets NO. OF REFLECTIONS, CUT-OFF, [FOm]

The first parameter is the number of reflections for which triplets will be generated. The default is:

No. of reflections =

4 * No. of atoms in asymmetric unit + 100 + 10 * Isymp + 50 * Idif + 150 * Ivdif

where:

Isymp = 0/1 for non symmorphic / symmorphic space group

Idif = 0/1 for a standard / hard structure

Ivdif = 0/1 for a standard / very hard structure

This is subject to a minimum of 250, or the number of E's greater than 1.0, whichever is the smaller. The maximum permitted is 800.

The second parameter is a cut-off, input as a fraction of the maximum sinθ^2/λ^2 of the data. This will prevent the use of triplets that contain two or more reflections with a resolution larger than the cut-off limit. The default is to use all reflections.

[FOm] is a keyword designed to cause the calculation of an extended triplet list, involving E's not used for phasing. This is used later in tangent refinement for the calculation of the LOGLIK figure of merit.

LIST

This causes a full list of each triplet to be output to the printer. In the interactive modes (LEVEL = 1, 2 or 3), this information is also output to the terminal. By using NOPRINT in conjunction with LIST it is possible to switch off the printing, but the user will still get the screen output. The print option uses one line per triplet, so use this option with care.

MDKS NUMBER OF CONTRIBUTORS REQUIRED FOR D AVERAGE

This invokes the MDKS calculations. The only parameter here specifies the maximum number of contributors required for the D average in the MDKS formula. The smaller this number is, the quicker the calculations will run, but the less reliable they will be. The default value is given by:

No. = max (5 * no. of atoms in asymmetric unit , 200)

Sometimes triplets appear where this number of contributors cannot be found. The program will accept such calculations provided there is at least half the required number. Any triplets with less than this are flagged as having no available MDKS estimates.

L.E.

This invokes a similar calculation to that of the MDKS formula, although it requires no scaling calculations. There are no parameters. Note the two periods in this command.

Both the MDKS and L.E. options are very slow and demanding; for this reason they are never called by default. Use them with caution.

WEIGht N

Use a weighting scheme based on sin θ^2/λ^2 to down-weight triplets that contain one or more reflections with a resolution greater than 0.95 times the resolution maximum. (See Gilmore & Brown, 1988)

N = 1 Print a list of the weighted E's

N = 0 Don’t print a list of the weighted E’s

As usual the commands TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY HARD, MODEL, SHOW and X are always available.

9.4.5 Quartets Module

Users unfamiliar with quartet theory are recommended to read a review by Hauptman (1980). Quartet invariants of the form:

$$\phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{-\underline{h}-\underline{k}-\underline{l}} = \Phi_4 \qquad \text{(EQ 3)}$$

are a very important component of Mithril94. Three types of quartet must be distinguished - that for which $\cos\Phi_4$ is estimated to be zero (a positive quartet), one in which $\cos\Phi_4$ is estimated to be 180 (a negative quartet), and the enantiomorph sensitive relationships between these two extremes. The negative quartets in particular are very useful. They can be used both as a figure of merit NQEST (DeTitta, Edmonds, Langs & Hauptman, 1975; Gilmore, 1977), and in an active mode to generate new phases (Freer & Gilmore, 1980). Two formulae are provided - the 7-magnitude, 2nd neighbourhood formula, and the more powerful 3rd neighbourhood, 13-magnitude formula (Hauptman, 1977a, 1977b). Missing members of these neighbourhoods are permitted, and the missing magnitudes are assigned values of unity. The use of the 2nd. neighbourhood formula is now widespread, but the 3rd. neighbourhood formula is still somewhat neglected.

Let us define the quartet which we wish to estimate as follows:

$$\phi = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{\underline{m}} \qquad \text{(EQ 4)}$$

such that:

$$\underline{h} + \underline{k} + \underline{l} + \underline{m} = 0 \qquad \text{(EQ 5)}$$

The first neighbourhood consists of $E_{\underline{h}}$,$E_{\underline{k}}$,$E_{\underline{l}}$ and $E_{\underline{m}}$, whilst the second neighbourhood adds the three cross terms $E_{\underline{h}+\underline{k}}$,$E_{\underline{k}+\underline{l}}$ and $E_{\underline{l}+\underline{h}}$. The addition of the third neighbourhood is accomplished by introducing an arbitrary vector $\underline{p}$ and its associated vector $\underline{q}$ such that:

$$\underline{h} + \underline{k} + \underline{p} + \underline{q} = 0 \qquad \text{(EQ 6)}$$

It is necessary that $E_{\underline{p}}$ and $E_{\underline{q}}$ are ‘large’. We now have a second quartet invariant:

$$\phi_{\underline{pq}} = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{p}} + \phi_{\underline{q}} \ (2) \qquad \text{(EQ 7)}$$

We have also indirectly defined a third quartet invariant:

$$\phi_{\underline{lm}} = \phi_{\underline{l}} + \phi_{\underline{m}} + \phi_{-\underline{p}} + \phi_{-\underline{q}} \qquad \text{(EQ 8)}$$

The quartet in Equation 5 has a second neighbourhood comprising:

$$E_{\underline{h}},\ E_{\underline{k}}\ ,\ E_{\underline{p}}\ ,\ E_{\underline{q}}\ ,\ E_{\underline{h}+\underline{k}}\ ,\ E_{\ \underline{k}+\underline{p}} \text{and } E_{\underline{p}+\underline{h}} \qquad \text{(EQ 9)}$$

The quartet in Equation 6 has a second neighbourhood comprising:

$$E_{\underline{l}}, E_{\underline{m}}, E_{\underline{p}}, E_{\underline{q}}, E_{\underline{l+m}}, E_{\underline{m-p}} \text{ and } E_{\underline{-p+q}} \qquad \text{(EQ 10)}$$

However, an identity exists between Equations 2, 5 & 6, such that:

$$\phi + \phi_{\underline{pq}} + \phi_{\underline{lm}} = 0 \qquad \text{(EQ 11)}$$

so that ϕ can be estimated not only by its own 7-magnitudes comprising its second neighbourhood, but also by the two invariants in Eq. 5 and Eq. 6. A total of 21 magnitudes are now involved in the estimation of ϕ, of which only 13 are unique. These 13-magnitudes define the third neighbourhood. The three quartets Eq. 1, Eq. 5 & Eq. 6 define a trio.

It is possible to construct a joint conditional probability distribution of the pair of structure invariants ϕ and $\phi_{\underline{pq}}$ given these 13 unique magnitudes. This is the $P_{2/13}$ distribution of Hauptman (1977a). It is possible to extract from this the $P_{1/13}$ distribution which gives ϕ as a function of 13 E- magnitudes.

Clearly, the third neighbourhood is not unique. There is a multiplicity of third neighbourhoods, each giving rise to an estimate for ϕ, and each estimate having its own variance. Direct methods will only allow us to use one of these. The best way to handle this situation is to use that estimate which has the lowest associated variance, at the same time discarding any invariants for which there is disagreement between the third neighbourhood calculations. It must also be stressed that the individual ϕ estimates are not independent because they will have many terms in common. The 3rd neighbourhood formula also goes some way towards alleviating the 1/N dependence of quartet reliability (N is the number of atoms -assumed equal- in the unit cell). It can also be very useful with weak or limited data sets (Gilmore, Hardy, MacNicol & Wilson, 1977).

In practical terms, it is necessary to restrict the vectors $\underline{p}$ and $\underline{q}$ to span the top 50-100 E-magnitudes, otherwise the calculations consume considerable amounts of computer time. There is very little loss of accuracy in doing this.

Positive quartets may also be generated if requested. They can sometimes be useful in situations where there is a deficiency in the number of triplets available, e.g. powder diffraction data. Positive quartets are correlated with triplets (Giacovazzo, 1980b) and this correlation is dependent on the E's involved. It is handled in the way described by Freer and Gilmore (1980). Higher invariants are put on the same scale as triplets using the concept of an equivalent $\underline{k}$ which is obtained from the variance of the appropriate probability distribution (Freer & Gilmore, 1980).

The negative quartet module is always called in situations where the space group is symmorphic. They are also generated in cases where difficulty in solving the structure is indicated by the user. As in triplets, there is an optional $\sin^2\theta/\lambda^2$ cut-off which eliminates all quartets that involve two or more reflections above a user specified limit, and a weighting scheme to down weight quartets that involve reflections with any reflections having a resolution above a specified limit.

This module generates the quartet invariants. It is always run in situations where the space group is symmorphic, or where the HARD or VERY_HARD command has been issued. The quartets are put on to the invariants file after the triplets. Any existing quartets will be overwritten. The user interface has the following appearance:

The available commands are as follows:

QUARtets NO. OF REFLECTIONS , CUT-OFF

The first parameter is the number of reflections for which quartets are to be generated. The default is:

No. of reflections = 100 + 15 * Isymp + 25 * Idif + 25 * Ivdif

where:

Isymp = 0 / 1 for non-symmorphic / symmorphic space group
Idif = 0 / 1 for standard / difficult structure
Ivdif = 0 / 1 for standard / very difficult structure

The maximum number of reflections is 256; it may not exceed the number of E’s greater than 1.0. The default is often unsuccessful and may result in either far too many or too few invariants. You may need to experiment with it.

The second parameter is a cut-off, input as a fraction of the current maximum sin θ^2/λ^2. It will prevent the use of any quartet that has two or more reflections with a resolution greater than that of the cut-off. The default uses all reflections.

POSItive

Usually the program generates only the negative quartets. This command will cause the positive invariants to be generated as well. It is a default option when the VERY_HARD instruction has been issued. It increases the computer time needed by a factor of at least two, and will generate a very large number of relationships - so use it with care.

LIST

This causes a listing of all the quartets to appear on the printer, and the terminal as well in the interactive mode. There is a considerable amount of output per invariant which will be even more extensive if the third neighbourhood has been invoked, so use this with care. It is possible just to get the screen output by using the NOPRINT and LIST commands together.

NEIGhbour SIZE OF 3rd NEIGHBOURHOOD, MAX NO. OF 3rd NEIGHBOURS TO EXPLORE

This command invokes the third neighbourhood calculations. There are two parameters:

The first dictates the range of the floating vector $\underline{p}$, which defines the neighbourhood. A value, for example, of 100 means that $\underline{p}$ is restricted to the top 100 E-magnitudes. The smaller this parameter is, the faster the calculation will run. The default is 100 which usually works well.

The second parameter specifies how many third neighbourhoods are to be found for a given quartet before moving on to the next. A full calculation in which all the neighbourhoods are explored is very time consuming, and may not be any better than a situation in which the search stops after 10-20 neighbours. The default is 10 unless the HARD or VERY_HARD command has been issued, in which case it is increased to 20.

MISSing MAX. NO. OF MISSING 2nd NEIGHBOURS, MAX. NO. OF MISSING 3rd NEIGHBOURS.

This command specifies how many missing reflections are allowed in the 2nd and 3rd neighbourhoods. The defaults are both zero unless the VERY_HARD command has been issued in which case they become 1 and 2 respectively. This command can be useful in situations where the data set is limited in some way, or in large structures. It increases the computer time required.

WEIGht N

Use a weighting scheme based on sin θ^2/λ^2 to down-weight quartets that contain one or more reflections with a resolution greater than 0.95 times the maximum resolution of the intensity data.(See Gilmore & Brown, 1988)

N = 1 Print a list of the weighted E’s

N = 0 Don’t print a list of the weighted E’s

As usual the commands TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY_HARD, MODEL, SHOW and X are always available.

9.4.6 Converge Module

The convergence mapping module follows invariant generation. It performs two quite distinct functions:

(1) The collection of invariants. All the invariants generated by previous modules are loaded for active use in the phasing procedures which follow, although it is possible to exclude higher invariants. Triplets are optionally weighted *via* their MDKS or LE cosine estimate, if they are available. A rather simpler method than that used by Bussetta and Comberton (1974) is employed. The invariants ϕ_3 are split into 4 classes:

(i) $\cos \phi_3 >= 0.7$	(ii) $0.7 > \cos \phi_3 >= 0.0$
(iii) $0.0 > \cos \phi_3 >= -0.7$	(iv) $\cos \phi_3 < -0.7$

Each class (i)-(iv) is assigned a weight greater than or equal to zero by the user, and this is used to multiply the κ value for the relationship, where:

$$\kappa = 2\, E_{\underline{h}}\, E_{\underline{k}}\, E_{\underline{h}-\underline{k}} / N^{1/2} \qquad \text{(EQ 12)}$$

($E_{\underline{h}}$, $E_{\underline{k}}$, $E_{\underline{h}-\underline{k}}$ are the E’s involved in the triplet).

By giving relationships in class (i) weights greater than unity, they can be upweighted and play a larger role in the phasing; in a similar way, those in class (iv) can be down-weighted or removed completely. The MDKS and LE formulae are unreliable in the two remaining classes and unit weights are usually used here. It is worth emphasising that this weighting scheme radically alters the phasing path as decided by the convergence method, and even if the MDKS estimates are unreliable, the resulting convergence map may be sufficiently different from the original that previous problems may disappear. In particular, unreliable triplets which appear early in the convergence map, even if not detected by the MDKS/LE tests, may now appear in less critical phasing areas.

Specific relationships may be deleted from or added to the list. Any 2-, 3- or 4-phase invariant or seminvariant may be added. This can be particularly useful as an adjunct to symbolic addition (Karle & Karle, 1966) where indications of relationships between symbolic phases or of potentially unreliable invariants can emerge. This information can now be supplied to the multisolution process in a simple way. The user must estimate the reliability of any input relationship by supplying a κ value.

(2) Convergence mapping itself. Once the relationships have been suitably ordered, the convergence map chooses suitable reflections which define the origin, the enantiomorph (if relevant), the known phases and the permuted phases. The user is provided with the usual options to define or partially define origin and enantiomorph, the permuted reflections, the maximum and minimum number of phase sets and the values of any known phases. Three criteria are applied in the acceptance of the Σ_1 results:

(i) A minimum probability.

(ii) A minimum number of indications (= (N+) + (N-) , where N+ is the number of zero indications and N- is the number of 180 degree indications)

(iii) A consistency ratio, c, defined as:

c = max(N+,N-) / ((N+) + (N-))

Caution in accepting a Σ_1 determined phase cannot be overemphasised.

Care is also needed with the active employment of quartets. It is obvious that with a triplet, two known phases can derive a third, whereas for a quartet, three are required to give a fourth. If the phase angles are very approximate, as they are in the early stages of direct methods, quartets will tend to propagate errors more than triplets. Accordingly, the number of higher invariants should be a maximum of ca. 20% of the total number of invariants used, although this rule may be broken for powder data sets.

At this point the negative invariants are assembled for use in NQEST. It is important to use reliable invariants, and thus only relationships with an equivalent κ greater than 1.0 are accepted. A minimum of 25 invariants is needed for NQEST. The ψ_0 figure of merit is also set up at this point, and again there is a minimum number of relationships needed for this figure of merit to be invoked.

The reflections which are eliminated last in the convergence procedure are those which will be phased first in the tangent refinement module. In other words, the map represents an inverted phasing path which is stored, and used by all subsequent modules. In difficult circumstances, it is necessary to examine this map carefully, and it should always warrant more than just a cursory glance.

Reading the convergence map output is straightforward. A typical line of output looks like this:

```
273 0 512.312 -4 0126.517 -6 804.8
-4 18 1 01.7 1 4 88 121.8
```

The reflection to be phased at this point is number 27 with indices 3 0 5. It has an alpha value of 12.3 at its point of elimination. Four relationships are shown as contributors. The first is a triplet:

$\phi_{12} - \phi_4 + (12/24) = \phi_{27}$

with a κ-value of 6.5, and a phase shift of (12/24) x 360 = 180˚. The second is a quartet involving reflections 17, 6, 8 and 27; the phase shift is zero and the κ-value is 4.8. The next relationship is also a quartet, involving reflections 1, 4, 18 and 27. This is followed by another triplet.

The phases to be permuted are listed at the end of the convergence output, along with origin and enantiomorph definitions. Permutation of general phases is carried out using magic integers, where each general phase (one with no phase restriction) is defined by:

$\phi = M_T X$

M_T is the magic integer, and X is a variable which takes on different values at equal intervals in the range 0-360˚. The interval of X is chosen to make the r.m.s. difference from one set to the next equal to the r.m.s. error in the ‘best’ phase set.

The user may specify the number of general reflections (Ngen), and the number of special reflections (Ns) the starting set. General reflections have no phase restrictions; special reflections have only two possible values. The number of starting sets so generated can be calculated by:

No. of sets = Nsets * 2 Ns * Ef

Nsets is the number of phase sets produced from the general reflections *via* magic integer phase permutation (Main, 1978). Ef is an enantiomorph factor. If the magic integer variable is set at 2 (the default), Nsets can be derived from the number of general reflections in the starting set (Ngen) from the following table:

TABLE 4.

Ngen	0	1	2	3	4	5	6	t7	8	9	10
Nsets	1	4	14	18	30	48	80	130	214	348	568
Rms error	0	26	28	39	42	45	47	48	49	50	50

The value of Ef depends on how the enantiomorph is fixed:

Ef = 1/2 if an unknown phase defines the enantiomorph - the range of permutation is restricted to 1/2 the phase circle.

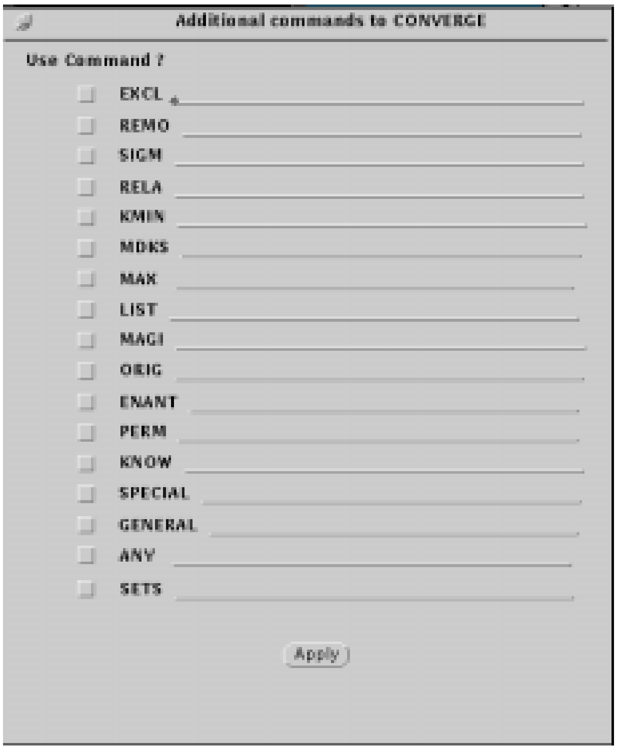
Ef = 1 if the enantiomorph is defined by restricting the origin defining phases to fixed values or the enantiomorph is defined by the space group.

Ef = 2 if the enantiomorph is defined by restricting the origin defining phases to two sets of values-one phase must take on two values.

Ef = 4 if the enantiomorph is defined by restricting the origin defining phases to four sets of values-two phases each take on two values.

The CONVERGE module carries out the dual functions of collecting together the invariants (and seminvariants), then selecting the reflections which comprise the starting set. Any complete direct methods analysis must include this module, and it will be run under defaults if not called explicitly. Press the CONVerge button. The following box appears:

The other commands button brings up the following dialog box.



The available instructions, which may appear in any order, are as follows:

CONVerge NDET, IQ4,

NDET is the number of reflections which are to be phased by tangent refinement. The default works as follows: The maximum possible value is defined by the number of reflections for which invariants were generated. This value is usually reduced by 10% to remove those reflections which are mostly weakly linked into the phasing procedure and this is the default unless:

- There are less than 101 reflections to be phased or
- The command HARD or VERY_HARD has been issued.

In both these cases the 10% reduction is not applied. If this procedure is unsatisfactory then enter your own value of NDET.

IQ4 =	0 for negative quartets to be used actively and passively; no positive quartets are to be used. This is the default.
	1 for use negative quartets as figures of merit only; no active use of positive or negative quartets.
	2 for use positive and negative quartets actively, and negative quartets as figures of merit as well. This is the default for VERY_HARD structures.

EXCLude H, K, L or REFLECTION NUMBER

All the invariants involving this reflection are removed from the input list so that the reflection is not phased. The reflection can either be specified by its serial number or by its indices.

REMOve RELATIONSHIP

This removes any specified phase relationship from the list of invariants. Any 1- , 2- , 3- , or 4-phase relationship can be deleted in this way. The relationship is specified by arranging the moduli of the serial numbers involved in descending order. The associated κ-value, and the phase shift are ignored. The serial numbers are then included as parameters of the REMOVE command. A limit of 100 REMOVE commands is imposed.

SIGMa [ALl] /[NOne] or MIN. PROBABILITY, MIN. CONSISTENCY RATIO.

This defines the criteria by which one-phase seminvariants (the Σ_1 reflections) are selected for the starting set. This command either uses the keywords “ALL” / “NONE” or utilises two numeric fields which define the acceptance parameters.

NONE requests that no Σ_1 type reflections are accepted. This is the default for symmorphic space groups.

ALL requests that all the Σ_1 reflections be put into the starting set suitably weighted.

MINIMUM PROBABILITY specifies the minimum acceptable probability. Default is 0.95 unless the space group is symmorphic, in which case default is set so that no reflections are accepted.

MINIMUM CONSISTENCY RATIO The default is 0.67. Note that a minimum of three contributors is also required for a Σ_1 reflection to be accepted.

RELAtionship SERIAL NUMBER(S), PHASE SHIFT, κ-VALUE

Inputs a 1-, 2-, 3- ,or 4-phase invariant or seminvariant into the list of relationships. The relationship is arranged in descending order of the moduli of the serial numbers involved; the phase shift is expressed in 24ths (e.g. a phase shift of 180° is 12); the κ-value is usually a guess at this point, but it must be included. The larger the value, the more reliable the relationship is deemed to be. If you want it to have a very large weight, then use a value of 10.0 - 20.0. The relationship is then input *via* the RELATIONSHIP command in which the parameters are the serial number(s) involved (complete with signs), the phase shift in 24ths, and the κ-value.

Suppose that symbolic addition strongly suggests that:

$$\phi_{12} = \phi_{23} + 180$$

rearranged it becomes:

$$\phi_{23} - \phi_{12} + 180 = 0$$

and it can now be input to CONVERGE *via*:

RELATIONSHIP 23 -12 12 10.0

KMIN MINIMUM ACCEPTABLE κ-VALUE

Specifies a minimum acceptable κ-value. The default is 0.6. Lower this value for difficult situations, or occasions where there is a paucity of phase relationships.

MDKS W1, W2, W3, W4

If the MDKS or L.E. command was issued when the TRIPLETS module was run, then a record of the estimate of the cosine of each invariant is stored on the invariants file. This information can be used to change the κ-values of the triplets. The estimated cosines are split into four classes (i) - (iv) and each class is assigned a weight which multiplies the κ-value of the triplet. The classes are:

- $\cos\phi_3 \geq 0.7$ Weight is W1, default = 1.2
- $0.7 > \cos\phi_3 \geq 0.0$ Weight is W2, default = 1.0
- $0.0 > \cos\phi_3 \geq 0.7$ Weight is W3, default = 1.0
- $\cos\phi_3 < -0.7$ Weight is W4, default = 0.0

A weight of 0.0 or less means that the relationship is deleted. Unless the MDKS command is issued, the cosine estimates from the TRIPLETS module are not used.

MAX_ MAXIMUM NUMBER OF PHASE RELATIONSHIPS

Usually the module uses all the available relationships up a limit of 28350, which is the program limit. If you want less than this then use this command. There is a minimum of 100.

LIST [ALL] / [NOne] / [PArtial]

This command uses keywords to specify how much of the convergence map is to be printed.

ALL prints the entire map.

NONE suspends all convergence map printing.

PARTIAL prints only the last 60 phasing steps of the map. This is the default.

MAGIc MAGIC INTEGER SEED

Specifies the number to be used as the seed for the series which acts as the magic integer sequence generator. The default is 2. Note that this command can decrease the number of phase sets generated, at the expense of reduced accuracy of the starting set.

ORIGin REFLECTION NUMBER

Usually, CONVERGE defines its own origin, but this command can be used to specify a reflection that is to be used for origin definition. The module will reject any reflection that is invalid. Note that it is not necessary to fully define the origin *via* a sequence of several ORIGIN commands. The module recognises partially defined origins, and will complete the definition during convergence.

ENANtiomorph REFLECTION NUMBER

Usually, CONVERGE defines its own enantiomorph, but this command can be used for a user defined enantiomorph. The module will reject an invalid reflection.

PERMute REFLECTION NUMBER

Specifies a reflection which is to be included in the starting set. Its phase is unknown, but it is given a sequence of values as generated by the magic integer routines in tangent refinement, unless it is a special reflection in which case it is given its two possible phase values in turn.

KNOWn REFLECTION NUMBER, PHASE (in DEGREES), WEIGHT

Allows a reflection of known phase to be included in the starting set. Its phase (in degrees) must be specified, and an associated weight in the range 0.0 - 1.0. The default value is 1.0.

SPECial NUMBER OF SPECIAL REFLECTIONS IN THE STARTING SET

This command defines the total number of special reflections to be included in the starting set, and whose phases are to be permuted. This includes any relevant reflections entered *via* the PERMUTE command.

GENEral NUMBER OF GENERAL REFLECTIONS IN THE STARTING SET

Specifies the total number of general reflections to be included in the starting set. The phases of these reflections are permuted in the usual way.

ANY_ NUMBER OF ANY SORT OF REFLECTIONS IN THE STARTING SET

Specifies the number of any sort of reflection, special or general, which are to be included in the starting set, and whose phases are to be permuted.

SETS MAXIMUM NO. OF PHASE SETS, MINIMUM NO. OF PHASE SETS

This command specifies the maximum and minimum number of phase sets to be generated. The maximum number of phase sets is only used if the user does not specify the starting set in any way *via* the GENERAL, SPECIAL or ANY commands. The default is 65 with a maximum of 2250, and a minimum of zero.

The minimum number of phase sets can also be specified. Again it is only used if the starting set has not been specified in any way. The default is calculated as follows:

Min. no. of sets = min(2*Nasu/3, Max. no. of sets/2 - 1) + 64*Ivdif + 32*Idif where:

Nasu = No. of atoms in asymmetric unit.

Ivdif = 0 / 1 for standard / very hard structure.

Idif = 0 / 1 for standard / hard structure.

This is subject to a minimum of 12. If the minimum number of phase sets is specified, but the maximum number is left as a default, then the latter is given the value:

Max. no. of sets = Min. no. of sets * 2 + 1

The general commands TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY_HARD, MODEL, SHOW and X can be issued at any time.

9.4.7 Sayre Module

The Sayre option acts as an alternative to tangent refinement. This uses both the standard triplets and the ψ_0 triplets actively in the tangent refinement. It can be very powerful, and is strongly recommended for difficult structures. Random phases are used throughout. It is used just like TANG or RANTAN and the available commands are as follows:

SAYRe No of phase sets.

Calls the SAYRE module. It must either follow CONVERGE, or the converge results must be available. The number of phase sets to generate has the default:

No of Phase Sets: 50 + 200 * Idif + 400 * Ivdif

where

Idif = 0 / 1 for Standard / Hard structure.

Ivdif = 0 / 1 for Standard / Very Hard structure.

The user interface is:

NOSTop

If the SAYRE module finds a solution with figures of merit that satisfy the conditions:

- Resid (R_{Karle}) less than 20.0
- ψ_0 less than 1.25
- NQEST less than -0.15.
- The figures of merit above are within 5% of the best so far

Then the module assumes that this is the correct solution and exits. Users in the interactive mode will be questioned first if they wish to accept this solution. The command NOSTOP switches off these tests. So do the HARD and VERY_HARD options. NOSTOP is often worthwhile as the correct solution can often be missed under the early stop algorithm.

ANEA

Requests simulated annealing

SNOAN

Switches off simulated annealing.

START IX,IY

Two odd integers used to seed the random number generator that generates phases for the SAYRE module. The defaults are 1,1. Even integers are converted to odd.

WTFOm W1, W2, W3, W4

This command defines the relative weights of the four figures of merit used by the TANGENT module when calculating a combined figure of merit (CFOM). The defaults are as follows:

TABLE 5.

Figure of Merit	Weight	Default	Default in symmorphic cases
ABS FOM	W1	1.0	0.6
ψ-ZERO	W2	1.0	1.2
RESID	W3	1.0	0.6
NQEST	W4	1.0	1.3

If there are no appropriate relationships for a particular figure of merit then a weight of zero is assigned. The relative weights are normalised such that the maximum CFOM value is equal to the total number of figures of merit contributing to it. The LOGLIK figure of merit, if calculated, is not used in the combined figure of merit.

WEIGHTS Small quartet weights, large quartet weights.

SAYRE uses two sorts of quartets:

1. Negative quartets generated from the ψ_0 relations.
2. Positive quartets generated from triplets.

These have a relative weights which are normally defaulted by the program but the user may override these if they so wish. The default for the positive quartet contribution is always zero.

For more information on the Sayre tangent formula, see:-

Debaerdemaeker, Tate and Woolfson, Acta Cryst (1986) ***A41****, 286-290*

9.4.8 X-Y Module

This is an alternative to tangent refinement. This can be called instead of RANTAN or TANG as follows:

X-Y No of phase sets.

Calls the X-Y module. It must either follow CONVERGE, or the converge results must be available. The number of phase sets to generate has the default:

No of Phase Sets: 50 + 200 * Idif + 400 * Ivdif

where

Idif = 0 / 1 for Standard / Hard structure

Ivdif= 0 / 1 for Standard / Very Hard structure.

The dialog box is:

The options are:

NOSTop

If the X-Y module finds a solution with figures of merit that satisfy the conditions:

- Resid (R_{Karle}) less than 20.0
- ψ_0 less than 1.25
- NQEST less than -0.15.

- The figures of merit above are within 5% of the best so far

Then the module assumes that this is the correct solution and exits. Users in the interactive mode will be questioned first if they wish to accept this solution. The command NOSTOP switches off these tests. So do the HARD and VERY_HARD options. NOSTOP is often worthwhile as the correct solution can often be missed under the early stop algorithm.

ANEA

Requests simulated annealing

SNOAN

Switches off simulated annealing.

SETS

Defines the number of phase sets to be produced.

STARt IX, IY

Two odd integers used to seed the random number generator. The default values are 1 and 1 (cf. YZARC).

WTFOm W1, W2, W3, W4

This command defines the relative weights of the four figures of merit used by the TANGENT module when calculating a combined figure of merit (CFOM). The defaults are as follows:

TABLE 6.

Figure of Merit	Weight	Default	Default in symmorphic cases
ABS FOM	W1	1.0	0.6
ψ-ZERO	W2	1.0	1.2
RESID	W3	1.0	0.6
NQEST	W4	1.0	1.3

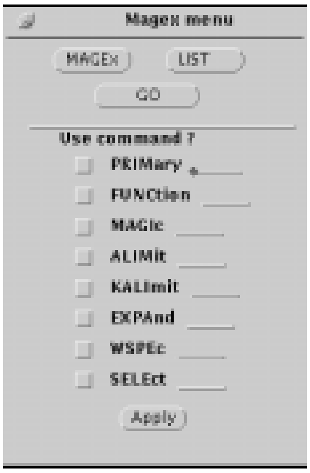
If there are no appropriate relationships for a particular figure of merit then a weight of zero is assigned. The relative weights are normalised such that the maximum CFOM value is equal to the total number of figures of merit contributing to it. The LOGLIK figure of merit, if calculated, is not used in the combined figure of merit.

9.4.9 Magex Module

MAGEX is another module run between convergence mapping and tangent refinement. Users unfamiliar with the concepts employed here should first read two papers - White and Woolfson (1975) and Declercq, Germain & Woolfson (1975). These deal with a three-dimensional form of MAGEX - the MAGIC procedure. A paper by Hull, Viterbo, Woolfson and Shao-Hui (1981) shows how the method is adapted to the one-dimensional case employed here.

MAGEX has been inserted in the Mithril94 package without any major modifications. Higher invariants are not employed here except that an NQEST value is assigned to each possible solution just as in YZARC.

MAGEX is always user called, and never entered by default. The relevant user interface is:



MAGEx NO. OF PHASE SETS TO GENERATE

This specifies how many phase sets are to be passed to tangent refinement. The default is:

No. of sets = 80 + 50 * Idif + 50 * Ivdif

where:

Idif = 0 / 1 for Standard / Hard structure.

Ivdif = 0 / 1 for Standard / Very hard structure.

LIST

This causes the primary and secondary reflections and the triplets linking them to be listed on the printer.

PRIMary NUMBER OF PRIMARY REFLECTIONS TO USE

This command defines how many primary reflections are to be used to initialise the MAGEX procedure. The default is:

No. of primaries = 10 + 6 * Max (Idif , Ivdif)

There is a maximum of 22 primaries.

FUNCtion FUNCTION TO USE IN PARAMETER SHIFT REFINEMENT (N)

There are four options here:

(1) N=1 uses $\sum_t \kappa_t \frac{I_1(\kappa_t)}{I_0(\kappa_t)} \cos\phi_{3_t}$

Special reflections are allowed to take any value, and then reset to the closest permitted value at the end of refinement. This is equation (16) in the paper by Hull, Viterbo, Woolfson and Shao-Hui (1981).

(2) N = -1. Uses option (1) above, but special reflections are only allowed to take one of their two possible values throughout.

(3) N = 2 uses $\sum_r \left(\sum_s K_{rs} \cos\phi_{3_{rs}} - \left| \sum_s K_{rs} \sin\phi_{3_{rs}} \right| \right)$

Special reflections are treated as in option (1). This is equation 17 in the paper of Hull et al. (1981).

(4) N = -2 uses the same function as option (3) but constrains the special reflections in the same way as option (2).

The default value is option (3) i.e. N = 2. For centrosymmetric cases only options (1) and (2) i.e. N = 1 or -1 are permitted.

MAGIc N

Defines the Magic Integer sequence. If N is a small positive integer then the integers are based on the sequence:

$F_n = F_{n-1} + F_{n-k}$

where k = N as defined above. A value of N = 1 gives a power of two series, whilst N = 2 gives the Fibonacci series which is the default. A value greater than 2 gives the hyper - Fibonacci series. If N is negative the sequence:

$F_n = 2 * F_{n-1} + F_{n-k}$

where k = N, is used instead.

ALIMit MINIMUM ALPHA VALUE FOR ACCEPTING A SECONDARY

This command defines the lowest estimated alpha for accepting a secondary reflection. The default is 1.7, but for centrosymmetric cases this should be increased to around 4.0. If the option chosen results in too many relationships, it will be automatically reduced by MAGEX.

KALImit LOWEST κ-VALUE FOR SECONDARY DEFINITION

This defines the minimum κ-value a triplet may have if it is to be used in secondary definition. The default is 1.0.

EXPAnd NO. OF CYCLES

This repeats the MAGEX process for the specified number of cycles. Use this option with care. It is not recommended that more than two cycles are performed, but more are possible if you wish. The default is zero.

WSPEc WEIGHT

Each potential primary reflection has an omega function associated with it (See equation (5) in Hull, Viterbo, Woolfson and Shao-Hui, 1981), and the reflections with the largest omega values are selected as working primaries. The weight defined on this command multiplies each omega value when a special reflection is involved. Setting the weight to a value less than unity can be used to avoid having too many special reflections as primaries. The default is 1.0.

SELEct IFUNC

This selects the function to use for ranking the solutions derived from the small ψ-maps. IFUNC can have only 2 possible values:

IFUNC = 1 uses function (1) defined in the FUNC command.

IFUNC = 0 uses function (2) defined in the FUNC command.

The default is function (2) i.e. IFUNC = 0.

As always the commands TITLE , END ,MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY_HARD, MODEL, SHOW and X are available.

9.4.10 Yzarc Module

Usually one proceeds directly to tangent refinement from convergence mapping, but Mithril94 offers two additional modules which can be run between CONVERGE and TANGENT. The first of these is YZARC. It uses sets of random phases as a starting point, and refines them *via* least-squares or steepest-descents. Users unfamiliar with these concepts should first read two papers - Baggio, Woolfson, Declercq and Germain (1978) and Declercq, Germain and Woolfson (1979). Neither of these papers discuss the steepest descents method. It is sufficient to state here that standard steepest-descents algebra is used with appropriate weighting schemes. Steepest descents and least squares will, in general, produce different phase sets even when all other conditions are the same.

Usually only a subset of reflections is phased, and these phases are passed to tangent refinement. Normally some 50-100 different sets are processed in this way. Some of the facilities offered in Mithril94 differ from the standard YZARC procedure:

(1) All the relationships collected during convergence are used. The inclusion of quartets alters the refinement; sometimes it improves the radius of convergence, but sometimes it does not.

(2) The problem of when to stop refinement has always been difficult in YZARC. An alternative method based on NQEST is offered as an option. After an initial round of n cycles, NQEST is calculated. If at this point the NQEST figure of merit is greater than a specified cut-off (e) refinement is terminated, otherwise it continues. It checks this figure of merit after each

cycle, and refinement continues whilst it continues to fall until it hits a minimum at which point it stops. Experience dictates that suitable values for n and e above are 7 and 0.0 respectively, but they are user options. This method has the effect of reducing the number of refinement cycles - usually by a factor of six - with a commensurate fall in the computer time required. As with the traditional techniques of stopping YZARC refinement, it is not always successful.

(3) YZARC normally phases the bottom 100 or so reflections from the convergence map. An option is provided whereby the convergence map is bypassed except for origin definition, and the top reflections ordered on E-magnitude alone are phased. This usually creates a singular matrix if least-squares is employed so that steepest descents is normally used. This can be useful in circumstances where the convergence map selects a subset of highly linked reflections in which certain parity groups are not represented. Bypassing the map can phase a more representative set of reflections.

(4) The final figures of merit are augmented by the inclusion of NQEST.

Yzarc is always user called and never entered automatically. The user interface is:

The following commands may be entered in any order:

YZARc NO. OF PHASE SETS, MAX. NO. OF CYCLES, MAX. MEAN SHIFT

This is the YZARC calling command. It has three parameters:

- The number of phase sets to pass to TANGENT. The default is calculated as:
  No. of sets = 60 + 50 * Idif + 50 * Ivdif

where:

Idif = 0 / 1 for Standard / Hard structure
Ivdif = 0 / 1 for Standard / Very hard structure.

- The maximum number of cycles of refinement for any phase set. The default is 70.

- The maximum value of the mean phase shift in degrees. If the mean phase shift falls below this value in any YZARC refinement cycle, then refinement ceases on this phase set and moves on to the next. The default is 4°.

Be careful about tampering with the latter two options - in general, they are set at their most useful values.

NREF NO. OF REFLECTIONS TO BE PHASED

This specifies the number of reflections to be phased by the YZARC procedure. The default is calculated *via*:

No. of reflections to phase = 100 + 50 * Idif + 50 * Ivdif

where Idif and Ivdif have already been defined.

L.S.

The module normally uses steepest-descents to refine the phase sets. This command invokes the standard least-squares procedure with its associated weighting scheme. Note the periods in the L.S. command.

STARt IX, IY

These are two odd integers which act as the seed for the random number generator which is used to give the reflections random phases. The default values are 190907 and 568835.

NQESt CYCLE NUMBER, MAXIMUM NQEST

This command causes NQEST to be used as the criterion for stopping refinement of a given phase set. The cycle number specifies the cycle at which the NQEST test is first applied. All previous cycles proceed untested. The default is 7. The second parameter defines the maximum allowed value of NQEST after these cycles. The default is zero. If a solution has a value greater than this, then refinement terminates, otherwise it continues as long as NQEST is falling. If there is an increase in NQEST, refinement stops. Refinement is still subject to any constraints input on the MAGEX calling command.

TOP_

Usually YZARC phases the bottom reflections in the convergence map; this command causes the map to be bypassed, apart from origin and enantiomorph definition, and the top reflections, based on E-magnitude alone, are phased. The number to be phased is that defined by the NREF command. This instruction should usually be used only in conjunction with steepest descents, since it is likely to give a singular matrix in a least-squares environment unless all the reflections are linked to each other by the available triplets. If least-squares has been specified and a singular matrix is found, the module will automatically change to steepest descents.

RANDom IX, IY

This instruction defines two odd integers which act as a seed for the random number generator. Unlike the START command, however, only a single phase set is produced, the module then expects another RANDOM command for another solution to be produce, and so on. In the interactive mode both these options are available, but the user can merely press a carriage return to terminate input of random numbers. This command is useful when re-running refinements in a situation where only certain solutions are to be investigated.

As usual the commands TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY_HARD, MODEL, SHOW and X are available as required.

9.4.11 Symbolic Addition Module

Mithril94 has a symbolic addition module, called SYMB derived from LSAM (Germain & Woolfson, 1968). Many modifications have been made including the use of quartets and new more powerful figures of merit.

SYMB accepts all triplet and quartet phase relationships (up to a maximum of 2000) which have probabilities greater than a minimum set by the user. These probabilities are calculated using the Cochran & Woolfson (1955) formula

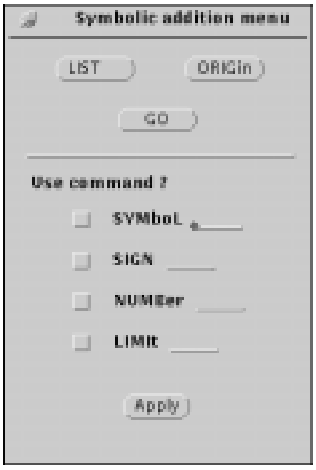
$$P = 1/2 + 1/2 \tan(\kappa/2)$$

where κ was as defined earlier.

Using program assigned symbols and origin reflections, signs are developed by symbolic addition (Karle & Karle 1966). Starting set signs are assumed true (i.e. Prob. = 1.0). Sign indications are then computed for all other reflections and the new reflection that is indicated with the highest probability is accepted. This is then used in another round of sign indication. In this iterative fashion all signs with a probability greater than a set minimum are found. When sign determination is done SYMB checks the consistency of the signs by rejecting all those with probabilities less than a specified cut-off and re-determining the signs that remain. This continues until there is no change in sign from one re-determination to the next. Unlike many symbolic addition programmes, SYMB allows for the failure of individual triplets.

Note that SYMB can only be used for centro-symmetric structures, and is most useful in case where tangent refinement causes phases to become over consistent, particularly heavy atom structures.

The SYMB module is only ever called by the user. The interface is:



The following commands may be entered in any order:

SYMB

Calls and initialises the SYMBOL module.

SYMboL NO. OF SYMBOLS

Allows the user to choose how many symbols, up to a maximum of 11, are to be used in the symbolic addition procedure. The default is four. Note that this command is SYML and not SYMB as may be expected.

SIGN MINIMUM PROBABILITY

This sets the minimum probability acceptable for a sign determination. The default is 0.95.

NUMBer NUMBER OF EQUATIONS

This determines the number of symbol equations to be considered in the solution for the symbols. It may be set at a number higher than the number of relationships in order to ensure that all relationships present are considered. The default is 10.

LIMIt PROB

All signs predicted with a probability greater than PROB will be included in the final output. The default is 0.8.

ORIGin Determine the origin by the SYMB module rather than accepting the origin determined by the CONVERGE module.

LIST

Output a table of signs developed by symbolic addition, and relationships between the symbols to be displayed in the output file.

As usual, the commands TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY_HARD, MODEL, SHOW and X are available as appropriate.

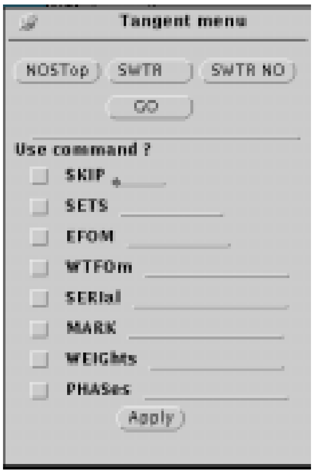
9.4.12 Tangent Module

The tangent formula carries out expansion and refinement of phase angles. The following options are provided:

- All relationships accepted by CONVERGE are used actively to generate new phase angles.
- Three early figures of merit - the two MULTAN80 options of ψ_0 combined with R_{Karle} and ψ_0 alone to which has been added NQEST. Any combination of these three can be specified. Early figures of merit are useful when it is proposed to generate large numbers of phase sets. There is a link here with SHEXTL which relies very heavily on NQEST to promote efficiency. These figures of merit can also be used to filter out unacceptable phase sets from YZARC and MAGEX after only a few cycles of tangent refinement and expansion.
- Two weighting schemes as in MULTAN80 - the traditional MULTAN scheme and Hull - Irwin statistical weights (Hull & Irwin, 1978). The latter is useful in cases of pseudo-symmetry, over consistent phase sets (as demonstrated by very low R_{Karle} values, and very high ABSFOM's), symmorphic space groups, and heavy-atom cases.
- In its interactive mode, the program will attempt to identify any solution with exceptional figures of merit. It is possible to stop at this point, compute an E-map then continue with tangent refinement if the map seems unpromising.
- Five figures of merit: ABSFOM, ψ_0, R_{Karle},,NQEST and LOGLIK combined together with user controlled weights give a single figure of merit and this is used to rank the solutions. The R_{Karle} is based on alpha values. LOGLIK is based on the agreement between observed and calculated magnitudes of unphased reflections has been implemented. It is based on the prediction of an E-magnitude from triplets, and the FOM keyword in the triplet call will allow calculation of this figure of merit.
- Both tangent and Hull-Irwin weighted tangent refinement are now performed without disc i/o to give a 50% reduction in the time taken for the refinement of one phase set, at the cost of increased memory requirements.

The TANGENT module performs the dual tasks of phase expansion and refinement. It must always be run except when Fourier recycling is being used, and will be entered and run with defaults if there is no explicit user call.

The interface is as follows:



The following commands can be entered in any order:

TANGent

Calls and initialises the TANGENT module.

SWTR [NO]

This command causes the Hull - Irwin weighting scheme to be used rather than the traditional MULTAN80 weights. This is a default when Karle recycling is used, otherwise the standard procedure is used. If for some reason the Hull - Irwin scheme is to be used and you wish to revert to the standard weighting method, then use the command:

```
SWTR NO
```

where “NO” is a keyword.

SKIP N

This command is usually used for restarts. It causes the first N phase sets to be skipped before starting tangent refinement. These N phase sets must already exist on the file on channel 11. When re-running a job which ran out of time, N should be the number of the last set output on the printer, but make allowances for buffering by reducing this. The exact extent of this reduction depends on your installation.

SKIP 0 has a special function. When either the YZARC or MAGEX modules have been run in a job, the TANGENT module automatically takes its input phases from those produced by them. Phase permutation is not carried out. The SKIP 0 command causes these YZARC / MAGEX calculations to be ignored, and routine tangent refinement to be carried out instead.

SETS N1, N2, N3, N4 .... etc.

With this option only the phase sets with numbers N1, N2 etc. are investigated by the TANGENT module. This is useful for re-runs. Unlike the SKIP command, the other phase sets do not need to be on file 11.

WTFOm W1, W2, W3, W4

This command defines the relative weights of the four figures of merit used by the TANGENT module when calculating a combined figure of merit (CFOM). The defaults are as follows:

TABLE 7.

Figure of Merit	Weight	Default	Default in symmorphic cases
ABS FOM	W1	1.0	0.6
ψ-ZERO	W2	1.0	1.2
RESID	W3	1.0	0.6
NQEST	W4	1.0	1.3

If there are no appropriate relationships for a particular figure of merit then a weight of zero is assigned. The relative weights are normalised such that the maximum CFOM value is equal to the total number of figures of merit contributing to it. The LOGLIK figure of merit, if calculated, is not used in the combined figure of merit.

NOSTop

If the TANGENT module finds a solution with figures of merit that satisfy the conditions:

- Resid (R_{Karle}) less than 20.0
- ψ_0 less than 1.25
- NQEST less than -0.15.
- The figures of merit above are within 5% of the best so far

Then the module assumes that this is the correct solution and exits. Users in the interactive mode will be questioned first if they wish to accept this solution The command NOSTOP switches off these tests. So do the HARD and VERY_HARD options. NOSTOP is often worthwhile as the correct solution can often be missed under the early stop algorithm.

SERIal SERIAL NUMBERS N1, N2, N3, N4 .... etc.

The commands SERIAL, MARK, WEIGHTS and PHASES are used together to input a set of known phases into tangent refinement. The SERIAL command gives the serial numbers (in any order) of the reflections whose phases are to be input.

MARK M1, M2, M3, M4 .... etc.

These are the markers associated with the reflection numbers entered on the SERIAL command. A value of 1 means that the input phase can be refined immediately, whereas a value of -1 signifies that the phase is to be held constant until the last two cycles. The default values are -1, so the MARK command is only needed if this is unsatisfactory.

WEIGhts W1, W2, W3, W4 .... etc.

The weights associated with each input phase. The default values are 0.9, so this command is only needed in cases where this is inappropriate.

PHASes PHASE 1, PHASE 2, PHASE 3, PHASE 4 .... etc.

Inputs the known phase angles in degrees. Each angle must appear in the same order as the serial numbers on the SERIAL command. You can input as many sets as you require; each set requires its own PHASES command and is refined immediately after it has been input. The first input phase set *via* the PHASES command terminates current input to the TANGENT module. In the interactive mode the user will be prompted for more phase sets as required-a simple carriage return or a call to another module is used to signify the end of input phase sets. As usual, the commands TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY_HARD, MODEL, SHOW and X are available as appropriate.

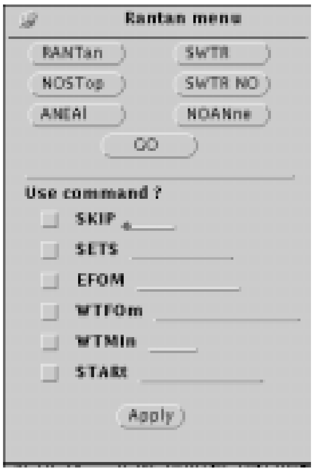
9.4.13 Rantan Module

The RANTAN module offers all the facilities provided by the tangent refinement module discussed above. It differs from the latter in that it does not use phase permutation, but assigns all unknown phases random values. There is thus no phase expansion, but only refinement. The phases are refined to convergence in the usual way. The theory of this technique is discussed by Yao Jia-Xing (1981). The LOGLIK figure of merit is available and run times are decreased by 50%.

- In addition, the use of simulated annealing to refine phases in both the regular tangent and Hull-Irwin weighted tangent refinement routines is provided. Only random phases can be used.

In one respect the RANTAN module is similar to TANGENT in that it refines phase angles. The difference is that RANTAN uses random phases for all the unknown phases and refines them using the tangent formula, rather than using the phase permutation techniques of TANGENT.

RANTAN is never called automatically.



The following commands, apart from the first, may be entered in any order:

RANTan NO. OF PHASE SETS TO GENERATE.

Calls and initialises the RANTAN module. The number of phase sets to generate is entered. The default is:

No. of phase sets = 100 + 50 * Idif + 100 * Ivdif

where:

Idif = 0 / 1 for Standard / Hard structure.

Ivdif = 0 / 1 for Standard / Very hard structure.

SWTR [NO]

This command causes the Hull - Irwin weighting scheme to be used rather than the traditional MULTAN80 weights. This is a default when Karle recycling is used, otherwise the standard procedure is used. If for some reason the Hull - Irwin scheme is to be used and you wish to revert to the standard weighting method, then use the command:

SWTR NO where “NO” is a keyword.

SKIP N

This command is usually used for restarts. It causes the first N phase sets to be skipped before starting tangent refinement. These N phase sets must already exist on the file on channel 11. When re-running a job which ran out of time, N should be the set number of the last set output on the printer, but make allowances for buffering by reducing this by 3 or 4 depending on your installation, and operating system.

WTMIn WEIGHTS OF UNKNOWN REFLECTIONS.

The random phases have weights of 0.25 assigned to them before refinement begins, but if a different value is wanted, then this command can be used. Some experimentation with these weights can be useful in difficult cases.

WTFOM W1, W2, W3, W4

This command defines the relative weights of the five figures of merit used by the RANTAN module when calculating a combined figure of merit (CFOM). The defaults are as follows:

TABLE 8.

Figure of Merit	Weight	Default	Default in symmorphic cases
ABS FOM	W1	1.0	0.6
ψ-ZERO	W2	1.0	1.2
RESID	W3	1.0	0.6
NQEST	W4	1.0	1.3

If there are no appropriate relationships for a particular figure of merit then a weight of zero is assigned. The relative weights are normalised such that the maximum CFOM value is equal to the total number of figures of merit contributing to it. The LOGLIK figure of merit, if calculated, is not used in the combined figure of merit.

NOSTop

If the RANTAN module finds a solution with figures of merit that satisfy the following conditions:

- Resid (R_{Karle}) less than 20.0
- ψ_0 less than 1.25
- NQEST less than -0.15
- The figures of merit above are within 5% of the best so far.

(assuming that these figures of merit are available), then the module assumes that this is the correct solution and exits. Users in the interactive modes will be questioned first if they wish to accept this solution. The command NOSTOP switches off these tests. So do the HARD and VERY_HARD options.

SETS N1, N2, N3, N4 .... etc.

With this option only the phase sets with numbers N1, N2 etc. are investigated *via* the RANTAN module. This is useful for re-runs. Unlike the SKIP command, the other phase sets do not need to be on file 11.

STARt IX, IY

Two odd integers used to seed the random number generator. The default values are 1 and 1 (cf. YZARC).

As usual, the commands TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY_HARD, MODEL, SHOW and X are available as appropriate.

ANEA

Requests simulated annealing

SNOAN

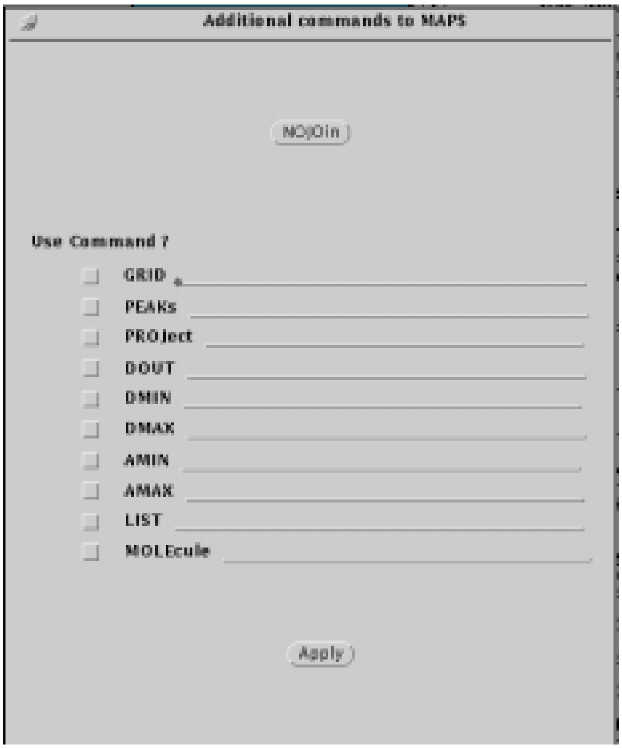
Switches off simulated annealing.

9.4.14 Maps Module

The final step of a direct methods analysis is the calculation of one or more E-maps. It is with E-map interpretation that the high level of user interaction which can be provided becomes important, particularly when there are many maps to search. It is possible to generate and search all the appropriate maps quickly, and reject unsuitable solutions on the basis of peak heights, peak positions or fragmentation patterns before attempting a full interpretation. Simple graphics facilities are also provided.

The MAPS module calculates Fourier maps, picks the peaks and attempts to identify chemically reasonable fragments (Main & Hull, 1978). Usually an E-map is computed, but where Fourier recycling has been requested a Sim-weighted electron density map is calculated and interpreted. Depressing the MAPS button the following window appears:

The sub menu is:



The following commands may appear in any order:

MAPS N1 [, N2] ... etc. or -N or [ALl]

Calls and initialises the MAPS module. There are three possibilities for parameters associated with this command:

1. One or more positive integers. Each of the specified solutions from tangent refinement will be used to compute an E-map. E.g.

    MAPS 2, 4, 1, 19

will compute the maps for solutions 2, 4, 1 and 19.

2. A single negative integer (-N) will compute maps for the +N best solutions as ranked on combined figure of merit.

3. The keyword “ALL” causes all the E-maps to be calculated. This is usually used in the interactive mode with LEVEL = 3 as a way of examining maps in a quick and efficient way. Under LEVEL 3, it is possible to exit from the MAPS module once a suitable map has been found. It is not necessary to actually investigate all the maps.

You may only choose one of these methods on a MAPS command. The default is to calculate the single best E-map. In the situation where Fourier recycling is being carried out, all parameters on the MAPS command are ignored.

GRID RESOLUTION, X LIMIT, Y LIMIT, Z LIMIT

This command defines the resolution and extent of the Fourier computation.

RESOLUTION is the resolution required in Å. The default is 0.333Å.

X LIMIT is the limit of the Fourier calculation along the x-axis. The calculation always starts at x = 0.0. The default is 1.0.

Y LIMIT operates like the X limit above but along y.

Z LIMIT operates like the X limit above but along z.

PEAKs MAX NO. OF PEAKS, NO. OF HEAVY ATOMS PRESENT, MIN. NO. OF PEAKS FOR PLOTTING

This command specifies how many peaks are to be picked from the Fourier map, and used in the interpretation routines. The default is calculated as follows:

Max. no. of peaks = (11*Nasu+13)/9 + 10*Idif + 10*Ivdif + 10*Iapx

where:

- Nasu = No. of atoms in asymmetric unit.
- Idif = 0 / 1 for Standard / Hard structure.
- Ivdif = 0 / 1 for Standard / Very Hard structure.
- Iapx = 0 / 1 for Yes / No CONTENTS command issued in the NORMAL module.

This default is increased when two clusters are within 2.8Å. of each other.

The number of heavy atoms present in the asymmetric unit (N) is used as follows: when carrying out an interpretation of a cluster, the MAPS module will assume that the N highest peaks correspond to these heavy atoms. They will not be included in the interpretation, and all peaks within the maximum bond length will be marked as spurious. This can be useful in heavy-atom cases to prevent diffraction ripples around the heavy atom(s) being treated as real atoms, but it does not always work well, so be wary of E-map interpretations that incorporate this option. The default is zero.

The final parameter defines the minimum size of cluster, defined by the number of atoms it contains, which will be plotted on the printer (and the terminal if graphics options are used). The default is four.

NOJOin

Stops all connectivity and peak interpretation calculations from being performed.

PROJect NO. OF PROJECTIONS, MAX. NO. OF INTERPRETATIONS

The first parameter here defines the number of orthogonal projections of each cluster to be plotted. The default is always to plot the least-squares projection first, and then to plot the projection orthogonal to the least- and most-squares planes if the cluster is spherical or cylindrical in shape. The maximum number of projections is three. The third projection is on to the most-squares plane. The second parameter specifies the maximum number of possible peak interpretations to be output for a cluster. The default is three.

DOUT MAXIMUM BOND LENGTH TO OUTPUT

All interpeak distances less than this parameter are tabulated. The default is 2.4Å.

DMIN MINIMUM ACCEPTABLE BOND LENGTH

This is the minimum allowed bond length for peak interpretation. The default is 1.1Å.

DMAX MAXIMUM ACCEPTABLE BOND LENGTH

This is the maximum allowed bond length for peak interpretation. The default is 1.95Å.

AMIN MINIMUM ACCEPTABLE BOND ANGLE

This is the minimum allowed bond angle for peak interpretation. The default is 85 degrees. You may need to lower this when three-membered rings are involved, although it can result in a rather more messy collection of peaks.

AMAX MAXIMUM ACCEPTABLE BOND ANGLE

This is the maximum allowed bond angle for peak interpretation. The default is 145 degrees This, too, will not always be a suitable figure.

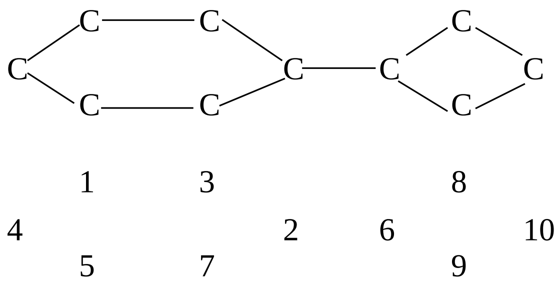
LIST N

This command causes the full Fourier map to be output on the printer. All the peaks greater than or equal to N are underlined with asterisks.

MOLEcule CONNECTIVITY

This instruction inputs a set of molecular connectivities which are used for comparison with the molecular fragments found in the map itself. They are not used by the MAPS module in its search for a fragment, and are therefore optional. The connectivities are input as follows:

Draw the molecule or fragment whose chemical structure and connectivity is known. The stereochemistry is not important. Then number the atoms in any order e.g.



Finally, specify the connectivities as follows:

MOLE 1 3 4 / 2 3 6 7 / 3 1 2 / 4 1 5 / 5 4 7 / 6 2 8/ =
7 2 5 / 8 6 10 / 9 6 10 / 10 8 9 ///

where atom 1 is joined to atoms 3 and 4, atom 2 to atoms 3, 6 and 7 etc.

Redundant information may be removed. For example the above sequence can be simplified to:

MOLE 1 3 4 / 2 3 6 7 / 4 5 / 5 7 / 6 8 9 / 8 10 / 9 10 /// or

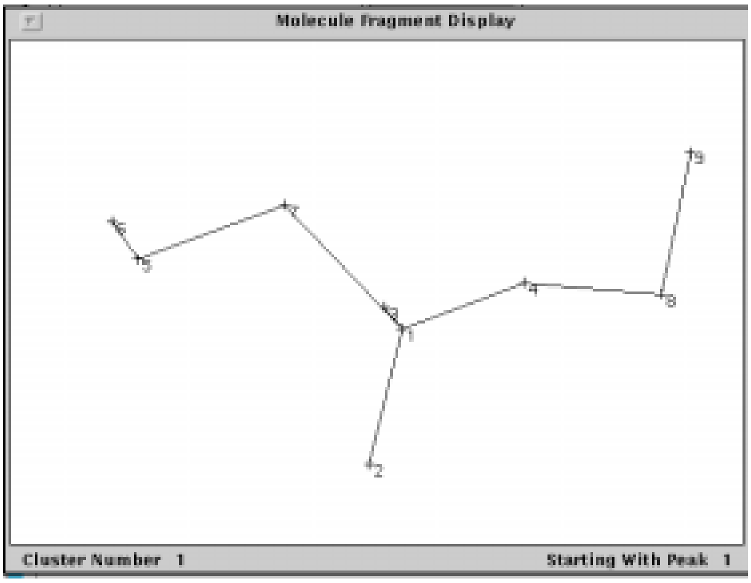
MOLE 1 3 4 / 2 3 6 7 / 5 4 7 / 8 6 10 / 9 6 10 ///

In both these cases each bond is specified only once.

The double slash (//) signifies the end of a fragment. Another can follow, although it must be part of the original MOLE command (only one such command is permitted). The maximum number of atoms which can be entered is 200. Up to five fragments or molecules can be entered each separated by the double slash (//). The end of the command is specified by a triple slash (///). This must appear.

The commands TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY_HARD, MODEL, SHOW and X are available as required.

When running a simple picture of each fragment appears:

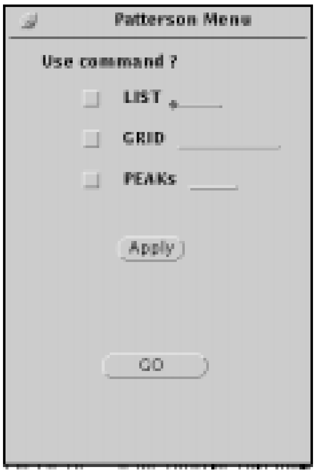


Click on the box to dismiss it.

9.4.15 Patterson Module

Provision is made for the calculation of an E^2 -1 vector map. User control of the sharpening procedure is possible. Examination of such maps can be useful in difficult cases (Nixon, 1978).

This module computes an E^2 -1 vector map followed by a peak search, but without peak interpretation. The user interface is:



The following commands are available:

PATTerson

Calls and initialises the Patterson map calculations.

LIST N

This command causes the map to be output on the printer. All the peaks greater than or equal to N are underlined.

GRID RESOLUTION, X LIMIT, Y LIMIT, Z LIMIT

This command defines the resolution and extent of the Patterson computation.

RESOLUTION is the resolution required in Angstroms. The default is 0.333A.

X LIMIT is the limit of the calculation along the x-axis. The calculation always starts at x = 0.0. The default is 1.0.

Y LIMIT operates like the X limit above but along y.

Z LIMIT operates like the X limit above but along z.

PEAKs MAX. NO. OF PEAKS TO LIST

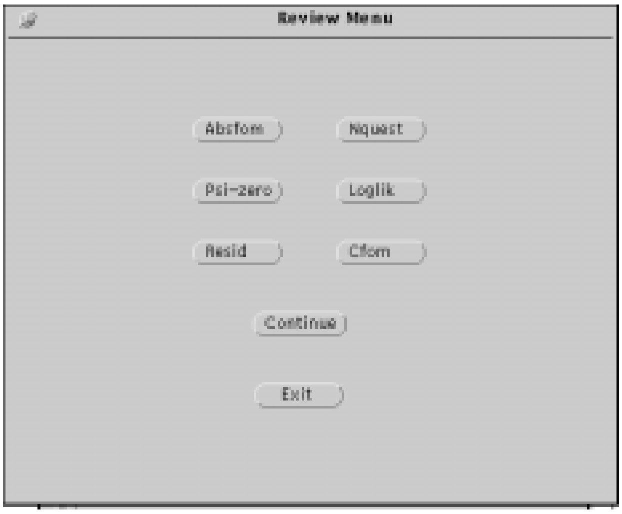
This specifies the maximum number of peaks to output. The default is 30.

As usual, the commands TITLE, END, MENU, LEVEL, NOPRINT, PRINT, DEFAULT, HARD, VERY_HARD, MODEL, SHOW and X are available as appropriate.

9.4.16 Review Module

The interactive module, Review, has been added to MITHRIL as a direct consequence of the experience gained in solving difficult structures. For some structures it has been found necessary to generate many phase sets (up to 2000 have been produced). To facilitate in the selection of a correct phase set Mithril94 will now sort the sets and rank them according to one of six figures of merit.

This module sorts and outputs the phase sets for each of the individual figures of merit as well as CFOM. The output will also give an indication as to whether a map has been investigated earlier or has an equivalent set. The menu is:



The following commands are available:

Absfom	Sort the sets on the Absolute figure of merit
Psi-zero	Sort the sets on the ψ_0 figure of merit
Resid	Sort the sets on the R_{Karle} figure of merit
Nqest	Sort the sets on the Nqest figure of merit
Loglik	Sort the sets on the Loglik figure of merit
Cfom	Sort the sets based on the Combined figure
Exit	Exits from the module and returns to the master Mithril94 menu.

Note: The above commands require only the first letter each command -any additional letters are ignored. This is the only module in which the general commands are not available.

9.5 Recycling Procedures

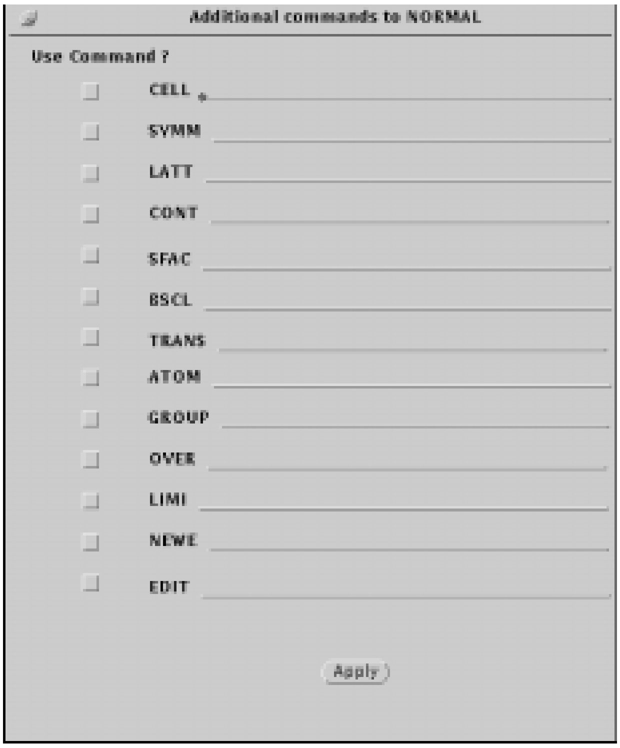
If a direct methods analysis provides only a partial fragment or fragments, some form of recycling is required to complete the structure. It is usually assumed initially that the fragment(s) are correctly oriented and positioned in the unit cell. Under these circumstances, four types of recycling are provided:

- Weighted Fourier calculations using the RECYCLE module. The fragment is recycled using a Sim weighted Fourier (Sim, 1959, 1960).
- Karle recycling using the RECYCLE module, the tangent formula is used to extend and refine phases (Karle, 1968).

- Karle recycling where the unknown phases are given random values using the RECYCLE module (Yao Jia-Xing, 1983).
- Using type 4 groups in NORMAL (Main, 1976).

We will deal with each of these in turn, and then the situation where the fragment(s) are in fact misplaced in the cell. The interface is:

The other commands appear on a sub-menu:



9.5.1 Weighted Fourier recycling

This is the recommended recycling procedure and the best way to proceed is as follows:

(a) Set up an ASCII file containing the relevant CELL, LATTICE, SYMM, CONTENTS, SFAC, LIMITS, EDIT, TRANS and NOCHECK commands as required. These are all commands to NORMAL. Add to this set the GROUP command without any parameters followed immediately by a set of atomic coordinates for the fragment which is to be recycled using the ATOM command. Store this as a model file.

(b) Now run RECYCLE with this file treated as a MODEL file. The RECYCLE command must contain the keyword “FOURIER”, all other parameters are ignored. RECYCLE calls module NORMAL.

(c) Either enter the command DEFAULT or call the MAPS module explicitly. You will get a weighted Fourier map. The whole sequence looks like this:

```
RECY FO         ! Requests Fourier recycling.
MODEL           ! Scans MODEL file-else insert CELL etc.
DEFAULT         ! Runs MAPS under default.
```

Note that for the MODEL command to work the MODEL file must be attached to the job at run time.

9.5.2 Karle recycling without random phases

The best way to proceed is as follows:

(a) Set up a model file containing the relevant CELL, LATTICE, SYMM, CONTENTS, SFAC, LIMITS, EDIT, TRANS and NOCHECK commands as required. These are all commands to NORMAL Add to this set the GROUP command without any parameters followed immediately by a set of atomic coordinates using the ATOM command. Store this as a model file.

(b) Now run RECYCLE with this file treated as a MODEL file. The RECYCLE command must contain the keyword “KARLE”, but unlike Fourier recycling, the parameters IK, NB, ISC and MAXDUP are available as on the NORMAL instruction.

(c) The program will now need to run the TRIPLETS, QUARTETS (if relevant), TANGENT and MAPS modules. Either call these explicitly or use the DEFAULT instruction in the usual way. You may control the refinement procedure in the same way as an a-priori analysis. You will finally obtain an E-map. E.g. for the simplest possible recycling enter the commands:

```
RECYCLE KARLE
MODEL           ! Scans MODEL file
DEFAULT         ! Runs necessary modules under default.
```

9.5.3 Karle recycling with random phases

This is run in exactly the same way as Karle recycling, but an explicit call to the RANTAN module is required in the correct place. For example, to run the recycling job outlined above but with random phases for the unknowns, the following sequence of commands could be used:

```
NORMAL KARLE
MODEL            ! Scans MODEL file
RANTAN 100       ! Asks for 100 phase sets to be generated
DEFAULT          ! MAPS will follow under default
```

9.5.4 Using type 4 groups in normal

The type 4 groups in NORMAL are those which are correctly oriented and placed in the cell. To recycle in this way it is sufficient to run a typical a-priori phasing calculation but include a

```
GROUP 4
```

instruction followed by a set of ATOM commands in the input to NORMAL. Then proceed in the usual way. The use of the MODEL file as an auxiliary input is much recommended since it makes the input of the atomic coordinates less prone to error.

9.5.5 Incorrectly placed groups

If none of these recycling procedures are effective, then it is reasonable to assume that the fragment(s) are incorrectly placed. If they make chemical sense, however, they are probably correctly oriented in the unit cell. Under these circumstances recycling using GROUPS of type 3 may well prove successful. If this does not work either, then relaxation to GROUPS of type 2 should be tried.

9.6 What to do when Mithril fails

This section outlines some of the options available to you if Mithril94 has failed to solve a structure. The possibilities are not quoted in any order of preference, except perhaps for the first set involving normalisation.

9.6.1 Normal

(1)Check the data. If there are many duplicates or systematic absences list them all by setting the MAXDUP parameter on the NORMAL command to 100000, or any large integer. Then investigate this list with suspicion-are you sure about the space group? Have you collected at least the unique data? How good is the resolution of the data? If there are missing reflections in your data re-insert them using the MISSING command.

(2) Read the output from NORMAL very carefully. Are the statistics sensible? Do the large E-magnitudes form a readily identifiable subset where some parity groups are missing? If some of the E-magnitudes are very large, it is often useful to use a LIMIT or EDIT command to reduce or remove them. The variation of E^2-1.0 as a function of Bragg angle is a good guide to the applicability of the calculated temperature factor. If there is a fall- off in this average as the angle increases, use the BSCL command to input a larger value of B than that calculated by the normalisation module. If some or all of the stereochemistry is known, then input this as a type 2 group. This is especially important with molecules containing planar fused rings.

(3) If you have generated the best set of E-magnitudes that you can, and the structure still will not solve, then a systematic distortion of the E's can be successful. Changes in the E-magnitudes cause changes in the invariants, and these in turn give rise to drastic modifications of the convergence map, and the subsequent phasing path. There are several ways of distorting the E's:

- Modify the unit cell contents using the CONTENTS command. Doubling the contents is the best starting point.
- Use artificially raised or lowered temperature factors *via* the BSCL instruction. Often only a small change in B gives rise to drastic changes in the E-magnitudes. This is especially recommended for situations where resolution is less than the Cu sphere.
- Insert a molecular fragment *via* the GROUP command that does not correspond to any group expected in the molecule.

9.6.2 Invariants - triplets, and quartets

There are two problems which can arise here. One concerns a paucity of suitable relationships which can be common in situations of low symmetry. The other concerns the accuracy of the invariants themselves.

(1) If there is a paucity of triplets:

- Use quartets as well. Try just the negative ones first, then add the positives if this is unsuccessful.
- Increase the number of reflections on the TRIPLETS command.
- Reduce the minimum κ-value from its default of 0.6 by using the KMIN command in CONVERGE. This will, however, introduce a number of very unreliable relationships.

(2) If there is a paucity of quartets:

- Increase the number of reflections on the QUARTETS command. This is usually the best way.
- Invoke the third neighbourhood *via* the NEIGH instruction.
- Ask for positive quartets as well. This can also be useful when triplets are scarce.
- Allow more missing second (and third) neighbours.

(3) If the problem is thought to be the reliability of the invariants then use the cut-off based on sin θ^2/λ^2, and/or the weighting scheme also based on sin θ^2/λ^2. These are less time consuming and have proven more useful than MDKS and L.E. when dealing with unreliable invariants.

(4) Even if the space group is not symmorphic, quartets often have a very beneficial effect on a direct methods analysis, and can be recommended as an option to try early in the list of weapons in the armoury. Only the negative quartets should be tried first, since they are independent of the triplets. Even a few four-phase invariants can drastically alter the phasing path.

(5) The MDKS and L.E. options coupled with convergence map weighting also has a drastic effect. The two options are different so if one is unsuccessful, it is worth trying the other. However, they are both very time consuming.

(6) Invariant generation is obviously of critical importance. Mithril94 allows a good deal of user control over the process, so use the options provided.

9.6.3 Convergence mapping

Convergence mapping lies at the very heart of the multisolution approach to direct methods employed by Mithril94. It is essential to examine the convergence map carefully in cases of difficulty.

(1) Make sure that the starting set is a good one with all the starting set reflections used early in the phase determination. This can be checked by examining the bottom of the convergence map. If a starting set reflection is not used at all early on, then a better starting point can often be obtained by including at least one other reflection whose phase depends on that of the late starter. Introducing quartets may also have a similar effect.

(2) If there are gaps near the bottom of the convergence map (i.e. reflections with a zero estimated alpha and no invariants contributing), or the map is very ‘thin’ with many phases determined by only one or two relationships, then the phasing often fails. This can be remedied by increasing the size of the starting set or introducing higher invariants, particularly quartets.

(3) Be wary of the Σ_1 determined phases. If they play a major role in the early stages of phasing, it is often worthwhile excluding them *via* the SIGMA NONE command. If MDKS or L.E. has been run, examine the Σ_1 triplet analysis. The triplets should have estimated cosines close to unity.

(4) If it still proves impossible to obtain a suitable convergence map without a massive amount of computer time, then several options are possible:

- Run MAGEX. Apart from origin and enantiomorph definition, it largely ignores the convergence map.
- Run YZARC. Try both least-squares and steepest descents - they give different results. The TOP_ instruction by-passes the convergence map.

- Run RANTAN instead of regular tangent refinement.

(5) Check to see if all the reflections at the bottom of the map have something in common e.g. they all have h even or k+l divisible by 3. If so, then make sure that the average value of E^2-1 is unity for such reflections. Renormalisation may be necessary. It may be possible to use the editing facilities of NORMAL to juggle these magnitudes. Try introducing new reflections into the starting set which do not belong to these groups.

(6) Altering the origin and enantiomorph is often unsuccessful, particularly if only small changes are made. The same relationships are still used in the early stages of phasing, but in a different form. For example, the triplet:

$$\phi_1 - \phi_2 + \phi_3$$

may appear in one map generating ϕ_1 from ϕ_2 and ϕ_3. If the origin is partially re-defined by the user, this triplet may well appear again in a critical place but this time generating ϕ_2 from ϕ_1 and ϕ_3. If this triplet is erroneous it will be erroneous however it is used. This said, juggling with the starting set can be successful on some occasions, and is worth a try.

(7) Hand applied symbolic addition, even in a limited form, can give rise to possible relationships between phases, and these can be introduced into the convergence map by the RELATIONSHIP command. The relationships linking two phases (the pair relationships) are the most valuable. The inclusion of only one or two with high associated κ-value will drastically alter a convergence map. There is the added bonus that symbolic addition can give valuable insights into the causes of phasing difficulties (Karle & Karle 1966). Do not use the convergence map for symbolic addition; get a list of triplets and work with this. The convergence procedure has too many weak relationships early in the phasing path.

9.6.4 Phase expansion and refinement

(1) The only way to monitor phase expansion and refinement is by inspecting the final figures of merit, so it is important to examine these closely in difficult cases. In particular do not just inspect the final CFOM's, but look also at the individual contributors:

- ABSFOM is the least reliable. If the ABSFOM values all tend to be large then the refined phases are over-consistent. Using Hull-Irwin weights will often give better results.
- NQEST and ψ_0 are the most reliable figures of merit provided that the weak reflections have been accurately measured. Do not expect NQEST to be very negative, particularly if quartets are being used actively in phase refinement. In these circumstances values around -0.1 are often satisfactory.
- Heavy atom cases often give extreme figures of merit. The correct solution may well be present even if the figures of merit seem unrealistic.
- Calculate the LOGLIK figure of merit. You do this by using the FOM keyword in the call to the TRIPLET module. It must never be examined on its own but always in conjunction with ψ_0.
- If all the phase sets have similar figures of merit too few invariants may be present.

(2) In case of pseudo-symmetry, the presence of heavy atoms or substantial planar moieties in the structure, use the Hull-Irwin scheme.

(3) Do not forget that there are two weighting schemes - if one does not work, the other may, even using the same starting set and convergence map.

(4) Be careful of the early stop option. It is often found that the set selected by the TANGENT module during refinement as being an obvious solution is not the correct one. The NOSTOP command will prevent this happening.

(5) If you are using early figures of merit, and most sets are getting rejected, it is possible that the correct set is also being discarded. It is an easy matter to turn off these figures of merit. Do not use the first EFOM for planar ring structures, and set the cut-off for the third EFOM to at least 1.7.

(6) Do not confine your attentions to the one or two phase sets with the highest CFOM’s. It may be necessary to examine maps with quite low associated CFOM’s.

(7) The CFOM’s are dependent on the relative weights of the individual figures of merit. Adjusting the weights to reflect your own intuition concerning the contributing figures of merit will often result in a drastic re-ranking of the solutions.

(8) In a case with over-consistent phases and a centrosymmetric structure it is strongly advised to use the SYMB module in preference to the TANGENT module.

9.6.5 E-maps

Look at the resulting E-maps carefully. Remember that the interpretation assumes that you have well-resolved peaks, and this may not be the case. The routines which perform the chemical interpretation are quite sophisticated, but they are never as good as a trained crystallographer. Do not, therefore, accept the given interpretations as the only possibilities. Some other points to note are:

- If the map contains one or two large peaks and no heavy atoms are expected, the phases are probably incorrect - but not always. Sometimes something can be salvaged. If heavy atoms are present direct methods will probably only produce these atoms.
- If the E-maps show pseudosymmetry switch to Hull-Irwin weights in TANGENT.
- One or two missing peaks coupled with one or more spurious ones, can quickly make a map uninterpretable. Increasing the number of peaks can make parts of the map more readily interpretable at the expense of producing more noise peaks.

In very difficult cases be tenacious. If just a small portion of the expected structure is found, stick with it through all the possible recycling schemes. It may be correctly oriented but misplaced, so the use of type 3 groups in NORMAL will be useful, but this procedure is not infallible. It sometimes does not work even with correct information.

9.7 References

Baggio, R. et al. (1978) *Acta Cryst.*, **A34**, 883-892

Busetta, B. et al. (1974) *Acta Cryst.*, **A30**, 564-568

Cochran, W. et al. (1955) *Acta Cryst.*, **8**, 473-478

Debaerdemaeker, Tate and Woolfson (1986) *Acta Cryst*., **A41**, 286-290

Declercq, J.P. et al. (1975) *Acta Cryst.*, **A31**, 367-372

Declercq, J.P. et al. (1979) *Acta Cryst.*, **A35**, 622-626

DeTitta, G.T. et al. (1975) *Acta Cryst.*, **A31**, 474-479

Freer, A.A. et al. (1980) *Acta Cryst.*, **A36**, 470-475

Germain, G. et al. (1968) *Acta Cryst.*, **B24**, 91

Giacovazzo, C. (1980a) *Direct Methods in Crystallography,* Academic Press

Giacovazzo, C. (1980b) *Acta Cryst.*, **A36**, 74-82

Gilmore, C.J. (1977) *Acta Cryst.*, **A33**, 712-716

Gilmore, C.J. et al. (1990) *Acta Cryst.*, **A46**, 297-308

Gilmore, C.J. et al. (1988) *Acta Cryst.*, **A44**, 1018-1021

Gilmore, C.J. et al. (1977), *JCS Perkin II*, 1427-1434

Gilmore, C.J. et al. (1985) *Acta Cryst.*, **A41**, 457-462

Hall, S.R. et al. (1982) *Acta Cryst*, **A38**, 598-608

Hauptman, H. (1972) *Crystal Structure Analysis: The Role of the Cosine Seminvariants*, Plenum Press, New York

Hauptman, H. (1977a), *Acta Cryst.*, **A33**, 556-564

Hauptman, H. (1977b), *Acta Cryst.*, **A33**, 565-568

Hauptman, H. (1980) in *Theory and Practice of Direct Methods in Crystallography,* 151-197, Editors M.F.C. Ladd and R.A. Palmer, Plenum Press

Henderson, R.K. et al. (1990) *J. Appl. Cryst*., **23**, 143-144

Hull, S.E. et al. (1978) *Acta Cryst.*, **A34**, 863-870

Hull, S.E. et al. (1981) *Acta Cryst.*, **A37**, 566-572

Karle, J. (1967) *Acta Cryst.*, **B24**, 182-186

Karle, J. et al. (1966) *Acta Cryst.*, **21**, 849-859

Main, P. (1976) in *Crystallographic Computing Techniques*, 97-105, Munksgaard, Copenhagen

Main, P. (1978) *Acta Cryst.*, **A34**, 31-38

Main, P. (1980) in *Computing in Crystallography,* 8.01-8.13, Indian Academy of Sciences

Main, P. et al. (1978) *Acta Cryst.*, **A34**, 353-361

Main, P. et al. (1980) *Multan-80 Manual*, University of York

Moore, F.H. (1963) *Acta Cryst.*, **16**, 1169-1175

Nixon, P.E. (1978) *Acta Cryst.*, **A34**, 450-453

Rodgers, D. (1965) in *Computing Methods in Crystallography*, 117-148, Editor J.S. Rollett, Pergamon Press

Rodgers, D. (1980) in *Theory and Practice of Direct Methods in Crystallography*, 82-92, Editors M.F.C. Ladd and R.A. Palmer, Plenum Press

Sim, G.A. (1959) *Acta Cryst.*, **12**, 813-815

Sim, G.A. (1960) *Acta Cryst.*, **13**, 511-512

White, P.S. et al. (1975) *Acta Cryst.*, **A31**, 53-56

Yao Jia-Xing (1981) *Acta Cryst.*, **A37**, 642-644

Yao Jia-Xing (1983) *Acta Cryst.*, **A39**, 35-37

10 SHELXS

10.1 Introduction

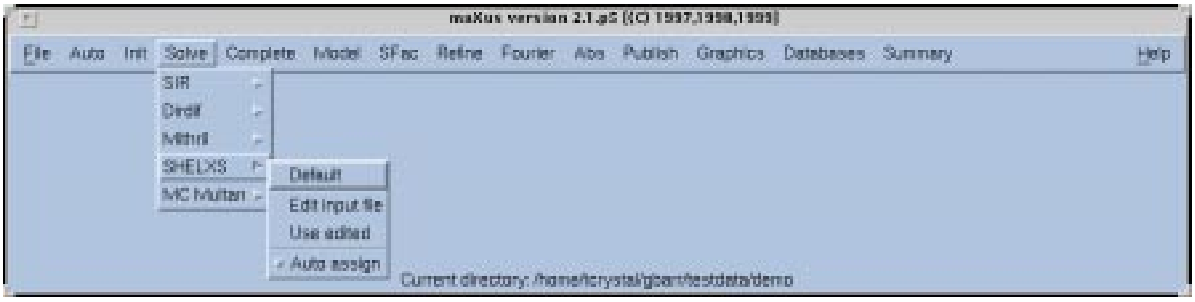
SHELXS is a crystallographic refinement program, written by Professor George Sheldrick and is not distributed with maXus. This manual details the maXus - SHELXS interface and will give guidance on how to obtain and integrate SHELXS with the maXus package.

The full SHELXS manual is only available from G. Sheldrick (for ordering information see Appendix A). This manual is written based on the assumption that the reader has a licenced copy of the full SHELXS manual.

SHELXS is used in maXus as one of several methods of solving crystal structures. It operates either in full black box mode; requiring no input from the user or the user may specify a set of commands which include automatic Patterson solution methods.

10.1.1 How to Access SHELXS *via* maXus

From the main menu:



Current directory: /home/lcrystal/gbart/testdata/demo

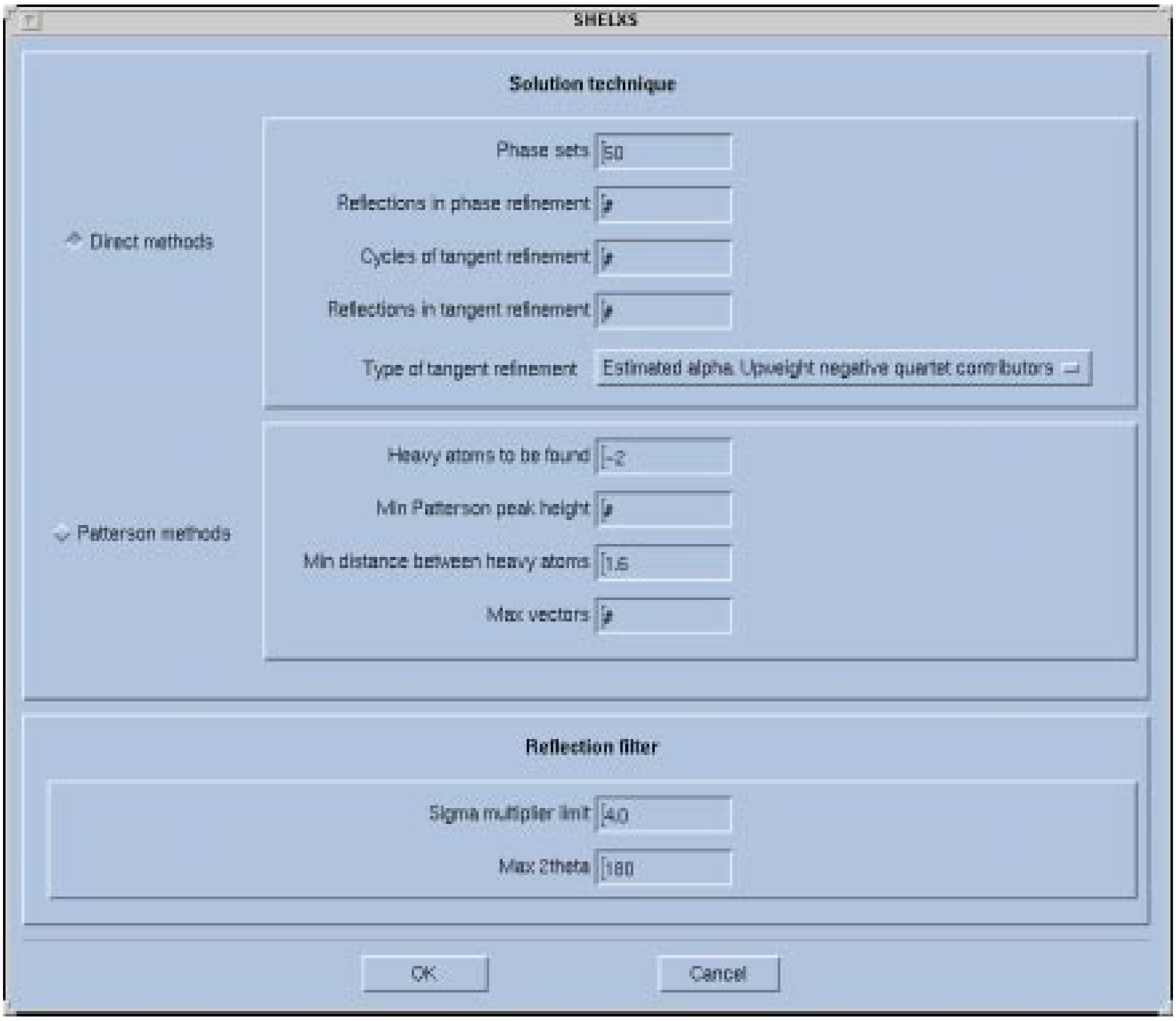
Choose the *Solve* option, followed by *SHELXS*. There are four options on the sub-menu:

• Default	See Section 10.2
• Edit input file	See Section 10.3
• Use edited	See Section 10.4
• ❑Auto assign	

The ❑Auto assign button is used as follows:

After the structure has been solved the atom types will still be unassigned and there will be no initial least squares and structure completion. By default, after each run of SHELXS the SIR92 auto-complete program is run, atoms are assigned to types and preliminary least-squares is carried out. If you select this option, the ❑ symbol disappears and SHELXS will run without the auto complete option. When you then enter the Model window all atoms will be unassigned, and coloured white - you will need to assign types yourself. In difficult cases this may be the preferred option. To restore the Auto assign option just select the menu option again; the ❑ symbol will return.

10.2 Default Option



SHELXS

Solution technique

● Direct methods

Phase sets: 50
Reflections in phase refinement:
Cycles of tangent refinement:
Reflections in tangent refinement:
Type of tangent refinement: Estimated alpha. Upweight negative quartet contributors

○ Patterson methods

Heavy atoms to be found: -2
Min Patterson peak height:
Min distance between heavy atoms: 1.6
Max vectors:

Reflection filter

Sigma multiplier limit: 4.0
Max 2theta: 180

10.2.1 Direct Methods

This allows the user to choose whether the structure will be solved using direct methods which will involve setting the parameters on the right of the button.

10.2.2 Phase Sets

This determines that number of phase sets that are to be evaluated. The default for this parameter is 50. For difficult structures it is recommended that this be increased.

10.2.3 Reflections in Phase Refinement

The default for this parameter is that SHELXS determines an optimum selection internally. If a negative number for this figure is input then extra output, to be used for diagnostic purposes, is produced.

10.2.4 Cycles Of Tangent Refinement

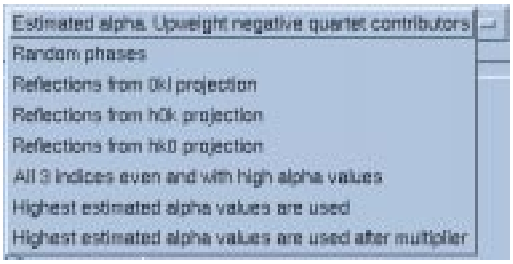
The default for this parameter is that SHELXS determines an optimum figure internally. A negative number input to this field will force SHELXS to use additional reflections in addition to those used for refinement, will be used in the calculation of R_α.

10.2.5 Reflections in Tangent Refinement

This refers to the number of reflection to be used in the initial subset of reflections to be used for tangent refinement.

10.2.6 Type of Tangent Refinement

Eight types of tangent subset selection for refinement are provided within SHELXS. To change the subset selection criteria, click on the option field. This will cause the following pop up menu to appear.



The following options are available from the pop up menu:

- Estimated alpha: Upweight negative quartet contributors. This will select reflections based on estimated alpha values but will also prioritise reflections that take part in the most negative quartets.
- Random Phases: This selects reflections based on estimated alpha values but does not refine the subset, thus resulting in random phases being passed to full tangent refinement.
- Reflections from 0kl projections: Reflections will be chosen from the 0kl projection based on estimated α.
- Reflections from h0l projections: Reflections will be chosen from the h0l projection based on estimated α.
- Reflections from hk0 projections: Reflections will be chosen from the hk0 projection based on estimated α.
- All 3 indices even with high alpha values: Reflection will only be chosen with even indices based on estimated α values.
- High estimated α values are used: The reflections with the highest estimated alpha are used.

- Highest estimated α values are used after multiplier: The reflections with the highest estimated alpha are used but the α value is corrected for temperature effects by multiplying with the equation:

$$e^{-Bsin^2\theta/\lambda^2}$$

10.2.7 Heavy Atoms to be Found

This is the number of heavy atoms to be found using the Patterson technique. If the number here is positive then that number of heavy atoms will be found. If the figure is negative, the default is -2, then all heavy atoms are found and the best solution found by the Patterson interpreting program will be used to further expand the structure.

10.2.8 Patterson Methods

This button allows the user to choose whether the structure will be solved using Patterson (heavy atom) methods which will involve setting the parameters on the right of the button.

10.2.9 Min. Patterson Peak Height

This parameter is used to determine the minimum height of the Patterson peak that will be accepted for interpretation. The default is set to # and thus SHELXS will determine a value for this based on the type of heavy atom based on the information given in the maXus model files.

10.2.10 Min. Distance Between Heavy Atoms

This is the minimum distance allowed between heavy atoms in Å.

10.2.11 Max Vectors

This allows the user to set the maximum number of Patterson vectors between pairs of atoms that are to be interpreted during the search for a solution.

10.2.12 Reflection Filter: Sigma Multiplier Limit

This allows the user to filter the data being used to determine the initial structural solution.

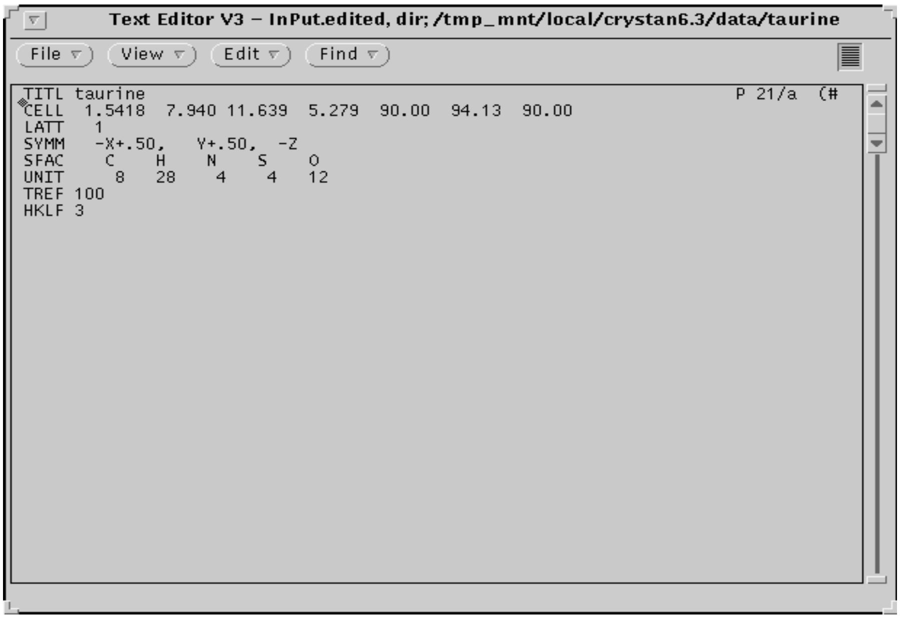
This option will flag all reflections with a magnitude less than this multiplier times the error as being unobserved. The default for this parameter is 4.0.

10.2.13 Reflection Filter: Max 2 Theta

Any reflections with a 2θ value greater than this threshold will be flagged as unobserved.

10.3 Edit Input File

This option works in conjunction with the *Use edited* command. It enables the user to enter their own SHELXS commands. When this option is chosen, a text editor pane appears:



```
TITL taurine                                                   P 21/a  (#
CELL  1.5418  7.940 11.639  5.279  90.00  94.13  90.00
LATT   1
SYMM   -X+.50,   Y+.50,  -Z
SFAC    C    H    N    S    O
UNIT     8   28    4    4   12
TREF 100
HKLF 3
```

The commands used in this example are:

TITL - A title
CELL - The cell dimensions
SYMM - The symmetry operations.
SFAC - The atom types in the unit cell.
UNIT - The number of atoms in the unit cell.
TREF
HKLF

These commands can be edited and new ones added. When you have finished, exit the editor in the usual way.

10.3.1 SHELXS Instruction Summary

For details on the various commands, their syntax and use, refer to the full SHELXS manual, supplied with the program. Basically, the instructions TITL ... UNIT must appear in the order given below, and HKLF 3 must be the last instruction. Default values a given in square brackets; ‘#’ indicates that the program sets a suitable value for the problem in question.

```
TITL[ ]
```

```
CELL a b c a alpha beta gamma
```

```
LATT N [1]
```

```
SYMM symmetry operation
```

```
SFAC elements
SFAC al bl a2 b2 a3 b3 a4 b4 c df′ df″ p r wt
UNIT nl n2 ...
OMITs [4] 2qmax[180]
OMIT h k l
ESELE Min [1.2] E max [5] dU [.005] renorm [.7] axis [0]
LIST m
FMAP code [#] axis [#] n1 [#]
GRID st [#] sa [#] sd [#] di [#] da [#] dd [#]
PLAN npeaks [#] d1[.5] d2 [1.5]
PATT m [-2] thresh [#] rmin [1.6] maxvecs [#]
PSEEm [200] 2qmax
PHAS h k l f
MOLE n [#]
MOVE dx [0] dy [0] dz [0] sign [1]
SPIN f1 [0] f2 [0] f3 [0]
FRAG code [0] a [1] b [1] c [1] a [90] b [90] g [90]
atom sfac xy z sof [1] U or U11 to U12
TEXP na [#] nH [0] Ek [1.5]
TREF np [50] nE [#] ntan [#] tw [0.1] wn [#]
SUBS type [4] ns [#]
END n
HKLF 3
```

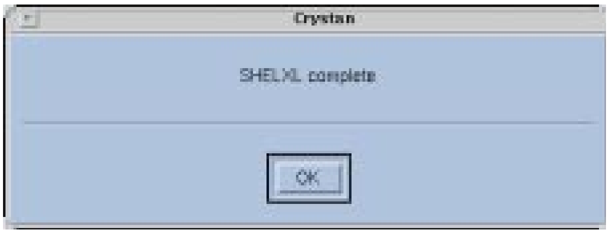
10.4 Use Edited Option

This will run SHELXS using the file edited by the user using the *Edit Input File* option previously.

10.5 End of SHELXS Processing

While SHELXS is running a performance meter will appear at the top of the screen. When SHELXS has completed the performance meter will be closed. The first area of the meter labelled Normal, will be highlighted while the conversion of file formats from maXus to SHELXS is underway. The second area labelled SHELXS will be highlighted while SHELXS is executing. The third area, labelled 2Peak, will be highlighted while the conversion of file formats from SHELXS back to maXus is underway. The fourth area, labelled Mkmol, will be highlighted when the peaks are being interpreted into a molecule or molecular fragment. It is possible to close this window while SHELXS is running. The performance meter window is shown below:

Once SHELXS has successfully completed the following window will appear:



SHELXL complete

Click on OK and maXus will return to the control of the main menu.

10.6 How SHELXS Solves Structures

SHELXS carries out the following steps:

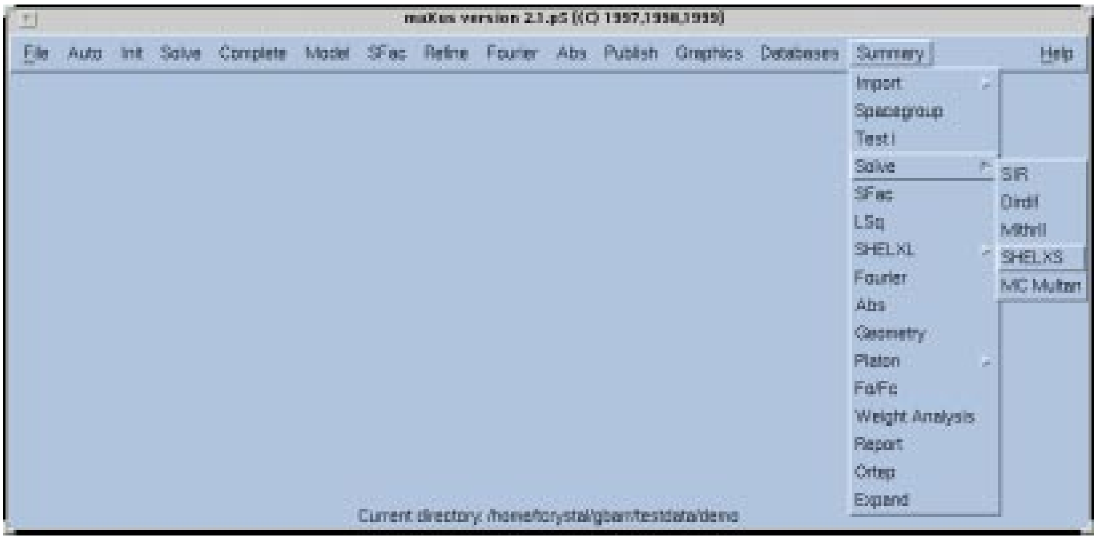
- Data normalisation.
- The generation of triplets and negative quartets.
- The generation of one-phase structure seminvariants.
- Tangent refinement using random phases, and filters based on the negative quartet figure of merit and the structure seminvariants.
- An E-map for the best solution
- Fragment interpretation.

10.7 SHELXS Output

While SHELXS is running, a window containing summary results will be displayed.

A detailed solution report from SHELXS can be accessed using the *Summary* option from the main menu.

This report will appear in an edit window and can be used review the solution process. This window can be dismissed in the normal manner. See below:



maXus version 2.1.p5 ((C) 1997,1998,1999)

Current directory: /home/crystal/gbam/testdata/demo

Choose the *Summary* option, followed by *Solve* followed by *SHELXS*.

APPENDIX A

Obtaining a Copy of SHELXS

APPLICATION FORM FOR SHELXS

SHELXS USER REGISTRATION FORM | Do not write here!

Title / Name:| Date Sent:

Full Postal Address:| Version:

Email address (if available):

Tel.:

Fax:

Please tick ALL relevant boxes, sign and return to: Prof. George Sheldrick, Institut fuer Anorg. Chemie, Tammannstrasse 4, D-37077 Goettingen, Germany. Fax: +49 551 393373; E-mail: gsheldr@shelx.ini-ac.gwdg.de

SHELXS is supplied ONLY on MSDOS format diskettes, in the form of self-extracting packed files containing MSDOS executables and sources for UNIX and VMS systems, documentation and test data. The programs PDBINS (PDS to SHELXS format conversion) and CIFTAB (tables from SHELXS CIF output) are included. The licence fee of DM 4999 for for-profit institutions covers use for an unlimited time on an unlimited number of computers at a specified firm or institution at a single geographical site. SHELXS is currently available free of charge to academics for non commercial use only; it may prove necessary to change this policy if the licence fees for for-profit institutions fail to cover the total costs involved. Academic institutions willing and able to contribute to the costs of developing and distributing the SHELX programs are of course welcome to do so (we suggest DM 99). Please make out checks to “Institut fuer Anorg. Chemie, Prof. Sheldrick”. If you wish to pay by direct bank transfer please ask us to send an invoice.

[ ] I wish to license SHELXS for use at the following for-profit firm or institution. I agree that within three months I will either destroy all copies of the program in my possession or pay the license fee of DM 4999.

[ ] The program will be used exclusively for non-commercial purposes at the following not-for-profit institution only.

[ ] Please sent me an invoice for DM.

[ ] Please send me a receipt for the enclosed payment of DM.

[ ] I agree to cite SHELXS in all publications reporting results obtained using it.

[ ] I accept that the author has no liabilities in respect of errors in the program or documentation.

Please supply SHELXS on:

[ ] 1.44 MB [ ] 1.2 MB [ ] either 1.2 or 1.44 MB MSDOS diskettes

[ ] I already possess a copy of SHELXS

APPENDIX B

Installing SHELXS in maXus

Once SHELXS has been obtained it must be compiled as per the instructions that come with the SHELXS source code.

The final executable object must be placed in the maXus binary directory, called “/usr/maXus/bin.<operating system name>” in a standard implementation. This module must be called shelxs97, with no capitalisation.

The operating system name can be: sunos5, linux, irix5 or irix6

Example:

Having obtained a copy of SHELXS and compiled it in a personal library to an executable object called a.out, you would perform the following UNIX command for moving the object from your own directories to the maXus executable directories on a SGI running IRIX5.

```
mv /local/usr/myname/shelxs/a.out /usr/maXus/bin.irix5/shelxs97
```

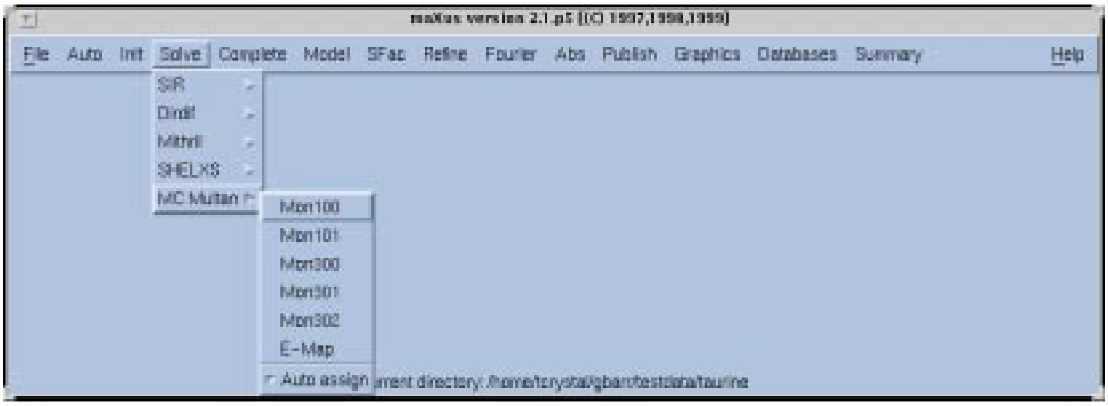
Once the executable is present in the appropriate directory maXus will recognise this and all the SHELXS functionality described in this manual will be available.

11 MC MULTAN

Monte-Carlo Structure Solution

11.1 Monte-Carlo MULTAN

To run this option, choose *MC Multan* from the *Solve* menu.



The following options are available:

- MON 100
- MON 101
- MON 300
- MON 301
- MON 301
- E-Map (see Section 11.3)
- ❑Auto assign

The ❑Auto assign button is used as follows:

After the structure has been solved the atom types will still be unassigned and there will be no initial least squares and structure completion. By default, after each run of MC Multan the SIR auto-complete program is entered, and atoms are assigned to types and preliminary least-squares is carried out. If you select this menu option, the ❑ symbol disappears and MC Multan will run without the auto complete option. When you then enter the Model window all atoms will be unassigned, and coloured white - you will need to assign types yourself. In difficult cases this may be the preferred option. To restore the Auto assign option just select the option again; the ❑ symbol will return.

11.2 Operation

These all run the following direct methods sequence:

- Data normalisation using the Wilson plot.
- Triplet generation
- Convergence mapping.

At this point tangent refinement commences with different options according to the MON option selected:

- MON 100 generates 100 random phase sets using a starting set that comprises 15 normalised reflections which are the 15 strongest.
- MON 101 generates 100 random phase sets using a starting set that comprises 15 normalised reflections which are those numbered 6-20 in the list of normalised structure factors.
- MON300 generates 300 random phase sets using a starting set that comprises 50 normalised reflections which are the 50 strongest.
- MON 301 generates 300 random phase sets using a starting set that comprises 50 normalised reflections which are those numbered 6-55 in the list of normalised structure factors.
- MON 303 generates 300 random phase sets using a starting set that comprises 50 normalised reflections which are those numbered 11-60 in the list of normalised structure factors.

Note that options MON 300, 301,302 take three times longer than the MON 100 and 101 options.

After selecting the desired option, a process indicator appears with the processes:

The program proceed quickly through Normal, but Multan will be much slower, and could take several minutes. This is followed by Exfft which computes an E-map for the best phase set; this is searched for peaks using the Search routines, and then interpreted chemically *via* Mkmol. The results of this can be viewed graphically from the MODEL window.

Detailed reports of Multan operation are available from the *Summary* main menu. Select *Solve*, and then *MC Multan*. A text window appears with the output file for viewing.

11.3 E-Maps

If you wish to view other E-maps generated by this procedure (if, for example, the map you have just viewed is incorrect or difficult to interpret) then select the *E-Map* option from the *MC Multan* sub-menu of *Solve*.

A process indicator appears:

The next map will be generated, and its results written to a file, which can then be examined.

12 Model

A Program to Manipulate Molecules and Prepare for Real-Time Least Squares.

12.1 Introduction

Three options are available from the Model menu:

- Display
- Display Extended
- Change Formula

The Model window represents the very heart of the maXus package, allowing atomic parameters and molecules to be visualised graphically. Its main purpose is to prepare a Model for least squares refinement. Other functions includes viewing the output from the direct methods programs and assigning atom types to the peaks. It carries out the following functions:

- It permits the manipulation of the molecules or fragments found from the structure solution packages (MITHRIL, SHELXS etc.)
- It allows the deletion of unwanted atoms.
- H atoms can be added in idealised geometry.
- Constraints, including rigid bodies, can be added for least squares.
- Individual atoms may be damped for least-squares.
- Fragments in different parts of the unit cell can be joined by application of space group symmetry.
- Thermal parameters can be set to isotropic or anisotropic.
- It gives a clear drawing of the molecule in three different formats.
- It can assign atoms as Deuterium.

Depending on the type of molecule you are studying, there are two different versions of the Model window available - the standard, and expanded displays..

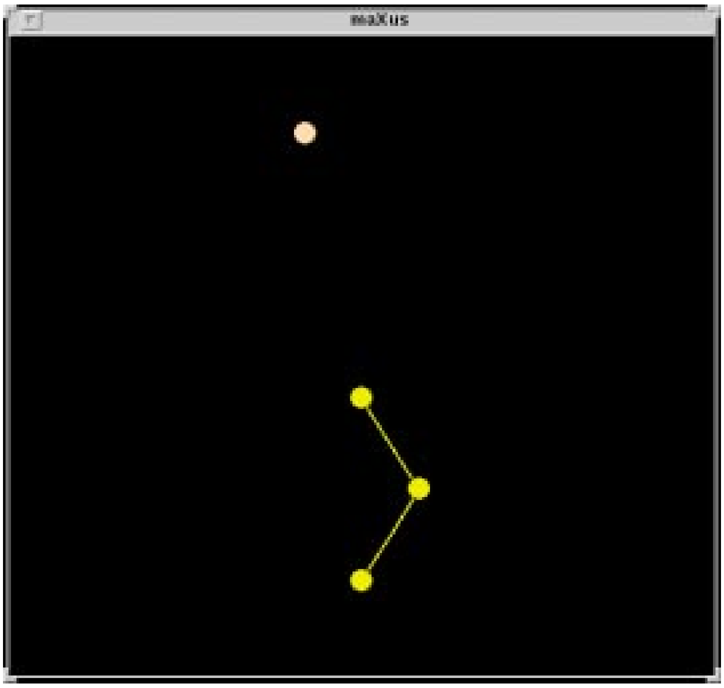
With small molecules of low symmetry, where the whole molecule is contained within the asymmetric unit, use the standard window. This has additional features not available in the expanded window.

The expanded model window is useful when you are dealing with molecules with inherent symmetry, where viewing just the asymmetric unit cell (as displayed by the standard model) is not hugely meaningful.

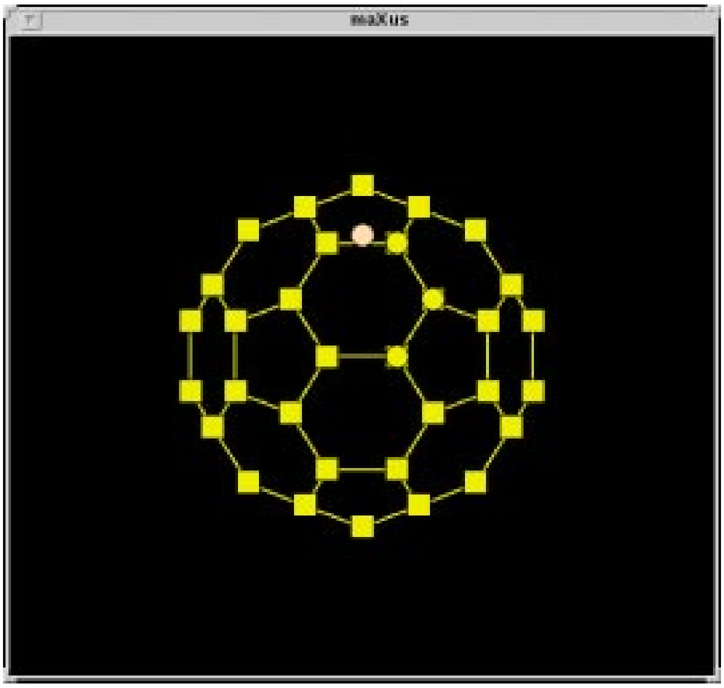
It attempts to expand the assymmetric unti cell to generate symmetry related atoms for the molecule you are studying, using routines developed for the Platon program (Spek, A.L. (1990), *Acta Cryst.* A46, C34).

For the example below the differences between the two display modes are obvious:

The standard window shows only four symmetry unique atoms (3 carbons and a potassium), making it very difficult to see what the full structure actually looks like:



Using the expanded display however, the full molecule can be easily seen:



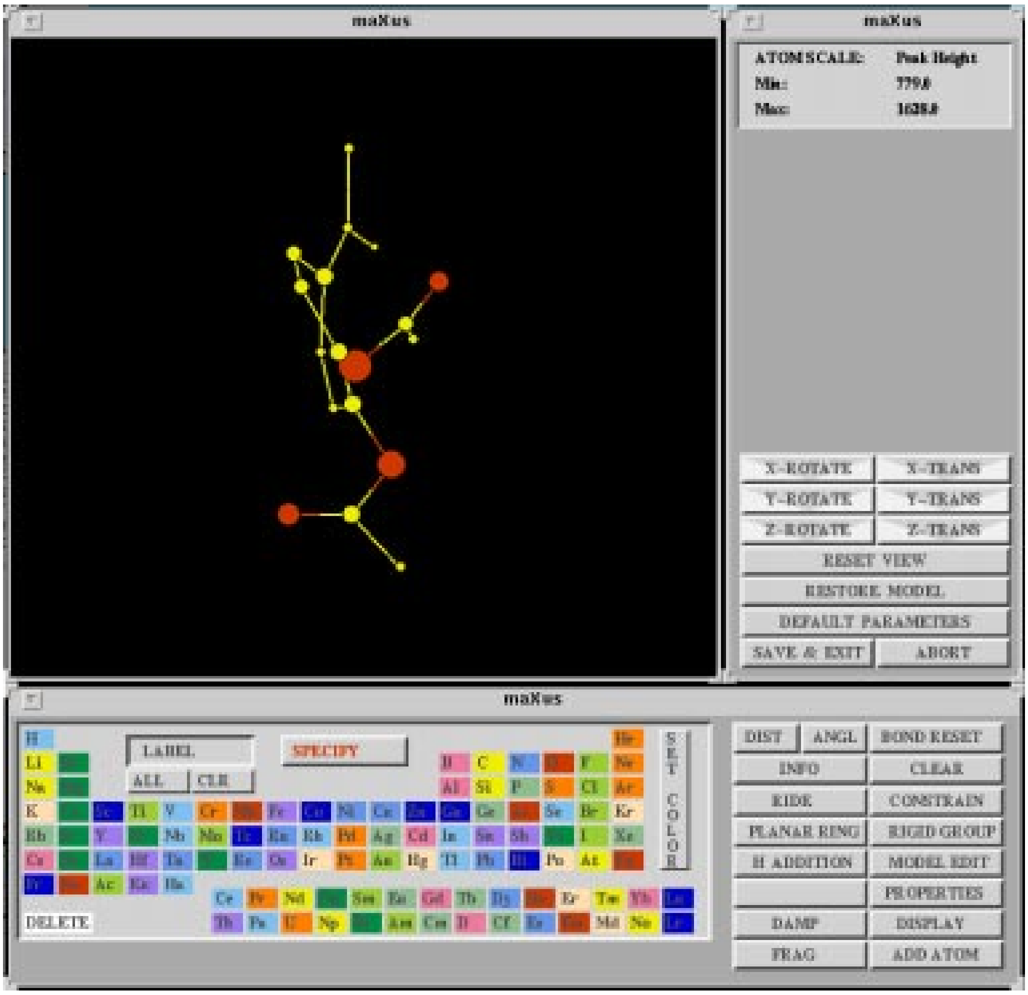
C_{60}, in all its glory.

Note that some features available in the normal window are not available in the expanded version. See below (Section 12.5) for full details.

12.2 Detailed Operation - Standard Window

Select the *Display* option from the *Model* menu.

The basic screen of Model will appear as shown in the example below:



ATOMSCALE: Peak Height
Min: 779.0
Max: 1628.0

The workspace is divided into three basic regions.

- The graphics window (upper left), where the molecule model is displayed.
- The menu window (upper right), where the key control buttons reside.
- The sub-menu window (lower window), where procedures invoked by control buttons from the menu window are executed.

12.3 The Menu Window

12.3.1 X-,Y- AND Z-ROTATION

These buttons rotate the molecule around the three axes. **The buttons are position sensitive. Starting from the left end of each button the molecule will rotate anticlockwise very quickly. As you move the mouse closer to the centre the rate falls, then it reaches zero before changing direction to clockwise and speeding up once more as you proceed further to the right.**

12.3.2 X-,Y- and Z TRANSLATION

These buttons translate the molecule along the three axes. **These buttons are also position sensitive. Starting from the left end of each button the molecule will translate in one direction very quickly. As you move the mouse closer to the centre the rate falls, then it reaches zero before changing direction and speeding up once more as you proceed further to the right.**

12.3.3 RESET VIEW

Resets the view to the original, default view i.e. normal to the least-squares plane.

12.3.4 RESTORE MODEL

This restores the model to its previous state as stored in a backup file. A dialog box appears:

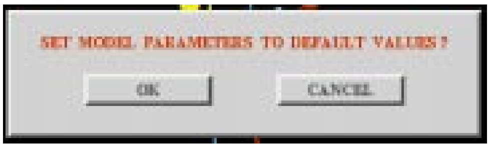
.

SET MODEL PARAMETERS TO DEFAULT VALUES ?

Click OK to use the backup.

12.3.5 DEFAULT PARAMETERS

This reset all the model parameters back to their defaults. A box appears:



SET MODEL PARAMETERS TO DEFAULT VALUES ?

Click OK to reset.

12.3.6 SAVE and EXIT

Saves all the current work/settings and exits.

12.3.7 ABORT

Quits without any saves - **any changes you have made are not kept!**

12.4 The Sub-menu Window

Each button in the menu window will be discussed in turn. **In what follows SELECT refers to the left mouse button.**

12.4.1 Atom Labelling & Specification

In the default view, a Periodic Table is displayed in the sub-menu.

To see the label for a particular atom, click *Label*, and then on the atom you wish to see the label for. To label all atoms, click *All*. To remove the labels, click *Clr*.

To change the element type assigned to a particular atom, click *Specify*. Choose the element you wish, and then click on the atom you wish to assign this element type to. It should change colour to match the colour assigned to the element chosen in the Periodic Table. If it is labelled the label will also change accordingly.

Note that it is possible to label an atom as Deuterium. This appears as **D** in the bottom row of the Periodic Table, 8th from the left, between **Cm** and **Cf**. (It appears here for complicated programming reasons.) It functions like any other element. Most maXus programs will recognise this as a separate atom type; any that do not treat it like Hydrogen.

When the *Delete* button (bottom left) is selected, clicking on atoms removes them from the display.

12.4.2 DIST

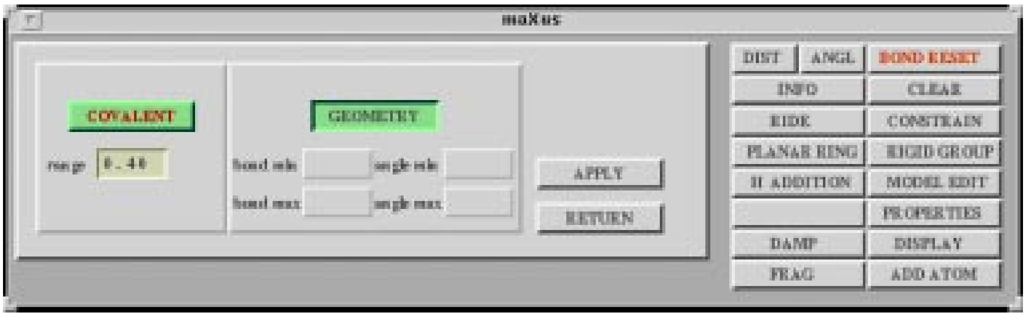
This is a function to determine the interatomic separation between any two atoms, within the asymmetric unit. Position the cursor over the 1st atom and press SELECT. Repeat over the second atom. The resulting separation is printed in the blank workspace above the main control buttons. This process can be repeated or control returned to the other control buttons by pressing CLEAR. The workspace will be returned to the sphere size chart.

12.4.3 ANGL

This works in the same fashion as DIST, but has the result of calculating the interatomic angle for any three atoms, in degrees. Note, the atoms do not have to be bonded for the calculation. CLEAR returns control.

12.4.4 BOND RESET

This function recalculates the connectivity of the molecule, based on covalent radii or a simple angle search as shown below:



range 0.40

In the COVALENT mode, unassigned peaks (white) will not be bonded to any atom/peak. In GEOMETRY mode the minimum and maximum bond lengths and angles can be set. This will allow bonding of unassigned peaks, to illustrate possible molecular fragments for atom assignment. This also allows control of bonding for molecules not handled well by the default covalent algorithm.

12.4.5 INFO

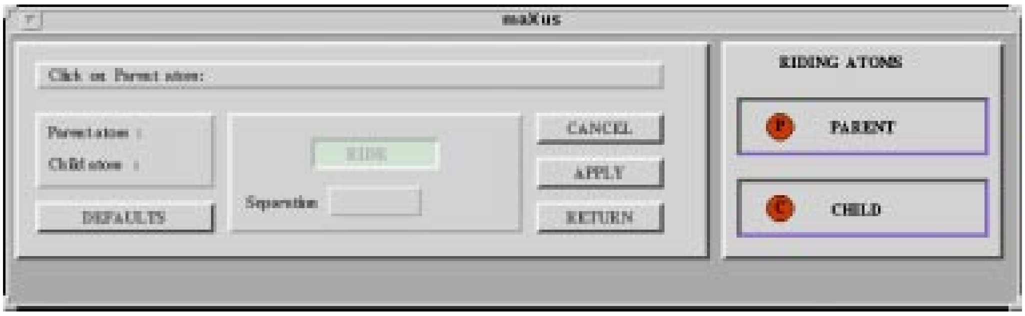
This lists the entire contents of the asymmetric unit in a scrolling window above the control buttons. The names and the peak heights are displayed in the default mode. By clicking on any of the atom names, the user can change the atom names. The entry in the scrolling list becomes countersunk and a cursor appears at the end of the present name. By typing in the new name or simply pressing RETURN will reset the name. If a simple RETURN is depressed without any editing, the record remains unchanged. Note that the atom names must be restricted to 5 characters.The scrolling list can be activated when there are more than 24 peaks and atoms. When this is the case, positioning the cursor over the scroll bar and dragging has the action of scrolling the atom list.

12.4.6 CLEAR

As mentioned in the DIST/ANGL function descriptions, CLEAR has the effect of returning control back to the other control buttons. At the same time it resets the peak height indicator.

12.4.7 RIDE

In refinement it is often necessary to restrict the relative movement of a pair of atoms. In the most common case, H atom bonding geometry may be required to remain constant during the least squares refinement. RIDE allows the user to maintain the relative positions of up to 500 pairs of atoms, during refinement. The least squares process will refine the positional parameters of the parent atom and the child atom will ride on it. The RIDE window appears as follows:



Click on Parent atom:

Parent atom :

Child atom :

The user is first prompted to select the parent atom, and then the child atom, which will be refined on the parent atom. The child and parent atoms' names and their atom-atom separation are displayed to clarify the selection. With the RIDE button selected, click APPLY to store the selection. The CANCEL button will abort from the current selection of selected atoms. CANCEL can be pressed at any time during the process.

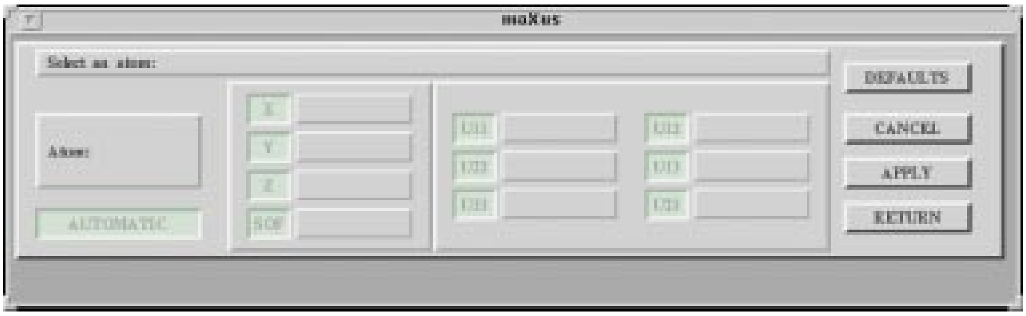
The separation distance can be set without enforcing a restraint by typing the desired distance, leaving the RIDE button deselected and pressing APPLY. This will shift the child atom to produce the specified distance.

Ensure that the order of selection is correct or the refinement will fail to produce the desired results.

The DEFAULTS button will undo all the restrained pairs which have been selected.

12.4.8 CONSTRAIN

In certain circumstances the user may wish to place constraints on specific atomic parameters in the asymmetric unit during cycles of least squares refinement. CONSTRAIN allows the user to do just that. By selecting CONSTRAIN, a new sub-menu appears, replacing the periodic chart. The sub-menu is shown below:



Select an atom:

Atom:

The user is first prompted to select an assigned peak (white unassigned peaks are not considered in the LS refinement). When selected the X, Y, Z and SOF (occupancy) buttons become active and the text fields will display the corresponding parameter values. By default the SOF value will be fixed, indicated by the button in the ON position. The other buttons will generally be in the OFF positions, unless the atoms position is subject to special symmetry operations. The occurrence of special positions will be discussed later.

To restrain any of the parameters, press the appropriate button(s) and follow this by APPLY. If an atom has a one or more coordinates with values close to a significant fraction, say X=0.2448 the value can be set to 0.2500 by editing the text in the text field.

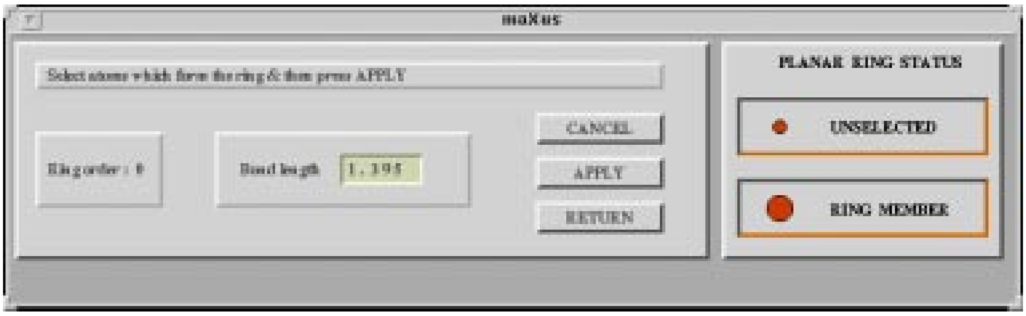
In the case where an atom is located on a special position certain parameters will be automatically constrained. Although not indicated where necessary, the thermal parameters will be constrained in accordance with the symmetry operators of the space group. The relevant parameter buttons will be de-activated and so the user will not be able to undo the constraints. If the user wishes to allow freedom of refinement for atoms on special positions, then the space group must be simplified to remove the special symmetry operations. Alternatively, the AUTOMATIC button may be turned off, allowing the user to override the special-position constraints. This should only be used with great care, as automatically included constraints may be required to ensure stable refinement (in the case of a floating origin for example).

The DEFAULTS button will, when pressed, undo all the constraints applied to all of the atoms, leaving them with default values. If atoms lie on special positions, the automatically enforced constraints will remain.

In general, constraints are automatically computed where they are imposed by crystallographic symmetry.

12.4.9 PLANAR RING

This function enables the user to select a set of atoms which form an approximately planar ring, then make it into a planar moiety ready for rigid body least-squares.



Select atoms which form the ring & then press APPLY

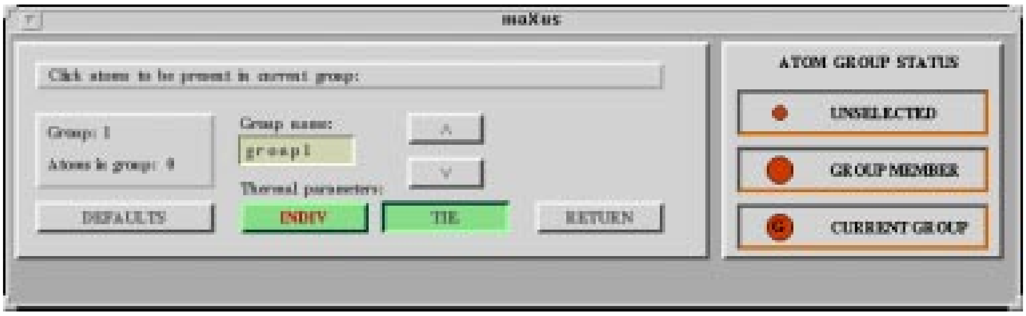
Ring order : 0

Bond length 1.395

To make a ring planar click on ALL the atoms that comprise the ring (e.g. all 6 for a phenyl group) then click APPLY. The selected atoms are made larger as they become selected. You can cancel using the CANCEL button, and change the bond length by clicking in the text box and typing a new value.

12.4.10 RIGID GROUP

This has the effect of instructing the least squares process to refine a group of atoms as one rigid group. This can be useful when a cyclic ring is required to retain its original geometry during refinement. The GROUP window has the following appearance:



Click atoms to be present in current group:

Group: 1
Atoms in group: 0

Group name: group1

The least squares program will permit up to 15 rigid groups, with up to 20 atoms per group. Each group is defined by a name and a number. By default the group names are “group#”, where # assumes numerical values ranging from 1 to 14. The user selects the group number using the /\ and \/ buttons on the control panel. Initially group1 will be selected and all atoms will be represented by small equally sized circles. To place a selection of atoms in the current group, select each of the atoms in turn. Note that on selection of each atom the atom radius will double and a “G” will appear in the centre. All atoms with the enlarged radii and “G” symbol indicate those atoms in the currently selected group.

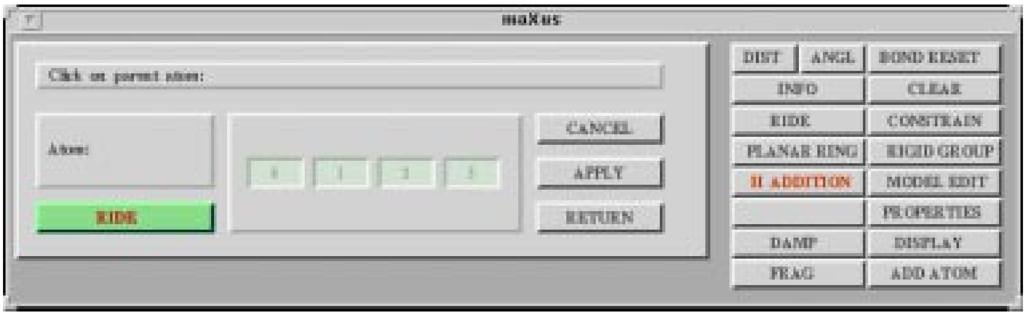
By changing the group number to two, the atoms belonging to the previous group and indeed all other groups will remain double sized, as coloured rings. This indicates that these atoms do in fact already belong to a group, so that no confusion over groups can occur.

> **Note: No atom is able to belong to more than one group. If an atom has been associated with a group and then it is added to another one, the previous association is lost.**

The DEFAULTS button has the effect of removing all group associations.

12.4.11 HYDROGEN ADDITION

Hydrogen atoms can be added to structures in calculated positions. The parent atom is first identified and the present H population is displayed on the buttons numbered 0 to 3. To change the H population for any atom, select the appropriate number and press APPLY. The new hydrogen atoms will be added to the molecule display.These calculated H atoms have no peak height and so are represented by very small spheres. In addition the new atoms are assigned default isotropic thermal parameter values of 0.05. The appropriate window has the following appearance:



Click on parent atom:

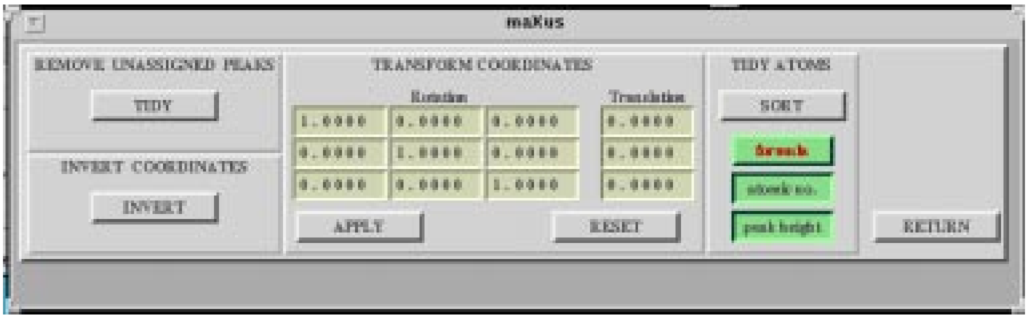
Atom:

In the situation where the proposed population number does not fit the model, the procedure will fail and no action will be taken. It is up to the user to determine the correct coordination of the parent atom. Again CANCEL and RETURN have the same effect.

In order to add a hydrogen to a terminal oxygen, you have to first add three, and the delete the two unwanted atoms.

12.4.12 MODEL EDIT

Selecting this gives the following options:



1.0000	0.0000	0.0000	0.0000
0.0000	1.0000	0.0000	0.0000
0.0000	0.0000	1.0000	0.0000

Clicking TIDY removes any unassigned (white) peaks from the display.

Clicking INVERT inverts all of the atomic co-ordinates in the Model file.

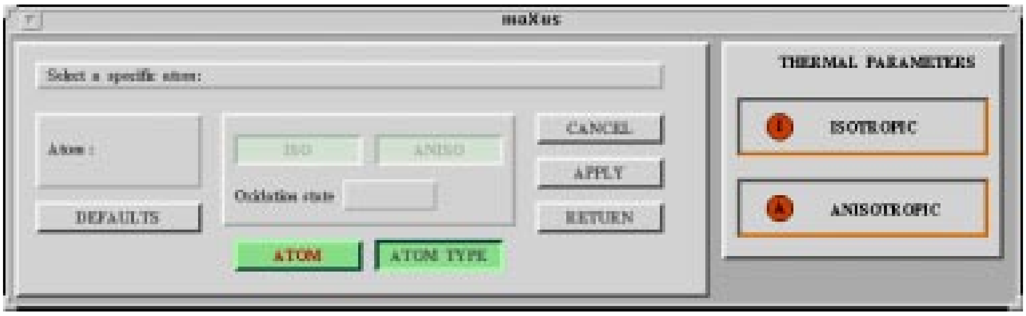
The *Transform Coordinates* window allows you to enter a rotation, or translation matrix, and then APPLY it to the Model file. Note that this does not merely change the molecule view; it alters the coordinates stored in the Model file.

Clicking SORT under *Tidy Atoms* sorts the list of atoms, and atomic coordinates, in the chosen order; either by Formula, Atomic Number, or Peak Height.

RETURN brings the main Model window back up.

12.4.13 PROPERTIES

The properties window has the following options:

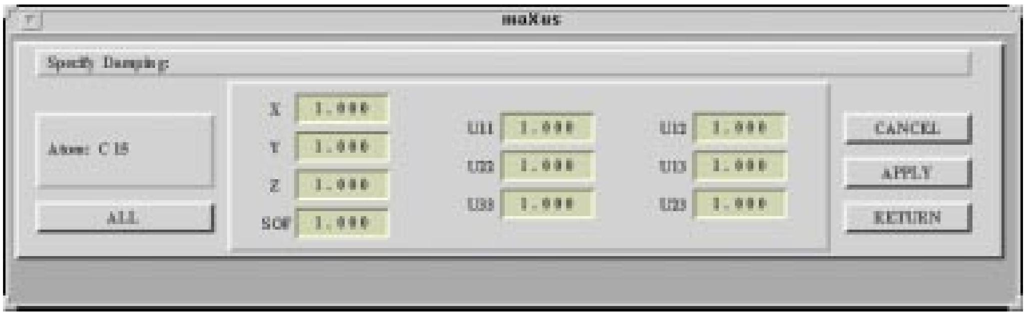


All atoms by default will be refined with isotropic thermal parameters, with U(iso) initially set to 0.05. However, when the user wishes to refine specific atom types anisotropically, this can be achieved by simply clicking on one atom belonging to that type. When an atom is selected, the ISO/ANISO buttons will reveal the current thermal parameter status. To change the status, simply press ATOM TYPE and then ISO or ANISO as required. Anisotropic atoms may be switched to isotropic and vice versa. The same process can be carried out for individual atoms by pressing the ATOM setting. To verify the decision, the APPLY button is pressed. The CANCEL and DEFAULTS buttons have their usual meaning.

To set all the non-hydrogen atoms anisotropic and refine them, the use of the *Auto/Aniso* option in the main maXus menu is recommended.

12.4.14 DAMP

A damping factor can be applied to any parameter of the refinement, except for the scale factors or the extinction parameter. The shift on the parameter during refinement is multiplied by the damping factor, which defaults to 1.0 (i.e. no damping). To apply a damping factor, click on an atom, alter the numbers corresponding to the desired parameter(s), then click APPLY (or CANCEL). Select another atom, or click RETURN to finish setting damping. Note that damping is very seldom needed, and users should be aware that a heavily damped refinement may not produce meaningful results.



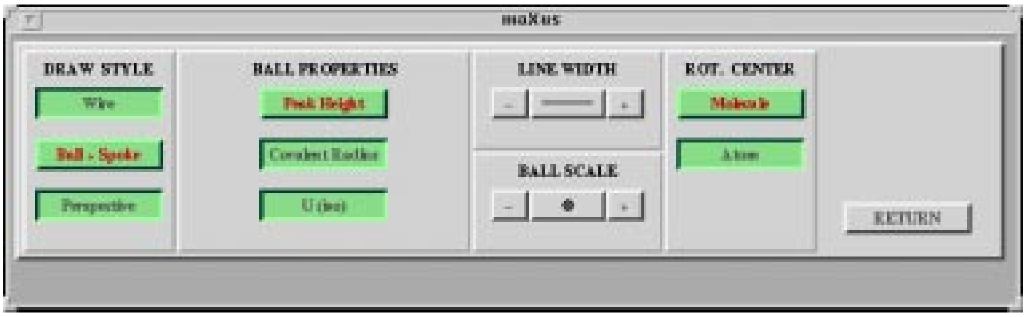
Specify Damping:

Atom: C 15

X 1.000
Y 1.000
Z 1.000
SOF 1.000
U11 1.000
U22 1.000
U33 1.000
U12 1.000
U13 1.000
U23 1.000

12.4.15 DISPLAY

This function allows the user to select one of several display modes. The figure below shows the control panel and the various options.



There are three basic molecule display modes: Ball & Spoke, Wire or Perspective. models.

The default mode is *Ball & Spoke*. When this mode is selected, there are three ways in which the ball radii can be scaled:

1. Peak Height - This is the default mode and the ball radii are scaled to the maximum peak height of the atoms from the structure solution process.
2. Covalent radius.
3. U(iso) i.e. scaled to the isotropic temperature factor. Atoms which have negative, or non-positive definite temperature factors will have '?' in them. Atoms with unusually large thermal parameters will appear with large spheres. These are the atoms which are probably wrongly assigned:

- The ? means that the atom should have an *increased* atomic number.
  - Large spheres mean that the atom should have an *decreased* atomic number.

Return to the model window and change the atom types by clicking on correct element in the periodic table, and then on the atom whose type is to be changed. Also, check the geometry around suspect atoms using the distance/angle options.

In addition, the user can select the bond radius using the line width adjuster in the line width box.

The **Wire option** draws simple wire diagrams. These are the quickest to manipulate, and for large complex molecules they can often be the clearest method.

Perspective is the most elaborate method. Using only native X graphics, it displays:

- Tapered bonds.
- Depth cueing.
- Shading and lighting.

This is the slowest to manipulate. Atom sizes are selected by the program.

It is possible to rotate around the molecule centroid or a given atom. For the latter option click on **Atom** then the required atom. You can rest this using the Molecule button.

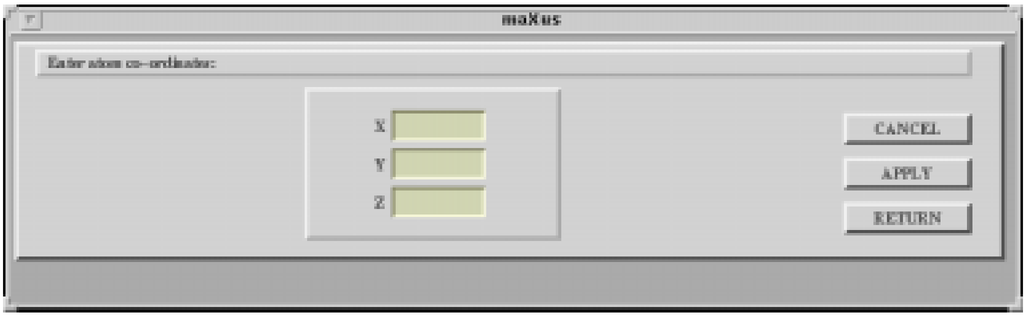
12.4.16 FRAGMENTS

This option allows you to look at only specific fragments or individual molecules when there is more than one molecule in the asymmetric unit. The dialog box tells you the number of fragments and allows you to select them using the alphameric boxes.

Fragments: 1

12.4.17 ADD ATOM

This option allows atomic coordinates to be input directly into maXus. A dialog box appears:

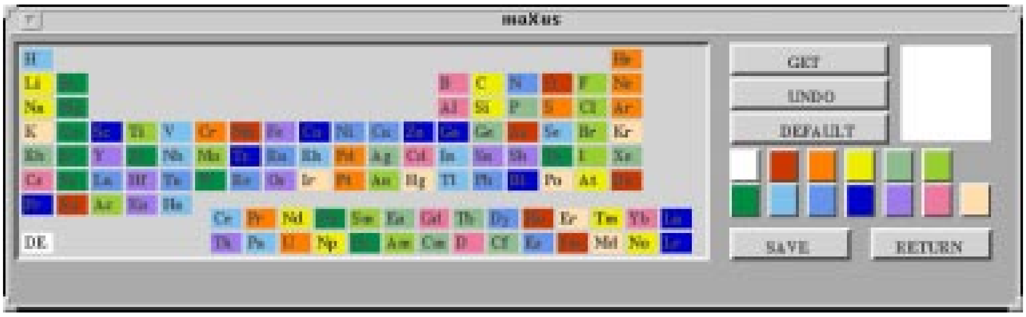


Enter atom co-ordinates:

Type in the coordinates as requested.

12.4.18 COLOUR SELECT

The button on the right side of the Periodic Table display allows the colours assigned to various elements to be altered. The colour select menu has the following appearance:



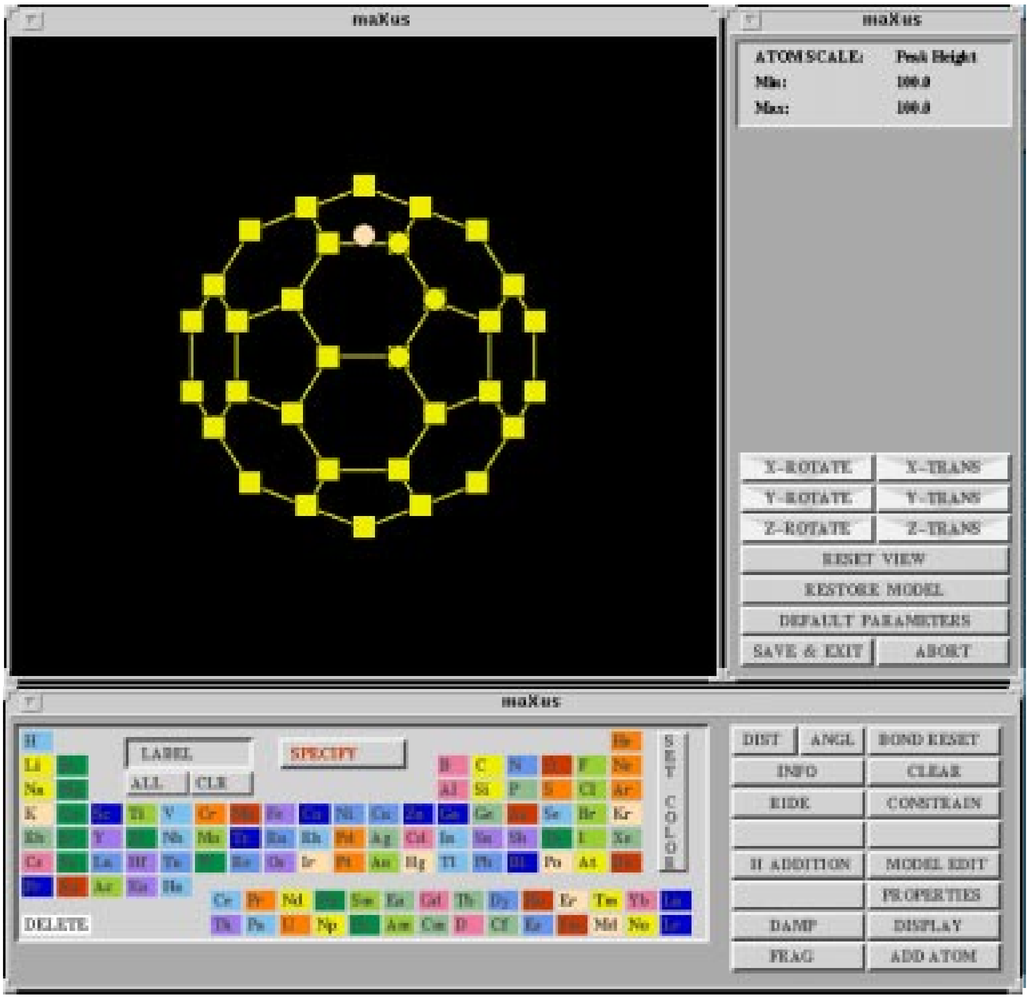
This option is used to customise the colour palette for the atoms. To change an atom colour, select an appropriate colour from the selection on the right. It will then appear in the top right box. Now click the corresponding box in the periodic table for the atom whose colour you wish to change, and click SAVE. UNDO undoes this effect. If you do not use SAVE, your changes are only temporary.

12.5 Expanded Model Window

12.5.1 Introduction

Select *Display Expanded* from the *Model* menu:

At first glance, the expanded display is very similar to the standard display:



Atoms generated by the program using symmetry are square, rather than the usual circles. Changing the label or atom type of any one atom should change all of the symmetry generated atoms automatically.

The majority of commands should work the same in this window as they did in the standard window. However, there are a few key differences. In particular, unlike the standard display, it is not possible to select Planar Ring or Rigid Group. This is because these features are too complex to be used on combinations of symmetry unique, and symmetry generated atoms.

The expanded model wites a file (*model.expand*) containing the symmetry-added atoms.

The results of the expansion can be viewed using the *Expand* option from the *Summary* menu (see Section 19.1).

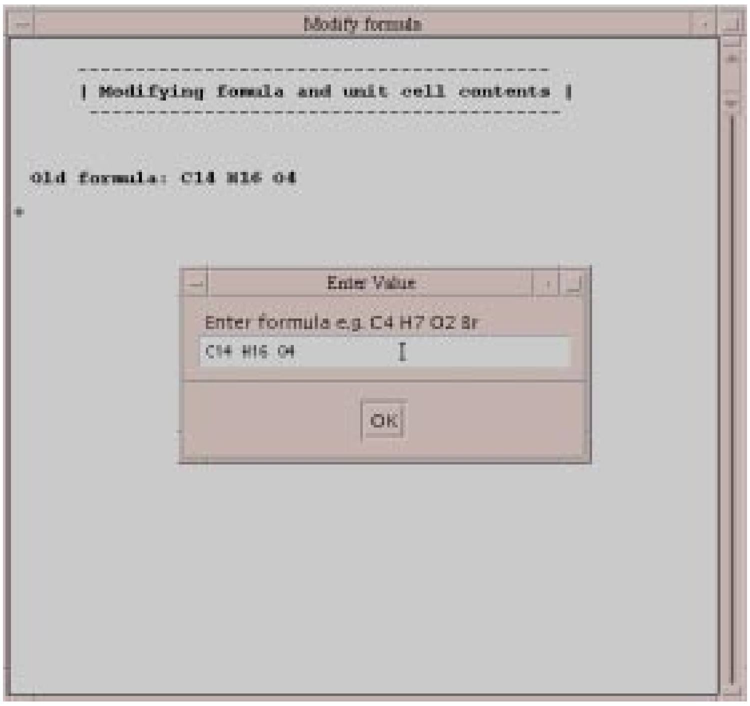
The expanded model file can be viewed in several of the graphics packages supplied with maXus, such as Ortep, RasMol, XXMol, and Pluto. Simply select the *View Expanded* option in the relevant graphics menu. See the Graphics chapter for further details.

12.6 Change Formula program

This option allows the chemical formula of the data set you are working on to be changed at any point. This is useful is you discover atoms present that you were not aware of at the start of the processing.

Select *Change Formula* from the *Model* menu.

The following windows will appear:



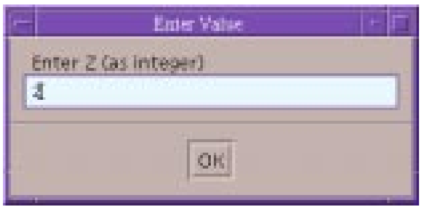
```
 ------------------------------------------
 | Modifying fomula and unit cell contents |
  ------------------------------------------

Old formula: C14 H16 O4
*
```

Enter formula e.g. C4 H7 O2 Br

C14 H16 O4

Simply enter the new formula in the text box, and click *OK*.

The new formula will appear in the display, and you are then asked for a value for Z:



Enter Z (as integer)

4

Enter it, and click *OK*.

The new calculated formula molecular weight, μ, and ρ_{calc} will be displayed in the text pane. You will then be asked:

Is this OK?

If the details are correct, click *Yes*. You are informed:

Files ammended for new cell contents

If the details were incorrect, click *No*. The program then asks you:

Do you want to quit?

Answering *Yes* to this question returns control to the main maXus menu.

Answering *No* starts the process over again.

13 Graphical Least Squares

and Calculation of Structure Factors

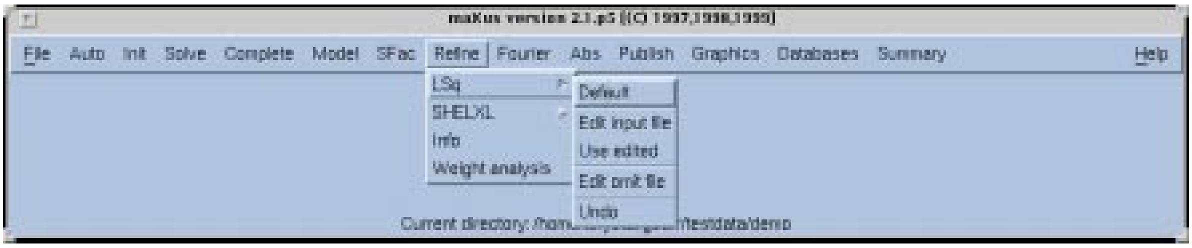
13.1 Introduction

There are two least squares refinement modules available in maXus: Lsq and SHELXL (although the latter is not supplied as standard - you need to have a separate licence). This chapter describes the Lsq program. SHELXL is described in the following chapter.

Structure factor calculations and least squares are both carried out by the same module in the maXus program. The program is nearly always called after MODEL.

13.2 Refine Menu

Selecting *Refine* from the main menu presents the following refinement options:



- Lsq - Least Squares program - see Section 13.3.
- SHELXL refinement program - see Chapter 14.
- Info - Graphical representation of Data - see Section 13.7.
- Weight Analysis - see Section 13.8.

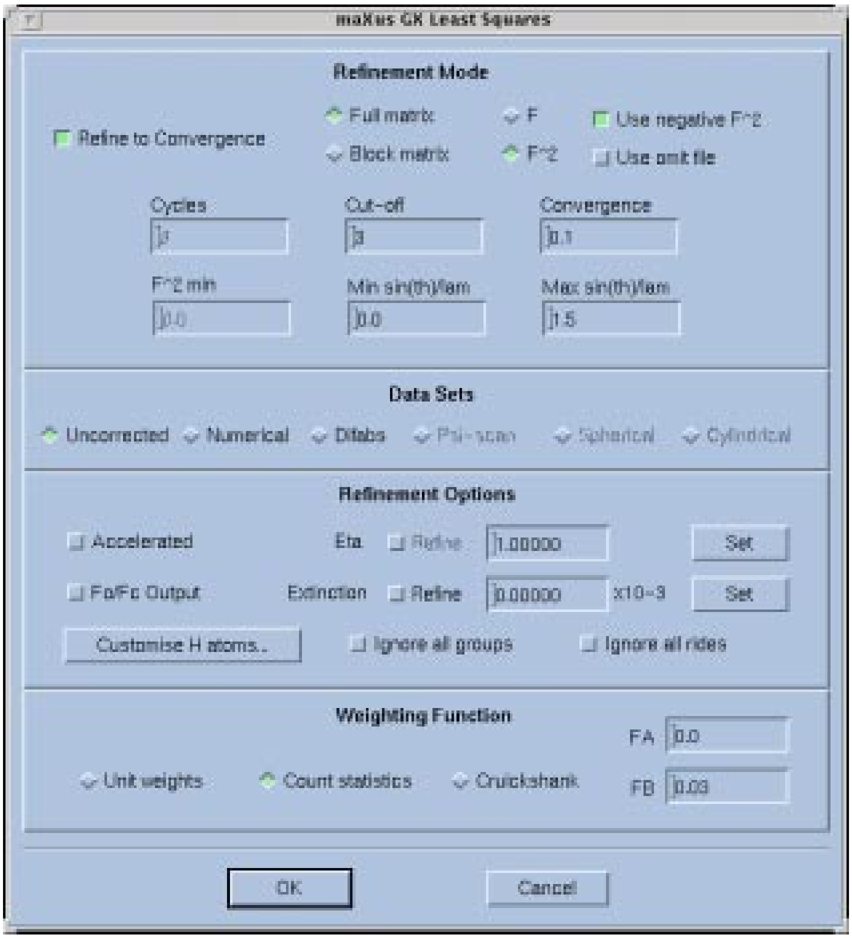
13.3 LSQ

Selecting *Lsq* from the *Refine* menu produces a list of further options:

- Default - see Section 13.3.1
- Edit Input File - see Section 13.5
- Use Edited - see Section 13.5
- Edit Omit File - see Section 13.9
- Undo - see Section 13.6

13.3.1 Default

Selecting *Default* causes the following window to appear:



maXus GX Least Squares

Refinement Mode

☑ Refine to Convergence
● Full matrix
○ Block matrix
○ F
● F^2
☑ Use negative F^2
☐ Use omit file

Cycles: [illegible]
Cut-off: 3
Convergence: 0.1
F^2 min: 0.0
Min sin(th)/lam: 0.0
Max sin(th)/lam: 1.5

Data Sets

● Uncorrected ○ Numerical ○ Difabs ○ Psi-scan ○ Spherical ○ Cylindrical

Refinement Options

☐ Accelerated
Eta ☐ Refine 1.00000 Set
☐ Fo/Fc Output
Extinction ☐ Refine 0.00000 x10-3 Set
☐ Ignore all groups
☐ Ignore all rides

Weighting Function

○ Unit weights ● Count statistics ○ Cruickshank
FA: 0.0
FB: 0.03

Note that all the necessary constraints and restraints, atom assignments, damping factors etc. have previously been set using the MODEL module. See the relevant section of the manual for further details. The options contained here are as follows:

13.3.2 Window Parameters: Refinement Mode

- Refine to Convergence: with this option on, the number of refinement cycles entered (see below) is overridden, and refinement continues until the convergence criterion is met.
- There are two main modes of operation: *Full matrix* and *Block matrix*. The former uses the full least-squares matrix, whereas the latter uses blocks containing the positional and thermal parameters with the scale factor for each atom. This latter is a much faster option, and can be used in the early stages of refinement. It is not recommended as a final refinement option.

- It is possible with full matrix refinement to refine on F or F^2. If you select F^2, you can then decide if negative values of F^2 should be included; the default for this button is on. (It is not possible to use F^2 with block diagonal least squares.). There is a continuing dispute as to which of F or F^2 is the best option. Be careful with F^2 if your data set contains a lot of weak reflections. You can include zero and negative intensities with this option.

  If you attempt to choose block matrix and F^2 a dialogue box will warn you:

Cannot refine on F squared with block

The other parameters are as follows:

- Cycles: The number of least squares cycles. The default is 3. The program will, however, terminate if the convergence criterion (see below) is met.
- Cut Off: Let this parameter be s. All reflections having $I<s.\sigma(I)$ are excluded. In terms of F this takes the form $F<2.s.\sigma(F)$. The default is 3.
- Convergence: If the maximum shift/error for all the refined parameters is less than this value, then least squares terminates. *This applies to full matrix least squares only.* The default value is 0.1.
- F^2min: All unscaled reflections with |F| less than this parameter are excluded from the least squares refinement. The default is 0.0. If you choose to include negative F^2, then this box is greyed out and cannot be used.
- Min sin(th)/lam: This is the minimum value of sinθ/λ for which data will be accepted. The default is no limits.
- Max sin(th)/lam: This is the maximum value of sinθ/λ for which data will be accepted. The default 1.5.
- Use Omit File button - default is off. You are given the option to create an Omit file when running a Weighting Scheme Analysis - see Section 13.8 - or you can enter reflection indices manually.

13.3.3 Window Parameters: Data Sets

- It is possible to choose from 6 possible input data sets for refinement:

(1) Data uncorrected for absorption.

(2) Data corrected for absorption using the numerical correction program (Chapter 6.2)

(3) Data corrected for absorption using DIFABS.(Chapter 6.7)

(4) Data corrected for absorption using PSI-SCAN.(Chapter 6.4)

(5) Data corrected for absorption using SPHERICAL.(Chapter 6.5)

(6) Data corrected for absorption using CYLINDRICAL(Chapter 6.6)

Select the option you require: (1) is the default. You can only select a box if you have already run the relevant absorption correction program. The program remembers your last choice.

13.3.4 Window Parameters: Refinement Options

- Accelerated Button: For full matrix refinement, the accelerated option may be chosen: this re-uses the least-squares matrix from cycle to cycle. Some care is needed with this, but it can be helpful with very large structures that use a lot of computer time. Default is off.
- Fo/Fc Output: Set this box to generate the output tables for structure factor tables. It is also needed for the weight analysis program. Default is off.
- The ETA Refine button can be pressed to initiate η-parameter refinement. This is used to check if your enantiomorph has been correctly assigned. Refine the structure with this parameter set, then look at the η-value:

  η=1.0 (approx.) means the enantiomorph is correct.

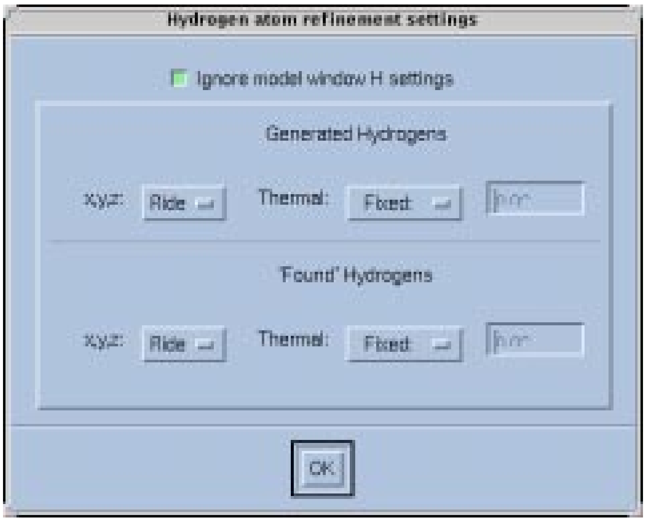
  η=-1.0 (approx.) means that the enantiomorph should be inverted using the MODEL options.

  η=0.0 for enantiomorph cannot be defined

  The η parameter can be set by entering a suitable value in the Eta box, and clicking the Set button. Refinement is only available in full matrix least squares, although the block matrix program can use a value defined by the user or set by the full matrix least squares program. Note that if your structure is centrosymmetric, this option is not available.
- Extinction: An isotropic extinction parameter may be refined by clicking the Extinction box. If numerical absorption corrections have been carried out, then the mean path length of the X-ray in the crystal will be used; if not a value of 10^{-3} will be used. The extinction parameter can be set by entering a suitable value in the Extinction box and clicking the *Set* button. Refinement is only available in full matrix least squares, although the block matrix program can use a value defined by the user or set by the full matrix least squares program. A starting value of 0.001 is suitable.

13.3.5 Window Parameters: Customise H Atoms

Clicking this button brings up the following dialog box:



Generated Hydrogens: x,y,z: Ride Thermal: Fixed

'Found' Hydrogens: x,y,z: Ride Thermal: Fixed

maXus recognises 2 sorts of H atoms:

- Those generated in theoretical positions,
- Those found in Fourier maps.

Lsq normally takes its H-atom settings from the MODEL program; however you can override these settings when the *Ignore Model Window H Settings* button is on.

For both generated hydrogens, and found hydrogens, you can set:

- *Free*: The coordinates are to be refined - i.e. refined separately of their parent atoms, or

  Ride: i.e. not refined separately, and can be considered only to be riding on their parent atom.
- The thermal parameter can be *Fixed* i.e. not altered by refinement, or *Free* to be refined, or set to the value of the *Parent* atom, plus a user-defined offset entered in the adjacent text field.

Two other MODEL parameters can be overridden from the main LSQ window:

- *Ignore Groups* button - if this is on, any settings you have made in MODEL to define areas of the molecule as groups are ignored, and such regions are calculated as individual atoms.
- *Ignore All Rides* button - if this is on, any atom parent/child relationships you have defined in the MODEL program are ignored during refinement.

13.3.6 Window Parameters: Weighting Functions

- Three weighting schemes are provided:

  (1) Unit weights

(2) Counting statistics weights of the form:

$$w=exp[FA.sin^2\theta/\lambda^2]/[\sigma^2(F_o)+FB.F_o^2]$$

where *FA* is a user defined variables entered in the relevant text box. The default value is 0 which gives simple counting statistics weights. For non-zero values of *FA* this scheme progressively up weights the higher angle reflections. Typical non-zero values of *FA* are usually in the range 10-25$A^{2.}$ Usually, a non-zero value of *FB i*s given [0.03 by default], this may need adjusting.Under the defaults, this weighting scheme reduces to:

$$w=1.0/[\sigma^2(F_o) + 0.03.F_o^2]$$

This scheme is the default.

When refining on F^2, the GUI prevents the an incorrect choice of weighting scheme - only Count Statistics or Unit Weights are available.

(3) A Cruickshank scheme of the form:

$$w=1/(1 + ((Fobs-FB) / FA)^2)^{2TANA}$$

Suitable values can be found by inspecting the results of the *Weight Analysis* program (See Section 13.8)**.**

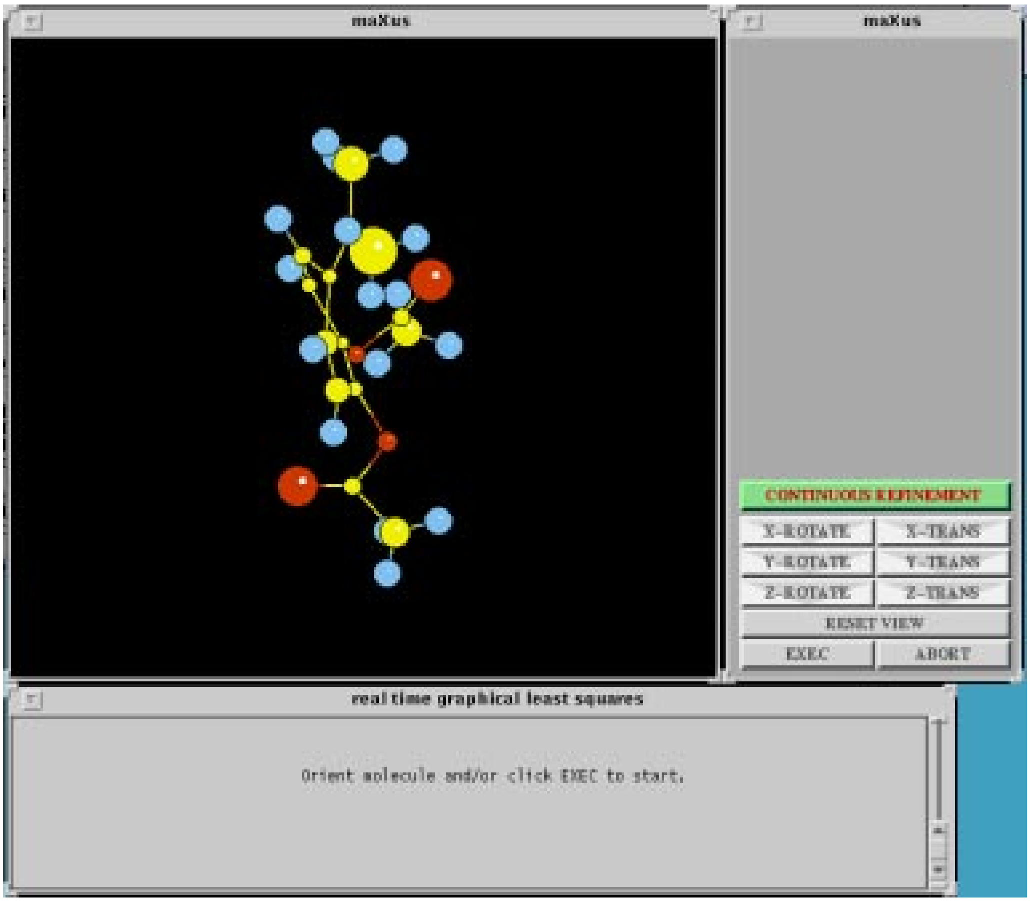
Do not use zero for FA when using the Cruickshank scheme.

13.3.7 Running the Program

Once everything has been set click *OK*.

A new progress indicator appears:

This is followed by the main LSQ process windows:



Orient molecule and/or click EXEC to start.

- Orient the molecule to useful view using the orientation buttons in the normal way
- You can choose from continuous refinement, or non-continuous refinement using the *CONTINUOUS REFINEMENT* button. In the latter case you can abort after any cycle by pressing the EXIT button; in continuous refinement this is not possible. **Note** that this is *not* the same as Refining to Convergence.
- Messages are displayed in the text pane.
- Once ready to begin, click on *EXEC*.

13.3.8 The Output

The performance monitor indicates the flow through the program. An initial structure factor calculation is always carried out to get the scale factor correct. The graphics window displaying the molecule is updated at every cycle. The sphere sizes are proportional to the isotropic temperature factors (or the equivalent in the isotropic case); atoms which go non-positive definite or for which Uiso<0.0 are marked with a question mark. The text window outputs a summary of the relevant statistics.

You can abort at the end of a refinement session (or after a given cycle if you are not in continous refinement mode) by clicking the *ABORT* button, in which case the results of the refinement are not kept.

Press the *EXIT* button to finally exit, and save the refinements. To view a full output choose *Lsq* from the *Summary* menu in the main maXus window.

13.3.9 Errors and Warnings

The program may generate an error message if there is a problem.

If no reflections greater than the cut-off limit were found, then a message is displayed:

No reflections > limits

If the R-factor increases by more than 0.1:

R factor increase > 0.1

It is advisable to *Quit* and recheck the structure, although it is possible to continue if you wish.

If the full matrix used is singular:

Singular Matrix

If this occurs, check for:

- Atoms very near special positions which have not been detected by the program.
- Highly correlated variables.

If the extinction refinement has given a negative mosaic spread:

Extinction refinement gives -ve r*

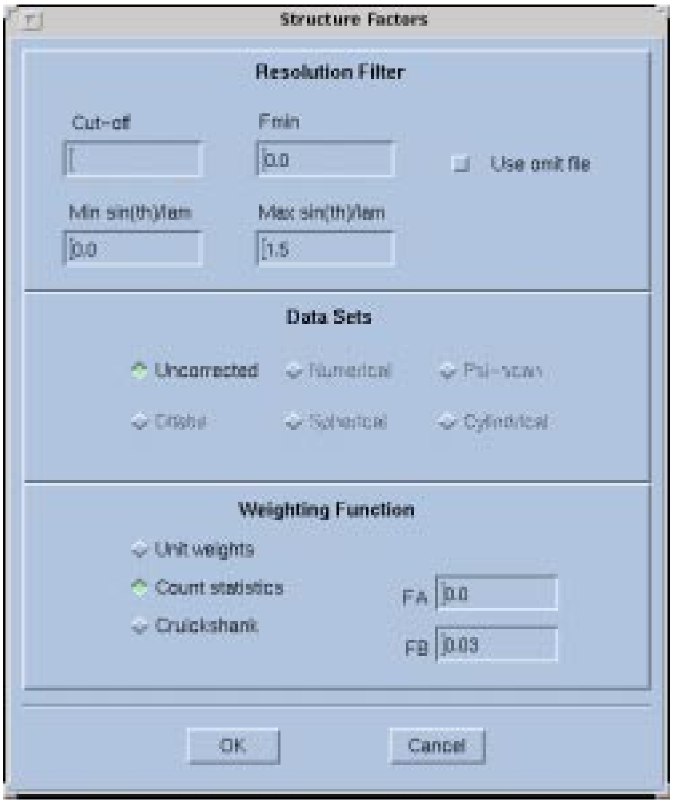
This is physically impossible, and may occur if the numerical absorption correction program has not been run, and all the mean path lengths have been set to 10^{-3} mm. If you click *Reset* the parameter is set to 0.0 and extinction refinement is switched off. If you click *Quit* then the program aborts.

13.3.10 Atoms in special positions

maXus correctly computes and imposes constraints on positional and thermal parameters for atoms in special positions. It is not necessary to impose such constraints manually.

13.4 Structure Factors

SFac is a program to carry out a structure factor calculation without invoking least squares. To invoke this module, select the *Structure factors* entry from the *SFac* menu. The following dialogue box appears:



Structure Factors

Resolution Filter

Cut-off:

Fmin: 0.0

☐ Use omit file

Min sin(th)/lam: 0.0

Max sin(th)/lam: 1.5

Data Sets

● Uncorrected ○ Numerical ○ Psi-scan
○ Difabs ○ Spherical ○ Cylindrical

Weighting Function

○ Unit weights
● Count statistics
○ Cruickshank

FA: 0.0

FB: 0.03

OK Cancel

The parameters are as follows:

13.4.1 Resolution Filter

- Cut Off: Let this parameter be s. All reflections having $I<s.\sigma(I)$ are excluded. In terms of F this takes the form $F<2.s.\sigma(F)$. The default is 3.
- Fmin: All unscaled reflections with |F| less than this parameter are excluded from the least squares refinement. The default is no cut off.
- Min sin(th)/lam: This is the minimum value of $\sin\theta/\lambda$ for which data will be accepted. The default is no limits.
- Max sin(th)/lam: This is the maximum value of $\sin\theta/\lambda$ for which data will be accepted. The default is 1.5.
- Use Omit File button - default is off. You are given the option to create an Omit file when running a Weighting Scheme Analysis - see Section 13.8.

13.4.2 Data Sets

- It is possible to select from 6 possible input data sets for refinement:

(1) Data uncorrected for absorption.

(2) Data corrected for absorption using the numerical correction program (Chapter 6.2)

(3) Data corrected for absorption using DIFABS.(Chapter 6.7)

(4) Data corrected for absorption using PSI-SCAN.(Chapter 6.4)

(5) Data corrected for absorption using SPHERICAL.(Chapter 6.5)

(6) Data corrected for absorption using CYLINDRICAL(Chapter 6.6)

Select the option you require: (1) is the default. You can only select a box if you have already run the relevant absorption correction program. The program remembers your last choice.

13.4.3 Weighting Functions

- Three weighting schemes are provided:

  (1) Unit weights

  (2) Counting statistics weights of the form:

  $w=exp[FA.sin^2\theta/\lambda^2]/[\sigma^2(F_o)+FB.F_o^2]$

  where *FA* is a user defined variables entered in the relevant text box. The default value is 0 which gives simple counting statistics weights. For non-zero values of *FA* this scheme progressively up weights the higher angle reflections. Typical non-zero values of *FA* are usually in the range 10-25A$^{2.}$ Usually, a non-zero value of *FB i*s given [0.03 by default], this may need adjusting.Under the defaults, this weighting scheme reduces to:

  $w=1.0/[\sigma^2(F_o) + 0.03.F_o^2]$

  This scheme is the default.

  (3) A Cruickshank scheme of the form:

  $w=1/(1 + ((Fobs-FB) / FA)^2)^{2TANA}$

  Suitable values can be found by inspecting the results of the *Weight Analysis* program (See Section 13.8)**.**

  Do not use zero for FA when running Cruickshank.

13.4.4 Running Structure Factors

When you are happy with all of the settings, click *OK*.

A progress indicator will appear while the program is running:

Once it is complete, the program terminates:

SFac Complete

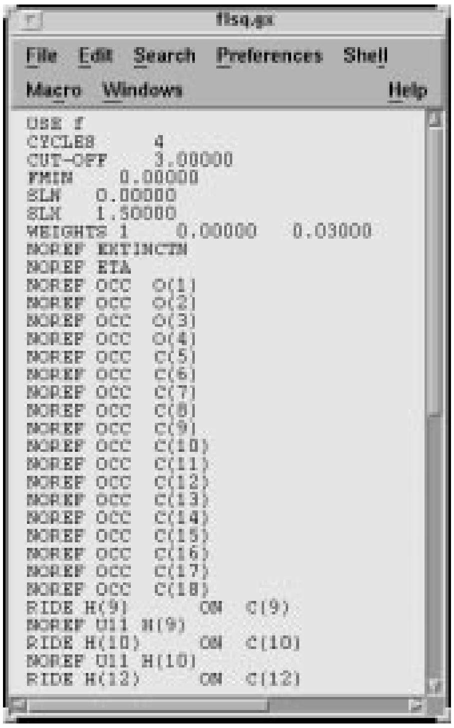
Output from the program may be examined by selecting *SFac* from the *Summary* menu in the main maXus window.

13.5 Using the *Edit Input File* and *Use Edited* Menu Options

13.5.1 Edit Input File

Not all the features you want in least-squares can be easily input using a graphical user interface. maXus offers the more experienced user a facility for controlling the refinement using line-by-line commands.

Under the *Refine* menu entry select *LSq*, and then *Edit input file*. The standard LSQ main window box will appear. Make your choices and click *OK*. A Text Editor box will now appear:



```
USE f
CYCLES      4
CUT-OFF     3.00000
FMIN     0.00000
SLN    0.00000
SLX    1.50000
WEIGHTS 1     0.00000    0.03000
NOREF EXTINCTN
NOREF ETA
NOREF OCC  O(1)
NOREF OCC  O(2)
NOREF OCC  O(3)
NOREF OCC  O(4)
NOREF OCC  C(5)
NOREF OCC  C(6)
NOREF OCC  C(7)
NOREF OCC  C(8)
NOREF OCC  C(9)
NOREF OCC  C(10)
NOREF OCC  C(11)
NOREF OCC  C(12)
NOREF OCC  C(13)
NOREF OCC  C(14)
NOREF OCC  C(15)
NOREF OCC  C(16)
NOREF OCC  C(17)
NOREF OCC  C(18)
RIDE H(9)       ON  C(9)
NOREF U11 H(9)
RIDE H(10)      ON  C(10)
NOREF U11 H(10)
RIDE H(12)      ON  C(12)
```

Now edit this box to your requirements, and exit the editor. See Appendix A to this chapter for details of all the Least-Squares commands.

13.5.2 Use Edited Menu option

Now click on this option. The Model window will appear directly and the least-squares procedures will run as usual, using the edited input file.

13.6 Undo

Every time you run least-squares, there is a possibility that the calculations will not proceed as you wish. To allow for this, maXus makes backup copies of critical files in a backup sub-directory each time least-squares is run. Selecting the *UNDO* option in the *LSq* submenu of *Refine* replaces the results of a bad least squares with those of the previous run. A dialogue box appears:

Revert to state prior to last Least Squares?

Make your decision. If the backup is successful, then a dialogue box appears:

Data restored

If there is no backup present, then the following dialogue box will appear:

Data has not been backed up

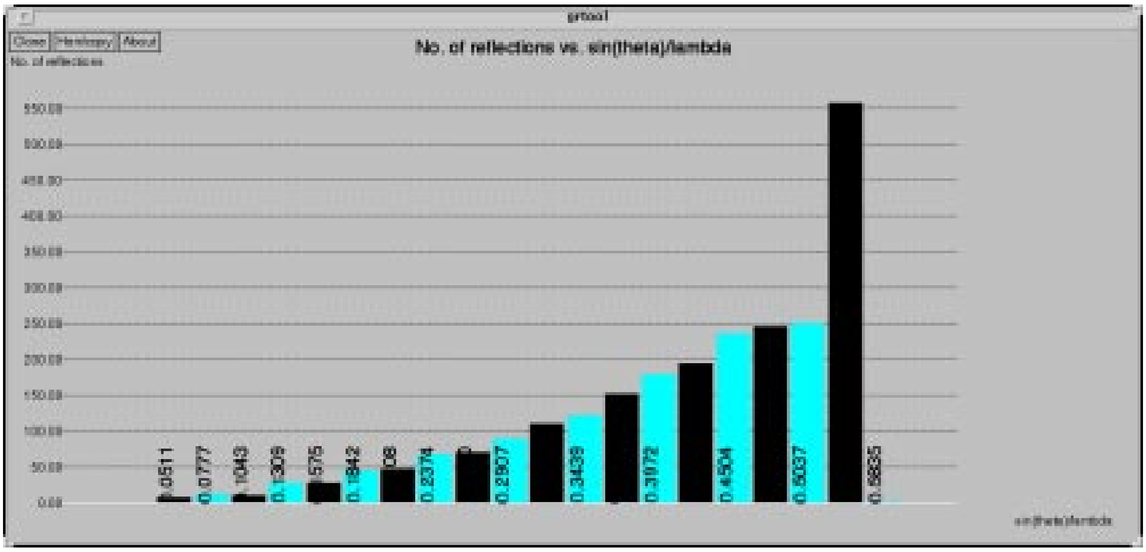
Note that the *UNDO* function does not currently work for SHELXL.

13.7 Info

Selecting this option from the *Refine* menu brings up 2 histograms of your data:

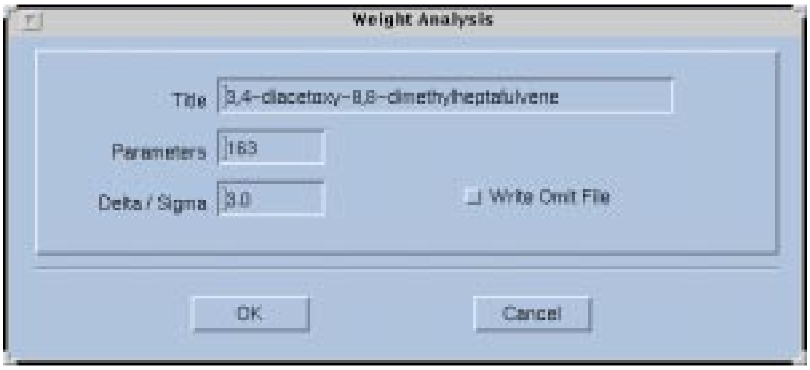
- Number of reflections vs. sin θ/λ
- Number of reflections (x10^3) vs. F_{obs}.

For example:



13.8 Weighting Scheme Analysis

Choose *Weight Analysis* from the *Refine* menu. This opens the weighting analysis window:



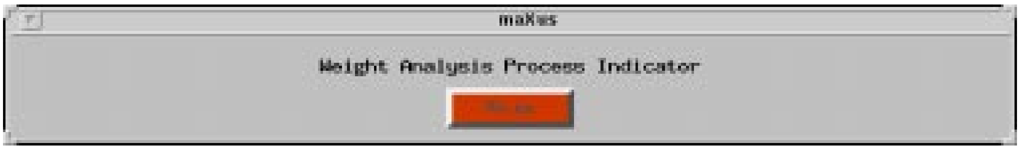
Title: 3,4-diacetoxy-8,8-dimethylheptafulvene
Parameters: 163
Delta / Sigma: 3.0
☐ Write Omit File

The parameters that the user can set here are:

- The number of parameter used in refinement. A suitable default is taken from your last least-squares, but this may be incorrect.

- The Delta / Sigma ($\Delta F/\sigma(F)$) ratio. Any reflections with ratios exceeding this value are listed in the text window and on the output file. This parameter may not exceed 10. You should not have very many reflections listed when using the default.
- Write Omit File button - if this is on, any reflections exceeding the $\Delta F/\sigma(F)$ are written to it. The next time you run the LSq program, you can set it to ignore the reflections listed in this file by selecting the *Use Omit File* option in the main dialog box.

When the settings are complete and you are ready to proceed, press *OK*. A progress indicator will appear.



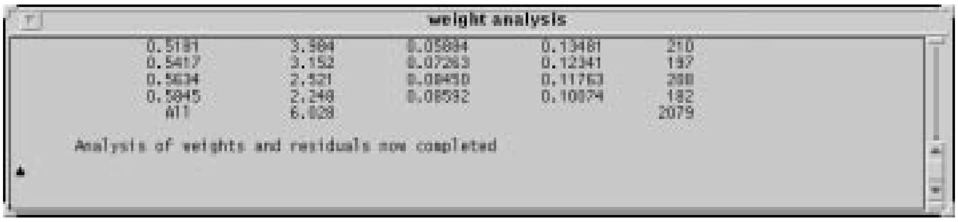
Weight Analysis Process Indicator

If you have not already run least squares, then a dialog box will appear:

LsqRpt doesn't exist - run Lsq

Click *Quit* to exit.

If there are no problems, a text pane appears with a summary of the results in table form:



```
          0.5181        3.384       0.05884       0.13481      210
          0.5417        3.152       0.07283       0.12341      197
          0.5634        2.521       0.08450       0.11763      208
          0.5845        2.248       0.08592       0.10874      182
             All        6.028                                2079

     Analysis of weights and residuals now completed
```

A dialogue box then appears:

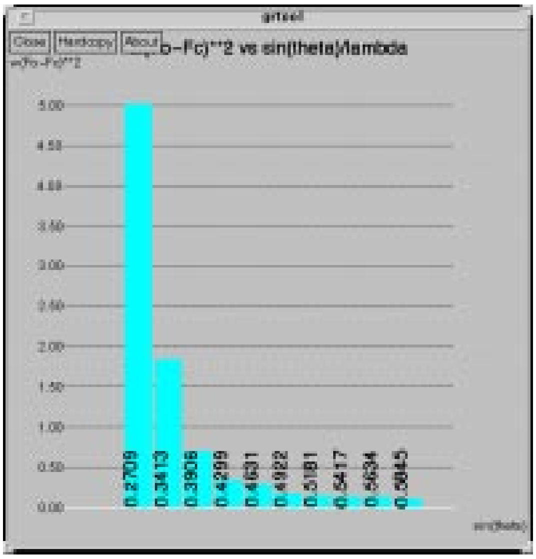
Wait for histograms

Click OK and six separate graphs will be displayed, showing relationships as follows:

- R-factor versus $\sin\theta/\lambda$

- w(Fo-Fc)^2 versus sinθ/λ
- R versus hkl parity
- w(Fo-Fc)^2 versus hkl parity
- w(Fo-Fc)^2 versus Fo
- R versus Fobs

The example shown below illustrates the layout of the graphs, with the range in the x-axis which is covered by each bar printed along the bar itself:



Finally, another dialogue box will appear:

Weight analysis complete

Click *OK*, and the output windows will close. The text output can be reviewed using the *Weight Analysis* entry in the *Summary* menu

13.9 Edit Omit File

If you have run Weight Analysis, and chosen to write an Omit File, this file can be edited here.

Selecting *Edit Omit File* from the *LSq* submenu of *Refine* brings up a standard text editor, where you can add, remove or edit reflections. If you select this option, but no Omit file exists, you will be given the opportunity to create one. These reflections can be input to the least squares program, where they are omitted from the refinement process.

13.10 Current Limitations on the Number of Atoms etc.

The current release has the following limits:

- 500 atoms
- 2000 refinable parameters. If this is too few, partition the molecule(s) into two or more fragments.

Appendix A

The Least-Squares Commands

- CYCLES Number of refinement cycles. However, there is more to it than this. It is possible to split a large calculation into blocks using the CYCLES / REF / NOREF sequence. The second appearance of this command causes the program to run with the commands it has been given up to that point- those commands appearing after the second cycles command are not read until the first part of the refinement has concluded.
- REF SCALE/EXTINCT/ETA/X/Y/Z/OCC/UIJ/ALL Atom label(s)

  These are the refinement flags which switches on the refinement of any parameter.

  SCALE is the scale parameter

  EXTINCT is the extinction parameter.

  ETA is the η parameter.

  X is the x-parameter, Y the y-parameter etc.

  OCC is the occupancy.

  UIJ is the single isotropic thermal parameter or all 6 U_{ij}'s for an anisotropic atom. You can refer to individual anisotropic components as U11, U22, U33, U12, U13, U23 if you wish.

  ALL switches on x, y, z, and isotropic or anisotropic refinement.

  Atom labels: If a single atom label is included here then this command refers to just that atom. If you put in 2 atom labels, then all the atoms between these to and including them are used. (Don't forget to give the full labels including the brackets.) If you want to see the order of the atoms in your list, it is best to use the Graphics/Ortep/Display model options to show this.

  By default, eta and extinction refinement are off.

  e.g.

  REF ALL C(1) C(6)

  will refine all the atomic coordinates and temperature factors for atoms between C(1) and C(6) in the list.
- NOREF SCALE/EXTINCT/ETA/X/Y/Z/OCC/UIJ/ALL Atom label(s)

  Just as for the REF command but it stops refinement.

  e.g.

  NOREF Y N(2)

  will switch off the refinement of the y coordinate of atom N(2)
- Example of multiple cycles:

  CYCLES 3

  NOREF Z SN C(2)

  NOREF X N

NOREF Y N

CYCLES 4

REF Z SN C(2)

REF X N

REF Y N

NOREF ALL C(3) C(8)

In this, 3 cycles of least squares are carried out holding the z coordinates of all atoms between SN and C(2) fixed, and the x and y coordinates of atom N fixed. Then 4 cycles of least squares let these parameters refine, and instead the x, y, and z coordinates of the atoms between C(3) and C(8) are held fixed. Further CYCLES commands can be included if you wish.

- GROUP GROUP NAME <CR> ATOM LABELS

  Specifies a group of atoms to be treated as a rigid group.

  GROUP NAME is the name of the group

  <CR > is a carriage control

  ATOM LABELS is a list of the labels of the atoms comprising the group. Up to 20 atoms are allowed. Continuation lines are allowed by using the '=' character (without the quotes).

- RIDE ATOM LABELS(S) ON ATOM LABELS

  Coordinate shifts for the atoms comprising the second set of entries are applied to atoms in the first group. ON is a keyword and must be present. E.g.

  RIDE H(1) ON C(1) rides H(1) on the C atom C(1)

  RIDE H(1) H(3) ON C(1) rides H(1) to H(3) on C(1)

  RIDE H(1) H(3) ON C(1) C(3) rides H(1) on C(1) and H(3) on C(3)

- CONSTRAIN X/Y/Z/OCC/UIJ ATOM LABEL 1 FACTOR X/Y/Z/OCC/UIJ ATOM LABEL 2 (CONSTANT)

  This establishes a dependent variable. The specified parameter of ATOM 1 has its derivative multiplied by FACTOR and added to the derivative of ATOM 2.

  e.g. to constrain the occupancy of C(7) to 1.0 - occupancy of C(8) use the command:

  CONSTRAIN OCC C(7) -1.0 OCC C(8) 1.0

- CUT-OFF s

  All reflections having $I<s.\sigma(I)$ are excluded.

- WEIGHTS scheme number FA FB

  Scheme no 1: Unit weights

Scheme no 2: Counting statistics weights of the form:
$w=exp[FA.sin^2\theta/\lambda^2]/[\sigma^2(F_o)+FB.F_o^2\ ]$
Scheme no 3: A Cruickshank scheme of the form:
$w=1/(1 + ((Fobs\text{-}FB) / FA)^2)^2$

- USE F/FSQ
  Use F or F^2 in refinement.

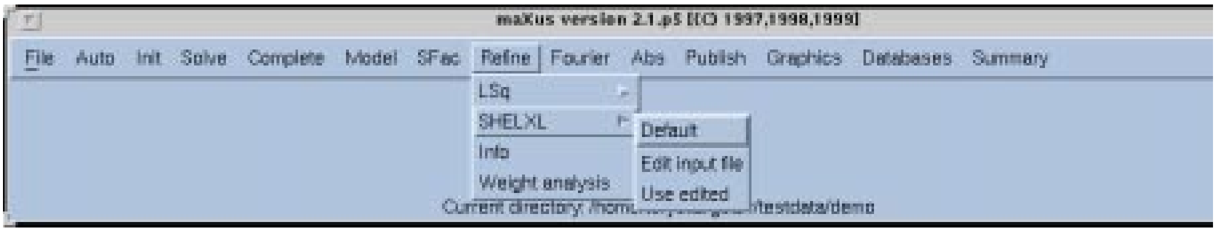
14 SHELXL

14.1 Introduction

SHELXL is a crystallographic refinement program, written by Professor George Sheldrick and is not distributed with maXus. This manual details the maXus - SHELXL interface and will give guidance on how to obtain and integrate SHELXL with the maXus package.

14.2 How to Access SHELXL *via* maXus

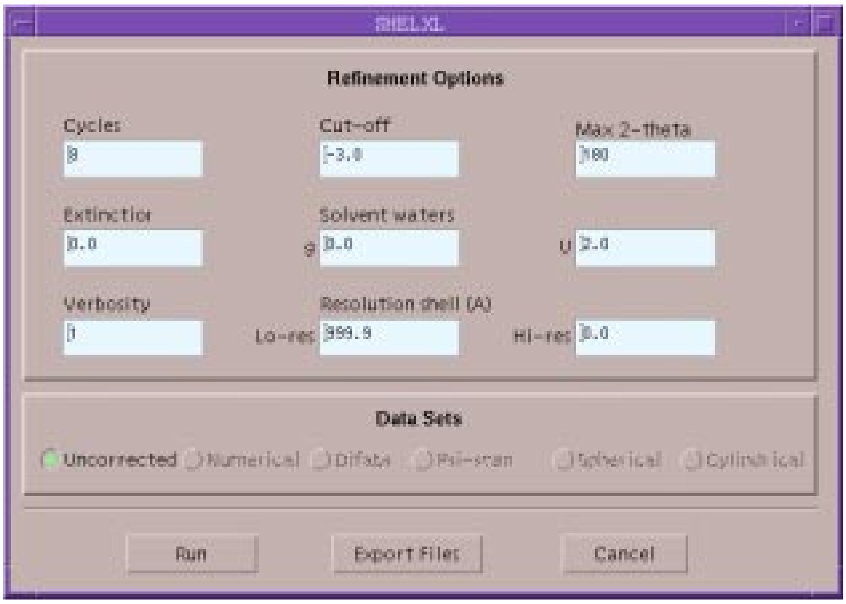
From the main menu:



Choose the *Refine* option, followed by *SHELXL*. This will allow access to the three options for command input and running SHELXL.

14.3 Default Option

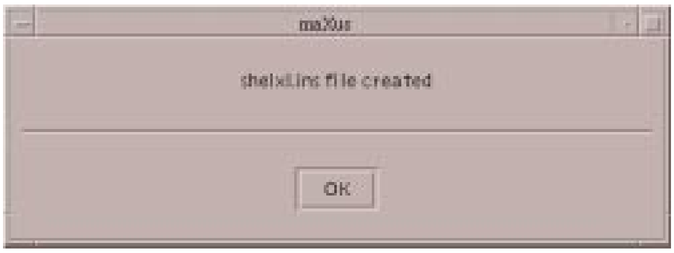
When selected the following window will appear:



Cycles: 8
Cut-off: -3.0
Max 2-theta: 180
Extinction: 0.0
Solvent waters: g 0.0, U 2.0
Verbosity: 1
Resolution shell (A): Lo-res 999.9, Hi-res 0.0

● Uncorrected ○ Numerical ○ Difabs ○ Psi-scan ○ Spherical ○ Cylindrical

Note that it is possible to save a SHELX file without actually having to run the program. This allows structures to be exported to external versions of SHELX. Selecting this option should bring up the following dialog:



shelxl.ins file created

The files will be in the working directory, entitled *shelx.hkl* and *shelx.ins*.

14.3.1 Data Sets

This allows the user to choose which dataset to refine the structure against. Only the data sets in normal type can be selected. The program will automatically decide which absorption corrections have been performed and therefore which data sets are available. These data sets are:

- Uncorrected: The observed structure factor file, with no absorption correction applied.
- Numerical: The structure factors have been corrected for absorption numerically.
- Difabs: The structure factors have been corrected for absorption using the Difabs program.
- Psi-scan: The structure factors have been corrected using the additional data collected using the Psi-scan.

- Spherical: The structure factors have been corrected using a spherical crystal model.
- Cylindrical: The structure factors have been corrected using a cylindrical crystal model.

14.3.2 Refinement Options: Cycles

This determines that number of cycles of refinement to be performed. The default is 8 and the value has a range of 0 - 20.

14.3.3 Refinement Options: Cut-Off

This is the multiplier for the standard deviation value, bellow which any reflection will be flagged as unobserved. i.e. $F_{obs}^2 < -3 \times \sigma(F_{obs}^2)$ will leave all large negative F^2 unobserved. The default for this parameter is -3 and it may lie in the range -3 to 0. *This value may not be positive*.

14.3.4 Refinement Options: Max 2-theta

This defines a maximum value of 2θ above which the reflections are ignored. The default value for this is 180^o.

14.3.5 Refinement Options: Extinction

This will cause an extinction parameter to be refined by least-squares. Default 0.0.

14.3.6 Refinement Options: Verbosity

This will change the verbosity of the final output of the refinement process. The least verbose report is produced using 0 and the most by 4. The default is 1.

14.3.7 Refinement Options: Solvent waters - g

This will refine a factor g that is used in the modification of non-hydrogen scattering factors. The value will be in the range 2-4 for proteins and should refine to zero for small molecules with no diffuse solvent. The default is zero.

14.3.8 Refinement Options: Solvent waters - U

This parameter will not be refines and is there to ensure that only low θ Fs are affected by and SWAT calculation. The default is two.

14.3.9 Refinement Options: Resolution Shell - Low Res Limit

Reflections with a resolution higher than that given in Å will be flagged as unobserved. This is only recommended with macromolecules. The default for this is 999.9Å.

14.3.10 Refinement Options: Resolution Shell - High Res Limit

Reflections with a resolution lower than that given in Å will be flagged as unobserved. This is only recommended with macromolecules. The default for this is 0.0Å.

14.4 Edit Input File Option

Selecting this option from the submenu causes an editor window to be opened which contains the SHELXL command file. The line commands for atomic position and general crystal properties will already be contained in this file when it is opened. Two commands will be inserted by maXus, these are:

- ACTA, produces a CIF file of the refinement
- LIST 3, produces a CIF file containing reflection details.

Both these commands must remain in the command file as the two files produced by them are used in the back conversion from SHELXL to maXus.

14.5 Use Edited Option

Selecting this option from the submenu will run SHELXL using the file previously edited by the user using the *Edit Input File* option.

14.6 End of SHELXL Processing

Once SHELX96 has successfully completed the following window will appear:

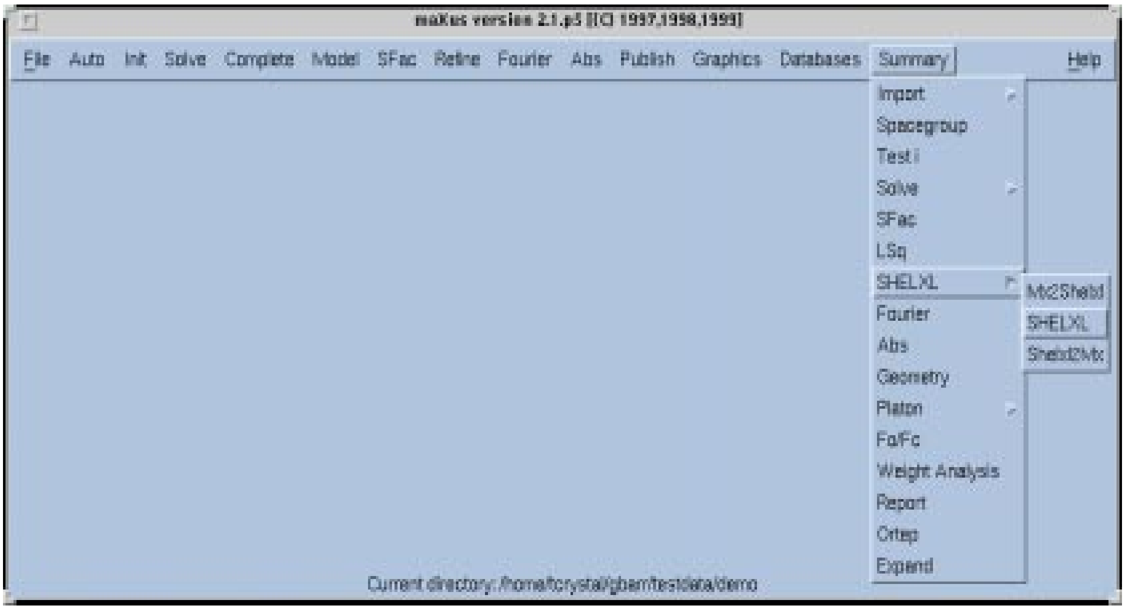
SHELXL complete

Click on OK and maXus will return to the control of the main menu.

14.7 Accessing SHELXL Output

While SHELXL is running, a window containing summary refinement results will be displayed on a cycle by cycle basis. On completion of refinement this summary window will be replaced by a text editor that contains the detailed refinement report from SHELXL. This edit window can be used review the refinement process and can be dismissed in the normal manner.

After the detail report has been dismissed it can be re-accessed if required using from the *Summary* main menu. See below:



Choose the *Summary* option, followed by *SHELXL*. This will cause a sub-menu to appear.

From this sub menu the option *Mx2Shelxl* will access file Sysout_mx2shel. This file contains any error messages produced by the conversion program in converting the maXus data to SHELXL format.

The *Shelxl2Mx* option will access file Sysout_shel2mx. This file contains the error messages produced by the program that back converts the SHELXL results to maXus format.

Both the above files need only be examined when a problem has occurred while running SHELXL.

The sub menu option *SHELXL* will access file shelxl.lis. This file contains the detailed results of the last SHELXL calculation.

14.8 The Performance Meter

While SHELXL is running a performance meter will appear at the top of the screen. When SHELXL has completed the performance meter will be closed. The first area labelled SHELXL will be highlighted while SHELXL is executing. The second area will be highlighted while the conversion of file formats from SHELXL back to maXus is underway. It is possible to close this window while SHELXL is running. The performance meter window is shown below:

14.9 Error Message Windows

Several data and parameter settings will cause error windows to appear during the file conversions between maXus and SHELXL. To remove all the error message windows click on the “OK” button.

14.9.1 MORE Error

If the field for output verbosity is set outside the range 0-4 than the following window will appear and the variable will be reset to 1. SHELXL will continue to run.

MORE was set outside range.This has been reset to 1

14.9.2 Cycles Errors

If the cycles parameter is set greater than 20 then the following window will appear and the number of cycles will be set to 20.

Cycles parameter was set >20. This has been reset to 20

If the number of cycles is set to a negative value then the following window will appear and the number of cycles to be performed will be set to zero.

Cycles parameter was set negative. This has been reset to 0

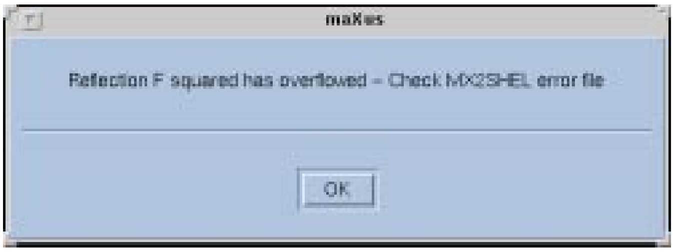
14.9.3 Cut-Off Error

If the Cut-Off parameter is set greater than zero then the following window will appear and the value of the cut off will be reset to -3. SHELXL will not allow a positive value for this parameter as the setting of these reflections to an unobserved status will bias the final refinement result and *Acta Cryst. C* will not accept structures that have been refined in this way.

CUTOFF parameter was positive. This has been reset to -3.0

14.9.4 F^2 Overflow

When the value of F^2 exceeds 9,999,999 then this value can no longer be passed from maXus to SHELXL. When this occurs the following window appears and the reflection is discarded from the SHELXL refinement processing. More information about this error is written to the report CM2SHEL accessible from the summary menu.



Reflection F squared has overflowed - Check MX2SHEL error file

14.9.5 Installation Error

If SHELXL is not installed properly, or maXus cannot find it in the correct directory, an error is generated:

SHELXL97 not installed

See Appendix B of this chapter for details on correct installation.

APPENDIX A

Obtaining a Copy of SHELXL

APPLICATION FORM FOR SHELXL

SHELXL USER REGISTRATION FORM | Do not write here!

Title / Name:| Date Sent:

Full Postal Address:| Version:

Email address (if available):

Tel.:

Fax:

Please tick ALL relevant boxes, sign and return to: Prof. George Sheldrick, Institut fuer Anorg. Chemie, Tammannstrasse 4, D-37077 Goettingen, Germany. Fax: +49 551 393373; E-mail: gsheldr@shelx.ini-ac.gwdg.de

SHELXL is supplied ONLY on MSDOS format diskettes, in the form of self-extracting packed files containing MSDOS executables and sources for UNIX and VMS systems, documentation and test data. The programs PDBINS (PDS to SHELXL format conversion) and CIFTAB (tables from SHELXL CIF output) are included. The licence fee of DM 4999 for for-profit institutions covers use for an unlimited time on an unlimited number of computers at a specified firm or institution at a single geographical site. SHELXL is currently available free of charge to academics for non commercial use only; it may prove necessary to change this policy if the licence fees for for-profit institutions fail to cover the total costs involved. Academic institutions willing and able to contribute to the costs of developing and distributing the SHELX programs are of course welcome to do so (we suggest DM 99). Please make out checks to “Institut fuer Anorg. Chemie, Prof. Sheldrick”. If you wish to pay by direct bank transfer please ask us to send an invoice.

[ ] I wish to license SHELXL for use at the following for-profit firm or institution. I agree that within three months I will either destroy all copies of the program in my possession or pay the license fee of DM 4999.

[ ] The program will be used exclusively for non-commercial purposes at the following not-for-profit institution only.

[ ] Please sent me an invoice for DM.

[ ] Please send me a receipt for the enclosed payment of DM.

[ ] I agree to cite SHELXL in all publications reporting results obtained using it.

[ ] I accept that the author has no liabilities in respect of errors in the program or documentation.

Please supply SHELXL on:

[ ] 1.44 MB [ ] 1.2 MB [ ] either 1.2 or 1.44 MB MSDOS diskettes

[ ] I already possess a copy of SHELXL

APPENDIX B

Installing SHELXL to maXus

Once SHELXL has been obtained it must be compiled as per the instructions that come with the SHELXL source code.

The final executable object must be placed in the maXus binary directory, called "/usr/maXus/bin.<operating system name>" in a standard implementation. This module must be called shelxl97, with no capitalisation.

The operating system name can be: sunos5, linux, irix5 or irix6

Example.

Having obtained a copy of SHELXL and compiled it in a personal library to an executable object called a.out, you would perform the following UNIX command for moving the object from your own directories to the maXus executable directories on a SGI running IRIX5.

```
mv /local/usr/myname/shelxl/a.out /usr/maXus/bin.irix5/shelxl97
```

Once the executable is present in the appropriate directory maXus will recognise this and all the SHELXL functionality described in this manual will be available.

15 Fourier

Calculation of Fourier Maps

15.1 Introduction

The Fourier program is nearly always called after the least squares program. Choose *Fourier* from the *Fourier* menu. The following window appears:

● Fo–Fc ○ 2Fo–Fc ○ Fo ○ Fc ○ Fsq ☑ SIM weight

Peaks:
Bond min: 1.00
Bond max: 1.65
Angle min: 85.0
Angle max: 145.0
Min peak dist: 0.5
Grid size: 0.333
Min sin(th)/lam: 0.0
Max sin(th)/lam: 1.5
Min I/sig(I): 0.0

- The following coefficients are available:

  (1) F_o - F_c

  (2) 2F_o - F_c

  (3) An F_o

  (4) An F_c

  (5) An F_o^2 (Sharpened Patterson maps, using coefficients based on E^2, can be calculated using MITHRIL or SIR)

  Select the option required.

- Options (1)-(4) are available with and without Sim weights. Use the Sim weighting button to select or de-select this option.

- Normally all the data are required to be used in the Fourier map, but three selection parameters are provided by which limits can be applied:

  (1) A minimum sinθ/λ. The default value is 0.0.

  (2) A maximum sinθ/λ. The default value is 1.0.

  (3) A minimum I/σ(I) ratio The default value is 0.0, but values of 2-3 may be appropriate especially when Sim weighting is not being used.

15.2 Peak Search

Every Fourier is followed by a peak search and peak interpretation. The numerical options for this part of the process are contained in seven boxes in the window:

- Peaks: This selects the number of peaks to search for. If left blank, the program uses a number based on the unit cell contents, and the number of atom currently in the model file. The program automatically identifies peaks corresponding to known atoms and excludes them from subsequent analysis, although they are listed.
- Bond min: This defines the minimum acceptable length for a bond.The default is 1.00A
- Bond max: This defines the maximum acceptable length for a bond. The default is 1.65A
- Angle min: This defines the minimum acceptable bond angle.The default is 145 degrees
- Angle max: This defines the maximum acceptable bond angle.The default is 85 degrees.
- Min Peak Distance: Default is 0.5.
- Grid Size: Default is 0.333.

Quite often these options will be unsuitable especially if you have organometallic compounds or 3-membered rings, so be careful.

Set the values you require then click on *OK*. Use *Cancel* to leave without running the program.

15.3 Program Output

There are four ways of examining the output from the Fourier module:

- Run the *Model* program. The current model is displayed with the new peaks from the Fourier in white with a size proportional to peak height. This is the simplest way, and allows easy visualisation of results.
- Examine the output by choosing *Fourier* from the *Summary* menu.
- Use PLOTQ to examine contoured maps graphically, in 2D or 3D. See Chapter 16.
- Use PReDS to examine contoured maps graphically in three-dimensions. See Chapter 16.

15.4 Errors and Warning Messages

The following error messages can be reported:

More than 200 points along x

or:

Map is too large

There are too many grid points: change the grid in the main menu. Consult support if the problem persists.

No reflections for Fourier

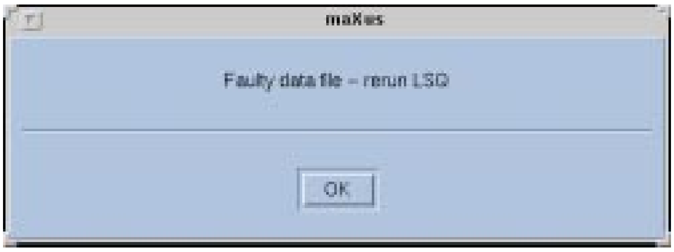
The conditions for accepting reflections are too restrictive: alter them accordingly.

Map has no peaks

The map is flat, check the input data. The best option for this is the *ListHKL/Calculated* option under the *File* menu.

All atoms found – no Sim Weights

Sim weights cannot be used if all the atoms are found: all the weights are set to unity.



Faulty data file – rerun LSQ

There is a problem with the data file. Try running LSq, and then Fourier again.

15.5 Limits

- Maximum number of reflections symmetry expanded in the Fourier program: 60,000.
- Maximum number of grid points along the x axis: 200.
- Minimum dynamic range of the map: 0.001. (The peak search will terminate if the difference between the maximum and minimum peak heights is less than this.)

16 Graphics

PLOTQ, ORTEP, XXMOL, RASMOL AND PLUTO

16.1 Graphics Packages

maXus provides the following graphics programs:

- Plotq for displaying electron density maps in 2- and 3-dimensions.
- Interactive ORTEP III.
- Two versions of XXMol: a state of the art display program.
- RasMol: a simple, industry standard display program.
- PLUTO.

Each of these options is accessed from the *Graphics* pull down menu. Each module is now described in detail.

Ortep, XXMol, Rasmol and Pluto have the additional choice between viewing the standard Model file, and the symmetry expanded file. This can be useful when dealing with molecules with high symmetry and small asymmetric unit cells.

For the graphics package of your choice, to view the standard model file, simply select *View Model* from the relevant submenu. To use the expanded model file in the graphics program, select the *View Expanded* option from the relevant *Graphics* submenu. For more information on the differences between the expanded and normal model files, see Section 12.1.

16.2 PlotQ: Density Contour Display In 2- and 3-Dimensions

Plotq is a program for the contouring and examination of electron density maps calculated by the maXus least squares and Mithril packages. It was considerably developed and expanded by Dr. Wei Dong from the original program written by Dr. Keith Henderson and Dr. Colin Bannister at the Chemistry Department, University of Glasgow, Scotland.

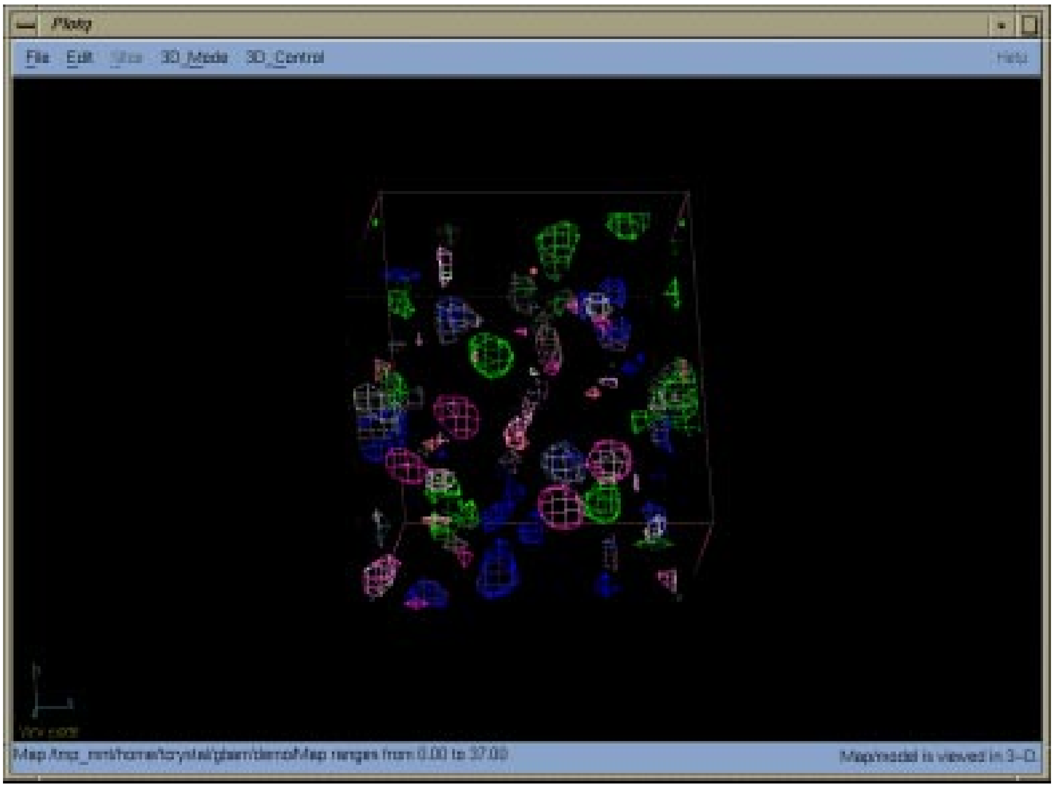
Maps can be viewed in 3D and freely rotated or translated with or without a Model. Model files can be displayed in a number of forms. They can be output for printing, or to VMRL files.

16.2.1 Starting the program

You must have calculated an electron density map using either:-

- **Fourier** The maXus Fourier module.
- **MITHRIL** The MAPS option in Mithril generates a map file every time an electron density map corresponding to a solution is calculated ready for a peak search. Only one map is ever held on file, corresponding to the last map requested. e.g. MAPS 5 generates the map for the 5th phase set (solution) whilst MAPS 3,5,7 generates and peak searches three maps in succession but only the map for phase set 7 will be held on file. See the MAPS section in the Mithril manual for more details of how to calculate Fourier maps.

Assuming a map has been calculated, select the *Graphics* option from the main maXus menu, then select either *Plotq (lsq)* or *Plotq (m94)* from the *PlotQ* submenu, depending on the type of map you wish to view. The program will automatically read the correct map file and the main window shown below will be displayed:



Map/model is viewed in 3-D

The density range of the map is shown at the bottom of the window. There are a wide range of options available for manipulating the map display and these are outlined in the subsequent sections.

16.2.2 Basic Image Manipulation

The map can be rotated freely by holding down the left mouse button and moving the mouse (or using the arrow keys).

It may also be translated across the screen using the middle mouse button (or SHIFT and the arrow keys).

If a Model file is open, showing atomic positions on the display, clicking the Right mouse button brings up a pop-up menu:

MODEL TYPE: Lines, Sticks, Ball&Stick, Spacefill

Selecting from these options controls how the atoms are displayed. Note that unique atoms are coloured differently according to atom-type, whereas symmetry generated atoms are all one colour (generally blue) to allow them to be distinguished.

16.2.3 File menu

This has the following options:

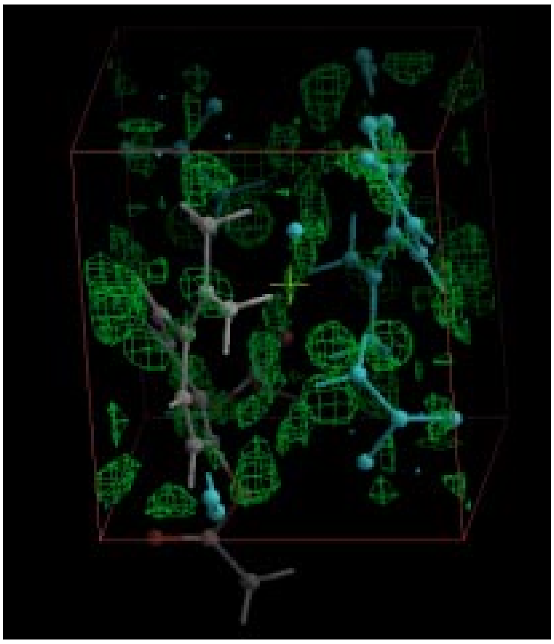
- **Open Map...**

  This brings up a standard File dialog box. In general, maXus will start plotq with the current map (generally a file called ‘Map’ in the working directory. However, the user may wish to view other Map files, in which case this option allows other maps to be viewed. Several different maps may be viewed simultaneously. To use this option, rename your Map files as necessary once they have been calculated.

- **Open Model...**

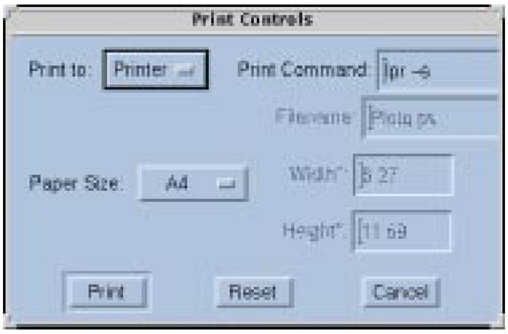
  This option brings up a standard File dialog box, allowing the user to select a maXus Model file. This will superimpose the positions of the atoms in the model file onto the map. Note that only one Model file may be viewed at a time.

For example:



- **Print...**

  This option outputs the display to a printer via a standard dialog:



Print to: Printer

Print Command: lpr -s

Filename: [illegible]

Paper Size: A4

Width": 8.27

Height": 11.69

  Note this option is only available when the program is in 2D mode. When in 3D mode, one of the other output options (see below) are required.

- **Screen Copy > as EPS...**

  as PPM...

  These options allow a copy of the current display to be output to a file, in either EPS or PPM format. A standard save-file dialog appears to allow the filename and destination to be selected.

- **Create VMRL...**

  This option allows the current map data (and model data if present) to be output to a VMRL file. This can then be viewed by any VMRL-equipped web-browser, for example.

  When first selected, the user is prompted for a filename and location to save the file.

  This is file empty at first. After the file is saved, any map or model files that are opened by the program will have data saved to this VMRL file. This continues until the menu option is reselected, at which point the VMRL file is closed.

If the filename entered is of a already-existing VMRL file, a dialog box will prompt the user to decide between either to completely *Overwrite* the file, or to just *Append* any new map or model information to the existing file.

- **Exit**

  This self-explanatory option quits the program and returns control to the main maXus menu.

16.2.4 Edit menu

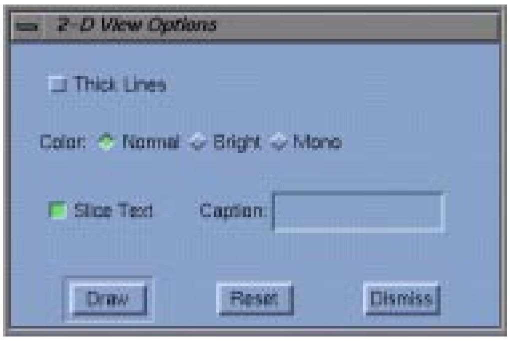
This has the following options:

- **View 2D / View 3D**

  This item switches between the two main display modes of the program - displaying the maps in either 2 or 3D. The default is 3D. Note that some menu items are only available in a particular mode; they are dimmed and unavailable in the other.

- **2D Viewing Options**

  This brings up an options dialog box:



☐ Thick Lines

Color: ● Normal ○ Bright ○ Mono

☑ Slice Text Caption:

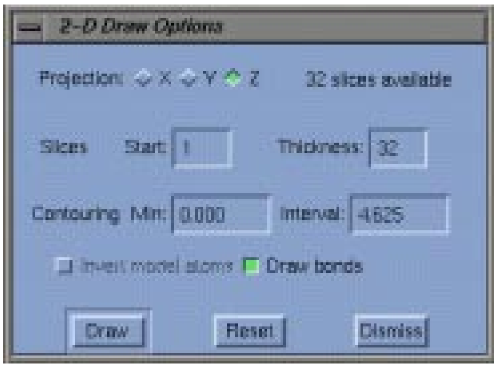
The **Normal** color scheme provides a reasonable screen display, and may provide good defaults for color printing. The **Bright** colors provide better choices for photographing the screen, when the user may also wish to select the **Thick Lines** checkbox, which doubles the line width. The **Mono** choice draws all contours in black on a white background and should be used when only a mono display or printer is available.

The user may type a **Caption** in the textfield provided, which will be displayed at the top of the graphics window. The indication of which slices are being viewed may be turned on or off, with the **Slice Text** checkbox. The slice information will always be displayed at the foot of the main window regardless of this settings.

Note that this menu option is unavailable in 3D mode.

- **2D Draw Options**

  This brings up an options dialog box (note that this menu option is unavailable in 3D mode):



Projection: ○ X ○ Y ● Z 32 slices available

Slices Start: 1 Thickness: 32

Contouring Min: 0.000 Interval: 4625

☐ Invert model atoms ☑ Draw bonds

Select **X**,**Y** or **Z** for a view down the X,Y or Z axes respectively. The default projection is down the Z axis. The number of slices available down the selected axis is displayed next to the selector.

Controlling the slice thickness

By default, a full projection down the selected axis is displayed i.e. the slice thickness is set equal to the total number of slices down that axis. In order to step through the map in thinner slices, the user should specify an integral number of slices which divides exactly into the total number. For example, if there are 20 slices down the Y axis and a thickness of 2 is specified, 10 sections will be drawn sequentially i.e. 1-2,3-4,5-6,............17-18,19-20. The user can then use the ***Slices*** menu to step forward or back though the slice display. By setting the start slice to a number other than one, the user can choose any combination of slices. A single slice can be displayed by setting the slice thickness to one.

Specifying the contour levels

PlotQ will contour a map automatically, selecting contour levels which span the range of density using relatively few lines. These default values will be displayed in the textfields, and the user can specify alternative values. As a general rule, maps have an interval such that

$$\frac{(\text{Maximum Density})}{(\text{Interval})} \cong 8$$

e.g With a map range is 0 to 1540, **Min** 0 **Interval** 200 would be appropriate.

Invert model atoms button

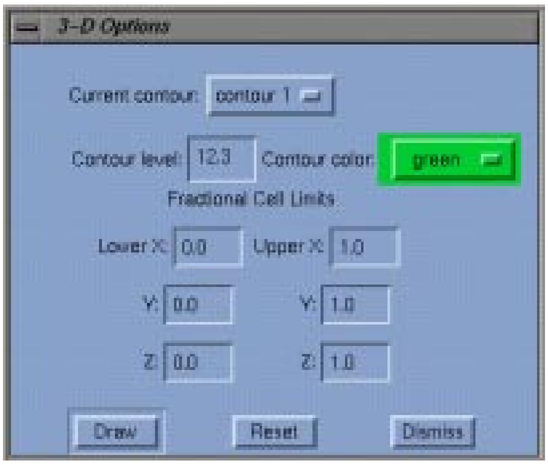
This option is only available when working with a structure which has a non-centrosymmetric spacegroup. It inverts all of the atomic coordinates.

Draw Bonds button

This option is used when a Model file is being displayed over a Map. With it on, lines are drawn between connected atoms to represent bonds.

- **3D Draw Options**

  This brings up an options dialog box:



Current contour: contour 1
Contour level: 12.3
Contour color: green
Fractional Cell Limits
Lower X: 0.0
Upper X: 1.0
Y: 0.0
Y: 1.0
Z: 0.0
Z: 1.0

For each of up to four different contours (selected from the *Current Contour* pop-up menu), the user may specify a **Contour level** lying between the minimum and maximum values of the map. These map limits will be visible at the lower left of the main window (in the “footer” area below the drawing pane). The “best” value must be found by trial and error, but the default contour level of one third of the maximum is usually a good starting place. The current contours will be displayed using the colour selected from the *Contour Colour* pop-up.

Once the map is displayed, the user may use the **Fractional Cell Limits** textfields to alter the display. The default is one unit cell, but fractions of a cell, or multiple cells can be drawn. Remember to click **Draw** or press return to view your changes.

Note that changes to the Fractional cell limits only affect the contour currently selected from the top pop-up menu. Consequently, it is possible to have several different contours displayed in varying fractions of the unit cell.

Note that this menu option is unavailable in 2D mode.

- **Dependant Model**

  This item is a toggle switch; when it is on (as it is as default) a green dot appears next to it in the menu. When on, the display uses the current size and symmetry of the Map when overlaying a new Model file on the map. When off, the and size and symmetry information contained in the Model file is used. The default is on, since Model files being viewed are generally of the same structure as the Map file. However, it is possible to view a Model file for a different structure, or of a larger structure that the Map may be only part of. In this situation, this option should be turned off.

- **Auto Reset**

  With this option on (default), every time a new Model file is read in, the program reverts back to all defaults in views, displays etc. When off, any current settings are maintained.

16.2.5 Slice menu

This has the following options (available only in 2D mode):

- **Slice NextCtrl-n**

  Slice Previous Ctrl-p

  These allow the user to easily step through all 2D slices of the map in order.

  This option is only available in 2D viewing mode.

16.2.6 3D Mode menu

This is a tear-off menu. Selecting the dotted line just below the menu title allows it to be ‘removed’ from the menubar, and placed in a pop-up window. This allows quick and easy access to all of its functions.

- **Show Model**

  This is on by default. If a Model file has been loaded, it will be visible in the display, superimposed over the Map. Turning this option off hides the Model data.

- **Show Bond**

  Default is on. With this option on, when a Model file is visible, bonds are drawn between connected atoms. When off, they appear individually with no connections.

- **Show Label**

  Default is off. When on, atoms present from an open Model file are displayed with their label. When off, the label is hidden.

- **Show Measurement**

  Default is on. The results of any geometry measurements performed (see later) are shown in the bottom left of the main window.

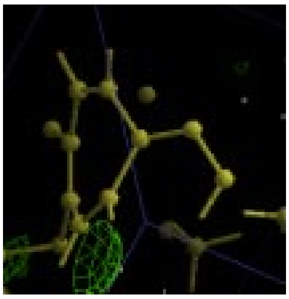
- **Select Atom**

  Default is off. With this option on, clicking on an atom in the display selects it (shown by yellow cross-hairs). Selecting atoms is used to be able to perform geometry calculations, and to re-centre the display. Details are given below.

- **Rotate Bond**

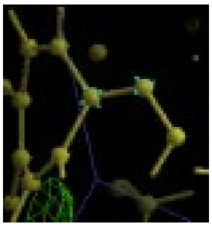
  With two bonded atoms selected, this option allows the bond to be rotated clockwise or anticlockwise.

  For example, to rotate a bond like:

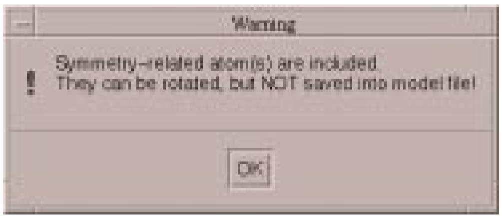


First turn on the *Select Atom* option of the *3D Mode* menu.

Then click on the two atoms that make up the bond to select them:



Finally, select *Bond Rotate* from the *3D Mode* Menu. If any of the selected atoms are symmetry related atoms, you will be warned:



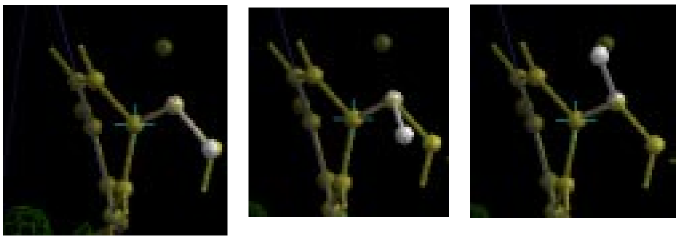
Warning

Symmetry–related atom(s) are included.
They can be rotated, but NOT saved into model file!

Witth this in mind, the bond can now be rotated, in small increments using:

Spacebar	<Clockwise>
Shift-Spacebar	<Anti-Clockwise>

A grey bond will appear to give visual feedback of the amount of rotation:



Once you are happy with the position of the grey bond, turn off the *Rotate Bond* menu option. You will then be asked:

Confirm

Update (NO undo) displaying model?

Click *Yes* if you are happy with the rotation, but be warned that there is no going back!

The Model file is then updated.

This is very useful with terminal methyl groups, for example, to derive accurate H coodinates for a terminal methyl group:

- Calculate structure factors without the H atoms on the CH_3
- Calculate an F_o - F_c Fourier map.
- Generate the terminal H atoms.

- Enter Plotq
- Rotate around the -C bond to optimally superimpose the H atoms and the electron density.

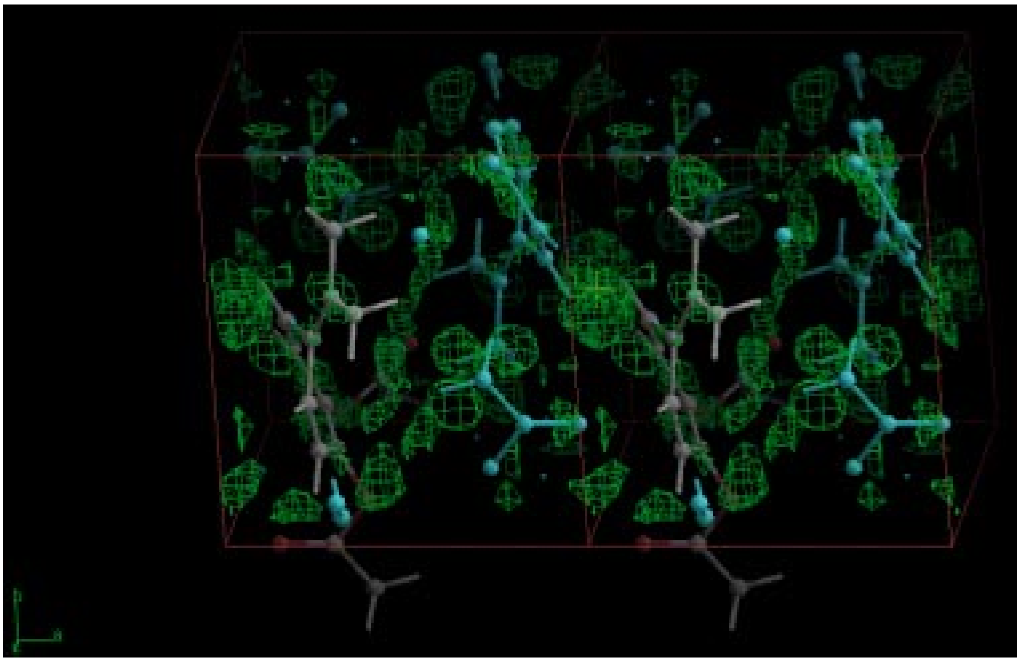
- **Show Contour of > Contour 1**
  Contour 2
  Contour 3
  Contour 4
  All

  PlotQ is capable of showing up to four different contour levels at once. This submenu lets the user control which of the set levels (set in the *3D Draw Options* dialog) are visible. The default is to show all contours.

- **Double Cell** > **Along a**
  Along b
  Along c
  All

  This submenu is useful to extend the map view by showing another unit cell along either a selected axis (a, b or c), or along all axes (All). This can help to see symmetry and how unit cells fit together.

  For example, selecting Double Cell Along a for the ‘Demo’ structure would give:



- **Depth Cueing**

This option (default is on) uses a ‘light source’ to simulate light, shadow and depth in the 3D display, thus improving the illusion of three-dimensionality. Atoms further away are drawn more faintly, for example. Turning it off may improve screen response times on slower computers.

- **White Background**

  Default is off. Normally, the Maps are shown against a black background. However, when pictures of maps are required to be output on white paper or transparencies, it is useful to be able to remove the dark background and view just the density map and model. This option allows this.

- **Clipping**

  Default is off. This menu item controls if a clipping plane is used along the z-axis. This means only atoms between the two planes are displayed. The plane thickness is controlled via the 3D Control Menu (see below).

- **Rocking**

  Default is off. Selecting this item causes the current display to rock gently from side to side through a few degrees. This increases the illusion of 3D, and allows easier visualisation of the relative position of atoms in the cell - for example, atoms that are ‘behind’ each other can be easily seen.

16.2.7 3D Control Menu

This is a tear-off menu. Selecting the dotted line just below the menu title allows it to be ‘removed’ from the menubar, and placed in a pop-up window. This allows quick and easy access to all of its functions.

- **Zoom In** **z**

  Zoom out **Shift-z**

  These control the zoom level in the main window. Selecting them repeatedly (or pressing z or Shift-z repeatedly) keeps zooming further and further in (or out).

- **Clip Thinner** **c**

  Clip Thicker **Shift-c**

  These control the thickness of the clipping plane used by the *Clipping* command described above.

- **Roll Clockwise** **s**

  Roll Anticlockwise **Shift-s**

  These control the direction the map is rotated in about the a-axis.

- **Geometry Measurement >** **Distance**

  Angle

List on console

Delete last entry

This submenu is only available when the *Select Atoms* option is checked. To obtain the distance between two atoms, click once one each atom so that both are selected. Then select the *Distance* option - the atom labels and the distance between them will be calculated and, if the *Show Measurement* option is on, this information will be shown in the bottom-left corner of the screen. To get an angle measurement, first select three atoms, and then choose *Angle*. If less than three atoms are selected, an error will be displayed.

When selected, the *List on Console* option outputs a copy of all measurement calculations taken at that point to your current active console window. From here it could be easily copied and pasted into another application such as a word processor.

Delete last entry removes the last measurement taken from the stored list of measurements kept.

- **Centring Selected Atom Home**

  This option is only available when a single atom is selected. It changes the centre of rotation of the map to around the chosen atom.

- **Reset View r**

  This option returns the Map view to its defaults, including resetting the centre of rotation to the middle of the Map.

16.2.8 Plotq Technical Appendix - Command-line arguments

maXus users will not normally need to make use of plotq on the command-line, but this information is included for completeness. In addition to reading files using the menus in plotq, a map file may be specified as an argument:

```
pq [-mice_map] [-capt_font name] [-axes_font name] [-atom_font name] [mapname]
```

-mice_map: For reading a different format of mapfile, not currently used by maXus
-capt_font: next argument is a fontname for the caption text
-axes_font: next argument is a fontname for the axis labelling text
-atom_font: next argument is a fontname for the atom number text
mapname: map file name

All of these are optional, and pq does not need to be given any arguments to start.

16.3 ORTEP

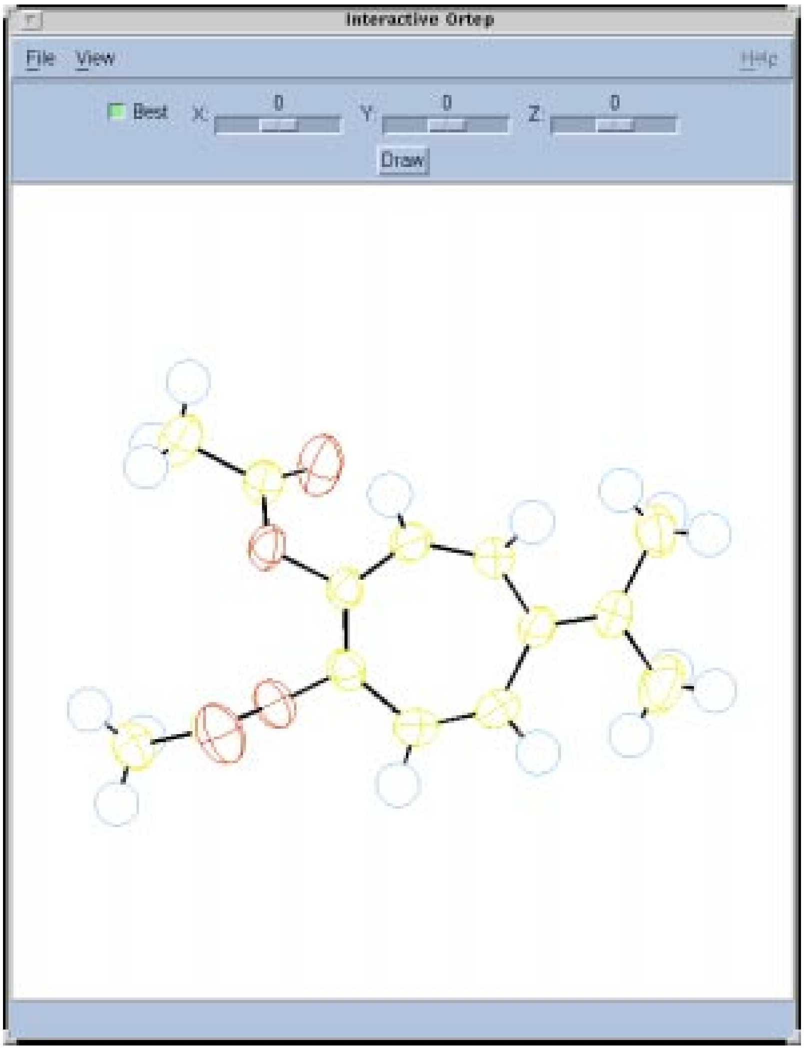
This program provides an X-window-based interactive method for drawing crystal structure illustrations featuring thermal ellipsoid plots, tapered bonds and perspective projections. The program automatically reads the maXus format of crystal model file and provides sensible defaults for rapid production of common diagram types. The controls are grouped in sub-windows according to function and there is a command editor for more sophisticated control of the drawing.

Select either *View Model*, or *View Expanded* from the *Ortep* menu to begin.

16.3.1 Main Ortep Window

The initial view of the Ortep window has a small command area at the top, and a large blank area in which is displayed the structure diagram. The window may be resized by dragging the corners of the window, as allowed by your window manager. The drawing will then be recalculated for the new window size. There are also several pop-up dialog windows with extra controls which may be called up and dismissed as required.

A default view of the molecule (as shown when the program is first launched) may be obtained simply by pressing clicking the **Draw** button. The view produced will be along a line perpendicular to the least squares plane of the molecule (the “**Best**” view) in which the hydrogen atoms are shown as plain ellipsoids and all other atoms are shown as unshaded ellipsoids. All bonds to H atoms are drawn as shaded, tapered cylinders, while all other bonds



are solid, tapered cylinders. The full set of atoms in the model file is displayed, without application of symmetry operations and the diagram will be scaled to fill the drawing area as far as possible.

<u>Rotation</u> of the view may be easily achieved by clicking and dragging on the diagram using the left mouse button.

Alternatively, drag the sliders marked X, Y and Z (or by clicking on the slider bar for fine adjustment). Once the desired angle has been set, click **Draw** to re-display the molecule in the new orientation. Rotations are relative to the **Best View** position, unless this button is turned off, in which case rotation is relative to a perpendicular view down the cell axes.

The remaining controls for setting viewing parameters and atom properties are located on dialog windows. These all have a **Reset** button which restores the default settings and a **Dismiss** button (which has the same function as the window manager Dismiss command). After altering settings in a window, the user must press the dialog's **Apply** button or the main **Draw** button to update the drawing.

16.3.2 Atom properties

These can be altered from the dialog window obtained when the **View/Atoms** menu item is selected.

Atom Controls

Atom Type: Hydrogen

Bond Style: Cylinder

Bond Radius: 0.02

Lines: 3

Atom Style: Plain

Atom Radius: ● Thermal% ○ Fixed

☑ Label

Atom Font Size: 12

Sphere Radius: 0.10

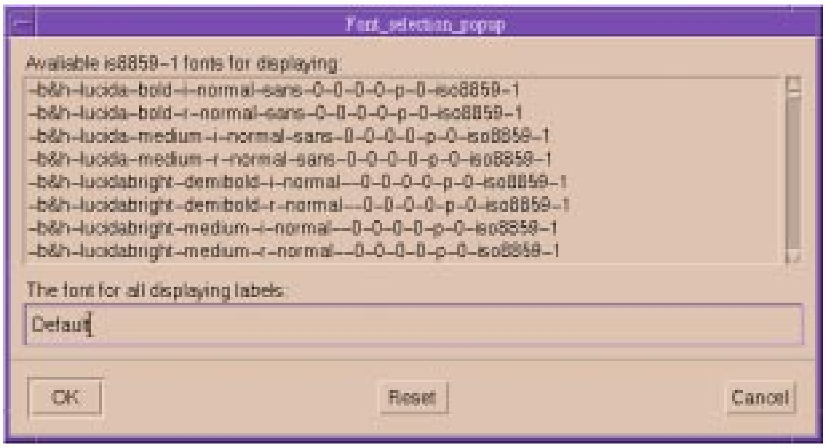
The dialog window has a menu for selecting the atom type to be acted upon; either Hydrogen or Non-Hydrogen. Setting you alter affect only the type of atom you have selected.

- **Bond Style** pop-up menu**:** choose from either tapered **Cylinder** (the default) or **Stick,** which draws a single line bond. Cylinder bonds may have the cylinder radius set (in Å) and the number of lines used to shade the bond can be selected from one to five.
- **Atom Style** pop-up menu: settings available are **Plain** (the default H atom style), **Unshaded** (non-H default) and **Shaded**, which adds octant shading to the thermal ellipsoids. In addition, there is an **Omit** option which removes the specified atoms from the diagram.
- **Atom Radius** buttons: Rather than displaying probability ellipsoids, the atoms may be drawn as fixed-radius spheres by changing the **Atom Radius** setting from **Thermal** to **Fixed**. The radius is set in the **Sphere Radius** text field, where the default value is 0.1Å.

- **Label** button: setting this will add atom labels to the drawing. The labels will be placed slightly to the left of the corresponding atom centre.

  Notes on Label Positioning: Once a labelled diagram is displayed, the labels may be moved for maximum visibility or to ensure that particular labels are near to the correct atoms. **To move a label, press and hold the middle mouse button over the label then drag the pointer, releasing the button when the label is in the desired position. If your mouse only has two buttons then press both simultaneously.** The diagram will then refresh to replace areas overwritten by moving the label. On SUN systems, the label will move with the pointer until the button is released, allowing accurate positioning first time. Labels can be deleted simply by dragging them outside the view window, preferably to the left (labels dragged out of view upwards or to the right may appear on print-out if the new label position fits onto the printed page). Note that re-drawing the diagram with the **Draw** or **Apply** button (for example to change the view direction) will reset all the label positions so the view direction and other parameters should be selected *before* labels are rearranged.

  The font size used for the labels is that entered in the *Atom Font Size* box. The Font used to display them may be changed using the **View** menu **Fonts...** option. This brings up a dialog box:



Available is8859-1 fonts for displaying:

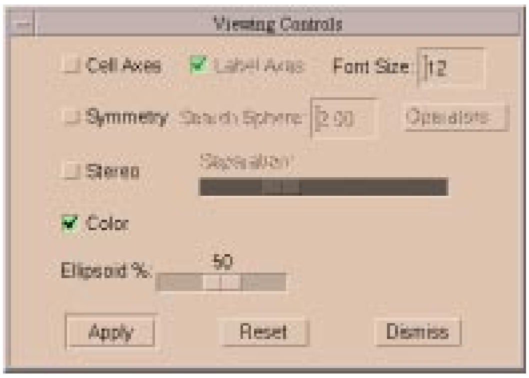
-b&h-lucida-bold-i-normal-sans-0-0-0-0-p-0-iso8859-1
-b&h-lucida-bold-r-normal-sans-0-0-0-0-p-0-iso8859-1
-b&h-lucida-medium-i-normal-sans-0-0-0-0-p-0-iso8859-1
-b&h-lucida-medium-r-normal-sans-0-0-0-0-p-0-iso8859-1
-b&h-lucidabright-demibold-i-normal--0-0-0-0-p-0-iso8859-1
-b&h-lucidabright-demibold-r-normal--0-0-0-0-p-0-iso8859-1
-b&h-lucidabright-medium-i-normal--0-0-0-0-p-0-iso8859-1
-b&h-lucidabright-medium-r-normal--0-0-0-0-p-0-iso8859-1

The font for all displaying labels: Default

  Simply click on the desired font and click *Apply*. Note however that any changes to the font will not be reflected to the output PostScript file.

16.3.3 General View Settings

The **View/Options** menu shows a window of settings for the general view of the model:



☐ Cell Axes ☑ Label Axes Font Size: 12

☐ Symmetry Search Sphere: 2.00

☐ Stereo

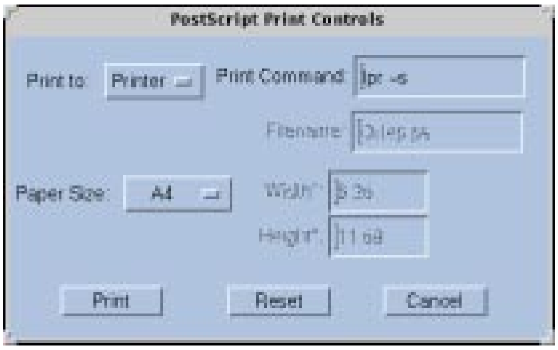
☑ Color

Ellipsoid %: 50

- The **Cell Axes** may be drawn, with or without axis **Labels** a b and c (which can be repositioned as for atom labelling). The Font Size may also be altered.
- The **Symmetry** checkbox will add complete molecules to make up the unit cell contents and is most useful in conjunction with the **Cell Axes** display. Initially, all the atoms within the unit cell (using all the symmetry operations) are added to the drawing, then a distance search is used to find all adjacent atoms within 2.0Å. If the drawing contains too many molecules then the search distance may be reduced using the text field provided. Alternatively if any molecules are incomplete then a value greater than 2.0Å may be required. For a model file with a large number of atoms or many molecules in the unit cell, the shading of the ellipsoids should be reduced to clarify the diagram and increase drawing speed. Fine detail will not be visible at the reduced scale required to display an entire cell. Similarly, there is likely to be little point in labelling the atoms in such a diagram. Remember that the window can be resized using the resize corners provided by your window manager. A larger window should make complex diagrams clearer.
- A “crossed-eyes” stereo image may be displayed with the **Stereo** setting. This works with all drawing styles, including symmetry. The stereo **Separation** and the width of the window may need to be adjusted to suit the image. Labelling or perspective may not produce satisfactory results on stereo diagrams.
- **Color** drawing is the default. Atoms are colored according to the scheme used in the model window of maXus. To alter the color of an atom, use the facilities in the Model window and the saved changes will be reflected when Ortep is restarted. Note that color printing is also the default, so users may want to turn off the color setting before printing to a monochrome printer.
- **Ellipsoid%** is a slider calibrated from 1 to 99% which represents the percentage probability enclosed by the thermal ellipsoids. It can be used to scale up or down all atom sizes. The default value is 50%. Note this differs from setting a fixed sphere size for all atoms, as different atoms may scale by different amounts with **Ellipsoid%**.

16.3.4 Printing

The **Print** facility is accessed from the **File/Print** menu which will show a dialog window with printing controls.



Print to: Printer

Print Command: lpr -s

Paper Size: A4

Height": 11.69

- The default action is to produce a PostScript printout using either “lpr” or the contents of the environment variable PRINT if it is set. This can be changed in the textfield provided. The printer to be used can be selected with the environment variable PRINTER if the system default printer is not PostScript compatible. Note that the labels will be printed in the positions they have been dragged to, and that the printout will normally be clearer and of much higher resolution than is seen on-screen. If the on-screen diagram is in color then color PostScript will be produced.
- The diagram can be saved to a PostScript file by changing the setting to “**File**”. The default filename is Ortep.ps (though this can be changed) in the same directory as the rest of the data.
- The default **Paper Size** is A4 which can be changed to A3 or A5 as desired. Selecting the **Custom** paper size allows the paper width and height (in inches) to be entered manually. The printing routine will attempt to position the diagram so as to fill the available paper size, with a suitable margin, using landscape or portrait mode as appropriate.

16.3.5 Exit

File/Exit will exit Ortep and return control to the maXus windows.

16.4 Ortep: More Advanced Features.

16.4.1 Extras

The **View/Extras** menu provides access to the dialog window labelled **Additional Controls** which contains some more advanced settings for users who wish to customise diagrams further:

Additional Controls

☐ Perspective View Dist

Overlap Gap: 2

Bonding Tolerance: 0.20

☐ Atom brackets

- The **Perspective** checkbox is off by default, giving an orthogonal projection, but it can be turned on and the slider moved to change the effective view distance (smaller distance implies greater perspective exaggeration).
- The **Overlap Gap** is a feature of the hidden line removal algorithm which allows a gap to be left between lines representing overlapping parts of the diagram. The slider defaults to 2 which is nominally the gap in mm (although the actual gap is affected by scaling etc.). The

overlap gap can be completely eliminated by setting the slider to 1, which some users may prefer for certain diagrams. The maximum setting of 9 is likely to be larger than required by most users.

- **Bonding Tolerance**: Controls how strict the bonding test is. Normally a bond is drawn between two atoms only if the inter-atomic distance is less than or equal to the sum of the covalent radii. Occasionally a user may wish to relax this constraint and allow bonds between atoms which are further apart. Enter a value in the number field. Default is 0.2.
- **Atom brackets button**: Default is on.

16.4.2 Command Edit Button

Command Editing is a further facility for users who wish to achieve effects not accessible from the normal window controls. This button invokes a text editor on the file of underlying commands which is used to communicate with the Ortep process. These allow a much greater degree of control. For example: a metal complex may require to be shown with a different bond style (or no bonds at all) to the metal atom; different atom styles may be needed for particular atom types (rather than simply H or Non-H); additional symmetry effects may be needed. A full reference of the commands available will be found in Section 16.5, and useful examples can be obtained simply by preparing a diagram using the window controls and then viewing the command file in the editor window.

Note 1. Once the command editor has been started, the window controls cannot be used. Even operations such as rotation must be carried out using the editor as the window controls will be inactivated to prevent discrepancies between the settings in the controls, and the settings in the editor. To return to the ordinary controls, both the editor and the Ortep window itself must be quit, and Ortep restarted from the maXus window.

Note 2. In order to redraw the diagram after changes have been made to the command file, the file should be saved using the menus provided by the text editor. The **Draw** button will then re-read the file and produce the corresponding diagram. The **Print** options are still available. Note that the **File/Exit** item will exit Ortep, but the text editor will have to be exited separately by using the window manager menu over the title bar of the editor.

Note 3. Both the GX commands listed in the next section and the original Ortep numerical commands may be used in the command editor. To allow numerical commands, syntax checking of the GX commands is turned off with NUMERICAL COMMANDS (Section 16.8.6). This means that the syntax should be carefully checked before saving the file, as the program will fail if there are syntax errors. Remember that the Summary feature in maXus allows viewing of the output files, so the Ortep output can be viewed if there are problems.

16.5 Ortep Text Commands

The text commands for the original GX version of Ortep are available in manual edit mode. They allow greater customisation of the diagram than can be obtained with predefined window controls alone, at the expense of greater effort on the part of the user. All of the original Ortep numerical commands can also be used, so experienced users of the Oak Ridge Ortep will be able to customise diagrams even further (with a little patience). To access text commands, click the **Command Edit** button in the *Additional Controls* dialog box.

16.6 Ortep Commands: Composing and arranging the illustration

16.6.1 MODEL_INSERT

Creates a list of atoms that are to be the subject of the illustration. All the atoms in the crystal model file are inserted in the list, with no symmetry transformations applied. This command should be used when the coordinates given for the asymmetric unit correspond to a connected set of atoms and no equivalent positions are to be generated. This is the default mode in Interactive Ortep.

16.6.2 MOLECULE_INSERT <dmax> [SYMMETRIC]

Creates a list of atoms that are to be the subject of the illustration. The first atom in the crystal model file is taken to be the centre of a sphere of enclosure of radius dmax Å (the longest bond length). Any atom in the file which is inside this sphere will be inserted in the list. The search is performed for all the symmetry-equivalent positions of the target atoms, so we are sure to generate the entire surroundings of the original atom. The process is automatically repeated for atoms added to the list until no new entries are found, so that the atom list then consists of a chemically bonded set. Omitting the dmax parameter gives a default value of 2.0Å. If the SYMMETRIC flag is given, the list will include atoms in more than one equivalent position, e.g. in a crystallographically symmetric molecule. This option should be used with caution, as an infinite chain can be generated if dmax is too large.

The MOLECULE_INSERT command should be used when the atomic positions given in the model file are not a connected set or when more than one asymmetric unit is required. Note that the first atom in the model file is used as a connectivity “seed” and should not therefore be structurally isolated from the rest of the molecule.

16.6.3 CELL_CONTENTS_INSERT <dmax> <fractional semi-dimensions>

Creates a list of atoms that are to be the subject of the illustration. Atoms in the model file are inserted in the list if, after any necessary symmetry transformations, they lie within a box of enclosure having the given fractional semi-dimensions whose centre is the point (1/2,1/2,1/2). The faces of the box are parallel to the unit cell faces. The default values of the semi-dimensions are 1/2, 1/2, 1/2 giving the contents of one full unit cell. Varying the dimensions permits selection of the contents to achieve maximum clarity. If dmax is smaller than the shortest bond length the list will contain just the atoms inside the box; otherwise atoms connected to these will be added until complete molecules are obtained. The “Search Sphere” textfield in the window interface corresponds to the dmax argument here, with a default of 2.0Å. The CELL_CONTENTS_INSERT command resets the reference axes orientation to the default position described under ROTATE below. The default CELL command is produced by the “Symmetry” check-box in Interactive Ortep.

16.6.4 DELETE <atom label(s)>

Removes one or more atoms from the figure subject list. The deletion applies to all equivalent positions of the given atom or atoms. If no atoms are given the list is cleared. If an atom type is given, all atoms of that type are deleted, e.g. DELETE H removes all hydrogen atoms from the figure subject list. This is the command produced by the “Omit” option on the Hydrogen atom control panel in Interactive Ortep.

16.6.5 BEST VIEW

Defines a set of reference axes in the crystal which parallel the graphical axes of the display window. The reference x and y axes lie in the least-squares plane through the atom in the figure subject list, with the z axis at right angles to x and y. The view direction is along the z-axis, with x horizontal and y vertical. The origin will be positioned in x and y so that the molecule is centred in the view window, and scaled so that the window is filled as far as possible. The atoms will be placed above and below the drawing plane, ready for projection. In Interactive Ortep the BEST VIEW option is selected by default.

16.6.6 ROTATE <axis> <angle>

Rotates the reference axes with respect to the crystal, so changing the viewpoint. The rotation axis must be given as X, Y or Z. A positive angle gives an anti-clockwise rotation of the crystal viewed from the positive end of the axis. Successive ROTATE commands are cumulative in effect. Initially (before any ROTATE or BEST VIEW commands), the x reference axis is parallel to a and z is parallel to c. The x, y and z directions on the screen are as defined under BEST VIEW above. In Interactive Ortep the rotations are controlled by sliders ranging from -180° to +180°.

16.6.7 SCALE <model scale> <ellipsoid scale

With no parameters, this scales the figure as described under BEST VIEW above. Alternatively, an absolute scale factor in mm/Å can be given and/or a scale factor for probability ellipsoids as detailed in Table 9, “Scale Factors for Probability Ellipsoids,” on page 17- 23. To change the ellipsoid scale without changing the existing model scale, a comma should be given in place of the first parameter. Interactive Ortep allows direct control of the ellipsoid scale as a percentage, from 1% to 99% probability.

TABLE 9. Scale Factors for Probability Ellipsoids

P%	Scale	P%	Scale	P%	Scale	P%	Scale
1	0.3389	31	1.2110	61	1.7351	91	2.5478
2	0.4299	32	1.2288	62	1.7540	92	2.5997
3	0.4951	33	1.2464	63	1.7730	93	2.6571
4	0.5479	34	1.2638	64	1.7924	94	2.7216
5	0.5932	35	1.2812	65	1.8119	95	2.7955
6	0.6334	36	1.2985	66	1.8318	96	2.8829
7	0.6699	37	1.3158	67	1.8519	97	2.9912
8	0.7035	38	1.3330	68	1.8724	98	3.1365
9	0.7349	39	1.3501	69	1.8932	99	3.3682
10	0.7644	40	1.3672	70	1.9144	99.1	3.4019
11	0.7924	41	1.3842	71	1.9360	99.2	3.4390
12	0.8192	42	1.4013	72	1.9580	99.3	3.4806
13	0.8447	43	1.4183	73	1.9804	99.4	3.5280
14	0.8694	44	1.4354	74	2.0034	99.5	3.5830
15	0.8932	45	1.4524	75	2.0269	99.6	3.6492
16	0.9162	46	1.4695	76	2.0510	99.7	3.7325
17	0.9386	47	1.4866	77	2.0757	99.8	3.8465
18	0.9605	48	1.5037	78	2.1012	99.9	4.0331
19	0.9818	49	1.5209	79	2.1274	99.91	4.0607
20	1.0026	50	1.5382	80	2.1544	99.92	4.0912
21	1.0230	51	1.5555	81	2.1824	99.93	4.1256
22	1.0430	52	1.5729	82	2.2114	99.94	4.1648
23	1.0627	53	1.5904	83	2.2416	99.95	4.2107
24	1.0821	54	1.6080	84	2.2730	99.96	4.2661
25	1.1012	55	1.6257	85	2.3059	99.97	4.3365
26	1.1200	56	1.6436	86	2.3404	99.98	4.4335
27	1.1386	57	1.6616	87	2.3767	99.99	4.5943
28	1.1570	58	1.6797	88	2.4153	99.999	5.0894
29	1.1751	59	1.6980	89	2.4563	99.9999	5.5376
30	1.1932	60	1.7164	90	2.5003	99.99999	5.9503

16.6.8 SIZE <element>/[ALL] radius

Sets a sphere size for the given atom type or for all atoms. The radius is in Å with a default of 0.1Å. This overrides the temperature factors from the model file. Interactive Ortep allows fixed spheres to be set for hydrogen or non-hydrogen types only.

16.7 Ortep Commands: Drawing the illustration.

These are most conveniently used in a “saved sequence” which is independent of rotations and scaling:

16.7.1 BONDS <shade> <half-thickness> <overlap gap> <tolerance> <CR> <element pair> <CR>

This causes bonds to be drawn between all pairs of atom types in the “element pair” list. The list consists of two atom types (C H etc.) on each line, terminated by a blank line. <CR> here represents “carriage return”. A bond is an interatomic vector shorter than the sum of the covalent radii and the tolerance. The radii are given in the crystal model file, or (more normally) looked up in the system databank file. Interactive Ortep automatically generates commands to bond all the atom types in the model file. This may not make chemical sense, but the bonding criteria normally produce a meaningful diagram, otherwise the bond commands may need to be edited manually.

If the “shade” and “half-thickness” parameters are given, bonds will be drawn as cylinders with a perspective taper, where the “half-thickness” provides the radius in Å of the cylinder. “Shade” is an integer in the range 1 to 5 which controls the number of lines used to draw the cylinder. These parameters can be controlled for bonds involving Hydrogen or non-Hydrogen atoms only in Interactive Ortep.

“Tolerance” controls how strict the bonding test is. Normally a bond is drawn between two atoms only if the inter-atomic distance is less than or equal to the sum of the covalent radii. Occasionally a user may wish to relax this constraint and allow bonds between atoms which are further apart. A tolerance of between 0.0 and 0.5Å is usually sufficient to achieve this. Too large a value will result in multiple unwanted bonds. NOTE that the tolerance may be a negative value which will restrict bonding to values smaller than the covalent radii sum. The tolerance is set in the “Extras” window in Interactive Ortep.

“Overlap gap” should normally be omitted (there is a suitable default). It provides control over the gap left around partially obscured atoms and bonds in the diagram and is equivalent to a gap in mm, less than 10mm. A value of 1.0 eliminates the gap completely, whereas 0 uses the default value. Overlap gap is controlled from the “Extras” popup in Interactive Ortep.

Note that the “shade”, “half-thickness” and “tolerance” parameters can also appear in the element pair list, following a pair of atom types on the same line. This allows bonds between particular atom types to be drawn in different styles or added/omitted using the tolerance value. Bonds to Hydrogen atoms are drawn with reduced shading by default using this technique. The BONDS command can also draw line bonds, by omitting “shade” and “half-thickness” or by setting the half-thickness to zero. Note that hidden line removal is carried out by the BONDS command, so it should always be used before ATOMS.

16.7.2 ATOMS [PLAIN]/[UNSHADED]/[SHADED] <element>

This causes the atoms in the figure subject list to be projected onto the drawing plane and drawn in the display area. PLAIN gives only the boundary ellipse (or circle) UNSHADED gives the boundary ellipse plus principal ellipses SHADED gives the UNSHADED style with octant shading The default style for Hydrogen is PLAIN, for non-Hydrogen, UNSHADED. These are controlled by pull-down menus in Interactive Ortep. Hidden line removal will be carried out if the ATOMS commands are preceded by BONDS. Note: Interactive Ortep expects atoms of the same type to be grouped together in the model, otherwise some bonds and atoms may be omitted. To achieve this the atom list is sorted by type when read from the file, regardless of sort order in the model window.

16.7.3 LABEL [AXES]/<element>/[ALL] <x offset> <y offset>

This places labels on the unit cell axes, on atoms of the given type, or on all atoms. The offsets are in mm along the screen x and y axes, from the end of the unit cell axis, or from the atom centres. Interactive Ortep uses a small default offset and then the user can interactively place or delete the labels to obtain maximum clarity before printing or saving the diagram as PostScript.

16.7.4 OUTLINE CELL

This draws the unit cell outline, provided it is within the drawing area on the current scale. Re-scaling may be achieved using OUTLINE followed by SCALE, since OUTLINE inserts dummy atoms at the cell corners in the figure-subject list. A second OUTLINE command will then produce a visible result, but the entire figure should be redrawn first, to ensure it is on the same scale as the outline. This procedure can cause users some confusion so the window commands should be used to set up the cell outline with correct scaling before the command files is edited.

16.7.5 PEN <number>

This command is used by default in producing a color diagram. The number corresponds to an index into the color table used by the maXus model window, so the actual colors will depend on the choices made in the model window color chooser. Color number one is the default foreground color (normally black) used for bonds and labels, or for monochrome (black/white) drawings.

16.8 Ortep Commands: Program control

16.8.1 SAVE SEQUENCE

This marks the start of a saved sequence, which can be executed as a block when desired. This will replace any previously-define sequence, or restart a currently open sequence.

16.8.2 END SAVE SEQUENCE

This marks the end of a save sequence.

16.8.3 EXECUTE SAVED SEQUENCE

This performs all commands previously defined between SAVE and END. Note that they are not performed when the save sequence is created, unlike the original version of Ortep.

16.8.4 STEREO <separation>

This executes the saved sequence twice, with a stereoscopic rotation of the viewpoint. The separation parameter is a distance in mm. A suitable model scale must be preset by the SCALE command to cause the entire diagram to appear in the view window. The window controls will automatically write the required scaling commands so the STEREO command is not one which users will normally require in command editing mode.

16.8.5 NEW_PLOT <view distance> <x origin>

This allows control of the view distance for perspective projection. The view distance is currently in the range 0 - 5000 in Interactive Ortep, which gives a reasonable range of control. A view distance of zero gives a parallel projection (no perspective) which is equivalent to turning off perspective in the Ortep “Extras” window. The “x origin” parameter is now redundant, but was previously used when output was directly to a plotter, to advance the plotter.

16.8.6 NUMERICAL COMMANDS

This permits the original Ortep instructions to be used (as described in ORNL-3794, Johnson, 1971). This command suppresses checking of command syntax and so care should be taken when manually editing command files unless “NUMERICAL COMMANDS” is removed. Interactive Ortep uses this command by default, to allow a “303” command in the original Ortep syntax, which is used internally by the X-graphics routines. If NUM is to be removed, then the 303 command must also be removed, and labels may appear more than once in a labelled diagram.

16.9 Ortep Bugs & Warnings

- An infinite network of molecules can sometimes be generated by the 2.0Å search distance used to generate symmetry-related molecules.In this case drawing may take a long time and produce too many molecules outside the cell. The drawing will stop at the maximum number of atoms (currently 1000). To avoid this problem the dmax search sphere value in the “View” window should be reduced slightly and the diagram redrawn.

- Complex diagrams may be slow to draw. This can be improved by reducing the drawing style to a simple stick model with plain ellipses for the atoms. Detailed drawing types such as octant-shaded ellipsoids are unlikely to be useful when symmetry-related atoms are drawn in a large unit cell.

16.10 Ortep Errors

There are some warning and error dialog boxes used by Ortep to inform the user of abnormal conditions. These are mostly related to checking the crystal model file for problems which would prevent the correct production of the drawing. These do not necessarily indicate a fault in the structure solution. For example Ortep will not run if there are any atoms with non-positive definite temperature factors, as can sometimes occur during refinement of a structure.

The following dialog boxes may be displayed, as they are part of the maXus syntax checking routines.

Initial indication of problems - click the button to list errors:

Model file faults

The model file contains only unassigned peaks, not atoms:

All atoms unassigned

There are no atoms (or peaks) in the model file:

No atoms

The model file does not end correctly:

Faulty model file end

There are non-positive definite temperature factors in the model file:

TFs are NPD

There are unreadable coordinates in the model file:

Coordinates unreadable

If the number of vectors involved in the drawing exceeds 100,000 the program will issue an error notice and fail to draw the structure. This should only occur for over-complex diagrams and can be avoided by using the PLAIN ellipsoid style for complex drawings (e.g. stereo, symmetry, large molecules, large unit cells).

If a symmetry-produced drawing would contain more than 1000 atoms, the drawing will cease at 1000 and display as much as has been completed. This is unlikely to occur unless the Search Sphere parameter is too large, generating an infinite network of molecules.

16.11 XXMOL - SG Version

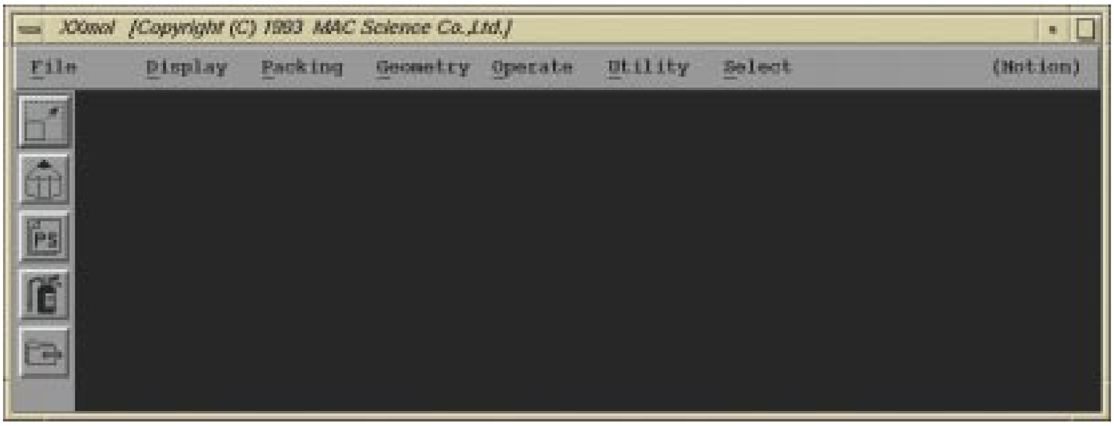
XXmol is a state-of-the-art molecular display program which comes in two separate versions: one only runs on computers supporting the GL graphics standard, notably Silicon Graphics. If it is available, the SpaceBall can be used as an interactive device. XXmol provides the user with an easy to use graphics program capable of displaying several structures simultaneously using different sources, and will give an extended display of the structure over many unit cells; there is an extensive an extensive choice of lighting and material effect making it suitable for displaying the structure for use in slides. If the SpaceBall is available, it can be used to control the molecule.

There is another simpler, more limited version for other UNIX computers - see Section 16.15.

16.11.1 Starting the program

Use the *Graphics/XXmol* menu option to select either *View Model* or *View Expanded*. The following window is then displayed. If the structure is already selected within maXus the window will automatically display the structure as wireframe. Otherwise the dialog box will appear - click on OK and XXmol empty window will open.

Draw file does not exist



16.11.2 General Operation

Mouse operations: If the left button is depressed while dragging it translates structure in the direction of the mouse movement; the middle mouse button if depressed while dragging rotates in direction of movement as long as button is depressed. If the mouse is still moving as the middle button is released, the structure will continue to rotate in the same direction until the middle button is clicked. The left and middle together (or shift key and middle) zooms the structure in and out of screen.

Window resizing: The window can be resized - the structure displayed will rescale and re-centre accordingly.

All pop-up windows are closed using *Dismiss*. Options within the popup menus can be selected using the mouse or by a combination of Tab and cursor keys.

The *Motion and Pick option* - This is toggled by clicking on the option at the far right of the menu bar. Pick is automatically selected under a command in which atoms are selected by the user e.g in calculating the distance between atoms, creating a bond or assigning a label to specific atoms. Under this condition the structure can still be rotated or zoomed using the mouse buttons, but not translated. For all movements the motion option must be active. A further click will return the selection from Motion back to Pick. Similarly, an atom cannot be selected from the structure in Motion mode, only in Pick. An atom is selected in pick using left mouse button. There is no acknowledgement when an atom is selected, although when the mouse is incorrectly placed for atom selection and clicked a warning beep is heard.

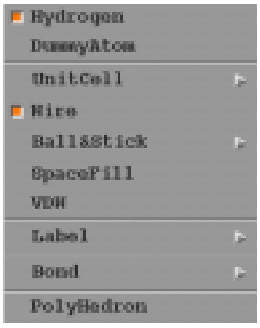
16.12 XXMol Menus

16.12.1 File

This option is used to **Open** or **Save** the selected structure. XXmol will accept many different formats of structure input files: the maXus DrawFile, CCD, PolyGraf, MOPAC input (cartesian), MOPAC output, MOPAC archive, Gaussian input, Gaussian output, MM2 input, MM2 output, PDB and Free format. A number of formats can also be saved from XXmol: DrawFile, CCD, PolyGraf, PDB and Free format. Note that this will only save information such as connectivity and not the actual display that is viewed. This can be saved using the Save/Load button down the left hand side of the XXmol window (see Section 16.13). The **Exit** option closes XXmol.

16.12.2 Display

This pull-down menu enables the user to visualize the structure using different display options such as wireframe or ball & stick. Atom labelling and the editing of bonds are also available. By default the structure is displayed using the wireframe model with hydrogen atoms included. Selected options are shown by indented squares that are switched on and off by highlighting using the mouse. Additional pull-down menus are also available on options where an arrow is displayed on the right side of the main display menu i.e. the UnitCell, Ball&Stick, Label and Bond options.



- **Hydrogen** displays all the hydrogen atoms in the structure that are given in the DrawFile. *This is not an option to add hydrogens to the structure in calculated positions.*
- **Dummy Atom** displays atoms that have no assigned atom type and hence contain no atom type, such as positions given by new peaks in a Fourier map.
- **UnitCell** displays the unit cell edges around the structure. An additional pull-down menu provides the option of highlighting the axis of the *abc* unit vectors from the origin of the unit cell. This axis is displayed by selection of **Axis On**, whereas **Axis Off** displays only the cell edges.
- **Wire** displays the structure as a wireframe model in which the atoms are represented as points at the ends of bonds. This allows a good view of the bonding geometry and is highly recommended for displaying molecular structures in which the bonds are normally well defined.
- **Ball&Stick** displays the atoms as spheres with the bonds drawn as sticks between them. The additional pull-down menu provides the further option of display using the stick only model.
- **SpaceFill** displays the structure using a space filling model in which the atoms are represented as spheres that interpenetrate so that the bonds are not shown.
- **VDW** displays the van der Waals radius of all atoms in the structure in the form of a dotted area around each atom.

Any combination of these display models can be used to represent the structure.

The sizes, color and resolution of the ball & stick, stick and spacefill models, and the density of points used to display the van der Waals radius can be changed using the *Utility/Model Parameter* and *Utility/Display Parameter* menus - see Section 16.12.9.

16.12.3 Display menu: Label

This submenu of the *Display* menu is used to display atom labels on the structure (according to the names given in the DrawFile).

- **Create all atoms** labels all atoms in both the original structure and any other symmetry equivalent molecules already displayed.

- **Delete all atoms** removes labels from all the atoms. Note that this does not actually delete any atoms - just their labels!
- **Create pick atom** labels atoms selected by the user by clicking the left mouse button on the chosen atom. Selection of an atom with a label will remove the label. Although atom labelling can be used with any display model representing the structure, it is recommended that the atom labels are viewed on a wireframe model as this gives the clearest display.

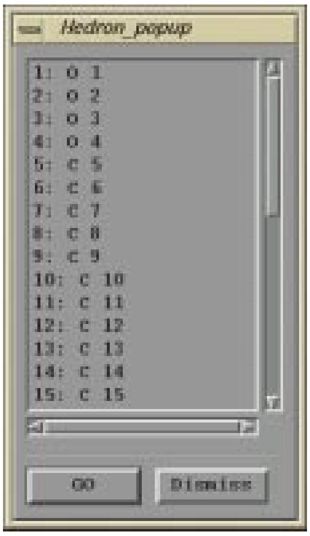
16.12.4 Display Menu: Bond

The structure is displayed as dictated by the connectivity in the DrawFile. This submenu of the *Display* menu allows bonds to be created or deleted.

- **Make all bonds** constructs all bonds between atoms which fall within the default bonding distance for the atoms under consideration. This will also delete any bonds in the display which are outside these default values even if they are dictated in the original DrawFile.
- **Make a bond** constructs a bond between chosen atoms selected by the user by clicking left mouse button on the two atoms to be bonded.
- **Delete a bond** removes the bond between the two selected atoms. XXmol will not automatically bond parts of a structure that are constructed from symmetry equivalents (e.g. two halves of a molecule), this must be done using the **make all bonds** option after the relevant symmetry equivalents have been constructed using the PACKING menu. Bonds are only created or deleted on the structure in the original asymmetric unit and will only effect symmetry equivalents if they are constructed afterwards.

16.12.5 Display Menu: Polyhedron

This option of the *Display* menu creates closed polyhedra around selected atoms chosen from the list of atoms.

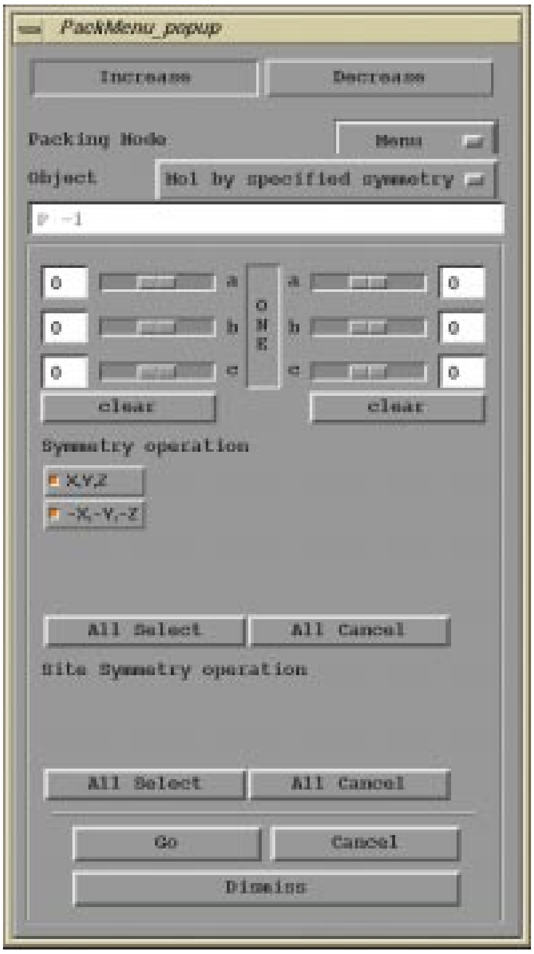


```
1: O 1
2: O 2
3: O 3
4: O 4
5: C 5
6: C 6
7: C 7
8: C 8
9: C 9
10: C 10
11: C 11
12: C 12
13: C 13
14: C 14
15: C 15
```

After the atoms at the centre of each polyhedra have been highlighted in the popup window using the mouse, the selection is applied to the display by pressing GO. This will also generate polyhedra on all symmetry equivalents and any number of polyhedra can be displayed. To add more or to delete existing polyhedra, pull down the display menu to the polyhedra option. When chosen this option will automatically switch off and all polyhedra will be removed from display. To access the polyhedron popup menu, the polyhedra option must be chosen again and atoms then chosen from the list as before. However, those polyhedra displayed in the previous cycle need to be selected again to be retained as the popup has no polyhedra selected by default and only those highlighted in the list will be displayed.

16.12.6 Packing Menu

This enables the display of multiple unit cells and symmetry related structures constructed from the asymmetric unit. Selection of this option initiates the following popup and the display of unit cell axes around the structure.



PackMenu_popup

Packing Mode: Menu

Object: Mol by specified symmetry

P -1

Symmetry operation:
- X,Y,Z
- -X,-Y,-Z

There are two options available for **Packing Mode**: *Menu or Box.* The default setting is *Menu* as shown above. This allows the construction of further structure packing based on the number of unit cells around the structure whereas the *Box* option builds structures based on a box drawn in the centre of the screen. The Box packing mode is described in detail in the next section.

However, in both packing modes the **Increase** button must be selected if adding more symmetry related structures to the display and the **Decrease** button must be selected to allow symmetry related structures to be removed from display. The space group of the structure under consideration is given in the textbox within the window. It cannot be altered here and is dictated by the DrawFile. Before any operation is applied to the structure the user must confirm by pressing *GO*.

There are three options available using the **Object** option:

- **Mol by specified symmetry** constructs structures directly from the asymmetric unit using the symmetry operators selected on the xyz position. These will not necessarily lie within the unit cell.
- **Mol in Cells** constructs all structures which lie partially within the unit cell i.e. will construct complete molecules even if only a fragment of the molecule is within the unit cell.
- **Atom in Cells** constructs all atoms within the unit cell using the symmetry operators selected and will not complete any fragments outside the unit cell axes.

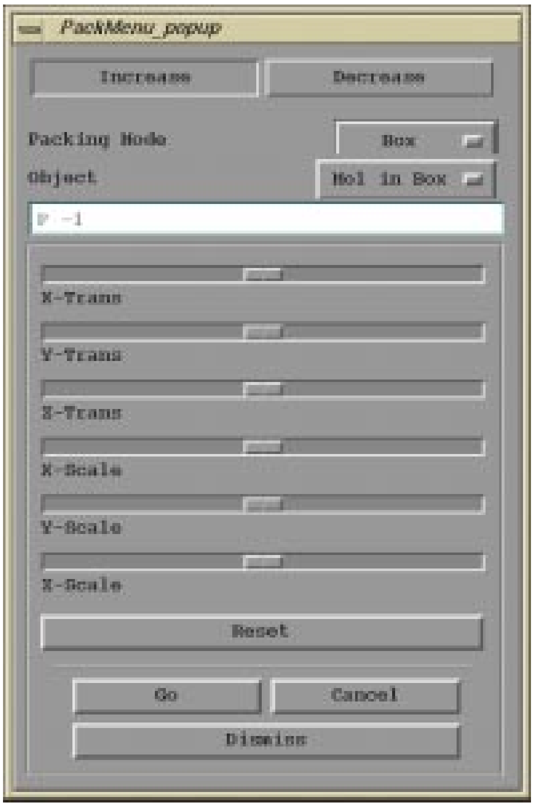
The **Symmetry operation** section lists all the symmetry operations allowed for the selected space group symmetry. These symmetry operations can be selected individually by highlighting the specific operation and clicking with the mouse. Alternatively, all the operations can be selected using **All Select** or all removed from the current selection using **All Cancel**. The Increase option automatically highlights all the symmetry operators whereas Decrease deselects all the symmetry operators.

The structure can be drawn in up to 11 unit cells along each direction using sliders along the a, b and c axes. The two columns of sliders are used to determine the position of the unit cell or range of the unit cells to be shown. These values can be altered either by dragging the sliders or manually in the textfield. If these numbers are changed using the textfield, the user must press <*CR*> after each change to implement any changes. These values signify the origin of the displayed cell (hence 0,0,0 for displaying the original unit cell) and range from -5 to +5 in each direction. There are also **clear** buttons available for both sides which will reset the respective side to all zeros. The vertical button between these button bars is used to display either a single unit cell by selecting **ONE** (as shown in the default popup window) or a number of cells by selecting **MULTI**. The left side controls the minimum value and the right side the maximum value. If the left side is selected greater than the right side the right side will automatically be set equal to the left hence displaying a single unit cell at that position (similarly for the right side). Changes to the number or position of the unit cell displayed are made automatically on selection, as soon as the button is released. When the new unit cells are displayed, the structure (and highlighted symmetries) are constructed in them using *GO*.

All unit cells and symmetry structures except the default are removed from the display using **Cancel**. All values given by the sliders are reset.

The complete display including both structure and unit cell axes can be translated and rotated in the display in the standard way.

Selection of the **Box** option in Packing mode will display a box in the centre of the screen and the following popup menu



Packing Mode: Box
Object: Mol in Box
P -1

The box can be translated in any of the x, y or z directions or scaled in any of these directions using the button bars in the window. These can be moved either by mouse or using the tab and cursor keys. The box must be moved in this way and cannot be altered by clicking and dragging with the mouse on the actual display. The box moves automatically with selection although the structures must be generated by pressing GO. The position of the box is returned to the original position (and scale) using **Reset**.

Rotation or translation using the standard mouse controls will only affect the structure on display and not the box. The box itself is translated and scaled as described above. However, the whole display (structure and box) can be rotated using either the Ctrl key and left mouse button or the Ctrl key and the middle mouse button. Subsequent translations of the structure alone using the mouse will go with the rotated axis system and not with the screen until the packing popup is dismissed. The combined structure and box can be scaled in the usual way.

Increase will add structure to the display wherever the box is positioned and **Decrease** will remove it (applies to mol and atom in box). Note that selection of Increase or Decrease will automatically reset the box to its original position and size so the option should be selected before positioning of the box.

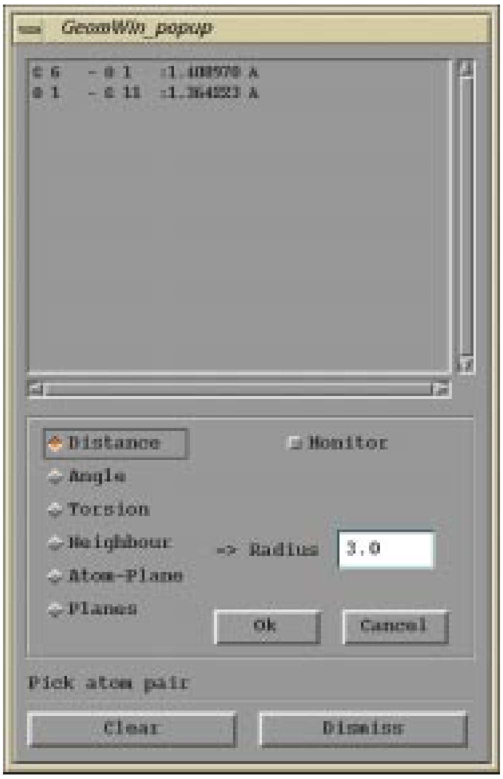
There are two options available under **Object**:

- **Mol in Box** constructs complete molecules that have any atoms within the box using all symmetry operators allowed by the space group. As described in the previous Packing Mode this completes fragments outside the box.
- **Atom in Box** constructs any atoms within the box using all symmetry operators allowed by the space group but will not complete fragments outside the box.

The box and structures except the default are removed from the display using **Cancel** and all values are reset. The default molecule that is originally displayed on selection of the popup menu will be deleted from the display on construction of further structure if it is not within the box.

16.12.7 Geometry Menu

Any distance or angle within the structure or between planes can be calculate and displayed.



```
C 6   - O 1   :1.408970 A
O 1   - C 11  :1.364223 A
```

Radius: 3.0

Pick atom pair

The option required is selected by left mouse button and is then highlighted. Atoms are then selected using left mouse button and selection prompted in window.

- **Distance** is displayed in Angstroms after selection of atom pair.
- **Angle** is displayed in degrees after selection of three atoms forming chosen angle.
- **Torsion** angle is displayed in degrees after selection of four atoms forming angle.
- **Neighbour** calculates the distances to all atoms not bonded to the selected atom within the radius specified.

The results are displayed immediately after selection of atoms. However, the user must click *OK* to signify end of selection of a plane.

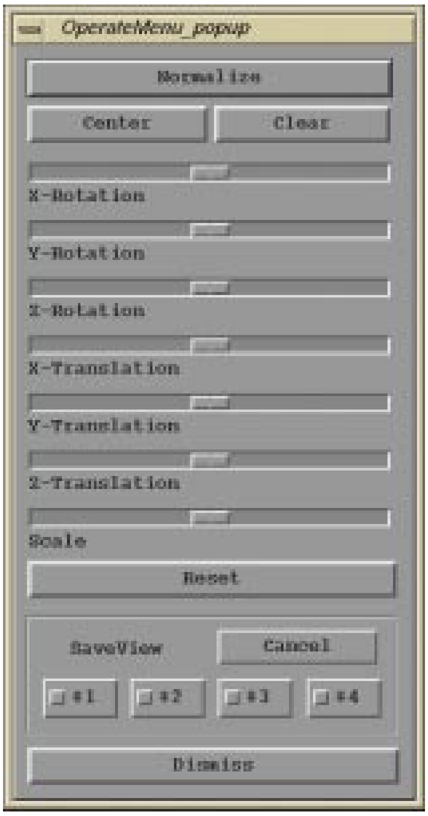
- **Atom-Plane** calculates the perpendicular distance between a chosen atom and a plane, and the angle that the atom forms to the centre of the plane. Firstly the atom under consideration must be selected, after which the user is prompted to select atoms (>2) defining the plane. When all atoms defining the plane have been selected press OK to signify the end of selection and the calculation is then carried out.
- **Plane** calculates distance between the centres of two selected planes and the angle between them. The user is prompted to select a number of atoms (>2) defining No.1 Plane after which press OK to signify the end of this selection and to start definition of No.2 plane. Click *OK* to end the selection process and begin calculation.
- **Monitor** is used to display selected distances (as dotted lines) and planes (as grey areas) - including atom-plane calculation. Distances listed using the NEIGHBOUR option are also displayed. The monitor option can be toggled on and off, and will display all the distances and planes listed in the window.

If an incorrect atom is selected for the geometry calculation, the last atom to be selected can be removed from the selection using **Cancel**. Note that this does not clear all atoms in the current calculation *e.g.*; if 3 atoms have been selected in calculation of a torsion angle, pressing Cancel will remove the 3rd atom from the calculation while the first two are retained. Two further atoms describing the angle must then be selected.

The list of distances and angles displayed in the window are removed using **Clear**. This does not remove any atoms currently selected in a calculation. Further calculations can be carried out although no prompt for picking atoms will be given.To get this prompt the geometry option within the window must not only be highlighted but must also have the box around it as shown above.

16.12.8 Operate Menu

This allows menu driven movement of the structure view and enables control without use of a mouse.



The structure can be rotated, translated and scaled using the sliders shown above. When the bar reaches the end of the track it is automatically reset to the centre to allow further movement in both directions. The end of each track represents an exact rotation of 180 degrees. Rotation is usually around the centre of the molecule.

The display can be **Reset** to the original view (given in the DrawFile) and all sliders reset to the centre. (This does not reset the centre of rotation to the default setting if an atom has been chosen as the centre).

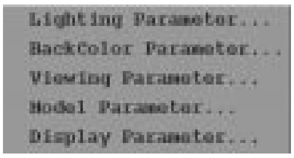
A view can be rescaled and re-translated back into the centre of the window using **Normalise**.

Center allows the user to select an atom around which the structure will subsequently rotate, by clicking on the chosen atom in the display. This selection can be removed using **Clear**, giving rotation around the centre of the structure or enabling the choice of another atom.

Up to four views can be temporarily saved using **Save View**. Once the chosen view has been obtained any unhighlighted button #1 to #4 can be selected - this will then retain the view. When highlighted, the view is re-displayed by simply pressing the button again. The views are cleared using **Cancel**. These views will be retained while XXmol is being used, however, once the program has been exited the views will be lost.

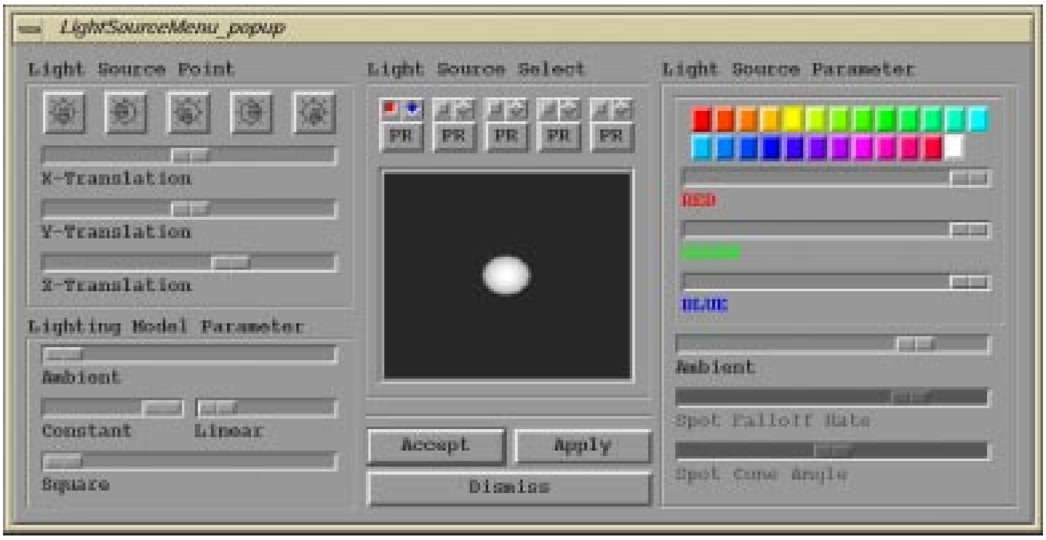
16.12.9 Utility Menu

This pull-down menu gives access to a number of popup menus used to control various graphics display capabilities such as lighting, color, resolution and the use of clipping planes.



- Lighting Parameter

  This option allows variation of the colour, position and various other properties of the light source used in the display:



The position of the light source is illustrated in the pop-up window as a small white circle lighting a sphere in the centre of the window. This gives an idea of the light source options that have been selected before they are implemented in the display using **Apply**. **Accept** will apply these changes to the display and automatically dismiss the popup window.

Various properties of the light source can be varied:

Light Source Point

This dictates the position of the light source and is controlled by buttons and bars. The buttons place the light source either in the centre or at a corner of the screen according to the symbol on the button i.e. centre, bottom right, top right, bottom left and top left respectively. The bars allow translation of the source in any of the x, y or z directions. The default position is shown at the centre of the screen and pointing from the monitor screen towards the back of the display.

Light Source Select

There are 5 possible light sources that can be used. These are switched on by highlighting the red square in the top left hand corner of each, then the properties of each can be altered one at a time selected using the blue diamond in the top right. The type of each light can also be selected:

PR (default setting) gives a directional light source, treated as though it is infinitely far away giving the effect that the rays of light can be considered parallel by the time they reach the object.

PT gives a positional light source as its exact position within the scene determines its effect and specifically the direction from which the light rays come. This positional light radiates in all directions.

SP gives a spot light which is effectively a positional light source with a restriction on the shape of the light so that it emits to a cone.

Each of these light sources can be moved and coloured independently. The source under consideration is displayed in the central window looking larger and brighter than the others.

Lighting Model Parameter

This allows variation of the amount of ambient light contributed from the light source, and the attenuation of the light intensity. This is varied using sliders controlling the constant, linear and quadratic parameters of the attenuation factor and multiplication of the contribution of the light source. Since a directional light source is infinitely far away, there is no point in attenuation of its intensity over distance, so this has no effect in the PR option, however, you might want to attenuate the light from a positional source.

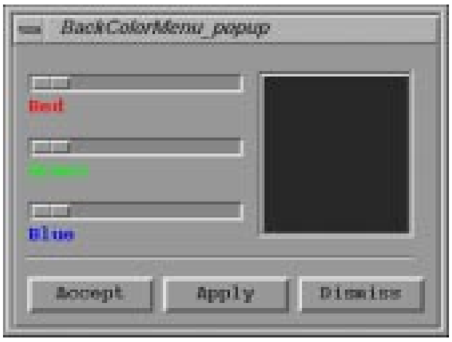
Light Source Parameter

This is used to alter the colour of the light source and the global ambience of the scene *i.e.* ambient light that is not from any particular source. The colour of the light source can be selected either using the range of coloured buttons to select one of the colours displayed or using the bars to vary the amount of red, green or blue used to generate the colour. The numbers of colour correspond to a percentage of full intensity for each colour. The full amount of all three generates brightest white whereas none of all three generates black. The default light source colour is white.

Spot Falloff Rate controls how concentrated the light is. The light intensity of the light is highest in the centre of the cone and is attenuated toward the edges of the cone. A higher falloff rate results in a more focused light source, and **Spot Cone Angle** (this specifies the angle between the axis of the cone and a ray along the edge of the cone and is restricted in the range 0 to 90 degrees. The angle of the cone at the apex is then twice this value.) are options if the light source under consideration has been specified a spotlight.

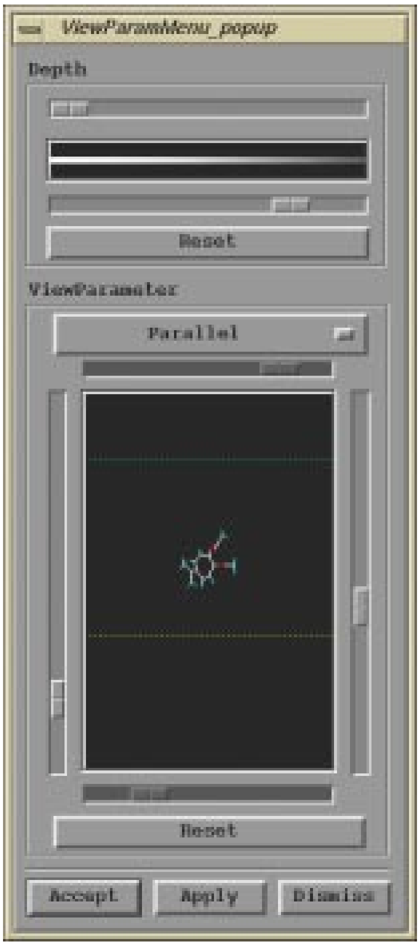
- **BackColor Parameter**

This option allows variation of the background colour of the display and is controlled by sliders to vary the percentage of full intensity for red, green and blue in a similar way to colour of the light source described above.



- **Viewing Parameter**

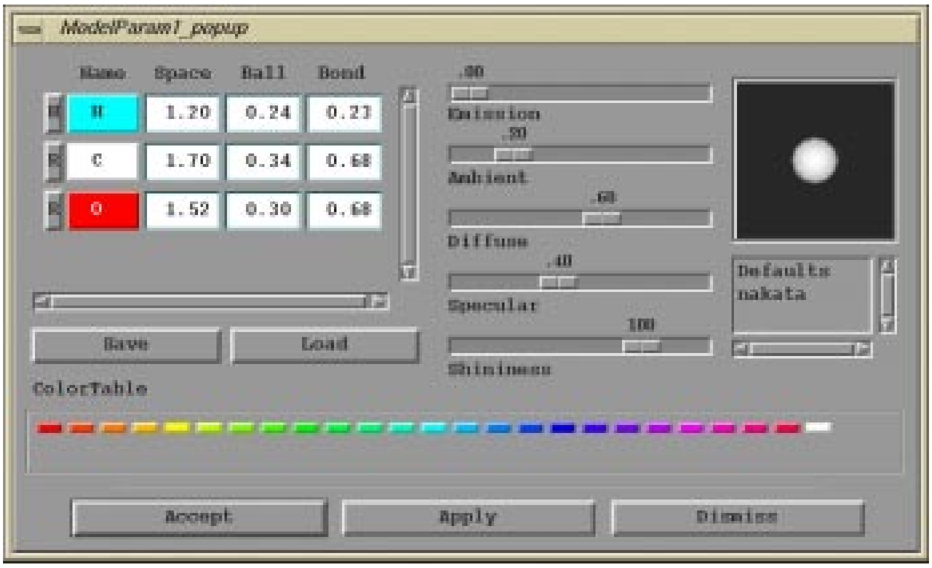
This popup menu controls the clipping planes used in display and provides a small view of the structure within the window showing the position of the clip planes being used.



By default the clip planes are set parallel with the top clip plane (green) controlled by the left slider and the lower clip plane (yellow) by the right slider. The view parameter can also be used to display the structure in perspective view. Depress the *Parallel* button to get the Perspective option. This displays two additional dotted lines showing the current perspective of the display. Two additional sliders become active, the top one alters the viewing distance whereas the lower alters the angle of the perspective. The **Reset** button can be used to place the clipping planes in a position where the complete molecule is displayed. You must click **Apply** before any changes are implemented in the display. **Accept** will apply these changes to the display and automatically dismiss the popup window.

- **Model Parameter**

This option allows the colours of the atoms, the size of balls and spacefill display and the lighting effects of the structural model to be varied. The view can first be seen in the display shown within the window before being applied.



ModelParam1_popup

Name	Space	Ball	Bond
H	1.20	0.24	0.23
C	1.70	0.34	0.68
O	1.52	0.30	0.68

Emission .00
Ambient .20
Diffuse .60
Specular .40
Shininess 100

Defaults
nakata

Each atom is automatically assigned an atom type (according to the DrawFile) as hydrogen (H), a regular atom type (R) or as a dummy atom (P) and can then be displayed (or not) using the main display pull-down menu. All three options are provided by clicking on the grey button next to each atom name. The colour of each atom type can be changed by clicking on the required colour in the atom table and then on the button with the atom name. Note that the colour selected for each atom corresponds to the percentage of the incoming red, green and blue light that it reflects. For example, a perfectly red atom reflects all the incoming red light and absorbs all the green and blue light that strikes it. Hence, if the atom is viewed in white light or red light, it appears red, whereas in pure green light it appears black (all incoming light is absorbed).

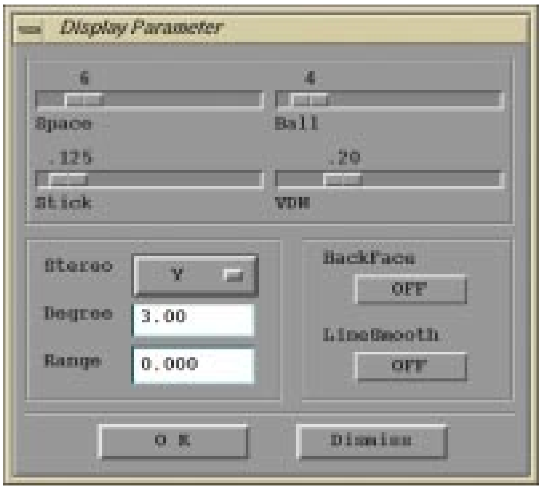
The size of the ball and spacefill used in display can also be altered for each atom type by alteration of the values listed (in Angstroms). For these changes to take effect, must press <*CR*> after each edit of text and then **Apply**.(If two textboxes are altered and <*CR*> pressed after only one, only that one will then take effect in the display. These settings for sizes and colours of the atom types can be saved and used in the display of other structures using the *Save* and *Load* options shown.

This menu also allows specification of how the structural model reflects light and hence what material it appears to be ‘made of’. This is done by variation of the model’s emitted, ambient, diffuse, and specular light, that define how shiny it is. *Emission* is used to make the structure appear to be giving off light as the light originates from the structure and is unaffected by any light sources. *Ambient* reflectance is affected by the global ambient light and ambient light from individual light sources whereas the *diffuse* reflectance is brightest where the structure is directly illuminated. *Specular* reflection from the structure produces highlights and depends on the location of the viewpoint as it is brightest along the direct angle of reflection. *Shininess* is then used to control the size and brightness of the highlight - the higher the value, the more focused the highlight.

Examples of these light settings have been saved in preset files - Defaults gives the settings shown in the window whereas nakata has more ambience and emission. These can be selected by highlighting and clicking with the mouse.

- **Display Parameter**

This allows the variation of parameters used in the modes of displaying the structure



Space 6
Ball 4
Stick .125
VDW .20
Stereo Y
Degree 3.00
Range 0.000
BackFace OFF
LineSmooth OFF

The default values are shown in the above window.

- **Space** is used to vary the resolution of the spacefill display.
- **Ball** varies the resolution of the balls in the ball & stick display.
- **Stick** used to vary the radius of the stick in the ball & stick display from 0100 to 0.500A.
- **VDW** varies the number of points (density) used in the display of van der Waals radius around each atom in the range 0.10 to 0.50. The lower the number the more points there are.

Stereo controls the degree of offset and range of stereo projections which are displayed from the ModelSelect menu. The **Degree** and **Range** values are altered by manual selection followed by *OK*. Stereo can be altered either before or after the stereo projections are actually displayed. **Range** controls the distance the images are apart - a larger number implies that they are to be closer together (Negative values are allowed). **Degree** controls the angle to which the left structure is offset, can be in X, Y or Z direction. If the stereo type is selected as **Crossed** rather than **Parallel,** the projection is rotated in the opposite direction. These values are then retained as stereo type is switched on and off and are only automatically reset as XXmol is restarted.

BackFace enables a view of the backface of bonds in stick mode. In this mode one can see the atom positions and wireframe within the cylinder and the backface of the bond.

LineSmooth smooths jagged looking lines of bonds and effects wireframe and stick bonds.

The display does not change automatically with selection so you must press *OK* to confirm and to implement changes.

16.12.10 Select Menu

This allows the display of stereographic projections and multiple structures within the same window.


Select Window type: Standard
Select Stereo type: Normal

More than one structure can be displayed in the window at one time. For each structure opened there are a row of buttons for independent selection, operation and display of the molecule. Each structure can be displayed or not by selection of **Vis** (the blue square) and any operation such as rotation or translation can apply to any number of the structures by selection of **Ope** (the pink square). Only one DrawFile can be selected at a time and this allows alteration of the display of the structure selected (such as van der Waals or Ball&Stick). **Whole** sets all operations (i.e. the movement of the display) to effect all structures although individual structures must still be selected for display variation. Lighting and other utility parameters will still effect the complete display.

There are a number of stereographic projections available using **Select Stereo type**.

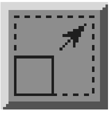
- **Normal** displays only the one structure and is the default setting.
- **Parallel** displays the standard stereographic projections
- **Crossed** displays the stereographic projections rotated in the opposite direction

  The distance, angle and direction of rotation between the projections is defined in the Display Parameter popup within the Utility menu.
- **RealVision** displays the stereographic projections in the form for viewing with LCD viewing glasses

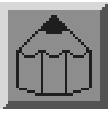
16.13 XXMol Buttons

A number of buttons are also available down the left side of the XXmol display window providing further functions.

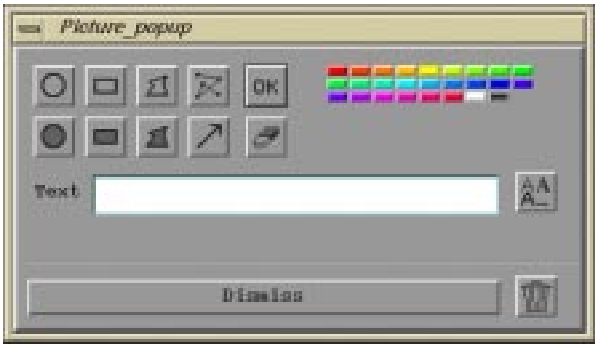
1) The following button makes the display full screen size allowing photography etc. Any mouse or keyboard operation returns the display to the normal display window.



2) This button permits the annotation popup to be used to add text or further graphics to the display. *In order to display this annotation, the display must be switched to a multi-type window using the Select popup menu.*



The following options are available:



Circle - after the circle button (boundary or filled) and colour have been selected the left mouse button is clicked first at the position in the display where the circle is to be centred, and a second time for the outer edge. The cursor position is then shown by a green cross. Another circle can then be draw in the same way. If the *OK* button is selected after only the centre of the circle has been chosen, the edge of the circle will automatically be drawn to the same as the last outer edge specified.

Rectangle - the position of the rectangle is defined by construction using the two positions selected by mouse as opposing corners of the rectangle. If *OK* is selected after only one position has been chosen, a rectangle will be drawn from the last corner selected.

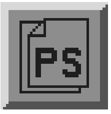
Polygon - the polygon can be defined using any number of points and is initially constructed from the first three points chosen.

Arrow - an arrow is constructed between two selected points with the latter selection acting as the head of the arrow.

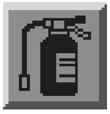
Text may also be added to the display in various fonts supplied by pressing the button next to the textfield. As in other textfields, the user must press <*CR*> after editing the textfield in order for the changes to take effect.

All annotation can be removed from the display using the waste button in the bottom right of the window.

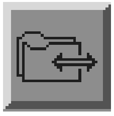
3) This button creates postscript file called XXmol.ps of the display and prints a hardcopy of the display on the default printing device.



4) The following dismisses all popup windows on display:



5) This saves or loads the view being displayed; i.e. stores orientation, lighting, packing settings, etc:



16.14 XXMol (SG) Known Problems

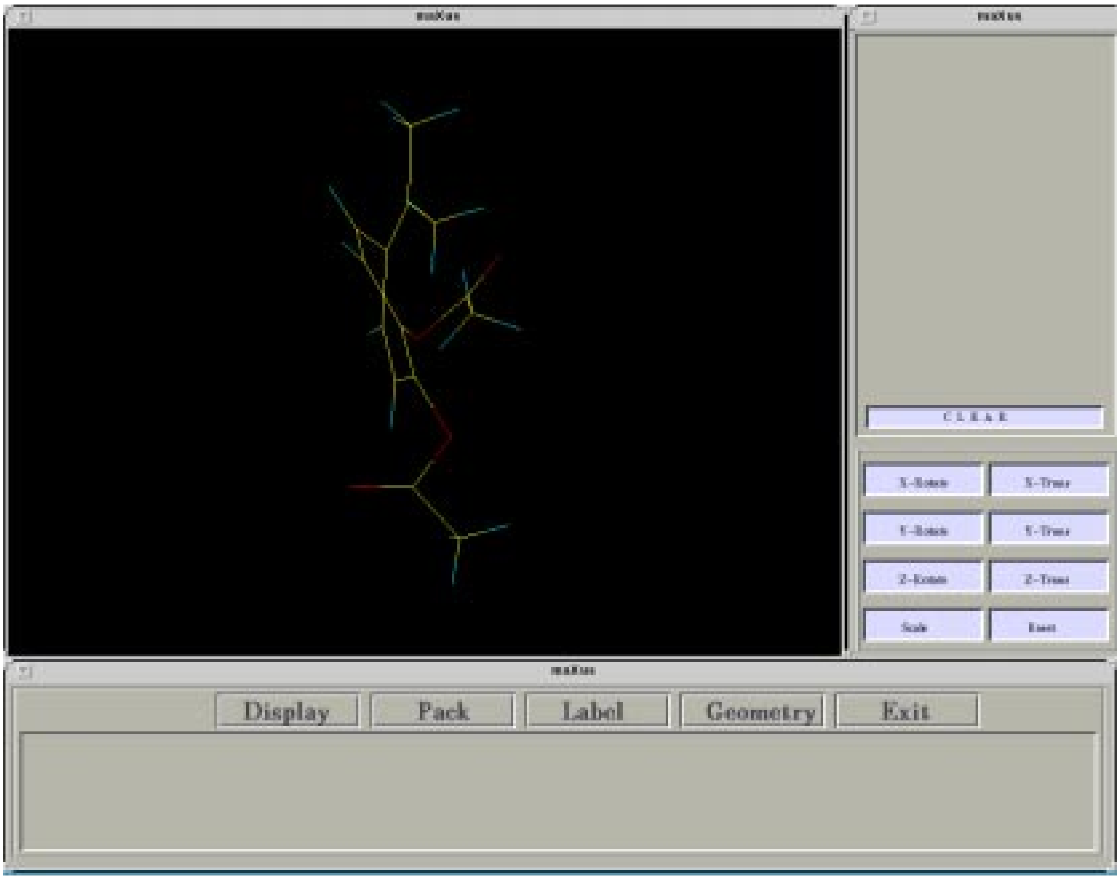
In some cases the default display of a structure directly from an input file may be incomplete or the display may even appear blank. This is caused by incorrect placement of the default clipping planes for the structure under consideration. The problem is easily corrected by resetting the clip planes in the viewing parameters in the Utility menu. The structure can then be reset to a reasonable size and position in the display using a combination of Normalize and the translation and scale sliders given in the Operate menu.

16.15 XXMOL - for LINUX and SUN

This version of XXMol is a version which does not use the GL graphics library. Consequently, it has much more limited features than the SG version; it does however run on Linux systems.

Launch XXMol by selecting either *View Model* or *View Expanded* from the *XXMol* submenu of the *Graphics* menu.

The following windows will appear:



16.15.1 Molecule Rotation

Using the buttons on the right-hand side of the main display window, the picture of the molecule can be manipulated in the normal fashion.

Note that where you click on the button affects what happens - clicking on the right side of the button moves the molecule in one direction, and the left side moves it in the other. The closer to the edge of the box you click, the faster the molecule moves/rotates etc.

X, Y, and Z-rotate buttons - rotate the molecule on the desired axis.

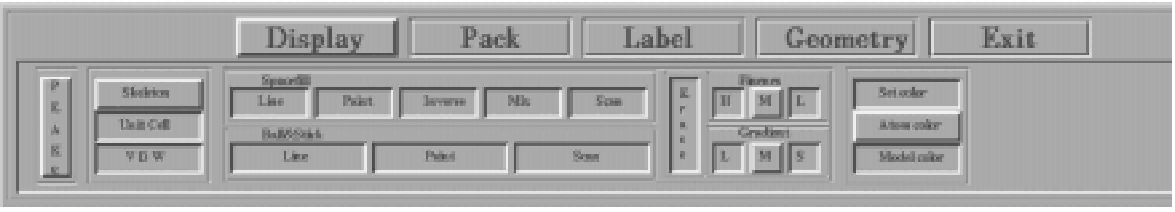
X, Y, and Z-Trans buttons - translates the molecule along the desired axis.

Scale button - zooms in and out of the molecule.

Reset button - returns to the default view of the molecule.

16.15.2 Display Options

Click on the *Display* button. The following options appear in the lower part of the window:



Peaks button: if you have unassigned peaks in the model, this sets whether or not they are displayed.

Skeleton / Unit Cell / VDW buttons:

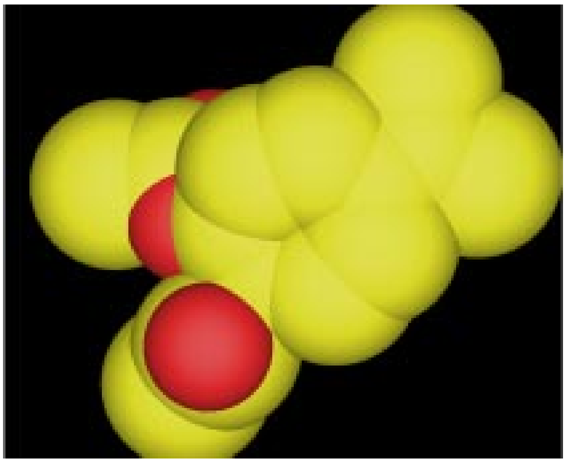
- **Skeleton** displays the structure as a wireframe model in which the atoms are represented as points at the ends of bonds. This allows a good view of the bonding geometry and is highly recommended for displaying molecular structures in which the bonds are normally well defined.
- **UnitCell** displays the unit cell edges around the structure.
- **VDW** displays the van der Waals radius of all atoms in the structure in the form of a dotted area around each atom.

Spacefill options:

SpaceFill displays the structure using a space filling model in which the atoms are represented as spheres that interpenetrate so that the bonds are not shown.

- Line: outlines the van der Waals radii of the atoms with a single, solid line.
- Paint: Lightly shades the van der Waals radii of the atoms with coloured dots.
- Inverse: Makes light colours dark & dark colours light.
- Mix: Combination of Line & Paint.
- Scan: Renders the surface of the van der Waals radii.

For example, the skeleton shown earlier would produce the following spacefilled picture:



Ball & Stick options:

Ball&Stick displays the atoms as spheres with the bonds drawn as sticks between them. The additional pull-down menu provides the further option of display using the stick only model.

- Line: shows the skeleton of the molecule surrounded by an outline of the ball & stick version.
- Paint: Sticks are crudely shaded; balls are still just outlines.
- Scan: Balls & sticks nicely rendered in full colour.

Erase button: resets the molecule to the default view.

Fineness control: controls the quality of the rendering; options are:

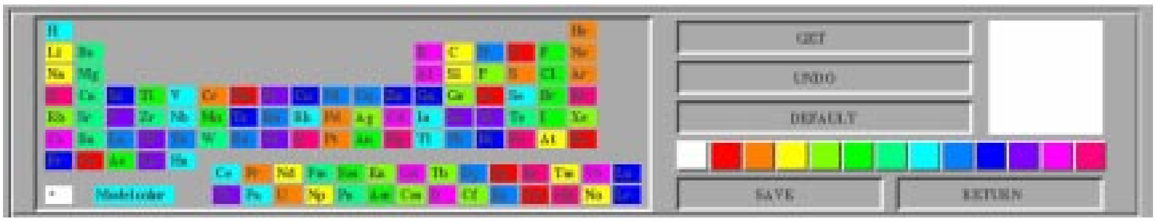
H - High, M - Medium, L - Low.

Gradient control: controls the shading and light level for the rendering; options are:

L - light, M - medium, S - strong.

Set Colour button:

This produces another options window:



To change the colour assigned to a particular element, click first on the colour you would like from the selection on the right hand side, and then on the symbol of the element you would like to set to that colour.

Clicking *Undo* undoes the last thing you changed.

Clicking *Save* stores the new colour settings for use.

Click *Return* when you are finished to go back to the main screen.

Atom Colour / Model Colour buttons:

These buttons are mutually exclusive.

If *Atom Colour* is on, the different atoms in the model are coloured according to the settings chosen in the *Set Colour* window.

If *Model Colour* is on, the entire model is displayed in the same colour. This colour can be altered from the *Set Colour* window, by clicking on the Model option rather than an element symbol.

16.15.3 Pack Options

Clicking on the *Pack* button produces the following:

This series of options enables the display of multiple unit cells and symmetry related molecules constructed from the asymmetric unit. Molecules that exploit crystallographic symmetry can also be constructed from their asymmetric unit.

The space group of the structure is displayed in the top leftmost box of this part of the screen (P-1 in this example).

To use the Pack option, first select from *Addition* or *Eliminate* -the first adds symmetry related atoms/molecules to the display, and the latter removes them.

Before any operation is actually applied to the structure, it must be confirmed by pressing *Go*.

The options in the left side of the window, below the space group indicator, control the number of adjacent unit cells to use.

The structure can be drawn in up to 9 unit cells along each direction using sliders along the a, b and c axes. The two sets of numbers are used to determine the position of the unit cell or range of the unit cells to be shown. These values can be altered as follows:

Left side: Clicking on the button with the left mouse button decreases the value in the box; clicking with the middle mouse button increases it.

Right side: Clicking on the button with the left mouse button increases the value in the box; clicking with the middle mouse button decreases it.

If these numbers are changed using the textfield, the user must press <*CR*> after each change to implement any changes. These values signify the origin of the displayed cell (hence 0,0,0 for displaying the original unit cell) and range from -4 to +4 in each direction. There are also *Clear* buttons available for both sides which will reset the respective side to all zeros. The vertical button between these button bars (*Same)* is used to set both sides to the same value. The left side controls the minimum value and the right side the maximum value. If the left side is selected greater than the right side the right side will automatically be set equal to the left, thus displaying a single unit cell at that position (similarly for the right side). Changes to the number or position of the unit cell displayed are made when you press *Go*.

All unit cells and symmetry structures except the default are removed from the display using *Cancel*. All values given by the buttons are reset.

The right hand side of the display is the symmetry operation section. It lists all the symmetry operations allowed for the selected space group symmetry. These symmetry operations can be selected individually by clicking on the specific operation with the mouse. Alternatively, all the operations can be selected using *All* or all removed from the current selection using *Clear.*

16.15.4 Label Options

Clicking on the *Label* button produces the following menu:

This controls how the individual atoms in the structure are labelled.

First select a Label colour: you can select from *White*, where the labels appear in white, or *By Model*, in which case the labels are the same colour as the atoms to which they are attached.

Now, you can choose from:

- *Atom Pick* - if this option is selected, you can label individual atoms simply by clicking on them in the main graphics window.
- *All Atoms* - selecting this option labels all the atoms in the molecule.

- *Clear* - All labels are removed.

16.15.5 Geometry Options

Clicking on the *Geometry* button produces the following:

Any distance or angle within the structure or between planes can be calculate and displayed.

The option you select is highlighted. Instructions as to how many atoms you then need to select appears in the centre panel. Atoms are then selected by clicking on them using the mouse.

- *Distance* is displayed in Å after selection of atom pair.
- *Angle* is displayed in degrees after selection of three atoms forming chosen angle.
- *Torsion* angle is displayed in degrees after selection of four atoms forming angle.
- *Plane* calculates distance between the centres of two selected planes and the angle between them. The user is prompted to select a number of atoms (>3) defining the first plane after which press the middle mouse button in the central text pane to signify the end of this selection and to start definition of the second plane. Click again with the middle button to end the selection process and begin calculation.
- *Atom-Plane* calculates the perpendicular distance between a chosen atom and a plane, and the angle that the atom forms to the centre of the plane. Firstly the atom under consideration must be selected, after which the user is prompted to select atoms (>3) defining the plane. When all atoms defining the plane have been selected press the middle mouse button to signify the end of selection, and the calculation is then carried out.
- *Neighbour* calculates the distances to all atoms not bonded to the selected atom within a specified radius. After clicking on your chosen atom, you will be prompted to enter a radius, in Angstroms, into a field that appears in the central pane.

The output appears in the top right window, and may be deleted by clicking on *Clear.*

16.15.6 Exit

Clicking on the *Exit* button returns you to the main maXus window.

16.15.7 Output

Currently, the only way to output graphics from this version of XXMol is to use a screen capture utility such as SnapShot.

16.16 RASMOL v2.6

This program is provided free of charge to all maXus users.

Launch RasMol by selecting either *View Model* or *View Expanded* from the *RasMol* submenu of the *Graphics* menu.

All rights reserved. Use of copyright notice does not imply publication or disclosure. The information supplied in this document is believed to be true but no liability is assumed for its use or for the infringements of the rights of the others resulting from its use. Information in this document is subject to change without notice and does not represent a commitment on the part of the supplier.

This package is sold/distributed subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out or otherwise circulated without the supplier's prior consent, in any form of packaging or cover other than that in which it was produced. No part of this manual or accompanying software may be reproduced, stored in a retrieval system on optical or magnetic disk, tape or any other medium, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise for any purpose other than the purchaser's personal use without contacting the author.
This product is not to be used in the planning, construction, maintenance, operation or use of any nuclear facility nor the flight, navigation or communication of aircraft or ground support equipment. The author, nor his employer, shall not be liable, in whole or in part, for any claims or damages arising from such use, including death, bankruptcy or outbreak of war.

16.17 RasMol Introduction

RasMol2 is a molecular graphics program intended for the visualisation of proteins, nucleic acids and small molecules. The program is aimed at display, teaching and generation of publication quality images. The program reads in molecular co-ordinate files and interactively displays the molecule on the screen in a variety of representations and colour schemes. Supported input file formats include Brookhaven Protein Databank (PDB), Tripos Associates' Alchemy and Sybyl Mol2 formats, Molecular Design Limited's (MDL) Mol file format, Minnesota Supercomputer Centre's (MSC) XYZ (XMol) format and CHARMm format files. If connectivity information is not contained in the file this is calculated automatically. The loaded molecule can be shown as wireframe bonds, cylinder 'Dreiding' stick bonds, alpha-carbon trace, space-filling (CPK) spheres, macromolecular ribbons (either smooth shaded solid ribbons or parallel strands), hydrogen bonding and dot surface representations. Different parts of the molecule may be represented and coloured independently of the rest of the molecule or displayed in several representations simultaneously. The displayed molecule may be rotated, translated, zoomed and z-clipped (slabbed) interactively using either the mouse, the scroll bars, the command line or an attached DialBox. RasMol can read a prepared list of commands from a 'script' file (or *via* inter-process communication) to allow a given image or viewpoint to be restored quickly. RasMol can also create a script file containing the commands required to regenerate the current image. Finally, the rendered image may be written out in a variety of formats including either raster or vector PostScript, GIF, PPM, BMP, PICT, Sun rasterfile or as a MolScript input script or Kinemage.

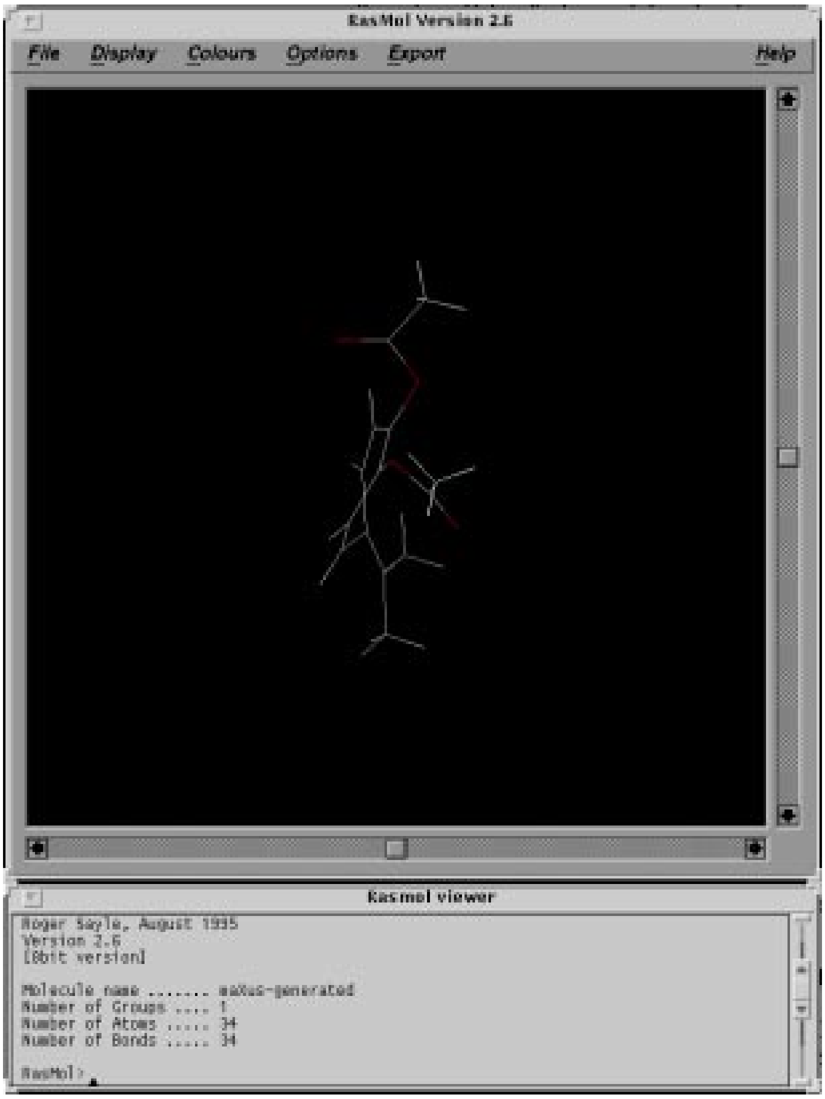
RasMol has been developed at the University of Edinburgh's Biocomputing Research Unit and the BioMolecular Structure Department, Glaxo Research and Development, Greenford, U.K. If you have any comments, questions, suggestions or complaints please do not hesitate to contact the author:

Roger Sayle, Email: **ras32425@ggr.co.uk**
Biomolecular Structure,**rasmol@dcs.ed.ac.uk**
Glaxo Research and Development,
Greenford Road, Greenford,Tel:(+44) (0)81 966 3567
Middlesex UB6 0HE.Fax:(+44) (0)81 966 4476
U.K.

16.18 RasMol General Operation

16.18.1 RasMol Windows

On all platforms RasMol displays two windows, the main graphics or canvas window with a black background and a command line or terminal window.



```
Roger Sayle, August 1995
Version 2.6
[8bit version]

Molecule name ....... maXus-generated
Number of Groups .... 1
Number of Atoms ..... 34
Number of Bonds ..... 34

RasMol>
```

At the top of the graphics window (or at the top of the screen for the Macintosh) is the RasMol menu bar. The contents of the menu bar change from platform to platform to support the local user interface guidelines, however all platforms support the 'File', 'Display', 'Colours', 'Export' and 'Options' pull-down menus. The Main graphics window also has two scroll bars, on the right and one at the bottom, that may be used to rotate the molecule interactively.

While the mouse pointer is located within the graphics area of the main display window, the mouse pointer is drawn as a cross-hair cursor, to enable the 'picking' of objects being displayed (see later); otherwise the mouse pointer is drawn as an arrowhead. Any characters that are typed at the keyboard while the display window is in 'focus' are redirected to the command line in the terminal window. Hence you do not need to continually switch focus between the command line and graphics windows.

The display window may be resized at any point during the session. This has the effect of simply rescaling the image displayed on the canvas. RasMol imposes limits on the size of the display window such that the window must be large enough to display the menu and scroll bars and yet small enough to fit on a single screen. Attempts to enlarge the screen may fail owing to insufficient memory on the host machine, in which case RasMol reports the error message '*Renderer Error: Unable to allocate frame buffer!*' or some similar error.

On eight bit displays, when the number of colours required by the program exceeds the number of free colours on the screen, the program uses its own colourmap. This has the effect of temporarily displaying all windows other than the display window in false colours while the mouse pointer is within the display windows. If the mouse pointer is moved outside the display windows, the original colours of the other windows return, and the image on the canvas is shown in 'false colour'. Once the number of colours required by the program drops again, the presentation of colours returns to normal.

16.18.2 RasMol Menus

Several options are available from the menus provided.

- **File Menu**

  Open: Although the current Model file is opened as a default when you launch RasMol from maXus, you can also view other files using this option. Note that if a file is already open in RasMol, you must *Close* it before another can be opened.

  Save As: This option allows you to save a copy of the file you are working on under a different name. Note that RasMol will ask you fto enter a file name in the text pane, and does not bring up any dialog boxes.

  Close: Closes the currently displayed molecule. RasMol then shows a blank screen.

  Exit: Quits the RasMol program, and returns to the maXus menu.

- **Display Menu**

  Wireframe: displays the molecule as in Fig. 1, below.

  Sticks: displays the molecule as in Fig. 2, below.

  Spacefill: displays the molecule as in Fig. 3, below.

  Ball & Stick: displayst the molecule as in Fig. 4, below.

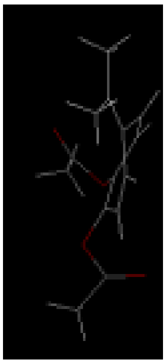


Fig. 1

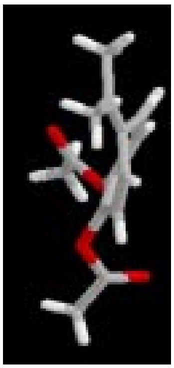


Fig. 2

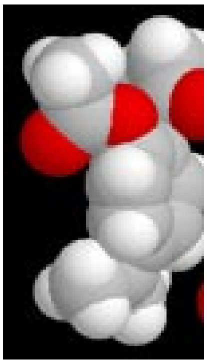


Fig. 3

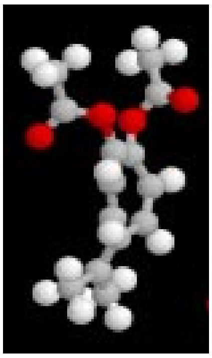


Fig. 4

The *Ribbons, Backbone, Strands,* and *Cartoons* options are only used with proteins.

- **Colours Menu**

  The options here control the colours used to display the molecule in RasMol; altering this does not affect the colours assigned in the Model window.

- **Options Menu**

  Slab Mode: toggles Z-clipping on or off

  Hydrogens: selecting this removes any hydrogens from the display. Note that to view the molecule with hydrogens again, you must exit and re-launch the program.

  Heteroatoms: see the RasMol command ‘set Hetero’ - Section 16.21.9.

  Specular: this causes highlights to appear as white reflections of the light source on the surface of a rendered molecule.

  Shadows: in Spacefill mode, renders the molecule with shadows if toggled on.

  Stereo: displays a stereo view of the molecule

  Labels: labels all non-C atoms.

- **Export Menu**

  This menu allows you to save a copy of the molecule image in one of a number of formats.

  Select from *Gif, PostScript, PPM, IRIS RGB, Sun Raster, or Pict*. You will be asked to enter an appropriate filename in the RasMol text pane. No dialog box is displayed.

16.18.3 Scroll Bars

The scroll bar across the bottom of the canvas area is used to rotate the molecule about the y-axis, i.e. to spin the nearest point on the molecule left or right; and the scroll bar to the right of the canvas rotates the molecule about the x-axis, i.e. the nearest point up or down. Each scroll bar has a 'indicator' to denote the relative orientation of the molecule, which is initially positioned in the centre of the scroll bar. These scroll bars may be operated in either of two ways. The first is by clicking any mouse button on the dotted scroll bar background to indicate a direct rotation relative to the current indictor position; the second is by clicking one of the arrows at either end of the scroll bar to rotate the molecule in fixed sized increments. Rotating the molecule by the second method may cause the indicators on the scroll bars to wrap around from one end of the bar to the other. A complete revolution is indicated by the indicator travelling the length of the scroll bar. The angle rotated by using the arrows depends upon the current size of the display window.

16.18.4 Picking

In order to identify a particular atom or bond being displayed, RasMol allows the users to 'pick' objects on the screen. The mouse is used to position the cross-hair cursor over the appropriate item, and then any of the mouse button is depressed. Provided that the pointer is located close enough to a visible object, the program determines the identity of the nearest atom to the point identified.

The program will display, in the terminal window, the atom's type, serial number, residue name and residue number. If the atom is a member of a named chain, the chain identifier is also displayed. Two examples of the output generated by selecting an atom are displayed below:

```
Atom: CA 349 Group: SER 70
Atom: O  526  Hetero: HOH 205  Chain: P
```

The first line describes the alpha carbon of the serine-70 amino acid in a protein. The unique Brookhaven serial number for this atom is 349. The following line describes the oxygen atom in a water molecule attached to the P chain of the main molecule. The word 'Hetero' distinguishes heterogeneous molecules (such as cofactors) from the residues in the main molecule, noted by 'Group'. [These two atoms are referred to by the two atom expressions "SER70.CA" and “HOH205:P.O”, respectively, when using the RasMol commands select and restrict.]

16.18.5 Dials Box

If RasMol detects a 'dials box' attached to the user's workstation, it also allows the molecule to be manipulated interactively by the dials. Once RasMol starts up, it labels the LED displays above each dial, “ROTATE X”, “ROTATE Y”, “ROTATE Z” and “ZOOM” across the top row from left to right, and “TRANS X”, “TRANS Y”, “TRANS Z” and “SLAB” from left to right across the bottom row. Rotating any of the knobs will automatically transform and redisplay the molecule interactively. The dials only have effect while the mouse pointer is within the display window. If more than one application is using the dials box at a time, care must be taken to remember the dial labels assigned by each program, as each application may overwrite the dial-label LEDs.

The rotation about the X and Y axes automatically updates the indicators on the appropriate scroll bars. All the rotation dials rotate the molecule 180 degrees for a complete revolution of the dial. All the remaining dials clamp their values to permissible ranges; turning these dials past their limits has no effect. The centre of rotation of the molecule may be changed using the 'centre' command on the command line.

The “ZOOM” dial allows the interactive zooming of the molecule between 10% and 200% of the original default magnification. Rotating the dial clockwise magnifies the molecule and anticlockwise shrinks it. A complete revolution of the dial corresponds to a 100% change in scale.

The “SLAB” dial, which is only effective when slabbing is enabled, allows the user to move the front z-clipping plane from the nearest point on the molecule to the furthest. A complete rotation of the SLAB dial corresponds to moving the clipping plane half the distance between the front and back of the molecule. Turning the SLAB knob clockwise moves the clipping plane closer to the viewer (increasing the number of objects displayed), and turning it anticlockwise moves it further away (preventing more objects from being displayed). Slabbing mode is enabled by typing the command 'slab on' on the command line or toggling the slab option on the options menu.

The translation along the X and Y axis allows the centre of the molecule to be moved within the canvas area of the screen. Rotation and zooming are still performed relative to the centre of rotation and the molecule respectively, which may often not be at the centre of the canvas. The TRANS Z dial currently has no effect.

16.19 RasMol Command Line Interface

RasMol allows the execution of interactive commands typed at the RasMol prompt in the terminal window. Characters typed into either the terminal or the display window are processed on the command line. Each command must be given on a separate line terminated by a new line or carriage return character. Keywords are case insensitive and may be entered in both lower and upper case letters. All whitespace (space, tab and formfeed) characters are ignored, except to separate the keyword and the arguments of a command. Blank lines (those containing only whitespace) are ignored. There is an internal restriction that command lines are limited to a maximum of 256 characters. Strings may be delimited by matching single or double quotation marks. Placing a hash '#' character anywhere outside quotes terminates the line. RasMol will ignore the rest of the line, which may be used to comment on the command.

If a syntax error is detected on entering an interactive command, RasMol indicates the location of the error on the command line by placing the '^' character under the offending word or character, and writing an error message on the following line. If a command is not recognised by RasMol, the program will generate an 'Unrecognised command!' error and redisplay the main prompt. If surplus information is given at the end of a command line, RasMol will execute the recognised command, but issue the warning message 'Warning: Ignoring rest of command!'. Some commands may prompt the user for more information. These commands display a different prompt and are discussed in the command reference.

Whenever RasMol outputs diagnostic or error messages to the screen owing to selecting options from the menu or picking objects on the screen, the current command line is cleared. And the prompt redisplayed after any text has been displayed.

16.19.1 Command Line Editing

RasMol allows basic editing of the command line. Pressing either backspace, delete or ^H will delete the previous character, and the key ^D may be used to delete the character under the cursor. Several characters may be used to move the cursor along the command line. The characters ^B, ^F, ^A and ^E move the cursor back a single character, forward a single character, to the beginning of the line and to the end of the line, respectively. When the cursor is not at the end of the command line, typed characters are inserted into the line and do not overwrite existing characters. After a command line has been edited, a newline or carriage return will enter the entire line, regardless of where the cursor is positioned. Because RasMol is unable to move the cursor up to the previous line, care must be taken when editing commands that wrap over several lines. In the event that another process overwrites or corrupts the command line, the character ^L may be used to redisplay the line on the screen.

RasMol maintains a history of recently used commands, so the that the user never needs to type the same commands repeatedly. Typing ^P on the command line will display the previous command in the history and ^N will display the following command. These commands may be edited using the features described above. Moving forward or backward through the command history undoes the modifications made to the current line. The number of commands retained in the history depends upon their length. RasMol can retain more short command lines and fewer long ones.

16.19.2 Dimensions within RasMol

All dimensions in RasMol, such as radii and distances, may be specified in either RasMol 'units' or Angstroms. The RasMol units were first introduced to allow reasonably sized values to be specified for most of the operations performed in RasMol. A single RasMol unit corresponds to 1/250th of an Angstrom, therefore the most frequently used values are in the hundreds. For this reason, if RasMol is given a distance parameter that doesn't contain a decimal point it is assumed to be in RasMol units. For example, the command “spacefill 300" specifies a sphere radius of 300 RasMol units, or 1.2 Angstroms.

However, dimensions within RasMol can also be specified in Angstroms by placing a decimal point in the number. For example, “spacefill 1.2" specifies a sphere radius in Angstroms. This is particularly useful for the cut-off distance parameter in 'within' expressions.

16.19.3 Start-up Initialisation Files

Each time RasMol is started, it searches for an initialisation file of commands to run before the command prompt is presented to the user. The file is called “.rasmolrc” on UNIX systems, and “RASMOL.INI” on VMS, MS-DOS and Microsoft Windows Systems. The format and execution of the this file is identical to that of the RasMol script command.

RasMol first looks for the initialisation file in the current directory, and if it is not found will look for it in the user's home directory. On all systems the environment variable 'HOME' may be used name the user's home appropriate directory. If no personal initialisation file is found the program looks for the file “rasmolrc” (or “RASMOLRC”) in the RasMol system directory pointed to by the environment variable RASMOLPATH. This directory should also contain the on-line help file “rasmol.hlp”. On UNIX systems RASMOLPATH is typically set to be “/usr/local/lib/rasmol”.

Unlike the command 'script “.rasmolrc”', the program will not generate an error message if the file is not found. The system rasmolrc file is commonly used by system managers to display information about the local installation and who to contact for help contain RasMol echo commands detailing a telephone number or e-mail address to be used for contacting

16.19.4 Inter-process Communication

RasMol supports Inter Process Communication (IPC) in one form or another on all platforms. Under Microsoft Windows, IPC is implemented using Dynamic Data Exchange (DDE), on the Macintosh IPC is implemented using Apple Events and on X Windows systems IPC is implemented using John Ousterhaut's Tcl/Tk communication protocol.

When RasMol starts up on an X window system it registers itself with the X Window Server as a Tcl interpreter. From within a Tcl application such as “wish”, you can use the Tcl command “winfo interps” to determine the currently register interpreters on that display. The first instance of RasMol registers itself as “rasmol”, the second as “rasmol #2”, the third as “rasmol #3” and so on. The Tcl interpreter can easily send a command to rasmol using the built-in 'send' command. RasMol interprets the string parameter to the send command not as a Tcl function to execute but as a RasMol command. Hence, typing “send {rasmol} {background red}” into the wish interpreter will cause RasMol's display window to change colour.

16.20 RasMol Commands

RasMol allows the execution of interactive commands typed at the “RasMol>” prompt in the terminal window. Each command must be given on a separate line. Keywords are case insensitive and may be entered in either upper or lower case letters. All whitespace characters are ignored except to separate keywords and their arguments.

The commands/keywords currently recognised by RasMol are given below.

backbone
background
centre
clipboard
colour
connect
cpk
dots
define
echo
exit
hbonds
help
label
load
print
quit
renumber
reset
restrict
ribbons
rotate
save
script
select
set
show
slab
source
spacefill
ssbonds
strands
structure
trace
translate
wireframe
write
zap
zoom

16.20.1 Backbone

Syntax: backbone {<boolean>}
backbone <value>

The RasMol 'backbone' command permits the representation of a polypeptide backbone as a series of bonds connecting the adjacent alpha carbons of each amino acid in a chain. The display of these backbone 'bonds' is turned on and off by the command parameter the same as the 'wireframe' command. The command 'backbone off' turns off the selected 'bonds', and 'backbone on' or with a number turns them on. The number can be used to determine the cylinder radius of the representation in either Angstrom or RasMol units. A parameter value of 500 (2.0 angstroms) or greater results in an “Parameter value too large” error. Backbone objects may be coloured using the RasMol 'colour backbone' command.

The reserved word 'backbone' is also used as a predefined set and as a parameter to the 'set hbond' and 'set ssbond' commands. The RasMol command 'trace' is synonymous with the command 'backbone'.

16.20.2 Background

Syntax: background <colour>

The RasMol 'background' command is used to set the colour of the “canvas” background. The colour may be given as either a colour name or a comma separated triple of Red, Green and Blue (RGB) components enclosed in square brackets. Typing the command 'help colours' will give a list of the predefined colour names recognised by RasMol. When running under X Windows, RasMol also recognises colours in the X server's colour name database.

The 'background' command is synonymous with the RasMol 'set background' command.

16.20.3 Centre

Syntax: center {<expression>}
centre {<expression>}

The RasMol 'centre' command defines the point about which the 'rotate' command and the scroll bars rotate the current molecule. Without a parameter the centre command resets the centre of rotation to be the centre of gravity of the molecule. If an atom expression is specified, RasMol rotates the molecule about the centre of gravity of the set of atoms specified by the expression. Hence, if a single atom is specified by the expression, that atom will remain 'stationary' during rotations.

Type 'help expression' for more information on RasMol atom expressions.

16.20.4 Clipboard

syntax:clipboard

The RasMol 'clipboard' command places a copy of the currently displayed image on the local graphics 'clipboard'. Note this command is not yet supported on UNIX or VMS systems. It is intended to make transferring of images between applications easier under Microsoft Windows or on an Apple Macintosh.

When using RasMol on a UNIX or VMS system this functionality may be achieved by generating a raster image in a format that can be read by the receiving program using the RasMol 'write' command.

16.20.5 Colour

Syntax: colour {<object>} <colour>
color {<object>} <colour>

Colour the atoms (or other objects) of the selected zone. The colour may be given as either a colour name or a comma separated triple of Red, Green and Blue (RGB) components enclosed in square brackets. Typing the command 'help colours' will give a list of all the predefined colour names recognised by RasMol.

Allowed objects are 'atoms,' 'bonds,' 'backbone,' 'ribbons', 'labels', 'dots,' 'hbonds' and 'ssbonds.' If no object is specified, the default keyword 'atom' is assumed. Some colour schemes are defined for certain object types. The colour scheme 'none' can be applied all objects accept atoms and dots, stating that the selected objects have no colour of their own, but use the colour of their associated atoms (i.e. the atoms they connect). 'Atom' objects can also be coloured by 'amino,' 'cpk,' 'chain,' 'group,' 'shapely,' 'structure,' 'temperature', 'charge' and 'user'. Hydrogen bonds can also be coloured by 'type' and dot surfaces also coloured by electrostatic 'potential'. For more information type 'help colour <colour>.'

16.20.6 Connect

Syntax:connect {<boolean>}

The RasMol 'connect' command is used to force RasMol to (re)calculate the connectivity of the current molecule. If the original input file contained connectivity information, this is discarded. The command 'connect false' uses a fast heuristic algorithm that is suitable for determining bonding in large bio-molecules such as proteins and nucleic acids. The command 'connect true' uses a slower more accurate algorithm based upon covalent radii that is more suitable to small molecules containing inorganic elements or strained rings. If no parameters are given, RasMol determines which algorithm to use based on the number of atoms in the input file. Greater than 255 atoms causes RasMol to use the faster implementation. This is the method used to determine bonding, if necessary, when a molecule is first read in using the 'load' command.

16.20.7 Define

Syntax: define <identifier> <expression>

The RasMol 'define' command allows the user to associate an arbitrary set of atoms with a unique identifier. This allows the definition of user-defined sets. These sets are declared statically, i.e. once defined the contents of the set do not change, even if the expression defining them depends on the current transformation and representation of the molecule.

16.20.8 Dots

Syntax: dots {<boolean>}
dots <value>

The RasMol 'dots' command is used to generate a Van der Waal's dot surface around the currently selected atoms. Dot surfaces display regularly spaced points on a sphere of Van der Waals' radius about each selected atom. Dots that would are 'buried' within the Van der Waal's radius of any other atom (selected or not) are not displayed. The command 'dots on' deletes any existing dot surface and generates a dots surface around the currently selected atom set with a default dot density of 100. The command 'dots off' deletes any existing dot surface. The dot density may be specified by providing a numeric parameter between 1 and 1000. This value approximately corresponds to the number of dots on the surface of a medium sized atom.

By default, the colour of each point on a dot surface is the colour of it's closest atom at the time the surface is generated. The colour of the whole dot surface may be changed using the 'colour dots' command.

16.20.9 Echo

Syntax: echo {<string>}

The RasMol 'echo' command is used to display a message in the RasMol command/terminal window. The string parameter may optionally be delimited in double quote characters. If no parameter is specified, the 'echo' command displays a blank line. This command is particularly useful for displaying text from within a RasMol 'script' file.

16.20.10 HBonds

Syntax: hbonds {<boolean>}
hbonds <value>

The RasMol 'hbond' command is used to represent the hydrogen bonding of the protein molecule's backbone. This information is useful in assessing the protein's secondary structure. Hydrogen bonds are represented as either dotted lines or cylinders between the donor and acceptor residues. The first time the 'hbond' command is used, the program searches the structure of the molecule to find hydrogen bonded residues and reports the number of bonds to the user. The command 'hbonds on' displays the selected 'bonds' as dotted lines, and the 'hbonds off' turns off their display. The colour of hbond objects may be changed by the 'colour hbond' command. Initially, each hydrogen bond has the colours of its connected atoms.

By default the dotted lines are drawn between the accepting oxygen and the donating nitrogen. By using the 'set hbonds' command the alpha carbon positions of the appropriate residues may be used instead. This is especially useful when examining proteins in backbone representation.

16.20.11 Help

Syntax: help {<topic> {<subtopic>}}
? {<topic> {<subtopic>}

The RasMol 'help' command provides on-line help on the given topic.

16.20.12 Label

Syntax:label {<string>}
label <boolean>

The RasMol 'label' command allows an arbitrary formatted text string to be associated with each currently selected atom. This string may contain embedded 'expansion specifiers' which display properties of the atom being labelled. An expansion specifier consists of a '%' character followed by a single alphabetic character specifying the property to be displayed (similar to C's printf syntax). An actual '%' character may be displayed by using the expansion specifier '%%'.

Atom labelling for the currently selected atoms may be turned off with the command 'label off.' By default, if no string is given as a parameter RasMol uses labels appropriate for the current molecule. RasMol uses the label "%n%r:%c.%a" if the molecule contains more than one chain, "%e%i" if the molecule has only a single residue (a small molecule) and "%n%r.%a" otherwise.

The colour of each label may be changed using the 'colour label' command. By default, each label is drawn in the same colour as the atom to which it is attached. The size of the displayed text may be changed using the 'set fontsize' command.

The following table lists the current expansion specifiers:

%a	Atom Name
%b %t	B-factor/Temperature
%c %s	Chain Identifier
%e	Element Atomic Symbol
%i	Atom Serial Number
%n	Residue Name
%r	Residue Number

16.20.13 Load

Syntax: load {<format>} <filename>

Load a molecule co-ordinate file into RasMol2. Valid molecule file formats are 'pdb' (Brookhaven Protein Databank), 'mdl' (Molecular Design Limited's MOL file format), 'alchemy' (Tripos' Alchemy file format), 'mol2' (Tripos' Sybyl Mol2 file format), 'charmm' (CHARMm file format) or 'xyz' (MSC's XMol XYZ file format). If no file format is specified, 'pdb' is assumed by default. Only a single molecule may be loaded at a time. To delete a molecule prior to loading another use the RasMol 'zap' command.

The 'load' command selects all the atoms in the molecule, centres it on the screen and renders it as a CPK coloured wireframe model. If the molecule contains no bonds (i.e. contains only alpha carbons), it is drawn as an alpha carbon backbone. If the file specifies less bonds than atoms, RasMol determines connectivity using the default 'connect' command.

16.20.14 Print

Syntax: print

The RasMol 'print' command sends the currently displayed image to the local default printer using the operating system's native printer driver. Note this command is not yet supported under UNIX or VMS. It is intended to take advantage of Microsoft Windows and Apple Macintosh printer drivers. For example, allowing images to be printed directly on a dot matrix printer.

When using RasMol on a UNIX or VMS system this functionality may be achieved by either generating a PostScript file using the RasMol 'write ps' or 'write vectps' commands and printing that or generating a raster image file and using a utility to dump that to the local printer.

16.20.15 Quit

Syntax: quit
exit

Exit from the RasMol program. The RasMol commands 'exit' and 'quit' are synonymous.

16.20.16 Renumber

Syntax: renumber {{-} <value>}

The RasMol 'renumber' command sequentially numbers the residues in a macromolecular chain. The optional parameter specifies the value of the first residue in the sequence. By default, this value is one. For proteins, each amino acid is numbered consecutively from the N terminus to the C terminus. For nucleic acids, each base is numbered from the 5' terminus to 3' terminus. All chains in the current database are renumbered and gaps in he original sequence are ignored. The starting value for numbering may be negative.

16.20.17 Reset

Syntax: reset

The RasMol 'reset' command restores the original viewing transformation and centre of rotation. The scale is set to it default value, 'zoom 100,' the centre of rotation is set to the geometric centre of the currently loaded molecule, 'centre all,' this centre is translated to the middle of the screen and the viewpoint set to the default orientation.

This command should not be mistaken for the RasMol 'zap' command which deletes the currently stored molecule, returning the program to its initial state.

16.20.18 Restrict

Syntax: restrict {<expression>}

The RasMol 'restrict' command both defines the currently active zone of the molecule and disables the representation of (most of) those parts of the molecule no longer selected. All subsequent RasMol commands that modify a molecule's colour or representation effect only the currently selected zone. The parameter of a 'restrict' command is a RasMol atom expression that is evaluated for every atom of the current molecule. This command is very similar to the RasMol 'select' command, except restrict disables the 'wireframe,' 'spacefill' and 'backbone' representations in the non-active zone.

Type “help expression” for more information on RasMol atom expressions.

16.20.19 Ribbons

Syntax: ribbons {<boolean>}
ribbons <value>

The RasMol 'ribbons' command displays the currently loaded protein or nucleic acid as a smooth solid “ribbon” surface passing along the backbone of the protein. The ribbon is drawn between each amino acid whose alpha carbon is currently selected. The colour of the ribbon is changed by the RasMol 'colour ribbon' command. If the current ribbon colour is 'none' (the default), the colour is taken from the alpha carbon at each position along its length.

The width of the ribbon at each position is determined by the optional parameter in the usual RasMol units. By default the width of the ribbon is taken from the secondary structure of the protein or a constant value of 720 for nucleic acids (which produces a ribbon 2.88 Angstroms wide). Default widths of protein alpha helices and beta sheets is 380 (1.52 Angstroms) and 100 (0.4 Angstroms) for turns and random coil. The secondary structure assignments is either from the PDB file or calculated using the DSSP algorithm as used by the 'structure' command. This command is similar to the RasMol command 'strands' which renders biomolecular ribbon as parallel strands.

16.20.20 Rotate

Syntax: rotate <axis> {-} <value>

Rotate the molecule about the specified axis. Permitted values for the axis parameter are “x”, “y” and “z”. The integer parameter states the angle in degrees for the structure to be rotated. For the X and Y axes, positive values move the closest point up and right, and negative values move it down and left respectively. For the Z axis, a positive rotation acts clockwise and a negative angle anti-clockwise.

16.20.21 Save

Syntax: save {pdb} <filename>
save alchemy <filename>

Save the currently selected set of atoms in either a Brookhaven Protein Database (PDB) or Alchemy(tm) format file. The distinction between this command and the RasMol 'write' command has been dropped. The only difference is that without a format specifier the 'save' command generates a PDB file and the 'write' command generates a GIF image.

16.20.22 Script

Syntax: script <filename>

The RasMol 'script' command reads a set of RasMol commands sequentially from a text file and executes them. This allows sequences of commonly used commands to be stored and performed by a single command. A RasMol script file may contain a further script command up to a maximum “depth” of 10, allowing complicated sequences of actions to be executed. RasMol ignores all characters after the first '#' character on each line allowing the scripts to be annotated. Script files are often also annotated using the RasMol 'echo' command.

The most common way to generate a RasMol script file is to use the 'write script' (or 'write rasmol') command to output the sequence of commands that are needed to regenerate the current view, representation and colouring of the currently displayed molecule.

The RasMol 'source' command is synonymous with the 'script' command.

16.20.23 Select

Syntax: select {<expression>}

Define the currently active zone of the molecule. All subsequent RasMol commands that manipulate a molecule or modify its colour or representation, only effects the currently selected zone. The parameter of a 'select' command is a RasMol expression that is evaluated for every atom of the current molecule. The currently selected (active) zone of the molecule are those atoms that cause the expression to evaluate true. To select the whole molecule use the RasMol command 'select all.' The behaviour of the 'select' command without any arguments is determined by the RasMol 'hetero' and 'hydrogen' internal parameters.

Type “help expression” for more information on RasMol atom expressions.

16.20.24 Set

Syntax: set <parameter> {<option>}

The RasMol 'set' command allows the user to alter various internal program parameters such as those controlling rendering options. Each parameter has its own set or permissible parameter options. Typically, omitting the parameter option resets that parameter to its default value. For more information on each internal parameter type 'help set parameter.' A list of valid parameter names is given below.

ambient **axes** **background** **bondmode**
boundbox **display** **fontsize** **hbonds**
hetero **hourglass** **hydrogen** **kinemage**
menus **mouse** **radius** **shadow**
slabmode **solvent** **specular** **specpower**
ssbonds **strands** **unitcell** **vectps**

16.20.25 Show

Syntax: show information
show sequence
show symmetry

The RasMol 'show' command display details of the status of the currently loaded molecule. The command 'show information' lists the molecule's name, classification, PDB code and the number of atoms, chains, groups it contains. If hydrogen bonding, disulphide bridges or secondary structure have been determined, the number of hbonds, ssbonds, helices, ladders and turns are also displayed respectively. The command 'show sequence' lists the residues that compose each chain of the molecule.

16.20.26 Slab

Syntax: slab {<boolean>}
slab <value>

The RasMol 'slab' command enables, disables or positions the z-clipping plane of the molecule. The program only draws those portions of the molecule that are further from the viewer than the slabbing plane. Integer values range from zero at the very back of the molecule to 100 which is completely in front of the molecule. Intermediate values determine the percentage of the molecule to be drawn.

16.20.27 Spacefill

Syntax: spacefill {<boolean>}
spacefill temperature
spacefill user
spacefill <value>

Represent the currently selected zone as a spacefilling union of spheres model. An integer parameter may be used to specify the radius of each atom given in 4nm units. If no parameter is given, each atom is drawn as a sphere of its van der Waals radius.

The 'temperature' option is used to set the radius of each selected sphere to the value in the temperature field of the molecule file. A zero or negative value causes no change in the selected atom. Temperature values greater than 2.00 are truncated to 2.00 Angstrom radius.

The 'user' option allows the radius of the selected spheres to be determined by matching each atom against optional lines in the input data file. Details of the wildcard pattern matching used by Raster3D's COLOR records is given in the manual.

The RasMol 'spacefill' command is used to represent all of the currently selected atoms as solid spheres. This command is used to produce both union-of-spheres and ball-and-stick models of a molecule. The command, 'spacefilll true,' the default, represents each atom as a sphere of Van der Waals radius. The command 'spacefill off' turns off the representation of the selected atom as spheres. A sphere radius may be specified as an integer in RasMol units (1/250th Angstrom) or a value containing a decimal point. A value of 500 (2.0 Angstroms) or greater results in a “Parameter value too large” error.

The 'temperature' option sets the radius of each sphere to the value stored in its temperature field. Zero or negative values causes have no effect and values greater than 2.0 are truncated to 2. The 'user' option allows the radius of each spheres to be specified by additional lines in the molecule's PDB file using Raster 3D's COLOR record extension.

The RasMol command 'cpk' is synonymous with the 'spacefill' command.

16.20.28 SSBonds

Syntax: ssbonds {<boolean>}
ssbonds <value>

The RasMol 'ssbonds' command is used to represent the disulphide bridges of the protein molecule as either dotted lines or cylinders between the connected systemise. The first time that the 'ssbonds' command is used, the program searches the structure of the protein to find half-celestine pairs (systemise whose sulphurous are within 3 angstroms of each other) and reports the number of bridges to the user. The command 'ssbonds on' displays the selected 'bonds' as dotted lines, and the command 'ssbonds off' disables the display of ssbonds in the currently selected area. Selection of disulphide bridges is identical to normal bonds, and may be adjusted using the RasMol 'set bondmode' command. The colour of disulphide bonds may be changed using the 'colour ssbonds' command. By default, each disulphide bond has the colours of its connected atoms.

By default disulphide bonds are drawn between the sulphur atoms within the celestine groups. By using the 'set ssbonds' command the position of the systemise alpha carbons may be used instead.

16.20.29 Strands

Syntax: strands {<boolean>}
strands <value>

The RasMol 'strands' command displays the currently loaded protein or nucleic acid as a smooth “ribbon” of depth-cued curves passing along the backbone of the protein. The ribbon is composed of a number of strands that run parallel to one another along the peptide plane or each residue. The ribbon is drawn between each amino acid whose alpha carbon is currently selected. The colour of the ribbon is changed by the RasMol 'colour ribbon' command. If the current ribbon colour is 'none' (the default), the colour is taken from the alpha carbon at each position along its length. The colour of the central and outermost strands may be coloured independently using the 'colour ribbon1' and 'colour ribbon2' commands respectively. The number of strands in the ribbon may be altered using the RasMol 'set strands' command.

The width of the ribbon at each position is determined by the optional parameter in the usual RasMol units. By default the width of the ribbon is taken from the secondary structure of the protein or a constant value of 720 for nucleic acids (which produces a ribbon 2.88 Angstroms wide). The default width of protein alpha helices and beta sheets is 380 (1.52 Angstroms) and 100 (0.4 Angstroms) for turns and random coil. The secondary structure assignments is either from the PDB file or calculated using the DSSP algorithm as used by the 'structure' command. This command is similar to the RasMol command 'ribbons' which renders the biomolecular ribbon as parallel strands.

16.20.30 Structure

Syntax: structure

The RasMol 'structure' command calculates secondary structure assignments for the currently loaded protein. If the original PDB file contained structural assignment records (HELIX and SHEET) these are discarded. Initially, the hydrogen bonds of the current molecule are found, if this hasn't been done already. The secondary structure is the determined using Kabsch and Sander's DSSP algorithm. Once finished the program reports the number of helices, strands and turns found.

16.20.31 Translate

Syntax: translate <axis> {-} <value>

The RasMol 'translate' command moves the position of the centre of the molecule on the screen. The axis parameter specifies along which axis the molecule is to be moved and the integer parameter specifies the absolute position of the molecule centre from the middle of the screen. Permitted values for the axis parameter are “x”, “y” and “z”. Displacement values must be between -100 and 100 which correspond to moving the current molecule just off the screen. A positive “x” displacement moves the molecule to the right, and a positive “y” displacement moves the molecule down the screen. The pair of commands 'translate x 0' and 'translate y 0' centres the molecule on the screen.

16.20.32 Wireframe

Syntax: wireframe {<boolean>}
wireframe <value>

The RasMol 'wireframe' command represents each bond within the selected region of the molecule as either a cylinder, a line or depth-cued vector. The display of bonds as depth-cued vectors (drawn darker the further away from the viewer) is turned on by the command 'wireframe' or 'wireframe on.' The selected bonds are displayed as cylinders by specifying a radius either as an integer in RasMol units or containing a decimal point as a value in Angstroms. A parameter value of 500 (2.0 angstroms) or above results in an “Parameter value too large” error. Bonds may be coloured using the 'colour bonds' command.

16.20.33 Write

Syntax: write {<format>} <filename>

Write the current image to a file in a standard raster format. Currently supported image file formats include “gif” (Compuserve GIF), “ppm” (Portable Pixmap), “ras” (Sun rasterfile), “ps” and “epsf” (Encapsulated PostScript), “monops” (Monochrome Encapsulated PostScript), “bmp” (Microsoft bitmap) and “pict” (Apple 'PICT'). The 'write' command may also be used to generate command scripts for other graphics programs. The format 'script' writes out a file containing the RasMol 'script' commands to reproduce the current image. The format 'molscript' writes out the commands required to render the current view of the molecule as ribbons in Per Kraulis' Molscript program and the format 'kinemage' the commands for David Richardson's program Mage.

The distinction between this command and the RasMol 'save' command has been dropped. The only difference is that without a format specifier the 'save' command generates a 'PDB' file and the 'write' command generates a 'GIF' image.

16.20.34 Zap

Syntax: zap

Deletes the contents of the current database and resets parameter variables to their initial default state.

16.20.35 Zoom

Syntax: zoom {<boolean>}
zoom <value>

Change the magnification of the currently displayed image. Boolean parameters either magnify or reset the scale of current molecule. An integer parameter between 10 and 200 specifies the desired magnification as a percentage of the default scale.

16.21 RasMol Internal Parameters

RasMol has a number of internal parameters that may be modified using the 'set' command. These parameters control a number of program options such as rendering options and mouse button mappings.

A complete list of internal parameter names is given below:

ambient
axes
background
bondmode
boundbox
display
fontsize
hbonds
hetero
hourglass
hydrogen
kinemage
menus
mouse
radius
shadow
slabmode
solvent
specular
specpower
ssbonds
strands
unitcell
vectps

16.21.1 Set Ambient

Syntax: set ambient {<value>}

The RasMol 'ambient' parameter is used to control the amount of ambient (or surrounding) light in the scene. The 'ambient' value must be between 0 and 100 that controls the percentage intensity of the darkest shade of an object. For a solid object, this is the intensity of surfaces facing away from the light source or in shadow. For depth-cued objects this is the intensity of objects furthest from the viewer.

This parameter is commonly used to correct for monitors with different “gamma values” (brightness), to change how light or dark a hardcopy image appears when printed or to alter the feeling of depth for wireframe or ribbon representations.

16.21.2 Set Axes

Syntax: set axes <boolean>

The RasMol 'axes' parameter controls the display of orthogonal co-ordinate axes on the current display. The co-ordinate axes are those used in the molecule data file, and the origin is the centre of the molecule's bounding box. The 'set axes' command is similar to the commands 'set boundbox' and 'set unitcell' that display the bounding box and the crystallographic unit cell respectively.

16.21.3 Set Background

Syntax: set background <colour>

The RasMol 'background' parameter is used to set the colour of the “canvas” background. The colour may be given as either a colour name or a comma separated triple of Red, Green, Blue (RGB) components enclosed in square brackets. Typing the command 'help colours' will give a list of the predefined colour names recognised by RasMol. When running under X Windows, RasMol also recognises colours in the X server's colour name database.

The command 'set background' is synonymous with the RasMol command 'background'.

16.21.4 Set BondMode

Syntax: set bondmode and
set bondmode or

The RasMol 'set bondmode' command controls the mechanism used to select individual bonds. When using the 'select' and 'restrict' commands, a given bond will be selected if i) the bondmode is 'or' and either of the connected atoms is selected, or ii) the bondmode is 'and' and both atoms connected by the bond are selected. Hence an individual bond may be uniquely identified by using the command “set bondmode and” and then uniquely selecting the atoms at both ends.

16.21.5 Set BoundBox

Syntax: set boundbox <boolean>

The RasMol 'boundbox' parameter controls the display of the current molecules bounding box on the display. The bounding box is orthogonal to the data file's original co-ordinate axes. The 'set boundbox' command is similar to the commands 'set axes' and 'set unitcell' that display orthogonal co-ordinate axes and the bounding box respectively.

16.21.6 Set Display

Syntax: set display selected
set display normal

This command controls the display mode within RasMol. By default, 'set display normal', RasMol displays the molecule in the representation specified by the user. The command 'set display selected' changes the display mode such that the molecule is temporarily drawn so as to indicate the currently selected portion of the molecule. The user specified colour scheme and representation remains unchanged. In this representation all selected atoms are shown in yellow and all non selected atoms are shown in blue. The colour of the background is also changed to a dark grey to indicate the change of display mode. This command is typically only used by external Graphical User Interfaces (GUIs).

16.21.7 Set HBonds

Syntax: set hbonds backbone
set hbonds sidechain

The RasMol 'hbonds' parameter determines whether hydrogen bonds are drawn between the donor and acceptor atoms of the hydrogen bond, 'set hbonds sidechain' or between the alpha carbon atoms of the protein backbone and between the phosphorous atoms of the nucleic acid backbone, 'set hbonds backbone'. The actual display of hydrogen bonds is controlled by the 'hbonds' command. Drawing hydrogen bonds between protein alpha carbons or nucleic acid phosphorous atoms is useful when the rest of the molecule is shown in only a schematic representation such as 'backbone', 'ribbons' or 'strands'. This parameter is similar the RasMol 'ssbonds' parameter.

16.21.8 Set FontSize

Syntax: set fontsize {<boolean>}

The RasMol 'set fontsize' command is used to control the size of the characters that form atom labels. This value corresponds to the height of the displayed character in pixels. The maximum value of 'fontsize' is 32 pixels, and the default value is 8 pixels high. To display atom labels on the screen use the RasMol 'label' command and to change the colour of displayed labels, use the 'colour labels' command.

16.21.9 Set Hetero

Syntax: set hetero <boolean>

The RasMol 'hetero' parameter is used to modify the 'default' behaviour of the RasMol 'select' command, i.e. the behaviour of 'select' without any parameters. When this value is 'false,' the default 'select' region does not include an heterogenous atoms (refer to the predefined set 'hetero'). When this value is 'true,' the default 'select' region may contain hetero atoms. This parameter is similar to the RasMol 'hydrogen' parameter which determines whether hydrogen atoms should be included in the default set. If both 'hetero' and 'hydrogen' are 'true,' 'select' without any parameters is equivalent to 'select all.'

16.21.10 Set HourGlass

Syntax: set hourglass <boolean>

The RasMol 'hourglass' parameter allows the user to enable and disable the use of the 'hour glass' cursor used by RasMol to indicate that the program is currently busy drawing the next frame. The command 'set hourglass on' enable the indicator, whilst 'set hourglass off' prevents RasMol from changing the cursor. This is useful when spinning the molecule, running a sequence of commands from a script file or using interprocess communication to execute complex sequences of commands. In these cases a 'flashing' cursor may be distracting.

16.21.11 Set Hydrogen

Syntax: set hydrogen <boolean>

The RasMol 'hydrogen' parameter is used to modify the 'default' behaviour of the RasMol 'select' command, i.e. the behaviour of 'select' without any parameters. When this value is 'false,' the default 'select' region does not include any hydrogen or deuterium atoms (refer to the predefined set 'hydrogen'). When this value is 'true,' the default 'select' region may contain hydrogen atoms. This parameter is similar to the RasMol 'hetero' parameter which determines whether heterogenous atoms should be included in the default set. If both 'hydrogen' and 'hetero' are 'true,' 'select' without any parameters is equivalent to 'select all.'

16.21.12 Set Kinemage

Syntax:set kinmage <boolean>

The RasMol 'set kinemage' command controls the amount of detail stored in a Kinemage output file generated by the RasMol 'write kinemage' command. The output kinemage files are intended to be displayed by David Richardson's Mage program. 'set kinemage false,' the default, only stores the currently displayed representation in the generated output file. The command 'set kinemage true,' generates a more complex Kinemage that contains both the wireframe and backbone representations as well as the co-ordinate axes, bounding box and crystal unit cell.

16.21.13 Set Menus

Syntax: set menus <boolean>

The RasMol 'set menus' command enables the canvas window's menu buttons or menu bar. This command is typically only used by graphical user interfaces or to create as large as image as possible when using Microsoft Windows.

16.21.14 Set Mouse

Syntax: set mouse rasmol
set mouse insight
set mouse quanta

The RasMol 'set mouse' command sets the rotation, translation, scaling and zooming mouse bindings. The default value is 'rasmol' which is suitable for two button mice (for three button mice the second and third buttons are synonymous); X-Y rotation is controlled by the first button, and X-Y translation by the second. Additional functions are controlled by holding a modifier key on the keyboard. [Shift] and the first button performs scaling, [shift] and the second button performs Z-rotation, and [control] and the first mouse button controls the clipping plane. The 'insight' and 'quanta' provide the same mouse bindings as other packages for experienced users.

16.21.15 Set Radius

Syntax: set radius {<value>}

The RasMol 'set radius' command is used to alter the behaviour of the RasMol 'dots' command depending upon the value of the 'solvent' parameter. When 'solvent' is 'true,' the 'radius' parameter controls whether a true Van der Waal's surface is generated by the 'dots' command. If the value of 'radius' is anything other than zero, that value is used as the radius of each atom instead of it true VdW value. When the value of 'solvent' is 'true,' this parameter determines the 'probe sphere' (solvent) radius. The parameter may be given as an integer in rasmol units or containing a decimal point in Angstroms. The default value of this parameter is determined by the value of 'solvent' and changing 'solvent' resets 'radius' to its new default value.

16.21.16 Set Shadow

Syntax: set shadow <boolean>

The RasMol 'set shadow' command enables and disables raytracing of the currently rendered image. Currently only the spacefilling representation is shadowed or can cast shadows. Enabling shadowing will automatically disable the Z-clipping (slabbing) plane using the command 'slab off.' Raytracing typically takes about 10s for a moderately sized protein. It is recommended that shadowing is normally disabled whilst the molecule is being transformed or manipulated, and only enabled once an appropriate viewpoint is selected, to provide a greater impression of depth.

16.21.17 Set SlabMode

Syntax: set slabmode <slabmode>

The RasMol 'slabmode' parameter controls the rendering method of objects cut by the slabbing (z-clipping) plane. Valid slabmode parameters are “reject”, “half”, “hollow”, “solid” and “section”.

16.21.18 Set Solvent

Syntax: set solvent <boolean>

The RasMol 'set solvent' command is used to control the behaviour of the RasMol 'dots' command. Depending upon the value of the 'solvent' parameter, the 'dots' command either generates a Van der Waal's or a solvent accessible surface around the currently selected set of atoms. Changing this parameter automatically resets the value of the RasMol 'radius' parameter. The command 'set solvent false,' the default value, indicates that a Van der Waal's surface should be generated and resets the value of 'radius' to zero. The command 'set solvent true' indicates that a 'Connolly' or 'Richards' solvent accessible surface should be drawn and sets the 'radius' parameter, the solvent radius, to 1.2 Angstroms (or 300 RasMol units).

16.21.19 Set Specular

Syntax: set specular <boolean>

The RasMol 'set specular' command enables and disables the display of specular highlights on solid objects drawn by RasMol. Specular highlights appear as white reflections of the light source on the surface of the object. The current RasMol implementation uses an approximation function to generate this highlight.

The specular highlights on the surfaces of solid objects may be altered by using the specular reflection coefficient, which is altered using the RasMol 'set specpower' command.

16.21.20 Set SpecPower

Syntax: set specpower {<value>}

The 'specpower' parameter determines the shininess of solid objects rendered by RasMol. This value between 0 and 100 adjusts the reflection coefficient used in specular highlight calculations. The specular highlights are enabled and disabled by the RasMol 'set specular' command. Values around 20 or 30 produce plastic looking surfaces. High values represent more shiny surfaces such as metals, while lower values produce more diffuse/dull surfaces.

16.21.21 Set SSBonds

Syntax: set ssbonds backbone
set ssbonds sidechain

The RasMol 'hbonds' parameter determines whether disulphide bridges are drawn between the sulphur atoms in the sidechain (the default) or between the alpha carbon atoms in the backbone of the cysteine residues. The actual display of disulphide bridges is controlled by the 'ssbonds' command. Drawing disulphide bridges between alpha carbons is useful when the rest of the protein is shown in only a schematic representation such as 'backbone', 'ribbons' or 'strands'. This parameter is similar the RasMol 'hbonds' parameter.

16.21.22 Set Strands

Syntax: set strands {<value>}

The RasMol 'strands' parameter controls the number of parallel strands that are displayed in the ribbon representations of proteins. The permissible values for this parameter are 1, 2, 3, 4, 5 and 9. The default value is 5. The number of strands is constant for all ribbons being displayed. However, the ribbon width (the separation between strands) may be controlled on a residue by residue basis using the RasMol 'ribbons' command.

16.21.23 Set UnitCell

Syntax: set unitcell <boolean>

The RasMol 'unitcell' parameter controls the display of the crystallographic unit cell on the current display. The crystal cell is only enabled if the appropriate crystal symmetry information is contained in the PDB data file. The RasMol command 'show symmetry' display details of the crystal's space group and unit cell axes. The 'set unitcell' command is similar to the commands 'set axes' and 'set boundbox' that display orthogonal co-ordinate axes and the bounding box respectively.

16.21.24 Set VectPS

Syntax: set vectps <boolean>

The RasMol 'vectps' parameter is use to control the way in which the RasMol 'write' command generates vector PostScript output files. The command 'set vectps on' enables to use of black outlines around spheres and cylinder bonds producing 'cartoon-like' high resolution output. However, the current implementation of RasMol incorrectly cartoons spheres that are intersected by more than one other sphere. Hence 'ball and stick' models are rendered correctly by not large spacefilling spheres models. Cartoon outlines can be disabled, the default, by the command 'set vectps off'

16.22 RasMol Atom Expressions

RasMol atom expressions uniquely identify an arbitrary group of atoms within a molecule. Atom expressions are composed of either primitive expressions, predefined sets, comparison operators, 'within' expressions, or logical (boolean) combinations of the above expression types.

The logical operators allow complex queries to be constructed out of simpler ones using the standard boolean connectives 'and, or' and 'not.' These may be abbreviated by the symbols “&”, “|” and “!” respectively. Parentheses (brackets) may be used to alter the precedence of the operators. For convenience, a comma may also be used for boolean disjunction.

The atom expression is evaluated for each atom, hence 'protein and backbone' selects protein backbone atoms, not the protein and [nucleic] acid backbone atoms!

16.22.1 Example Expressions

The following table gives some useful examples of RasMol atom expressions.

Examples:	**backbone**
	within(8.0, ser70)
	not (hydrogen or hetero)
	not *.FE and hetero
	8, 12, 16, 20-28
	arg, his, lys

16.22.2 Primitive Expressions

RasMol primitive expressions are the fundamental building blocks of atom expressions. There are two types of primitive expression. The first type is used to identify a given residue number or range of residue numbers. A single residue is identified by its number (position in the sequence), and a range is specified by lower and upper bounds separated by a hyphen character. For example 'select 5,6,7,8' is also 'select 5-8.' Note that this selects the given residue numbers in all macromolecule chains.

The second type of primitive expression specifies a sequence of fields that must match for a given atom. The first part specifies a residue (or group of residues) and an optional second part specifies the atoms within those residues. The first part consists of a residue name, optionally followed by a residue number and/or chain identifier.

A residue name typically consists of up to three alphabetic characters, which are case insensitive. Hence the primitive expressions 'SER' and 'ser' are equivalent, identifying all cysteine residues. Residue names that contain non-alphabetic characters, such as sulphate groups, may be delimited using square brackets, i.e. '[SO4]'

The residue number is the residue's position in the macromolecule sequence. Negative sequence numbers are permitted. For example, 'SER70' Care must be taken when specifying both residue name and number, it the group at the specified position isn't the specified residue no atoms are selected.

The chain identifier is typically a single case-insensitive alphabetic or numeric character. Numeric chain identifiers must be distinguished or separated from residue numbers by a colon character. For example, 'SER70A' or 'SER70:1'

The second part consists of a period character followed by an atom name. An atom name may be up to four alphabetic or numeric characters.

An asterisk may be used as a wild card for a whole field and a question mark as a single character wildcard.

16.22.3 Comparison Operators

Parts of a molecule may also be distinguished using equality, inequality and ordering operators on their properties. The format of such comparison expression is a property name, followed by a comparison operator and then an integer value.

The atom properties that may be used in RasMol are 'atomno' for the atom serial number, 'elemno' for the atom's atomic number (element type), 'resno' for the residue number, 'radius' for the spacefill radius in RasMol units (or zero if not represented as a sphere) and 'temperature' for the PDB anisotropic temperature value.

The equality operator is denoted either "=" or "==". The inequality operator as either “<>”, "!=" or “/=”. The ordering operators are “<“ for less than, “<=” for less than or equal to, “>” for greater than, and “>” for greater than or equal to.

16.22.4 Within Expressions

A RasMol 'within' expression allows atoms to be selected on their proximity to another set of atoms. A 'within' expression takes two parameters separated by a comma and surrounded by parenthesis. The first argument is an integer value called the “cut-off” distance of the within expression and the second argument is any valid atom expression. The cut-off distance is expressed in either integer RasMol units or Angstroms containing a decimal point. An atom is selected if it is within the cut-off distance of any of the atoms defined by the second argument. This allows complex expressions to be constructed containing nested 'within' expressions.

For example, the command 'select within(3.2,backbone)' selects any atom within a 3.2

Angstrom radius of any atom in a protein or nucleic acid backbone. 'Within' expressions are particularly useful for selecting the atoms around an active site.

16.23 RasMol Predefined Sets

RasMol atom expressions may contain predefined sets. These sets are single keywords that represent portions of a molecule of interest. Predefined sets are often abbreviations primitive atom expressions, and in some cases of selecting areas of a molecule that could not otherwise be distinguished. A list of the currently predefined sets is given below. In addition to the sets listed here, RasMol also treats element names (and their plurals) as predefined sets containing all atoms of that element type, i.e. the command “select oxygen” is equivalent to the command “select atomno=8”.

16.23.1 AT Set

This set contains the atoms in the complementary nucleotides adenosine and thymidine (A and T respectively). All nucleotides are classified as either the set 'at' or the set 'cg' This set is equivalent to the RasMol atom expressions “a,t” and “nucleic and not cg”

16.23.2 Acidic Set

The set of acidic amino acids. These are the residue types Asp and Glu. All amino acids are classified as either 'acidic,' 'basic' 'or' 'neutral.' This set is equivalent to the RasMol atom expressions "asp, glu" and “amino and not (basic or neutral)”

16.23.3 Acyclic Set

The set of atoms in amino acids not containing a cycle or ring. All amino acids are classified as either 'cyclic' or 'acyclic.' This set is equivalent to the RasMol atom expression “amino and not cyclic”

16.23.4 Aliphatic Set

This set contains the aliphatic amino acids. These are the amino acids Ala, Gly, Ile, Leu and Val. This set is equivalent to the RasMol atom expression “ala, gly, ile, leu, val”

16.23.5 Alpha Set

The set of alpha carbons in the protein molecule. This set is approximately equivalent to the RasMol atom expression “*.CA” This command should not be confused with the predefined set 'helix' which contains the atoms in the amino acids of the protein's alpha helices.

16.23.6 Amino Set

This set contains all the atoms contained in amino acid residues. This is useful for distinguishing the protein from the nucleic acid and heterogenous atoms in the current molecule database.

16.23.7 Aromatic Set

The set of atoms in amino acids containing aromatic rings. These are the amino acids His, Phe, Trp and Tyr. Because they contain aromatic rings all members of this set are member of the predefined set 'cyclic.' This set is equivalent to the RasMol atom expressions "his, phe, trp, tyr" and “cyclic and not pro.”

16.23.8 Backbone Set

This set contains the four atoms of each amino acid that form the polypeptide N-C-C-O backbone of proteins, and the atoms the sugar phosphate backbone of nucleic acids. Use the RasMol predefined sets 'protein' and 'nucleic' to distinguish between the two forms of backbone. Atoms in nucleic acids and proteins are either 'backbone' or 'sidechain.' This set is equivalent to the RasMol expression "(protein or nucleic) and not sidechain"

The predefined set 'mainchain' is synonymous with the set 'backbone'.

16.23.9 Basic Set

The set of basic amino acids. These are the residue types Arg, His and Lys. All amino acids are classified as either 'acidic,' 'basic' or 'neutral.' This set is equivalent to the RasMol atom expressions "arg, his, lys" and "amino and not (acidic or neutral)"

16.23.10 Bonded Set

This set contain all the atoms in the current molecule database that are bonded to at least one other atom.

16.23.11 Buried Set

This set contains the atoms in those amino acids that tend (prefer) to buried inside protein, away from contact with solvent molecules. This set refers to the amino acids preference and not the actual solvent accessibility for the current protein. All amino acids are classified as either 'surface' or 'buried.' This set is equivalent to the RasMol atom expression “amino and not surface”

16.23.12 CG Set

This set contains the atoms in the complementary nucleotides cytidine and guanoine (C and G respectively). All nucleotides are classified as either the set 'at' or the set 'cg' This set is equivalent to the RasMol atom expressions “c,g” and “nucleic and not at”

16.23.13 Charged Set

This set contains the charged amino acids. These are the amino acids that are either 'acidic' or 'basic.' Amino acids are classified as being either 'charged' or 'neutral.' This set is equivalent to the RasMol atom expressions “acidic or basic” and “amino and not neutral”

16.23.14 Cyclic Set

The set of atoms in amino acids containing a cycle or rings. All amino acids are classified as either 'cyclic' or 'acyclic.' This set consists of the amino acids His, Phe, Pro, Trp and Tyr. The members of the predefined set 'aromatic' are members of this set. The only cyclic but non-aromatic amino acid is proline. This set is equivalent to the RasMol atom expressions “his, phe, pro, trp, tyr” and “aromatic or pro” and “amino and not acyclic”

16.23.15 Cystine Set

This set contains the atoms of cysteine residues that form part of a disulphide bridge, i.e. half cystines. RasMol automatically determines disulphide bridges, if neither the predefined set 'cystine' nor the RasMol 'ssbonds' command have been used since the molecule was loaded. The set of free cysteines may be determined using the RasMol atom expression "cys and not cystine”

16.23.16 Helix Set

This set contains all atoms that form part of a protein alpha helix as determined by either the PDB file author or Kabsch and Sander's DSSP algorithm. By default, RasMol uses the secondary structure determination given in the PDB file if it exists. Otherwise, it uses the DSSP algorithm as used by the RasMol 'structure' command.

This predefined set should not be confused with the predefined set 'alpha' which contains the alpha carbon atoms of a protein.

16.23.17 Hetero Set

This set contains all the heterogenous atoms in the molecule. These are the atoms described by HETATM entries in the PDB file. These typically contain water, cofactors and other solvents and ligands. All 'hetero' atoms are classified as either 'ligand' or 'solvent' atoms. These heterogenous 'solvent' atoms are further classified as either 'water' or 'ions.'

16.23.18 Hydrogen Set

This predefined set contains all the hydrogen and deuterium atoms of the current molecule. This predefined set is equivalent to the RasMol atom expression “elemno=1”.

16.23.19 Hydrophobic Set

This set contains all the hydrophobic amino acids. These are the amino acids Ala, Leu, Val, Ile, Pro, Phe, Met and Trp. All amino acids are classified as either 'hydrophobic' or 'polar.' This set is equivalent to the RasMol atom expressions "ala, leu, val, ile, pro, phe, met, trp" and “amino and not polar”

16.23.20 Ions Set

This set contains all the heterogenous phosphate and sulphate ions in the current molecule data file. A large number of these ions are sometimes associated with protein and nucleic acid structures determined by X-ray crystallography. These atoms tend to clutter an image. All 'hetero' atoms are classified as either 'ligand' or 'solvent' atoms. All 'solvent' atoms are classified as either 'water' or 'ions.'

16.23.21 Large Set

All amino acids are classified as either 'small,' 'medium' or 'large.' This set is equivalent to the RasMol atom expression “amino and not (small or medium)”

16.23.22 Ligand Set

This set contains all the heterogenous cofactor and ligand moieties that are contained in the current molecule data file. At this set is defined to be all 'hetero' atoms that are not 'solvent' atoms. Hence this set is equivalent to the RasMol atom expression “hetero and not solvent”

16.24 RasMol Medium Set

All amino acids are classified as either 'small,' 'medium' or 'large.' This set is equivalent to the RasMol atom expression “amino and not (large or small)”

16.24.1 Neutral Set

The set of neutral amino acids. All amino acids are classified as either 'acidic,' 'basic' or 'neutral.' This set is equivalent to the RasMol atom expression “amino and not (acidic or basic)”

16.24.2 Nucleic Set

The set of all atoms in nucleic acids, which consists of the four nucleotide bases adenosine, cytidine, guanosine and thymidine (A, C, G and T respectively). All nucleotides are classified as either 'purine' or 'pyrimidine'. This set is equivalent to the RasMol atom expressions “a,c,g,t” and “purine or pyrimidine”.

16.24.3 Polar Set

This set contains the polar amino acids. All amino acids are classified as either 'hydrophobic' or 'polar.' This set is equivalent to the RasMol atom expression “amino and not hydrophobic”

16.24.4 Protein Set

The set of all atoms in proteins. This consists of the RasMol predefined set 'amino' and common post-translation modifications.

16.24.5 Purine Set

The set of purine nucleotides. These are the bases adenosine and guanosine (A and G respectively). All nucleotides are either 'purines' or 'pyrimidines.' This set is equivalent to the RasMol atom expressions “a,g” and “nucleic and not purine"

16.24.6 Pyrimidine Set

The set of pyrimidine nucleotides. These are the bases cytidine and thymidine (C and T respectively). All nucleotides are either 'purines' or 'pyrimidines.' This set is equivalent to the RasMol atom expressions "c,t" and "nucleic and not pyrimidine"

16.24.7 Selected Set

This set contains the set of atoms in the currently active zone. The currently active zone is defined by the preceding 'select' or 'restrict' command and not the atom expression containing the 'selected' keyword.

16.24.8 Sheet Set

This set contains all atoms that form part of a protein beta sheet as determined by either the PDB file author or Kabsch and Sander's DSSP algorithm. By default, RasMol uses the secondary structure determination given in the PDB file if it exists. Otherwise, it uses the DSSP algorithm as used by the RasMol 'structure' command.

16.24.9 Sidechain Set

This set contains the functional sidechains of any amino acids and the base of each nucleotide. These are the atoms not part of the polypeptide N-C-C-O backbone of proteins or the sugar phosphate backbone of nucleic acids. Use the RasMol predefined sets 'protein' and 'nucleic' to distinguish between the two forms of sidechain. Atoms in nucleic acids and proteins are either 'backbone' or 'sidechain.' This set is equivalent to the RasMol expression "(protein or nucleic) and not backbone"

16.24.10 Small Set

All amino acids are classified as either 'small,' 'medium' or 'large.' This set is equivalent to the RasMol atom expression "amino and not (medium or large)"

16.24.11 Solvent Set

This set contains the solvent atoms in the molecule co-ordinate file. These are the heterogenous water molecules, phosphate and sulphate ions. All 'hetero' atoms are classified as either 'ligand' or 'solvent' atoms. All 'solvent' atoms are classified as either 'water' or 'ions.' This set is equivalent to the RasMol atom expressions "hetero and not ligand" and "water or ions"

16.24.12 Surface Set

This set contains the atoms in those amino acids that tend (prefer) to be on the surface of proteins, in contact with solvent molecules. This set refers to the amino acids preference and not the actual solvent accessibility for the current protein. All amino acids are classified as either 'surface' or 'buried.' This set is equivalent to the RasMol atom expression "amino and not buried"

16.24.13 Turn Set

This set contains all atoms that form part of a protein turns as determined by either the PDB file author or Kabsch and Sander's DSSP algorithm. By default, RasMol uses the secondary structure determination given in the PDB file if it exists. Otherwise, it uses the DSSP algorithm as used by the RasMol 'structure' command.

16.24.14 Water Set

This set contains all the heterogenous water molecules in the current database. A large number of water molecules are sometimes associated with protein and nucleic acid structures determined by X-ray crystallography. These atoms tend to clutter an image. All 'hetero' atoms are classified as either 'ligand' or 'solvent' atoms. The 'solvent' atoms are further classified as either 'water' or 'ions.'

16.25 RasMol Classification Summary

The table below summarises RasMol's classification of the common amino acids.

Predefined set	ALA	ARG	ASN	ASP	CYS	GLU	GLN	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
	A	**R**	**N**	**D**	**C**	**E**	**Q**	**G**	**H**	**I**	**L**	**K**	**M**	**F**	**P**	**S**	**T**	**W**	**Y**	**V**
acidic				●		●														
acyclic	●	●	●	●	●	●	●	●		●	●	●	●			●	●			●
aliphatic	●							●		●	●									●
aromatic									●					●				●	●	
basic		●							●			●								
buried	●				●					●	●		●	●				●		●
charged		●		●		●			●			●								
cyclic									●					●	●			●	●	
hydrophobic	●							●		●	●		●	●	●			●	●	●
large		●				●	●		●	●	●	●	●	●				●	●	
medium			●	●	●										●		●			●
negative				●		●														
neutral	●		●		●		●	●	●	●	●		●	●	●	●	●	●	●	●
polar		●	●	●	●	●	●		●			●				●	●			
positive		●							●			●								
small	●							●								●				
surface		●	●	●		●	●	●	●			●			●	●	●		●	

16.26 RasMol Colour Schemes

The RasMol 'colour' command allows different objects (such as atoms, bonds and ribbon segments) to be given a specified colour. Typically this colour is either a RasMol predefined colour name or an RGB triple. Additionally RasMol also supports 'amino,' 'chain,' 'group,' 'shapely,' 'structure,' 'temperature,' 'charge,' 'user', 'hbond type' and 'dot potential' colour schemes. The currently predefined colour names are given in the table below.

blue	[0,0,256]	black	[0,0,0]
cyan	[0,255,255]	green	[0,255,0]
greenblue	[46,139,87]	magenta	[255,0,255]
orange	[255,165,0]	purple	[160,32,24]
red	[255,0,0]	redorange	[255,69,0]

violet	[238,130,23]	white	[255,255,2]
yellow	[255,255,0]		

16.26.1 Amino Colours

The RasMol 'amino' colour scheme colours amino acids according to traditional amino acid properties. The purpose of colouring is to identify amino acids in an unusual or surprising environment. The outer parts of a protein that are polar are visible (bright) colours and non-polar residues darker. Most colours are hallowed by tradition. This colour scheme is similar to the 'shapely' scheme.

Amino Acid	Colour	Triple
ASP, GLU	bright red	[230,10,10]
LYS, ARG	blue	[20,90,255]
CYS, MET	yellow	[230,230,0]
SER, THR	orange	[250,150,0]
PHE, TYR	mid blue	[50,50,170]
ASN, GLN	cyan	[230,230,0]
GLY	light grey	[235,235,235]
LEU, VAL, ILE	green	[15,130,15]
ALA	dark grey	[200,200,200]
TRP	pink	[180,90,180]
HIS	pale blue	[130,130,210]
PRO	flesh	[220,150,130]

16.26.2 Chain Colours

The RasMol 'chain' colour scheme assigns each macromolecular chain a unique colour. This colour scheme is particularly useful for distinguishing the parts of multimeric structure or the individual 'strands' of a DNA chain.

16.26.3 CPK Colours

The RasMol 'cpk' colour scheme is based upon the colours of the popular plastic spacefilling models which were developed by Corey, Pauling and later improved by Kultun. This colour scheme colour 'atom' objects by the atom (element) type. This is the scheme conventionally used by chemists.

Element	Colour	Triple
Carbon	light grey	[200,200,200]
Oxygen	red	[240,0,0]
Hydrogen	white	[255,255,255]
Nitrogen	light blue	[143,143,255]
Sulphur	yellow	[255,200,50]
Phosphorous	orange	[255,165,0]
Chlorine	green	[0,255,0]
Bromine, Zinc	brown	[165,42,42]
Sodium	blue	[0,0,255]
Iron	purple	[160,32,240]
Calcium, Metals	dark grey	[128,128,144]
unknown	deep pink	[255,20,147]

16.26.4 Group Colours

The RasMol 'group' colour scheme colour codes residues by their position in a macromolecular chain. Each chain is drawn as a smooth spectrum from blue through green, yellow and orange to red. Hence the N terminus of proteins and 5' terminus of nucleic acids are coloured red and the C terminus of proteins and 3' terminus of nucleic acids are drawn in blue. If a chain has a large number of heterogenous molecules associated with it, the macromolecule may not be drawn in the full 'range' of the spectrum.

16.26.5 Shapely Colours

The RasMol 'shapely' colour scheme colour codes residues by amino acid property. This scheme is based upon Bob Fletterick's "Shapely Models". Each amino acid and nucleic acid residue is given a unique colour. The 'shapely' colour scheme is used by David Bacon's Raster3D program. This colour scheme is similar to the 'amino' colour scheme.

16.26.6 Structure Colours

The RasMol 'structure' colour scheme colours the molecule by protein secondary structure. Alpha helices are coloured magenta, beta sheets are coloured yellow, turns are coloured pale blue, [96,128,255] and all other residues are coloured white. The secondary structure is either read from the PDB file (HELIX and SHEET records), if available, or determined using Kabsch and Sander's DSSP algorithm. The RasMol 'structure' command may be used to force DSSP's structure assignment to be used.

16.26.7 Temperature Colours

The RasMol 'temperature' colour scheme colour codes each atom according to the anisotropic temperature (beta) value stored in the PDB file. Typically this gives a measure of the mobility/ uncertainty of a given atom's position. High values are coloured in warmer (red) colours and lower values in colder (blue) colours. This feature is often used to associate a "scale" value [such as amino acid variability in viral mutants] with each atom in a PDB file, and colour the molecule appropriately.

The difference between the 'temperature' and 'charge' colour schemes is that increasing temperature values proceed from blue to red, whereas increasing charge values go from red to blue.

16.26.8 Charge Colours

The RasMol 'charge' colour scheme colour codes each atom according to the charge value stored in the input file (or beta factor field of PDB files). High values are coloured in blue (positive) and lower values coloured in red (negative). Rather than use a fixed scale this scheme determines the maximum and minimum values of the charge/temperature field and interpolates from red to blue appropriately. Hence, green cannot be assumed to be 'no net charge' charge.

The difference between the 'charge' and 'temperature' colour schemes is that increasing temperature values proceed from blue to red, whereas increasing charge values go from red to blue.

If the charge/temperature field stores reasonable values it is possible to use the RasMol 'colour dots potential' command to colour code a dot surface (generated by the 'dots' command) by electrostatic potential.

16.26.9 User Colours

The RasMol 'user' colour scheme allows RasMol to use the colour scheme stored in the PDB file. The colours for each atom are stored in COLO records placed in the PDB data file. This convention was introduced by David Bacon's Raster3D program.

16.26.10 HBond Type Colours

The RasMol 'type' colour scheme applies only to hydrogen bonds, hence is used in the command “colour hbonds type” This scheme colour codes each hydrogen bond according to the distance along a protein chain between hydrogen bond donor and acceptor. This schematic representation was introduced by Belhadj-Mostefa and Milner-White. This representation gives a good insight into protein secondary structure (hbonds forming alpha helices appear red, those forming sheets appear yellow and those forming turns appear magenta).

Offset	Colour	Triple
+2	white	[255,255,255]
+3	magenta	[255,0,255]
+4	red	[255,0,0]
+5	orange	[255,165,0]
-3	cyan	[0,255,255]
-4	green	[0,255,0]
default	yellow	[255,255,0]

16.26.11 Potential Colours

The RasMol 'potential' colour scheme applies only to dot surfaces, hence is used in the command “colour dots potential” This scheme colours each currently displayed dot by the electrostatic potential at that point in space. This potential is calculated using Coulomb's law taking the temperature/charge field of the input file to be the charge associated with that atom. This is the same interpretation used by the 'colour charge' command. Like the 'charge' colour scheme low values are blue/white and high values are red. The table below shows the static assignment of colours using a dielectric constant value of 10.

25 < V	red	[255,0,0]
10 < V < 25	orange	[255,165,0]
3 < V < 10	yellow	[255,255,0]
0 < V < 3	green	[0,255,0]
-3 < V < 0	cyan	[0,255,255]
-10 < V < -3	blue	[0,0,255]
-25 < V < -10	purple	[160,32,240]
V < -25	white	[255,255,255]

16.27 RasMol File Formats

16.27.1 Brookhaven Data Bank Files

If you do not have the Brookhaven documentation, you may find the following summary of the PDB file format useful. The Protein Data Bank is a computer-based archival database for macromolecular structures. The database was established in 1971 by the Brookhaven National Laboratory, New York, as a public domain repository for resolved crystallographic structures. The Bank uses a uniform format to store atomic co-ordinates and partial bond connectivities as derived from crystallographic studies.

PDB file entries consist of records of 80 characters each. Using the punched card analogy, columns 1 to 6 contain a record-type identifier, the columns 7 to 70 contain data. Columns 71 to 80 are normally blank, but may contain sequence information added by library management programs. The first four characters of the record identifier are sufficient to identify the type of record uniquely, and the syntax of each record is independent of the order of records within any entry for a particular macromolecule. The only record types that are of major interest to the RasMol program are the ATOM and HETATM records which describe the position of each atom. ATOM/HETATM records contain standard atom names and residue abbreviations, along with sequence identifiers, co-ordinates in Angstrom units, occupancies and thermal motion factors. The exact details are given below as a FORTRAN format statement.[FORMAT(6A1,I5,1X,A4,A1,A3,1X,A1,I4,A1,3X,3F8.3,2F6.2,1X,I3)]

Column	Content
1-6	'AOM' or 'HETAM'
7-11	Atom serial number (may have gaps)
13-16	Atom name, in IUPAC standard format
17	Alternate location indicator indicated by A, B or C
18-20	Residue name, in IUPAC standard format
23-26	Residue sequence number (ordered as below)
27	Code for insertions of residues (i.e. 66A & 66B)
31-38	X co-ordinate
39-46	Y co-ordinate
47-54	Z co-ordinate
55-60	Occupancy
61-66	Temperature factor
68-70	Footnote number

Residues occur in order of their sequence numbers, which always increase starting from the N-terminal residue for proteins and 5'-terminus for nucleic acids. If the residue sequence is known, certain atom serial numbers may be omitted to allow for future insertion of any missing atoms. Within each residue, atoms are ordered in a standard manner, starting with the backbone (N-C-C-O for proteins) and proceeding in increasing remoteness from the alpha carbon, along the side chain.

HETATM records are used to define post-translational modifications and cofactors associated with the main molecule. Optional TER records are interpreted as breaks in the main molecule's backbone.

If present, RasMol also inspects HEADER, COMPND, HELIX, SHEET, TURN, CONECT, CRYST1, MODEL, ENDM and END records. Information such as the name, Brookhaven code, revision date and classification of the molecule are extracted from HEADER and COMPND records, initial secondary structure assignments are taken from HELIX, SHEET and TURN records, and the end of the file may be indicated by an END record.

16.27.2 RasMol Interpretation of PDB fields

Atoms located at 9999.000, 9999.000, 9999.000 are assumed to be Insight pseudo atoms and are ignored by RasMol. Atom names beginning ' Q' are also assumed to be pseudo atoms or position markers.

When a data file contains an NMR structure, multiple conformations may be placed in a single PDB file delimited by several MODEL and ENDM records. In this case, RasMol only displays the first NMR model displayed in the file.

Residue names "CSH", "CYH" and "CSM" are considered pseudonyms for cysteine "CYS". Residue names "WAT", "H20", "SOL" and "TIP" are considered pseudonyms for water "HOH". The residue name "D20" is consider heavy water "DOD". The residue name "SUL" is considered a sulphate ion "SO4". The residue name "CPR" is considered to be cis-proline and is translated as "PRO". The residue name "TRY" is considered a pseudonym for tryptophan "TRY".

RasMol uses the HETATM fields to define the sets hetero, water, solvent and ligand. Any group with the name "HOH", "DOD", "SO4" or "PO4" (or aliased to one of these names by the preceding rules) is considered a solvent and is considered to be defined by a HETATM field.

RasMol only respect CONECT connectivity records in PDB files containing less than 256 atoms. This is explained in more detail in the section on determining molecule connectivity. CONECT records that define a bond more than once are interpreted as specifying the bond order of that bond, i.e. a bond specified twice is a double bond and a bond specified three (or more) times is a triple bond.

16.27.3 PDB Colour Scheme Specification

RasMol also accepts the supplementary COLO record type in the PDB files. This record format was introduced by David Bacon's Raster3D program for specifying the colour scheme to be used when rendering the molecule. This extension is not currently supported by Brookhaven. The COLO record has the same basic record type as the ATOM and HETATM records described above.

Colours are assigned to atoms using a matching process. The Mask field is used in the matching process as follows. First RasMol reads in and remembers all the ATOM, HETA and COLO records in input order. When the user-defined ('User') colour scheme is selected, RasMol goes through each remembered ATOM/HETA record in turn, and searches for a COLO record that matches in all of columns 7 through 30. The first such COLO record to be found determines the colour and radius of the atom.

Column	Content
1-6	'COLOR' or 'COLOUR'
7-30	Mask (described below)
31-38	Red component
39-46	Green component
47-54	Blue component
55-60	Sphere radius in Angstroms
61-70	Comments

Note that the Red, Green and Blue components are in the same positions as the X, Y, and Z components of an ATOM or HETA record, and the van der Waals radius goes in the place of the Occupancy. The Red, Green and Blue components must all be in the range 0 to 1.

In order that one COLO record can provide colour and radius specifications for more than one atom (e.g. based on residue, atom type, or any other criterion for which labels can be given somewhere in columns 7 through 30), a 'don't-care' character, the hash mark "#" (number or sharp sign) is used. This character, when found in a COLO record, matches any character in the corresponding column in a ATOM/HETA record. All other characters must match identically to count as a match. As an extension to the specification, any atom that fails to match a COLO record is displayed in white.

16.28 RasMol Bibliography

16.28.1 Molecular Graphics

[1] Nelson Max, "**Computer Representation of Molecular Surfaces**", *IEEE Computer Graphics and Applications*, pp.21-29, August 1983.

[2] Arthur M. Lesk, "**Protein Architecture: A Practical Approach**", IRL Press Publishers, 1991.

16.28.2 Molecular Graphics Programs

[3] Per J. Kraulis, "**MOLSCRIPT: A Program to Produce both Detailed and Schematic Plots of Protein Structures**", *Journal of Applied Crystallography,* Vol.24, pp.946-950, 1991.

[4] David Bacon and Wayne F. Anderson, "**A Fast Algorithm for Rendering Space-Filling Molecule Pictures**", *Journal of Molecular Graphics,* Vol.6, No.4, pp.219-220, December 1988.

[5] David C. Richardson and Jane S. Richardson, "**The Kinemage: A tool for Scientific Communication**", *Protein Science,* Vol.1, No.1,pp.3-9, January 1992.

[6] Mike Carson, "**RIBBONS 2.0**", *Journal of Applied Crystallography,* Vol.24, pp.958-961, 1991.

[7] Conrad C. Huang, Eric F. Pettersen, Teri E. Klein, Thomas E. Ferrinand Robert Langridge, "**Conic: A Fast Renderer for Space-FillingMolecules with Shadows**", *Journal of Molecular Graphics,* Vol.9, No.4, pp.230-236, December 1991.

16.28.3 Molecular Biology Algorithms

[8] Wolfgang Kabsch and Christian Sander, "**Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features**", *Biopolymers,* Vol.22, pp.2577-2637, 1983.

[9] M. Connolly, "**Solvent-Accessible Surfaces of Proteins and Nucleic Acids**", *Science,* Vol.221, No.4612, pp.709-713, August 1983.

[10] Khaled Belhadj-Mostefa, Ron Poet and E. James Milner-White, "**Displaying Inter-Main Chain Hydrogen Bond Patterns in Proteins**", *Journal of Molecular Graphics,* Vol.9, No.3, pp.194-197, September 1991.

[11] Mike Carson, "**Ribbon Models of Macromolecules**", *Journal of Molecular Graphics,* Vol.5, No.2, pp.103-106, June 1987.

[12] Mike Carson and Charles E. Bugg, "**Algorithm for Ribbon Models of Proteins**", *Journal of Molecular Graphics,* Vol.4, No.2, pp.121-122, June 1986.

[13] H. Iijima, J. B. Dunbar Jr. and G. Marshall, "**Calibration of Effective Van der Waals Atomic Contact Radii for Proteins and Peptides**", *Proteins: Structure, Functions and Genetics,* Vol.2, pp.330-339,1987.

16.28.4 Graphics Algorithms

[14] J. Foley, A. van Dam, S. Feiner and J. Hughes, "**Computer Graphics: Principles and Practice**", 2nd Edition, Addison Wesley Publishers, 1990.

[15] J. Cleary and G. Wyvill, "**Analysis of an Algorithm for Fast Ray Tracing using**

Uniform Space Subdivision", *The Visual Computer*, Vol.4, pp.65-83, 1988.

[16] Thomas Porter,"**Spherical Shading**", *Computer Graphics* Vol.12, ACM SIGGRAPH, pp.282-285, 1978.

[17] Jean-Michel Cense, "**Exact Visibility Calculation for Space-Filling Molecular Models**", *Journal of Molecular Graphics*, Vol.9, No.3, pp.191-193, September 1991.

[18] Chris Schafmeister, "**Fast Algorithm for Generating CPK Images on Graphics Workstations**", *Journal of Molecular Graphics*, Vol.8, No.4, pp.201-206, December 1990.

[19] Bruce A. Johnson, "**MSURF: A Rapid and General Program for the Representation of Molecular Surfaces**", *Journal of Molecular Graphics*, Vol.5, No.3, pp.167-169, September 1987.

16.28.5 File Formats

[20] Frances C. Bernstein et al., "**The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures**", *Journal of Molecular Biology*, Vol.112, pp.535-542, 1977.

[21] Arthur Dalby, Jmes G. Nourse, W. Douglas Hounshell, Ann K. I. Gushurst,David L. Grier, Burton A. Leland and John Laufer, "**Description of Several Chemical File Formats Used by Computer Programs Developed at Molecular Design Limited**", *Journal of Chemical Information and Computer Sciences*, Vol.32, No.3, pp.244-255, 1992.

[22] Adobe Systems Inc., "**PostScript Language Reference Manual**", Addison-Wesley Publishers, Reading, Mass., 1985.

16.29 PLUTO

16.29.1 The Pluto Interface

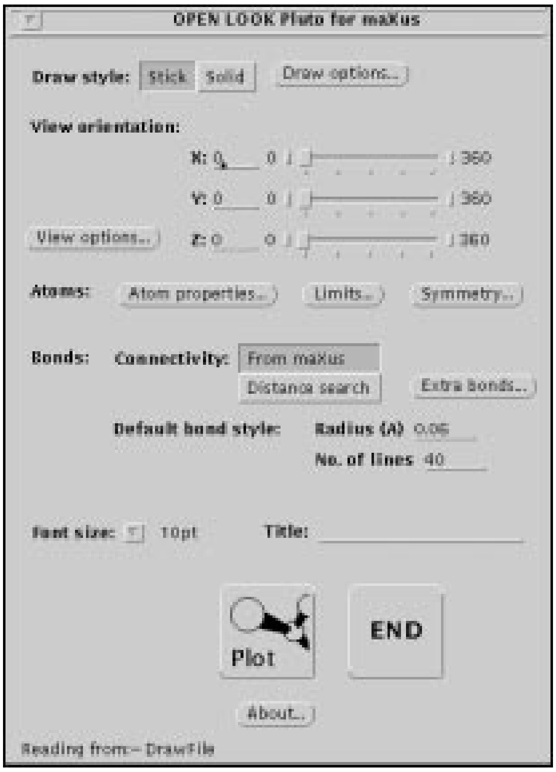
Pluto is a program for viewing molecules but efficient manipulation of the diagrams requires a rather complicated command sequence and a good deal of familiarity with the package. The graphics interface to Pluto is intended to overcome these problems, allowing the user very rapid access to the numerous display modes that Pluto offers. An inexperienced user can quickly see the results of changing program parameters and learn to use the full facilities of the program in a very short time. Extensive use of “active” and “inactive” controls is made to help guide the user through the many available options.

16.29.2 Interface to maXus

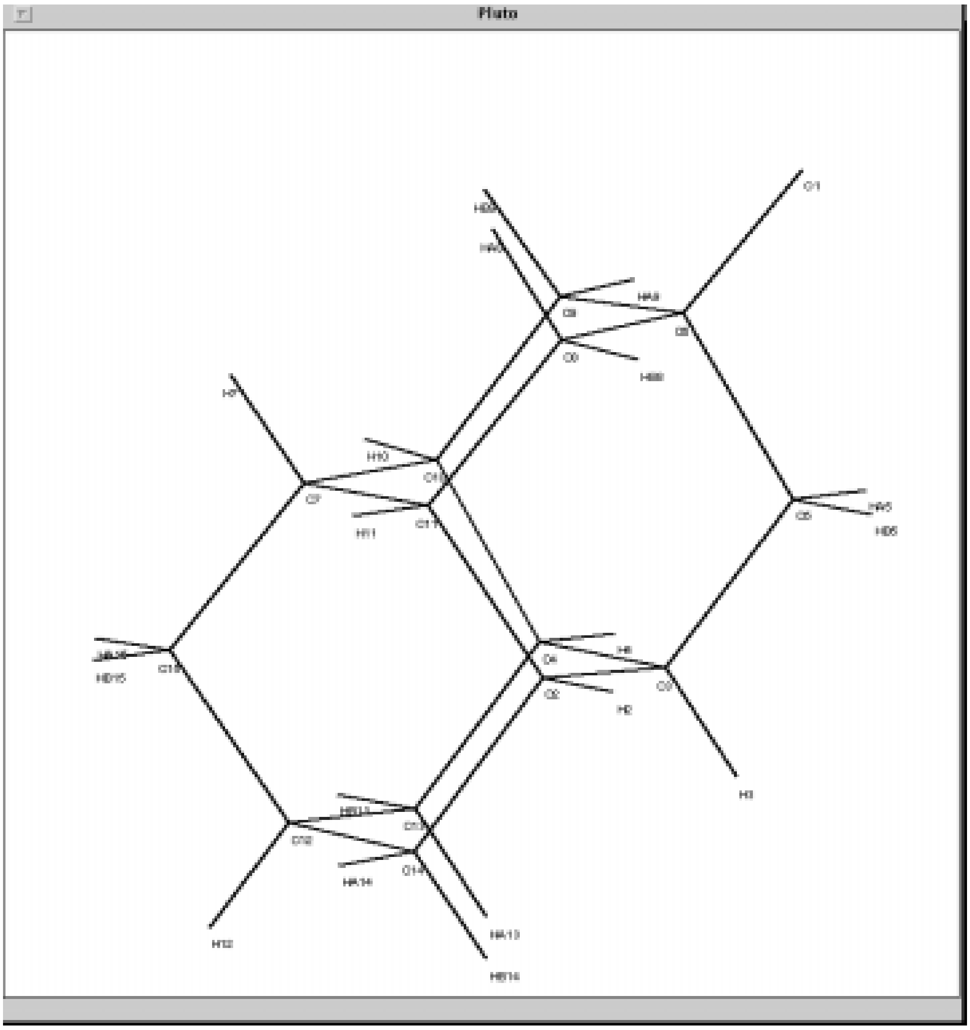
Pluto reads the DrawFile created by maXus and takes from it the atom positions and connectivities, along with the symmetry operators.

16.29.3 Starting the program

Assuming a structure has been refined, choose either *View Model* or *View Expanded* from the *Pluto* submenu of the *Graphics* menu. The following dialogue box appears:



The molecule can be drawn immediately by clicking on the **Plot** button. A stick diagram will appear:



16.29.4 Basic Operation

All mouse operations are carried out using the left mouse button except for the “Label font” abbreviated menu button, where the menu is displayed by clicking right over the button and releasing once the required choice has been made. Pop-up windows are dismissed by pulling out the small pushpin in the top left of the frame.

Once the main Pluto window has appeared, the user can click on the **Plot** button to display a stick molecule. The default view is currently set to view along the x-axis towards the origin of the unit cell. The displayed molecule can be dismissed by dismissing the frame around the molecule in the usual manner. The display can be printed to a PostScript printer from the pop-up which appears when the right mouse button is clicked over the molecule display window.

After making any alteration to the default settings in the main window or any of the pop-up windows, the results of the change can be displayed by again clicking on **Plot**. Note that previous diagrams are not cleared, so that comparisons may be made between different views and styles. As many diagrams may be displayed as are required, though it is the user's responsibility to quit from each diagram window; Pluto **will not** clear them automatically when it terminates.

Alterations to the basic display fall into the following categories:-

Drawing style	Stick or solid, stereo, perspective, shaded, tapering of bonds, display of unit cell, colour or black and white display.
View orientation	The exact point from which a diagram is seen
Limits	Restrict the viewing area by a box, sphere or cell subsection
Atom properties	Atom labelling, display size
Symmetry	Select symmetry operators to apply to molecule
Bond properties	Connectivity, appearance, specific bonds
Miscellaneous	Label font size, title for diagram

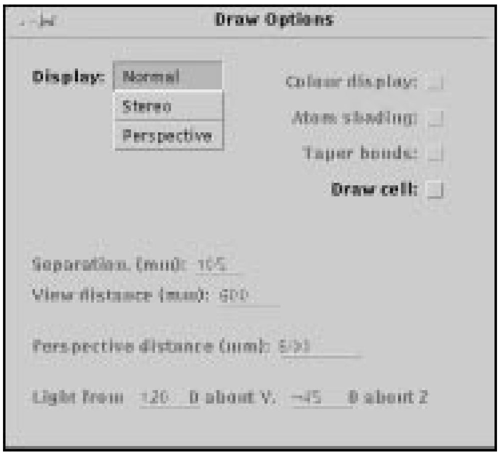
The various menus associated with the above categories are described in subsequent sections.

16.29.5 Drawing Style

Main options are either:

Stick	Draw a black on white stick model, or
Solid	Draw a ball and stick model in colour on a black background. Ball appearance can be altered from the **Atom Properties** pop-up menu, as can the bond appearance to particular atoms or element types. **Default bond style** can be altered on the main menu.

Further options are available by clicking *Draw Options* button, giving:



Display: Normal
Colour display: ☐
Atom shading: ☐
Taper bonds: ☐
Draw cell: ☐
Separation (mm): 105
View distance (mm): 600
Perspective distance (mm): 500
Light from 120 θ about Y: −45 θ about Z

Stereo	Draws a stereo pair of either a stick or ball and stick model. Default separation of the pair is 105mm with a view distance of 600mm. Setting the separation to an excessively large value will cause the molecules to slip off the edge of the window - a separation of 105mm is usually sufficient.

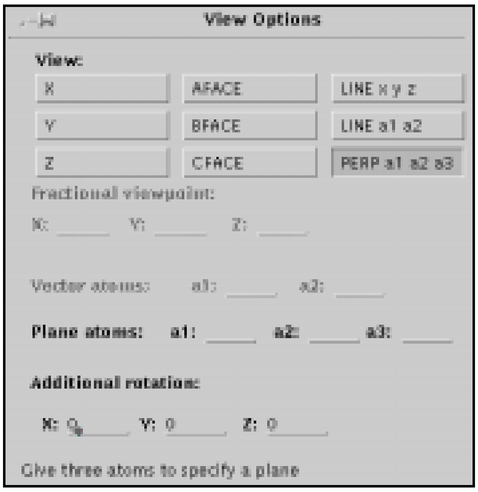
Perspective	Draws a perspective view of a stick or ball and stick model. Default view distance is 600mm, but this can be altered.
Atom shading	Applicable only to ball and stick models, this option draws a line across the atoms to simulate the effect of a light source coming from 120 degrees about Y, -45 degrees about Z.
Draw cell	Draws the unit cell as well as the displayed molecule.
Colour display	When turned off, the ball and stick model will be displayed as black on white. This mode should be selected if the display is to be printed on a black and white printer.
Taper bonds	Draws bonds as tapered cones.

16.29.6 View orientation

Options:

View orientation	This is the only orientation option available on the main Pluto window. Specify three rotation angles using the sliders to define a viewpoint relative to looking along the x-axis towards the origin. The sliders can be operated in the usual manner i.e. Click left over the slider handle and move it to the desired position *or* Click left over either of the end bars *or* Enter a value in the associated text field and press return

The following options are all found by clicking on the *View options* button:



X: 0 Y: 0 Z: 0

Give three atoms to specify a plane

Options available:

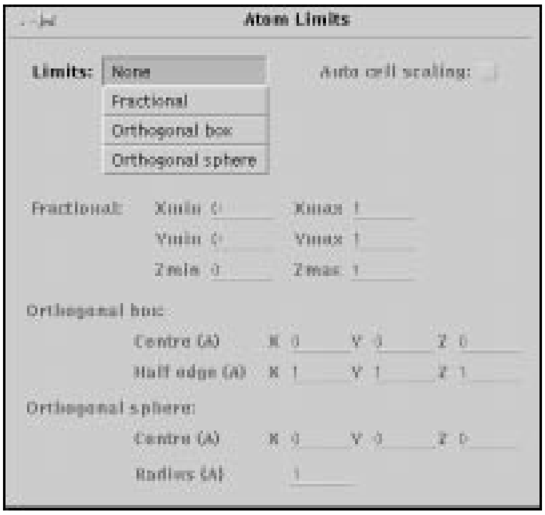
X	View along X axis towards the origin
Y	View along Y axis towards the origin
Z	View along Z axis towards the origin

AFACE	View from 1,0.5,0.5 towards the origin
BFACE	View from 0.5,1,0.5 towards the origin
CFACE	View from 0.5,0.5,1 towards the origin
LINE x y z	View from fractional point x,y,z towards the origin
LINE a1 a2	View along line between atom ‘a1’ and atom ‘a2’
PERP a1 a2 a3	View perpendicular to the plane defined by atom ‘a1’, atom ‘a2’ and atom ‘a3’

Note that rotations about any or all of the axes may be applied *in addition* to the view currently selected by using the **Additional rotation** textfields.

16.29.7 Limits

Distance search mode is automatically selected when a limit is activated, as Pluto uses symmetry operators to transform atoms in the specified region of the unit cell. Click the *Limits* button in the main window:



Limits: None

Fractional: Xmin 0 Xmax 1
Ymin 0 Ymax 1
Zmin 0 Zmax 1

Orthogonal box:
Centre (A) X 0 Y 0 Z 0
Half edge (A) X 1 Y 1 Z 1

Orthogonal sphere:
Centre (A) X 0 Y 0 Z 0
Radius (A) 1

Options available:

None	Use the default display of all atoms in the unit cell
Fractional	Specify the atom limits in fractional coordinates. Only atoms within these limits are plotted. Unit cell transformations are applied if necessary to put the atom into this volume.
Orthogonal box	Select atoms in a box centred at x,y,z orthogonal Angstrom coordinates with half-edges x,y,z in Angstroms.
Orthogonal sphere	Select atoms in a sphere centred at x,y,z orthogonal sphere Angstrom coordinates with a radius specified in Angstroms.
Auto cell scaling	If the unit cell is displayed, scale the diagram to the new fractional bounds

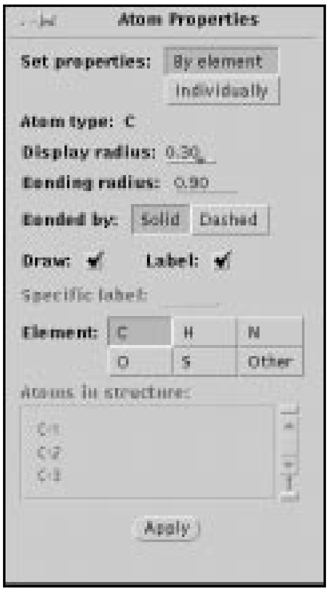
NOTE

If Plot fails to produce a diagram when certain limits have been specified, it is because there are no atoms in the specified area.

If you need to see orthogonal Å coordinates, generate any Pluto diagram and then look in the current maXus directory (i.e. the one where the DrawFile resides) in the file called Pluto.pdb.

16.29.8 Atom properties

Click the *Atom Properties* button in the main window:



Atom type: C

Display radius: 0.30

Bonding radius: 0.90

- **Setting properties by element**

  Properties are assigned to a particular element type. The element types available are C,H,N,O,S and Other. The program is aware of what atom types have been read from the maXus DrawFile. Changes made to the display radius, bonding radius, bond style, draw or label options apply to ALL atoms of that type. Changes are applied to an element type when a new element type is selected, when the Apply button is pressed or when the selection mode is changed to “Individually”. If an undesirable change is made, it may be cancelled by pulling out the pushpin of the pop-up **before** any of the above actions are taken.

- **Setting individual properties**

  This option is useful when specific properties have to be assigned to one or more particular atoms. Atoms can be selected from the scrolling list, at which point the properties associated with that atom will be displayed. If no specific properties have yet been assigned, those displayed will be for the element type into which the atom has been automatically classified (i.e. C,H,N,O,S or Other). A particular label of up to six characters can be associated with the selected atom. Changes will **only** be made when the Apply button is pressed. If this atom is subsequently selected, a footnote will be displayed to remind the user that specific properties have already been assigned to this atom.

Options available:

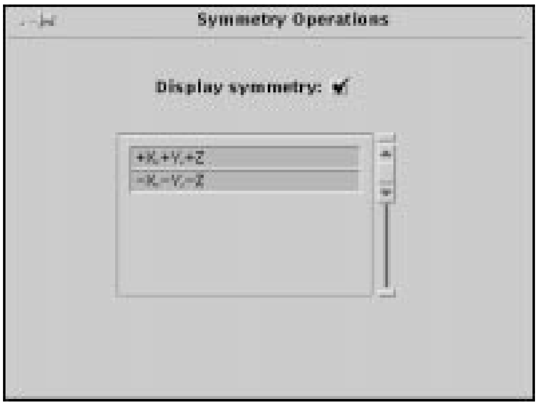
Display radius	Set the radius of the sphere representing the atom(s).
Bonding radius	Set the radius used by program to calculate the connectivity when the connectivity method is set to “Distance search”.
Draw	When tickmark is displayed, the atom(s) will be drawn in the diagram.
Label	Applicable only to atoms selected by element type. When tickmark is displayed, atom name is drawn next to the atom.
Specific label	Applicable only to specifically selected atoms. A label up to six characters long can be specified.

NOTE

Changes to atom properties are not carried over to symmetry copies of the molecule. This is a restriction of the original Pluto code.

16.29.9 Symmetry

Clicking the Symmetry button in the main window gives:



Display symmetry: ✓

+X,+Y,+Z
-X,-Y,-Z

If the `Display symmetry’ check box is ticked, the symmetry operators to be used are displayed in the scrolling list. These are all initially selected and may be deselected/selected by clicking with the left mouse button. Only those operators that are highlighted will be used. The Distance Search mode is automatically selected when symmetry related copies are displayed.

NOTE:

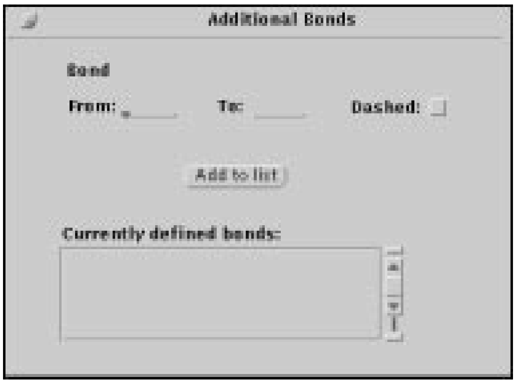
Changes to atom properties are not carried over to symmetry copies of the molecule. This is a restriction of the original Pluto code.

16.29.10 Bond Properties

The connectivity method is selectable from the main window. It is recommended that the connectivity is taken automatically from the maXus DrawFile. If a distance search is specified, connectivity will be calculated using the default bonding radii, which can be altered *via* the **Atom Properties** pop-up.

The default bond style can be varied from the main window. The default options produce a bond that appears solid. If 30 or fewer lines are requested per bond, the lines drawn will be single thickness. If 31 or more bonds are requested, a double thickness line will be used to give the bond an appearance of solidity.

If it is required to place a bond that is not present in the DrawFile, this can be done from the **Extra Bonds** pop-up.



The two atoms to be bonded are entered, the Dashed line style selected if required, and the Apply button clicked to enter the bond in the scrolling list. Entries can be deleted from the list by the following procedure.

Click the right mouse button over the scrolling list and select the “Edit List” option. Select the bond to be deleted, which will be highlighted. Click right again and select “Delete”. The bond will disappear when the mouse button is released. Click right again and select “End Editing”. The scrolling list will then return to normal operation.

Note that no checks are performed upon the validity of the atom names entered into the list. It is the user’s responsibility to ensure that valid atom names are entered - these can be obtained from scrolling list in the **Atom Properties** popup. Failure will cause an error with unpredictable results.

16.29.11 Miscellaneous

Other options:

Font size	Changes the font style used for labelling. This applies to all labels and the title. The font used is Helvetica and this must be available to the server.
Title	Specify a title of up to 40 characters long - the field will scroll as necessary.

About	Displays a pop-up containing information about the authors of the interface to Pluto.

16.29.12 Printing

To print a molecule display, plot the molecule on screen as described and, with the cursor over the display, click right to reveal the Printers pop-up. Whilst holding the right button down, select a printer and release the button. The cursor will change to a small cross. Click the left mouse button over the display. A beep will sound twice when the screen snap is completed.

16.29.13 Printing process

The picture displayed is captured with ‘xwd’ and piped into ‘xpr’ with the appropriate filter (PostScript or LaserJet).

16.29.14 Known problems

The LaserJet output option is untested and therefore not recommended.

17 Publish

Programs to Prepare Results for Publication: Geometries, CIF files, and Archiving

17.1 Introduction

The PUBLISH menu offers the following facilities:

- Geometry:

  The production of tables for publication containing:

  Atomic coordinates

  Anisotropic and isotropic temperature factors

  Bond lengths (with or without hydrogen atoms).

  Bond angles (with or without hydrogen atoms).

  Bond lengths and angles involving more than one asymmetric unit (with or without hydrogen atoms).

  Torsion angles (with or without hydrogen atoms).

  Non bonded contacts (with or without hydrogen atoms).

  Mean planes

The user can select options to view these tables, and print them in Postscript.

- Structure factor tables, with a wide variety of options.
- The production of CIF files for direct submission of papers to *Acta Crystallographica*, or *Zeitschift für Kristallograpie*.
- Facilities to edit and check the syntax of the CIF files that have been edited.
- A program to convert to a cssr type file for use with other modelling packages (especially Cerius™.

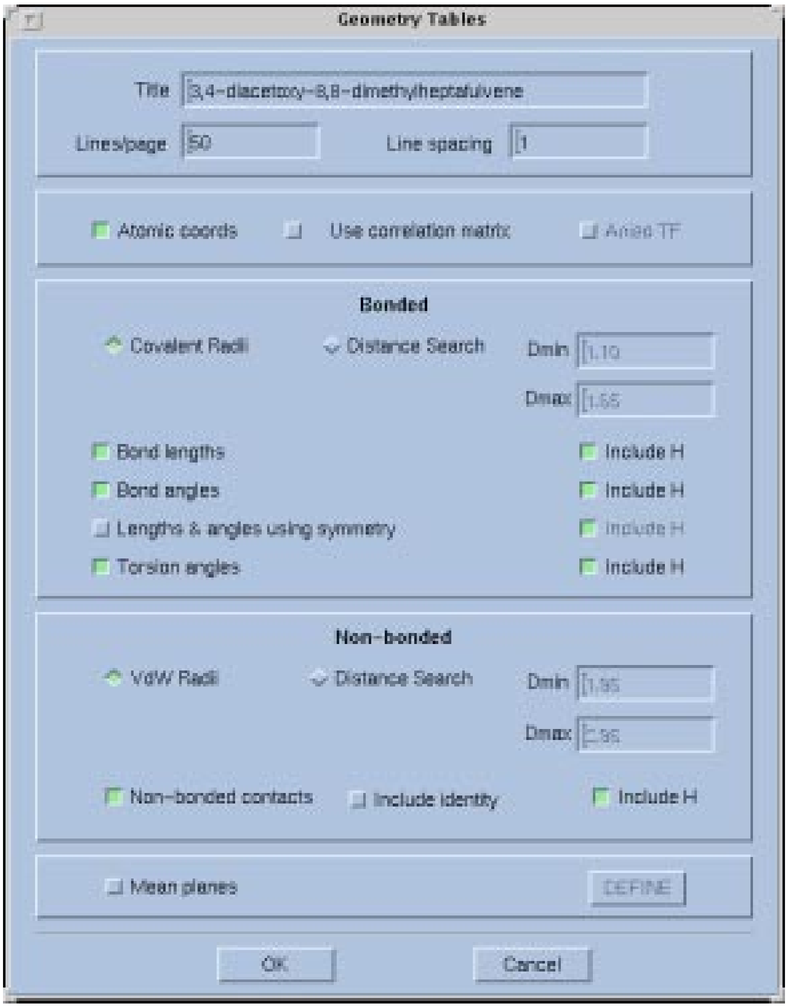
17.2 Geometry

17.2.1 Geometry Tables

To run the GEOMETRY program, choose *Geometry* from the *Publish* menu. There are four options in the sub-menu:

- Calculate
- View/Edit (Text)
- View (Postscript)
- Print (Postscript)

Selecting the *Calculate* option causes a new window to appear:



Geometry Tables

Title: 3,4-diacetoxy-8,8-dimethylheptafulvene
Lines/page: 50
Line spacing: 1

☑ Atomic coords
☐ Use correlation matrix
☐ Aniso TF

Bonded

● Covalent Radii
○ Distance Search
Dmin: 1.10
Dmax: 1.55

☑ Bond lengths ☑ Include H
☑ Bond angles ☑ Include H
☐ Lengths & angles using symmetry ☑ Include H
☑ Torsion angles ☑ Include H

Non-bonded

● VdW Radii
○ Distance Search
Dmin: 1.95
Dmax: 2.95

☑ Non-bonded contacts
☐ Include identity
☑ Include H

☐ Mean planes

The options are as follows:

- **Title**: Enter the title you need for publication. A default will be taken from the Original file header.
- **Lines per page**: Enter the number of lines per page for printing. The default is 50.
- **Line spacing**: The default is 1.
- **Atomic Coordinates**: If this button is highlighted then a table of atomic coordinates and equivalent isotropic temperature factors will be produced.
- **Use Correlation matrix**: If this button is highlighted then the full correlation matrix from least squares will be used in the estimation of errors in geometries. If not, then cell angle cosines are used as covariances, and variances are taken from the least squares. The default is *ON*.
- **Aniso TF**: If this button is highlighted then a table of anisotropic temperature factors will be produced if present.

Bonded

- **Covalent radii**: This option uses covalent radii to decide on what atoms are bonded.
- **Distance Search**: When deciding which atoms are bonded when calculating bond lengths, bond angles and torsion angles there are two choices - the program can use the limits given in the boxes (Dmin, Dmax, defaults 1.10 and 1.65A respectively) or use covalent radii. If the Distance search button is pressed then all interatomic distances between Dmin and Dmax are treated as bonds.
- **Bond lengths**: If this button is highlighted then a table of bond lengths will be produced for all bonds satisfying the bonding conditions set above. By default the **Include H** button is on, but it can be turned off so that the tables will not contain bond lengths involving H atoms.
- **Bond angles**: If this button is highlighted then a table of bond angles will be produced for all bonds satisfying the bonding conditions set above. By default the **Include H** button is on, but it can be turned off so that the tables will not contain bond angles involving H atoms.
- **Lengths and Angles Using Symmetry**: This is an alternative to using the separate lengths and angles options above. When this is selected, the other two options will become unavailable. When the bond calculations are carried out, each symmetry operator will be applied in turn, so that bonds between atoms in different asymmetric units will be evaluated. This is most useful when only part of a molecule is in a single asymmetric unit. The **Include H** button works in the normal way.
- **Torsion angles**: If this button is highlighted then a table of torsion angles will be produced for all bonds satisfying the bonding conditions set above. The **Include H** button works in the normal way.

Non-bonded

- **Distance Search**: Interatomic distances between Dmin and Dmax will be calculated. The default values are 1.95Å and 2.95Å.
- **VdW Radii**: This option calculates interatomic distances greater than the sum of the covalent radii, but less then the sum of the van der Waals radii.
- **Non-bonded contacts**: If this button is highlighted then a table of non-bonded lengths will be produced for all contacts according to the conditions above. The **Include H** button works in the normal way.

Mean Planes

- **Mean planes**: Selecting this option, and clicking *Define* opens a Model window displaying the molecule. All atoms are set to be the same size at this point. There is an ^ and v option which selects the plane number. Use this to select the relevant number, click on the atoms for which the mean plane is required, the atoms thus selected go to double size. Use the ^ and v to select a new plane, then use RETURN to return to the Tables window.

IMPORTANT: Some of the non-bonded contacts, and torsion angle options can be very slow and produce a lot of output unless dmin and dmax are sensibly set!

When all options have been set then click OK. There may be a long pause. A peformance indicator appears at the top of the window:

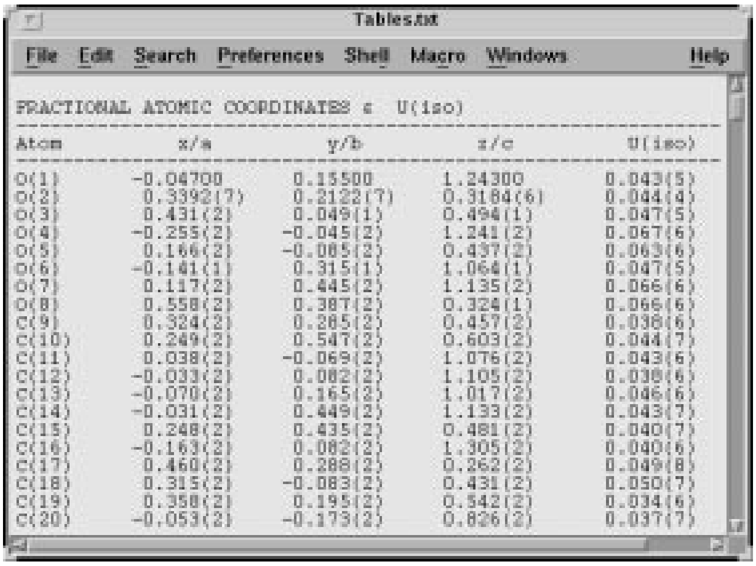
When the program has completed, the dialogue box:

Geometry Complete

appears. Click OK, and the calculations are complete.

17.2.2 View/Edit Text

If you select this option from the *Geometry* sub-menu, a text editor appears with the results in ASCII text of the geometry calculations.

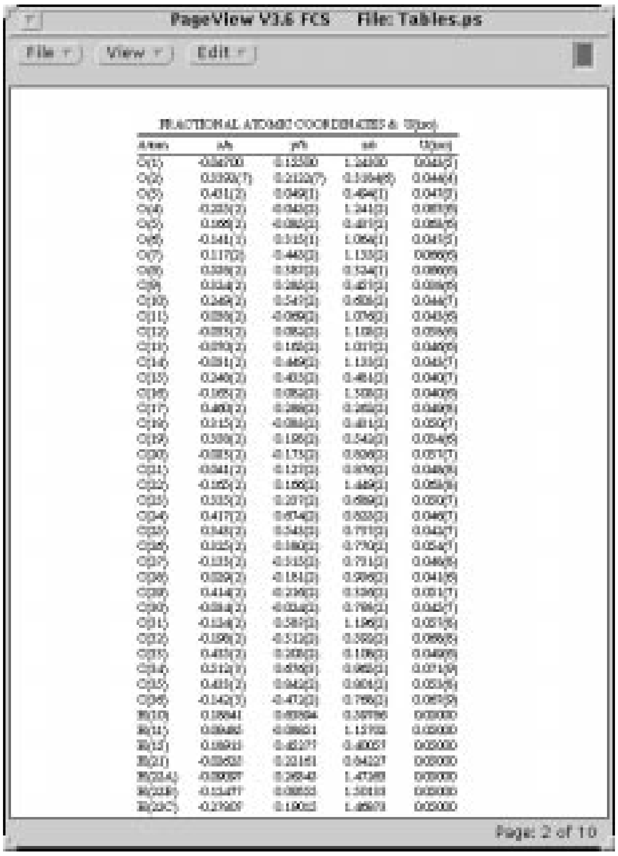


```
FRACTIONAL ATOMIC COORDINATES &  U(iso)
----------------------------------------------------------------------
Atom           x/a           y/b           z/c            U(iso)
----------------------------------------------------------------------
O(1)       -0.04700       0.15500       1.24300        0.043(5)
O(2)        0.3392(7)     0.2122(7)     0.3184(6)      0.044(4)
O(3)        0.431(2)      0.049(1)      0.494(1)       0.047(5)
O(4)       -0.255(2)     -0.045(2)      1.241(2)       0.067(6)
O(5)        0.166(2)     -0.085(2)      0.437(2)       0.063(6)
O(6)       -0.141(1)      0.315(1)      1.064(1)       0.047(5)
O(7)        0.117(2)      0.445(2)      1.135(2)       0.066(6)
O(8)        0.558(2)      0.387(2)      0.324(1)       0.066(6)
C(9)        0.324(2)      0.285(2)      0.457(2)       0.038(6)
C(10)       0.249(2)      0.547(2)      0.603(2)       0.044(7)
C(11)       0.038(2)     -0.069(2)      1.076(2)       0.043(6)
C(12)      -0.033(2)      0.082(2)      1.105(2)       0.038(6)
C(13)      -0.070(2)      0.165(2)      1.017(2)       0.046(6)
C(14)      -0.031(2)      0.449(2)      1.133(2)       0.043(7)
C(15)       0.248(2)      0.435(2)      0.481(2)       0.040(7)
C(16)      -0.163(2)      0.082(2)      1.305(2)       0.040(6)
C(17)       0.460(2)      0.288(2)      0.262(2)       0.049(8)
C(18)       0.315(2)     -0.083(2)      0.431(2)       0.050(7)
C(19)       0.358(2)      0.195(2)      0.542(2)       0.034(6)
C(20)      -0.053(2)     -0.173(2)      0.826(2)       0.037(7)
```

This file is called Tables.txt, and can be used to produce any tables you wish in ASCII.

17.2.3 View Postscript

If you select this option from the *Geometry* sub-menu, a Postscript viewer appears with the results in Postscript of the geometry calculations.



This file is called Tables.ps. It cannot be edited.

17.2.4 Print Postscript

This prints the Postscript tables.

17.3 PLATON

Platon is a geometry program. The implementation in maXus provides functions in PLATON that are not available in the GEOMETRY program. These functions include the searching of the unit cell for voids, for solvent spaces, the calculation and plotting of Newman projections, and the analysis of rigid body thermal motion parameters.

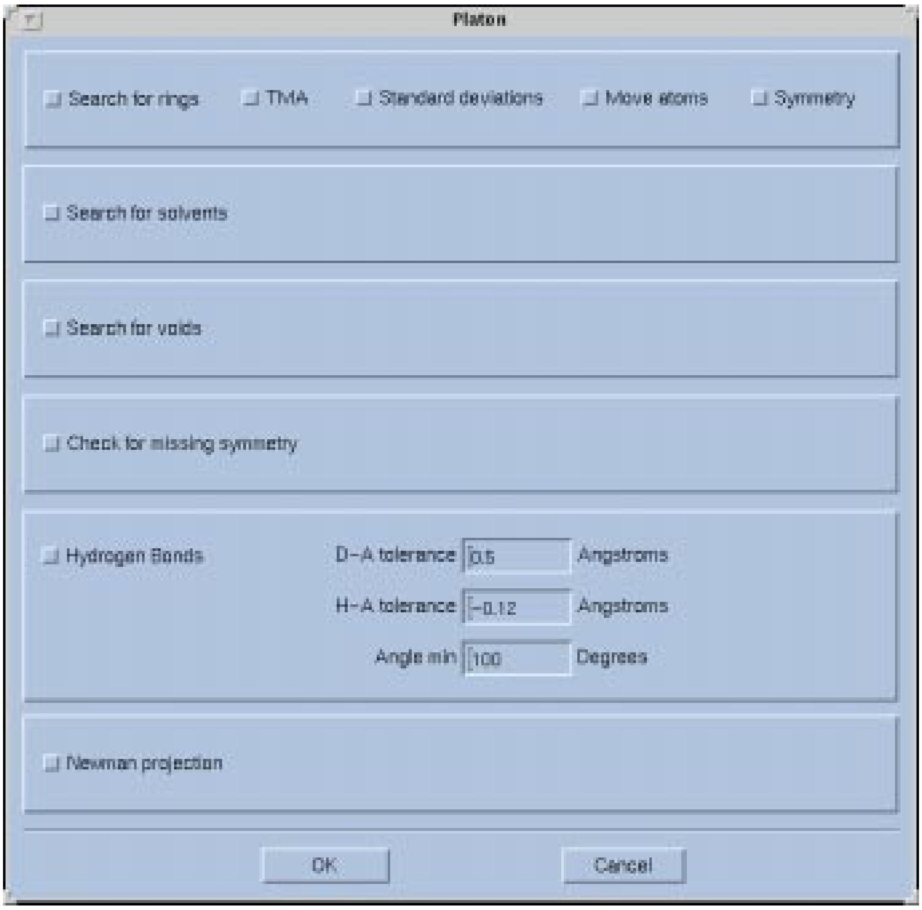
For users familiar with the PLATON program, a facility to prepare user edited command files is available, and the full PLATON documentation is included in this manual as Appendix B.

17.4 PLATON in Default Graphical Interface Mode

From the main menu, choose the *Publish* menu, followed by *Platon*, followed by *Default*:

17.4.1 The Main PLATON Window

This will bring up the main Platon window:



D-A tolerance 0.5 Angstroms
H-A tolerance -0.12 Angstroms
Angle min 100 Degrees

Once all parameters that require calculation have been selected click on the *OK* button at the bottom of the PLATON window. This will initiate processing.

Each of the options is explained below:

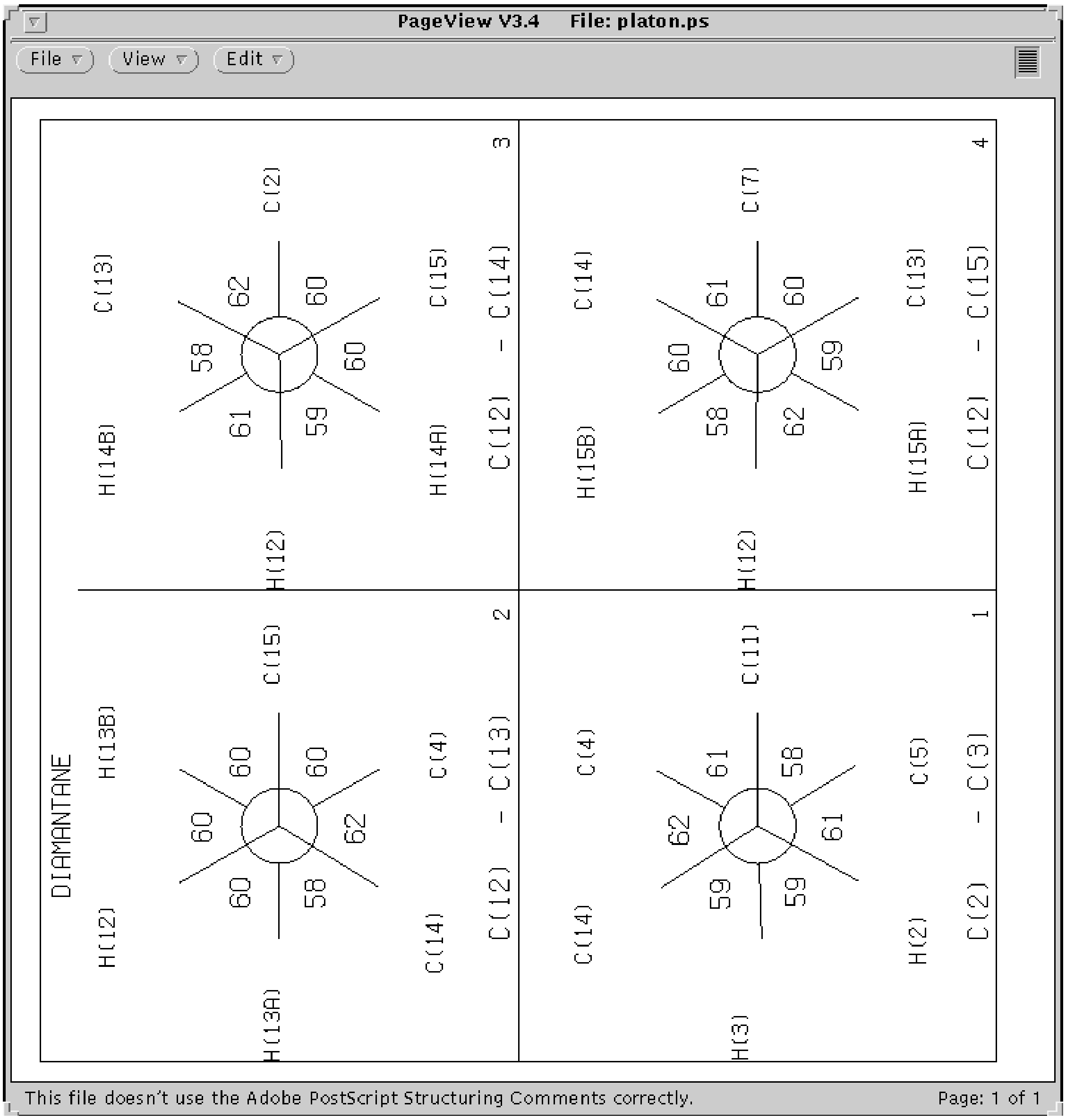
- **Search For Rings**: This option will search for and characterise rings of up to 12 atoms. Atoms with a separation of less than the sum of their covalent radii plus a tolerance limit of 0.4Å are considered bonded for the purposes of connectivity. The results of the ring search will be reported on in the detailed Summary file.
- **TMA:** A Thermal motion analysis will be carried out and reported on in the detail file in the summary section.
- **Standard Deviations**: Standard deviations will be calculated on all values. These will be reported on in the detail file in the summary section.
- **Move Atoms**: Atoms will be moved during the construction of a connected set. Any movement will be reported on in the detail file in the summary section.
- **Symmetry**: This will allow the expansion of the input dataset of atomic coordinates by translational and rotational symmetry, when considering the connectivity of a set.

- **Search For Solvents**: The unit cell will be checked for any voids capable of containing a solvent molecule of radius 1.2Å. The unit cell will be checked at a grid spacing of 0.2Å.
- **Search For Voids**: The unit cell will be checked for any voids of radius 1.2Å. The unit cell will be checked at a grid spacing of 0.2Å.

  Warning: This option is very slow.
- **Newman Projection**: A postscript file of four Newman projections applicable to the molecule are produced. After calculation a Postscript viewer window will appear containing the Newman projection graphics. An example of this is shown on the next page. Only four projections can be shown. If a specific Newman projection is required, you have to edit the file manually and insert the appropriate command:

  ```
  PLOT NEWMAN META at1 at2.
  ```

  Where `at1` and `at2` are the atom names that specify the central bond.



This file doesn't use the Adobe PostScript Structuring Comments correctly.

Page: 1 of 1

17.4.2 Platon Text Output

After Platon has completed all calculations, an edit window containing the detailed report will appear. This detail report is contained within file platon.lis and is 132 bytes wide.

Please note that if using an A4 printer this report must be printed in Landscape format.

A summary report of what calculations were performed along with summary results is also produced and is held in file Sysout_Platon. This file is accessible *via* the *Summary/Platon* option from the main menu.

17.5 Platon in User Edit Mode

From the main menu, choose *Publish*, followed by *Platon*, followed by *Edit Input File*:

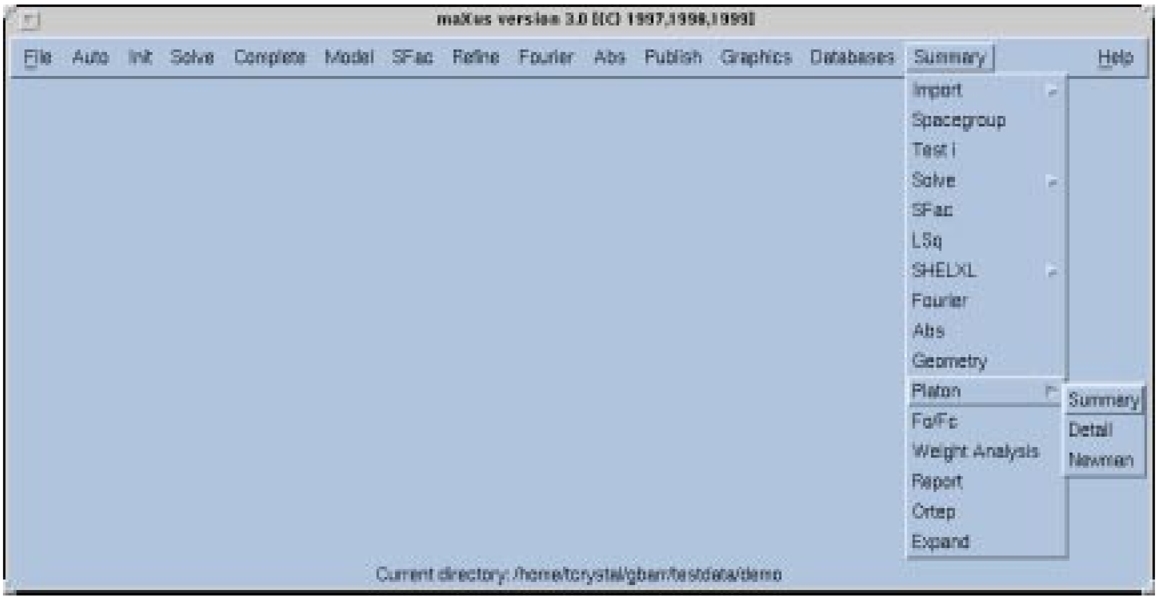
This will call up a text editor.

Using the information in Appendix B - “Documentation for Platon ’92”, the user can then enter commands for Platon in batch mode. Once the commands have been entered, save the file in the editor and exit.

To use this file, choose *Publish* from the main menu, followed by *Platon*, followed by *Use Edited*. Platon will now execute in batch mode using the file the user has entered as the controlling input. Running Platon in the User Edit Mode allows access to all of the PLATON functions from within maXus.

17.5.1 Accessing Platon Output

Summary information appears at the end of the Platon processing automatically. After this has been reviewed it is possible to return to this information *via* the *Summary* option from the main maXus menu.



Choose the *Summary* option, followed by *Platon*.

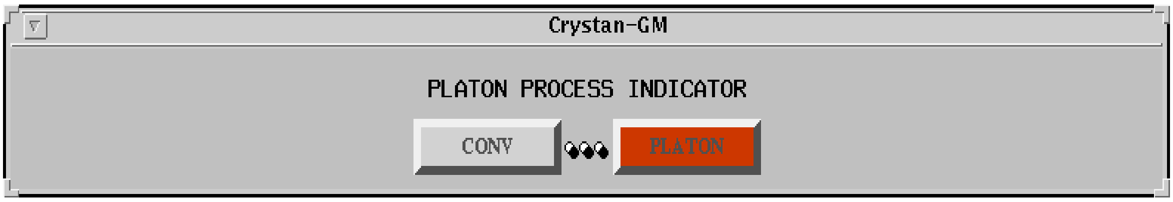
From this sub-menu the option *Summary* will access the file Sysout_Platon. This file contains summary information about the Platon run, including the options chosen, basic molecular information and run times. This file is a maximum of 80 bytes wide and is suitable for portrait printing on an A4 printer.

The sub menu option *Detail* will access file `platon.lis`. This file contains the detailed results of the last Platon calculation. This will include a torsion angle analysis if you have requested a Newman projection. This file is a maximum of 132 bytes wide and must be printed in landscape format on an A4 printer.

The sub menu option *Newman* will access file `platon.ps.` This file contains the postscript image produced after Platon is run with the Newman Projection request. A Postscript viewer will be used to display the image.

17.5.2 The Performance Meter

While PLATON is running a performance meter will appear at the top of the screen. When Platon has completed the performance meter will be closed. The first area of the meter will be highlighted while the conversion of file formats from maXus to Platon is underway, and the second area will be highlighted while Platon is executing. It is possible to close this window while Platon is running. The performance meter window is shown below:



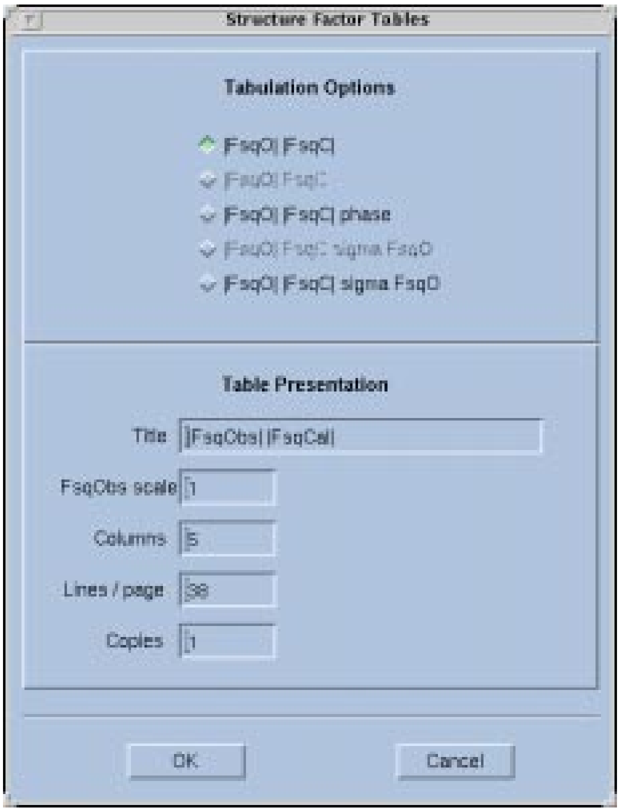
17.6 Structure Factor Tables

Choose ‘Structure factor tables’ in the ‘Publish’ menu. There are three options:

- Calculate.
- View
- Print

17.6.1 Calculate

Select the Calculate sub-menu; the following window will appear:



Title: |FsqObs| |FsqCal|

FsqObs scale: 1

Columns: 5

Lines / page: 38

Copies: 1

The top section allows you to select from the possible options to be tabulated:

- **|FsqO| |FsqC|** - i.e. a table of $|F^2_{obs}|$ $|F^2_{calc}|$
- **|FsqO| FsqC** - i.e. a table of $|F^2_{obs}|$ F^2_{calc}
- **|FsqO| |FsqC| phase** - i.e. a table of $|F^2_{obs}|$ $|F^2_{calc}|$ ϕ
- **|FsqO| FsqC sigma(FsqO)** - i.e. a table of $|F^2_{obs}|$ F^2_{calc} $\sigma(F^2_{obs})$
- **|FsqO| |FsqC| sigma(FsqO)** - i.e. a table of $|F^2_{obs}|$ $|F^2_{calc}|$ $\sigma(F^2_{obs})$

Two of the options ($|F^2_{obs}|$ F^2_{calc} and $|F^2_{obs}|$ F^2_{calc} $\sigma(F^2_{obs})$) are only available when working with a centrosymmetric structure.

In the lower half of the window, enter:

- The **Title**; the default is |FsqObs| |FsqCal|
- The **FsqObs scale factor**: the default is 1.
- The number of **Columns** per page: the default is 5.
- The number of **lines per page** for printing: the default is 38.
- The number of **copies** required: the default is 1.

Select the required options, and click *OK*. Click *Cancel* to exit. Once complete, the following dialogue box will appear:

Structure Factor tables complete

17.6.2 View

This brings up a text editor to view the tables just calculated.

17.6.3 Print

To print these tables, use this command.

17.7 CIF File Generation

To generate CIF files, choose *CIF* from the *Publish* menu. This produces a CIF file using Version 2.0 of the CIF dictionary. Four options are available through sub-menus:

17.7.1 Create

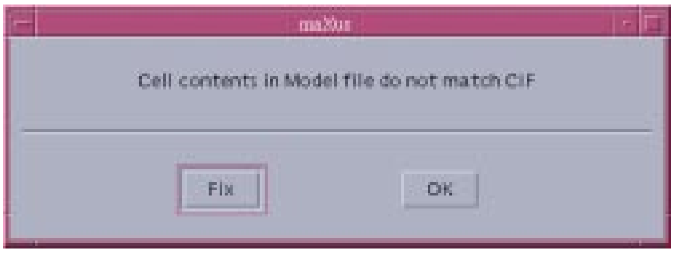
This option generates a new CIF file automatically.

The process of CIF generation is quite complex. (See flow chart of operation in Appendix C.) The CIF generator uses the *maXus.history* file to determine which options the user has selected at a particular stage, and collates the relevant CIF files to give a final complete file called *full.cif*.

If the history file has been reset, the CIF generator will be unable to function properly.

It works backwords throught the history file to see what modules have been run, and also automatically includes relevant references for those programs.

The Spacegroup program CIF is read in and checked for the chemical formula. This is then cross-checked with the Model file to see if the user has altered the unit cell contents (for example, by adding new atom types). If so, a dialog box will appear:



Cell contents in Model file do not match CIF

The sensible course at this point is to click *Fix*. This brings up a text-entry box:

Enter formula (e.g. C4 H7 O2 Br)

Enter the new complete formula in the box, in the format required. Finally:

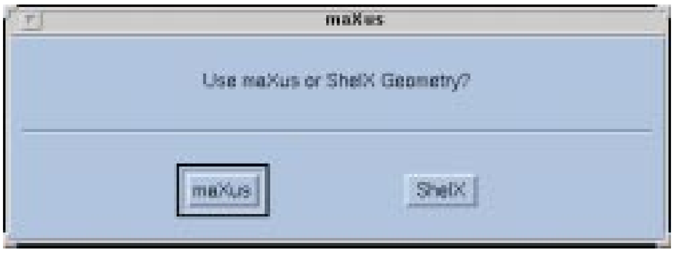
Enter Z (as integer)

Enter the Z value as asked. CIF generation then proceeds normally.

It then checks to find the last least squares refinement run, and see if Shelxl or the standard maXus LSQ program was used.

The history file is checked to see which, if any, absorbtion correction programs have been run, and read in the relevant CIF file, and add this to the *full.cif*.

If Shelxl least squares was used, the CIF file from this includes geometry and atomic coordinates. The CIF generators edit and substitute entries in this as necessary to extract the required information. If you have used Shelxl, but have run the maXus Geometry program as well, you will be asked which Geometry information you wish to include:



Use maXus or ShelX Geometry?

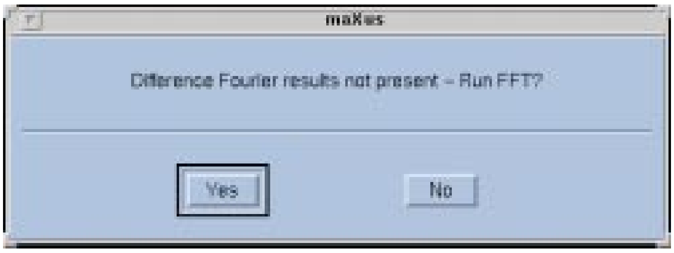
If maXus least squares was used, it checks if an up-to-date geometry output is present.

If the molecule has been refined since the last Geometry run was made, you will be prompted to re-run the module so as to get the most up-to-date results:

Out of date Geom file - Run Geom?

Note: this option runs the Geometry module using its default settings. These may not be ideal for your case, so running the module manually (Section 17.2) before generating your CIF file is recommended.

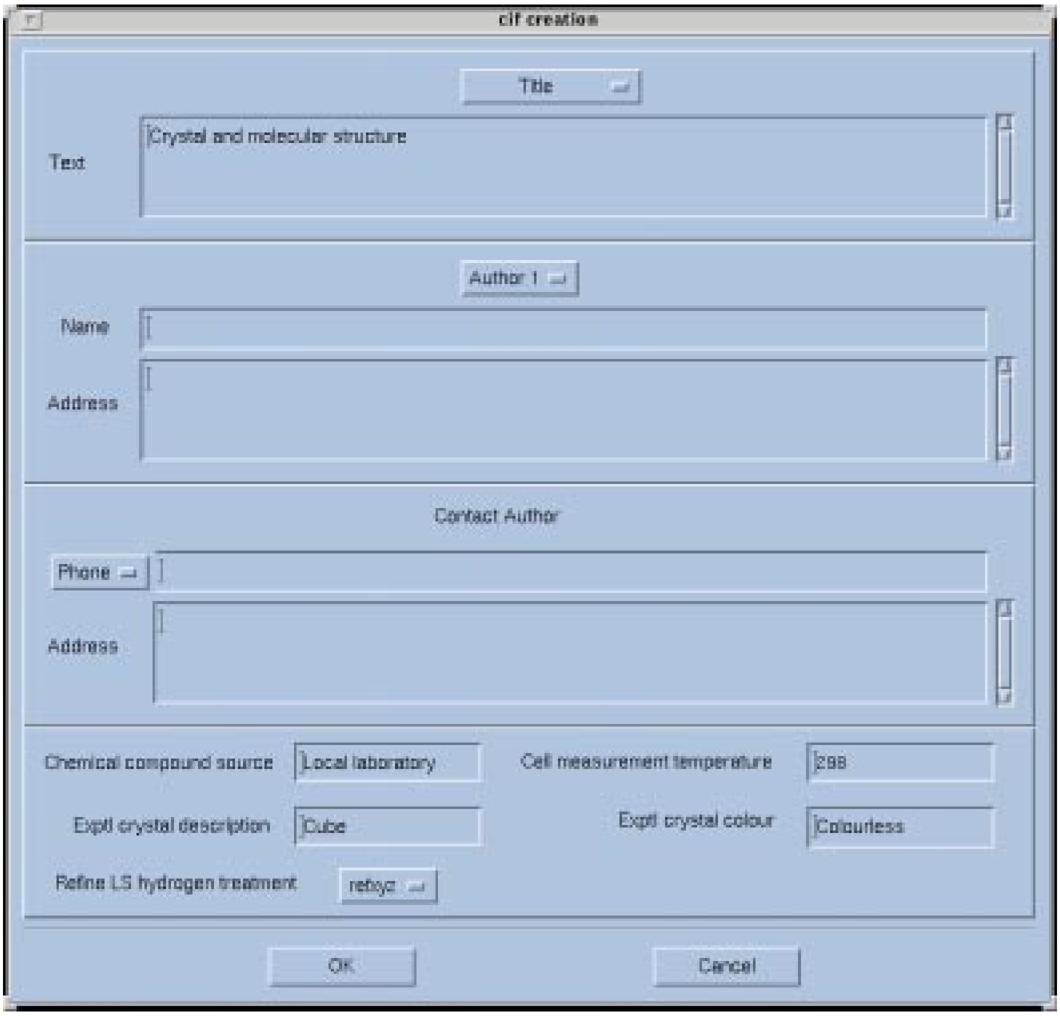
The program then checks to see if a Fourier map has been has been calculated; if not, you will be prompted to allow the program to generate one:



Difference Fourier results not present - Run FFT?

All this information is collated into a single CIF file called *full.cif*. This generally contains some duplicate entries, which are removed by the automatic checking process. As a result of this, when the file is being checked, some error messages about duplicates are generally produced. This is normal and can be safely ignored.

At this point, a window will appear:



Text: Crystal and molecular structure

Chemical compound source: Local laboratory

Cell measurement temperature: 298

Exptl crystal description: Cube

Exptl crystal colour: Colourless

Refine LS hydrogen treatment: refxyz

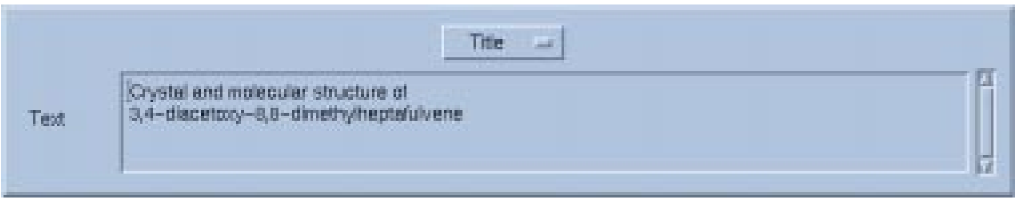
The user must enter parameters not available from the output files of the various maXus modules. The features of the window are as follows:

- **Title** Pop-Up Menu

  There are six options pulled down from the Title box:

  Title, *Abstract*, *Comment*, *Synopsis*, *Exptl Prep*, *Expt Refinement*, *Figure Captions*.

  For the first three there is a brief default in a Text box below. Edit and scroll these boxes as you wish.

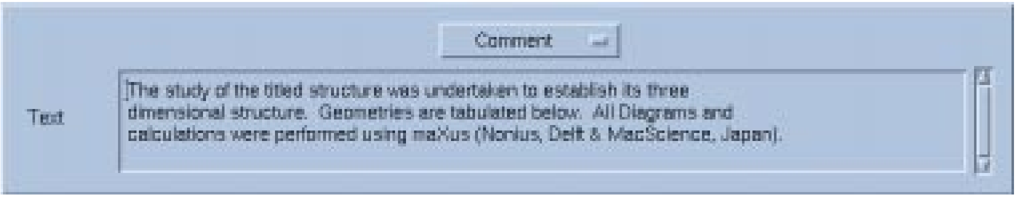


Crystal and molecular structure of
3,4-diacetoxy-8,8-dimethylheptafulvene

or:

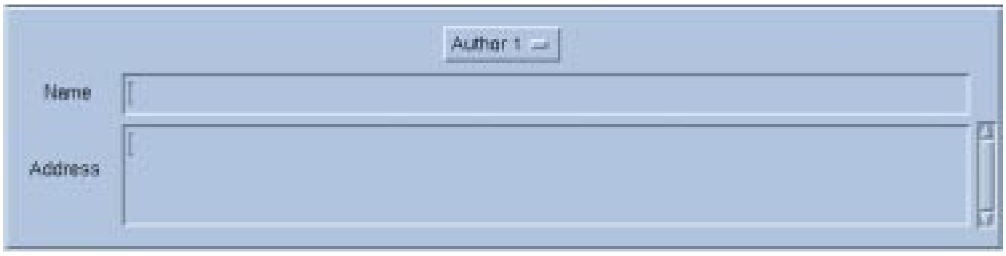
We present the crystal and molecular structure of
3,4-diacetoxy-8,8-dimethylheptafulvene.

or:



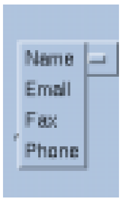
The study of the titled structure was undertaken to establish its three
dimensional structure. Geometries are tabulated below. All Diagrams and
calculations were performed using maXus (Nonius, Delft & MacScience, Japan).

- **Authors** (up to a maximum of 6) are entered *via* the Author menu.



- **Contact Author**

  The name and address of the Contact Author should be entered here. In addition, by using the Name pop-up menu to the left, contact email address, fax and phone numbers may be included:



Other options in the dialog box are:

- **Chemical compound source**: Where did the crystal come from? Default: Local laboratory.
- **Exptl crystal description**: What is the shape of the crystal? Default: Cube
- **Exptl crystal colour**: What colour is the crystal? Default: Colourless.

- **Refine LS hydrogen treatment**: How were the H atoms refined? (If no H atoms are present in your molecule, this option is ignored).

  The pull down menu options are:

  refxyz: The coordinates were refined. This is the default.

  refall: All the parameters were refined.

  refU: The temperature factors were refined.

  noref: No parameters were refined.
- **Cell measurement temperature**: At what temperature was the cell measured at? Default: is 298K.

The defaults here can be altered by the user; see Section 18.4.2.

Enter the requested parameters and click *OK*.

Finally, the program will check if you wish reflection information included in the CIF file:

Do you want to include reflection data in the cif

A CIF file is now generated, containing the information you have entered, and the information the collated from several CIF files produced at different stages of the structure solution process. The geometry information such as bond lengths, angles, torsion angles etc. which has been calculated is included, as will crystallographic parameters such as orientation matrix, cell parameters and density information from the last Fourier run.

The generated file has its syntax checked automatically, and additional checking sources are suggested in the text pane:

```
Cif generation complete: e-mail the cif to: checkcif@iucr.org for full checking
Or use web submission via: http://www.iucr.org/iucr-top/journals/acta/c/services/checkcif.html
```

On completion, a file *full.cif* is written to the working directory, and a message is displayed:

Cif file generation complete

The CIF file thus produced is the minimum requirement for submission to *Acta Crystallographica*, and may still need to be edited if you wish to add new information not covered by the defaults. In particular, you may need to add additional references unique to your compound.

If in doubt, use the email or web checking sources mentioned above to see if it meets the required standards for publication.

If there are problems found when the syntax generated file is checked, the display will show:

```
Faulty syntax on cif file.
Check temp.cif file
```

and a dialog box informing you that CIF generation has failed. In this case, a temporary file produced by the program, *temp.cif* will not be deleted as normal, but left so the user may check it for errors. Any error line-numbers mentioned in the output pane refer to this temporary file. Examining this file may give valuable clues as to why the CIF generation failed.

17.7.2 View/Edit

Opens a standard edit window which can be used to view and edit the generated CIF file in the usual way.

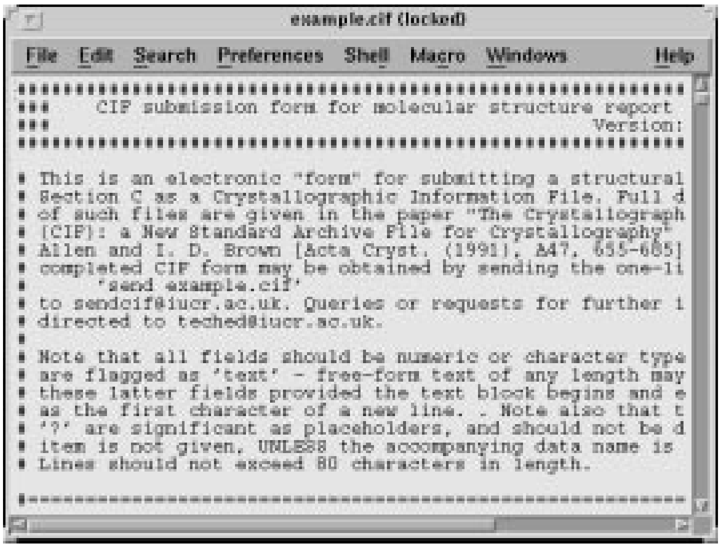
```
data_demo
_publ_section_title
;
Crystal and molecular structure of
3,4-diacetoxy-8,8-dimethylheptafulvene
;
_publ_section_abstract
;
We present the crystal and molecular structure of
3,4-diacetoxy-8,8-dimethylheptafulvene.
;
_publ_section_comment
;
The study of the titled structure was undertaken to es
dimensional structure.  Geometries are tabulated below
calculations were performed using maXus (MacScience, J
;
_chemical_compound_source        'Local laboratory'
_exptl_crystal_description       'Cube'
_exptl_crystal_colour            'Colourless'
_exptl_crystal_size_max           0.4
_exptl_crystal_size_mid           0.3
_exptl_crystal_size_min           0.1
_chemical_formula_sum            'C14 H16 O4
_chemical_formula_weight          248.30
_cell_length_a                    7.940(2)
_cell_length_b                    8.571(1)
_cell_length_c                    10.215(2)
_cell_angle_alpha                 105.03(1)
_cell_angle_beta                  95.56(2)
_cell_angle_gamma                 94.24(1)
_cell_volume                      664.7(2)
```

17.7.3 Check

Calls a program to check the syntax of your edited CIF file. Errors are output to a text pane. If there are errors that require attention, then re-edit the file with the View/Edit option.

17.7.4 View example

This brings up a text pane with an example CIF file to view.



```
##########################################################
###    CIF submission form for molecular structure report
###                                               Version:
##########################################################

# This is an electronic "form" for submitting a structural
# Section C as a Crystallographic Information File. Full d
# of such files are given in the paper "The Crystallograph
# (CIF): a New Standard Archive File for Crystallography"
# Allen and I. D. Brown [Acta Cryst. (1991), A47, 655-685]
# completed CIF form may be obtained by sending the one-li
#      'send example.cif'
# to sendcif@iucr.ac.uk. Queries or requests for further i
# directed to teched@iucr.ac.uk.
#
# Note that all fields should be numeric or character type
# are flagged as 'text' - free-form text of any length may
# these latter fields provided the text block begins and e
# as the first character of a new line. . Note also that t
# '?' are significant as placeholders, and should not be d
# item is not given, UNLESS the accompanying data name is
# Lines should not exceed 80 characters in length.

#=========================================================
```

17.7.5 Creating typeset output

It is possible to obtain formatted output from your CIF files. This can let you see the format a paper would have in in *Acta Crystallographica C*. This utilises the **printcif** utility from the IUCr. Output can be in PDF format or PostScript.

Submit your final full.cif file to the IUCr either through their webpage:

http://www.iucr.org/iucr-top/journals/acta/c/services/printcif.html

or via e-mail to:

printcif@iucr.org

17.8 CSSR Output

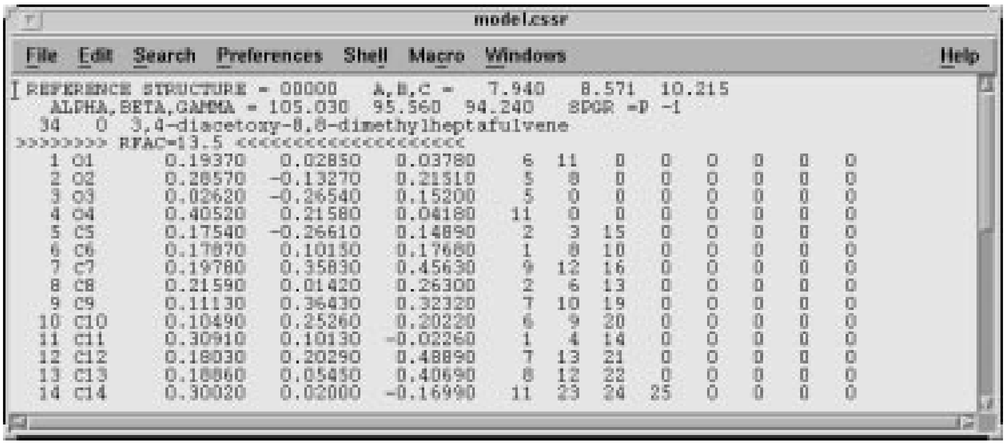
This module converts your coordinates into CSSR format so that they can be used by other viewing and modelling programs such as Cerius. A dialogue box appears:

Coordinates: ● Fractional ○ Cartesian

Filename: model.cssr

The options are as follows:

- **Coordinates**. Choose from either *Fractional* or *Cartesian* for the output file.
- **File name**: Enter the cssr file name you want here. The default is model.cssr.
- **Select**: This brings up a file browser.
- **Edit**: This brings up an standard edit window:



```
REFERENCE STRUCTURE = 00000   A,B,C =   7.940   8.571  10.215
  ALPHA,BETA,GAMMA = 105.030  95.560  94.240    SPGR =P -1
 34   0  3,4-diacetoxy-8,8-dimethylheptafulvene
>>>>>>>> RFAC=13.5 <<<<<<<<<<<<<<<<<<<<<
   1 O1       0.19370   0.02850   0.03780    6  11   0   0   0   0   0   0
   2 O2       0.28570  -0.13270   0.21510    5   8   0   0   0   0   0   0
   3 O3       0.02620  -0.26540   0.15200    5   0   0   0   0   0   0   0
   4 O4       0.40520   0.21580   0.04180   11   0   0   0   0   0   0   0
   5 C5       0.17540  -0.26610   0.14890    2   3  15   0   0   0   0   0
   6 C6       0.17870   0.10150   0.17680    1   8  10   0   0   0   0   0
   7 C7       0.19780   0.35830   0.45630    9  12  16   0   0   0   0   0
   8 C8       0.21590   0.01420   0.26300    2   6  13   0   0   0   0   0
   9 C9       0.11130   0.36430   0.32320    7  10  19   0   0   0   0   0
  10 C10      0.10490   0.25260   0.20220    6   9  20   0   0   0   0   0
  11 C11      0.30910   0.10130  -0.02260    1   4  14   0   0   0   0   0
  12 C12      0.18030   0.20290   0.48890    7  13  21   0   0   0   0   0
  13 C13      0.18860   0.05450   0.40690    8  12  22   0   0   0   0   0
  14 C14      0.30020   0.02000  -0.16990   11  23  24  25   0   0   0   0
```

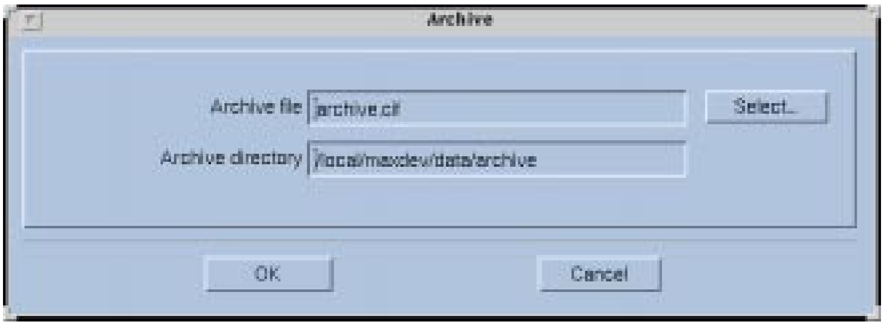
If you have not already created a .cssr file of this name then an error warning appears:

cssr file does not exist

17.9 Archive

This option generates an archive file (a CIF file with reflection data) this is stored in a pre-defined archive directory. It can then be used to input data *via* the *File/Import/Archive* option. To use this facility:

- Generate a CIF file using *CIF/Create*. **Make sure you include the reflections**.
- Select *Publish/Archive*. A dialog box appears:



Archive file archive.cif
Archive directory /local/maxdev/data/archive

Enter the Archive file name, and change the directory it is to be save in if you wish.

Clicking *Select* allows you to browse through files & directories. Click *OK*. If the run was successful then a box appears:

Archiving Complete

If there is no CIF file present, then the following dialogue box appears:

No CIF file available for archiving

Generate the CIF file using *Publish/CIF/Create* then try again.

If no intensity data are present then the following message appears:

No reflection data on CIF file

Followed by:

Archiving Failed

Re-run the CIF generator but this time make sure the *Include reflections* button is on. Then run *Archive* again.

Appendix A

A typical CIF file

A typical CIF file is as follows:

```
_publ_contact_author_name     ’Dr Ulrich Fl'orke’
_publ_contact_author_address
; Anorganische und Analytische Chemie
  Universit'at-GH Paderborn
  Warburgerstr. 100
  D-4790 Paderborn
  Germany
;
_publ_contact_author_email    floe@mvaxac.uni-paderborn.de
_publ_contact_author_fax          ’49(5251)603423’
_publ_contact_author_phone        ’49(5251)602496’

_publ_contact_letter
; Please consider this CIF submission for publication in
  Acta Cryst. C. This file is a resubmission of a revised
  version of the paper SE1031, as requested by the Coeditor.
;
_publ_requested_journal    ’Acta Crystallographica Section C’
_publ_requested_category          FM


# TITLE AND AUTHOR LIST

_publ_section_title
; The First Dinuclear Cobalt--Rhenium Cluster Compound:
  Hexacarbonylbis(\m-dicyclohexylphosphanido)cobaltrhenium
;
_publ_section_title_footnote
; Contribution No. ...
;


loop_
_publ_author_name
_publ_author_footnote
_publ_author_address
 ’Fl'orke, Ulrich’  .
; Anorganische und Analytische Chemie
  Universit'at-GH Paderborn
  Warburgerstrasse 100
  D-4790 Paderborn
  Germany
;
 ’Haupt, Hans-J'urgen’
; Current address: Department of Chemistry ...
;
; Anorganische und Analytische Chemie
  Universit'at-GH Paderborn
```

```
  Warburgerstrasse 100
  D-4790 Paderborn
  Germany
;


# TEXT

_publ_section_abstract
; The title compound [hexacarbonyl-1˛^4^C,2˛^2^C-bis(\m-
dicyclohexylphosphido-1:2˛^2^P)cobaltrhenium(Co---Re) 0.5-
methanol solvate], [CoRe(C~12~H~22~P)~2~(CO)~6~].0.5CH~3~OH,
belongs to a type of edge-linked tetrahedron--octahedron
coordination complex. The molecule has local C~2v~ symmetry
with ecliptic conformation of the carbonyl ligands at both
metal centres, and a Co---Re single-bond length of 2.786(1) ‰A.
;


_publ_section_comment
; In the course of our work on phosphanido-bridged homo- and
heteronuclear metal-atom cluster compounds (Haupt et al., 1988,
1989; Fl'orke & Haupt, 1993), we isolated the title compound
(I) in which a Co---Re bond is symmetrically bridged by two
dicyclohexylphosphanido groups.

The Co atom has distorted tetrahedral coordination from two
carbonyl ligands and the bridging P atoms. These bridging atoms
and four of the carbonyl groups give rise to distorted
octahedral coordination at the Re atom. The central CoReP~2~
ring is nearly planar; the maximum deviation from the best
plane is 0.02 ‰A with a dihedral angle of 2.2‰. With respect
to the different metal atomic radii, the ring may be regarded
as regular. It shows two equal M---P bond lengths for Co
[2.111(1) and 2.116(1) ‰A] as well as for Re [2.541(1) and
2.544(1) ‰A]. The enclosed ring angles at both P atoms are
acute [72.8(1) and 72.9(1)‰] and the P---M---P angles reflect
the distorted coordination polyhedron of each metal atom
[121.3(1) and 92.9(1)‰ for Co and Re, respectively]. The
most interesting structural feature is the Co---Re single bond
which meets the requirement of 18 valence electrons for each
metal atom and has a length of 2.786(1) ‰A. We have
established, by use of the Cambridge Structural Database
(Allen et al., 1979), that the only other cobalt--rhenium
cluster reported so far is
[Co~2~Re(\m~3~-CC~6~H~4~Me-4)(CO)~10~] (Jeffery et al., 1985),
with Co---Re bond lengths of 2.686(1) and
2.720(1) ‰A. This triangular cluster has distinctly different
bonding and bridging patterns, so direct comparison of the
heteronuclear bond lengths of both compounds is not possible.
;
_publ_section_acknowledgements                  ?

_publ_section_references
; Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A.,
Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters,
B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. &
Watson, D. G. (1979). Acta Cryst. B35, 2331--2339.

Fl'orke, U. & Haupt, H.-J. (1993). Acta Cryst. C49, 374--376.
```

```
Haupt, H.-J., Balsaa, P. & Fl'orke, U. (1988).
Inorg. Chem. 27, 280--286.

Haupt, H.-J., Heinekamp, C. & Fl'orke, U. (1989).
Inorg. Chem. 29, 2955--2963.

Jeffery, J. C., Lewis, D. B., Lewis, G. E. & Stone,
F. G. A. (1985). J. Chem. Soc.  Dalton Trans. pp. 2001--2007.

North, A. C. T., Phillips, D. C. & Mathews, F. S. (1968) Acta
Cryst. A24, 351--359.

Sheldrick, G. M. (1990). SHELXTL-Plus. Structure Determination
Software Programs. Siemens Analytical X-ray Instruments Inc.,
Madison, Wisconsin, USA.

Siemens (1990). XP. Molecular Graphics Program. Siemens
Analytical X-ray Instruments Inc., Madison, Wisconsin, USA.

Siemens (1991). XSCANS Users Manual. Siemens Analytical X-ray
Instruments Inc., Madison, Wisconsin, USA.
;

_publ_section_figure_captions
; Fig. 1. Molecular structure showing 50% probability
displacement ellipsoids. H atoms are omitted for clarity.

Fig. 2. Packing diagram viewed down the a axis. Note the
solvent molecule in the centre of the cell.
;

_publ_section_exptl_prep
; Synthesis was carried out by reaction of Re~2~(CO)~10~,
Co~2~(CO)~8~ and HP(C~6~H~11~)~2~ (molar ratio 1:1:2) in xylene
solution for 10 h at 423 K in a glass tube. Recrystallization
was from MeOH.
;

_publ_section_exptl_refinement
; The enclosed CH~3~OH solvent molecule had a site occupation
factor of 0.5. Cyclohexyl H atoms were fixed at ideal positions
with common isotropic displacement parameters (U~iso~ = 0.08
‰A^2^).
;

#############################################################

data_(I)

# CHEMICAL DATA

_chemical_name_systematic
; Hexacarbonyl-1˛^4^C,2˛^2^C-bis(\m-dicyclohexylphosphido-
  1:2˛^2^P)-cobaltrhenium(Co---Re) 0.5-methanol solvate
;
_chemical_formula_moiety  ’C30 H44 Co O6 P2 Re,0.5(C H4 O)’
_chemical_formula_sum             ’C30.5 H46 Co1 O6.5 P2 Re1’
_chemical_formula_iupac
           ’[Co Re (C12 H22 P)2 (C O)6].0.5C H3 O H’
```

```
_chemical_formula_weight           823.7

# CRYSTAL DATA
_symmetry_cell_setting             triclinic
_symmetry_space_group_name_H-M     ’P -1’

loop_
    _symmetry_equiv_pos_as_xyz
’x,y,z’
’-x,-y,-z’

_cell_length_a                      10.452(3)
_cell_length_b                      11.664(4)
_cell_length_c                      15.641(4)
_cell_angle_alpha                    94.37(2)
_cell_angle_beta                     89.75(2)
_cell_angle_gamma                   111.87(2)
_cell_volume                       1763.8(8)
_cell_formula_units_Z              2
_cell_measurement_reflns_used      40
_cell_measurement_theta_min        7
_cell_measurement_theta_max        16
_cell_measurement_temperature      293

_exptl_crystal_description         prism
_exptl_crystal_colour              red
_exptl_crystal_size_max            0.50
_exptl_crystal_size_mid            0.34
_exptl_crystal_size_min            0.28
_exptl_crystal_size_rad            ?
_exptl_crystal_density_diffrn      1.551
_exptl_crystal_density_meas        ?
_exptl_crystal_density_method     ’not measured’
_exptl_crystal_F_000               826
_exptl_absorpt_coefficient_mu      4.03
_exptl_absorpt_correction_type     psi-scan
_exptl_absorpt_process_details
           ’(North et al., 1968)’
_exptl_absorpt_correction_T_min    0.131
_exptl_absorpt_correction_T_max    0.320

# EXPERIMENTAL DATA

_diffrn_radiation_type             ’Mo K\a’
_diffrn_radiation_wavelength       0.71073
_diffrn_measurement_device_type    ’Siemens R3m/V’
_diffrn_measurement_method         \w--2\q
_diffrn_reflns_number              15189
_diffrn_reflns_av_R_equivalents    0.022
_diffrn_reflns_theta_max           27.5
_diffrn_reflns_limit_h_min         -13
_diffrn_reflns_limit_h_max         13
_diffrn_reflns_limit_k_min         -15
_diffrn_reflns_limit_k_max         15
_diffrn_reflns_limit_l_min         -21
_diffrn_reflns_limit_l_max         21
_diffrn_standards_number           4
_diffrn_standards_interval_count   400
_diffrn_standards_interval_time    ?
_diffrn_standards_decay_%          0
```

```
# REFINEMENT DATA

_refine_special_details            ?

_reflns_number_total               8161
_reflns_number_gt                  6813
_reflns_threshold_expression       F>4\s(F)

_refine_ls_structure_factor_coef   F
_refine_ls_R_factor_gt             0.038
_refine_ls_wR_factor_ref           0.034
_refine_ls_goodness_of_fit_ref     1.583
_refine_ls_number_reflns           6813
_refine_ls_number_parameters       379
_refine_ls_hydrogen_treatment      noref
_refine_ls_weighting_scheme        calc
_refine_ls_weighting_details  'w = 1/[\s^2^(F) + 0.0001F^2^]'

_refine_ls_shift/su_max            0.001
_refine_diff_density_max           0.95
_refine_diff_density_min           -0.80
_refine_ls_extinction_method       none
_refine_ls_extinction_coef         ?

loop_
 _atom_type_symbol
 _atom_type_description
 _atom_type_scat_dispersion_real
 _atom_type_scat_dispersion_imag
 _atom_type_scat_source
 'C'  'C'   0.0033   0.0016
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
 'H'  'H'   0.0000   0.0000
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
 'Co'  'Co'   0.3494   0.9721
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
 'O'  'O'   0.0106   0.0060
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
 'P'  'P'   0.1023   0.0942
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
 'Re'  'Re'  -1.0185   7.2310
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'

_computing_data_collection        'XSCANS (Siemens, 1991)'
_computing_cell_refinement        'XSCANS (Siemens, 1991)'
_computing_data_reduction         'XSCANS (Siemens, 1991)'
_computing_structure_solution     'SHELXTL-Plus (Sheldrick, 1990)'
_computing_structure_refinement   'SHELXTL-Plus (Sheldrick, 1990)'
_computing_molecular_graphics     'XP (Siemens, 1990)'
_computing_publication_material   'SHELXTL-Plus (Sheldrick, 1990)'

# ATOMIC COORDINATES AND DISPLACEMENT PARAMETERS

loop_
     _atom_site_label
     _atom_site_fract_x
     _atom_site_fract_y
     _atom_site_fract_z
     _atom_site_U_iso_or_equiv
```

```
     _atom_site_adp_type
     _atom_site_type_symbol
Re   0.2227(1)  -0.0032(1)   0.1460(1)    0.042(1)   Uani   Re
Co   0.2347(1)   0.1392(1)   0.2992(1)    0.046(1)   Uani   Co

<Cut for space.>

H46A   -0.1014     -0.0125      0.3981       0.080       Uiso   H
H46B   -0.1432     -0.1551      0.3997       0.080       Uiso   H


loop_
     _atom_site_aniso_label
     _atom_site_aniso_U_11
     _atom_site_aniso_U_22
     _atom_site_aniso_U_33
     _atom_site_aniso_U_23
     _atom_site_aniso_U_13
     _atom_site_aniso_U_12
     _atom_site_aniso_type_symbol
Re     0.045(1)  0.045(1)  0.036(1)  0.001(1)  0.002(1)  0.021(1)   Re
Co     0.050(1)  0.048(1)  0.042(1) -0.001(1)  0.006(1)  0.015(1)   Co
P1     0.044(1)  0.045(1)  0.044(1)  0.004(1)  0.002(1)  0.018(1)   P


<Cut for space.>

O100   0.123(10) 0.140(11) 0.129(11) -0.053(9) -0.028(8)  0.102(9)   O

# MOLECULAR GEOMETRY

loop_
     _geom_bond_atom_site_label_1
     _geom_bond_atom_site_label_2
     _geom_bond_site_symmetry_1
     _geom_bond_site_symmetry_2
     _geom_bond_distance
     _geom_bond_publ_flag
Re Co   .   .  2.786(1)   yes
Re P1   .   .  2.544(1)   yes
Re P2   .   .  2.541(1)   yes
Re C1   .   .  1.955(6)   no
Re C2   .   .  1.967(4)   no
Re C3   .   .  1.987(5)   no
Re C4   .   .  1.945(6)   no
Co P1   .   .  2.111(1)   yes
Co P2   .   .  2.116(1)   yes
Co C5   .   .  1.750(5)   no
Co C6   .   .  1.731(6)   no
P1 C11   .   .  1.859(4)   no
P1 C21   .   .  1.864(6)   no
P2 C31   .   .  1.840(6)   no
P2 C41   .   .  1.858(4)   no
C1 O1   .   .  1.134(7)   no
C2 O2   .   .  1.146(5)   no
C3 O3   .   .  1.136(6)   no
C4 O4   .   .  1.139(8)   no
C5 O5   .   .  1.150(7)   no
C6 O6   .   .  1.147(8)   no
C11 C12   .   .  1.518(8)   no
```

```
C11 C16   .   .  1.471(9)   no


<Cut for space.>

C43 C44   .   .  1.457(12)  no
C44 C45   .   .  1.503(12)  no
C45 C46   .   .  1.532(5)   no




loop_
     _geom_angle_atom_site_label_1
     _geom_angle_atom_site_label_2
     _geom_angle_atom_site_label_3
     _geom_angle_site_symmetry_1
     _geom_angle_site_symmetry_2
     _geom_angle_site_symmetry_3
     _geom_angle
     _geom_angle_publ_flag
Co Re P1  .  .  .   46.4(1)   yes
Co Re P2  .  .  .   46.5(1)   yes

<Cut for space.>

C42 C43 C44  .  .  .  113.1(6)   no
C43 C44 C45  .  .  .  114.6(6)   no
C44 C45 C46  .  .  .  112.3(6)   no
C41 C46 C45  .  .  .  110.9(6)   no





loop_
    _geom_hbond_atom_site_label_D
    _geom_hbond_atom_site_label_H
    _geom_hbond_atom_site_label_A
    _geom_hbond_site_symmetry_A
    _geom_hbond_distance_DH
    _geom_hbond_distance_HA
    _geom_hbond_distance_DA
    _geom_hbond_angle_DHA
    _geom_hbond_publ_flag
O6  H6  O3  2_675  0.98(3)  1.69(3)  2.664(3)  172(3)  yes
O6  H6  O4  .      0.98(3)  1.71(2)  2.661(3)  166(3)  yes
```

APPENDIX B

PLATON - 92 MANUAL

The manual for Platon ‘92 (Version 25-Jan.-92) is presented here as produced by A. L. Spek.

Sections that are not relevant to the implementation in maXus are presented in *`italicised courier font`*.

A.L.SPEK,
VAKGROEP KRISTAL- EN STRUCTUURCHEMIE,
UNIVERSITY OF UTRECHT,
PADUALAAN 8,
3584 CH UTRECHT,
THE NETHERLANDS

CONTENTS

INTRODUCTION TO PLATON-92

PLATON-92 brings together, in the context of a single program, a variety of computational and graphical procedures for the analysis of the results of a crystal structure determination. The program may be used in conjunction with structure determination and refinement programs such as SHELXS-86 and SHELX-76. The companion program PLUTON-92 may be used with the same data files for more elaborate graphics such as molecule packing plots.

PLATON-92 is designed for the automated generation of a variety of geometrical entities such as bond distances, bond angles, torsion angles, least-squares planes and ring-puckering parameters. All derived parameters are accompanied by standard deviations that are calculated by numerical methods from the supplied standard deviations in the primary input parameters.

The free format input data generally consist of two types of data that are normally but not necessarily supplied separately to the program. The crystallographic or molecular data such as coordinates, thermal parameters, cell dimensions and symmetry, possibly generated by the structure determination or refinement program, may be read from a file. Instructions are conveniently entered directly from the keyboard.

In general only global instructions are necessary to obtain tables of the required derived molecular geometry data. Specific information such as atomic radii and other properties related with the element types involved are by default drawn from internal tables. The simple instruction CALC will automatically execute virtually all calculations that are of interest for the supplied parameter set.

The set of available analysis procedures includes those for rigid-body thermal motion with bond distance correction, puckering analysis (including the Cremer & Pople variety), hydrogen bonds (including an analysis in terms of networks), voids in the structure and five-coordination (Berry pseudorotation). Appropriate references to the literature are given in the output listing.

The calculations are complemented with graphics facilities such as thermal motion ellipsoid plots, Newman projection plots, space-filling plots and projection plots of the structure on the various least-squares and ring planes.

The program is written in FORTRAN-77 and easily implemented on mainframes such as CDC-CYBER under NOS/VE, on VAX/VMS systems and on UNIX systems (COVEX, Silicon Graphics, DEC).

Note: PLATON is not a finished program. It is constantly improved with new facilities as their need arises in the course of the large variety of structure determinations that are carried out in our laboratory. In view of the extremely large number of options of the program, combined with the unique characteristics of each new crystal structure examined with the program, problems may arise in non-standard cases. The author will be interested in any user comment and suggestions for extensions for future releases.

Projected future extensions includes R/S assignments, generation of H-atom positions (in particular those involved in H-bonding)

INTRODUCTORY EXAMPLE

The following example, assumed to run on a VAX/VMS system (for other systems see APPENDIX II), of the structure of SUCROSE should provide an introduction to the use of this program and its potential. The structural parameters are assumed to reside on a disk file named SUCROSE.SPF for which the free format contents are listed in part below and included in full on the release tape:

```
TITL SUCROSE (ACTA CRYST. (1973),B29,790-797)
CELL 1.5418 10.8633 8.7050 7.7585 90 102.945 90
CESD 0.0005 0.0004 0.0004 0 0.006 0
SPGR P21
ATOM C1 0.29961 0.35792 0.48487 0.00008 0.00000 0.00012
BIJ C1 0.00274 .00376 .00584 .00004 .00094 .00006
SBIJ C1 .00006 .00009 .00012 .00009 .00007 .00006
ATOM C2 0.31253 0.47474 0.63600 0.00009 0.00015 0.00012
BIJ C2 .00304 .00498 .00641 -.00063 .00073 -.00043
SBIJ C2 .00006 .00010 .00013 .00009 .00007 .00007
ATOM C3 0.28545 0.63673 0.56447 0.00009 0.00015 0.00013
BIJ C3 .00321 .00437 .00965 -.00071 .00196 -.00017
SBIJ C3 .00007 .00010 .00015 .00010 .00008 .00007
ATOM C4 0.37404 0.67095 0.44198 0.00010 0.00015 0.00014
BIJ C4 .00400 .00403 .01003 .00006 .00243 -.00021
```

```
SBIJ C4 .00007 .00010 .00015 .00011 .00009 .00007
ATOM C5 0.35925 0.55107 0.29529 0.00009 0.00016 0.00013
BIJ C5 .00314 .00525 .00756 .00064 .00133 -.00022
SBIJ C5 .00007 .00011 .00014 .00010 .00008 .00007
```

etc. etc.

The program PLATON may be invoked for this data set with the command PLATON SUCROSE. As a result the data set SUCROSE is loaded and, since this file does not contain an END instruction at the end of the file, the program comes, after that the end-of- file has been reached, with the prompt >> to receive more data and/or instructions. A calculation of the intra-molecular geometry may now be invoked with the instruction CALC INTRA. The results are written to a disk file (in this case SUCROSE.LIS). An analysis of short inter-molecular contacts is performed with a subsequent CALC INTER instruction. The analysis may be completed with a CALC COORDN instruction that gives a listing of all bonds and angles about all atoms (excluding C and H) involving atoms within a 3.2 Angstrom coordination sphere. A thermal motion ellipsoid plot is obtained with PLOT TME. The plot may be rotated over 45 degrees about the vertical Y-axis with the instruction VIEW YROT 45. A labelled stick plot in the same orientation is obtained with PLOT STICK. The session may be closed with the instruction END.

ON HOW IT WORKS

This section on the program internals should provide a framework to understand the effects of the various available instructions. The input atomic coordinates (X, Y, Z) are with reference to user-defined axes (a, b, c), which will usually be either crystallographic unit cell axes or an arbitrary orthogonal set; these coordinates are input as fractions of the unit cell edges or as Angstrom units (in the latter case they are converted and stored as fractions of dummy cell edges). A second, orthogonal system (A, B, C) with coordinates (XO, YO, ZO) in Angstroms is set up internally (see J.D. Dunitz, X-Ray analysis and structure of Organic molecules, p236): A is a unit vector along a, B is a unit vector normal to a in the ab-plane, and C is normal to A and B. B will coincide with b in monoclinic cells in the b-setting. If the input axes are orthogonal, the two sets of axes a,b,c and A,B,C are coincident. The third system is the plotting coordinate system in cm: XP across the picture from left to right, YP up the picture from bottom to top and ZP out of the paper. All these axial sets are right-handed and absolute configuration is preserved in all rotations. As atoms are input to the program, they are stored in the X,Y,Z and XO,YO,ZO axes systems. Each atom also has additional information stored for it such as estimated standard deviations, thermal motion parameters, a name (the embedded element name is used by default to set various radii to be used during the subsequent calculations) and various bit flags such as the inclusion bit. Coordinate data are checked for duplications on input. Atom labels not conforming to the required format are renamed. A CALC instruction generally initiates a distance search on the basis of the previously set or supplied radii. In the INTRA mode this results in the setup of an array that stores per atom all connections that are found. This list is used subsequently by a geometry listing routine that generates all unique bond distances, bond angles and torsion angles. Simultaneously with the setup of the connectivity array all atoms are transformed (when necessary) to obtain a connected set. In addition, in the case that the molecule lies on a special position, the primary coordinate list is expanded with additional symmetry generated atoms in order to handle the geometry of the complete molecule.

See also Appendix I: Terms & Notions.

INSTRUCTION SUMMARY

This chapter provides a description of the available instructions. They are grouped together as compound specific, calculation, plot, list and general instructions. The logical order of calculations is intra-molecular, inter- molecular and coordination geometry. In the description of individual instructions below the following applies: - (Sub)keywords are in uppercase and user data in lower case - Data in parentheses are optional. - Choices are separated by a slash.

Note: parentheses in atom names (on input) are ignored except for that Ag denotes the atom type and Ag() the individual atom.

Lower case input is automatically converted to upper case.

Lines with a blank character in position 1 are ignored.

Input lines may be continued with data on the next line by placing the symbol = at the end of the line.

INSTRUCTIONS PRECEDING ALL CALCULATIONS

These instructions will be necessary only in special situations.

ROUND (ON/OFF)

This option defines whether primary input data and derived geometrical parameter values will be rounded based on their standard deviations or not. The default setting is ON.

Example: ROUND OFF PARENTHESES (ON/OFF)

By default, the numerical part of an atomic label will be enclosed within parentheses.
Example: PARENTHESES OFF

INCLUDE El1 El2 ...

Only the elements specified in the include list will be included in the calculations.

Example: INCLUDE C N O

EXCLUDE El1 El2 ...

The elements in the exclude list will be excluded from all calculations.

Example: EXCLUDE H

DOAC El1 El2 ....

The elements N, O, Cl, S, F and Br are treated as potential donor/acceptor atoms for hydrogen bonding by the program. This list will be replaced by the one specified in the instruction. Example: DOAC N O

HBOND p1 p2 p3

Default criteria for hydrogen bonds are: distance between donor and acceptor atom less than the sum of their van der Waals radii + p1 (= 0.5 angstrom); distance H to acceptor atom less than sum of corresponding van der Waals radii + p2 (= -0.12 angstrom) and angle D-H...A greater than p3 (= 100 degree). The default values may be changed with the HBOND instruction.

LSPL atom_name1 atom_name2 ..

This instruction specifies the set of atoms for which a least-squares plane should be calculated. In this way it is possible to include special planes in the following calculations that include the generation of least-squares planes for planar parts in the structure.

RING atom_name1 atom_name2 ...

Rings in the structure up to 8 membered are found automatically. This instruction provides a facility to include larger rings (up to 30 membered) in the calculations. The atoms should be specified in bonded order.

FIT atom_name1 atom_name2 ....

(Implementation in progress)

CALCULATIONS

The full range of molecular geometry calculations will be carried out automatically with a the single keyword instruction CALC. This includes all the calculations that may be executed alternatively with the instruction sequence CALC INTRA, CALC INTER, CALC COORDN and CALC METAL.

CALC INTRA ((El1 r1 El2 r2 ..)/(TOLA p1)) (NOBOND) (NOANG) (NOTOR) (NOLSPL) (NORING) (NOTMA) (NOBPA) (NOSTD) (WLSPL) (NOPESD) (NOMOVE) (NOSYMM) (VERIFY) (TOLP t2)

The default instruction CALC INTRA produces a full calculation and listing of all relevant intra-molecular geometrical parameter options using default covalent radii drawn from internal tables. Atoms with distances less than the sum of their covalent radii plus a tolerance (TOLA = 0.4 Angstrom) are considered to be bonded. The default radii values may be modified with their explicit specification (in which case TOLA is set to zero, unless specified explicitly). Alternatively the parameter TOLA may be modified. In the automatic radii mode an additional 0.6 Angstrom is added to the tolerance to catch (Earth)alkali to non-metal contacts. The calculation and listing of bonds, bond angles, torsion angles, least-squares planes, rings, angles between bonds and least-squares planes and thermal motion analysis may be suppressed with the specification of the sub-keywords NOBOND, NOANG, NOTOR, NOLSPL, NORING, NOBPA and/or NOTMA. The calculation of standard deviations may be suppressed with NOSTD. The NOMOVE sub-keyword has the effect that atoms are left at their input positions in the course of the generation of a connected set. The sub-keyword VERIFY may be used to control the inclusion of a connection in the bond list for plotting. TOLP is an out-of-plane deviation parameter (by default 0.1 Angstrom) that determines the inclusion of an atom in the process of automatic least-squares plane search. NOPESD , when specified, has the effect that the e.s.d. of the plane parameters is not included in the calculation of the e.s.d. in out-of-plane deviations. WLSPL invokes mass-weighted least-squares plane calculations as opposed to unit weighted. NOSYMM limits the search for connections within the input coordinate set without the application of translation or rotation symmetry.

Example: CALC INTRA NOLSPL NORING

CALC GEOM (SHELX/OMEGA/MOGLI/EUCLID) (VERIFY) (NOMOVE) (EXPAND)

This instruction executes a short intra calculation, mainly producing a list of bond distances, bond angles and torsion angles, as an alternative for the exhaustive CALC INTRA calculations. The sub-keyword SHELX may be used to generate an ordered coordinate file suitable for SHELX; OMEGA generates a file suitable for the tabulation of primary and derived parameters; MOGLI results in a DGE-file suitable for the program MOGLI and EUCLID gives a new SPF style file. The NOMOVE sub-keyword has the effect that atoms are left at their input positions in the course of the generation of a connected set. The EXPAND option may be useful for the generation of a file with the complete molecule as opposed to just the unique part.

Example: CALC GEOM EUCLID EXPAND

CALC TMA

This invokes the execution of a rigid-body thermal motion analysis and the calculation of derived quantities. It is automatically included in a CALC INTRA calculation. Note: No TMA analysis is done when the residue contains too few atoms or when the R-index of the observed and calculated Uij’s is too high.

CALC INTER (El1 p1 El2 p2 ..)/(TOLR p1)

Short inter-molecular contacts are listed with this instruction. By default van der Waals radii drawn from internal tables are used in conjunction with a default tolerance (TOLR = 0.2 Angstrom). Hydrogen bonds are automatically found and analysed.

CALC HBOND (p1 p2 p3)

This instruction provides a subset of the information generated with the CALC INTER instruction and may be of use when interest is concentrated on H-bonds.

CALC COORDN (p1/El1 r1 El2 r2 .. (NOANG) FIVE (TBA))

This instruction provides for the analysis of coordination spheres. Bond distances and bond angles are calculated for atoms within the specified sphere. By default such a calculation is done for all atoms (excluding C and H) and with radius 3.2 Angstrom. This default may be changed with the specification of the desired value. Alternatively a list of selected elements and their corresponding coordination radii may be specified for the coordination geometry calculations. Bond angles may be excluded from the listings with the NOANG sub-keyword. A Berry pseudo rotation analysis is carried out automatically when an atom is found to be bonded to exactly 5 atoms. Such a calculation may be enforced for the five shortest contacts with the sub-keyword FIVE optionally followed with the value for the trans-basal-angle (default 150 degree).

CALC COORDN atom_name p1

The coordination geometry about a single atom may be examined with this instruction. Example:CALC COORDN O3 3.2

CALC METAL (p1)

Distances between metal atoms less than p1 (default 10 Angstrom) are calculated. This option is included in the default CALC calculations.

CALC VOID (LIST) (GRID p1) (TOLV p2)

This option may be used to check the structure for voids as possible sites for solvents. The GRID (default value 0.4 Angstrom) and the minimum VOID radius (1.2 + p2 Angstrom) may be changed (default p2 = 0.0). The LIST option gives a map on the line printer. Positions with a shortest contact distance to the van der Waals surface of at least 1.2 + p2 Angstrom are indicated with >. Solvent accessible areas are indicated with a dot. Blank areas indicate small voids, all other gridpoints are within the molecular van der Waals volume. Note: This option may also be used to study cases where the unit cell contents are misplaced with respect to the symmetry elements, since this fault will generally result in both areas with short molecular contacts and areas with voids. The VOID option is more compute intensive than the rest of the instructions. It is advised to run this option in BATCH mode.

CALC DIST (eltype p1)

A distance scan is done for all vectors between the specified element and within the specified radius. By default a scan is done for H-atoms.

Example: CALC DIST I 4.0

DIST Atom_name1 Atom_name2

With this option a distance between two specified and not necessarily bonded atoms may be calculated between atoms in the atom_array.

ANGL Atom_name1 Atom_name2 Atom_name3

The angle between the specified and not necessarily bonded atoms is calculated.

TORS Atom_name1 Atom_name2 Atom_name3 Atom_name4

The dihedral angle involving the four specified atoms (not necessarily bonded) is calculated.

LSPL Atom_name1 Atom_name2 Atom_name3 Atom_name4 ...

The least-squares plane determined by the specified atoms is calculated.

PLOT

The program provides graphics options to support the geometry analysis.

PLOT (LSPL/PLAN/RING/RESD) (ALONG/PERP) (HPD/HPP/CAL)

Plots of the structure viewed perpendicular to or along the various least-squares planes may be produced for inspection. PLOT NEWMAN (DISPLAY/PLOTTER/META) (at1 at2)

Newman plots are produced, provided that a CALC INTRA instruction was carried out previously in order to prepare a file with the relevant data for all Newman projections. The Newman plots may be examined sequentially or for an individual one to be selected by specifying the relevant central bond.

PLOT TME (nr) (COLOR) (VERIFY) (DISPLAY/PLOTTER/META)

A thermal motion ellipsoid plot is produced for residue number nr. The COLOR option provides for the distinction of atom types in the plot (oxygen RED, Nitrogen BLUE and halogens GREEN). The three plot angles xr, yr and zr to reconstruct the present orientation are plotted in the lower right corner, upper left corner and lower left corner respectively. The probability level of the ellipsoid surfaces is shown in the upper right corner. When no VIEW instruction was given previously, the program will calculate a minimum overlap view. Example: PLOT TME 3 COLOR

PLOT (STICK/SOLID/CPK) (MONO/STEREO) (HATCH/CONTOUR) (nr) (SHN/SHH/SHF/SHC) (VERIFY) (COLOR) (DISPLAY/PLOTTER/META)

Plots of the molecule in various styles are produced in a orientation that may be modified with VIEW instructions. The STICK style provides a fast plot with labelled atoms. The SOLID sub-keyword results in a ball-and-stick plot and the CPK sub- keyword in a spacefilling plot. The COLOR option colours oxygen atoms RED Nitrogen atoms BLUE and Halogen atoms GREEN. nr specifies the number of the residue to be plotted. The three plot angles xr, yr and zr to reconstruct the present orientation are plotted in the lower right, upper left and lower left corners respectively.

BOX (ON/OFF[ON]) (RATIO ratio[1.333])

By default the drawing will be surrounded with a rectangular box outline. This setting may be changed with the ON and OFF sub- keywords. The three numbers shown in the bottom right, top left and bottom left corner of the box are the rotation angles xr, yr and zr respectively. These numbers may be used to reconstruct this particular orientation directly from the default UNIT orientation *via* a VIEW XR xr YR yr ZR zr instruction. The default horizontal to vertical size ratio of the box for a TME plot is 4/3. A ratio of 1 produces a square box. Example: BOX ON RATIO 1.0 VIEW (XR xr) (YR yr) (ZR zr) ...

The current orientation of the molecule for plotting may be modified with a VIEW instruction: VIEW XR 45 YR -55 will rotate the molecule first clockwise about the horizontal X-axis, followed by an anti-clockwise rotation by 55 degrees about the vertical Y-axis. VIEW instructions are accumulative. The single keyword instruction VIEW will bring the molecule back in the default orientation. SET PROB (30/40/50)

The probability level for the ellipsoid surfaces is set by default to 50%. Example: SET PROB 30

GENERAL INSTRUCTIONS

HELP (SPGR)

This instruction provides an on-line HELP facility. The SPGR option lists all space groups known to PLATON.

LIST BONDS/ATOMS/SYMM/CELL/RADII

This provides for on-line inspection of BOND and ATOM tables, the current symmetry, CELL dimensions and default radii.

LIST IPR/PAR (ival1 (ival2))

Internal parameter values (see Appendix VII) may be inspected with this instruction. A range will be listed when two values are specified and the full range when none is given.

Example: LIST PAR 3 5

SET PAR p1 p2

This instruction is not meant for general use. It provides a facility to modify internal parameter values, in particular those with no equivalent (sub)keyword. p1 is the parameter number and p2 the new value.

SET IPR p1 p2

This instruction is not meant for general use. It provides a facility to modify internal parameter values, in particular those with no equivalent (sub)keyword. p1 is the parameter number and p2 the new value.

SAVE

This instruction causes the saving of subsequent instructions on a file to be executed on all data sets, separated by ENDS cards, on the parameter file.

END

This results in a normal end of program when the .SPF file contains only one data set, otherwise the program restarts for the next data set on the file.

QUIT

This results in an immediate stop of the program, ignoring possible further datasets on the input file.

STOP

This results in an immediate stop of the program, ignoring possible further datasets on the input file.

APPENDIX - I: TERMS AND NOTIONS

The n.ijk symmetry operation on input

Atomic coordinates as input will be transformed in general by symmetry operations following certain rules. In the default automatic mode this will result in a connected set with residues properly positioned within the unit cell range. The symmetry operation applied to the input data will be listed under the header trans in the atomic coordinates listing and is encoded as n.ijk. n stands for the number of the symmetry operation as specified on the first page of the output listing and ijk for the unit cell translations in the three directions relative to 555: ijk=564 means 1 positive translation in the b direction, 1 negative translation in the c direction and none in the a direction.

The automatic mode transformation may be overruled for a given atom by preceding the data for that particular atom by a TRNS instruction e.g. TRNS 3.564. This facility may be used to determine the part of the molecule that is to be considered as the asymmetric part of a symmetrical molecule.

The transformation to be applied only to the first atom as a starting point of a new residue can be forced with a negative symmetry transformation code e.g. TRNS -5.354. Its position in the input stream determines the atoms to which it will apply. The input stream may contain several of such instructions, each apply to the atoms that follow until overruled by a new one. Their effect will only be on atoms that are chosen to start a new residue.

Disorder

The program attempts to manage the problems that are encountered with several types of disorder. Only two-fold disorder is allowed. Populations higher than 0.5 are understood as major disorder components and those less than 0.5 as minor disorder components. The usual transformations on input coordinates are restricted. In general it will be necessary to supply disordered molecules as connected sets. The calculation of distances and angles etc. will extend only to entities involving the major disorder component or the minor disorder component but not both.

Molecules and Residues

The concepts of molecules and residues are related but not always synonymous within the PLATON realm. A residue is defined as a part of the structure that is connected by intra-molecular bonds only and is associated with a number. A structure may thus contain one or more residues. Residues may be chemically equivalent or chemically distinct. A molecule is defined as an asymmetric part of the structure connected by intra-molecular bonds only. Several molecules may join by crystallographic symmetry into one residue. A particular molecule is designated by a code: [nijk.rr] where n denotes the symmetry operation with respect to the basic molecule, ijk the translation with respect to 555 and rr the residue number. The structure of sucrose thus consists of two molecules (e.g. [1555.01] and [2545.01]) but only one residue.

APPENDIX - II: IMPLEMENTATIONS

PLATON consists of a main program, over 123 dedicated subroutines, a BLOCK DATA statement, a general SPACE GROUP SYMMETRY handler and a graphics driver package (that differ for PC and non-PC versions). It is written in FORTRAN-77 and contains only a few system dependent features that are isolated in routines that are marked as such.

VAX/VMS - Implementation: (2047 atoms)

*The program source may be compiled with the DCL command: FORTRAN PLATON and linked with the DCL command: LINK PLATON in much the same way as SHELXS-86. A symbol like PLA*TON == “$ $DISK1:[USER.PLATON]PLATON” should be defined.*

The program may then be invoked with PLATON COMPOUND where the file COMPOUND.SPF contains pertinent structural parameters for COMPOUND. All files associated with COMPOUND will have this name with varying extensions. The default filename is ATOM.SPF.

VAX/VMS-workstation Implementation: Similar to VAX/VMS with workstation graphics facilities included.

CDC-CYBER NOS/VE - Implementation: (250 atoms) The program source may be compiled with:

```
FTN I=PLATON_FOR BO=PLATON_LGO FS=ON
```

and executed with:

EXET PLATON_LGO

The parameter file is expected on the file ATOM_SPF.

PC-MS/DOS - Implementation: (250 atoms)

In this case the program is supplied as an executable. It was compiled with the Ryan-McFarland version 2.43 compiler under MS-DOS 3.3 and depends on some assembler routines for graphics support. The program is invoked with the command: PLATON COMPOUND. The program depends on the availability of a co-processor: it will terminate with error message 4001 otherwise. The program is dimensioned to fit in 512K MS-DOS machines. The number of files and buffers in CONFIG.SYS should be raised to 20. The program requests for the available graphics type and stores it in a diskfile named CONFIG.PLT for future automatic reference. The void option is not available in this version.

note: A COMPOUND-name should not exceed 7 characters.

APPENDIX - III: PARAMETER FILE

The atomic parameters (including unit cell parameters, coordinates and temperature parameters) for a given structure may be inputted in various ways:

a - The STANDARD PARAMETER FILE Format (SPF).

b - SHELX-76 and SHELXS-86 type Format (possibly edited)

c - Simple Coordinate Format

The SPF-format is card image oriented. The first four characters on a card specify the nature of the data that follow on that card. Data that are not needed for the current program are simply skipped. All data are free format.

CARD TYPES

TITL text

This text may be used for various titleing purposes. It may be overridden at any time by another TITL instruction.

CELL (wavelength) a b c alpha beta gamma

Optional wavelength and cell parameters in Angstroms and degrees respectively. No CELL card is needed for Angstrom data input. The wavelength is used for the calculation of the linear absorption coefficient.

CESD sig(a) sig(b) sig(c) sig(alpha) sig(beta) sig(gamma)

This optional card specifies standard deviations in the cell parameters. No CESD card is needed for Angstrom data. The cell e.s.d. is combined with the coordinate e.s.d. for the calculation of the e.s.d. in derived parameters.

SPGR space-group-name

Space group symbol. See Appendix-IV for more details.

LATT (P/A/B/C/I/F) (A/C)

First parameter specifies the Bravais lattice type and the second whether the lattice is acentric or centric.

SYMM symmetry-operation

Symmetry operation. See appendix - IV.

ATOM atom_name x y z (pop) (sig(x) sig(y) sig(z)) (spop)

This specifies the positional parameters, the population and their estimated standard deviations. The atom_name should conform some rules in order to be acceptable since it is interpreted. The first one or two characters should correspond to an element name known to the program (see Appendix V). The number of characters of the element type and the attached digital number cannot exceed four. ‘and “ are allowed as part of an atom name. Labels not conforming with the PLATON-rules are modified in a new label including the symbol #. The atom-name may contain parentheses enclosing the numerical part.

UIJ atom_name U11 U22 U33 U23 U13 U12

Anisotropic thermal parameters. Note the order of the components that is the same as in SHELX but often different in other systems (such as the XRAY and XTAL systems). TF = exp[-2*pi**2(U11*H**2(A*)**2+...+2*U12*H*K*(A*)(B*)+...)]

SUIJ atom_name sig(U11) sig(U22) sig(U33) sig(U23).. sig(U12)

Estimated standard deviations for the anisotropic thermal parameters.

U atom_name U sig(U)

Isotropic temperature factor along with its associate standard deviation.

BIJ atom_name Beta11 Beta22 Beta33 Beta23 Beta13 Beta12

Anisotropic thermal parameters. Note the order of the components.

TF = exp[-(Beta11*H**2+Beta22*K**2+...+2*Beta12*H*K+...)]

Definition: Beta11 = 2*pi**2*astar**2; Beta12 = 2*pi**2*astar*bstar.

SBIJ atom_name sig(Beta11) .. sig(Beta23) .. sig(Beta12)

Estimated standard deviations for the anisotropic thermal parameters.

B atom_name B sig(B)

Isotropic temperature factor along with its associate standard deviation.

Definition: B = 8*pi**2*U

TRNS -n.klm

Facility to influence the applied symmetry operation for the first atom in a new residue. (see appendix I)

TRNS n.klm

When placed in front of an ATOM card this instruction will transform the input coordinates on that card by the named symmetry operation: n is the number of the symmetry operation and k,l,m are the translations. (see appendix I)

TRNS T11 T12 T13 T21 T22 T23 T31 T32 T33 (SH1 SH2 SH3)

Transformation matrix on cell axis and origin shift to be applied to the data following (CELL parameters, atomic coordinates and thermal parameters).

Example:

TITL NI-COMPOUND

CELL NI .123 .544 -.176 1 .001 .002 .001 0.0

UIJ NI .011 .013 .025 -.011 .004 .009

SUIJ NI .001 .001 .002 .002 .002 .001

ATOM C1 .345 .675 -.334 1 .010 .009 .005 0.0

U C1 0.04 0.01

(etc)

THE SHELX-76 FORMAT

Most SHELX-76 type files (possibly edited with TITL, CELL, LATT, SYMM information will be acceptable as well. In any case the atomic parameters should be preceded with an FVAR card since this triggers the program to expect this type of input. An END card on a SHELX file will be ignored. Possibly some file editing may be necessary. note: A LATT card should always be included to overrule the PLATON default acentric primitive lattice type.

An example input file named NAME.SPF, that has been edited from a NAME.RES file, is given below:

```
TITL C14H15N9S2ZN
CELL 0.71073 10.202 14.730 90 112.11 90
LATT 1
SYMM -X,0.5+Y,0.5-Z
ZN 5 0.18575 0.13934 0.25302 11.00000 0.06160 0.04669 =
0.04998 0.00183 0.02235 0.00576 =
0001 .00006 .00008 .0007 .0005 .0006 .0006 .0005 .0007
S42 4 0.55652 0.32960 0.46431 11.00000 0.07026 0.06705 =
0.07565 -0.00393 0.02317 -0.00608
S52 4 -0.21825 0.21554 0.30836 11.00000 0.08508 0.07061 =
0.09007 0.01831 0.04312 0.00509
(etc.)
END
```

note: e.s.d.s on refined parameters may be obtained by putting one of the SHELX parameters negative: LIST -n.

ANGSTROM DATA FORMAT

Files with just positional parameters, not preceded by CELL and symmetry cards are understood to be angstrom data. Coordinate data may be preceded by an ANGSTROM card with optionally a multiplication factor to transform the data to angstrom units. ATOM cards may be as: C1 1.123 1.456 1.789.

APPENDIX - IV: SPACE GROUP SYMMETRY

Space group symmetry is handled in PLATON with a general space group symmetry management routine that permits the specification of the symmetry either explicitly in terms of the general equivalent positions as presented in the International Tables or implicitly in terms of space group generators. The generators for all space groups in their standard setting and some commonly used non-standard settings are also implicitly retrievable by the program from internal tables (see tables below) on the basis of the specified name of the space group (e.g. R-3m)

EXAMPLE: The symmetry for space group nr. 19 (P212121) may be specified either as: LATT P A

SYMM X,Y,Z

SYMM 1/2 + X, 1/2 - Y, -Z

SYMM -X, 1/2 + Y, 1/2 - Z

SYMM 1/2 - X, - Y, 1/2 + Z or

LATT P A

SYMM 1/2 + X, 1/2 - Y, -Z

SYMM -X, 1/2 + Y, 1/2 - Z or

SPGR P212121

A LATT card should precede any SYMM card in order that the symmetry arrays are initialized to either, by default, a primitive non-centrosymmetric lattice or to the specified lattice type: (P/A/B/C/I/F) and (A)Centric type (A/C).

The general equivalent positions should be given as specified in International Tables and should have the centre of symmetry at the origin, in the case that the space group is centrosymmetric. The symmetry operation SYMM X,Y,Z is always implicitly assumed as the first symmetry operation and needs not be given although any redundancy in the symmetry input will be ignored.

Note: Rhombohedral lattice types (in hexagonal setting) should be specified explicitly using an extra symmetry generator. Thus the generators for space group R3 are:

LATT P A

SYMM -Y, X-Y, Z

SYMM 1/3+X, 2/3+Y, 2/3+Z

The same space group on rhombohedral axes should be specified as R3R.

The translation part may be specified either as a ratio or as a real (e.g. 1/4 or 0.25). Monoclinic-b is taken as the standard setting for monoclinic space groups. Other settings are to be specified by the full space group name: e.g. P112 for the monoclinic-c setting of P2.

Non-standard orthorhombic settings such as space group A2aa may be handled by specifying Ccc2 -cba on the SPGR card (see International Tables Vol A). In fact the program automatically modifies the input line accordingly for non-standard settings (see table below). The standard setting symmetry is than transformed accordingly.

Note: Symmetry may also be presented in the SHELX-76 style. However a LATT card should always be supplied since the default symmetry of PLATON is always P1 whereas SHELX defaults to P-1.

The names of the space groups known to the program are given in the following table and are in accordance with the usage in the CAMBRIDGE STRUCTURAL DATABASE files.

SPACE GROUP NAMES KNOWN TO THE PROGRAM

TABLE 10.

P1	P-1	P2	P112	P21	P1121	C2	A2
B112	PM	P11M	PC	PA	PN	P11B	CM
AM	B11M	CC	IC	IA	AA	B11B	P2/M
P112/M	P21/M	P1121/M	C2/M	A2/M	B112/M	P2/C	P2/A
P112/B	P21/C	P21/A	P2/N	P1121/B	P21/N	P21/N11	C2/C
A2/A	C2/N	I2/C	I2/N	I2/M	I2/A	B112/B	P222
P2221	P2212	P2122	P21212	P212121	C2221	B2212	C222
F222	I222	I212121	PMM2	PMC21	PCC2	PMA2	PCA21
PNC2	PMN21	PBA2	PNA21	P21NB	PC21N	PN21A	PBN21
P21CN	PNN2	CMM2	CMC21	CCC2	AMM2	ABM2	AMA2
ABA2	FMM2	FDD2	IMM2	IBA2	IMA2	PMMM	PNNN
PCCM	PBAN	PMMA	PNNA	PMNA	PCCA	PBAM	PCCN
PBCM	PNNM	PMMN	PBCN	PBCA	PCAB	PNMA	PBNM
PMCN	PNAM	PMNB	PCMN	CMCM	CMCA	CMMM	CCCM
CMMA	CCCA	FMMM	FDDD	IMMM	IBAM	IBCA	IMMA
P4	P41	P42	P43	I4	I41	P-4	I-4
P4/M	P42/M	P4/N	P42/N	I4/M	I41/A	P422	P4212
P4122	P41212	P4222	P42212	P4322	P43212	I422	I4122
P4MM	P4BM	P42CM	P42NM	P4CC	P4NC	P42MC	P42BC
I4MM	I4CM	I41MD	I41CD	P-42M	P-42C	P-421M	P-421C
P-4M2	P-4C2	P-4B2	P-4N2	I-4M2	I-4C2	I-42M	I-42D
P4/MMM	P4/MCC	P4/NBM	P4/NNC	P4/MBM	P4/MNC	P4/NMM	P4/NCC
P42/MMC	P42/MCM	P42/NBC	P42/NNM	P42/MBC	P42/MNM	P42/NMC	P42/NCM
I4/MMM	I4/MCM	I41/AMD	I41/ACD	P3	P31	P32	R3
R3R	P-3	R-3	R-3R	P312	P321	P3112	P3121
P3212	P3221	R32	R32R	P3M1	P31M	P3C1	P31C
R3M	R3MR	R3C	R3CR	P-31M	P-31C	P-3M1	P-3C1
R-3M	R-3MR	R-3C	R-3CR	P6	P61	P65	P62
P64	P63	P-6	P6/M	P63/M	P622	P6122	P6522
P6222	P6422	P6322	P6MM	P6CC	P63CM	P63MC	P-6M2

TABLE 10.

P1	P-1	P2	P112	P21	P1121	C2	A2
P-6C2	P-62M	P-62C	P6/MMM	P6/MCC	P63/MCM	P63/MMC	P23
F23	I23	P213	I213	PM3	PM-3	PN3	PN-3
FM3	FM-3	FD3	FD-3	IM3	IM-3	PA3	PA-3
IA3	IA-3	P432	P4232	F432	F4132	I432	P4332
P4132	I4132	P-43M	F-43M	I43M	P-43N	F-43C	I-43D
PM3M	PM-3M	PN3N	PN-3N	PM3N	PM-3N	PN3M	PN-3M
FM3M	FM-3M	FM3C	FM-3C	FD3M	FD-3M	FD3C	FD-3C
IM3M	IM-3M	IA3D	IA-3D				

NONSTANDARD SETTINGS WITH CELL TRANSFORMATION

TABLE 11.

ABC	BA-C	CAB	-CBA	BCA	A-CB	P2221	P2221
P2122	P2122	P2212	P2212	P21212	P21212	P22121	P22121
P21221	P21221	C2221	C2221	A2122	A2122	B2212	B2212
C222	C222	A222	A222	B222	B222	PMM2	PMM2
P2MM	P2MM	PM2M	PM2M	PMC21	PCM21	P21MA	P21AM
PB21M	PM21B	PCC2	PCC2	P2AA	P2AA	PBAB	PBAB
PMA2	PBM2	P2MB	P2CM	PC2M	PM2A	PCA21	PBC21
P21AB	P21CA	PC21B	PB21A	PNC2	PCN2	P2NA	P2AN
PB2N	PN2B	PMN21	PNM21	P21MN	P21NM	PN21M	PM21N
PBA2	PBA2	P2CB	P2CB	PC2A	PC2A	PNA21	PBN21
P21NB	P21CN	PC21N	PN21A	PNN2	PNN2	P2NN	P2NN
PN2N	PN2N	CMM2	CMM2	A2MM	A2MM	BM2M	BM2M
CMC21	CCM21	A21MA	A21AM	BB21M	BM21B	CCC2	CCC2
A2AA	A2AA	BB2B	BB2B	AMM2	BMM2	B2MM	C2MM
CM2M	AM2M	ABM2	BMA2	B2CM	C2MB	CM2A	AC2M
AMA2	BBM2	B2MB	C2CM	CC2M	AM2A	ABA2	BBA2
B2CB	C2CB	CC2A	AC2A	FMM2	FMM2	F2MM	F2MM
FM2M	FM2M	FDD2	FDD2	F2DD	F2DD	FD2D	FD2D
IMM2	IMM2	I2MM	I2MM	IM2M	IM2M	IBA2	IBA2
I2CB	I2CB	IC2A	IC2A	IMA2	IBM2	I2MB	I2CM
IC2M	IM2A	PCCM	PCCM	PMAA	PMAA	PBMB	PBMB
PBAN	PBAN	PNCB	PNCB	PCNA	PCNA	PMMA	PMMB
PBMM	PCMM	PMCM	PMAM	PNNA	PNNB	PBNN	PCNN
PNCN	PNAN	PMNA	PNMB	PBMN	PCNM	PNCM	PMAN
PCCA	PCCB	PBAA	PCAA	PBCB	PBAB	PBAM	PBAM
PMCB	PMCB	PCMA	PCMA	PCCN	PCCN	PNAA	PNAA

TABLE 11.

ABC	BA-C	CAB	-CBA	BCA	A-CB	P2221	P2221
PBNB	PBNB	PBCM	PCAM	PMCA	PMAB	PBMA	PCMB
PNNM	PNNM	PMNN	PMNN	PNMN	PNMN	PMMN	PMMN
PNMM	PNMM	PMNM	PMNM	PBCN	PCAN	PNCA	PNAB
PBNA	PCNB	PBCA	PCAB	PBCA	PCAB	PBCA	PCAB
PNMA	PMNB	PBNM	PCMN	PMCN	PNAM	CMCM	CCMM
AMMA	AMAM	BBMM	BMMB	CMCA	CCMB	ABMA	ACAM
BBCM	BMAB	CMMM	CMMM	AMMM	AMMM	BMMM	BMMM
CCCM	CCCM	AMAA	AMAA	BBMB	BBMB	CMMA	CMMB
ABMM	ACMM	BMCM	BMAM	CCCA	CCCB	ABAA	ACAA
BBCB	BBAB	IBAM	IBAM	IMCB	IMCB	ICMA	ICMA
IBCA	ICAB	IBCA	ICAB	IBCA	ICAB	IMMA	IMMB
IBMM	ICMM	IMCM	IMAM				

APPENDIX - V: ATOMIC RADII

Atomic radii used for covalent bonding etc.

TABLE 12.

Ac	**1.88**	**Er**	**1.73**	**Na**	**0.97**	**Sb**	**1.46**
Ag	1.59	Eu	1.99	Nb	1.48	Sc	1.44
Al	1.35	F	0.64	Nd	1.81	Se	1.22
Am	1.51	Fe	1.34	Ni	1.50	Si	1.20
As	1.21	Ga	1.22	Np	1.55	Sm	1.80
Au	1.50	Gd	1.79	O	0.68	Sn	1.46
B	0.83	Ge	1.17	Os	1.37	Sr	1.12
Ba	1.34	H	0.23	P	1.05	Ta	1.43
Be	0.35	Hf	1.57	Pa	1.61	Tb	1.76
Bi	1.54	Hg	1.70	Pb	1.54	Tc	1.35
Br	1.21	Ho	1.74	Pd	1.50	Te	1.47
C	0.68	I	1.40	Pm	1.80	Th	1.79
Ca	0.99	In	1.63	Po	1.68	Ti	1.47
Cd	1.69	Ir	1.32	Pr	1.82	Tl	1.55
Ce	1.83	K	1.33	Pt	1.50	Tm	1.72
Cl	0.99	La	1.87	Pu	1.53	U	1.58
Co	1.33	Li	0.68	Ra	1.90	V	1.33
Cr	1.35	Lu	1.72	Rr	1.47	W	1.37
Cs	1.67	Mg	1.10	Re	1.35	Y	1.78
Cu	1.52	Mn	1.35	Rh	1.45	Yb	1.94

TABLE 12.

Ac	**1.88**	**Er**	**1.73**	**Na**	**0.97**	**Sb**	**1.46**
D	0.23	Mo	1.47	Ru	1.40	Zn	1.45
Dy	1.75	N	0.68	S	1.02	Zr	1.56

Note: OW is equivalent to O and Q1 is equivalent to C1.
Covalent radii are those given in the Cambridge Structural database manual.

APPENDIX - VI: THE GRAPHICS PACKAGE

The graphics drivers provide for three types of output:

HPD - Interactive graphics output on the display. HPP - Interactive graphics output on the plotter. CAL - A meta-graphics file with graphics instructions for later use.

The type of graphics output is managed by the user with instructions of the type: HPP ON to open the graphics output channel to the plotter; HPD ON to (re)open the graphics channel to the display; CAL ON to generate a graphics file. Both the CAL and either HPP or HPD may be open but not all three. HPP and HPD are mutually exclusive. The graphics interface package is implementation dependent i.e. different for the PC and non-PC implementations. Some details pertinent to the implementation that is used is available though the instruction HELP GRAPHICS.

The PC-package supports various display adapters, including CGA, EGA, MCGA, PGA, VGA, Hercules, and HP-GL plotters (either on-line or off-line). The off-line code is easily transferred in text processors that can accommodate graphics. A file named CONFIG.PLT stores implementation dependent parameters so that they need not be specified at each program start-up. The non-PC version supports Hewlett-Packard and Tektronix displays, VAX-workstations, HP plotters or HP-GL and Tektronix compatible laser printers.

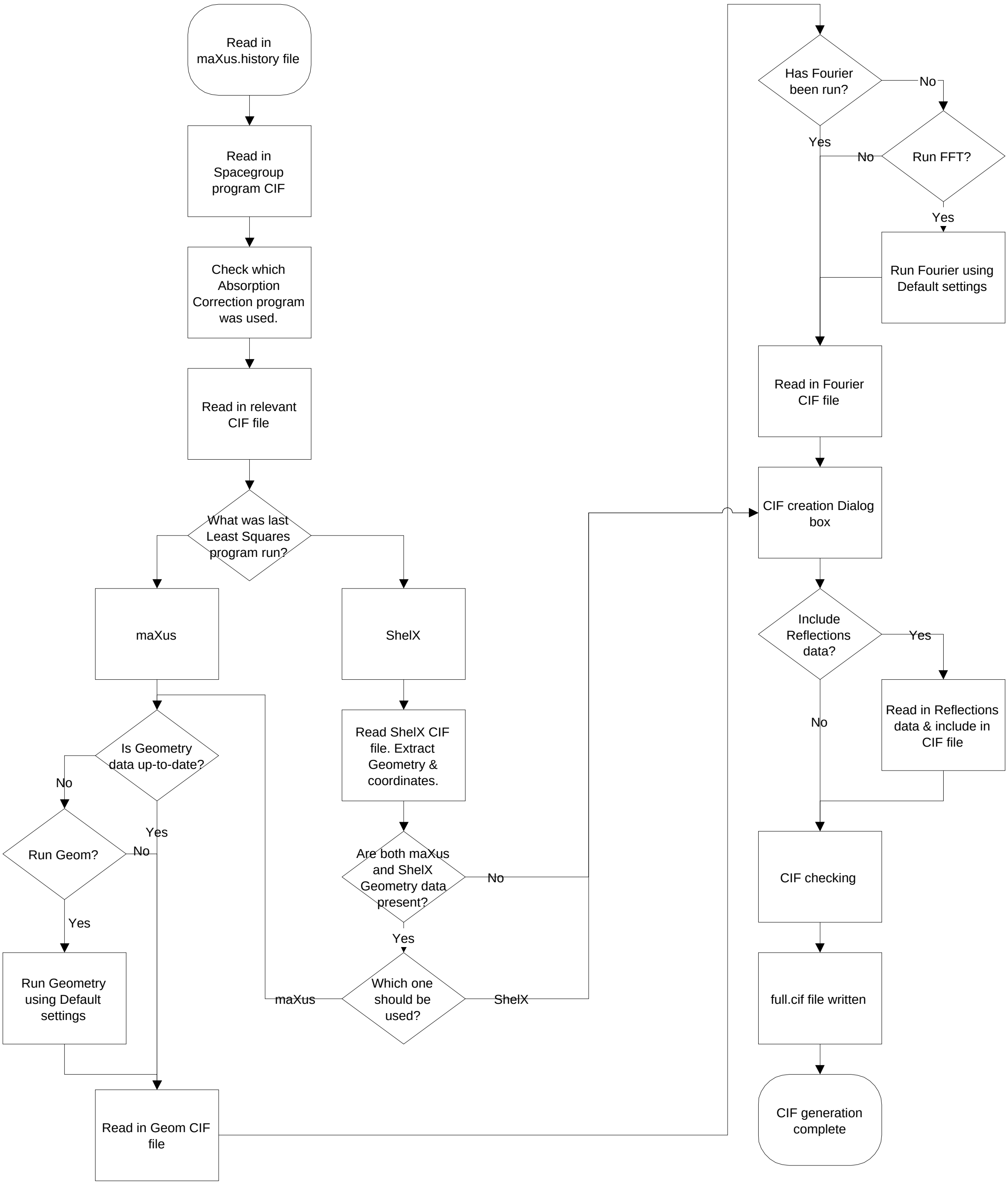
Non-PC Graphics:
The kind of code that may be generated for the various media is:
HPD : TEK4010 (default),
HP2648A,
VAX-workstation
HPP : HP7221A (default) or HP-GL CAL
TEK4014 (default),
HP-GL or CARTH (=3D CODE)

Type to change the currently generated display code: SET DISPLAY TEK or SET DISPLAY HP and to change the currently generated plotter code: SET PLOTTER HPBIN or SET PLOTTER HPGL Default plotter format is A4(vertical); other available formats are: 3 = A3(horizontal), -3 = A3(vertical), 4 = A4(vertical) and -4 = A4(horizontal).

E.g. to change the default setting to A3 type: SET PLOTTER FORMAT 3.

APPENDIX C

Flowchart of CIF Generation program



18 Advanced User Options

Databases and Defaults, Files, Known Bugs and Current Limits

18.1 Introduction

This chapter describes several options which the average user of maXus will use only very rarely; some advanced users may find some of the extra options available useful, and thus they are documented here.

This chapter covers the following topics:

- Editing the maXus databases
- Chaning the behaviour of the pattern precession photo program.
- Changing application defaults
- Files written and read by various sections of maXus
- Miscellaneous known bugs and features of maXus
- Current limits in maXus

With the possible exception of the last topic, these are not sections of the manual that would need to be referred to in the day-to-day running of the program.

18.2 Databases

There are two databases used in maXus:

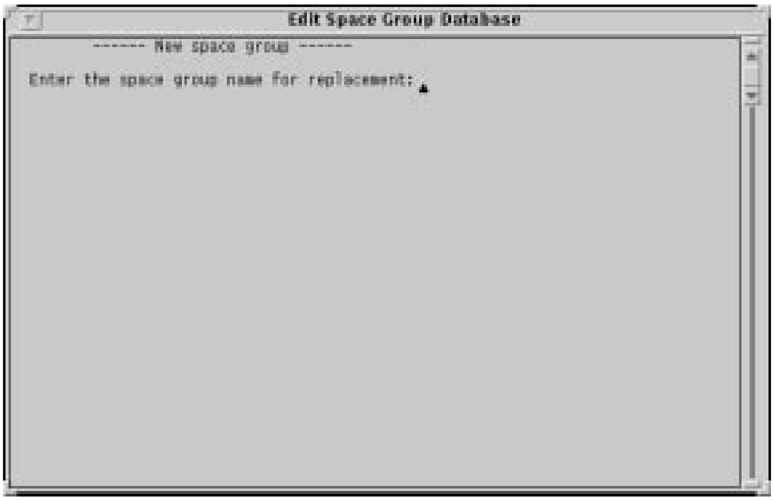
- Space group information
- Scattering factors, covalent radii, μ, df′, df″ etc.

Selecting *Databases* from the main maXus menu allows you to edit the two different databases. This is not something that requires to be done by the average user. In some specific situations, it may become necessary to add or alter entries in one of the databases used by maXus.

18.2.1 Space Groups Database.

maXus employs a database of space groups stored in an appropriate library directory which cannot be directly edited by the user. It contains a large number of standard and non-standard space groups, used by the Spacegroup program. You may wish to use a non-standard setting that is not in this list. This is easily done:

(1) Make sure you are in the directory which contains your Original file. In the *Databases* menu on the main maXus window, there is an entry *Edit Space Group*. Select this, and a dialogue box appears:



```
------ New space group ------

Enter the space group name for replacement:
```

(2) Enter the name of the space group you wish to *replace*. i.e. the space group for which you are entering a non-standard setting. You must put spaces between the first symbol and all the symmetry elements e.g. Pna2(1) must be typed P n a 2(1).

(3) Enter the new space group name as requested; this should also use spaces as in (2).

(4) Enter the symmetry operations in free format, one per line. *Do not include operations arising from the centre of symmetry, or lattice centerings*. Press return <cr> to terminate each line and to terminate the list.

(5) A new version of the database called SpGroup will be copied into your current directory. Every program which accesses this will search your local directory for this file first and use it if it is found, otherwise the standard database is used.

(6) If the program fails then no database is written. You can remove the database *via* the UNIX command:

```
rm SpGroup
```

(7) You can now work in the usual way. Use the space group browser in the Spacegroup program to select the new, non-standard space group. When the new space group is selected a message will appear:

```
This space group has a non-standard setting and
was manually added to the space group database
```

Successful completion gives:

All Done

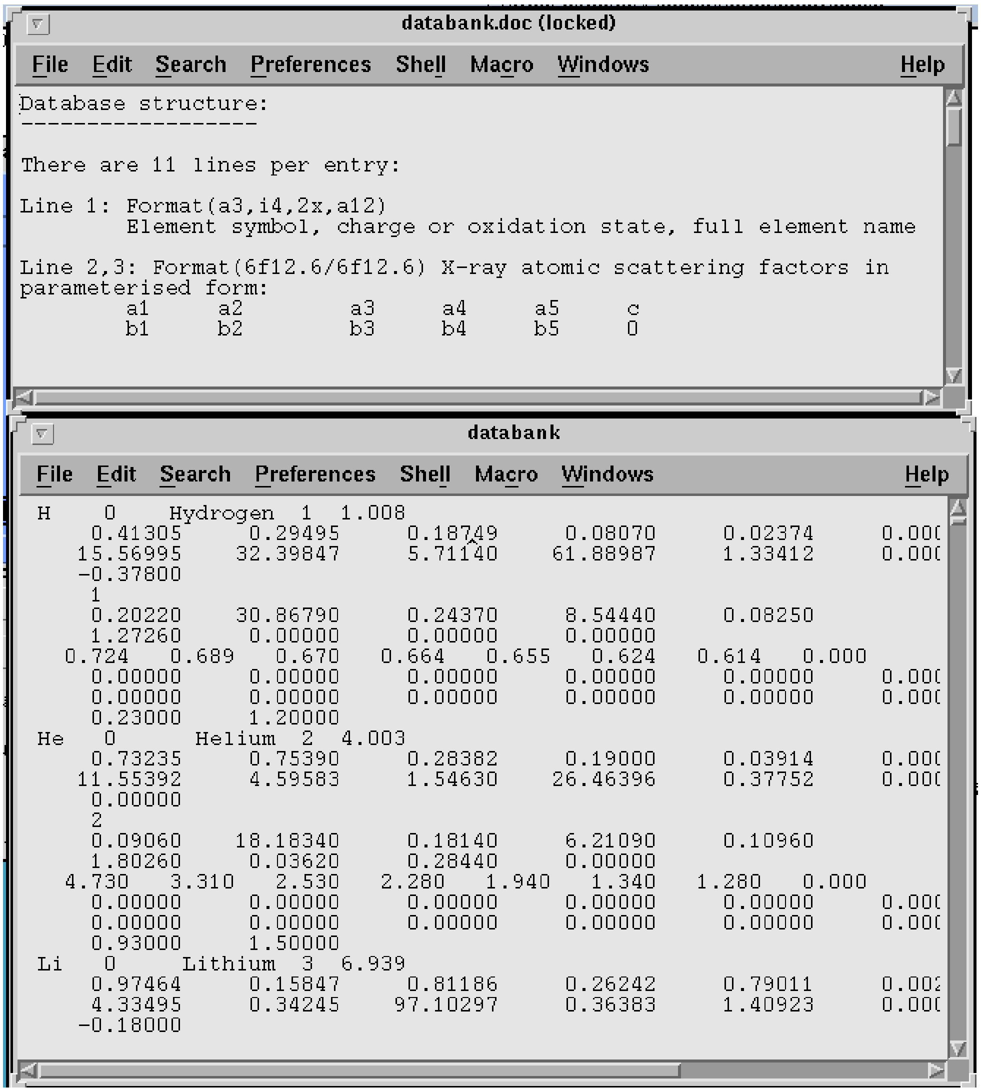
An error gives:

No match

This means that the space group name for which a replacement was to be entered does not exist.

18.2.2 Elements Database

In the *Databases* menu on the main maXus window, there is an entry *Edit Elements*. Select this, and two text boxes appear:



```
Database structure:
------------------

There are 11 lines per entry:

Line 1: Format(a3,i4,2x,a12)
        Element symbol, charge or oxidation state, full element name

Line 2,3: Format(6f12.6/6f12.6) X-ray atomic scattering factors in
parameterised form:
        a1      a2        a3      a4      a5      c
        b1      b2        b3      b4      b5      0
```

```
H    0    Hydrogen  1  1.008
    0.41305     0.29495     0.18749     0.08070     0.02374     0.000
   15.56995    32.39847     5.71140    61.88987     1.33412     0.000
   -0.37800
    1
    0.20220    30.86790     0.24370     8.54440     0.08250
    1.27260     0.00000     0.00000     0.00000
  0.724   0.689   0.670   0.664   0.655   0.624   0.614   0.000
    0.00000     0.00000     0.00000     0.00000     0.00000     0.000
    0.00000     0.00000     0.00000     0.00000     0.00000     0.000
    0.23000     1.20000
He   0      Helium  2  4.003
    0.73235     0.75390     0.28382     0.19000     0.03914     0.000
   11.55392     4.59583     1.54630    26.46396     0.37752     0.000
    0.00000
    2
    0.09060    18.18340     0.18140     6.21090     0.10960
    1.80260     0.03620     0.28440     0.00000
  4.730   3.310   2.530   2.280   1.940   1.340   1.280   0.000
    0.00000     0.00000     0.00000     0.00000     0.00000     0.000
    0.00000     0.00000     0.00000     0.00000     0.00000     0.000
    0.93000     1.50000
Li   0     Lithium  3  6.939
    0.97464     0.15847     0.81186     0.26242     0.79011     0.002
    4.33495     0.34245    97.10297     0.36383     1.40923     0.000
   -0.18000
```

The top box describes the format & layout of the data, and what each field represents.

The lower text pane is the actual databank. Entries can be edited or added to as required, although caution is strongly recommended since the data here are used extensively in calculations throughout maXus. The default values should not be changed unless you know exactly what you are doing.

Edit the data bank and save the edited file. If you do this a local copy is included in the directory you are using. Every time maXus tries to access the databank it looks for this local version first. If it is found then it is used, if not the global databank is used. Delete the local copy of the databank to stop this activity: `rm databank` is the UNIX command.

18.2.3 References

“New Analytical Scattering Factor Functions for Free Atoms and Ions”, D. Waasmaier & A. Kirfel, *Acta Cryst.* 1995, **A51**, 416-431)

(D.REZ, P.REZ & I.GRANT, *Acta Cryst.* (1994), **A50**, 481-497)

Smith, G.H. & Burge, R.E. (1962) *Acta Cryst.* **15**, 182-186

Doyle, P.A. & Turner, P.S. (1968)*Acta Cryst.* **A24**, 390-397

Jiang, J.S. & Fang-Hua, L. (1984) *Acta Physica Sinica* **33**, 845-849

μ & electron scattering factors are taken from International Tables Volume C 193-199

df′ and df″ are taken from International Tables Volume C 219-221.

N.W.Alcock ‘*Bonding and Structure in Organic Chemistry*’ Ellis Horwood (1990)

A Bondi, *J. Phys. Chem.*, 68, (1963), pp441-51.

Cambridge Structural Database (CSD)

Open University Second Level Chemistry Data Book (1976)

18.3 The ‘Pattern’ precession photo generator

This program can be run as an option during spacegroup determination. The default options produce 6 images - hk0,h0l,0kl,hk1,h1l,1kl. This can be altered by using the **Databases->Edit Pattern Template** menu item. This provides a text editor on the template file to which is added the cell parameter line and compound title, at runtime. The edited pattern.tem file will be left in the user data directory. Deleting this file will then revert to the original unedited template file ($CRYSTANHOME/lib/pattern.tem).

A copy of any generated photos is saved in the file *pattern.ps* in the user data directory.

There follows the description of the program and commands provided by the author.

This program is free distributed without any restriction. The author is Guoguang Lu,
Dept.of.Molelcular.Biology,Box590,75124,Uppsala,Sweden
e-mail guoguang@xray.bmc.uu.se
fax (46)-18-536971.

This is a program to plot a diffraction pattern from a dataset (any ascii format reflection file) by generating a postscript file. The picture is in the style of a precession film. The black paint degree and radius of the spots is linearly correlated with log intensity. In all the tested cases, if data quality is OK, it is very similar to a real procession film. It can plot any section. Users can directly supply a spacegroup number or name or give symmetry operations to define the symmetry information. Possible uses are:

1. check the space group (if data collected as lower space group)
2. check the quality of the dataset by comparing it with precession photo,
3. monitor the completeness when collecting data from area detector.

4. I think that this might help some people to develop new approach in protein crystallography by directly “seeing” the reciprocal space.
5. Visualise twinning problems.

18.3.1 File Description

- FORTRAN unit 1 Input format reflection data file. File name and format can be specified when running program.
- SYMOP Input symmetry operation library file in ccp4 style File name should be defined in the command file by $ASSIGN or setenv before running program
- post Output postscript file, file name should be defined in the command file by $ASSIGN or setenv before running program
- Unit 41--->40+number of section: temporary working file.

18.3.2 Pattern command list (alphabetical order)

- BACKGROUND   background_amplitude.

  when the fobs is less than this value, the program will treat it as background and ignore them. The remaining reflections will have stronger contrast. In some cases, changing this value may produce a better match between the generated image and a real image.
- BEAM   beam_radii (in mm)

  this is to control the spot size when the reflection is strongest, the spot radii is beam_radii+ least_radii. If not given or set to zero, the program will calculate a default value according to the space group and cell dimensions to avoid overlap.
- CELL a,b,c,alpha,beta,gamma
- DIRECTION [def l 0]

  This means you scan layer (h k 0). Now you can only scan along h, k, or l direction. If you want to scan along (111) direction, you have use command: SECTION h+k+l=0 , etc.
- EXPOSE_TIME   100 (hours) (should not be changed)

  The full value is 100 hours.
- FILE reflection_filename

  The program read H K L FOBS SIGF (or, I SigI) in order, but not use SIGF, i.e. it just need a file with FOBS. If input is H K L I SigI, it gives you exactly same picture This parameter should be left unchanged for use inside maXus.
- FORMAT “format” --- example (3I4 2F10.2)

  Format of the reflection. H K L FOBS SIGF (or I SIGI). This parameter should be left unchanged for use inside maXus.
- GRAYgray [def, 0.96]

  Background gray level. The correct value must be between 1.0-0.0, If it is 0.0, the whole picture will become completely black. It is 1.0, there will be no gray background in the picture. The recommended value is 0.9-0.99. Default is 0.96

- NOCENtral_symmetry_operation_in_reciprocal_space

  If this card is not given, central symmetry in reciprocal space will be automatically applied.

- RADiu least_radii

  when the reflection is weakest, the spot radii is least_radii.

- RESOLution resolution

  Default is 4.62 Å. Any value could be given. But if it is too high, the pattern might be out of your "film". In this case change screen_distance.

- SCREEN_DISTANCE [default 250 (mm)]

  This is the distance between crystal and film in your precession camera. It is used to control the scale.

- SECTION [def l=0]

  Example section h+2k-l=3, or section h=0, or section h-k=1, etc. In principle, one could use ah+bk+cl=d (where a,b,c, and d are integer) to describe any section. Using this expression, it is possible to check symmetry along any direction. Input should be in a order of h,k,l If this command is repeated, several sections can be plotted on separate pages.

- SPACEgroup Number or Name or PG+point group name

  Example:SPACE 19(spacegroup number)

  SPACE P212121(spacegroup name)

  SPACE PG222(point group name)

  If your space group is not inside the SYMOP file like P6222, you have to give SYMM command or edit the SYMOP file.

  Note that this function is not used by maXus, and the ccp4 symop library is not supplied, so spacegroups names and numbers are not supported in the pattern program.

- SYMMetry symmetric_operation such as -x,y,-z etc.

  This symmetric operation must be in real space and be compatible with the cell you defined. Translation operations will be ignored. If you have not merged the data, you could consider not to use this card to show the symmetry and find the space group. In that case, there is only h,k,l operation in reciprocal space.

- TITLE text

  This title will be printed in the picture. It is added by maXus at runtime

- WAVE_LENGTH 1.54

  You can ignore this card unless you want to plot a picture which has same scale with you real precession film. In that case you must give same screen_distance.

- Note:

1. Lines with first character '!' or ';' will be ignored.
2. Central symmetry in reciprocal space is automatically added unless you give a card NOCENter
3. The program reads free format input and is case-insensitive.
4. If you feel spacegroup name is strange, it is better to give a number.

18.4 Changing Application Defaults

Some programs and modules within maXus allow the user to alter certain default options that the program presents in windows and dialog boxes.

This feature is still under development, so only certain defaults can be altered by the user at present. Further options will be introduced at a later date.

Currently, defaults may be altered for several main maXus windows, as well as for Plotq and Ortep.

Note that for buttons, off = false, on = true.

18.4.1 maXus Defaults

User-alterable program defaults are contained in the maXus directory in:

```
/lib/app-defaults/mc
```

This is a plain text file, which can be altered in any standard text editor. It is heavily commented and changes may sometime be made in the defaults file which are not shown here in the documentation. Remember to always keep a pristine copy in case of emergency. Changes take effect when maXus is restarted.

The file begins:

```
! maXus MAIN PROGRAM defaults are in this file
! First, set the defaults for the whole program ----------------
*background:lightsteelblue
*FontList:*helvetica-medium-r-normal--14-*
*SelectColor:light green
! The time for which the ’about’ window displays on startup, in
secs
*start_delay:4
```

The only value here which it might be useful to alter is the ‘About’ window start delay.

This controls how long the About window is displayed for when maXus is launched. Setting the value for this to 0 prevents the window from being shown.

Defaults for the CIF import window dialog:

```
! Import cif window defaults follow  --------------------------
! Filename of the cif file to import
*.impcif_file.value: import.cif
! ------------------------------------------------------------
```

The last entry controls the default filename (import.cif) displayed in the dialog. Altering this can be useful if you always import CIF files with the same name.

Defaults for the CAD4 import window dialog:

```
! Import CAD4 window defaults follow  ------------------------
! Filename of the cad4 file to import
*.cad4_dataname.value: CAD4.DAT
! -------------------------------------------------------------
```

The last entry controls the default filename (CAD4.DAT) displayed in the dialog. Altering this can be useful if you always import CAD4 files with the same name.

Defaults for the ASCII import window dialog:

```
! Import ASCII window defaults follow  -----------------------
! Filename to import
*.impascii_file.value: ascii.dat
! Free format file by default
*.impascii_free_format_toggle.set: true
! If not free format, use this f77-style fixed format
*.impascii_format_text.value: (3I4,2F8.2)
! -------------------------------------------------------------
```

The `*.impascii_file.value` entry controls the default filename (ascii.dat) displayed in the dialog. Altering this can be useful if you always import ASCII files with the same name.

`*.impascii_free_format_toggle.set` should be set to either true or false; this controls whether the Free Format switch is set on or off as a default.

If not using Free Format, but Fortran 77 style instead, altering `*.impascii_format_text.value` lets you change the default expected format.

Defaults for the Spacegroup dialog window:

```
! SpaceGroup window defaults follow  --------------------------
! Display simulated precession photos
*.sg_win*sg_display_precess_toggle.set: true
! I/sigmaI cutoff for photo display
*.sg_precess_limit.value: 0.0
! Display reflection intensity histograms
*.sg_win*sg_display_graphs_toggle.set: true
! Accelerating Voltage
*.sg_win*sg_av.value: 100
! Correct for decomposition if possible (on by default)
*.sg_correct.set: true
! prompt user for a transformation matrix when starting sg
determination (off by default)
*.sg_transform.set: false
```

```
! Delta limit for metric symmetry check. Do not change unless
you are sure.
*.sg_delta.value: 0.3
! --------------------------------------------------------------
```

The first toggles set whether the Precession photos are set to on or off as a default. The next controls the default value of the I/σ(I) cutoff value. Next is the control for the Intensity histograms toggle, followed by the value for the accelerating voltage used. The next toggle controls the Decomposition correction button, follwed by the button to control if the user is asked to enter a Transformation matrix. Finally, you can alter the Delta limit for the LePage metric symmetry check.

Defaults for the Least Squares refinement dialog:

```
! Lsq (GX) window defaults follow  ----------------------------
! reflection inclusion sigma cutoff value
*.lsq_win*lsq_cutoff.value:3
! Refine to convergence ? (number of cycles will be ignored if
true)
*.lsq_convergence_toggle.set: true
! Number of cycles to run
*.lsq_cycles.value: 8
! shift/sigma convergence criterion
*.lsq_convergence.value: 0.1
! Whether to use the reflection omit file
*.lsq_use_omit_file_toggle.set: false
! Ignore all groups set in the model window ?
*.lsq_ignore_groups_toggle.set: false
! Ignore all restraints (_not_ constraints) set in model window
?
*.lsq_ignore_restr_toggle.set: false
! --------------------------------------------------------------
```

The first variable sets the entry for the Sigma cutoff value; the next option controls if the Refine to convergence button is toggled on or off. The next option controls the default number of cycles run (although this is ignored if the previous option is on), followed by the shift/σ convergence criterion. The next option controls if the Use Omit File button is toggled on or off. The next two toggles control if the Groups and Restraint settings specified in the Model window are ignored (true) or used (false).

Defaults for the Least Squares: Hydrogen Options Sub-Window:

```
! Lsq (GX) Hydrogen sub-window defaults follow ------
! Ignore model window refinement settings for H, and use custom
settings ?
! If this is false, all the following will be ignored
*.lsq_h_ignore_model_toggle.set: false
! xyz constraints for ’generated’ H atoms [*lsq_h_gen_ride |
*lsq_h_gen_free]
```

```
*.lsq_h_gen_xyz_opt: *lsq_h_gen_ride
! adp constraints for 'generated' H atoms
! Values are [*lsq_h_gen_adp_fixed | *lsq_h_gen_adp_offset |
*lsq_h_gen_adp_free]
*.lsq_h_gen_adp_opt: *lsq_h_gen_adp_fixed
! xyz constraints for 'found' H atoms [*lsq_h_fnd_ride |
*lsq_h_fnd_free]
*.lsq_h_fnd_xyz_opt: *lsq_h_fnd_ride
! adp constraints for 'found' H atoms
! Values are [*lsq_h_fnd_adp_fixed | *lsq_h_fnd_adp_offset |
*lsq_h_fnd_adp_free]
*.lsq_h_fnd_adp_opt: *lsq_h_fnd_adp_fixed
!
! Fixed adp (or offset from parent) value for generated H
*.lsq_h_gen_adp_text.value: 0.02
! Fixed adp (or offset from parent) value for found H
*.lsq_h_fnd_adp_text.value: 0.02
! ---------------------------------------------------
```

If the first toggle is set to false, all of the settings that follow it will be ignored.

Several of the rest of the setting control which option from a pop-up menu is selected as the default option. This possible options for each menu are listed above the line for the variable.

The only numerical values which can be altered are for the distance offset from the parent atom for both generated and found hydrogens, which are default set to 0.02.

Defaults for the ShelXL refinement window dialog:

```
! Shelxl refinement window defaults follow --------------------
! Number of cycles of refinement
*.shelxl_cycles.value: 8
! Maximum 2theta
*.shelxl_max2theta.value: 180
! Reflection selection sigma cutoff
*.shelxl_cutoff.value: -3.0
! Extinction
*.shelxl_extinction.value: 0.0
! Solvent waters g value
*.shelxl_swatg.value: 0.0
! Solvent waters g value
! Solvent waters U value
*.shelxl_swatu.value: 2.0
! Verbosity level
*.shelxl_more.value: 1
! Resolution shell low-resolution limit (A)
*.shelxl_lores.value: 999.9
! Resolution shell high-resolution limit (A)
*.shelxl_hires.value: 0.0
```

```
! --------------------------------------------------------------
```

Defaults for the Fourier window dialog:

```
! Weighted Fourier window defaults follow ---------------------
! Number of peaks to find - blank means auto-determine number
*.wtf_peaks.value:
! Bonding parameters - maximum bond angle
*.wtf_anglemax.value: 145.0
! Bonding parameters - minimum bond angle
*.wtf_anglemin.value: 85.0
! Bonding parameters - maximum bond length
*.wtf_bondmax.value: 1.65
! Bonding parameters - minimum bond length
*.wtf_bondmin.value: 1.00
! Minimum atom-peak separation
*.wtf_minpeakdist.value: 0.5
! Minumum I/sig(I)
*.wtf_minisigi.value: 0.0
! FFT gris size (A)
*.wtf_grid.value: 0.333
! Minimum sin(theta)/lambda
*.wtf_minsinthlam.value: 0.0
! Maximum sin(theta)/lambda
*.wtf_maxsinthlam.value: 1.0
! -------------------------------------------------------------
```

Defaults for the Structure Factor calculation dialog:

```
! SFac (structure factor calculation) defaults follow ----------
! Whether to use the reflection omit file
*.sfac_use_omit_file_toggle.set: false
! Sigma cutoff value (compare Lsq above)
*.sfac_cutoff.value: 3
! --------------------------------------------------------------
```

The only options which can currently be altered here are whether the Use Omit File option is checked or unchecked, and fhe value for the Sigma cutoff.

Defaults for the Weight Analysis program:

```
! Weight analysis (wtan) defaults follow ----------------------
! Create omit file to exclude certain reflections see above
*.wtan_write_omit_file_toggle.set: false
! --------------------------------------------------------------
```

The single user-settable option here is to control the Write Omit File button.

Defaults for the Sortav dialog:

```
! sortav absorption corrections defaults follow ---------------
! Program previously known as eqv
! test run only
*.eqv_test_only.set: true
! ---------------------------------------------------
```

18.4.2 CIF Generation options

The default options for the CIF generation window (Publish menu) are contained in the file in the maXus directory:

```
/lib/app-defaults/cifgen
```

This file contains the following:

```
! This is the defaults file for the cifgen window
! Lines beginning with a '!' are comments
! The values you may wish to change are after the colon on each
line
*background:lightsteelblue
*FontList:*helvetica-medium-r-normal--14-*
*SelectColor:   light green
! Default for _chemical_compound_source
*.ccs_tf.value: Local laboratory
```

Obviously, you may wish to change the compound source from ‘Local laboratory’ to , for example, the name of your workplace.

```
! Default for _exptl_crystal_description
*.ecd_tf.value: Cube
! Default for _exptl_crystal_colour
*.ecc_tf.value: Colourless
! The default contents of the _cell_measurement_temperature text
field
*.cmt_tf.value: 298
! The default for _cell_measurement_theta_min
*.cmtmin_tf.value: 1
```

The crystal description and colour variables can also be altered as desired as can the temperature and theta min fields.

18.4.3 Plotq Defaults

In the maXus directory, the file:

```
lib/app-defaults/Pq
```

contains the user-definable defaults for the Plotq graphics program. These defaults are also built into the program, so it will continue to work if the file is missing. This standard X11 application default process provides a powerful way of customising the behaviour of Plotq. The current contents are as follows:

```
*fontList:-*-helvetica-medium-r-normal--14-*-*-*-*-*-*-*
*background:lightsteelblue
*SelectColor:light green
*capt_font:-*-helvetica-medium-r-*-*-20-*-*-*-*-*-*-*
*axes_font:-*-helvetica-medium-r-*-*-17-*-*-*-*-*-*-*
*atom_font:-*-helvetica-medium-r-*-*-14-*-*-*-*-*-*-*
*normal_contour0:white
*normal_contour1:black
*normal_contour2:#00AAFF
*normal_contour3:#006699
*normal_contour4:#009966
*normal_contour5:#00CC33
*normal_contour6:#00FF00
*normal_contour7:#44BB00
*normal_contour8:#669900
*normal_contour9:#887700
*normal_contour10:#AA6600
*normal_contour11:#BB5500
*normal_contour12:#CC3300
*normal_contour13:#FF0000
*normal_contour14:#00FF00
*bright_contour0:white
*bright_contour1:black
*bright_contour2:#BBBB00
*bright_contour3:#DDDD00
*bright_contour4:#FFFF00
*bright_contour5:#FFDD11
*bright_contour6:#FFBB33
*bright_contour7:#FF9955
*bright_contour8:#FF7755
*bright_contour9:#FF5577
*bright_contour10:#FF3399
*bright_contour11:#FF11BB
*bright_contour12:#FF00DD
*bright_contour13:#FF00FF
*bright_contour14:#00FF00
```

The fontlist, background and SelectColor resources change the Motif interface appearance and should not normally be altered as they are intended to be consistent throughout maXus. There are three font specifications which can also be altered on the command line, and should be valid X11 font names, available on the target system

The remaining resources are color specifications for contour colors. There are two sets, for the “Normal” color scheme and for the “Bright” color scheme as selected in the *View Options* dialog window. Most of the colors are given as hexadecimal numbers, indicating Red, Green Blue components as #RRGGBB. They can also be given as color names, in the standard X11 way. “contour0” is the background color, “contour1” is the foreground, used to draw axes and text. These should not normally be altered. Experienced users may also override X defaults in their own environment using mechanisms such as the personal Xdefaults file but a discussion of such means is outwith the scope of this manual.

18.4.4 Ortep Defaults

In the maXus directory, the file:

/lib/app-defaults/Ortep

contains the program and user-alterable defaults for the Ortep program. The file contains the following:

```
! DTB_USER_RES_START - vvv Add file header below vvv
! DTB_USER_RES_END - ^^^ Add file header above ^^^

! File: Ortep
!
! Resource file for application ortep
! This file was generated by dtcodegen, from project ortep
!
! Any text may be added between the DTB_USER_RES_START
! and DTB_USER_RES_END comments. Descriptive comments are
! provided only as an aid.
!
!  ** EDIT ONLY WITHIN SECTIONS MARKED WITH DTB_USER_ COMMENTS.      **
!  ** ALL OTHER MODIFICATIONS WILL BE OVERWRITTEN. DO NOT MODIFY OR  **
!  ** DELETE THE GENERATED COMMENTS! **
!

! All CDE applications should include the standard Dt resource file
!#include "Dt"


! Resource list for AppBuilder project ortep
!

! Resource list for AppBuilder module ortep
!
Ortep*dtb_ortep_print_dialog.background:white
Ortep*dtb_ortep_print_dialog_shellform.background:white
Ortep*dtb_ortep_print_dialog_shellform*print_dialog_panedwin.background:
white
Ortep*dtb_ortep_print_dialog_shellform*print_dialog_form.background:
white
Ortep*dtb_ortep_atoms_dialog.background:white
Ortep*dtb_ortep_atoms_dialog_shellform.background:white
Ortep*dtb_ortep_atoms_dialog_shellform*atoms_dialog_panedwin.background:
white
```

```
Ortep*dtb_ortep_atoms_dialog_shellform*atoms_dialog_form.background:
white
Ortep*dtb_ortep_view_dialog.background:white
Ortep*dtb_ortep_view_dialog_shellform.background:white
Ortep*dtb_ortep_view_dialog_shellform*view_dialog_panedwin.background:
white
Ortep*dtb_ortep_view_dialog_shellform*view_dialog_form.background:white
Ortep*dtb_ortep_extras_dialog.background:white
Ortep*dtb_ortep_extras_dialog_shellform.background:white
Ortep*dtb_ortep_extras_dialog_shellform*extras_dialog_panedwin.backgroun
d:white
Ortep*dtb_ortep_extras_dialog_shellform*extras_dialog_form.background:
white
Ortep*dtb_ortep_mainwindow_mainwin*drawpane_frame.background:white
Ortep*dtb_ortep_mainwindow_mainwin*drawpane.background:white
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
! The remainder of this file (following this comment) may be modified.
! Add new resources or override values defined above.
!
! DTB_USER_RES_START
Ortep*fontList: -*-helvetica-medium-r-normal--14-*-*-*-*-*-*-*
Ortep*background:lightsteelblue
Ortep*SelectColor:light green
```

As warned in the file, do not alter any of the defaults in the main body of the file. If you wish to override any of the settings, repeat the particular line with altered variable, after the `DTB_USER_RES_START` line at the bottom of the file.

18.5 Files

This section details all of the various files which maXus reads, writes and modifies during the course of its operation.

Starting with:

`Original`

After running Auto:

`AUTOsolve.status`
`Model`
`CRYSIN`
`Cormat`
`DrawFile`
`DrawFile%`
`HklCal`
`HklObs`
`Map`
`demo.bin`
`Model%`
`MulRef`
`Multan78`
`Original`
`Tables`
`autosolve.conf`
`RefBin`
`maXus.history`
`Sysout_Init`
`Sysout_Lsq`
`Sysout_SFac`
`Sysout_SIR`
`sg.cif`
`sir92.in`
`cols.bysym`
`refine.cif`
`demo.ins`
`demo.plt`
`flsq.gx`
`fort.31`
`fort.40`

LsqRpt
PeakData
blsq.gx

Starting with :

Original

After running Spacegroup:

CARD3
CRYSIN
DrawFile
HklObs
Model
MulRef
Multan78
Original
Original.nodecay
Original.original
laue2rint
Sysout_Init
sg.cif
maXus.history

After running Testi:

New file: Sysout_Testi

After running Solve: SIR:

New Files:

demo.plt
Sysout_SIR
Model%
blsq.gx
Tables
LsqRpt
demo.ins
RefBin
Drawfile%
PeakData
Cormat
HklCal
demo.bin
cols.bysym
flsq.gx
Map
fort.31
Sysout_SFac

Modified Files:

Model
DrawFile

After running Complete: SIR:

Extra files:

sir92.frag
sir92.in

Modified files:

demo.bin
demo.ins
Model%
blsq.gx
HklCal
demo.plt
RefBin
DrawFile%
Cormat
Model
Sysout_SIR
cols.bysym
flsq.gx
Sysout_SFac
DrawFile
LsqRpt

After Model has been run (and changes made):

Modified Files:

`Model`
`DrawFile%`
`Model%`
`RefBin`
`DrawFile`
`cols.bysym`

After running Structure Factors:

Modified files:

`blsq.gx`
`Model%`
`HklCal`
`LsqRpt`
`flsq.gx`
`Cormat`
`Model`
`DrawFile%`
`Sysout_Sfac`
`DrawFile`

New files:

`meter.stat`

After running Refine:LSq

New files: `Sysout_Lsq`

Modified files:

`Model%`
`blsq.gx`
`Model`
`Drawfile%`
`Cormat`
`flsq.gx`
`HklCal`

After running Refine: Info:

New files: `hist1` `hist2`

After running Refine: Weight Anal

New files:

`omitgx.hkl`
`Sysout_Wnal`
`Summary`

After running Fourier:

New files:

`Sysout_Wtf`
`fort.60`
`FftRpt`
`dfft.gx`
`Wtf.lsq.out`
`scratch.gx`

Modified files:

`DrawFile%`
`DrawFile`
`Model%`
`Summary`
`Map`

After running Solve:Mithril (Auto-assign):

New/modified files:

MITH.I
Multan78
MITH.i
MITH.P
Sysout_Mithril
demo.plt
demo.ins
flsq.gx
DrawFile%
fort.31
LsqRpt
DrawFile
MITH.r
MITH.e
MITH.p
MITH.M
sir92.in
demo.bin
RefBin
blsq.gx
Tables
Map
HklCal
Sysout_SFac
MITH.m
MITH.c
MITH.h
PeakFile
sir92.frag
Sysout_SIR
cols.bysym
Model%
Cormat
PeakData
Model

After running Refine:SHELXS97: (direct methods)

New/Modified:

shelxs97.ins
transfer
shelxs97.lst
meter.stat
shelxs97.hkl
Multan78
Sysout_SHELXS
shelxs2mx.err
binPhase2
shelxs97.res
DrawFile

After running Refine: MC Multan: Mon100 (auto-assign):

New/modified:

Multan78
binPhase1
LsResult
sir92.in
demo.bin
RefBin
DrawFile%
PEakData
transfer
binPhase
fourierMap
sir92.frag
Sysout_SIR
cols.bysym
flsq.gx
Map
bibPhase2
PeakFile
Sysout_Multan
demo.plt
demo.ins
Model%
blsq.gx
Tables

`Cormat` `fort.31` `HklCal`
`Model` `DrawFile` `LsqRpt`
`Sysout_SFac`

After running Complete: DIRDIF:

New/mod:

`ATMOD` `DDJOB` `CRYSDA`
`BINFO` `CONDA2` `DDLOG`
`ATOMS` `ATOLD` `DDSYST`
`IPR2` `XYZN` `Model`
`DrawFile` `Sysout_Dirdif`

After running Refine: SHELXL: Default:

New/Mod:

`Sysout_mx2` `shelxl` `shelxl.hkl`
`shelxs.lst` `shelxl.res` `DrawFile`
`Model` `Sysout_shelxl2mx` `shelxl.ins`
`RefBin` `meter.stat` `Model%`
`DrawFile%` `cols.bysym`

After running Abs: Difabs

New/Mod:

`Sysout_Abs` `HklAbsD`

After running Abs:Spherical:

New/Mod:

`Sysout_Abs` `HklAbsS`

After running Abs:Cylindrical:

New/mod files:

`Sysout_Abs` `HklAbsC`

After running Model: Expand:

New/Modified:

`Model` `DrawFile`
`Model.expand` `DrawFile.expand`

After running Geometry: Calculate:

New/Modified:

`Sysout_Geo` `Tables.ps` `Tables.tbl`
`Tables.txt` `geom.cif` `planes.gx`
`scratch3` `geom.gx`

After running Platon: Default (all options on):

New/modified:

`platon.com` `convplat.err` `platon.spf`
`Sysout_Platon` `platon.lis` `platon.par`

After running Publish: Calculate: Structure Factor Tables:

`ftab.gx` `Summary` `Sysout_Sft`

After running Publish: CIF: Create:

`full.cif` `refine.cif` `temp.cif`

After running Pub: CSSR:

`model.cssr`

After running Ortep:

`NewModel` `ortepdata` `OrthoCo`
`SavSeq` `Sysout_Ortep`

After running RasMol:

`drawfile.pdb`

After running Pluto:

`Sysout_Pluto` `Pluto.Com` `Pluto.pdb`

18.6 Known Bugs & Problems

This section details some of the known problems, bugs, features and strange behaviour of maXus in certain conditions. It is not intended to be an exhaustive list, and its compilation is also very much a work in progress.

18.6.1 Model

- In the Model window, if any of the modifier keys - i.e. CAPS LOCK, NUM LOCK etc. are on, the molecule will spin wildly and continually if any of the rotation control buttons are used. Turn off all modfier keys when using Model to prevent this from occuring.
- Hydrogen addition has some inconsistencies, particularly between the *Auto* Add-H and the manual method. Choice of constraints etc. for H addition is poor, and sometimes existing H atoms are removed when new ones are added. Note that the HFIX command in shelxl may provide a workaround for some problems, and atoms can be manually added in the Model window using fractional coordinates if known.

18.6.2 Ortep

- An infinite network of molecules can sometimes be generated by the 2.0Å search distance used to generate symmetry-related molecules. In this case drawing may take a long time and produce too many molecules outside the cell. The drawing will stop at the maximum number of atoms (currently 1000). To avoid this problem the dmax search sphere value in the “View” window should be reduced slightly and the diagram redrawn.
- Complex diagrams may be slow to draw. This can be improved by reducing the drawing style to a simple stick model with plain ellipses for the atoms. Detailed drawing types such as octant-shaded ellipsoids are unlikely to be useful when symmetry-related atoms are drawn in a large unit cell.

18.6.3 XXMol (SG Version)

- In some cases the default display of a structure directly from an input file may be incomplete or the display may even appear blank. This is caused by incorrect placement of the default clipping planes for the structure under consideration. The problem is easily corrected by resetting the clip planes in the viewing parameters in the Utility menu. The structure can then be reset to a reasonable size and position in the display using a combination of Normalize and the translation and scale sliders given in the Operate menu.

18.6.4 Plotq

- There is no way to remove the Model atoms once displayed, except by re-reading a Map file or restarting Plotq.
- Note that re-reading a map file resets all the default values, which some users have found inconvenient.

18.6.5 Pluto

- The LaserJet output option is untested and therefore not recommended.
- Changes to atom properties are not carried over to symmetry copies of the molecule. This is a restriction of the original Pluto code.
- Atom names: Note that no checks are performed upon the validity of the atom names entered into the list. It is the user's responsibility to ensure that valid atom names are entered - these can be obtained from scrolling list in the **Atom Properties** popup. Failure will cause an error with unpredictable results.

18.6.6 Geometry

This program may have trouble correctly calculating values for structures with a large amount of symmetry. This will in turn cause CIF generation to fail.

18.7 Current limits in maXus

18.7.1 Atoms

- Maximum number of atoms: 500.

18.7.2 Parameters

- Maximum number of refined parameters in full matrix least squares (before blocking): 2,000.
- Maximum number of refined parameters in block-diagonal least squares: 2,000.

18.7.3 Reflections

- Maximum number of reflections in least squares: No limit.
- Maximum number of reflections in Original file, before merging: 400,000.

18.7.4 Fourier

- Maximum number of reflections symmetry expanded in the Fourier program: 150,000.
- Maximum number of grid points along x axis: 200.
- Minimum Dynamic range of the map: 0.001. (The peak search will terminate if the difference between the maximum and minimum peak heights is less than this.)

19 Summary Files,
Tidy and Manuals

A set of utilities for maXus

19.1 Summary Files

Select the *Summary* option from the main maXus menu bar.

Each option in this menu refers to a summary file, which stores the results of the last run of a specified program. These are as follows:

- *CAD4 Import*: see Chapter 3.11
- *CIF Import*: see Chapter 3.7
- *ASCII Import*: results of importing an ASCII file.
- *Spacegroup*: Output data from the Spacegroup program.
- *Test i*: Output from the Test i program.
- *Solve*: The program used to solve the crystal structure. Options are:
  - SIR
  - Mithril
  - Dirdif
  - SHELXS
  - MC Multan
- *SFac*: Output from the Structure Factor program.
- *LSq*: The last run of the least-squares.
- *SHELXL*: This has a submenu offering:
  - Mx2ShelXL and Shelxl2xMx - info on conversion between maXus and SHELXL.
  - SHELXL
- *Fourier*: The last run of the Fourier program.
- *Abs*: Output from the absorption correction program.
- *Geometry*: The calculated bond lengths, angles etc.
- *Platon*: This has a submenu offering:
  - Summary
  - Detail
  - Newman
- *Fo/Fc*: The structure factor tables.
- *Weight Analysis*: The weighting scheme analysis.
- *Report*: The final report file.
- *Ortep*: The output from the ORTEP program. It contains useful information about thermal parameters.
- *Expand*: The output from the process that expands the molecule using symmetry.

If you have not run a particular program, and ask for a summary, or if a particular summary file is missing, an error is generated:

maXus

Summary file does not exist

19.2 The Tidy Facility

The maXus program generates a large number of files. When the structure analysis has finished you may wish to reduce this to a bare minimum, and the tidy facility offers this option to the user. Select the *Tidy* option from the *File* menu.

The *KeepFile* stores all the names of the relevant files to be retained after the Tidy function.The *KeepFile* is to be found in /usr/maXus/lib. A command window now opens and you are asked to choose one of the following options:

- a - delete all the files listed in the command tool.
- i - interactively select files for deletion.
- q - quits with no action

Take careful note of the directory listed at the top of the window before you make your choice - running Tidy on the wrong directory will delete most of the files in it!

The name *KeepFile* is listed as a necessary/useful file as a default.

An example of the contents of the *KeepFile* after using the *a* option above are given below:-

```
KeepFile
Original
Model
DrawFileaa
Sysout_Init
Sysout_Sol
Sysout_Abs
Sysout_Wtf
Sysout_Lsq
Sysout_Pub
Sysout_Wnal
Sysout_Ftab
Sysout_Geom
Sysout_Mopac
Sysout_Rpt
Sysout_Cif
PsiScan
FaceDat
LsqRpt
FftRpt
```

It is possible to re-construct your work from these files, but it will not be possible to carry on from where you left off.

If necessary, you can manually edit the KeepFile to add the names of other files you wish to keep.

19.3 Help

19.3.1 On-Line Manuals

maXus provides a full set of on-line manuals, in PDF format. A PDF viewer (Adobe Acrobat 3.0) is supplied with the program. Choose the *Help* menu from the main window, and select *View Manual*.

Select the chapter you require from the scrolling list on the left; it is displayed in the large window on the right. They can be browsed page by page, and printed if required. They are hyperlinked, so clicking on a section heading in the Table of Contents, or a page number in the Index, takes you directly to the relevant page.

19.3.2 Tutorial

Selecting *View Tutorial* from the *Help* menu brings up a window that displays a basic tutorial for maXus, in PDF format.This has a few different molecules to work through to solve using maXus.

19.3.3 Release Notes

Select the *Release Notes* submenu from the main *Help* menu. It lists options for the various versions of maXus. Selecting one brings up a window displaying notes on that particular version.

Index