Sir2014 - on line help

Copyright © 2014,Institure of Crystallography, CNR - Italy

Table of contents

General information 3
Notes on distribution 6
Installing the program 7
Getting Started 9
Description of Sir2014 27
  Data module 27
  Invariants module 31
  Phase module 31
  Simulated Annealing 37
  Molecular Replacement 50
  Substructure determination 56
Sir2014 strategy for ab initio phasing of crystal structures 57
Completion and refinement of the structure 58
When default Sir2014 fails: some advices 59
The Graphic Interface 60
  File 61
  Modify 75
  Settings 78
  View 85
  Solve 89
  Refine 95
  Info 111
  Help 117
Commands and their use 118
Directives 119
Examples 131
References 136

General information

Sir2014

Sir2014 is the latest product of the Sir family

Sir2014 is available as a stand-alone program or included in the package ***Il Milione*** (Burla *et al*., 2007).

The **SIR** (SEMI-INVARIANTS REPRESENTATION) package was originally developed for *ab initio* crystal structures solution by Direct Methods (DM). 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.

Sir2014 is designed for the automatic *ab-initio* and non *ab-initio* crystal structure solution of small, medium and macromolecules according to the following scheme:

Ab initio phasing approach:

- Modern Direct Methods (MDM)
- Standard Direct Methods (SDM)
- Vive la Difference (VLD)
- Patterson techniques

Defaults:

MDM is the default choice for small (up to 80 non-hydrogen atoms in asymmetric unit) and medium size structures (up to 300 non-hydrogen atoms in asymmetric unit).
SDM is the default choice for electron diffraction data.
For proteins, if heavy atoms are contained in the unit cell, the default choice is the Patterson approach.

Non ab-initio phasing approach:

- Simulated Annealing
- Molecular Replacement

It is also possible to locate the anomalous scatterers or the heavy atoms when SAD/MAD data or SIR/MIR data are available. Data can be collected with X-Ray, electron or neutron sources. There is no limit to the number of reflections and to the number of atoms in the asymmetric unit. The maximum value allowed for |h|, |k|, |l| is 512. The maximum number of different atomic species is 18.

The program has been designed to:

- require a minimal information in input;

- work automatically;
- reduce the user intervention and facilitate the interaction by means of a friendly graphic interface.

Expert crystallographers will have a wide range of options to solve their crystal structures.

What's new

New features have been added to Sir2014 with respect to the previous version (Sir2011). Among them we quote:

- **Modern Direct Method (MDM) procedure**
- **Molecular Replacement pipeline**
- **Patterson approach pipeline**

A new program named Jav has been realized. It is not only a visualizer for models and maps, but it can be a useful tool to edit the model. It can be used or as an external program started from the main toolbar of Sir2014 or as an alternative to the classical visualization of the model by Sir2014. In the first case also the electron density map is drawn, in the second case the model can be modified and updated in Sir2014.

Moreover It is now possible to start Sir2014 using a *.ins* file under the following conditions:

- the *.ins* file must contain at least the keywords: CELL, UNIT, LATT, SFAC (only atomic species), HKLF;
- in the same directory of the *.ins* file the *.hkl* file must exist with the same name.

The corresponding *.sir* file is automatically created.

The program

The program uses some CCP4 (Collaborative Computer Project 4, Daresbury Laboratory, UK) libraries in accordance with the CCP4 License Agreement (clause 2.1.2) and it is linked to external programs to face with protein structure solution (Refmac, Buccaneer, Nautilus, ARP/wARP, Phenix). The program uses FFTw (Frigo and Johnson, 2005) and GTK libraries which are distributed under the GPL (GNU General Public License).

The program is available for Microsoft Windows, Linux and Mac platforms (see the Installing the program section).

Authors:

M.C. Burla(1,2), R. Caliandro(1), B. Carrozzini(1), G.L. Cascarano(1), C. Cuocci(1), C. Giacovazzo(1), M. Mallamo, A. Mazzone(1), G. Polidori(1)
(1) Istituto di Cristallografia, CNR, Bari, Italy.
(2) Dipartimento di Fisica e Geologia, Università di Perugia, Perugia, Italy.

Support:

E-mail: sirmail@ic.cnr.it

References:

Please refer to:

Maria Cristina Burla, Rocco Caliandro, Benedetta Carrozzini, Giovanni Luca Cascarano, Corrado Cuocci, Carmelo Giacovazzo, Mariarosaria Mallamo, Annamaria Mazzone and Giampiero Polidori (2014),

Crystal structure determination and refinement *via* Sir2014

J. App. Cryst. submitted.

Created with the Personal Edition of HelpNDoc: Full-featured Kindle eBooks generator

Notes on distribution

Depending on the platform required (MS Windows, Linux, Mac) different files are

supplied, **Jav** (**J**ust **a**nother **v**iewer) is always included:

- Windows: the program files are distributed as exe file Sir2014SetUp-XX.exe (or IlMilioneSetup-XX.exe, **Sir2014** is included);

- Unix / Linux / Mac: the program files are distributed as gzipped file sir2014-XX.tar.gz (or ilmilione-XX.tar.gz);

- for some linux distributions pre-compiled version are available (.rpm, .deb ...);

- for Mac systems pre-compiled version is available (.dmg).

Since source files are supplied, the program can be compiled and linked. In order to use the graphical interface some public domain libraries need to be linked (details on libraries are available in README file included in tar file). If a Fortran 90 compiler is not available, *gfortran* can be freely downloaded from http://www.gfortran.org/

Created with the Personal Edition of HelpNDoc: Easily create EPub books

Installing the program

On PC-Windows platforms

Refer to Sir2014Readme.txt file for updated information

- Run Sir2014SetUp-XX.exe
- Run the program by clicking on the Sir2014 (or Jav) icon

Sir2014 icon

Jav icon

On Linux / Unix / MacOS platforms

Refer to README file for updated information

- To install Sir2014 included in ***ilmilione*** you need to execute the following commands:

    1. tar xvfz ilmilione-XX.tar.gz
    2. cd ilmilione-XX
    3. ./configure --prefix=***destination_directory*** --srcdir=***ilmilione_source_directory***
    4. make clean
    5. make install
    6. ilmilione setup

- To install Sir2014 (stand-alone) you need to execute the following commands

    7. tar xvfz sir2014-XX.tar.gz
    8. cd sir2014-XX
    9. ./configure --prefix=***destination_directory*** --srcdir=***sir2014_source_directory***
    10. make clean
    11. make install

    Executable files will be stored in ***destination_directory***/bin
    Be sure that the ***destination_directory***/bin is in the **PATH**

- To start the program

    Click on Sir2014 icon or start sir2014 in a terminal window

Configure process

The configure procedure will use the first Fortran compiler found in the following list:

ifort gfortran g95 lf95 pgf90 pgf95 ifc efc gfort xlf95 xlf f95 f90

It is possible to modify the selection adding to the configure command the option

FC=Fortran_compiler_name

If no Fortran90/95 compiler is installed, it is possible to download the GFortran compiler

If no GTK libraries are installed, it is possible to download them from Internet

Some examples:

Default
./configure --prefix=$HOME --srcdir=$PWD

gfortran
./configure --prefix=$HOME --srcdir=$PWD FC=gfortran

Intel Fortran
./configure --prefix=$HOME --srcdir=$PWD FC=ifort

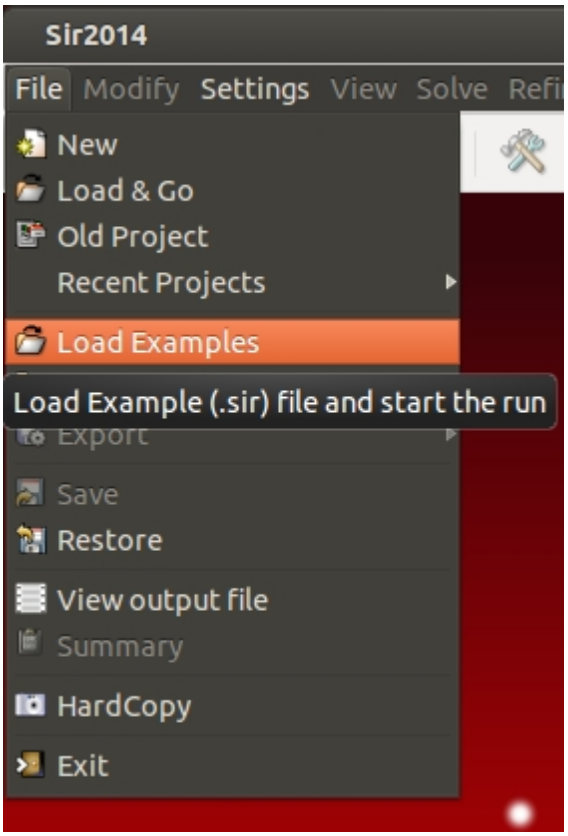
NAG
./configure --prefix=$HOME --srcdir=$PWD FC=f95

Lahey
./configure --prefix=$HOME --srcdir=$PWD FC=lf95

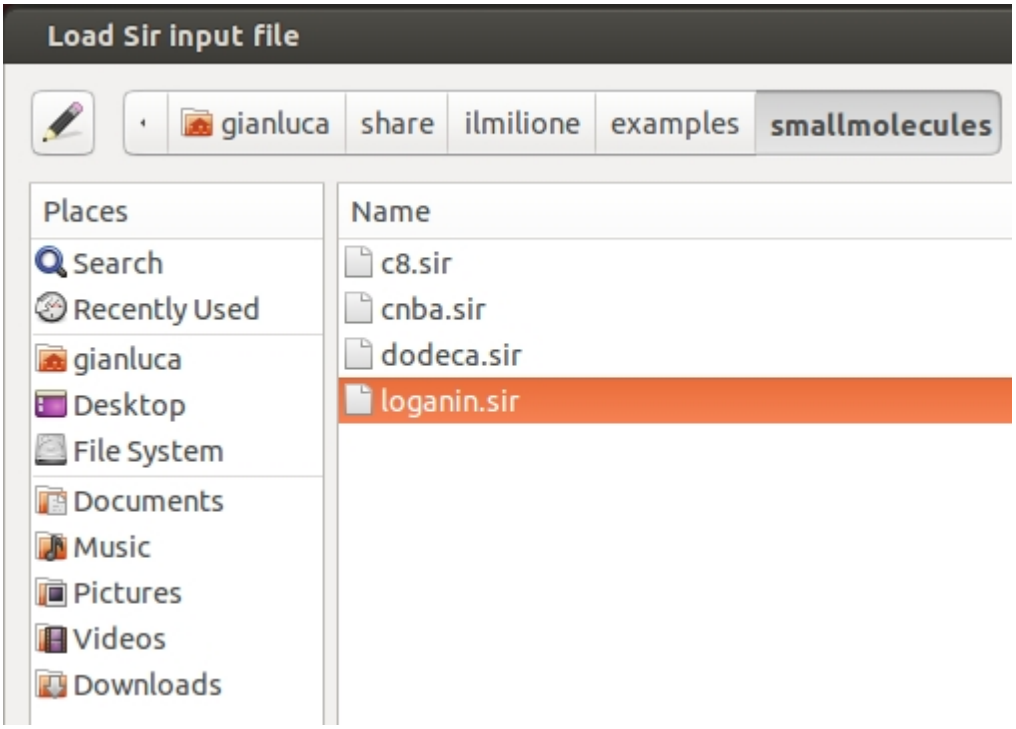
Created with the Personal Edition of HelpNDoc: Easy to use tool to create HTML Help files and Help web sites

Getting Started

Let's use, as an example, a simple small molecule (loganin) supplied with the program. To load it:



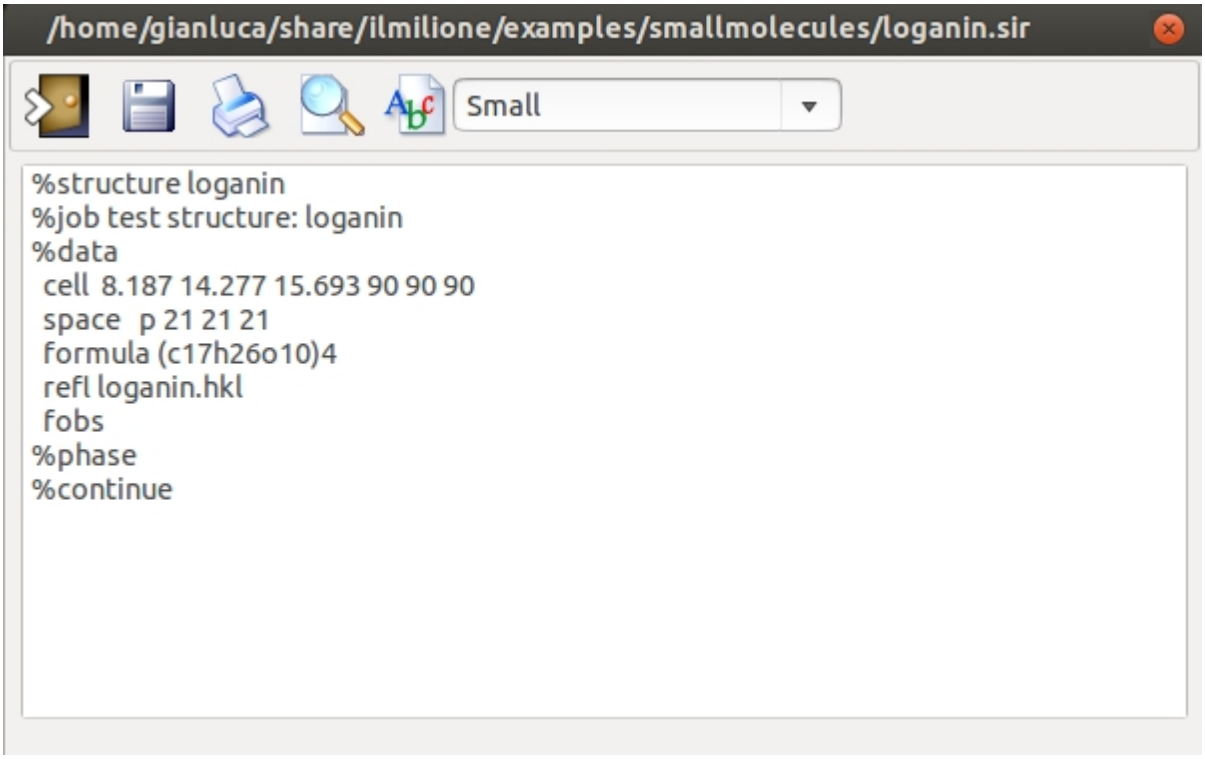
Select the input file for the test structure: **loganin.sir**



It is possible to visualize and to edit this file using the

button.

The file loganin.sir contains the minimal input instructions to make Sir2014 run. Only few commands and directives are mandatory: pay attention to structural info (i.e. cell parameters, space group, ...), chemical composition (i.e. chemical content in the unit cell, ...) and experimental data (i.e. reflection file, type of data - structure factors or intensities, ...).

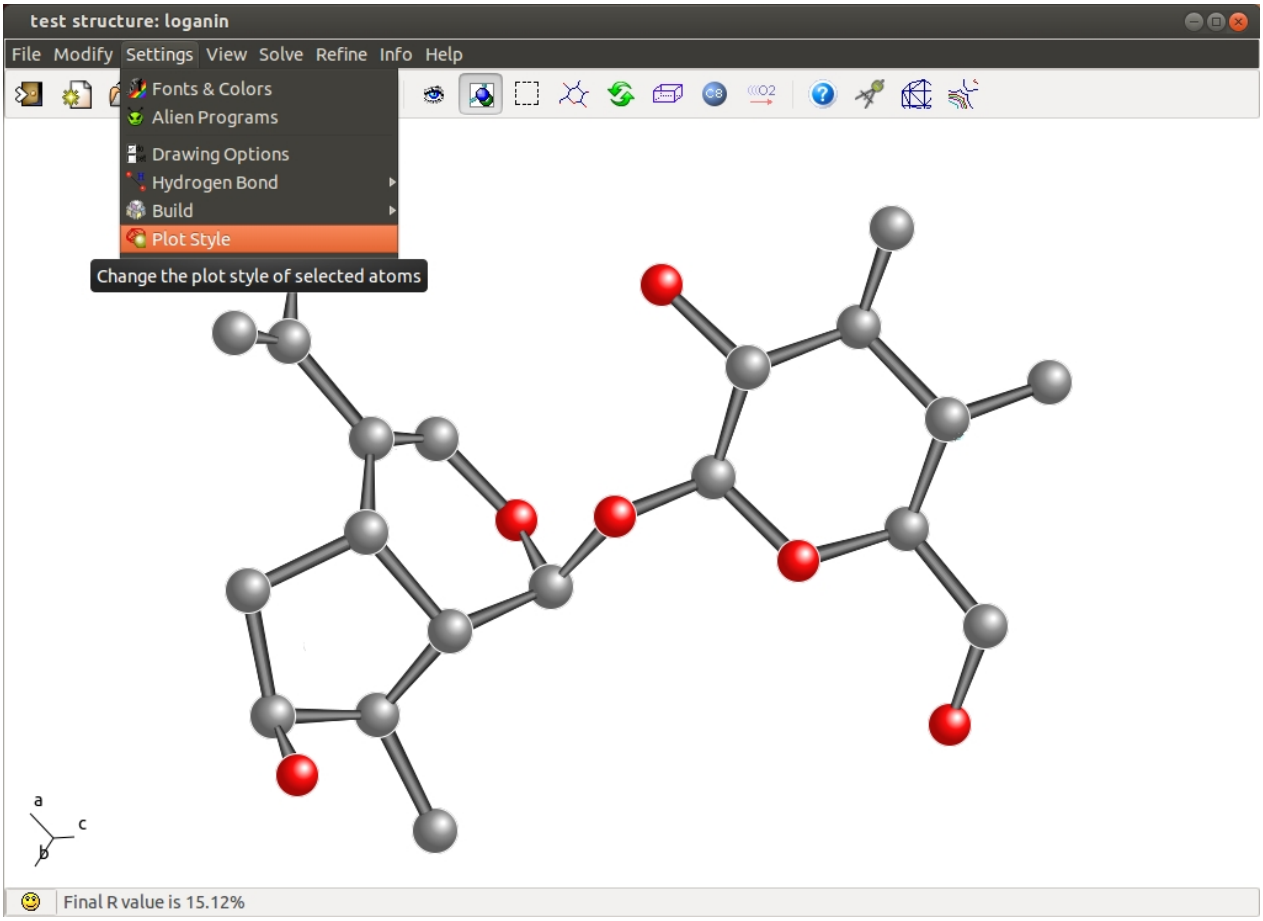


To start the program, click on

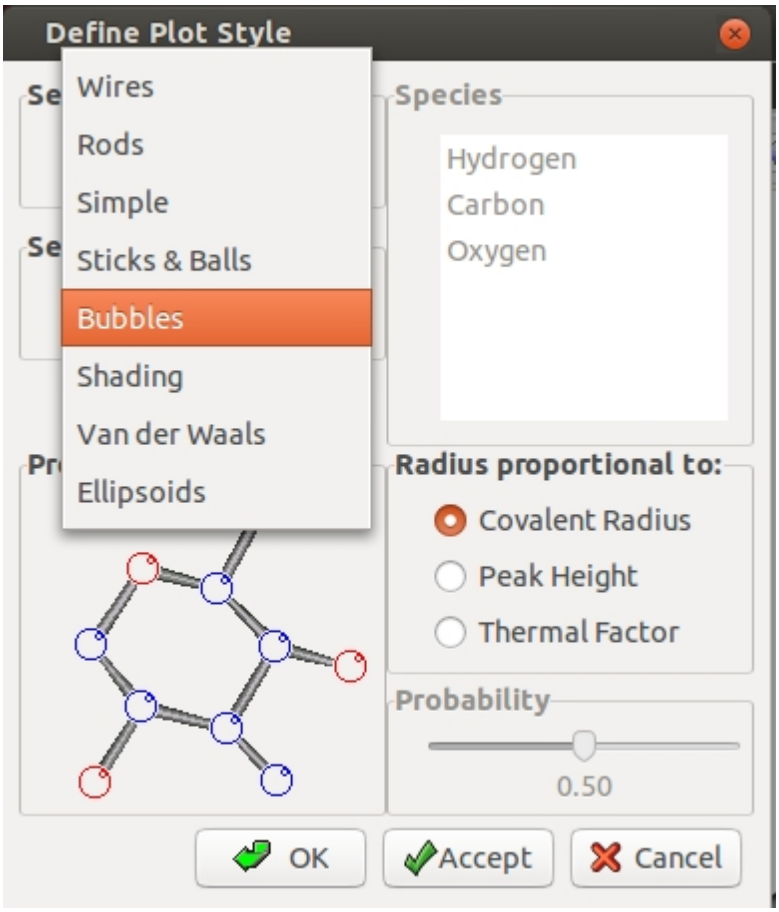
button.

At the end of the run, Sir2014 shows the "best" model of the structure; the final Rf% = 15.12 is quoted in the status bar (at the bottom of the program window).

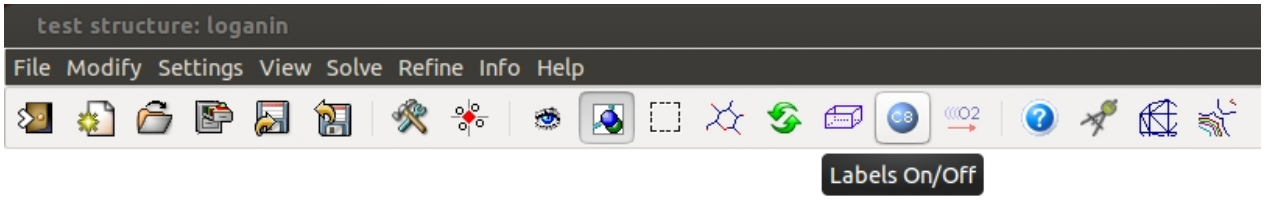
To modify the style go to menu item “**Settings**” on the main toolbar and select “Plot Style”option.



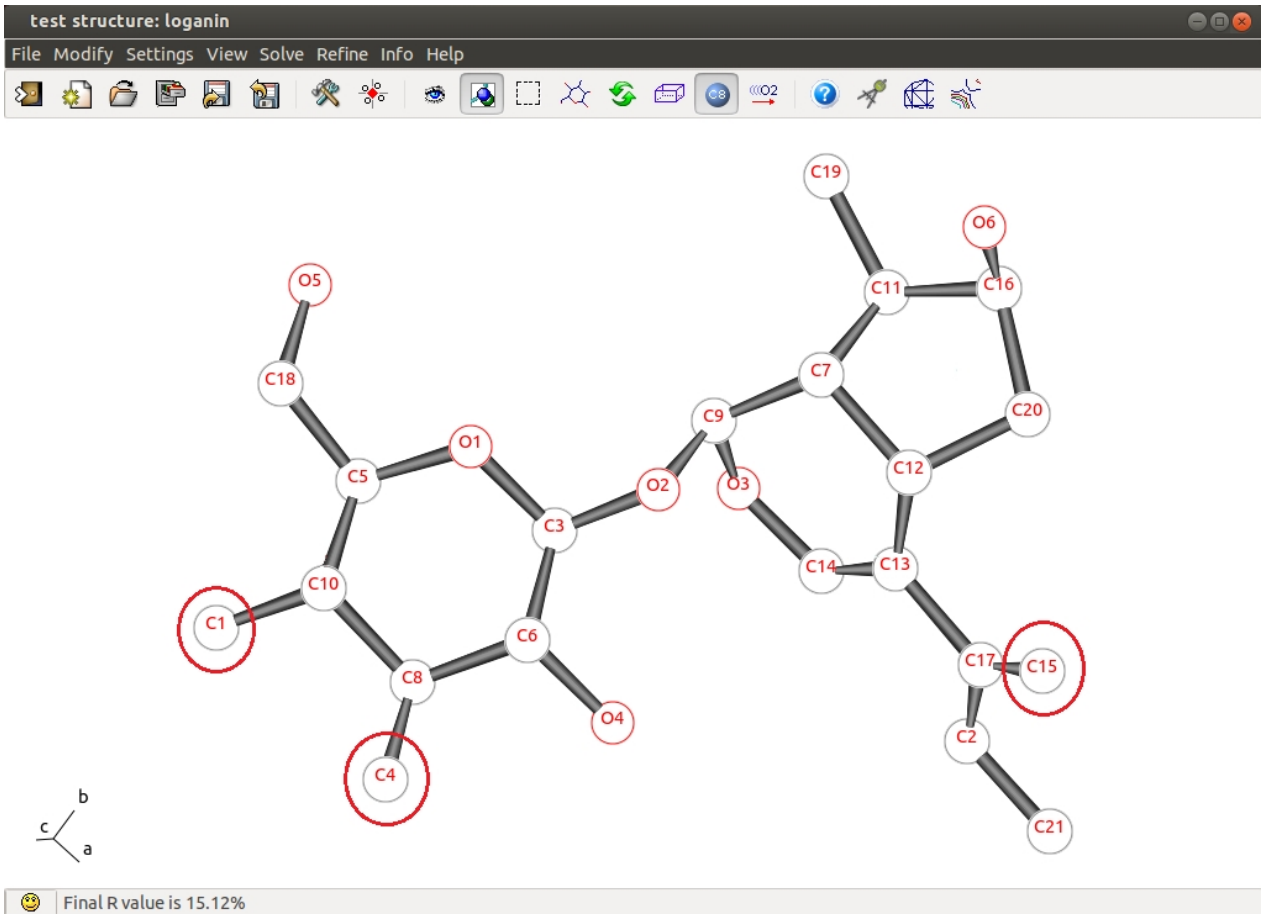
Change the plot style for the model (i.e. Bubbles) and accept it (click on Accept or OK buttons).



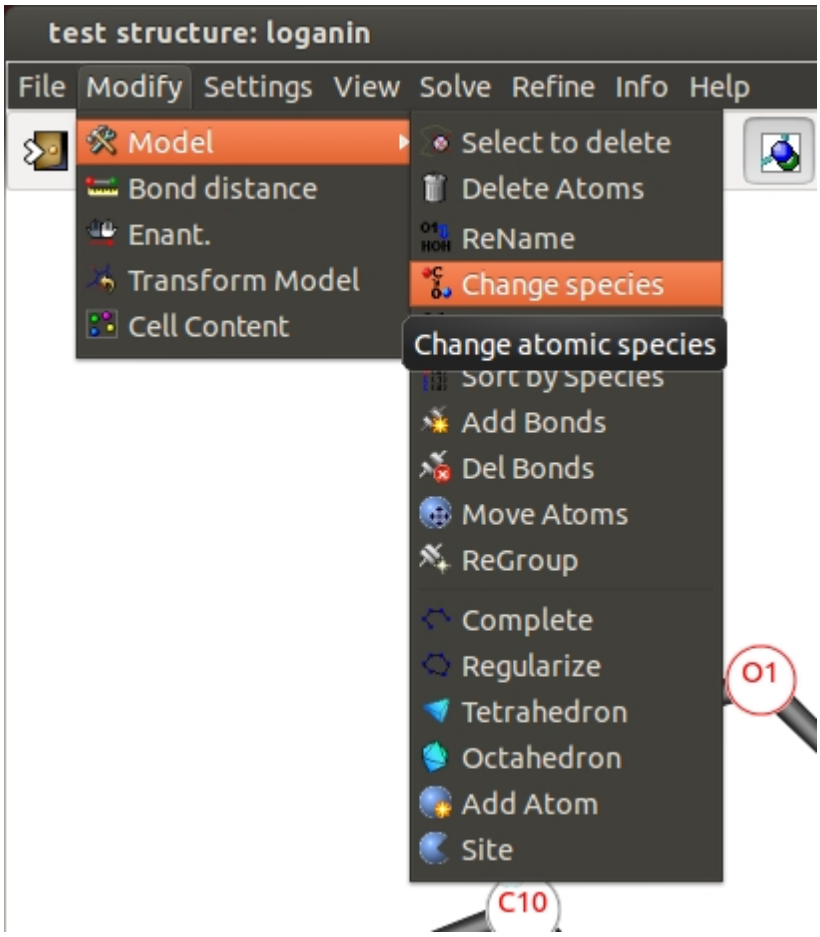
Add the labels (chemical species and progressive numbers) to the atoms of the model clicking on Labels button:



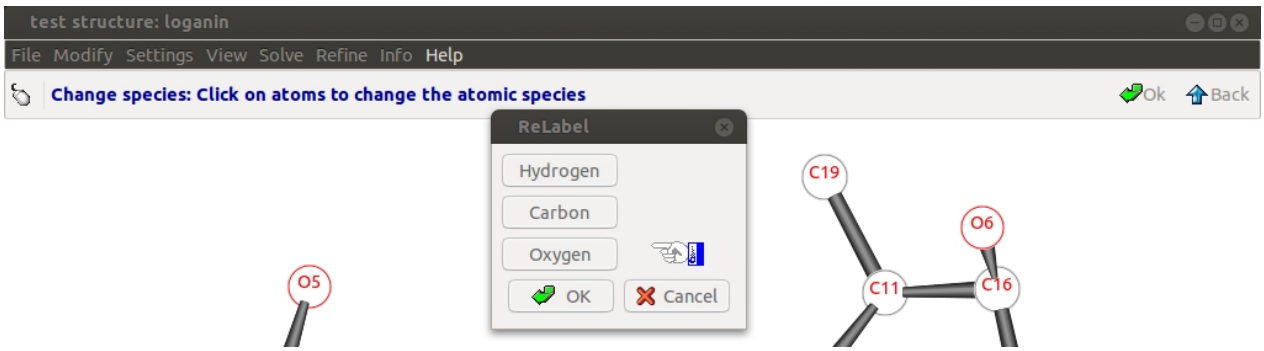
The new plot is the following:



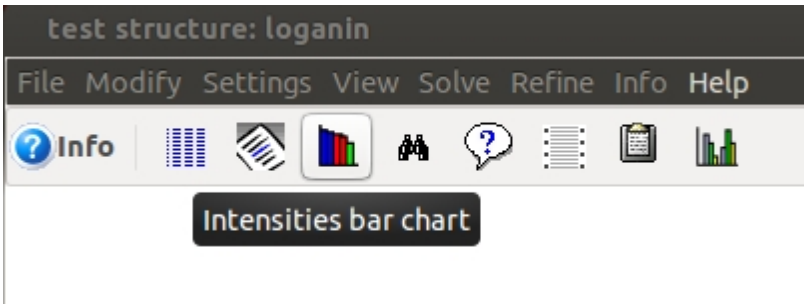
Atoms C1, C4 and C15 (inside red circles) show wrong labelling, they should be oxigens. Change the labels for C1, C4 and C15: go to menu item "**Modify**" on the main toolbar; select Model option and then Change species option.

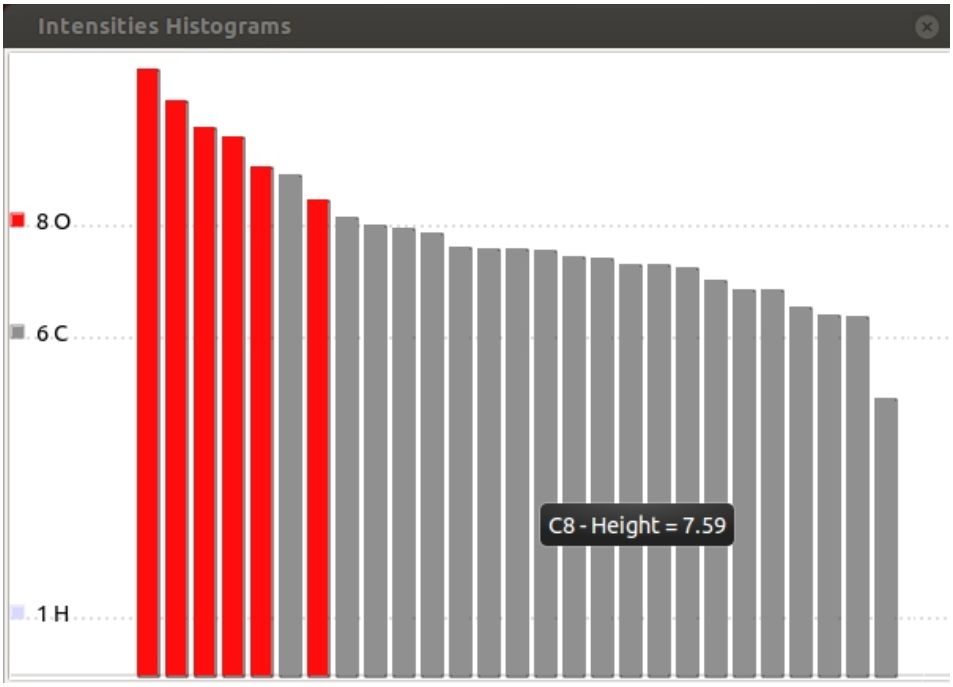


Select the atoms to change the atomic species and choose the correct chemical element (click on the related button).

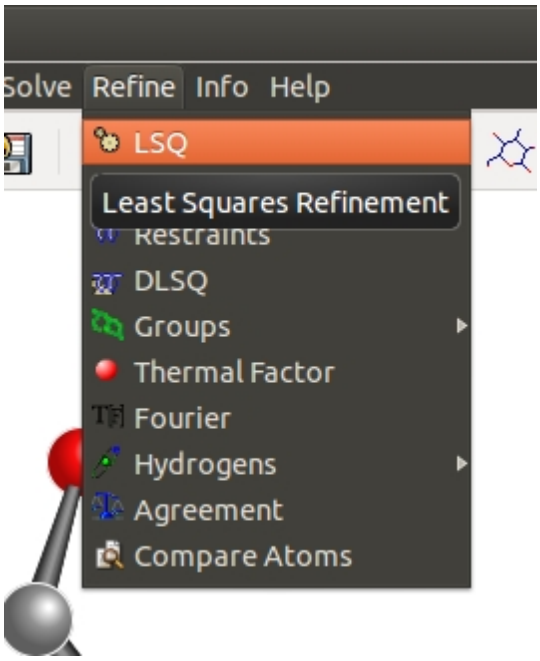


Click on Intensity bar chart chemical species button to show the intensity histogram for the atoms of the model and the associated atomic species.





Start the refining process: go to menu item “**Refine**” on the main toolbar and select LSQ option: the Least Square Refinement Menu will be opened.

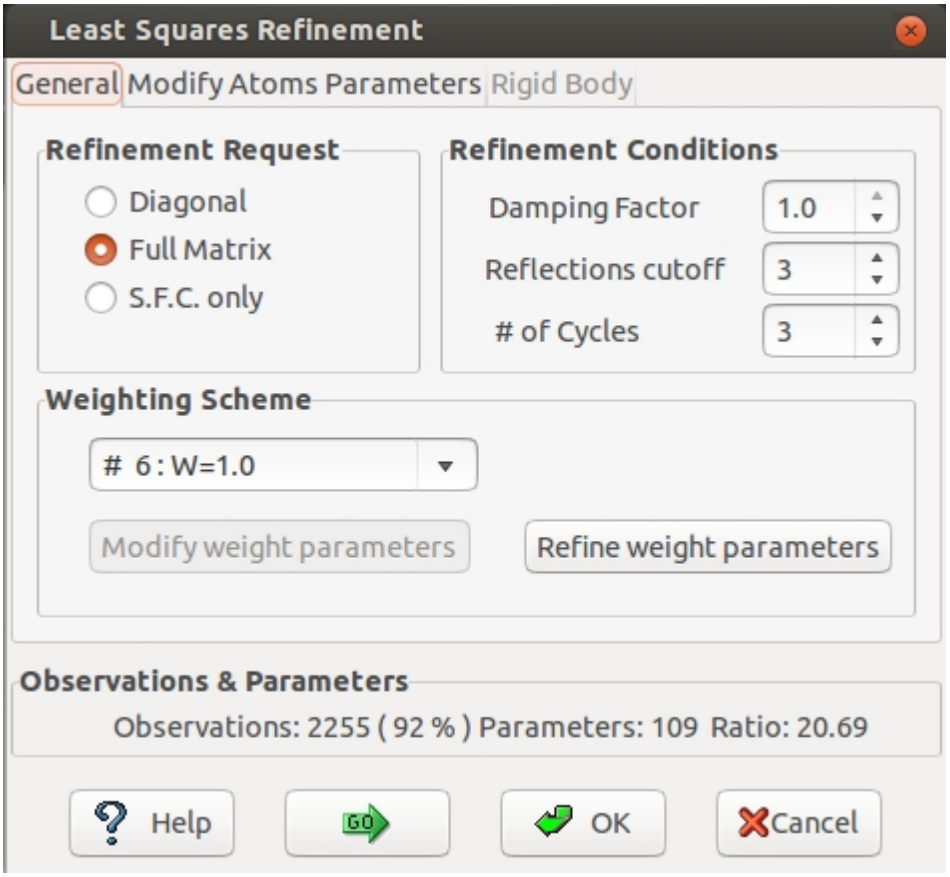


Go to LSQ sub-menu item "**General**" and set the LSQ options (i.e. Full Matrix, sigma(F_o) cutoff, number of LSQ cycles, weighting scheme, ...).

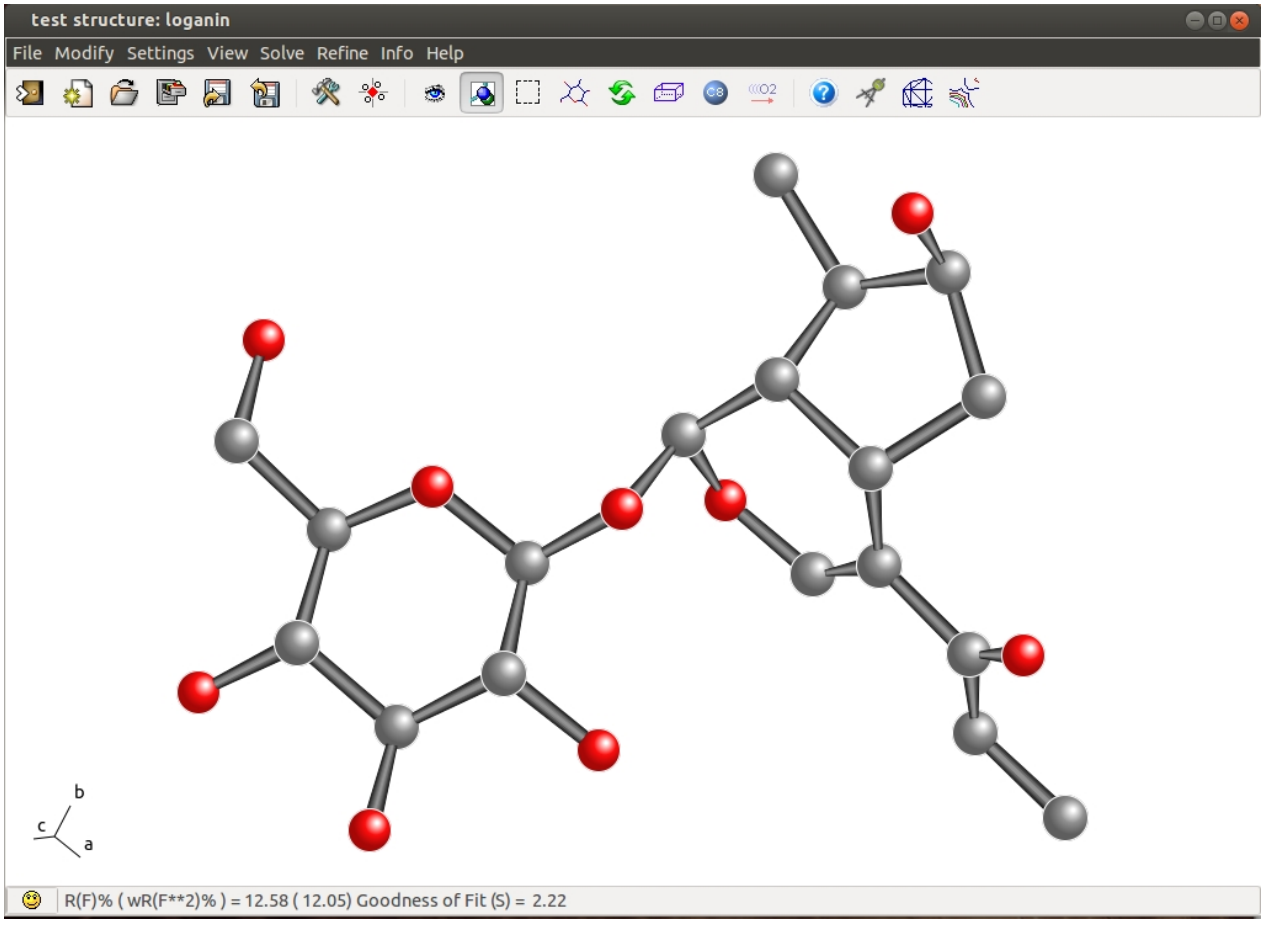
Be careful to the ratio Observation / (number of refining) Parameters. For this example, the ratio value (20.69) is high enough to lead to a reliable LSQ refinement; otherwise if this value is lower than 10, the refining process could be meaningless.

Accept the choices clicking on button; the LSQ refinement will start.

The first step (after the changes of atomic species when necessary), is an isotropic LSQ refinement for all the model atoms (only coordinates and isotropic thermal factor will be refined: four parameters per atom) plus the scale factor.

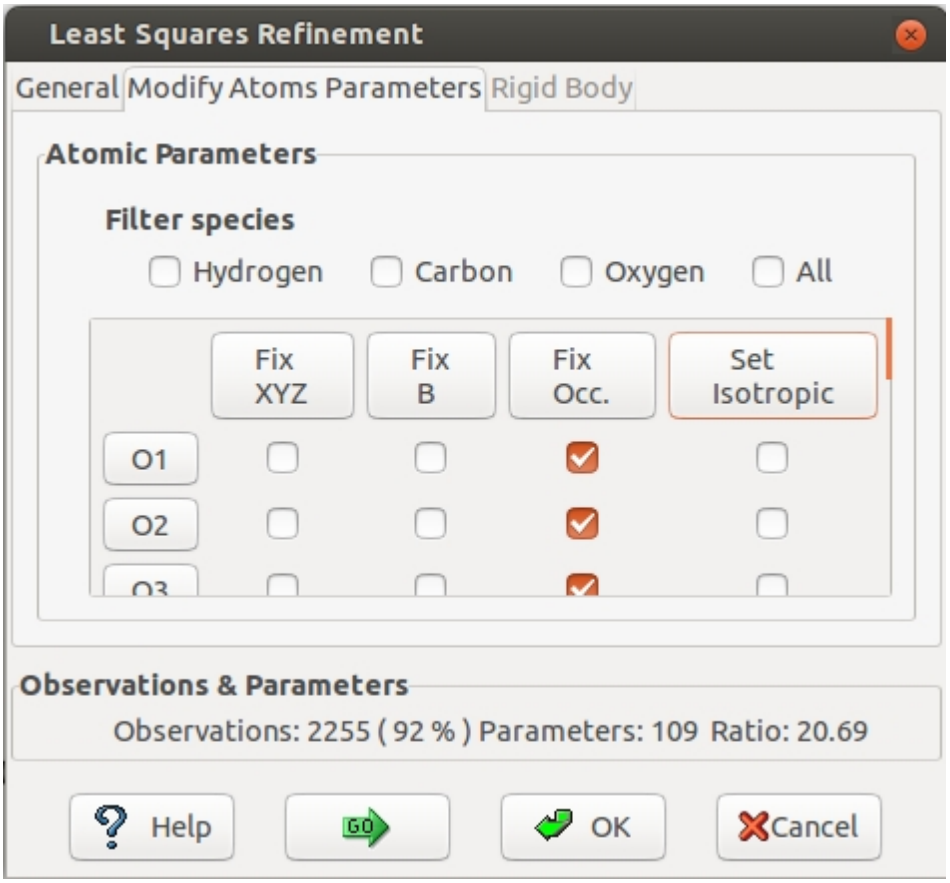


Updated Rf% value can be found in the status bar (Rf% = 12.58). The new Rf% value is lower than the previous one (Rf% = 15.12). This mean that modified labeling is correct; otherwise the new Rf% value should be larger than the previous one.

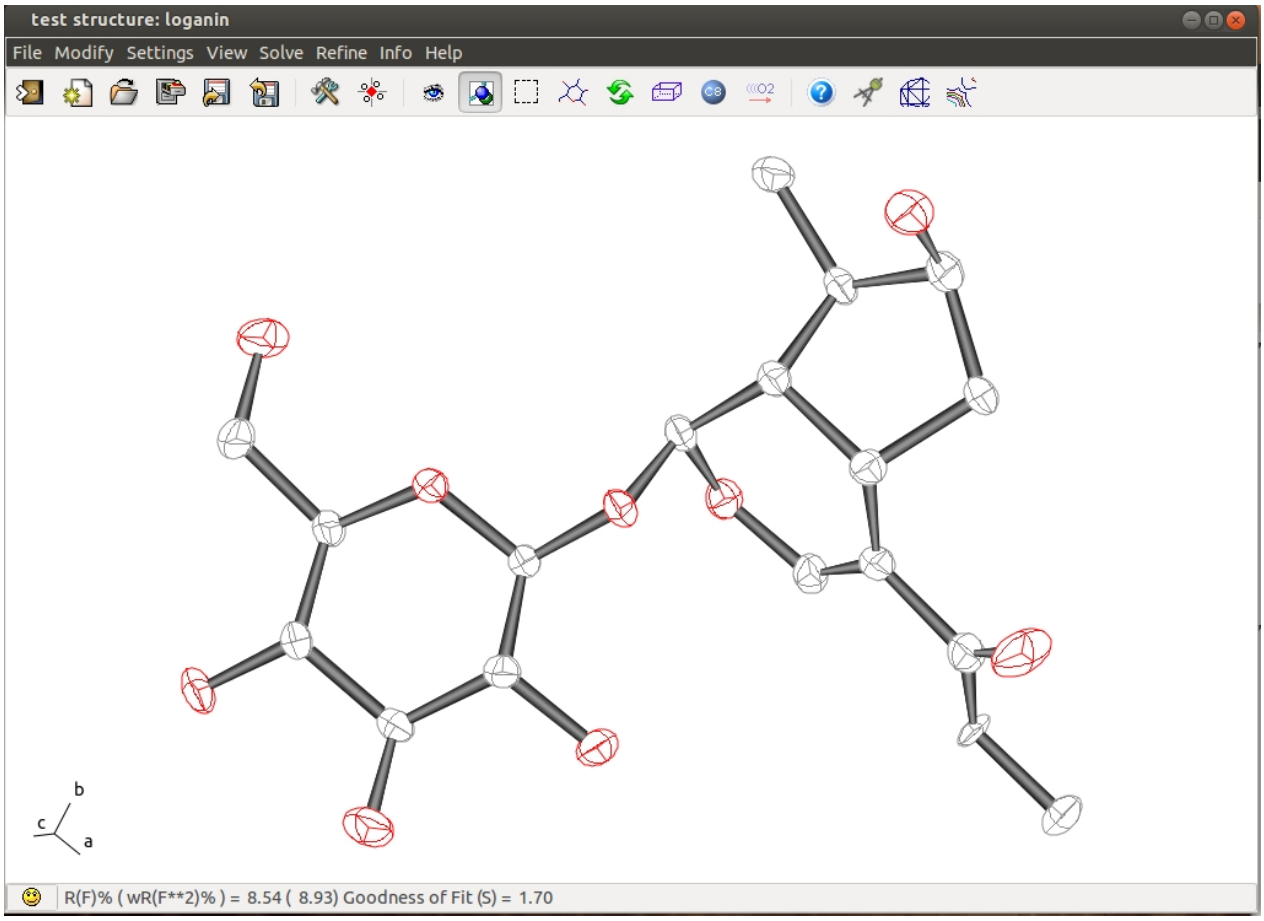


To improve the reliability of the atomic model, we have to apply a thermal anisotropic LSQ refinement. If necessary it is possible to fix some specific atomic parameters.

Go to LSQ sub-menu item “**Modify Atoms Parameters**”; choose the filter (i.e. All) and select the option (i.e Set Isotropic: No). Atomic coordinates and anisotropic thermal parameters will be refined: nine parameters per atom).

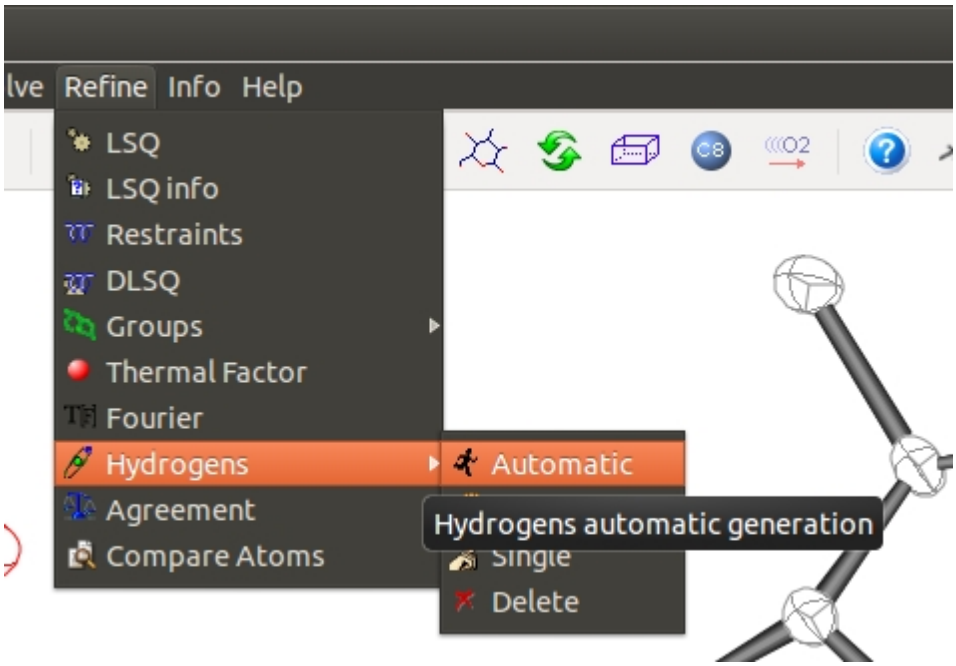


Change the plot style for the model: go to menu item "**Settings**" and select "Plot Style"option: select Radius proportional to Thermal Factor (atom thermal ellipsoids will be plotted, proportional to anisotropic thermal parameters). Updated Rf% value can be found in the status bar (Rf% = 8.54). The Rf% value is clearly lower than the previous one

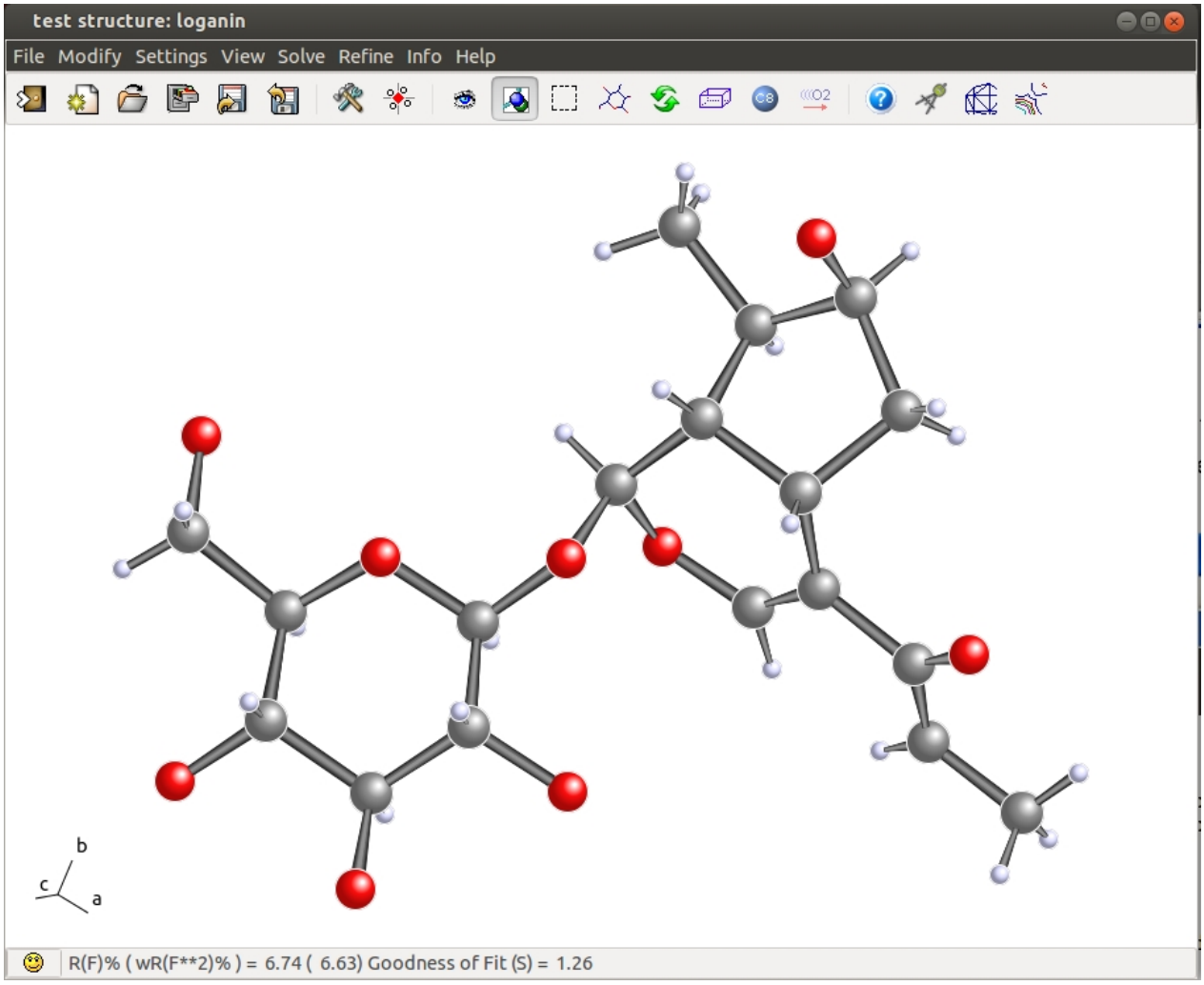


Complete the model by adding hydrogen atoms.

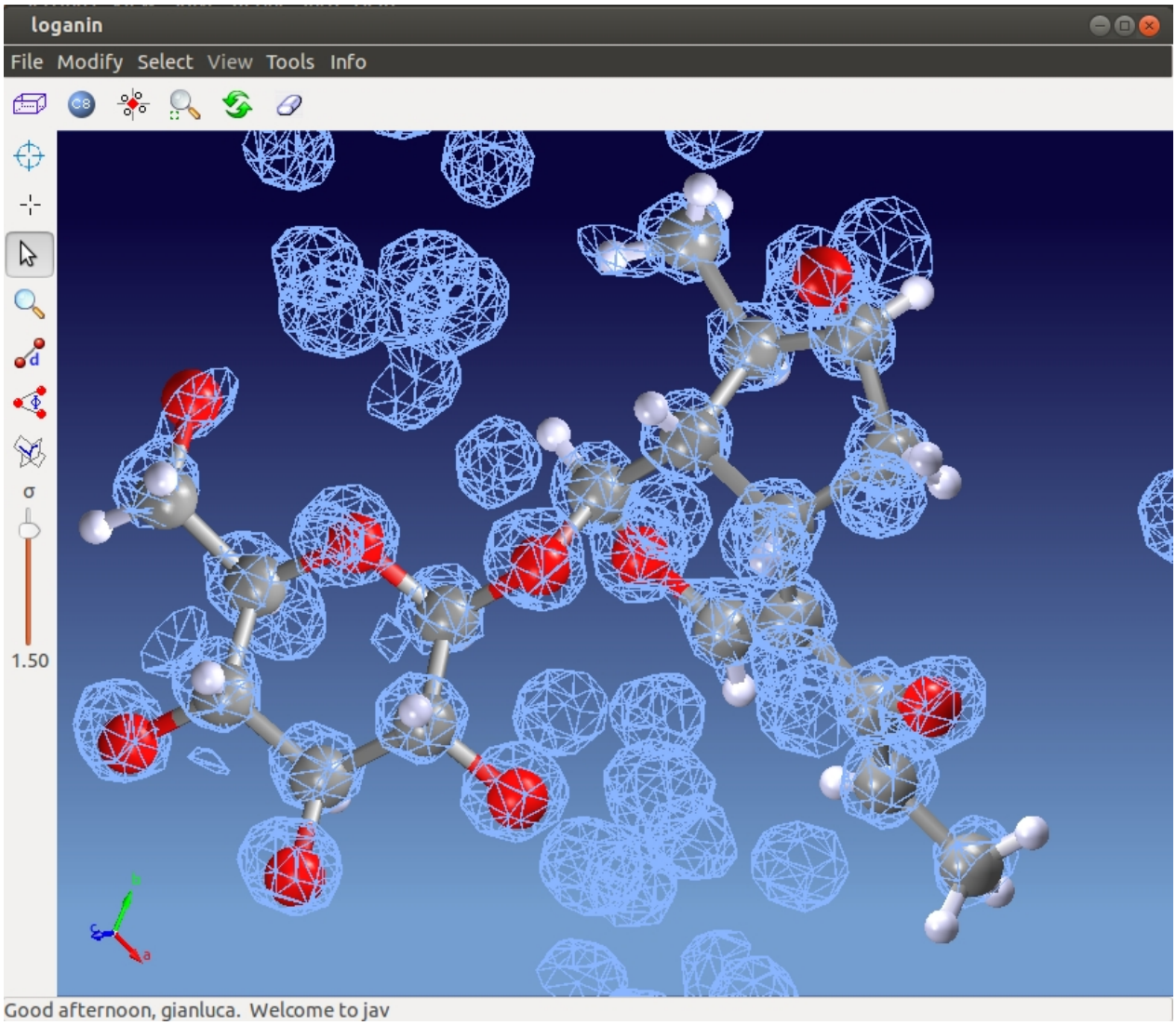
Go to menu item “**Refine**”, select Hydrogens option and then Automatic option (hydrogens automatic generation).



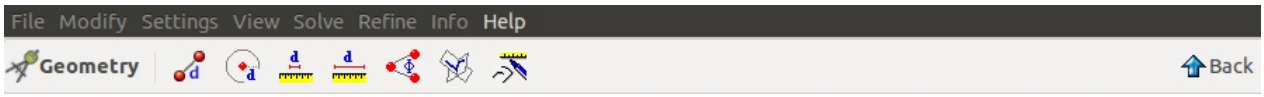
The next step is to start a new LSQ refiniment. Updated Rf% value is shown in the status bar (Rf% = 6.74). This Rf% value is low enough to suggest that the atomic model is correct.



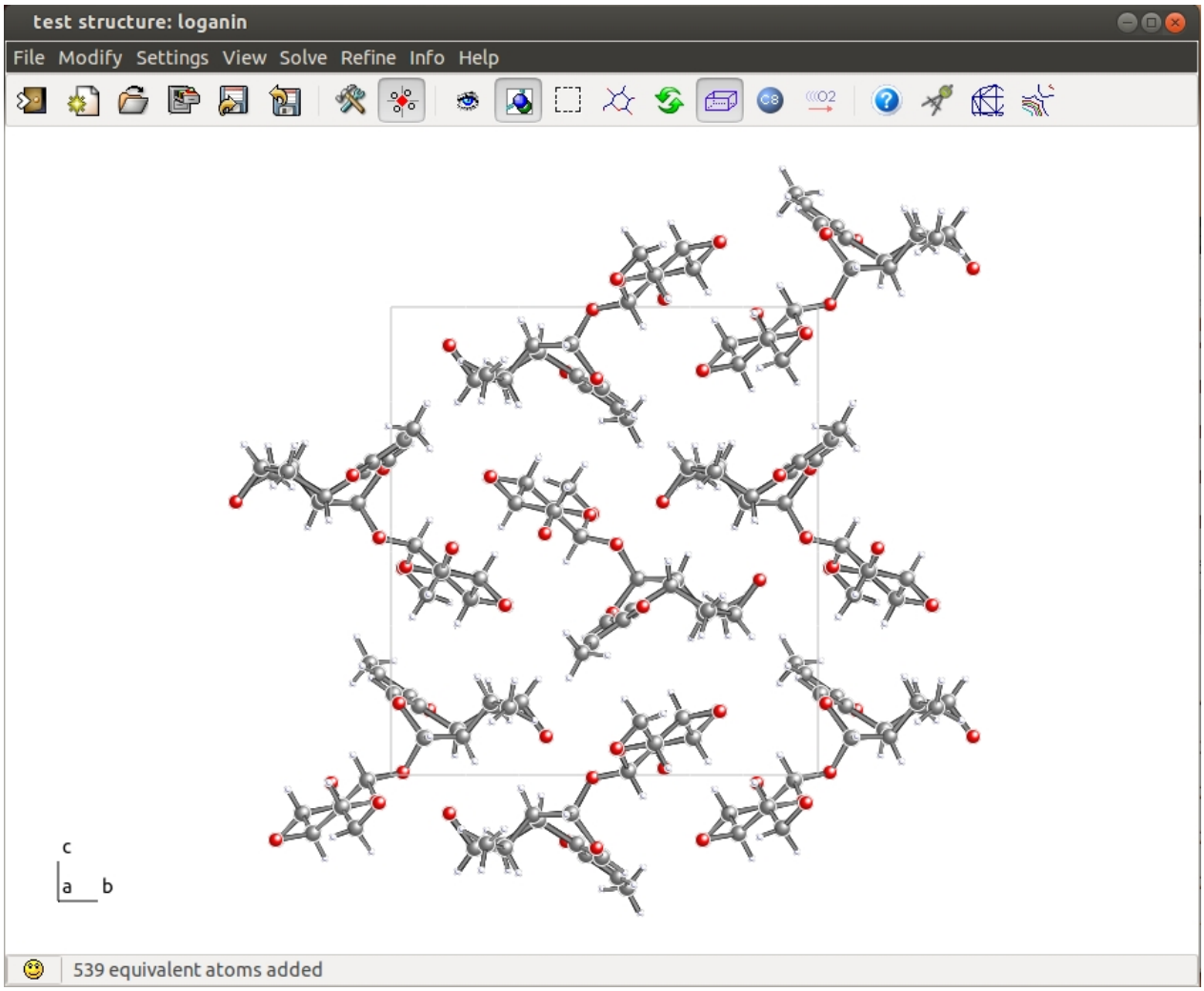
Using **Jav**, it is possible to visualize the electron density map and the related model; click on the button in the main toolbar:



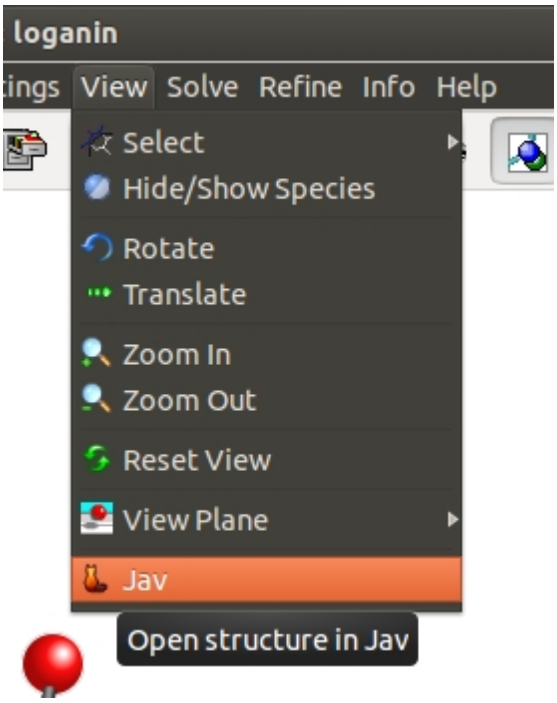
It is possible to compute some quantities (distances, angles, torsion angles, ...) trough the "**Geometry**" menu; access to it, clicking on the button of the main toolbar:

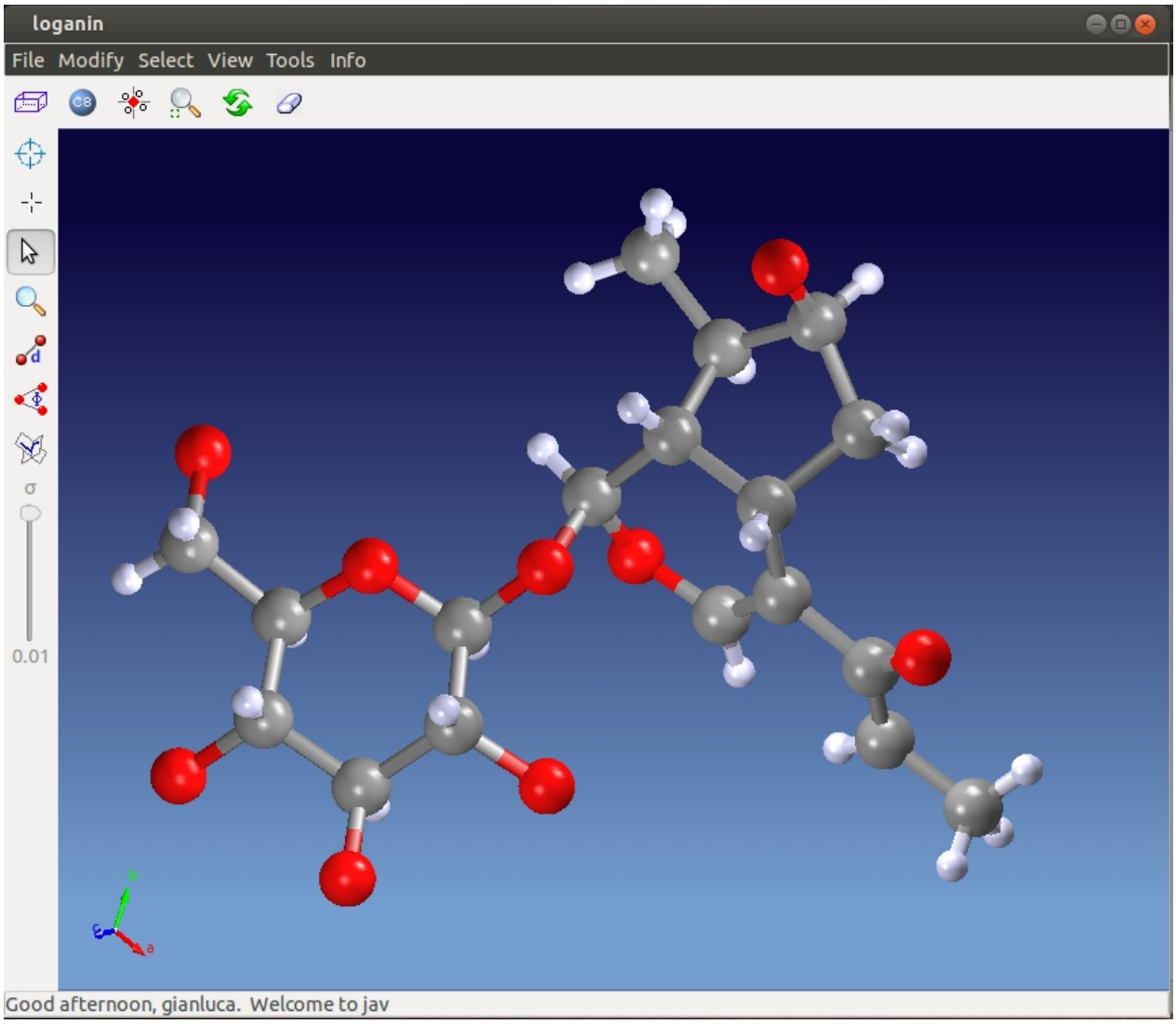


Apply the symmetry using the button on the toolbar; different views are possible using the button.



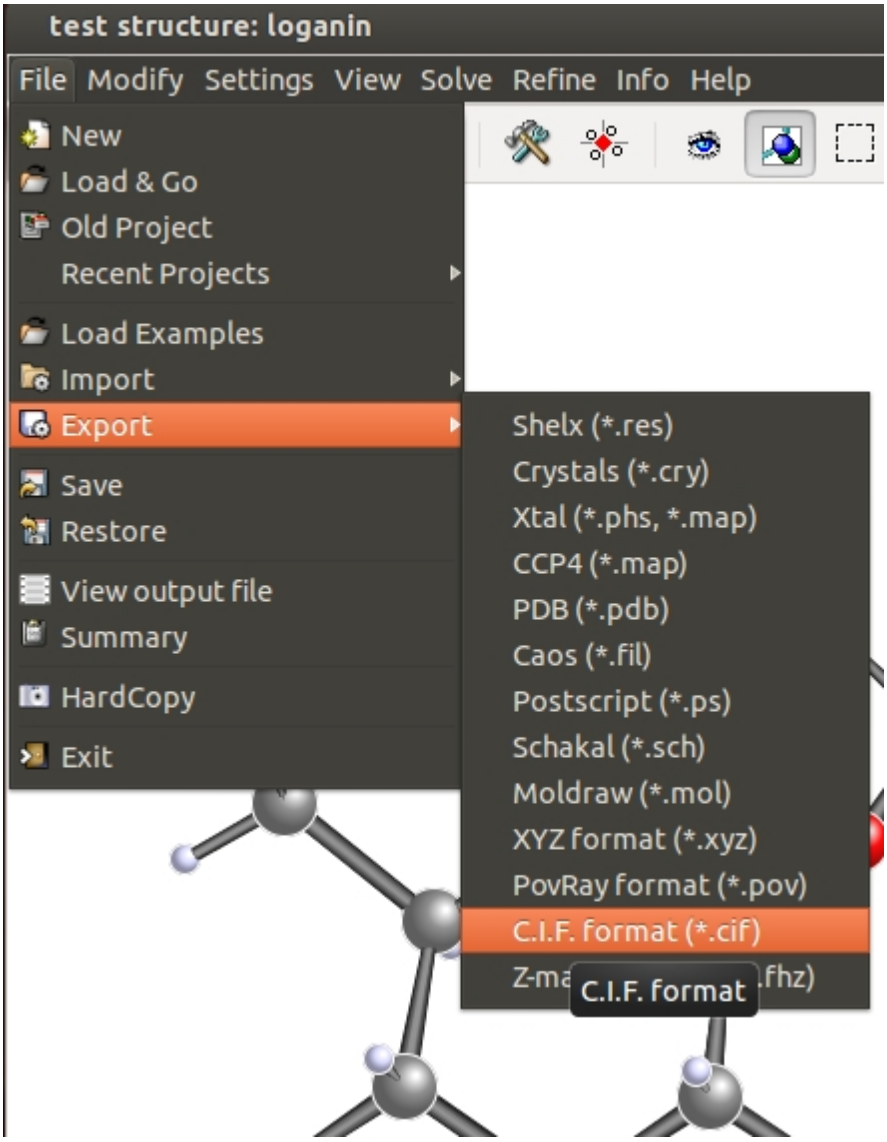
Jav can be used also as an integrated interface to study the model and to modify it:


loganin
ings
View
Solve
Refine
Info
Help
Select
Hide/Show Species
Rotate
Translate
Zoom In
Zoom Out
Reset View
View Plane
Jav
Open structure in Jav


loganin
File
Modify
Select
View
Tools
Info
0.01
Good afternoon, gianluca. Welcome to jav

It is possible to export data and results in order to use them in another program; in particular, when the refinement is complete and the structure is ready for publication, it is possible to export data and results in CIF format.

Go to menu item “**File**” on the main toolbar, select Export option and then CIF option (other formats are available: i.e. PDB format, for proteins).



Created with the Personal Edition of HelpNDoc: Generate EPub eBooks with ease

Description of Sir2014

The main modules of the program are: DATA, INVARIANTS (if MDM or SDM is required), PHASE, Simulated Annealing and Molecular Replacement (REMO09).
If a graphic interface is active, it is possible to interact with the model in order to complete and refine it. This option is available only if the number of atoms in the asymmetric unit does not exceed 500.

The graphic interface incorporates an online help in order to describe all the features and tools available via graphics. The graphical user interface has been created by M. Mallamo.

Thanks are due to J. Gonzalez-Platas (2D Contouring tools) for the code integrated in Sir2014.

Created with the Personal Edition of HelpNDoc: Single source CHM, PDF, DOC and HTML Help creation

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 (Burzlaff & Hountas, 1982). Symmetry operators and information necessary to identify structure invariants (estimated in INVARIANTS module) are directly derived from the space group symbol.

Diffraction data are checked in order to merge equivalent reflections, to find out systematically absent reflections (which are then excluded from the data set) and, eventually, (weak) reflections not included in the data set (Cascarano *et al*., 1991).
Diffraction intensities are normalized using the Wilson Method (1942). Statistical analysis of intensities is made in order to check the space group correctness, to suggest the presence or absence of the inversion centre and to identify the possible presence and type of pseudotranslational symmetry (Cascarano *et al*., 1988 a,b; Fan *et al*., 1988). Possible deviations (of displacive type) from ideal pseudotranslational symmetry are also detected. Up to 4000 large reflections (those with the largest $|E_o|$ values) are selected for invariants

calculations; their number (Nlarge) is automatically settled by the program.

When the space group is ambiguous it is possible to derive the Laue class, the extinction group and the space group symbol through the statistical analysis of the observed structure factors.

Procedure to select the space group

An algorithm (Camalli *et al.,* 2011) has been developed for the automatic identification of the Laue group and of the extinction symbol from diffraction intensities, suitably specialized in case of electron data.

The algorithm has a statistical basis, and tries to face the severe problems arising from the often non-kinematical nature of the diffraction intensities, from the limited accuracy of the lattice parameters determined *via* electron diffraction and from the limited amount of measured intensities.

Since the accepted unit cell may show a lattice symmetry higher than the true one, all the crystal systems with unit cells compatible with that experimentally estimated (we will call them *feasible systems*) are taken into considerations for the next steps. For example, in case of a cell with cubic geometry, tetragonal, orthorhombic, monoclinic and triclinic systems are considered *feasible*, hexagonal and trigonal are excluded.

Than a list of the Laue groups compatible with each feasible crystal system is automatically generated (denoted as *admitted* Laue groups). For each feasible crystal system the admitted Laue group with minimal symmetry (say *LGmin*) is considered. For each *i-th* admitted Laue group the internal residual factor *R*int(i) is calculated:

$$R_{\mathrm{int}}(i) = \frac{\sum \left| \| F_{obs\,\mathbf{h}} | - \langle | F_{obs\,\mathbf{h}} | \rangle \right|}{\sum \left| F_{obs\,\mathbf{h}} \right|}$$

The probability of the *i-th* Laue group is:

$$P_L(i) = \left\{ [0.85 - \mathrm{R}_{\mathrm{int}}(\mathrm{i})](\mathrm{n}/2)^{1/6} \right\}^2$$

The final probabilistic formula for identifying the most probable extinction symbol, and simultaneously, for confirming the most probable Laue group and crystal system is:

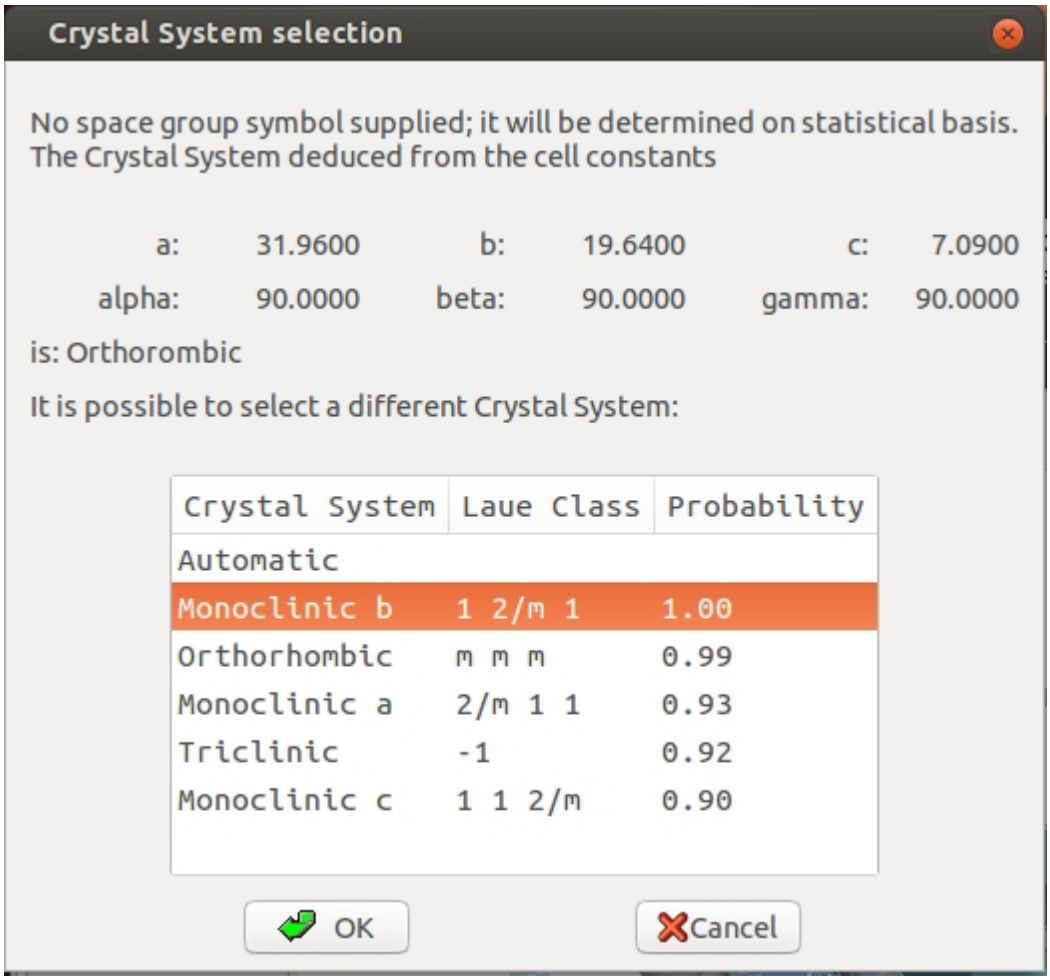
$$[P_{EX}(i,j)]_{sc} = P_L(i)[P_{EX}(j)]^{1/m}$$

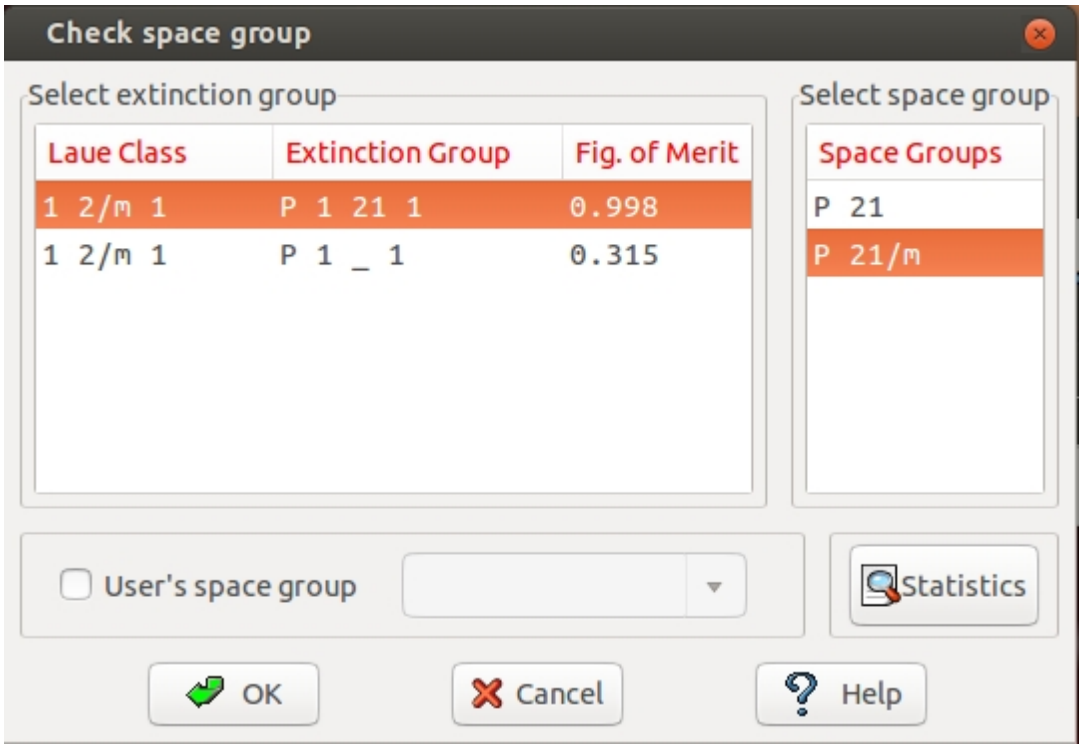
The formula is associated to the *i-th* admitted Laue group and the *j-th* extinction symbol.

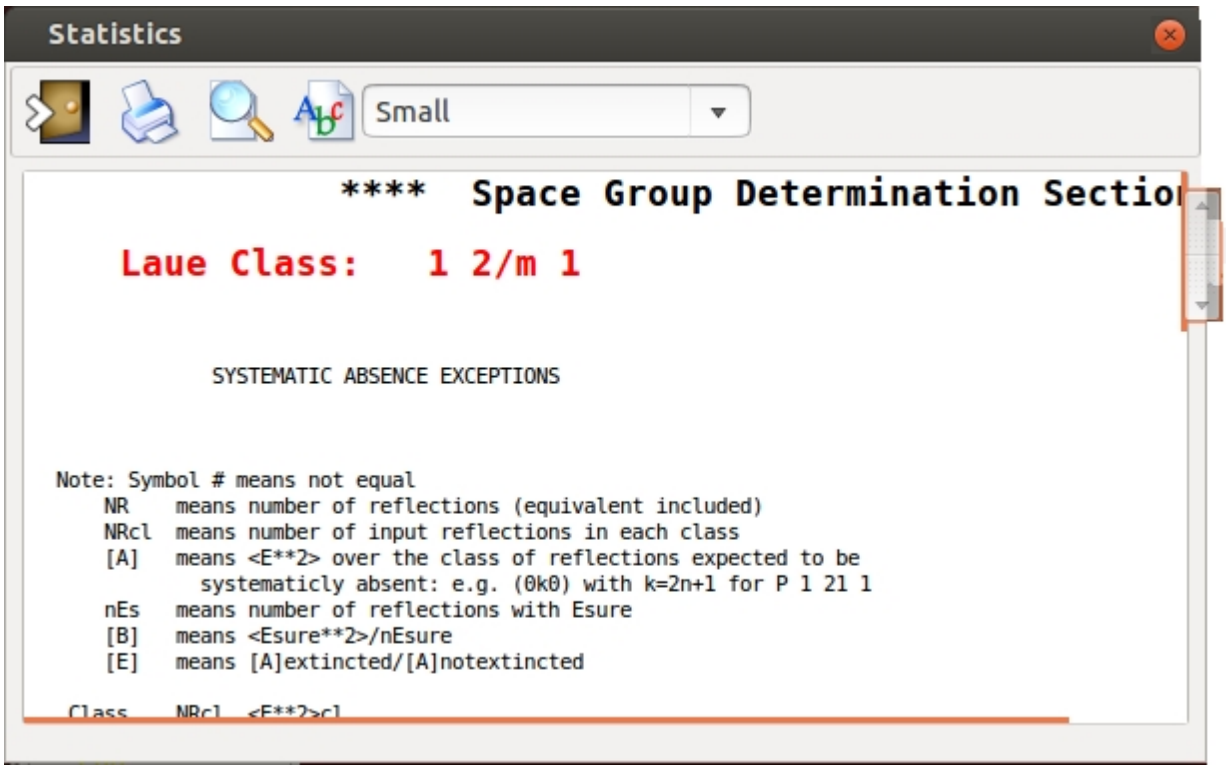
In the following example the cell constants are orthorombic, but the true space group, as suggested by the program, is P 2_1/m.

In the first image, the correct Crystal System has the maximum value of probability.

In the second image the Laue Class is "1 2/m 1" and the Extinction Symbol is "P 1 21 1" have the higher value of the Figure of Merit; the corresponding possible space groups are shown. The button "Statistics" provides details about the calculations related to all the possible Laue Classes for this structure.







Often the same extintion symbol is compatible with several space groups. A special directive (i.e. "*SPACEGROUP SCAN",* available also through the graphic interface) has been introduced, which enables to consecutively explore all the allowed groups (belonging to the most probable extintion group) to make faster the crystal structure solution. The

routine ends with a list of ordered crystallographic residuals, to allow the user to definitively choose the space group and pass to the structure refinement step.

Created with the Personal Edition of HelpNDoc: Easy to use tool to create HTML Help files and Help web sites

Invariants module

Up to 300000 triplets relating the Nlarge reflections are stored for active use in the phasing process.

Triplets are estimated according to their second representation (i.e. P10 formula, as described by Cascarano *et al*., 1984). The concentration parameter of the von Mises distribution is given by

$$G = C\,(1 + q)$$

where C is the concentration parameter of the Cochran's distribution

$$C = 2|E_{\mathbf{h}}||E_{\mathbf{k}}||E_{\mathbf{h-k}}|\,/\sqrt{N} \quad \text{(Cochran, 1955)}$$

and 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 ranked relationships represent a better selection of triplets (with phase value close to zero) than that obtained sorting triplets according to C. Triplets estimated with a negative G value represent a sufficiently good selection of relationships close to 180 degrees. Positive and negative triplets will be actively used in the phase determination process (Giacovazzo *et al*., 1992).

Negative quartets are generated by combining the PSI-0 triplets (relating two reflections with large |E| and one with |E| close to zero) in pairs; 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 actively used in the phasing process (Giacovazzo *et al*., 1992).

To improve the efficiency of the tangent formula in Sir2014, the information contained in the PSI-0 triplets has been actively combined with the P10 relationships and negative quartets, according to Cascarano & Giacovazzo (1995).

Created with the Personal Edition of HelpNDoc: Single source CHM, PDF, DOC and HTML Help creation

Phase module

Phasing Tools

According to circumstances, the *ab-initio* Sir2014 phasing process may apply Modern Direct Methods (MDM), Standard Direct Methods (SDM), VLD or Patterson procedures. Phase extension and refinement are achieved by Direct Space Refinement (DSR) techniques. To preserve its efficiency the program distinguishes different categories of structures: XSmall molecules (up to 5 atoms per asymmetric unit), Small (up to 80), Medium (up to 300), Large (up to 600), XLarge (up to 1000), XXLarge (up to 1750), Huge (no upper limit).

According to user preferences and to the presence or absence of heavy atoms, a different approach (Direct Methods, VLD or Patterson techniques) may be applied. The DSR procedure, mainly based on electron density modification (EDM) techniques, may be accomplished in different ways (it requires increasing computing times moving from Xsmall to Huge structures). We shortly describe the various tools the user may employ in the phasing process.

Modern Direct Methods (MDM) procedure

The approach is of multisolution type. It starts from random phases and applies a tangent like formula simultaneously using:

a) the P_{10} formula, based on the second representation of the triplet invariant. For each triplet the reciprocal space is explored, to provide a contribution which is summed to the Cochran contribution. Negative estimated triplets are then available.
b) the PSI-0 triplets, which in other programs are passively employed as a figure of merit for recognizing the correct solution.
c) negative quartet invariants *via* their first representation.

The phases obtained *via* the tangent step are then submitted to Direct Space Refinement, *via* EDM cycles and RELAX procedure.

The approach is sequential. Each trial starting set is first submitted to the tangent procedure and then to direct space procedures.

The program stops when a figure of merit is larger than a given threshold.

In Sir2014 the MDM procedure is the default choice for structures up to 300 non-hydrogen atoms in asymmetric unit.

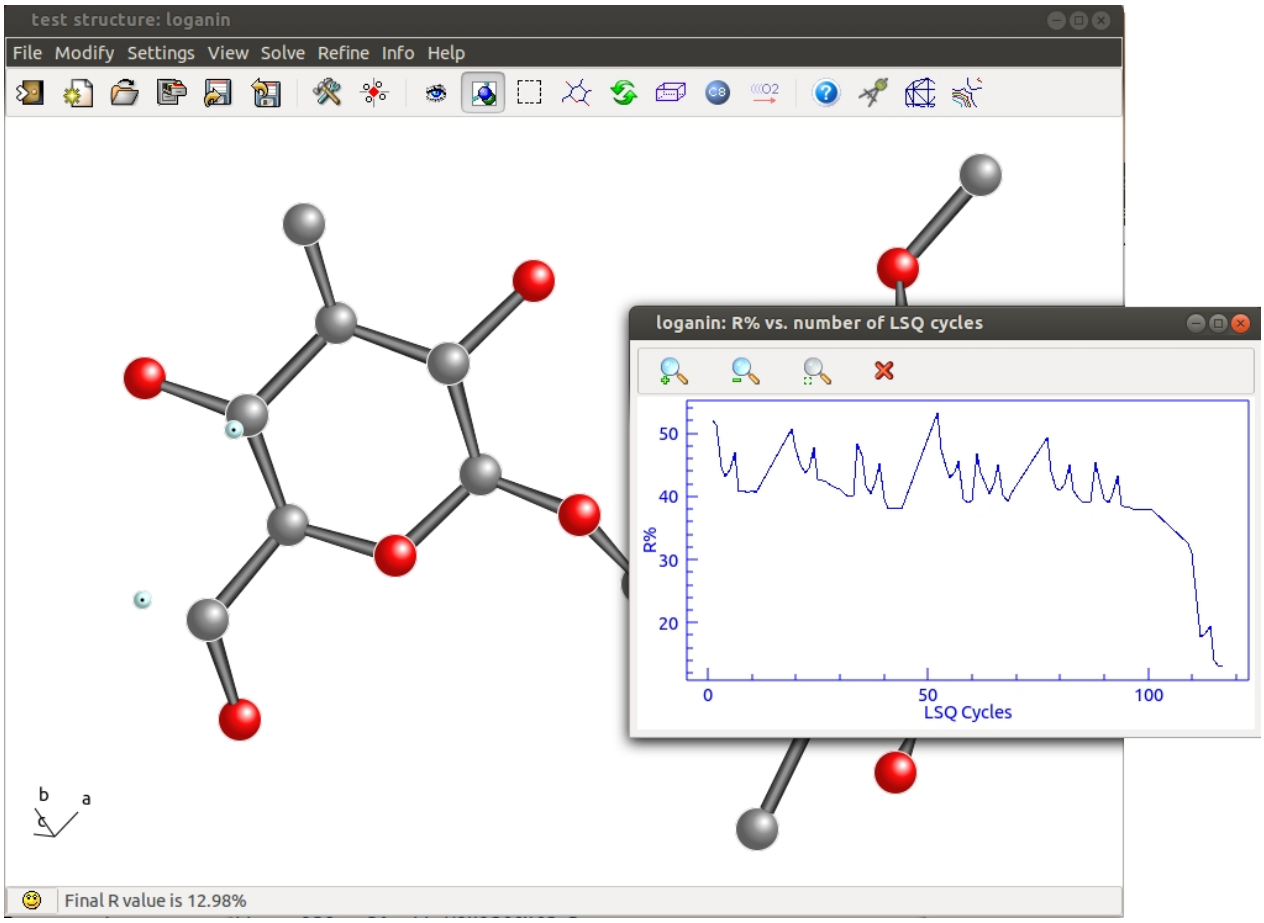
Standard Direct Methods (SDM) procedure

Even in this case a multisolution approach is used; the phasing procedures involves, for each trial, the application of the tangent step to Nlarge reflections (those selected by the normalization process), starting from a subset of random phases. The P_{10} formula is applied and the most reliable negative quartets are actively used during the phasing process.
An early figure of merit (eFOM) is computed for each tangent trial: eFOM (Burla *et al*. 2003, 2004b) is expected to be a maximum for the most promising phase sets. Only the best trial solutions, sorted by eFOM, are submitted to the DSR procedure.

The VLD (Vive la Difference) procedure

A new cyclic phasing algorithm which does not use Direct or Patterson methods has been developed: it is based only on properties of a new difference electron density and of the observed Fourier synthesis.
Random phases are assigned to the target structure factors *F*, an observed Fourier synthesis is calculated and suitably modified (2.5% of the pixels, those with the largest density are selected, the rest is set to zero) to obtain a starting random model. The Fourier inversion of such modified map provides the model structure factors F_p. A new type of difference Fourier synthesis (Burla, Caliandro *et al.,* 2010) is calculated, which, suitably modified and inverted, provides the best estimates of the structure factors of the difference electron density (say F_q).
By definition F_p + F_q is the current best approximation of *F*, by which a new observed electron density is calculated. Cycles of electron density modifications are applied in order to provide, at the end, a new electron density from which a model density is derived as before described.
The procedure is cyclic and stops when a figure of merit suggests that the correct solution has been found (Burla, Giacovazzo *et al.,* 2010, 2011). More recently (Burla, Carrozzini *et al.,* 2011) the efficiency of the algorithm has been improved, by integrating the RELAX procedure (see below), and applied to proteins at atomic resolution.

As an example the screenshots of the crystal structure solution obtained using VLD is reported. In the small window the behaviour of the crystallographic residual during the VLD cycles.



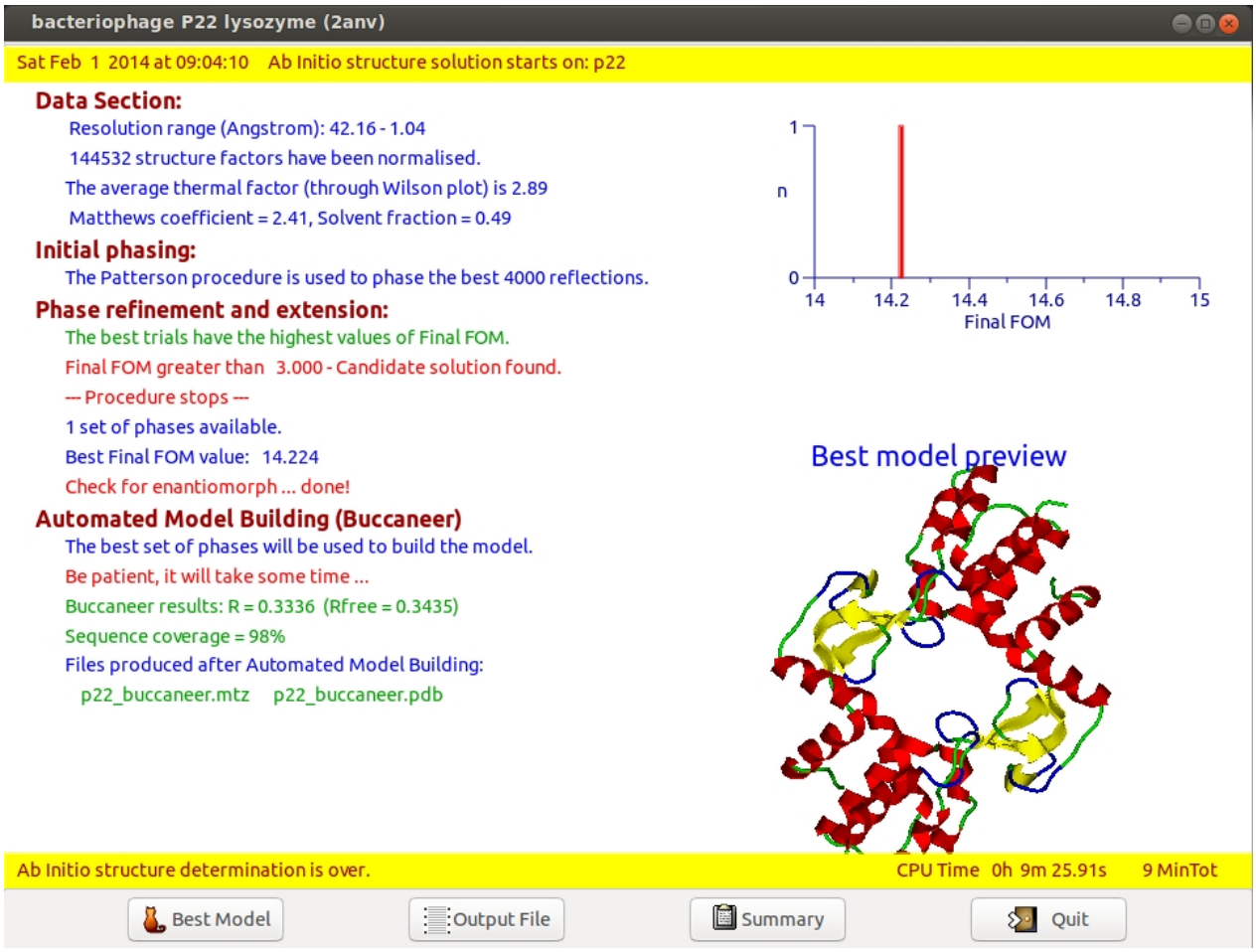
Patterson deconvolution procedure

Sir2014 uses the Patterson approach (Bürger, 1959; Richardson & Jacobson, 1987; Sheldrick, 1992) mostly for the *ab initio* solution of large molecules and proteins.

The Patterson deconvolution techniques are based on the use of implication transformations, of the minimum superposition function (SMF) and on the superposition techniques. The procedure, described by Burla *et al*. (2004a) and Burla *et al.* (2006a,b), includes filtering algorithms (to eliminate the residual Patterson symmetry, to break off non-crystallographic symmetries produced by the deconvolution process, and to restate the non-centrosymmetric nature of the electron density map), EDM and VLD techniques (in the DSR procedure) for improving and extending the phases .

Several new algorithms like C-MAP and SNIP (Caliandro *et al.,* 2014) have been devised and implemented in Sir2014. An Automated Model Building procedure (*via* Buccaneer, ARP/wARP or Phenix) of the final electron density map is now part of the procedure.

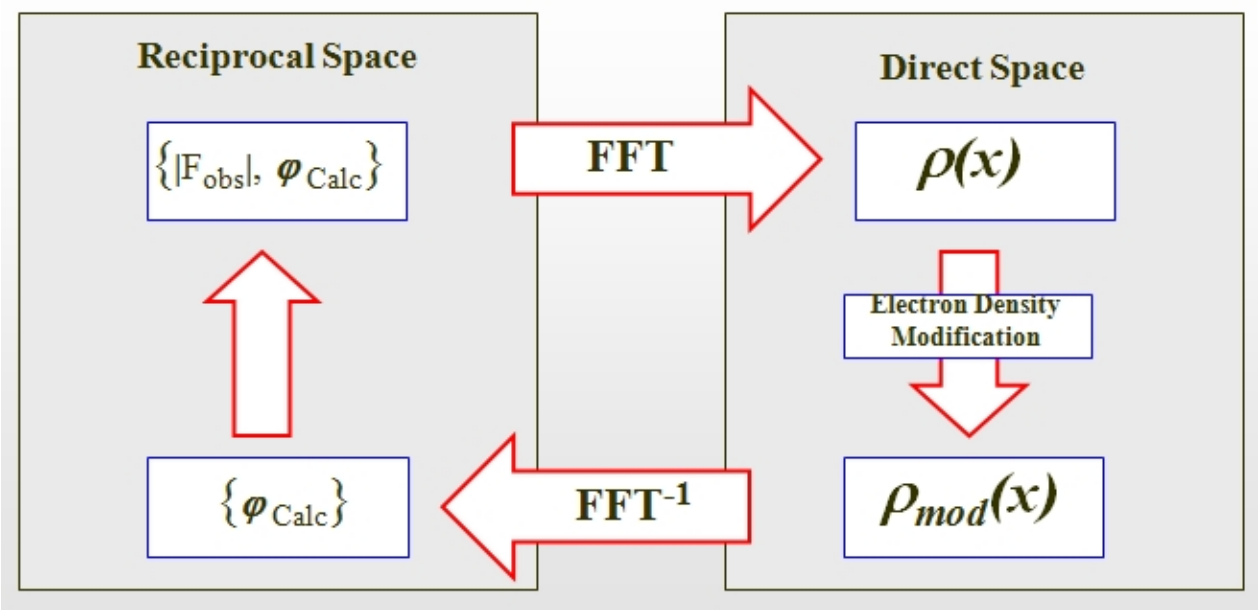
An example of Patterson approach pipeline using the P22 structure (PDB code 2ANV) follows:



Direct Space Refinement (DSR)

The DSR procedure (Burla *et al.,* 2003) is mainly based on *s* supercycles of electron density modification (EDM step), each constituted by *t* microcycles $\rho \rightarrow \{\varphi\} \rightarrow \rho$ (see Fig. 2). The default values of *s* and *t* change with the structural complexity. The modification of the electron density map includes powering (Refaat & Woolfson, 1993) and the inversion of small negative domains (Burla, *et al*., 2003). *n* cycles of difference electron density modification (VLD step) have been implemented in DSR procedure, to improve the phasing process. New phases and normalized structure factor modules (R_c) are obtained by inversion of a changing percentage (up to 10% for small/medium-sized molecules or up to 40% for proteins) of the electron density map: such modules are rescaled by histogram matching with respect to the distribution of the observed ones (R_o). Proper weights are supplied to the new phases, appling the classical Sim-like scheme (Sim, 1959, 1960) or the popular σ_A scheme (Read, 1986; Caliandro *et al*., 2005c) for macromolecules.

Figure 2 - The Electron Density Modification cycle.



The molecular envelope of the protein (Wang, 1985; Leslie, 1987) is used, in Sir2014, as a mask in the EDM step, in order to improve its efficiency (Burla *et al*., 2003). The protein volume is estimated through the Matthews (1968) formula and the envelope is calculated, for each trial solution, from the current phases. The electron density map is modified by assigning different weights to pixels falling inside or outside the envelope, so tentatively depleting the intensities of the false peaks. The map is then inverted and the resulting phases may improve their values.

The whole DSR procedure could automatically be iterated for the same trial, restarting each time from the current phases (Burla *et al*., 2003). The number of total DSR iterations is automatically defined by the program, according to the structural complexity and to the experimental data resolution. This iterative process, although time consuming, allows to solve also resistant molecules (i.e. protein structures diffracting at non-atomic resolution). The DSR procedure is also integrated, at the end of the final iteration, with the RELAX procedure (for structures with experimental data up to 1.5Å resolution).

In order to improve the efficiency of the phasing process in cases where the experimental data have quasi-atomic resolution or poor completeness (e.i. macromolecular crystallography), the *Free Lunch* algorithm (Caliandro *et al*., 2005a,b) is actively used in Sir2014. From an approximate electron density map, the procedure, combined with classical EDM techniques, is able to extrapolate moduli and phases of

unmeasured reflections, with resolution lower or higher than the experimental one, and to actively use them in the *ab initio* solving process. As a consequence, the phase estimates of the observed reflections are subsequently improved and the interpretability of the corresponding electron density map increases. Finally, the *Free Lunch* algorithm makes easier the recognition of the correct solution, by means of a suitable figure of merit (*fFOM*).

The RELAX procedure

The problem of well oriented but misplaced molecular fragments may be solved by the RELAX procedure (Burla *et al*., 2002; Caliandro *et al.,* 2007). Data and corresponding phases of a trial solution obtained in the correct space group are extended and refined in P1 and a reduced number of EDM cycles are performed with no space-group symmetry imposed. Some figures of merit enable the determination of the shift to apply in the correct space group in order to re-establish the original space group symmetry.
The RELAX procedure is automatically applied in a default run of Sir2014, even to Patterson phase sets.

The identification of the correct solution: the final Fom (fFOM)

For small/medium-sized structures the correctness of a solution is assessed, at the end of the DSR process, by the crystallographic residual factor (*Rf*): if the final value of *Rf* is smaller than a given threshold (default value 25%) the program stops; otherwise, the program explores the next ranked phase set.
For large sized molecules and proteins the least squares are very time consuming: furthermore they cannot be applied to non-atomic resolution data. To recognise the correct solution a suitable figure of merit (*fFOM*) has been devised and applied at the end of the *DSR* process. Further details are quoted in Caliandro *et al*. (2014).
The correct solution should be identified by large values of *fFOM.*

Created with the Personal Edition of HelpNDoc: Easy EBook and documentation generator

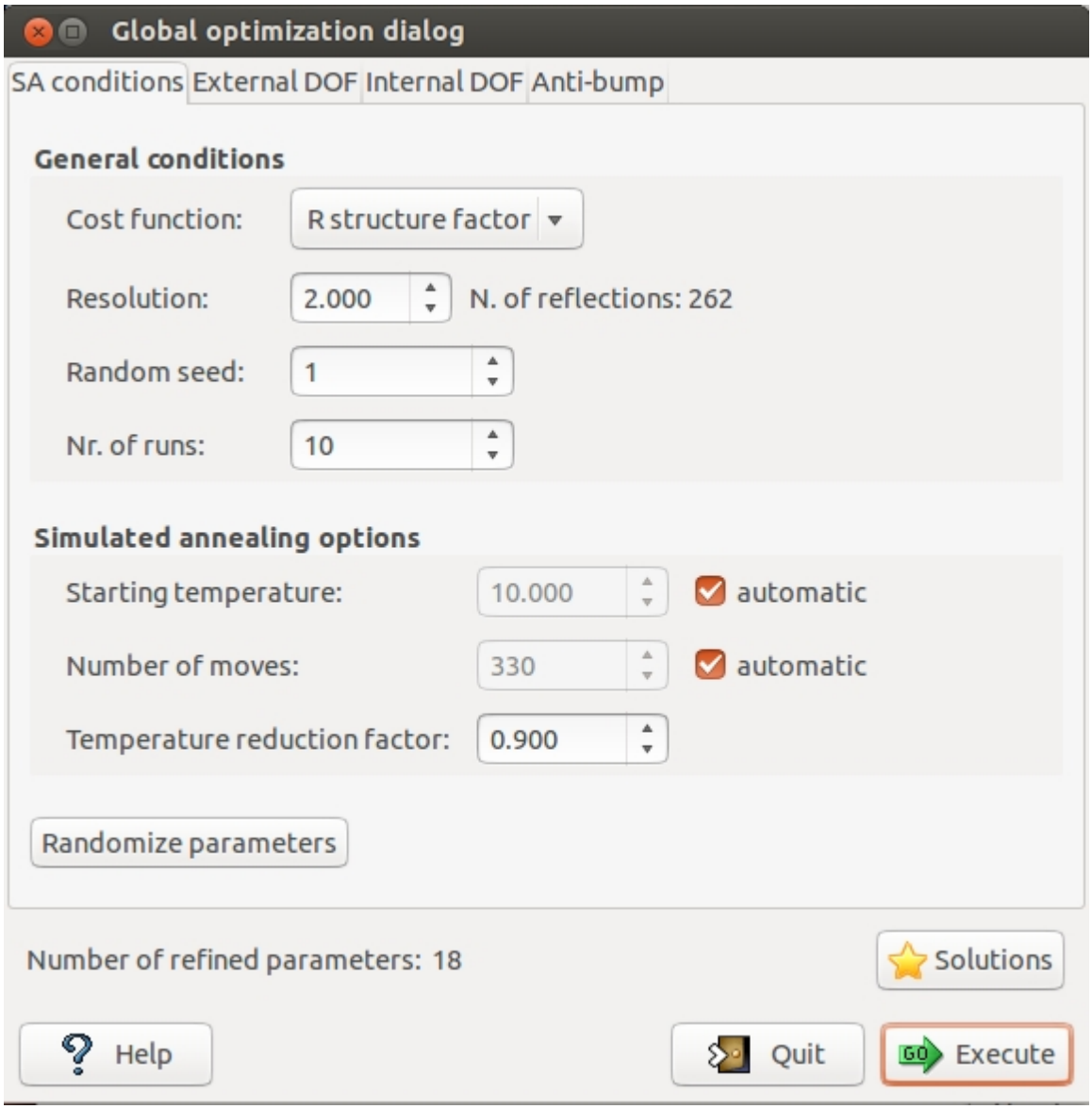
Simulated Annealing

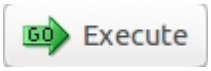
- **Basics**
- **Create input file for Simulated Annealing**
- **Graphic Interface of Simulated Annealing**
- **Output of Simulated Annealing**
- **When Simulated Annealing fails**
- **The directives in Simulated Annealing**

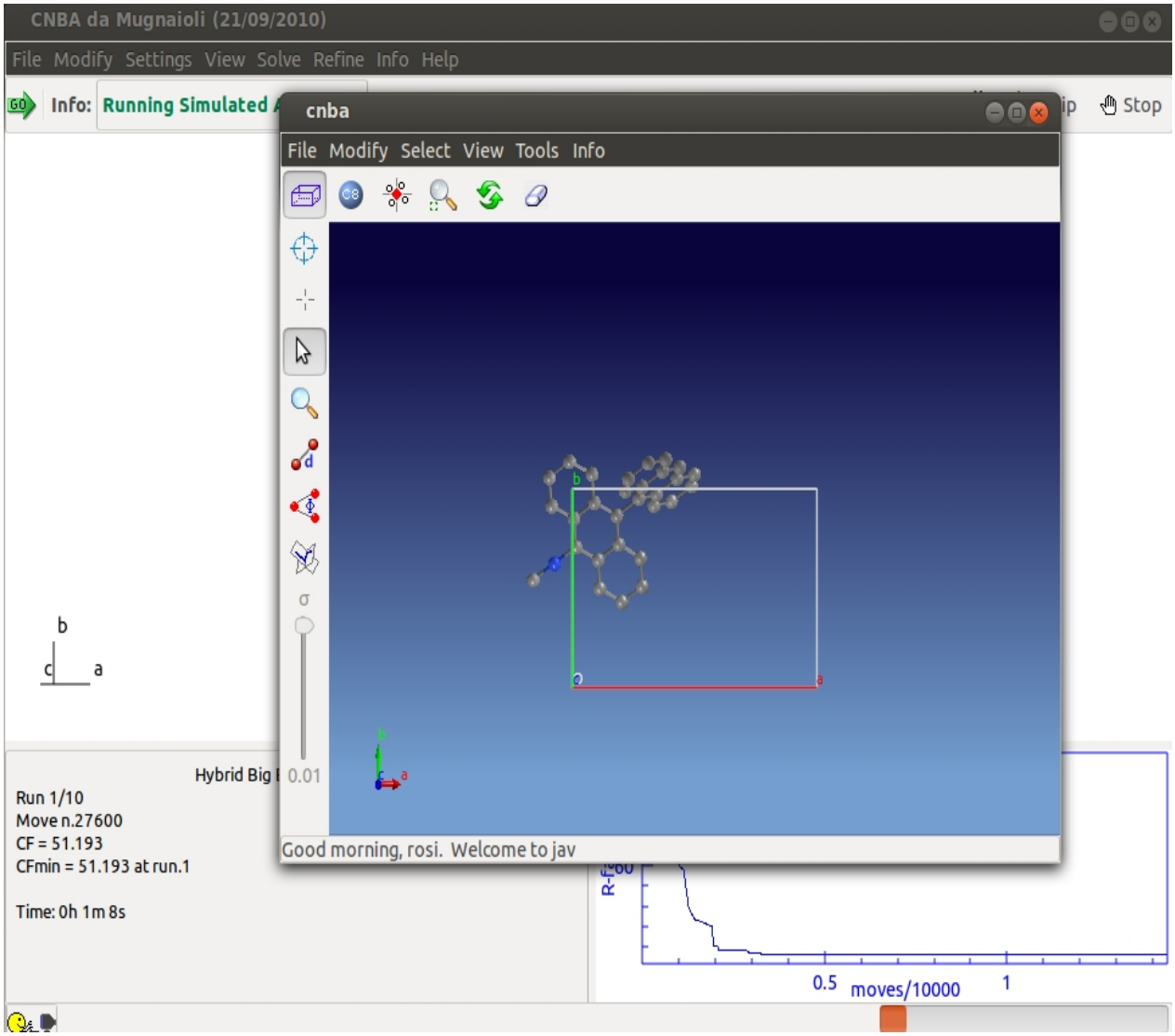
- **Contact**

Basics

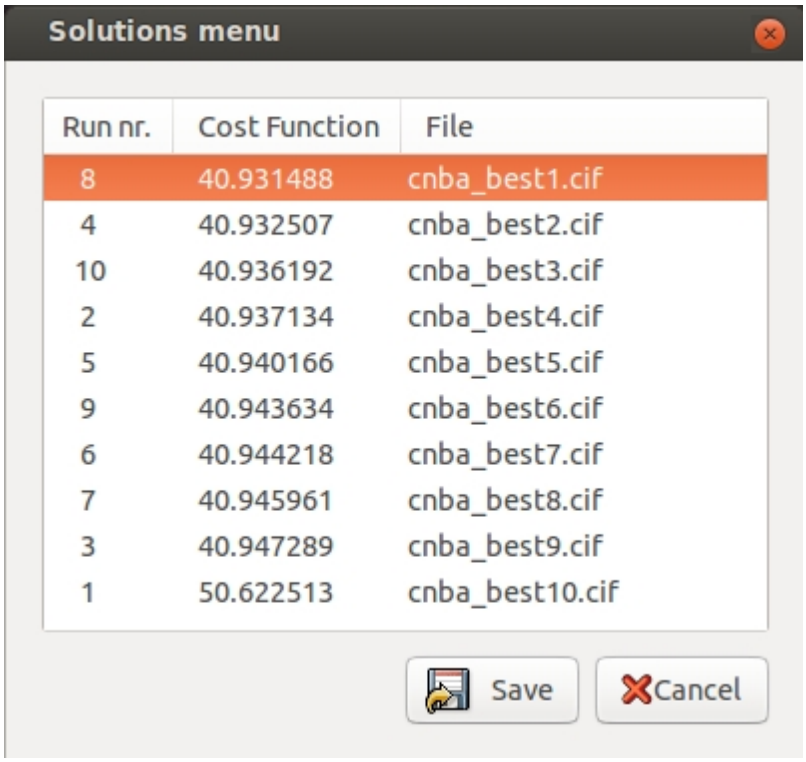
Simulated Annealing (SA), as well as similar procedures like grid search, Monte Carlo, parallel tempering, genetic algorithm, etc., involves the generation of a random sequence of trial structures starting from an appropriate 3D model. The parameters defing the model are modified until a good match between the calculated and the observed pattern is found. The information about chemical knowledge of molecules is actively used to reduce the number of parameters to be changed: bond distances and angles are usually known and kept fixed. Global parameters defining the orientation and the postion of each fragment are varied during the procedure, together with internal parameters like torsion angles.
The dynamical occupancy correction for atoms lying on special positions and atoms sharing the same position in the crystal structure may be taken into account.
In the following images, the Simulated Annealing dialog, the Sir2014 window during the run and the solution menu are reported.



To start click on 

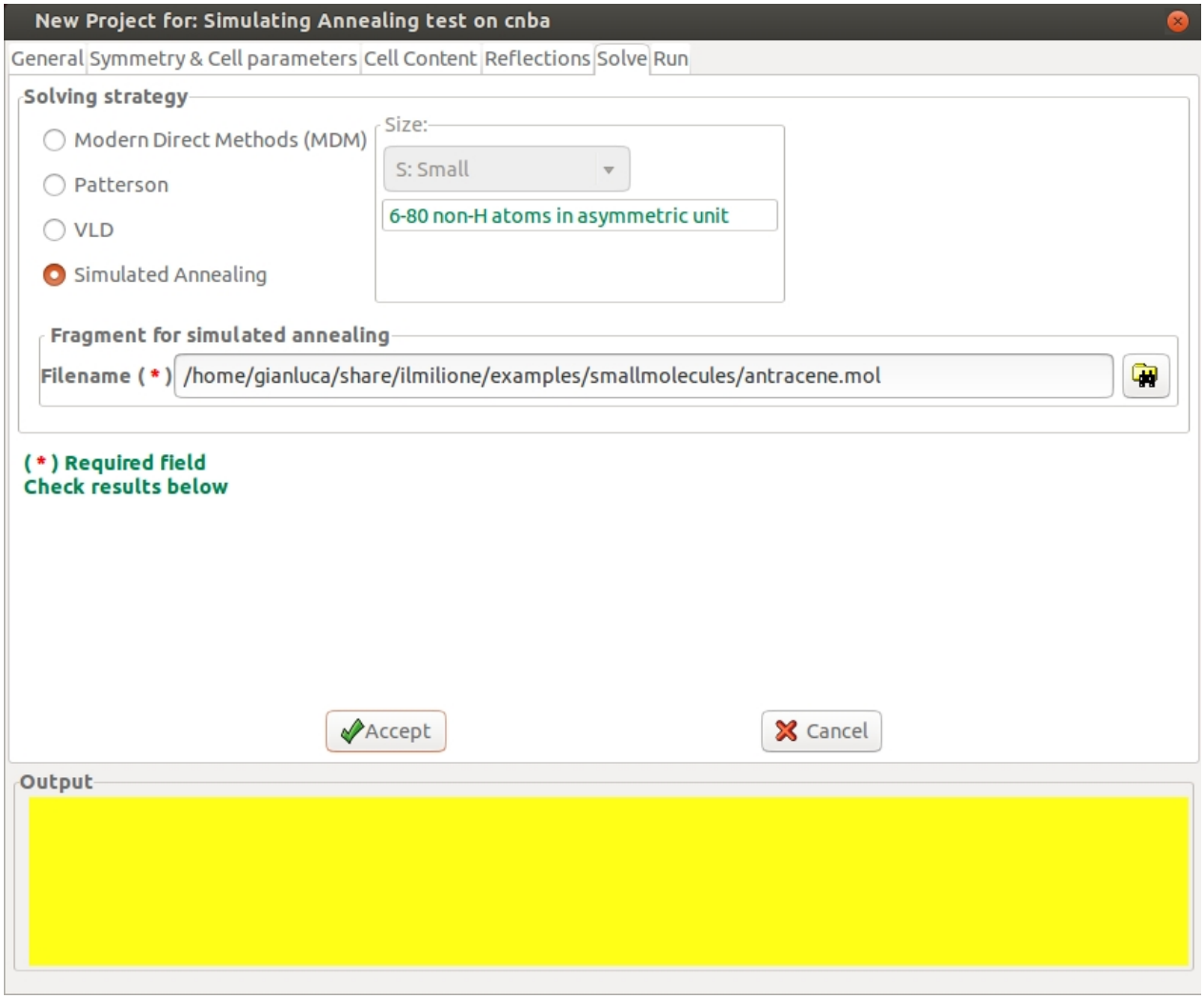


To look at solutions use



Create input file for Simulated Annealing

To run Sir2014 for structure solution by Simulated Annealing you need to create an input file (*.sir) using a text editor or using the program interface. File, New.
In the tab '**Solve**' select 'Simulated Annealing' in the frame **Solving strategy** and import the starting model. The following picture is an example of the interface for the crystal structure determination of the structure *cnba* using Simulated Annealing.



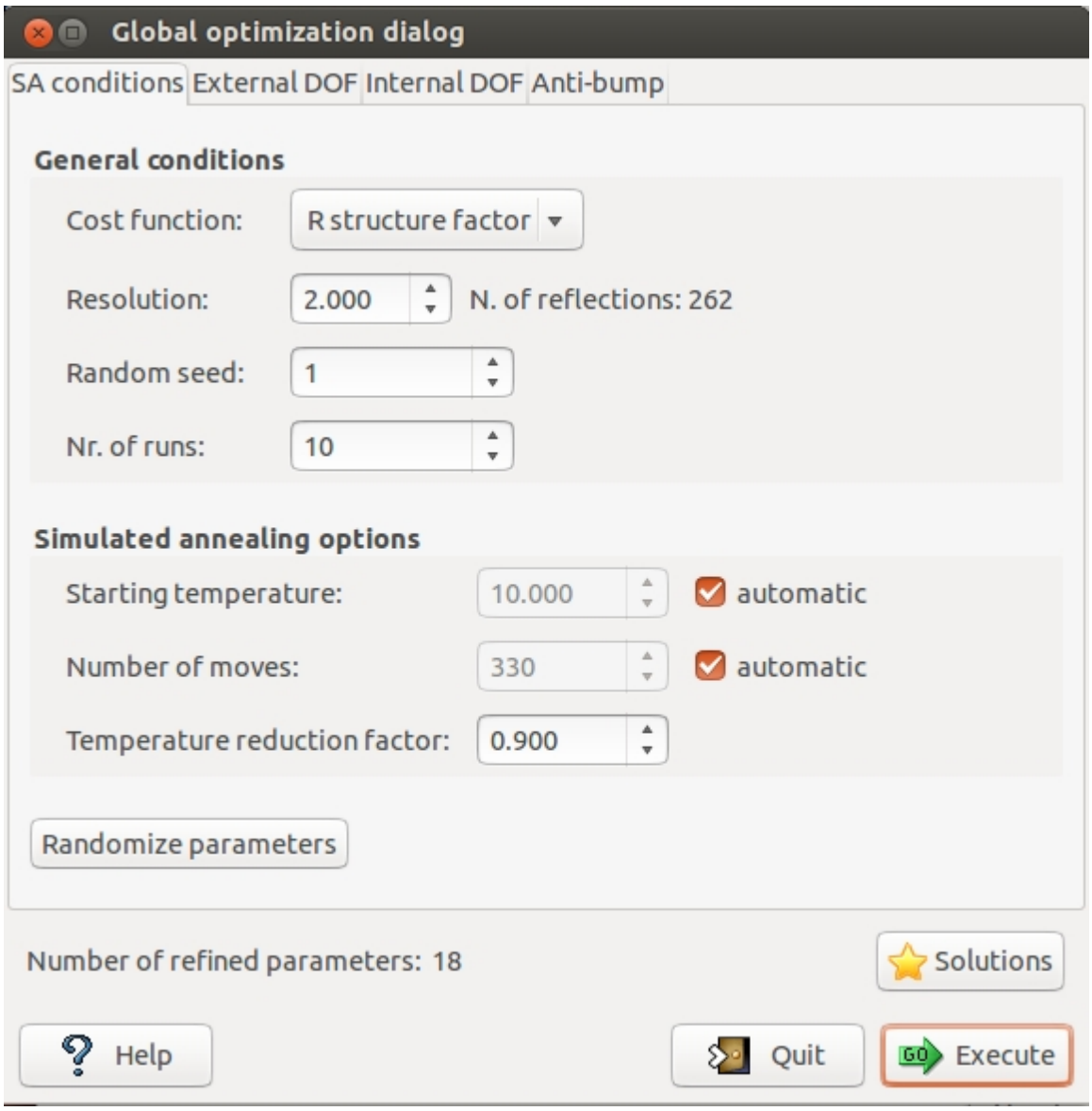
"%sannel" is a command to access to the simulated annealing graphical interface."% fragment struct.mol" is the command to import the starting structural model for simulated annealing in MDL Molfile format (*.mol). This command can be repeted several times to import more than one fragment. Some other file type can be import in the same way: MOPAC file (*.mop), Tripos Sybyl file (*.mol2,*.ml2), C.I.F. file (*.cif), Protein Data Bank file (*.pdb), Sir fragment or Free Fractional Format (*.frac), XYZ format (*.xyz). Refer to openbabel wiki or http://www.ch.ic.ac.uk/chemime for information about common molecular file formats. Refer to http://en.wikipedia.org/wiki/Molecule_editor and http://www.ccp14.ac.uk/solution/2d_3d_model_builders/index.html for an exhaustive list of programs for building 3D starting models. Suggested free available programs for building 3D models are ACD/ChemSketch (windows), Avogadro (all platforms), Marvin (all platforms), ghemical (linux).

NIST Chemisty webbook and Drugbank are useful links to look up fragments.

Instead of command "%sannel", you can access directly to the graphical interface from menu "Solve" > "Simulated Annealing".

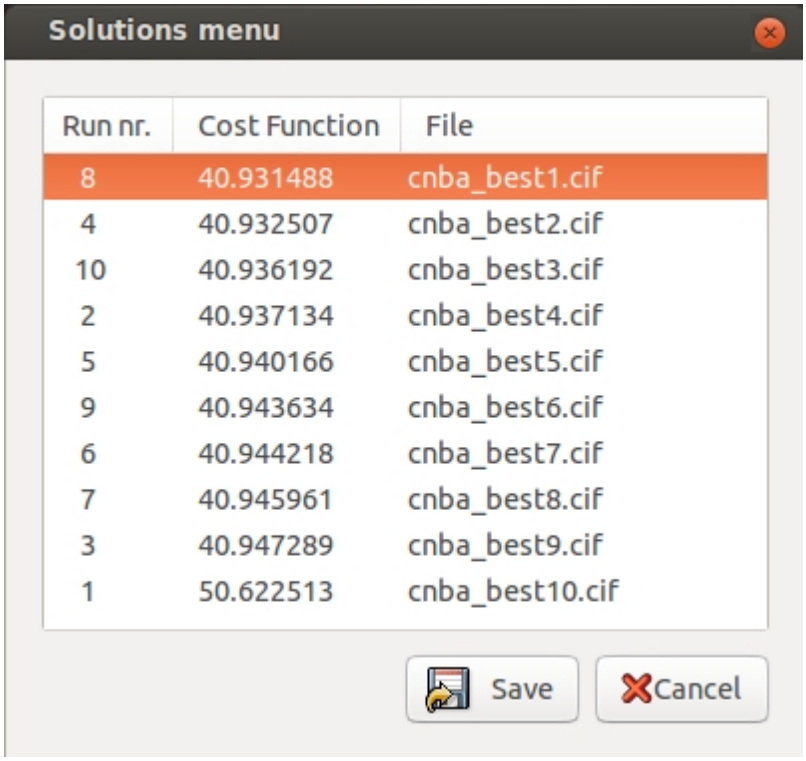
Graphic interface of Simulated Annealing

The dialog of Simulated Anneling in Sir2014 is composed by 4 pages (tabs).
The first page '**SA conditions**' contains the general settings of the Simulated Annealing algorithm:

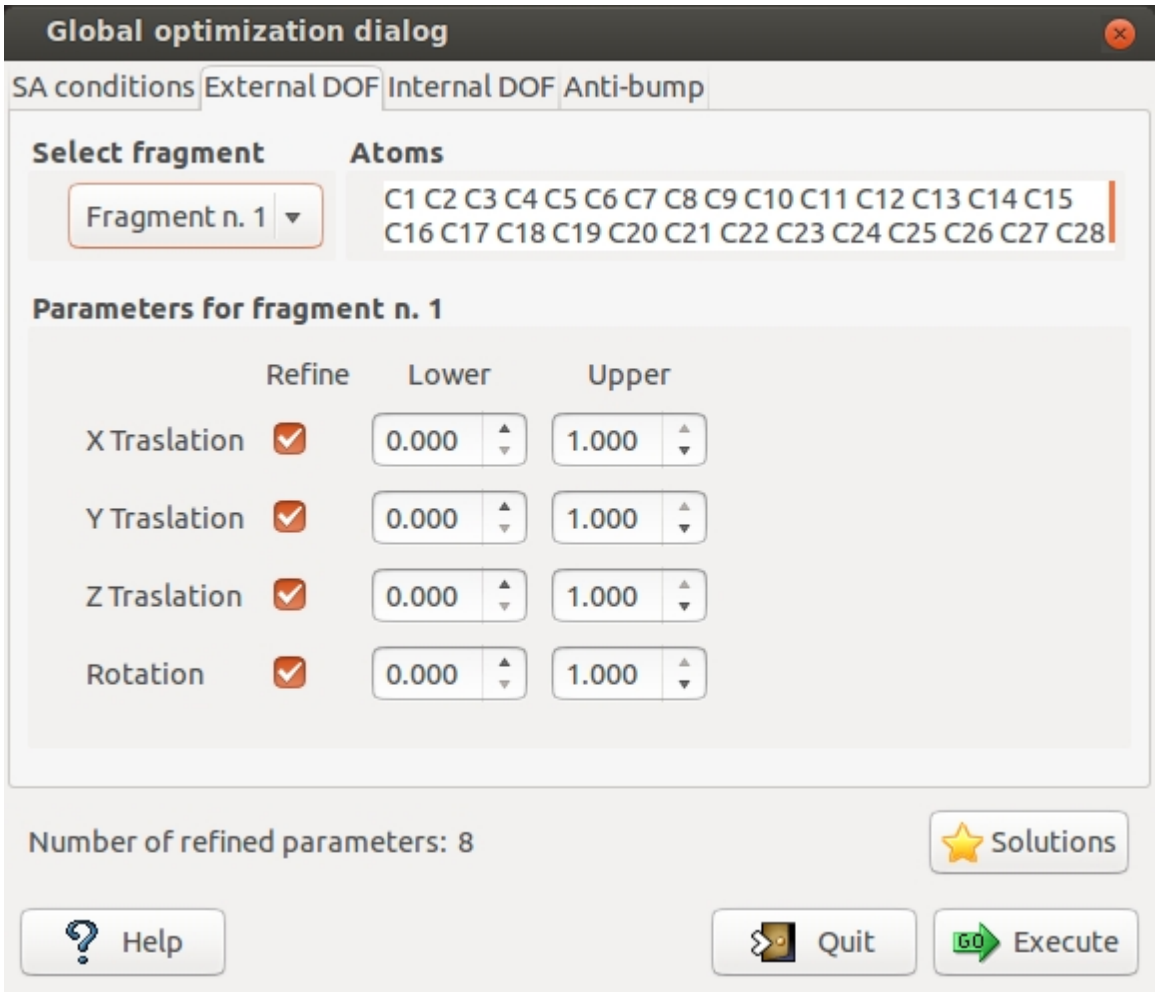


Cost function: 2 cost functions can be selected: R structure factor (default cost function), R intensities.

Resolution: define the maximum resolution used by Simulated Annealing procedure.
Random seed: select the value determining the sequence of random numbers used from the algorithm. When it is set to 0 the random seed will be calculated by the system clock.
Nr. of runs: select the number of Simulated Annealing (SA) runs. At the beginning of each run a new value of random seed is calculated.
Solutions: browse the best solutions saved at the end of each run.



The second page '**External DOF**' contains information about the external degrees of freedom (DOF)

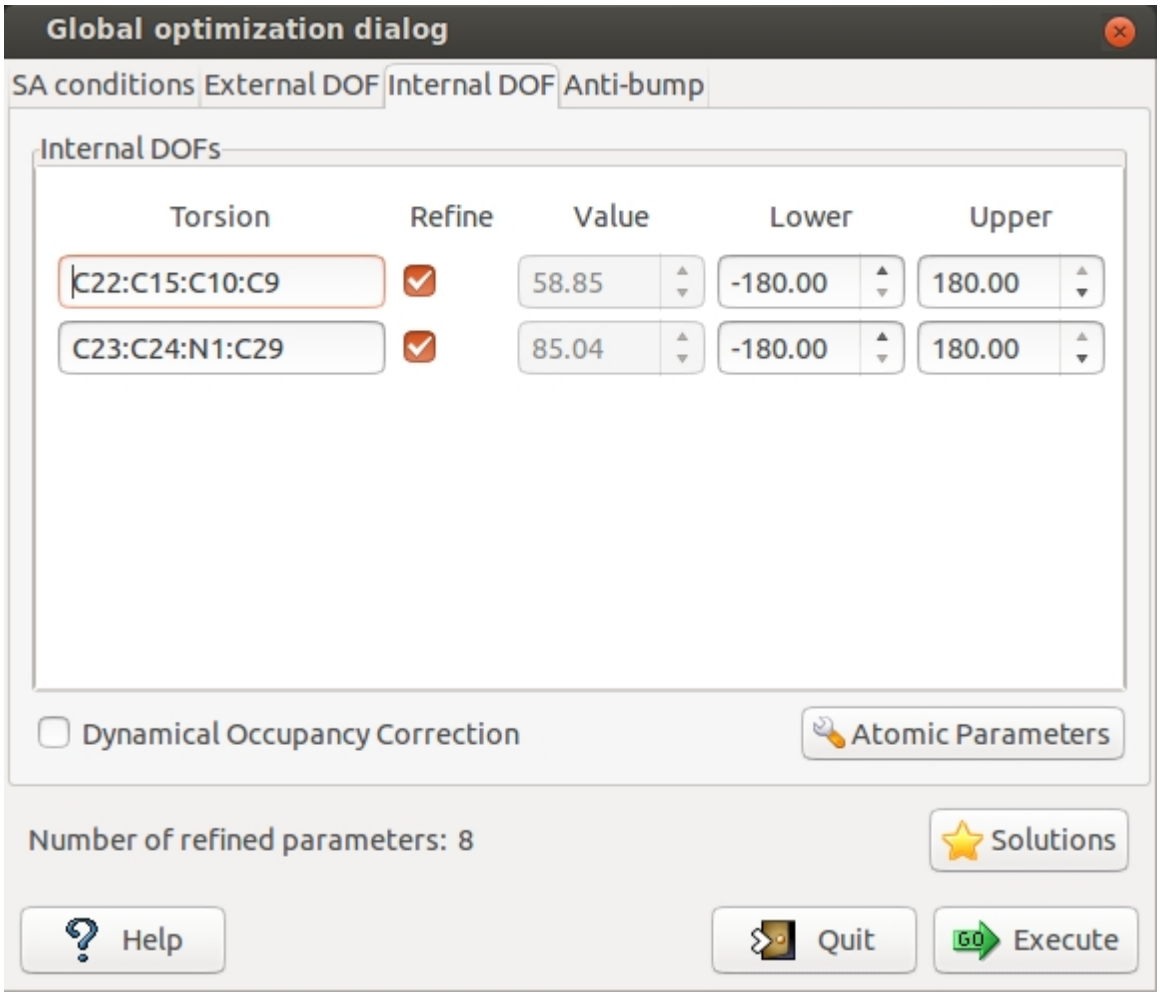


Select fragment: select the fragment and visualize the corresponding structure information.

Atoms: list of the atoms in the selected fragment

Parameter for fragment: list of the external DOFs for the selected fragment. Check the parameters to refine, enter the lower and upper bounds of the parameters.

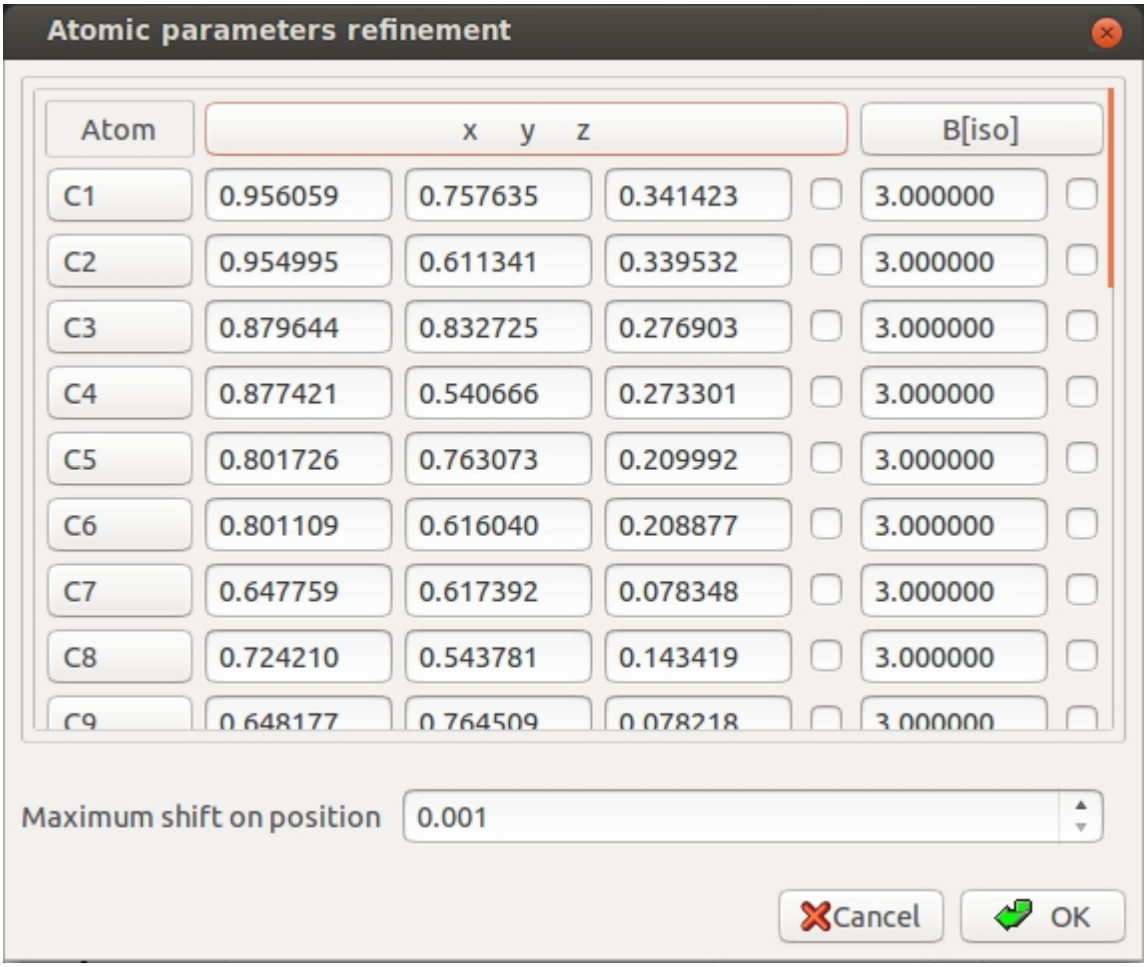
The third page '**Internal DOF**' contains information about the internal DOFs



Internal DOFs: list of the torsions associated to each refinable internal DOF. Check the parameters to refine, enter the lower and upper bounds of parameters.

Dynamical Occupancy Correction: automatic detection of atoms in special position and atoms that share the same position. If this option is not active all refined atoms are considered in general position (default).

The '**Atomic parameters**' button give access to the following dialog

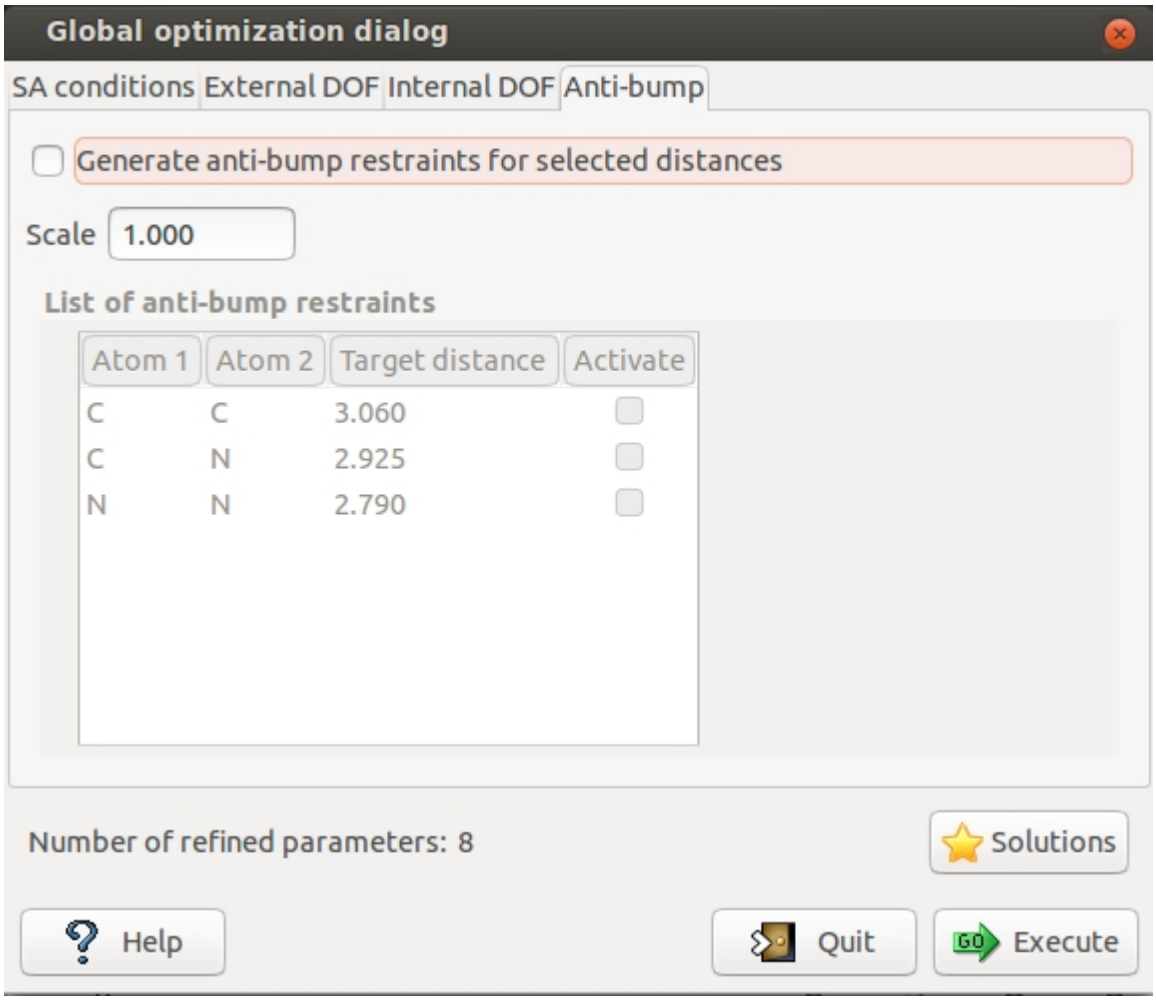


X Y Z: check the buttons on column 'X Y Z' to refine the position of atoms, in this case the fragment is not rigid but the position of atoms are shifted respect to the barycentre. Click on label 'X Y Z' to select all checks in the column.

B[iso]: check the buttons on column 'Thermal parameters' to refine the thermal parameters of atoms (B). Click on label 'Thermal parameter' to select all checks in the column. If the thermal parameters are non indicated in the imported fragment file, the program assigns the default values of B=3.0 for non-hydrogen atoms and B=6.0 for hydrogen atoms. These default values can be changed editing the new values in the column.

Maximum shift on position: enter the value of the maximum shift of the atoms position. 0.020 is the suggested value. Increasing this value the space explored by the algorithm becomes wider and the probability of falling in a local minimum increases.

The fourth page '**Anti-bump**' contains information about the anti-bump restraints.



When simulated annealing is running the progress of structure solution is monitored and the user can examine the graph of the minimum values of cost function (CF) *vs*. the number of moves. The crystal packing is plotted in the JAV viewer.

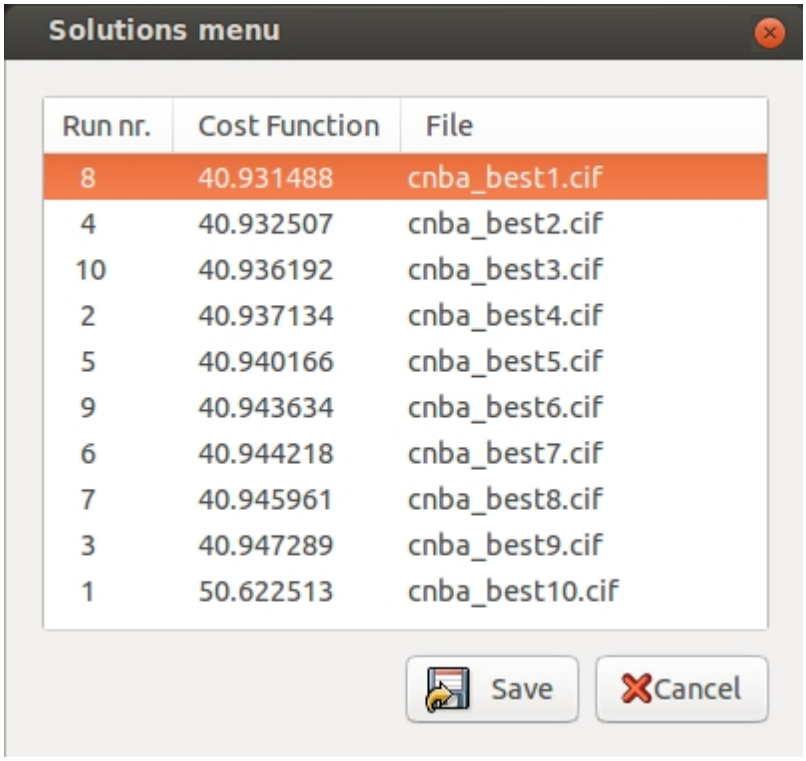
Output of Simulated Annealing

The output file contains information on:

1) the starting structure
2) the volume per atom
2) the connectivity
3) the internal and external DOFs
4) the refined parameters by SA
5) the information about each SA run
6) the summary of all the SA runs

7) the selected structure

At the end of each run the structure coordinates are saved in CIF file with name created by project name with suffix _best1,_best2, ... (e.g. cnba_best1.cif is the best solution, cnba_best2.cif is the second best solution, ...). The order number in the name represents the position in the list of structures ordered according to the cost function.



When Simulated Annealing fails

If Simulated Annealing fails:

1. Starting model is incorrect: bond distances and angle are not entirely accurate, number of building blocks is wrong. Improve your model with Cambridge Structural Database (CSD) or building packages (Avogadro, ChemSketch, ChemDraw, ...). Check for the volume per atom in the output file (about 15-20 Å/atom).
2. The assumption about thermal factors are invalid. Check thermal factors for similar structures.
3. Space group and cell are not correct. Additionally, in many cases it may be necessary

to carry out a series of independent calculations to test different potential space groups and/or unit cell choices.

4. For complex structures the default SA conditions are not sufficient. Try with slower temperature reduction, increase the number of moves and/or runs.

Simulated Annealing Directives

Usually you don't need to know the directives unless you are interested to run Simulated Annealing without interaction with the graphic interface. The list of directives is available here.

Contact

For suggestions and bugs contact corrado.cuocci@ic.cnr.it

Created with the Personal Edition of HelpNDoc: Free Web Help generator

Molecular Replacement

- **Basics**
- **Molecular Replacement Interface & Directives**

Basics

Molecular Replacement (MR) is the most popular method for solving protein crystal structures. Sir2014 uses REMO09 (Caliandro *et al*., 2009) for orienting and locating in the unit cell of the target structure the various monomers constituting the model. REMO09 breaks the six dimensional search into 3-dimensional rotation and translation steps as follows:

a) the space group of the model structure is assumed to be the symmorphic variant of the protein space group;

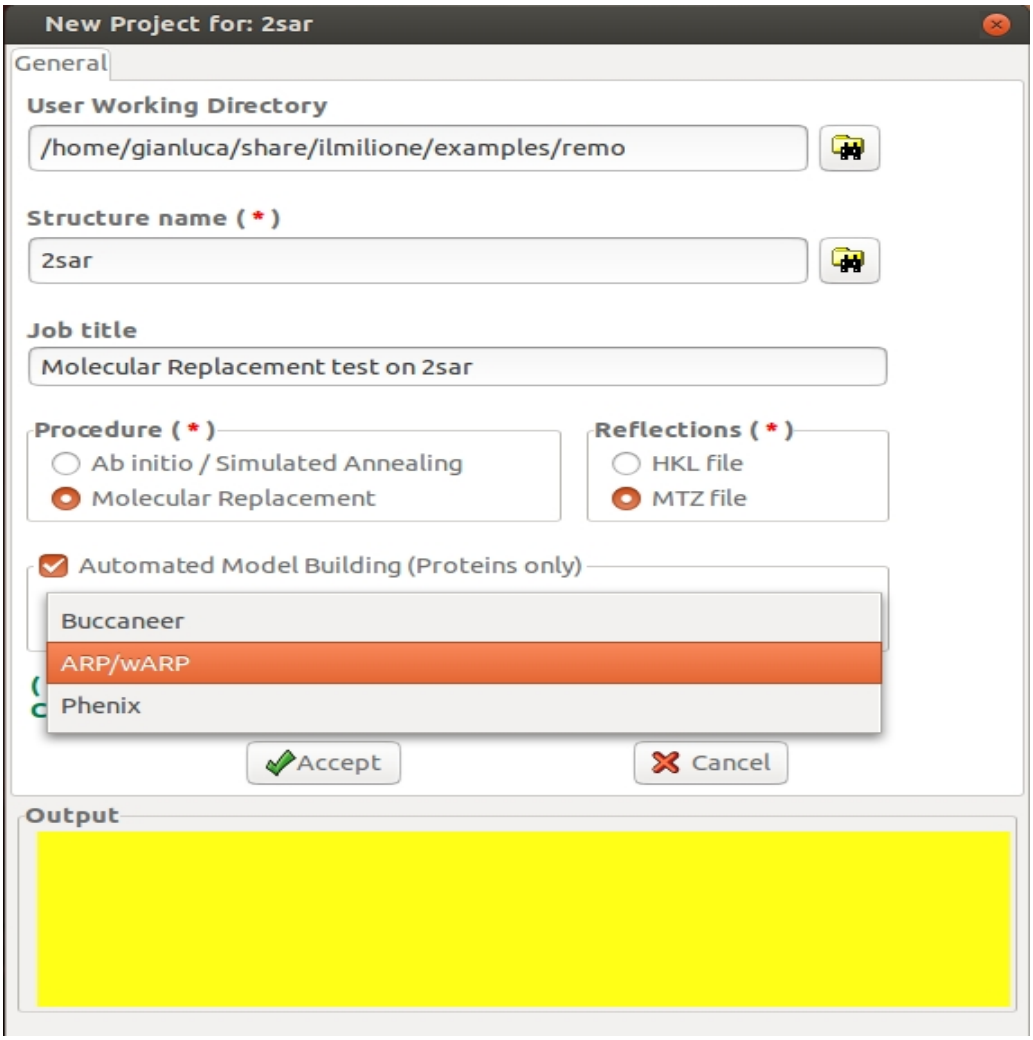
b) fitting between model and target structure is searched in reciprocal space by rotating the reciprocal lattice of the protein with respect to the calculated transform of the model structure;

c) the oriented model molecule is located in the target unit cell by using correlation functions calculated by Fast Fourier Transform.

The approach is purely probabilistic: the various figures of merit are all derived by means of the method of joint probability distribution functions.
REMO09 has been integrated into a pipeline including VLD, EDM, *Free Lunch* algorithms and the program Refmac.
If an external Automated Model Building program is available (Buccaneer, ARP/wARP or Phenix) the program starts it in an automatic way in order to supply a refined and interpreted solution.

Molecular Replacement Interface & Directives

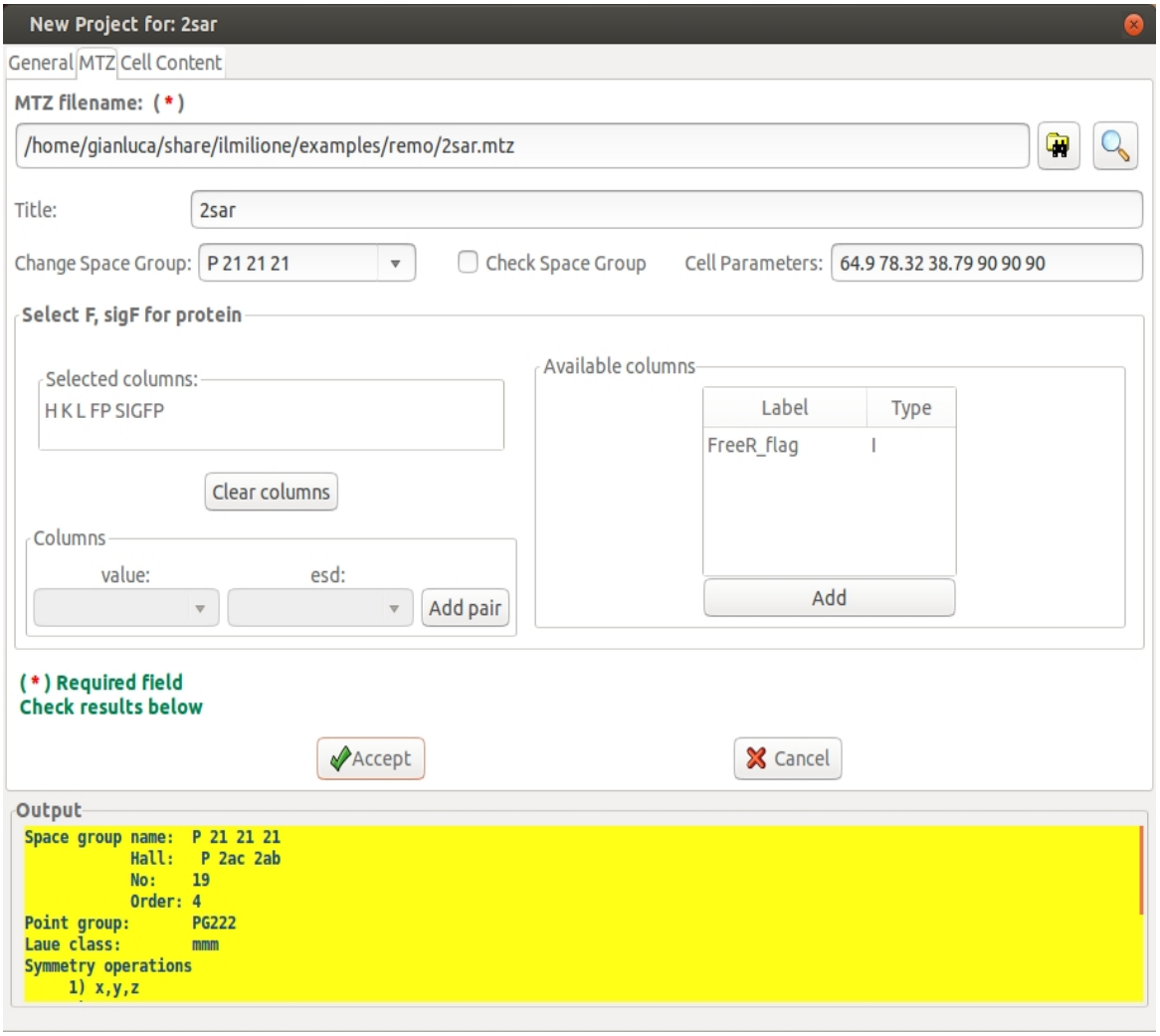
A graphic interface is available in Sir2014 to work with Moleculare Replacement by means of the *new* item of the menu *File* :



In this window it has been specified, among others, the working directory, the procedure to

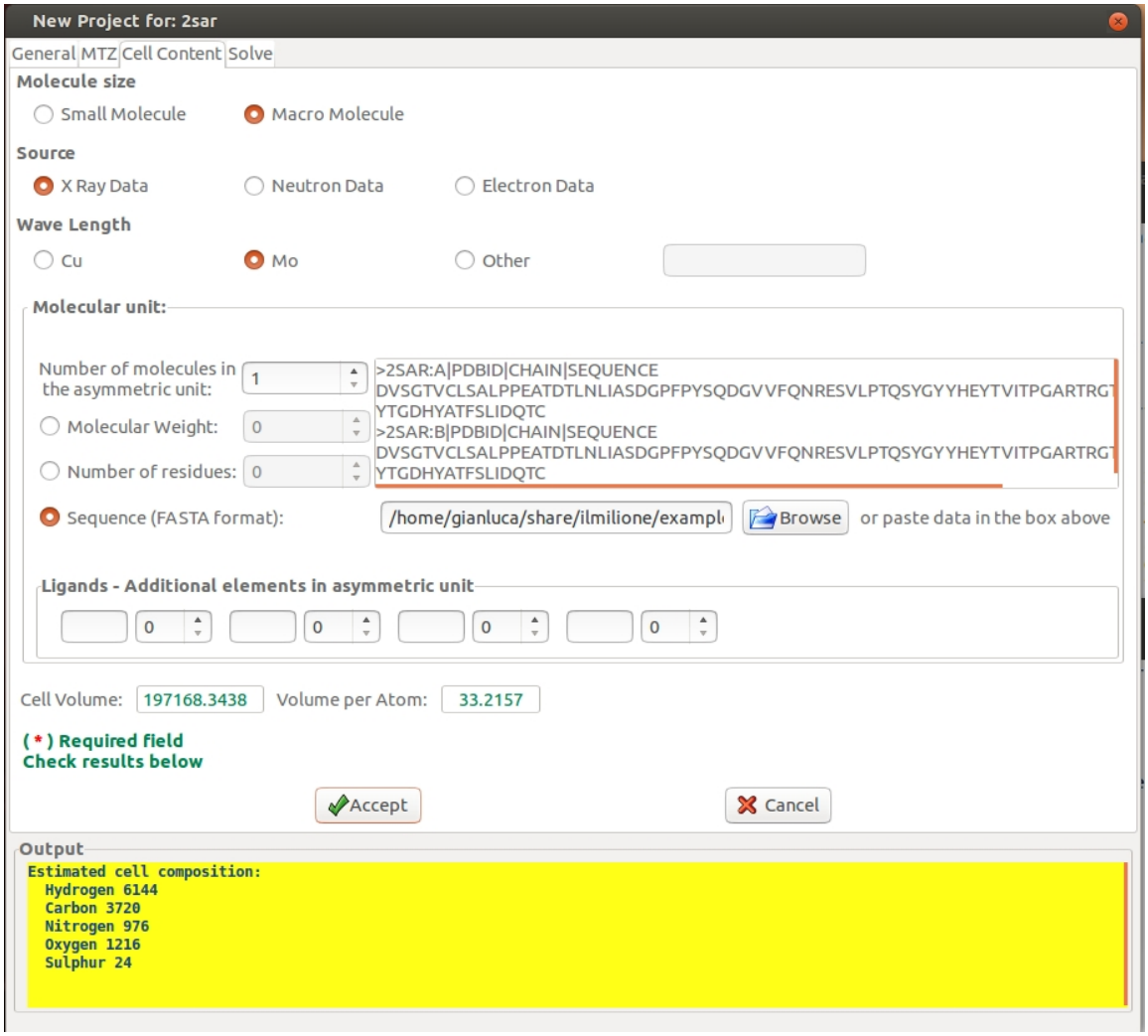
use (Molecular Replacement) and the reflection file format (MTZ). If an Automated Model Building program is installed, it is possible to use it by selecting Buccaneer or ARP/wARP or Phenix.

Once clicked on ***Accept***, it is possible to access the *MTZ* section:



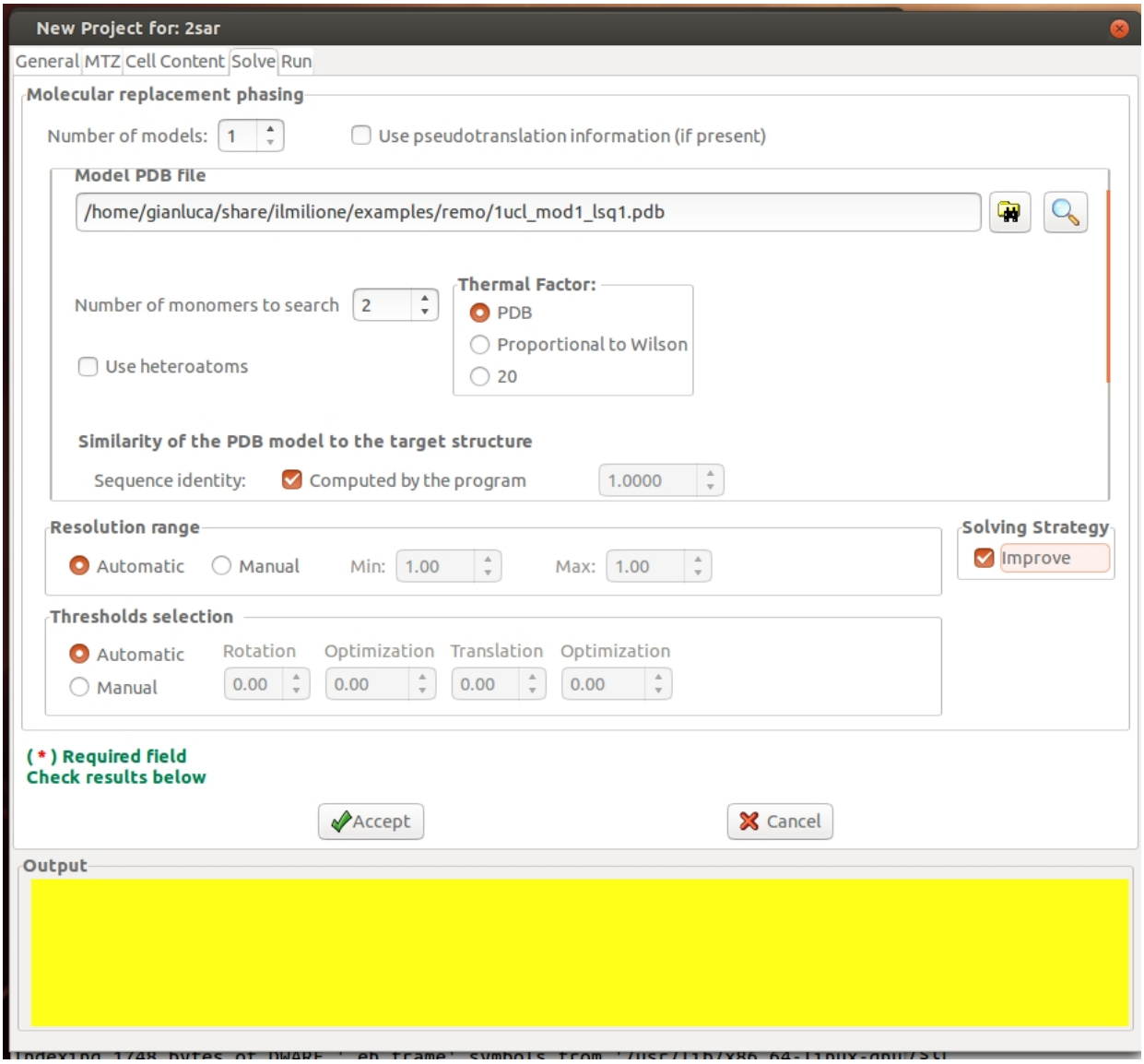
Using this window it has been supplied the name of the MTZ file and have been selected the necessary columns (H K L FP SIGFP).

Once clicked on ***Accept***, it is possible to access the *Cell Content* section:

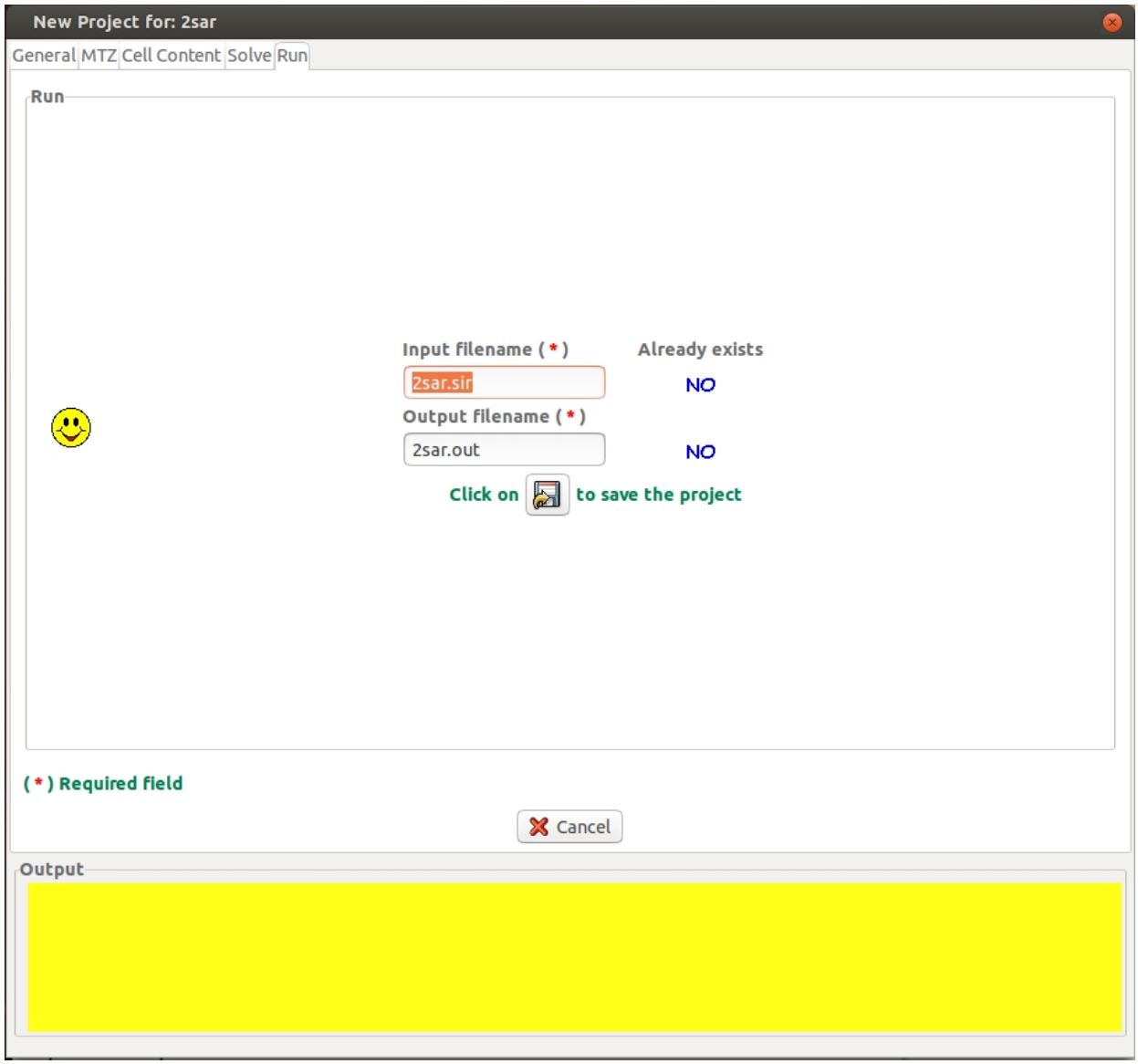


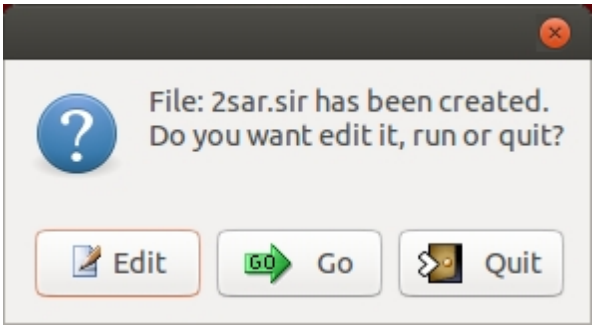
In this example, the sequence (in FASTA format) has been supplied.

After these data have been accepted, the *Solve* interface is available:

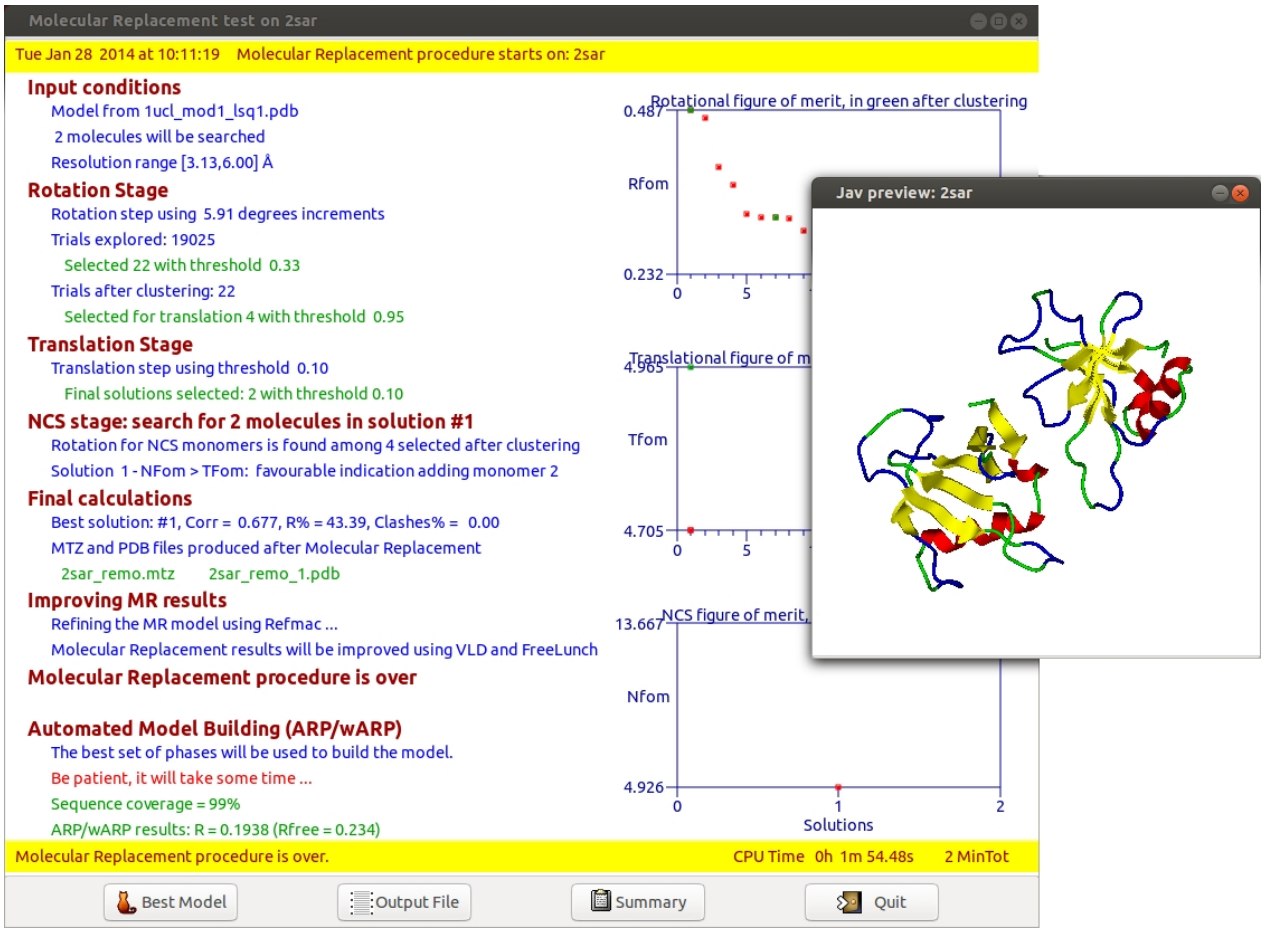


Once clicked on ***Accept***, it is possible to access the *Run* section:





When the program is over, the resulting structure is plotted in a preview window by Jav. It is possible to interact with the structure using the button Best Model



The list of directives is available here. The directives for %data in case of proteins are available here

Created with the Personal Edition of HelpNDoc: Free PDF documentation generator

Substructure determination

When SIR-MIR, SAD-MAD or SIRAS/MIRAS data are available, the first step of the phasing procedure aims at locating the substructure of the heavy or the anomalously scattering atoms, respectively. The experimental isomorphous data are used to derive, *via* probabilistic methods, the most probable values of the normalized structure factor moduli of the heavy or anomalously scattering structure: this is then identified *via* Patterson techniques. The second step of the phasing procedure is entrusted to ***Il Milione***, in which, by a probabilistic approach using the substructure parameters as prior information, the full structure is recovered.

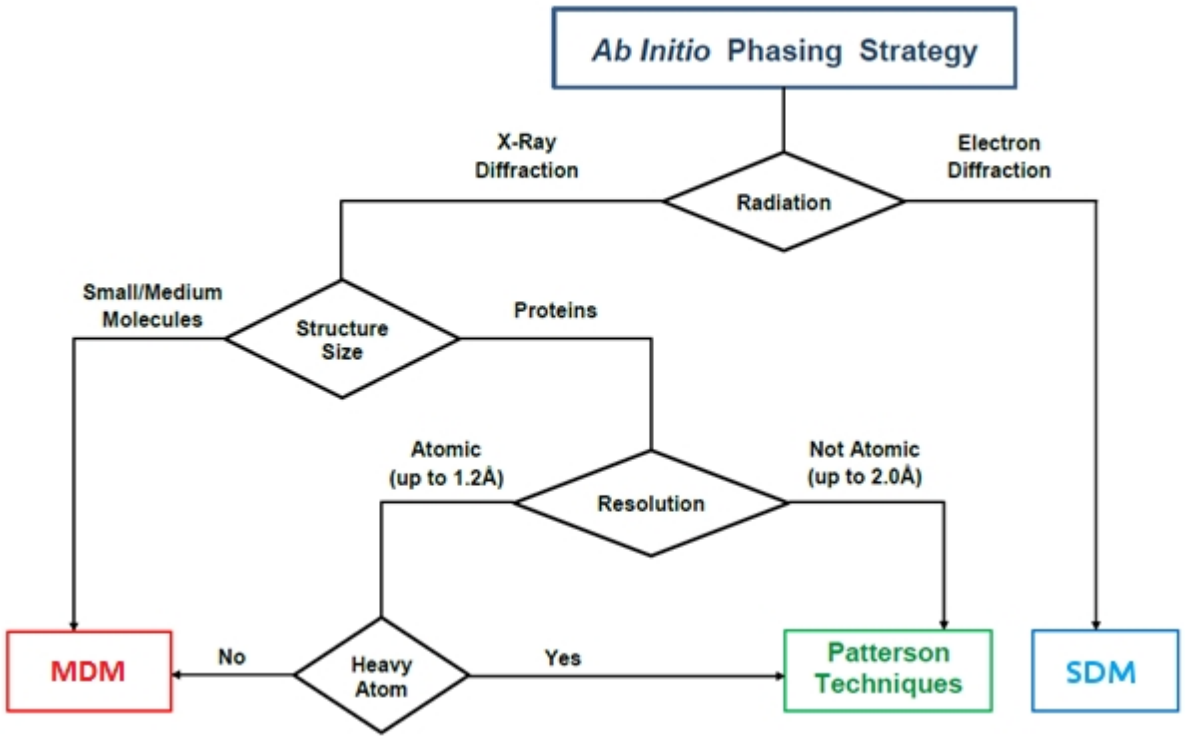
Substructure Determination Directives

No graphic interface is available in Sir2014 to determine the substructure while it is implemented in ***Il Milione***. The list of directives for %data in case of proteins are available here

Created with the Personal Edition of HelpNDoc: Single source CHM, PDF, DOC and HTML Help creation

Sir2014 strategy for ab initio phasing of crystal structures

The Sir2014 flow diagram, shown in below, is a useful guide for understanding the program strategy for *ab-initio* structure solution.



The MDM approach is particularly efficient for small and medium-sized molecules; the application of the RELAX procedure is a very important tool: it is often able to drive phases with large errors to their correct values. VLD may be used as a valid alternative. SDM is the default phasing approach for electron diffraction data. For proteins the tangent procedures are rather inefficient: a large set of trials have to be explored before finding the useful one. The Patterson deconvolution techniques are frequently able to find the correct solution in few trials, provided some heavy atoms are present in the structure.

In order to reduce CPU times, different strategies have been implemented in Sir2014, depending on the starting phasing approach and structural complexity.

The program stops when a proper *fFOM* is larger than a given threshold or, for small molecules, the R value is smaller than a threshold.

Created with the Personal Edition of HelpNDoc: Produce electronic books easily

Completion and refinement of the structure

As mentioned before, for small/medium-sized molecules, a preliminary diagonal least squares refinement is automatically performed at the end of the DSR process to recognize the correct solution. The least squares procedure is also suitable for the complete crystal structure refinement. Its specific features are:

1) The full matrix or diagonal matrix may be used.

2) 18 weighting schemes are available. If the weighting scheme contains adjustable parameters, the program refines the values to obtain a good distribution of $\langle w\Delta^2 \rangle$ against $|F_o|$ and resolution, and the value of the goodness of fit close to one (Spagna & Camalli, 1999).

3) The program generates constraints for the parameters of atoms on special positions in all space groups.

4) Automatic or guided generation of hydrogen atoms. Their contributions are included in the refinement, by allowing the positional parameters to ride on the corresponding parent atom.

5) Floating origin is restrained automatically by setting the restrain on the sum of the appropriate coordinates.

6) The possibility to impose conditions (constraints) or additional information (restraints).

The constrained atoms are regularized to an ideal model structure of known geometry (i.e. benzene ring) and this rigid body is refined as compact unit assuming three translational parameters and three angles which define its orientation. The method used to compute the coordinates of the model follows the approach described by Arnott & Wonacott (1966). In order to build the internal Cartesian coordinates, the program uses the ASCII file

Sir2014.gru which contains models described by the Z-matrix formalism.

The following restraints are available: bond distances, bond angles, planarity.

Fourier, least squares, hydrogens and restraints tools are accessible through the Graphical User Interface (GUI).

Created with the Personal Edition of HelpNDoc: Write EPub books for the iPad

When default Sir2014 fails: some advices

Sir2014 has developed automatic strategies to find the correct solution among the various trials. In addition the user can adopt several options to choose his own phasing pathway.

- A different phasing approach can be used.
- An alternative space group should be carefully considered.
- The number of reflections (Nlarge) actively used in the MDM or SDM phasing process is fixed by the program. For some structures, the ratio "number of active triplets/Nlarge" is too small (e.g., less than 15). Larger values of Nlarge (NREF directive) may improve the phasing procedure. The same result may be achieved modifying (decreasing) the G threshold (GMIN directive) for triplets reliability.
- High (or low) resolution reflections may occasionally play a too important role in the first steps of the phasing process. Fixing an overall thermal factor lower or larger (BFAC directive) than that provided by the normalization routine may successfully change the phase extension and refinement procedures.
- Experimental data of good quality allow the success of structure solution. If in some extreme (very low or very high) resolution shells, equivalent reflections show large internal *Rf* values (*Rmerge* > 25-30%), the user can truncate (MINRES o RESMIN directives) the data submitted to the phasing process.
- The user can modify the DSR strategy (DSR directive) to improve the efficiency of the phasing procedures and of the structure refinement process.
- Low data resolution is an unavoidable feature in macromolecular crystallography: it may limit the efficiency of crystal structure determination. Extrapolating the moduli and phases of unobserved reflections beyond and behind the experimental resolution limit (GENERATE directive) makes the *ab initio* phasing procedures and structure solution process more efficient.

Created with the Personal Edition of HelpNDoc: Produce Kindle eBooks easily

The Graphic Interface

The graphic interface provides a tool to monitor the process of solution and refinement of the structure. The user can interact with the program using the facilities accessible through the menu and the toolbars.

The following icons belongs to the main toolbar; click on them for more info:

Terminates the program.

A new input file is created.

An old input file is opened and the run started. Input file can be edited.

An old project (.BIN) file is opened.

The project binary file or the model can be saved (.SAV or .MDL extension).

A previously saved project file or model is restored.

Offer many possibilities to modify the model.

Apply/Remove symmetry.

Supply different views of the model.

Set the background color (black or white). In alternative Ctrl-b can be used.

Define part of the model to zoom.

Show all atoms (when some of them has been hidden).

Let the model fit the drawing window..

Show/Hide cell edges.

Show/Hide atom labels.

Move atom labels.

Info on the model.

Distances and Angles calculation.

Start Jav loading the current model and the related map.

2D contour facility.

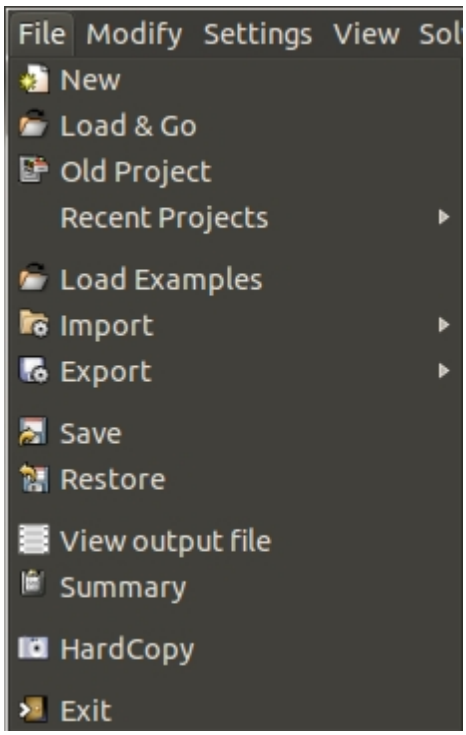
The menu items follow:

- File
- Modify
- Settings
- View
- Solve
- Refine
- Info
- Help

Created with the Personal Edition of HelpNDoc: Full-featured Help generator

File

Possibilities offered by the sub-menu File shown on the left; some of these can be accessed directly using icons in the main toolbar.



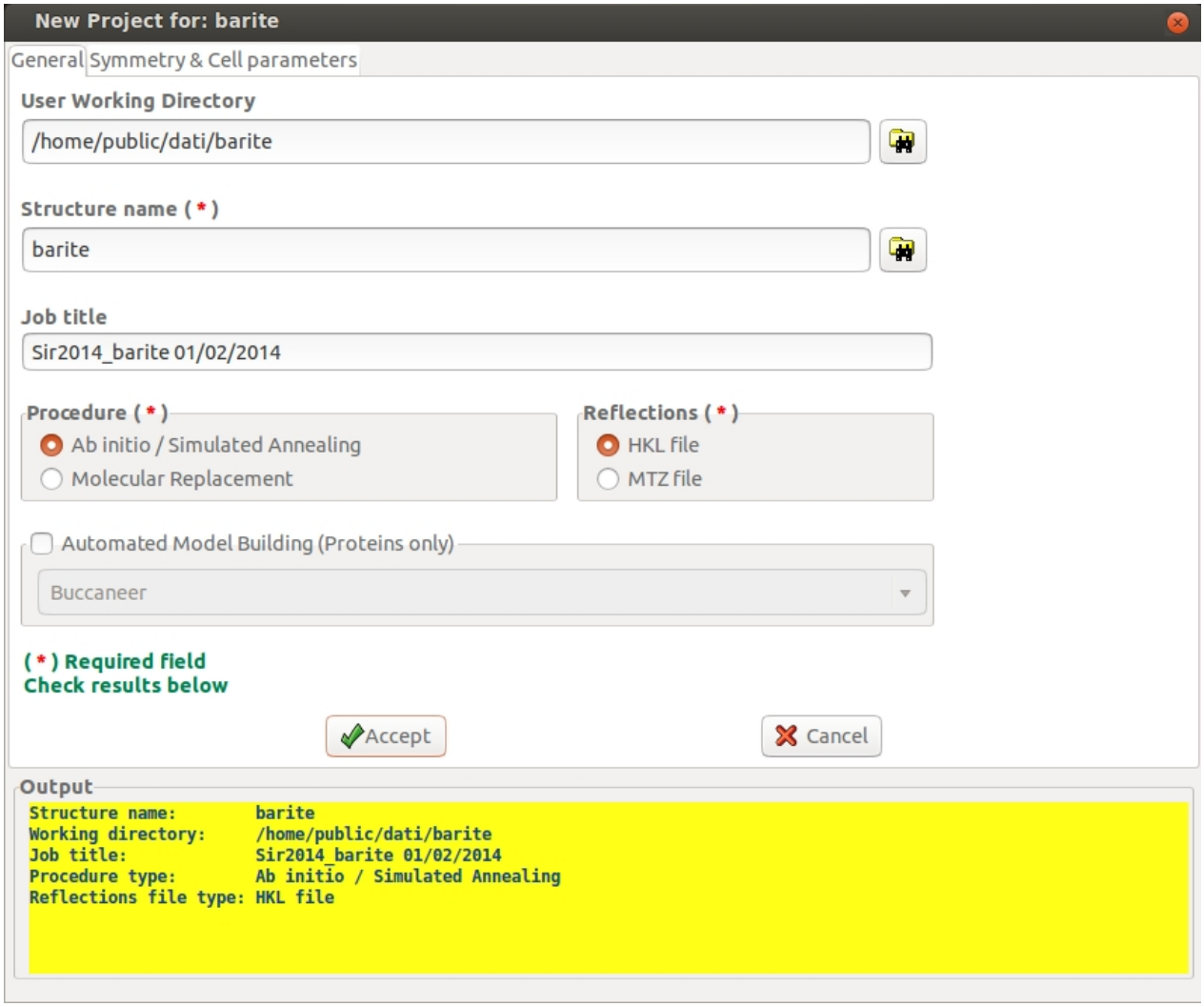
New	A new input file is created.
Load & Go	An old input file is opened and the run started; input file can be edited.
Old Project	An old project (.BIN) file is opened.
Recent Projects	The selection among the 5 most recent projects is allowed.
Load Examples	Some examples are provided.
Import	Data in different format can be imported.
Export	Files for different programs are created.
Save	The project binary file or the model can be saved (.SAV or .MDL extension).
Restore	A previously saved project file or model is restored.
View output file	Provide access to the output file.
Summary	Main results are summarized.
Hard Copy	Create an image file of the program window.

Exit Terminates the program.

New

A sequence of windows helps the user to set up an input file for the program. In these windows the user has to supply the information required, click on the button **Accept** and check the results in the yellow text area at the bottom of the window. As soon as the data have been accepted, the next window will be accessible.

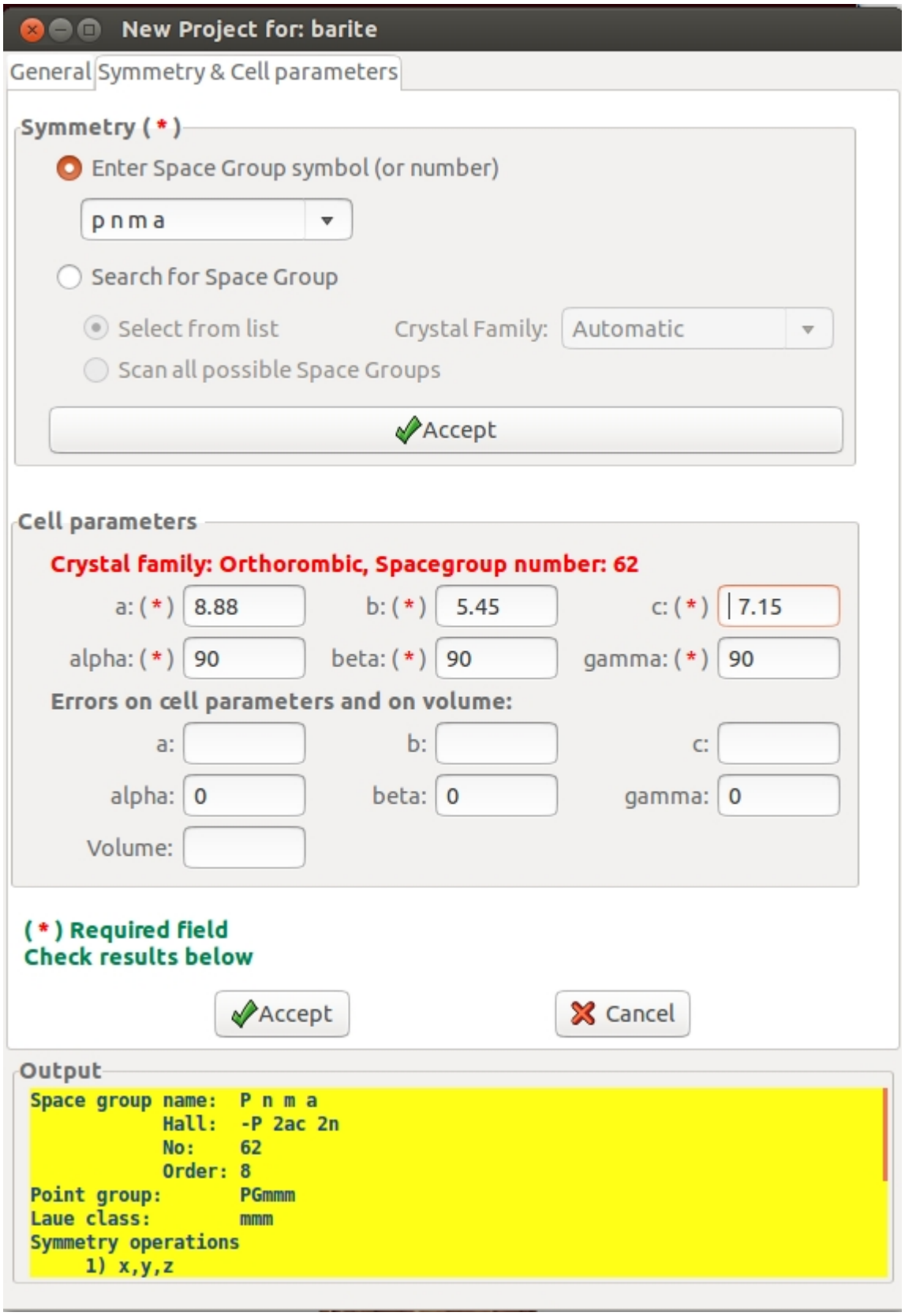
In the first window the user can select the working directory, assign a name to the structure or recover a previously created New Project file and supply a *title* string.



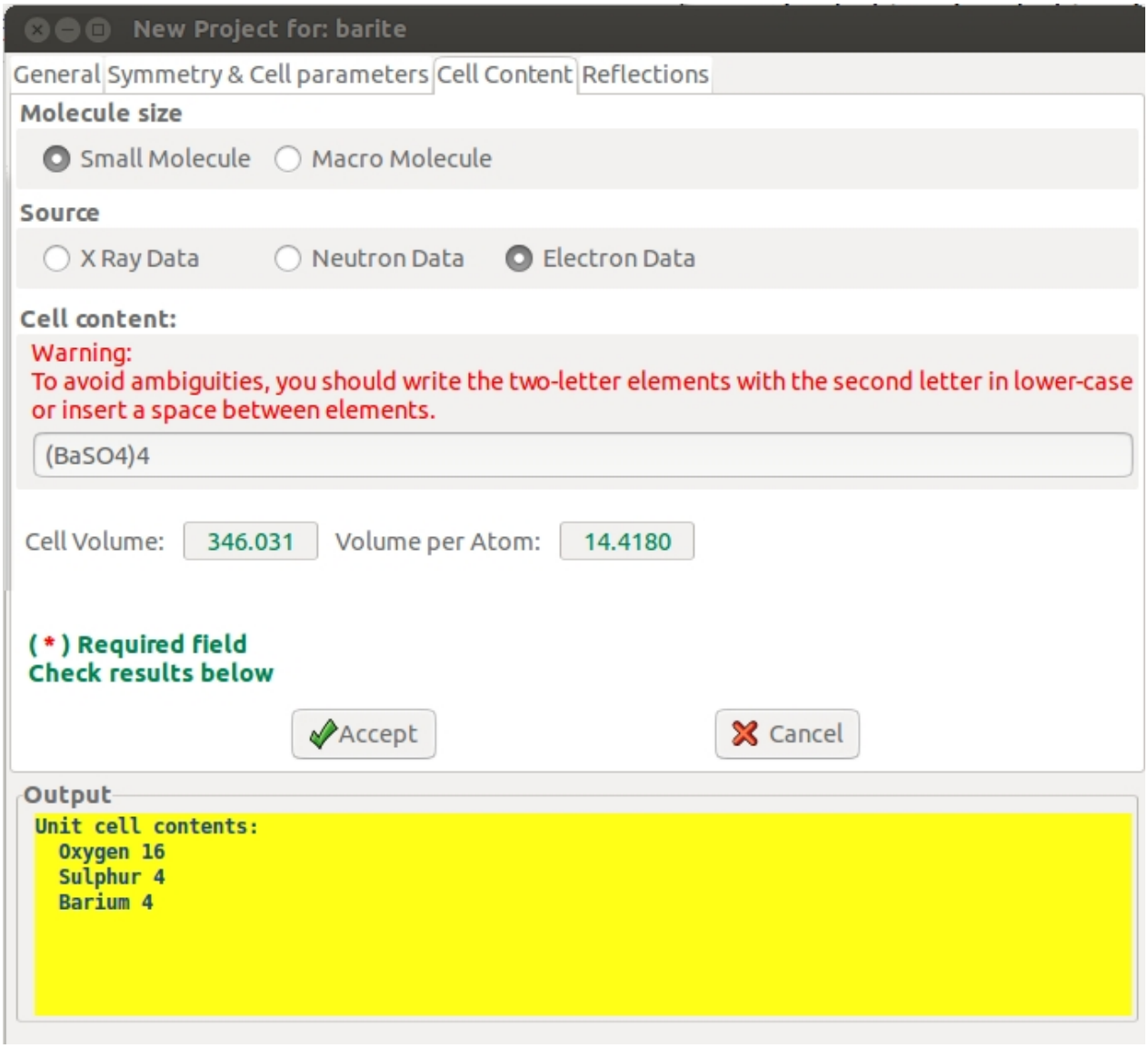
In the second window the user can select the space group symbol from a list, write it or its number in International Tables directly in the entry field.
If the user is not sure about the space group, it is possible to search for a possible space group (*Search for Space Group*) or to apply the program to all the most probable ones

(*Scan all possible Space Groups*).

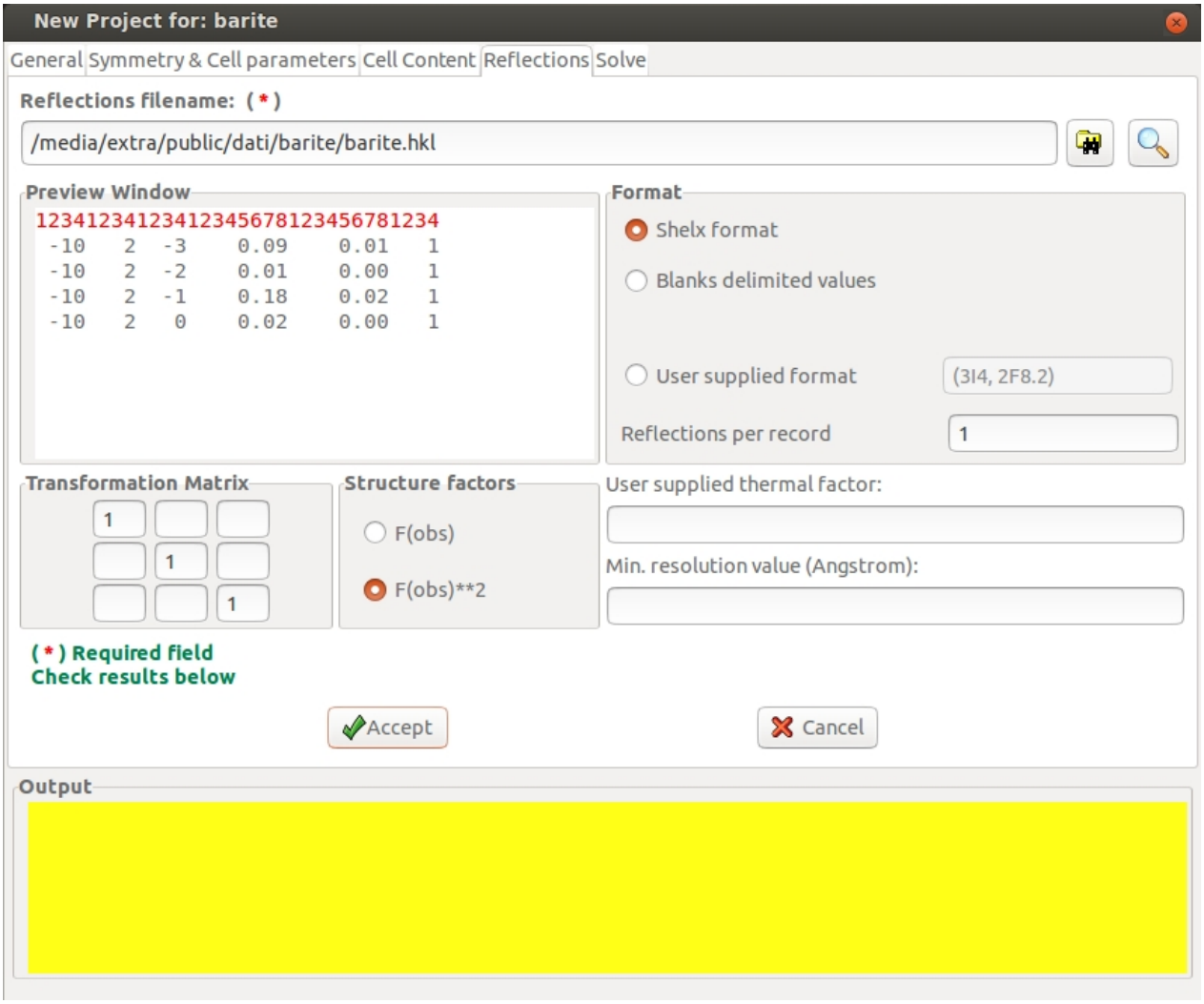
By clicking on **Accept** button the program will define the crystal family and provide the corresponding default values. Errors on cell parameters and on volume are optional.



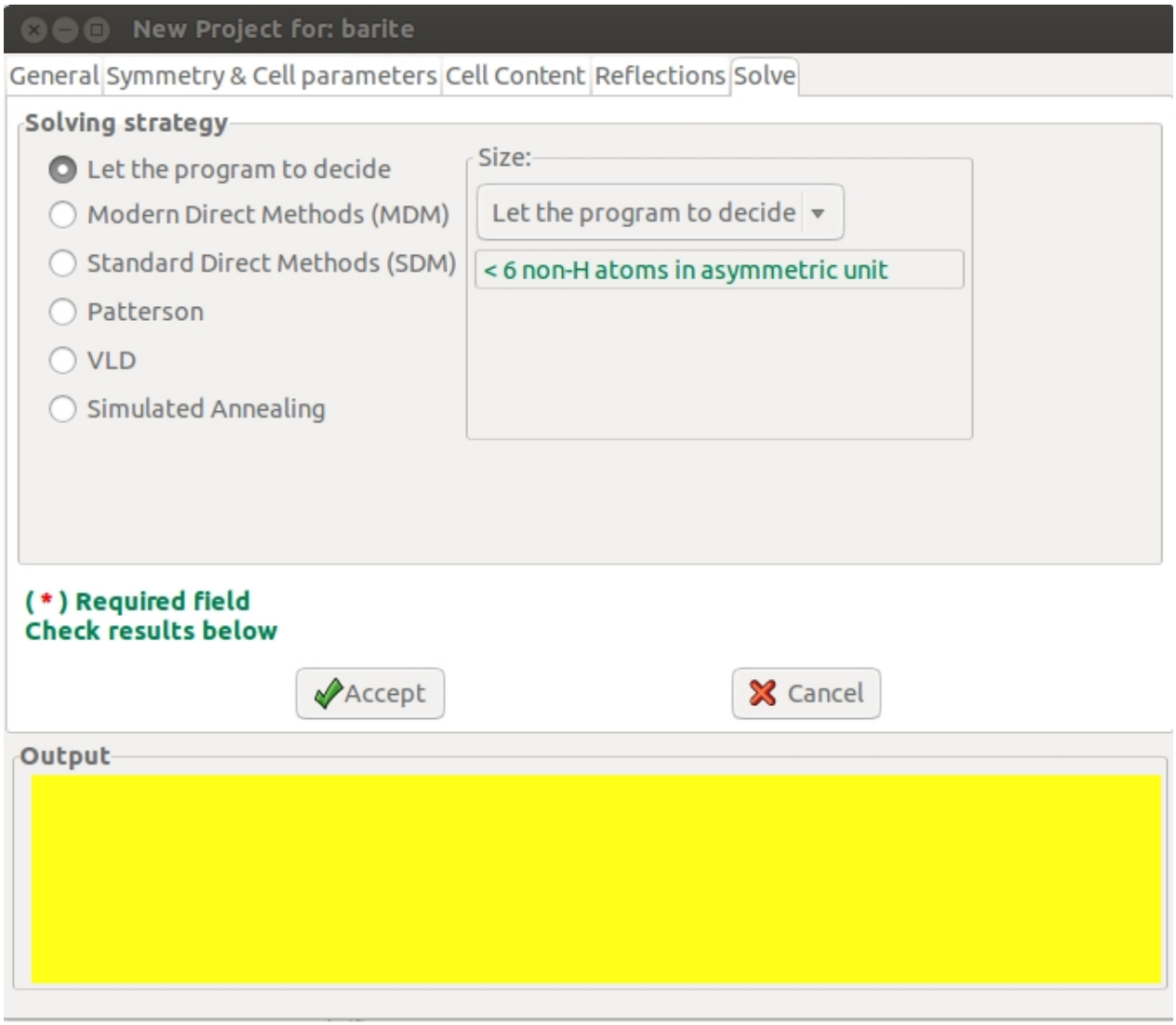
In the third window the user provides the unit cell content. The values for Delta f', f" and Wave Length are optional. In the following example electron diffraction area used, so "**Electron data**" has to be checked.



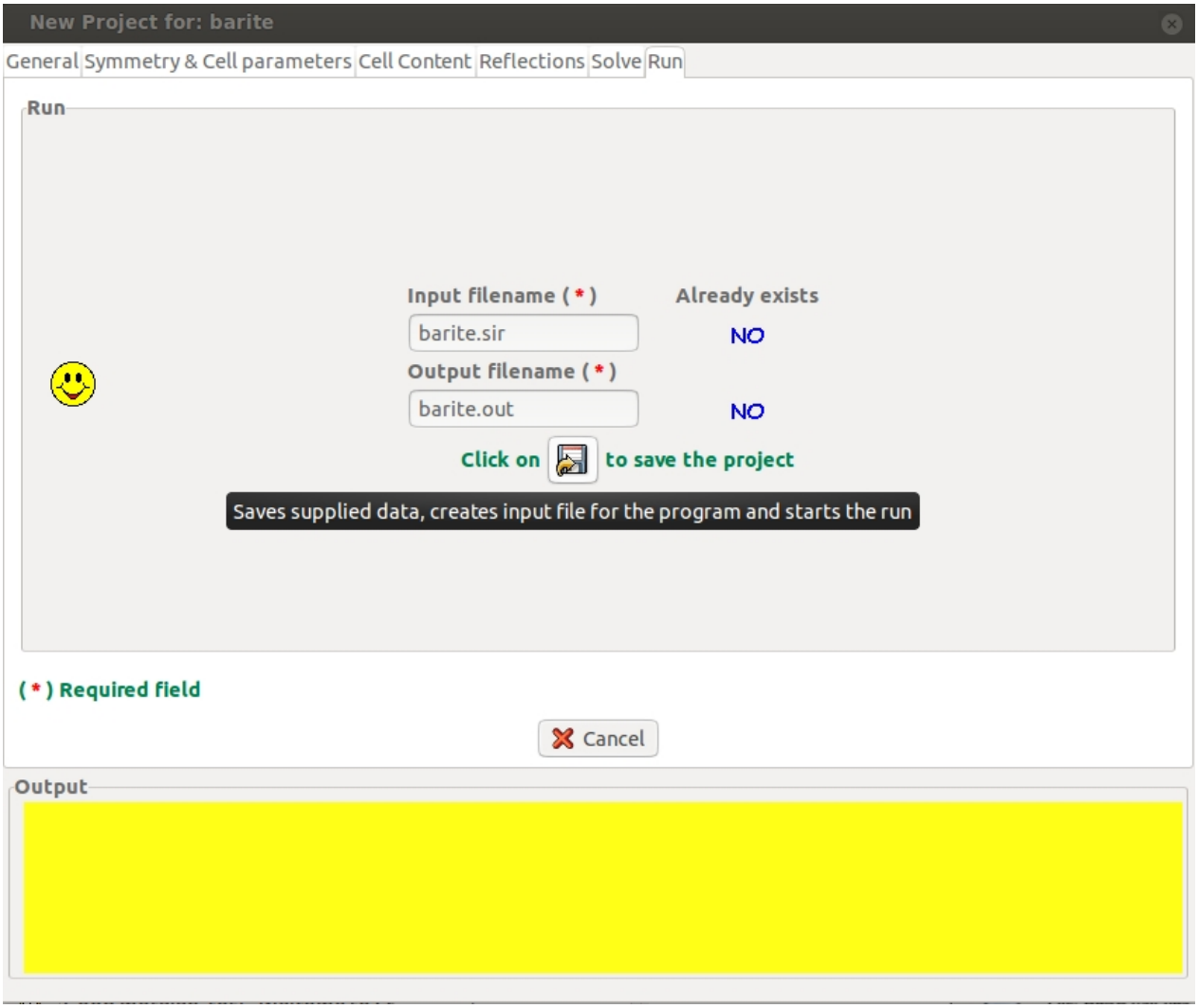
In the fourth window information about reflections has to be provided. Details about the possibilities offered in this window can be obtained looking at Directives: Data Routine. To define correctly the Format information, a preview Window can be activated by clicking on 🔍 button.



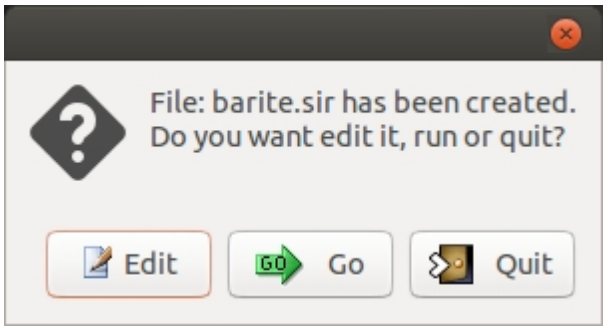
The fifth window requires information about the phasing procedure. Details about the possibilities offered in this window can be obtained looking in Description of Sir2014: Phase module or in Directives: Phase Routine.



The last window informs the user that an input file is going to be created. The user, by clicking on button, can save the input file for Sir2014

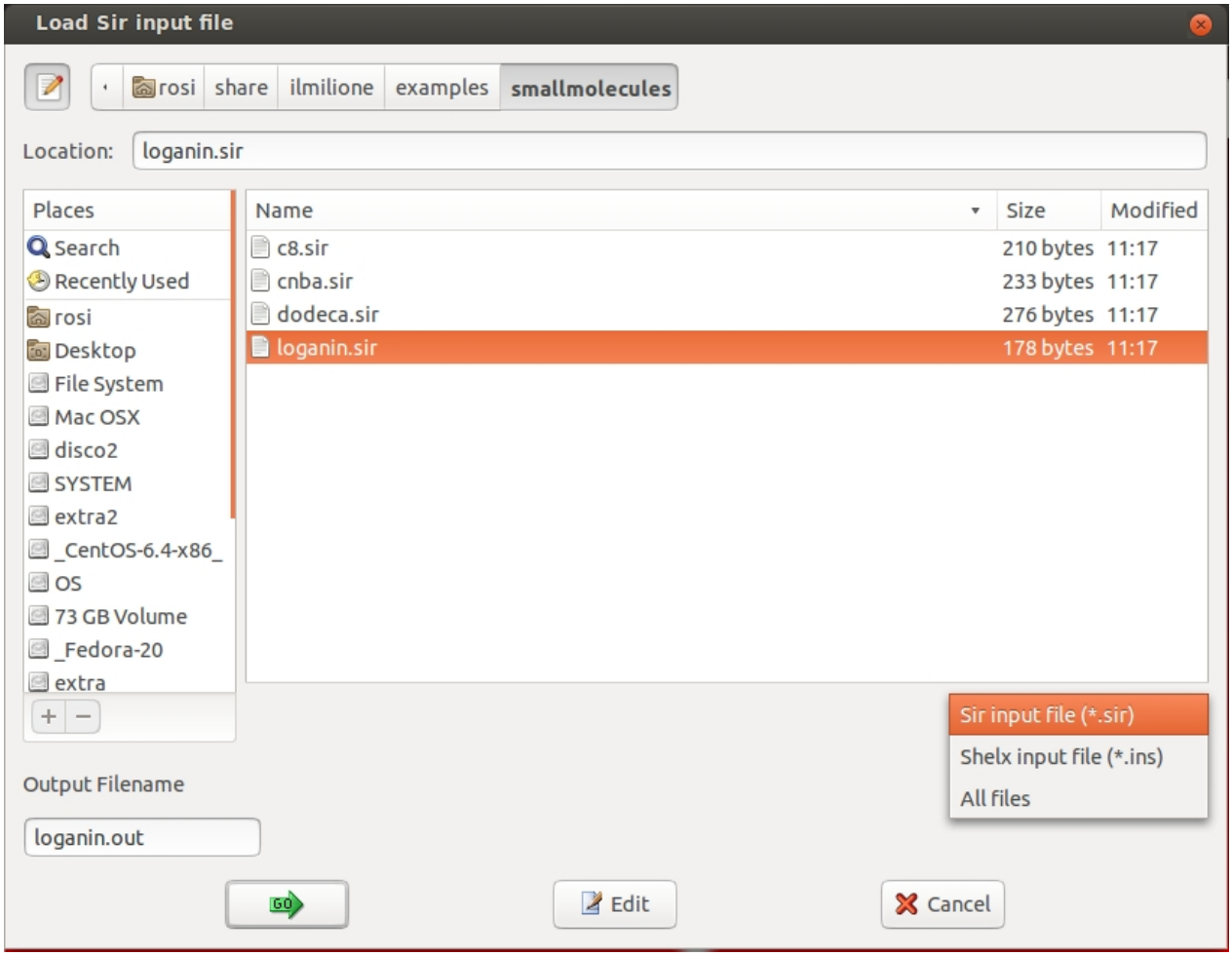


At this stage, the user can edit the input file just created, can start the program on this file or quit.



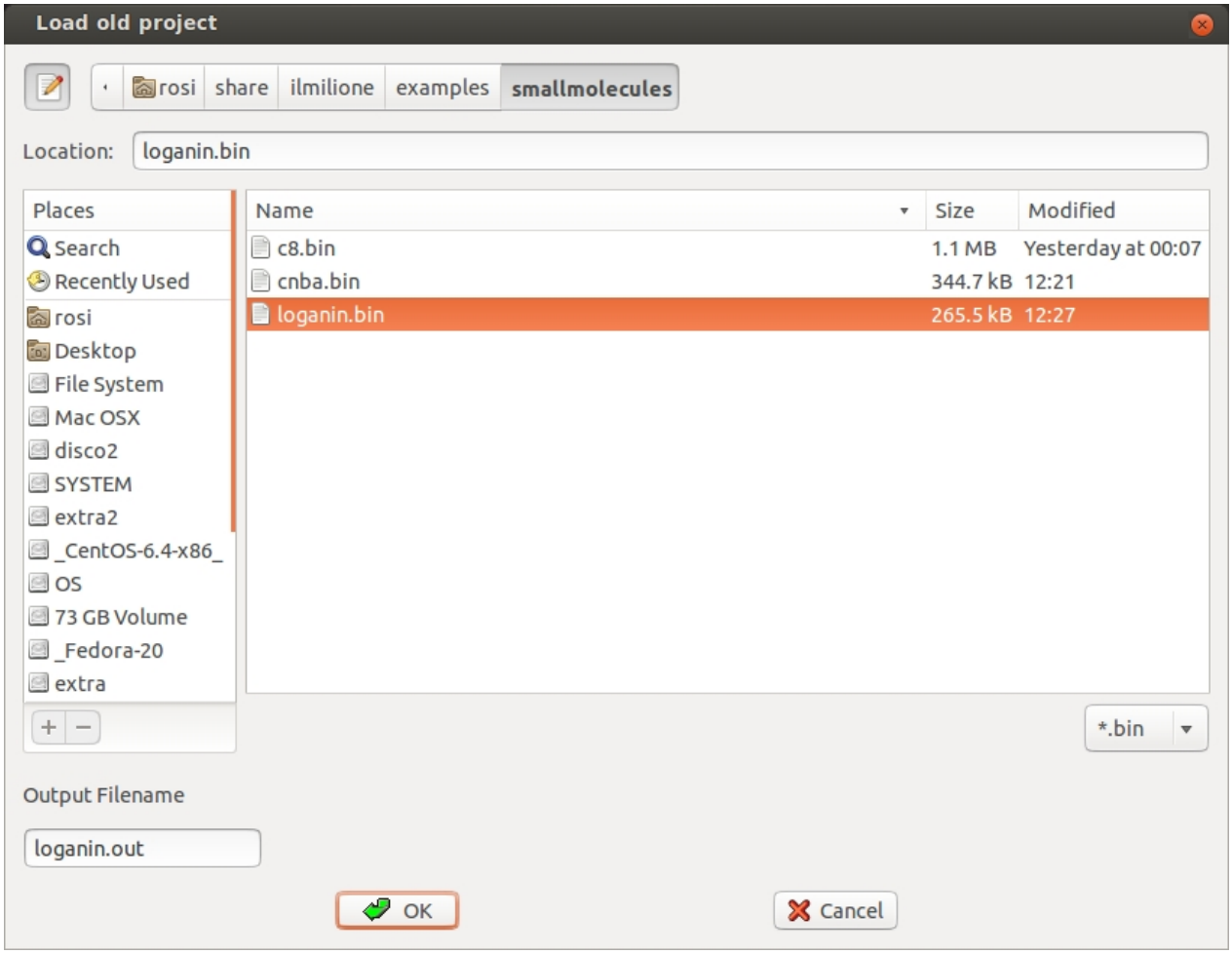
Load & Go

The file selection window will be opened in order to select the input file (.sir). The selected file can be edited using an internal editor.



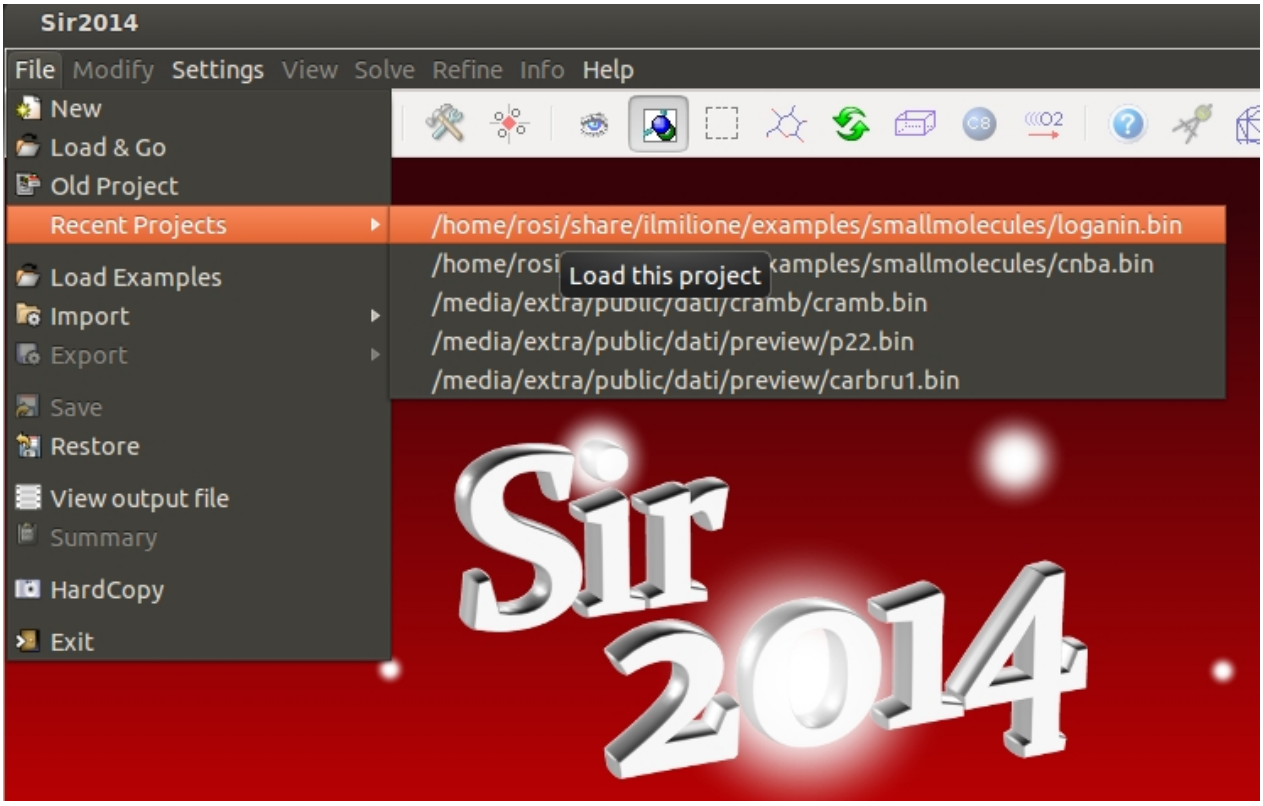
Old Project

The file selection window should be opened in order to select an old project file:



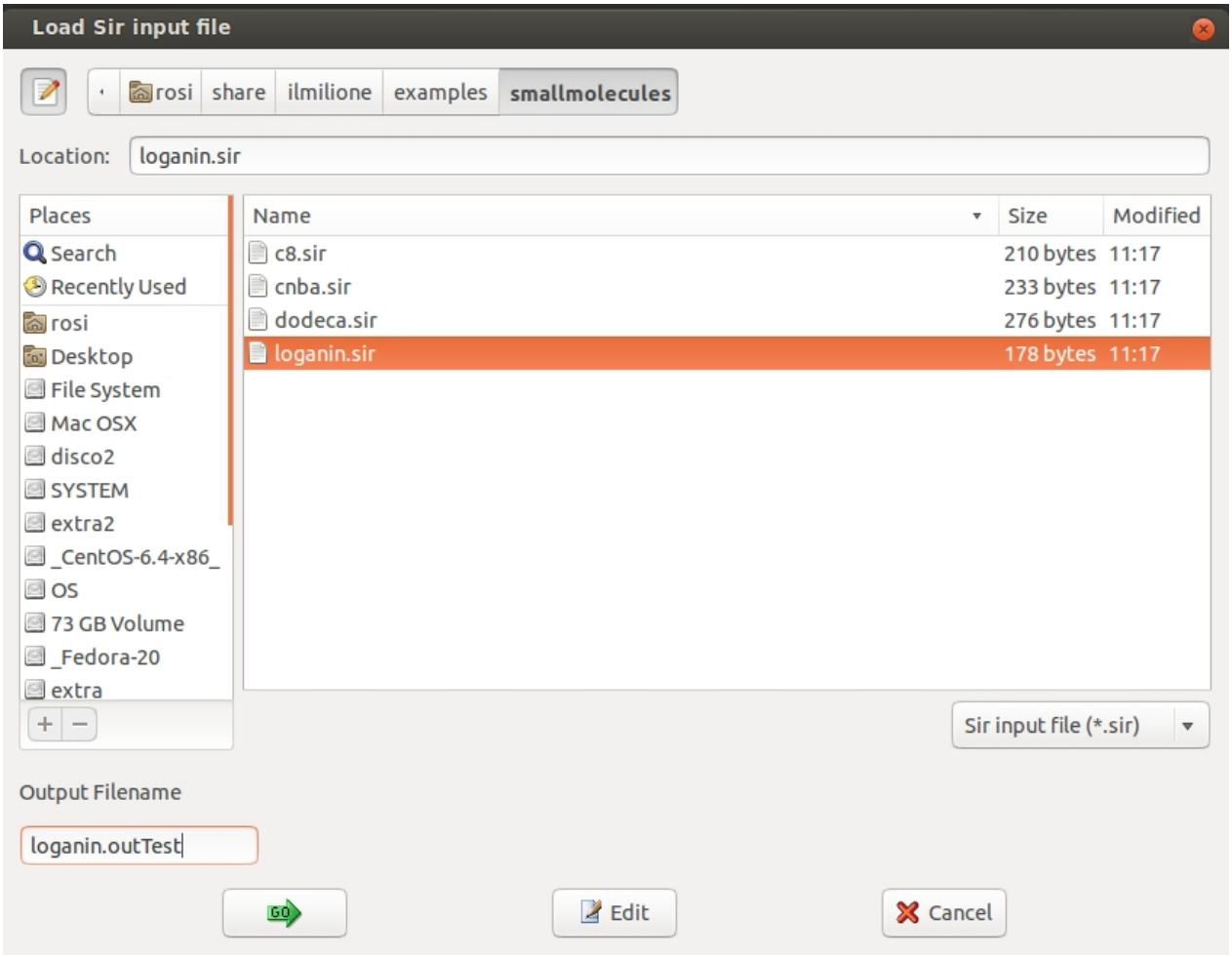
Recent Projects

The program stores the name of the 5 most recent projects opened by the program. The user can select among them.



Load Examples

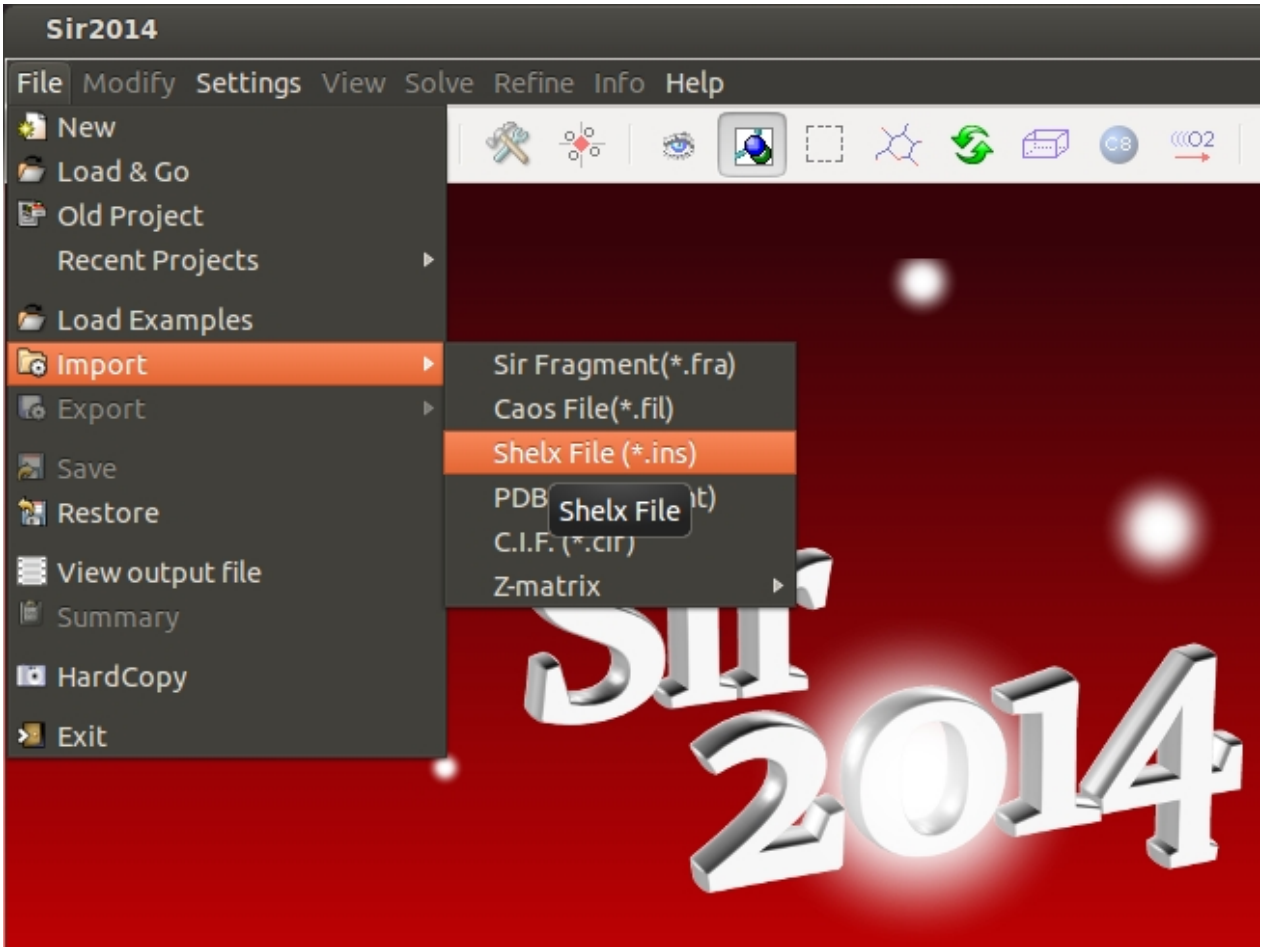
The file selection window will be opened in order to select the input file provided as example.



Import

Data from files in different format can be imported:

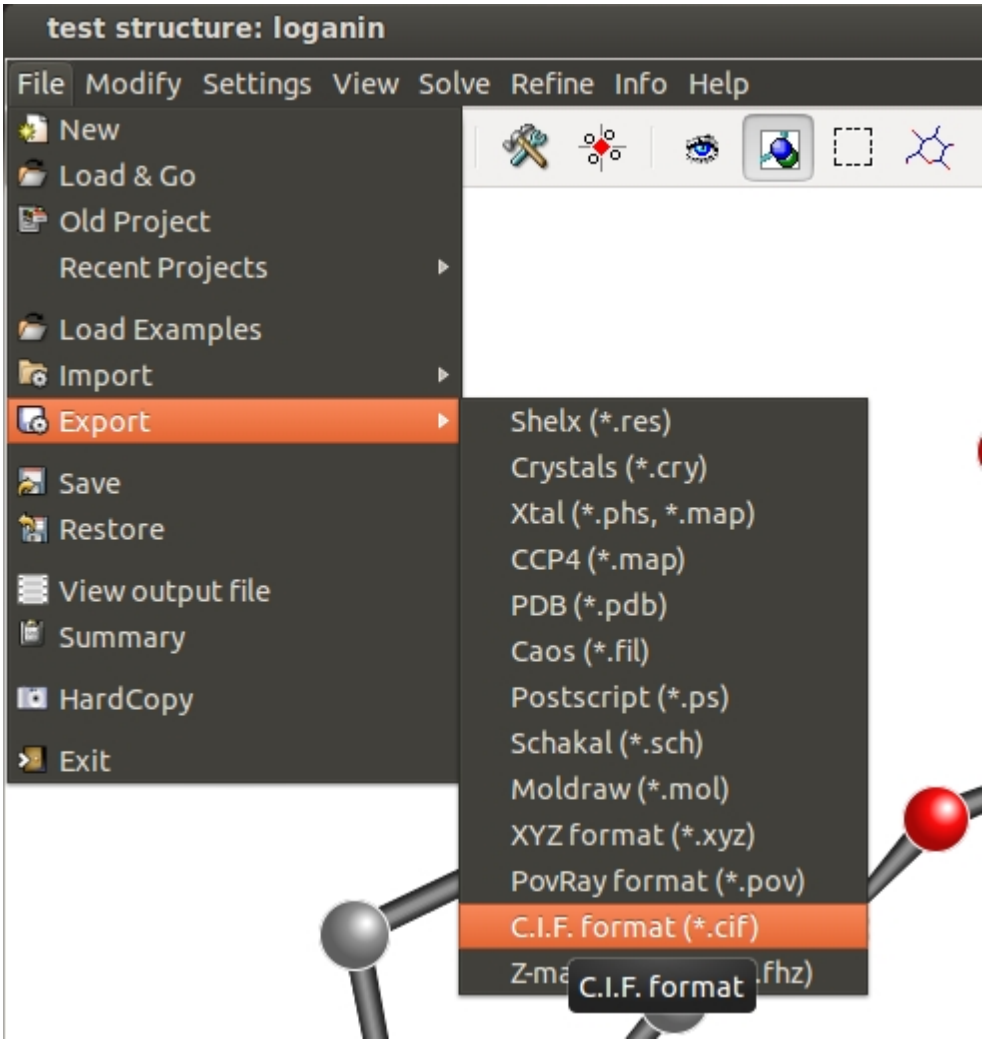
- Sir fragment (*.fra)
- Caos file (*.fil)
- Shelx file (*.ins)
- PDB file (*.pdb)
- CIF file (*.cif)
- Z-matrix file (*.zmt or *.mop)



Export

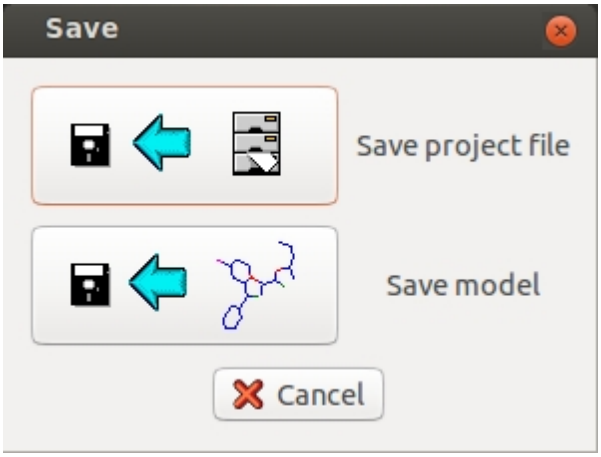
Sir2014 can produce files for different programs or applications:

- Shelx (a complete input file)
- Crystals (coordinate file)
- XtalView (reflections and map files)
- CCP4 (map file)
- PDB (coordinate file)
- CAOS (coordinate file)
- PostScript (image of the structure)
- HPGL (image of the structure)
- Schakal (a complete input file)
- Moldraw (a complete input file)
- XYZ (coordinate file)
- PovRay (a complete input file)
- C.I.F.(complete file)
- Z-matrix.(*.zmt)



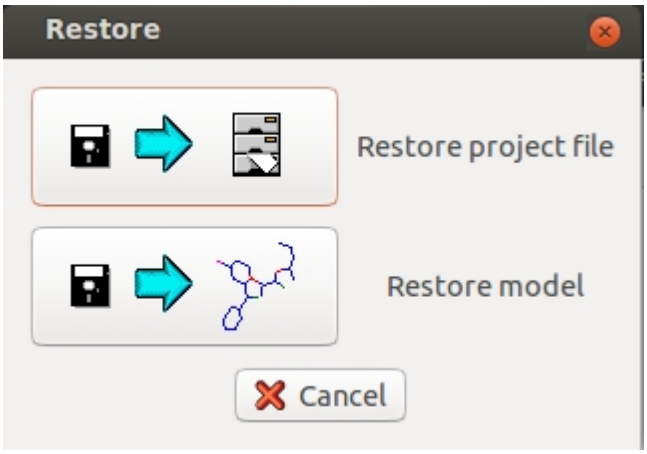
Save

The current project file or model is saved. It could be restored, in case of troubles, using the Restore feature



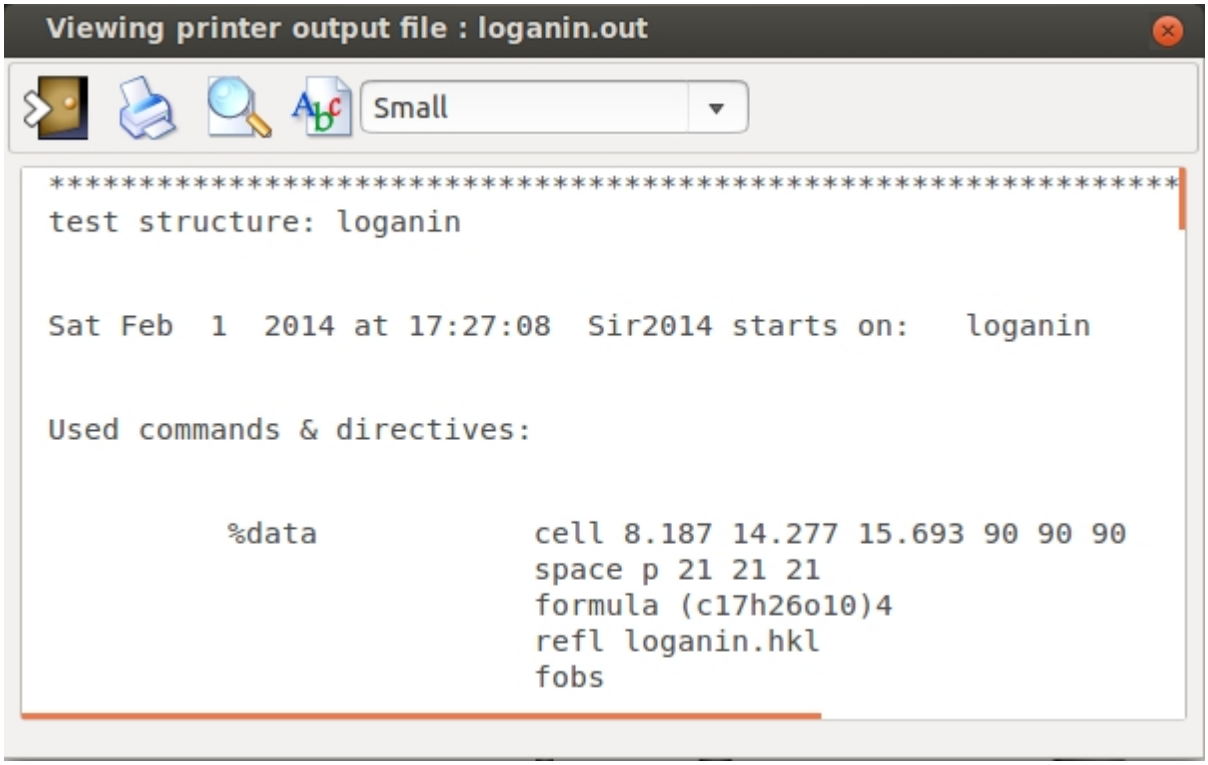
Restore

A previously saved project file or model is restored.



View Output File

It can be used to view the output file during the program run.

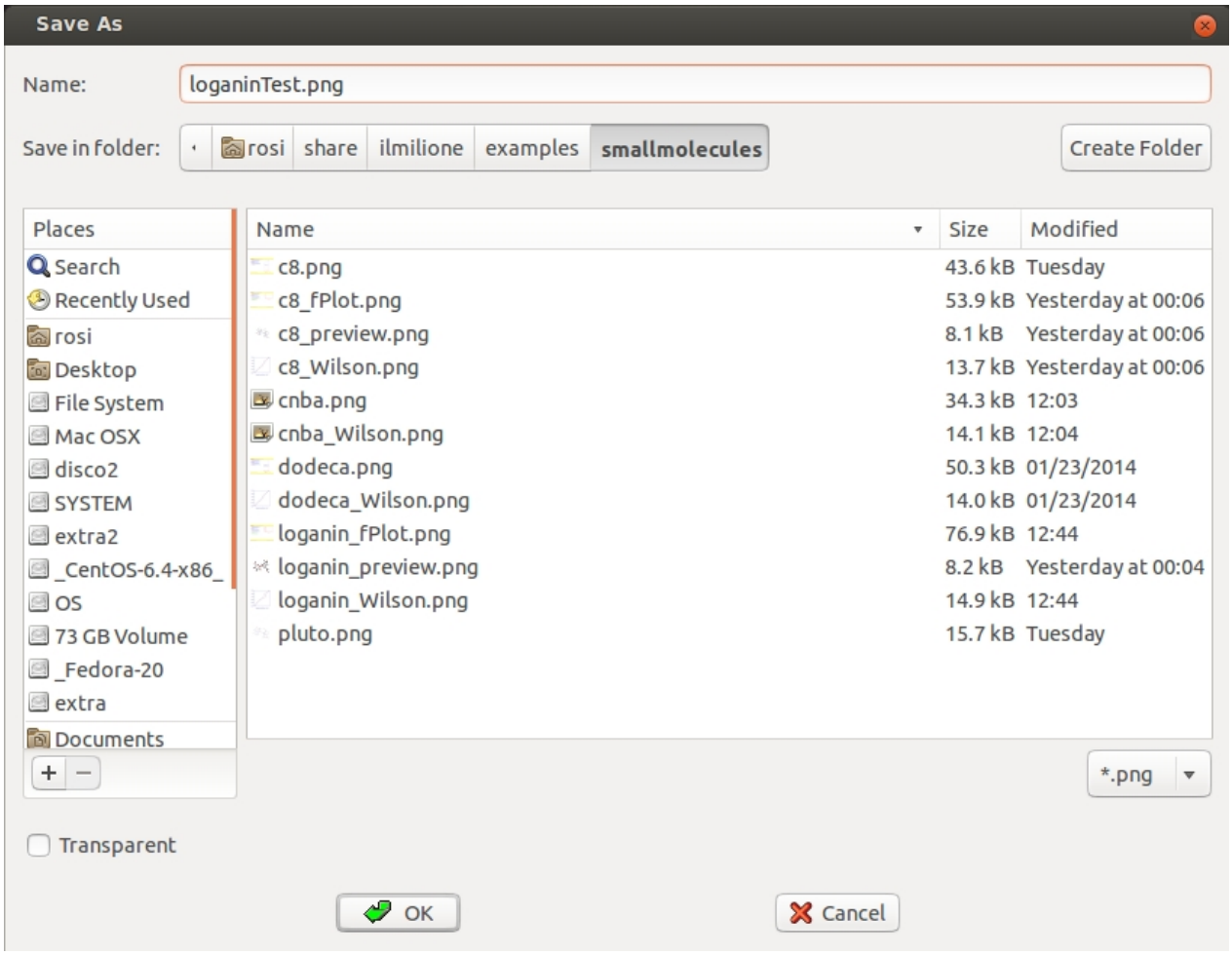


Summary

Starts the Internet Browser on a page containing a summary of the results.

Hard Copy

An image file (PNG) is created with the program window content.



Exit

This is used to exit the program.

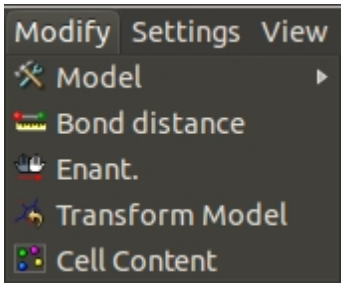
All relevant information will be stored and saved in the project (.BIN) file:

Created with the Personal Edition of HelpNDoc: Write eBooks for the Kindle

Modify

Possibilities offered by the sub-menu Modify shown on the left; some of these can be

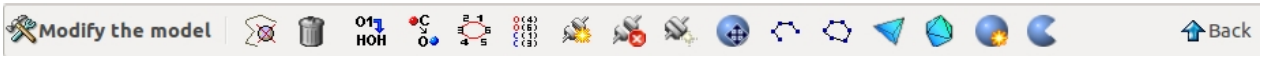
accessed directly using icons in the main toolbar.



Model	Offer many possibilities to modify the model.
Bond distance	Change the default bond distances.
Enantiomorph	Switch to the other enantiomorph.
Transform model	Apply symmetry transformations to the model.
Cell Content	Modify the unit cell content.

Model

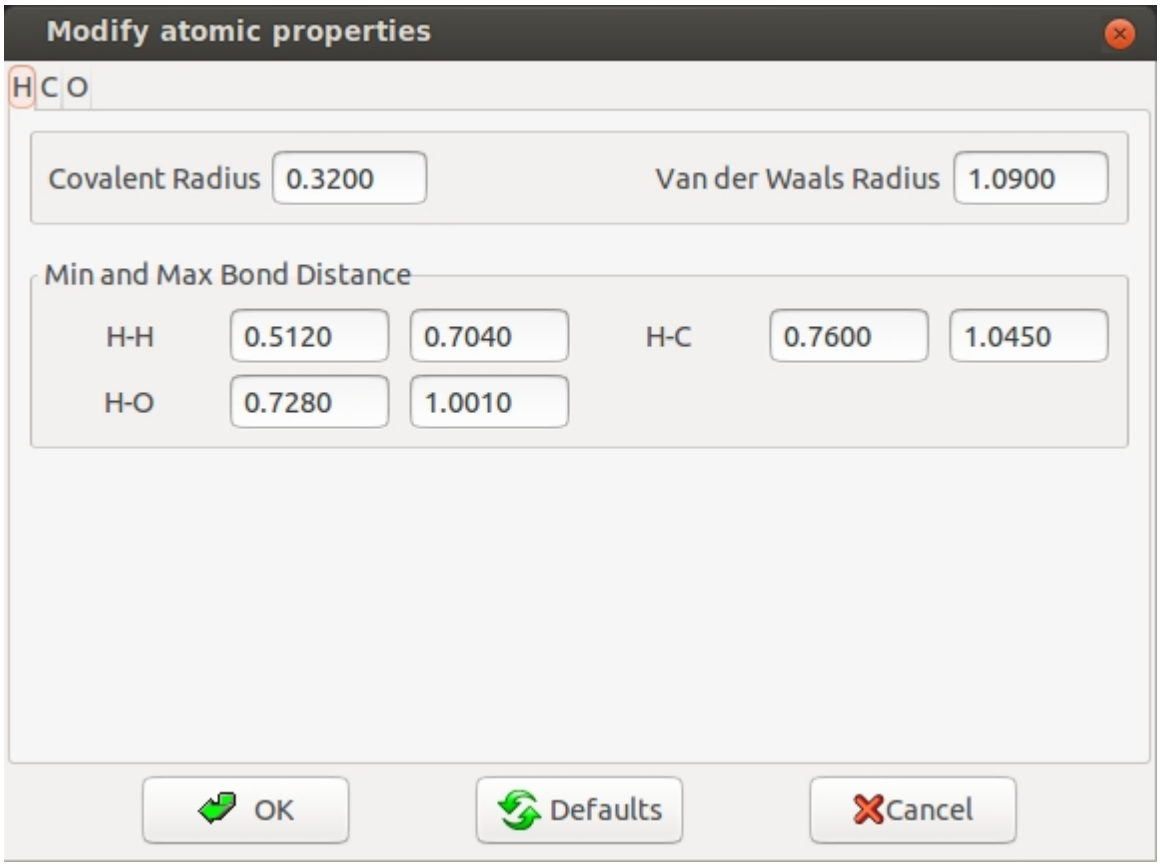
Tools are supplied in order to modify the model. It is possible to access them directly using the toolbar:



- Enclose in a poligon the atoms to be deleted.
- Click on atoms to be deleted.
- Click on atoms to change their label.
- Click on atoms to change their species.
- Click on atoms to change their sequence number.
- Sort atoms with respect to the atomic species.
- Click on couples of atoms to be bonded.
- Click on couples of atoms to delete their bond.
- Re-compute connectivity.
- Move a group of atoms in a symmetrically equivalent position.
- Complete 6-membered ring.
- Regularize 6-membered ring.
- Add atoms to complete a Tetrahedron.
- Add one atom to the model.
- Manage atom site (reduced occupancy, disorder, ...).

Bond Distance

This window is devoted to the modification of the bond distances between couples of atomic species:

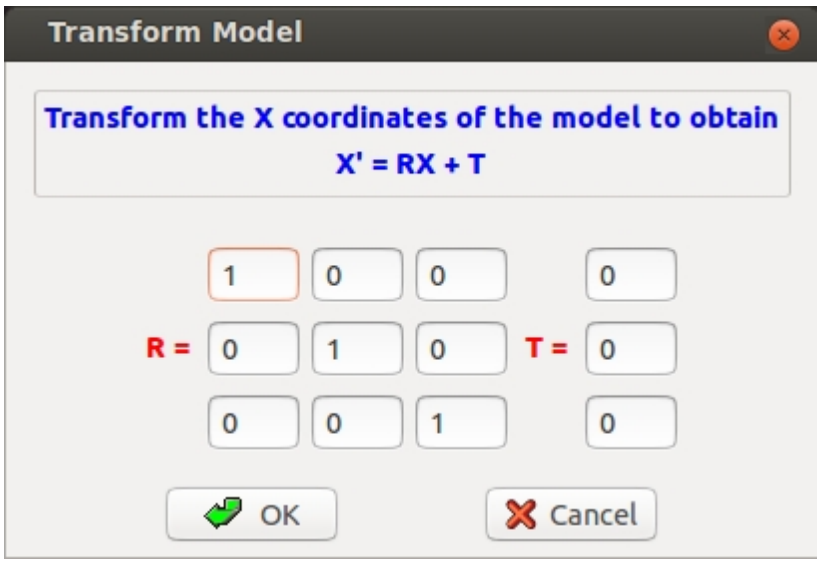


Enantiomorph

Makes it possible to switch between one enantiomorph and the other.

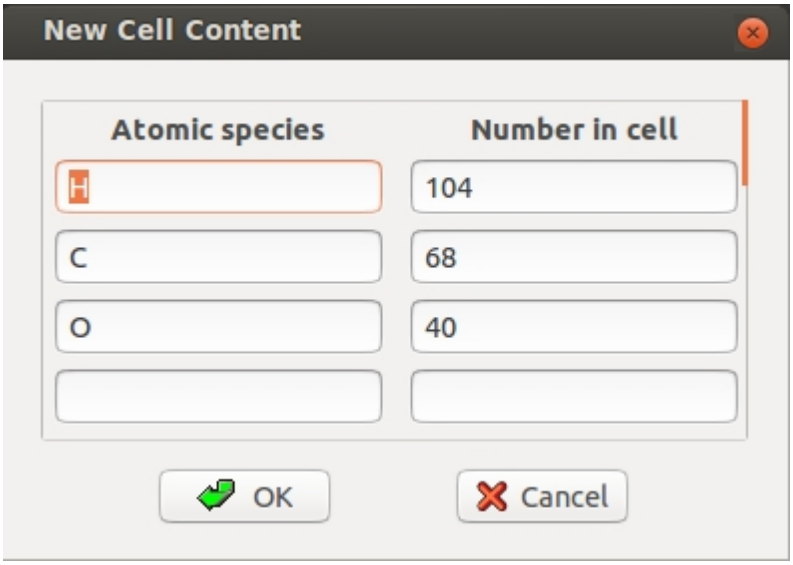
Transform Model

Make it possible to modify the asymmetric coordinates of the model applying a rotation and a translation operator.



Cell Content

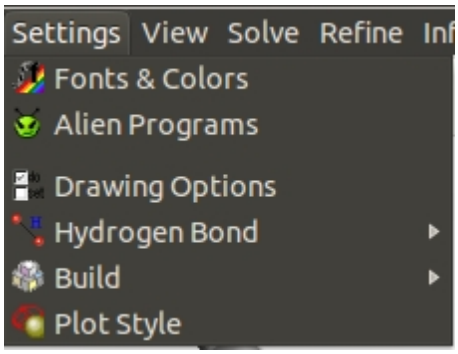
During the refinement the unit cell content declared can change (solvent molecules etc.) By means of this window the user can add/delete atomic species and modify the number of atoms in the unit cell.



Created with the Personal Edition of HelpNDoc: Easy EPub and documentation editor

Settings

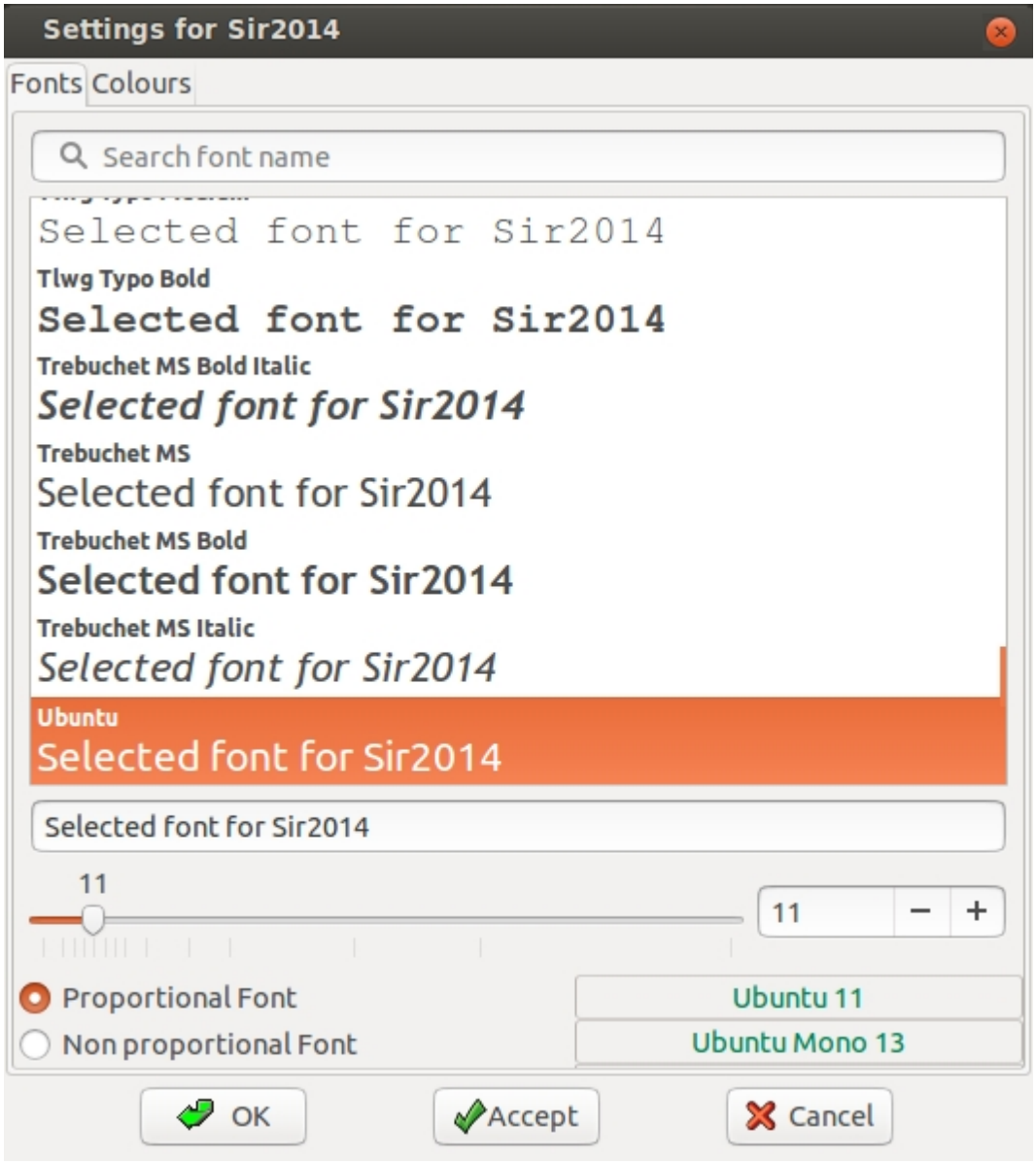
The sub-menu Settings is shown on the left. It allows to set:.



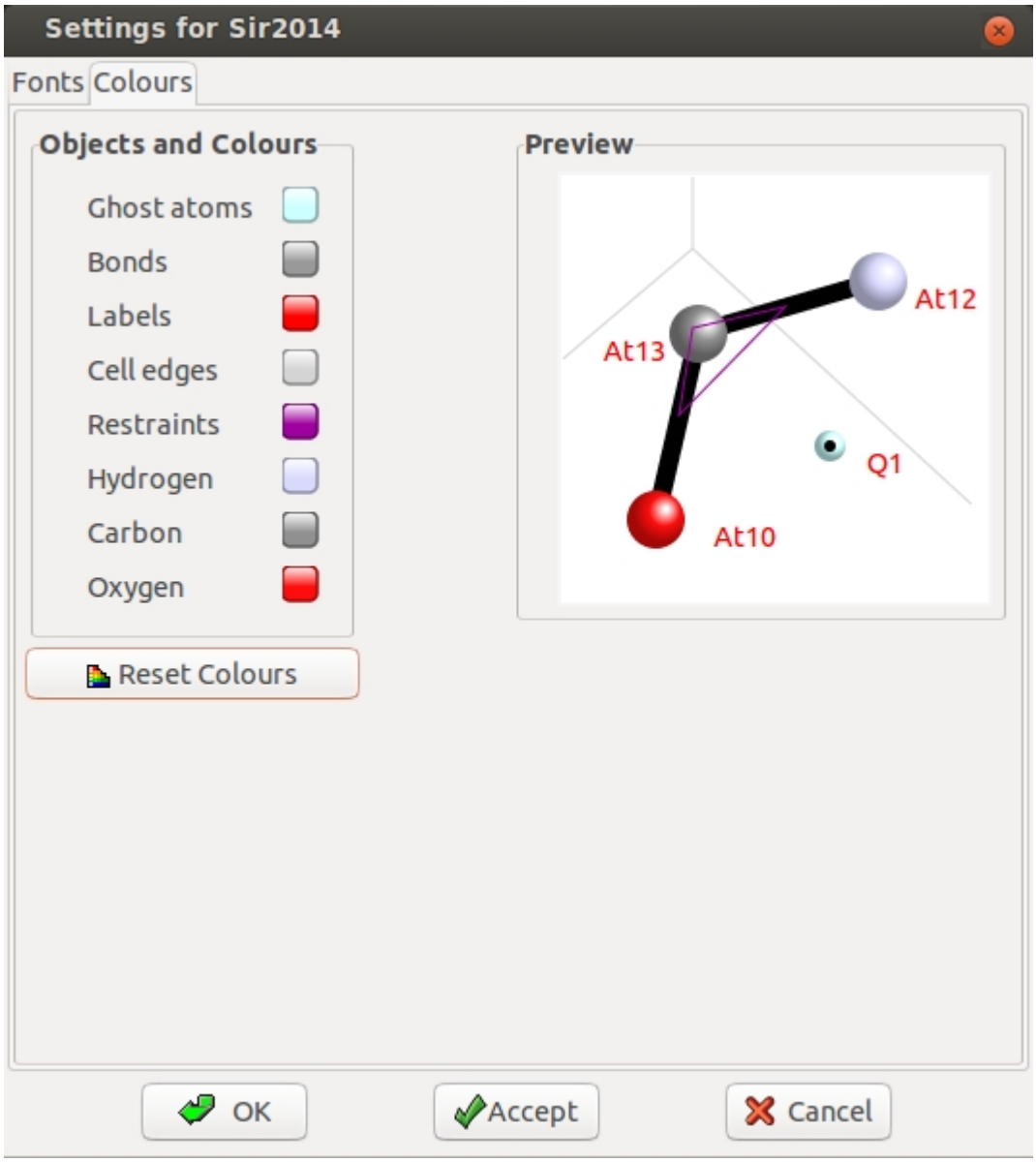
Fonts & Colors	Set fonts and colors used by the program.
Alien Programs	Install alien programs (Internet browser, PovRay).
Drawing Options	Many features for drawings.
Hydrogen Bond	Add or delete hydrogen bonds.
Build	Complete molecule using symmetry information.
Plot Style	Preferred representation for all or for selected atomic species.

Fonts & Colors

The user can select Proportional and non-Proportional Fonts as well as preferred colors using this menu item. To select fonts use the following window:



To change the default colors for the objects in the list use the following window:

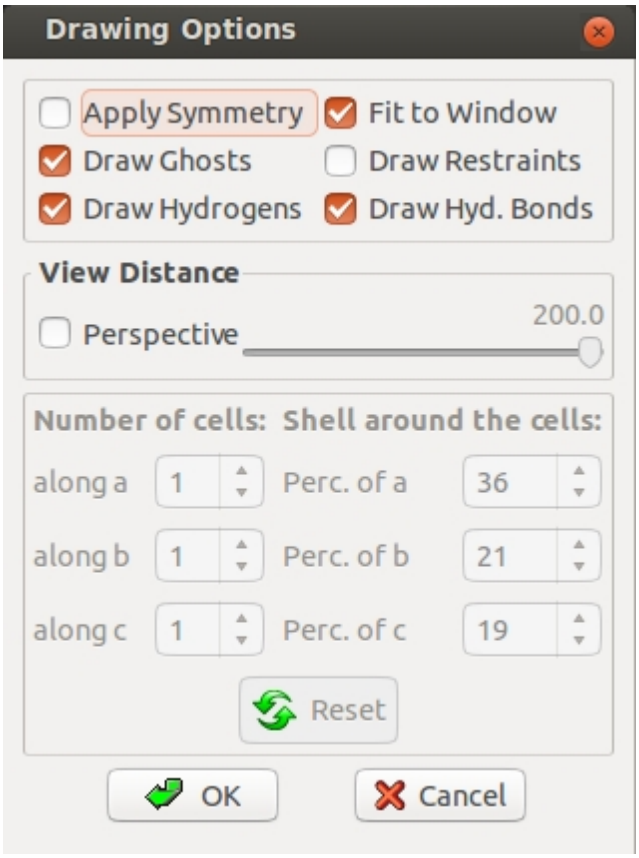


Alien Programs

The user can specify the path of ***PovRay*** (if installed) and change the default Internet Browser by means of the following window:

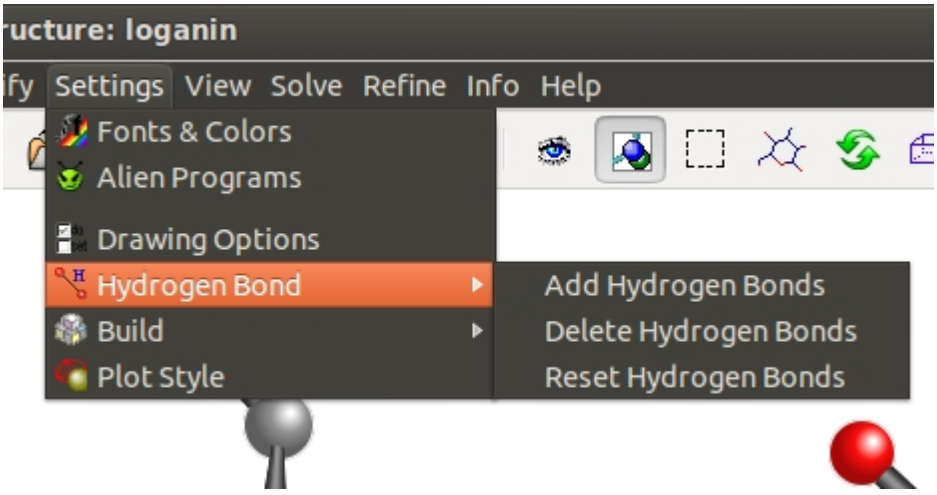
Drawing Options

Many features can be switched on or off such as: symmetry equivalent atoms, rescaling of the drawing with respect to the plot area etc. If perspective is checked, the View Distance can be changed in order to have a 3D view of the model. The number of cells along the three directions a, b and c and the percentage of area around the cells can be decided by the user.

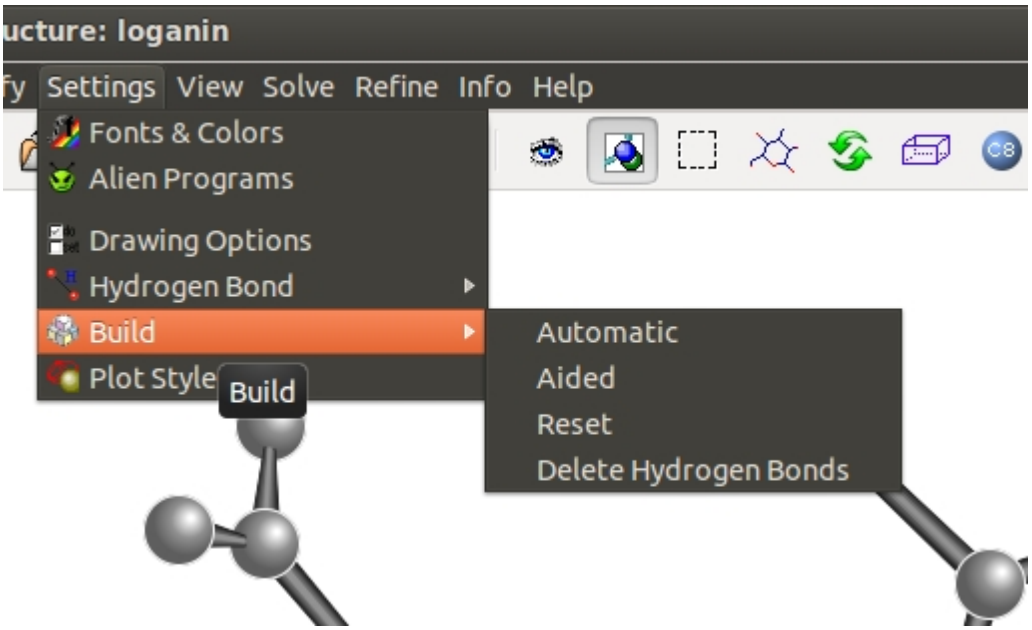


Hydrogen Bond

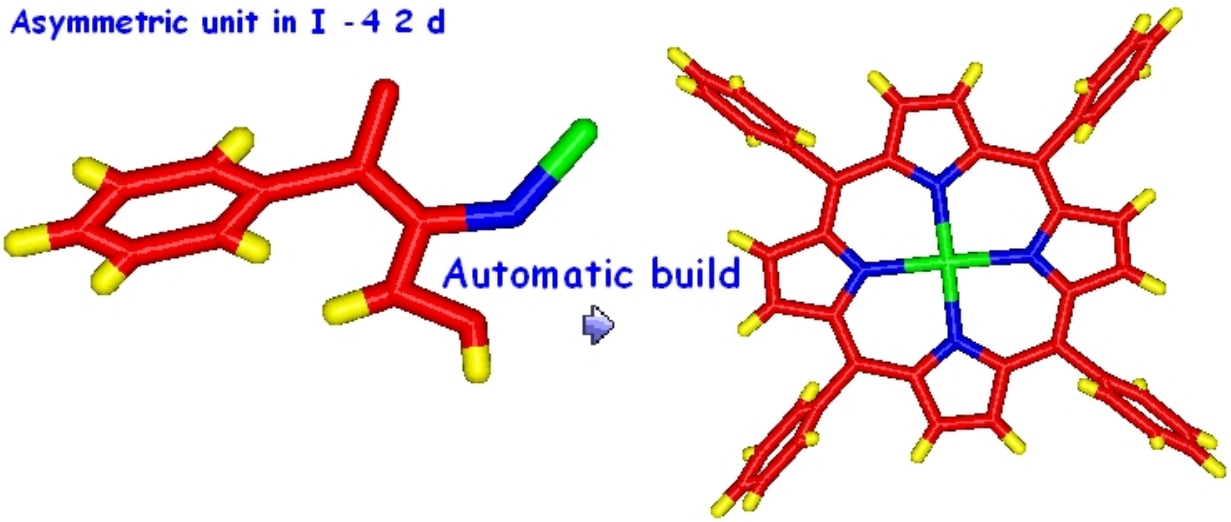
Make it possible to draw Hydrogen Bonds; the graphical representation of these bonds (dotted lines) can be used to produce suitable images.



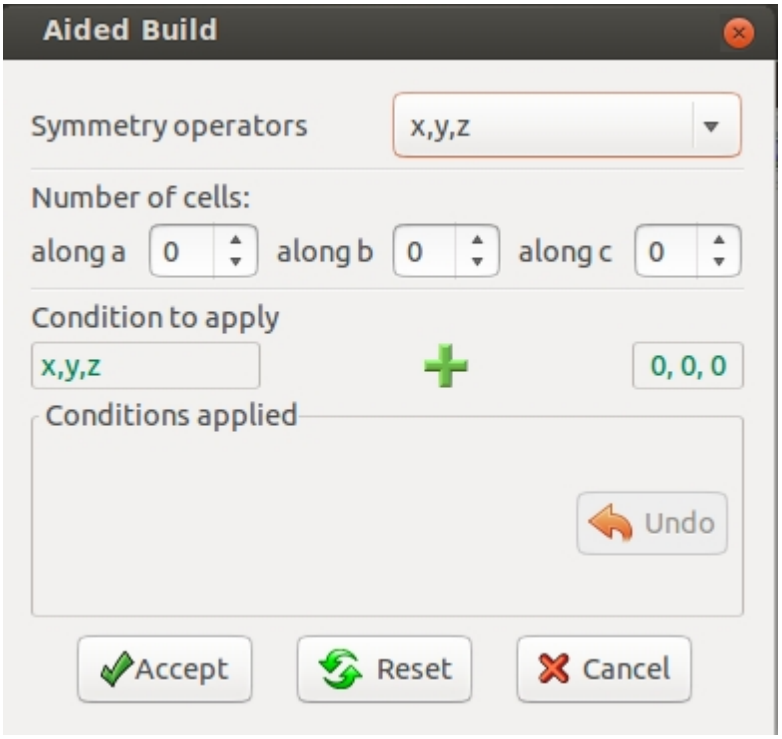
Build



It is possible to build, in a automatic way, the complete molecule, i.e.:

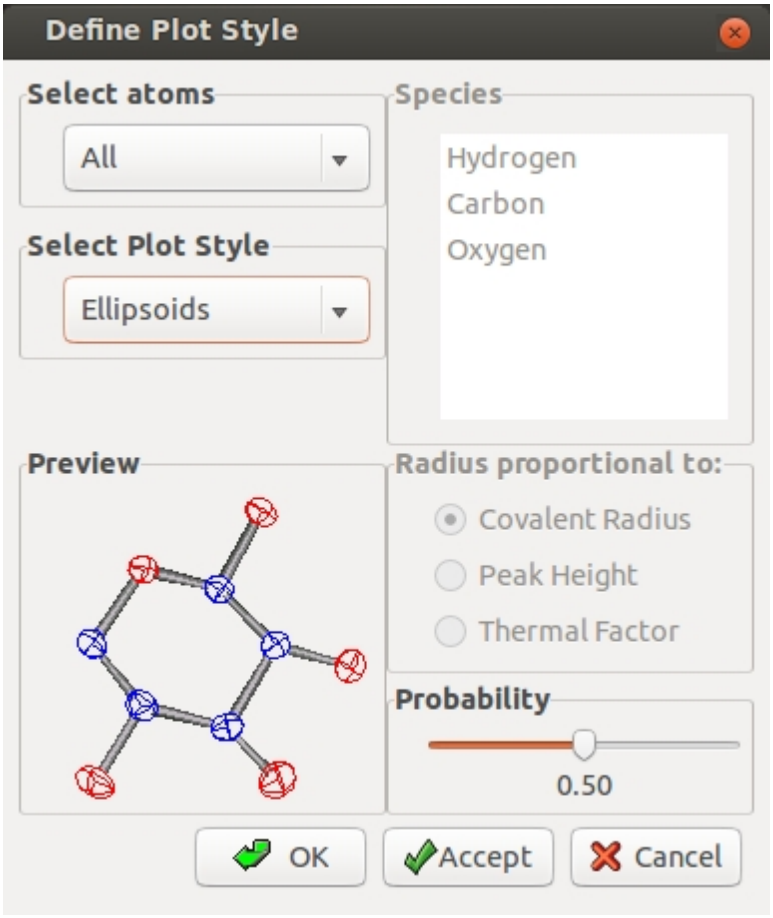


or to add different asymmetric units:



Plot Style

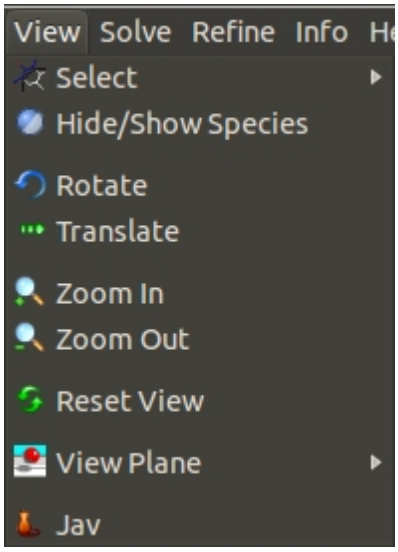
The user can select the preferred representation for all or for selected atomic species. The possible plot styles are: Sticks, Rods, Simple, Sticks & Balls, Bubbles, Shading, Van der Waals, Ellipsoids, Polyhedra. The effect is checked in the **Preview** area. It is possible also to draw atoms using different radius.



Created with the Personal Edition of HelpNDoc: Free EPub producer

View

Possibilities offered by the sub-menu View shown on the left; some of these can be accessed directly using icons in the main or in the view toolbar.

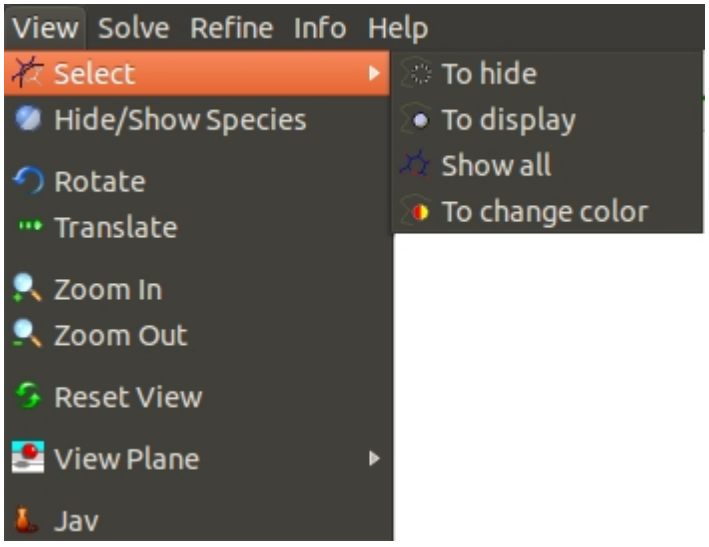


Select	Select part of the model.
Hide/Show Species	Hide/Show the atoms corresponding to the selected species.
Rotate	Rotate the model.
Translate	Translate the model.
Zoom In	Magnify the model.
Zoom Out	Un-magnify the model.
Reset View	Let the model fit the drawing window.
View Plane	Select View Plane.

Jav Start the Jav interface.

Select

In order to investigate part of the structure, the user has the possibility to use the following tools:



Hide atoms inside a polygon.

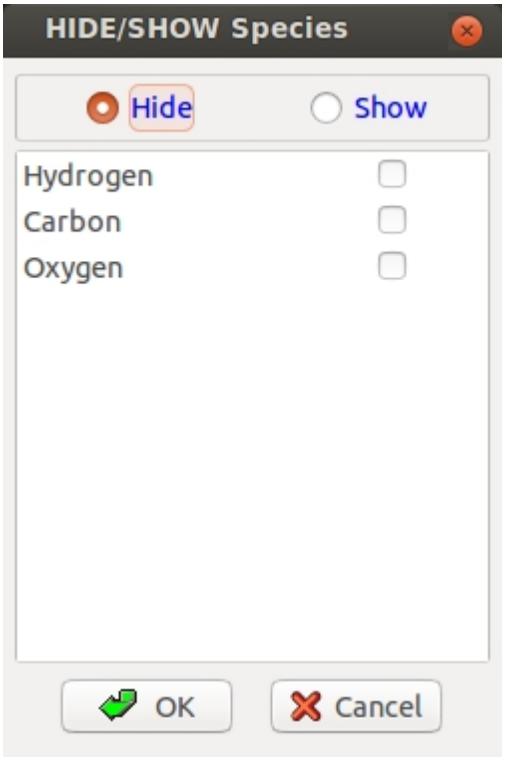
Show only atoms inside a polygon.

Show all atoms of the model.

Change color to atoms inside a polygon.

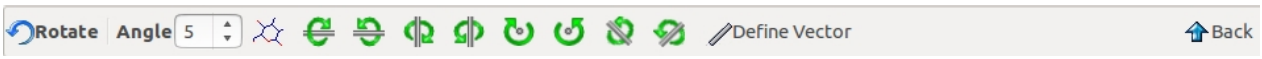
Hide/Show Species

Hide/Show the atoms corresponding to the selected species.



Rotate

The user can, by means of the following toolbar, define a rotation angle and rotate the model (using positive or negative angles) around X, Y or Z. It is possible also to define a vector around which the model can be rotated (by clicking on the "**Define Vector**" button and by selecting two atoms).



Translate

The user can, by means of the following toolbar, define a shift as a percentage of the cell and translate the model (in the positive or negative direction) along X, Y or Z.

Zoom In

The structure can be moved forward with respect to the screen.

Zoom Out

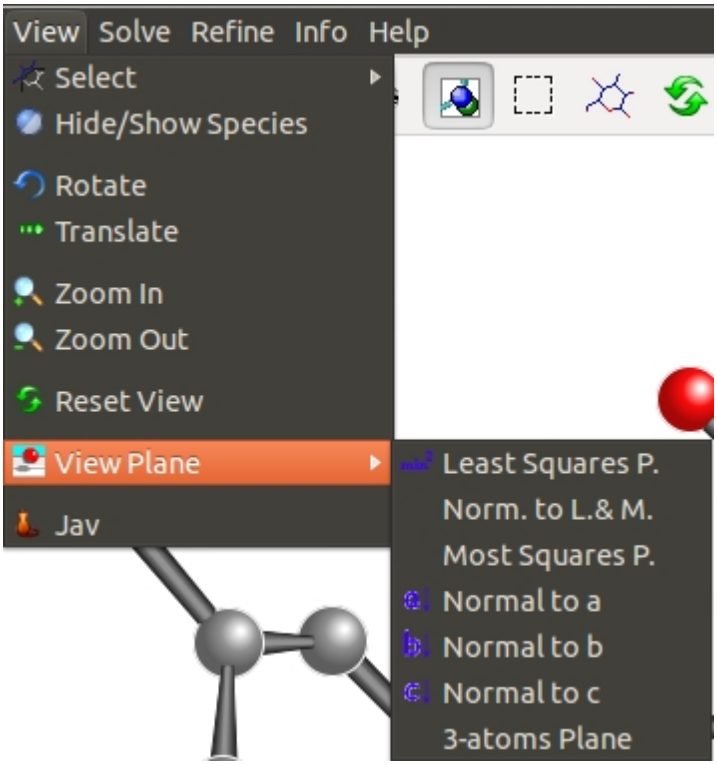
The structure can be moved backward with respect to the screen.

Reset View

The structure can be rescaled in order to fit the drawing area.

View Plane

View plane definition:



min^2 Project model on Least Squares Plane.

Project model on plane normal to Least Squares Plane.

Project model on Most Squares Plane.

a↓ Project model on the plane normal to a.

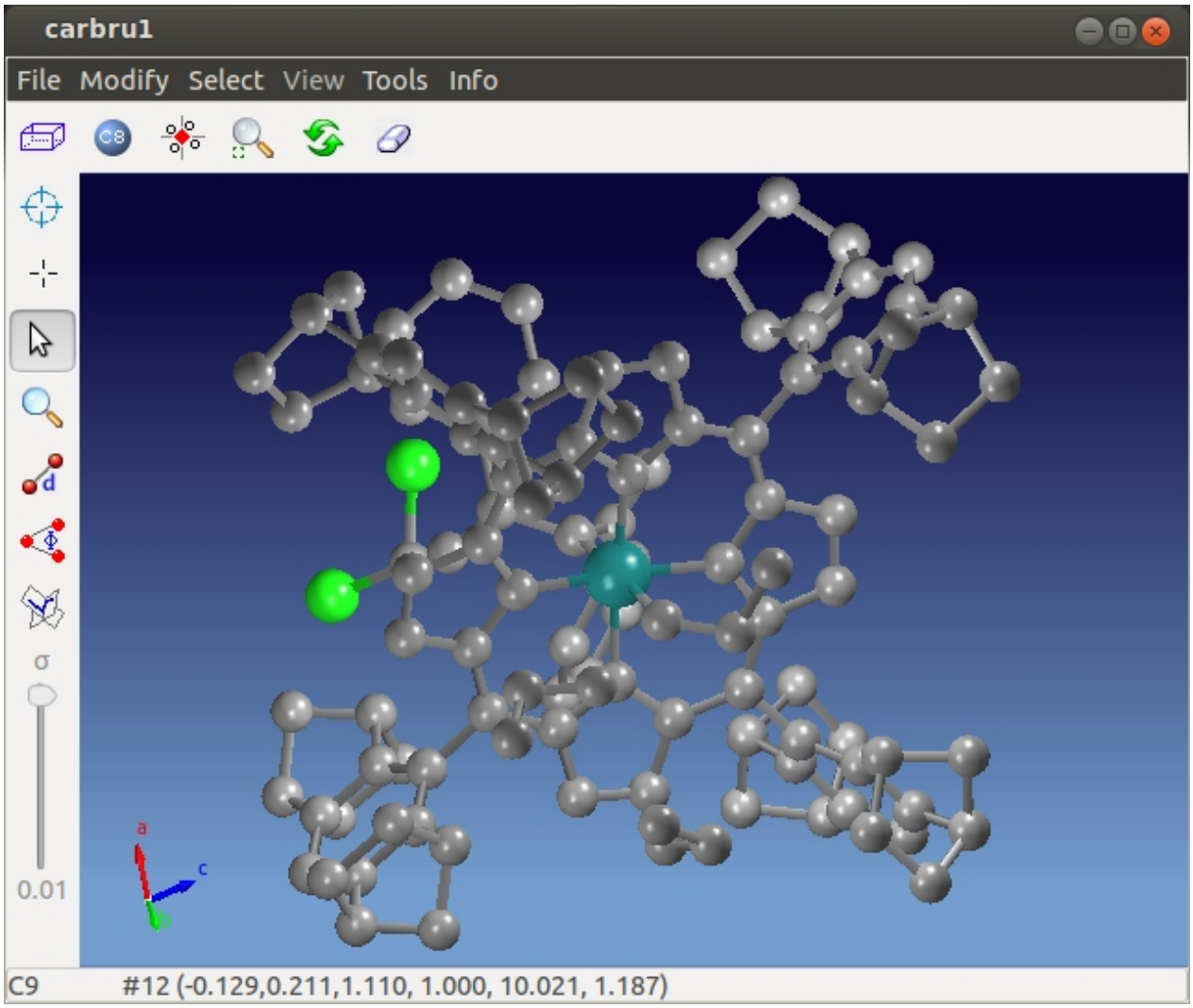
b↓ Project model on the plane normal to b.

c↓ Project model on the plane normal to c.

Project model on the plane defined by 3 or more atoms.

Jav

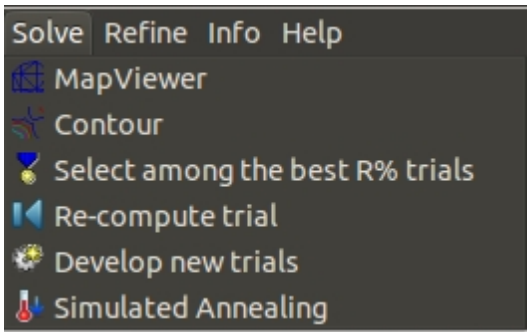
The Jav interface is started. For details see the help on-line of Jav.



Created with the Personal Edition of HelpNDoc: Easily create EBooks

Solve

The sub-menu Solve is shown on the left. It allows: .

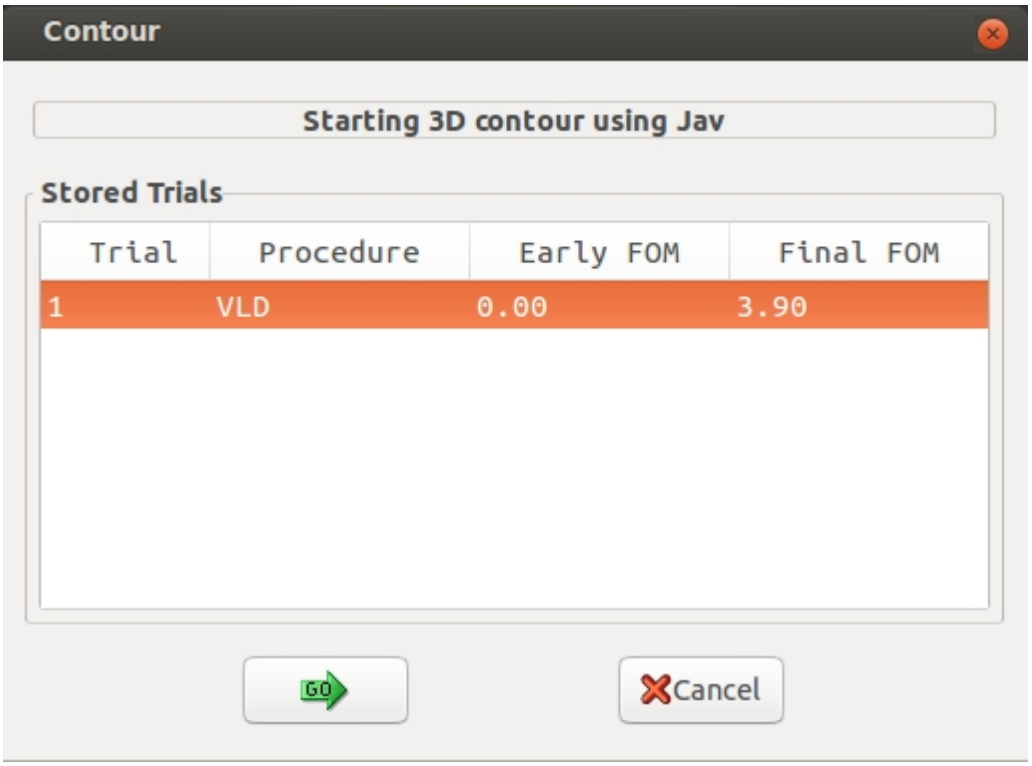


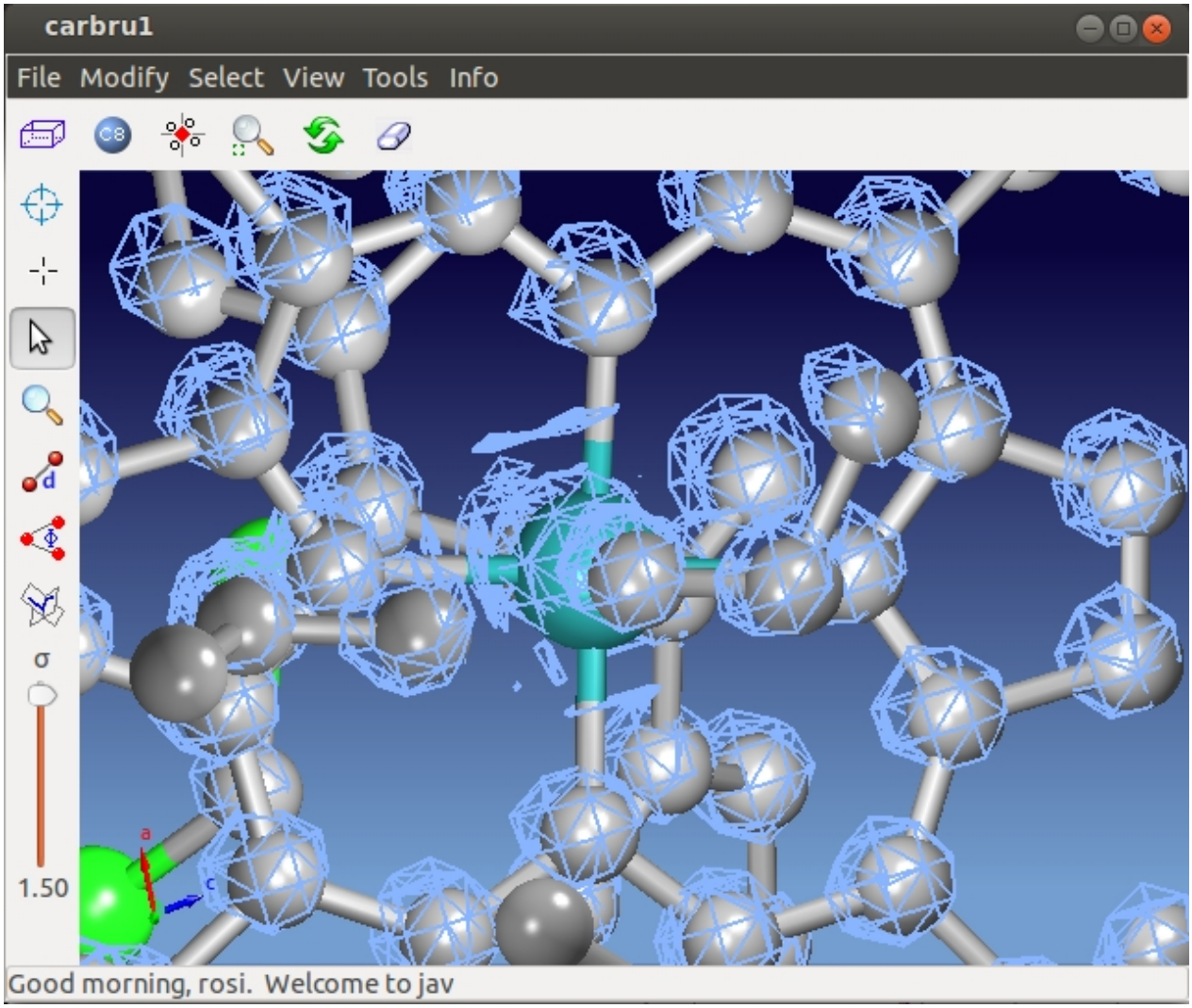
Map Viewer	Start Jav loading the current model and the related map.
Contour	2D contour facility.
Select among the best R% trials	Select and recover the model among the lowest R% trials.
Re-compute trial	Re-compute the atomic model using a previous trial.
Develop new trials	Develop new trials.
Simulated	Start the Simulated Annealing

Annealing procedure.

Map Viewer

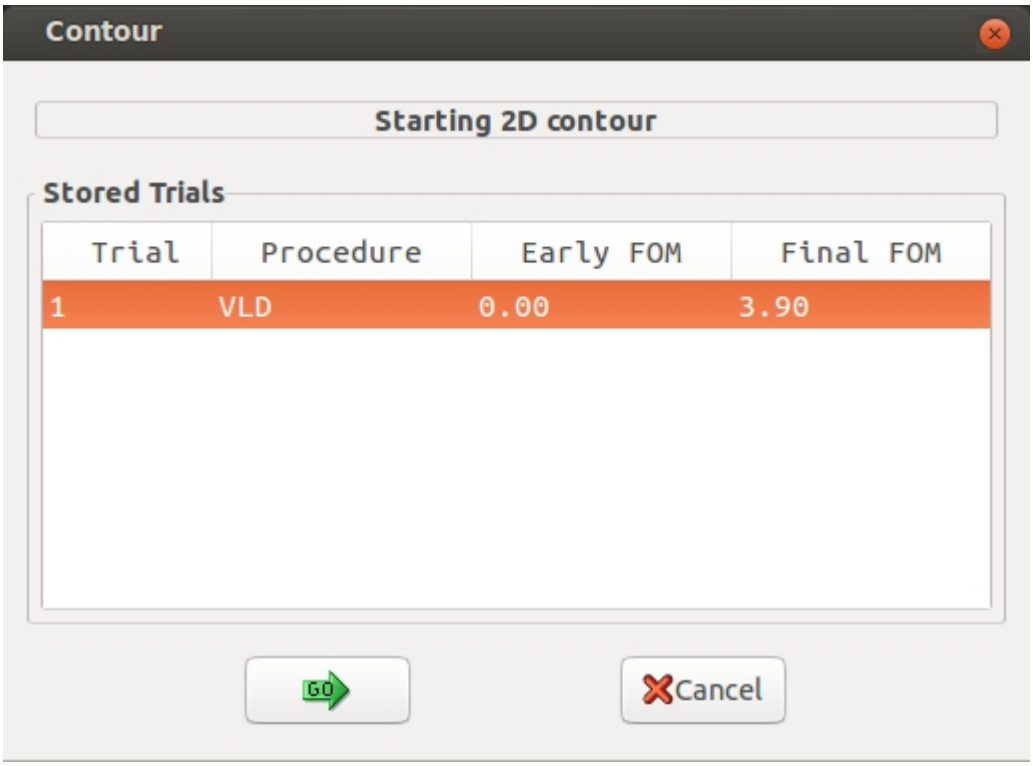
Start Jav on the map computed using the phases related to the selected trial. If a model is avaible, it will be included in the plot. For details see the help on-line of Jav.

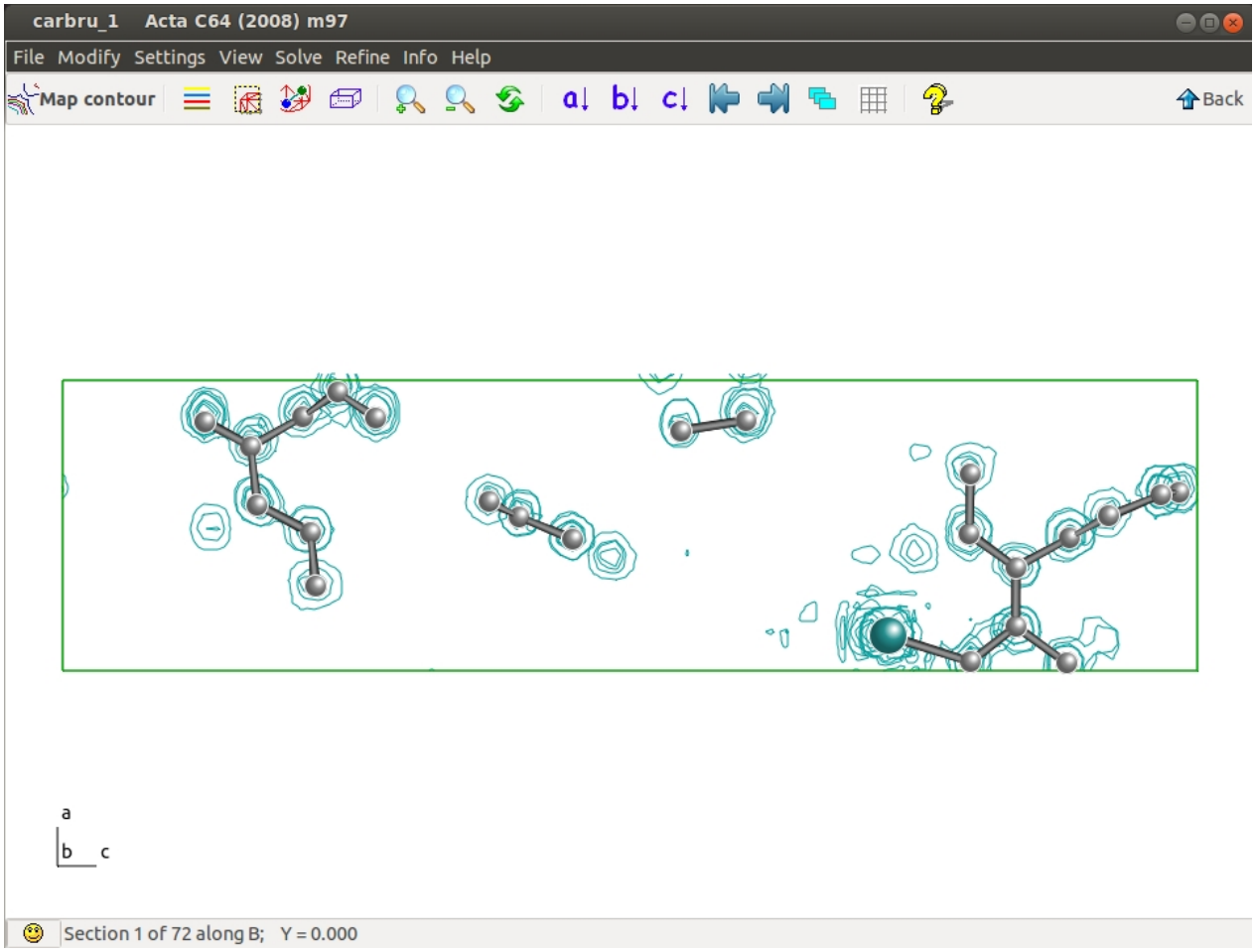




Contour

Start the 2D contour procedure on the map computed using the phases related to the selected trial.





Manage contouring levels.

Set limit & graphic resolution.

Show / Hide atomic model.

Show / Hide unit cell.

Magnify the plot.

Un-magnify the plot.

Let the plot fit the drawing window.

Project plot on the plane normal to a.

Project plot on the plane normal to b.

Project plot on the plane normal to c.

Plot previous section(s).

Plot next section(s).

Plot more than one section per time.

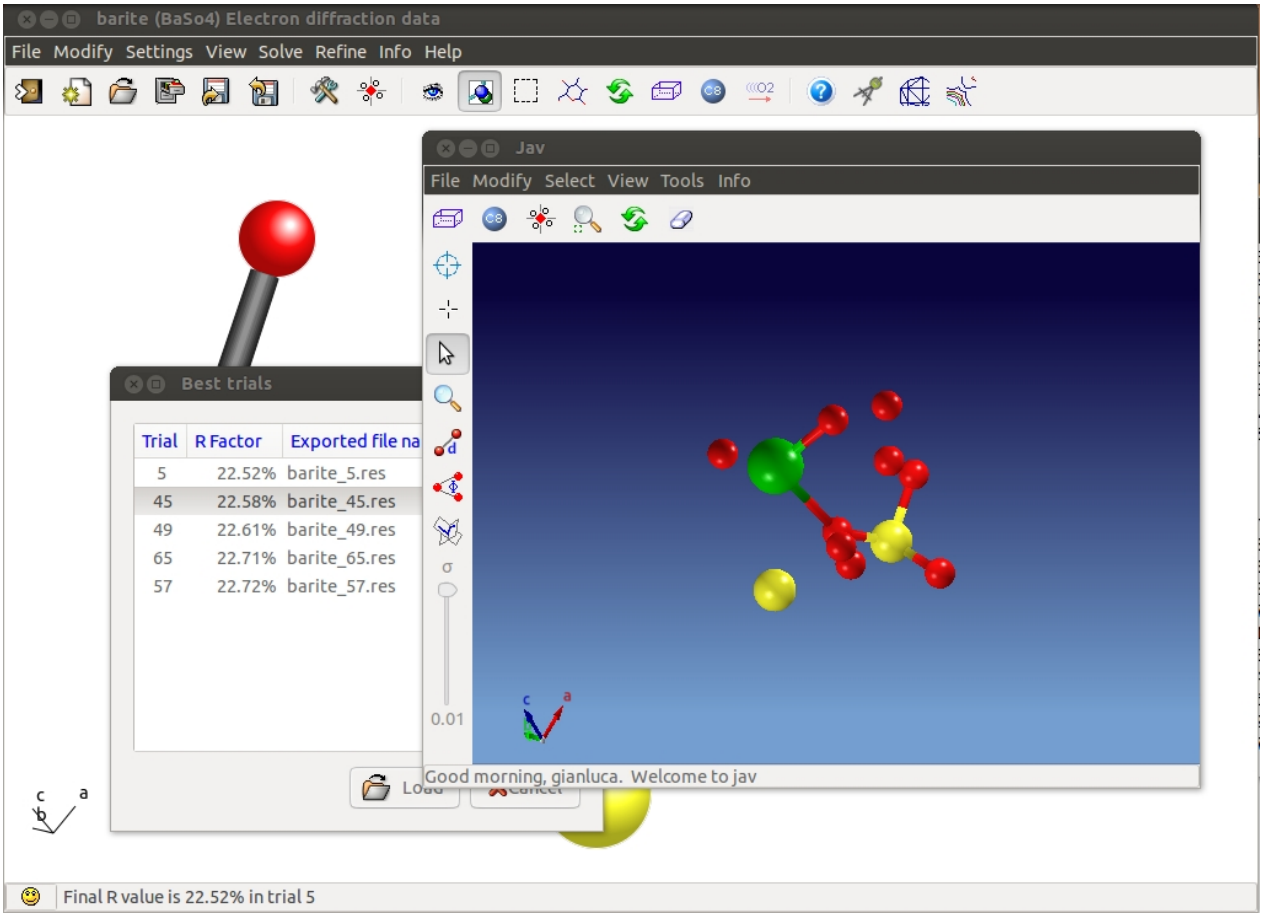
Show / Hide grid.

Credits.

Select among the best R% trials

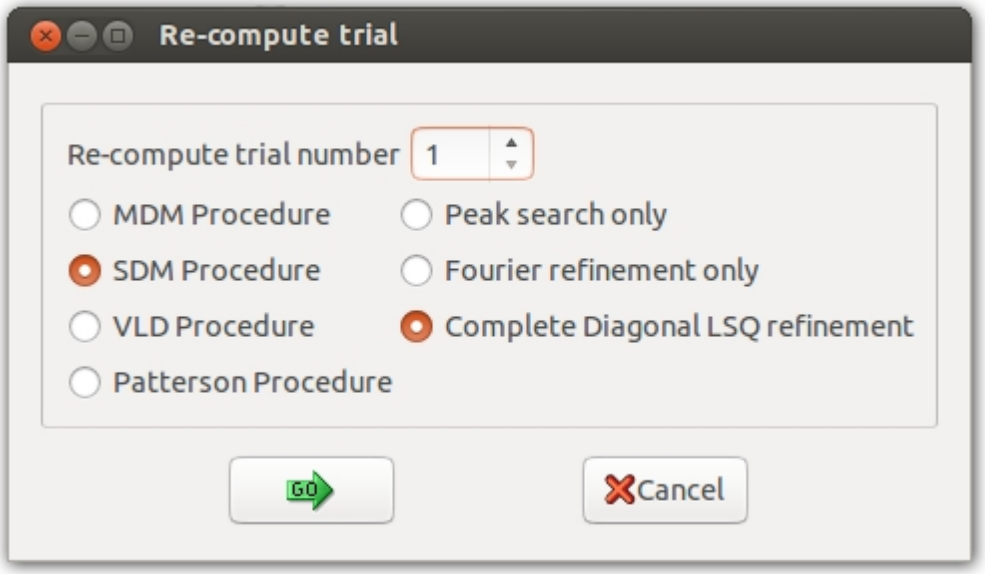
The best *n* trials (i.e. the trials with the lowest values of the crystallographic residual) have been stored in the project file. In default *n* is equal to 5 for electron diffraction data, equal to 1 otherwise.

By means of this menu item the user can view and restore the model corresponding to the selected trial.



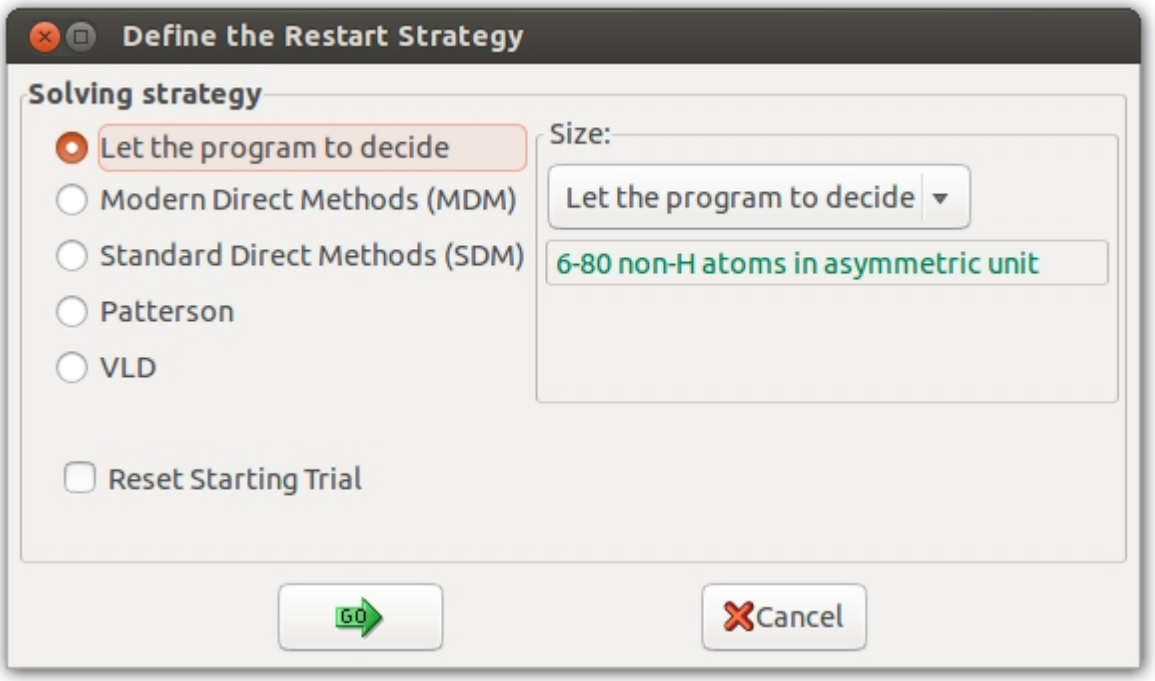
Re-compute trial

The following window allows to re-compute phases for a (previously explored) trial. It is possible to select the starting phases procedure (MDM Patterson or VLD) and to define the final steps of Direct Space Refinement (see Description of Sir2014, Phase Module).



Develop new trials

When the solution is not satisfactory, the user can explore new trials defining the **Solving Strategy**, threshold on R% and number of trials to explore.



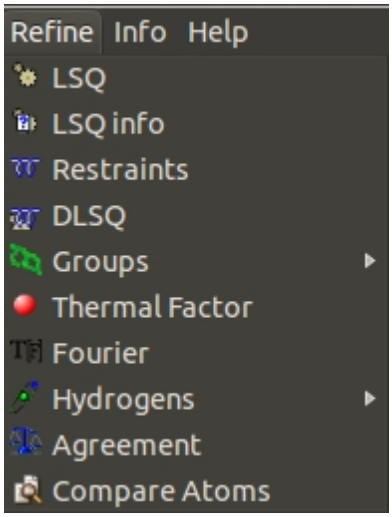
Simulated Annealing

The dialog of Simulated Anneling in Sir2014 is described here.

Created with the Personal Edition of HelpNDoc: Easily create EPub books

Refine

The sub-menu Refine is shown on the left. It allows:



LSQ	Least Squares Refinement.
LSQ info	Information about Least Squares calculations.
Restraints	Define and reset restraints.
DLSQ	Distance Least Squares Refinement.
Groups	Manage rigid body groups.
Thermal Factor	Select and delete atoms with a very high thermal factor.
Fourier	Fourier synthesis.
Hydrogens	Add or delete hydrogens.
Agreement	Produce statistics about agreement.
Compare Atoms	Compare coordinates of the current model with coordinates from an external file.

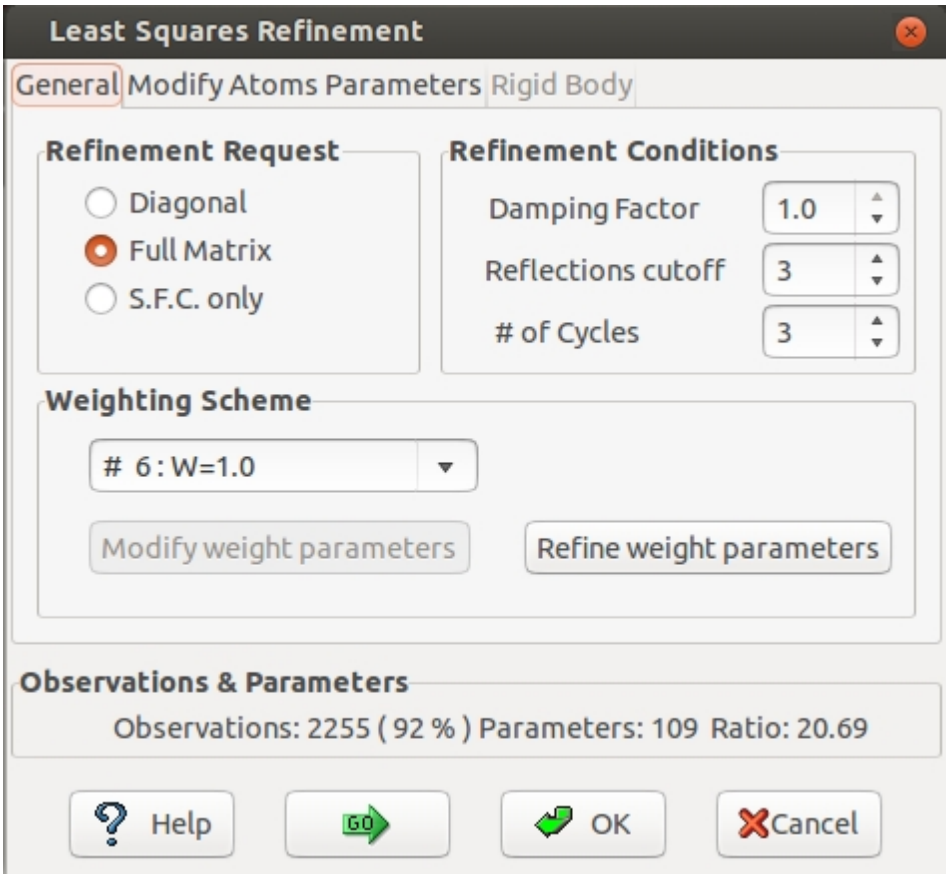
LSQ

Least Squares graphical interface is divided in 3 sections:

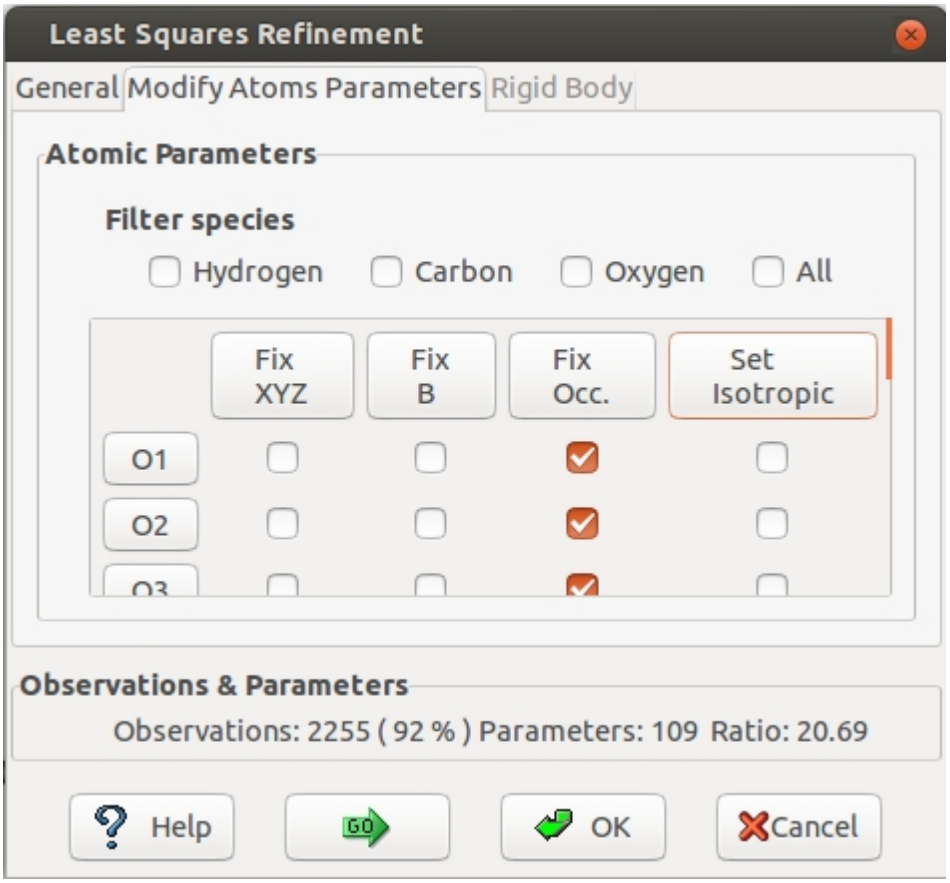
- **General**
- **Modify Atoms Parameters**
- **Rigid Body**

The program provides default values but the user can change them.

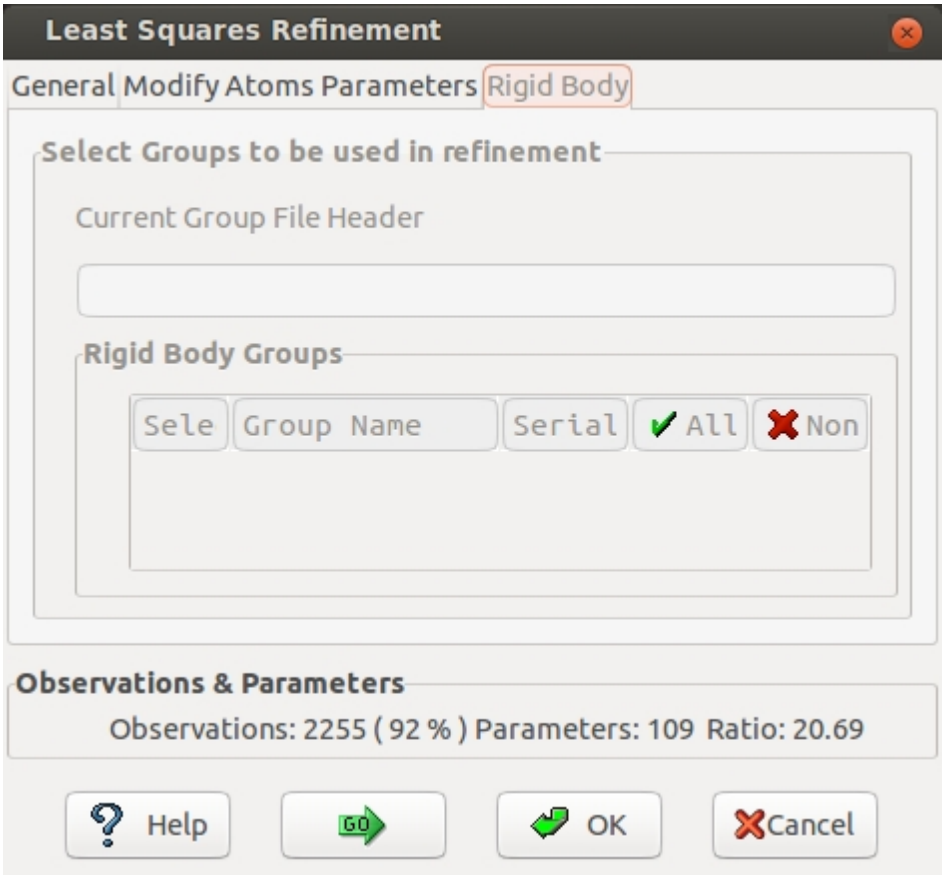
In the first section the user defines the Least Squares strategy (Diagonal Matrix, Full Matrix or Structure Factor Calculation only) and the Refinement Conditions such as Damping Factor, Reflections cutoff, number of cycles. 18 Weighting Schemes are available and their parameters can be modified. At the bottom of this window information about the ratio Observations/Parameters is updated depending on user choices. The **GO** button will start the calculations.



In this section the user can fix or relax the atom parameters (coordinates, thermal factor) and set the thermal factor isotropic or anisotropic.

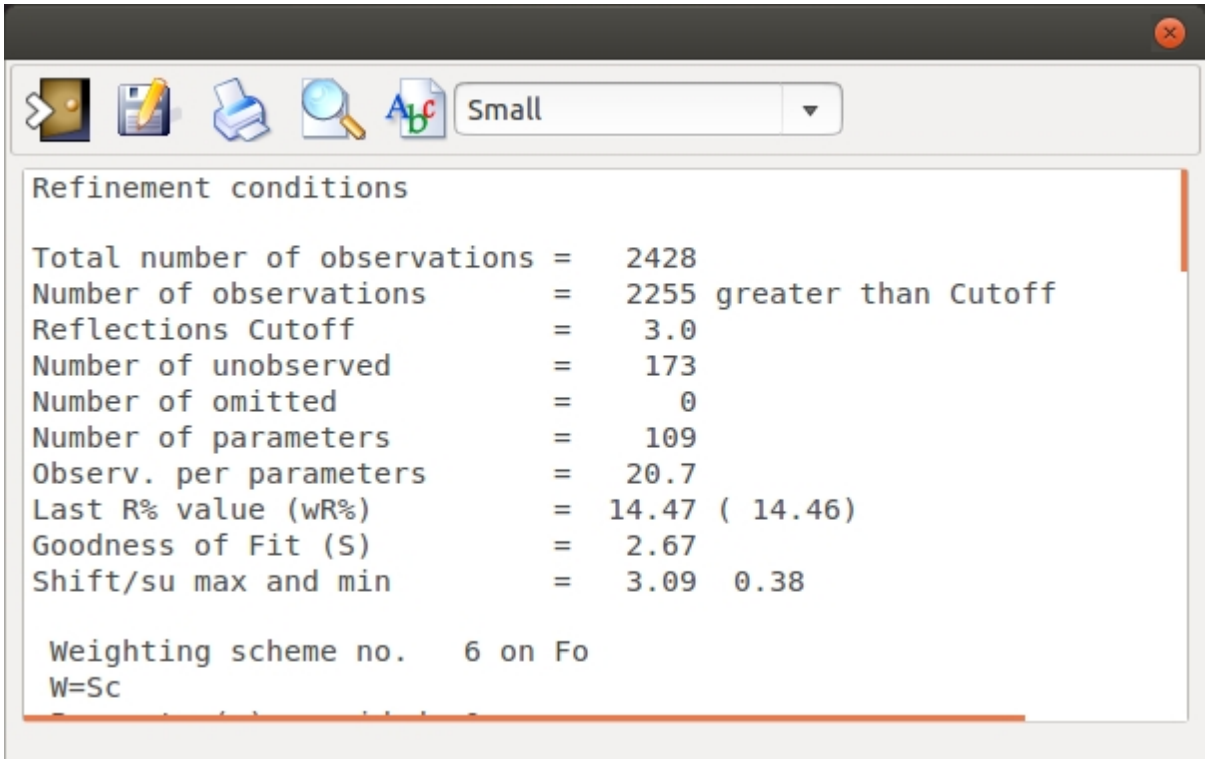


In the **Rigid Body** section the user can refine (or not) as Rigid Body the groups listed.



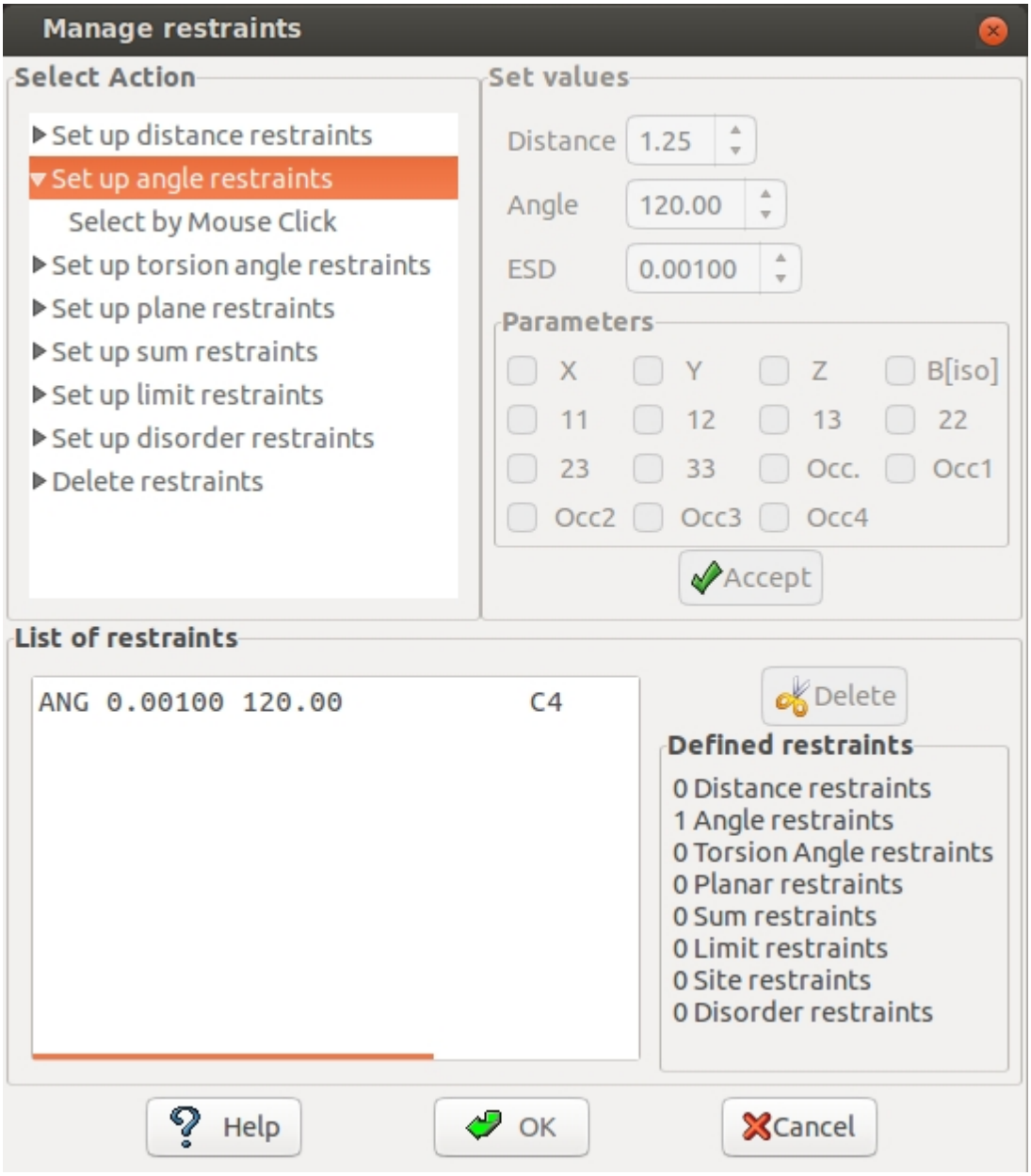
LSQ info

Information about Least Squares calculations such as atom coordinates, isotropic/ anisotropic thermal factor, conditions on atom parameters (fixed, equivalence,..), last value of the residual are displayed.



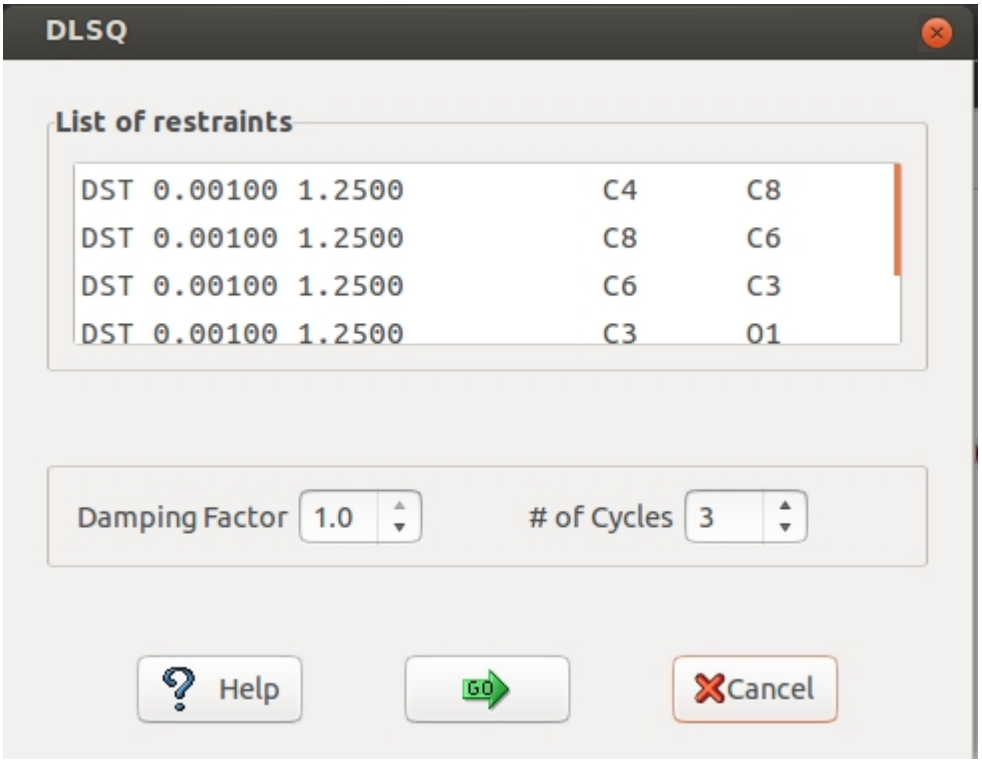
Restraints

An efficient way to improve the ratio Observations/Parameters during Least Squares calculations is the use of Restraints. They are managed through the following interface. Once some restraint has been defined, it is shown on the graphic window. As an example a restrained angle has been defined below.

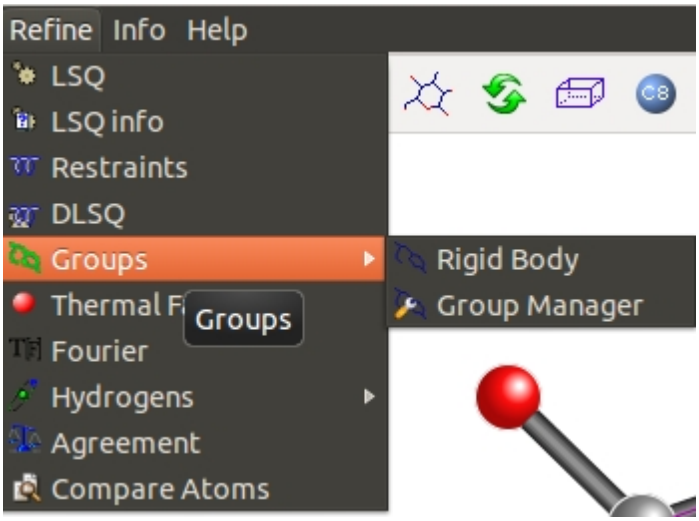


DLSQ

Least Squares Calculation based on restrained distances

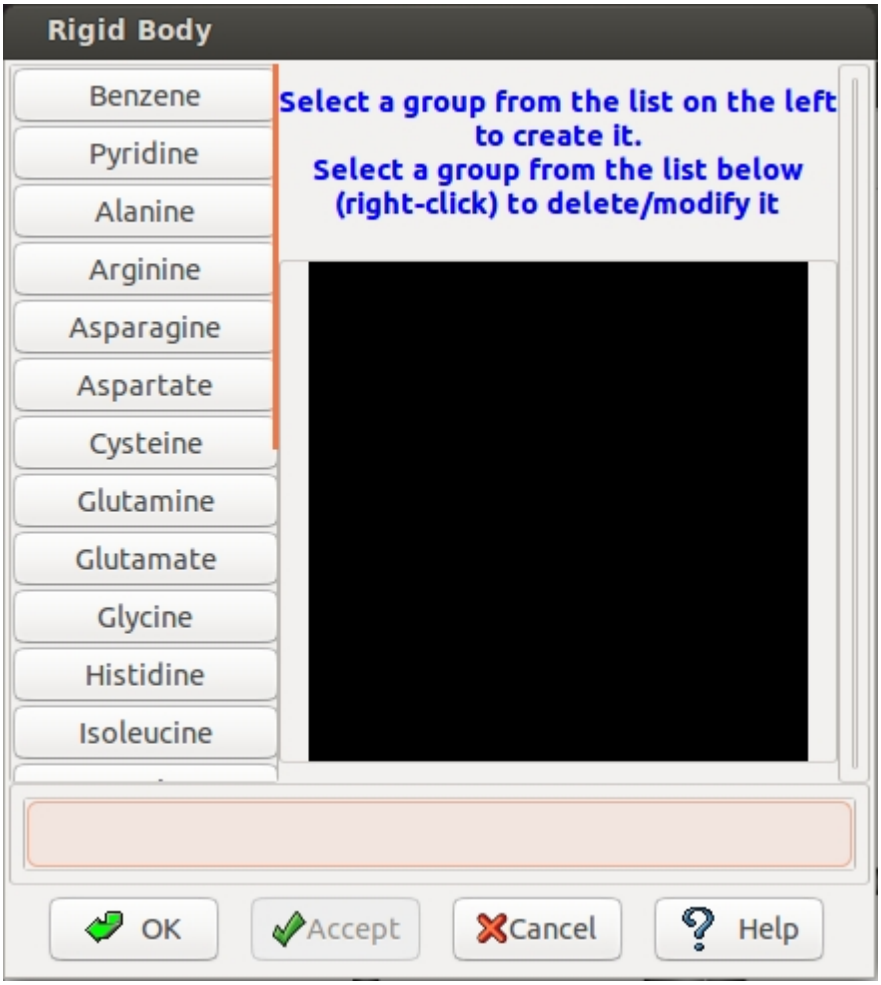


Groups



These menu items are related to the definition of groups of atoms to be refined as Rigid Body.

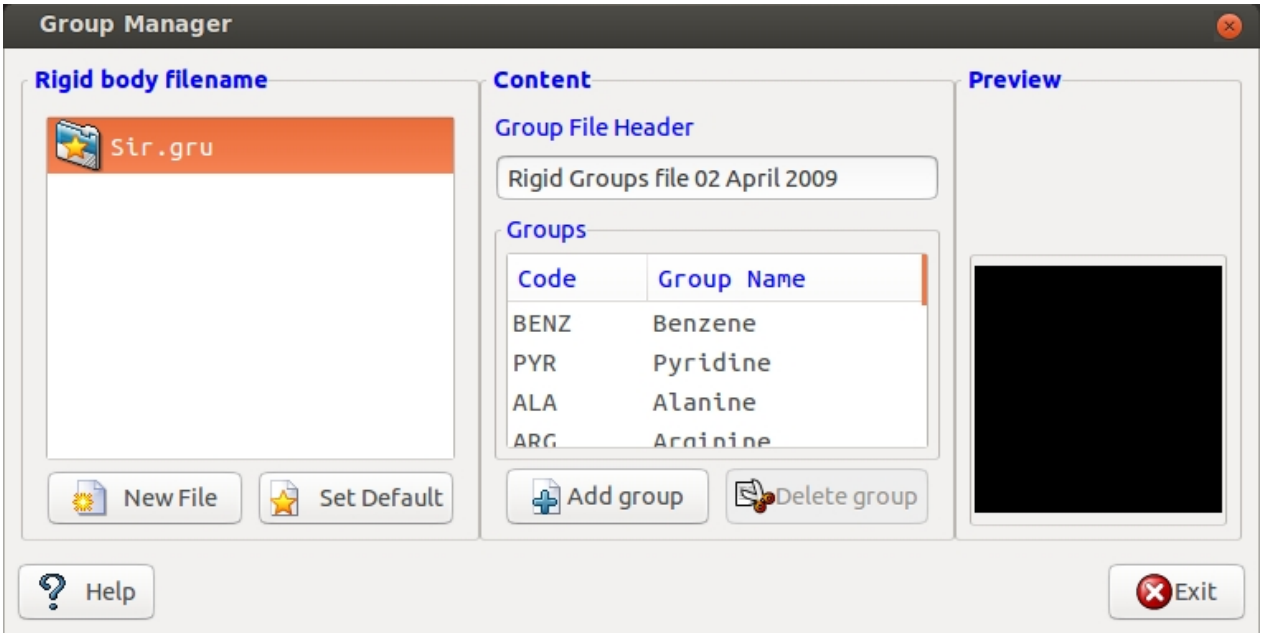
The **Rigid Body** interface allows, once a group of the list on the left has been selected, to associate the atoms of the group with the atoms of the current model.



Group Manager

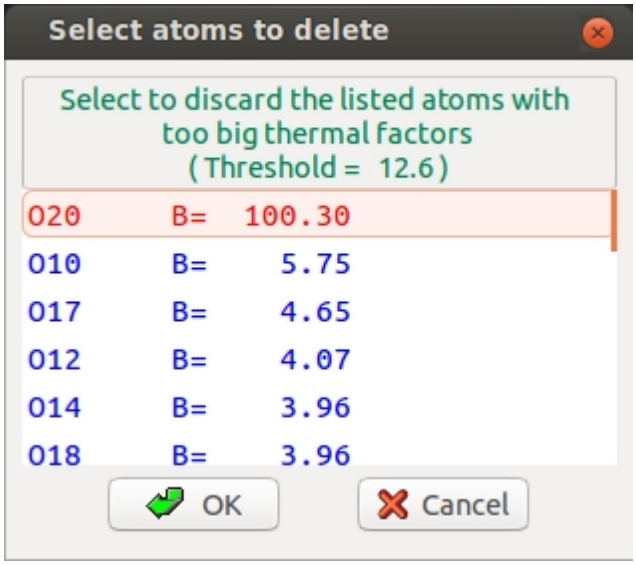
At installation time a file, called Sir.gru, is created in the Home Directory of the user. It contains the groups of benzene and residues.

The user can add some customized group of atoms to be refined as Rigid Body.



Thermal Factor

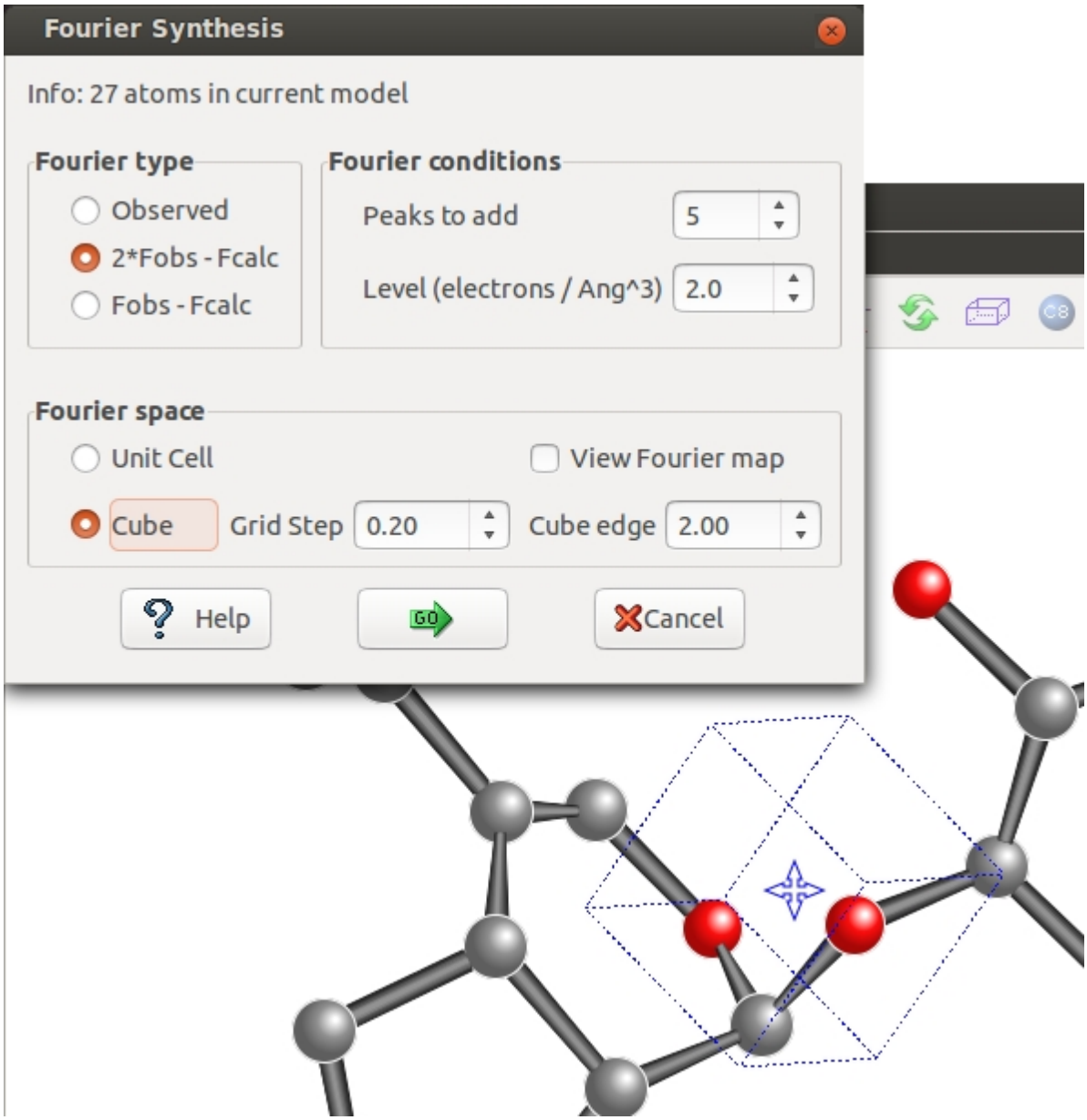
During Least Squares calculations the value of the thermal factor for some atoms can assume too large values. This window can be used to analyse the values of the thermal factor and, eventually, delete the atoms whose value is too high.



Fourier

In order to locate new atoms, a Fourier interface is available.

Once the **Fourier type** has been selected and the number of new peaks has been decided, the calculations are started by clicking on button.



Buttons

Help: this output.

Go: execute the Fourier synthesis.

Cancel: exit from Fourier interface.

Fourier type

Observed: The coefficients of the Fourier synthesis are the observed structure factors (Fobs).

2*Fobs – Fcalc: The coefficients are two times the Fobs minus the calculated structure factors (Fcalc).

Fobs – Fcalc: The coefficients are the difference between the observed minus the calculated structure factors.

Fourier Conditions

Peaks to add: The user can select the maximum number of peaks to be located.

Level (electrons/Å^3): The peaks to be located must have the height greater than this value of the electron density.

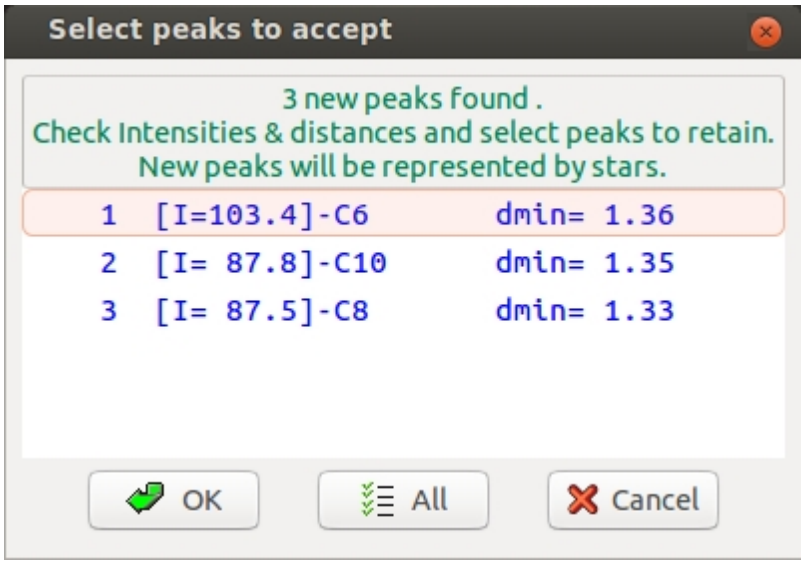
Fourier Space

Unit Cell: The asymmetric part defined by the Space Group is the space where the Fourier synthesis has to be calculated.

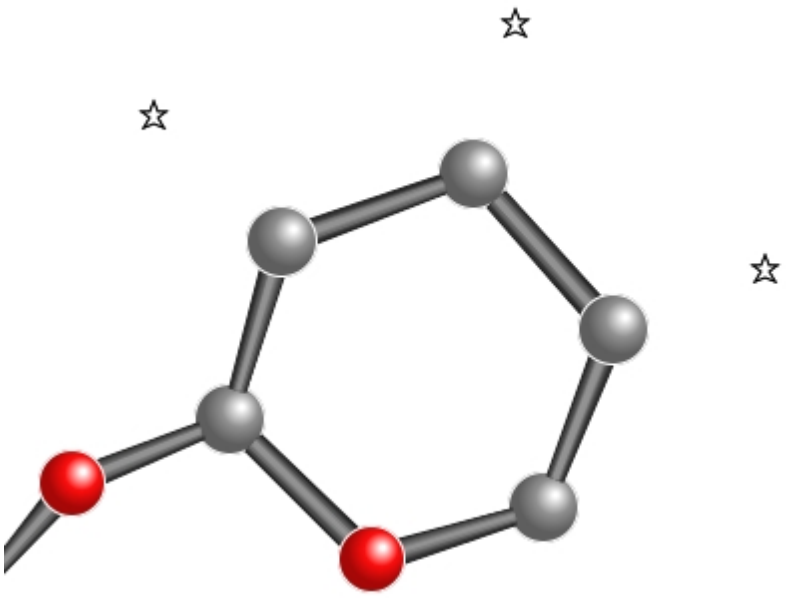
Cube: The user can define a cube, having the faces parallel to the cell faces, where the Fourier synthesis has to be calculated. The dimension in Angstrom of the cube edge (default 2.0 Å) and the grid stepof the Fourier synthesis (default 0.20 Å) can be defined by the user. Using the mouse and moving the model, the user can put the cube where he wants.

If '**View Fourier map**' is checked, the Fourier map and the model are shown by the program Jav.

The peak search routine provides a number of new peaks. Information about peak intensity and distance from the closest atom is provided by means of the window below, so the user can select peaks to retain:

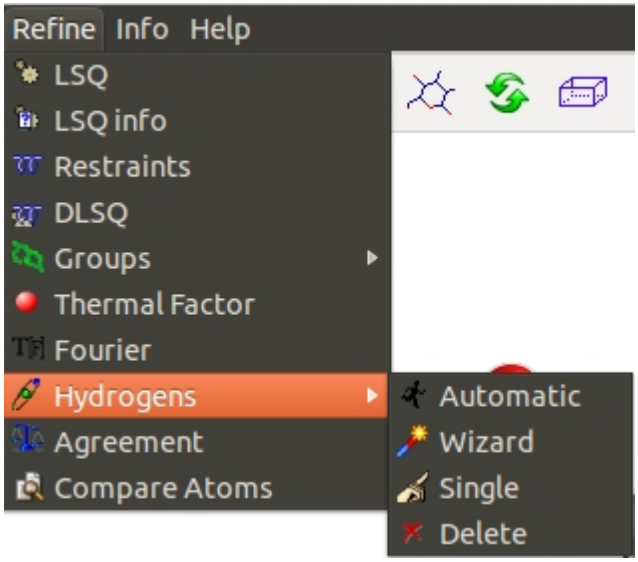


New peaks are represented by stars:



Hydrogens

This menu item provides different approaches to add hydrogens to the model:



Hydrogens automatic generation.

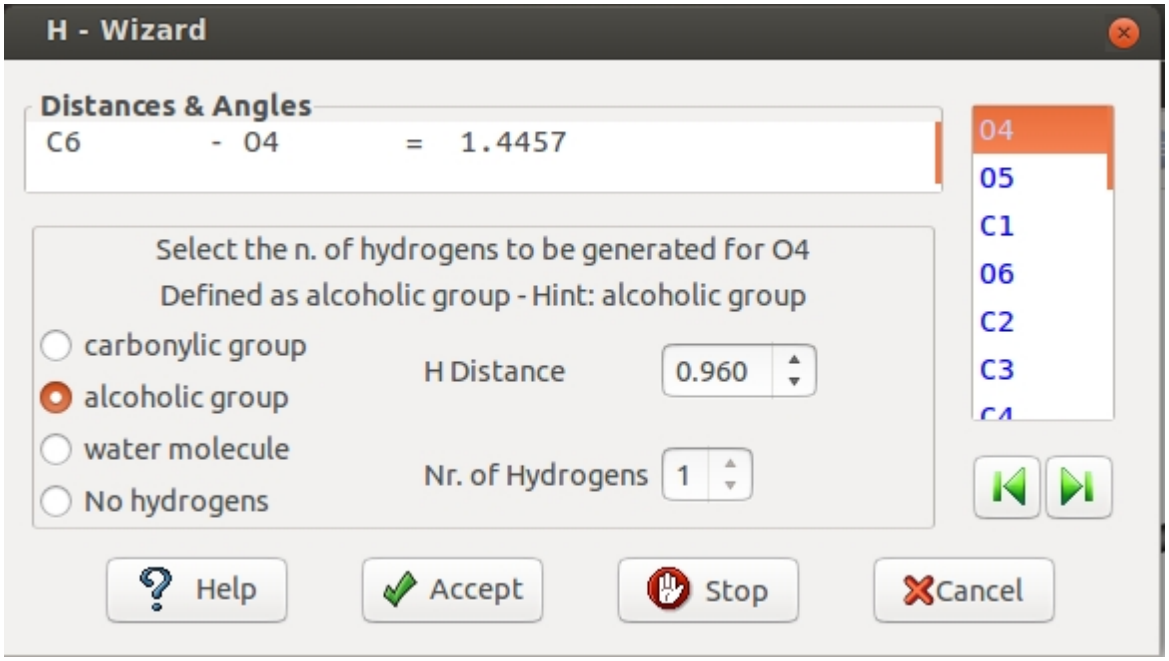
Hydrogens aided generation.

Add hydrogens to a selected atom.

Delete all (generated) hydrogens.

If the automatic choice is selected, no user intervention is required. All the possible hydrogens will be generated on the basis of geometrical criteria.

Using the ***Wizard***, the program checks distances and angles around atoms, supplies this information and suggests the possible hybridization; the user can accept the hint or modify the hybridization, can skip the current atom or abort the Wizard procedure.



Buttons

Help: this output

Accept: the current definitions are accepted and the next atom is selected

Stop: stops the modifications and generates hydrogens

Cancel: exit from the Wizard interface

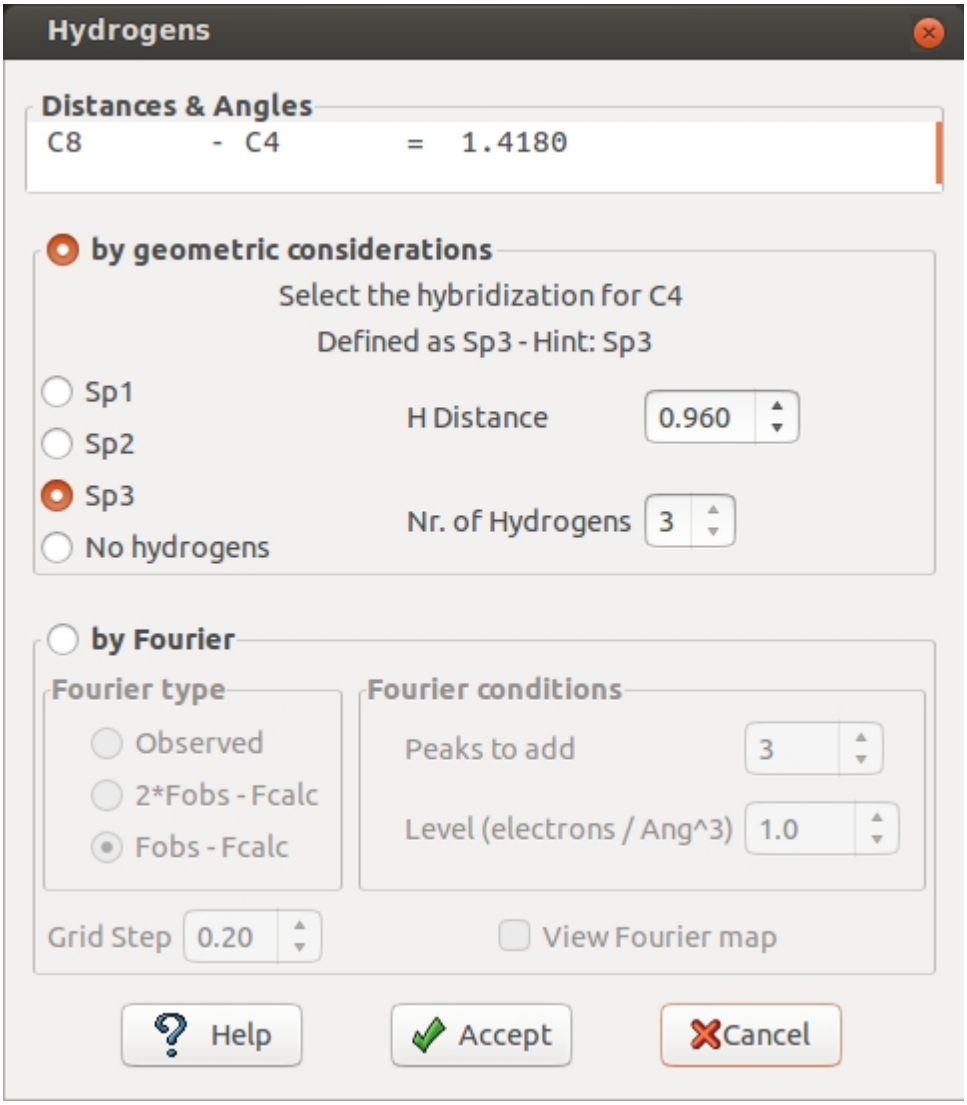
On the right of the window it is supplied the list of candidate atoms for the H atoms generation. Only O, C and N atoms are considered.
The atom can be selected by clicking on a list item or using the buttons

On the left (**Distances & Angles**) , the geometry of the bonds corresponding to the current atom is visualized. The program determines and shows the hybriditation and the number of H atoms to be generated. It is possible to change the hybriditation, different from the suggested choice.
The number of H atoms to be generated can be changed only for N atoms.
It is possible to modify the default bond distances "current atom - H" (default: 0.960 Å).

By clicking on "Accept" the next atom will be considered.

The process ends when all atoms have been considered or when the "Stop" button is clicked.

Single: using this feature it is possible to generate hydrogen(s) bonded to a single atom. It is possible to use geometric considerations (in this case the program will suggest the possible hybridization) or it is possible to search hydrogens in a small volume of the map, around the selected atom.

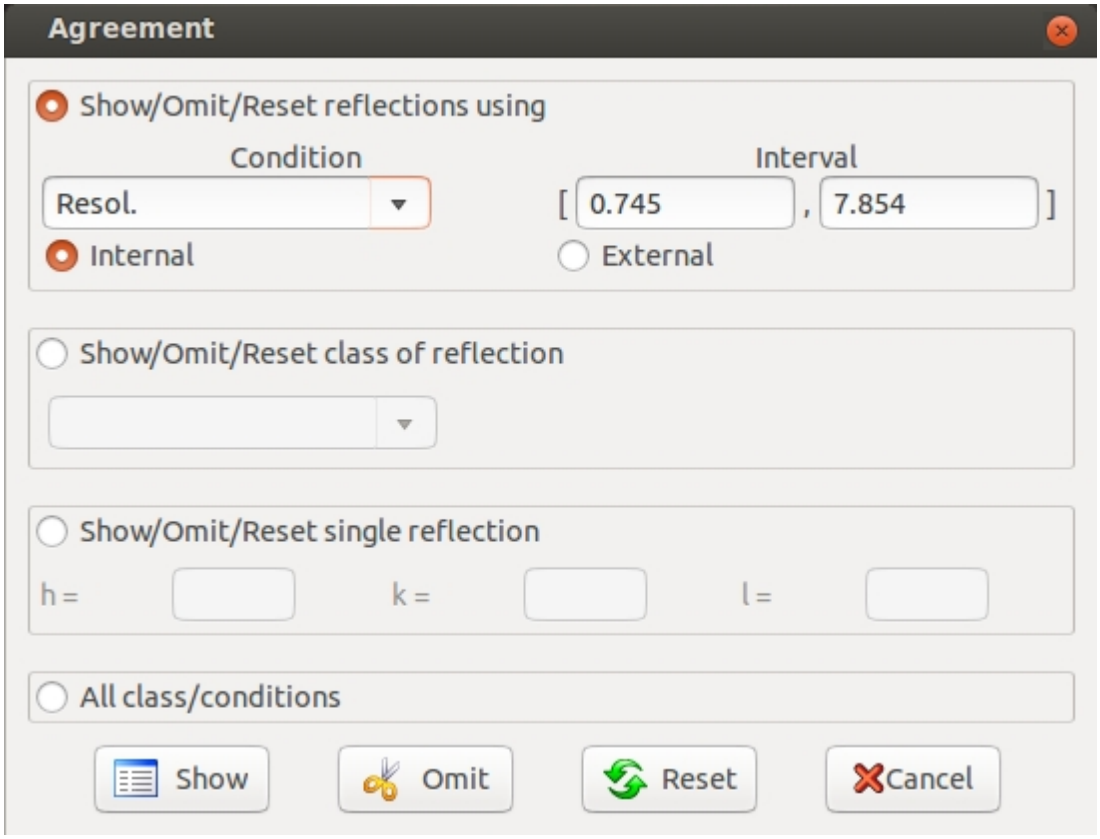


If '**View Fourier map**' is checked, the Fourier map and the model are shown by the program Jav.

Hydrogens can also be generated using Jav.

Agreement

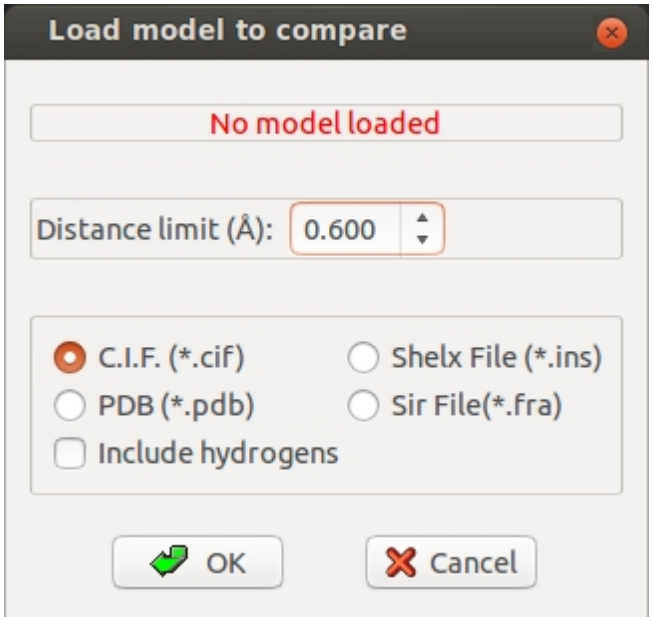
Statistic on agreement are important to improve the refinement. It is possible to show or omit reflections with respect to a selected condition or class.



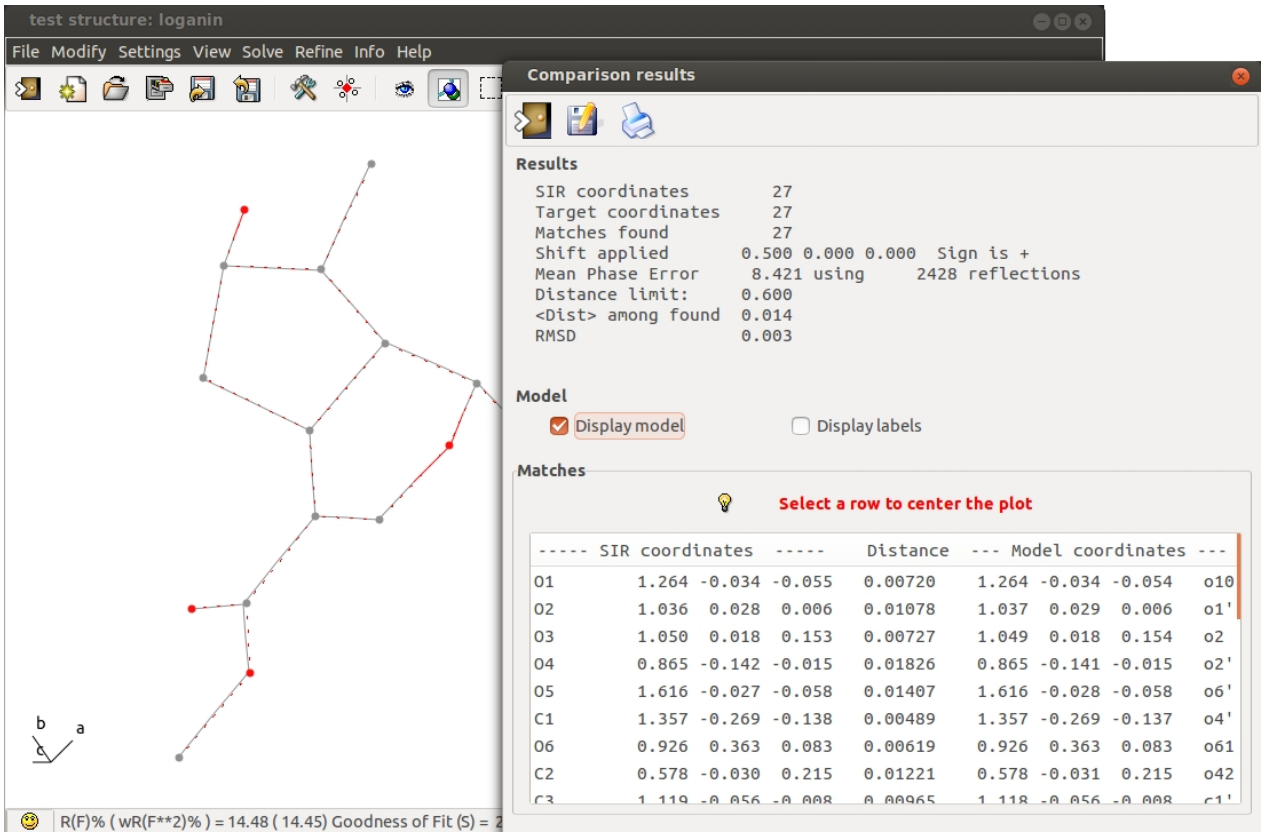
Compare Atoms

If a structural model is available (i.e. solving a structure using electron diffraction data an X-Ray model could be available), it is possible to compare the coordinates provided by the program with the model.

Using the following window it is possible to load the model to compare:



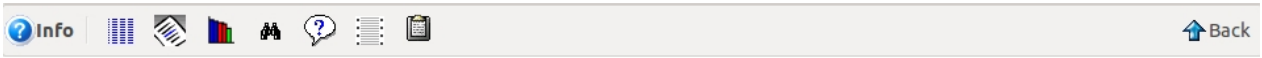
The result of the comparison is shown below; the model is superimposed and drawn using dotted lines.



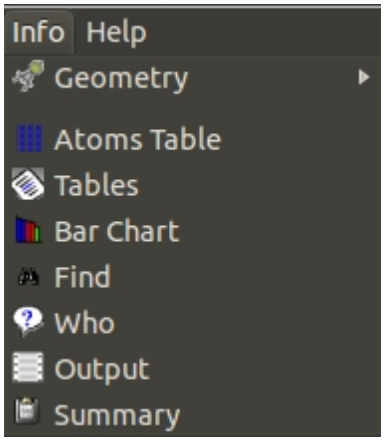
Created with the Personal Edition of HelpNDoc: Free iPhone documentation generator

Info

Possibilities offered by the sub-menu Info shown on the left; they can be accessed directly using icons in the Info toolbar



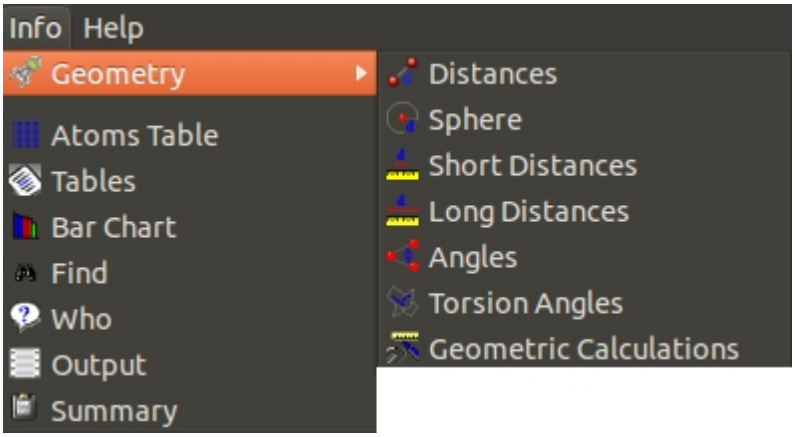
or in the Geometry toolbar.

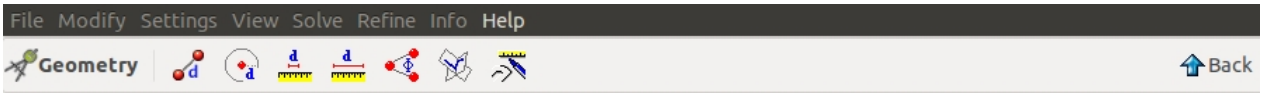


Geometry	Distances and Angles calculation.
Atoms Table	Show a numeric table of atoms.
Tables	Produce some useful tables for publications.
Bar Chart	Show the bar chart of intensities (peak height of the Fourier map).
Find	Mark with a cross the required atom.
Who	Info on selected atoms.
Output	Provide a short summary of the results.
Summary	Main results are summarized.

Geometry

By means of the ***Geometry*** menu or toolbar it is possible to compute distances (between two atoms or all distances around a single atom), angles and torsion etc.





Compute distances between 2 atoms or around a single atom.

Compute distances between a selected atom and all the other atoms in a sphere.

Create a table of all short distances.

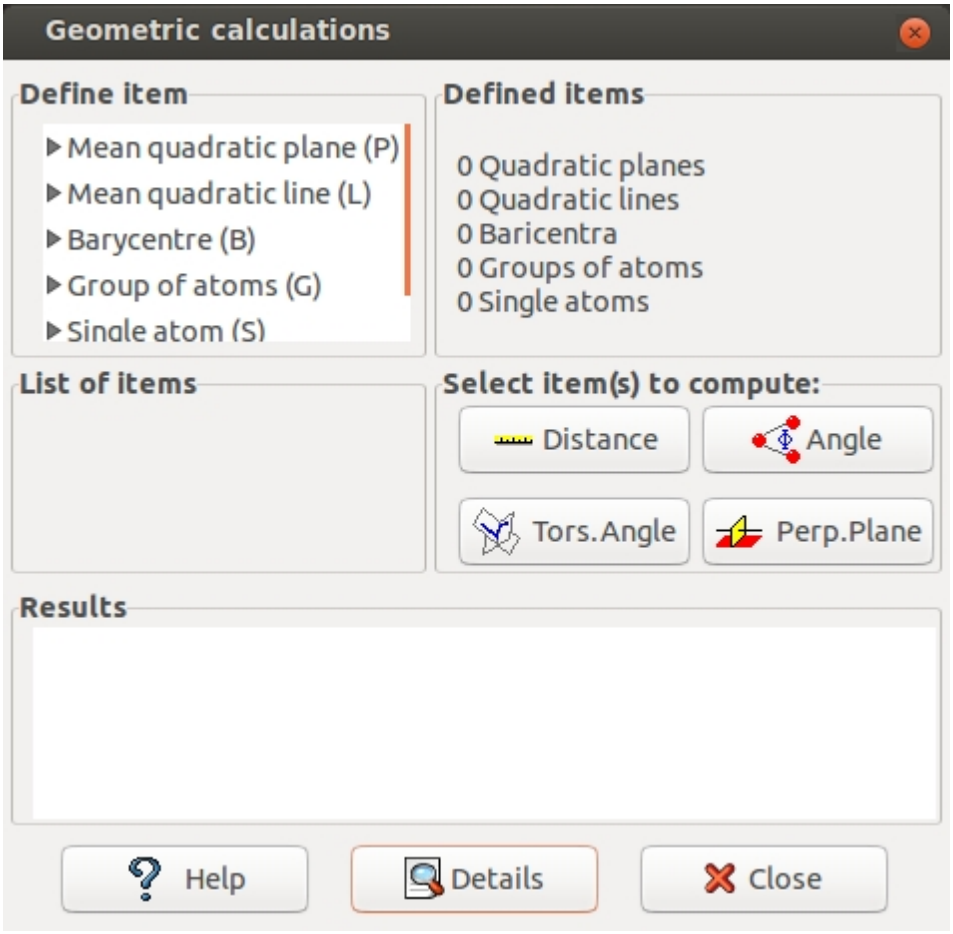
Create a table of all long distances.

Compute angle among 3 atoms.

Compute torsion angle among 4 atoms.

Compute several geometric quantities.

The Geometric calculations window follow:



Buttons

Help : this output.

Details : to show the full results of the calculations.

Close : to exit from Geometry Calculations interface.

Define item

Results of geometric calculations or definitions of one or many atoms are defined as items. The user selects the atoms to be included in the calculation by the list of atoms, by mouse click on each atom or by successive mouse clicking outside the atoms forming a polygon which surrounds the requested atoms. In particularly, the allowed items are the following:

1. Mean quadratic plane (P). The equation of the mean quadratic plane through more than three atoms is calculated (if the number of atoms is three, the unique plane is calculated).

2. Mean quadratic line (L). The line is described parametrically [xg(i) and u(i)] as the

set of points with coordinates x(i) for some values τ: x(i) = xg(i) + u(i) * τ

3. Barycentre (B). The program calculates the coordinates of the mean location of all the selected atoms.
4. Group of atoms (G). The user defines some atoms.
5. Single atom (S). The user defines one atom.

There is also the possibility to delete items.

Defined items

List of items already defined.

List of items

The program shows the items. The user can select one or two of them to be used in subsequent calculations.

Select item(s) to compute

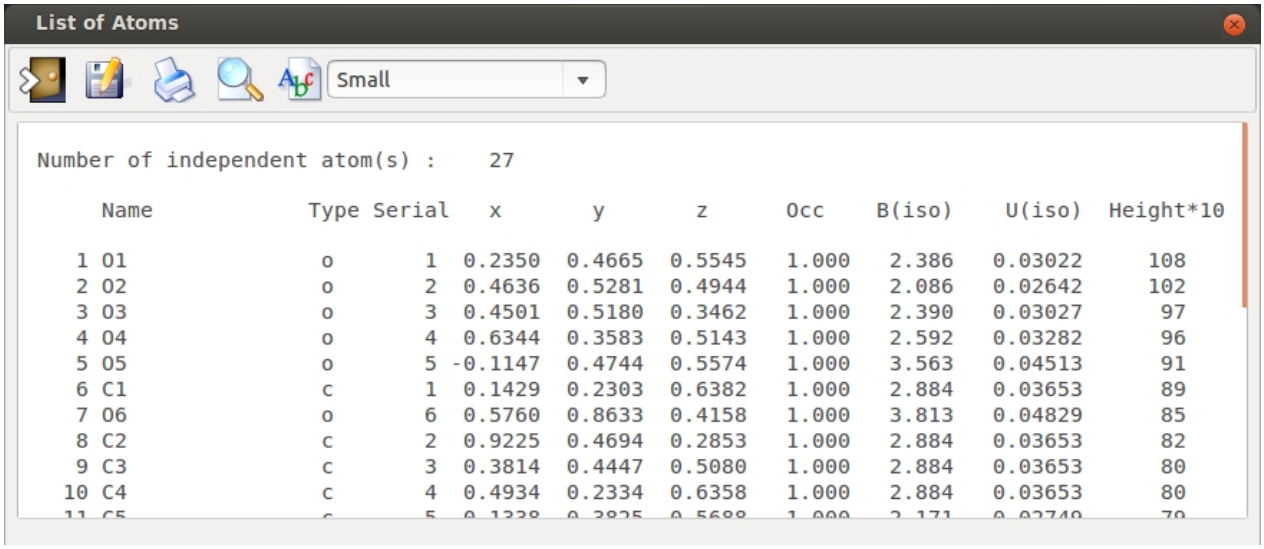
After the selection of one or two items, the user can execute the following calculations:

1. **Distances** between the atoms of an item or between the first item and the second item.
2. **Angles** using the atoms of an item (G) or of two items (G, S or B) with the central position occupied by the atoms of the first item and the other atoms of the second item.
3. **Angle** between plane (P) or line (L).
4. **Torsion Angle**: the torsion angles along the sequence of atoms of the item (G) picked up four by four.
5. **Perpendicular Plane**: The program calculates the equation of the plane normal to a line (L) and passing through one atom (S).

Results

The concise results of the calculations are shown.

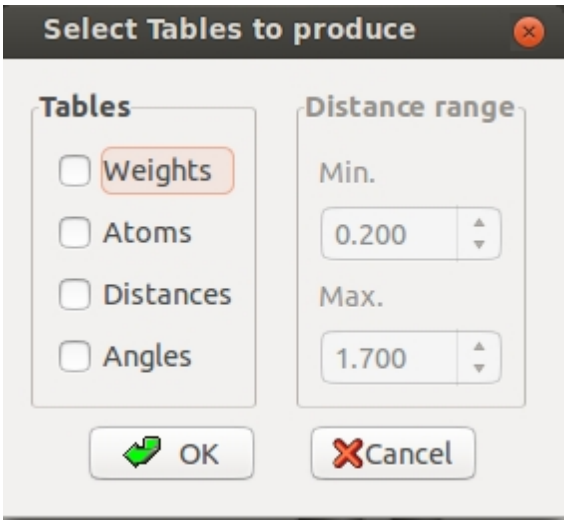
Atoms Table



The atom list of the model is shown. The user can save this information in a file or print it:

Tables

It is possible to prepare tables suitable for publications



Bar Chart

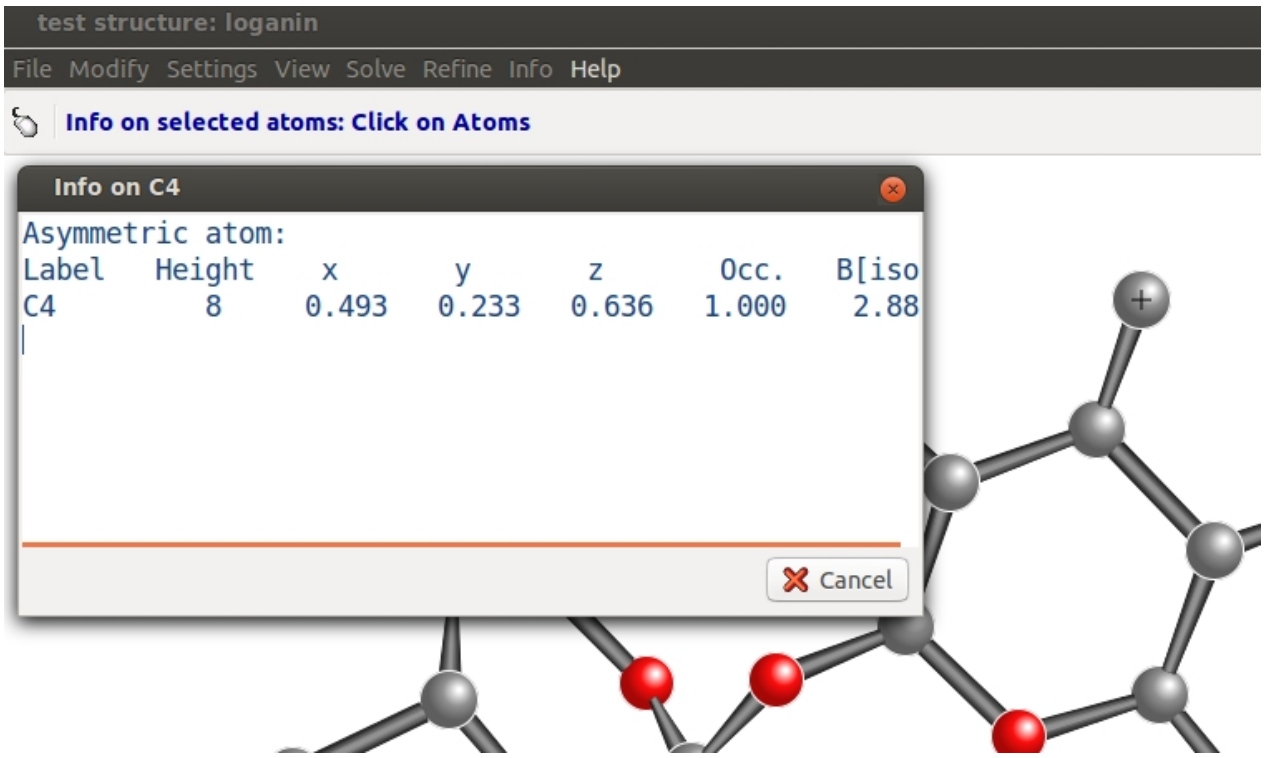
The bar chart of intensities (peak height of the Fourier map) is shown; different colors correspond to different atomic species. This diagram is useful to check the presence of false atoms or atoms not declared in the unit cell content.

Find

The program will mark with a cross the atom entered using the ***Find*** toolbar.

Who

The program will provide information about the atoms selected by mouse click. Such information is displayed in the following window:



Output

See ***File*** section

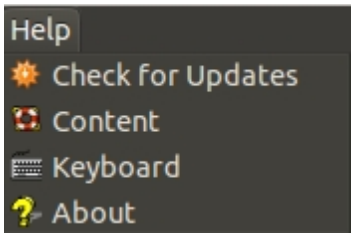
Summary

See ***File*** section

Created with the Personal Edition of HelpNDoc: Easy EBook and documentation generator

Help

The sub-menu Help is shown on the left. It allows:



Check for Updates	Check if an updated version of the program is available.
Content	Starts the Internet Browser on these pages.
Keyboard	Info on shortcuts and mouse usage
About	Shows the Sir2008 About window.

Check for Updates

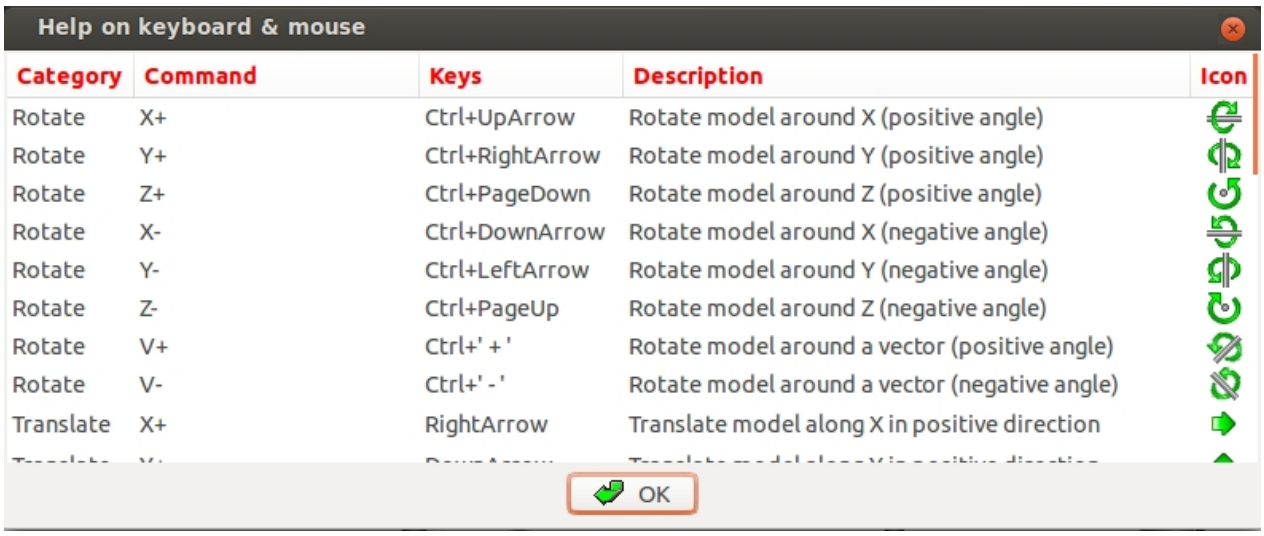
Check if an updated version of the program is available on the IC server (www.ic.cnr.it).

Content

Starts the Internet Browser on these pages

Keyboard

Short-cuts and mouse usage:



About

Short information such as authors, contact addresses and libraries versions are supplied.

Created with the Personal Edition of HelpNDoc: Free Web Help generator

Commands and their use

The input consists of a sequence of comments, commands and directives.
The commands are headed by '%' character and directives must follow the related command.

Sir2014 recognizes the following commands:

%DATA	Data input routine.
%INVARIANTS	Invariants routine.
%PHASE	Phasing routine + Direct Space Refinement routine.
%IMPORT *file*	Used to import a structural model (for Simulated Annealing).
%SANNEL	Simulated Annealing routine.
%REMO	Molecular Replacement routine.
%END	End of the input file
%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.
%AMB *string*	This command is used to specify the name of the Automated Model Building program (Buccaneer, Nautilus, ARP/wARP, Phenix) to be used (proteins only)
%CUDA	If a powerful graphic card is available, it is possible to use some CUDA code to speed up the calculations. CUDA specific drivers have to be installed..

Warning: a patch for nautilus_pipeline and buccaneer_pipeline is needed in CCP4 6.4.

The program creates the name of some needed files, adding the appropriate extension to the structure name. The file names are:

string.bin → direct access file

string.res → file in SHELX format

string.plt → file for graphics

If **%STRUCTURE** command is not used the default string "struct" (instead of the name of the structure) is used to create file names.

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 ">" or "!" then the record is interpreted as a comment; characters following ">" or "!" will be ignored.

Sir2014 preserves intermediate results. The first routine used must be DATA, if it has not been used in a previous run.

The minimal information needed by Sir2014 is constituted by:

- cell parameters;
- cell content;
- reflections.

Created with the Personal Edition of HelpNDoc: Produce electronic books easily

Directives

List of the directives associated to the commands

%Data

%Invariants

%Phase

%Sannel

%Remo

Data routine

Only directives marked in **red** are mandatory.

CELL *a b c* α β γ

Cell dimensions: *a, b* and *c* are in Å, α, β and γ in degrees.

ERRORS esd(*a*) esd(*b*) esd(*c*) esd(α) esd(β) esd(γ)

Estimated standard deviations for the unit cell dimensions.

SPACEGROUP *string*

String is the symbol or the number of the space group, according to International Tables (1974). Blanks are necessary among the terms constituting the space group symbol (see examples). If no space group is supplied, the program performs the automatic identification of the Laue group and of the extinction symbol from diffraction intensities. If the *string* SCAN is supplied, the program consecutively explores (performing a complete solution run) all the allowed space groups, related to the most probable extinction group.

FORMULA $(El_1n_1El_2n_2El_3n_3 \ldots)Z$

Unit cell content. El_i is the chemical symbol of atomic type i, n_i is the corresponding number of atoms in the chemical formula, Z is the number of molecules in the unit cell. To avid ambiguities, the two-characters elements must have the second character in lower case i.e. formula (BaSO4)4.

OR

CONTENT *El*$_1$ *n*$_1$ *El*$_2$ *n*$_2$ *El*$_3$ *n*$_3$ ....

Unit cell content. El$_i$ is the chemical symbol of atomic type *i*, n$_i$ is the corresponding number of atoms in the unit cell. The items in this directive must be separated by blanks. i.e. CONTENT C 68 H 104 O 40

Up to a maximum of 18 atomic species are allowed. For each chemical element up to Californium (Cf, Z=98) X-Ray, electrons and neutrons scattering factor constants for are stored, together with information on the atomic number and weight, covalent and Van-der-Waals radii, etc. in the file Sir.xen provided with the program.

SFACTORS *El a1 b1 a2 b2 a3 b3 a4 b4 c*

Scattering factors for species *El*. If more lines are necessary, use character = at the end of the line.

REFLECTIONS *string*

String is the name of the reflections file. Each records contains: Miller indices h,k,l, $|F_o|$, $\sigma(F_o)$ (or $|F_o|^2$, $\sigma(F_o^2)$ where h,k,l are integer (up to 512). The end of reflections is detected using one of the following:

- blank record.

- end of file.

Negative values of $|F_o|$ are allowed; negative values of $\sigma(F_o)$ are forbidden.

FORMAT *string*

String is the run time format to read reflections. Default value for *string* is (3I4,2F8.2). If *String* is equal to **free**, blank delimited values are expected in the reflections file.

NOSIGMA

To be used when $\sigma(F_o)$ values are meaningless or not available.

FOBSERVED

Program assumes h,k,l, $|F_o|$, $\sigma(F_o)$. $|F_o|^2$ and $\sigma(F_o^2)$ are expected as default.

FOSQUARED

Program assumes h,k,l, $|F_o|^2$and $\sigma(F_o^2)$ (default choice).

SIZE *String*

To set a suitable phasing strategy. The possible values for *String* are: s, m, l. Default procedure is automatically chosen by the program, on the basis of the structural complexity.

RHOMAX *x*

Maximum value of $(\sin\theta/\lambda)^2$ accepted for reflections to be used. In default all the data are accepted.

RESMAX *x*

Maximum value of resolution (in Å) accepted for reflections to be used. In default all the data are accepted.

MINRES *x*

Minimum value of resolution (in Å) accepted for reflections to be used. In default all the data are accepted.

GENERATE *x*

The program generates non-measured reflections beyond and behind the experimental resolution limit up to the specified maximum resolution value (in Å). Unobserved values are actively used in the DSR procedures. If *x* value is not supplied, the default resolution limit is settled by the program. Generate is a default option for macromolecules, if data at non atomic resolution are supplied.

NOGENERATE

The program does not generate (and use in active way) non-measured reflections beyond and behind the experimental resolution limit (as request by the default *DSR* procedures).

WAVE *string* or *x*

Nevertheless the wavelength is not used during the structure solution stage, it is necessary for the LSQ-refinement and its value is also written in CIF file produced by the program. Possible values for *string* are: Cu, Mo or a numeric value. Default wavelength is Mo. It is possible to supply a numeric value for the wavelength used (in Å).

NREFLECTION *n*

Number of active reflections with largest |E| values (up to 4000) subject to a minimum value of |E| = 1.2. Default number is computed by the program.

BFACTOR *x*

Overall isotropic thermal factor, if the user wants to supply it. (The scale factor is assumed equal to 1).

ELECTRONS

This directive specifies that electron diffraction data will be used.

NEUTRONS

This directive specifies that neutron diffraction data will be used.

Specific directives for macromolecules:

MTZ *string*

String is the name of the reflections file in MTZ format.

LABELS *strings*

Strings are the names of columns in the MTZ reflections file.

SEQUENCE *string*

String is the name of the sequence file (FASTA format).

RESIDUES *n*

Number of residues in the asymmetric unit.

WEIGHT *x*

Molecular weight (in Dalton) in the asymmetric unit.

NMOLECULES *n*

Number of equal molecules in the asymmetric unit if **SEQUENCE** has been used.

HETATOMS El_1 n_1 El_2 n_2 El_3 n_3 ....

Heterogenic atoms content. El_i is the chemical symbol of atomic type i, n_i is the corresponding number of atoms in the asymmetric unit.

Specific directives for substructure determination:

ANOMALOUS *El n f'1 f"1* [*f'2 f"2* ...]

El is the chemical symbol of the anomalous scatterer, *n* is the corresponding number of atoms in the asymmetric unit cell, *f'* and *f"* are the values corresponding to the first wavelength used. Up to four wavelengths are allowed. This directive is used for SAD/MAD data. If more than one anomalous scatterer is used, the ANOMALOUS directive has to be repeated for all of them.

DERIVATIVE *El n f' f"*

El is the chemical symbol of the heavy atom used, *n* is the corresponding number of atoms in the asymmetric unit cell. This directive is used for SIR/MIR data. If *f'* and *f"* are different from zero, they correspond to the wavelength used in a SIRAS/MIRAS experiment. If more than one heavy atom is used, the DERIVATIVE directive has to be repeated for all of them.

NATIVE

Used when also the native data are supplied in the reflections file.

Invariants routine

The following directives can be used also in Phase routine

GMIN *x*

Triplets with G < *x* are not actively used. Default threshold value is computed by the program (in any case *x* > 0.1).

COCHRAN

The program uses the Cochran distribution. (P_{10} formula is used by default).

NOQUARTETS

The program does not actively use negative quartet relationships.

NOPSI0

The program does not actively use PSI-0 triplet relationships.

Phase routine

MDM [n_1 n_2]

The program applies the Modern Direct Methods procedure to n_1 sets of random phases. Default n_1 value is automatically chosen by the program, on the basis of the structural complexity. The program starts from the MDM n_2^{th} trial (default n_2 = 1).

SDM [n_{11} n_{12} .... n_{i1} n_{i2}]

The program applies the Standard Direct Methods procedure. If n_{11} and n_{12} (or more pairs) are supplied by the user, the program explores blocks (up to 10) of n_1 tangent trials and then selects the most promising n_2 (up to 500) phase sets (those with the highest eFOM score) to perform DSR procedure. Default n_1 and n_2 values are automatically selected by the program, on the basis of the structural complexity.

VLD [n_1 n_2 n_3]

The program applies the VLD procedure. n_1 is the number of cycles per trial, n_2 is the number of trials, n_3 is starting trial. Default values for n_1 , n_2 , n_3 are 300, 20, 1 respectively.

PATTERSON [n_1 n_2]

The program applies the Patterson procedure. n_1 is the total number of Patterson trials explored by the program. Default n_1 value is automatically chosen by the program, on the basis of the structural complexity. The program starts from the Patterson n_2^{th} trial (default n_2 = 1).

BEST *n*

The program stores the *n* best trials in terms of R factor. Default *n* = 1 for X-Ray data, *n* = 5 for electron diffraction data.

RESIDUAL *x*

The program stops if the final crystallographic residual factor (Rf%) is less than *x*. The default value is 25% for X-Ray data, is 0% for electron diffraction data. This directive is meaningless for macromolecules.

FRAGMENT *string*

Used to visualize a known model. *String* is the name of the coordinates file (*.cif, *.res, *.ins, *.pdb).

RECYCLE *string*

Used to refine and complete a known fragment. *String* is the name of the coordinates file (*.cif, *.res, *.ins, *.pdb).

SHELX

The program will produce the coordinates output file in SHELX format.

CRYSTALS

The program will produce the coordinates output file in CRYSTALS format.

Simulated Annealing

NRUN *n*

To modify the number of Simulated Annealing runs. The default is 10 runs.

NITER *n*

To modify the number of moves for each temperature step. In default the number of moves is automatically calculated.

RESMAX *x*

To define the maximum resolution used by Simulated Annealing. The default is x=2.0 Å.

TEMPER *temp*

To modify the initial temperature. In default the initial temperature is automatically calculated.

TFACTOR *x*

x ranges between 0 and 1 and determines the cooling rate. The default value is 0.90.

FIXROTATION n_1 n_2

To fix the internal DOF. n_1 and n_2 are the atoms associated to the DOF. This directive is useful when relying on the prior chemical knowledge of the structure to fix some of its internal DOFs. For example, double bonds in conjugated acyclic systems, triple bonds, nitrilic bonds, etc.

FFTRANSLATION *n*

To fix the translation of fragment containing *n* atom, where *n* is the order number of any atom in the fragment. This directive is repeated in case of fixing the translation of different fragments.

FFROTATION *n*

To fix the rotation of fragment containing *n* atom, where *n* is the order number of any atom in the fragment. This directive is repeated in case of fixing the rotation of different fragments.

REFINETF

To refine the thermal parameters.

SHIFT_ATOM *x* or **SHIFT_ATOM** *x* n_1 n_2 n_3 …

To refine the atomic parameters by applying shifts (up to *x*) on the atoms with respect to the center of gravity of the fragment. The suggested *x* is 0.02. Add n_1 , n_2 , n_3 ,… to refine only some specific atoms.

COST_FUNCTION *n*

To choose the cost function: 1 for R structure factor, 2 for R intensities. The default is 1.

RANDOMIZE *n*

To randomize the internal and external DOFs and the atomic parameters (if refined). *n* is an optional parameters used as seed of random generator.

ROTAT n_1 n_2

To specify the internal DOF using the input file.

Molecular Replacement

PSEUDO

The pseudotranslational information will be actively used to solve the protein structure.

APPROT

The program just finds the orientation of a model and stops.

SHELX

The program export the output coordinates files in SHELX format (PDB format is the default).

MODEL

Starts the block of directives related to each model.

ENDMODEL

Ends the block of directives related to each model.

FRAGMENT *[HETAM] string*

String is the name of the coordinates file (*.pdb). HETAM allows the program to use Heterogenic atoms content.

IMPROVE

The MR model will be improved using VLD and, if the resolution is worse than 1.25 Å, by means of the Free Lunch algorithm.

REFINE *n*

The MR model will be refined by Refmac (if installed) using batches on *n* refinement cycles. The default is REFINE 10.

NOREFINE

No refinement by Refmac of the MR model.

NOFREELUNCH

The Free Lunch algorithm will not be used.

ALIGNMENT *string*

String is the name of the file (in FASTA format) containing the alignment between the sequence of the target and the sequence of the model.

POLY

The polyalanine model will be applied.

BFACTOR *x / wilson / pdb*

The atomic B factors are set to x or to the Wilson B-factor or to the PDB values.

CUT n_1 n_2

The program removes from the model some terminal residues if they lie 'outside' n_1 and n_2 limits.

RANGE *x y*

Resolution range (in Å) the program will use. If only one limit is given, it is the high resolution limit; otherwise must be $x < y$.

THRESHOLD *x y z w*

The program selects the solutions to be analyzed after the rotation, the clustering, the translation and the optimization stage; the cutoff value is the maximum normalized distance from the best solution FOM. The default value is 0.1 for all the stages. In case of difficult structures with sequence identity less than 45% the values is automatically set to 0.15.

MONOMERS *n*

The number of monomer to search is set to *n*.

IDSEQUENCY *x*

Sequence identity with the target sequence (between 0.0 and 1.0). In default it will be calculated by the program.

POSITION r_1 r_2 r_3 t_1 t_2 t_3

The rotation (r_1 , r_2 , r_3 eulerian angles) and the translation (t_1 , t_2 , t_3 fractional values) are applied to the model.

ROTSELECTED r_1 r_2 r_3

The rotation (r_1 , r_2 , r_3 eulerian angles) is applied to the model; the procedure continues normally.

TRASELECTED t_1 t_2 t_3

The translation (t_1 , t_2 , t_3 fractional values) is applied to the model; the procedure continues normally.

Created with the Personal Edition of HelpNDoc: Full-featured Help generator

Examples

Example 1

The following example shows the maximum default use of Sir2014. Most of the structures can be solved in this way. Diffraction data are in the file xanthene.hkl, in format (3i4, 2f8.2), one reflection per record. This structure has been downloaded from the **Crystallography Open Database** (**COD**, Grazulis *et al*., 2009).

```
%structure xanthene
%job xanthene (COD 2224920)
%data
  cell 11.2438  20.1671  23.6474  90.0000  90.0000  90.0000
  space P n a 21
  cont C 220 H 320 N 4 O 44
  refl xanthene.hkl
%continue
```

Example 2

In the following example the experimental data are stored as $|F_o|$ (not $|F_o|^2$), using the free format.

```
%Structure    iled
%Job          lsoleucinomycin
%Data
      cell           11.516  15.705  39.310  90.00  90.00  90.00
      spacegroup     P 21 21 21
      content        C 240  H 408  N 24  O 72
      reflections    iled.hkl
      format         free
      fobs
%Continue
```

Example 3

In the following example, no space group is supplied; the program performs the automatic identification of the Laue group and of the extinction symbol from diffraction

intensities.

```
%structure xanthene
%job xanthene (COD 2224920)
%data
  cell 11.2438  20.1671  23.6474  90.0000  90.0000  90.0000
  cont C 220 H 320 N 4 O 44
  refl xanthene.hkl
%continue
```

Example 4

In the following example, the directive spacegroup SCAN is used; the program performs the automatic identification of the Laue group and of the extinction symbol and consecutively explores (performing a complete solution run) all the selected groups.

```
%structure xanthene
%job xanthene (COD 2224920)
%data
  cell 11.2438  20.1671  23.6474  90.0000  90.0000  90.0000
 Space  scan
  cont c 220 h 320 n 4 o 44
  refl xanthene.hkl
%continue
```

Example 5

The user wants to supply the value for the isotropic thermal factor and to set the number of strong ($|E_o|$ value) reflections. Formula directive is used to supply the chemical cell content.

```
%Structure loganin
%Job    loganin
%Data
      cell             8.187  14.277  15.693  90.00  90.00  90.00
      spacegroup     P 21 21 21
      formula         (C17  H26  O10)4
      reflections      loganin.hkl
      fobs
      bfac             3.5
      nref             200
%Continue
```

Example 6

In the following example, the Cochran formula is applied and all triplets with a concentration parameter greater than 0.4 are actively used in the phasing process, as requested by the user. The binary file "loganin.bin" must exist. Commands or directives following ">" or "!" are interpreted as a comment and will be ignored.

```
%STRUCTURE LOGANIN
      >%DATA
         >CELL    8.187  14.277  15.693  90.00  90.00  90.00
      ! spacegroup     P 21 21 21
      ! content          C 68  H 104  O 40
      >reflections       loganin.hkl
         >FOBS
%PHASE
      GMIN 0.4
      COCHRAN
%END
```

Example 7

The user wants to explore 100 trials, starting from the trial number 132, from the phasing process up to the Fourier-Least Squares refinement. The binary file "iled.bin" must exist.

```
%Structure iled
%Phase
   MDM   100   132
 %End
```

Example 8

The user wants to explore only trial number 154 using 300 cycles of VLD instead of the default procedure (MDM).

```
%Structure loganin
%Phase
      VLD 300 1 154
%End
```

Example 9

The program explores only the Patterson trial associated with the peak number 5; it applies the Direct Space Refinement strategy for medium size structures and stops if the final residual value (Rf %) is less than the threshold indicated by the user.

```
%Structure conotoxin
%Phase
      Patterson 1   5
      Residual 10.0
      Size m
%End
```

Example 10

In the following example the user knows a fragment and wants to complete it using the Fourier-Least Squares procedure. The file "choror.cif" must exist.

```
%STRUCUTRE  choror
%Phase
    recycle   choror.cif
%CONTINUE
```

Example 11

The user wants to apply the VLD phasing approach; for proteins, like in this example, the default is Patterson.

```
%Structure crambin
%job  Crambin - PDB Code 1CNR
%Data
   Cell   40.763 18.492 22.333 90.00 90.61 90.00
   SpaceGroup  P 21
   Content    C 406 H 776 N 110 O 131 S 12
   Reflections  crambin.hkl
%Phase
   VLD
%End
```

Example 12

Example of Sir2014 input file for electron diffraction data.

```
%Job charoite - Courtesy of Enrico Mugnaioli
%Structure charoite
%Data
        Cell   31.9600   19.6400    7.0900   90.0000   90.0000   90.0000
        Spacegroup  P 21/m
        Content     O 186 Na 10 Si 72 K 14 Ca 24
        Reflections charoite.hkl
        Format(3i4, 2f8.2)
        Fosq
        Wavelength  0.019700
        Electrons
  %Phase
  %End
```

Example 13

Example of Sir2014 input file for Simulated Annealing using electron diffraction data.

```
%structure cnba
%job CNBA Courtesy of Enrico Mugnaioli
%data
  cell 14.70 9.47 15.42 90 112 90
  spacegroup P21/c
  electrons
  content C 116 N 4 H 68
  reflections CNBA.hkl
%import antracene.mol
%sannel
%end
```

Example 14

Example of Sir2014 input file for Molecular Replacement.

```
%structure 2sar
%job test structure: pdb code 2sar
%data
    mtz 2sar.mtz
   label H K L FP SIGFP
   Sequence 2sar.seq
   nMolecules 1
   hetatm S 10 P 1
%remo
   model
      fragment  hetam 1ucl_mod1_lsq1.pdb
```

```
      monomers   2
      BFactor PDB
    endmodel
%end
```

Example 15

Example of Sir2014 input file for MAD (4 Se atoms) substructure determination.

```
%structure ttg_task1
%initialize
%job THERMUS THERMOPHILUS GROEL - PDB: 1SRV
%data
  mtz ttg.mtz
  label H K L FP SIGFP F(+)SE1 SIGF(+)SE1 F(-)SE1 SIGF(-)SE1 F(+)SE2 SIGF(+)
SE2 F(-)SE2 SIGF(-)SE2 F(+)SE3 SIGF(+)SE3 F(-)SE3 SIGF(-)SE3 F(+)SE4 SIGF(+)
SE4 F(-)SE4 SIGF(-)SE4
  nResidues 145
  nMolecules 1
  anomalous Se 3 -1.805 0.646 -8.582 3.843 -7.663 3.841 -2.618 3.578
  native
%phase
  Patterson
%end
```

Created with the Personal Edition of HelpNDoc: Easily create PDF Help documents

References

Adams, P.D., Afonine, P.V., Chen, V., Echols, N., Headd, J. J., Hung, L.-W., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. and Zwart, P. H. (2010), *Acta Cryst.*, D**66**, 213-221.

Altomare, A., Camalli, M. Cuocci, C., Giacovazzo, C., Moliterni, A & Rizzi, R. (2009), *J. Appl. Cryst.*, **42**, 1197-1202.

Altomare, A., Cascarano, G., Giacovazzo, C. & Guagliardi, A. (1993), *J. Appl. Cryst.*, **26**, 343-350.

Altomare, A., Cuocci, C., Giacovazzo, C., Kamel, G.S., Moliterni, A & Rizzi, R. (2008), *Acta Cryst.*, A**64**, 326-336.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A & Rizzi, R. (2008), *J. Appl. Cryst.*, **41**, 592-599.

Arnott, S.& Wonacott, A.J. (1966), *Polymer*, **7**, 157-166.

Baggio, R., Woolfson, M.M., Declerq, J.P. & Germain, G.(1978), *Acta Cryst.*, A**34**, 883-892.

Burla, M.C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G.L., De Caro L., Giacovazzo, C. Polidori, G. & Spagna, R. (2005), *J. Appl. Cryst.*, **38**, 381-388.

Burla, M.C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G.L., De Caro L.,

Giacovazzo, C., Polidori, G., Siliqi, D. & Spagna, R. (2007), *J. Appl. Cryst.*, **40**, 609-613.
Burla, M.C., Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacovazzo, C. & Polidori, G. (2004a), *J. Appl. Cryst.*, **37**, 258-264.
Burla, M.C., Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacovazzo, C. & Polidori, G. (2004b), *J. Appl. Cryst.*, **37**, 791-801.
Burla, M.C., Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacovazzo, C., Polidori, G. & Siliqi, D.(2006a), *J. Appl. Cryst.*, **39**, 527-535.
Burla, M.C., Caliandro, R., Carrozzini, B., Cascarano, G.L:, De Caro L., Giacovazzo, C., Polidori, G. & Siliqi, D.(2006b), *J. Appl. Cryst.*, **39**, 728-734.
Burla, M.C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G.L:, Giacovazzo, C., Mallamo, M., Mazzone, A., Polidori, G. & Spagna, R. (2012), *J. Appl. Cryst.*, **45**, 357–361.
Burla, M.C., Caliandro, R., Giacovazzo, C. & Polidori, G. (2010a), *Acta Cryst.,* A**66**, 347-361.
Burla, M.C., Carrozzini, B., Cascarano, G.L., De Caro, L., Giacovazzo, C., & Polidori, G. (2003), *Acta Cryst.*, A**59**, 245-249.
Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C. & Polidori, G. (2002), *Z. Kristallogr.*, **217**, 629-635.
Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C. & Polidori, G. (2011), *J. Appl. Cryst.*, in preparation
Burla, M.C., Cascarano, G.L. & Giacovazzo, C. (1992), *Acta Cryst.*, A**48**, 906-912.
Burla, M.C., Giacovazzo, C. & Polidori, G. (2010), *J. Appl. Cryst.* **43**, 825-836.
Burla, M.C., Giacovazzo, C. & Polidori, G. (2011), *J. Appl. Cryst.* **44**, 193-199.
Burla, M.C., Giacovazzo, C. & Polidori, G. (2013), *J. Appl. Cryst.* **46**, 1592-1602.
Bürger, M. J. (1959), *Vector Space*, chapter 11. Wiley, New York.
Burzlaff, H. & Hountas, A. (1982), *J. Appl. Cryst.*, **15**, 464-467.
Caliandro, R., Carrozzini, B., Cascarano, G.L., Comunale, G. & Giacovazzo C. (2013), *Acta Cryst.*, A**69**, 98–107.
Caliandro, R., Carrozzini, B., Cascarano, G.L., Comunale, G., Giacovazzo, C. & Mazzone A. (2014), *J. Appl. Cryst.*, submitted.
Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacovazzo, C. & Siliqi, D. (2005a), *Acta Cryst.*, A**61**, 556-565.
Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacovazzo, C. & Siliqi, D. (2005b), *Acta Cryst.*, A**61**, 1080-1087.
Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacovazzo, C., Moustiakimov, M. & Siliqi, D. (2005), *Acta Cryst.*, A**61**, 343-349.
Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacovazzo, C. & Siliqi, D. (2007), *J. Appl. Cryst.*, **40**, 883-890.
Caliandro, R., Carrozzini, B., Cascarano, G.L., Giacovazzo, C., Mazzone, A. & Siliqi, D. (2009), *Acta Cryst.*, A**65**, 512-527.
Camalli, M., Carrozzini, B., Cascarano, G.L., & Giacovazzo, C. (2011), *Cryst. Res. Technol.* 46, No. 6, 555 – 560.
Carrozzini, B., Cascarano, G.L., Comunale, G., Giacovazzo, C. & Mazzone, A. (2013), *Acta Cryst.* D**69**, 1038–1044.
Cascarano,G.L. & Giacovazzo, C. (1995), *Acta Cryst.*, A**51**, 820-825.
Cascarano,G.L., Giacovazzo, C. & Carrozzini, B. (2010), *Ultramicroscopy*, **111**, 56-61.
Cascarano,G.L., Giacovazzo, C. & Guagliardi, A. (1991), *Acta Cryst.*, A**47**, 698-702.
Cascarano, G.L., Giacovazzo, C. & Luić, M. (1988a), *Acta Cryst.*, A**44**, 176-183.
Cascarano, G.L., Giacovazzo, C. & Luić, M. (1988b), *Acta Cryst.*, A**44**, 183-188.
Cascarano, G.L., Giacovazzo, C., Burla, M.C., Nunzi, A. & Polidori, G. (1984), *Acta Cryst.*, A**40**, 389-394.

Cochran, W. (1955), *Acta Cryst*., **8**, 473-478.
Cowtan, K. (2006) *Acta Cryst*., D**62**, 1002-1011.
Fan, Hai-Fu, Yao, Jia-Xing & Qian, Jin-Zi (1988), *Acta Cryst*., A**44**, 688-691.
Frigo, M. and Johnson, S.G. (2005) *Proceedings of the IEEE* **93** (2), 216–231
Giacovazzo, C. (1976), *Acta Cryst*., A**32**, 958-966.
Giacovazzo, C. (1977), *Acta Cryst*. A**33**, 933-944.
Giacovazzo, C. (1980), *Acta Cryst*. A**36**, 362-372.
Giacovazzo, C., Burla, M.C. & Cascarano, G. (1992), *Acta Cryst*., A**48**, 901-906.
Grazulis, S., Chateigner, D., Downs, R.T., Yokochi, A.T., Quirós, M., Lutterotti, L., Manakova, E., Butkus, J., Moeck, P. & Le Bail, A. (2009), *J. Appl. Cryst*., **42**, 726-729.
Harris, K.D., Tremayne, M. & Kariuki, B.M. (2001), *Angew. Chem. Int*., **40**, 1626-1651.
Leslie, A.G.W. (1987), *Acta Cryst*., A**43**, 134-136.
Matthews, B.W. (1968), *J. Mol. Biol*., **33**, 491-497.
Mooers, B.H., Matthews, B.W. (2006) *Acta Cryst*., D**62**: 165-176.
Perrakis, A., Morris, R.M. & Lamzin, V.S. (1999), *Nature Struct. Biol*., **6**, 458-463.
Refaat, L.S. & Woolfson, M.M. (1993), *Acta Cryst*., D**49**, 367-371.
Richardson, J.W. & Jacobson, R.A. (1987), In: *Patterson and Pattersons*. Ed. by Glusker, J.P., Patterson, B.K. &. Rossi, M, pp. 310-317. Oxford University Press.
Sheldrick, G.M. (1992), In: *Crystallographic Computing **5***. Ed. by Moras, D., Podjarny, A.D. & Thierry, J.C., pp. 145-157. Oxford University Press.
Sheldrick, G.M. & Gould, R.O. (1995), *Acta Cryst*., B**51**, 423-431.
Sim, G.A. (1959), *Acta Cryst*., **12**, 813-815.
Sim, G.A. (1960), *Acta Cryst*., **13**, 511-512.
Spagna, R. & Camalli, M. (1999), *J. Appl. Cryst*., **32**, 934-942.
Wang, B.C. (1985), *Methods Enzymol*., **115**, 90-112.
Wilson, A.J.C. (1942), *Nature*, **150**, 152.

Created with the Personal Edition of HelpNDoc: Easily create CHM Help documents