

Sir2019

Crystal structures solution by single crystal data

PDF generated March 04, 2019 by Kalin's PDF Creation Station WordPress plugin

Table Of Contents

Commands and their use	3
Data module	5
Description of Sir2019	9
Directives	11
Examples	21
File	27
General information	41
Getting started (small molecules)	44
Installing the program	61
Invariants module	64
Molecular Replacement	65
Notes on distribution	78
Phase module	79
Phase refinement	86
References	96
Refine	101
SAD/MAD structure determination	109
Simulated Annealing	121
Solve	134
The Graphic Interface	138
View	140
When default Sir2019 fails: some advices	148

Commands and their use

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/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.

Sir2019 recognizes the following commands:

%WINDOW	A window is opened to monitor the process of solution and refinement of the structure (default).
%NOWINDOW	The program is forced to run without graphics.
%DATA	Data input routine.
%INVARIANTS	Invariants routine.
%PHASE	Phasing routine + Direct Space Refinement routine.
%IMPORT <i>file</i>	Used to import a structural model (for Simulated Annealing).
%SANNEL	Simulated Annealing routine.
%REMO	Molecular Replacement routine.
%MAD	Protein structure determination routine using SAD/MAD data.
%END	End of the input file
%JOB <i>string</i>	A caption is printed in the output.
%CONTINUE	The program runs in default conditions from the last given command up to the end.
%STRUCTURE <i>string</i>	This command is used to specify the name of the structure to investigate.
%AMB <i>string</i>	This command is used to specify the name of the Automated Model Building program (Buccaneer, Nautilus, ARP/wARP, Phenix) to be used (proteins only).
%CAB <i>string</i>	This command is used to specify the name of the Automated Model Building program (Buccaneer, Nautilus, ARP/wARP, Phenix) to be used in the CAB (cyclic automated model-building); proteins only.
%REFINE	This command is used to refine phases produced in a previous run of the program or by a different program. Phantom Derivative procedure is 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. Add the blue line at the beginning of these two files:

import sys

`sys.path.append("/home/john/ccp4-7.0/lib/py2/")`

from CCP4pipeline import Control

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.plt ? coordinate file for graphics

string.pdb ? coordinate file in PDB format (proteins only)

string.mtz ? reflections file in MTZ format (proteins only)

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 [here](#), 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.

Sir2019 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 Sir2019 is constituted by:

cell parameters;

cell content;

- reflections.

[Previous page](#)

[Next page](#)

Data module

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/data-module/>

This routine reads the basic crystallographic information like cell parameters, space group symbol, unit cell and/or sequence content and diffraction data. It includes a modified version of the subroutine SYMM (Burzlaff & Hountas, 1982). Symmetry operators and information necessary to identify structure invariants (estimated in the 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 (Casarano *et al.*, 1991).

Diffraction intensities are normalized using the Wilson Method (Wilson, 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 (Casarano *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 $|E|$ values) are selected for invariants calculations (if MDM or SDM approach is required); 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

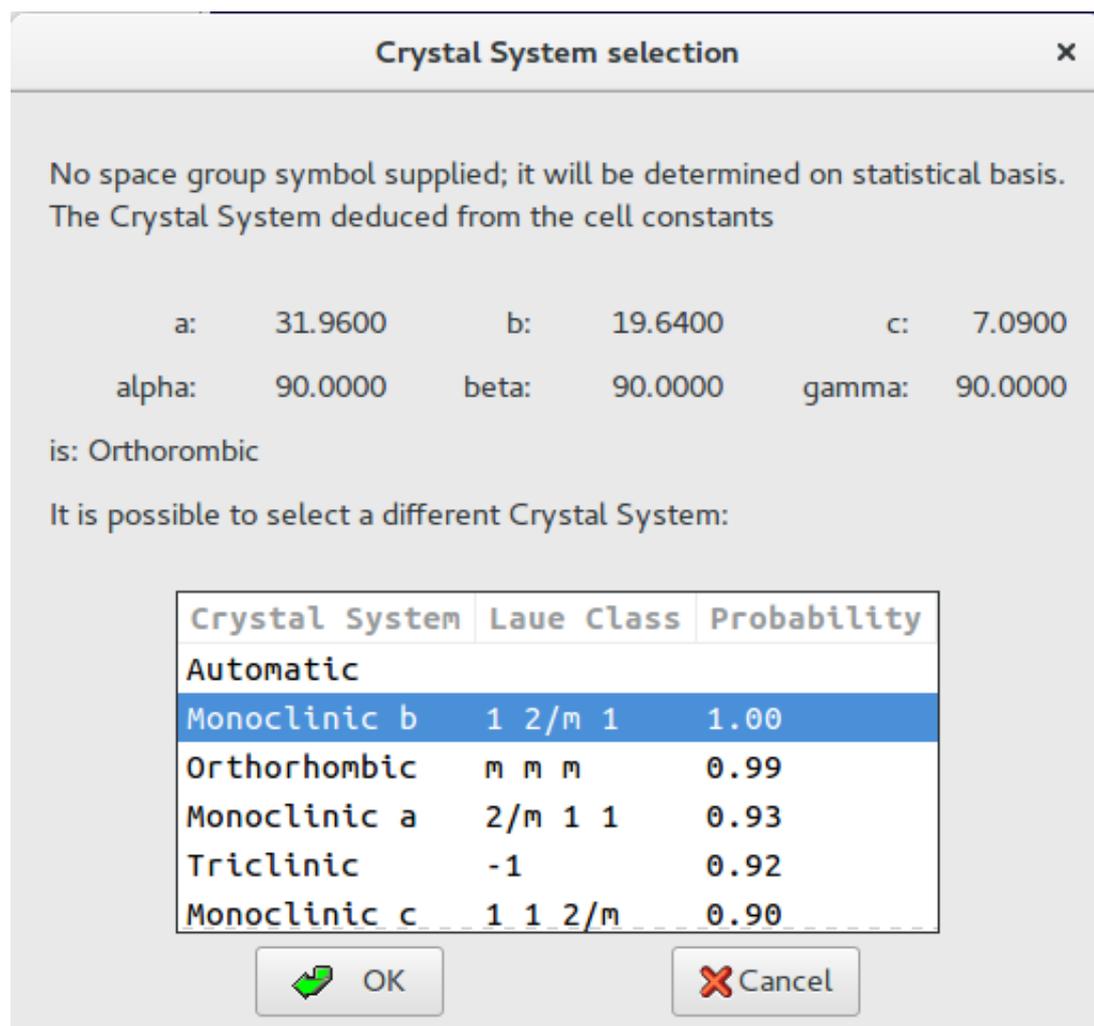
An algorithm (Camalli *et al.*, 2011) has been developed for the automatic identification of the Laue group and of the extinction symbol from diffraction intensities. Its efficient use requires the presence of all the measured reflections in the experimental data, weak and potentially extincted reflections included.

The algorithm has statistical basis, is of general use, and has been specialized for electron data. It 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.

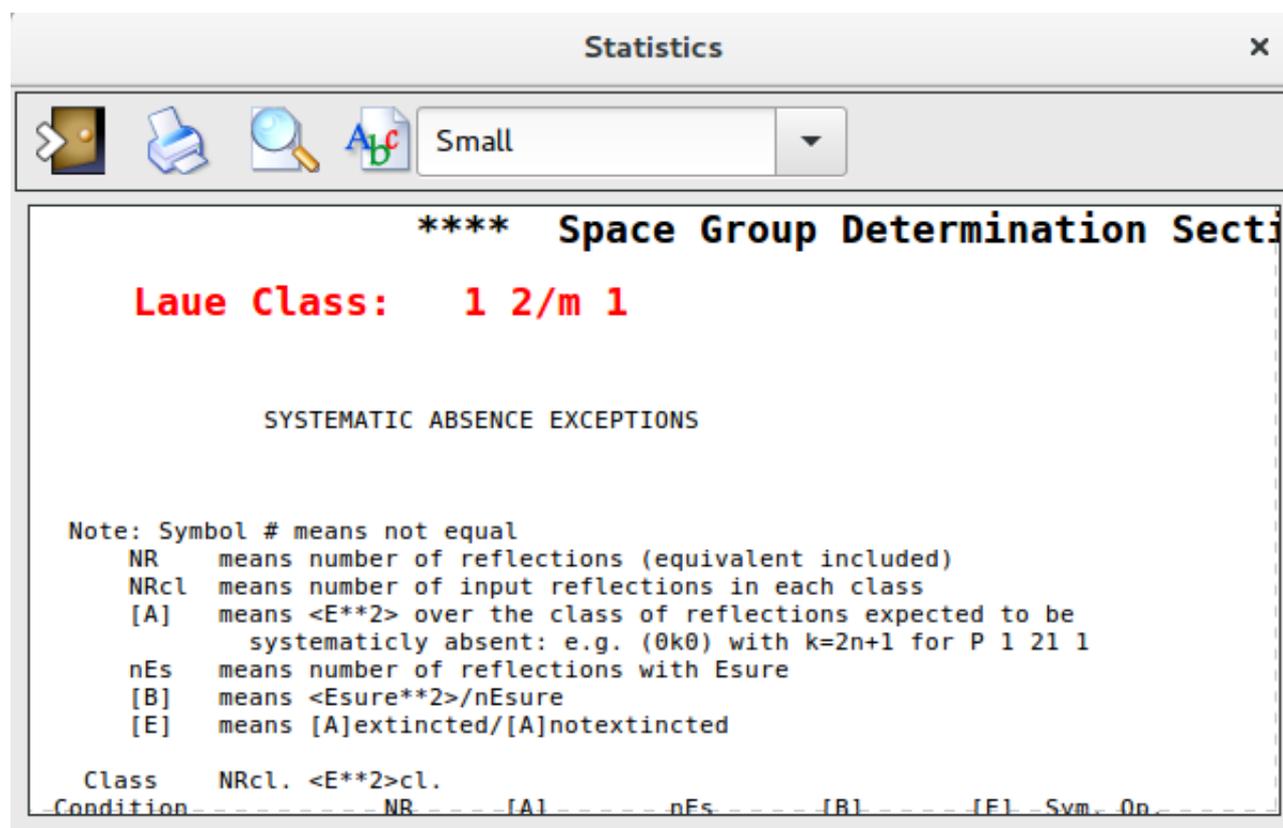
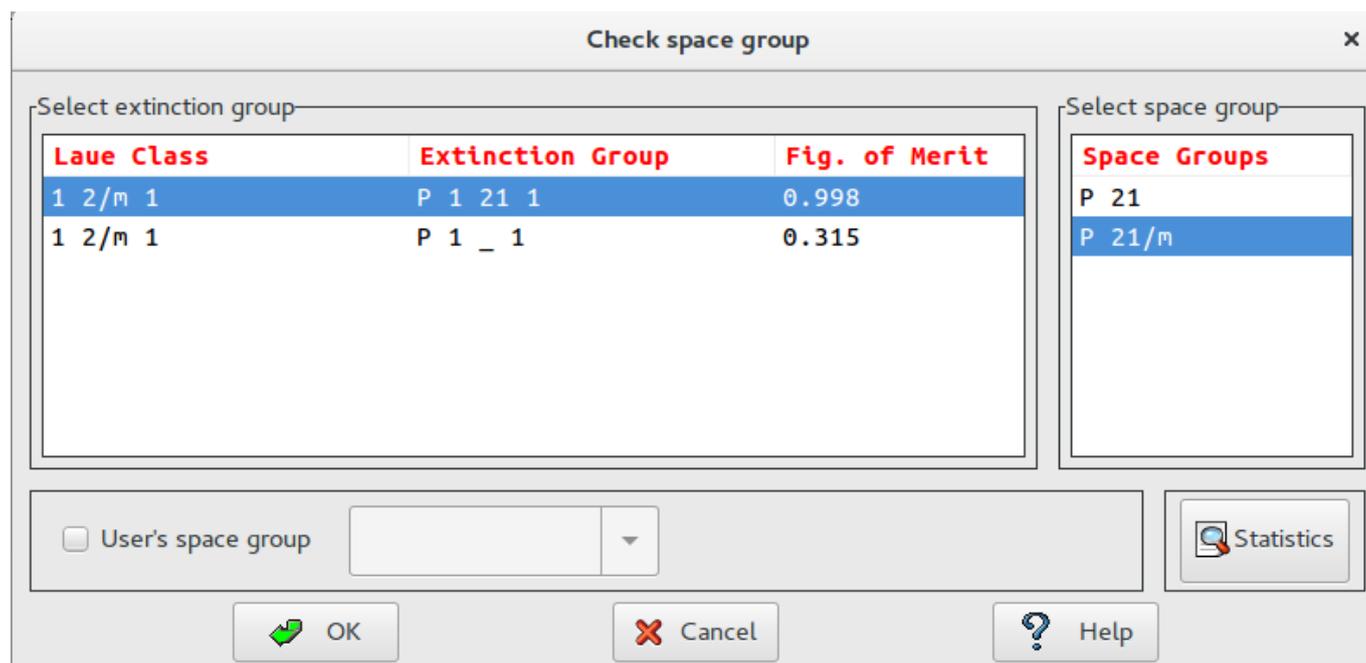
In the following example, the cell constants are orthorhombic, but the true space group, as suggested by the program, is $P 2_1/m$.

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



In the second image, the Laue Class is “1 2/m 1” and the Extinction Symbol “P 1 21 1” has 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.

To get help click on 



Often the same extinction 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 extinction 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.

[Previous page](#)
[Next page](#)

Description of Sir2019

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/description-of-sir2018/>

The main modules of the program are: [DATA](#), [INVARIANTS](#) (if MDM or SDM is required), [PHASE](#), [Simulated Annealing](#) (SANNEL), [Molecular Replacement](#) (REMO) and [SAD/MAD structure determination](#) (MAD).

If the [graphic interface](#) is active, it is possible to visualize the model using  Jav.

If the number of atoms in the asymmetric unit does not exceed 500 it is possible to modify, complete and refine the model using the tools provided by Jav.

A web based help is available in order to describe the program and all the features and tools accessible *via* graphics.

Information about the program version can be obtained through the About window if graphics is active



or using the command

sir2019 -version

from the terminal window.

Thanks are due to J. Gonzalez-Platas (2D Contouring tools) for the code integrated in Sir2019.

[Previous page](#) [Next page](#)

Directives

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/directives/>

List of the directives associated to the commands

[%Data](#)

[%Invariants](#)

[%Phase](#)

[%Sannel](#)

[%Remo](#)

[%MAD](#)

[%Refine](#)

Data routine

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

CELL a b c ? ? ?

Cell dimensions: a, b and c are in Å, ?, ? and ? in degrees.

ERRORS esd(a) esd(b) esd(c) esd(?) esd(?) esd(?)

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 (E1n1E12n2E13n3 ...)Z

Unit cell content. Eli is the chemical symbol of atomic type i, ni is the corresponding number of atoms in the chemical formula, Z is the number of molecules in the unit cell. To avoid ambiguities, the two-characters elements

must have the second character in lower case i.e. formula (BaSO4)4.

OR

CONTENT El1 n1 El2 n2 El3 n3

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, |Fo|, ?(Fo) (or |Fo|², ?(Fo)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 |Fo| are allowed; negative values of ?(Fo) 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 ?(Fo) values are meaningless or not available.

FOBSERVED

Program assumes h,k,l, |Fo|, ?(Fo). |Fo|² and ?(Fo)² are expected as default.

FOSQUARED

Program assumes h,k,l, |Fo|² and ?(Fo)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 $(\sin^2\theta)^2$ 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 E11 n1 E12 n2 E13 n3

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

Specific directives for SAD/MAD structure determination:

ANOMALOUS E1 n f'1 f''1 [f'2 f''2 ...]

E1 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.

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. (P10 formula is used by default).

NOQUARTETS

The program does not actively use negative quartet relationships.

NOPSI

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

Phase routine

MDM [n1 n2]

The program applies the Modern Direct Methods procedure to n1 sets of random phases. Default n1 value is automatically chosen by the program, on the basis of the structural complexity. The program starts from the MDM n2th trial (default n2 = 1).

SDM [n11 n12 ni1 ni2]

The program applies the Standard Direct Methods procedure. If n11 and n12 (or more pairs) are supplied by the user, the program explores blocks (up to 10) of n1 tangent trials and then selects the most promising n2 (up to 500) phase sets (those with the highest eFOM score) to perform DSR procedure. Default n1 and n2 values are automatically selected by the program, on the basis of the structural complexity.

VLD [n1 n2 n3]

The program applies the VLD procedure. n1 is the number of cycles per trial, n2 is the number of trials, n3 is starting trial. Default values for n1 , n2 , n3 are 300, 20, 1 respectively.

PATTERSON [n1 n2]

The program applies the Patterson procedure. n1 is the total number of Patterson trials explored by the program. Default n1 value is automatically chosen by the program, on the basis of the structural complexity. The program starts from the Patterson n2th trial (default n2 = 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.

Sannel (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 $x=2.0 \text{ \AA}$.

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 n1 n2

To fix the internal DOF. n1 and n2 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.

REFINETE

To refine the thermal parameters.

SHIFT_ATOM x or SHIFT_ATOM x n1 n2 n3 ...

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 n1 , n2 , n3 ,... 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 n1 n2

To specify the internal DOF using the input file.

Remo (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 n1 n2

The program removes from the model some terminal residues if they lie 'outside' n1 and n2 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 r1 r2 r3 t1 t2 t3

The rotation (r1 , r2 , r3 eulerian angles) and the translation (t1 , t2 , t3 fractional values) are applied to the model.

ROTSELECTED r1 r2 r3

The rotation (r1 , r2 , r3 eulerian angles) is applied to the model; the procedure continues normally.

TRASELECTED t1 t2 t3

The translation (t1 , t2 , t3 fractional values) is applied to the model; the procedure continues normally.

MAD (Protein structure determination)**WAVE** n1 n2 n3 n4

Which wavelengths will be used for substructure refinement and phase extension. The wavelength number order corresponds to that specified in %Data with the directive ANOMALOUS

For example:

WAVE 1 3 Only the first and the third wavelength will be used.

If the directive WAVE is absent all the wavelengths will be used.

SET n1 n2

Specifies the number (the range) of the substructures previously obtained.

n1, n2 = 1, 2 ... 10 (ordered by FOM_frag)

For example:

SET 1 2 the first two sets (in order of FOMfrag) will be used.

The default is SET 1 - 1, i.e. all fragments will be used.

LSQ nm

Conditions for the refinement (LSQ) of the substructure/wavelength parameters;

n=0,1,..5 specifies the options for substructure parameters (x,y,z, B-factor, sof).

0 no refinement

1 only B-factor

2 x, y,z

3 x, y, z, B-factor

4 only sof

5 x, y, z, B-factor, sof

The default value is 5.

m specifies the options for wavelength parameter (f' , df').

0 no refinement

1 f'

2 df'

3 f' , df'

The default value is 3.

For example:

LSQ 50 Refine only substructure parameters x,y,z, B-factor, sof

LSQ 53 Refine all parameters: x, y, z, B-factor, sof, f' and df'

POSITIONS string

String is the name of the file (*.pdb) which contains the coordinates of the anomalous substructure.

Default: the program will use the file *structurename_SIR200x.fra* written during the anomalous scatterer substructure determination.

Refine routine

MODEL string

String is the name of the coordinates file (*.pdb) where a previously created model is stored . (i.e. a Molecular Replacement result).

MTZ string

String is the name of the reflections file in MTZ format, if not already defined in %Data. Phase and fom columns are necessary to perform refinement.

LABELS strings

Strings are the names of columns in the MTZ reflections file, if not already defined in %Data.

MR

Phases have been produced by a Molecular Replacement procedure.

MAD

Phases have been produced by a MAD/SAD procedure.

ABI

Phases have been produced by an *ab initio* procedure.

[Previous page](#) [Next page](#)

Examples

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/examples/>

Example 1

The following example shows the maximum default use of Sir2019. 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](#)).

```
% window
% 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 |Fo| (not |Fo|²), using the free format.

```
% window
% Structure iled
% Job Isoleucinomycin
% 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.

```
% window
% 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.

```
%window
%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.

```
%window
%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.

```
%window
%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.

```
% window
%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).

```
% window
%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 R(F)% is less than the threshold (Residual) indicated by the user.

```
% window
%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.

```
% window
%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.

```
% window
%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 input file for electron diffraction data.

```
% window
%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 input file for Simulated Annealing using electron diffraction data.

```
% window
% 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 14**

Example of 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 input file for MAD (3 Se atoms) data. Buccaneer will be used to build the structure.

```
% structure ttg
% amb Buccaneer
% 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
```

%mad

%end

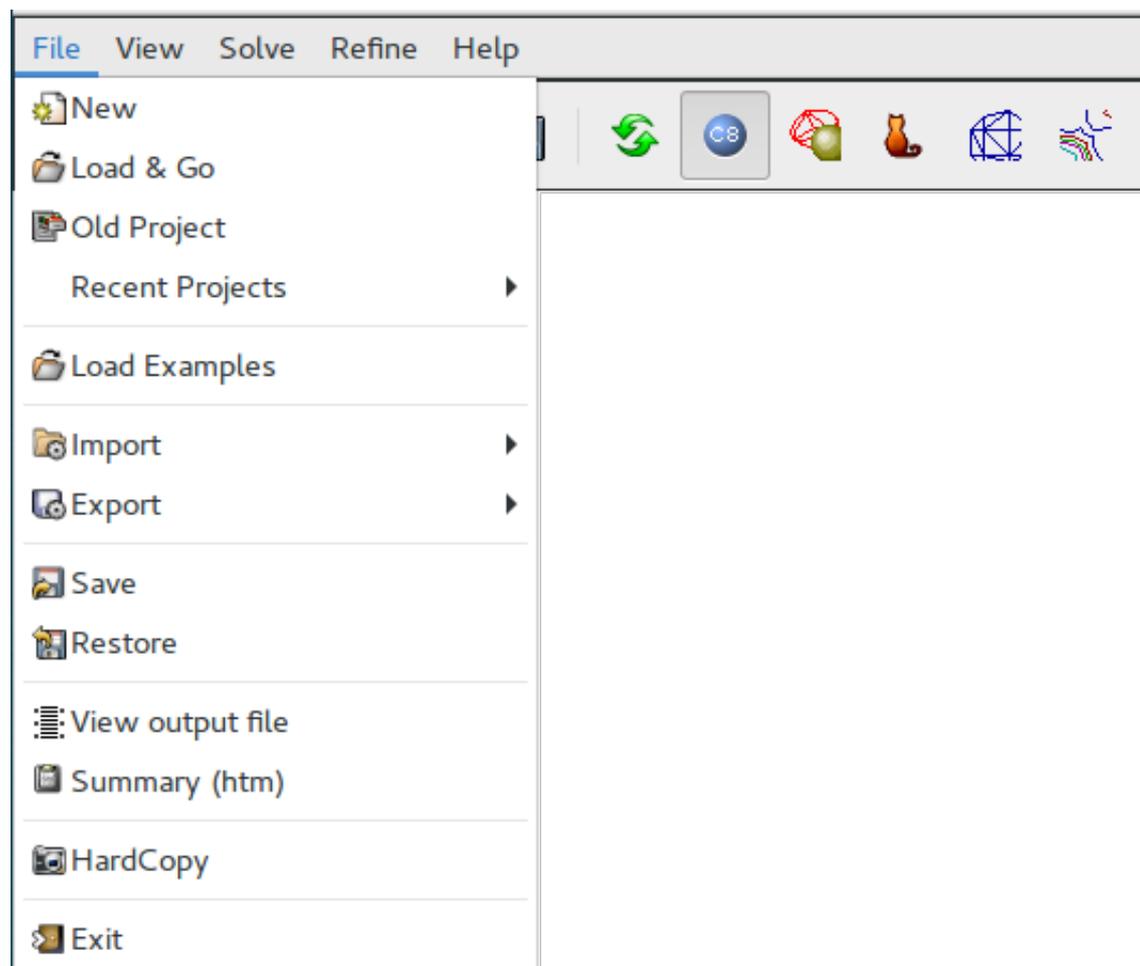
[Previous page](#) [Next page](#)

File

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/file/>

Sub-menu File:



It allows

Some of these can be accessed directly using icons in the toolbar.

 [New](#)

 [Load & Go](#)

 [Old Project](#)

[Recent Projects](#)

[Load Examples](#)[Import](#)[Export](#)[Save](#)[Restore](#)[View output file](#)[Summary \(htm\)](#)[Hard Copy](#)[Exit](#)

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. The size of the structure and the type of reflections file has to be set.

New Project for: barite ✕

General | Symmetry & Cell parameters

User Working Directory
 

Structure name or previous project (*)
 

Job title

Molecule size (*)
 Small Molecule Macro Molecule

Procedure (*)
 Molecular Replacement
 SAD/MAD
 Ab initio / Simulated Annealing
 Phase Refinement

Reflections (*)
 HKL file
 MTZ file

Automated Model Building (Macromolecules only)
 
 Use CAB procedure

(*) Required field

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.

New Project for: barite ✕

General Symmetry & Cell parameters

Symmetry (*)

Enter Space Group symbol (or number)

▼

Search for Space Group

Select from list Crystal Family: ▼

Scan all possible Space Groups

Cell parameters

Crystal family: Orthorombic, Spacegroup number: 62

a: (*) <input type="text" value="8.88"/>	b: (*) <input type="text" value="5.45"/>	c: (*) <input style="border: 2px solid blue;" type="text" value="7.15"/>
alpha: (*) <input type="text" value="90"/>	beta: (*) <input type="text" value="90"/>	gamma: (*) <input type="text" value="90"/>

(*) Required field

In the third window the user provides the unit cell content. In the following example electron diffraction data are used, so “*Electron data*” has to be checked.

New Project for: barite ✕

GeneralSymmetry & Cell parameters**Cell Content**Reflections

Source

X Ray Data Neutron Data Electron Data

Cell content:

Warning:
To avoid ambiguities, you should write the two-letter elements with the second letter in lower-case or insert a space between elements.

(BaSO4)4

Cell Volume: Volume per Atom:

(*) Required field

Accept Cancel

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](#). A preview Window is shown to define correctly the Format information.

New Project for: barite ✕

General Symmetry & Cell parameters **Cell Content** **Reflections** **Solve**

Reflections filename: (*)



Preview Window

```
1234123412341234567812345678
-10  2  -3  0.09  0.01
-10  2  -2  0.01  0.00
-10  2  -1  0.18  0.02
-10  2   0  0.02  0.00
```

Format

Assumed format

Blanks delimited values

Shelx format (3i4, 2F8.2)

sigma(F) not supplied

Reflections per record

Transformation Matrix

1		
	1	
		1

Structure factors

F(obs)

F(obs)**2

User supplied thermal factor:

Min. resolution value (Angstrom):

(*) Required field

 Accept
 Cancel

The fifth window requires information about the phasing procedure. Details about the possibilities offered in this window can be obtained looking in [Description of Sir2019: Phase module](#) or in [Directives: Phase Routine](#).

New Project for: barite ✕

General Symmetry & Cell parameters Cell Content Reflections **Solve**

Solving strategy

- Let the program to decide
- Modern Direct Methods (MDM)
- Standard Direct Methods (SDM)
- Patterson
- VLD
- Simulated Annealing

Size:

Let the program to decide ▼

< 6 non-H atoms in asymmetric unit

(*) Required field

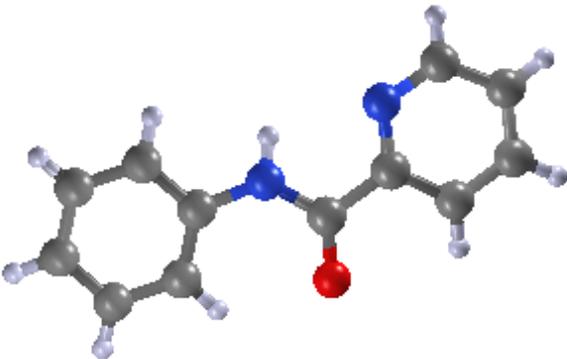
The last window informs the user that an input file is going to be created. The user, by clicking on  button, can save the input file for Sir2019.

New Project for: barite

General Symmetry & Cell parameters Cell Content Reflections Solve Run

Run

Graphic window required



Input filename (*) Already exists

barite.sir YES

Output filename (*)

barite.out YES

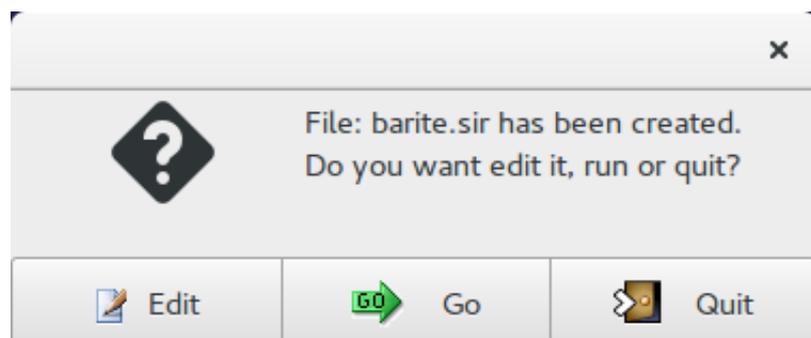
Click on  to save the project

(*) Required field

If **Graphic window required** is checked, a graphical output will be produced.

It is possible to change the default names for the input (project) and for the output file.

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



Load & Go

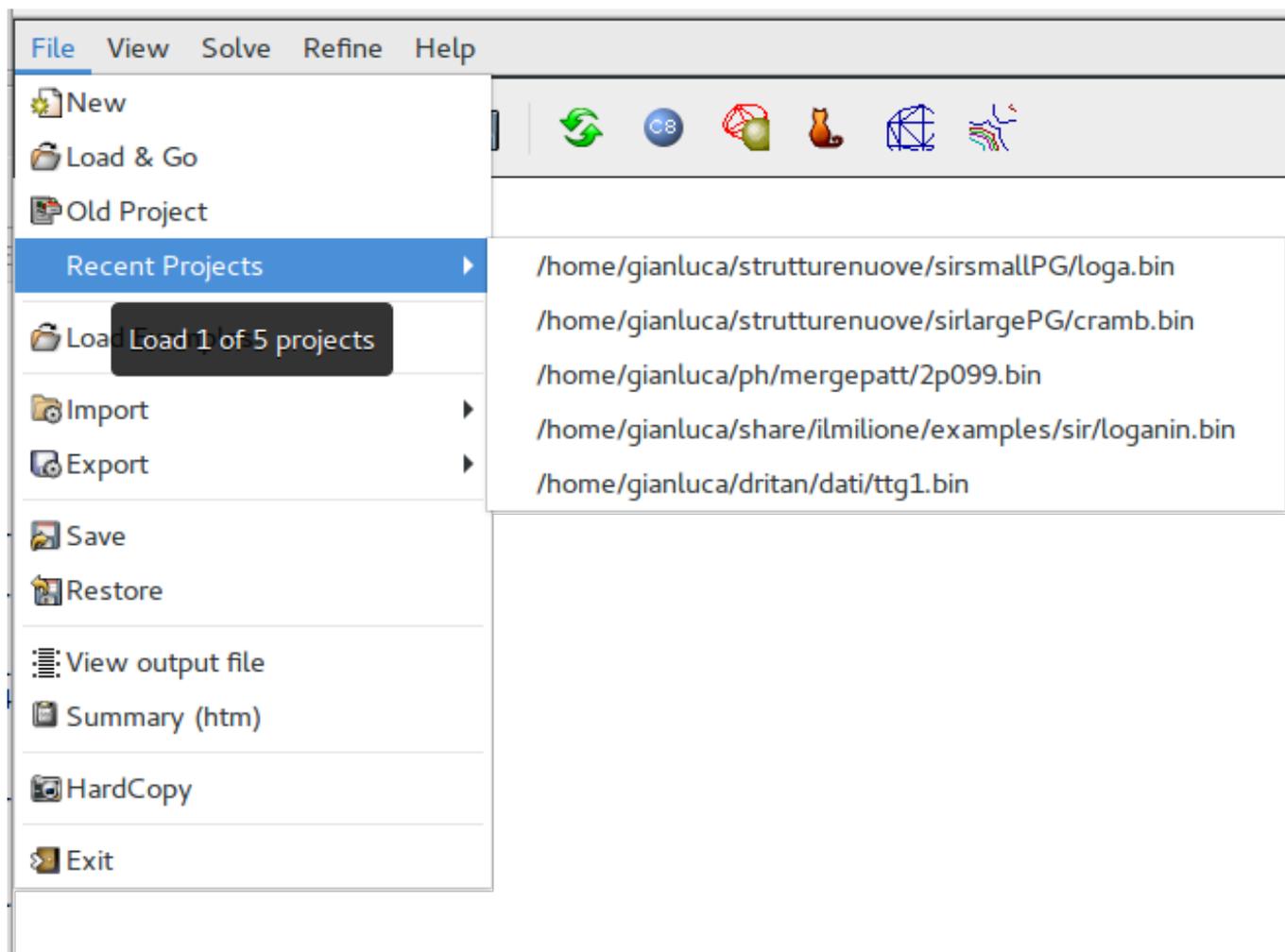
The file selection window will be opened in order to select the input file (.sir). The selected file can be edited using an internal editor.

Old Project

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

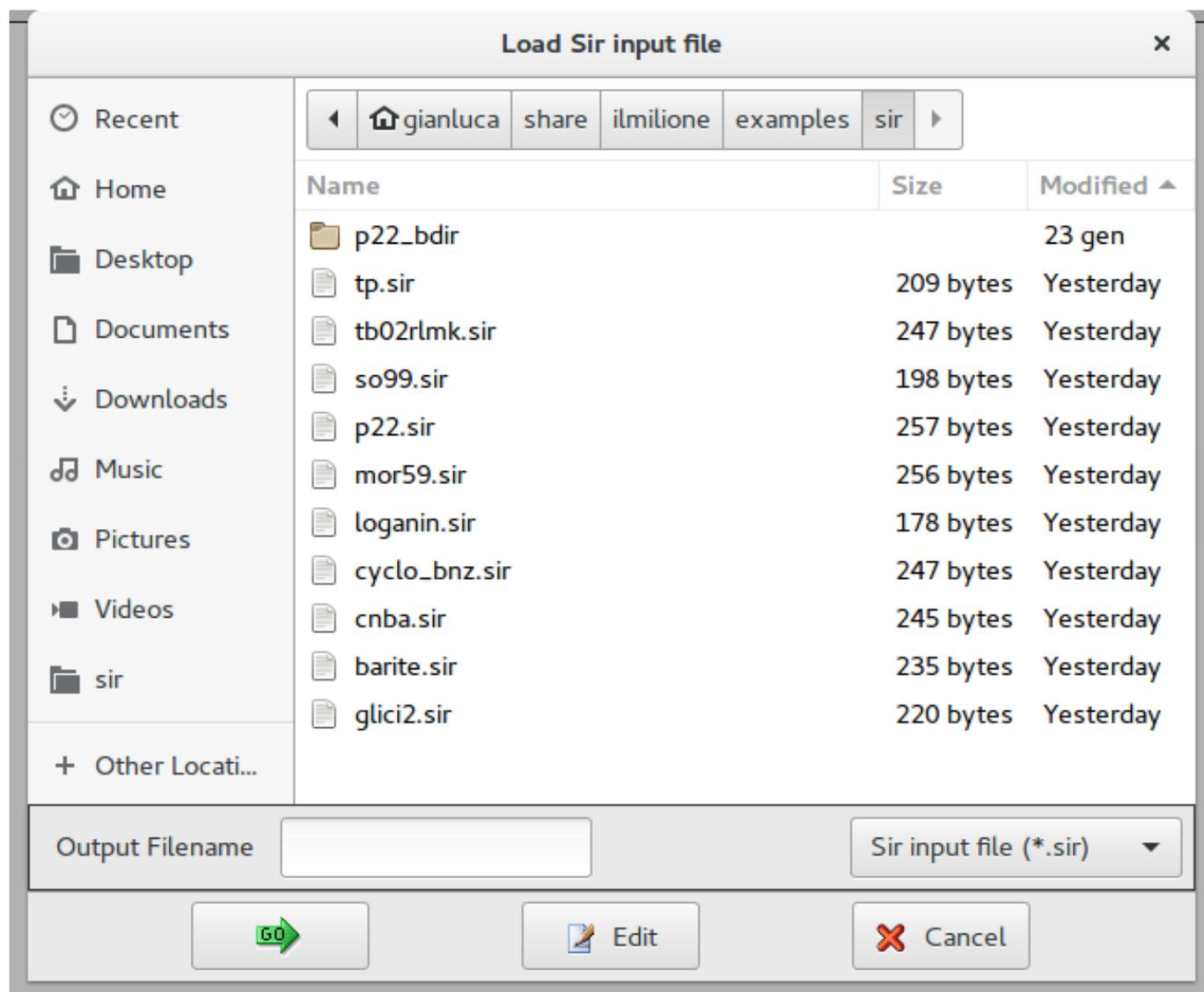
Recent Projects

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 an input file provided as example.



Import

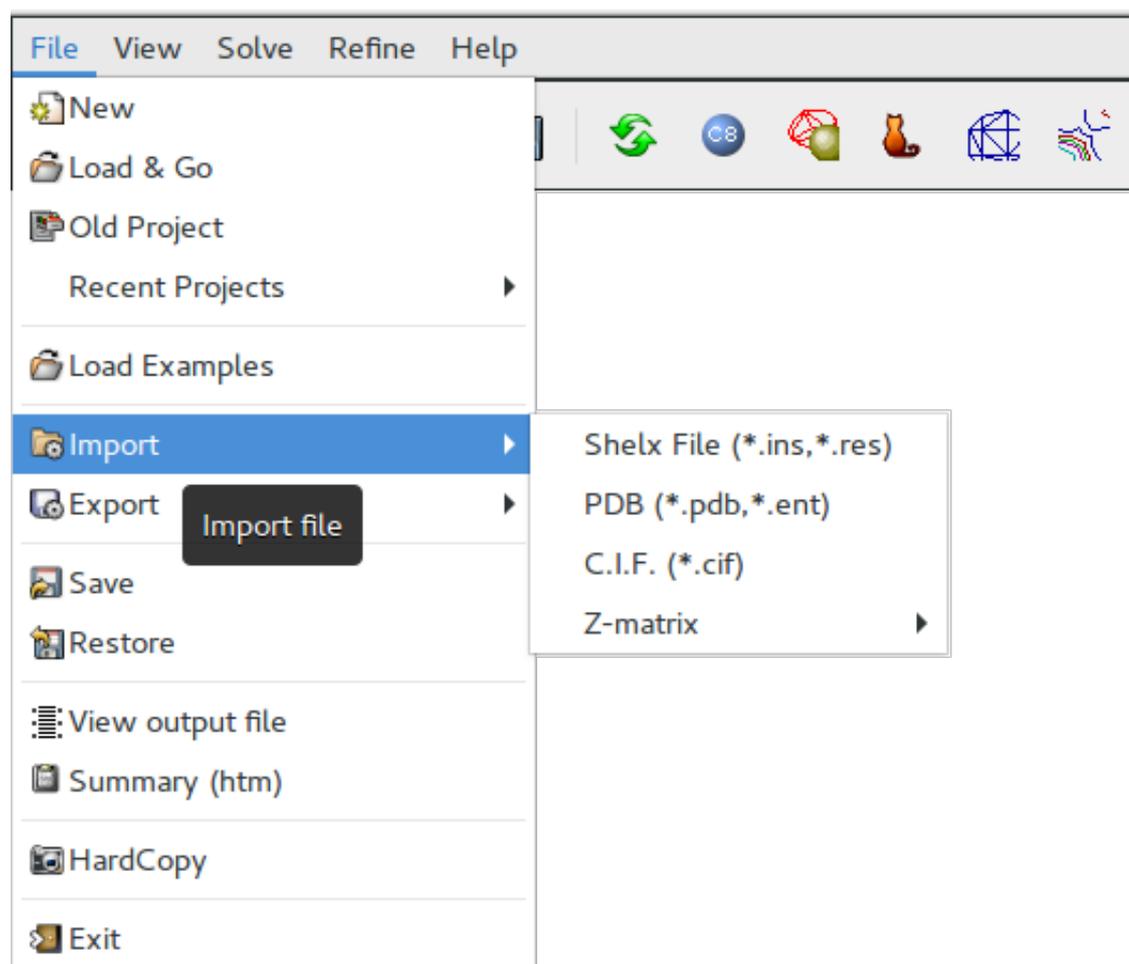
Data from files in different format can be imported:

Shelx file (*.ins)

PDB file (*.pdb)

CIF file (*.cif)

- Z-matrix file (*.zmt or *.mop)



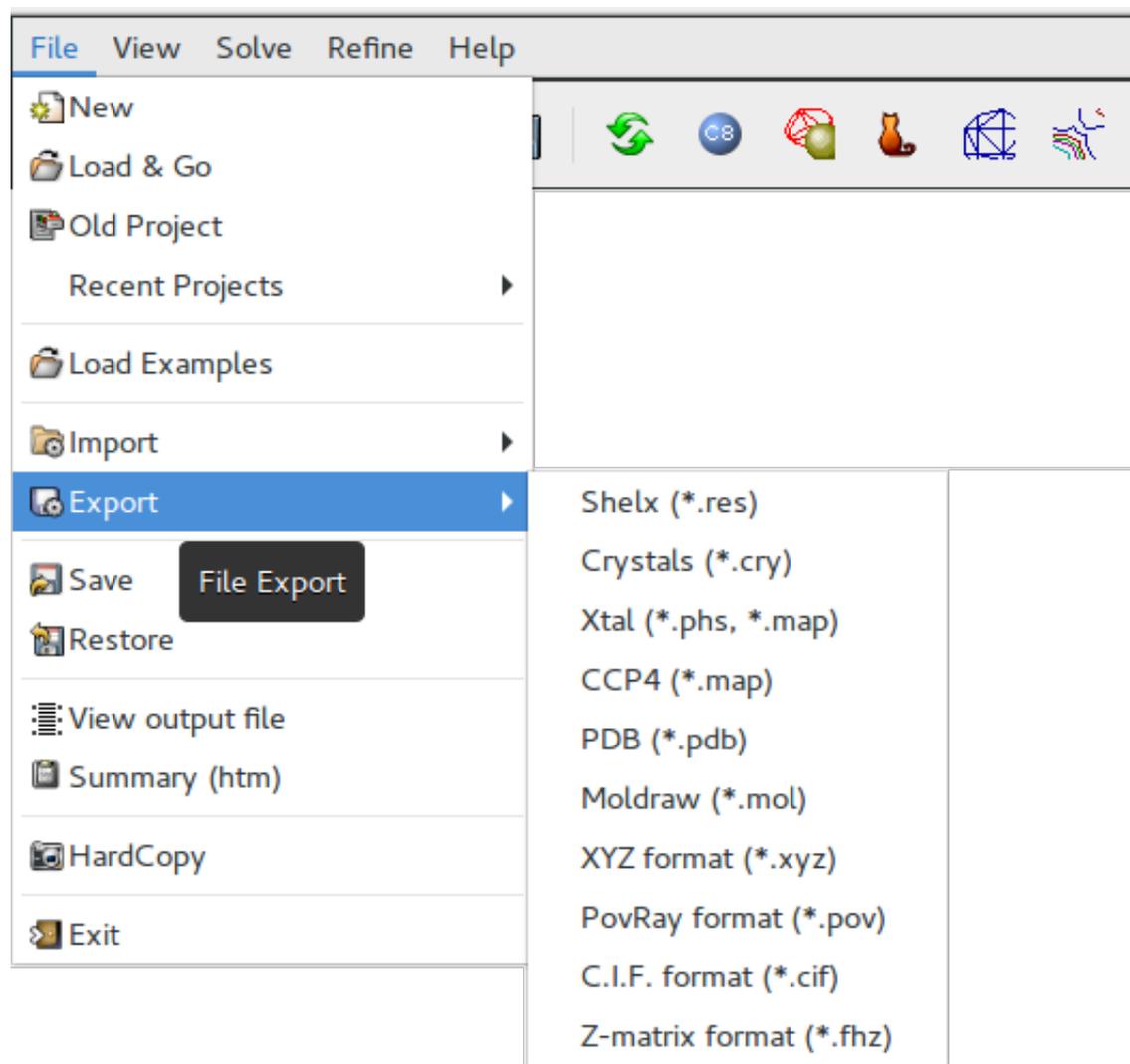
Export

Sir2019 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)
- 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.

Summary

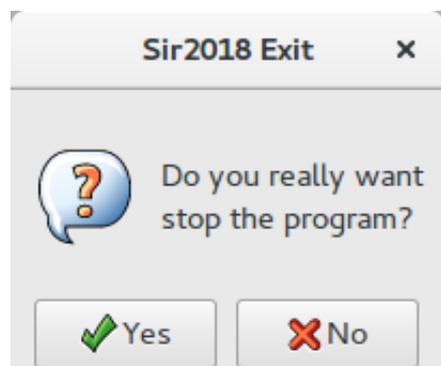
Starts the Internet Browser on a page containing the summary of the results.

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.



[Previous page](#)

[Next page](#)

General information

by admin - Thursday, December 22, 2016

<http://www.ba.ic.cnr.it/softwareic/sir/>



Sir2019, the evolution of Sir2014, is the latest product of the Sir family.

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.

Data

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.

Tasks

Sir2019 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 approaches:**

Modern Direct Methods (MDM)

Standard Direct Methods (SDM)

Vive la Difference (VLD)

Patterson techniques

Default strategies

MDM is the default choice for small (up to 80 non-hydrogen atoms in asymmetric unit) and medium sized 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. MDM is the suggested second choice.

Non *ab initio* phasing approach:

Simulated Annealing (for s/m molecules)

Molecular Replacement
SAD/MAD

Simulated Annealing is applied when some geometrical information about the target molecule is *a priori* available.

Molecular Replacement is applied when a model structure is available: the relative routines perform the rotation and translation steps, automatically followed by phase extension and refinement.

In SAD/MAD cases, the anomalous scatterer substructure is found and refined first. Then such information is actively and automatically used to phase the protein and for final phase extension and refinement.

Automatism

The program has been designed to:

require a minimal input information;

work automatically;

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

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

What's new

New features have been added to Sir2019 with respect to the previous version (Sir2014). From the practical point of view the user may exploit:

a new Molecular Replacement pipeline (from MR application to the automatic model building);
a SAD/MAD pipeline (from SAD/MAD application to the automatic model building);
an *ab initio* Patterson pipeline (from the application of Patterson techniques to automatic model building)

The three pipelines are available because of recent advances in phase extension and refinement (see the description

of SYNERGY, in [Phase Module](#)) and in automatic model building (see the description of CAB in [Phase Module](#))

Jav (Just another viewer)

It is not only a visualizer for models and maps: it can also be a useful tool to edit the model. It can be used as an external program (started from its icon or from terminal) or as an alternative to the classical visualization of the model by Sir2019.

Sir2019 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 & 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. Caliendo(1), B. Carrozzini(1), G.L. Cascarano(1), C. Cuocci(1), C. Giacovazzo(1), M. Mallamo, A. Mazzone(1), G. Polidori(1), D. Siliqi(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:

M.C. Burla, R. Caliendo, B. Carrozzini, G. L. Cascarano, C. Cuocci, C. Giacovazzo, M. Mallamo, A. Mazzone and G. Polidori

“Crystal structure determination and refinement via SIR2014”,
[J. Appl. Cryst. \(2015\). 48, 306–309](#)

[Next page](#)

Getting started (small molecules)

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/getting-started/>

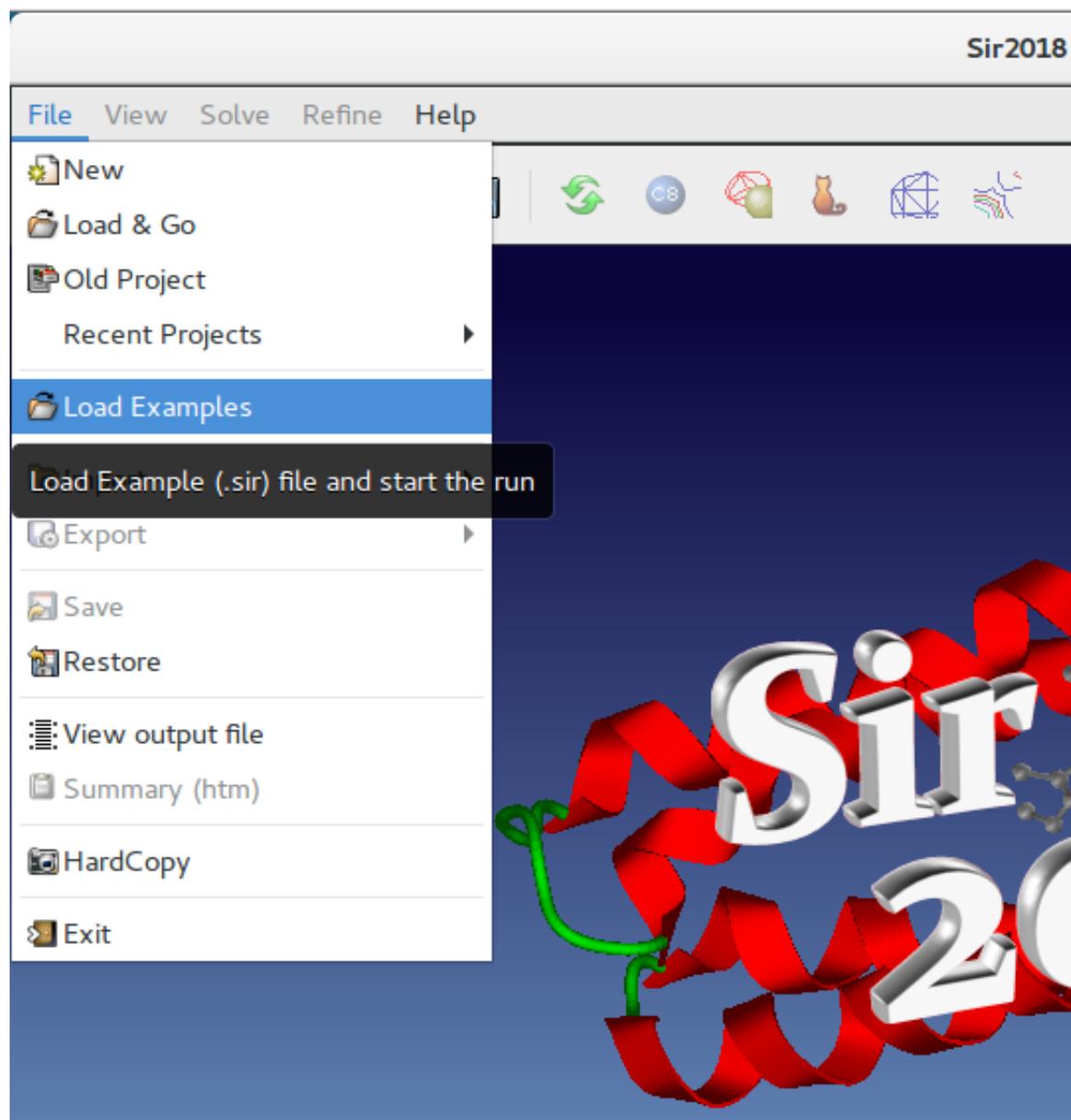
To run the program it is necessary to create an input file with .sir extension using a text editor or the graphical tools (see The Graphic Interface).

Moreover it is now possible to start Sir2019 using a .ins file under the following conditions:

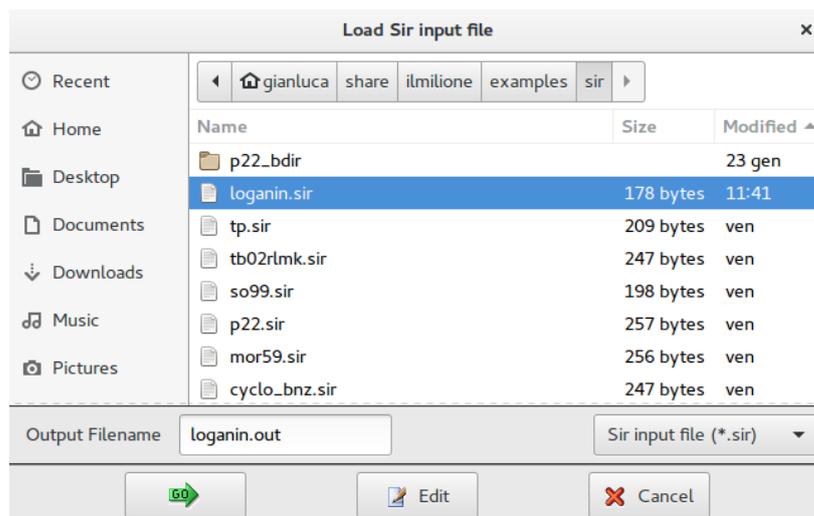
- the .ins file must contain at least the keywords: CELL, UNIT, LATT, SFAC (only atomic species), HKLF;
- in the same directory of the .ins file the .hkl file must exist with the same name.

The corresponding .sir file is automatically created by the program.

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



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



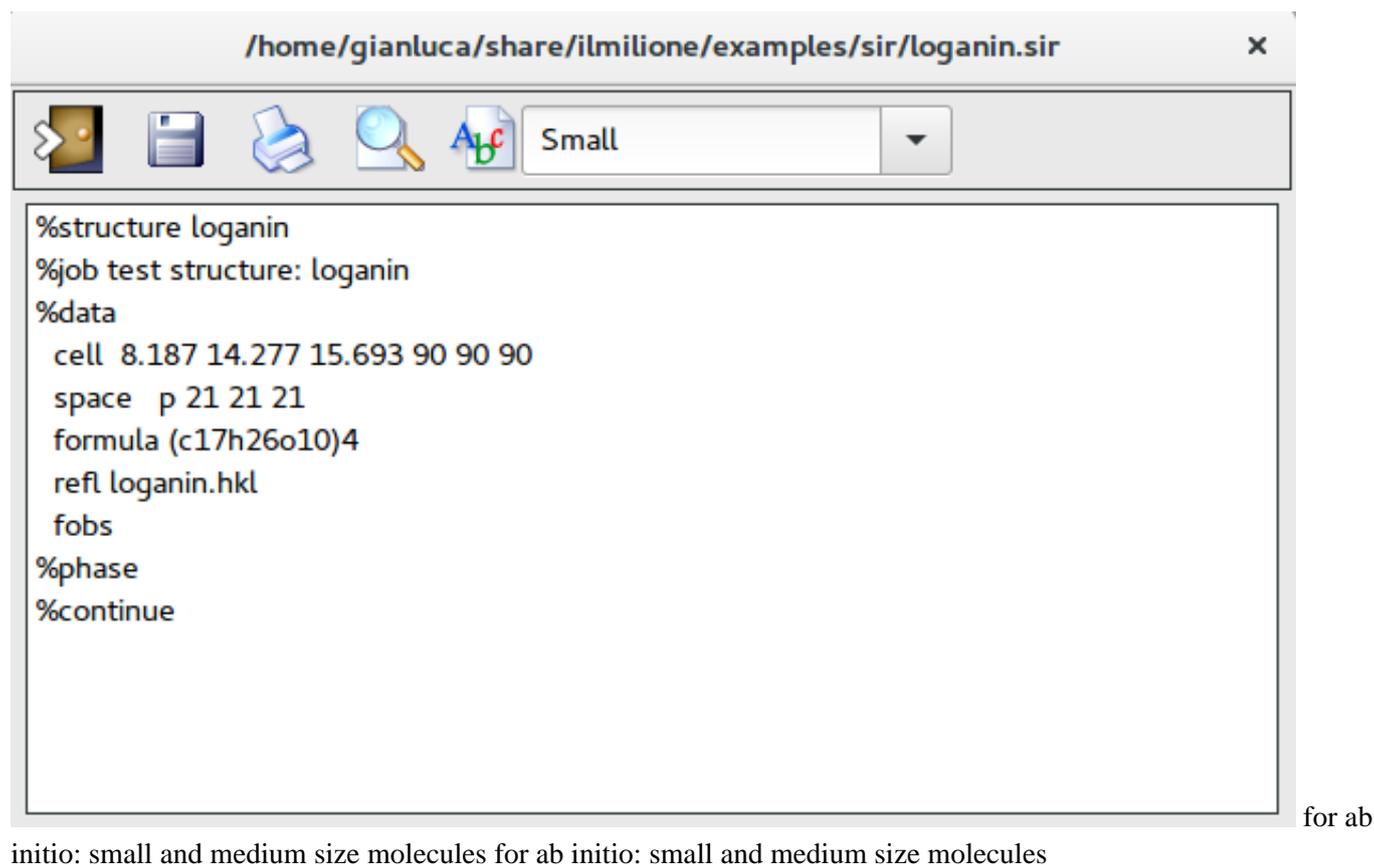
for ab initio: small and medium size molecules

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



button. for ab initio: small and medium size molecules

The file loganin.sir contains the minimal input instructions to make Sir2019 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, ...).



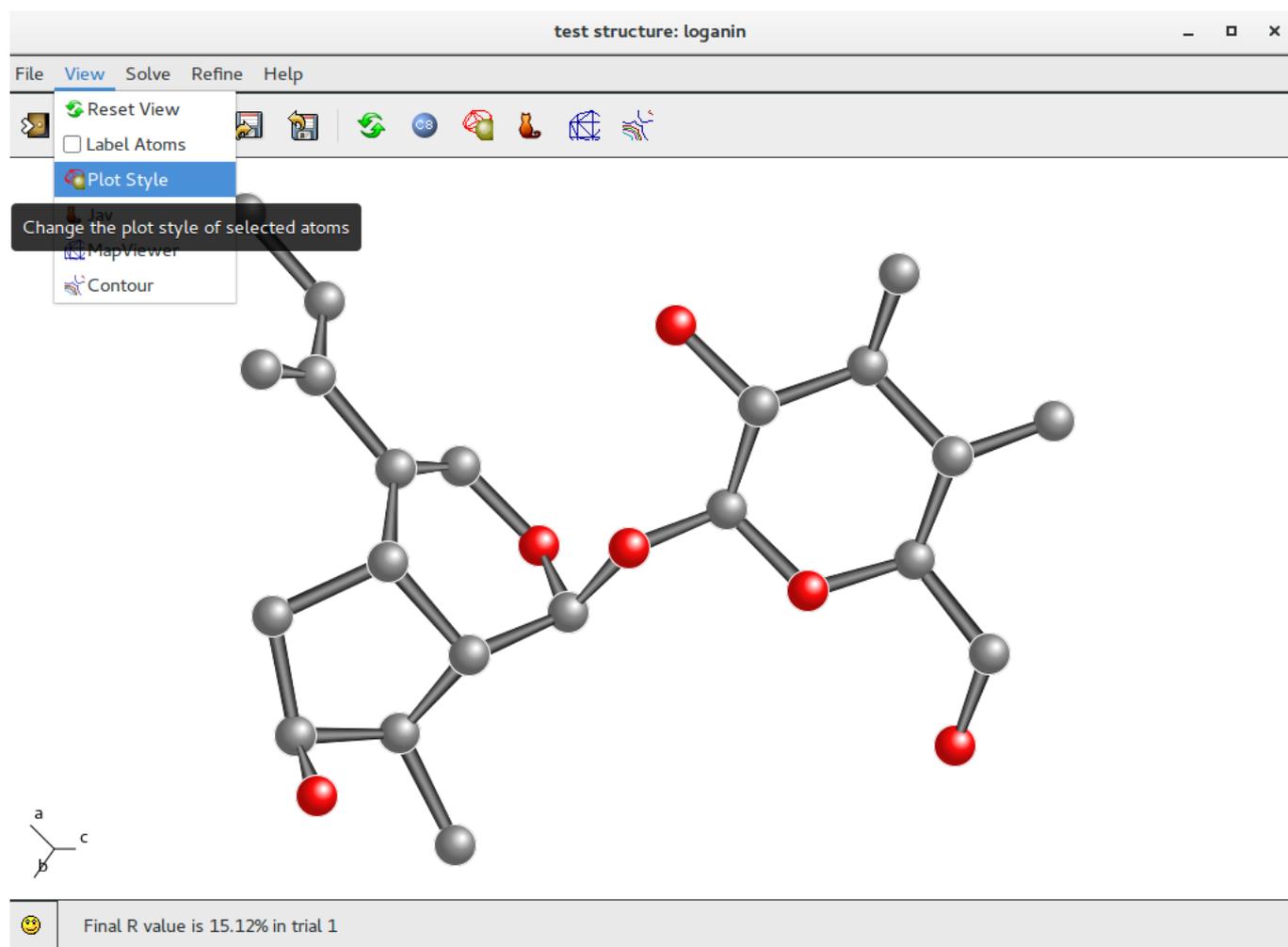
```
%structure loganin
%job test structure: loganin
%data
cell 8.187 14.277 15.693 90 90 90
space p 21 21 21
formula (c17h26o10)4
refl loganin.hkl
fobs
%phase
%continue
```

for ab

initio: small and medium size molecules for ab initio: small and medium size molecules

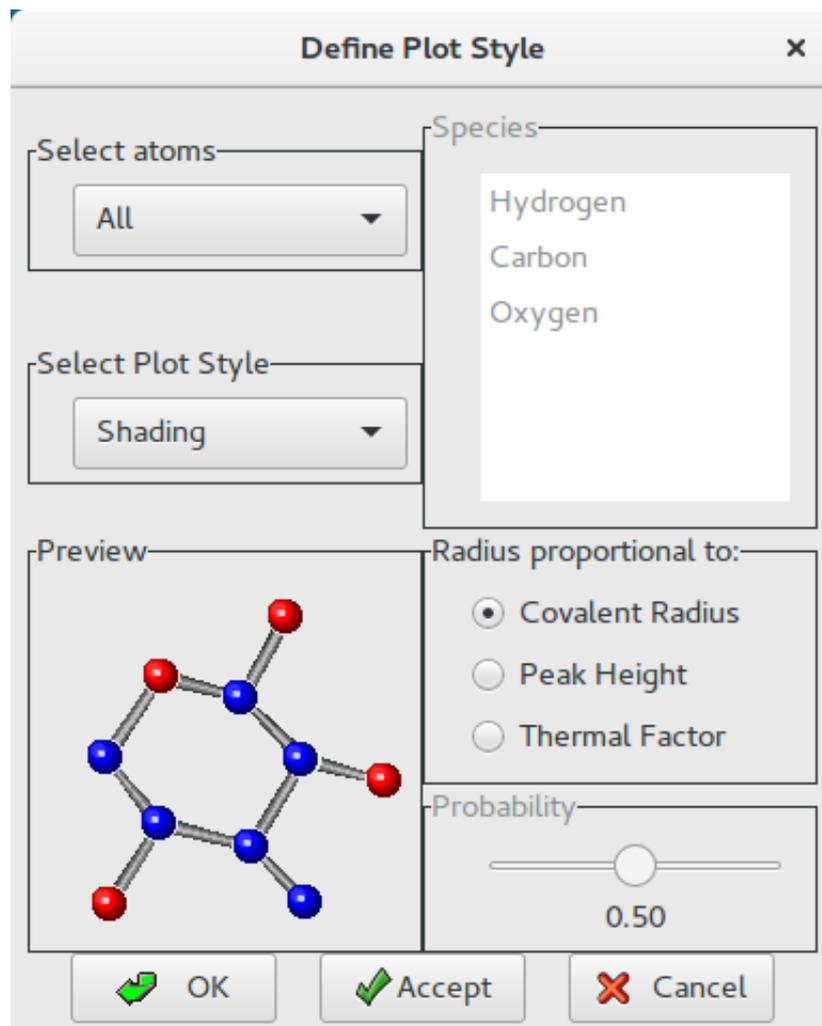
To save changes click on ; leave this window using . To start the program, click on  button.

At the end of the run, Sir2019 shows the “best” model of the structure; the final $R(F)\% = 15.12$ is quoted in the status bar (at the bottom of the program window).

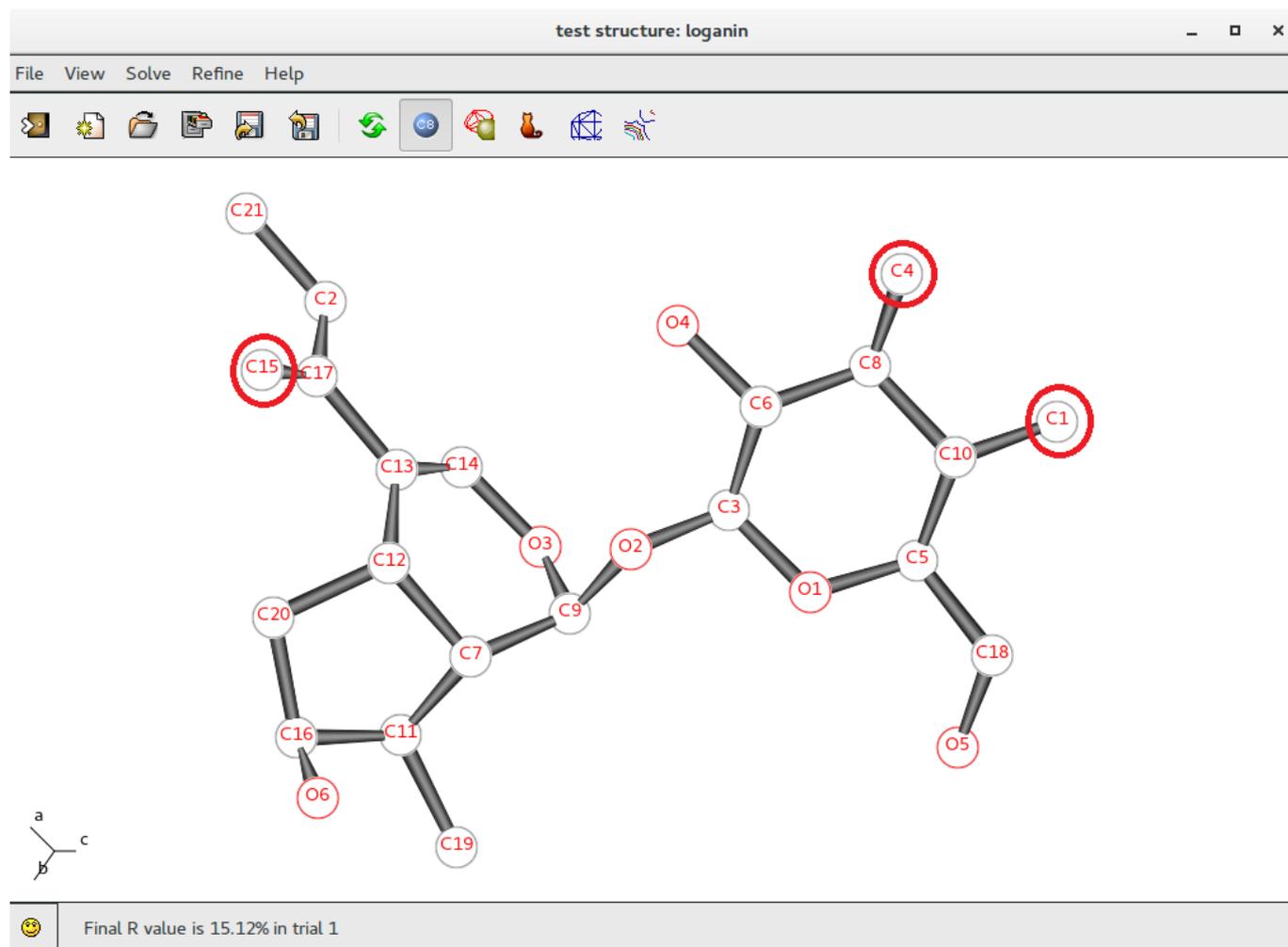


for ab initio: small and medium size molecules for ab initio: small and medium size molecules

To modify the style go to menu item "View" on the main toolbar and select "Plot Style" option.

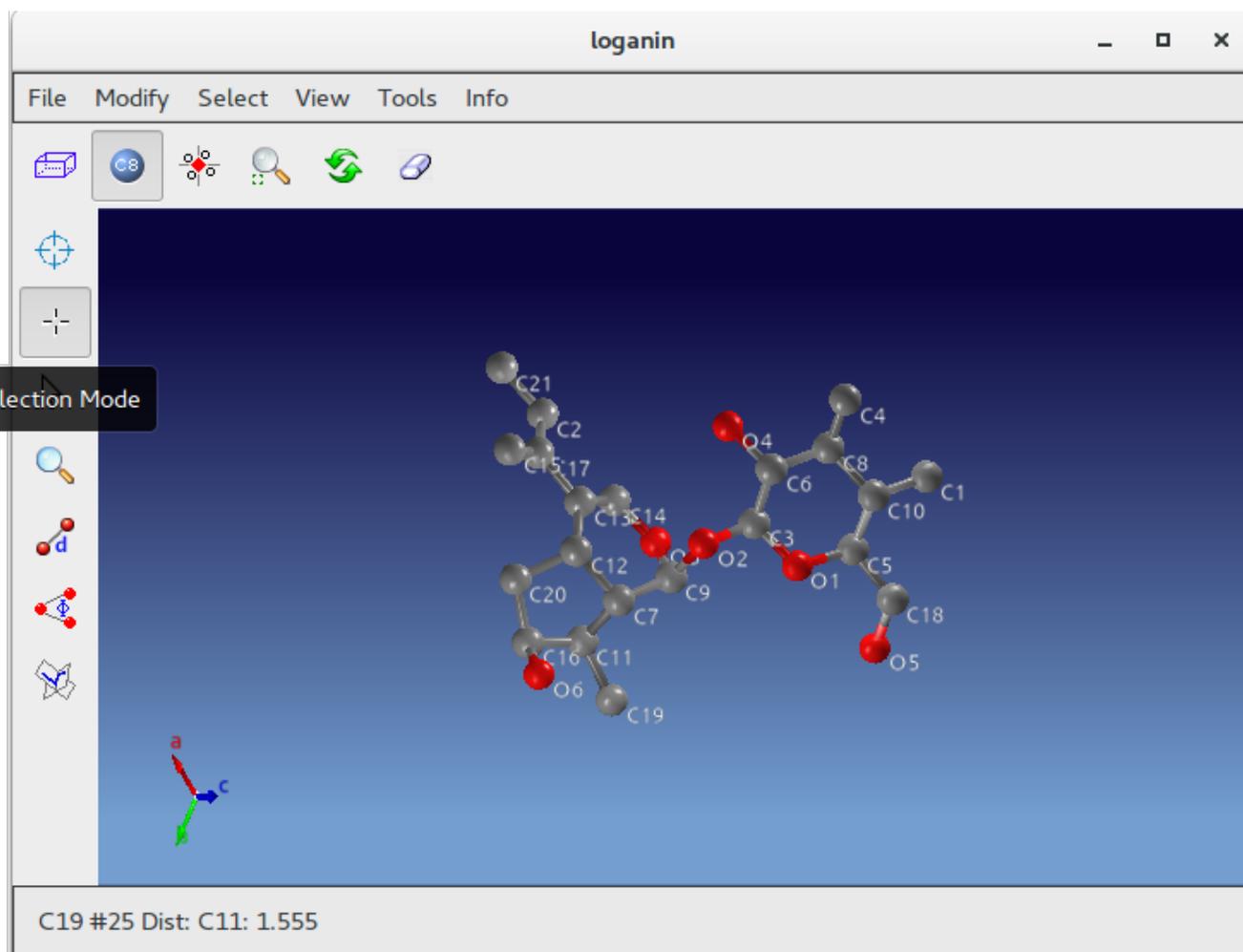


Change the plot style for the model (i.e. Bubbles) and accept it (click on Accept or OK buttons); add the labels (chemical species and progressive numbers) to the atoms of the model by clicking on Labels button :

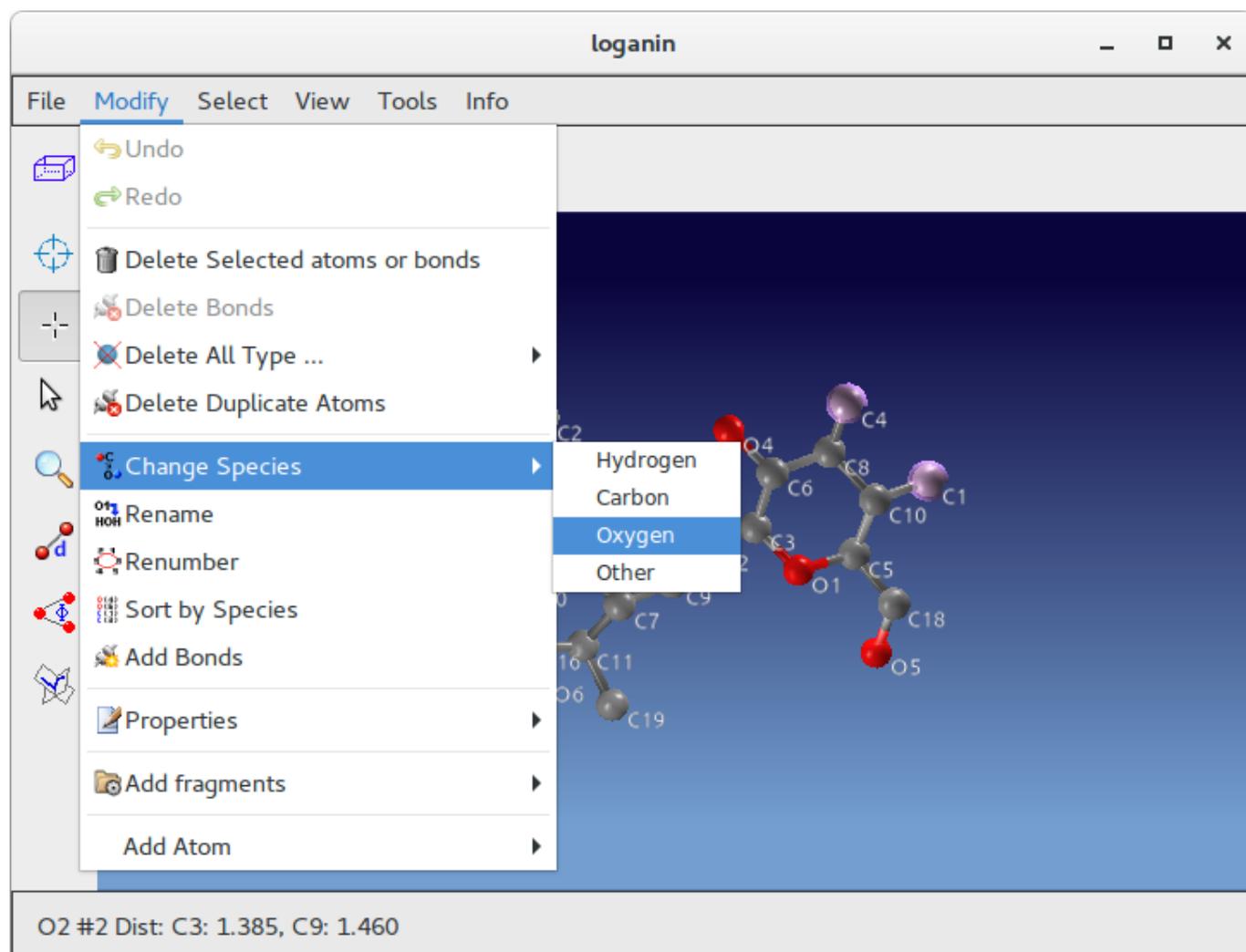


for ab initio: small and medium size molecules

Atoms C1, C4 and C15 (inside red circles) show wrong labelling, they should be oxygens. To change their atomic species start  Jav and click on Selection Mode button:



Using the mouse select C1, C4 and C15; go to menu item “[Modify](#)” on the toolbar; select “[Change Species](#)” and then “[Oxygen](#)”.



Start the refining process: go to menu item "[Refine](#)" on the Sir2019 toolbar and select [LSQ](#) option: the Least Square Refinement interface will be opened.

Least Squares Refinement

General Modify Atoms Parameters

Refinement Request

- Diagonal
- Full Matrix
- S.F.C. only

Refinement Conditions

Damping Factor: 1.0 - +

Reflections cutoff: 3 - +

of Cycles: 3 - +

Weighting Scheme

6 : W=1.0

Modify weight parameters Refine weight parameters

Observations & Parameters

Observations: 2255 (92 %) Parameters: 109 Ratio: 20.69

Help GO OK Cancel

Set the LSQ options (i.e. Full Matrix, sigma(Fo) cutoff, number of LSQ cycles, weighting scheme, ...)

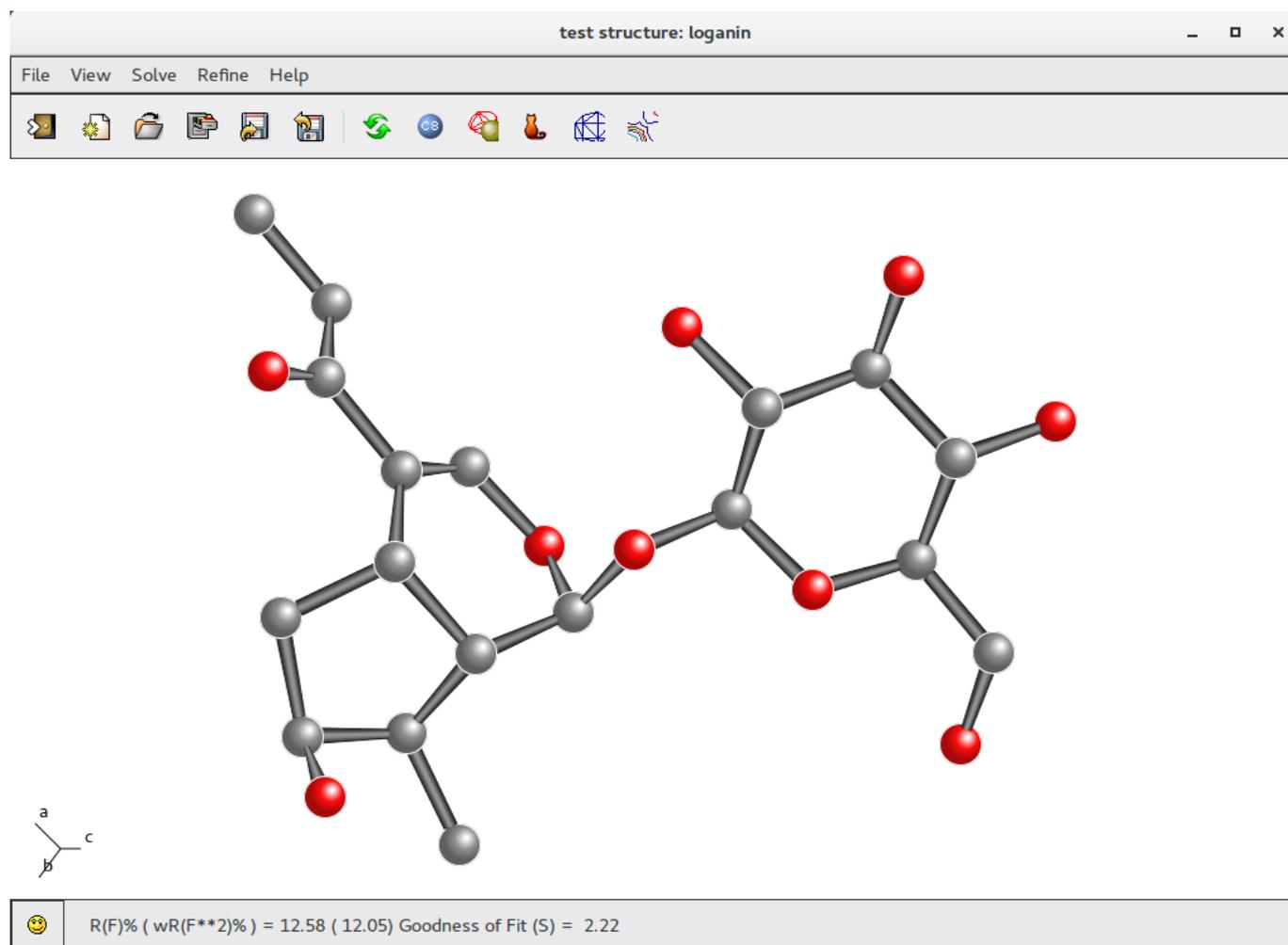
Be careful to the ratio Observation / (number of refined) Parameters. For this example the ratio value (20.69) is high enough to lead to a reliable LSQ refinement. In general, the variance of the final LSQ parameters increases with decreasing values of the ratio: too small values of it can make the refining process meaningless even if the crystallographic residual is small.



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: R(F)% = 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, it has to be applied a thermal anisotropic LSQ refinement. If necessary it is possible to fix some specific atomic parameters.

Restart the LSQ interface and select “[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).  to restart calculations.

Least Squares Refinement ×

General Modify Atoms Parameters

Atomic Parameters

Filter species

Hydrogen Carbon Oxygen All

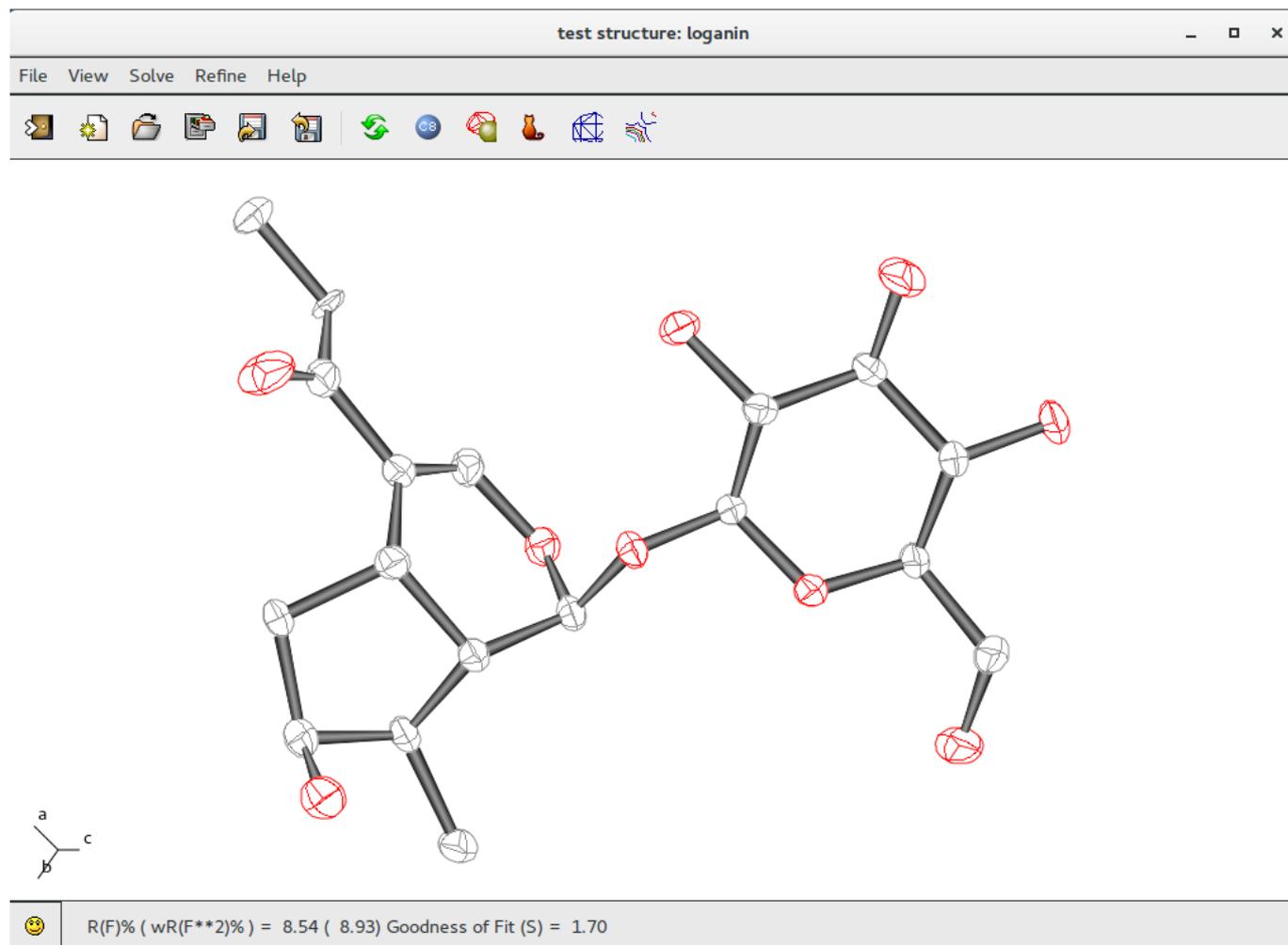
	Fix XYZ	Fix B	Fix Occ.	Set Isotropic
O1	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
O2	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
O3	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Observations & Parameters

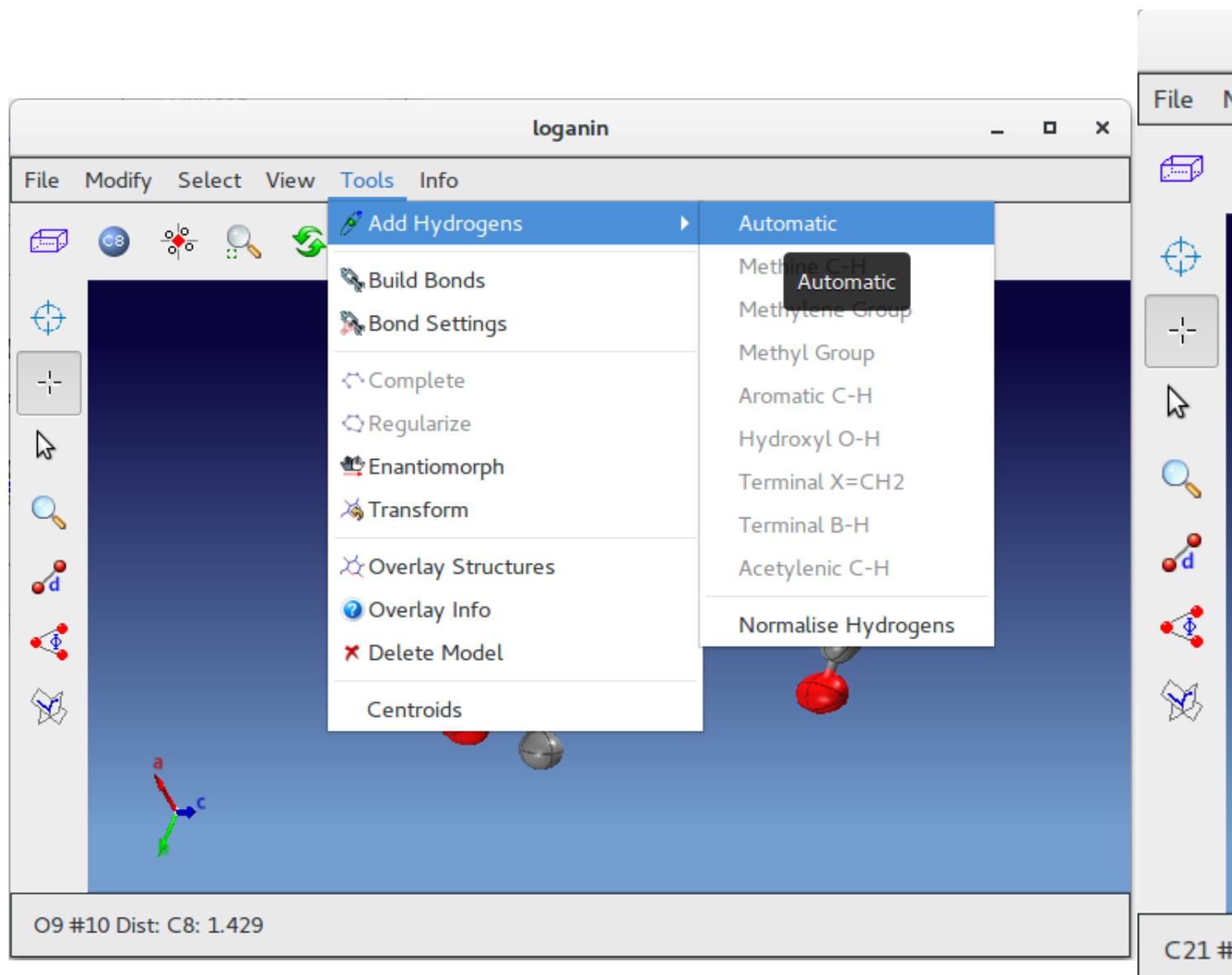
Observations: 2255 (92 %) Parameters: 109 Ratio: 20.69

 Help  GO  OK  Cancel

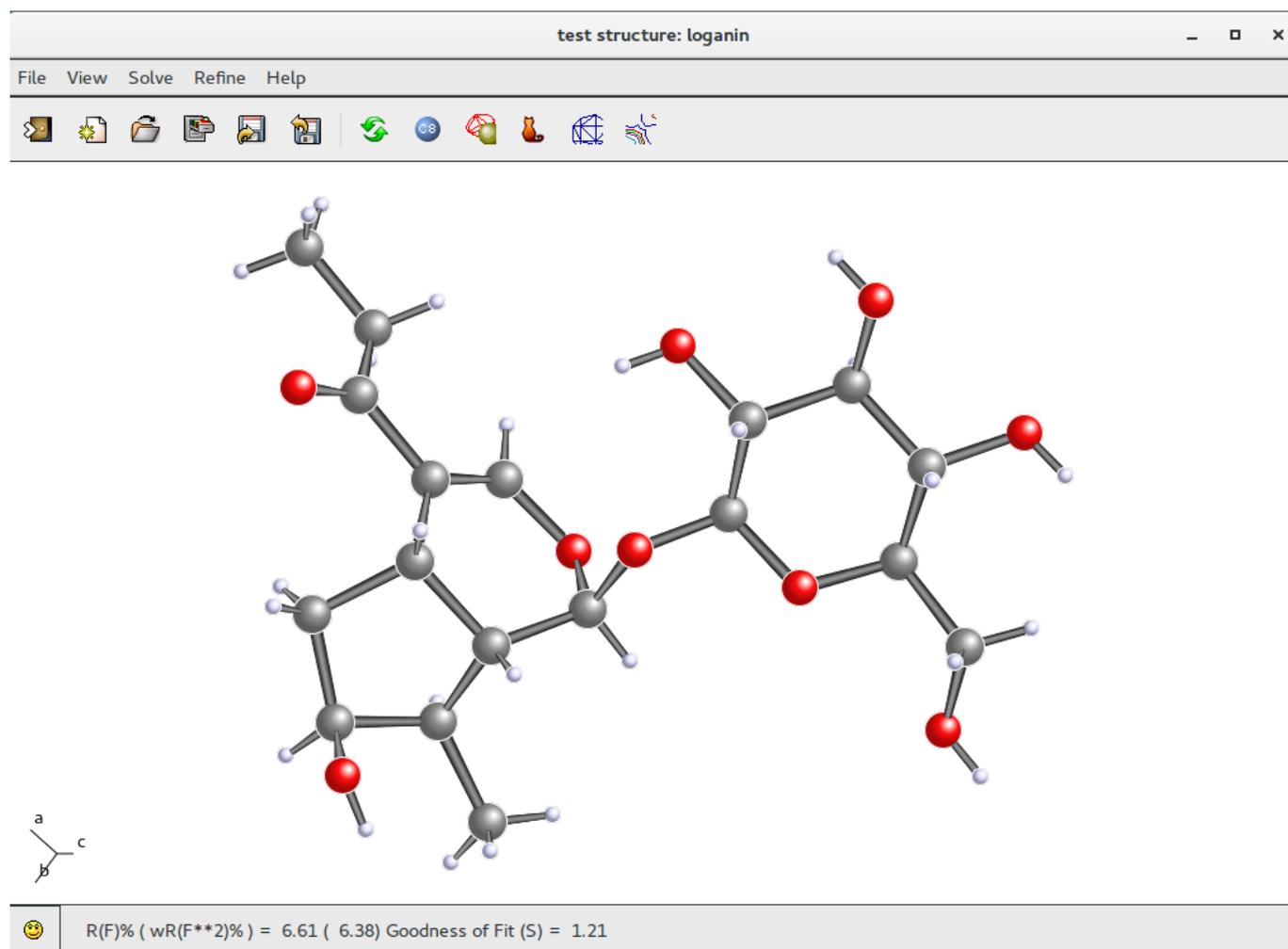
Change the plot style for the model: go to menu item “View” and select “Plot Style” option: select “Ellipsoids” and radius proportional to Thermal Factor (atom thermal ellipsoids will be plotted, proportional to anisotropic thermal parameters). Updated R(F)% = 8.54 can be found in the status bar .



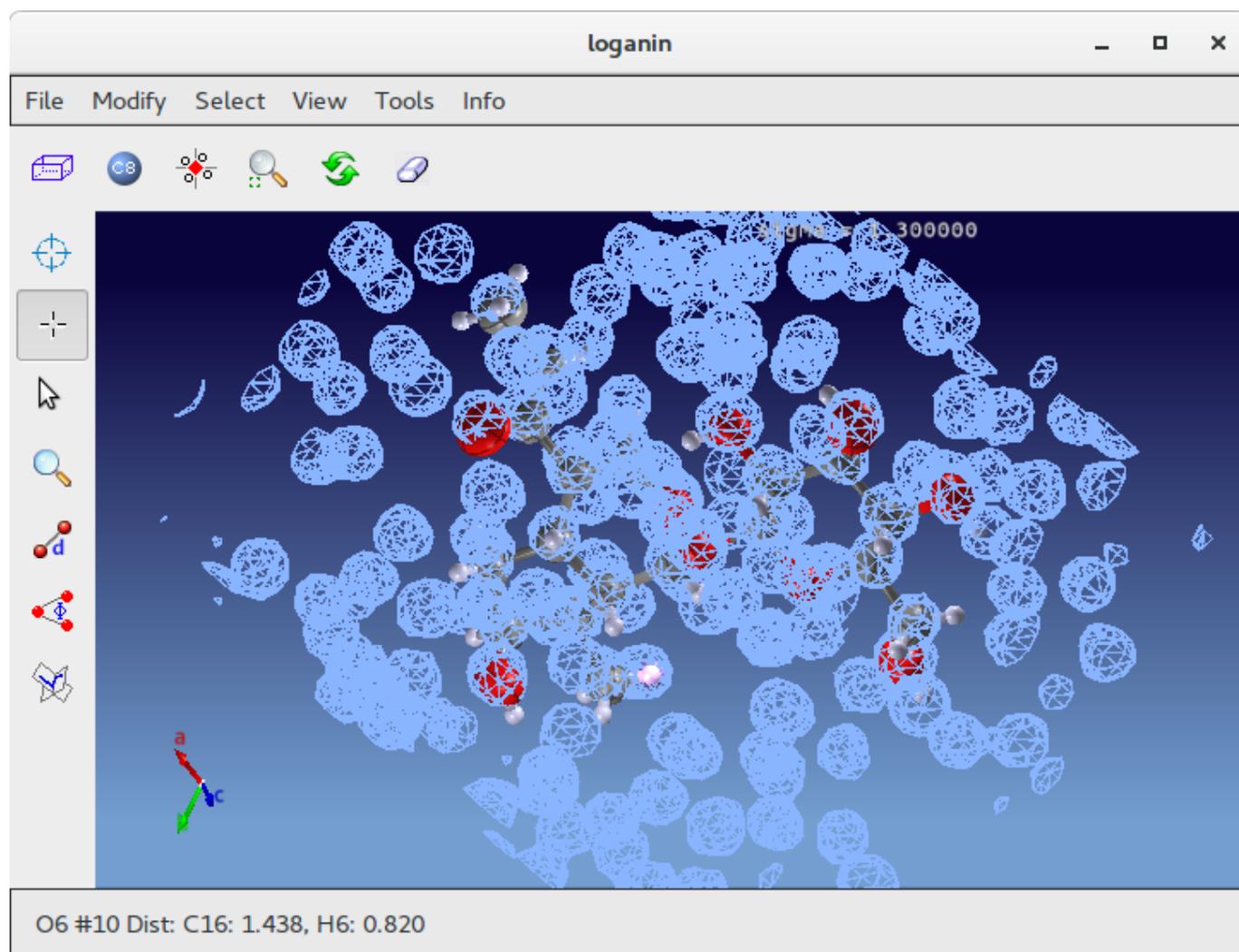
Complete the model by adding hydrogen atoms using the “Tools” menu in  Jav:



The next step is to start a new LSQ refinement. Updated R(F)% value is shown in the status bar. This R(F)% value (6.61) 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 toolbar :



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.

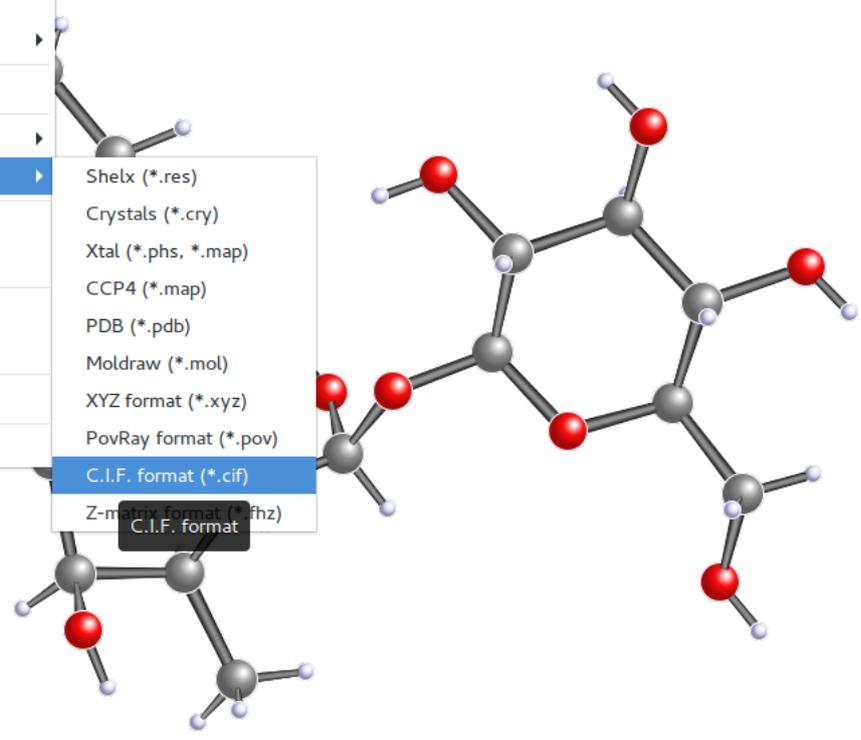
Go to menu item "File" on the toolbar, select "Export" option and then CIF option (other formats are available: i.e. PDB format, for proteins).

test structure: loganin

File View Solve Refine Help

New
Load & Go
Old Project
Recent Projects
Load Examples
Import
Export
Save
Restore
View output file
Summary (htm)
HardCopy
Exit

Shelx (*.res)
Crystals (*.cry)
Xtal (*.phs, *.map)
CCP4 (*.map)
PDB (*.pdb)
Moldraw (*.mol)
XYZ format (*.xyz)
PovRay format (*.pov)
C.I.F. format (*.cif)
Z-matrix format (*.fhz)
C.I.F. format



a
b c

R(F)% (wR(F**2)%) = 6.61 (6.38) Goodness of Fit (S) = 1.21

[Previous page](#)[Next page](#)

Installing the program

by sir - Tuesday, January 23, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/installing-the-program/>

On Windows platforms

Refer to Sir2019Readme.txt file for updated information.

Run Sir2019SetUp-XX.exe

- Run the program by clicking on the Sir2019 (or Jav) icon



?Sir2019 icon



Jav icon

On Linux / Unix / MacOS platforms

Refer to README file for updated information.

- To install Sir2019 you need to execute the following commands:

```
tar xvfz sir2019-XX.tar.gz
```

```
cd sir2019-XX
```

```
./configure --prefix=destination_directory --srcdir=sir2019_source_directory
```

```
make clean
```

```
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 Sir2019 icon or start sir2019 in a terminal window.

Installing dependencies:

1. Install build tools

Most Linux distributions come with the core set of development tools already installed but, if not, you will need to install the following packages *via* the appropriate package manager.

Autoconf Program which writes configure scripts.

automake Program which writes Makefiles.

libtool Generic library support script.

g++ C++ compiler from the GNU compiler collection

Fortran compiler compliant with Fortran 95 standard: Sir2019 has written and tested using the Intel Fortran Compilers (ifort) but, if IFORT is not available, GNU Fortran (gfortran) could be also used.

2. Install graphic development packages

Sir2019 requires development files for gtk+ library, Mesa OpenGL utility library, multiprotocol file transfer library libcurl and FFTW.

The name of the libraries could be slightly different depending on your GNU/Linux distribution. The names of the packages is reported for some tested distribution.

Debian distributions: libgtk2.0-dev (or libgtk-3-dev), libgl1-mesa-dev, (Ubuntu, Mint) libglu1-mesa-dev, libcurl4-gnutls-dev, libfftw3-dev

RedHat distributions: gtk2-devel-..... (or gtk3-devel-.....), (Fedora, CentOS) mesa-libGL-devel-....., mesa-libGLU-devel-....., libcurl-devel-....., fftw-devel
Mandriva: libgtk+2.0-devel (or libgtk+3.0-devel), libmesag11-devel, libmesaglu1-devel, libcurl-devel
Suse: gtk2-devel (or gtk3-devel), mesa-devel, libcurl-devel

Configure process

The configure procedure 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 compiler 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
```

[Previous page](#)

[Next page](#)

Invariants module

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/invariants-module/>

Up to 300000 triplets relating the Nlarge reflections are stored for active use in the Direct Methods 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

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

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 Sir2019, the information contained in the PSI-0 triplets has been actively combined with the P10 relationships and negative quartets, according to Cascarano & Giacovazzo (1995).

[Previous page](#)

[Next page](#)

Molecular Replacement

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/molecular-replacement/>

Basics

- [Molecular Replacement Interface](#)
- [Molecular Replacement directives](#)

Basics

Molecular Replacement (*MR*) is the most popular method for solving protein crystal structures. Sir2019 uses an updated version of 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; for both of them the program automatically decides the working resolution. The program may be schematized in the following three steps:

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

Rotations and translation parameters are refined *via* simplex method. The phase values so obtained are extended *via* a modified version of the pipeline REVAN (Carrozzini *et al.*, 2015), combining REMO09, REFMAC, REFMAC phase driven, *DM*, *VLD*, *FL* and BUCCANEER (Cowtan, 2006), as default choice; if other Automated Model Building (*AMB*) programs are available (Nautilus, ARP/wARP or Phenix) the program can use them in an automatic way in order to supply a refined and interpreted solution.. Models with sequence similarity index below 0.40 are automatically submitted to mutations, to make the model more similar to the target. The only permitted mutation is the trivial truncation of the residues to alanine. As in the standard REVAN version, when data resolution is smaller than 1.25 Å, *LDT* is preferred to *DM* because of its larger effectiveness. The modified model is cyclically resubmitted to *REFMAC* phase driven- *EDM-VLD-FL* cycles.

The *AMB* program is used not only to check if the resulting model may be automatically interpreted, but also as intermediate step for improving phases which are resubmitted to cycle of *EDM-VLD-FL*.

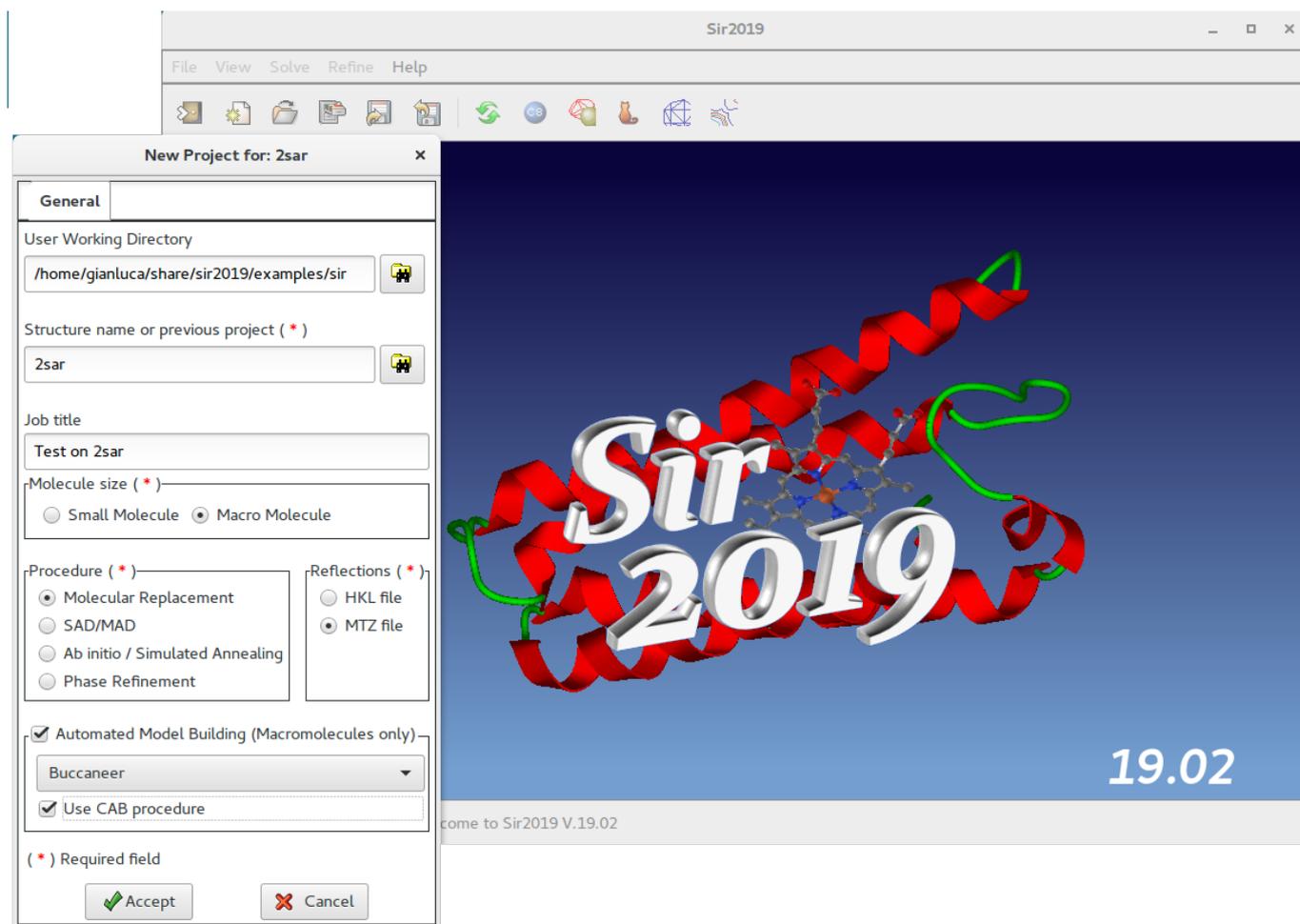
It is possible to create the input file for Sir2019 using a text editor; as an example:

```
%Structure 2sar
%Job Test on MR
%Amb Buccaneer
%Data
  Mtz /home/gianluca/share/ilmilione/examples/sir/2sar.mtz
  Label H K L FP SIGFP
  Sequence /home/gianluca/share/ilmilione/examples/sir/2sar.seq
  nMolecules 1
%Remo
  Improve
  Model
    Fragment /home/gianluca/share/ilmilione/examples/sir/1ucl_mod1_lsq1.pdb
    Monomers 2
    BFactor PDB
  EndModel
%End
```

The list of directives is available at the end of this page or [here](#). The directives for %data in case of proteins, are available [here](#)

Molecular Replacement Interface

A graphic interface is available in Sir2019 to work with Molecular Replacement by means of the *New* item  in the toolbar or in the menu *File* :



In this window the user should specify, among others, the working directory, the molecule size (small or macro), the procedure to use (Molecular Replacement) and the reflection file format (MTZ).

Proteins only: if an Automated Model Building program is installed, it is possible to use it by selecting Buccaneer or Nautilus or ARP/wARP or Phenix. If selected the CAB procedure will be applied at the end of the Molecular Replacement.

Once clicked on 

it is possible to access the **MTZ** section:

New Project for: 2sar ✕

General MTZ

MTZ filename: (*)

Title:

Change Space Group: Check Space Group

Cell Parameters:

Select F, sigF for protein

Selected columns:

H K L

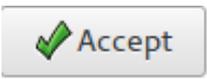
Available columns

Label	Type
FreeR_flag	I
FP	F
SIGFP	Q

Columns

value:	esd:	
<input type="text" value="FP"/>	<input type="text" value="SIGFP"/>	<input type="button" value="Add pair"/>

Using this window it is possible to supply the name of the MTZ file and to select the necessary columns (H K L FP SIGFP).

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

New Project for: 2sar

General MTZ Cell Content Solve

Source

X Ray Data Neutron Data Electron Data

Molecular unit:

Number of molecules in the asymmetric unit: -

Molecular Weight: -

Number of residues: -

Sequence (FASTA format): or paste data in the box above

```
>2SAR:A|PDBID|CHAIN|SEQUENCE
DVSGTVCLSALPPEATDTLNLIASDGPFY
YTGDHYATFSLIDQTC
>2SAR:B|PDBID|CHAIN|SEQUENCE
DVSGTVCLSALPPEATDTLNLIASDGPFY
YTGDHYATFSLIDQTC
```

Ligands - Additional elements and corresponding number in asymmetric unit

<input type="text"/>							
----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

Cell Volume: Volume per Atom:

(*) Required field

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

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

New Project for: 2sar

General MTZ Cell Content **Solve**

Molecular replacement phasing

Number of models: - Use pseudotranslation information (if present)
 Improve MR results

Model n. 1

Model PDB file
  

Number of monomers to search - Thermal Factor:
 PDB
 Proportional to Wilson
 20

Use heteroatoms

Similarity of the PDB model to the target structure

Resolution range
 Automatic Manual Max Res.: Min Res.:

Thresholds selection
 Automatic Rotation Optimization Translation Optimization
 Manual

(*) Required field

It is possible to specify the Sequence identity (otherwise it is computed by the program).

The model can be modified transforming it in a sequence of alanine residues or it can be cutted.

Similarity of the PDB model to the target structure

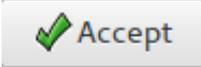
Sequence identity: Computed by the program - +

Modify model

Polyalanine model Cut model: use residues from to

Rotation:

Fixed Eulerian angles (degrees)

Once clicked on 

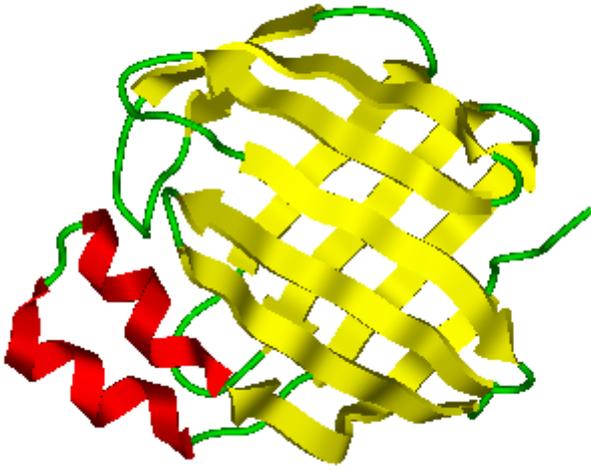
it is possible to access the **Run** section:

x**New Project for: 2sar**

GeneralMTZCell ContentSolveRun

Run

Graphic window required



Input filename (*) Already exists

YES

Output filename (*)

YES

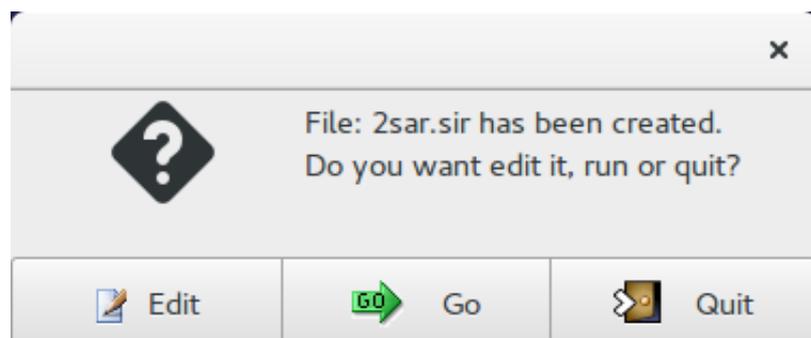
Click on  to save the project

(*) Required field

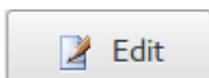
X Cancel

If **Graphic window required** is checked, a graphical output will be produced (see below).

In this section it is possible to change the default names for the input (project) and for the output file. The user, by clicking on  button, can save the input file for Sir2019.



Now it is possible to edit



the input file (2sar.sir); to start calculations click on



Once the program is over, by means of the “View output file”  feature in “File” it is possible to access the complete output file.

By means of the “Summary”  feature in “File” menu is possible to get a synthetic output and also the graphic windows produced by the program during the run (in HTML format). An example of graphical output follows.

First part: application of Molecular Replacement procedure

Thu Feb 8 2018 at 17:56:32 Molecular Replacement procedure starts on: 2sar

Input conditions

Model from 1ucl_mod1_lsq1.pdb
2 molecules will be searched
Sequence identity: 0.979
Used resolution range [6.00,3.13] Å

Rotation Stage using 5.91 degrees increments

Explored trials: 19025
Selected 20 trials for clustering (threshold: 0.32)
Marked trials after clustering: 5

Selected 4 trials for translation (threshold 0.92)

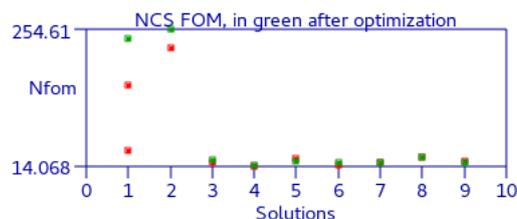
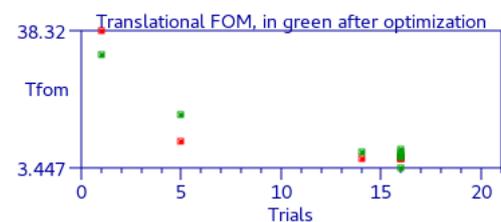
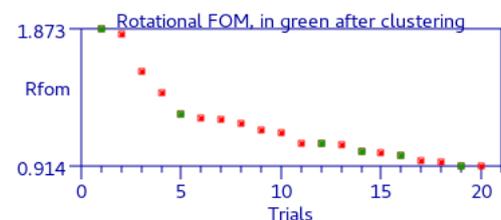
Translation Stage using threshold 0.10

Final selected solutions: 9 (threshold 1.00)

NCS stage: search for 2 molecules in solutions [1-9]

Solution 9 - NFom > TFom: favourable indication adding monomer 2
Best solution: #2, Corr = 0.687, R% = 43.39, Clashes% = 0.00
MTZ and PDB files produced after Molecular Replacement

2sar_remo_1.mtz 2sar_remo_1.pdb



Second part:

improving MR results, protein phase refinement and CAB procedure application using Buccaneer.

MR test structure 2sar

File View Solve Refine Help



Fri Feb 22 2019 at 10:11:21 Phantom Derivative procedure starts on: 2sar

Improving MR results
Molecular Replacement results will be improved using
VLD, Free Lunch, Refmac
R = 0.2154 (Rfree = 0.2521) after 45 Refmac cycles

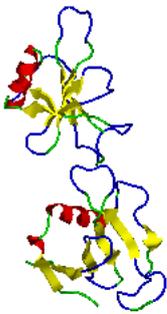
Protein phase refinement and automatic model building
Edm, Vld, Free Lunch and Phantom Derivative techniques
are used to extend and refine phases.
Phased reflections 17024 over 17301 from MTZ file
Running PhD ... Cycle 2/ 2 derivative 7/ 7

CAB procedure using Buccaneer
R = 0.2489 (Rfree = 0.2892) after 2 CAB cycles

Final CAB results:
R = 0.2449 (Rfree = 0.2948) Coverage = 100%

Files produced at the end of the run:
MTZ format PDB format
2sar_buccaneer.mtz 2sar_buccaneer.pdb

Best model preview

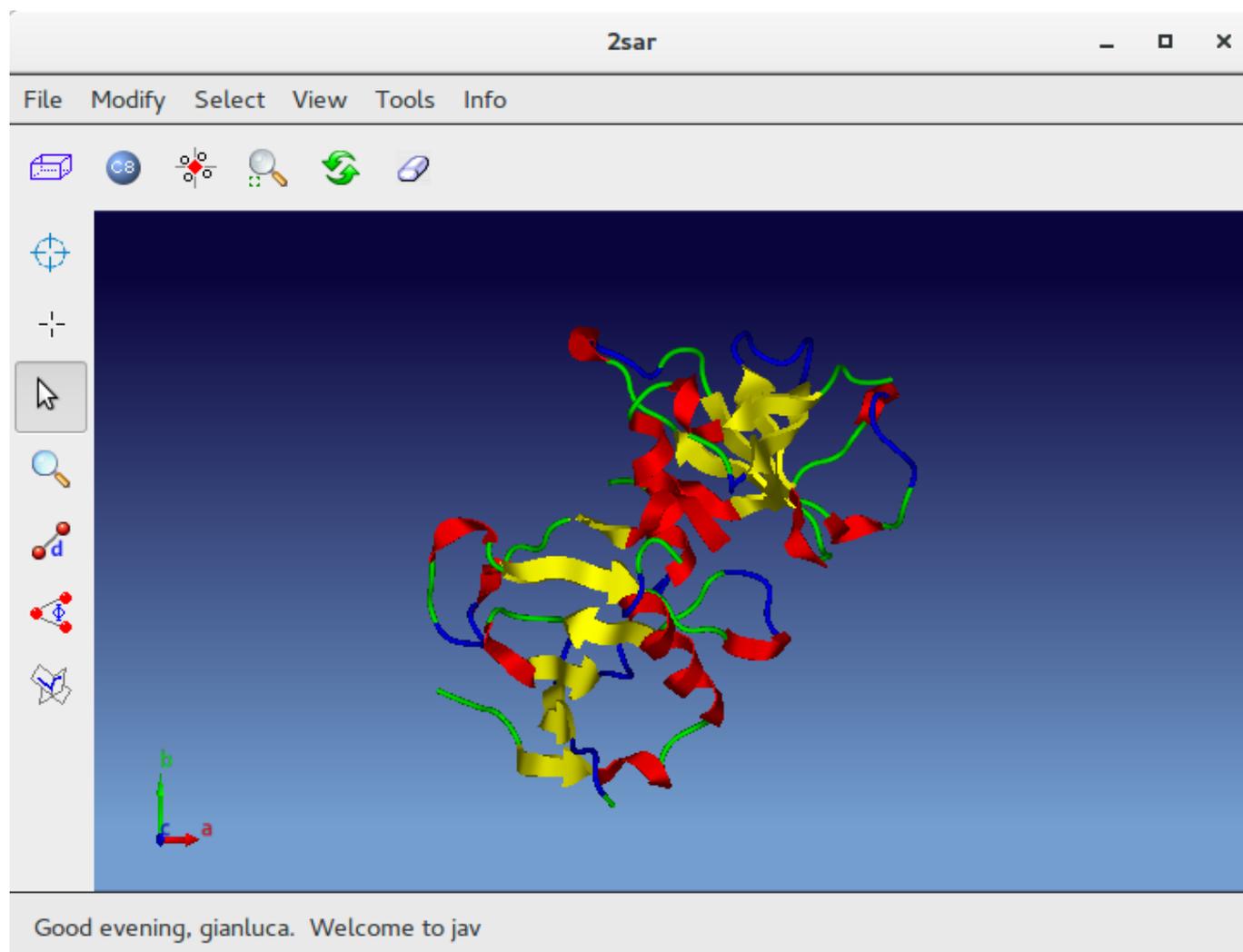


Sir2019 is overCPU Time 0h 10m 56.04s 11 MinTot



When the program is over it is possible to interact with the structure using 

Jav:



Molecular Replacement directives

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). HETATM 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 n1 n2

The program removes from the model some terminal residues if they lie 'outside' n1 and n2 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 threshold values are set automatically by the program and are shown on the graphic window and in the output file.

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 r1 r2 r3 t1 t2 t3

The rotation (r1 , r2 , r3 eulerian angles) and the translation (t1 , t2 , t3 fractional values) are applied to the model.

ROTSELECTED r1 r2 r3

The rotation (r1 , r2 , r3 eulerian angles) is applied to the model; the procedure continues normally.

TRASELECTED t1 t2 t3

The translation (t1 , t2 , t3 fractional values) is applied to the model; the procedure continues normally.

[Previous page](#)

[Next page](#)

Notes on distribution

by sir - Tuesday, January 23, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/notes-on-distribution/>

Depending on the platform required (MS Windows, Linux, Mac) different files are supplied, Jav (Just another viewer) is always included:

- Windows: the program files are distributed as exe file Sir2019SetUp-XX.exe;
- Unix / Linux / Mac: the program files are distributed as gzipped file sir2019-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/>

[Previous page](#) [Next page](#)

Phase module

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/phase-module/>

As previously established, Sir2019 allows *ab initio* and non *ab initio* phasing, followed by phase extension and refinement. Tools for such purposes are the most popular phasing techniques like:

Direct Methods, *VLD*, Patterson techniques for *ab initio* phasing;

MR, *SAD/MAD*, techniques for non *ab initio* phasing;

- Phase extension and refinement methods.

The last category is coupled to all the remaining categories and will be described first.

Phase extension and refinement methods

EDM techniques

Their basic procedure may be described in three steps:

Phases and weights, coupled with observed amplitudes, are used to calculate an electron density map.

The Fourier inversion of the modified map should lead to new and more accurate target phase estimates and to their corresponding weights.

For high resolution data, Sir2019 adopted the FFTW routines by Frigo & Johnson (2005) coupled with a simple low density modification algorithm (Giacovazzo & Siliqi, 1997) similar to that suggested by Shiono & Woolfson (1992). Only a very small percentage (from 2.5% to 10%) of the pixels, those with the largest density values, are used for the Fourier inversion of the electron density map, the rest is set to zero (*LDT* modification). For lower resolution data, Sir2019 prefers to use the *DM* routine by Cowtan (1998), except when the *FL* (*Free Lunch*, see below) techniques are used.

The *LDT* technique above described is applied to small molecules for extending and refining phases obtained by Direct Methods or by Patterson Techniques. More complete models are so obtained which may then be submitted to least squares.

For proteins, owing to the larger size of the molecules and to the limited data resolution, the above technique is not sufficient, and a more complex strategy, including a modified *EDM* procedure, is necessary.

Indeed, during density modification cycles, the molecular envelope of the protein (Wang, 1985; Leslie, 1987) is used as a mask in order to improve *EDM* efficiency (Burla, Carrozzini *et al.*, 2003a). 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.

1.

RELAX procedure

The problem of well oriented but misplaced molecular fragments may be solved by the *RELAX* procedure (Burla, Carrozzini *et al.*, 2002a; 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 default *ab initio* runs of Sir2019.

Free lunch (FL)

Any modified electron density map allows the extrapolation of non-observed structure factors, beyond and behind the experimental resolution limit (Caliandro *et al.*, 2005 a,b). The method is applied for two different purposes:

- i) for increasing the rate of success of the *ab initio* phasing procedures during the phasing process, *via* the active use of the extrapolated reflections (Patterson techniques only);
- ii) for improving the best electron density achievable by current phasing techniques. The combined active use of observed and extrapolated structure factors may lead to improved electron density maps and therefore, after their modification and Fourier inversion, to better phase estimates.

FL is automatically applied when data resolution is worse than 1.25 Å.

Vive la difference (VLD)

This algorithm, coupled with *EDM* techniques, may be used for *ab initio* phasing as well as for phase extension and refinement: it is based only on properties of a new difference electron density and on the observed Fourier synthesis.

In *ab initio* phasing (crystal structure solution of small and medium size molecules, as well as of proteins with atomic resolution data) 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 above), and applied to proteins at atomic resolution.

VLD may be applied, in combination with *EDM* techniques, for non *ab initio* phasing (Burla, Carrozzini *et al.* 2011, 2012; Carrozzini *et al.*, 2013a; Caliandro *et al.*, 2014) to perform phase refinement of initial models arising from *ab initio* Patterson techniques or from *MR* and *SAD/MAD*.

Synergy – Phantom Derivative (*PhD*)

This approach (Giacovazzo, 2015) is applied for non *ab initio* approaches. It involves the creation of a large number of structures (called *ancils*), having random atomic positions, the same unit cell and the same space group of the target structure. The founding *PhD* conjecture was the following: random structures like the ancils, even if completely uncorrelated with a given target structure, may usefully contribute to refine its phases. The conjecture has been experimentally proved by Burla, Carrozzini *et al.* (2015a) when *PhD* was applied to electron density maps provided by *MR* techniques, and theoretically justified by Burla, Cascarano *et al.* (2017).

A simple perspective for understanding the method is the following.

From n ancils n derivatives may be created according to

$$\rho_d(j) = \rho + \rho_a(j), \quad j = 1, 2, \dots, n$$

where ρ is the target density, $\rho_a(j)$ is the j th ancil density, and $\rho_d(j)$ is the electron density of the j th derivative (it is called *phantom* since it is devoid of chemical meaning).

If a model density map ρ_p is available, the $\rho_d(j)$'s may be approximated by n derivative density estimates $\rho_p + \rho_a(j)$, $j = 1, 2, \dots, n$, which, submitted to *EDM* techniques, may provide more sound derivative densities $\rho_{\text{dest}}(j)$. Then the simple *sum function*

$$\rho_{\text{dsum}} = \sum_{j=1}^n \rho_{\text{dest}}(j) = \sum_{j=1}^n \rho_{\text{est}}(j) + \sum_{j=1}^n \rho_a(j)$$

may provide a better estimate of ρ because the sum of the $\rho_{\text{est}}(j)$ will emphasize the target features and the sum of the $\rho_a(j)$ will contribute to the background.

Later on (Carrozzini *et al.*, 2016) the concept of ancil was further on extended: an ancil may also be the same target structure shifted by a permissible translation. The advantage of this choice is that amplitudes and phases of the derivative (target + shifted target) are related *via* simple trigonometric relations to amplitudes and phases of the target structure:

$$|F_{\text{dh}}| = 2|F_{\text{h}}|\cos(\delta t), \quad \delta_{\text{dh}} = \delta_{\text{h}} + \delta t.$$

The use of such new type of ancil is fully competitive with the results obtained when the ancils are randomly fixed and is much faster: therefore it is the default choice of Sir2019.

Phase driven refinement of a molecular model (*PDRF*)

Current crystallographic least squares minimize the quantity

$$\sum (|F_{\text{h}}| - |F_{\text{ph}}|)^2 = \min$$

If some previous information on δ_{h} is available, then the more general expression

$$\sum ||F_{\text{h}}| \exp(i\varphi_{\text{h}}) - |F_{\text{ph}}| \exp(i\varphi_{\text{ph}})|^2 = \sum |F_{\text{qh}}|^2 = \min$$

which minimizes the $|F_{\text{qh}}|$ amplitudes, and is referred in literature (Arnold & Rossmann, 1988) as *vector refinement*.

Its doubles the number of observational equations, because it minimizes functions of amplitudes and of phases. Vector refinement has been implemented in *REFMAC* (Murshudov *et al.*, 1997) *via* maximum likelihood techniques.

If the available molecular model structure is improved by *EDM-VLD-FL-PhD* techniques, then the application of *PDRF* would lead to a model automatically fitting the best available electron density, without passing through the model building step which, if the model is still inaccurate, is expected to fail. *PDRF* may be cyclically applied for improving the efficiency of the phase refinement after the *MR* step.

PDRF has been successfully introduced in the Sir2019 pipeline for *MR* (Carrozzini *et al.*, 2015).

CAB: a cyclic automatic model-building procedure

BUCCANEER (Cowtan K., 2006) has been included in a cyclic procedure (CAB) aiming at increasing its rate of success and the quality of the provided molecular models without modifying its basic algorithms. The model phases provided by the first application of *Buccaneer* are modified in order to allow next *Buccaneer* applications more deeply exploiting the information provided by the phases originally used as input for CAB. The success of the experimental tests suggested to extend CAB to nucleic acids: the procedure allows the use of *PHENIX.AUTOBUILD* (Terwilliger *et al.*, 2008), *NAUTILUS* (Cowtan K., 1994) and *ARP/wARP* (Perrakis *et al.*, 1999), again without modifying their basic algorithms.

STOP to Phase Refinement procedure

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) calculated *via* a diagonal least squares procedure: 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 without the use of supplementary information on the molecular geometry (here not considered). To recognize the correct solution a suitable figure of merit (fFOM) has been devised and applied at the end of the *DSR* process. The correct solution should be identified by large values of fFOM. Further details are quoted in Caliandro *et al.* (2014).

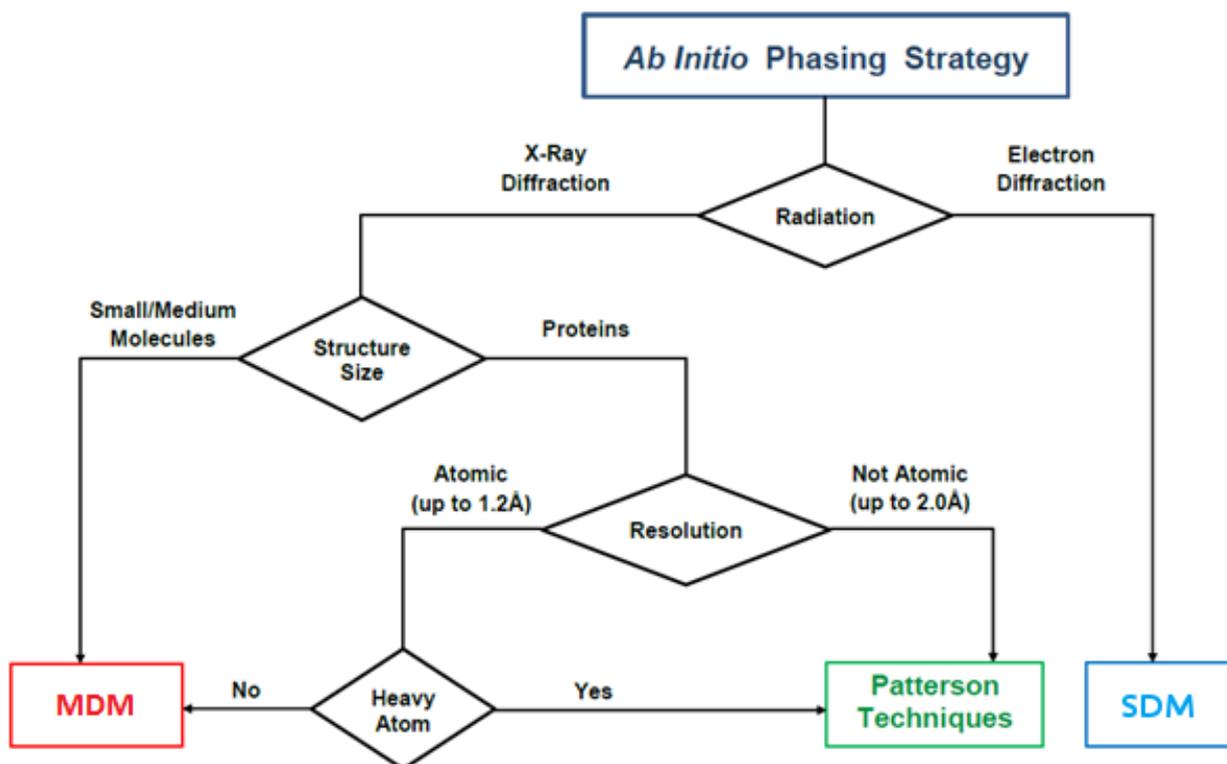
***Ab initio* Phasing**

According to circumstances, the *ab initio* Sir2019 phasing process may apply Modern Direct Methods (*MDM*), Standard Direct Methods (*SDM*), *VLD* or Patterson procedures. 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.

The Sir2019 flow diagram, shown below, is a useful guide for understanding the program strategy for *ab*

initio structure solution.



Modern Direct Methods (*MDM*) procedure

The approach is of multisolution type (Burla, Carrozzini *et al.*, 2015b, 2017). It starts from random phases and applies a tangent like formula simultaneously using:

1. the P10 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;
2. the PSI-0 triplets, which in other programs are passively employed as a figure of merit for recognizing the correct solution;
3. negative quartet invariants *via* their first representation.

The phases obtained *via* the tangent step are then submitted to the *DSR* process, *via EDM* cycles and *RELAX* procedure.

The approach is sequential. Each trial starting set is submitted first 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 Sir2019 the *MDM* procedure is the default choice for structures up to 300 non-hydrogen atoms in asymmetric unit.

Standard Direct Methods (*SDM*) procedure

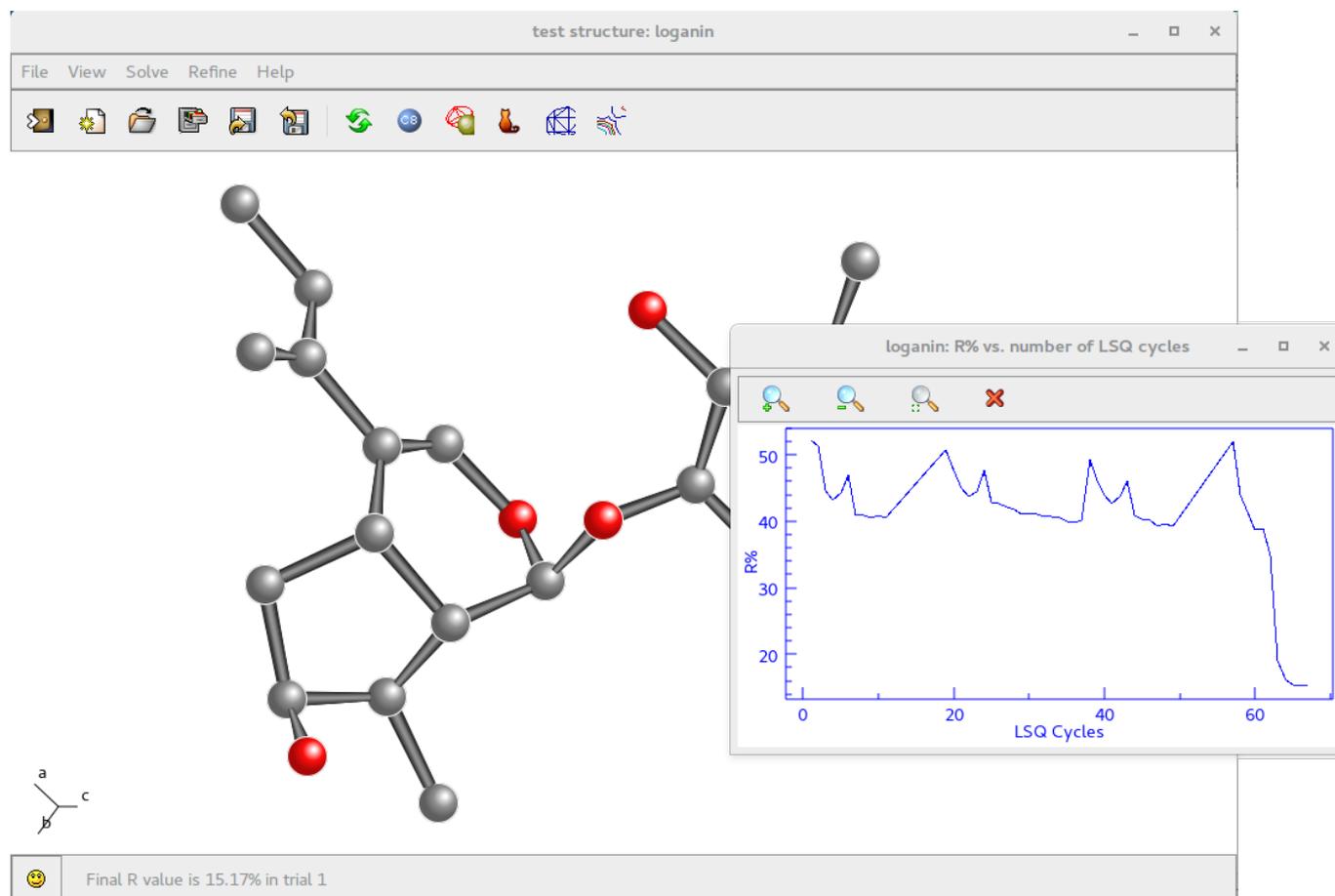
Even in this case a multisolution approach is used; the phasing procedure involves, for each trial, the application of the tangent step to *N* large reflections (those selected by the normalization process), starting from a subset of random phases. The P10 formula is applied and the most reliable negative quartets are actively used during the phasing process.

An early figure of merit (eFOM) is computed for each tangent trial: eFOM (Burla, Carrozzini *et al.* 2003a, Burla, Caliandro *et al.* 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

As previously described in the section Phase extension and refinement methods, random phases are assigned to the target structure factors *F*. Then *n* *EDM* cycles are alternated with a *DEDM* (difference electron density) calculation and modification to recover new estimates of the model structure factors. The procedure is cyclic until a suitable figure of merit stops the calculations.

As an example the screenshot of the crystal structure solution obtained using *VLD* is reported. The *R_f* vs *VLD* cycle trend is shown in the small window.



Patterson deconvolution procedure

Sir2019 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, Caliandro *et al.* (2004a; 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*, *VLD* and *FL* techniques (included 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 Sir2019. An Automated Model Building procedure (*via* Buccaneer, Nautilus, ARP/wARP or Phenix) of the final electron density map is now part of the procedure.

Identifying the correct enantiomorph

When MR techniques are used the correct enantiomorph of the protein is automatically fixed by the model. In *ab initio* and SAD/MAD cases the final phases and their enantiomorph one at the time are automatically submitted to Buccaneer: alpha-helices are used to identify the correct enantiomorph.

[Previous page](#)

[Next page](#)

Phase refinement

by admin - Wednesday, February 20, 2019

<http://www.ba.ic.cnr.it/softwareic/sir/refine-2/>

-

Basics

The command %refine is used to improve phases produced by a previous run of Sir2019 or by a different program.

It is possible to create the input file for Sir2019 using a text editor; as an example:

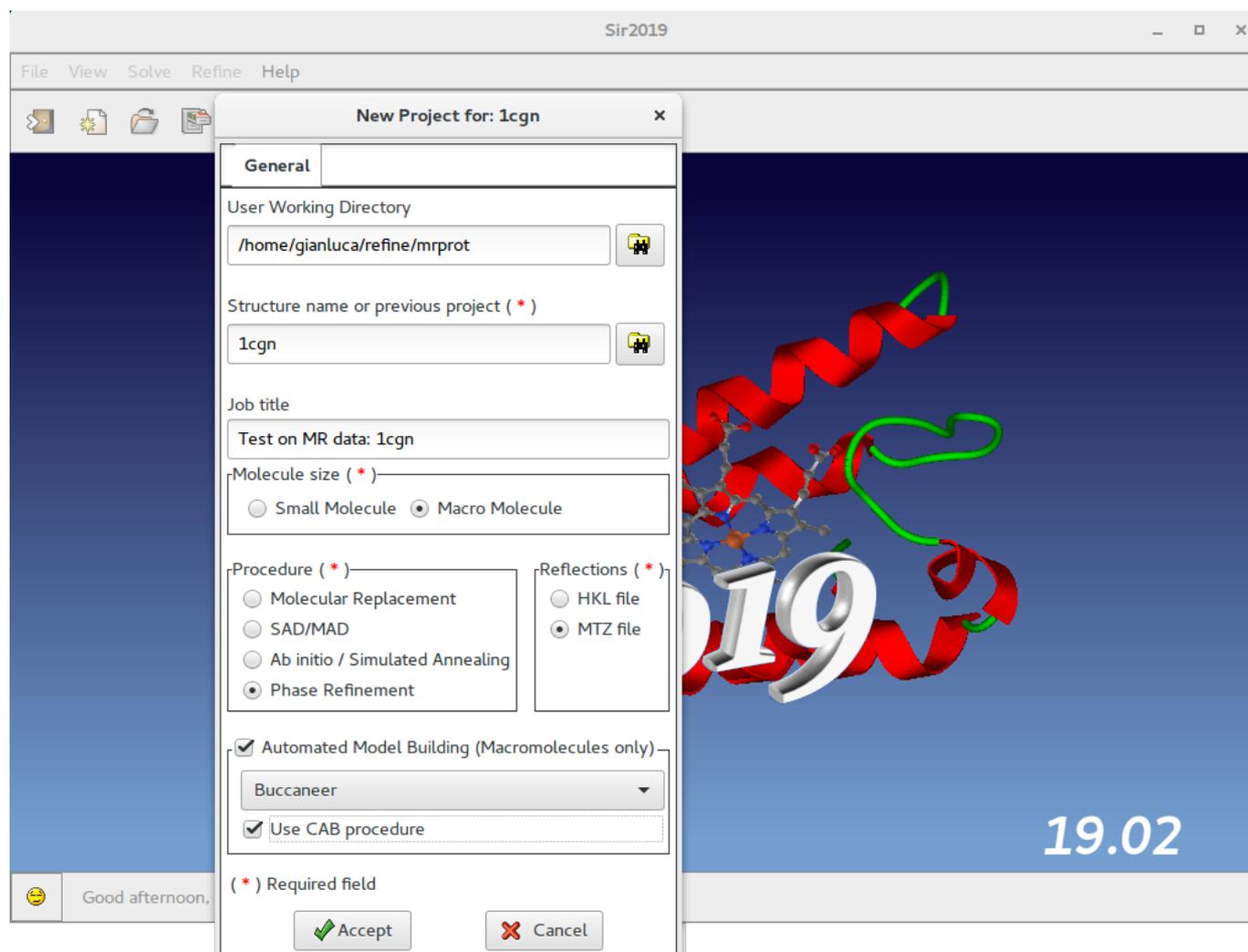
```
%Window
%CAB Buccaneer
%Structure caufd
%job pdb code : 2FDN
%data
  sequence 2fdn.seq
  mtz caufd_phase.mtz
  label H K L F SIGF PHIC FOM
%refine
  mad
%end
```

The full list of directives is available at the end of this page or [here](#). The directives for %data, in case of macromolecules, are available [here](#)

Refine Interface

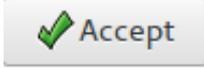
The user may be interested to refine by Sir2019 phases obtained by other programs: e.g., phases obtained by a previous application of MR tool, or obtained as output of some SAD-MAD program, or otherwise previously refined. The instructions in this sections, valid for macromolecules only, help to perform the refinement task.

A graphic interface is available in Sir2019 to refine phases by means of the *New* item  in the toolbar or in the menu *File* :



In this window the user should specify, among others, the working directory, the molecule size, the procedure to use (Phase Refinement) and the reflection file format (MTZ).

If an Automated Model Building program is installed, it is possible to use it by selecting Buccaneer or Nautilus or ARP/wARP or Phenix. If selected the CAB procedure will be applied at the end of this procedure.

Once clicked on 

it is possible to access the **MTZ** section:

New Project for: 1cgn
✕

General
MTZ

MTZ filename: (*)

/home/gianluca/refine/mrprot/1cgn_improved.mtz

Title: Output mtz file from reffmac

Change Space Group: P 65 2 2 ▼ Check Space Group

Cell Parameters: 54.6 54.6 180.4 90 90 120

Select F, sigF, PhiCalc and Fom

Selected columns:

H K L F SIGF PHIC FOM

Clear columns

Columns

value:

esd:

Add pair

FC ▼

▼

Add pair

Available columns

Label	Type
FC	F
FC_ALL	F
PHIC_ALL	P
FWT	F
PHWT	P

Add

(*) Required field

Accept

Cancel

Using this window it is possible to supply the name of the MTZ file and to select the necessary columns (H K L F SigF, PhiCalc and Fom).

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

New Project for: 1cgn
×

General
MTZ
Cell Content
Refine

Source

X Ray Data
 Neutron Data
 Electron Data

Molecular unit:

Number of molecules in the asymmetric unit:

Molecular Weight:

Number of residues:

Sequence (FASTA format): or paste data in the box above

```
>1CGN:_|PDBID|CHAIN|SEQUENCE
EFAKPEDAVKYRQSALTMASHFGRMAPY
QKQQAFQDNIVKLSAADAGDLDKLRAA
```

Ligands - Additional elements and corresponding number in asymmetric unit

Fe	1		0		0		0
----	---	--	---	--	---	--	---

Cell Volume: Volume per Atom:

(*) Required field

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

After data have been accepted, the **Refine** interface below is available, where the user may specify the method by which phases to refine were obtained. In case of Molecular Replacement a preliminary model may be supplied.

New Project for: 1cgn

General MTZ Cell Content Refine Run

Phase Refinement

Phases from:

- Molecular Replacement
- SAD/MAD
- Ab initio
- Let the program to decide

Model PDB File (when available)

(*) Required field

Once clicked on

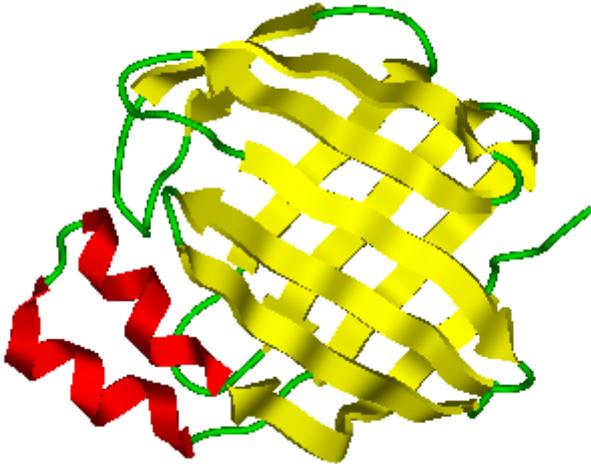
it is possible to access the **Run** section:

New Project for: 1cgn

General MTZ Cell Content Refine Run

Run

Graphic window required



Input filename (*) Already exists

1cgn.sir YES

Output filename (*) NO

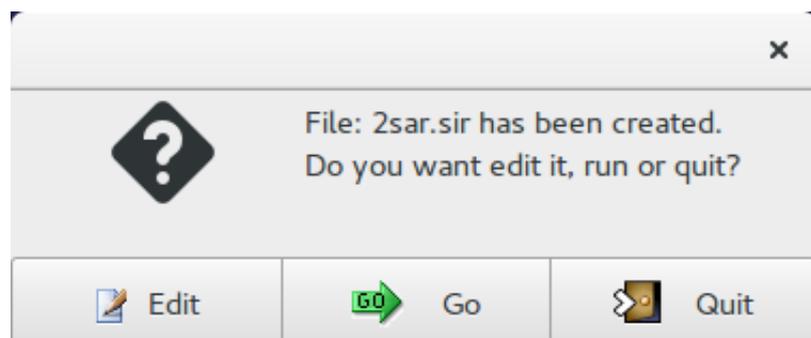
1cgn.out

Click on  to save the project

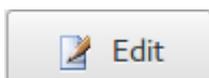
(*) Required field

If **Graphic window required** is checked, a graphical output will be produced (see below).

In this section it is possible to change the default names for the input (project) and for the output file. The user, by clicking on  button, can save the input file for Sir2019.



Now it is possible to edit



the input file (2sar.sir); to start calculations click on



Once the program is over, by means of the “[View output file](#)”  feature in “[File](#)” it is possible to access the complete output file.

By means of the “[Summary](#)”  feature in “[File](#)” menu is possible to get a synthetic output and also the graphic windows produced by the program during the run (in HTML format). An example of graphical output follows.

Test on MR data: 1cgn _ □ ×

File View Solve Refine Help



Wed Feb 20 2019 at 15:36:08 Phantom Derivative procedure starts on: 1cgn

Data Section:
8220 observed reflections
Observed resolution range: 19.78 - 2.15 Å .
26103 reflections extrapolated up to 1.37 Å .

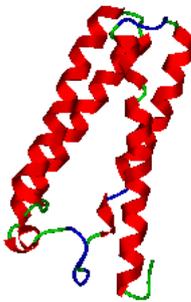
Protein phase refinement and automatic model building
Edm, Vld, Free Lunch and Phantom Derivative techniques
are used to extend and refine phases.
Phased reflections 7806 over 8220 from MTZ file
Running PhD ... Cycle 2/ 2 derivative 1/ 1

CAB procedure using Buccaneer
R = 0.2146 (Rfree = 0.2545) after 2 CAB cycles
R = 0.1876 (Rfree = 0.2093) after 30 Refmac cycles

Final CAB results:
R = 0.1861 (Rfree = 0.2059) Coverage = 100%

Files produced at the end of the run:
MTZ format PDB format
1cgn_buccaneer.mtz 1cgn_buccaneer.pdb

Best model preview



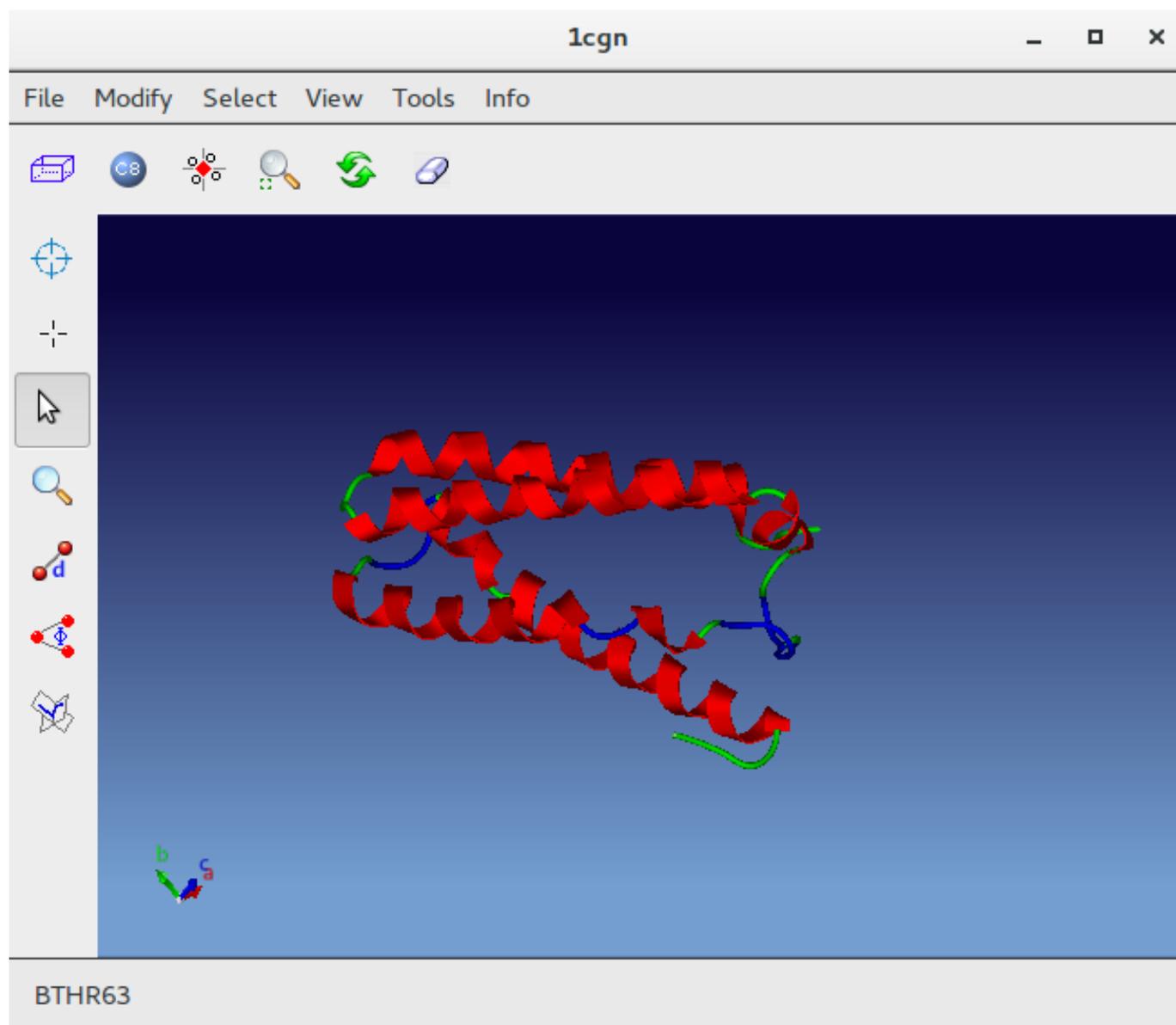
Sir2019 is over1cgn_fPlot.png



Final Refmac refinement

When the program is over it is possible to interact with the structure using 

Jav:



Refine directives

MODEL string

String is the name of the coordinates file (*.pdb) where a previously created model is stored . (i.e. a Molecular Replacement result).

MTZ string

String is the name of the reflections file in MTZ format, if not already defined in %Data. Phase and fom columns are necessary to perform refinement.

LABELS strings

Strings are the names of columns in the MTZ reflections file, if not already defined in %Data.

MR

Phases have been produced by a Molecular Replacement procedure.

MAD

Phases have been produced by a MAD/SAD procedure.

ABI

Phases have been produced by an *ab initio* procedure.

[Previous page](#)

[Next page](#)

References

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/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.*, D66, 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.*, A64, 326-336.
- Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A & Rizzi, R. (2008), *J. Appl. Cryst.*, 41, 592-599.
- Arnold, E. & Rossmann, M. G. (1988). *Acta Cryst.*, A44, 270–283.
- Arnott, S. & Wonacott, A.J. (1966), *Polymer*, 7, 157-166.
- Baggio, R., Woolfson, M.M., Declerq, J.P. & Germain, G. (1978), *Acta Cryst.*, A34, 883-892.
- Burla, M.C., Caliendo, 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., Caliendo, 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., Caliendo, 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., Caliendo, R., Carrozzini, B., Cascarano, G.L., Cuocci, C., Giacovazzo, C., Mallamo, M. Mazzone A. & Polidori, G. (2015), *J. Appl. Cryst.*, 48, 306-309.
- Burla, M.C., Caliendo, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacovazzo, C. & Polidori, G. (2004a), *J. Appl. Cryst.*, 37, 258-264.
- Burla, M.C., Caliendo, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacovazzo, C. & Polidori, G. (2004b), *J. Appl. Cryst.*, 37, 791-801.
- Burla, M.C., Caliendo, 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., Caliendo, 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., Caliendo, R., Giacovazzo, C. & Polidori, G. (2010), *Acta Cryst.*, A66, 347-361.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., De Caro, L., Giacovazzo, C. & Polidori, G. (2003a), *Acta Cryst.*, A59, 245-249.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C. & Polidori, G. (2002a), *Z. Kristallogr.*, 217, 629-635.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C., Polidori, G. & Siliqi D. (2002b), *Acta Cryst.*, D58, 928-935.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C. & Polidori, G. (2003b), *Acta Cryst.*, D59, 662-669.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C. & Polidori, G. (2011), *J. Appl. Cryst.*, 44, 1143-1151.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C. & Polidori, G. (2012), *J. Appl. Cryst.*, 45, 1287-1294.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C. & Polidori, G. (2015a), *Acta Cryst.*, D71, 1864-1871.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C. & Polidori, G. (2015b), *J. Appl. Cryst.*, 48, 1692-1698.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C. & Polidori, G. (2017), *J. Appl. Cryst.*, 50, 1048-1055.
- Burla, M.C., Cascarano, G.L. & Giacovazzo, C. (1992), *Acta Cryst.*, A48, 906-912.
- Burla, M.C., Cascarano, G.L., Giacovazzo, C. & Polidori, G. (2017), *Acta Cryst.*, A73, 218-226.
- 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.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., Giacovazzo, C., Polidori, G. (2017), *J. Acta Cryst*, D73, 877-888.
- Burla, M.C., Carrozzini, B., Cascarano, G.L., Polidori, G. and Giacovazzo, C. (2018), *J. Acta Cryst*, D74, 1096-1104.
- 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.*, A69, 98-107.
- Caliandro, R., Carrozzini, B., Cascarano, G.L., Comunale, G., Giacovazzo, C. & Mazzone A. (2014), *Acta Cryst.*, D70, 1994-2006.

- Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacobazzo, C. & Siliqi, D. (2005a), *Acta Cryst.*, A61, 556-565.
- Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacobazzo, C. & Siliqi, D. (2005b), *Acta Cryst.*, A61, 1080-1087.
- Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacobazzo, C., Moustiakimov, M. & Siliqi, D. (2005), *Acta Cryst.*, A61, 343-349.
- Caliandro, R., Carrozzini, B., Cascarano, G.L., De Caro L., Giacobazzo, C. & Siliqi, D. (2007), *J. Appl. Cryst.*, 40, 883-890.
- Caliandro, R., Carrozzini, B., Cascarano, G.L., Giacobazzo, C., Mazzone, A. & Siliqi, D. (2009), *Acta Cryst.*, A65, 512-527.
- Camalli, M., Carrozzini, B., Cascarano, G.L. & Giacobazzo, C. (2011), *Cryst. Res. Technol.*, 46, No. 6, 555-560.
- Carrozzini, B., Cascarano, G.L., Comunale, G., Giacobazzo, C. & Mazzone, A. (2013), *Acta Cryst.*, D69, 1038-1044.
- Carrozzini, B., Cascarano, G.L. & Giacobazzo, C. (2016), *Acta Cryst.*, D72, 551-557.
- Carrozzini, B., Cascarano, G.L., Giacobazzo, C. & Mazzone, A. (2015), *Acta Cryst.*, D71, 1856-1863.
- Cascarano, G.L. & Giacobazzo, C. (1995), *Acta Cryst.*, A51, 820-825.
- Cascarano, G.L., Giacobazzo, C. & Carrozzini, B. (2010), *Ultramicroscopy*, 111, 56-61.
- Cascarano, G.L., Giacobazzo, C. & Guagliardi, A. (1991), *Acta Cryst.*, A47, 698-702.
- Cascarano, G.L., Giacobazzo, C. & Lui?, M. (1988a), *Acta Cryst.*, A44, 176-183.
- Cascarano, G.L., Giacobazzo, C. & Lui?, M. (1988b), *Acta Cryst.*, A44, 183-188.
- Cascarano, G.L., Giacobazzo, C., Burla, M.C., Nunzi, A. & Polidori, G. (1984), *Acta Cryst.*, A40, 389-394.
- Cochran, W. (1955), *Acta Cryst.*, 8, 473-478.
- Cowtan, K. D. (1998). *dm. Software for Density Modification, Versions 1.8-2.0*. Department of Chemistry, University of York, Heslington, York, England.
- Cowtan, K. (1994) *Acta Cryst.*, D50, 760-763.
- Cowtan, K. (2006) *Acta Cryst.*, D62, 1002-1011.
- Fan, Hai-Fu, Yao, Jia-Xing & Qian, Jin-Zi (1988), *Acta Cryst.*, A44, 688-691.
- Frigo, M. and Johnson, S.G. (2005) *Proceedings of the IEEE* 93 (2), 216-231.
- Giacobazzo, C. (1976), *Acta Cryst.*, A32, 958-966.

Giacovazzo, C. (1977), *Acta Cryst.* A33, 933-944.

Giacovazzo, C. (1980), *Acta Cryst.* A36, 362-372.

Giacovazzo, C. (2015), *Acta Cryst.* A71, 483-512.

Giacovazzo, C. & Siliqi, D. (1997), *Acta Cryst.* A53, 789-798.

Giacovazzo, C. & Siliqi, D. (2004), *Acta Cryst.* D60, 73-82

Giacovazzo, C., Burla, M.C. & Cascarano, G. (1992), *Acta Cryst.*, A48, 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.*, A43, 134-136.

Matthews, B.W. (1968), *J. Mol. Biol.*, 33, 491-497.

Mooers, B.H. & Matthews, B.W. (2006) *Acta Cryst.*, D62, 165-176.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.*, D67, 355–367.

Otwinowski, Z. (1991). In: *Isomorphous Replacement and Anomalous Scattering, Proceedings of the CCP4 Study Weekend*; eds. Wolf, W., Evans, P.R. and Leslie, A.G.W., (Daresbury Laboratory), pp. 80-86.

Pähler, A., Smith, J.L. & Hendrickson, W.A. (1990). *Acta Cryst.*, A46, 537-540.

Perrakis, A., Morris, R.M. & Lamzin, V.S. (1999), *Nature Struct. Biol.*, 6, 458-463.

Refaat, L.S. & Woolfson, M.M. (1993), *Acta Cryst.*, D49, 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. (1998). *Direct Methods for Solving Macromolecular Structures*; ed. Fortier, S., pp. 131-141. Dordrecht: Kluwer Academic Publishers.

Sheldrick, G.M. & Gould, R.O. (1995), *Acta Cryst.*, B51, 423-431.

Shiono, M. & Woolfson, M.M. (1992), *Acta Cryst.*, A48, 451-456.

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.

Terwilliger, T.C. (1994a). *Acta Cryst.*, D50, 11-16.

Terwilliger, T.C. (1994b). *Acta Cryst.*, D50, 17-23.

Terwilliger, T.C., Grosse-Kunstleve, R.W. , Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.-W. Read, R.J. and Adams P.D. (2008). *Acta Cryst.*, D64, 61-69.

Wang, B.C. (1985), *Methods Enzymol.*, 115, 90-112.

Wilson, A.J.C. (1942), *Nature*, 150, 152.

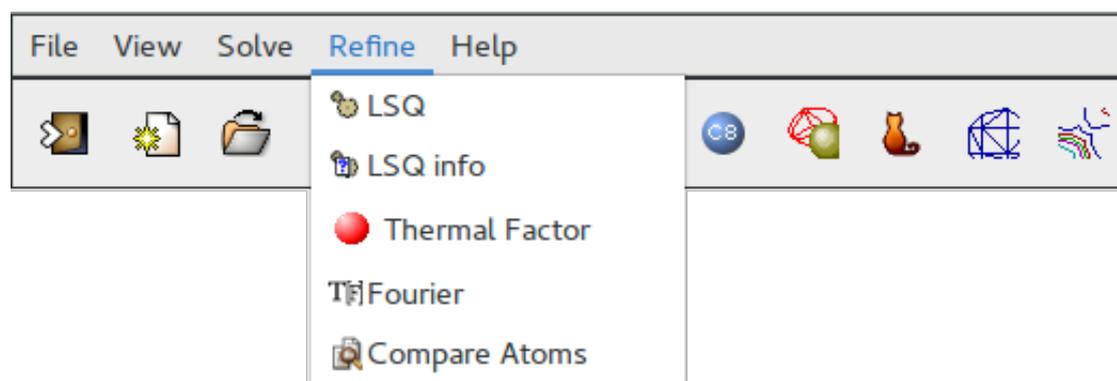
[Previous page](#)

Refine

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/refine/>

Sub-menu Refine:



It allows:

 [LSQ](#)

 [LSQ info](#)

 [Thermal Factor](#)

 [Fourier](#)

 [Compare Atoms](#)

LSQ

Least Squares graphical interface is divided in 2 sections:

- General
- Modify Atoms Parameters

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.

To start the calculations click on  To get help click on 

Least Squares Refinement

General Modify Atoms Parameters

Refinement Request	Refinement Conditions
<input type="radio"/> Diagonal	Damping Factor <input type="text" value="1.0"/> - +
<input checked="" type="radio"/> Full Matrix	Reflections cutoff <input type="text" value="3"/> - +
<input type="radio"/> S.F.C. only	# of Cycles <input type="text" value="3"/> - +

Weighting Scheme

6 : W=1.0

Observations & Parameters

Observations: 2255 (92 %) Parameters: 109 Ratio: 20.69



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

Least Squares Refinement ✕

General Modify Atoms Parameters

Atomic Parameters

Filter species

Hydrogen Carbon Oxygen All

	Fix XYZ	Fix B	Fix Occ.	Set Isotropic
O1	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
O2	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
O3	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

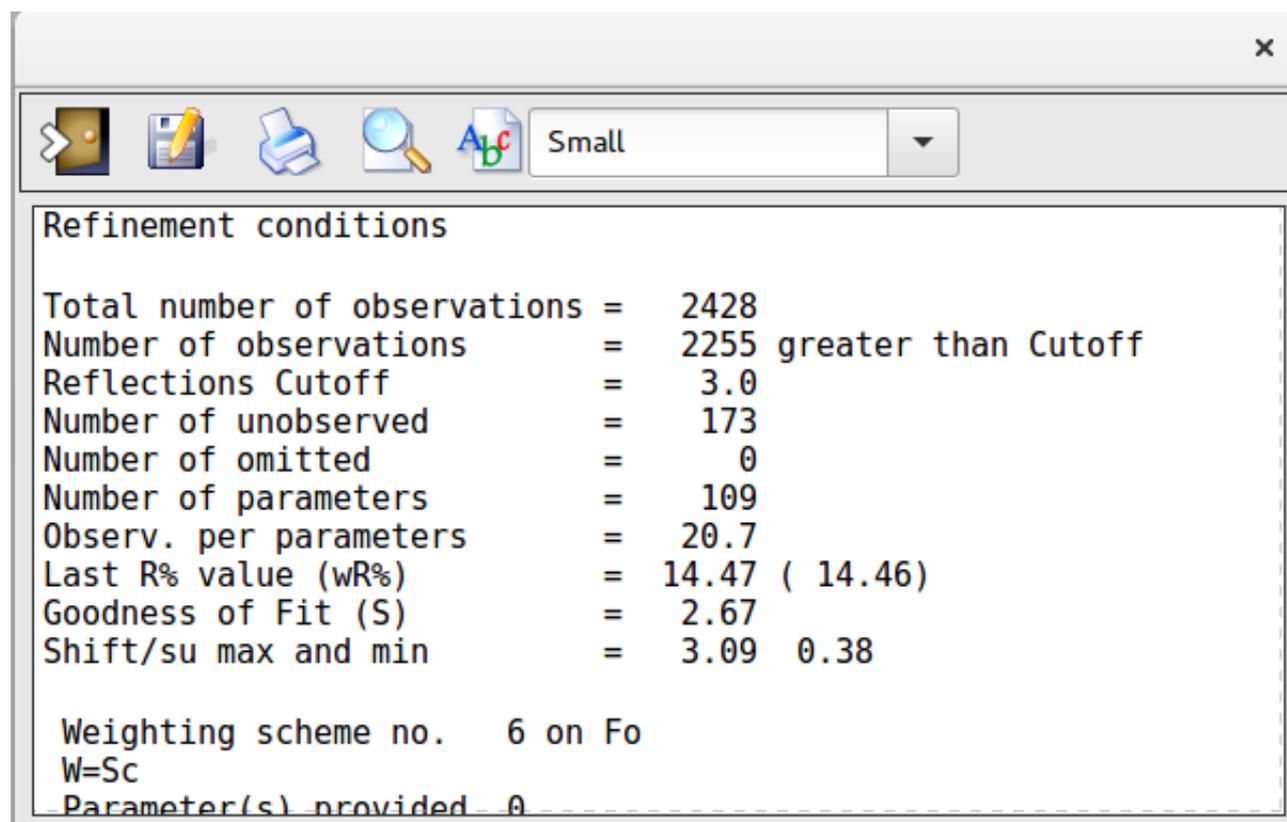
Observations & Parameters

Observations: 2255 (92 %) Parameters: 109 Ratio: 20.69

 Help  GO  OK  Cancel

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.



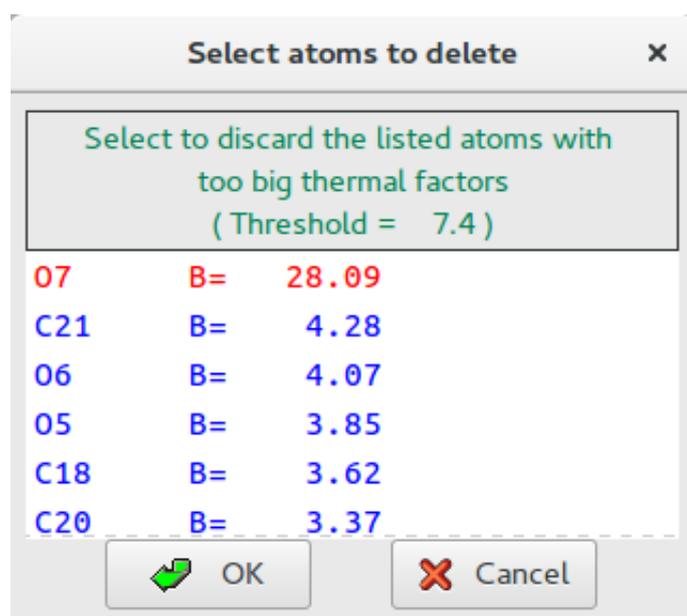
Refinement conditions

Total number of observations	=	2428
Number of observations	=	2255 greater than Cutoff
Reflections Cutoff	=	3.0
Number of unobserved	=	173
Number of omitted	=	0
Number of parameters	=	109
Observ. per parameters	=	20.7
Last R% value (wR%)	=	14.47 (14.46)
Goodness of Fit (S)	=	2.67
Shift/su max and min	=	3.09 0.38

Weighting scheme no. 6 on Fo
W=Sc
Parameter(s) provided 0

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.



Select atoms to delete

Select to discard the listed atoms with too big thermal factors (Threshold = 7.4)

07	B=	28.09
C21	B=	4.28
06	B=	4.07
05	B=	3.85
C18	B=	3.62
C20	B=	3.37

OK Cancel

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.

To get help click on .

Fourier Synthesis ✕

Info: 25 atoms in current model

Fourier type <input type="radio"/> Observed <input checked="" type="radio"/> 2*Fobs - Fcalc <input type="radio"/> Fobs - Fcalc	Fourier conditions Peaks to add 5 - + Level (electrons / Ang ³) 0.8 - +
--	--

Fourier space <input checked="" type="radio"/> Unit Cell <input type="checkbox"/> View Fourier map <input type="radio"/> Cube	Grid Step 0.20 - +	Cube edge 2.00 - +
---	---	---





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 and the calculated structure factors.

Fourier Conditions

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

- **Level (electrons/Å³):** 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 step of the Fourier synthesis (default 0.20 Å) can be defined by the user. Using the mouse and moving the model, the user can place the cube where it is needed.

If 'View Fourier map' is checked, the Fourier map and the model are shown by 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 "Select peaks to accept", so the user can select peaks to retain. New peaks (two in this example) are coloured in light blue in the Jav window.

The screenshot displays the Sir software interface. The main window, titled 'test structure: loganin', shows a 3D ball-and-stick model of the loganin molecule. A smaller window titled 'loganin' provides a different view of the molecule. A dialog box titled 'Select peaks to accept' is open, displaying the following table:

Label	Intensity	Minimum distance	From
Q1	139	1.427	C16
Q2	127	1.530	C11

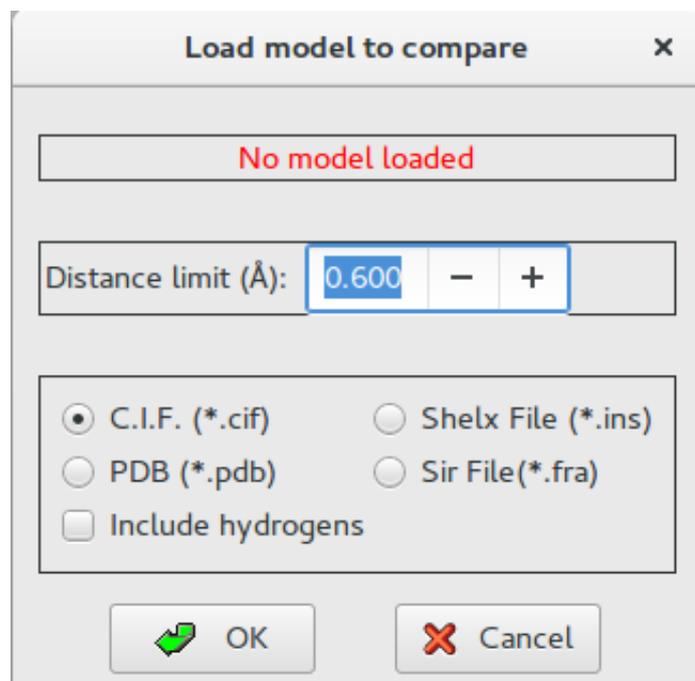
The dialog box also includes buttons for 'OK', 'All', and 'Cancel'. A status bar at the bottom of the dialog box reads 'Good morning, gianluca. Welcome to jav'.

By means of the Jav interface is possible to delete them or to assign an atomic species.

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. The list of the atoms matching the coordinates of the reference model is produced.

Test structure: barite
_ □ ×

File View Solve Refine Help

Comparison results
×

Results

SIR coordinates 10
 Target coordinates 5
 Matches found 4
 Shift applied 0.000 0.500 0.500 Sign is +
 Mean Phase Error 33.972 using 355 reflections
 Distance limit: 0.600
 <Dist> among found 0.088
 RMSD 0.215

Model

Display model Display labels

Matches

💡 Select a row to center the plot

	----- SIR coordinates -----	Distance	---- Model coordinates ----	
Ba1	0.182 0.250 0.155	0.03048	0.184 0.250 0.159	Ba1
S1	0.431 0.750 0.190	0.05983	0.437 0.750 0.191	S1
O1	0.422 0.982 0.310	0.06954	0.420 0.970 0.312	O3
O2	0.605 0.750 0.089	0.19276	0.589 0.750 0.107	O1

Final R value is 22.52% in trial 5

[Previous page](#)

[Next page](#)

108 / 149

SAD/MAD structure determination

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/substructure-determination/>

Basics

- [SAD/MAD graphical interface](#)
- [SAD/MAD directives](#)

Basics

Sir2019 uses a two step phasing technique (Pälher *et al.*, 1990; Terwilliger, 1994a,b; Sheldrick, 1998; Burla *et al.*, 2002b; Burla *et al.* 2007). First, the parameters defining the anomalous scatterer substructure are determined, and then the protein phases are assigned and refined.

The program may treat up to four wavelengths. The phasing procedure is based on the following mathematical techniques.

The expected conditional value

$\langle R_{0a} | R_1^+, \dots, R_n^+, R_1^-, \dots, R_n^- \rangle$ is derived from the joint probability distribution (Burla *et al.*, 2002b, 2003b)

$$P(R_{0a}, R_1^+, \dots, R_n^+, R_1^-, \dots, R_n^-, \varphi_{0a}, \varphi_1^+, \dots, \varphi_n^+, \varphi_1^-, \dots, \varphi_n^-)$$

where R 's are the amplitudes of the normalized structure factors at the various wavelengths and φ are the corresponding phases. R_{0a} is the normalized amplitude of the structure factor corresponding to the anomalous scatterer substructure, calculated by neglecting the anomalous dispersion effects on the scattering factors; φ_{0a} is the corresponding phase value.

$\langle R_{0a} | \dots \rangle$ is considered to be the best estimate of R_{0a} , and therefore may be used in two ways for recovering the anomalous substructure: by tangent procedures or by Patterson deconvolution techniques. In our tests we always use the second approach, which is the default choice of Sir2019.

An algorithm is automatically used (default mode) for limiting data resolution and for predicting the most informative wavelength combinations in a MAD experiment.

Once a model structure is found (actually the program provides 30 model candidates ranked by a FOM, but only the best ten will be used for phasing), the program applies an automatic least squares procedure for improving the substructure parameters and for refining φ^+ and φ^- values (Giacovazzo & Siliqi, 2004; see also Otwinowski,

1991). The information so gained is used in the second phasing step, where the protein phases are estimated from the experimental anomalous dispersion effects. Such phases are assigned according to the probability distribution function

$$P(\varphi_p | \{R_j^+, R_j^-, R_{aj}^+, R_{aj}^-, \varphi_{aj}^+, \varphi_{aj}^-, \text{for } j = 1, \dots, n\}).$$

where φ_p is the protein phase to assign, R_a and φ_a (supposed to be well estimated at this step) are the normalized amplitudes and phases of the anomalous substructures.

The protein phases φ_p so estimated are used as starting set for the application of the *EDM-VLD-FL* techniques before described.

In order to automatically run the full process the user is only asked to submit the experimental data in terms of all the data sets (four columns for each wavelength and the native dataset when available), the atomic species of the anomalous scatterers, the estimated/expected number of them in the asymmetric unit, the corresponding f' and f'' values expected at the experimental wavelengths, and the information about the unit cell content of the protein. This last may be given by specifying the number of residues, or the protein sequence or the overall molecular weight.

It is also possible to decide the best strategy for the parameters to be refined.

It is possible to create the input file for Sir2019 using a text editor; in the following example the CAB procedure will use Buccaneer to complete and refine the model:

```
%structure ttg
% CAB buccaneer
%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
  sequence 1srv.seq
  nMolecules 1
  anomalous Se 3 -1.805 0.646 -8.582 3.843 -7.663 3.841 -2.618 3.578
  native
%mad
%end
```

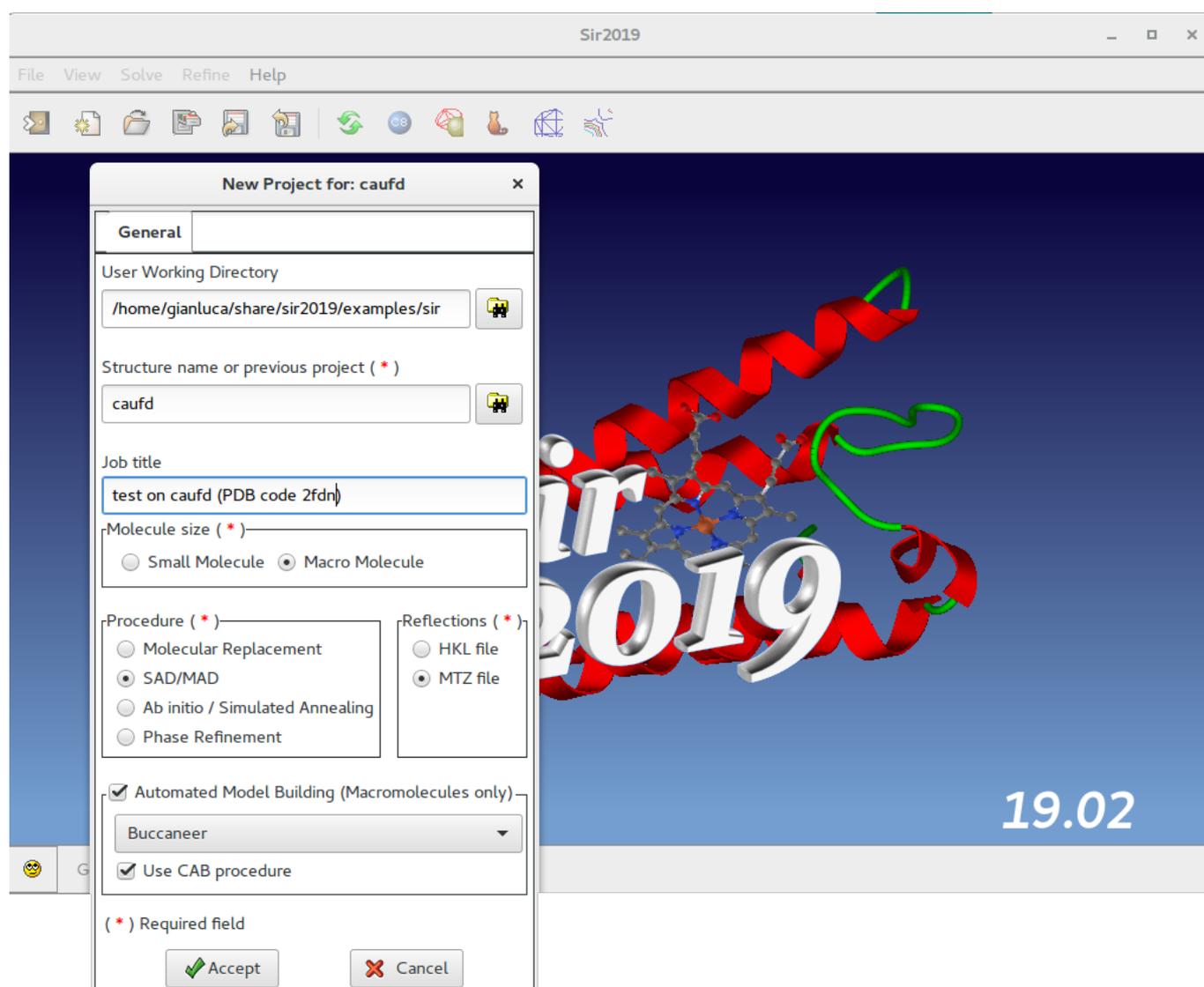
An example of SAD data follows:

```
%struct caufd
%job SAD test structure caufd (pdb code 2FDN)
%data
  mtz caufd.mtz
  label H K L FP SIGFP F(+) SIGF(+) F(-) SIGF(-)
  native
  sequence caufd.seq
  nMolecules 1
  anom fe 8 0.26 1.25
%mad
```

%end

SAD/MAD graphical interface

A graphic interface is available in Sir2019 to work with SAD/MAD structure determination by means of the **New** item  in the toolbar or in the menu **File** :



In this window the user should specify, among others, the working directory, the molecule size, the procedure to use (SAD/MAD) and the reflection file format (MTZ).

Macromolecules only: if an Automated Model Building (AMB) program is installed, it is possible to use it by

selecting Buccaneer or Nautilus or ARP/wARP or Phenix. It will be used in CAB procedure if selected.

Once clicked on 

it is possible to access the **MTZ** section:

New Project for: caufd x

General MTZ Cell Content

MTZ filename: (*)

Title:

Change Space Group: Check Space Group

Cell Parameters:

Select F, sigF for protein if available and F+,sigF+, F-,sigF- for every wavelength

Selected columns:

H K L FP SIGFP F(+) SIGF(+) F(-) SIGF(-)

Available columns

Label	Type
DANO	D

Columns

value:

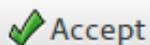
esd:

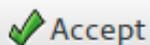
(*) Required field

This form is used to get all the information from an MTZ file. Once browsed, it is possible to view the available columns.

The user can add, when available, the two columns for F and sigma(F) with data without anomalous information (native date). For every wavelength the pair columns F+ / sig(F+) and F- / sig(F-) are necessary (four columns per wavelength).

If necessary it is possible to specify a different space group; the user can just write the new symbol or to select it from a list. It is also possible to check the possible space group (see [here](#)).



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

New Project for: caufd
×

General MTZ
Cell Content
Solve

Source

X Ray Data
 Neutron Data
 Electron Data

Molecular unit:

Number of molecules in the asymmetric unit:

Molecular Weight:

Number of residues:

Sequence (FASTA format): or paste data in the box above

```
>2FDNA:FERREDOXIN
AYVINEACISCGACEPECPVNAISSGDDRYVIDATC
```

Additional elements in asymmetric unit with anomalous dispersion coefficients

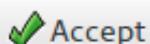
Element	Amount	Wave 1: f' and f''	Wave 2: f' and f''	Wave 3: f' and f''	Wave 4: f' and f''
fe	8	0.26	1.25		

Cell Volume: Volume per Atom:

(*) Required field

This form is used to supply the cell content. It is possible to specify the number of molecules in asymmetric unit. The chemical content can be specified through the weight or the number of residues or the sequence (in FASTA format); the sequence can be stored in a file or pasted directly in the box (up right part of the form). In this example, the sequence (in FASTA format) has been supplied. The program needs to know an approximate number of anomalous species and the value for f' and f'' for all the wavelengths supplied in the previous window.



Once clicked on  it is possible to access the **Solve** section:

New Project for: caufd
×

General MTZ Cell Content Solve Run

SAD/MAD phasing

Select waves: Wave 1 Wave 2 Wave 3 Wave 4

N. trials to generate : Use default values

Anomalous scatterers positions file: (PDB)

Best trials All from : to :

LSQ: parameters to refine

Atomic parameters: position B-factor sof

Wave parameters: $\Delta f'$ f''

(*) Required field

The strategy for the structure solution is defined in this form. It is possible to select which wavelength has to be used (all in default). The program will use the Patterson approach to find the substructures; it is possible to specify the number of trials to be used or rely on its default value.

As an alternative to the substructures calculation it is possible to supply a PDB file.

Once clicked on

it is possible to access the **Run** section:

New Project for: caufd

General MTZ Cell Content Solve **Run**

Run

☺

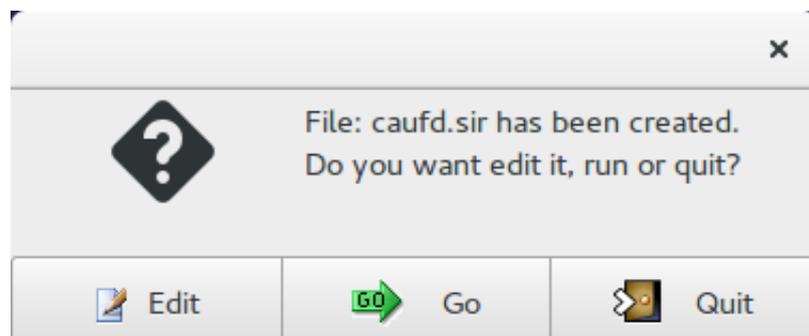
Input filename (*)	Already exists
<input type="text" value="caufd.sir"/>	NO
Output filename (*)	
<input type="text" value="caufd.out"/>	NO

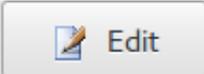
Click on  to save the project

(*) Required field

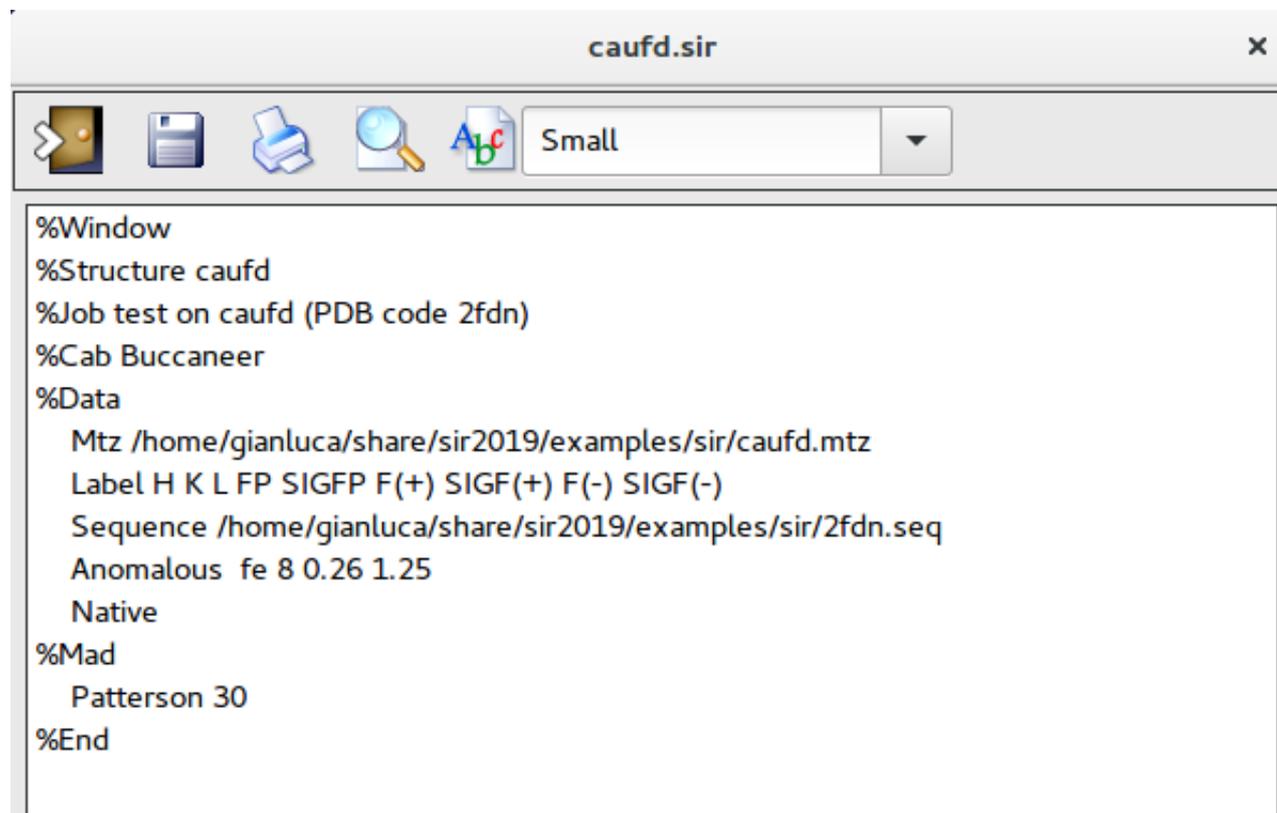
If **Graphic window required** is checked, a graphical output will be produced (see below).

In this section it is possible to change the default names for the input (project) and for the output file. The user, by clicking on  button, can save the input file for Sir2019.



Now it is possible to edit 

the input file (caufd.sir in this example):



To start calculations click on 

Once the program is over, by means of the “[View output file](#)”  feature in “[File](#)” it is possible to access the complete output file.

By means of the “[Summary](#)”  feature in “[File](#)” menu is possible to get a synthetic output and also the graphic windows produced by the program during the run (in HTML format). An example of graphical output follows.

First part: substructure determination

Fri Feb 9 2018 at 15:24:03 Mad/Sad substructure procedure starts on: ttg

Data Section:

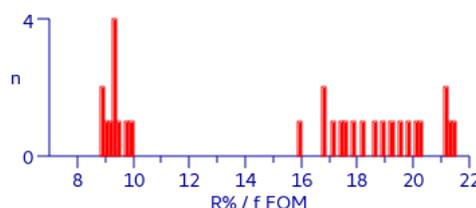
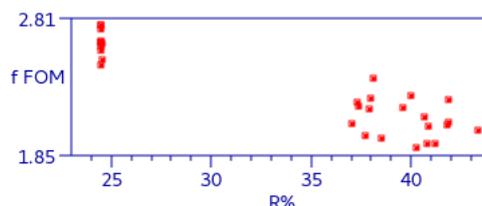
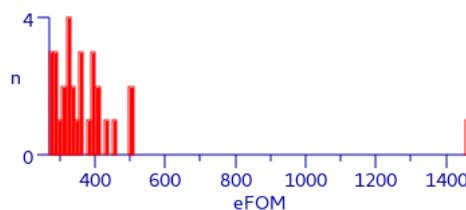
3 anomalous scatterers to locate.
 The input file contains native and data measured at 4 wavelengths
 The wavelength combination 2 3 4 has the highest correlation coefficient
 between the Δ_{ano} values and will be used to estimate $|E_{0a}|$ values
 (normalized amplitudes of the anomalous scatterer substructure).
 5618 structure factors have been normalised.
 The average thermal factor (through Wilson plot) is : 25.28

Initial phasing:

$|E_{0a}|$ amplitudes are used to calculate a Patterson map.
 Then Patterson superposition techniques are used to locate
 the anomalous scatterer substructure.
 30 trials requested and 30 trials stored.
 They will be used to locate the anomalous substructure.
 For each trial, 1310 reflections are used to locate and refine
 the anomalous substructure.

Phase refinement:

The best trials are expected to have the lowest value of R%/fFOM.
 At the end of the procedure 30 substructure models have been stored.



**Second part:
 anomalous substructure refinement and phase extension**

Wed Feb 14 2018 at 11:42:06 Mad/Sad substructure procedure starts on: ttg

Anomalous substructure refinement

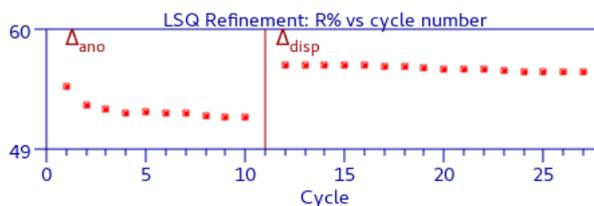
Reflection Data:

Wavelength:	λ_1	λ_2	λ_3	λ_4
Nr. Reflections:	7044	6848	7258	7181
Resolution Range:	2.3- 19.9	2.3- 20.1	2.3- 20.0	2.3- 20.1
Completeness(%):	99.7	99.7	99.7	99.7
Thermal Factor:	28.0	24.9	25.8	24.5

Refinement/Phasing

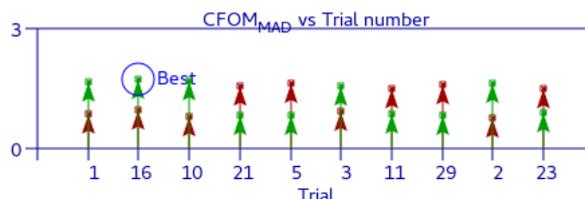
LSQ of the 10 selected substructure models in order of FOM_{FRAG}
 (Otwinowski,1991)
 After each LSQ both the enantiomorphs will be tested.
 The most probable solution will be indicated by the highest $CFOM_{MAD}$.
 Protein phases are calculated given observed amplitudes
 and substructure phases up to SAD/MAD resolution.

Current Trial = 23 (10/10) $FOM_{FRAG} = 9.75$



LSQ Refinement:

Number of Anomalous Scatterers: 3
 Parameters to be refined: x y z occ bt P
 Nr of Reflection Used: 4487 3952 3772 3706 R% : 51.97
 Parameters to be refined: occ Δf
 Nr of Reflection Used: 785 705 692 0 R% : 56.09



Third part:

protein phase refinement and automatic model building using CAB

test on caufd (PDB code 2fdn) _ □ ×

File View Solve Refine Help

Tue Feb 19 2019 at 17:03:15 Phantom Derivative procedure starts on: caufd

Protein phase refinement and automatic model building
 Edm, Vld, Free Lunch and Phantom Derivative techniques
 are used to extend and refine phases up to native resolution.
 SAD/MAD phased reflections 17131 over 29094 for the most promising trial.
 In PhD 2 cycles and 3 derivatives will be used
 Running PhD ... Cycle 2/ 2 derivative 3/ 3

Enantiomorph check using Buccaneer ... done!

CAB procedure using Buccaneer
 R = 0.4497 (Rfree = 0.4874) after 3 CAB cycles
 R = 0.2423 (Rfree = 0.2600) after 155 Refmac cycles

Final CAB results:
 R = 0.2423 (Rfree = 0.2582) Coverage = 100%

Files produced at the end of the run:
 MTZ format PDB format
 caufd_buccaneer.mtz caufd_buccaneer.pdb

Best model preview

Sir2019 is over caufd_fPlot.png

😊

SAD/MAD directives

The list of directives is also available [here](#).

The following two directives are part of the %DATA section

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.

NATIVE

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

The following directives are part of the %MAD section

WAVE n1 n2 n3 n4

Which wavelengths will be used for substructure refinement and phase extension. The wavelength number order corresponds to that specified in %Data with the directive ANOMALOUS

For example:

WAVE 1 3 Only the first and the third wavelength will be used.

If the directive WAVE is absent all the wavelengths will be used.

SET n1 n2

Specifies the number (the range) of the substructures previously obtained.

n1, n2 = 1 , 2 ... 10 (ordered by FOM_frag)

For example:

SET 1 2 the first two sets (in order of FOMfrag) will be used.

The default is SET 1 – 1, i.e. all fragments will be used.

LSQ nm

Conditions for the refinement (LSQ) of the substructure/wavelength parameters;

n=0,1,..5 specifies the options for substructure parameters (x,y,z, B-factor, sof).

0 no refinement

1 only B-factor

2 x, y,z

3 x, y, z , B-factor

4 only sof

5 x, y, z, B-factor, sof

The default value is 5.

m specifies the options for wavelength parameter (f' , df').

0 no refinement

1 f'

2 df'

3 f' , df'

The default value is 3.

For example:

LSQ 50 Refine only substructure parameters x, y, z , B-factor, sof

LSQ 53 Refine all parameters: x, y, z , B-factor, sof, f' and df'

POSITIONS string

String is the name of the file (*.pdb) which contains the coordinates of the anomalous substructure.

Default: the program will use the file *structurename_SIR200x.fra* written during the anomalous scatterer substructure determination.

[Previous page](#)

[Next page](#)

Simulated Annealing

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/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 defining the model are modified until a good match between calculated and observed structure factors 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 position 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 Sir2019 window during the run and the solution menu are reported.

Global optimization dialog ×

SA conditions External DOF Internal DOF Anti-bump

General conditions

Cost function: R structure factor ▼

Resolution: 2.000 - + N. of reflections: 239

Random seed: 1 - +

Nr. of runs: 10 - +

Global optimization conditions

Number of cycles: 20 - + automatic

Randomize parameters

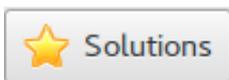
Number of refined parameters: 7 ★ Solutions

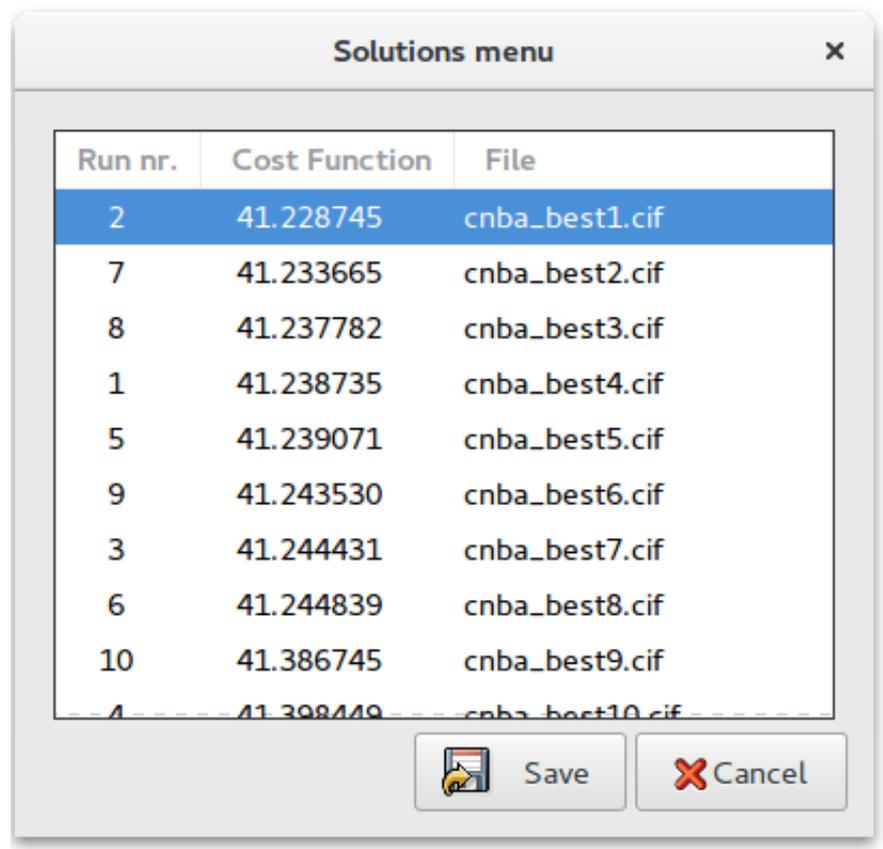
? Help  Quit  Execute

To get help click on 

To start click on 

To look at solutions use





Create input file for Simulated Annealing

To run Sir2019 for structure solution by Simulated Annealing you need to create an input file (*.sir) using a text editor:

```
%structure cnba
%job CNBA Courtesy of E. Mugnaioli (21/09/2010)
%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
  Format(3i4, 2f8.2)
%import antracene.mol
%sannel
%end
```

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.

New Project for: cnba

General Symmetry & Cell parameters Cell Content Reflections **Solve Run**

Solving strategy

- Let the program to decide
- Modern Direct Methods (MDM)
- Standard Direct Methods (SDM)
- Patterson
- VLD
- Simulated Annealing

Size: Let the program to decide ▾
6-80 non-H atoms in asymmetric unit

Fragment for simulated annealing

Number of fragments: 1 - +

Filename (*) /home/gianluca/share/ilmilione/examples/sir/antracene.mol

(*) Required field
Check results below

Accept Cancel

Output

“%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 repeated 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

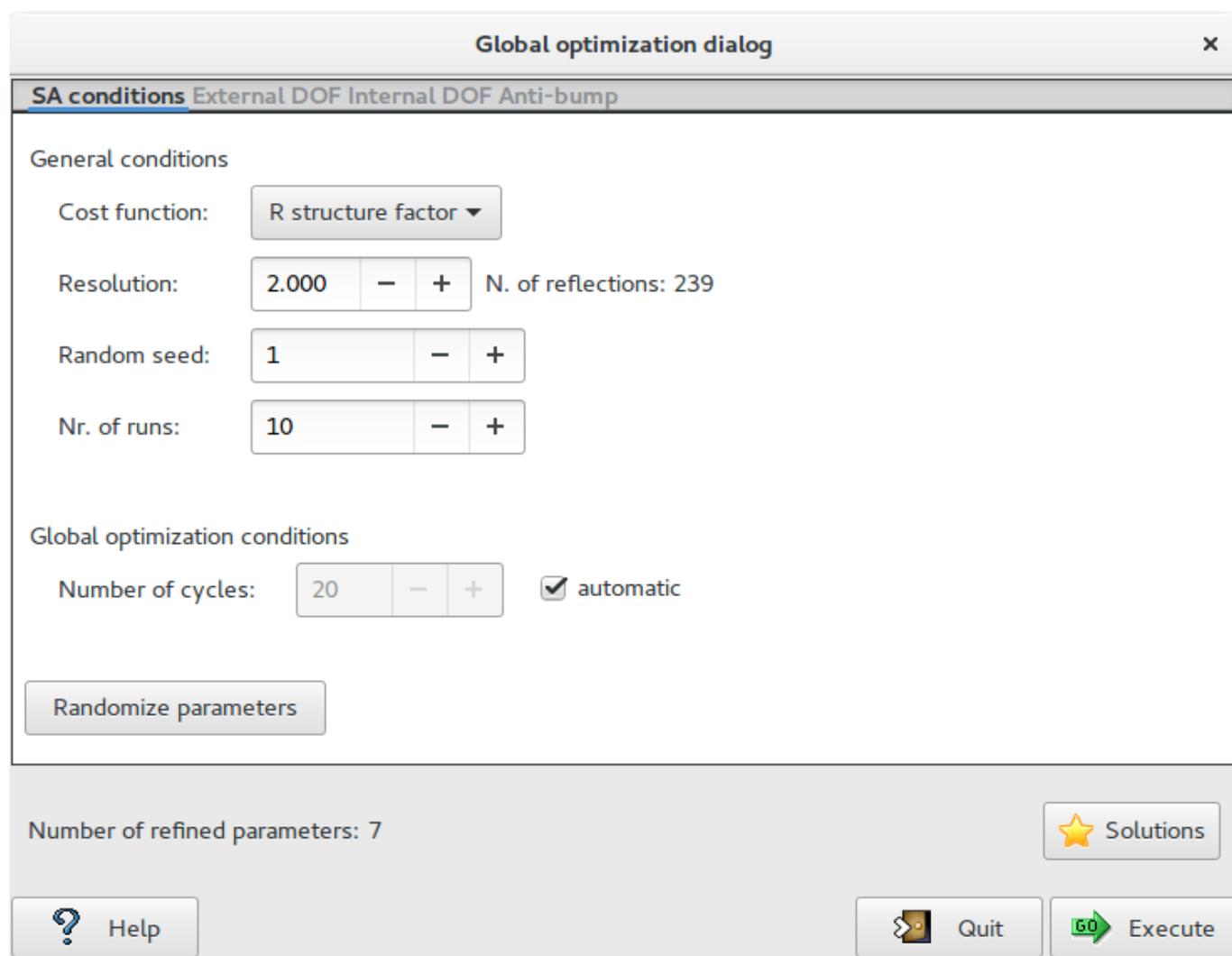
3D starting models. Suggested free available programs for building 3D models are [ACD/ChemSketch](#) (windows), [Avogadro](#) (all platforms), [Marvin](#) (all platforms), [ghemical](#) (linux). [NIST Chemistry webbook](#) and [Drugbank](#) are useful links to look up fragments.

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 Annealing in Sir2019 is composed by 4 pages (tabs).

The first page ‘SA conditions’ contains the general settings of the Simulated Annealing algorithm:



The screenshot shows the 'Global optimization dialog' window with the 'SA conditions' tab selected. The window has a title bar with a close button (X) and a tab bar with four tabs: 'SA conditions', 'External DOF', 'Internal DOF', and 'Anti-bump'. The 'SA conditions' tab is active and contains the following settings:

- General conditions**
 - Cost function: R structure factor (dropdown menu)
 - Resolution: 2.000 (input field with minus and plus buttons) N. of reflections: 239
 - Random seed: 1 (input field with minus and plus buttons)
 - Nr. of runs: 10 (input field with minus and plus buttons)
- Global optimization conditions**
 - Number of cycles: 20 (input field with minus and plus buttons) automatic

At the bottom of the dialog, there is a 'Randomize parameters' button. The status bar at the bottom shows 'Number of refined parameters: 7' and three buttons: 'Help' (with a question mark icon), 'Quit' (with a window icon), and 'Execute' (with a green arrow icon). There is also a 'Solutions' button with a star icon.

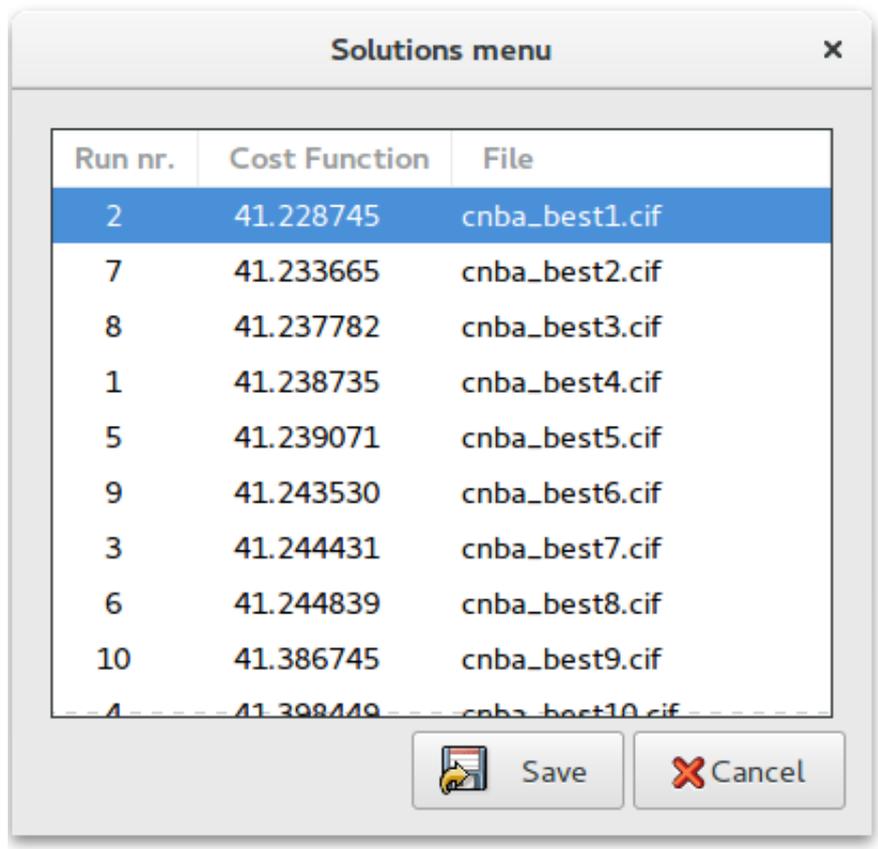
Cost function: 2 cost functions can be selected: R structure factor (default cost function), R intensities.

Resolution: define the maximum resolution used by Simulated Annealing procedure. Usually the algorithm works well with data up to 2 Å resolution.

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.



Run nr.	Cost Function	File
2	41.228745	cnba_best1.cif
7	41.233665	cnba_best2.cif
8	41.237782	cnba_best3.cif
1	41.238735	cnba_best4.cif
5	41.239071	cnba_best5.cif
9	41.243530	cnba_best6.cif
3	41.244431	cnba_best7.cif
6	41.244839	cnba_best8.cif
10	41.386745	cnba_best9.cif
4	41.398449	cnba_best10.cif

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

Global optimization dialog ✕

SA conditions **External DOF** Internal DOF Anti-bump

Select fragment Atoms

Fragment n. 1 ▾ c1 c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12 c13 c14 c15 c16 c17 c18 c19 c20 c21 c22
c23 c24 c25 c26 c27 c28 c29 n1

Parameters for fragment n. 1

	Refine	Lower	Upper
X Traslation	<input checked="" type="checkbox"/>	0.000 - +	1.000 - +
Y Traslation	<input checked="" type="checkbox"/>	0.000 - +	1.000 - +
Z Traslation	<input checked="" type="checkbox"/>	0.000 - +	1.000 - +
Rotation	<input checked="" type="checkbox"/>	0.000 - +	1.000 - +

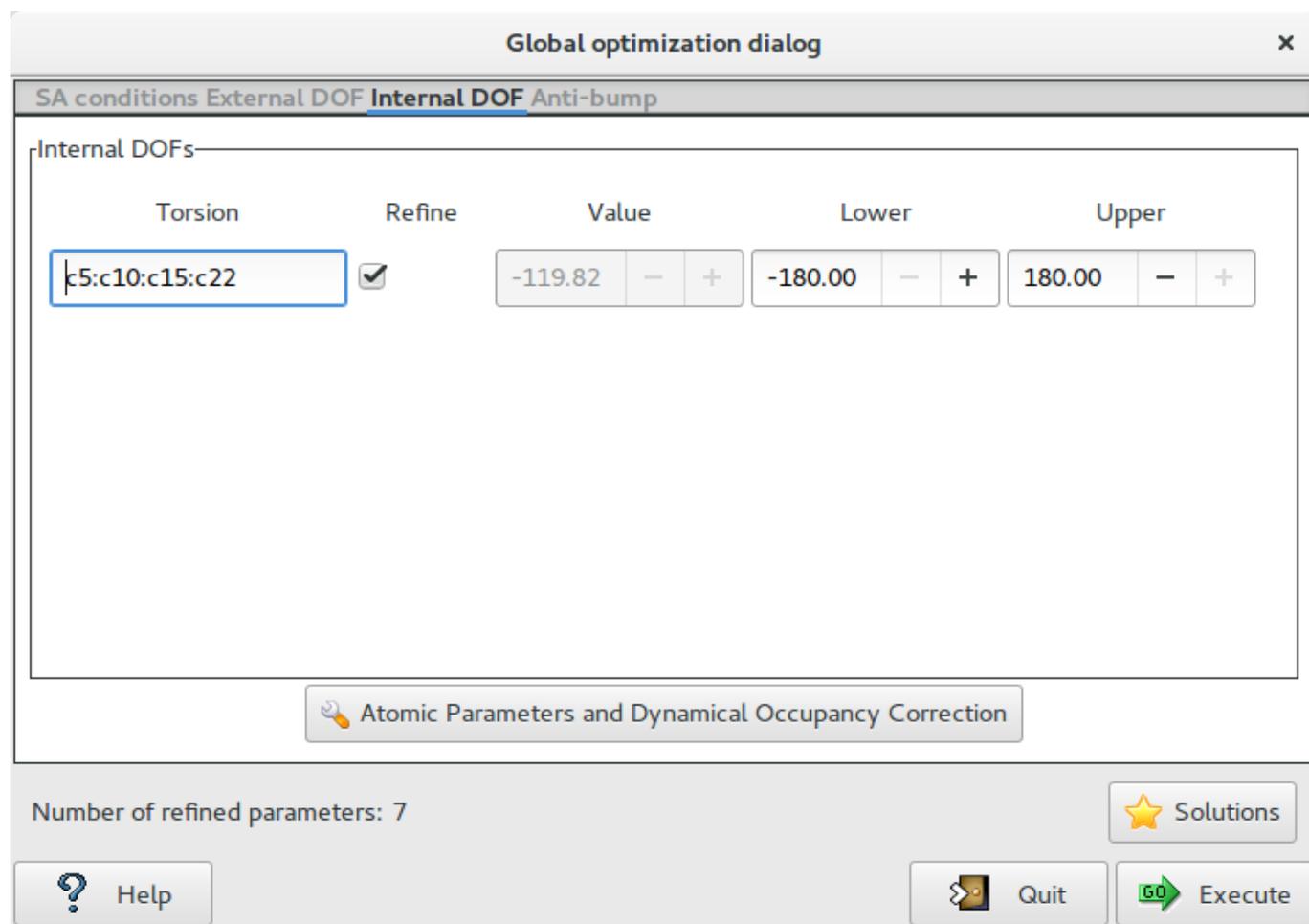
Number of refined parameters: 7 ★ Solutions

? Help ▶ Quit GO Execute

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

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 and Dynamical Occupancy Correction](#)” button



give access to the following dialog

Atomic parameters refinement ✕

Atom	Occ.	B[iso]	<input type="checkbox"/> Refine B[iso]	<input type="checkbox"/> Shift on xyz	<input type="checkbox"/> D.O.C.
c1	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c2	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c3	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c4	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c5	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c6	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c7	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c8	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c9	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c10	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c11	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c12	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Maximum shift on position - +

✕ Cancel
✔ OK

Thermal parameters: check the buttons on column 'B[iso]' to refine the thermal parameters of atoms. Click on label 'B[iso]' to select all checks in the column. If the thermal parameters are not 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. The refinement of the thermal parameters is discouraged because usually doesn't improve the results.

X Y Z: check the buttons on column 'Shift on xyz' to refine the atom positions, in this case the fragment is not rigid but the atom positions are shifted respect to the barycentre. Click on label 'Shift on xyz' to select all checks in the column. The positions refinement (x,y,z) is generally discouraged because can considerably increase the time to reach the global minimum.

Maximum shift on position: enter the value of the maximum shift of the atomic positions. 0.020 is the suggested value. Increasing this value the explored parameter space becomes wider so increasing the probability of falling into a local minimum.

Dynamical Occupancy Correction (DOC): 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). DOC will be applied only on the atoms selected in the column D.O.C., select all atoms if you don't know exactly which atoms will fall into special positions. DOC slows down the computation time so it should be avoid if non special position or shared atoms are expected

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

Global optimization dialog ✕

SA conditions External DOF Internal DOF Anti-bump

None
 Only for C,N,O,S
 Select from list

Weight
 Scale factor on distances

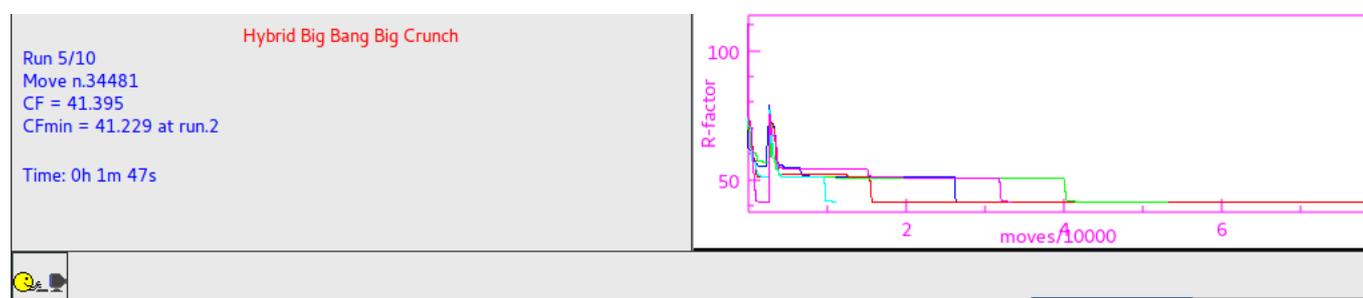
List of anti-bump restraints

Atom 1	Atom 2	Distance	Activate
C	C	3.400	<input type="checkbox"/>
C	N	3.250	<input type="checkbox"/>
N	N	3.100	<input type="checkbox"/>

Number of refined parameters: 7 ★ Solutions

? Help
Quit
GO Execute

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.

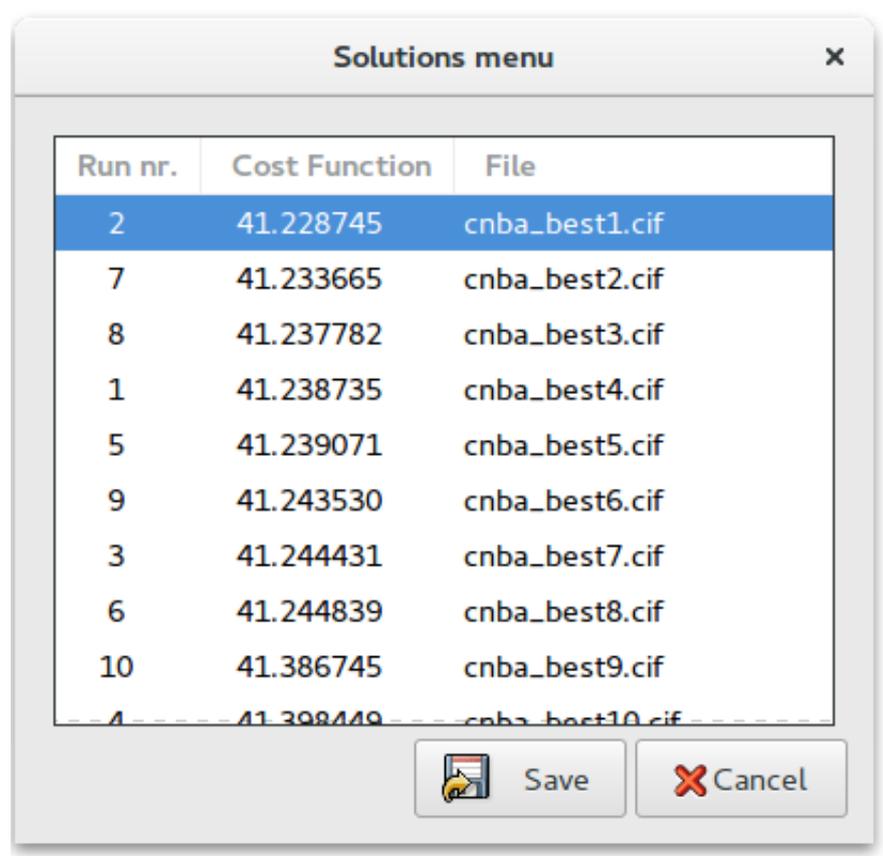


Output of Simulated Annealing

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.



Run nr.	Cost Function	File
2	41.228745	cnba_best1.cif
7	41.233665	cnba_best2.cif
8	41.237782	cnba_best3.cif
1	41.238735	cnba_best4.cif
5	41.239071	cnba_best5.cif
9	41.243530	cnba_best6.cif
3	41.244431	cnba_best7.cif
6	41.244839	cnba_best8.cif
10	41.386745	cnba_best9.cif
4	41.398449	cnba_best10.cif

Possible reasons for a Simulated Annealing failure

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 Å³/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 the user does not need to know the directives unless he is 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

[Previous page](#)

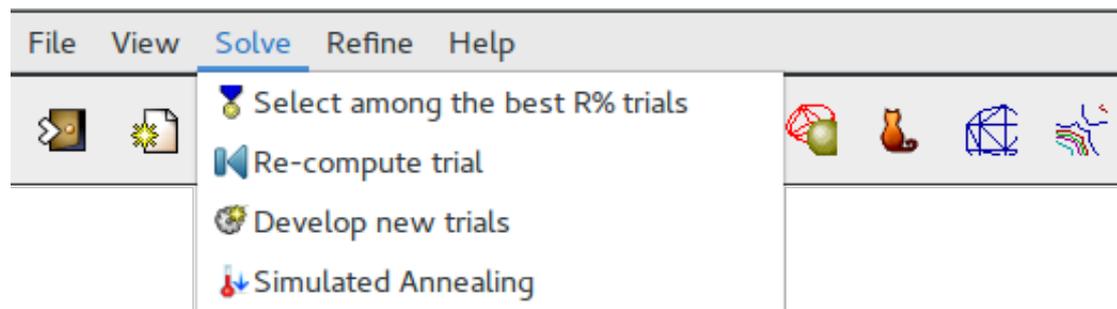
[Next page](#)

Solve

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/solve/>

Sub-menu Solve:



It allows:



[Select among the best R% trials](#)



[Re-compute trial](#)



[Develop new trials](#)

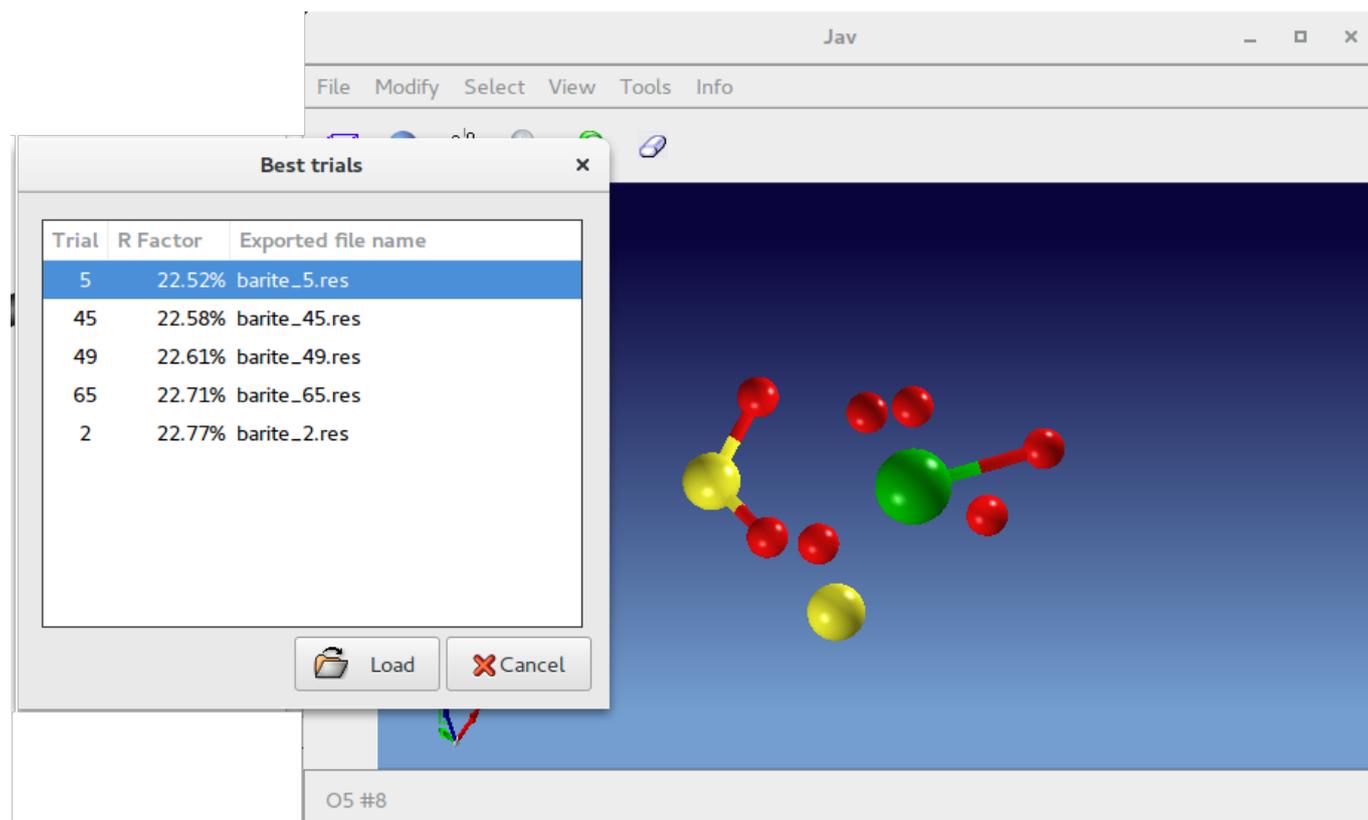


[Simulated Annealing](#)

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.



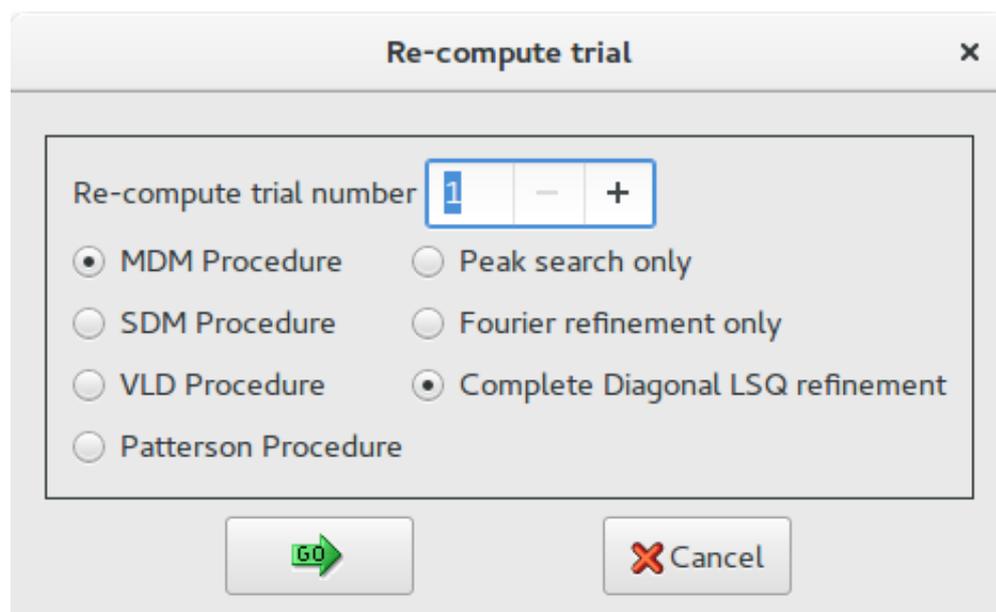
The screenshot displays the Sir2019 software interface. A 'Best trials' dialog box is open, showing a table of trial results. The table has three columns: 'Trial', 'R Factor', and 'Exported file name'. The data is as follows:

Trial	R Factor	Exported file name
5	22.52%	barite_5.res
45	22.58%	barite_45.res
49	22.61%	barite_49.res
65	22.71%	barite_65.res
2	22.77%	barite_2.res

Below the table are 'Load' and 'Cancel' buttons. In the background, a 3D molecular model is visible, showing a complex structure with red, yellow, and green spheres. The window title is 'Jav' and the status bar shows 'O5 #8'.

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 Sir2019, 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.

Define the Restart Strategy ✕

Solving strategy

Let the program to decide

Modern Direct Methods (MDM)

Standard Direct Methods (SDM)

Patterson

VLD

Reset Starting Trial

Size:

Let the program to decide ▾

6-80 non-H atoms in asymmetric unit

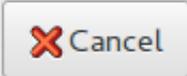
Parameters

Trials to generate

100 - +

Use default values

Stop when R% is less than 25 - +

Simulated Annealing

The dialog of Simulated Annealing in Sir2019 is described [here](#).

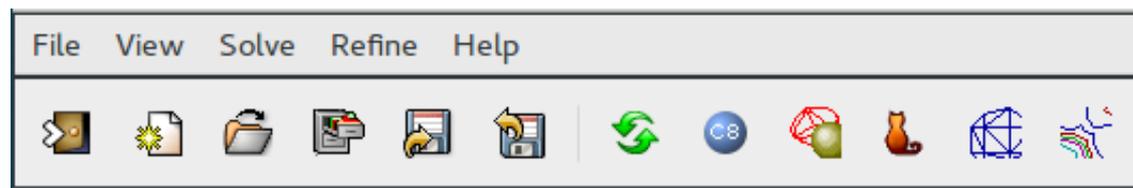
[Previous page](#) [Next page](#)

The Graphic Interface

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/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 toolbar.



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.](#)
-  [Reset the view for the model.](#)
-  [Show/Hide atom labels.](#)
-  [Preferred representation for the model.](#)
-  [Start Jav loading the current model and the related map.](#)
-  [3D map viewer.](#)
-  [2D contour facility.](#)

The menu items follow:

[File](#)

[View](#)

[Solve](#)

[Refine](#)

- [Help](#)

[Previous page](#)

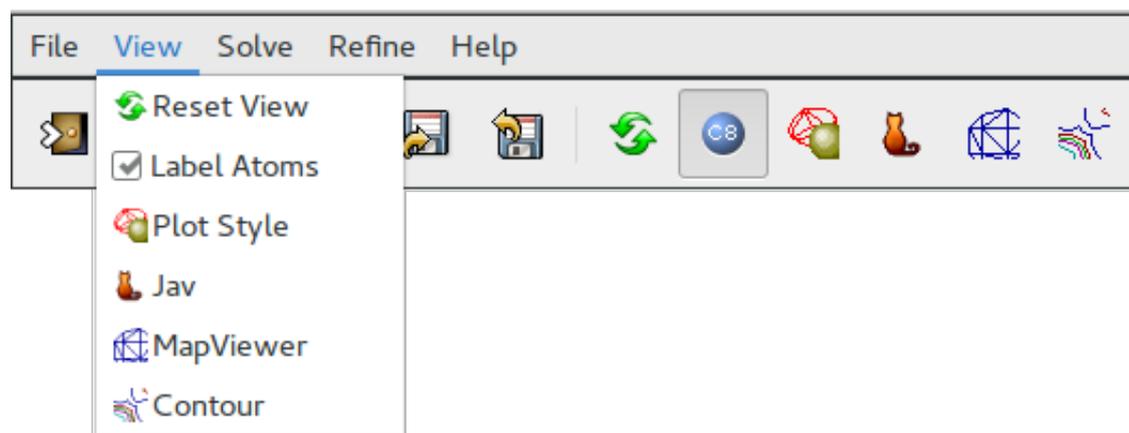
[Next page](#)

View

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/view/>

Sub-menu View:



These features are directly accessible through the toolbar

It allows:

 [Reset View](#)

[Label atoms](#)

 [Plot Style](#)

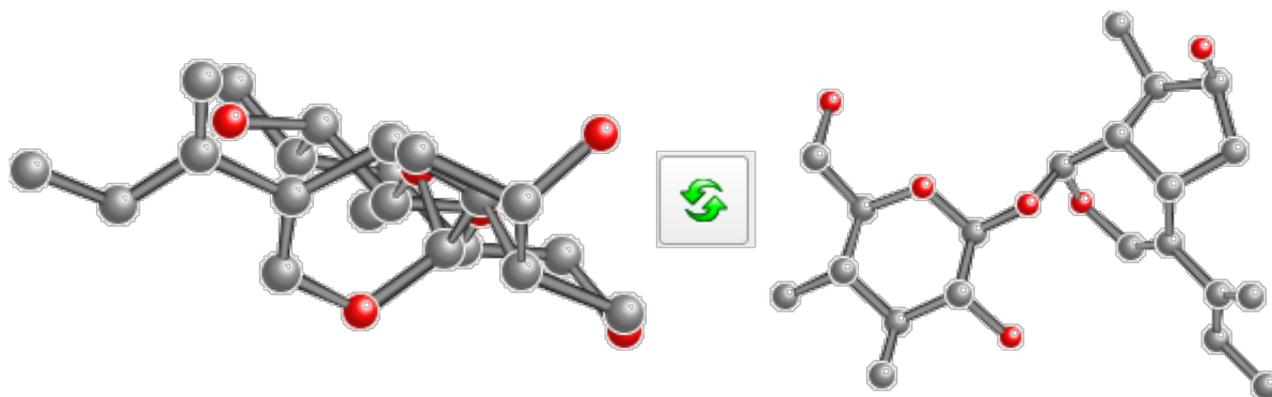
 [Jav](#)

 [Map Viewer](#)

 [Contour](#)

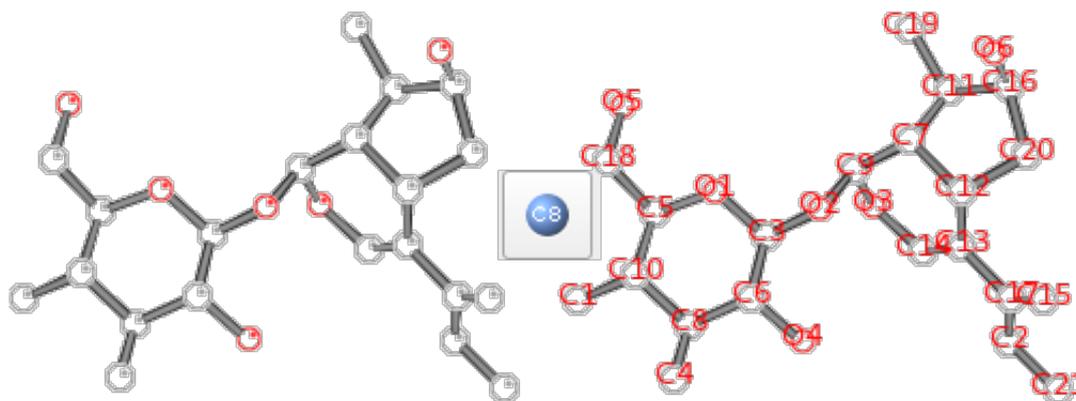
Reset

The structure is projected on LSQ plane and rescaled in order to fit the drawing area.



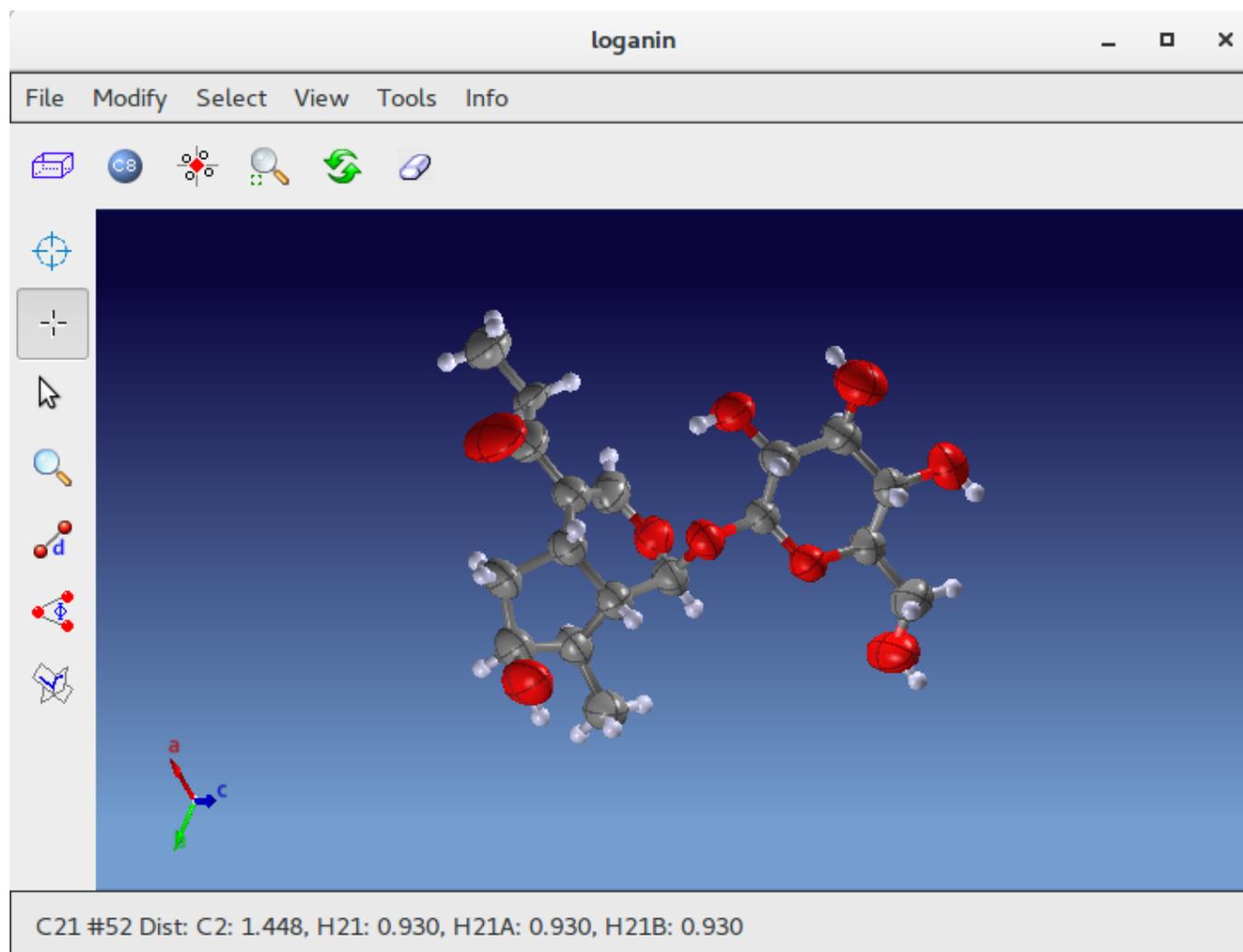
Labels

Show/Hide atom labels.



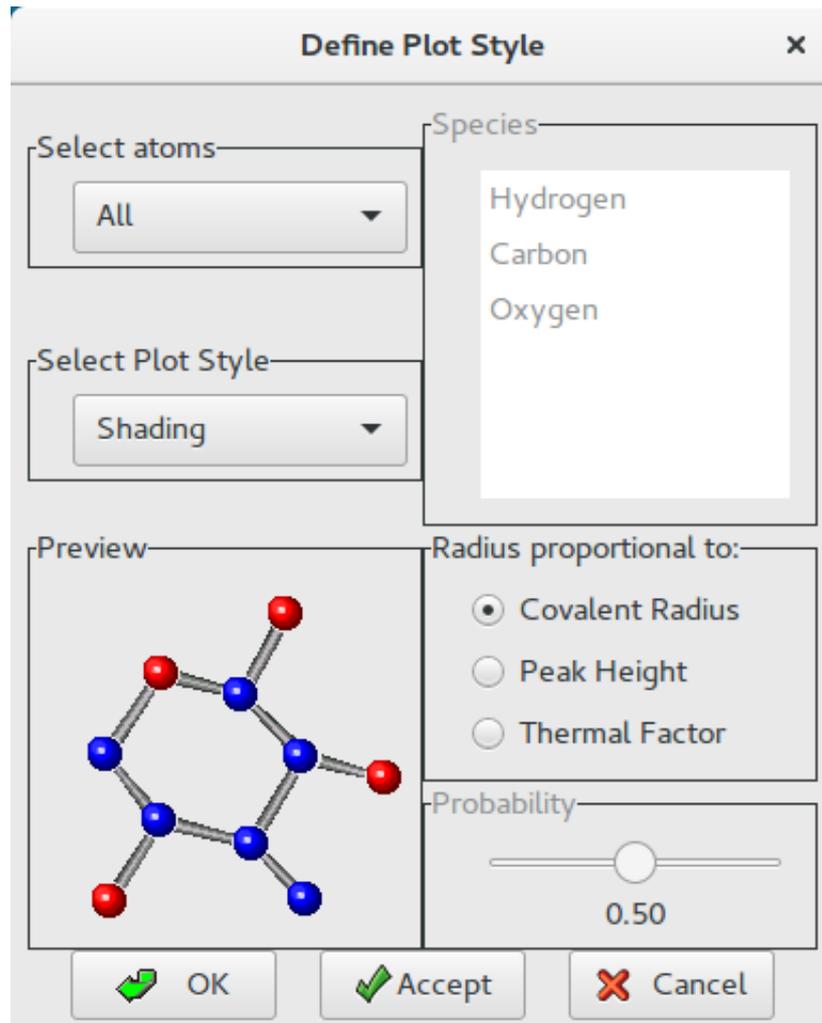
Jav

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



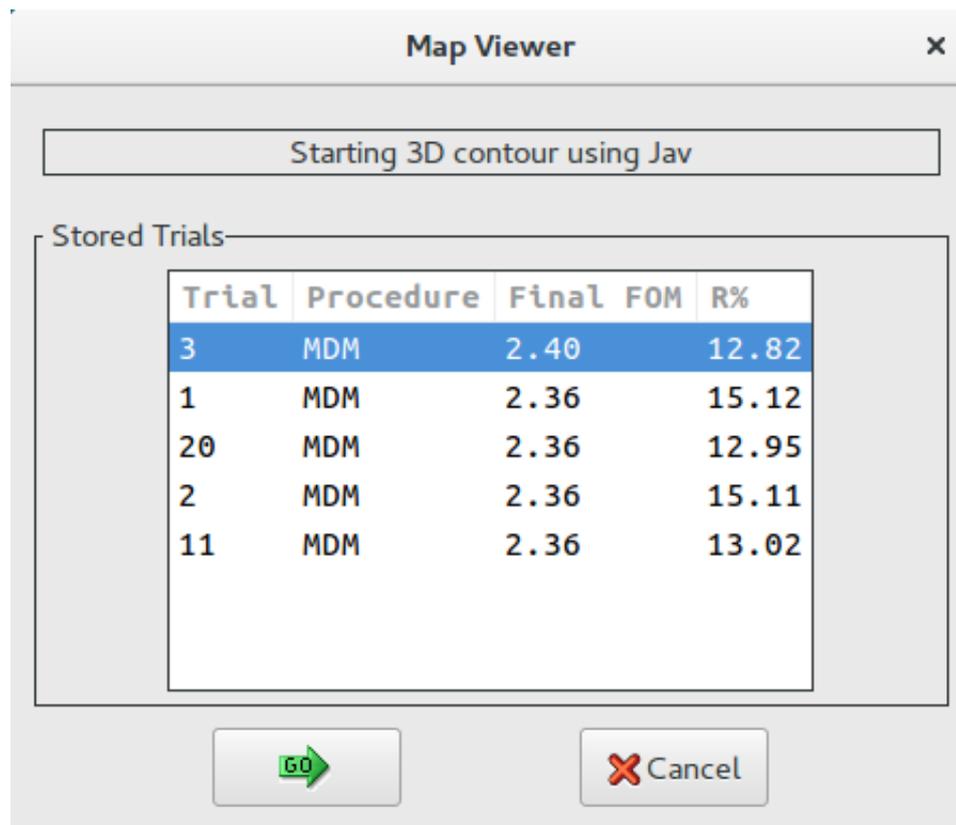
Plot style

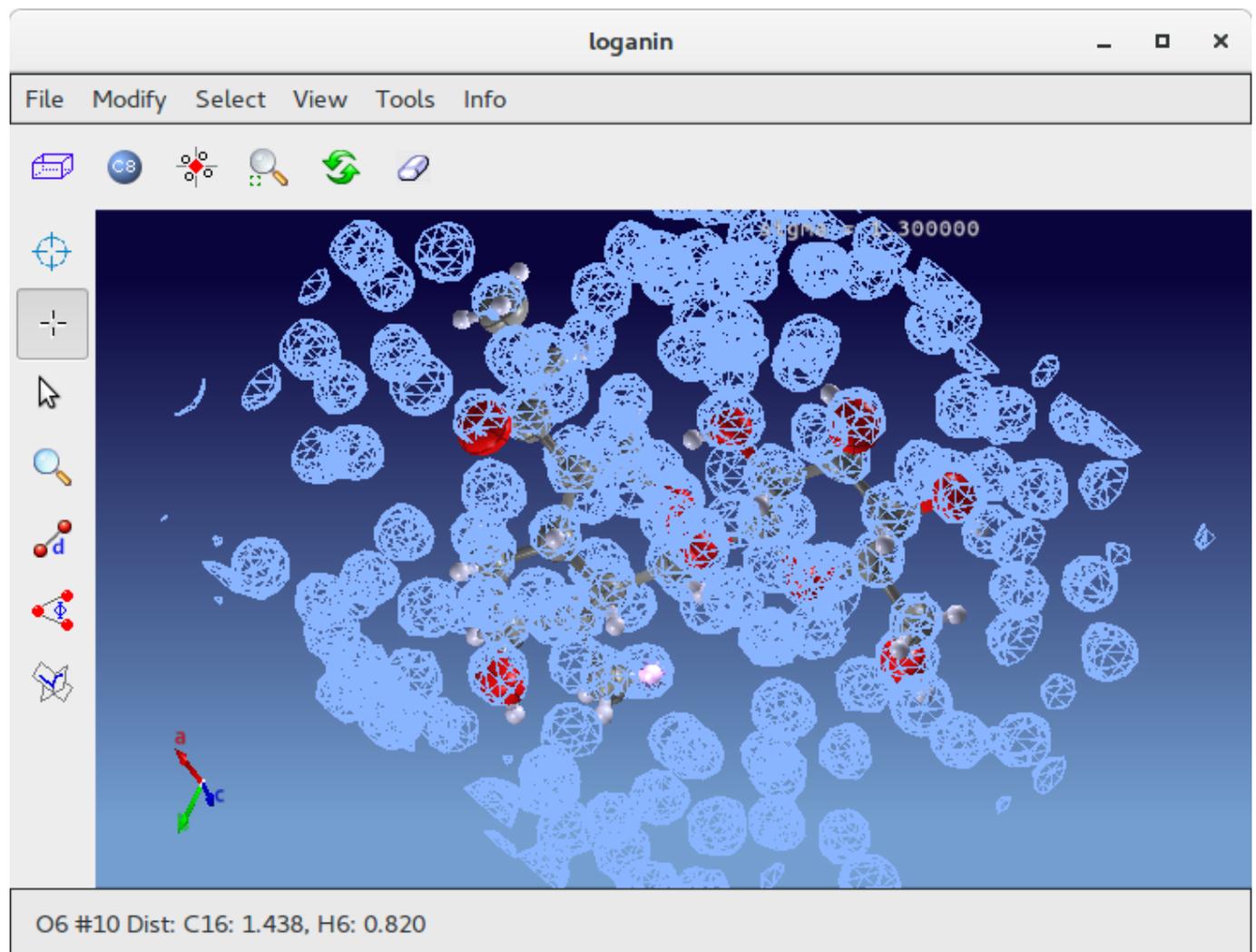
The user can select the preferred representation for all or for selected atomic species. The possible plot styles are: Sticks, Rods, Simple, Sticks & Balls, Bubbles, Shading, Van der Waals, Ellipsoids, Polyhedra. The effect is checked in the Preview area. It is possible also to draw atoms using different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



Map Viewer

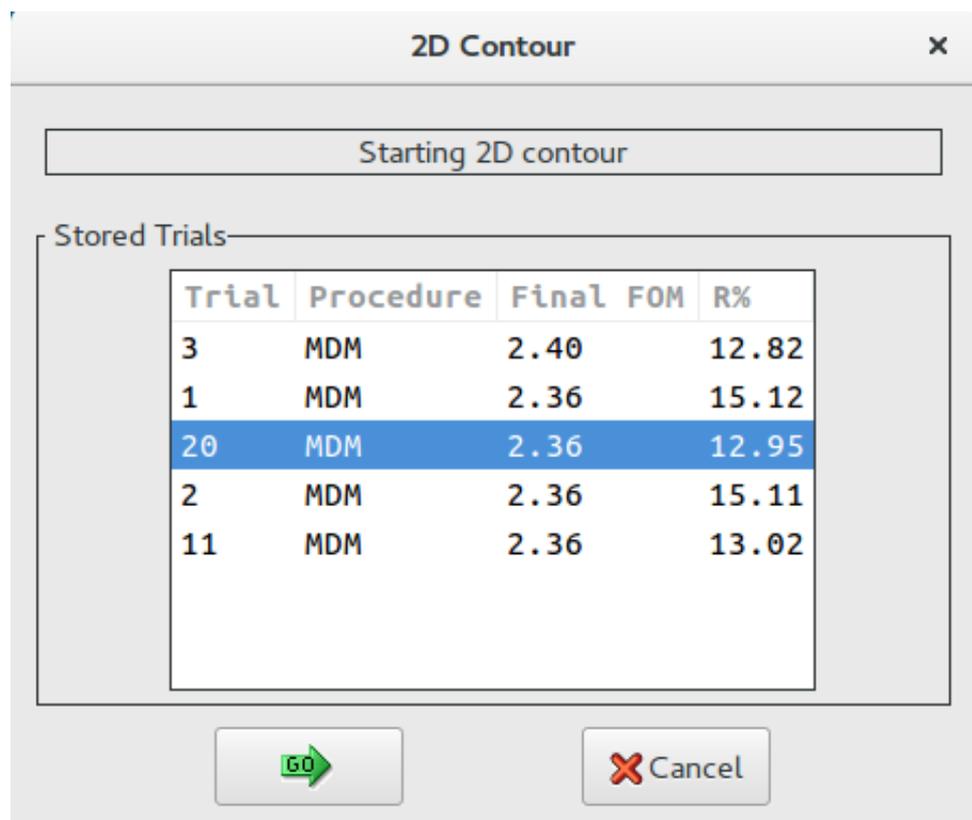
Start Jav on the map computed using the phases related to the selected trial. If a model is available, it will be included in the plot. For details see the help on-line of Jav.





Contour

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



Different possibilities are accessible using the icons on the toolbar:



 Manage contouring levels

 Set limit & graphic resolution

 Show / Hide atomic model

 Show / Hide unit cell

  Magnify / un-magnify the plot

 Let the plot fit the drawing window.

   Project plot on the plane normal to a, b, c

  Plot previous / next section(s)



Plot more than one section per time



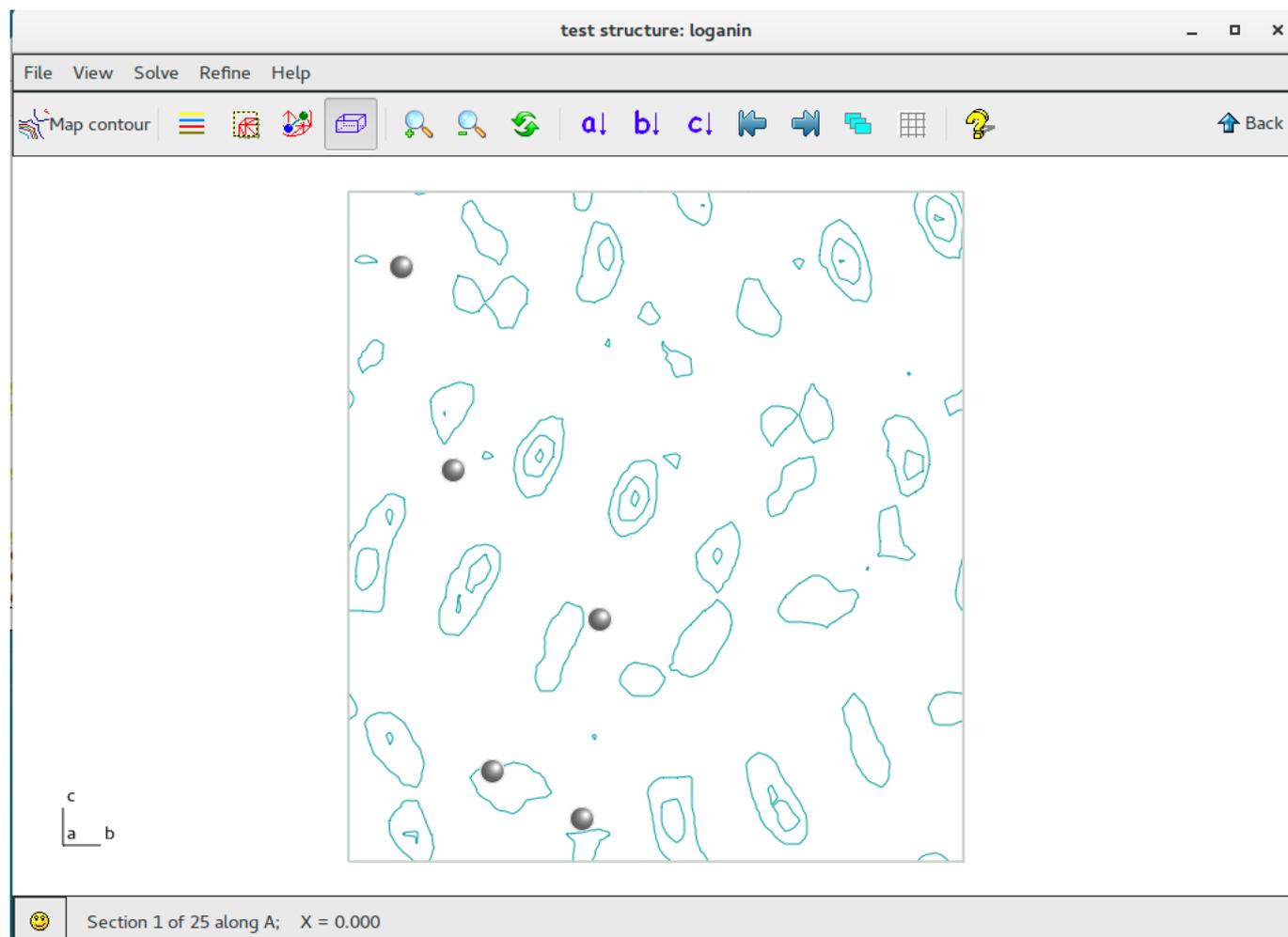
Plot more than one section per time



Credits



Leave 2D contour interface



[Previous page](#)

[Next page](#)

When default Sir2019 fails: some advices

by sir - Wednesday, January 24, 2018

<http://www.ba.ic.cnr.it/softwareic/sir/when-default-sir2014-fails-some-advices/>

Sir2019 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 (N_{large}) actively used in the MDM or SDM phasing process is fixed by the program. For some structures, the ratio “number of active triplets/ N_{large} ” is too small (e.g., less than 15). Larger values of N_{large} (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 R(F) values ($R_{\text{merge}} > 25\text{-}30\%$), 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.

[Previous page](#)

[Next page](#)

Sir2019

Crystal structures solution by single crystal data

PDF generated March 04, 2019 at 3:31 PM by Kalin's PDF Creation Station WordPress plugin