Expo2014

Crystal structures solution by powder diffraction data

PDF generated May 30, 2024 by Kalin's PDF Creation Station WordPress plugin

Table Of Contents

EXPO2014 documentation

http://www.ba.ic.cnr.it/softwareic/expo/

EXPO2014 [A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni, R. Rizzi, N. Corriero and A. Falcicchio, (2013). *J. Appl. Cryst.* **46**, 1231-1235], the latest version of EXPO2009 [A. Altomare, M. Camalli, C. Cuocci, C. Giacovazzo, A. Moliterni, R. Rizzi, (2009). *J. Appl. Cryst.* **42**, 1197-1202.] is able to perform all the steps of the structure solution process by powder diffraction data: indexing, space group determination, estimation of the reflection integrated intensities, structure solution by Direct Methods or/and by direct space/hybrid approaches, model refinement by Rietveld technique.

Authors

A. Altomare*, C. Cuocci*, C. Giacovazzo*, A. Moliterni*, R. Rizzi*, N. Corriero* and A. Falcicchio*

*Institute of Crystallography-CNR, via Amendola 122/o 70126 Bari, Italy

angela.altomare@ic.cnr.it
corrado.cuocci@ic.cnr.it
carmelo.giacovazzo@ic.cnr.it
annagrazia.moliterni@ic.cnr.it
rosanna.rizzi@ic.cnr.it
nicola.corriero@ic.cnr.it
aurelia.falcicchio@ic.cnr.it

Contact

For questions, problems and suggestions, please, visit http://www.ba.ic.cnr.it/softwareic/expo/contact-us/

We kindly ask you to fill the survey.

Main Reference

If EXPO2014 is used in the solution process of a structure, whose results are published in a scientific paper, please, include the following reference:

A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni, R. Rizzi, N. Corriero and A. Falcicchio, (2013). J. Appl. Cryst. **46**, 1231-1235.

_____

General information

http://www.ba.ic.cnr.it/softwareic/expo/general-information/

EXPO2014 is able to perform all the steps of the structure solution process by powder diffraction data: indexing, space group determination, estimation of the reflection integrated intensities, structure solution by Direct Methods or/and by direct space/hybrid approaches, model refinement by Rietveld technique.

The program has been designed to:

- require a minimal information as input;
- work automatically;
- reduce the user intervention and facilitate the interaction by means of a user-friendly graphic interface.

<div align="center">

Program specification

</div>

 X-ray, synchrotron or neutron diffraction data can be used. There are no limits on the number of counts in the pattern.

<div align="center">

The background of system

</div>

 EXPO2014 is the updated version of EXPO2009. Its main steps are (please, select each topic for more details):

Indexing
  An automatic peak search routine generates a list of "d" values corresponding to the experimental diffraction peaks. The values are processed for providing the cell parameters.

Space Group Determination
A statistical analysis of the experimental diffraction intensities compatible with the identified crystal system is performed. This step is able to provide a list of all the possible extinction symbols ranked according to calculated probability values. For each extinction symbol, the corresponding list of compatible space groups (one or more than one) is supplied via graphic interface.

Structure Solution
Structure solution can be carried out by adopting one of the two alternative following approaches:

- Direct Methods for phasing the reflections. The electron density map is calculated and then optimized.
- Direct Space methods for generating trial structure models (compatible with the expected molecular geometry) and selecting that one providing the best agreement with the experimental data.

Rietveld Refinement

The structure model is refined by the Rietveld method.

Each procedure of EXPO2014 can be activated by using a specified command or directive in the input file, or by selecting a specific graphic option. The information about commands, directives and graphic options are contained in this manual.

EXPO2014 creates the following files (in the work directory):
 – a direct access file to preserve data and results (*name.bin*);

– an ascii file containing the Miller indices (*hkl*) and the experimental structure factor moduli of the generated reflections (*name.hkl*);

– an ascii file which stores diffraction peak positions if the indexing process has been carried out (*name.pea*);

– an output file containing information about all the steps of the structure determination process (*name.out*).

Other files associated to the application of particular strategies can be also generated by EXPO2014.

Authors

A. Altomare*, C. Cuocci*, C. Giacovazzo*, A. Moliterni*, R. Rizzi*, N. Corriero* and A. Falcicchio*

*Institute of Crystallography-CNR, via Amendola 122/o 70126 Bari, Italy

angela.altomare@ic.cnr.it
corrado.cuocci@ic.cnr.it
carmelo.giacovazzo@ic.cnr.it
annagrazia.moliterni@ic.cnr.it
rosanna.rizzi@ic.cnr.it
nicola.corriero@ic.cnr.it
aurelia.falcicchio@ic.cnr.it

Contact

For questions, problems and suggestions, please, visit http://www.ba.ic.cnr.it/softwareic/expo/contact-us/

We kindly ask you to fill the survey.

Acknowledgements

The authors are indebted to Giovanni Luca Cascarano and Mariarosaria Mallamo for their help in the development of the graphic environment.

Main References

If EXPO2014 is used in the solution process of a structure, whose results are published in a scientific paper, please, include the following reference:

A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni, R. Rizzi, N. Corriero and A. Falcicchio, (2013). *J. Appl. Cryst.* 46, 1231-1235 .

_____

Tutorials

http://www.ba.ic.cnr.it/softwareic/expo/tutorials/

Tutorial 1 – Default crystal structure determination of cimetidine compound by direct methods

The first tutorial is supplied for carrying out the full pathway of the ab-initio crystal structure solution process, from indexing up to the structure solution by Direct Methods, in the case of cimetidine.
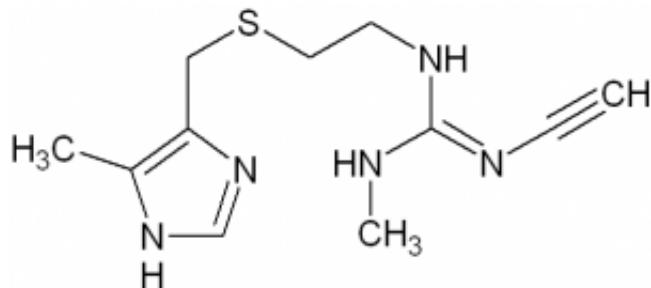
The only required information by EXPO is the experimental powder diffraction pattern and the chemical formula $C_{10}H_{16}N_6S$.

The examples directory contains the necessary files:

- `cime.exp`: input file for EXPO
- `cime.dat`: synchrotron X-ray diffraction data file containing the initial 2-theta, the 2-theta step and the final 2-theta values (in the first line) and the experimental diffraction counts (in the rest of lines).

To access to `examples` directory:

- Open EXPO by double clicking on Expo2 icon.
- Click on **File > Load Examples**
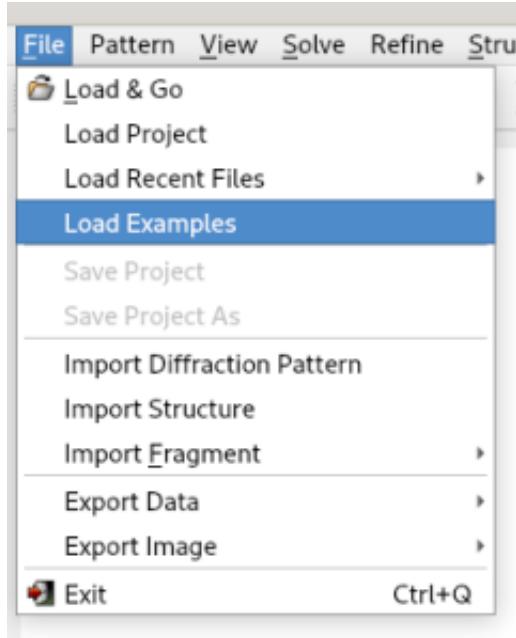
and select the file `cime.exp`.

The content of the file `cime.exp` follows.

```
%structure cime
%job cimetidine (C10H16N6S) -- Synchrotron data
%data
pattern cime.pow
wavelength 1.52904
synchrotron
%ntreor
%continue
```

It contains the command (%data) for providing the minimal required information (name of data file, wavelength and type of radiation used to collect data, if different from laboratory X-rays) and the commands for activating the indexing process (%ntreor) and the ab-initio structure solution (%continue). This file can be visualized and eventually modified by the 'Edit' button.
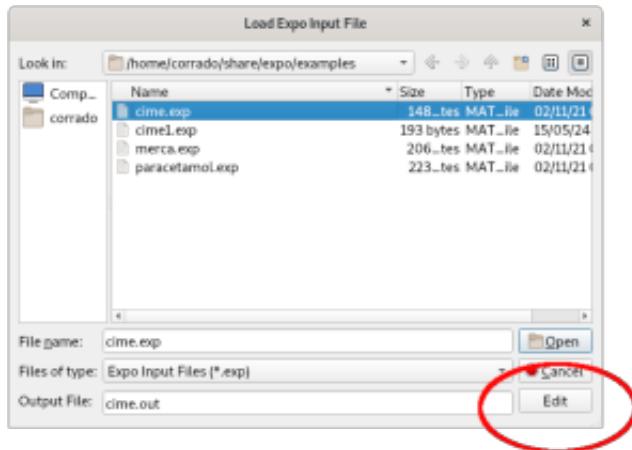
More information about the creation of input file can be found in the section Preparing input file.

Clicking on **OK**, the powder diffraction data is displayed.
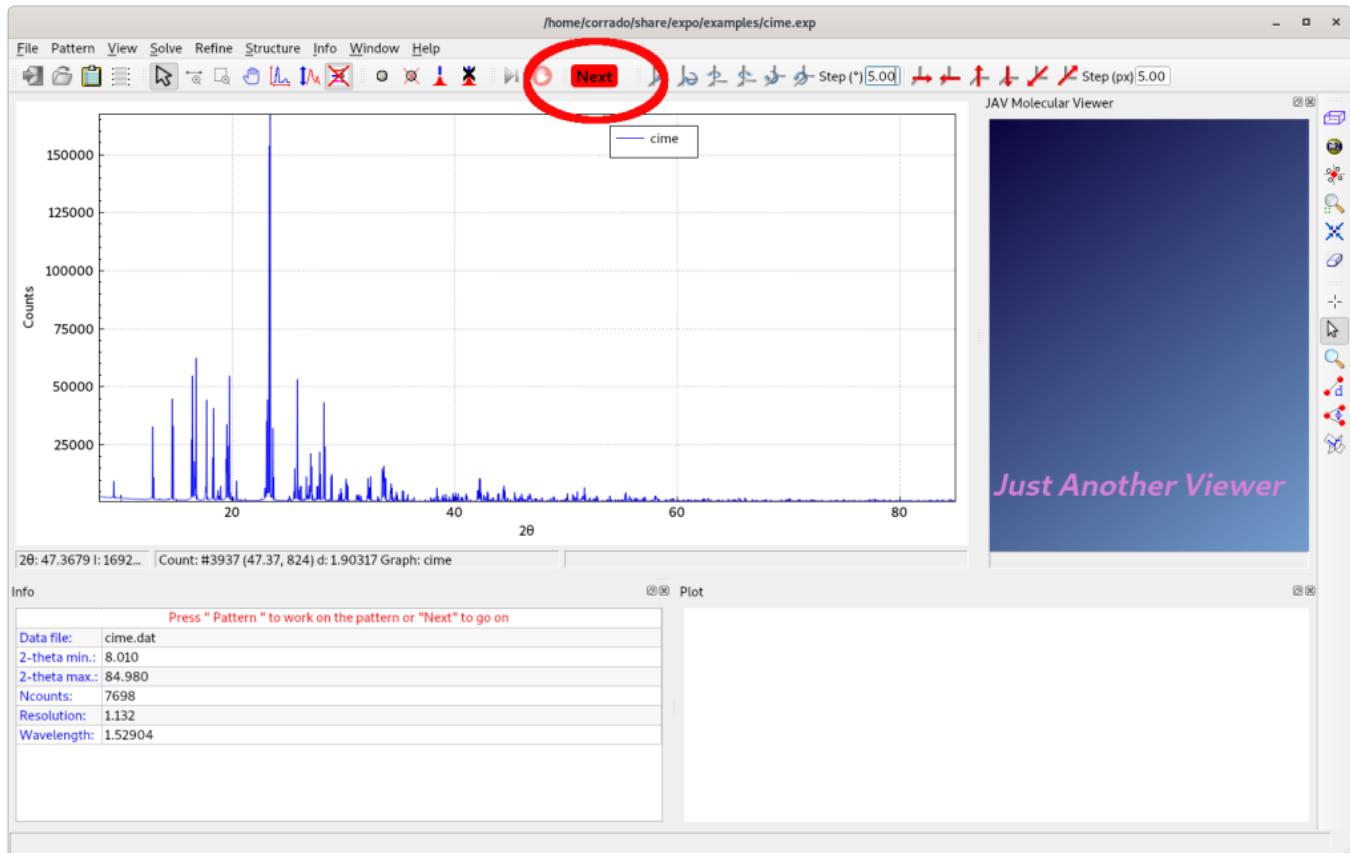
Selecting **Next**, an automatic peak search procedure is executed. The d-values corresponding to the detected experimental peaks are automatically supplied to the indexing procedure. At the end of this process, EXPO provides the indexing results by graphic interface:

Nr. ▲	Prog.	a	b	c	alpha	beta	gamma	Vol.	M20	FOMnew	McM20	Shift	NIX	Symmetry Info
1	N	10.39358	18.81631	6.82405	90.000	106.438	90.000	1280.02	110.00	6.08	-	0.000	0	P 1 21/a 1
2	N	10.40310	18.83935	6.83171	90.000	106.423	90.000	1284.30	21.00	2.34	-	0.036	0	P 1 21/a 1
3	N	10.38221	18.79205	6.81667	90.000	106.452	90.000	1275.50	18.00	1.96	-	-0.036	0	P 1 21/a 1

Click on **OK** to accept the selected plausible cell parameters and to move to the next step: the space group determination.

The space group determination step, as well as direct methods, requires the unit cell content information. It can be easily derived by taking into account the structure formula and the cell volume (Vcell) which is evaluated, by EXPO, from the cell parameters. (The cell volume is graphically provided and/or contained in the output file of EXPO).

The cell content is obtained by multiplying Z by the structure formula. You can calculate Z as the ratio between Vcell and the molecular volume (Vmol): Z = Vcell/Vmol.
Vmol can be calculated by one of the two following alternative ways:

1. Using the 18 Å^3 rule: counting all the non-hydrogen atoms in the chemical formula and multiplying this number by 18 Å^3;
2. Using the following approximate atomic volumes: 14 Å^3 for C, 12 Å^3 for N, 11 Å^3 for O, 5 Å^3 for H, 26 Å^3 for Cl, 25 Å^3 for S.

In case of cimetidine compound, the chemical formula is $C_{10}H_{16}N_6S$ and the cell volume is 1280 Å^3. If the first criterion is applied, the molecular volume is given by: Vmol = 18 x 17 = 306 Å^3. Using the second approach Vmol = 14 x 10 + 5 x 16 + 12 x 6 + 26 x 1 = 318 Å^3.

The corresponding Z value is: Z = Vcell/Vmol = 1280/306 = 4.18, that can be approximated to 4. So 4 molecules of Cimetidine are in the unit cell and the unit cell content to be supplied to EXPO by graphic interface is `(C10H16N6S)4` or `C40H64N24S4`.

Click on **OK** in order to perform the space group determination and then on the **Next** button to go on continuously. At the end of the procedure, a list of the space groups and extinction groups, compatible with the lattice symmetry and ranked according probability values, is given:

The most plausible extinction group $P \ 1 \ 21/a \ 1$ is automatically selected and by clicking on **OK**, the space group $P21/a$, corresponding to the chosen extinction group, is considered.

After the determination of the space group, EXPO passes to the extraction step, which is carried out by using the selected space group. Click on the **Next** button continuously. At the end of the process, a list of integrated intensity values, extracted from the experimental powder pattern, is obtained and automatically supplied to the Direct

[Methods process](#) for structure solution.

The final result is visualized on JAV Molecular Viewer.
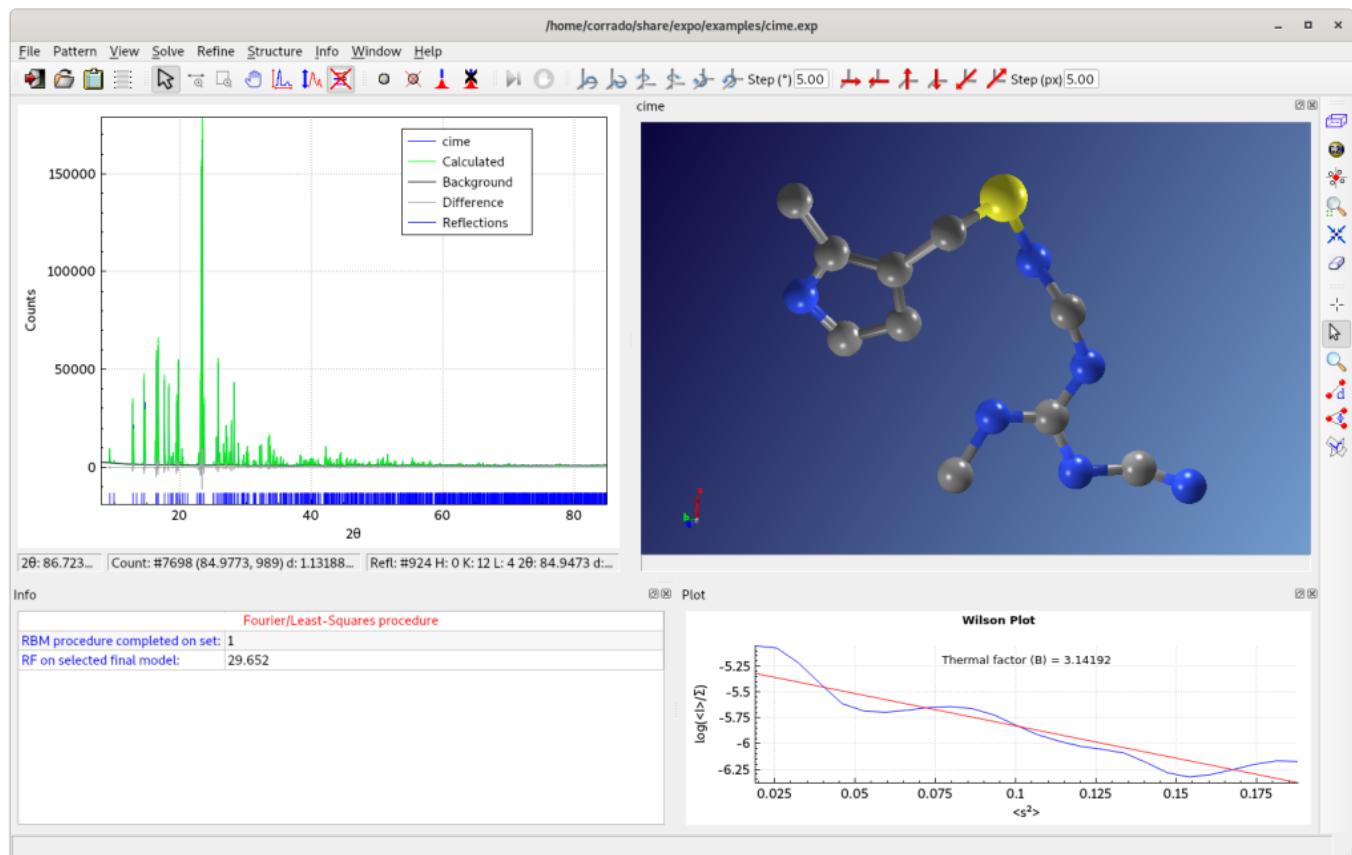
Some errors in the assignment of carbon or nitrogen atomic species can be easily recognized and corrected by using the graphic option 'Modify' from JAV menu.

Tutorial 2 – Non-default crystal structure solution of 2-mercaptobenzoic acid compound by direct methods
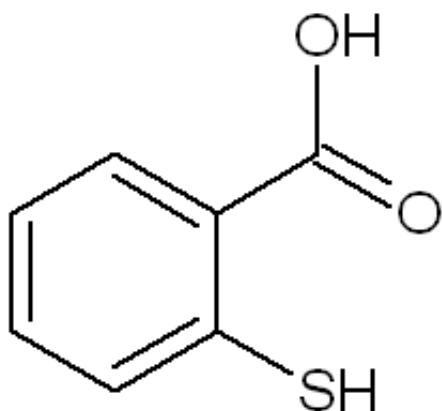
The aim of this example is to guide the user in case of failure of a default run of EXPO.

Open EXPO by double click on Expo icon.

Click on **File > Load Examples**

Select the input file `merca.exp`.

```
%structure merca
%job 2-Mercaptobenzoic acid
%data
cell        7.885  5.976   14.949  90.0  100.48  90
spacegroup p 21/c
content    (C 7 O 2 S 1 H 6)4
pattern    merca.xy
wavelength 1.54056
%continue
```

In this example, it is assumed that the cell and space group have been already determined and the unit cell content corresponds to 4 molecule of $C_7H_6O_2S$.

Click on the **Next** button to go on continuously.

The final structure model provided by EXPO could not correspond to the correct crystal structure of 2-Mercaptobenzoic acid (the model is not chemically interpreted). This can occur when the most plausible set of phases, provided by Direct Methods, is unreliable.

In a default run of EXPO, Direct Methods generate several phasing trials; not more than 20 of them are stored and ranked according to decreasing values of a suitable combined figure of merit (CFOM). The largest CFOM set of phases is automatically selected for structure solution. In case of failure, the procedure ALLTRIALS can be applied: all the rest of stored phasing trials are explored in order to find the correct solution.

This can be automatically performed by graphic interface:

a) Selecting from the top-level menu **Solve** and then **Explore Trials**

b) Clicking on the **Select all new trials** button

Explore Trials

Fourier Procedures

RBM

☐ Shift_and_Fix

☑ Explore trials non processed yet

Explore	Set	cfom	done	RF
☐	3	0.999	Yes	28.437
☐	14	0.981	Yes	29.625
☐	1	0.999	Yes	30.439
☐	13	0.981	Yes	37.928
☐	12	0.985	Yes	39.382
☐	20	0.957	Yes	40.946
☐	11	0.985	Yes	41.607
☐	10	0.985	Yes	42.805
☐	4	0.999	Yes	43.952
☐	16	0.974	Yes	45.250
☐	15	0.977	Yes	46.544
☐	9	0.990	Yes	46.550
☐	19	0.959	Yes	46.810
☐	5	0.998	Yes	48.642
☐	8	0.994	Yes	49.419
☐	7	0.997	Yes	49.769

🔵 Explore   ✖ Close

c) Clicking on **GO**

At the end of the procedure, EXPO ranks the phasing trials (and the corresponding structure models) according to the increasing R_F agreement factor value.

The best solution (corresponding to the lowest R_F value) is selected and displayed.

Tutorial 3 – Crystal structure solution of paracetamol compound by direct space methods

The purpose of this tutorial is to guide the user through the crystal structure solution process by Direct Space methods of paracetamol (form I polymorph) compound.
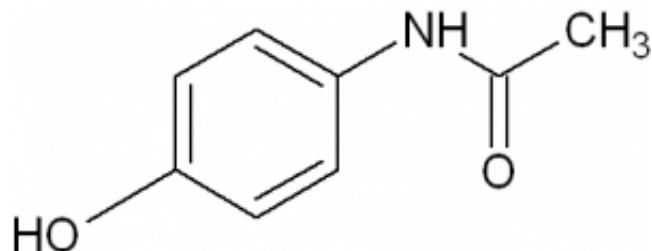
Open EXPO, select **Load Examples.**

Select the input file `paracetamol.exp`.

Clicking on the **Next** button, the file `paracetamol.mol`, containing a 3D molecular description of paracetamol, is loaded and displayed and the following window appears:
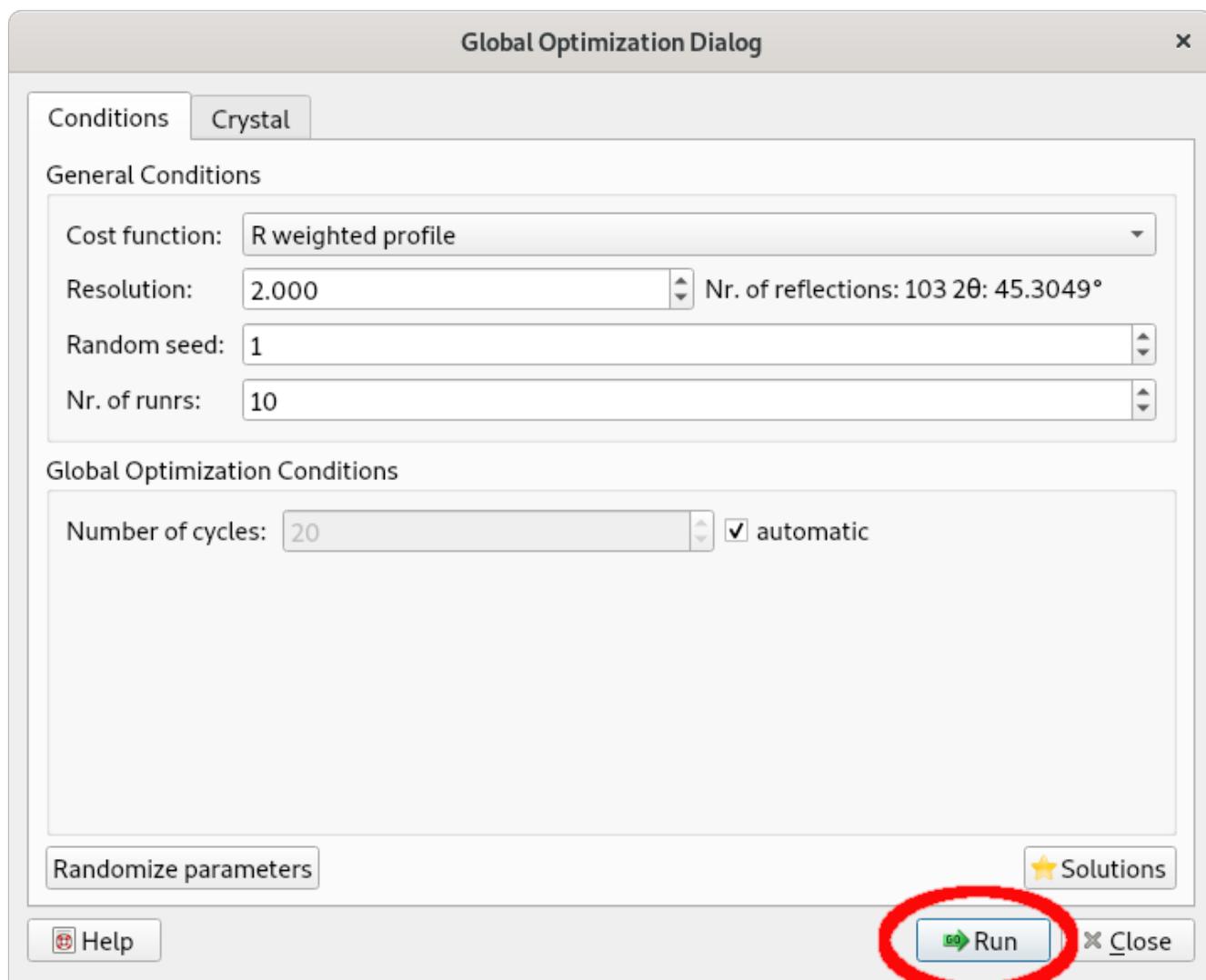
Select **Run**; the Simulated Annealing procedure is applied performing 10 runs whose corresponding final structure models can be explored by the button **Solutions.**
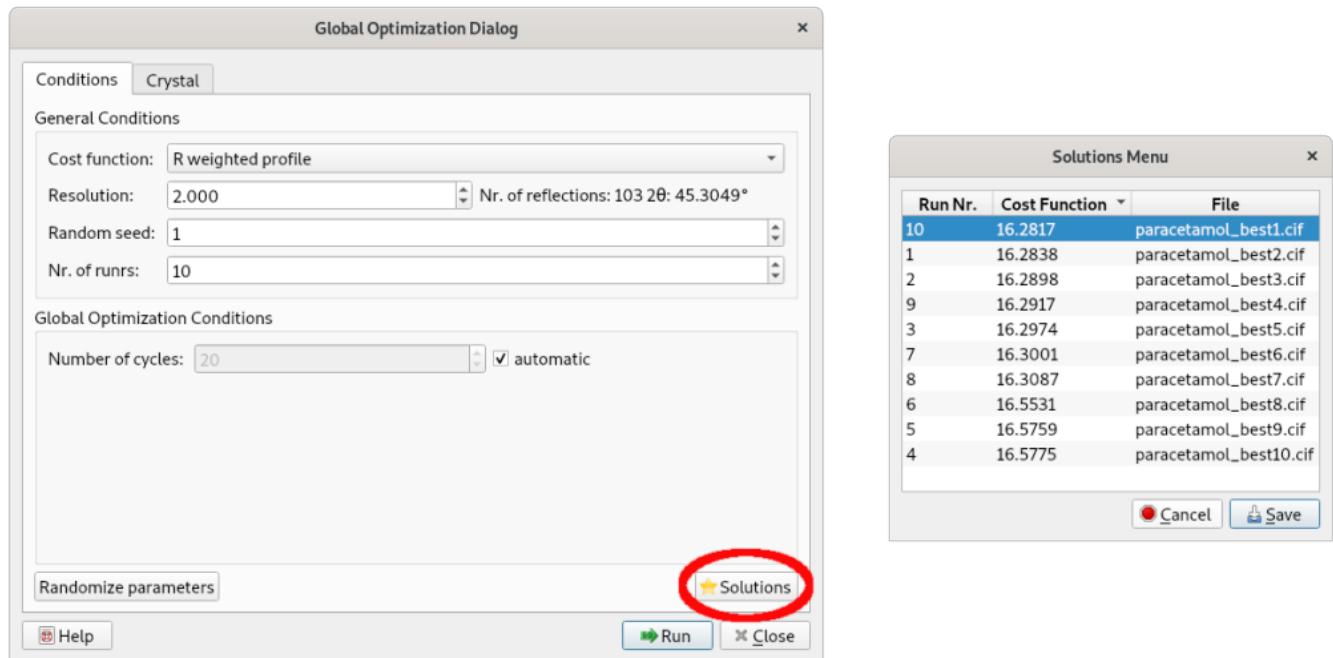
References

Cimetidine – Cernik, R.J., Cheetham, A.K., Prout, C.K., Watkin, D.J., Wilkinson, A.P., Willis, B.T.M. (1991). *J. Appl. Cryst.* **24,** 222-226.

2-Mercaptobenzoic acid – Steiner, T. (2000). *Acta Cryst.* C**56,** 876–877.

Paracetamol (form I polymorph) – Nichols, C. & Frampton, C. S. (1998). *J. Pharm. Sci.***87,** 684–693.

_____

Preparing input file

http://www.ba.ic.cnr.it/softwareic/expo/preparing-input-file/

Preparing input file for EXPO2014

To run EXPO2014 you need an input file (`*.exp`). Two possibilities are available: a) to create a new input file; b) to use an already existing input file.

a) Creation of a new input file by selecting, *via* graphic interface, the option of a new project by the menu **File > New**.
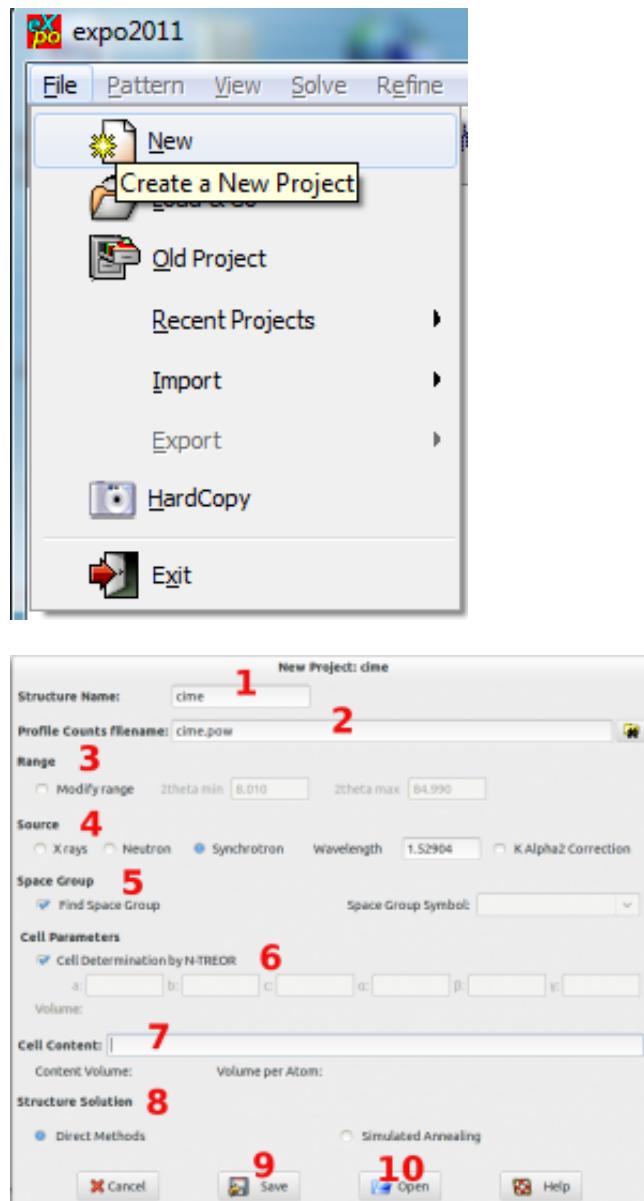
The required information are hereinafter listed. You have to

1. Specify the structure name. This name will be used to create the name of the output file (`name.out`);
2. Load the profile counts file. The following powder diffraction data ASCII files can be imported: **XY**

file (the filename extension should be `.xy`), **DAT file** (the filename extension should be `*.dat`), **CIF powder file** (the filename extension should be `*.rtv` or `*.cif`), **GSAS file** (the filename extension should be `*.gda`), **CCDC Mercury xye file** (the filename extension should be `*.xye`). The program recognizes automatically the file format by the filename extension. In case of filename extension different from the previous ones, the program tries to read the counts file according to the format of XY or DAT file. For more information about the formats of the counts files readable by the program, see the directive `pattern` of the command `%data`;

3. Select the 'Range' button if you want to change the 2? minimum and/or the 2? maximum;
4. Input the wavelength and the type of radiation source (laboratory X-rays source is the default choice);
5. Activate the check button 'Find Space Group' to carry out the space group determination;
6. Click on the 'Cell Determination by N-TREOR' button to carry out the indexation;
7. Provide the content of unit cell. This information is important for the space group determination and for the structure solution by Direct Methods (8). A valid molecular formula for the unit cell content must meet the criteria adopted for *string* in case of the directive `content` of the command `%data`.
   If your formula doesn't follow the required criteria, the input string will appear written in red.

When you press the button 'Save' (9), an input file for EXPO2014 (`name.exp`) is created in the same folder of the powder data file and automatically loaded by the indexing program. The button 'Open' (10) is used to open (and possibly modify with a text editor) an already existing input file.

b) An already existing `*.exp` input file (created by EXPO2014 or by a text editor) can be loaded from the menu **File > Load & Go**.
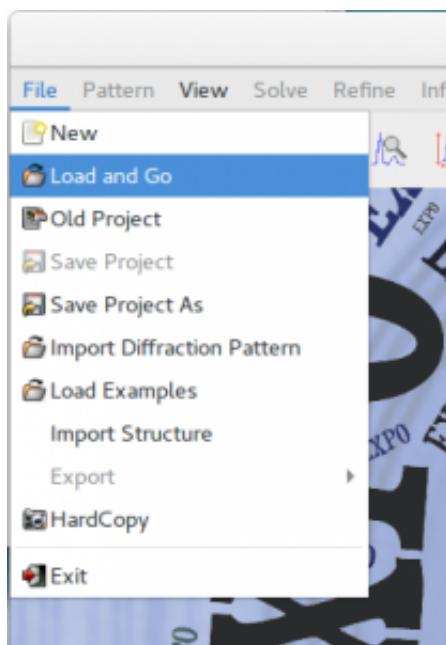
The content of the input file for EXPO2014

The input file for EXPO2014, created by EXPO2014 graphic interface or by a text editor, contains commands (the character '%' comes first in each command line) and directives (the directives must follow the corresponding command) for activating the default and non-default strategies we would like to apply. To each specific process is associated a command and/or directives.

The input file for EXPO2014 must contain at least two main commands:

%structure followed by a name (usually the name of the structure to be investigated) for identifying the solution process.

%data followed by directives for providing the minimal required information on the experimental powder diffraction data (name of data file, wavelength and type of radiation used to collect data, if different from laboratory X-rays).

Other commands (and directives) can be used depending on the process to be carried out: mainly, indexing, space group determination, structure solution by Direct Methods or structure solution by Direct Space methods and/or and any other non-default strategy.

It is here given an example of input file for EXPO2014, for carrying out the following sequential processes in a default way: indexing, space group determination and structure solution by Direct Methods. In this example, the command %job is used for printing a caption line in the output file, the command %ntreor is necessary to activate the indexation and the command %continue to execute all the rest of steps of the structure solution process.

```
%structure nbpo
%job title nbpo powder data
%data
pattern nbpo.pow
wavelength 1.0001
synchrotron
%ntreor
%continue
```

Commands of EXPO2014

http://www.ba.ic.cnr.it/softwareic/expo/commands-of-expo2014-and-their-use/

EXPO2014 recognizes the following commands:

`%window`	Graphic window is required (default choice)
`%nowindow`	Graphic window is suppressed
`%structure` *string*	This command is used to specify the name of the structure to investigate. The program creates the name of the output file adding the appropriate extension to the structure name: `string.out` If this command is not used the default string "STRUCT" (instead of the name of the structure) is used to create the name of the output file
`%job caption`	A caption is printed in the output file
%alltrials	The ALLTRIALS procedure
%data	Preparation of data
%ntreor	Indexing by N-TREOR09
%dicvol	Indexing by DICVOL06
%mcmaille	Indexing by McMaille
%extraction	Extraction of the integrated intensities
%normal	Normalization of the integrated intensities
%invariants	Structure invariants estimate
%phase	Phasing by tangent formula
%fourier	Fourier map calculation and crystal structure optimization
%polyhedra	Structure model completion in case of polyhedral and/or tetrahedral coordinated structure by POLPO1 and POLPO2 procedures
`%end`	End of the input file
`%continue`	The program runs in default conditions from the last given command up to the end. (If located after the directives of the `%data` command, it is equivalent to the following sequence of commands: `%extraction`, `%normal`, `%invariants`, `%phase`, `%fourier`). If *reflections* or *ref2* directives are used in `%data` command, `%continue` corresponds to sequential execution of `%normal`, `%invariants`, `%phase`, `%fourier`
%fragment *filename.xxx*	Importing a structure model in a format (*xxx=cif*, *mol*, ...).
%patterson	Patterson map calculation
%randomsolve	The RAMM procedure

%changelabel	The peak labelling procedure
%sannel	Structure solution by Direct Space methods
%sdirect	Direct Methods combined with Simulated Annealing
%export *filename.xxx*	Exporting files (*xxx=hkl,* cif, xyz, …)

To each command some directives correspond, usually used for activating non-default procedures. The description of the directives is given in each section corresponding to a command.

All commands and directives are in free format (between columns 1-80) and are cases independent. Only the first four characters are significant. The keywords can start in any position. If the first non-blank character is '>' or '!' or '#', then the record is interpreted as a comment. Blank line are permitted and tab characters are treated as whitespace.

Commands must be given under the following conditions:

- %data must be present;
- %extraction must follow %data;
- %normal must follow %extraction;
- %invariants must follow %normal;
- %phase must follow %invariants;
- %fourier must follow %phase;
- %patterson must follow %normal;
- %sannel must follow %data;
- %sdirect must follow %data;
- %polyhedra must follow %extraction and %fragment;
- %label must follow %extraction and %fragment.

The minimal information needed by EXPO2014 for a full run, from indexing to structure solution by Direct Methods, is given in Tutorial 1.

The minimal information needed by EXPO2014 for solving a structure by direct space methods is given in Tutorial 3.

EXPO2014 prints results in the *string.out* file.

Example 1.
Use of the command %export in the structure solution process by Direct Methods

```
%structure cime
%job structure cime
%data
 pattern cime.pow
 content s 4 c 40 n 24 h 64
 wavelength 1.52904
```

```
 cell      10.6986  18.8181 6.8246  90.000    111.284     90.000
 spacegroup p 21/n
 synchrotron
%continue
```
%export cime.hkl
%export cime.cif

Example 2

Use of the command %export in the structure solution process by Direct Space methods

```
%Structure paracetamol
%Job paracetamol
%Data
 Cell  7.100  9.380  11.708  90.0  97.42  90.0
 SpaceGroup p 21/n
 Content    c 32  o 8 n 4 h 36
 Pattern    paracetamol.pow
 Wavelength 1.54056
%fragment paracetamol.mol
%sannel
```
%export paracetamol.hkl
%export paracetamol.cif

_____

Preparation of data

http://www.ba.ic.cnr.it/softwareic/expo/preparation-of-data/

<div style="text-align:center; background-color:cyan;">

Preparation of data (`%data` command)

</div>

In this step the program reads the basic crystallographic information like unit cell parameters (if available, otherwise they can be determined by the command `%ntreor`, see Indexing), space group symbol (if available, otherwise it can be determined by the directive `findspace`, see Space group determination), the experimental wavelength, the unit cell content and the counts (or reflections). Symmetry operators are directly derived from the space group symbol. EXPO2014 uses a modified version of the subroutine SYMM Burzlaff & Hountas (1982).

An external file (containing counts or reflections) is provided by the user. If a reflection file is supplied, the extraction step is skipped and diffraction data are checked in order to find out equivalent reflections or systematically absent reflections (which are then excluded from the data set).

<div style="text-align:center;">

Directives in `%data` command

</div>

The following directives must be added after the command `%data` in the input file to supply basic crystallographic information and/or to activate specific non-default procedures:

alpha2

To carry out the K?2 stripping. It can be used only if the directive `pattern` is supplied.

cell *a b c ? ? ?*

To supply the unit cell parameters. Cell dimensions *a*, *b*, and *c* are in angstroms, *?*, *?* and *?* in degrees.

content *string*

string is the unit cell content and has to satisfy the following criteria:

– Chemical elements are represented by valid atomic symbols (IUPAC-nomenclature), each, optionally, followed by a number (omitted if it is one);

– both upper- and lower-case letters are allowed. To avoid ambiguity, in case of a two-letter symbol the second letter should be a lower-case letter; *e.g.*, *string* "Sb" is interpreted as one antimony atom, while "SB" as one antimony atom and one boron atom. *string* can contain blanks (e.g., 'S B', is interpreted as one antimony atom and one boron atom) and/or brackets ( ), [ ], { }.

Some examples of valid unit cell contents are the following one:

```
 (C10H16N6S)4
(C10 H16 N6 S)4
s 4 c 40 n 24 h 64
[(C7H8N4O2)(C7H7NO)]8
```

The first three *strings* indicate the same unit cell content.

For each chemical element up to Cf (Z=98) X-ray scattering factor constants are stored, together with information on the atomic number and weight, covalent and Van-der-Waals radii, in the file `expo.xen`. Neutron scattering factors are coded in the program for elements up to Pu (Z=94).

equiportion

To equiportion the overall intensity of a group of overlapping reflections among the reflections in the group (equal E-values are obtained).

If an external reflection file is supplied, this directive has to be used if the file contains the information on the

FWHM of reflections (*i.e.*, only if the directive <u>ref2</u> is introduced).

extefile *string*

To supply an external file containing "*d*" or "2?" values of peak positions to be used by the indexing program. In the latter case the additional directive <u>2-theta</u> has to be introduced in the input file.
If this directive is used the program will skip the automatic peak search procedure.

filetype *double*

To be used in case of a counts file containing, in the first line, the minimum 2?, 2? step and maximum 2? values and, in the rest of file, intensities and standard deviations on separated lines (see the corresponding example of counts file described in case of the directive `pattern`).

findspace

To find a possible space group when the unit cell parameters are known.

fobserved

This directive has to be used in combination with the directive <u>reflections</u> or <u>ref2</u>, if the external reflection file contains |Fobs| instead of |Fobs|**2.

fwhm *x*

To supply the value of *x* for the overlapping calculation. Two reflections are assumed to be overlapping if 2?1 -2?2 < *x* * FWHM, where 2?j is the theoretical angular position of j-*th* reflection (j=1,2) and FWHM is the full width at half-maximum of the reflection nr. 2. Default is *x* = 0.1.

histogram *n*

n is the number of bars in the histogram for the K?2 stripping procedure (only if the directive <u>alpha2</u> is used). The allowed values for *n* are 3, 5, 7 (default is *n* = 3).

isotope *El sf*

To modify the value of the neutron scattering factor for the chemical element *El* using the value *sf*. It must be introduced if the directive `neutron` is given.

neutron

To specify that neutron radiation has been used to collect the powder pattern.

norefine

To skip the cluster refinement step in the peak search procedure.

pattern *filename*

Filename is the name of the profile counts file.
The following powder diffraction data ASCII files can be imported:

- Free format X,Y data file (`*.xy`). File contains 2? in the first column and intensities in the second one. # could be used in any position for comment line or to exclude data ranges. Additional column(s) after the second one will be ignored. Optionally, the wavelength can be supplied in the first line of the file.
- Free format DAT file (`*.dat`). File contains minimum 2?, 2? step, maximum 2? values in the first row and one or more counts in free format in the subsequent lines. Caption lines can be introduced at the beginning of data file provided the first character is '#' or '!' or '>'.
- CCDC Mercury xye file (`*.xye`).
- Data file from GSAS software, format STD and ESD (`*.gda`).

- CIF powder file (`*.rtv, *.cif`).
- Siemens data file (`*.uxd`).
- Sietronics Sieray data file (`*.cpi`).
- XDD data file (`*.xdd`).
- DBWS data file (`*.dbw`).
- XDA data file (`*.xda`).
- Philips UDF data file (`*.udf`).
- PANalytical XRDML data file (`*.xrdml`).

The program recognizes automatically the file format by the filename extension. The more general extension `.pow` can be associated to a file `.xy` and `.dat`. Two example are reported here for the `.xy` and `.dat` format files.

```
#An example of XY file
  4.700    4769.000
  4.714    4860.000
  4.729    4794.000
  4.743    4857.000
  4.758    4837.000
  4.772    4891.000
  4.787    4788.000
  4.801    4886.000
  4.816    4720.000
  4.830    4688.000
  4.845    4690.000
  4.859    4867.000
  4.874    4800.000
  4.888    4863.000
```

....
....

```
# An example of DAT file
  7.00   0.020 100.00
     3017    2925    3056    2983    3008    2953    2875    2976    2886
2928
     2973    2831    2960    2821    2854    2900    2857    2798    2876
2798
     2819    2865    2830    2735    2771    2813    2806    2704    2823
2878
     2793    2795    2756    2722    2705    2736    2802    2657    2811
2736
     2753    2722    2780    2734    2795    2741    2812    2808    2713
2746
     2696    2747    2799    2876    2741    2921    2776    2966    3017
3082
     3058    3193    3147    3309    3319    3404    3490    3554    3684
3845
     4015    4069    4293    4450    4657    4905    5201    5587    5796
```

```
6172
     6646     7225     7834     8918     9767    10641    11891    13411    15091     1
7902
```

```
.................................................................................
..........................
```

If your data file format doesn't appear in the list of files supported by expo, you might use external programs for file format conversion, *e.g.*, PowDLL, POWF, xyconv.

range *thmin thmax*
Only experimental counts included in the range *thmin–thmax* are considered.

reflections *string*
To supply an external (and single-crystal like) reflection file. Each record should contain: Miller indices *h,k,l*, |Fobs|**2 and ?(|Fobs|**2) values. The default format is (3i4,2f8.2); free format (with fields separate by blanks) is also allowed and automatically read by the program.
If the reflection file contains |Fobs| instead of |Fobs|**2, the additional directive `fobserved` has to be introduced in the input file (default values are |Fobs|**2).
If the directive `reflections` is used, the directive `pattern` and all the directives introduced in combination with this last one (*e.g.*, `range`, `alpha2`, …) must to be absent in the input file.
The next command to be used is `%normal` or `%continue`.

ref2 *string*
To supply an external reflection file containing, for each record, Miller indices *h,k,l*, FWHM and |Fobs|**2.
FWHM is the full width at half-maximum of reflection. Reflection file in free format (with fields separate by blanks) can be read by the program.
If the reflection file contains |Fobs| instead of |Fobs|**2, the additional directive `fobserved` has to be introduced in the input file (default values are |Fobs|**2).
If the directive `ref2` is used, the directive `pattern` and all the directives used in combination with this last one (*e.g.*, `range`, `alpha2`, …) must to be absent in the input file.
The next command to be used is `%normal` or `%continue`.

search *thmin   thmax npeaks thre*
To change the search conditions applied to locate the peaks supplied to the indexing process.
thmin and *thmax* set, respectively, the minimum and maximum 2? value of the range to be considered for the peak search (the default values of *thmin* and *thmax* are the minimum and maximum experimental 2?,
respectively). *npeaks* is the number of the most intense peaks for which the intensity percentage is larger than the intensity percentage threshold *thre* (the default value of *thre* is 1.5).
If negative or zero values are assigned to *thmin*, *thmax* and *thre* they are ignored and replaced by their default values.
It is not necessary to specify all values (*e.g.*, it is possible to set *thmin* and *thmax* only or *thmin*, *thmax* and *npeaks* only).
Some examples of use of the search directive follow:
• to locate the 25 most intense peaks in the 2? range 5-80:
   search 5 80 25
• to search the 20 most intense peaks belonging to the full experimental pattern:

```
search 0 0 20
```
• to search all the peaks for which the intensity percentage threshold is larger than 2.5%:
```
search 0 0 0 2.5
```

shift *sx sy sz*

To supply origin shift components. They must be integral multiples of 1/24.

spacegroup *string*

String is the symbol of the space group, according to International Tables for Crystallography .

standard *string*

Standardize the space group

synchrotron

To specify that synchrotron radiation has been used to collect the powder pattern.

thcorr *x*

x is the ?2? value that is used to correct by zeropoint shift the set of 2? peak positions (to be supplied to the indexing process), according to the procedure described in the paper by Altomare *et al.* (2000). If this directive is not introduced the ??0 default value is the experimental 2? step.

wavelength *x*

x is the wavelength (in Angstrom) used to collect the powder pattern.

2-theta

This directive is used in combination with the directive <u>extefile</u>, to supply to the indexing program an external file containing the 2? values of peak positions.

Examples

Example 1

To skip the peak search procedure and supply to the <u>Indexing</u> step a set of 2? peak positions (see the ASCII file mes.pea).


```
%structure mes
%job MES - data from home diffractometer
%data
        pattern mes.pow
        wavelength 1.5406
        2-theta
        extefile mes.pea
%ntreor
%continue
```


mes.pea content:


```
12.0014
```

```
13.6797
15.5003
16.0599
16.2003
17.8815
18.5194
19.6196
20.5805
20.7582
21.6204
...............
```

Example 2

To activate the space group determination procedure in case of neutron data.

```
%structure cfcl
%job CFCL - data from neutron source
%data
        pattern cfcl.pow
        cell 10.168 14.964 5.100 90.000 90.000 90.000
        content C 8 F 16 Cl 16
        findspace
        wavelength 1.595
        neutron
%continue
```

Example 3

Minimal crystallographic information to be provided to EXPO2014 if unit cell parameters and space group are known.

```
%structure crox
%job CROX- data from home diffractometer
%data
        pattern crox.pow
        cell 5.447 6.5576 12.1147 106.382 95.715 77.970
        content Cr 8 O 21
        wavelength 1.3922
        spacegroup p -1
 %continue
```

Example 4

The K?2 stripping with a 5 bars histogram is requested.

```
%structure dada
%job DADA- data from home diffractometer
%data
        spacegroup p 21 21 21
        cell 7.13115 9.9077 12.9256 90.0 90.0 90.0
        content Si 12 Ti 4 K 8 O 40
        wavelength 1.540562
        pattern dada.dat
        alpha2
        histogram 5
%continue
```

Example 5

To supply a reflection file containing Miller indices *h,k,l*, |Fobs|**2 and ?(|Fobs|**2) values (see the ASCII file mes.hkl; the extraction step is skipped).

```
%structure mes
%job MES - data from home diffractometer
%data
        reflections mes.hkl
        wavelength 1.5406
        content c 24 n 4 o 20 s 4 h 52
        cell 8.5884 9.9309 11.1046 90.000 93.754 90.000
        spacegroup p 21/c
%normal
%continue
```

mes.hkl content:

```
>   H   K   L   Fobs**2  sigma(Fobs**2)
    1   0   0    0.01       0.01
    0   1   1    0.17       0.01
    1   1   0    0.29       0.01
    1   1  -1    0.41       0.01
    0   0   2    0.36       0.01
    1   1   1    0.41       0.01
    0   2   0    1.34       0.01
    0   1   2    0.00       0.01
    1   0  -2    0.59       0.01
    0   2   1    0.63       0.01
    1   0   2    0.04       0.01
    1   1  -2    1.81       0.01
.............................
```

Example 6

To supply a reflection file containing Miller indices *h,k,l*, FWHM and |Fobs|**2 values (see the ASCII file
mes_ref2.hkl; the extraction step is skipped).

```
%structure mes
%job MES - data from home diffractometer
%data
       ref2 mes_ref2.hkl
       wavelength 1.5406
       content c 24 n 4 o 20 s 4 h 52
       cell 8.5884 9.9309 11.1046 90.000 93.754 90.000
       spacegroup p 21/c
%normal
%continue
```

mes_ref2.hkl content:

```
>    H    K    L      FWHM        Fobs**2
     1    0    0     0.1064      0.0108
     0    1    1     0.1065      0.1671
     1    1    0     0.1066      0.2948
     1    1   -1     0.1066      0.4085
     0    0    2     0.1066      0.3635
     1    1    1     0.1067      0.4123
     0    2    0     0.1067      1.3356
     0    1    2     0.1068      0.0014
     1    0   -2     0.1068      0.5944
     0    2    1     0.1068      0.6280
     1    0    2     0.1068      0.0363
     1    1   -2     0.1069      1.8067
...........................
```

<div align="center">

References

</div>

Altomare, A., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Rizzi, R. & Werner, P.-E. (2000). J.
Appl. Cryst. **33**, 1180-1186.
Burzlaff H. & Hountas A. (1982). *J. Appl. Cryst.* **15**, 464-467.

_____

Indexing

http://www.ba.ic.cnr.it/softwareic/expo/indexing-by-expo2014/

The primary step of the solution process by powder diffraction data is the indexing process. Its main goal is to recover the unit cell parameters (a, b, c, ?, ?, ?), by assigning the appropriate triple of Miller indices (*hkl* or **h**) to each observed interplanar spacing (d_h). Indeed, the indexing step enables to reconstruct the three-dimensional elementary cell by the information on the d_h values extracted from the experimental one-dimensional pattern. A wrong unit cell is responsible for the failure of the structure solution process, consequently finding the correct unit cell is a mandatory request.

The indexing process is performed by EXPO2014 *via* the program N-TREOR09 (Altomare *et al.*, 2009), the evolution of the N-TREOR software (Altomare *et al.*, 2000). In addition, EXPO2014 can determine the unit cell by two widely used and freely available indexing programs: DICVOL06 (Boultif & Louër, 2004), McMaille (Le Bail, 2004).

<div style="background-color:cyan">

Indexing by EXPO2014

</div>

EXPO2014 can index a powder diffraction pattern by one of the three following ways:

1) by graphic interface, by creating a new EXPO2014 input file or loading an existing one (`.exp`);
2) by graphic interface, by loading a powder diffraction pattern file;
3) by a command line.

1) by grafic interface, by creating a new EXPO2014 input file or loading an existing one (`.exp`)

To carry out the indexing process by EXPO2014 is it possible to edit a new input file (saving it with the extension `.exp`) or load/modify an existing one, created by EXPO2014 or by a text editor (see the chapter Preparing input file).

In case of an already created input file, load it from the menu **File > Load & Go**.
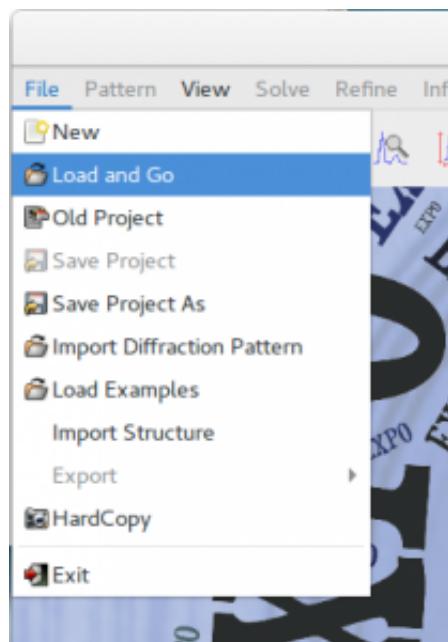
Use the button   of the toolbar to carry out the steps of the indexing procedure.

An example of input file for indexing the powder pattern of cimetidine is `cime.exp`, available at the directory `examples` (see the 'Running examples' at the [EXPO2014 installation](#) section):

```
%structure cime
%job cimetidine (C10H16N6S) -- Synchrotron data
%data
pattern cime.dat
wavelength 1.52904
synchrotron
%ntreor
%continue
```

The command that has to be introduced in the input file for carrying out the indexing process by N-TREOR09 is `%ntreor`. It must be added after the command `%data`. Add `%continue` if you want to carry out all the next steps of the structure determination process (see the section [Commands of EXPO2014](#) for the description of the useful commands that can be introduced in the EXPO2014 input file).

The commands `%dicvol` and `%mcmaille` have to be used to perform the indexing by the program DICVOL06 and McMaille, respectively. For example, by loading the following EXPO2014 input file the cell determination step will be carried out, in sequence, by N-TREOR09, DICVOL06 and McMaille.

```
%structure cime
%job cimetidine (C10H16N6S) -- Synchrotron data
%data
pattern cime.dat
wavelength 1.52904
synchrotron
%ntreor
%dicvol
%mcmaille
%continue
```

2)Indexing by graphic interface, by loading a powder diffraction pattern file
In order to determine the unit cell by graphic interface, it is necessary:
a) to load the powder diffraction data file by the menu **File > Import Diffraction Pattern** (for the description of the formats of the counts files readable by EXPO2014 see the directive [pattern](#) of the

command %data);
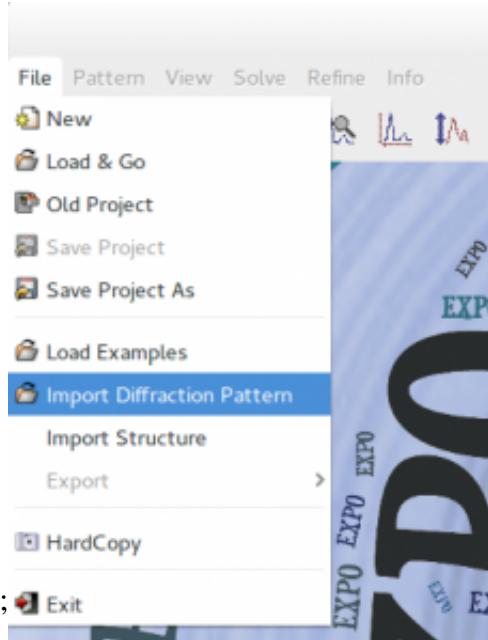
b) to set the type and the value of the experimental wavelength by the EXPO2014 dialogue window.

c) to select the indexing program from **Pattern > Indexing**.

As an example, load the counts file merca.xy, available at the directory examples, and supply the wavelength value 1.54056 Å.

If the program N-TREOR09 is chosen from **Pattern > Indexing**, the dialogue box shown in the following figure will be opened.
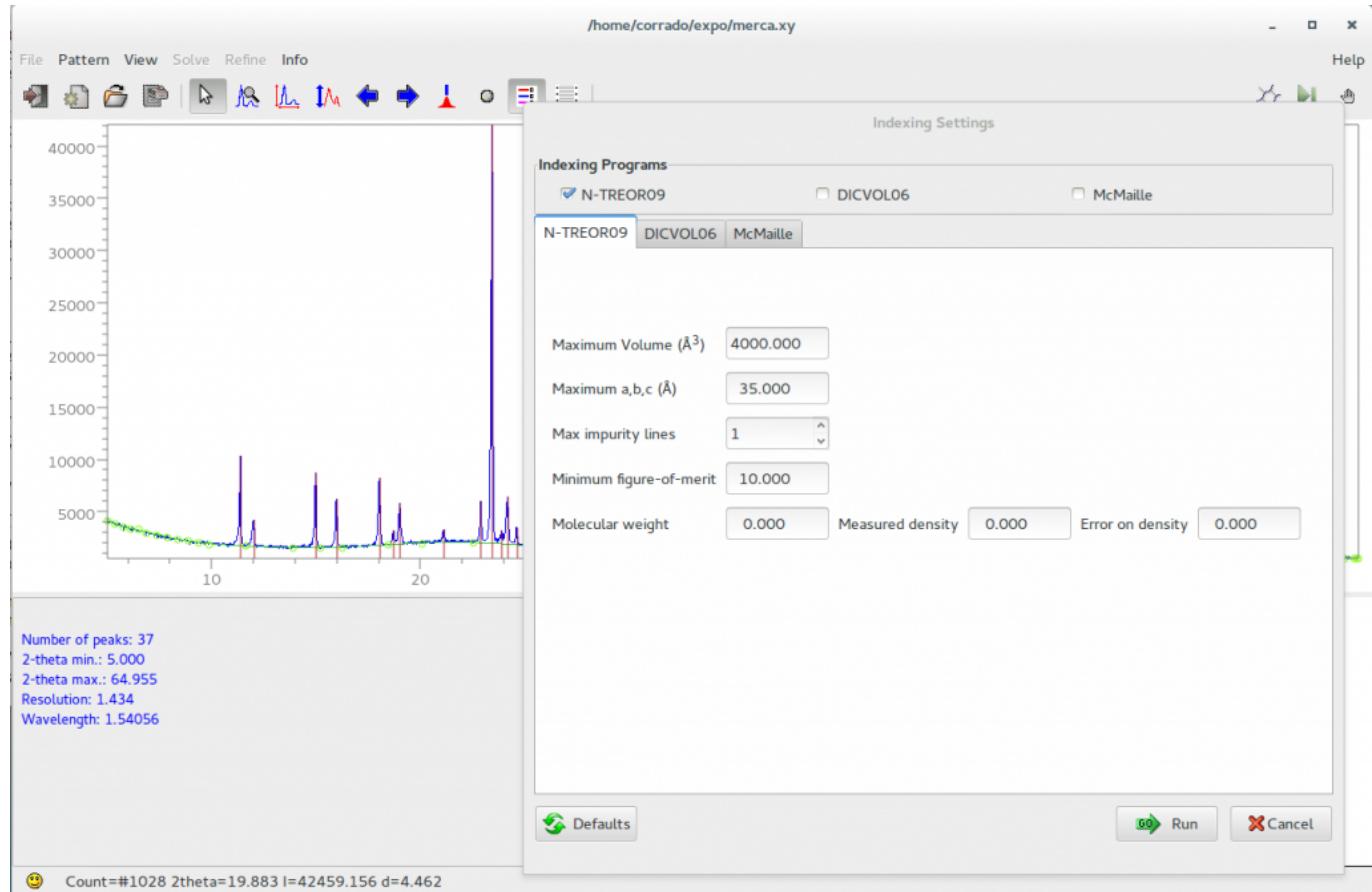
The default values of the maximum cell volume, the maximum cell axes, the maximum number of allowed impurity lines, the minimum figure of merit M20 (de Wolff, 1968) characterizing the candidate cells, can be changed by graphic interface. If available, it is possible to supply the measured density together with the corresponding error. Molecular weight is used only if the density value is not equal to zero.

N-TREOR09 documentation is available at http://www.ba.ic.cnr.it/softwareic/expo/ntreor/. By clicking on 'Run', EXPO2014 automatically creates a N-TREOR09 input file and executes N-TREOR09.

By ticking 'DICVOL06' in the frame 'Indexing Programs' it is also possible to activate the graphic interface for the indexing by DICVOL06 (see the following figure). In this case, by clicking on 'Run', EXPO2014 automatically creates a DICVOL06 input file and runs DICVOL06.

On Windows operating system the executable file `dicvol06.exe` is tipically located in the folder `C:\Program Files\Expo2014\bin` and corresponds to the program freely available at http://www.ccp14.ac.uk/ccp/web-mirrors/dicvol/. On Linux OS `dicvol06.exe` is located in `expo_install_dir\bin` (`expo_install_dir` is the directory where EXPO2014 has been installed). To run DICVOL06 under Linux OS, the Windows Emulator Wine is required. Wine is available at the software center of the most common linux distributions. DICVOL06 documentation can be found at http://www.ba.ic.cnr.it/softwareic/expo/dicvol06/.

Indexing Settings

Indexing Programs

☐ N-TREOR09   ☑ DICVOL06   ☐ McMaille

| N-TREOR09 | **DICVOL06** | McMaille |

Number of lines used for searching solutions 20 ⌃⌄   Peak position error 0.020

Volume (Å³) Min 0.000   Max 4000.000

a,b,c (Å)   Min 0.000   Max 30.000

B (°)   Min 90.000 Max 125.000

Max impurity lines 0 ⌃⌄   Minimum figure-of-merit 10.000

Molecular weight 0.000   Measured density 0.000   Error on density 0.000

Crystal System

☑ Cubic   ☑ Hexagonal   ☑ Monoclinic

☑ Tetragonal   ☑ Orthorhombic   ☐ Triclinic

Zero-point

☑ Refine zero-point   ☐ Estimate zero-point   Experimental zero-point 0.000

☐ Exhaustive search

🔄 Defaults   GO ▶ Run   ✖ Cancel

By ticking 'McMaille' in the frame 'Indexing Programs', EXPO2014 activates the graphic interface for the indexing program McMaille version 4.00. Its documentation is available at http://www.cristal.org/McMaille/McMaille-v4.html.

Whatever the indexing program is graphically chosen, EXPO2014 automatically locates and visualizes peaks that will be used for finding the unit cell. The automatic peak search is skipped if peaks positions (in terms of "*d*" or 2? values) are already available, *e.g*., because, before the selection of the indexing

program, an external file of peaks has been graphically loaded (*i.e.*, by selecting **Pattern > Peaks > Load Peaks**).

If the located peaks are reliable, clicking on 'Run' they are automatically refined and supplied to the indexing program to determine the unit cell. On the contrary, if changes of peaks are necessary press 'Cancel' to leave the 'Indexing Setting' window and modify the peaks selection by graphic interface (see the section Selecting peaks for indexing).

3) Indexing by using a command line

To execute the indexing procedure by a command line, the following sintax should be used:

`expo` followed by the `filename` of the `powder diffraction data` and the option `-index=indexing_program`, where `indexing_program` indicates the program(s) to be used for the indexing process (the variable `indexing_program` has to be equal to n in case of N-TREOR09, to d in case of DICVOL06 and to m in case of McMaille). For more infomation about the use of a command line see the section Command line usage of the **EXPO2014 installation**.

For example, the following command line

`expo merca.xy -index=ndm`

has to be typed to carry out the indexing of the powder diffraction pattern stored in the file `merca.xy`, by running, in sequence, N-TREOR09 (n), DICVOL06 (d) and McMaille (m).

Add the option `-wavel=wavelength` to supply the value of the wavelength used to collect the powder pattern. If `-wavel` is omitted the default wavelength is Cu$K?_1$ (1.54056 Å).

For example, if the following command line is used:

`expo cime.dat -index=nd -wavel=1.52904`

EXPO2014 carries out the indexing of the powder diffraction pattern stored in the file `cime.dat` (collected by using an experimental wavelength of 1.52904 Å) by executing N-TREOR09 (n) and DICVOL06 (d), in sequence.

The following additional options can be introduced in the command line:

`-auto` to avoid the user interaction with the EXPO2014 graphic interface and run the program in automatic way;

`-nogui` to run the indexing program(s) by deactivating the graphical interface. For example, by using the following command line:

`expo cime.dat -index=n -wavel=1.52904 -nogui`

EXPO2014 performs the indexing process by N-TREOR09 with no graphic interface.

Selecting peaks for indexing

Accurate and reliable peak detection is fundamental for the success of the powder pattern indexing. To detect peaks, EXPO2014 uses an approach based on background subtraction, profile smoothing and application of the second derivative method. The automatic peak search is extremely sensitive to noise and may suffer from misinterpretation of the background that can make inefficient the algorithm.

In order to check the reliability of the automatic peak search results, use the toolbar buttons to explore the pattern. If the automatic peak search outcome is not reliable, the peak selection can be modified in different ways:

1.By changing the curve describing the background. This is possible adding or removing the background points by pressing on the toolbar button and/or using left mouse click to add points and right mouse click to delete points. Background function can be also modified by the button of the 'Pattern' option. When background is changed, the peak positions are automatically updated and visualized.

2.By choosing the toolbar button to add peaks (by left click of mouse) and/or remove peaks (by right click of mouse).

3.By the menu **Pattern > Peaks > Peak Search Conditions**. The button 'Peak Search Conditions' opens a window that enables to modify the main search conditions. All peaks with percentage intensity larger than or equal to '% Int. Threshold' will be selected.
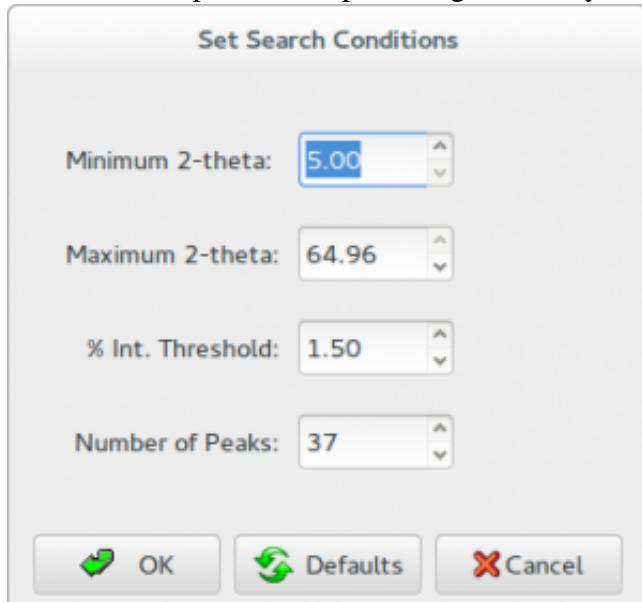
Set Search Conditions	
Minimum 2-theta:	5.00
Maximum 2-theta:	64.96
% Int. Threshold:	1.50
Number of Peaks:	37
OK	Defaults	Cancel

4.Select **Pattern > Peaks > Peak List** to have access to the list of peaks; for each of them the following values are provided: 2?, *d*, intensity (calculated removing the contribution of background), full width at half maximum (FWHM). This window can be used to edit the 2? or *d* values (peak intensity and FWHM values cannot be modified) and to add or delete peaks. If you click on a row in the list, the corresponding peak will be marked with a tick.

5. Use **Pattern > Peaks > Load Peaks / Save Peaks** to load or save peaks as 2? or *d* values.
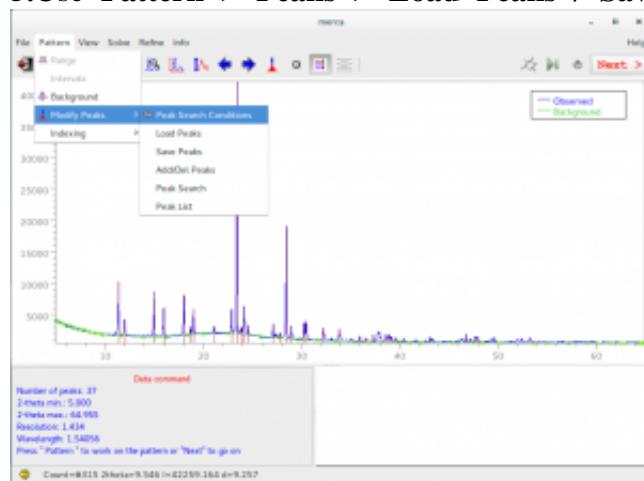
Output of indexing

To have access to the output file it is possible to

– select **Info > View Output file,** or

– open the file `name.out` (where `name` is the name defined by the command `%structure`, if an input file `.exp` has been loaded, or is the filename of the powder diffraction file without its extension, if the indexing has been carried out by loading this file by graphic interface) with a text editor, or

– click on the button of the main menu.
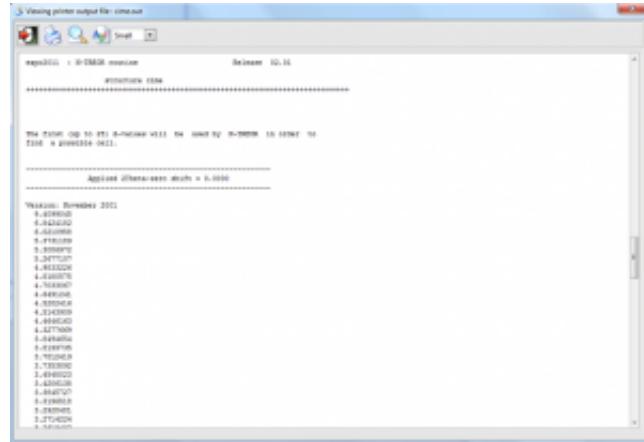
Additional output files are created by the specific indexing programs and are accessible through the menu **Pattern > Indexing > Output**. Among the additional input and output files that can be found in the working directory: `name.new` (where `name` is the name defined by the command `%structure`, if an input file `.exp` has been loaded, or is the filename of the powder diffraction file without its extension, if the indexing has been carried out by loading this file by graphic interface), `condens.out` and `ntreor.out` in case of N-TREOR09; `dicvol_inp`, `dicvol_out` in case of DICVOL06; `mcmaille.dat`, `mcmaille.imp`, `mcmaille.mcm` in

case of McMaille.

<div align="center">

Directives in the command `%ntreor`

</div>

The default choices of N-TREOR09 can be changed *via* suitable directives of the command `%ntreor` .
All the directives of the indexing program TREOR (Werner, Eriksson & Westdahl, 1985) are valid for
the command `%ntreor` and have to be used with the same syntax (see the <u>N-TREOR09 manual</u> for details).

<div align="center">

Directives in the command `%data,` useful for the indexing process

</div>

**The following directives of the command `%data` can be used to supply a suitable list of peaks to
the indexing process:**

- **extefile**
- **2-theta**
- **search**

<div align="center">

Remarks and suggestions

</div>

The indexing problem can be usually solved by using the first (in terms of 2θ values) 25 peaks (even
very weak peaks included). Peaks at low angle not clearly resolved and with an appreciable peak width
should be not included to avoid to supply to the indexing program not well positioned peaks.
In case of indexing failure try by different attempts, gradually eliminating suspicious Bragg peaks
(*e.g.*, the weakest peaks or peaks on a shoulder) and check if it results in a success of the indexing
process. The failure could be due to the presence of impurity peaks and/or to the absence of one or more
useful peaks. Sometimes, if weak peaks are missed, it could be useful to change the search conditions by
reducing the intensity threshold value. On the other hand, in case of presence of an impurity phase, if
it is a minority phase, it could be useful to select the most intense peaks (about 20 peaks are usually
enough). Indexing program like N-TREOR09 and DICVOL06 can successfully index a powder
diffraction pattern even if some peaks are omitted but they can fail if a lot of wrong or misplaced peaks
are selected for the cell determination process.

Once the peaks have been located, the next step is the automatic refinement of peak positions and profile
parameters (a Pearson VII profile function is used). At the end of the refinement a file
named `name.pea` (where `name` is the structure name), containing the refined *d* values associated to the
selected peaks, is automatically created by the program. This file can be used to skip the peak search
procedure (see the directive <u>`extefile`</u>) if a new indexing process is performed.

At the end of the indexing procedure the graphic interface supplies a list containing, for each
solution, the used indexing program (**N** for N-TREOR09, **D** for DICVOL06, **M** for McMaille), the cell
parameters and the cell volume, the classical figure of merit M20 (de Wolff, 1968), the applied zero-shift,

the number of unindexed lines (NIX), the symmetry information. In case of N-TREOR09, the figure of merit FOMnew (Altomare *et al.*, 2009) is provided and the symmetry information concerns the most probable extinction symbol. In case of indexing by McMaille the figure of merit McM20 is displayed (see the following figure).

Nr.	**Prog.**	**a**	**b**	**c**	**alpha**	**beta**	**gamma**	**Vol.**	**M20**	**FOMnew**	**Mc20**	**shift**	**NIX**	**Symmetry Info**
1	N	14.93486	5.97028	7.87733	90.000	100.502	90.000	690.6	38.00	3.829	-	-0.029	0	P 1 21/a 1
2	N	14.96515	5.97891	7.89357	90.000	100.510	90.000	694.4	21.00	2.743	-	0.029	0	P 1 21/a 1
3	D	14.94120	5.97260	7.88120	90.000	100.511	90.000	691.5	46.40	-	-	0.015	0	Mono
4	M	14.95360	5.97590	7.88700	90.000	100.554	90.000	692.9	45.04	-	116.22	0.000	0	Mono
5	M	7.75300	14.70340	11.94790	90.000	90.000	90.000	1362.0	33.26	-	29.46	0.000	1	Ortho
6	M	9.81870	5.97190	11.82060	90.000	93.047	90.000	692.1	45.09	-	21.81	0.000	2	Mono
7	M	6.28440	14.69450	9.58530	90.000	107.791	90.000	842.8	26.63	-	19.78	0.000	3	Mono
8	M	15.50610	15.15650	11.90570	90.000	90.000	90.000	2798.1	10.97	-	17.54	0.000	2	Ortho
9	M	15.40430	15.15640	11.84900	90.000	90.000	90.000	2766.5	21.03	-	16.73	0.000	1	Ortho
10	M	15.78000	13.61530	11.82230	90.000	90.000	90.000	2540.0	13.57	-	16.37	0.000	3	Ortho
11	M	15.77260	13.69110	11.80230	90.000	90.000	90.000	2548.6	20.64	-	15.98	0.000	3	Ortho

If more plausible cells are suggested and more than one indexing program has been applied, for each indexing program the cells are ranked according to its reference figure of merit. When the quality of data is not good, uncertainty on the indexing procedure can lead to more than one solution that should be carefully examined, in particular if their figures of merit are not encouraging. Clicking on each suggested cell of the list, the tick marks at the bottom of the experimental pattern, indicating the calculated positions of the Bragg reflections, will be displayed. The tick marks must correspond to peaks in the pattern. When no experimental peaks are associated to the tick mark, the corresponding reflection is probably systematically absent. On the other side, if a diffraction peak in the pattern is not associated to any tick, the peak can be referred to an impurity phase or the candidate cell is wrong. The unindexed lines are highlighted by small full circles above the peak.

If more cells are available at the end of indexing they are saved and always accessible from **Pattern> Indexing> Cell List**.

 The reliability of indexing is assessed by means of numerical figures of merit. The figures of merit reported by EXPO2014 are M20, FOMnew only for N-TREOR09, M20 for DICVOL06 and McM20 only for McMaille. Consult the output file of the specific program (Select **Pattern** > **Indexing** > **Output**) to have a look at additional figures of merit (*e.g.*, F20) for each solution.

In case of M20, the rule suggested by de Wolff is the following one: if M20 > 10 and the number of unindexed lines, among the first twenty ones, is not larger than 2 the indexing process is reliable and substantially correct; exceptions to this rule can be possible. In case of high quality data (e.g., synchrotron data) the M20 of the correct cell is usually much larger than 10; in case of laboratory data the correct solution can be characterized by M20 value not so larger than 10.

The possibility of running more than one indexing program enhances the indexing power of EXPO2014: if one of the indexing programs fails it is possible that another one (at least) can succeed.

In addition, similar results obtained by different programs strengthen the confidence in the correctness of the cell. N-TREOR09 and DICVOL06 usually run much faster than McMaille, it is better to apply them

first.

N-TREOR09 includes some features in the default strategy:
– it performs an automatic correction of the 2? zero-shift;
– it may successfully index by default powder diffraction patterns containing one impurity peak;
– starting the search in the cubic system, if an acceptable solution is not found the program explores all the rest of crystal systems up to triclinic and stops as soon as a plausible solution has been found.

In case of DICVOL06 and McMaille these features are not included in the default strategy and should be enabled by user in the graphical interface when a preliminary indexing fails.

The search of triclinic cell can take a long time depending on the indexing program.

Once the indexing step has been carried out, you may perform the space group determination step by

selecting the cell parameters and by clicking on the OK button. The following dialog window will appear:

Before clicking OK, in this window you must specify the unit cell content. In the case of the 2-Mercaptobenzoic acid structure the cell volume suggests that the cell can accommodate 4 molecules of $C_7H_6O_2S$ (Z=4) and the cell content may be written as `(C7 H6 O2 S)4`.

During the editing of the cell content, the program calculates and displays a rough estimate of the following quantities:

1. Content Volume: an approximate value of the cell volume calculated as summation of the average atomic volumes associated to the supplied atoms. Its value is expected to be close the cell volume calculated directly by cell dimensions and reported in the window below the cell parameters. A table of the average volume of the atoms is reported in Appendix A.

2. Volume per Atom: the average volume for non-hydrogen atoms in the crystal structure calculated as ratio between the cell volume and the number of non-hydrogen atoms. Its value is expected to be close to 18 Å^3 (18 Å^3 rule) but sensible deviations from this ideal value can be observed due, for example, to the presence of H-bonding or large void spaces (*e.g.*, in case of zeolites) in the crystal structure. A reasonable range is 15-22 Å^3.

3. Density: is the density calculated by the mass of supplied atoms and the cell volume (g/cm^3). Keep in mind that the density of molecular compound is relatively low and lying to the range 0.9-2.0 g/cm^3.

Examples

Example 1:

The following example concerns a zeolite structure, characterized by a large cell axis (73.6505 Å) and cell volume (7194.74 Å^3).

`vni.pea` is an ASCII file containing the interplanar distances (*d*) to be supplied to N-TREOR09 (the peak-search step is skipped).

A default indexing run is not able to find the correct cell, due to the default limits concerning the maximum value for the cell axes and cell volume (*i.e.*, 35 Å and 4000 Å^3, respectively). In order to increase these limits the directives `cem = 80,` (*i.e.*, maximum allowed value for cell axes set to 80 Å) and `vol = 10000,` (*i.e.*, maximum allowed value for the cell volume set to 10000 Å^3) have been introduced in the input file, (see the documentation of N-TREOR09, available at http://www.ba.ic.cnr.it/softwareic/expo/ntreor/)

The additional directive `nix=0,` states that the maximum allowed number of unindexed lines (`nix`) should be zero (the default value for `nix` is one), *i.e.*, only cells with no unindexed lines will be considered as plausible by N-TREOR09.

```
%structure vni
%job vni - Synchrotron data
%data
pattern vni.pow
wavelength 0.99975
synchrotron
extefile vni.pea
%ntreor
vol=10000,cem = 80, nix=0,
%continue
```

Example 2:

The maximum number of the allowed unindexed lines is set to 3 (nix=3,). The counts file format is not the default one.

```
%structure norbo_ag
%job norbo_ag - C7H10 - T = 85 K
%data
pattern norbo_ag.pow
```

```
filetype double
wavelength 0.85041
synchrotron
%ntreor
nix=3,
%continue
```

References

Altomare, A., Campi, G., Cuocci, C., Eriksson, L., Giacovazzo, C., Moliterni, A., Rizzi, R. & Werner, P.-E. (2009). *J. Appl. Cryst.* **42**, 768-775.

Altomare, A., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Rizzi, R. & Werner, P.-E. (2000). J. Appl. Cryst. **33**, 1180-1186.

Boultif, A., Louër, D., (2004). *J. Appl. Cryst.* **37**, 724-731.

Le Bail, A., (2004). *Powder Diffraction* **19**, 249-254.

Le Bail, A., Clearfield, J. H. Reibenspies & N. Bhuvanesh, (2008). Principles and Applications of Powder Diffraction 261–309.

De Wolff, P. M.(1968). J. Appl. Cryst. **1**, 108-113.

Werner, P.-E., Eriksson, L., & Westdahl, M. (1985). J. Appl. Cryst. **18**, 367-370.

_____

Space group determination

http://www.ba.ic.cnr.it/softwareic/expo/space-group-determination-2/

In the crystal structure solution process the determination of the space group usually follows the indexing step. This task is carried out by careful investigation of the experimental intensities, aimed at detecting the extinction group. Because of the collapse of the experimental data into the Bragg angle one-dimension space, the experimental pattern suffers by peak overlap and, sometimes, non-trivial background description. As a consequence, the judge about the 'presence' or 'absence' of reflections may provide ambiguous results. A new probabilistic procedure aiming at recognizing the extinction symbol has been implemented in EXPO2014. It uses the statistics of the normalized intensities (as extracted by the Le Bail method) to calculate the probability of each extinction symbol compatible with the crystal system suggested by the indexation process of EXPO2014. A suitable weight is associated to the intensity of each reflection, in order to take into account the uncertainty about its estimate and avoiding that large errors on the experimental intensities compromise the identification of the extinction group and, consequently, of the space group. The EXPO2014 approach has been tested on a large number of test structures and on all the crystal systems, providing good results (Altomare *et al.*, 2004; Altomare *et. al.*, 2005; Altomare *et al.*, 2007).

<div style="background-color:cyan; text-align:center; font-weight:bold;">Preparing input file for space group determination</div>

To run EXPO2014 for space group determination you can follow two different approach

- **Creating an input file (`*.exp`).** You can edit a new input file or modify an exiting input file by a text editor and save it with extension `.exp`.

A typical `.exp` input file for space group determination is here reported in case of the crystal structure determination of cimetidine from the indexed pattern.

```
%structure cime
%job cimetidine (C10H16N6S) -- Synchrotron data
%data
pattern cime.dat
wavelength 1.52904
synchrotron
cell 10.69808 18.81786 6.82267 90.00000 111.28911 90.00000
findspace
content (C10H16N6S)4
%continue
```

The directive **findspace** enables the space group determination procedure. It is a directive of the command **%data**. The unit cell is assumed to be determined by the indexing step, the directive **synchrotron** is

necessary only in case of X-ray synchrotron diffraction data. The cell content (directive `content`) is mandatory for the space group determination. To know the meaning of the other lines of the input file, see the chapter command and their use of this manual. For more information about how you can create an `.exp` input file visit the Preparing input file section.

Once a time the input file has been created, load the `.exp` input file from menu **File > Load & Go**.
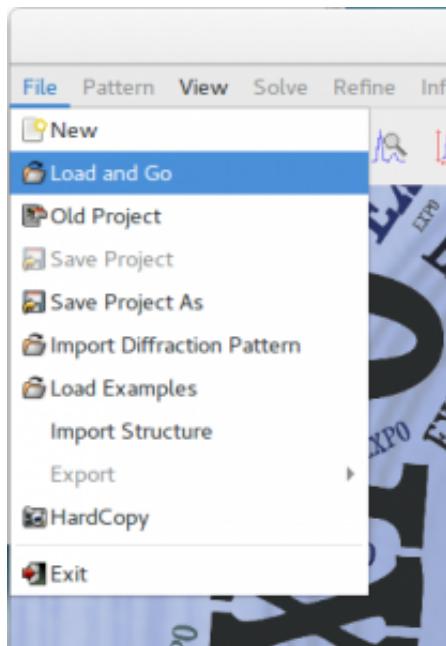
- **By graphical interface.** Load the powder diffraction data file selecting **Import Diffraction Pattern** in the menu **File**, then select **Pattern > Space Group > Space Group Determination**. Fill the frame 'Cell Parameters' and 'Cell Content'.
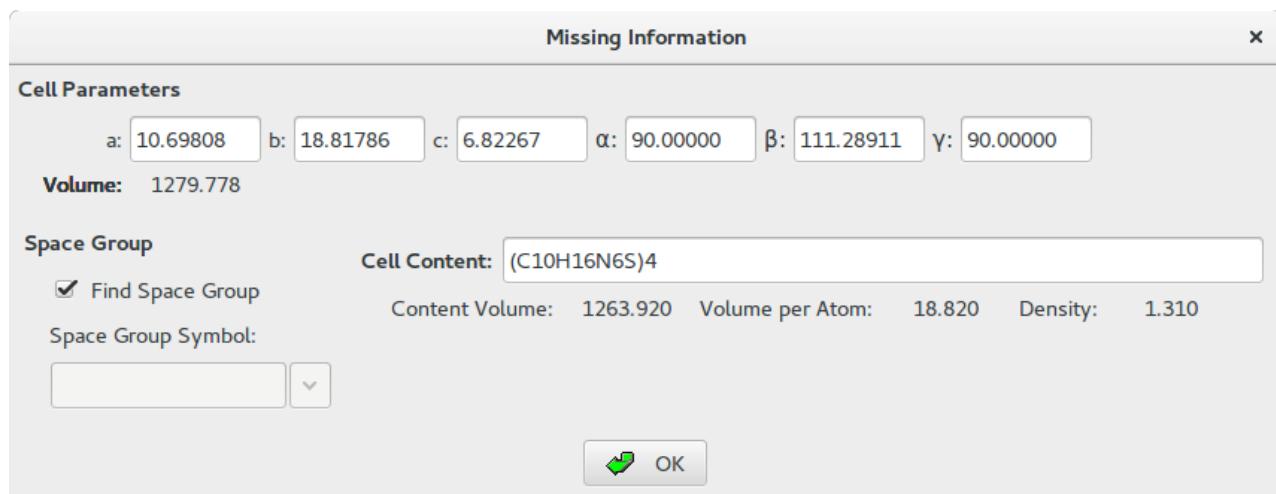
Press the **OK** button to launch the space group determination step.

Space group determination

The different steps of the algorithm are here schematized.

1) The full pattern decomposition is performed, after that the integrated intensities have been extracted from the experimental pattern according to the Le Bail algorithm, in the space group having the largest Laue symmetry and no extinction conditions (*e.g.*, `P 1 2/m 1` in the monoclinic system, `P 2/m 2/m 2/m` for the orthorhombic, `P 4/m m m` for the tetragonal, `P 6/m m m` for the trigonal-hexagonal, `P m -3 m` for the cubic systems).

2) The extracted intensities are then <u>normalized</u>, according to the classical Wilson method, by using the space group mentioned at the point 1).

3) The statistic of the suitably weighted integrated intensities is used to define the probability of the extinction symbol.

4) For each crystal system the probability corresponding to the different possible extinction symbols is calculated.

Use the button [Next >] on the toolbar to carry out all the next steps. At the end of the procedure the following window will be displayed containing the list of the most plausible space groups and the corresponding extinction symbol and probability (FoM).

Find space group

Space Group	Extinction symbol	FoM	Nabs	Nasym	No. in CSD	% of CSD	Rank	Chiral
P 21/n	P 1 21/n 1	0.659	20	17	279041	34.57	1	no
P 2/n	P 1 n 1	0.155	18	17	5232	0.65	14	no
P n	P 1 n 1	0.155	18	34	3447	0.43	18	no
P 21	P 1 21 1	0.150	2	34	41791	5.18	5	yes
P 21/m	P 1 21 1	0.150	2	17	4023	0.50	17	no
P 2	P 1 - 1	0.035	0	34	142	0.02	96	yes
P 2/m	P 1 - 1	0.035	0	17	110	0.01	111	no
P m	P 1 - 1	0.035	0	34	21	0.00	202	no
P 21/c	P 1 21/c 1	0.003	24	17	279041	34.57	1	no
P 21/a	P 1 21/a 1	0.003	18	17	279041	34.57	1	no
P 2/a	P 1 a 1	0.001	16	17	5232	0.65	14	no

[List] [OK] [Cancel]

In general, more space groups are associated to the same extinction symbol (e.g., *P* 2/*n* and *P n* in case of the extinction symbol *P* 1 *n* 1) but occasionally just one space group can correspond to an extinction symbol (e.g., *P* 21/*n* in case of the extinction symbol *P* 1 21/*n* 1).

Additional information reported in the window above can be useful in case of space group ambiguities: the number of systematically absent reflections not in overlap (Nabs), the expected number of atoms in the asymmetric unit (Nasym), the frequency of the selected space group in the Cambridge Structural Database (CSD) in terms of the number of occurrences, the percentage of occurrences in the CSD (% of CSD) and the space group frequency ranking (Rank). In the last column is reported 'yes' if the space group is a chiral (Sohnke) space group; this information is useful because enantiomerically pure compound can crystallize only in a chiral space group.

Press the button [List] to access to a list of reflections that, according to the selected extinction symbol, should be systematically absent. For each reflection the following specifications are given: the order number of the

reflection, the Miller indices, the symbol of the symmetry operator responsible for the extinction and the corresponding extinction condition, the probability of the extinction, the type of reflection (single or in overlap).

By default the reflections are ordered according to the extinction condition but you can change the order selecting a different ordering criterion by clicking on the title column. The selected reflection on the list will be marked on the pattern.

The reflections that are expected to be absent according to the selected extinction symbol are marked by vertical bars located below the observed pattern, just above the set of bars showing the positions of reflections. They have different colors depending on whether the reflection is single (or belongs to a whole extinct cluster) or in overlapping. By default, a black vertical bar marks single systematically absent reflections, while the rest of the absent reflections (those in overlap) are marked with a purple bar. This graphical support helps the user to discard unreliable extinction symbols even if estimated with large probability.
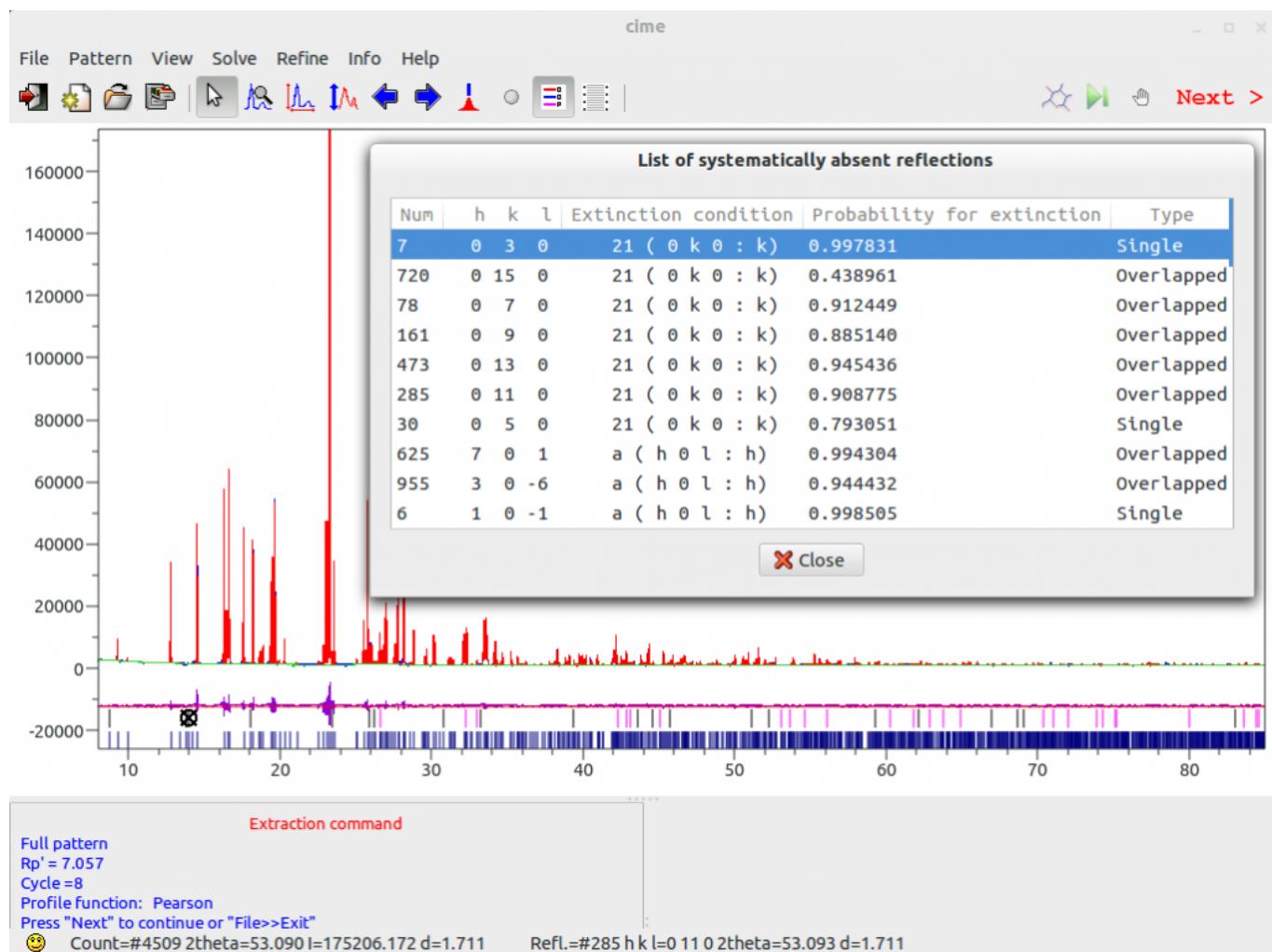
Sometimes this procedure doesn't guarantee that the space group corresponding to the highest FoM is the correct one, at the same time space groups having identical systematic absences (corresponding to the same extinction symbol) cannot be distinguished. In case of ambiguities more space groups should be for the structure solution step. A good choice is to start from the space group with the highest frequency.

Press the button [OK] to accept the selected space group and continue with the structure solution process. You can also save the table of results in an .expo project file selecting **File > Save Project As**. To reload the .expo file click on **File > Old Project**.

Output of space group determination

The output file contains details about the space group determination procedure. Use **Info > View output file** to access to the output file or locate the file `structure_name.out` in your working directory and open it with a text editor. At the end of the procedure a new input file, called `structure_name1.exp` (*e.g.* `cime1.exp`) will be created. This file contains the selected space group and can be loaded to skip the space group determination and to perform the structure solution process. The table of space groups is accessible by selecting **Pattern > Space Group > Space Group List**.

References

Altomare, A., Caliandro, R., Camalli, M., Cuocci, C., da Silva, I., Giacovazzo, C., Moliterni, A. G. G. & Spagna, R. (2004). *Space-group determination from powder diffraction data: a probabilistic approach.* J. Appl. Cryst. **37**, 957-966.

Altomare, A., Camalli, M., Cuocci, C., da Silva, I. Giacovazzo, C., Moliterni, A. G. G & Rizzi, R. (2005). *Space group determination: improvements in EXPO2004.* J. Appl. Cryst. **38**, 760-767.

Altomare, A., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A. G. G. and Rizzi, R. (2007). *Advances in space-group determination from powder diffraction data.* J. Appl. Cryst. **40**, 743-748.

_____

Structure solution by Direct Methods (DM)

http://www.ba.ic.cnr.it/softwareic/expo/structure-solution-by-direct-methods/

Typical X-ray diffraction experiments provide the structure factor moduli, while the relative phases are lost. Recovery of the phase information is crucial for the *ab-initio* crystal structure solution and is referred in Crystallography as the *phase problem*. In EXPO2014, the default strategy used to solve this problem is based on the application of the Direct Methods (DM) theory (Giacovazzo, 2013). DM are able to estimate phases directly from the structure factor amplitudes. The application of DM to powder data requires the previous decomposition of the full experimental powder pattern to extract single diffraction intensities for each reflection in the measured 2theta range. Owing to the peak overlap the estimates of the diffraction moduli are affected by unavoidable errors: this weakens the efficiency of DM (naively, wrong moduli will produce wrong phases), and still today makes crystal structure solution from powder data a challenge.

Load input file for Structure Solution by Direct Methods

To run EXPO2014 for structure solution you need to create an input file (`*.exp`). You can create an input file accessing by graphic interface by the menu **File > New**.
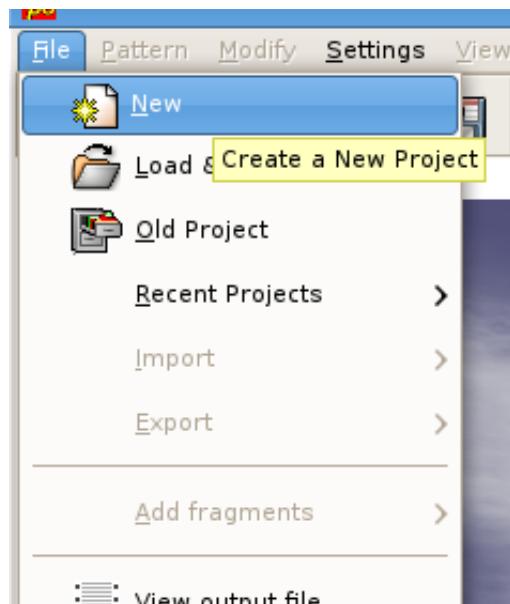
You must specify the structure name, the profile count filename and its format, the type of radiation source. It is supposed that the cell parameters and the space group have been determined before, so fill the frame 'Cell Parameters' and 'Space Group'. Activate the **Structure Solution** check button if you intend to perform the crystal structure solution process by Direct Methods. Otherwise, use the button **Open** to import an existing input file `.exp`; for example you could load the file `cime.exp` already existing in the directory 'examples'. The following picture is an example of the input file for crystal structure determination of the cimetidine compound.

When you press the button **Save,** an input file `cime.exp` is created and automatically loaded by the program for the structure solution process by DM.

```
%structure cime
%job structure cime
%data
pattern cime.pow
content (C10H16N6S)4
wavelength 1.52904
cell 10.6986  18.8181  6.8246  90.000  111.284  90.000
spacegroup p 21/n
synchrotron
%continue
```

The meaning of lines in the input file is contained in the chapter Commands of Expo2014 of this manual.

Alternatively to the creation of the input file by graphic interface, you can edit a new input file or modify an exiting one with a text editor and save it with extension `.exp`. Once a time the input file has been created, load the `.exp` input file from menu **File > Load & Go**.

Press the button [Next >] in the toolbar and the following steps of the structure solution process will be automatically performed:

1) Extraction of the integrated intensities from the powder diffraction pattern;

2) Normalization of the integrated intensities *via* Wilson method (Wilson, 1942);

3) Triplet invariants calculation and estimate of their phases *via* P10 formula (Cascarano *et al.*, 1984);

4) Phasing of structure factors by using the tangent formula: more plausible sets of phases are generated, whose reliability is assessed *via* the CFOM figure of merit (Cascarano *et al.*, 1992). In a default run, not more than twenty sets of phases with the largest CFOM values are stored and ranked in decreasing CFOM order.

5) Fourier map (E-map) calculation and crystal structure optimization by using the normalized structure factor moduli and the set of phases corresponding to the largest CFOM value. The most intense peaks in the E-map are located and chemically interpreted. The obtained structure model is optimized and completed *via* automatic

structure model optimization strategies.

More information about the structure solution process by Direct Methods are given in the chapter The steps of structure solution by DM.

Each step of the solution process by DM corresponds to a specific command in the input file for EXPO2014.
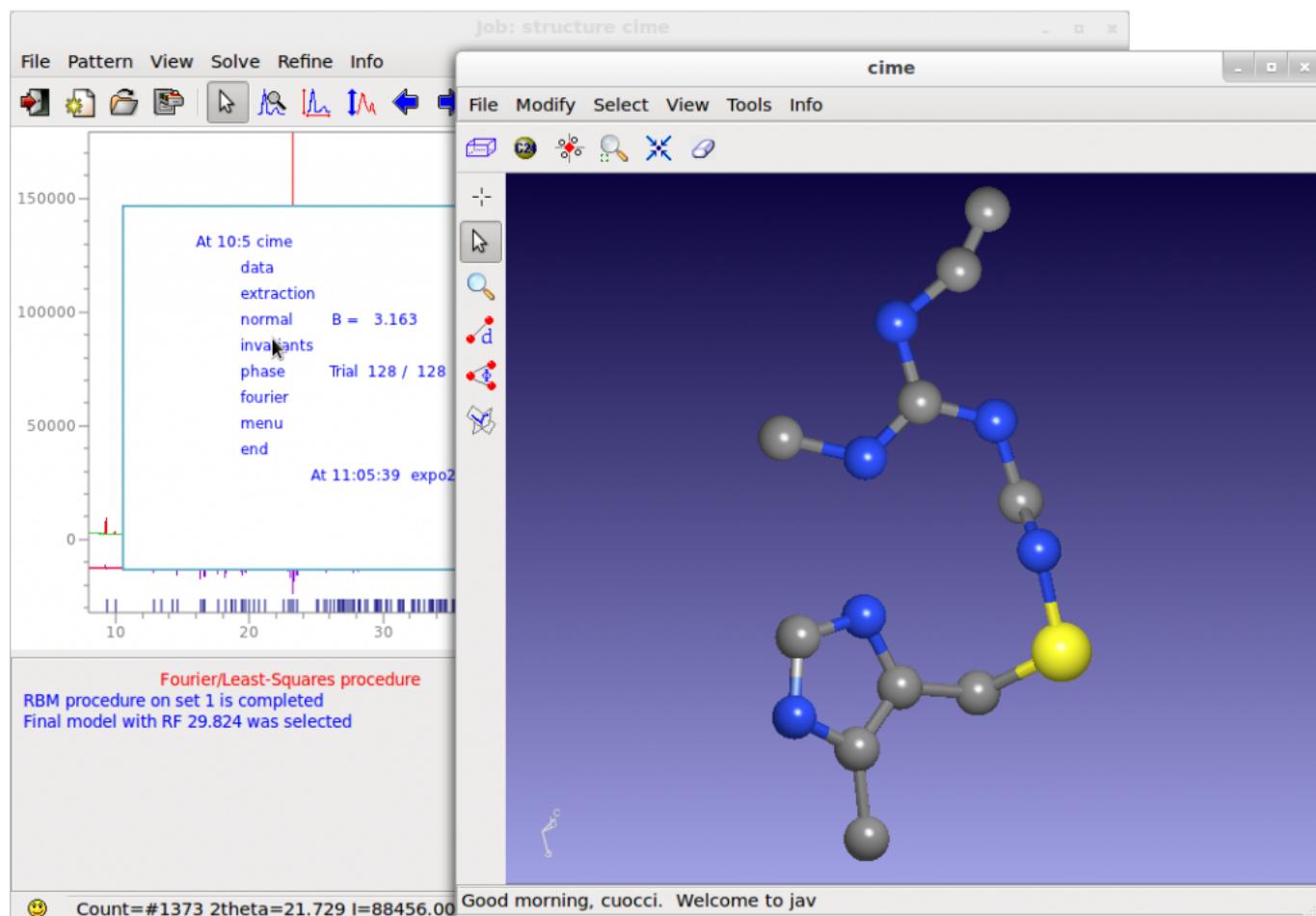
When default EXPO2014 fails: strategies

A lot of crystal structures are solved by EXPO2014 by default settings in few minutes and in completely automatic way.

Unfortunately you can't always obtain the correct structure solution in default. In this situation two different strategies are suggested and can be applied at the end of the default run.

The ALLTRIALS procedure (Altomare *at al.,* 2013)

In a typical Direct Methods procedure more sets of phases (trials) are generated and no more than 20 of them are saved and ranked according to the combined CFOM figure of merit (estimating the quality of the trial) and only the best trial corresponding to the largest CFOM value is used to calculate a Fourier map. Because of unavoidable errors in the phases, the figure of merit can fail to evaluate the best trial, so it is strictly suggested to process the other trials by menu **Solve > Explore Trials**.
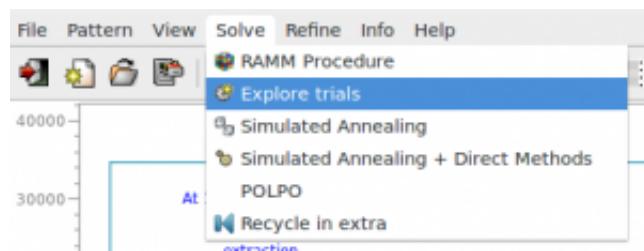
The following window is opened (the procedure has been applied to 2-Mercaptobenzoic acid compound, whose input file, merca.exp, is in the directory 'examples')

and the CFOM values, for each saved trial, can be read. The highest CFOM value could not correspond to the correct solution while subsequent different trials may be successful. This is particularly true when several nearly equivalent CFOM's are available for different trials as in the picture. If the trial has been already developed (done='yes'), you can read the RF figure of merit value that quantifies the agreement between the structure factors extracted from the experimental pattern and calculated by the model corresponding to the current trial.

Check the button in the first column to decide which trial will be developed and press the button [GO]. If you check the button **Explore trials not processed yet**, all the trials not already explored will be automatically selected and developed.

You can choose three different actions for the structure model optimization (see later):

1. RBM: Resolution Bias Modification is the default action for organic and metal organic compounds (Altomare *et al.*, 2008*a*, 2008*b*; 2009, 2010*a*, 2010*b*).
2. Fourier recycling: classical Fourier refinement, default choice for inorganic compounds (Altomare *et al.*, 2006).
3. E-map: the first electron density map calculated directly after the phasing process by Direct Methods.
4. COVMAP: the covariance principle based completion (Altomare *et al.*, 2012).
5. Shift_and_Fix: the method based on random shift of part of the model for improving the solutions (Altomare *et al.*, 2017).

At the end of the procedure, all the trials appear ordered according to the RF value and the user can also view the model corresponding to each trial by mouse selection.
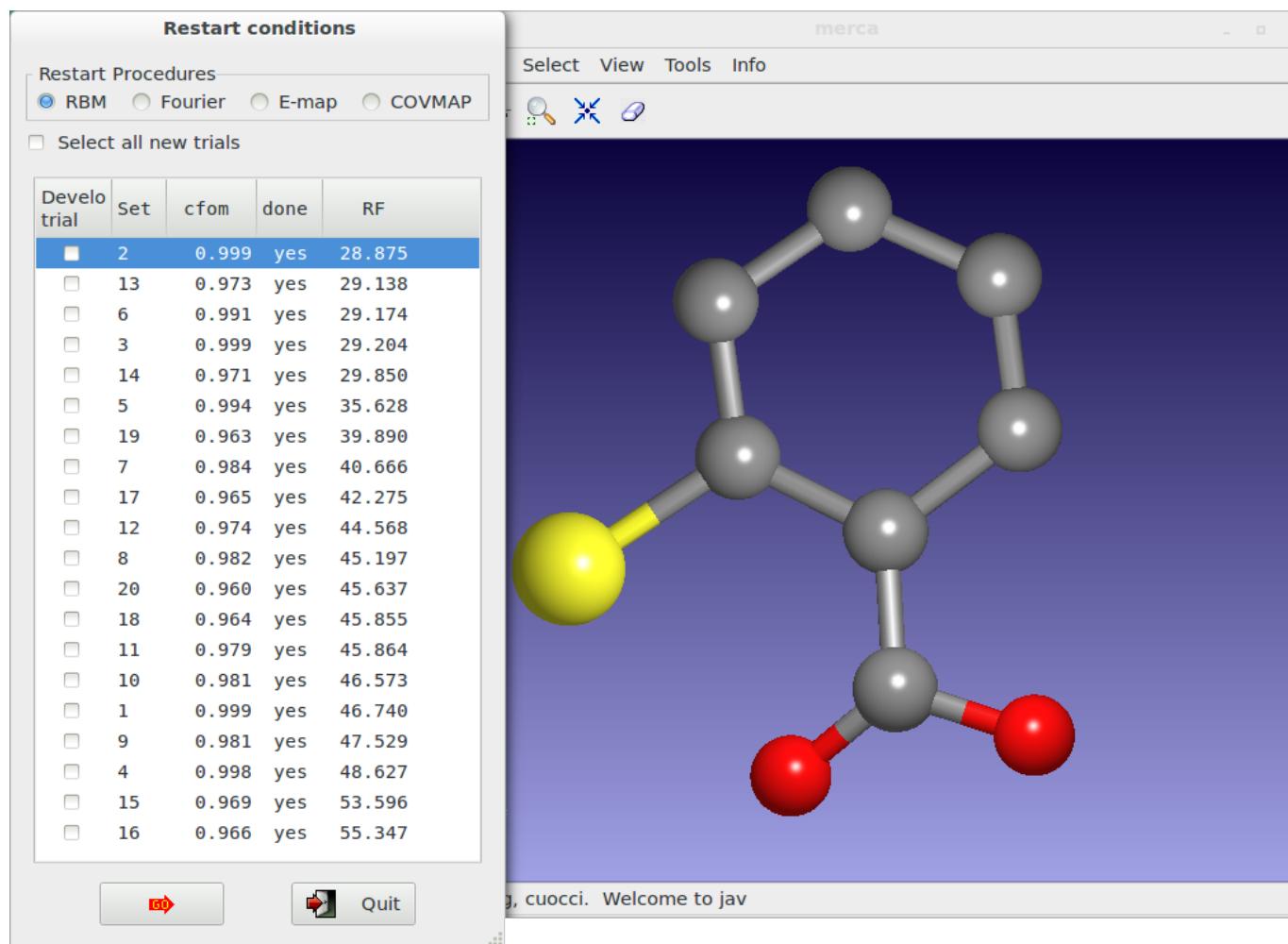
Alternatively to the graphical option, the use of the command `%alltrials` in the input file automatically activate the 'Explore trials' approach. An example of input file is here reported:

```
%structure merca
%job merca
%data
cell       7.885  5.976   14.949  90.0  100.48  90
space p 21/c
content    c 28 O 8 S 4 h 24
pattern    pd_0002.xy
wavelength 1.54056
%alltrials
```

<div align="center">

Structure model optimization

</div>

The structure model obtained at the end of the default strategy, generally is a partially correct representation of the true model: some atoms are in correct positions but other ones are moved with respect to the correct positions or completely false. In this situation the user can try to improve the quality of the Fourier map obtained at the end of Direct Methods, by using different optimization strategies implemented in the EXPO2014 program:

1. suitably weighted least squares (*wLSQ*) (Altomare *et al.*, 2006), able to compensate the low accuracy of the intensities of strongly overlapped reflections. The procedure is automatically applied in case of inorganic compounds.
2. the resolution bias correction algorithm (RBM) (Altomare *et al.*, 2008*a*, 2008*b*; 2009, 2010*a*, 2010*b*) for reducing, in the electron density map, the errors caused by the limited experimental resolution: peak broadening, peak shift, intensity distortion. The correction has been developed in direct space (it represents the default choice in case of organic and metal organic compounds), in reciprocal space and in both direct and reciprocal space;
3. the procedure of electron density modification based on the concept of covariance between points of the map (COVMAP) (Altomare *et al.*, 2012). The procedure executes the following strategy: the structural model provided by the RBM procedure is submitted to COVMAP approach which suitably modifies the electron density map and successively submits the improved model to *wLSQ* analysis. The resulting new model is again modified by COVMAP and cyclically processed in the other two steps. The number of cycles depends on the structure complexity.

It is possible to graphically choose different RBM procedures for model optimization by **Refine > Resolution Bias Modification (RBM)** from the upper EXPO2014 menu.

<div align="center">
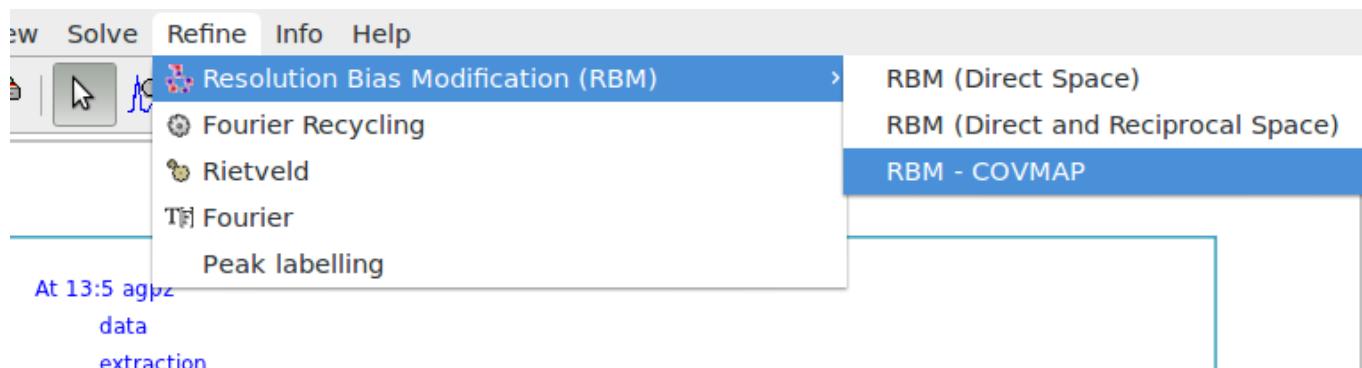
Apply the RAMM (RAndom Model based Method) procedure

</div>

The new method RAMM (Altomare *et al.*, 2013) has been developed and implemented in the EXPO2014 program for improving the *ab-initio* crystal structure solution process. When the available information consists of only the experimental powder diffraction pattern and the chemical formula of the compound under study, the structure solution classical approach follows two main steps: 1) phasing by Direct Methods and obtaining a structure model (this last is usually uncompleted and/or approximate); 2) improving the model by structure optimization techniques. The alternative RAMM approach skips step 1) and supplies a fully random model to step 2). Such model is then submitted to important structure optimization tools present in EXPO2014: *wLSQ* , RBM and COVMAP which are able to lead to the correct structure. RAMM is based on a cyclic process generating several random models which are then optimized. The process stops automatically when recognizes the correct structure.

Click 'RAMM' on the menu 'Solve' to activate this alternative strategy for structure solution at the end of the default run of EXPO2014. The following pictures are related to 2-Mercaptobenzoic acid, whose input file is in the directory `examples`.
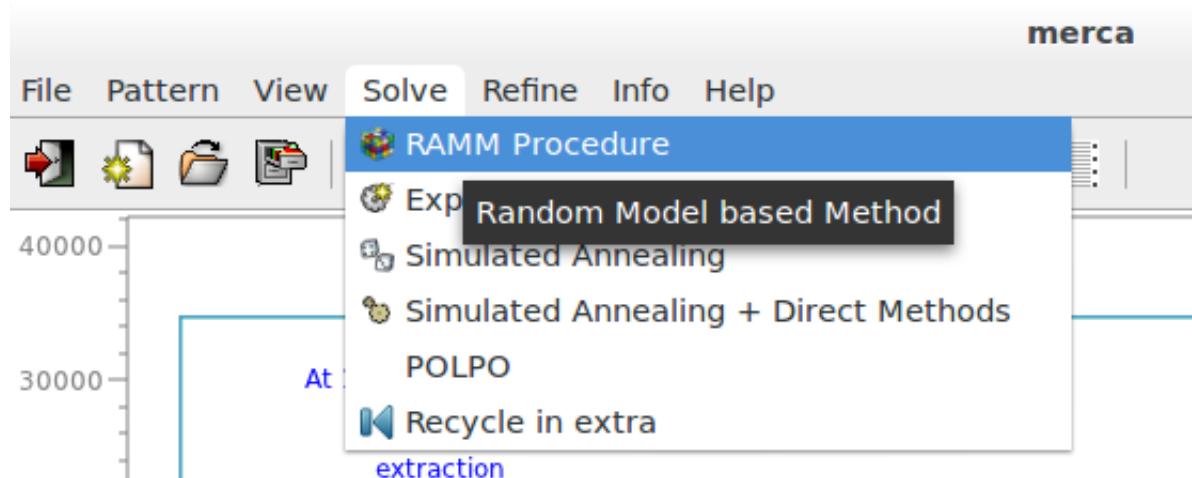
The procedure can take from some minutes to several hours depending on the complexity of problem (number of atoms, quality of data, data resolution). If more than one plausible solution is found they are ordered according to the RF value in an interactive list.

RAMM procedure

Select model generated by RAMM procedure

Run.	Cost function (RF)
14	0.538
6	0.546
2	0.566
13	0.570
5	0.570
9	0.581
8	0.592
3	0.600
15	0.600
11	0.602
12	0.620
10	0.625
4	0.626
1	0.635

OK   Cancel
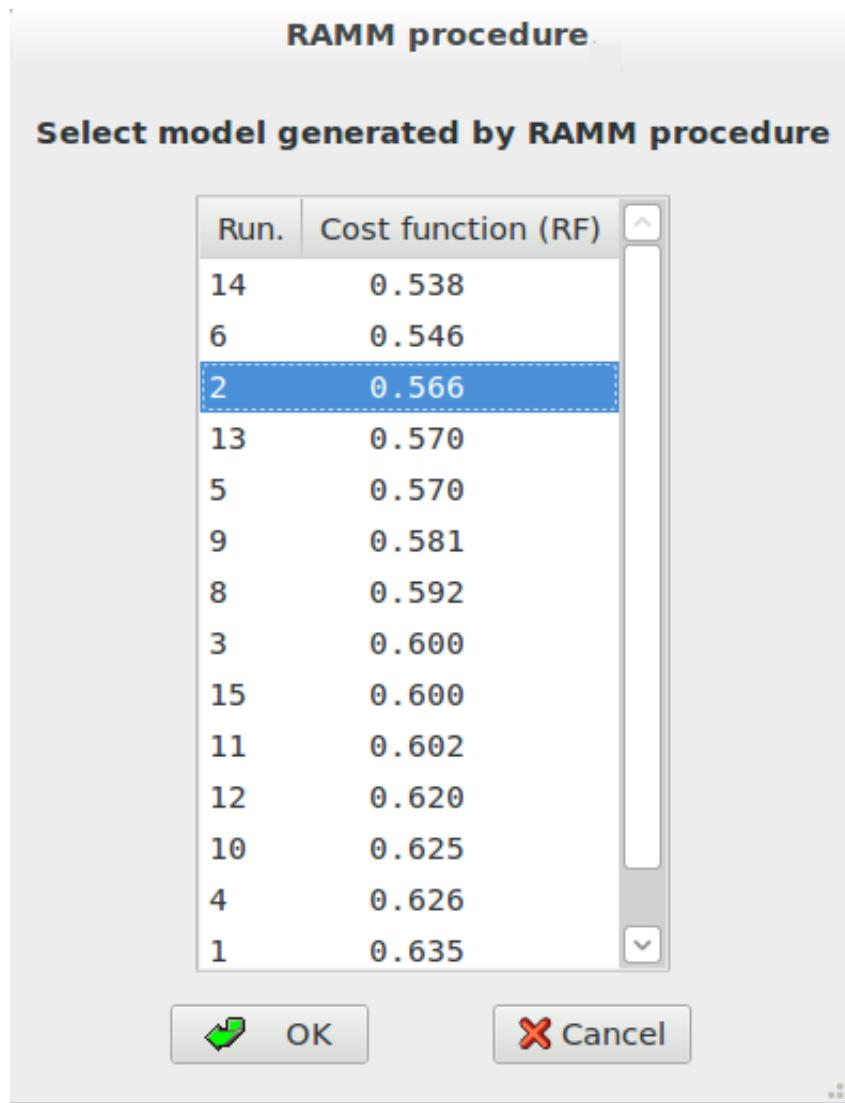
The selected model is visualized on the molecular viewer. Press 'OK' to accept the selected model.
When Direct Methods procedure fails, the RAMM approach can be a very useful tool to find the correct solution.
The use of the command `%randomsolve` in the input file automatically activate the RAMM approach. An example of input file is here reported:

```
%structure merca
%job merca
%data
cell       7.885  5.976   14.949  90.0  100.48  90
spacegroup p 21/c
content    c 28 O 8 S 4 h 24
pattern    pd_0002.xy
wavelength 1.54056
%randomsolve
```

Apply the SHIFT_and_FIX procedure

The Shift_and_Fix method (Altomare et al. 2017) has been implemented in EXPO2014 with the aim of obtaining a complete and correct solution starting from an approximate, sometimes very rough, structure model. It can be activated by menu **Solve>Explore Trials** by selecting the corresponding button in combination with RBM (for organic/metal-organic structure) or Fourier recycling (for inorganic structures) selection. It can be applied to all the Direct Methods trial (suggested) or to preferred trial(s).

For the current trial, it starts from the model generated by Direct Methods and by cycling actions: shifts partly and randomly the model (possibly the most unreliable part of the model); combines Fourier map calculation and least?squares cycles for relocating the shifted atoms onto positions which can finally be moved onto the true ones by the standard model optimization approaches; calculates the Fourier map by using coefficients which depend on the chemical content of the compound under study.

Click on 'Explore Trials' on the menu 'Solve' and select 'Shift_and_Fix' at the end of the default run of EXPO2014.

Suggestions

The number of options in the program is quite large. The optional strategies regard the improvement of both the decomposition process for carrying out more reliable reflection intensities and the Direct Methods performances. We give information about some other secondary strategies.

The improvement of the decomposition process

- It is important to provide good quality diffraction pattern. For example, it is recommended to reduce the pattern range in case of high noise signal at large 2theta values. In the input file for Expo2014, use the directive `range` in the command `%data` to select a portion of pattern or select **Pattern > Range** by the graphical interface.
- If a partial structure fragment has been located by the default run, it may be recycled in the extraction process to improve the extraction of the experimental integrated intensities (a graphic option must be selected). Click on **Solve > Recycle in extra**.
- The use of the `random` directive in the command `%extraction`, in the input file for Expo2014, activates the procedure for overcoming the Le Bail method tendency to equipartition the intensity of a group of strongly overlapped reflections. Based on a Monte Carlo approach, it may provide a set of more statistically meaningful extracted intensities.
- If the default run has revealed pseudo symmetry effects and, in particular, if the percentage of electron density affected by pseudo is large, it may be very useful to exploit that information in the extraction process: a graphic option is made available for activating the procedure.

The optimization of the Direct Methods procedure

- Magic integers procedure is an effective and simple way for starting a multi-solution process.
  A random approach may be alternatively chosen in the phasing process by using the directive `random` in the command `%phase` (in the input file for Expo2014).
- Second representation formula for triplet invariants proved much more powerful than first representation

formula therefore it is used as a default for estimating triplet invariants. In case of failure, the user can choose P3 formula (directive `cochran` in the command `%invariant` in the input file for Expo2014) and/or conveniently increase the number of symbols in the phasing step (directive `symbols` followed by the number of symbols to be used in the command `%phase`).

- A check for possible pseudotranslational symmetry is always made by EXPO2014: if PSEUDO keyword is not activated the information is given but not used in the normalization routine and in the phasing process. The user can decide to use it by the directive `pseudo` in the command `%normal` in the input file for Expo2014.

- The value of the number of reflections actively used in the phasing process is fixed by EXPO2014 (nref). For some special structures the ratio "number of active triplets/nref" is too small (less than 10). Larger values of nref may improve the phasing procedure (directive `nreflections` followed by the number of reflections in the command `%normal`).

- High (or low) resolution reflections may occasionally play an important role in the phasing process by Direct Methods. Fixing an isotropical thermal factor lower (or larger) than that automatically provided by the Wilson plot in the normalization process by Direct Methods may be successful (directive `bfactor` followed by the isotropic thermal factor in the command `%normal`).

- An alternative space group should be carefully considered (directive `spacegroup` in the command `%data`).

- When the complete and correct solution is not attained, successive combined Fourier map calculations (F^o-F^c, $2F^o$-F^c or F^o) and/or Rietveld refinement may be conveniently selected by graphical options. Click on **Refine > Fourier** and select the action.

Contacts

For suggestions and bugs contact:
angela.altomare@ic.cnr.it
annagrazia.moliterni@ic.cnr.it
rosanna.rizzi@ic.cnr.it
corrado.cuocci@ic.cnr.it

References

Altomare, A., Cuocci, C., Moliterni, A., Rizzi, R., Corriero, N., Falcicchio, A. (2017). J. Appl. Cryst. **50** 1812-1820.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A.G.G., Rizzi, R. *(2006).* J. Appl. Cryst. **39** 558-562.

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

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

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2009). *Acta Cryst.* A**65,** 183-189.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2010a). J. Appl. Cryst. **43**, 798-804.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2010). *Z. Kristallogr.* **225**, 548-551.

A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni and R. Rizzi (2012). Acta Cryst. A**68**, 244-255.

A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni, R. Rizzi, N. Corriero and A. Falcicchio (2013). *J. Appl. Cryst.* **46**, 1231-1235.

A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni and R. Rizzi (2013). J. Appl. Cryst. **46**, 476-482.

Cascarano, G., Giacovazzo, Camalli, M., Spagna, R., Burla, M. C, Nunzi, A. & Polidori, G. (1984). *Acta Cryst.*

A**40**, 278-283.

Cascarano G., Giacovazzo C. & Guagliardi A. (1992). *Acta Cryst.* A**48**, 859-865.

Giacovazzo, C. (2013). Oxford: IUCr/Oxford University Press.

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

_____

The steps of structure solution by DM

http://www.ba.ic.cnr.it/softwareic/expo/the-steps-of-structure-solution-procedure/

<div style="background:cyan">**The steps of structure solution by DM**</div>

Using the command %continue, the program runs in default conditions carrying out the commands

%extraction
%normal
%invar
%phase
%fourier
%menu
%end

and each command corresponds to a step in the structure solution with Direct Methods. All the following steps are described.

1)Extraction of the integrated intensities (`%extraction` command)

2)Normalization of the integrated intensities (`%normal` command)

3)Triplet invariants estimate (`%invariants` command)

4)Phasing by the tangent formula (`%phase` command)

5)Fourier map calculation and crystal structure optimization (`%fourier` command)

_____

Extraction of the integrated intensities

http://www.ba.ic.cnr.it/softwareic/expo/extraction-of-the-integrated-intensities/

Extraction of the integrated intensities (`%extraction` command)

The extraction process is devoted to the decomposition of the measured powder diffraction pattern for extracting the single integrated intensities associated to each reflection in the pattern (Altomare et al., 1995). From the integrated intensities the experimental structure factor moduli are derived. To calculate the integrated intensities, EXPO2014 uses the Le Bail algorithm (Le Bail *et al.*, 1988), suitably combined with a least squares procedure minimizing the quantity:

$$\sum_i w_i (y_{oi} - y_{ci})^2 \tag{1}$$

The summation in (1) goes over all the data points,

$w_i = \dfrac{1}{y_i}$

y_{oi} = observed intensity at the i-th step

y_{ci} = calculated intensity at the i-th step

$$y_{ci} = s \sum_h LP_h m_h |F_h|^2 P(2\theta_i - 2\theta_h) A(2\theta_i - 2\theta_h) + y_{bi} \tag{2}$$

The summation over *h* includes neighbouring Bragg **h** reflections,

s =   scale factor

LP_h =   Lorentz – polarization factor

m_h =   reflection multiplicity

F_h =   structure factor

P =   profile function

A =   asymmetry function

y_{bi} =   background intensity at the i-th step.

The background in (2) is automatically estimated and described by the polynomial function (Young, 1993)

$$y_{bi} = \sum_{m=0} bkg_m \left[\frac{2\theta_i}{2\theta_M} - 1\right]^m$$

m, the polynomial degree, is automatically defined by the program as the number which provides the best fit of the background interpolation points.

From the pattern a standard reference peak is automatically selected. The standard peak fitting is performed to provide the starting values of the variables to refine in the least squares procedure.
The full pattern is automatically divided into intervals. The refined parameters derived by the fitting of each interval will be taken into account for the final full pattern fitting.
The following profile shape functions are available for user choice:

$$\frac{\Gamma(\beta_h)}{\Gamma(\beta_h - 0.5)}\left(\frac{C_2}{M}\right)^{\frac{1}{2}}\frac{2}{H_b}(1 + 4C_2 X_{ih}^2)^{\beta_h} \qquad \text{Pearson VII}$$

$$\frac{C_o^{1/2}}{\sqrt{\pi}H_h}exp(-C_o X_{ih}^2) \qquad \text{Gaussian}$$

$$\frac{C_1^{1/2}}{\pi H_h}(1 + C_1 X_{ih}^2)^{-1} \qquad \text{Lorentzian}$$

$$\eta\frac{C_1^{1/2}}{\pi H_h}(1 + C_1 X_{ih}^2)^{-1} + (1 - \eta)\frac{C_o^{1/2}}{\sqrt{\pi}H_h}exp(-C_o X_{ih}^2) \qquad \text{Pseudo-Voigt}$$

where

$$C_0 = 4ln2 \qquad C_1 = 4$$

$$X_{ih} = (\Delta\theta_{ih}/H_h)$$

$$H_h = [w + v \cdot \tan\theta_h + u \cdot \tan^2\theta_h]^{1/2} \quad H'_h = Z\tan\theta_h + T/\cos\theta$$

$$\eta = \eta_0 + \eta_1 \cdot \tan \theta_h + \eta_2 \cdot \tan^2 \theta_h$$
$$\beta_h = \beta_0 + \frac{\beta_1}{2\theta_h} + \frac{\beta_2}{(2\theta_h)^2}$$

The default choice is the Pearson VII function. In this case

$$A(i, \mathbf{h}) = \frac{a \cdot (2\theta_i - 2\theta_\mathbf{h})^3}{[\frac{H_\mathbf{h}^2}{C_3} + (2\theta_i - 2\theta_\mathbf{h})^2]^{3/2}} + 1$$

otherwise

$$A(i, \mathbf{h}) = 1 - \frac{a \cdot sign(2\theta_i - 2\theta_\mathbf{h}) \cdot (2\theta_i 2\theta_\mathbf{h})^2}{tg\theta_\mathbf{h}}$$

The peak shape, the asymmetry, the cell and the background parameters are all refined by the program.

<div align="center">

Directives in the `%extraction` command

</div>

The following directives must be added after the command `%extraction` in the input file to activate some specific non-default procedures:

all
To activate the procedure of selection of the profile function type (among Pearson VII, Gaussian, Lorentzian, Pseudo-Voigt) from the best fit of each interval.

first
To activate the procedure of selection of the profile function type (among Pearson VII, Gaussian, Lorentzian, Pseudo-Voigt) from the best fit of the first interval.

standard
To activate the procedure of selection of the profile function type (among Pearson VII, Gaussian, Lorentzian, Pseudo-Voigt) from the best fit of the standard peak.

gaussian
To select the gaussian function.

lorentzian
To select the lorentzian function.

pvoigt

To select the Pseudo-Voigt function.

interval *n*

n is the number of intervals in which the experimental powder pattern is divided (max 8). *n* is only given if the user wants to change the automatic division of the pattern.

ncycle *n*

n is the number of cycles for each interval (default value is 60).

nosingle

To omit the preliminary fitting of the standard peak (automatically selected by the program). In this case a preliminary fitting of the first interval is carried out.

lowb

To reduce the background by a constant factor equal to the minimum value of the observed counts in the experimental pattern.

noba

To skip the background coefficients refinement.

polynomial *k*

k is the background polynomial degree (max 10). *k* is given if the user wants to use a fixed degree in each interval and not the degree automatically evaluated by the program.

smoothing

To activate the smoothing procedure described by Brückner (Brückner, 2000) for estimating the background. In this case the background coefficients are not refined.

nocell

To skip the cell parameters refinement.

random *p*

To activate the random procedure (Altomare *et al*., 2001) in order to improve the quality of the extracted structure factor moduli values. *p* is the coefficient for the reflection overlapping definition (default value is 0.5).

frecycle *filename.hkl*

If the file `filename.hkl,` containing hkl and the corresponding F-squared values, is available from a previous run of EXPO2014, or from any other source, the Le Bail algorithm starts, for a new decomposition process, by using the provided F-squared values as starting values.

damping *x*

x is the damping factor used in the least squares routine (default value 0.1). (It is convenient to use a small *x* value).

width *x*

x is the factor which, multiplied by the FWHM of each peak in the experimental pattern, gives the total peak width. (In this case the automatic calculation of *x* is not taken into account).

zeroprofile *x*

x is the 2? zero shift value supplied by the user (it is used to correct the 2? observed positions of the experimental

counts). In this case the automatic calculation of the 2? zero shift is skipped.

slow

To stop the run of EXPO2014 after some relevant calculations and wait for the user graphic interaction to go on. This directive can be used if the graphic interface is active.

verbose

To have a more detailed print in the output file. In this case, the Rp and Rwp values and the values of the refined variables are printed interval by interval.

Examples

Example 1

A Pseudo-Voigt function is used and the background in each interval is described by a constant, the graphic window is suppressed.

```
%nowindow
%structure mes
%job MES - data from home diffractometer
%data
pattern mes.pow
cell 8.588 9.931 11.105 90.0 93.754 90.0
content C 24 N 4 O 20 S 4 H 52
spacegroup p 21/c
wavelength 1.5406
%extraction
pvoigt
polynomial 0
%continue
```

Example 2

The starting values of the integrated intensities in the extraction step by the Le Bail formula are supplied by the user in the file "*yono.hkl*" and the number of least-square cycles is set to 40.

```
%structure yono
%job YONO - Synchrotron data
%data
pattern yono.pow
cell 9.385 16.394 3.627 90.000 101.067 90.000
content Y 8 O 26 N 2 H 18
spacegroup p 21
wavelength 1.2323
synchrotron
%extraction
frecycle yono.hkl
ncycle 40
%continue
```

Example 3

The preliminary fitting of the standard peak is skipped and the starting 2? zero shift value is supplied by the user (the program refines it).

```
%structure agpz
%job AGPZ - data from home diffractometer
%data
pattern agpz.pow
cell 6.526 20.059 6.464 90.000 90.000 90.000
spacegroup p b c a
content Ag 8 N 16 C 24 H 24
wavelength 1.54056
%extraction
nosing
zeroprof 0.02
%continue
```

References

Altomare A., Burla M.C., Cascarano G., Giacovazzo C., Guagliardi A., Moliterni A.G.G. and Polidori G. (1995). *J. Appl. Cryst.* **28**, 842-846.

Altomare A. , Giacovazzo C., Moliterni A. G. G. and Rizzi R. (2001). *J. Appl. Cryst.* **46**, 476-482.

Brückner S. (2000). *J. Appl. Cryst.* **33**, 977-979.

Le Bail, A., Duroy, H., Fourquet, J.L. (1988). Math. Res. Bull. **23**, 447-452.

_____

Normalization of the integrated intensities

http://www.ba.ic.cnr.it/softwareic/expo/normalization-of-the-integrated-intensities/

Normalization of the integrated intensities (`%normal` command)

In the normalization procedure the diffraction intensities are normalized using the Wilson method (Wilson, 1942). Statistical analysis of the intensities is made in order to suggest the presence or absence of the inversion center, to identify the possible presence and type of pseudotranslational symmetry (Cascarano *et al.* 1988 a,b; Fan, Yao *et al.* 1988) and to detect preferred orientation effects (Altomare *et al.* 1994; Altomare *et al.* 1996). Possible deviations (of displacive type) from ideal pseudotranslational symmetry are also detected. Information about the presence of pseudo symmetry, can be used as prior information for a new integrated intensities extraction process.

Directives in the `%normal` command

The following directives must be added after the command `%normal` in the input file to activate specific non-default procedures:

bfactor *x*
Temperature factor if the user wants to supply it. (The scale factor is assumed equal to 1).

nopreferred
The program does not apply the correction for preferred orientation (if detected).

yespseudo
The program recycles the information about pseudotranslational symmetry for a new extraction process (see Example 1 for its use).

nreflections *n*
Number of active reflections with the largest E-values subject to a minimum value of E(obs) = 1.0.
Default = 4 * number of independent atoms + 50 if centrosymmetric + 50 if triclinic + 150 to a maximum of 499. If **pseudo** directive is on, the program tries to use all the |E|'s down to 1.0.

nzro *n*
Number of the smallest E-values (maximum E-value=0.3) to calculate psi-zero triplets (used in the calculation of the figures of merit in the phasing process). The default (and the maximum) is 1/3 of the strong reflections. (The directive allows to decrease that number).

npla *n*
n is the number of candidate planes for preferred orientation analysis (default 10; maximum 30).
The number includes the equivalent planes which will be not considered.

plane *h k l G*
Preferred orientation direction supplied by the user; *G* is the correction factor.

pseudotranslation *n(1,1) n(2,1) n(3,1) n(4,1) n(1,2) . . . n(4,3)*
If parameters are equal to zero, or omitted, the normalization routine performs statistics for the pseudotraslation

effects and renormalizes the reflections on assuming the most probable pseudotranslation as prior information. If user wants to supply a specific pseudotranslational symmetry, and this corresponds to a class given in Table 1, $n(1,1)$ can be set to the class number and the other values must be omitted.

If the pseudotranslation is more complex, it should be specified by up to 3 sets of 4 values $n(i,j)$ so that:

$$[n(1,j)*h + n(2,j)*k + n(3,j)*l = n(4,j)*m] \qquad j=1,..,3$$

where m is an integer number.

thrp x

Threshold value of the mean fractionary scattering power to recycle the pseudotranslational symmetry information for a new extraction process.

Patterson map calculation (`%patterson` command)

In this module the Patterson map can be calculated by using the normalized (or non-normalized) structure factor moduli as coefficients. By using the directive **inversion,** it is possible to exploit the information of positivity of the electron density (in the direct space) in the extraction routine: the Patterson map is modified and inverted, providing a set of structure factor moduli values to use for a new extraction process.

Directives in the `%patterson` command

The following directives must be added after the command `%patterson` in the input file to activate specific non-default procedures:

e*f (or f*e)

Coefficients to be used in the Patterson synthesis (**e**2** is the default).

f**2 (or f*f)

Coefficients to be used in the Patterson synthesis (**e**2** is the default).

inverse

To activate the procedure of calculating the |F| values from an inverted suitably modified Patterson map (Altomare *et al*., 1998). The values are then used as starting point in the Le Bail algorithm for extracting new structure factor moduli from the experimental pattern (see Example 2 for its application).

map, layx, layy, layz, peaks, limit, grid

See the corresponding directives in the `%fourier` command.

npat n

The number of cycles of successive inversions of the Patterson map. (Default values: 6 for X-ray data, 5 for neutron data).

Examples

Example 1

Use of the pseudotranslational symmetry (if it is present) as prior information for a new extraction process.

```
%structure agpz
%job AGPZ - data from home diffractometer
%data
    pattern agpz.pow
    cell 6.526 20.059 6.464 90.000 90.000 90.000
    spacegroup p b c a
    content Ag 8 N 16 C 24 H 24
    wavelength 1.54056
%extraction
%normal
    yespseudo
%continue
```

Example 2

Use of the Patterson map inversion as prior information for a new extraction process.

```
%structure lasi
%job lasi - Neutron data
%data
        pattern lasi.pow
        cell 5.4059 8.7934 14.2754 90.000 112.731 90.00
        content La 8 Si 8 O 8
        spacegroup p 21/c
        wavelength 2.3400
        neutron
%extraction
%normal
%patterson
        inverse
%continue
```

Table 1

Classes of reflections corresponding to the low index pseudotranslational symmetry.

1)	h+k+l= 1n	2)	h= 2n	3)	k= 2n
4)	l= 2n	5)	h+k+l= 2n	6)	h+k= 2n
7)	h+l= 2n	8)	k+l= 2n	9)	h= 3n
10)	k= 3n	11)	l= 3n	12)	h+k= 3n
13)	h+l= 3n	14)	k+l= 3n	15)	h+k+l= 3n
16)	h+k+2l= 3n	17)	h+2k+l= 3n	18)	2h+k+l= 3n
19)	h+2k= 3n	20)	h+2l= 3n	21)	k+2l= 3n

22)	$l = 4n$	23)	$k = 4n$	24)	$h = 4n$
25)	$h+k = 4n$	26)	$h+l = 4n$	27)	$k+l = 4n$
28)	$h+k+l = 4n$	29)	$2h+2k+l = 4n$	30)	$2h+k+2l = 4n$
31)	$h+2k+2l = 4n$	32)	$2h+k+l = 4n$	33)	$h+2k+l = 4n$
34)	$h+k+2l = 4n$	35)	$h+2k = 4n$	36)	$h+2l = 4n$
37)	$k+2l = 4n$	38)	$2h+k = 4n$	39)	$2h+l = 4n$
40)	$2k+l = 4n$	41)	$3h+3k+l = 4n$	42)	$3h+k+3l = 4n$
43)	$h+3k+3l = 4n$	44)	$h+2k+3l = 4n$	45)	$h+3k+2l = 4n$
46)	$3h+k+2l = 4n$	47)	$h+3k = 4n$	48)	$h+3l = 4n$
49)	$k+3l = 4n$	50)	$3k+2l = 6n$	51)	$2k+3l = 6n$
52)	$2h+3k = 6n$	53)	$3h+2k = 6n$	54)	$3h+2l = 6n$
55)	$2h+3l = 6n$	56)	$2h+2k+3l = 6n$	57)	$3h+2k+3l = 6n$
58)	$3h+3k+2l = 6n$	59)	$4k+3l = 12n$	60)	$4h+3l = 12n$
61)	$4h+3k = 12n$	62)	$3k+4l = 12n$	63)	$3h+4k = 12n$
64)	$3h+4l = 12n$	65)	$h = 2n \ \& \ k = 2n$	66)	$h = 2n \ \& \ l = 2n$
67)	$k = 2n \ \& \ l = 2n$	68)	$h = 2n \ \& \ k+l = 2n$	69)	$k = 2n \ \& \ h+l = 2n$
70)	$l = 2n \ \& \ h+k = 2n$	71)	$h+k = 2n \ \& \ h+l = 2n$	72)	$h = 2n \ \& \ k = 2n \ \& \ l = 2n$

References

Altomare A., Cascarano G., Giacovazzo C., Guagliardi A., Burla M.C., Polidori G. & Camalli M. (1994). *J. Appl. Cryst.* **27**, 435 – 436.

Altomare A., Burla M.C., Cascarano G., Giacovazzo C., Guagliardi A., Moliterni, A.G.G. & Polidori, G. (1996). *J. Appl. Cryst.* **29**, 341-345.

Altomare A., Foadi J., Giacovazzo C., Moliterni A.G.G., Burla M.C. and Polidori G. (1998). *J. Appl. Cryst.* **31**, 74 – 77.

Cascarano G., Giacovazzo C. & Luic' M. (1988a). *Acta Cryst.* A**44**, 176-183.

Cascarano G., Giacovazzo C. & Luic' M. (1988b). *Acta Cryst.* A**44**, 183-189.

Fan H., Yao J. & Qian J. (1988). *Acta Cryst.* A**44**, 688-691.

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

_____

Structure invariants estimate

http://www.ba.ic.cnr.it/softwareic/expo/estimate-of-structure-invariants/

Structure invariants estimate (`%invariants` command)

Up to 10000 triplets relating reflections with normalized E values greater than a given threshold (strong triplets) are stored for active use in the phasing process. Also triplets (psi-zero triplets) relating two reflections with large |E| and one with |E| close to zero are generated: they are used to define a special figure of merit (PSCOMB).

Negative quartets are generated by combining the psi-zero triplets in pairs, and those with cross-magnitudes smaller than a given threshold are estimated by means of their first representation, as described by Giacovazzo (1976). These quartets are used to provide an important contribution to the CPHASE figure of merit.

Active triplets may be estimated according to Cochran's (1955) P3 distribution; the concentration parameter of the von Mises distribution is given by:

$$C = \frac{2 \cdot |E_h E_k \cdot E_{h-k}|}{\sqrt{N}}$$

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

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

where q is a function (positive or negative) of all the magnitudes in the second representation of the triplet. The G values are rescaled on the C values and the triplets are ranked in decreasing order of G. The top relationships represent a better selection of triplets with phase value close to zero than that obtained when ranking according to C. These triplets will be actively used in the phase determination process.

Triplets characterized by a negative G represent a sufficiently good selection of relationships close to 180 degrees, to be used for the calculation of a powerful FOM (CPHASE). Triplets with G close to zero are expected to have values widely dispersed around 90 or 270 degrees and are used to compute an enantiomorph sensitive FOM. A similar FOM is also computed using quartets estimated with a very small concentration parameter.

As a default, triplets are estimated according to P10 formula.

The parameter C of the Cochran distribution (say P3) is suitably modified when the pseudo information in the section Normalization procedure is available. Then triplet phases are no longer expected to be around zero (see quoted references) and may lie anywhere between 0 and 2?.

Directives in the `%invariants` command

The following directives must be added after the command `%invariants` in the input file to activate specific non-default procedures:

gmin *x*

Positive triplets with G < x are not actively used.

Default value of x is 0.6 (in any case 0.2 < x ? 1.0 is an accepted range of value) (see the <u>Example 1</u> for its use).

cochran

To use the P3 formula. (By default P10 formula is used).

emax x

Maximum value of E for the weak reflections for the psi-zero search (up to a maximum of **nrpsizero** reflections). x > 0.3 not accepted.

Minimum value for the normalized structure factors in the sigma2 search. x < 1.0 not accepted.

The default value is the minimum value of |E| for the strongest reflections selected by the normalization procedure.

npsizero

To not calculate psi-zero triplets (default uses them).

nrpsizero n

The number of weakest reflections for the psi-zero search.

Default = number computed by normalization routine.

nrtriplets n

The number of strongest reflections for the sigma2 search (up to Emin = 1.0).

Default = number computed by normalization routine (nstrong).

Value less than nstrong/2 is not accepted.

nlarge n

To supply the number of largest reflections for the psi-zero search (default = number of strongest reflections used for sigma2 search).

nlarge > nrtriplets is not accepted.

nquartets

To not calculate negative quartets (default uses them).

numk n

The number of free-vectors used by P10 formula.

Default value is provided by EXPO2014 according to the structural complexity:

n = min (natom / msym + 30, 70)

where natom is the number of atoms in the cell and msym is the number of symmetry operators of the space group.

Example

Example 1

Use of gmin directive.

```
%structure mes
%job MES - data from home diffractometer
%data
        pattern mes.pow
        cell 8.588 9.931 11.105 90.0 93.754 90.0
```

```
      content C 24 N 4 O 20 S 4 H 52
      spacegroup p 21/c
      wavelength 1.5406
%extraction
%normal
%invariants
      gmin 0.4
%continue
```

References

Cascarano G., Giacovazzo C., Burla M.C., Nunzi A. & Polidori G. (1984). *Acta Cryst.* A**40**, 389-394.
Cochran W.(1955). *Acta Cryst.* **8**, 473-478.
Giacovazzo C. (1976). *Acta Cryst.* A**32**, 958-966.

_____

Phasing by tangent formula

http://www.ba.ic.cnr.it/softwareic/expo/application-of-the-tangent-formula/

Phasing by the tangent formula (`%phase` command)

In the EXPO2014 program the most reliable one-phase structure seminvariants (s.s.) are treated as known phases. Besides triplets, also the most reliable negative quartets and two-phase s.s. may be actively used.

Each relationship is used with its proper weight: the concentration parameter of the first representation for quartets and two-phase s.s., and C or G for triplets.

– CONVERGENCE/DIVERGENCE PROCEDURE –

The convergence procedure (Germain *et al.* 1970) is a convenient way of defining an optimum starting set of phases to be expanded by the tangent formula or by any other algorithm.

When the P10 formula is used, as a default, a special convergence process is devised which chooses the starting set according to

$$\langle \alpha \rangle = \sum_j [G_j \cdot D_1(G_j)] \cdot D_1(\langle \alpha_{h-k} \rangle)$$

(1)

as suggested by Giacovazzo (1979) and by Burla *et al.* (1987), with

$$D_1(G_j) = \frac{I_1(G_j)}{I_0(G_j)}$$

I1 and I0 represent modified Bessel functions of order one and zero respectively. The summation in (1) is over all relationships defining the reflection **h**. If P3 formula is used the default choice is

$$\langle \alpha \rangle = \sum_j [C_j \cdot D_1(C_j)]$$

Once the starting set has been defined, a good pathway for phase expansion is determined by a divergence procedure. In the divergence map, starting from the reflections in the starting set, each new reflection is linked to the preceding ones with the highest value of $\langle \alpha \rangle$.

– PHASE EXTENSION AND REFINEMENT –

The starting set defined by the preceding step is usually formed by the origin (and enantiomorph) fixing reflections, a few one phase s.s. and a number of other phases which may be obtained:

a) by magic integer permutation (White *et al.* 1975; Main, 1978),

b) by a random approach (Baggio *et al.* 1978; Burla *et al.* 1992).

The option a) is the default, b) runs if the directive RANDOM is used. In this last case a large number (depending on the available computer time) of trials can be requested.

Directives SYMBOLS, SPECIALS, MAXTRIAL may be used to change default values.

Phase expansion and refinement are carried out by means of a tangent formula using triplets, negative quartets, psizero triplets and the most reliable two-phase structure seminvariants. In the weighting scheme the experimental distributions of the alpha parameters are forced to match with the theoretical ones (Burla *et al.* 1987).

For each phase set, several FOM's are computed using all invariants and seminvariants estimated by means of the representation method. Their meaning and an optimized way of combining all the computed FOM's to give a highly selective combined figure of merit (CFOM) is described in the papers by Cascarano *et al.* (1987, 1992).

All FOM's, as well as the combined CFOM, are expected to be equal to 1.0 for correct solutions. CFOM larger than 0.5 can be considered encouraging.

If pseudotranslational symmetry is present, CFOM>0.3 may characterize the correct solution.

Directives in the %phase command

The following directives must be added after the command `%phase` in the input file to activate specific non-default procedures:

enantiomorph *n*
Code of the reflection chosen by the user to fix the enantiomorph (by default the enantiomorph is chosen by the program).

list *n*
The number of reflections at the top of the divergence map for which a list of the map is obtained. Default = no printout of the map.

maxtrials *n*
Maximum number of trials when random approach is used (default value = 100). Only if RANDOM directive is on. (See Example 1 for its use).

minfom *x n*
x: the program automatically stops when a solution is found with CFOM > x . Default value = 1.00.
n: is the maximum number of phase sets (with the largest combined figure of merit – CFOM) to retain (default value = 20). (See Example 1 for its use).

noreject
Used to retain equivalent sets of phases.

origin *n(i) phi(i)*
Codes and phases of the reflections chosen by the user to fix the origin. The total number of reflections (max 3) needed to fix the origin must be given (by default the origin is chosen by the program).

permute *n(i)*
Codes of the permuted reflections chosen by the user (i.e. included in the number of required symbols, max 12).

phase *n(i) phi(i) wt(i)*
Codes, phases and weights of reflections with known phase (max 200). If one card is not sufficient, the directive **phase** must be repeated in the other card(s).

random *n*

To use random phases starting sets. (See Example 1 for its use).

n: is the number of random phases. If n is omitted, it is equal to one half of the strong reflections (nstrong).

Maximum value allowed for n = nstrong – 4

Minimum value allowed for n = nstrong / 4

Default value for random generation seed is 67543.

(Not consistent with SYMBOLS directive)

specials *n*

The number of reflections with restricted phase to be permuted (i.e. included in the number of required symbols, max 12).

symbols *n*

The number of permuted phases (max 12). Default value = 5.

(Not consistent with RANDOM directive)

table

To print the table of alpha values (at the beginning of PHASE procedure).

trial *n*

To print the final phases for the specified trial n .

yneg

To actively use negative triplets in the phasing process (default does not use them).

Example

Example 1

A random approach is used and the best 100 sets of phases, over 250 trials, are retained.

```
%structure nizr
%job nizr structure
%data
pattern nizr.pow
cell 12.388  8.9269  8.8449  90.0  90.550  90.0
cont zr 8 p 12 o 48 ni 4
space p 21/n
wave 1.0302
synch
%extraction
%normal
%invar
%phase
    random
    maxtrials 250
    minfom 1.0 100
%continue
```

References

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

Burla M.C., Cascarano G., Giacovazzo C., Nunzi A. & Polidori G. (1987). *Acta Cryst.* A**43**, 370-374.

Burla M.C., Cascarano G. & Giacovazzo C. (1992). *Acta Cryst.* A**48**, 906-912.

Cascarano G., Giacovazzo C. & Viterbo D. (1987). *Acta Cryst.* A**43**, 22-29.

Cascarano G., Giacovazzo C. & Guagliardi A. (1992). *Acta Cryst.* A**48**, 859-865.

Germain G., Main P. & Woolfson M.M. (1970). *Acta Cryst.* B**26**, 274-285.

Giacovazzo C., (1979). Acta Cryst. **A35**, 757-764.

Main P. (1978). *Acta Cryst.* A**34**, 31-38.

White P.S. & Woolfson M.M. (1975). *Acta Cryst.* A**31**, 53-56.

_____

Fourier map calculation and crystal structure optimization

http://www.ba.ic.cnr.it/softwareic/expo/crystal-structure-completion-and-refinement/

Fourier map calculation and crystal structure optimization (`%fourier` command)

The sets of phases generated by the tangent routine are passed to the fast fourier transform routine written by L.F.Ten Eick (1977) and subsequently modified by the MULTAN team (Main, 1978; Main *et al.,* 1980).

In the default run, the best phasing set selected by CFOM is processed. When the first ranked CFOM set does not provide a reasonable structure model, see the ALLTRIALS strategy for exploring all the stored phasing sets.

The E-map, whose coefficients are the normalized experimental structure factor moduli and the phases are obtained by Direct Methods, is calculated. The map is then chemically interpreted and the structure model is recovered.

Special positions are handled, peaks very close to symmetry elements are moved onto symmetry elements, the site symmetry is defined and the crystallographic occupancy factor is evaluated.

As a next step, for improving the structure model derived from the E-map, a Fourier recycling strategy is applied and then a structure model optimization process is carried in a default way by

1. suitably weighted least squares (*w*LSQ) (Altomare *et al.*, 2006). The procedure is automatically applied in case of inorganic compounds;

2. the resolution bias correction algorithm (RBM) (Altomare *et al.*, 2008a, 2008b; 2009, 2010a, 2010b). It represents the default choice in case of organic and metal organic compounds.

At the end of the Expo2014 run, non default model optimization strategies can be applied if the model is incomplete and/or very rough.

It is possible to graphically choose different RBM procedures for model optimization by **Refine > Resolution Bias Modification (RBM)** from the upper EXPO2014 menu.

A non standard approach (directive **edmo**) has been introduced in EXPO2014 to improve the phasing process and consequently for optimizing the electron density map. Based on real space techniques (Altomare *et al.*, 2002) originally used for phase extension and refinement of proteins (Giacovazzo & Siliqi, 1997; Burla *et al.*, 2000), it combines an Electron Density Modification (EDM) procedure with Fourier map calculations, enabling to reduce the errors on the Direct Methods phases of the strong reflections (*i.e.*, with E>1.0) and to extend the phases also to the subset of reflections with E larger than 0.8. The structure model obtained by the E-map calculated with all the phased reflections is automatically refined by the structure model optimization process. The procedure can be carried out both for X-ray and neutron data; in this last case it has to be applied only in case of crystal structures with not negative scattering factor atoms.

Directives in the `%fourier` command

The following directives must be added after the command `%fourier` in the input file to activate specific non-default procedures:

edmo
To enable the Electron Density Modification (EDM) approach, followed by an automatic structure model optimization procedure. Except for the Fourier directives **set** and **recycle 0**, all the rest of the Fourier directives are not valid for the EDM procedure. The directive **recycle 0** has to be used to require the E-map calculation only (the automatic structure model optimization is not carried out).

fragment *string*

To supply a known structural fragment to be completed. s*tring* is the name of the file in which, for each atom, are stored: chemical element, atomic coordinates (x, y, z), isotropic thermal factor.

fwhm *D*

To change the value of *D* which is the parameter for the reflection overlapping condition: two reflections are considered to be in overlapping if $2\theta_1 - 2\theta_2 < D*FWHM$ ($2\theta_1$ and $2\theta_2$ are the peak positions of the two reflections).

Default value is approximately 0.9 (depending on the ratio between the number of observations and the number of parameters to refine).

grid *x*

To define the grid of the Fourier map. The default value of *x* is computed by the program (approximately 1/3 Å).

integration *x n*

x is the radius of integration for searching peaks in the electron density map. (default value is 0.5). *n* may be 1 for using the gaussian fitting of the Fourier peak in the map or 2 for using the peak average density. (default value is 2).

layx

To print the Fourier map in section of constant x. This directive can be used only if **map** directive is given.

layy

To print the Fourier map in section of constant y. This directive can be used only if **map** directive is given.

layz

To print the Fourier map in section of constant z. This directive can be used only if **map** directive is given.

level *n*

All numerical values in the E-map greater than *n* will be underlined with **** in the line-printer output to facilitate contouring.

The default is 100. The map is automatically scaled between ± 999 (approximately) at the grid points at which it is calculated.

Note, however, that it is printed between ± 99. (This directive can be used only if associated to **map** directive).

limits *x1 x2 x3*

To print the E-map from 0 to *x1* along x, from 0 to *x2* along y, from 0 to *x3* along z. (This directive can be used only if associated to the **map** directive).

map

To print the E-map on the line-printer. The printing of the E-map is rarely used.

The output is in sections of constant y with x across the page and z running down it on a grid of about three points per angstrom.

The maximum number of grid points across the page is 36 (i.e. about 12 angstroms in length). If more are needed, a new page is started.

peaks *n*

The number of peaks to be searched for in the E-map. The default is the number of peaks which

$\sum(occ(i)*m) = 1.3 * n$

where occ(i) is the crystallographic site occupancy factor of the i-th peak, m is the number of symmetry operators,

n is the number of non-hydrogen atoms in the unit cell.

radius *El x*

The value *x* of the radius assigned to the chemical element *El* to evaluate the connectivity.

recycle *n*

To stop the Fourier recycling procedure after the cycle number *n*. If *n* is equal to zero only the E-map is calculated.

set *n*

The serial number of the phasing set from which the E-map is calculated. If '**set** *n*' is not specified, the set with the highest CFOM combined figure of merit is used, and then if the %fourier command is used again r, the set with the next highest combined CFOM will be used.

sogf *x*

The percentage of the reflections with the largest |F| structure factor moduli actively used in the Fourier recycling routine.

The reflections with the smallest |F| values corresponding to the rejected percentage are considered with zero weights in the refinement process (default value is 100.0).

sogi *x*

The percentage of the most intense groups of overlapping reflections actively used in the Fourier recycling routine (default value is 91.0).

thre *x*

x is the threshold for the E value of reflections used in the Electron Density Modification procedure. x should be positive (its default value is 0.8).

thrm *x*

x is the fraction of the largest intensity Fourier map points used in the Electron Density Modification procedure (0< x ? 1.0). Its default value is 0.025.

<div align="center">

Examples

</div>

Example 1

Only the E-map corresponding to the set number 7 and the complete E-map calculation and structure model optimization procedure for set number 10 are requested by the user.

```
%structure nbpo
%job titolo nbpo.dat
%data
pattern nbpo.pow
content Nb 20 O 120 P 28
wavelength 1.0001
cell    29.8661   8.7215    8.7860   90.000    91.769    90.000
spacegroup c 2/c
synchrotron
%extraction
%normal
%invar
```

```
%phase
%fourier
set 7
recyc 0
%fourier
set 10
%continue
```

Peak labelling procedure (`%changelabel` command)

The frequent incorrect labelling (in term of atomic species) of the atomic positions, as obtained for example at the end of a Direct Methods procedure, can strongly affect the efficiency of the approaches used for crystal structure refinement. In this situation, the correct structural model is difficult to establish. The atomic label can be automatically modified using chemical information about the structure. The procedure (Altomare *et al.*, 2002) requires: 1) the prior information on the number and coordination of the heavy atoms (the procedure can only handle tetrahedral and octahedral coordinations); 2) the range of the typical distances between heavy atoms; 3) the range of the typical distances between heavy and light atoms.

The peak labelling procedure starts only from a list of atomic positions, which are supplied by an external file (they can be selected from the Fourier map of Expo2014).

The command `%changelabel` must follow the commands `%extraction` and `%fragment`.

Directives in the `%changelabel` command

The following directives must be added after the command `%changelabel` in the input file to activate the peak labelling procedure:

dlight *d tol*

d is the expected heavy atom (*i.e.* atomic species for which Z>10) to light atom distance;

tol is the tolerance around the ideal distances. Default value = 0.5.

The program uses only one value of 'd' for all types of light atoms.

dweight *d tol*

d is the expected heavy atom to heavy atom distance;

tol defines the tolerance around the ideal distances. Default value = 0.5.

The program uses only one value of 'd' for all types of heavy atoms.

label *n1 El1 n2 El2*

n is the number of atoms to relabel using the El element;

El is the atomic species to assign to the heavy atoms.

Only a maximum of two combination (n, El) can be used.

coord *xi*

x specifies the expected number of heavy atoms surrounding each i-th heavy atom (the coordination type) (i = 1,2).

Examples

Example 1

Use of the peak labelling procedure to relabel the atoms contained in the `vfi.fra` file, supplied by using the command `%fragment`. The procedure requires that the extraction process must be carried out (`%extraction`).

```
% window
% structure vfi
% job vfi - Synchrotron data
%data
cell 18.9752 18.9752 8.1044 90.0 90.0 120.0
space p 63
wave 1.5193
cont al 18 p 18 o 114
pattern vfi.pow
synchrotron
%extraction
%fragment vfi.fra
%changelabel
label 3 Al 3 P
coord 4 4
dwei 3.1 0.5
dlig 1.9 0.5
% continue
```

```
The vfi.fra file contains the chemical labels, the fractional coordinates and
the isotropic thermal factors for each atom in the supplied model.
```

The POLPO procedures for polyhedrally coordinated structure (`%polyhedra` command)

The POLPO1 (Altomare *et al.*, 2000) and POLPO2 (Giacovazzo *et al.*, 2002) procedures can be applied for completing polyhedral (tetrahedral or octahedral) coordinated structures when the cations have been located and the coordination is *a priori* known. The procedures, which exploit the experimental information about the cation connectivity, are based on the generation of several structure models compatible with the coordination information, by using the Monte Carlo algorithm. POLPO1 is suggested when all the cations are rightly positioned and labelled. POLPO2 is able to position one or more missing cations and the surrounding anions. The POLPO procedures start only from the positions of cations, which are supplied by an external file (they can be selected from the Fourier map of Expo2014). The command `%polyhedra` must follow the commands <u>%extraction</u> and <u>%fragment</u>.

Directives in the `%polyhedra` command

The following directives must be added after the command `%polyhedra` in the input file to activate specific non-default procedures:

missing *n1 n2 El tetrahedron (octahedron) d tol1 tol2*

n1 and *n2* are the two order numbers (taken from an external file or from the list of atom positions in a Fourier map) of the cations between which the missing cation(s) must be located;

El is the missing cation label, *tetrahedron* (*octahedron*) is the type of coordination of the missing cation. (The procedure can only handle octahedron and tetrahedron);

d is the expected cation-anion distance;

tol1, *tol2* are the tolerance parameters for the tetrahedron (octahedron) distances and angles, respectively. Default values are 0.3.

octahedron *El d tol1 tol2*

d is the expected octahedral coordination distance around the *El* chemical element;

tol1 and *tol2* are the tolerance parameters for distances and angles, respectively.

tetrahedron *El d tol1 tol2*

d is the expected tetrahedral coordination distance around the *El* chemical element;

tol1 and *tol2* are the tolerance parameters for the distances and angles, respectively.

<div align="center">

Examples

</div>

Example 1

The POLPO1 procedure is applied to locate the light atoms around the specified heavy atoms contained in the `crox_cr.fra` file, supplied by using the command `%fragment`. The POLPO1 procedure requires that the extraction process must be carried out (`%extraction`).


```
%structure crox
%job CROX
%data
pattern crox.pow
content cr 8 o 21
wavelength 1.39222
cell 5.447 6.5576 12.1147 106.382 95.715 77.970
spacegroup p -1
%extraction
%fragment crox_cr.fra
%polyhedra
octahedron 1 1.92 0.2 0.2
tetrahedron 2 1.75 0.2 0.2
tetrahedron 3 1.75 0.2 0.2
tetrahedron 4 1.75 0.2 0.2
%continue
```


The `crox_cr.fra` file contains the chemical labels, the fractional coordinates and the isotropic thermal factors of the four located chromium cations.


```
cr .859 .211 .068 1.500
cr .280 .234 .898 1.700
```

```
cr .358 .377 .258 2.100
cr .721 .799 .528 3.200
```

Example 2

The POLPO2 procedure is applied to locate the two missing silicon cations and the polyhedral coordinated oxygen anions. The POLPO2 procedure requires that the extraction process must be carried out (`%extraction`).

```
%structure sapo
%job  SAPO
%data
cell    21.9410 13.6912  7.1243  90.000  90.000  90.000
spacegroup p m m n
wavelength 1.54056
content si 32 o 64 n 2  c 48
pattern sapo.pow
%extraction
%fragment sapo_si.fra
%polyhedra
tetrahedron 1 1.6 0.2 0.2
tetrahedron 2 1.6 0.2 0.2
missing 1 2 Si tetrahedron 1.6 0.2 0.2
missing 1 2 Si tetrahedron 1.6 0.2 0.2
%continue
```

The `sapo_si.fra` file contains the chemical labels and the fractional coordinates of the two located silicon cations.

```
Si    1.17900    0.13400    0.11500
Si    0.98700    0.14200    0.11900
```

<div style="background:cyan">

Direct Methods combined with Simulated Annealing (`%sdirect` command)

</div>

The crystal structure solution by powder diffraction data via Direct Methods can be very difficult if the diffraction data are of low quality and if no heavy atoms are present in the molecule. On the contrary, the use of direct space methods does not require good quality diffraction data but only the expected molecular model, and the structure solution is limited principally by the number of degrees of freedom used to describe the model.

The two approaches can be combined: the information contained in the electron density map as calculated by Direct Methods, is associate with the Simulated Annealing (SA) algorithm (direct space techniques), to be a powerful tool for crystal structure solution, especially for organic structures (Altomare *et al*., 2003, 2008c).

The combined procedure starts from a list of atomic positions (they can be selected from the Fourier map of Expo2014), supplied by an external file via the %fragment command, and used by the procedure as pivot peaks around which a flexible external model can be accommodated.

The command `%sdirect` must follow the commands `%extraction` and <u>`%fragment`</u>.

<div align="center">

Directives in the `%sdirect` command

</div>

The following directives must be added after the command `%sdirect` in the input file to activate the Direct Methods combined with SA procedure:

extmod *string*

To supply the atomic positions of the external model. s*tring* is the name of the file in which, for each atom, are stored: chemical element and atomic coordinates (*x*, *y*, *z*).

Example 1

The procedure requires that the extraction process must be carried out (`%extraction`).

```
%structure alpha_lga
%job  alpha_lga
%data
cell     10.282  8.779 7.068 90.000 90.000  90.000
spacegroup p 21 21 21
wavelength 1.54056
content o 16 c 20 n 4
pattern alpha_lga.pow
%extraction
%fragment alpha.fra
%sdirect
extmod alpha_inizmod.fra
%continue
```

References

Altomare A., Cascarano G., Giacovazzo C. & Viterbo D. (1991). *Acta Cryst.* A**47**, 744-748.

Altomare, A., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G. & Rizzi, R. (2000). *J. Appl. Cryst.* **33,** 1305–1310.

Altomare A., Giacovazzo C., Ianigro M., Moliterni A.G.G., Rizzi R. (2002). J. Appl. Cryst. **35**, 21-27.

Altomare A., Giacovazzo C., Ianigro M., Moliterni A.G.G., Rizzi R. (2003). J. Appl. Cryst. **36**, 230-238.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A.G.G., Rizzi, R. (2006). *J. Appl. Cryst.* **39,** 558-562.

Altomare, A., Cuocci, C., Giacovazzo, C., Kamel, G. S., Moliterni, A. & Rizzi, R. (2008a). Acta Cryst. **A64**, 326-336.

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

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A. & Rizzi, R. (2008c). *J. Appl. Cryst.* **41,** 56-51.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2009). *Acta Cryst.* **A65,** 183-189.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2010a). *J. Appl. Cryst.* **43,** 798-804.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2010b). *Z. Kristallogr.* **225,** 548-551.

Giacovazzo, C., Altomare, A., Cuocci, C., Moliterni,A.G.G and Rizzi, R. (2002). *J. Appl. Cryst.* **35**, 422÷429.

Main P. (1978). *Acta Cryst.* A**34**, 31-38.

Main P., Fiske S.J., Hull S.E., Lessinger L., Germain G., Declercq J.P. & Woolfson M.M. (1980) – MULTAN80, a system of computer programs for the automatic solution of crystal structures from x-ray diffraction data – Univ. of York, England.

Ten Eick L.F. (1977). *Acta Cryst.* A**33**, 486-492.

_____

Structure solution by direct space methods

http://www.ba.ic.cnr.it/softwareic/expo/structure-solution-by-direct-space-methods/

Direct space methods for crystal structure determination from powder diffraction data have become widely available and popular in recent years and have successfully been applied to solve the structures of organic, inorganic and organometallic materials. Different but similar procedures can be realized: grid search, Monte Carlo, simulated annealing (SA), parallel tempering, genetic algorithm, particle swarm. Each method involves the generation of a random sequence of trial structures starting from an appropriate 3D model and moving it until a good match between the calculated and the observed pattern is found. The information about chemical knowledge of molecules is actively used to reduce the number of parameters to be varied: bond distances, bond angles and ring conformation are usually known and kept fixed while only the torsion angles are varied during the procedure.

Two Direct Space algorithms are available in the EXPO2014 program, for crystal structure solution.

They are: 1) a classical Simulated Annealing (SA) approach; 2) the Hybrid Big Bang Big Crunch (*H*BB-BC) algorithm (Altomare *et al*., 2013). It results from an appropriate combination of BB-BC approach with SA and relies on one of the evolutionary theory of the universe consisting of two successive phases: 1) the Big Bang, corresponding to an energy dissipation procedure for creating a completely random initial population; 2) the Big Crunch, corresponding to a contraction procedure for converging to a global optimum point. This procedure is automatically applied by the program for compounds described by less than 6 internal degrees of freedom (torsion angles) and constituted by less than 3 fragments. In all the other cases the classical SA approach is used.

<div style="background:cyan;text-align:center;font-weight:bold">Preparing Input file</div>

To perform EXPO for structure solution using direct space methods, you'll first need to create an input file. It's assumed that the cell parameters and space group have already been determined.

The example input file, `paracetamol.exp`, is shown below. You can use this file for the structure solution of paracetamol using simulated annealing within EXPO.

```
%Structure paracetamol
%Job paracetamol
%Data
        Cell  7.100  9.380  11.708  90.0  97.42  90.0
        SpaceGroup p 21/n
        Pattern   paracetamol.xy
        Wavelength 1.54056
 %fragment paracetamol.mol
 %sannel
```

You can import this file by going to **File > Load & GO** in the main menu. The file `paracetamol.exp` is already located in the example folder.
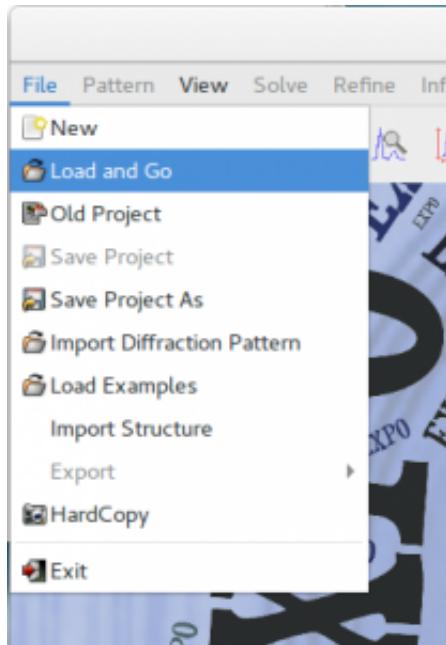
`%sannel` is a command to access to the Simulated Annealing graphic interface.

`%fragment paracetamol.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 (see the section [Usage of command %fragment](#)). Some other file types can be imported in the same way: MOPAC file (`*.mop`), Tripos Sybyl file (`*.mol2,*.ml2`), [C.I.F.](#) file (`*.cif`), Protein Data Bank file (`*.pdb`), Fenske-Hall Z-matrix (`*.zmt`), EXPO fragment or Free Fractional Format (`*.frac`), XYZ format (`*.xyz`), Tripos SYBYL (`*.mol2,  *.ml2`), Shelx File (`*.ins, *.res`). A complete list of the supported input file can be find in section [Import](#) of this document.

Instead of command `%fragment name_starting_model.ext`, with ext=mol, frac, cif etc, you can import the starting structural model directly by the graphic interface from the menu **File > Import**.
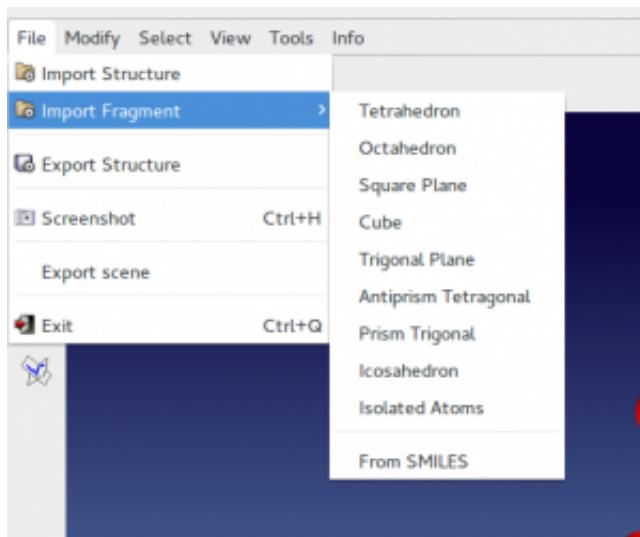
Modify > Add fragments to add fragment to an existing partial structural model.

Instead of command `%sannel`, you can access directly to the graphic interface from menu **Solve > Simulated**

Annealing.

Solve Refine Info Help
- RAMM Procedure
- Explore trials
- Simulated Annealing
- Simulated Annealing + Direct Methods
- POLPO
- Recycle in extra
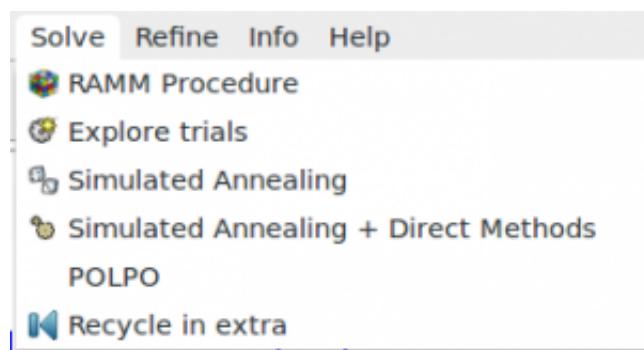
Press the button [GO Execute] in the dialog window to run the Simulated Annealing procedure.

Molecular description

The required starting model can be described in terms of combination of scattering objects that are molecular fragments *e.g.*, isolated atoms, molecules, coordination polyhedra. For each molecular fragment, the position may be defined by the fractional coordinates (*x, y, z*) of the center-of-mass or a predefined pivot atom, the orientation may be defined by rotation angles (*?, ?, ?*) around a set of orthogonal axes, and the intramolecular geometry may be specified by a set of n variable torsion angles $\{?_1, ?_2, \ldots, ?_n\}$. These concepts may be extended to the case of two or more (identical or nonidentical) molecular fragments within the asymmetric unit (a.u.). In general bond lengths, bond angles and ring conformation are not considered as variables during the optimization procedure, but they should match as closely as possible that of the studied compound. One of the drawbacks of global optimization methods is to input an accurate 3D starting structural model of the a.u.

Starting model for direct-space method can be prepared by the following two strategies, or by a combination of both approaches.

1) The user is strongly advised to check for similar molecules in the Cambridge Structure Database (CSD) (organics & organometallics) or in the Inorganic Crystal Structure Database (ICSD) (inorganics, elements, minerals & intermetallics) or in the Crystallography Open Database (COD: http://www.crystallography.net/) or in the crystallographic literature. If a new structure is being studied, one can frequently find significant molecular parts of the structure in existing crystal structures. Molecular fragment found in database can be modified and optimized by quantum-chemistry package to finally generate the desired structure.

Some open chemistry database where look up molecules are:

1. NIST Chemistry WebBook: http://webbook.nist.gov/chemistry/
2. PubChem: https://pubchem.ncbi.nlm.nih.gov/
2. Drugbank: http://www.drugbank.ca/

Calculated 3D molecules in `sdf`, `pdb`, `mol` format can be find in these database and imported in EXPO2014. Computed Simplified Molecular-Input Line-Entry System (SMILES), when available, can be converted in 3D molecule by **File > Import Fragment > From SMILES** or **Modify > Add Fragments > From SMILES**. You can verify that the bond lengths have reasonable values using a tables of standard bond lengths available in Volume C of *International Tables for X-Ray Crystallography*.

2) Starting molecular model can be created by geometry optimization using quantum-chemistry package, *e.g.*, MOPAC, Gaussian, Gamess, NWChem, etc. Nowadays this is usually done by building the molecule with an interactive builder in a graphical user interface, then optimizing it with forcefield method with the click of mouse. The resulting structure is then subjected to an *ab initio*, semi-empirical, or DFT calculations.

Some free available software that can be used to sketch molecules, optimize the geometry by forcefield method and create input file for the quantum-chemistry calculations are:

1. Avogadro: http://avogadro.cc/wiki/Main_Page
2. Gabedit: http://gabedit.sourceforge.net/
3. ACD/ChemSketch: http://www.acdlabs.com/resources/freeware/chemsketch/
4. MarvinSketch: https://www.chemaxon.com/products/marvin/marvinsketch/

See the section Geometry Optimization for more information about the support provided by EXPO2014 for theoretical calculations.

The description of crystal structure is straightforward in the case of molecular organic crystals: isolated molecules with known chemical connectivity are packed together by weak intermolecular forces. Because the starting model should be a tridimensional representation of the a.u., is important to known how many molecules are contained in the a.u. (Z'). In case of molecular compounds the a.u. is usually a molecule or occasionally two or more molecules which differ from one another in orientation or conformation (Z'>1). When a molecule posses symmetry coincident with a crystallographic symmetry element, it may occupy a special position, and the a.u. will then be a half molecule or even some smaller fraction (Z'<1). Inorganic materials often consist of connected polyhedra and the topology of this connectivity is not generally known a priori and thus building a model of inorganic crystal structure for global optimization is often less straightforward and clear. Software provides tool to import some regular polyhedra or you can find the type of polyhedra you need from literature or from database. The choice of the number of each polyhedra to use must be made, talking into account how many atoms are expected per unit cell (volume of atom in the cell). These non?molecular crystals usually crystallize with higher symmetries, and atoms often occupy special positions. At the same time it is necessary to take into account the possibility that polyhedra share atoms at the corners. The use of dynamical occupancy correction (DOC) is able to automatically correct the occupancy of atoms if they are close to a special position or in case of overlap between atoms of the same atomic type. DOC could be very useful when the exact composition of the studied compound is a priori not known exactly. Theoretically this means that it is possible to add more atoms than initially deemed necessary, expecting the DOC to artificially merge the excess atoms. It is also possible to solve structures putting into the cell independent atoms at random positions, but the observation/parameter ratio ('observation' here means the number of reflections) should be at least 8 or more depending on the quality of diffraction data.

Graphic Interface of Simulated Annealing

The dialog window of Direct Space algorithms in EXPO2014 is composed by 4 pages.

The first page contains the general settings of Simulated Annealing algorithm:

Global optimization dialog

SA conditions | External DOF | Internal DOF | Pattern

General conditions

Cost function: R weighted profile

Resolution: 2.000   N. of reflections: 103   2thmax: 45.305

Random seed: 1

Nr. of runs: 10

Simulated annealing options

Starting temperature: 10.000   ☑ automatic

Number of moves: 60   ☑ automatic

Temperature reduction factor: 0.900

Number of refined parameters: 8   ⭐ Solutions

🛟 Help   🚪 Quit   GO⟩ Execute

Cost function: 2 cost functions can be selected: R weighted profile (R_{wp}), R-Bragg intensities (R_B). 1.The default cost function is R_{wp}, which is the agreement factor usually used in the Rietveld refinement

$$R_{wp} = \sqrt{\frac{\sum_i w_i (y_{exp}(\theta_i) - y_{calc}(\theta_i))^2}{\sum_i w_i y_{exp}(\theta_i)^2}}$$

where $y_{exp}(\theta_i)$ and $y_{calcl}(\theta_i)$ are the observed and calculated profile values at the θ_i value of the i-th experimental step, respectively, and $w_i = \frac{1}{y_{exp}(\theta_i)}$. If R_{wp} is used there is no need to extract the structure factor moduli and a profile fitting procedure must be carried out. This task is automatically performed by EXPO2014 before starting SA.

2.R_B factor that compares the experimental integrated intensities I_h^{exp} with the intensities I_h^{calc} calculated by

the model

$$R_B = \frac{\sum_{h} |I_h^{exp} - I_h^{calc}|}{\sum_{h} I_h^{exp}}$$

The preliminary extraction of experimental integrated intensities is automatically performed by EXPO2014 by using the Le Bail algorithm, before starting SA. The advantage of using R_B with respect to R_{wp} is that it offers a significant reduction of the computational time, but, if the overlap is severe, R_B can be unreliable because the I_h^{exp} values can be affected by large errors.

Resolution: defines the maximum resolution used by Simulated Annealing procedure. Direct space methods do not require the use of the entire pattern: usually the algorithm works well with data up to 2 Å resolution.

Random seed: selects the value determining the sequence of random numbers used from the algorithm. When is set to 0 the random seed will be calculated by the system clock.

Nr. of runs: select the number of Simulated Annealing runs. At the beginning of each run a new value of random seed is calculated.

Starting temperature: selects the starting temperature. Check on 'automatic' and the program automatically will find the starting temperature at the beginning of the procedure.

Number of cycles: the number of moves for each step of temperature is Np*N*20 where Np is the number of refined parameters and N is a number set by the user in the entry box. Choosing 'automatic' the program automatically will determine the value of N by taking into account the external and internal DOFs and the flexibility of the molecule. This number can be modified also by directive `niter` (see above).

Temperature reduction factor: the reduction factor applied to the temperature at each step in the annealing schedule. The default value is 0.90. Increasing this value, the chance to find the global minimum can be improved even if a longer execution time will be taken by the procedure.

Solutions: browse the best solutions saved at the end of each run.

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

Global optimization dialog

| SA conditions | External DOF | Internal DOF | Pattern |

Select fragment

Fragment n. 1

Atoms

O1 C1 C2 C3 C4 N1 C5 C6 C7 C8 O2

Parameter for fragment n. 1

	Refine	Lower	Upper
X Traslation	☑	0.000	1.000
Y Traslation	☑	0.000	1.000
Z Traslation	☑	0.000	1.000
Rotation	☑	0.000	1.000

Number of refined parameters: 8

Solutions

Help

Quit

Execute
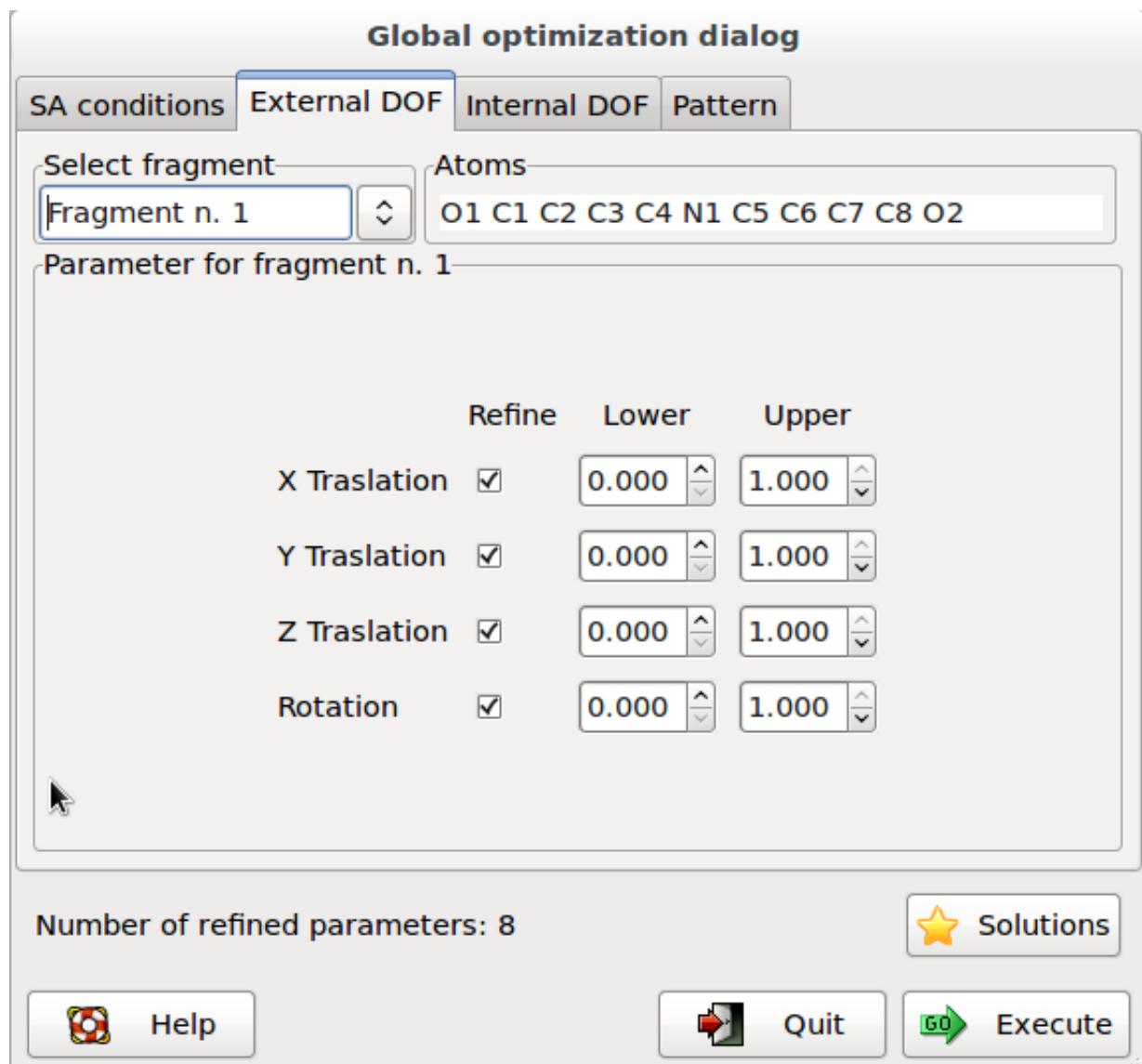
Select fragment: selects the fragment and visualizes the corresponding structure information. Atoms: list of the atoms in the selected fragment. Parameter for fragment: list of the external DOFs for the selected fragment. Check the parameters to refine, enter the lower and upper bounds of the parameters.
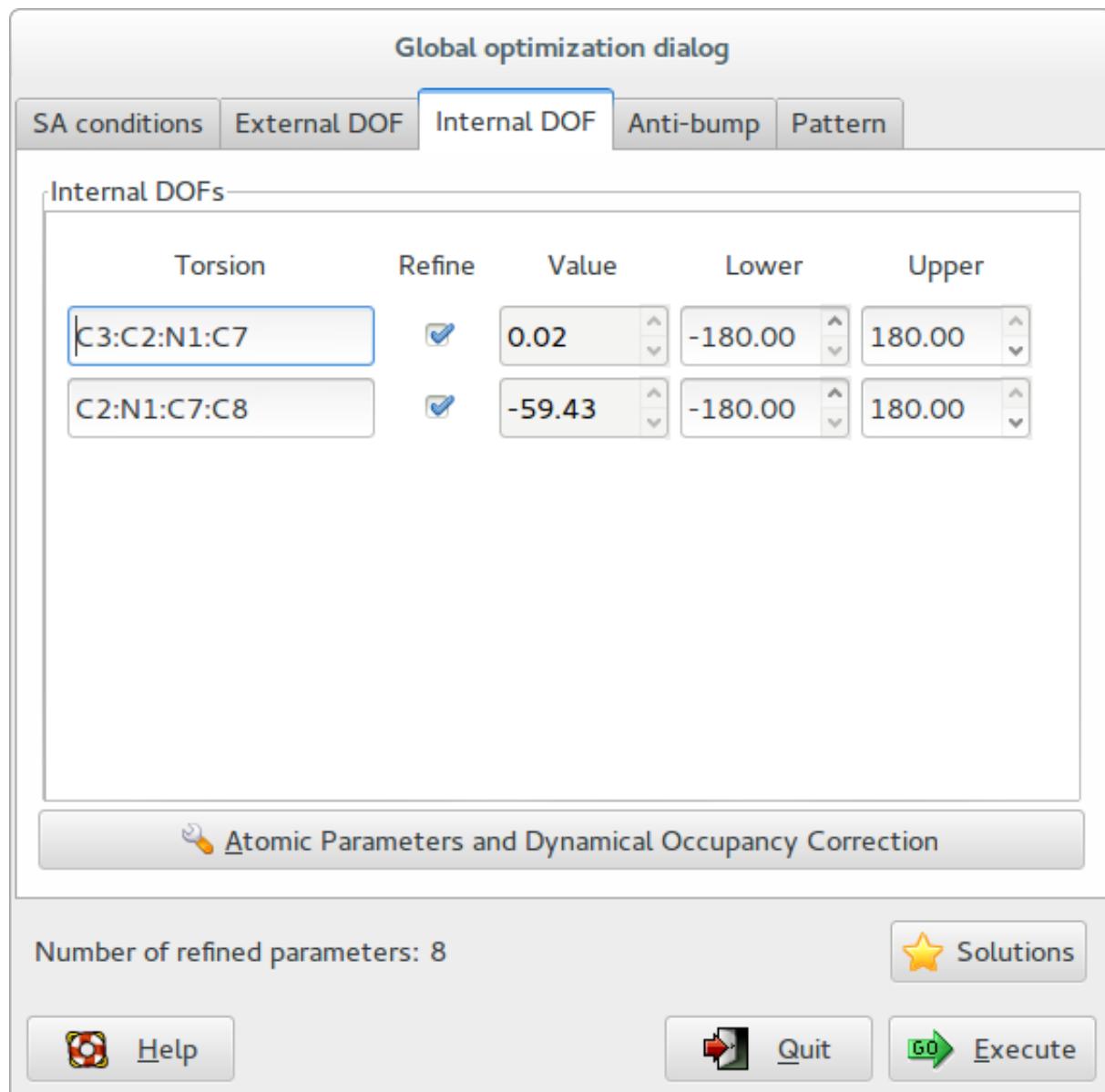
The third page **Internal DOF** contains information about the internal degrees of freedom (DOFS):

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.

Press the button 'Atomic parameters and Dynamical Occupancy Correction' to access to the following dialog window:

Atomic parameters refinement

Atom	Occ.	B[iso]	☐ Refine B[iso]	☐ Shift on xyz	☐ D.O.C.
O1	1.00000	3.00000	☐	☐	☐
C1	1.00000	3.00000	☐	☐	☐
C2	1.00000	3.00000	☐	☐	☐
C3	1.00000	3.00000	☐	☐	☐
C4	1.00000	3.00000	☐	☐	☐
N1	1.00000	3.00000	☐	☐	☐
C5	1.00000	3.00000	☐	☐	☐
C6	1.00000	3.00000	☐	☐	☐
C7	1.00000	3.00000	☐	☐	☐

Maximum shift on position   `0.500`

✗ 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 syz' 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

Pattern: Enter the direction of the preferred orientation and check 'G Factor'. The March-Dollase correction will be applied and the magnitude of the preferred orientation will be optimized.

A visual match between observed and calculated powder pattern is plotted when Simulated Annealing is running, the progress of structure solution is monitored and the user can examine: 1) the graph of the minimum values of the cost function (CF) vs. the number of moves; 3) the crystal packing by using the button [button icon] on the JAV viewer.
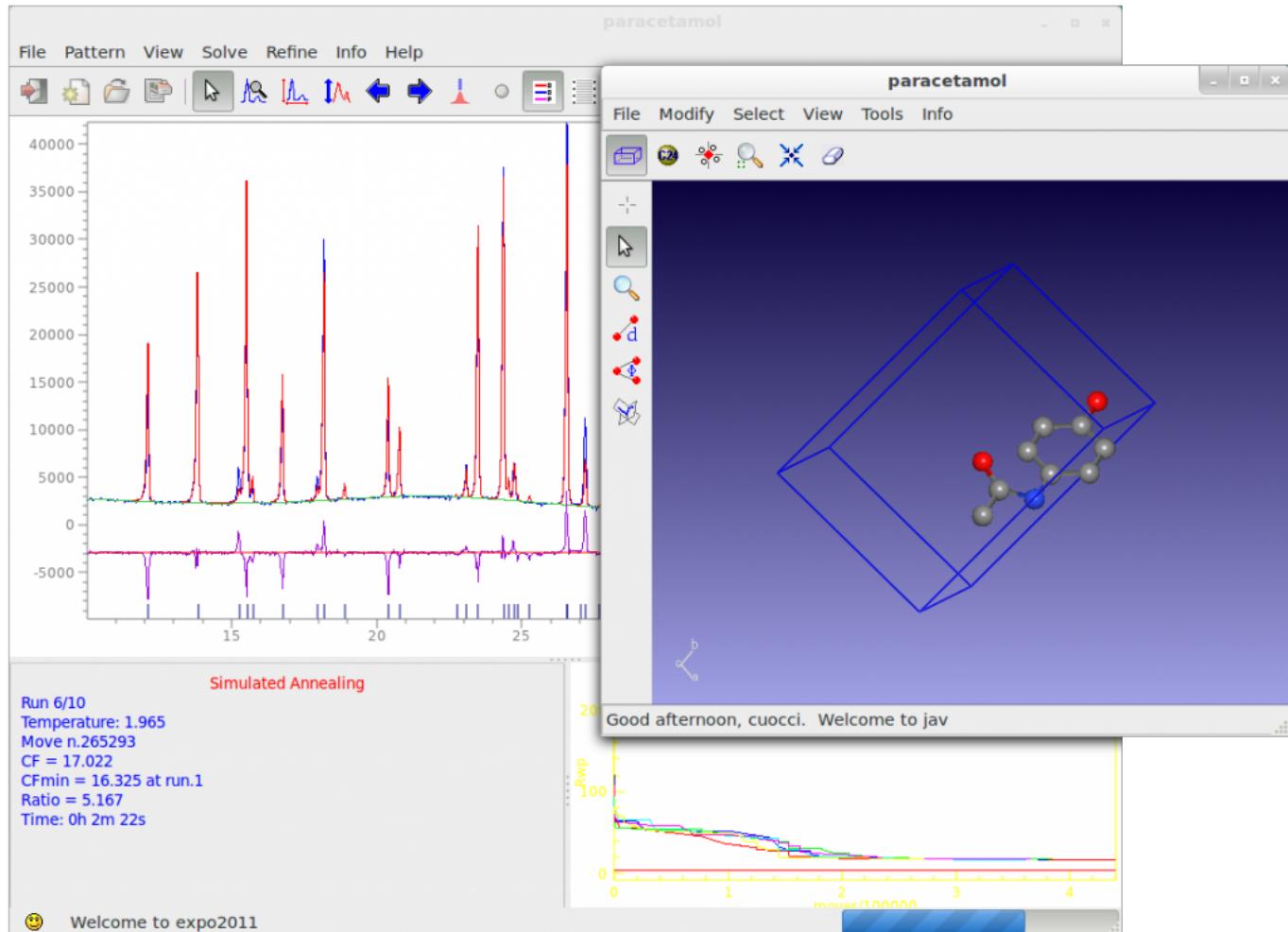
During the Direct Space runs, three buttons are active on the toolbar of the main window: [icon] opens the JAV viewer for crystal structure visualization. [icon] skips the current run. [icon] stops the procedure.

Anti-bumping restraints

When the quantity of information available from a powder diffraction pattern is limited (due to, *e.g.*, severe peak overlap, broad peaks, preferred orientation, presence of weak scatters) and/or the number of degrees of freedom is large, it may be necessary to add extra chemical information to the optimization process in order to obtain the correct solution. In fact, in this situation, the correct structure may not correspond to the minimum of CF. The use of restraints on bond distances and angles or the application of bond valence restraints are approaches to increase the probability to obtain only chemically plausible models (Falcioni & Deem, 1999; Favre-Nicolin & ?ern?, 2002) . In particular if the final global solution contains atoms colliding each other, anti-bumping restraints can be used to

avoid this solution. They are relationships between atoms to prevent atomic group overlapping with other parts of structure. The restraint can be implemented as normal distance restraint, which is only applied if the interatomic distance becomes less than some threshold. The contribution of the anti-bumping restraints to the global cost function is measured by the espression

$$CF_{bump} = \sum_{ij}^{n} w_{ij}(d_{ij}^{min} - d_{ij}^{model})^{2k}$$

with the condition $d_{ij}^{model} < d_{ij}^{min}$, where d_{ij}^{min} and d_{ij}^{model} are the minimum ideal distances and the model distances, respectively, between pairs of atoms i and j, and summation is over n contacts, k is equal to 2 (Hendrickson, 1985). d_{ij}^{min} depends on the atomic elements in contact and on the type of contact, in particular if the two atoms are hydrogen-bonded atom pairs. d_{ij}^{min} in default are defined as the sum of the atomic radii of the two atoms multiplied by a scale factor ϵ (0.7 is the default value)

$$d_{ij}^{min} = \epsilon(R_i^{vdW} + R_j^{vdW})$$

and R_i^{vdW} and R_j^{vdW} are the van der Waals radii (Feng *et al.*, 2009). This minimum distance d_{ij}^{min} can also be modified by the user for each pair of atom types by using the directive bump. The directive bscale allows the modification of the scale factor ϵ on distances d_{ij}^{min} (o.7 is the default value). Increase ϵ if the default value is not able to avoid structures with overlapped atoms. $\epsilon = 0.6 \div 1.0$ is recommended value. Small value of ϵ (*e.g.*, $0.3 \div 0.6$) are recommended if the anti-bumping restraints are applied between isolated atoms in the starting model but that are expected to be connected in the final structure. In principle, the weight w_{ij} associated to restrains varies with the type of restrains, but in practice a uniform value is used (1.0 is the default value). You could change this value.

Use the directives bump and/or nobump if you are interested in applying restraints between specific atoms and to between groups of atoms of the same species. The weight w_{ij} and scale factor ϵ can be modified by using the directives bweight and bscale.

Output file contains a list of all applied anti-bumping restraints with the distances d_{ij}^{min} and d_{ij}^{model} .

All pairs of atoms separated by a rigid sequence of bonds (*e.g.*, two atoms in a rings) are excluded from the anti-bumping interactions. In case of intramolecular contacts it is also advantageous to exclude atoms that are separated by a short contiguous flexible chain of covalent bonds. The applied rule consists to exclude all atoms that are separated by a path of bonds containing 4 rotable bonds or less.

The application of restraints is relatively time-consuming because requires that all symmetry equivalent atoms are taken into account. Strictly speaking, it should be unnecessary and it should be use only when the quality of the diffraction pattern is insufficient to avoid that the final solution contains overlapping atoms.

Bond valence restraints

Bond valence sums can be introduced as restraints in the optimization process adding to the χ^2 sum another term with the form:

$$CF_{VB} = \sum_{n}^{N_R} w_n(V_n - V_n^0)^2$$

where the sum is done over N_R bond valence restraints, w_n is a suitable weight assigned to each restraint, V_n is the bond valence sum of an atom with the expected formal charge V_n^0. The directives `bvres` and `bvpar` are available to set bond valence restraints.

Output of SA

The output file generated from the procedure, contains information on:

- Starting crystal structure
- Volume per atom
- Connectivity
- Internal and external DOFs
- List of restraints
- List of optimized parameters
- Each SA or *H*BB—BC run
- Summary of all the SA or *H*BB—BC runs
- The selected final structures 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.*, `paracetamol_best1.cif` is the best solution, `paracetamol_best2.cif` is the second best solution, etc.). The order number in the name represents the position in the list of structures ordered according to the cost function.

When SA fails

When Direct Space approaches fail:

1) The default conditions, in particular for complex structure, could not be appropriate. Increase 'Number of moves' and 'Nr. of runs' in the tab 'SA conditions', these are the most important parameters in determining the success of the optimization process. You can double or, if necessary, triple the default value of 'Number of moves' to be sure to find the global minimum. The directives `niter` and `nrun` can be used to set these parameters directly in the `.exp` input file. Some examples about the application to complex problems are reported at this link.

2) The quality of data is not good and the diffraction pattern is not suitable for the extraction of the integrated intensities. In this case can be convenient to perform the optimization by using the R weighted profile cost function.

3) The starting model is incorrect: bond distances and angles are not entirely accurate, the number of building blocks is wrong. Improve your model with Cambridge Structural Database (CSD) or building packages (Avogadro, ChemSketch, ChemDraw, …), check the volume per atom in the output file (about 15-20 Ang/atom).

4) The assumption about thermal factors is invalid. Check thermal factors from similar structures and/or include them in the optimization process by the window dialog 'Atomic parameters and dynamical Occupancy Correction '. See also the directive refinetf.

5) Space group and cell parameters 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.

Directives of SA

Usually you don't need to read this paragraph unless you are interested to run Simulated Annealing without interaction with graphic interface. In this case use the command %sannel to run Simulated Annealing from the input file (*.exp). Use command %automatic to skip the interaction with the graphical interface and the program will perform the Simulated Annealing using the default values. An example of input file with command %automatic and %sannel is the following:

```
%automatic
%Structure paracetamol
%Job paracetamol
%Data
        Cell  7.100  9.380  11.708  90.0  97.42  90.0
        SpaceGroup p 21/n
        Pattern   paracetamol.xy
        Wavelength 1.54056
%fragment paracetamol.mol
%sannel
```

To modify the default values of SA, some directives can be used after the command %sannel in the .exp input file.

An example of application of directive niter and nrun to increase the percentage of success in case of structure solution of the largely flexible molecule verapamil hydrochloride

```
%Structure verapamil
%Job Verapamil hydrochloride (C27H39N2O4.Cl)
%Data
        Cell   7.089  10.593  19.207  100.11  93.75  101.56
        SpaceGroup p -1
        Pattern    verapamil_hydrochloride.xy
        Wavelength 1.54056
%fragment verapamil_hydrochloride.mol
%sannel
        niter 2000
        nrun 20
```

The following directives must be added after the command `%sannel` in the input file to activate some specific features of the simulated annealing procedure

Abbreviations for directives name at least 4 character are permissible, *i.e.* `cost 2` instead of `cost_function 2`.

`bump`
or
`bump atoms1 atoms2 [dist wei]`
`bump` directive is used for the generation of anti-bumping restraints. If this directive is used without additional parameters anti-bumping restraints are automatically generated and extended to all C, N, O and S atoms. The `atoms1` and `atoms2` parameters must be used to apply restraints between specific atoms or between groups of atoms of the same species. `atoms1` and `atoms2` are the label of atom. If S is an atomic specie, the symbol S* should be used to refer to all the atoms of the same species. The special symbol * select all atoms. Use `frag1`, `frag2`, … to refer to entire molecular fragment. `dist` and `wei` are optional parameters. `dist` can be used to specify the minimum distance at which the restraint is active. `wei` is the weight associated to the restraint. Negative values of `dist` and `wei` will be ignored applying the default values that are: sum of Van der Waals radii for `dist`, 1.0 for `wei`. Complete list of generated anti-bumping restraints is available in the output file.

E.g.
bump
anti-bumping restraints are applied only to all C,N,O,S atoms in the structure

bump C* S*
apply anti-bumping restraints only between C and S atoms. Default values are assigned to `dist` and `wei`.

bump C1 S1 3.5 10.
apply anti-bumping restraint only between C1 and S1 atoms

bump C* S1 -1 10.
generate anti-bumping restraints between all C atoms and the S1 atom. The weight is set to 10. and the default value is used for distance.

bump C1 S1 3.5 10.
bump C1 C2 -1 10.
`bump` directives can be combined to generate specific restraints lists.

bump
bump C1 S* -1 10.
anti-bumping restraints are applied only to all C,N,O,S atoms in the structure. The weight for contact between C1 and all sulfur atoms is set to 10.

bump C1 *
generate anti-bumping restraints between atom C1 and all other atoms.

```
bump frag1 frag2 -1 10.
```
generate anti-bumping restraints between the atoms in the molecular fragment 1 and the atoms in the molecular fragment 2.

nobump atoms1 atoms2

Delete anti—bumping restraints. This directive makes sense if applied in combination with `bump` directive to remove previously set restraints.

E.g.
```
bump * *
nobump Ag1 N1
nobump Ag1 N2
```
This combination of 3 directives generates anti-bumping between all atoms except between atom Ag1 and N1, Ag2 and N2.

```
bump
nobump O5 *
```
This combination of 2 directives generates anti-bumping between all O,C,N,S atoms except O5 oxygen atom.

bweight w

This directive controls the weight on all anti-bumping restraints. The default weight is 1.0.

E.g.
```
bweight 10.
bump
```

bscale s

This directive allows the modification of the scale factor ϵ applied to a sum of Van der Waals radii in the definition of the default minimum distance between two atoms for anti-bumping restraints. 0.7 is the default value. If you increase this value, you increase the minimum distance d_{ij}^{min} and the number of contacts that contribute to the cost function.

cost_function n

To choose the cost function: 1 for Rw-profile, 2 for RF, 3 for RI. The default choice is 1.

doc
or
doc atom1 atom2 atom3 …

Activate the dynamical occupancy correction (DOC). By default DOC is applied to all the atoms unless you specify a list of atoms.

E.g.
```
doc Ni1
```
DOC will be applied only to the atom `Ni1`.

centre_of_rotation atc
or
centre_of_rotation atc atf

This directive is used to specify the origin of rotation. `atc` is the label of the atom chosen as centre of rotation and the rotation will be applied to the molecular fragment containing the atc atom. A second generic atom `atf`

included in the molecular fragment may be indicated when the centre of rotation `atc` is not a part of the fragment. As default choice, the centre of rotation is the central atom in case of polyhedra; the center of rotation is the center of mass for all other type of scattering objects. The keyword `@com` can be used to change the default choice in case polyhedra. The origin of rotation for each molecular fragments is reported in the output file in the section 'List of fragments'.

E.g.
`centre_of_rotation P1`
The atom P1 is the origin of rotation of the molecule $P(C_6H_5)_3$ containing P1

`centre_of_rotation Ni1 P1`
The molecule $P(C_6H_5)_3$ containing the atom P1 will be rotated around Ni1

`centre_of_rotation @com Si1`
The polyhedron containing the Si1 atom will be rotated around the center of mass. This directive modify the default choice of program.

extdof `atom [translation_code rotation_code]`
To fix external DOFs of a specific molecular fragment. `atom` is any atom belonging to the molecular fragment. `translation_code` and `rotation_code` must be 0 to fix the translation and rotation respectively and 1 to refine it. The default is 1.

E.g.
`extdof Si1 0 0`
Translation and rotation parameters of the molecular fragment containing the Si1 atom will be not optimised.

intdof `atom optim_code`
or
intdof `atom1 atom2 [optim_code]`
To fix the optimization of a specific internal DOF. This directive can be used in two different ways.
a) To select all the internal DOFs of a molecular fragment where `atom` is any atom in the fragment.
b) To select a specific internal DOF. In this case `atom1` and `atom2` must be the 2 atoms that define the rotation axis. This directive is useful when, relying on the prior chemical knowledge of the structure, some internal DOFs can be fixed, for example double bonds.
`optim_code` must be 0 to fix and 1 to enable the optimization. The default is 1.

E.g.
`intdof C1 C2 0`
Fix the internal DOF corresponding to the rotation around the C1-C2 axis.

`intdof C1 0`
Fix all internal DOFs of the molecular fragment containing the C1 atom.

`extdof Si1 0 0`
`intdof Si1 0`
`intdof` is used in combination with `extdof` directive to fix an entire framework of silicon atoms.

intdof_disp `atom1 atom2 atom3 atom4 disp [mode]`
These directive can be used to apply a specific displacement to the rotation associated to the torsion
`atom1-atom2-atom3-atom4`. atom1, atom2, atom3, atom4 are the atoms that define a specific torsion. The

torsion will be sampled in the interval $\theta_0 - disp \leq \theta_0 \leq \theta_0 + disp$ where θ_0 is the initial value of the torsion.

`mode` is an optional parameter that can be used to specify whether the torsion is bimodal (`mode=2`) or trimodal (`mode=3`), default is mode=1. This information can be very useful in case of expected bimodal planar torsion ($\theta_0 = 0^\circ$); for example, if `disp` is equal to 30°, the torsion will be sampled in two range if `mode=2` is specified: $-30^\circ \div 30^\circ$ and the complementary range $-170^\circ \div 170^\circ$.

If the torsion is not planar and the starting value is 60°, the sampling ranges will be: $30^\circ \div 90^\circ$ and $-90^\circ \div -30^\circ$. In case of trimodal torsion, 3 ranges will be sampled: the range around the starting value and two additional range at +120° and -120°.

E.g.
```
intdof_disp N1 C2 C2 B2 30
```
The directive is used to limit the sampling of the specified torsion.

```
intdof_disp C1 C2 C3 C4 20 2
```
The torsion C1-C2-C3-C4 is expected to be bimodal and planar. The starting value of the torsion must be zero or close to zero.

move `atom1 @(x y z)`
The `move` directive can be used to move the position of a molecular fragment in a specific position.
`atom1` is the label of an atom contained in the molecular fragment and `x y z` are the final fractional coordinates of `atom1` after translation of the entire fragment.
This directive can be useful when a position of an atom is already known, for example an heavy atom located by direct methods.

E.g.
```
move Ni1 @(0.5 0.286 0.75)
extdof Ni1 0 1
centre_of_rotation Ni1
```
The move directive is used to move the molecular fragment with `Ni1` to a specific position. The combination of the directives `extdof` and `centre_of_rotation` activates the rotation around `Ni1` fixing the translation.

niter `n`
To modify the number of moves for each temperature step. In a default run the number of moves is automatically calculated. This parameter is important in determining the success percentage of the optimization procedure.

nrun `n`
To modify the number of simulated annealing runs. The default is 10.

resmax `val`
To define the maximum resolution used by simulated annealing. The default value is 2.0 Å.

temper `temp`
To modify the initial temperature. In a default run the initial temperature is automatically calculated.

tfactor `val`

Determines the temperature reduction factor. `val` ranges between 0 and 1 and its default value is 0.90.

refinetf `overall`
refinetf `species`
refinetf `molecules`
refinetf `atoms atoml atom2 atom3 …`
refinetf `groups atoml atom2 atom3 …`

The directive `refinetf` can be used for the optimization of the thermal factors. This directive should be followed by a keyword that specifies the type of used constraint. The thermal factors of the hydrogens are constrained to be 1.2 times of that of the heavy atom to which they are attached. The `overall` keyword activates the optimization of an overall thermal factor. The `species` keyword activates the optimization of a thermal factor for each atomic specie. The `molecules` keyword activates the optimization of a thermal factor for each molecular fragment in the crystal structure. The `atoms` keyword allows the optimization of thermal factors of specific atoms, independently. `atom1`, `atom2`, `atom3`, … are the atomic label of the atoms. Groups of atoms can be selected by using the `*` character (*e.g.*, `C*` to select all C atoms). The groups keyword can be used to indicate that the specified atoms (`atom1`, `atom2`, `atom3`, …) must have the same thermal factor.
E.g.
`refinetf atoms *`
All thermal factors are refined independently.

rest `atom1 atom2 [target_dist tol weight inter]`
rest `atom1 atom2 atom3 [target_angle tol weight inter]`

Apply restraints between 2 atoms. atom1 and atom2 are the labels of atoms. A third `atom3` atom must be specified to apply restraints on the angles. The other parameters are optional and can be omitted. `target_dist` is the ideal distance in Angstrom between the pair of atoms, when omitted the distance is automatically deduced by using an internal table of distances. `target_angle` is the ideal angle in degrees. `tol` is a permitted tolerance, when omitted the default value is 0.2 Å for distance and 2.0° for angle. `weight` is a user supplied weight. A default weight is specified as 100. `inter` is an optional keyword that force the restraint only between molecules (intermolecular) and not inside the molecules (intramolecular). The default behaviour is this:
a) if the two atoms `atom1` and `atom2` are in the same molecule, symmetrically equivalent atoms are non considered and the restraint is automatically set as intramolecular. This setting is useful in a common application of restraints with direct space method: when it is necessary to force the closure of a flexible ring.
b) If the two atoms are in separated molecules, expansion of symmetry is performed and all atom1-atom2 intermolecular and intramolecular bonds in the cell are considered. The keyword 'inter' apply the restraints only between molecules excluding the intramolecular bonds. Therefore the keyword `inter` is required if you want to impose the chain linearity in a polymeric structure.
E.g.
`rest C11 C16 1.54`
`rest C2 C1 C6 110`

rotate_around_axis `Ax1 Ax2 At1 At2 At3 …`
or
rotate_around_axis `Ax1 Ax2 At1 At2 At3 … theta`

Rotate atoms At1, At2, At3, … around a rotation axis defined by atoms Ax1 and Ax2. theta is an optional value (degrees) used to specify the limits of rotation angle from -theta to +theta. When `rotate_around_axis` directive is used the specified rotation will be included in the panel **Internal DOF** of the graphical interface.
E.g.
`rotate C1 C4 C5 C6 C7 C8 C9 C10`
the phenyl ring C5-C6-C7-C8-C9-C10 will be rigidly rotated around the axis C1-C4

shift_atom val

  or

shift_atom atom1 atom2 atom3 …. val

To optimize the atomic parameters by applying shifts (up to val) on the atoms with respect to the centre of gravity of the fragment. The default val is 0.5. Add `atom1`, `atom2`, `atom3`,… to refine only some specific atoms.

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.

po H K L

Activate the *March-Dollase* preferred orientation correction. `H K L` are three integer numbers necessary to specify the direction of the preferred orientation

bvres atom1 valence [sigma wei]

Set bond valence restraint specifying the atom name and the expected oxidation state omitting the sign. The deviation from the expected value `sigma` and the weight of restraint `wei` can be optionally indicated. The default values for `sigma` and `wei` are 0.3 and 1, respectively. The special character `*` can be used to apply the same restraints to group of atoms of the same atom type.
E.g.
Apply the bond valence restraints on all gallium atoms. The expected formal valence of `Ga1`, `Ga2`, `Ga3`, `Ga4` is 3.
```
bvres Ga1 3
bvres Ga2 3
bvres Ga3 3
bvres Ga4 3
```
the same directives can be written in a more synthetic way as
```
bvres Ga* 3
```

bvpar El1 El2 Ro B [rmin rmax]

For each pair of atom types `El1` and `El2` the program searches in the database the corresponding bond valence parameters (R_0 and B) and assign a bond windows, *i.e.* the range of distances between any two atoms that are to be considered as bonds. These range is defined in terms of minimum and maximum distance `rmin` and `rmax`. The table of bond valence parameters is written in the output file. Use this directive only if the bond valence parameters and the bond window assigned in default by the program should be modified.

Usage of command %fragment

The command `%fragment` followed by the name of a file can be used to import molecular starting model from external input file. The file extension must be mandatory specified to enable the program to identify the type of file. A complete list of the supported input file can be find in section Import of this document.

The command `%fragment` may also be associated with specific keywords to generate some predefined molecular fragments without the use of external files. In this case the general form of the command is: `%fragment keyword optionl [option2] [option3]`

The keyword must be followed by at least one option to define further details of the molecular fragment. All possible keywords are listed below.

- **%fragment** `tetra AtC AtV [dist]`
  %fragment `octa AtC AtV [dist]`
  %fragment `square AtC AtV [dist]`
  %fragment `cube AtC AtV [dist]`
  %fragment `trigonal AtC AtV [dist]`
  %fragment `prism_tetra AtC AtV [dist]`
  %fragment `prism_trig AtC AtV [dist]`
  %fragment `icosa AtC AtV [dist]`

  The keywords `tetra, octa, square, cube, trigonal, prism_tetra, prism_trig, icosa` can be used to generate polyhedra, respectively tetrahedron, octahedron, square planar molecular geometry, cube, trigonal planar molecular geometry, tetragonal prism, trigonal prism, icosahedron. The keyword must be followed by the atomic specie `AtC` in the center of the geometry and the second atomic specie `AtV` at the vertices of the geometry. Optionally you can specify the distance dist between the two atomic species. If `dist` is absent the distance will be assigned by accessing an internal database of average distances.

  E.g., `%fragment tetra Si O` import a silicon-oxygen tetrahedron SiO_4 with default distance Si-O of 1.645. `%fragment tetra Si O 1.597` also specifies a Si-O distance of 1.5972 Å.

- **%fragment** `atoms chem_formula`

  The keyword atoms allows you to import isolated atoms in asymmetric unit. The keyword must be followed by the list of atomic species and their number as a chemical formula. Examples of valid formula: `[C3H5(OH)3]4, C 28 O 8 S 4 H 24`.

  E.g., `%fragment atoms Pb` import an isolated atom of Pb in random position. `%fragment atoms (Sb2O3)2` import 4 Sb atoms and 6 O atoms in random positions.

- **%fragment** `smiles SMILES_string`

  The keyword `smiles` allows you to import molecule from an ASCII string written by using the SMILES notation. The following input file can be used to load the starting model by using the keyword `smiles` and solve the crystal structure of paracetamol.

```
%Structure paracetamol
%Job Paracetamol (C8H9NO2)
%Data
Cell 7.100 9.380 11.708 90.0 97.42 90.0
SpaceGroup p 21/n
Pattern paracetamol.xy
Wavelength 1.54056
%fragment smiles CC(=O)NC1=CC=C(C=C1)O
%sannel
```

The command `%fragment` can be repeated several times to import more than one molecules. Some examples are reported.

Input file for the structure solution of the creatine monohydrate

```
%Structure creatinem
%Job Creatine monohydrate (C4H9N3O2.H2O)
%Data
        Cell   12.506  5.046  12.169  90  108.88  90
        SpaceGroup P 21/c
        Pattern    creatinem.xy
        Wavelength 1.54056
%fragment creatine.cif
%fragment H2O.sdf
%sannel
```

The command `%fragment` was used to import the cif file of the creatine molecule (`creatine.cif`) and the sdf file of the water molecule (`H2O.sdf`). The cell parameters and space group reported in the creatine.cif were ignored and the creatine molecule was imported in the cell and space group indicated by the user in the input file by directive `Cell` and `SpaceGroup`.

Input file for structure solution of S-Ibuprofen containing two molecules of Ibuprofen in the asymmetric unit (Z'=2)

```
%Structure S-Ibuprofen (C13 H18 O2)
%Job Structure solution of S-Ibuprofen
%Data
        Cell  12.463  8.029  13.538  90  112.93  90
        SpaceGroup p 21
        Pattern    pd_0034.pow
        Wavelength 1.54056
%fragment Structure3D_CID_3672.sdf
%fragment Structure3D_CID_3672.sdf
%sannel
```

Two molecules of ibuprofen were imported using the file `Structure3D_CID_3672.sdf` downloaded from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/compound/962#section=Top).

Input file for structure solution of lead(II) sulfate (PbSO$_4$).

```
%structure pbso4
%job Lead(II) sulfate (PbSO4)
%data
pattern pbso4.dat
wave 1.54056
cell 6.95802 8.48024 5.39754 90 90 90
```

```
space p b n m
%fragment tetra S O
%fragment atoms Pb
%sannel
```

The command `%fragment` was used to import a SO_4 tetrahedron and an isolated atom of Pb.

References

Altomare, A., Corriero, N., Cuocci, C., Moliterni, A., Rizzi, R. (2013). *J. Appl. Cryst.*, **46,**779-787.

Falcioni, M., Deem, MW. (1999).*J. Chem. Phys.***110, 1754-1766**.

Favre-Nicolin, V. & ?erný, R. (2002). *J.Appl. Cryst.***35, 734-743**.

Feng, Z.J., Jia, R.R., Dong, C., Cao, S.X. & Zhang, J.C. (2010). *J.Appl. Cryst.***43, 179—180**.

Wayne A. Hendrickson, Stereochemically restrained refinement of macromolecular structures, Methods in Enzymology, Academic Press, Volume 115, 1985, Pages 252-270.

_____

Examples of structure solution by direct space method

http://www.ba.ic.cnr.it/softwareic/expo/examples-of-structure-solution-by-direct-space-method/

All examples can be downloaded from this link and are provided as supplementary information for the paper: 'Direct space approach in action: Challenging structure solution of microcrystalline materials using the EXPO software', Computational Materials Science 210 (2022) 111465.

Example 1: structure solution of diltiazem hydrochloride (CSD reference code: CEYHUJ01)

- Folder `examples/Example 1 – diltiazem hydrochloride/test1` contains the following files:
  `diltia.exp`: EXPO input file for structure solution by direct space method starting from a model with an open ring at bond C4-C5.
  `diltiazem_nw_break.mol`: starting model.
  `pd_0029.xy`: powder diffraction data file (available to the public at http://www.powderdata.net).

The input file called `diltia.exp` is here reported:

```
%Structure diltia
%Job CSD refcode: CEYHUJ01

%Data
Cell   42.190  9.075  6.037  90  90  90
SpaceGroup p 21 21 21
Pattern    pd_0029.xy
Wavelength 1.54056

%fragment diltiazem_nw_break.mol
%fragment atoms Cl
deletehydro

%sannel
nrun 100
niter 5000
```

- Folder `examples/Example 1 – diltiazem hydrochloride/test2` contains the following files:
  `diltia.exp`: EXPO input file for structure solution by direct space method starting from a model with an open ring at bond C4-C5. Restraint is applied to the bond C4-C5.
  `diltiazem_nw_break.mol`: starting model.
  `pd_0029.xy`: powder diffraction data file.

The input file called `diltia.exp` is here reported:

```
%Structure diltia
%Job CSD refcode: CEYHUJ01

%Data
Cell 42.190 9.075 6.037 90 90 90
SpaceGroup p 21 21 21
Pattern pd_0029.xy
Wavelength 1.54056

%fragment diltiazem_nw_break.mol
%frag atoms Cl
deletehydro

%sannel
nrun 100
niter 1000
rest C4 C5
```

- Folder `examples/Example 1 – diltiazem hydrochloride/test3/conformer_test` contains files for the structure solution by direct space method applied to 100 conformers generated by [conformer generator tool RDKit](#):
  `scanconformers`: bash shell script used to generate automatically `.exp` file for each conformer, run the structure solution and create result files.
  `diltia_templ`: `.exp` input file used as a template by the script `scanconformers`.
  `molecules.txt`: file used by the script `scanconformers` and containing the list of conformers to test for structure solution.
  `diltia1.mol, diltia2.mol, …`: all conformer files in MOL format. They were generated using the python script `rdk_confgen.py` available in the folder `examples/Example 1 – diltiazem hydrochloride/test3/conformer_generation`. If [RDKit](#) and [Open Babel](#) are installed, you can generate conformers running the following commands:

  ```
  python3 rdk_confgen.py --input diltia_nw.mol --numconf 100
  ```

  ```
  obabel gen_confs.sdf -O diltia1.mol -m
  ```

Openbabel is used to split the SDF multi-file generated by `rdk_confgen.py` in single files: `diltia1.mol`, `diltia2.mol`, `diltia3.mol`, … The `rdk_confgen.py` is a modified version of the python script available at: https://github.com/iwatobipen/rdk_confgen.

The file `scanconformers` is a bash script and is reported here:

To use this file you have to make it executable (`chmod +x scanconformers`) and run it (`./scanconformers`). Modify the line 16 of the script to change the expo installation folder and the number of used CPU-cores. We run expo in parallel on 20 CPU-cores but if the serial version of the program is used, replace the line 16 with the commented line 15.
In the lines 18-20, records in the output file tagged with the '@@' are filtered and redirect on the files `summary.dat` and `results.dat`. When the script is ended `results.dat` will contain a report that can be consulted to quickly identify the most reasonable structure. In addition, for each conformer other files are generated:

- output file (`.out`) with general information about the procedure;
- CIF files with the solutions found during SA;
- project file (`.expo`) that can be used for a visual inspection of the models. Load the files `.expo` in EXPO by **File > Old Project** and access to the list of solutions selecting **Solve > Simulated Annealing** and clicking on the [Solutions] button.
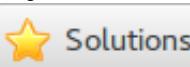
Example 2: structure solution of tetracycline hydrochloride (CSD reference code: XAYCAB)

Folder `examples/Example 2 – tetracycline hydrochloride` contains files for the structure solution by direct space method applied to 300 conformers generated by conformer generator tool RDKit:
`scanconformers`: bash shell script used to generate automatically `.exp` file for each conformer, run the structure solution, create a result file.
`aldx_templ`: `.exp` input file used as a template by the script `scanconformers`.
`molecules.txt`: file used by the script `scanconformers` and containing the list of conformers to test for structure solution.
`aldx1.mol`, `aldx2.mol`, ...: all conformer files in `MOL` format. They were generated using the python script `rdk_confgen.py` available in the folder `/examples/Example 2 – tetracycline hydrochloride/conformer_test`.
`aldx.dat`: powder diffraction data file.

For more information about the meaning and the use of these files see the Example 1 (`examples/Example 1 – diltiazem hydrochloride/conformer_test`).

Example 3: structure solution of selexipag form I (CSD reference code: VOHVIA)

Folder `examples/Example 3 – selexipag` contains the following files:
`selexipag.exp`: EXPO input file for structure solution by direct space method.
`selexipag_mopac.mol`: starting model.
`VOHVIA.cif`: CIF file containing powder diffraction data.

The input file called `selexipag.exp` is here reported:

```
%structure selexipag
%job CSD refcode: VOHVIA

%data
pattern VOHVIA.cif
synchrotron
cell 37.96347 6.110426 22.47454 90 98.3273 90
space P21/c

%fragment selexipag_mopac.mol
%fragment selexipag_mopac.mol
deletehydro
```

```
%sannel
nrun 100
niter 2000
resm 2.5
```

Example 4: structure solution of ALPO-M (CSD reference code: RAMXUZ)

Folder `examples/Example 4 - ALPO-M` contains the following files:
`alpo.exp`: EXPO input file for structure solution by direct space method.
`alpo.xy`: synchrotron powder diffraction data file.

The input file called `alpo.exp` is here reported:

```
%structure alpo
%job AlPO-M L.B. McCusker - (dati SLS) - (CSD refcode: RAMXUZ)

%data
pattern alpo.xy
cell 9.7493 29.1668 9.3528 90. 90. 90.
space p b c a
wave 0.9187
synchrotron

%frag tetra P O
%frag tetra P O
%frag atoms Al2O
%frag smiles C(CO)N
%frag smiles C(CO)N
deletehydro

%sannel
nrun 100
niter 3000
doc
nodoc C* N* O10 O11
po 0 1 0
```

Example 5: structure solution of calcium glycinate trihydrate (CSD reference code: ZINXAX)

Folder `examples/Example 5 - calcium glycinate trihydrate` contains the following files:

`cagly.exp`: EXPO input file for structure solution by direct space method.
`sk3493Isup2.rtv`: file containing powder diffraction data.

The input file called `cagly.exp` is here reported:

```
%job calcium glycinate trihydrate (CSD refcode: ZINXAX)
%structure cagly

%data
pattern sk3493Isup2.rtv
cell 9.6572 9.6878 5.7627 90.588 76.997 97.467
space P-1
wave 1.540560

%frag smiles C(C(=O)[O-])N
%frag smiles C(C(=O)[O-])N
%frag atoms CaO3
deletehydro

%sannel
nrun 100
niter 20000
refinetf molecules
```

Example 6: structure solution of LaTi$_2$Al$_9$O$_{19}$ (ICSD 262089)

Folder `examples/Example 6 - LaTi2Al9O19` contains the following files:
`LaTi.exp`: EXPO input file for structure solution by direct space method.
`hw5019Isup2.rtv`: file containing powder diffraction data.

The input file called `LaTi.exp` is here reported:

```
%job La Ti2 Al9 O19 (ICSD 262089)
%structure LaTi

%data
pattern hw5019Isup2.rtv
cell 22.59355 10.99919 9.72968 90 98.5634 90
space C2/c
wave 1.54059

%fragment atoms La
%fragment octa Ti O
%fragment octa Ti O
%fragment tetra Al O
%fragment tetra Al O
%fragment octa Al O
```

```
%fragment octa Al O
%fragment octa Al O
%fragment octa Al O
%fragment octa Al O
%fragment octa Al O
%fragment octa Al O
%fragment octa Al O

%sannel
nrun 100
doc
```

_____

Parallelism in Expo2014 for structure solution by direct space method

http://www.ba.ic.cnr.it/softwareic/expo/parallelism-in-expo2014-for-structure-solution-by-direct-space-method/

A parallel version of Expo2014 is available if you want to speed up your crystal structure solution by direct space method (DSM) using more than one processor and exploiting the whole available computing power. Different parallelization paradigms are available to benefit from parallelism but in Expo2014 only Message Passing Interface (MPI) is at moment implemented and only DSM have been parallelized.

Running DSM in parallel permits tackling structures with great complexity in a reasonable time. Structures with more than 15-20 degrees of freedom, several molecular fragments, presence of largely flexible atomic chains represent a complex problem for the crystal structure solution by DSM: large number of SA moves per run, large number of runs are required to guarantee finding the global minimum and increasing the frequency of correct solutions. In addition, the use of dynamical occupancy correction in the case of non-molecular compounds or the introduction of constraints and restrains, requires the computation of geometrical parameters (distances and angles) between neighboring atoms including all symmetry equivalent positions. These expensive calculations could increase dramatically the computation time when there is a large number of atoms.

Fortunately the type of calculation for DSM can be easily distributed between more CPU-core. In a MPI job, different program copies are created, each copy has the own private memory and execute the same code independently from others but with the possibility to communicate with other by set of routines that make up the MPI library. In the parallel Expo2014 all the global optimization trials are distributed between the CPU-cores and each CPU-core runs a proper independent trials of structure solution. At the end the solutions found by the CPU-cores are communicated to the master process that orders all the structural models based on the cost function and performs output operations. In the current version of the program the CPU-cores do not exchange any data during the simulation or, in other words, the CPU-cores do not collaborate with each other. Because the overhead due to communications is practically zero, the parallel speedup (S_p) is almost identical to the number of CPU-cores ($S_p \lesssim N_{CPU}$). Implementations of the algorithm with periodical cooperation by exchanging the best solutions found between CPU-cores are under testing.

Running the parallel version of EXPO requires:

- Computer with multi-core CPUs: laptop, desktop PC, workstation, supercomputer.
- Linux operating system installed.
- Open MPI installed.
- Install EXPO from source and linking with MPI libraries or from debian package.
- Run EXPO by using the launcher `mpirun` with the appropriate options.

Installing the parallel version of Expo2014

The debian binary packages (see the download page) contain Expo2014 compiled with Open MPI. Alternatively, you can compile Expo2014 from source and link with MPI libraries. Read and follow the instructions in the section Install build tools and graphic development packages.

Parallel execution

The typical way to run Expo2014 under MPI is by using `mpirun` that is invoked from the command line with the arguments consisting of the number of MPI processes, the name of the program executable followed by the name of the input file, *e.g.*

```
mpirun -np 10 expo input_file.exp
```

Here the option `-np` specifies the number of MPI processes to run. The MPI process are the instances of program that are running and should be not more than the total number of CPU-cores to avoid degradation of performance. 10 processes are used in this case. The calculation will be distributed between 10 CPU-cores in a parallel computer with 10 or more CPU-cores. During a default DSM 10 run of the global optimization algorithm are performed so each CPU-core will perform 1 run. The number of runs can be increased by using the directive [nrun](#) in the input file. If the number of processes required exceeds the number of runs, the number of runs will be automatically extended to the number of processes exploiting all the hardware resources specified by the user with the `np` option.

The graphical user interface (GUI) is not displayed when Expo2014 runs in parallel, in order to monitor the progress of job, the user can consult the `.out` output file. When the calculation is completed, the program creates a project file `structure_name.expo` that can be used to display the results by GUI. In order to examine the results of the job, the `.expo` project file can be opened using Expo2014 from **File > Old Project** or typing the following command:

```
expo structure_name.expo
```

The best solution will be displayed. To access to the list of all structural models generated during the DSM select
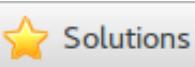
Solve > Simulated Annealing and click on the [Solutions] button.

You can test the parallel version of Expo2014 on the structure paracetamol reported in the [examples directory](#). Copy the file `paracetamol.exp`, `paracetamol.xy` and `paracetamol.mol` in a directory whose name might be `test_mpi`, change the working direcory in `test_mpi`.

```
mkdir test_mpi
cd test_mpi
cp /usr/local/share/expo/examples/paracetamol.*  .
```

In the working directory `test_mpi` run the following command using a PC-biprocessor machine:

```
mpirun -np 2 expo paracetamol
```

where instead of `expo` you might declare the complete path of the folder where Expo2014 parallel is installed. In few minutes you should get something like this picture

```
                          corrado@Z420-1: ~/test_mpi                    _  □  ✕

File  Edit  View  Search  Terminal  Help

corrado@Z420-1:~/test_mpi$ mpirun -np 2 ~/work/expo_mpi/expo paracetamol

  Expo2014 will go on without graphics.
  Results will be written in the file
        paracetamol.out

  Direct space method will run on 2 processors ...

     --------------------------------------------------------------------
        List of structural models sorted according to cost function
     RUN        CF              Ratio          time          saved on
     --------------------------------------------------------------------
      1       16.293200         4.990      0h 0m 21.544s    paracetamol_best1.cif
      4       16.295530         4.991      0h 0m 21.392s    paracetamol_best2.cif
      8       16.295530         4.991      0h 0m 21.368s    paracetamol_best3.cif
      2       16.296677         4.991      0h 0m 21.396s    paracetamol_best4.cif
      6       16.296677         4.991      0h 0m 21.372s    paracetamol_best5.cif
      5       16.301003         4.993      0h 0m 20.104s    paracetamol_best6.cif
      9       16.301003         4.993      0h 0m 20.084s    paracetamol_best7.cif
      3       16.306234         4.994      0h 0m 20.676s    paracetamol_best8.cif
      7       16.306234         4.994      0h 0m 20.656s    paracetamol_best9.cif
     10       17.091347         5.235      0h 0m 22.264s    paracetamol_best10.cif
corrado@Z420-1:~/test_mpi$ □
```

If you look the content of the working directory `test_mpi` (use command `ls`), some output files have been created: `paracetamol.out` containing general information about the job, CIF files `paracetamol_best1.cif`, `paracetamol_best2.cif`, … with the best structural models, a project file called `paracetamol.expo`. Load `paracetamol.expo` from **File > Old Project** or by the following command:

`expo paracetamol.expo`

Then **Solve > Simulated Annealing** and click on the [⭐ Solutions] button for a visual inspection of the best solutions.
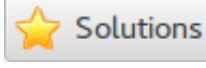
_____

Rietveld Refinement

http://www.ba.ic.cnr.it/softwareic/expo/rietveld-refinement/

Structure refinement by the Rietveld method (Rietveld, 1969) can be carried out by EXPO2014. Graphical and computing tools are available in the program in order to attain a successful automatic or user-defined Rietveld refinement.

The refined variables include peak shape parameters, unit cell dimensions, fractional coordinates of atoms in the structure model, background coefficients, scale factor, isotropic displacement parameters. Atomic displacement parameters can be refined individually or in a group of atoms with the same atomic type or the same environment. Correction for preferred orientation can be achieved by the March-Dollase approach. The background is described using empirical functions: classical polynomial function, Chebyshev polynomial function, cosine Fourier series. The background can also modeled by mouse-click selection of pattern points. The available peak shape functions are: Pseudo-Voigt, Pearson VII, TCH pseudo-Voigt.

The non-linear least squares are carried out by employing the damped Gauss-Newton method. The refinement convergence condition is reached when the increments on parameters become smaller with respect to their standard deviations or when the relative gradient of $?^2$ minimized function is less then a tolerance value. Tolerance value and the maximum number of cycles could be suitably modified by the user.

The Le Bail technique can be adopted to perform a full pattern decomposition prior to Rietveld refinement in order to determine starting values of parameters (background, peak shape, zero shift and unit cell dimensions) to refine. This strategy is suggested especially if the available structure model is not completed (McCusker *et al.*, 1999) or when the starting model is too different from the target model.

The refinement can be carried out by following two alternative approaches: 1) the user can decide the refinement strategy *via* graphic interface; 2) an automatic refinement schedule can be applied. In the automatic mode, groups of parameters are refined according to a fixed sequence as established in the Rietveld guidelines (Young, 1995). In the last step of refinement all parameters are refined simultaneously.

The knowledge of molecular geometry can be exploited in the refinement in form of restraints on bond distances, angles and planes. To simplify the setting of restraints, the program is able to extract from the connectivity of the initial model the possible restraints providing a list of current and target values. The user can select the restraints to be included in the refinement procedure and eventually modify the target value.

Some figures of merit are available in order to follow the progress and evaluate the quality of refinement: the weighted profile R-factor and the unweighted profile R-factor calculated for the full pattern (R_{wp}, R_p); the corresponding figures background-subtracted (R_{wp}', R_p'); the statistically expected R value (R_{exp}); the goodness-of-fit (R_{wp}/R_{exp}); the Durbin-Watson d-statistic. R values similar to those used in case of single crystal data are also available: R_F and the Bragg-intensity R which use the structure factor moduli extracted from the experimental profile.

$$R_p = \frac{\sum\limits_{i}^{N} |y_{i,obs} - y_{i,calc}|}{\sum\limits_{i}^{N} y_{i,obs}} \times 100$$

$$R_{wp} = \sqrt{\frac{\sum\limits_{i}^{N} w_i \cdot (y_{i,obs} - y_{i,calc})^2}{\sum\limits_{i}^{N} w_i \cdot (y_{i,obs})^2}} \times 100$$

$$R_{exp} = \sqrt{\frac{N-p}{\sum\limits_{i}^{N} w_i \cdot (y_{i,obs})^2}} \times 100$$

$$R'_{wp} = \left[\frac{\sum\limits_{i}^{N} w_i (y_i - Y_{ci})^2}{\sum\limits_{i}^{N} w_i (y_i - y_{bi})^2} \right]^{1/2}$$

$$R_{exp} = \left[\frac{(N-P)}{\sum\limits_{i}^{N} w_i y_i^2} \right]^{1/2}$$

$$\chi^2 = \frac{\sum\limits_{i}^{N} w_i \cdot (y_{i,obs} - y_{i,calc})^2}{N-P} = \left[\frac{R_{wp}}{R_{exp}} \right]^2$$

$$R_F = \frac{\sum\limits_{j}^{m} |F_{j,obs} - F_{j,calc}|}{\sum\limits_{j}^{m} F_{j,obs}} \times 100$$

$$R_B = \frac{\sum\limits_{j}^{m} |I_{j,obs} - I_{j,calc}|}{\sum\limits_{j}^{m} I_{j,obs}} \times 100$$

where N is the total number of counts in the pattern and P is the number of refined parameters.

In addition, graphical tools are available for visualizing: 1) the observed, calculated and difference pattern and the cumulative $?^2$ value; 2) the evolution of the structure model during the refinement in real time.

Preparing input file for Rietveld refinement

To run EXPO2014 for Rietveld refinement, first you need to create the input file. This is an example of the typical input file:

```
%Structure paracetamol
%Job paracetamol
%Data
Cell  7.100  9.380  11.708  90.0  97.42  90.0
SpaceGroup p 21/n
Pattern   paracetamol.xy
Wavelength 1.54056
%fragment paracetamol.cif
%rietveld
```

It is supposed that the cell parameters and the space group have been determined before and a structure model is available and imported by using the command `%fragment` followed by the file name containing the structure. **`%rietveld`** is the command to access to the Rietveld Refinement graphical interface.
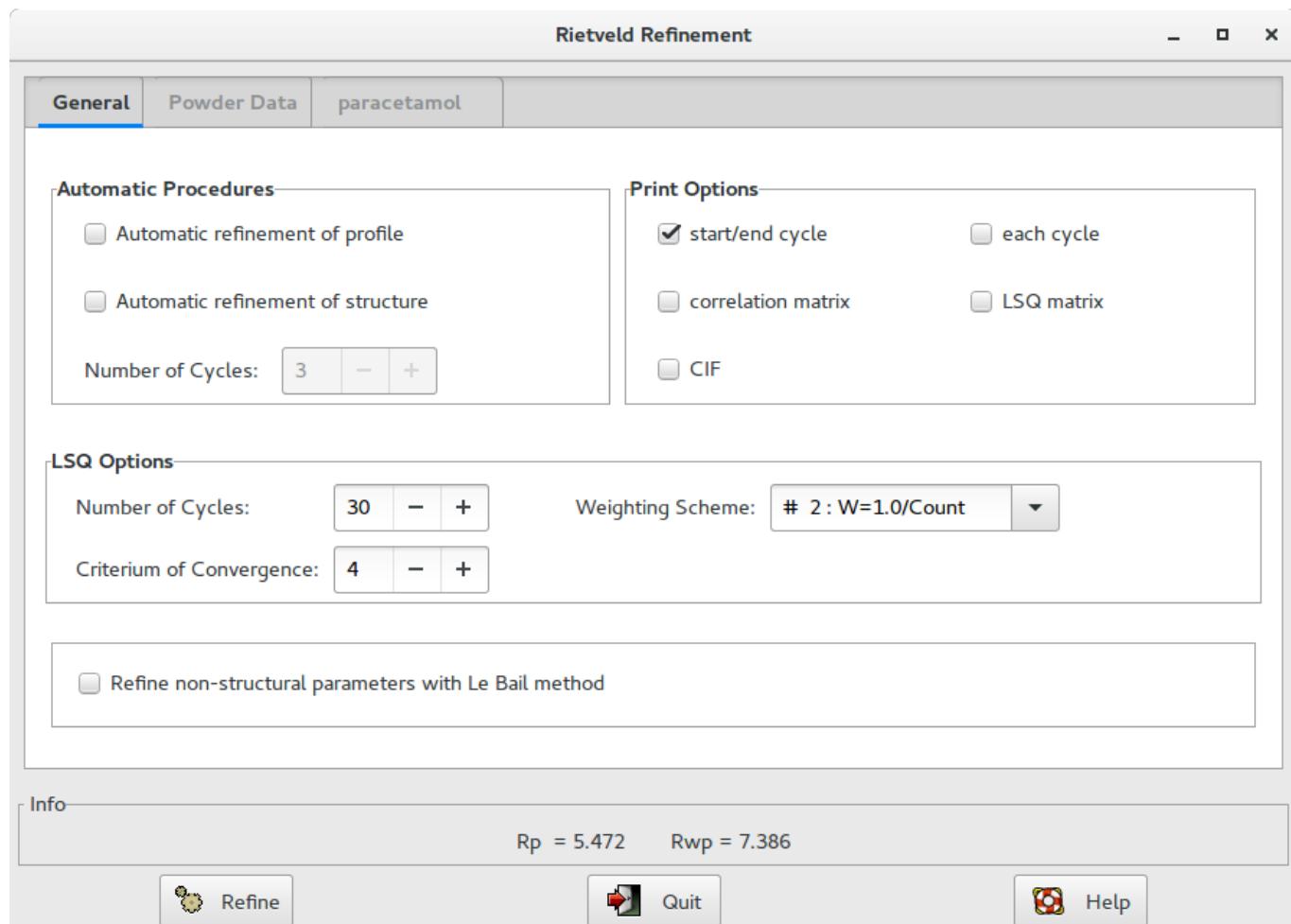
Alternatively, when a structure model is available, you can access directly to the graphic interface from menu **Refine > Rietveld**.
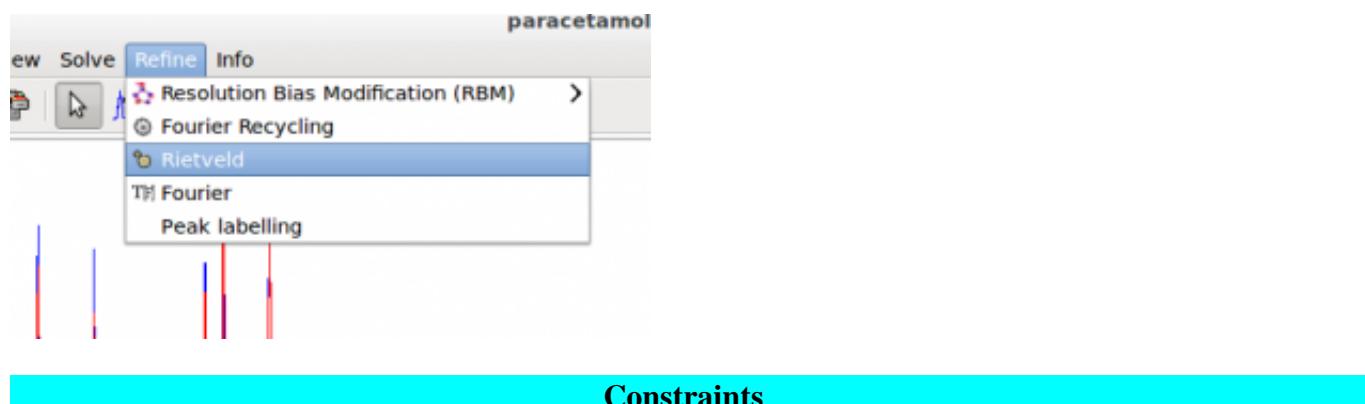
Constraints

Constraints are defined as exact mathematical relationships between least squares parameters and can be used to reduce the number of parameters.

Symmetry constraints, required to conserve the space-group symmetry requirements, are mandatory and automatically deduced and imposed by the program. For example, an atom on the special position (0.5 , 0.5 , 0.5) in P-1 should not be refined at all and its parameters are constrained to these special values. Alternatively, an atom on

the special position (x, x, x) in space-group P23 can be refined but must be constrained to have equal shifts applied in the x, y and z directions. Lattice symmetry establishes specific relationships between the unit cell dimensions (*e.g.*, in cubic a=b=c and alpha=beta=gamma). EXPO2014 automatically deduces and imposes these constraints.

Constraints are often employed to control **site occupancies** when disorder is present in one or more sites. It is not uncommon in minerals to find an atomic site that can be occupied by either of two element types. For example, if three atoms A, B, and C occupy the same site then, assuming the full total occupancy of the site, the following relationship (or constraint) results: $occ_A + occ_B + occ_C = 1$. Thus, only two of the fractional occupancies should be refined, for example, occ_A and occ_B are free least squares variables but $occ_C = 1 - occ_A - occ_B$. Atomic displacement parameters (ADP) are made to shift synchronously. EXPO2014 automatically identifies atoms that occupy the same site and impose the constrains on site occupancies and on ADPs.

Other constraints may be imposed by the user. Clicking on the button **Constraints** in the **Structure** section under the tab with the structure name (*paracetamol* in figure above) is possible to access to a dialogue window to define group of ADPs that can have the same shift.
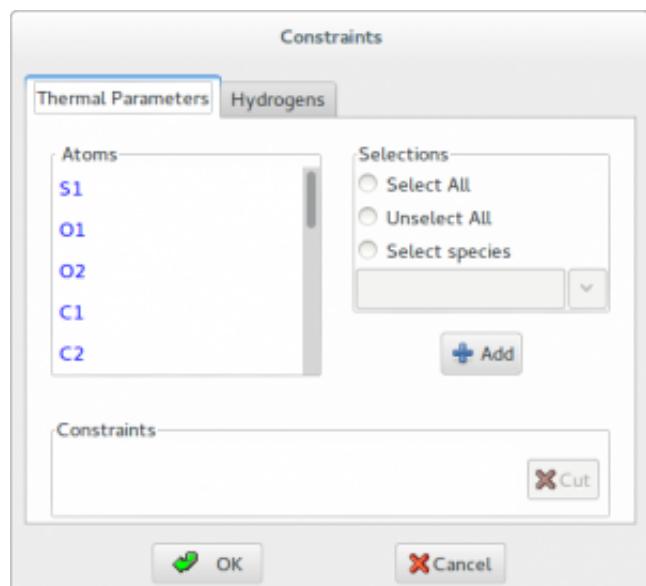
Contraints are applied to H atoms by using the **riding model**. This technique is used to move H atoms synchronously with the C atoms to which they are bonded, thereby preserving the bond length and direction. The isotropic displacement parameters U_{iso} of the hydrogens are constrained to be 1.2 times of that of the heavy atom to which they are attached. The factor 1.2 can be changed by graphic interface.
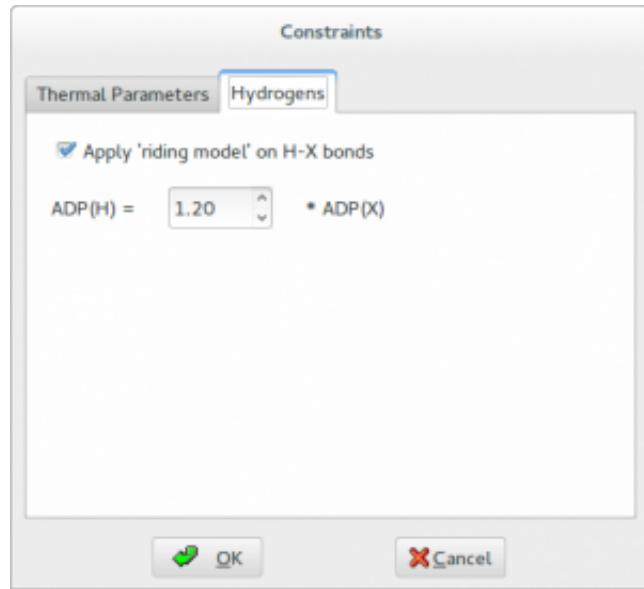
Constraints on atomic parameters can be introduced by using the directive constr that has the following syntax:

```
constr constraint_equation
```

For example, if you want apply constraints on isotropic atomic displacement parameters use the following directive
```
constr B(At1) = B(At2) = B(At3) = …
```
in which the equation is chain of equalities. At1, At2, At3, … are the atomic labels or the order number of atoms

E.g.,
```
constr B(La1) = B(La2)
```
and the atomic displacement parameters of lanthanum atoms La1 and La2 are constrained at identical values.
The special character * can be used to specify more atoms in a constraint equation, for example, if you want apply a constraint on the displacement parameters of all 6 carbon atoms labelled C1, C2, C3, C4, C5, C6 you can write
```
constr B(C1) = B(C2) = B(C3) = B(C4) = B(C5) = B(C6)
```
or in more sintetic way
```
constr B(C*)
```

To constrain the isotropic atomic displacement parameters of all C and O atoms
```
constr B(C*) = B(O*)
```

If different atoms occupy the same crystallographic site is necessary apply a restrictions on the coordinates and on the occupation factor. Generally also the displacement parameters should be constrained.

E.g.,
```
constr xyz(La1) = xyz(Sn1)
constr occ(La1) + occ(Sn1) = 1
constr B(La1) = B(Sn1)
```
if La1 and Sn1 occupy the same site, assuming the full total occupancy of the site.

Restraints

The knowledge of molecular geometry can be exploited in the refinement in form of restraints on bond distances,

angles and planes. To simplify the setting of restraints, the program is able to extract from the connectivity of the initial model the possible restraints providing a list of current and target values. The user can select the restraints to be included in the refinement procedure and eventually modify the target value.
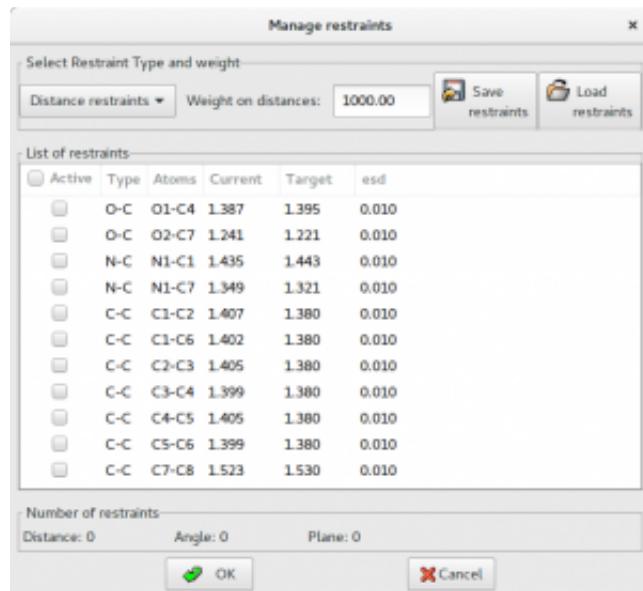
In case of structure solved by using direct space methods, bond distances and angles are generally accurate and the target value could be set equal to the current. The directive `res_target_type current` after the command `%rietveld` can be used to enable this particular setting.

<div style="background:cyan">

References

</div>

Rietveld, H.M. *(1969)*. J. Appl. Cryst. **2** 65–71.

McCusker, L.B.; Von Dreele, R.B.; Cox, D.E.; Louër, D.; Scardi, P. (1999). *J. Appl. Cryst.* **32**, 36-50.

The Rietveld Method. Edited by R. A. Young, Oxford University Press, Oxford (1995).

———————————————————————————

Crystal Structure Visualization

http://www.ba.ic.cnr.it/softwareic/expo/crystal-structure-visualization/

EXPO2014 includes a viewer for 3D crystal structure visualization. This viewer is a modified version of **Jav** (**J**ust **A**nother **V**iewer), program to visualize structural models and electron density maps.
The program uses GTK Cairo and OpenGL libraries which are distributed under the GPL (GNU General Public License).

_____

EXPO2014 commands
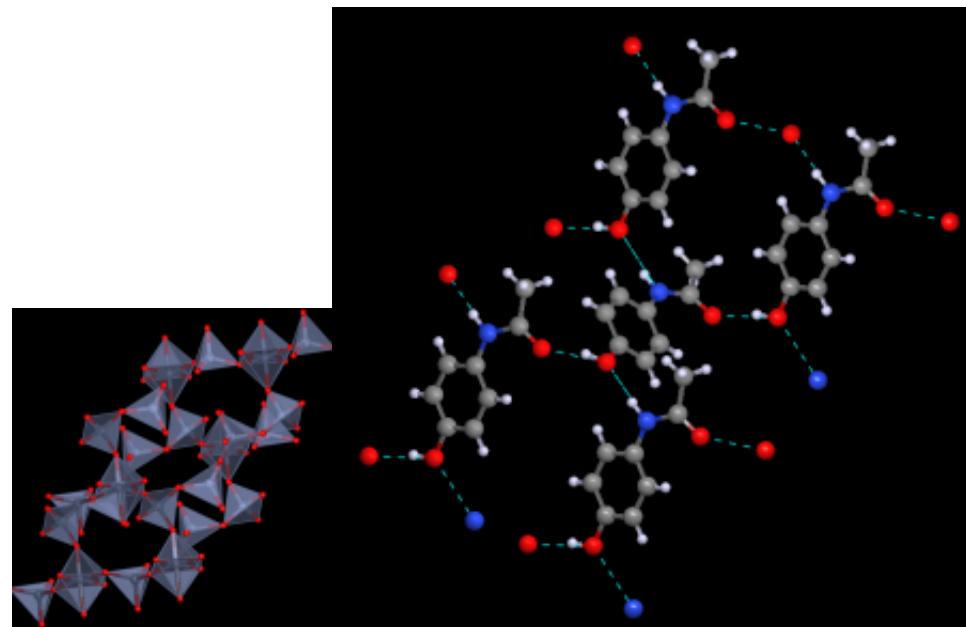
http://www.ba.ic.cnr.it/softwareic/expo/description-of-expo2014-commands/

The main modules of the program are:

EXPO, LIST, DATA, NTREOR, EXTRACTION, NORMAL, INVARIANTS, PHASE, FOURIER/LEAST-SQUARES, DIRECT, PATTERSON.

EXPO module

It interprets commands and calls desired routines.

LIST module

It is the software interface between EXPO2013 and the direct access file on which data and results are stored.

DATA module

This routine reads the basic crystallographic informations like cell parameters, space group symbol (it may be also found by the program, using "findspace" directive), wavelength, unit cell content and the counts (or reflections). It includes a modified version of the subroutine SYMM Burzlaff & Hountas (1982). Symmetry operators are directly derived from the space group symbol.

An external file (containing counts or reflections) is supplied by the user. If an external reflection file is supplied (avoiding the extraction step), diffraction data are checked in order to find out equivalent reflections or systematically absent reflections (which are then excluded from the data set).

N-TREOR module

This routine performs the indexing step (a peak search procedure is automatically carried out by the program before the indexing step). If available, an external peak positions file can be supplied to the program and the automatic peak search procedure will be skipped by EXPO2014. The indexing procedure retains the basic search algorithm of TREOR90, (Werner *et al.*, 1985) and has been strengthened in the last version of N-TREOR09 (Altomare et al., 2009).

EXTRACTION module

In this module diffraction integrated intensities are extracted from the powder diffraction pattern. To calculate the integrated intensities, the program uses the Le Bail algorithm (Le Bail, Duroy & Fourquet, 1988) and combines it with a least squares procedure minimizing the quantity:

$$?w_i \, (y_{oi} - y_{ci})^2 \tag{1}$$

NORMAL module

In this module diffraction intensities are normalized using the Wilson method (Wilson, 1942). Statistical analysis of the intensities is made in order to suggest the presence or absence of the inversion centre, to identify the possible presence and type of pseudotranslational symmetry (Cascarano *et al.*, 1988 a,b; Fan, Yao & Qian, 1988) and to detect preferred orientation effects (Altomare *et al.*, 1994; Altomare *et al.*, 1996). Possible deviations (of displacive type) from ideal pseudotranslational symmetry are also detected. All the above information is used as prior information in the integrated intensities extraction process, unless the directive NOPSEUDO is given to the program. When some additional prior information, besides positivity and atomicity of electron density, is available, then a suitable renormalization of structure factors is made.

`INVARIANTS` module

Up to 10000 triplets relating reflections with normalized E values greater than a given threshold (strong triplets) are stored for active use in the phasing process. Also triplets (PSI-ZERO triplets) relating two reflections with large E and one with E close to zero are generated: they are used to define a special figure of merit (PSCOMB). Special types of triplets (PSI-E triplets) based on two strong and one intermediate reflections (just below the threshold of strong reflections) are calculated and used in the FOURIER/LEAST-SQUARES module in order to extend phase information (Altomare *et al.*, 1991). Negative quartets are generated by combining the psi-zero triplets in pairs, and those with cross-magnitudes smaller than a given threshold are estimated by means of their first representation, as described by Giacovazzo (1976). These quartets are used to provide an important contribution to the CPHASE FOM.

`PHASE` module

In the EXPO2014 program the most reliable one-phase s.s. are treated as known phases. Besides triplets, also the most reliable negative quartets and two-phase s.s. may be actively used. Each relationship is used with its proper weight: the concentration parameter of the first representation for quartets and two-phase s.s., and C or G for triplets.

– CONVERGENCE/DIVERGENCE PROCEDURE –

The convergence procedure (Germain *et al.*, 1970) is a convenient way of defining an optimum starting set of phases to be expanded by the tangent formula or by any other algorithm. When the P10 formula is used, as a default, a special convergence process is devised which chooses the starting set.
Once the starting set has been defined, a good pathway for phase expansion is determined by a divergence procedure. In the divergence map, starting from the reflections in the starting set, each new reflection is linked to the preceding ones with the highest value of $< ? >$.

– PHASE EXTENSION AND REFINEMENT –

The starting set defined by the preceding step is usually formed by the origin (and enantiomorph) fixing reflections, a few one phase s.s. and a number of other phases which may be obtained:
a) by magic integer permutation (White & Woolfson, 1975; Main, 1978),
b) by a random approach (Baggio *et al.*, 1978; Burla *et al.*, 1992).
The option a) is the default, b) runs if the directive RANDOM is used. In this last case a large number (depending on the available computer time) of trials can be requested.

FOURIER/LEAST-SQUARES module

The sets of phases generated by the tangent routine are first expanded through psi-E relationships and then passed to the fast fourier transform routine written by L.F.Ten Eick (1977) and subsequently modified by the MULTAN team (Main et al., 1980).

DIRECT module

In this module some "direct-space" approaches for solving crystal structures from powder diffraction data, can be used. The basic idea is to actively use structural and geometrical information, available a priori, during the structure solution process.

LABELLING procedure

The atomic species assigned to the peaks of an electron density map obtained at the end of a direct method procedure, can be modified using chemical information about the structure. The procedure (Altomare *et al.*, 2002) requires: 1) the prior information on the number and coordination of the heavy atoms (the procedure can only handle tetrahedral and octohedral coordinations); 2) the range of the typical distances between heavy atoms; 3) the

range of the typical distances between heavy and light atoms.

POLPO procedure

Using the "%polyhedra" command, two procedures can be activated. Both of them use a Monte Carlo approach and exploit the experimental information about the heavy – atom connectivity. The first one (POLPO1) (Altomare et al., 2000) is suggested when all the heavy atoms are rightly positioned and labelled. It is able to correctly locate the light atoms. The second one (POLPO2) (Giacovazzo et al., 2002) is able to position one or more missing cations and surrounding anions.

PATTERSON module

In the EXPO2014 program it is possible to compute a PATTERSON map using various coefficients. By using a special directive (INVERSION) it is possible to exploit the positivity of the electron density (in the direct space) in the extraction routine. The so-modified Patterson map is inverted providing a set of structure factor values to use in a new extraction process.

References

Altomare A., Burla M.C., Cascarano G., Giacovazzo C., Guagliardi A., Moliterni, A.G.G. & Polidori, G. (1996). *J. Appl. Cryst.* **29**, 341-345.

Altomare A., Cascarano G., Giacovazzo C., Guagliardi A., Burla M.C., Polidori G. & Camalli M. (1994). *J. Appl. Cryst.* **27**, 435 – 436.

Altomare A., Cascarano G., Giacovazzo C. & Viterbo D. (1991). *Acta Cryst.* A**47**, 744-748.

Altomare, A., Campi, G., Cuocci, C., Eriksson, L., Giacovazzo, C., Moliterni, A., Rizzi, R. & Werner, P.-E. (2009). *J. Appl. Cryst.* **42**, 768-775.+

Altomare A., Giacovazzo C., Guagliardi A. Moliterni A.G.G., Rizzi R. (2000). *J. Appl. Cryst.* **33**, 1305-1310.

Altomare A., Giacovazzo C., Ianigro M., Moliterni A.G.G., Rizzi R. (2002). *J. Appl. Cryst.* **35**, 21-27.

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

Burla M.C., Cascarano G. & Giacovazzo C. (1992). *Acta Cryst.* A**48**, 906-912.

Burzlaff H. & Hountas A. (1982). J. Appl. Cryst. **15**, 464-467.

Cascarano G., Giacovazzo C. & Luic' M. (1988a). *Acta Cryst.* A**44**, 176-183.

Cascarano G., Giacovazzo C. & Luic' M. (1988b). *Acta Cryst.* A**44**, 183-188.

Fan H., Yao J. & Qian J. (1988). Acta Cryst. A**44**, 688-691.

Germain G., Main P. & Woolfson M.M. (1970). *Acta Cryst.* B**26**, 274-285.

Giacovazzo C. (1976). *Acta Cryst.* A**32**, 958-966.

Giacovazzo C., Altomare A., Cuocci C., Moliterni A.G.G., Rizzi R. (2002). *J. Appl. Cryst.* **35**, 422-429.

Le Bail A., Duroy H. & Fourquet J.L. (1988). *Math. Res. Bull.* **23**, 447-452.

Main P. (1978). *Acta Cryst.* A**34**, 31-38.

Main P., Fiske S.J., Hull S.E., Lessinger L., Germain G., Declercq J.P. & Woolfson M.M. (1980) – MULTAN80, a system of computer programs for the automatic solution of crystal structures from x-ray diffraction data – Univ. of York, England.

Ten Eick L.F. (1977). *Acta Cryst.* A**33**, 486-496.

Werner P.-E., Eriksson L. & Westdahl M. (1985). *J. Appl.Cryst.* **18**, 367-370.

White P.S. & Woolfson M.M. (1975). *Acta Cryst.* A**31**, 53-56.

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

_____

How to

http://www.ba.ic.cnr.it/softwareic/expo/how-to/

Many possibilities are available to interact with the program: through the toolbar, the menu and using keyboard/mouse.

The menu items follow:

- File
- Modify
- Select
- View
- Tools
- Info

_____

File

http://www.ba.ic.cnr.it/softwareic/expo/file/

The sub-menu File is shown.
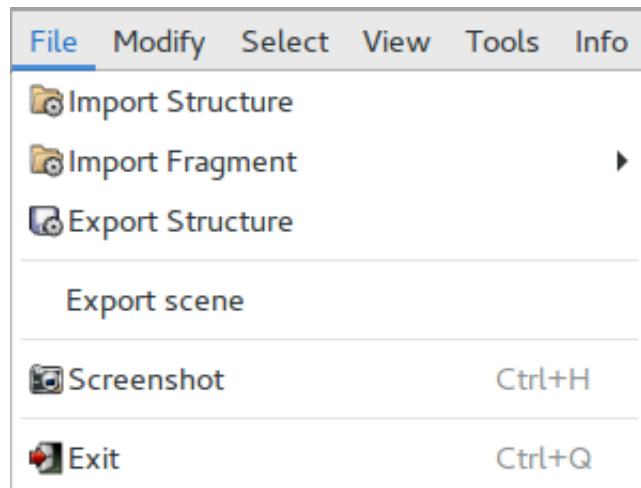
It allows:

Import Structure   Load a structure from file

Import Fragment   Load a molecular fragment from file

Export   Export current crystal data to a file

Export Scene   Export STL file for 3D printing

Screenshot   Produce an image file of Jav screen

Exit   Close the viewer

Import
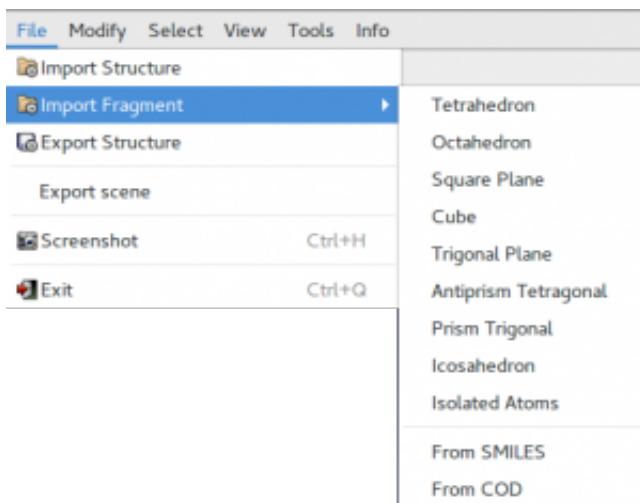
It is possible to read crystal structure from the following types of file:

- Crystallographic Information File (`*.cif`) http://www.iucr.org/resources/cif
- MDL Molfile (`*.mol`) http://accelrys.com/products/informatics/cheminformatics/ctfile-formats/no-fee.php
- MOPAC File Formats. It can handle data structure in Cartesian coordinates or in internal coordinates. http://openmopac.net/manual/index.html
- Tripos SYBYL Mol2 File (`*.mol2, *.ml2`) http://tripos.com/data/support/mol2.pdf
- Protein Data Bata Bank (`*.pdb`) http://www.wwpdb.org/
- Fenske-Hall Z-matrix (`*.zmt`), Z-matrix file created by Open Babel
- Free Form Fractional (`*.fra, *.frac`) http://openbabel.org/wiki/Free_Form_Fractional
- XYZ Cartesian Format (`*.xyz`) http://en.wikipedia.org/wiki/XYZ_file_format
- Input/Output files of SHELXL (`*.ins,*.res`) http://shelx.uni-ac.gwdg.de/SHELX/

EXPO2014 can also read the coordinates of the final molecular geometry in the output files generated by well known quantum chemistry programs:

- GAMESS output file (`*.out,*.log`) http://www.msg.ameslab.gov/gamess/
- NWCHEM output file (`*.out,*.log`) http://www.nwchem-sw.org/index.php/Main_Page
- MOPAC output file (`*.out`) http://openmopac.net/MOPAC2016.html
- ABINIT output file (`*.out`) http://www.abinit.org/
- CRYSTAL output file (`*.out`) http://www.crystal.unito.it/index.php
- Quantum ESPRESSO output file (`*.out`) https://www.quantum-espresso.org/

In default the program export in the external file the asymmetric unit.

Isolated atoms and some common chemical geometries (tetrahedron, octahedron, square plane) can also imported. They are useful building blocks in structure solution by global optimization.

The program can read molecule from specified SMILES string and has the ability to import CIF files by querying the Crystallographic Open Database (COD). The button **From COD** allows the access to a graphical interface useful to build query to COD from appropriate search parameters (text contained in the metadata of the entry, chemical formula of the crystal, type and number of chemical element symbols). HTTP requests to get data from COD are performed by using RESTful API made available by COD developers to facilitate searches in the COD.
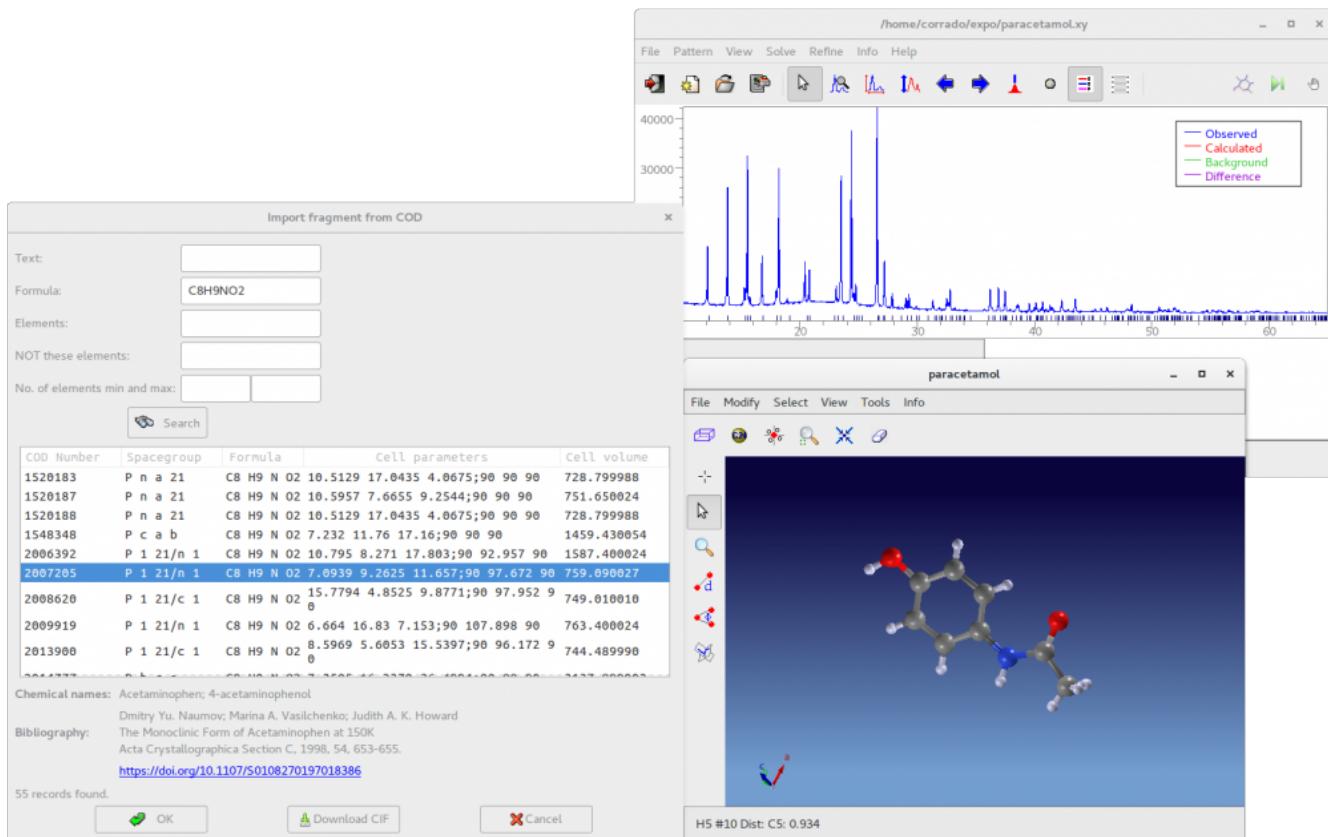
The picture above shows the result of the query to COD searching for the chemical formula of paracetamol ($C_8H_9NO_2$). All the molecules having the requested formula are listed and the user can visualize the 3D structure by selecting a row in the list. If powder data has been previously loaded, the program computes the reflections of the selected structure and displays them below the pattern in the main window, allowing a visual match between the experimental pattern and the reflections calculated from structure in the COD. Before the query, if the JAV viewer is closed, select from the menubar of the main window **View > JAV Molecular Viewer**.

Export

Information about the currently displayed crystal structure can be exported in the following files:

- Crystallographic Information File (`*.cif`) http://www.iucr.org/resources/cif
- Input/Output files of SHELXL (`*.ins`,`*.res`) http://shelx.uni-ac.gwdg.de/SHELX/
- Free Form Fractional (`*.fra, *.frac`) http://openbabel.org/wiki/Free_Form_Fractional
- Protein Data Bata Bank (`*.pdb`) http://www.wwpdb.org/
- POV-Ray input file (`*.pov`) input file for the freeware program POV-Ray. http://www.povray.org/download/
- MDL Molfile (`*.mol`) http://accelrys.com/products/informatics/cheminformatics/ctfile-formats/no-fee.php
- Fenske-Hall Z-matrix (*.zmt), Z-matrix file created by Open Babel
- MDL Molfile (`*.mol`) http://accelrys.com/products/informatics/cheminformatics/ctfile-formats/no-fee.php
- XYZ Cartesian Format (`*.xyz`) http://en.wikipedia.org/wiki/XYZ_file_format
- MOPAC File Formats with cartesian coordinates. http://openmopac.net/manual/index.html

EXPO2014 create input files for well known quantum chemistry programs. Input files for GAMESS and NWCHEM contain keywords for geometry optimization by density functional theory (DFT) calculations to perform using the B3LYP functional and a standard 6-31G* basis set. This approach is useful to generate molecular

model with accurate angles and bond lengths, suitable as a starting model for structure solution by real space methods

- GAMESS input file (`*.inp`,`*.gamin`) http://www.msg.ameslab.gov/gamess/
- NWCHEM input file (`*.out`,`*.log`) http://www.nwchem-sw.org/index.php/Main_Page
- GAUSSIAN Cartesian Input (`*.com, *.gau`) http://www.gaussian.com/
- GAUSSIAN Z-matrix Input (`*.gzmat`) http://www.gaussian.com/
- ABINIT input file (`*.in`) http://www.abinit.org/
- CRYSTAL input file (`*.d12`) http://www.crystal.unito.it/index.php
- Quantum ESPRESSO input file (`*.in`) https://www.quantum-espresso.org/

In default, the program exports only the coordinates of the atoms in the asymmetric unit to a file. To export all displayed atoms select the 'Displayed atoms' option in the file selection dialog box.

Screenshot

Images of crystal structure displayed in the graphic area can be saved in files of different following formats:

- Bitmap File (`*.bmp`)
- ICO file (`*.ico`)
- jpeg file (`*.jpg`)
- Portable Network Graphics file (`*.png`)
- Tagged Image File (`*.tif`)

Check the option 'Transparent' in the file selection dialog box the obtain images with transparent background.

Exit

This is used to exit the program.

_____

Modify

http://www.ba.ic.cnr.it/softwareic/expo/modify/

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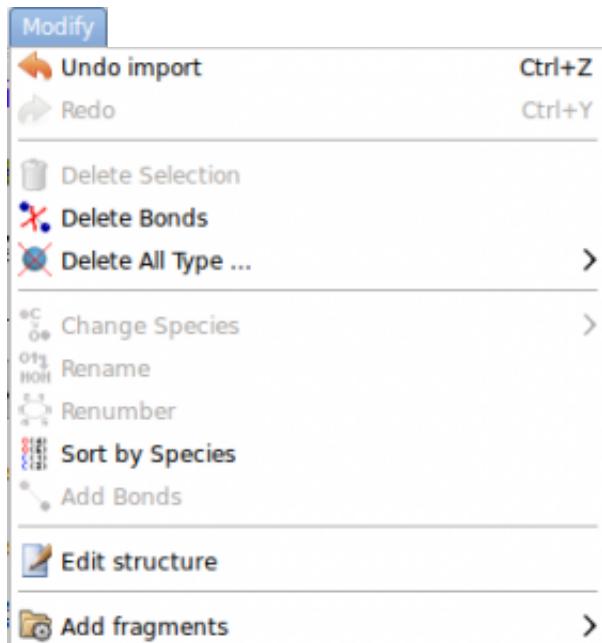
It allows:

Undo/Redo

Delete Selection

Delete Bonds

Delete All Type

Change Species

Rename

Renumber

Sort by Species

Edit structure

Add fragments

_____

Select

http://www.ba.ic.cnr.it/softwareic/expo/select/

The sub-menu select is shownon the left. It allows:
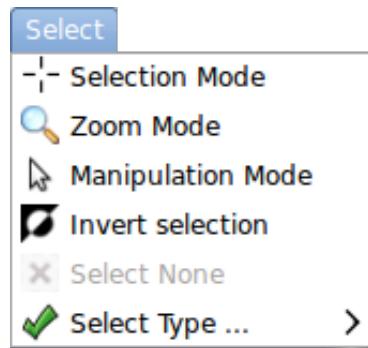
Selection Mode

Zoom Mode

Manipulation Mode

Invert selection

Select none

Select Type

_____

View

http://www.ba.ic.cnr.it/softwareic/expo/view/

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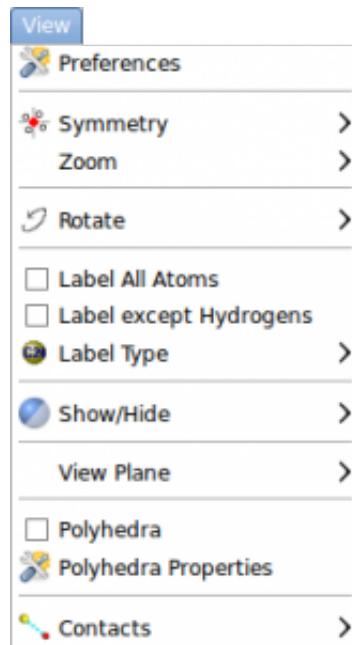
It allows:

Preferences

Symmetry

Zoom

Rotate

Label All Atoms

Label except Hydrogens

Label Type

Show/Hide

View Plane

Polyhedra

Polyhedra Properties

Contacts

Preferences

Some details about models are reported. The following window shows what modifications con be applied to the model.
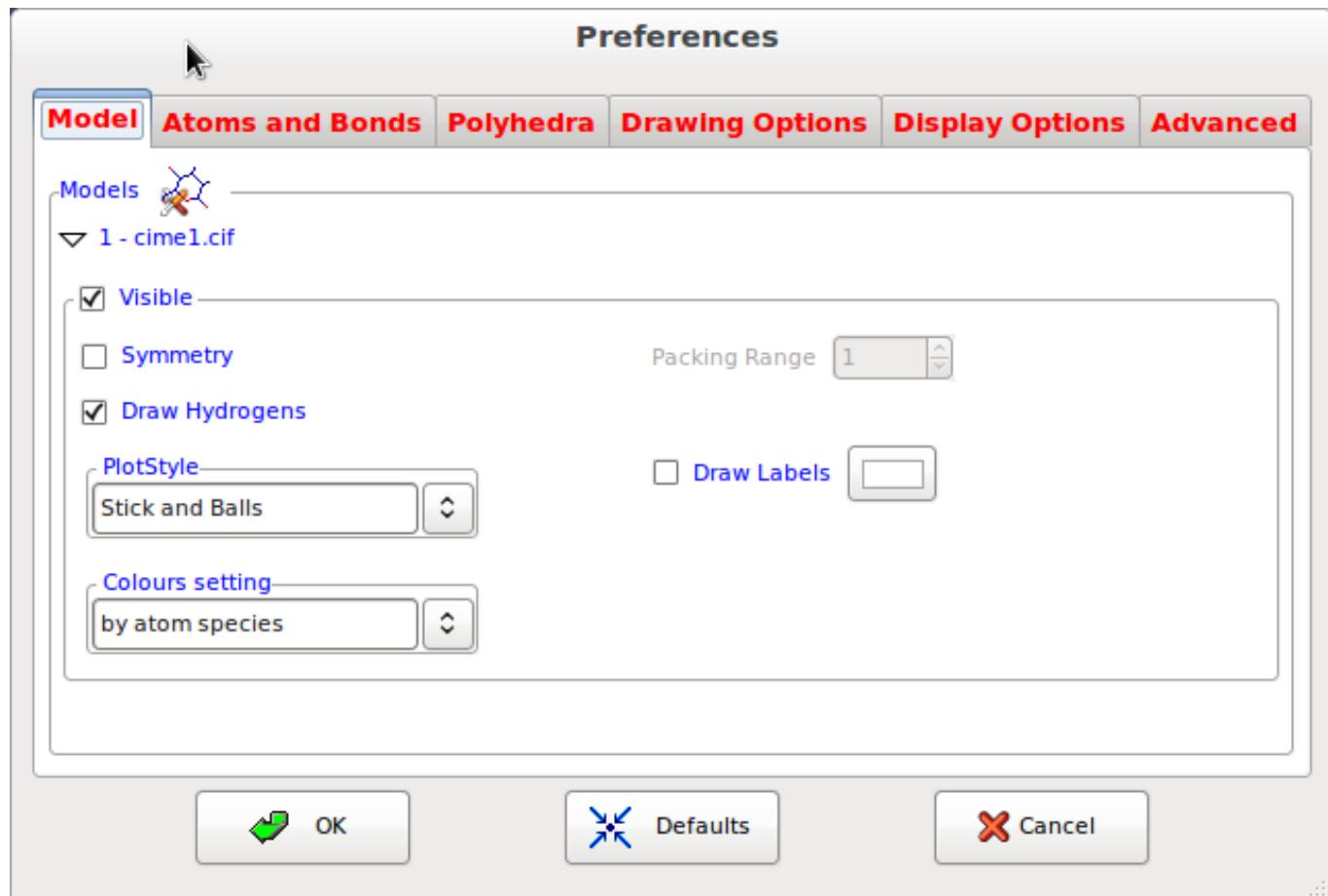
In order to get a better view of a model it is possible to tune the radius of the spheres representing the atoms and the radius of the cylinders representing the bonds. The atom colors can be modified.
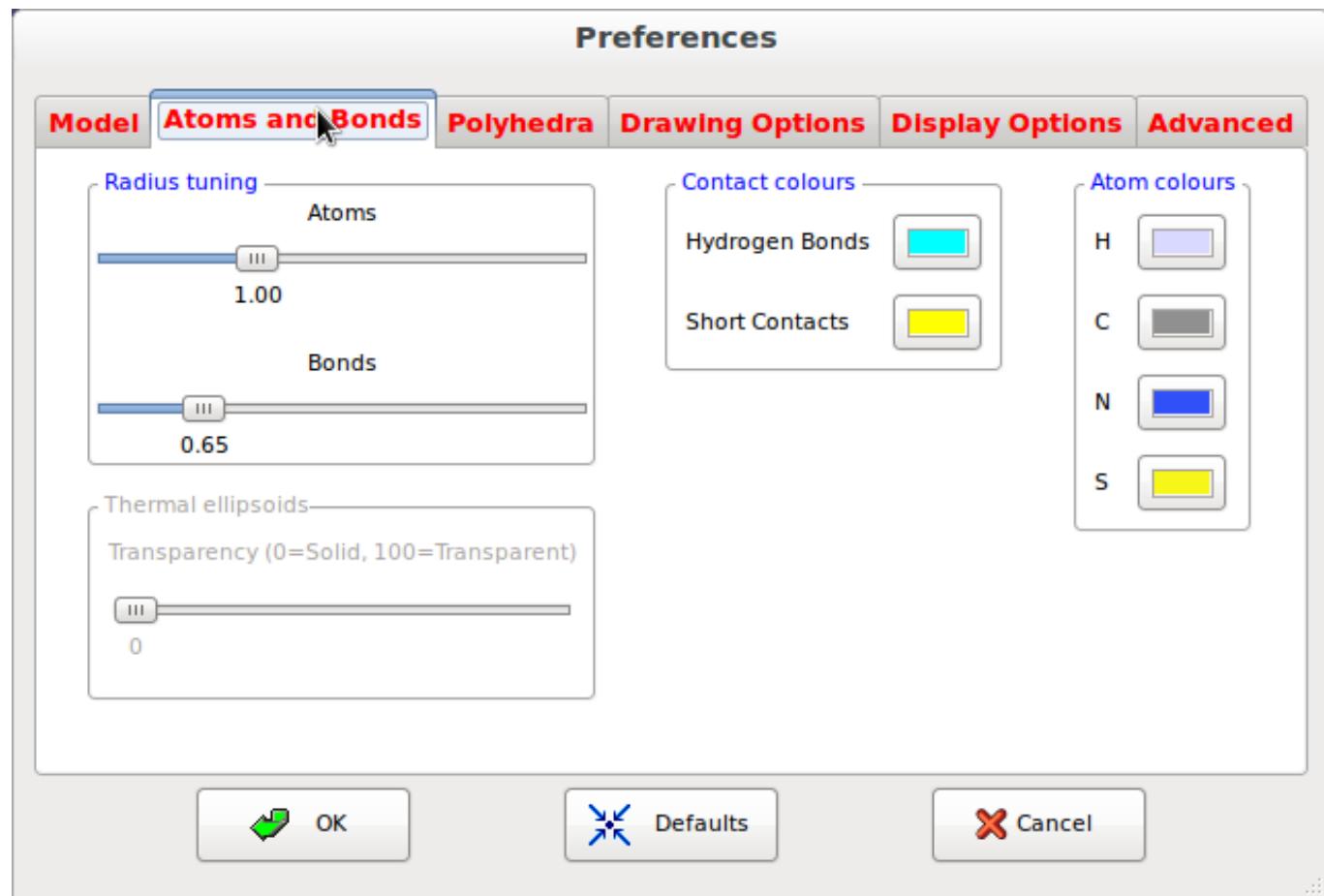
Polyhedra Properties

This window provides some tools to set the properties of the polyhedra; it is possible to visualize or to hide some of them, it is possible to change the colors and the transparency, to show or to hide the edges etc.
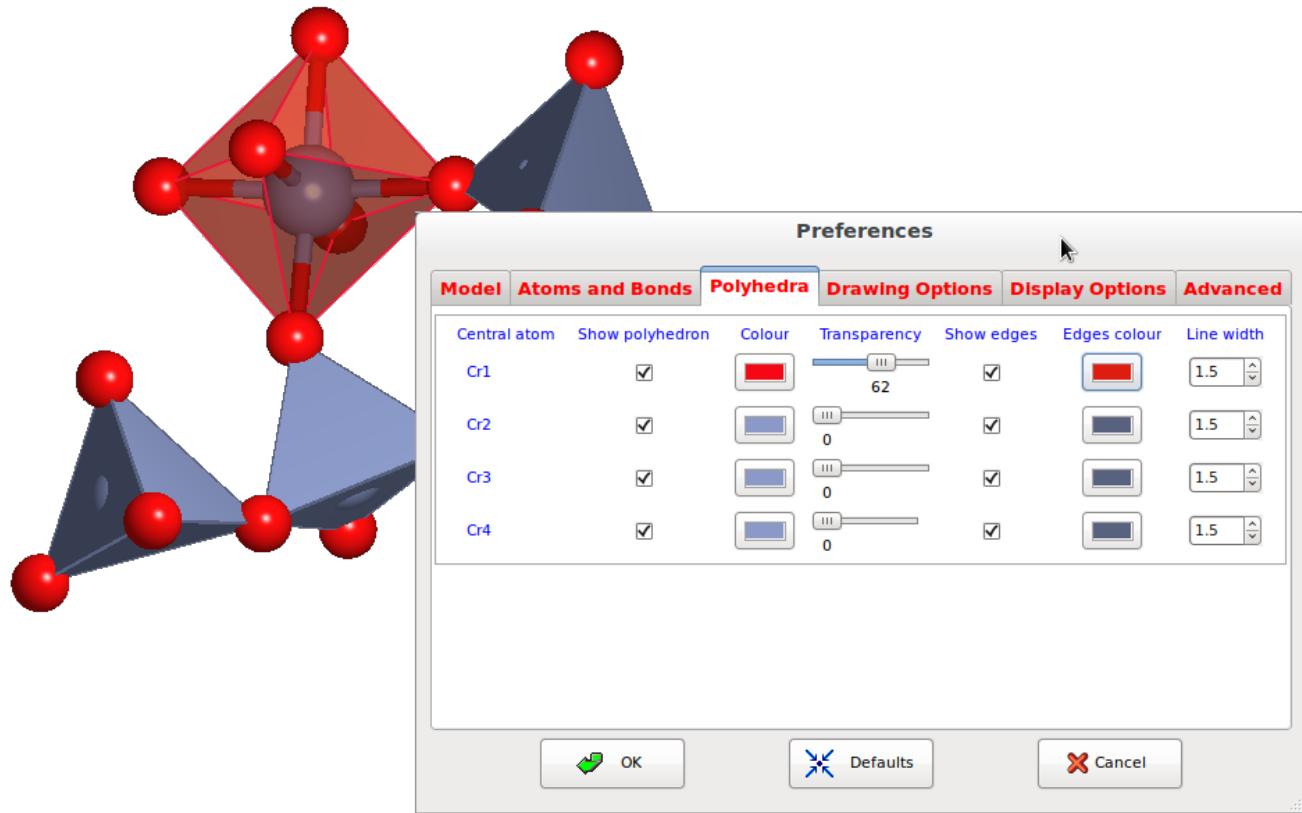
View plane

Set the preferred view plane.

Tools
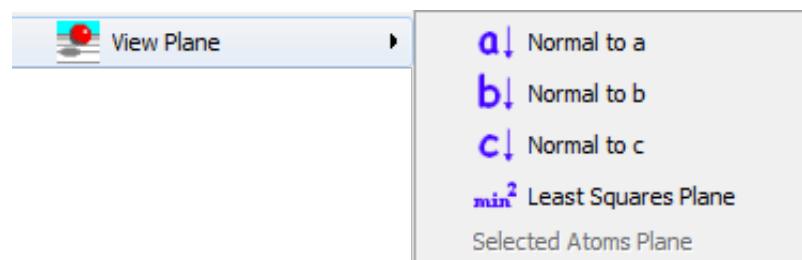
http://www.ba.ic.cnr.it/softwareic/expo/tools/

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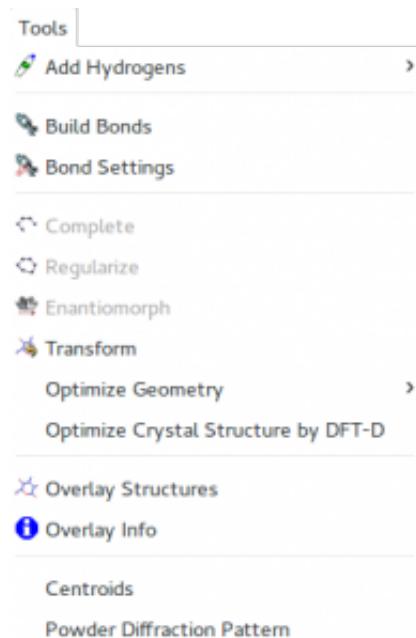
It allows:

Geometry optimization

Bond valence calculation

_____

Geometry Optimization

http://www.ba.ic.cnr.it/softwareic/expo/geometry-optimization/

<div style="background-color:cyan; text-align:center;">**Geometry Optimization**</div>

Quantum-chemical calculations have been successfully used to complement the experimental X-ray powder diffraction (XRPD) data at several stages of the structure solution process:

- To optimize the molecular geometry in the gas phase to obtain accurate starting molecular structure suitable for structure solution by real space methods. In this stage different levels of theory are applied: molecular mechanics, semi-empirical methods, Hartree-Fock methods, density functional theory (DFT).
- To minimize the energy of crystal structure and provide reasonable bond lengths and angles to input into the Rietveld refinement as chemical restraints. Calculations are performed using plane wave DFT with dispersion correction (DFT-D).
- Because it is difficult to accurately locate the positions of H atoms from the XRPD data, energy optimization provides the most reasonable approach to compute optimal positions for the H atoms as the final step, with unit cell and non-H atomic positions fixed to those established by Rietveld refinement.
- To clarify an ambiguity concerning the orientation of functional group that could not be distinguished on the basis of the XRPD data alone.
- Geometry optimization with DFT-D approach in the solid state has been applied to refine the crystal structure when Rietveld refinement yielded inaccurate molecular geometry, providing results whose accuracy is comparable to that of single-crystal refinement.
- To assess the correctness of experimental organic crystal structures

EXPO2014 provides crystallographers with tools to perform theoretical calculations for the geometry optimization. The first geometry optimization should be done with a faster level of theory, such as molecular mechanics or a semiempirical method. Once a geometry close to the correct geometry has been obtained with this lower level of theory, it is used as the starting geometry for a second optimization at
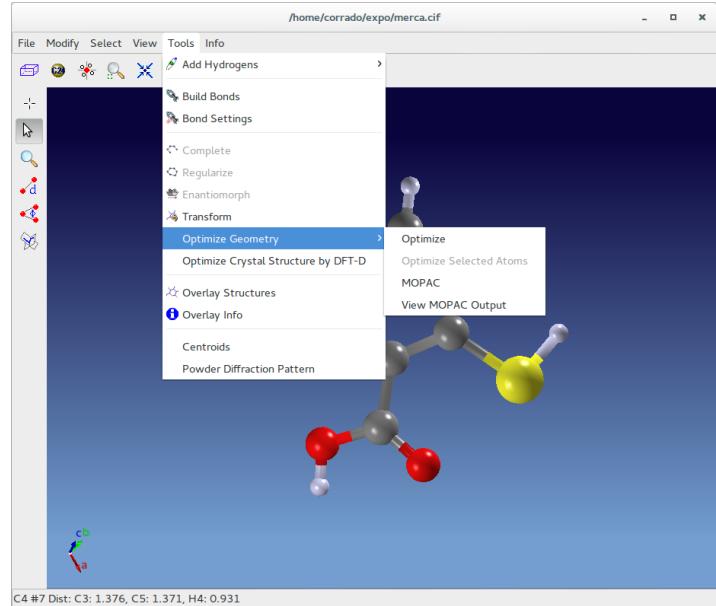
the final, more accurate level of theory (*e.g.*, DFT).

Geometry optimization by molecular-mechanics force fields

In molecular mechanics (MM), the energy of a compound consists of the sum of simple classical equations and the molecule is described as a collection of balls (corresponding to the atoms) held together by springs (corresponding to the bonds). The molecular mechanics model clearly does not use wave function and electrons are not explicitly included.

The total potential energy is typically taken to be the sum of the bond stretching energy E_{str}, the bending energy E_{bend} , the twisting (or torsion) energy E_{tor}, and the energy of interaction between non-bonded atoms $E_{non-bond}$. The last contribution includes van der Waals, steric, and electrostatic interactions between atoms not chemically bound.

$$E_{tot} = E_{str} + E_{bend} + E_{tor} + E_{non-bond}$$

The equations for the potential energy terms contain parameters. The specified mathematical form of these equations and the parameters in them constitute a particular *forcefield,* and MM methods are sometimes called forcefield methods. The constants in this equation are obtained either from spectroscopic data or *ab initio* calculations. Many good force fields have already been developed and they differ in the number of terms in the energy expression, the complexity of those terms, and the way in which the constants were obtained.

To optimize the geometry and find the more stable conformation, each of the forces f_i acting on a nucleus by electrons and other nuclei must vanish:

$f_i = - ?E_{tot}/?q_i$

where q_i are the nuclear coordinates of i-th atom.

In EXPO2014 two molecular mechanics force fields are used:

- The Merck molecular force field (MMFF) can be used for organic molecules and biomolecules [1]
- The Universal force field (UFF) is a full periodic table force field. This is widely used for system containing inorganic elements [2]

In EXPO2014 the support for molecular mechanics is provided using the class OBForceField from the library Open Babel [3]

Clicking on **Tools > Optimize Geometry > Optimize** MMFF94 force field will be applied and in case of failure (e.g. inorganic elements) UFF will be executed. Use **Modify> Undo Optimize Geometry** to restore the initial molecular geometry.

The geometry optimization can also be applied only on selected atoms:

- click on ⊹ to activate the selection mode,
- select by mouse atoms to apply the geometry optimization,
- click on **Tools > Optimize Geometry > Optimize Selected Atoms**.

E.g., you are interested to optimize only the position of H atoms, fixing the position of all other atoms: click on

Select > Select Type > H atoms and then **Tools > Optimize Geometry > Optimize Selected Atoms**.

The most frequent use of MM is undoubtedly to obtain reasonable structures for other kinds of calculations: semi-empirical (SE), *ab initio* or density functional theory (DFT). The two salient features of MM calculations on small to medium-sized molecules is that they are fast and they can be very accurate. MM calculations might still take only seconds. MM, even when it does not provide very accurate geometries, can supply reasonably good input geometries for SE, ab initio or DFT, and this is one of its main applications.

Geometry optimization by MOPAC

There are two general approaches to solving the Schrödinger equation of a molecular system: semi-empirical and *ab initio* methods. The semi-empirical methods assume an approximate Hamiltonian operator and the calculations are further simplified by approximating integrals with various experimental data such as ionization energies, electronic spectral transition energies, and bond energies. On the other hand, *ab initio* methods use a "correct" Hamiltonian operator, which includes kinetic energy of the electrons, attractions between electrons and nuclei, and repulsions between electrons and those between nuclei, to calculate all integrals without making use of any experimental data other than the values of the fundamental constants. An example of these methods is the self-consistent field (SCF) method first introduced by D. R. Hartree and V. Fock in the 1920s.

MOPAC (Molecular Orbital PACkage) is one of the most widely used semi-empirical software packages and is designed for a wide range of functionality. We tested version 2012 and 2016. MOPAC is not distributed with EXPO, so a copy of the software needs to be separately downloaded. For more information or to obtain a copy, please visit http://www.openmopac.net/MOPAC2016.html, the software is currently free for academic use.

An interface of MOPAC is provided in EXPO and allows users to perform a gas-phase geometry optimization. This approach is useful to generate molecular model with accurate angles and bond lengths, suitable as a starting model for structure solution by real space methods. The advantage of semi-empirical calculations is that they are much faster than SCF or DFT methods and can be applied routinely to large system.

A MOPAC optimization is performed selecting **Tools > Optimize Geometry > MOPAC** from the principal menu.

The calculation is performed by choosing the following options in the MOPAC dialogue window:

Title: define the title of the calculations, *e.g.*, the name of molecule. It is reported in the second line of the MOPAC input file

Method: Different semi-empirical Hamiltonians are available and are used in the electronic part of the calculation to obtain molecular orbitals and its derivative with respect to molecular geometry. Not all chemical elements are supported by the selected Hamiltonian. For more information about the elements available for the specified method see the geometry section in MOPAC manual: http://openmopac.net/manual/

Charge: Write the charge of studied system

Multiplicity: Allow the user to specify the number of singly occupied orbitals. Singlet is the default, and specify a closed shell that includes the most of organic molecules; doublet specify one singly occupied orbital (e.g. organic radicals); triplet specifies two singly occupied orbitals; and so forth. If the multiplicity is unknown, you have to try all.

Format: Geometries can be specified within MOPAC using the Cartesian coordinate definition or the Gaussian Z-matrix format. The position of molecule in the cell unit will be lost if the geometry is defined in Z-matrix format.

Mopac Program: EXPO2014 automatically locates the position where MOPAC exists, if MOPAC is installed in a standard directory (C:/Program Files/MOPAC/ on Windows, /opt/mopac/ on Linux) or the environmental variable PATH includes the installation directory, otherwise the MOPAC executable must be specified directly by the user.

If you click on **Submit**, MOPAC will be launched and at the end of the geometry optimization, the new geometry will be displayed and an output file, called structure_name_mopac.out, will be produced in the working directory. The output file can be viewed from **Tools> Optimize Geometry> View Mopac**

Output. You can read previously generated MOPAC output files and display the molecular geometry from **File > Import Structure** and selecting 'MOPAC Output File (`*.out`)' in the file extension menu.

Before performing the calculation, add the missing hydrogen. You can use the option **Tools>Add Hydrogens**. Wrong starting geometry, missing hydrogens or wrong multiplicity are possible causes of an incorrect result. In this situation use **Modify> Undo Optimize Geometry** to restore the initial geometry, then read the output file to find a possible cause of error.

Additional calculations can be performed by clicking the preview button and changing the input file. For example, it might be particularly important to calculate the bond order: add the keyword **BONDS** in the first line and bond order between all pairs of atoms will be printed in the output file. The keyword **NOOPT** can be used to perform only an energy calculation; the calculated energy will be written in the output file. Use the keyword `NOOPT OPT-H` to optimize the positions of hydrogen atoms while freezing the positions of all other atoms. A complete list of keywords used in MOPAC2016 is available at link http://openmopac.net/manual/allkeys.html

Geometry optimization by density functional theory (DFT)

Another approach closely related to the ab initio methods that has gained increasing prominence in recent years is the DFT. This method bypasses the determination of the wavefunction ?. Instead, it determines the molecular electronic probability density ? directly and then calculates the energy of the system from ?.

EXPO2014 creates input files for well known quantum chemistry programs, clicking on **File > Export Structure**. Input files for GAMESS [4] and NWCHEM [5] contain keywords for geometry optimization by density functional theory (DFT) calculations to perform using the B3LYP functional and a standard 6-31G* basis set.

In addition **Tools > Optimize Crystal Structure by DFT-D** can create an NWCHEM input file for the optimization of crystal structure by plane wave density function theory with Grimme dispersion correction. In order to obtain a good accuracy on the interatomic distances, you should do a convergence test with respect the cutoff energy and a convergence study associated with the sampling of the Brillouin zone.

At the end of the DFT-optimization, the output file generated by the quantum chemistry software can be imported in EXPO2014 and the calculated crystal structure can be visually compared and superimposed with the experimental one by using **Tools > Overlay Structures** to generate an overlaid.
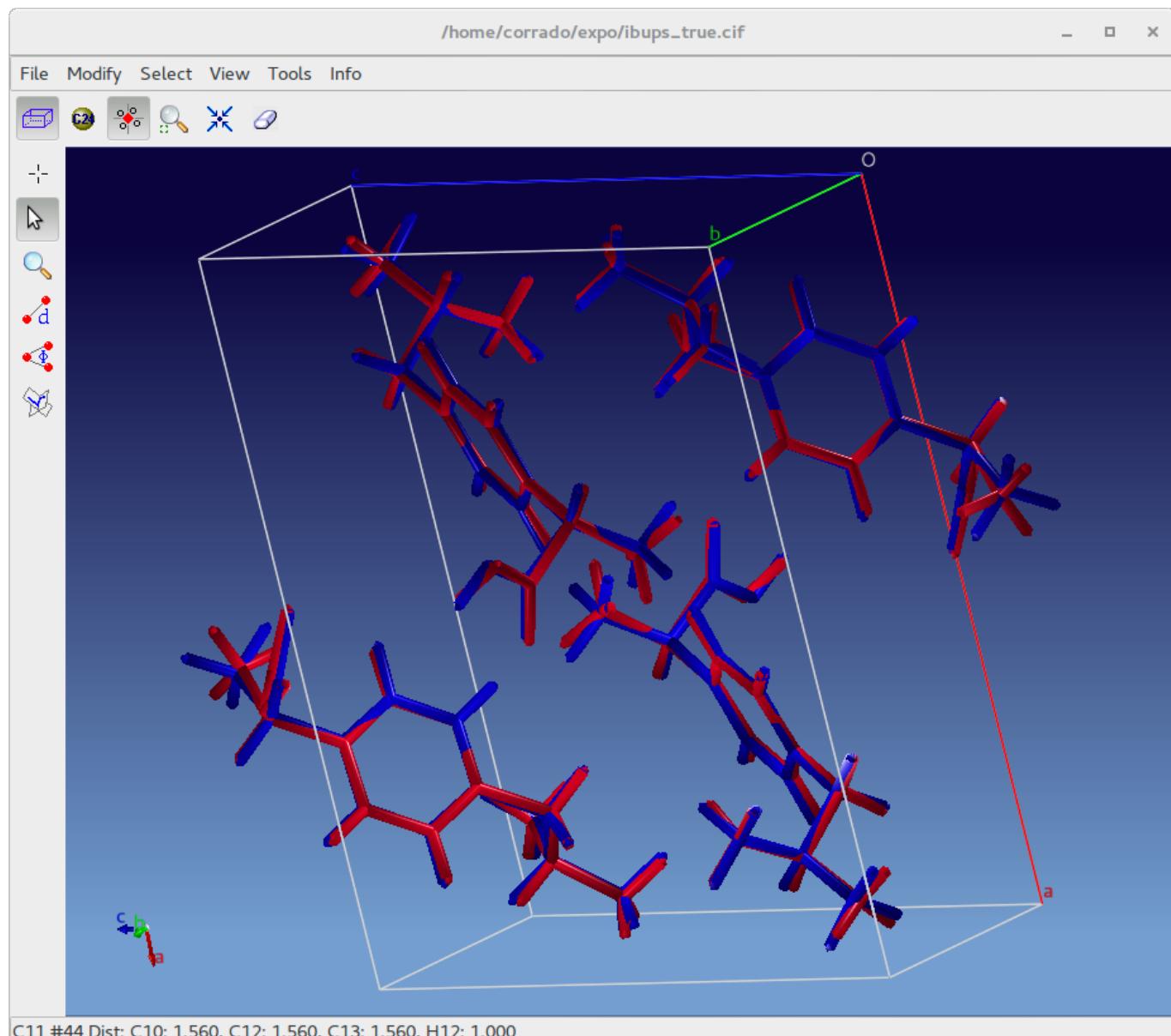
Example of overlaid between experimental (red) and optimized (blue) crystal structure of ibuprofen. The optimization was performed by using DFT-D3 in NWChem.

The quality of the agreement between the experimental and the theoretical structure can be evaluated in terms of root mean square displacement (RMSD) calculated for all non-H atoms. In a recent paper 215 organic crystal structures determined from X-ray powder diffraction data and published in an IUCr journal were energy-minimized with DFT-D and compared to the single crystal structure (van de Streek & Neumann, 2014). According to this survey the value of RMSD never exceeds 0.35 Å for correct crystal structure. RMSD information is also available in EXPO2014 by **Tools > Overlay info**.

References

[1] Thomas A. Halgren, J. Comput. Chem., 17, 490-519 (1996)

[2] J. Am. Chem. Soc. 1992, Vol. 114, No. 25, 10024-10035

[3] N M O'Boyle, M Banck, C A James, C Morley, T Vandermeersch, and G R Hutchison. "Open Babel: An open chemical toolbox." J. Cheminf. (2011), 3, 33.

[4] http://www.msg.ameslab.gov/gamess/

[5] http://www.nwchem-sw.org/index.php/Main_Page

van de Streek, J. and Neumann, M.A. "Validation of molecular crystal structures from powder diffraction data with dispersion-corrected density functional theory (DFT-D)" *Acta Cryst.* (2014). **B70**, 1020–1032.

———————————————————————————————

Bond Valence Sum

http://www.ba.ic.cnr.it/softwareic/expo/bond-valence-sum/

<div style="background:cyan;text-align:center;">

Introduction to Bond Valence Method

</div>

The Bond Valence Model (Brese & O'Keeffe, 1991; Brown, 2009) is an important approach to assessing the chemical plausibility of inorganic crystal structure. The model is based on the *valence sum rule* from the Pauling's electrostatic valence concept (Pauling, 1929): the sum of bond valences S_{ij} around any ion, i, is equal to its valence, V_i (the formal oxidation state)

$$V_i = \sum_j S_{ij} \tag{1}$$

where the sum goes over all neighboring atoms j of the atom i.

The bond valence S_{ij} can be considered a measure of the electrostatic flux between a cation and an anion, directly related with the strength of the bond and correlated inversely with bond length. It can be approximated by

$$S_{ij} = exp\left(\frac{R_0 - R_{ij}}{B}\right) \tag{2}$$

where R_0 and B are constant parameters and R_{ij} is the interatomic distance between atoms i and j. For many bonds B is close to 0.37. The bond valence parameters (R_0 and B) used by the program were taken from a list compiled by I. D. Brown from various references and stored in a CIF file, `bvparm2016.cif`, available for download from https://www.iucr.org/resources/data/datasets/bond-valence-parameters. It should be noted, that bond valence parameters are not available for all pairs of atoms.

The most common application of the bond valence model is for checking and validating new crystal structures by comparing oxidation states (formal ionic charges) V_i^0 with bond valence sums V_i calculated from the equations (1) and (2). The agreement is generally good if the structure is correct, but when the agreement is poor and large deviations are observed the structure determination might be wrong (*e.g.,* incorrect symmetry or missing atoms) or inappropriate bond valence parameters were used.

Bond valence model is often used to explain the structural distortions from the ideal three-dimensional structure. These distortions are usually measured by using the global instability index, G_{ii}, which is the root-mean-square value over the deviations $V_i - V_i^0$ of the atomic valence of all atoms from the bond valence sum:

$$G_{ii} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (V_i - V_i^0)^2}$$

The G_{ii} index is usually less than 0.1 v.u. in stable and well determined structures. Values between 0.1 and 0,2 v.u. can be associated to strained structures. Typically, room-temperature structures with G_{ii} values greater than

0,2 v.u. are rare.

Bond valence sums are incorporated as restraints in the structure solution by <u>real space methods</u>.

<div style="background:cyan">

Check the bond valence sums

</div>

To check the bond valence sums (BVS) of your crystal structure choose **Bond Valence Sum** in the **Tools** menu. In the case of the phosphate $AgTbP_2O_7$ the test produce the following result.
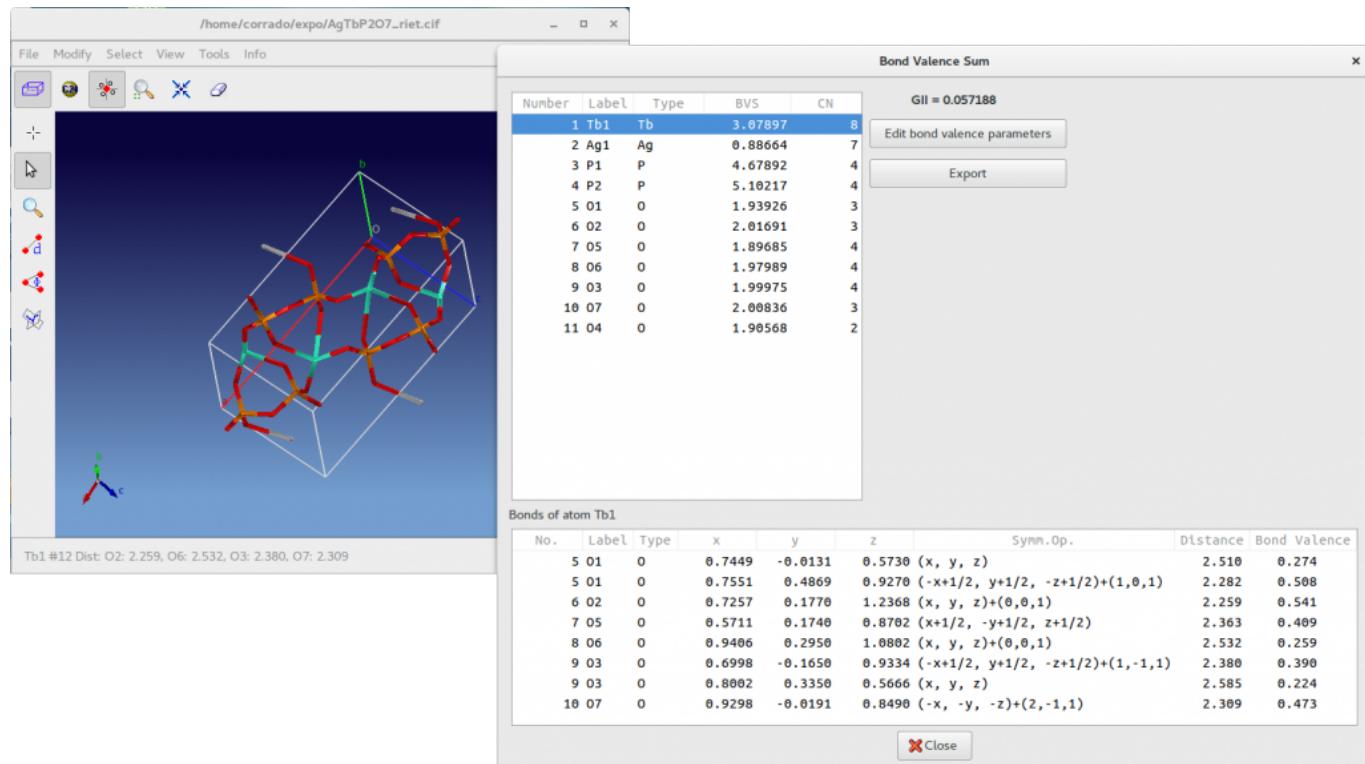
For each atom the BVS as well as the coordination numbers (CN) are reported in the above list box. The below list box contains the bond lengths and bond valences of all bonds that contribute to BVS of the selected atom. For example, in the above picture for the selected atom `Tb1` you can see all the bond `Tb-O` contributing to the bond valence sum 3.08. In addition, fractional coordinates and the symmetry operator applied to the atom in the asymmetric unit are reported. The G_{ii} is displayed and all the information accessible by this window can be exported in a text file pressing the button **Export**. The button **Edit bond valence parameters** enables the access to the bond valence parameters (R_0 and B) corresponding to each pair of atom types.

Edit bond valence parameters

Atom1	Ox.St.1	Atom2	Ox.St.2	Ro	B	dmin	dmax
O	-2	O	-2	1.406	0.370	1.000	0.000
P	5	O	-2	1.617	0.370	1.007	2.231
P	5	P	5	2.220	0.350	1.000	0.000
Ag	1	O	-2	1.842	0.370	1.232	3.052
Ag	9	P	-3	2.220	0.370	1.474	0.000
Ag	0	Ag	0	0.000	0.000	1.000	0.000
Tb	3	O	-2	2.032	0.370	1.422	2.835
Tb	3	P	-3	2.590	0.370	1.844	0.000
Ag	0	Tb	0	0.000	0.000	1.000	0.000
Tb	0	Tb	0	0.000	0.000	1.000	0.000

OK   Defaults   Cancel

You will ensure the all information specified in this window are reasonable and if necessary modify the entries that are not valid.

For each pair the program search in the database the corresponding bond valence parameters and assign a bond windows, *i.e.* the range of distances between any two atoms that are to be considered as bonds. These range is defined in terms of d_{min} and d_{max} following some of the rules described by Altermatt and Brown (Altermatt & Brown, 1985).

It is assumed that the bonding occurs only between atoms of different electronegativity. This is the reason why bonds like P-P or O-O are excluded assigning value of $d_{min} = 1$ and $d_{max} = 0$.

If the bond valence parameters corresponding to a pair is not available in the database, the parameters are set to $R_0 = 0$, $B = 0$, $d_{min} = 1$ and $d_{max} = 0$ generating a contribution equal 0 to the BVS.

When the oxidation state is non specified by the user the program try to assign the most probable.

References

Altermatt, D & Brown, I.D. (1985). Acta Cryst. B41, 240-244.

Brese, N. E. & O'Keeffe, M. (1991). *Acta Cryst.* B**47**, 192–197.

Brown, I. D. (2009). *Chem. Rev.* **109**, 6858–6919.

Pauling, L. (1929). *J. Am. Chem. Soc.*, **51**,1010.

Info

http://www.ba.ic.cnr.it/softwareic/expo/info/

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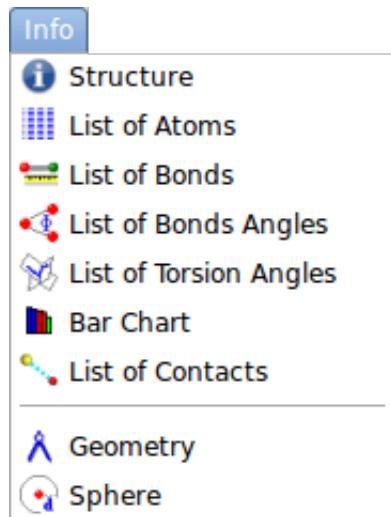
It allows:

Structure	Info on the structure.
List of Atoms	List of Atoms.
List of Bonds	List of Bond Distances.
List of Bond Angles	List of Bond Angles.
List of Torsion Angles	List of Torsion Angles.
Bar Chart	Histogram of the peak intensities.
List of Contacts	
Geometry	Some geometric calculations can be performed.
Sphere	list of atoms around a selected atom

Structure

In the following tabs detailed information of the model are reported.

cimetidine

| Structure | Atoms | Bonds | Angles | Torsion Angles | Bar Chart | Contacts |

```
Name                cimetidine
Cell Lengths        a       10.699  b       18.818  c        6.825
Cell Angles         alpha   90.000  beta  111.284  gamma   90.000
Space Group         P 21/n              (centrosymmetric)
Symmetry operations
                    1) x,y,z
                    2) 1/2-x,1/2+y,1/2-z
                    3) -x,-y,-z
                    4) 1/2+x,1/2-y,1/2+z
Cell Volume         1280.230
Cell Content        H       64
                    C       40
                    N       24
                    S        4
Volume per atom     18.827
R-Factor            28.393
Atoms in model      17
```

Save   Close

List of Atoms

The second tab contains the information related to the atoms.

cimetidine

| Structure | Atoms | Bonds | Angles | Torsion Angles | Bar Chart | Contacts |

Number	Label	Type	Intensity	Xfrac	Yfrac	Zfrac	B	s.o.f.	Symm.Op
1	S1	S	16719.00000	0.48185	0.41469	0.19725	4.51421	1.00000	x,y,z
2	N1	N	6562.00000	0.62974	0.23371	0.72680	3.36343	1.00000	x,y,z
3	N2	N	6279.00000	0.95063	0.37542	0.54685	3.40571	1.00000	x,y,z
4	N3	N	6153.00000	0.87073	0.44743	0.24875	4.44055	1.00000	x,y,z
5	N4	N	5543.00000	1.09869	0.40861	0.39000	3.95594	1.00000	x,y,z
6	N5	N	5413.00000	1.16586	0.48553	0.15358	5.43250	1.00000	x,y,z
7	N6	N	5401.00000	0.71092	0.31939	0.58086	4.04593	1.00000	x,y,z
8	C1	C	5332.00000	0.63725	0.39924	0.15331	4.48181	1.00000	x,y,z
9	C2	C	5078.00000	0.53363	0.28805	0.66299	4.04910	1.00000	x,y,z
10	C3	C	4986.00000	0.40380	0.28380	0.70951	4.72578	1.00000	x,y,z
11	C4	C	4944.00000	0.73520	0.25285	0.67636	4.39471	1.00000	x,y,z
12	C5	C	4923.00000	0.97095	0.40969	0.39064	2.77809	1.00000	x,y,z
13	C6	C	4871.00000	1.12231	0.45105	0.25634	3.85322	1.00000	x,y,z
14	C7	C	4615.00000	0.58344	0.33910	0.57113	5.01616	1.00000	x,y,z
15	C8	C	4447.00000	0.73739	0.46058	0.26631	3.08464	1.00000	x,y,z
16	C9	C	4369.00000	0.52202	0.41038	0.47842	4.07738	1.00000	x,y,z
17	C10	C	3997.00000	1.06300	0.34428	0.72908	4.82915	1.00000	x,y,z

Save   Close

List of Bonds

The third tab contains the list of the bond distances.

cimetidine

| Structure | Atoms | Bonds | Angles | Torsion Angles | Bar Chart | Contacts |

Number	Atom 1	Atom 2	Distance	Symm.Op1	Symm.Op2
1	S1	C1	1.81675	x,y,z	x,y,z
2	S1	C9	1.80959	x,y,z	x,y,z
3	N1	C2	1.40276	x,y,z	x,y,z
4	N1	C4	1.34281	x,y,z	x,y,z
5	N2	C5	1.32952	x,y,z	x,y,z
6	N2	C10	1.50002	x,y,z	x,y,z
7	N3	C5	1.35573	x,y,z	x,y,z
8	N3	C8	1.49493	x,y,z	x,y,z
9	N4	C5	1.36840	x,y,z	x,y,z
10	N4	C6	1.30425	x,y,z	x,y,z
11	N5	C6	1.16922	x,y,z	x,y,z
12	N6	C4	1.39193	x,y,z	x,y,z
13	N6	C7	1.39146	x,y,z	x,y,z
14	C1	C8	1.57330	x,y,z	x,y,z
15	C2	C3	1.53509	x,y,z	x,y,z
16	C2	C7	1.35730	x,y,z	x,y,z
17	C7	C9	1.52636	x,y,z	x,y,z

Save   Close

List of Bond Angles

The fourth tab contains the list of the bond angles.

cimetidine

| Structure | Atoms | Bonds | Angles | Torsion Angles | Bar Chart | Contacts |

Number	Atom 1	Atom 2	Atom 3	Angle
1	C1	S1	C9	106.77
2	S1	C1	C8	107.63
3	S1	C9	C7	112.31
4	C2	N1	C4	108.32
5	N1	C2	C3	120.94
6	N1	C2	C7	107.03
7	N1	C4	N6	109.27
8	C5	N2	C10	122.63
9	N2	C5	N3	120.64
10	N2	C5	N4	116.71
11	C5	N3	C8	124.04
12	N3	C5	N4	122.50
13	N3	C8	C1	109.82
14	C5	N4	C6	116.13
15	N4	C6	N5	168.47
16	C4	N6	C7	105.97
17	N6	C7	C2	109.37
18	N6	C7	C9	121.40
19	C2	C2	C7	121.07

Save   Close

List of Torsion Angles

The fifth tab contains the list of the torsion angles.

cimetidine

Structure	Atoms	Bonds	Angles	Torsion Angles	Bar Chart	Contacts

Number	Atom 1	Atom 2	Atom 3	Atom4	Angle
1	C9	S1	C1	C8	-61.77
2	C1	S1	C9	C7	-58.42
3	C4	N1	C2	C3	-176.71
4	C4	N1	C2	C7	0.91
5	C2	N1	C4	N6	0.33
6	C10	N2	C5	N3	-168.53
7	C10	N2	C5	N4	7.24
8	C8	N3	C5	N2	6.88
9	C8	N3	C5	N4	-168.63
10	C5	N3	C8	C1	-89.79
11	C6	N4	C5	N2	-169.82
12	C6	N4	C5	N3	5.86
13	C5	N4	C6	N5	172.94
14	C7	N6	C4	N1	-1.41
15	C4	N6	C7	C2	1.99
16	C4	N6	C7	C9	179.00
17	S1	C1	C8	N3	175.76
18	N1	C2	C7	N6	-1.80
19	N1	C2	C7	C9	178.51

[Save]   [Close]

Bar Chart

The tab Bar Chart contains the histograms of the fourier intensities of the atoms.
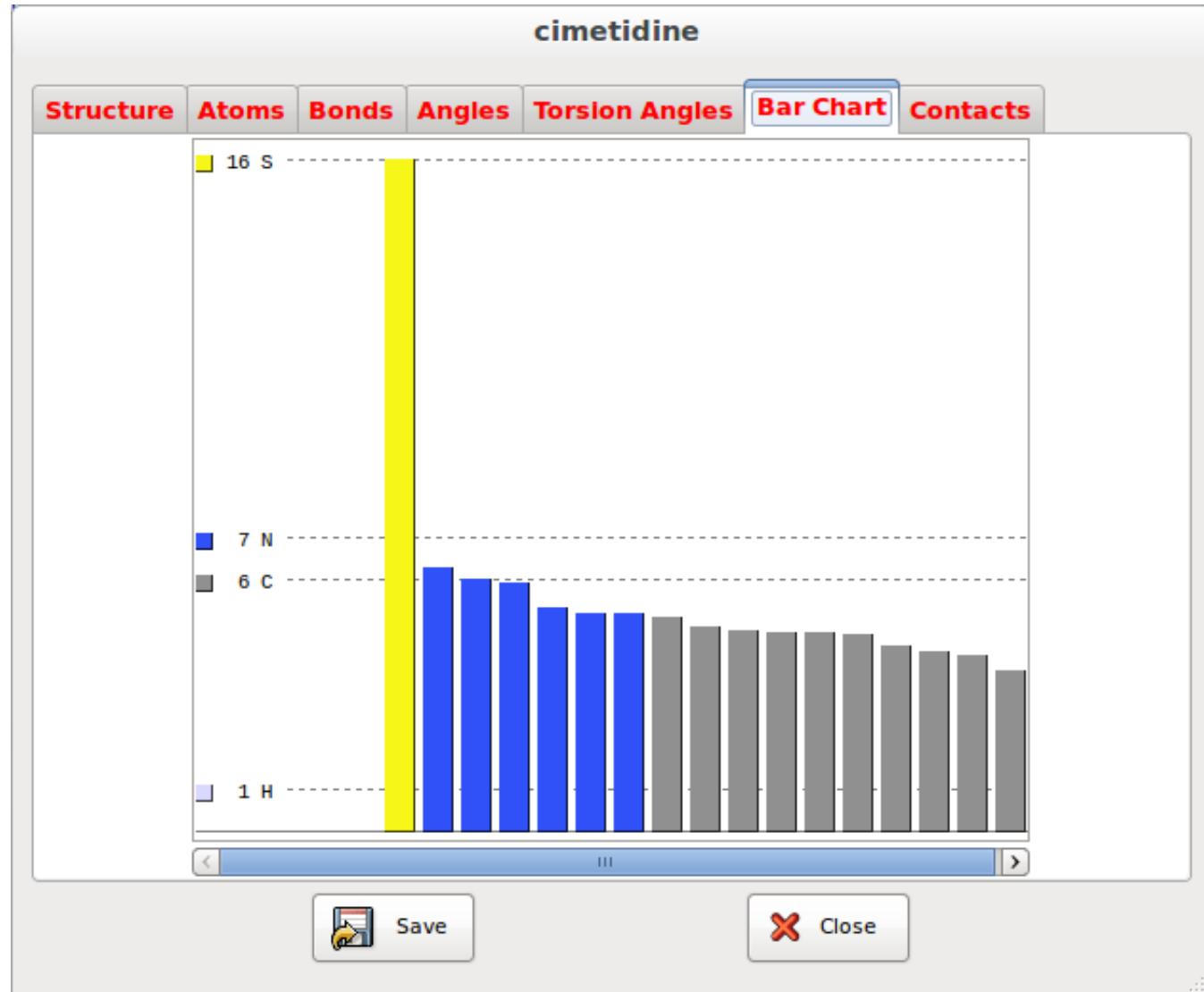
List of Contacts

The tab Contacts contains a list of non bonded contacts.

cimetidine

Number	Atom 1	Atom 2	Distance	Symm.Op1	Symm.Op2
1	N4	N1	2.96500 x,y,z	1/2+x,1/2-y,1/2+z
2	N5	N3	2.91607 x,y,z	-x,-y,-z
3	N1	N4	2.96500 x,y,z	1/2+x,1/2-y,1/2+z
4	N3	N5	2.91607 x,y,z	-x,-y,-z
5	N2	N6	2.85911 x,y,z	x,y,z

Save   Close

Geometry

The value of the distances, angles and torsion angles among the atoms can be calculated.

Geometric calculations

○ Distance   ⦿ Angle   ○ Tors.Angle

Results

Angle (C3, C8, C5) = 110.297 °

Close

_____

Appendix A: Average volume of the atoms

http://www.ba.ic.cnr.it/softwareic/expo/average-volume-of-the-atoms/

The average volume of elements at 298 K (Acta Cryst. (2002). **B57**, 489-493).

Atom number	Element	Average volume (Å^3)
1	H	5.08
2	He	10.00
3	Li	22.60
4	Be	36.00
5	B	13.24
6	C	13.87
7	N	11.80
8	O	11.39
9	F	11.17
10	Ne	20.00
11	Na	26.00
12	Mg	36.00
13	Al	39.60
14	Si	37.30
15	P	29.50
16	S	25.20
17	Cl	25.80
18	Ar	30.00
19	K	36.00
20	Ca	45.00
21	Sc	42.00
22	Ti	27.30
23	V	24.00
24	Cr	28.10
25	Mn	31.90
26	Fe	30.40
27	Co	29.40
28	Ni	26.00
29	Cu	26.90
30	Zn	39.00
31	Ga	37.80
32	Ge	41.60
33	As	36.40
34	Se	30.30
35	Br	32.70
36	Kr	40.00
37	Rb	42.00
38	Sr	47.00
39	Y	44.00
40	Zr	27.00
41	Nb	37.00
42	Mo	38.00

43	Tc	38.00
44	Ru	37.30
45	Rh	31.20
46	Pd	35.00
47	Ag	35.00
48	Cd	51.00
49	In	55.00
50	Sn	52.80
51	Sb	48.00
52	Te	46.70
53	I	46.20
54	Xe	45.00
55	Cs	46.00
56	Ba	66.00
57	La	58.00
58	Ce	54.00
59	Pr	57.00
60	Nd	50.00
61	Pm	55.00
62	Sm	50.00
63	Eu	53.00
64	Gd	56.00
65	Tb	45.00
66	Dy	50.00
67	Ho	42.00
68	Er	54.00
69	Tm	49.00
70	Yb	59.00
71	Lu	35.00
72	Hf	40.00
73	Ta	43.00
74	W	38.80
75	Re	42.70
76	Os	41.90
77	Ir	34.30
78	Pt	38.00
79	Au	43.00
80	Hg	38.00
81	Tl	54.00
82	Pb	52.00
83	Bi	60.00
84	Po	50.00
85	At	55.00
86	Rn	60.00
87	Fr	70.00
88	Ra	60.00
89	Ac	74.00
90	Th	56.00
91	Pa	60.00
92	U	58.00

93	Np	45.00
94	Pu	70.00
95	Am	17.00
96	Cm	70.00
97	Bk	70.00
98	Cf	70.00
99	Es	70.00
100	Fm	70.00
101	Md	70.00
102	No	70.00
103	Lr	70.00
104	Rf	70.00
105	Db	70.00
106	Sg	70.00
107	Bh	70.00
108	Hs	70.00
109	Mt	70.00
110	Ds	70.00
111	Rg	70.00

_____

Appendix B: Extinction symbols and corresponding space groups

http://www.ba.ic.cnr.it/softwareic/expo/extinction_symbols/

@page { }
table { border-collapse:collapse; border-spacing:0; empty-cells:show }
td, th { vertical-align:top; font-size:12pt;}
h1, h2, h3, h4, h5, h6 { clear:both;}
ol, ul { margin:0; padding:0;}
li { list-style: none; margin:0; padding:0;}
/* "li span.odfLiEnd" - IE 7 issue*/
li span. { clear: both; line-height:0; width:0; height:0; margin:0; padding:0; }
span.footnodeNumber { padding-right:1em; }
span.annotation_style_by_filter { font-size:95%; font-family:Arial; background-color:#fff000; margin:0; border:0; padding:0; }
span.heading_numbering { margin-right: 0.8rem; }* { margin:0;}
.P1 { font-size:10pt; margin-bottom:0in; margin-top:0in; font-family:Liberation Mono; writing-mode:horizontal-tb; direction:ltr;font-weight:bold; }
.Preformatted_20_Text { font-size:10pt; font-family:Liberation Mono; writing-mode:horizontal-tb; direction:ltr;margin-top:0in; margin-bottom:0in; }
.Table1 { width:6.6931in; float:none; }
.Table1_A { width:2.2313in; }
/* ODF styles with no properties representable as CSS */
{ }

#	Extinction Symbol	Space Groups
Monoclinic		
1	P – 1 1	P211(3), Pm11(6), P2/m11(10
2	P 21 1 1	P2111(4), P21/m11(11)
3	P b 1 1	Pb11(7), P2/b11(13)
4	P 21/b 1 1	P21/b11(14)
5	P c 1 1	Pc11(7), P2/c11(13)
6	P 21/c 1 1	P21/c11(14)
7	P n 1 1	Pn11(7), P2/n11(13)
8	P 21/n 1 1	P21/n11(14)
9	C – 1 1	C211(5), Cm11(8), C2/m11(1
10	C n 1 1	Cn11(9), C2/n11(15)
11	B – 1 1	B211(5), Bm11(8), B2/m11(1
12	B b 1 1	Bb11(9), B2/b11(15)
13	I – 1 1	I211(5), Im11(8), I2/m11(12)
14	I c 1 1	Ic11(9), I2/c11(15)
15	P 1 – 1	P2(3), Pm(6), P2/m(10)
16	P 1 21 1	P21(4), P21/m(11)
17	P 1 a 1	Pa(7), P2/a(13)
18	P 1 21/a 1	P21/a(14)

19	P 1 c 1	Pc(7), P2/c(13)
20	P 1 21/c 1	P21/c(14)
21	P 1 n 1	Pn(7), P2/n(13)
22	P 1 21/n 1	P21/n(14)
23	C 1 – 1	C2(5), Cm(8), C2/m(12)
24	C 1 c 1	Cc(9), C2/c(15)
25	A 1 – 1	A2(5), Am(8), A2/m(12)
26	A 1 n 1	An(9), A2/n(15)
27	I 1 – 1	I2(5), Im(8), I2/m(12)
28	I 1 a 1	Ia(9), I2/a(15)
29	P 1 1 –	P112(3), P11m(6), P112/m(10
30	P 1 1 21	P1121(4), P1121/m(11)
31	P 1 1 a	P11a(7), P112/a(13)
32	P 1 1 21/a	P1121/a(14)
33	P 1 1 b	P11b(7), P112/b(13)
34	P 1 1 21/b	P1121/b(14)
35	P 1 1 n	P11n(7), P112/n(13)
36	P 1 1 21/n	P1121/n(14)
37	B 1 1 –	B112(5), B11m(8), B112/m(1
38	B 1 1 n	B11n(9), B112/n(15)
39	A 1 1 –	A112(5), A11m(8), A112/m(1
40	A 1 1 a	A11a(9), A112/a(15)
41	I 1 1 –	I112(5), I11m(8), I112/m(12)
42	I 1 1 b	I11b(9), I112/b(15)

Orthorhombic

43	P – – –	P222(16), Pmm2(25), Pmmm P2mm(25)
44	P – – 21	P2221(17)
45	P – 21 –	P2212(17)
46	P – 21 21	P22121(18)
47	P 21 – –	P2122(17)
48	P 21 – 21	P21221(18)
49	P 21 21 –	P21212(18)
50	P 21 21 21	P212121(19)
51	P – – a	Pm2a(28), P21ma(26), Pmma
52	P – – b	Pm21b(26), P2mb(28), Pmmb
53	P – – n	Pm21n(31), P21mn(31), Pmm
54	P – a –	Pma2(28), Pmam(51), P21am
55	P – a a	P2aa(27), Pmaa(49)
56	P – a b	P21ab(29), Pmab(57)
57	P – a n	P2an(30), Pman(53)
58	P – c –	Pmc21(26), P2cm(28), Pmcm
59	P – c a	P21ca(29), Pmca(57)
60	P – c b	P2cb(32), Pmcb(55)
61	P – c n	P21cn(33), Pmcn(62)
62	P – n –	Pmn21(31), P21nm(31), Pmn
63	P – n a	P2na(30), Pmna(53)
64	P – n b	P21nb(33), Pmnb(62)
65	P – n n	P2nn(34), Pmnn(58)
66	P b – –	Pbm2(28), Pb21m(26), Pbmm

67	P b – a	Pb21a(29), Pbma(57)
68	P b – b	Pb2b(27), Pbmb(49)
69	P b – n	Pb2n(30), Pbmn(53)
70	P b a –	Pba2(32), Pbam(55)
71	P b a a	Pbaa(54)
72	P b a b	Pbab(54)
73	P b a n	Pban(50)
74	P b c –	Pbc21(29), Pbcm(57)
75	P b c a	Pbca(61)
76	P b c b	Pbcb(54)
77	P b c n	Pbcn(60)
78	P b n –	Pbn21(33), Pbnm(62)
79	P b n a	Pbna(60)
80	P b n b	Pbnb(56)
81	P b n n	Pbnn(52)
82	P c – –	Pcm21(26), Pc2m(28), Pcmm
83	P c – a	Pc2a(32), Pcma(55)
84	P c – b	Pc21b(29), Pcmb(57)
85	P c – n	Pc21n(33), Pcmn(62)
86	P c a –	Pca21(29), Pcam(57)
87	P c a a	Pcaa(54)
88	P c a b	Pcab(61)
89	P c a n	Pcan(60)
90	P c c –	Pcc2(27), Pccm(49)
91	P c c a	Pcca(54)
92	P c c b	Pccb(54)
93	P c c n	Pccn(56)
94	P c n –	Pcn2(30), Pcnm(53)
95	P c n a	Pcna(50)
96	P c n b	Pcnb(60)
97	P c n n	Pcnn(52)
98	P n – –	Pnm21(31), Pnmm(59), Pn21
99	P n – a	Pn21a(33), Pnma(62)
100	P n – b	Pn2b(30), Pnmb(53)
101	P n – n	Pn2n(34), Pnmn(58)
102	P n a –	Pna21(33), Pnam(62)
103	P n a a	Pnaa(56)
104	P n a b	Pnab(60)
105	P n a n	Pnan(52)
106	P n c –	Pnc2(30), Pncm(53)
107	P n c a	Pnca(60)
108	P n c b	Pncb(50)
109	P n c n	Pncn(52)
110	P n n –	Pnn2(34), Pnnm(58)
111	P n n a	Pnna(52)
112	P n n b	Pnnb(52)
113	P n n n	Pnnn(48)
114	C – – –	C222(21), Cmm2(35), Cmmn C2mm(38)
115	C – – 21	C2221(20)

116	C – – (ab)	Cm2a(39), Cmma(67), C2mb(
117	C – c –	Cmc21(36), Cmcm(63), C2cm
118	C – c (ab)	C2cb(41), Cmca(64)
119	C c – –	Ccm21(36), Ccmm(63), Cc2m
120	C c – (ab)	Cc2a(41), Ccmb(64)
121	C c c –	Ccc2(37), Cccm(66)
122	C c c (ab)	Ccca(68), Cccb(68)
123	B – – –	B222(21), Bmm2(38), Bmmm
		B2mm(38)
124	B – 21 –	B2212(20)
125	B – – b	Bm21b(36), Bmmb(63), B2m
126	B – (ac)-	Bma2(39), Bmam(67), B2cm(
127	B – (ac)b	B2cb(41), Bmab(64)
128	B b – –	Bbm2(40), Bbmm(63), Bb21r
129	B b – b	Bb2b(37), Bbmb(66)
130	B b (ac)-	Bba2(41), Bbcm(64)
131	B b (ac)b	Bbab(68), Bbcb(68)
132	A – – –	A222(21), Amm2(38), Ammr
		A2mm(35)
133	A 21 – –	A2122(20)
134	A – – a	Am2a(40), Amma(63), A21m
135	A – a –	Ama2(40), Amam(63), A21ar
136	A – a a	A2aa(37), Amaa(66)
137	A(bc)- –	Abm2(39), Abmm(67), Ac2m
138	A(bc)- a	Ac2a(41), Abma(64)
139	A(bc)a –	Aba2(41), Acam(64)
140	A(bc)a a	Abaa(68), Acaa(68)
141	I – – –	I222(23), Imm2(44), Immm(7
		Im2m(44), I2mm(44)
142	I – – (ab)	Im2a(46), Imma(74), I2mb(46
143	I – (ac)-	Ima2(46), Imam(74), I2cm(46
144	I – c b	I2cb(45), Imcb(72)
145	I(bc)- –	Ibm2(46), Ibmm(74), Ic2m(46
146	I c – a	Ic2a(45), Icma(72)
147	I b a –	Iba2(45), Ibam(72)
148	I b c a	Ibca(73)
149	F – – –	F222(22), Fmm2(42), Fmmm
		F2mm(42)
150	F – d d	F2dd(43)
151	F d – d	Fd2d(43)
152	F d d –	Fdd2(43)
153	F d d d	Fddd(70)

Tetragonal

154	P – – –	P4(75), P-4(81), P4/m(83), P4
		P4mm(99), P-42m(111), P4/m
		P-4m2(115)
155	P – 21 –	P4212(90), P-421m(113)
156	P 42 – –	P42(77), P42/m(84), P4222(9
157	P 42 21 –	P42212(94)
158	P 41 – –	P41(76), P43(78), P4122(91),

159	P 41 21 –	P41212(92), P43212(96)
160	P – – c	P42mc(105), P-42c(112), P42
161	P – 21 c	P-421c(114)
162	P – b –	P4bm(100), P-4b2(117), P4/m
163	P – b c	P42bc(106), P42/mbc(135)
164	P – c –	P42cm(101), P-4c2(116), P42
165	P – c c	P4cc(103), P4/mcc(124)
166	P – n –	P42nm(102), P-4n2(118), P42
167	P – n c	P4nc(104), P4/mnc(128)
168	P n – –	P4/n(85), P4/nmm(129)
169	P 42/n – –	P42/n(86)
170	P n – c	P42/nmc(137)
171	P n b –	P4/nbm(125)
172	P n b c	P42/nbc(133)
173	P n c –	P42/ncm(138)
174	P n c c	P4/ncc(130)
175	P n n –	P42/nnm(134)
176	P n n c	P4/nnc(126)
177	I – – –	I4(79), I-4(82), I4/m(87), I422
		I-42m(121), I4/mmm(139), I-4
178	I 41 – –	I41(80), I4122(98)
179	I – – d	I41md(109), I-42d(122)
180	I – c –	I4cm(108), I-4c2(120), I4/mc
181	I – c d	I41cd(110)
182	I 41/a – –	I41/a(88)
183	I a – d	I41/amd(141)
184	I a c d	I41/acd(142)
Trigonal/Hexagonal		
185	P – – –	P3(143), P-3(147), P321(150),
		P-3m1(164), P312(149), P31m
		P-31m(162), P6(168), P-6(174
		P622(177), P6mm(183), P-62
		P6/mmm(191), P-6m2(187)
186	P 31 – –	P31(144), P3121(152), P3112
		P3221(154), P3212(153)
187	P – – c	P31c(159), P-31c(163), P63m
		P-62c(190), P63/mmc(194)
188	P – c –	P3c1(158), P-3c1(165), P63cr
		P-6c2(188), P63/mcm(193)
189	R (obv) – –	R3(146), R-3(148), R32(155),
		R-3m(166)
190	R (obv)- – c	R3c(161), R-3c(167)
191	R (rev) – –	R3(146), R-3(148), R32(155),
		R-3m(166)
192	R (rev)- – c	R3c(161), R-3c(167)
193	R – – –	R3(146), R-3(148), R32(155),
		R-3m(166)
194	R – – c	R3c(161), R-3c(167)
195	P 63 – –	P63(173), P63/m(176), P6322
196	P 62 – –	P62(171), P6222(180), P64(17

| 197 | P 61 – – | P61(169), P6122(178), P65(1 |
| 198 | P – c c | P6cc(184), P6/mcc(192) |

Cubic

199	P – – –	P23(195), Pm-3(200), P432(2
		Pm-3m(221)
200	P 21 – –	P213(198)
201	P 42 – –	P4232(208)
202	P 41 – –	P4132(213), P4332(212)
203	P – – n	P-43n(218), Pm-3n(223)
204	P a – –	Pa-3(205)
205	P n – –	Pn-3(201), Pn-3m(224)
206	P n – n	Pn-3n(222)
207	I – – –	I23(197), I213(199), Im-3(204
		I-43m(217), Im-3m(229)
208	I 41 – –	I4132(214)
209	I – – d	I-43d(220)
210	I a – –	Ia-3(206)
211	I a – d	Ia-3d(230)
212	F – – –	F23(196), Fm-3(202), F432(2
		Fm-3m(225)
213	F 41 – –	F4132(210)
214	F – – c	F-43c(219), Fm-3c(226)
215	F d – –	Fd-3(203), Fd-3m(227)
216	F d – c	Fd-3c(228)

Triclinic

| 217 | P – | P1(1), P-1(2) |

_____

Expo2014

Crystal structures solution by powder diffraction data

PDF generated May 30, 2024 at 8:41 PM by Kalin's PDF Creation Station WordPress plugin