EXPO2014

Table of contents

General Information

EXPO2014

The program EXPO2014 is the evolution of EXPO2009. EXPO2014 performs all the steps of the *ab-initio* structure solution by powder data: indexing, space group determination, estimation of the reflection integrated intensities, structure solution by Direct/Patterson 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.

Program specification

X-ray, synchrotron or neutron diffraction data can be used. There are no limits on the number of counts in the pattern. The maximum number of reflections is 3000.

The background of system

EXPO2014, the updated version of EXPO2009 is able to solve crystal structures from unindexed powder diffraction pattern. Its main steps are (select the topic for more details):

Indexing

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

Space Group Determination

A statistical analysis is performed by using the experimental diffraction intensities compatible with the identified crystal system. The procedure is able to provide a list of all the possible extinction symbols ranked according to their probability value. For each of them, the corresponding list of compatible space groups is supplied *via* graphic interface.

Structure Solution

Structure solution can be carried out by the two alternative following methods:

- 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 Rietveld method.

Files generated by EXPO2014

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 integrated intensities of the generated reflections (name.hkl);
- an ascii file to store peak positions if the indexing process has been carried out (name.pea);
- output file containing information about all the steps of structure determination (name.out).

Authors

A. Altomare[*],C. Cuocci[*], C. Giacovazzo[*], A. Moliterni[*], R. Rizzi[*].

[*] Institute of Crystallography CNR, (Bari) - Italy

Support

sirmail@ic.cnr.it
http://www.ba.ic.cnr.it/contact

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 a publication please include in your citations the reference to following work:
A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni, R. Rizzi, N. Corriero and A. Falcicchio, "EXPO2013: a kit of tools for phasing crystal structures from powder data", J. Appl. Cryst. (2013). 46, 1231-1235

EXPO2014 installation

Hardware and software environment

EXPO2014 is built utilizing OpenGL for three-dimensional molecular visualization and the GTK for the graphic user interface. All binary packages are pre-compiled by using the Intel

Fortran Composer XE 2013 and the GNU C++ compiler. Expo2014 runs on the most popular operating systems: Windows, Apple's Mac OS and some Linux distributions.

On Windows platforms

Both 32- and 64-bit versions of EXPO2014 are available for Windows. They were tested on Windows 7, Vista, and XP.
To get the 32-bit or 64-bit version of EXPO2014 for Windows, download the file expo-020513_install.exe or expo-020513-x64_install.exe from
http://www.ba.ic.cnr.it/content/expo2013
Note that the 32-bit version of EXPO2014 can also be used on 64-bit Windows. To install the software, double click on the downloaded file and follow the instructions.
The program can be launched by double click on the icon of EXPO2014

Mac OS X

The Mac OS X version of EXPO2014 runs on Mac OS X 10.5 (Leopard), Mac OS X (Snow Leopard), Mac OS X 10.7 (Lion), OS X 10.8 (Mountain Lion) . The program was not tested on on Mac OS X 10.4 (Tiger) or older.

The Mac OS X version is contained in a file .dmg with a disc image format.

1. Download the file .dmg from
http://www.ba.ic.cnr.it/content/expo2013
2. Double-click the .dmg file to mount it. A new Finder window containing the EXPO2014's icon should appear.
3. Open a new Finder window and then drag the program EXPO2014 over the "Applications" in the left side of window.
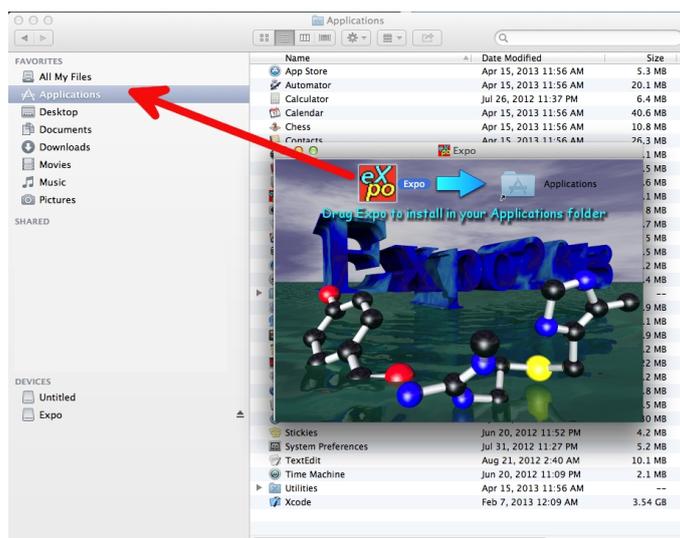
To do cleaning:
· Close any Finder Windows
· Eject the disk image (not the .dmg file).
· Delete the .dmg file by dragging it to the trash.

In order to add EXPO2014 to the dock, follow these steps:
- Open up a new Finder window
- Click on "Applications".
- Locate the EXPO2014's icon.
- Drag the icon to your Dock, and drop it wherever you like.

EXPO2014 can be launched by Double click on the icon of EXPO2014.

On Linux platforms

Installing from binary package

Debian and Red Hat RPM packages are available via the download page.

- Install the .deb package typing from the terminal

```
sudo dpkg -i expo_1.13.05_ubuntu12.10_amd64.deb
```

- To install RPM package, enter the command:

```
rpm -i expo-1.13.05.fc18.x86_64.rpm
```

EXPO2014 can be launched by
1. Double click the icon of the EXPO2014 icon

2. Type the `expo` command. Read the section for more details.

Compiling from source

- Install build tools

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

   Autoconf Program which writes configure scripts
   automake Program which writes Makefiles
   libtool Generic library support script
   g++ C++ compiler from the GNU compiler collection
   Fortran compiler compliant with Fortran 95 standard: EXPO has been written and tested using the Intel Fortran Compilers (IFORT) but if the IFORT is not available, GNU Fortran (Gfortran) could be also used.

- Install graphic development packages
   Expo requires development files for gtk+ library, Mesa OpenGL utility library. The name of the libraries could be slightly different depending on your GNU/Linux distribution. The names of the packages is reported for some tested distribution.
   - Debian distributions: `libgtk2.0-dev, libgl1-mesa-dev, libglu1-mesa-dev, libcurl4-gnutls-dev`
   - RedHat distributions: `gtk2-devel-....., mesa-libGL-devel-....., mesa-libGLU-devel-....., libcurl-devel-....`
   - Mandriva: `libgtk+2.0-devel, libmesagl1-devel, libmesaglu1-devel,libcurl-devel`

o Suse: `gtk2-devel`, `mesa-devel`, `libcurl-devel`

· Download the file *expo_1.13.05.tar.gz*
Install "expo" using the following commands

```
tar xvfz expo_1.13.05.tar.gz
cd expo_1.13.05
sh ./autogen.sh
./configure         --prefix=destination_directory        --
srcdir=expo_source_directory
make clean
make install
```

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

For more information read the file INSTALL in the expo directory.

Command line usage

When typing the `expo` command, you can supply different types of input file

Type the following command in a command line:

```
expo merca.exp merca.out
```

This command will start EXPO2014 running the expo input file `merca.exp` and generating the output file `merca.out`.
The output file may be omitted:

```
expo merca.exp
```

or simply

```
expo
```

In addition, EXPO2014 can read crystal structures from different types of file (e.g. `*.cif`). You can supply the name of the crystal structure file on the command line, and EXPO2014 will open the file when it starts, e.g.

```
expo paracetamol_best1.cif
```

You can also type the name of a powder diffraxtion data file (e.g. `*.xy`, `*.dat`)

```
expo merca.xy
```

Use the option `-help` to displays a brief help message, showing the main command-line options

```
expo -help
```

On a Windows PC to run EXPO2014 simply typing `expo`, the environmental variable PATH must be set to include absolute path of the directory where EXPO2014 is placed.

On Linux, if the program has been installed from source files (file expo_1.13.05.tar.gz) create a symbolic link with 'ln -s' of EXPO2014 in a directory where environment variable path is set, or declare the path with the shell command `export`

`PATH=destination_directory/bin:$PATH`, `destination_directory` is the directory where EXPO2014 was installed.

Running examples

Test structures can be found in the folder `examples`. On Windows platform this folder is typically located in `C:\Users\username\share\expo\examples` and on linux operating system in `/home/username/share/examples`. To run the program select the test structures by means of the menu **File > Load Examples**

Click on `Next >` on the toolbar to perform the steps of structure solution.

Three test structures are supplied and described in the section Tutorials.

Contact

If you are having trouble during the installation contact corrado.cuocci@ic.cnr.it

Tutorials

Tutorial 1 - Crystal structure determination of Cimetidine compound

The first tutorial is supplied for carrying out the full pathway of the ab-initio structure solution process (from indexing up to the structure solution) in case of cimetidine.
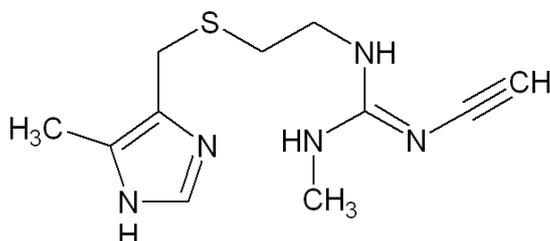
The only required information by EXPO2014 is the diffraction data and the chemical formula ($C_{10}H_{16}N_6S$).
The *examples* directory contains the necessary files:
- *cime.exp*: input file for EXPO2014
- *cime.dat*: a synchrotron X-ray diffraction data file containing the initial 2-theta, the 2-theta step and the final 2-theta values (in the first line) and the experimental counts (in the rest of lines).

To access to *examples* directory:
- Open EXPO2014 by double clicking on Expo2014 icon.
- Click on **File > Load Examples**
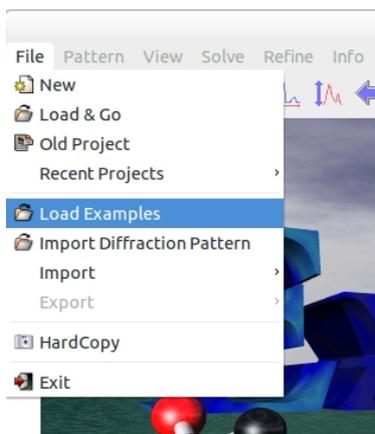
and select the file cime.exp.

The content of file cime.exp is here reported

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

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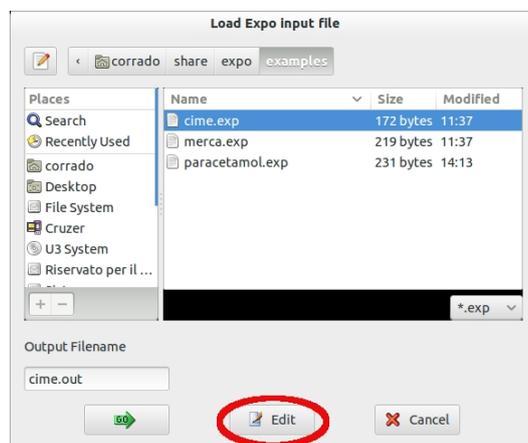
More information about the creation of input file can be found in the section Preparing input file.

Clicking on **OK** the powder diffraction data will be displayed.
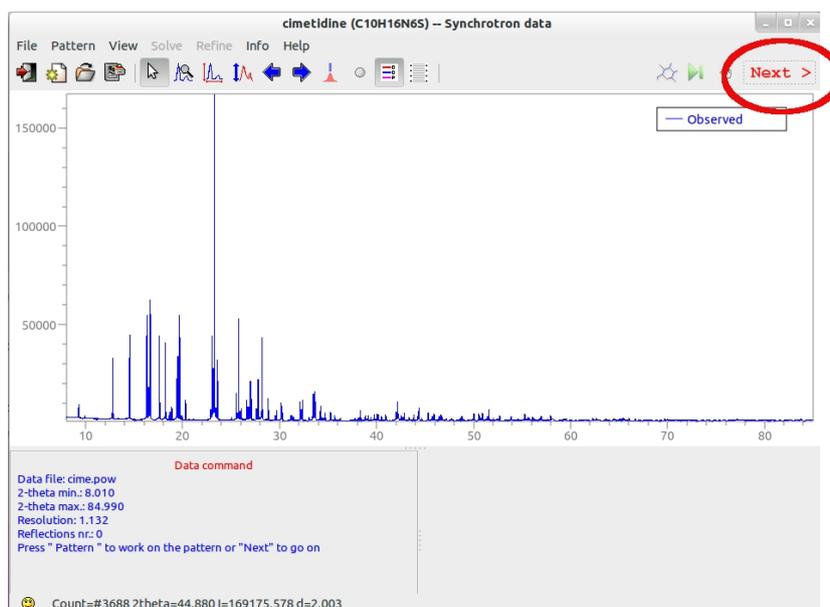
Selecting **Next,** an automatic peak search procedure will be applied and the *d*-values corresponding to the selected experimental peaks will be automatically supplied to the indexing procedure. At the end of this process EXPO2014 will provide the indexing results by graphic interface:

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) provided by the window above or by the output file of EXPO2014.
The cell content is given by Z multiplied 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 $C_{10}H_{16}N_6S_4$ and the cell volume is 1280 Å^3. If the fist criterion is applied, the molecular volume is given by: Vmol = 20 x 18 = 360 Å^3.

The corresponding Z value is: Z = Vcell/Vmol = 1280/360 = 3.55, 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 EXPO2014 by graphic interface is `(C10H16N6S)4` or `C40H64N24S4`

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

the most plausibile Extinction Group *P 1 21/a 1* is automatically selected and by clicking on **OK** the space group P *21/a,* corresponding to this extinction group, will be chosen.

After the determination of the space group, EXPO2014 starts with the extraction step which is carried out in the selected space group. Click on **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 Direct Methods for structure solution.

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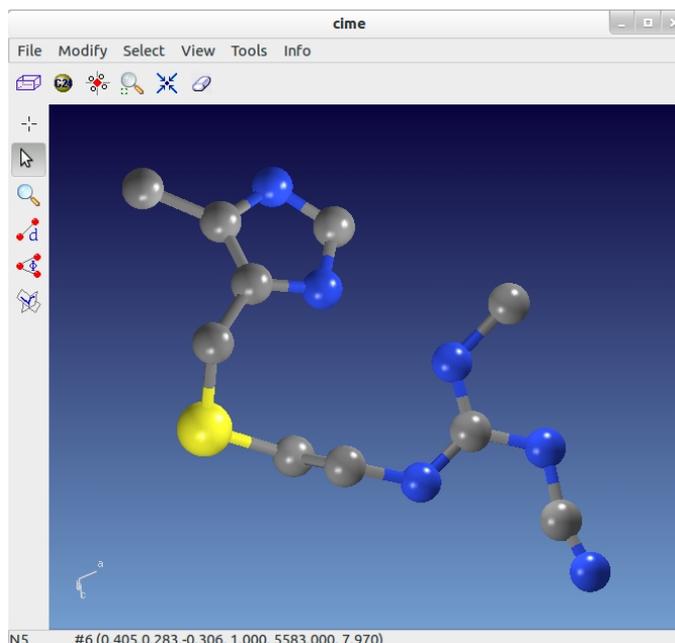
Some errors in the assignment of carbon or nitrogen atomic species can be easily recognized and corrected by using the option 'Modify' from JAV menu.

Tutorial 2 - Crystal structure solution of 2-Mercaptobenzoic acid compound
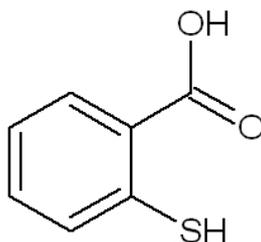
The aim of this example is to guide the user in case of failure of a default run of EXPO2014.

Open EXPO2014 by double click on Expo2014 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 28 O 8 S 4 H 24
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 **Next** button to go on continuously.

The final structure model provided by EXPO2014 could not correspond to the correct crystal structure of 2-Mercaptobenzoic acid. This can occur when the most plausible set of phases, provided by Direct Methods, is unreliable.

In a default run of EXPO2014, Direct Methods generate several phasing trials; 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, all the rest of stored phasing trials should be explored in order to find the 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
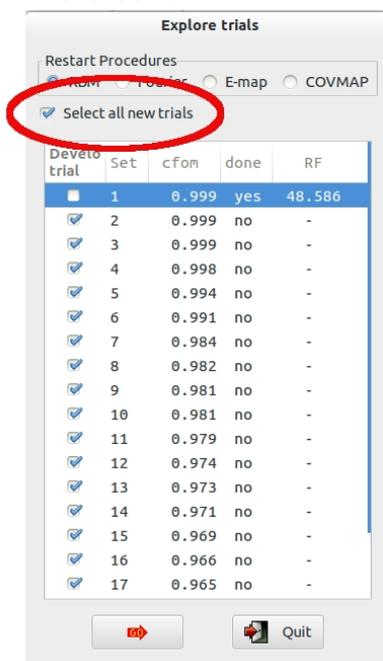
c) Clicking on **GO**

At the end of the procedure, EXPO2014 will rank the phasing trials (and the corresponding structure models) according to the increasing R_F agreement factor values.

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

More information about the strategies to apply when Direct Methods fail are described in the next section of this manual.

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 by direct space methods of paracetamol (form I polymorph) compound.
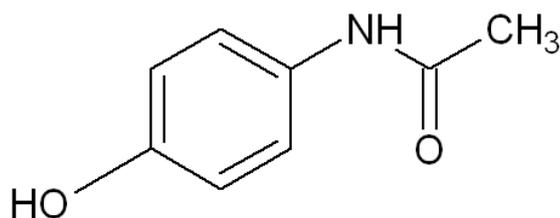
Open EXPO2014, select **Load Examples** and load the input file 'paracetamol.exp'. Clicking on **Next** button, the file paracetamol.mol, containing a 3D molecular description of paracetamol, will be loaded and displayed and the following window will appear:
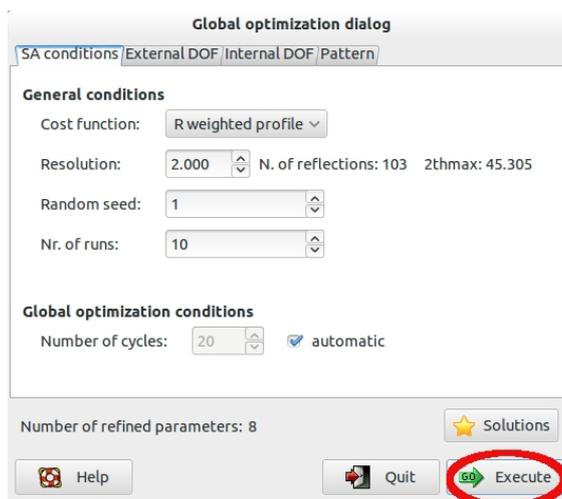
Select **Execute**; the Simulated Annealing procedure will be applied performing 10 runs whose corresponding final structure models can be explored by the button **Solutions**
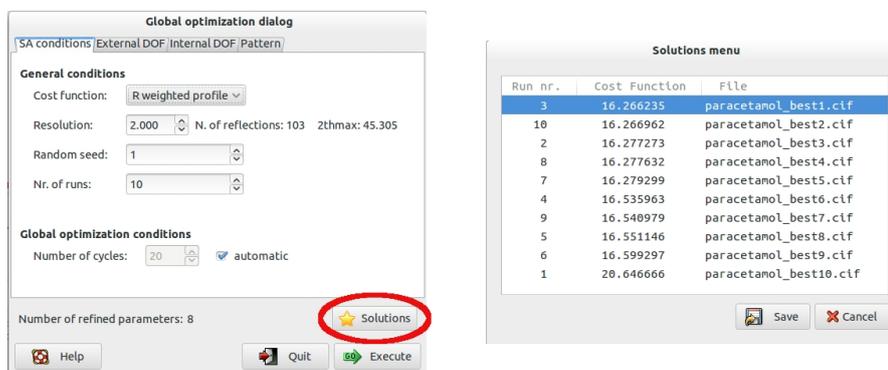
References

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 - Florence, A. J., Shankland, N., Shankland, K., David, W. I. F, Pidcock, E., Xu, X., Johnston, A., Kennedy, A. R., Cox, P. J., Evans, J. S. O., Steele, G., Cosgrove, S. D. & Frampton, C. S. (2005). *J. Appl. Cryst.* **38**, 249-259.

Paracetamol (form I polymorph) - Florence, A. J., Shankland, N., Shankland, K., David, W. I. F, Pidcock, E., Xu, X., Johnston, A., Kennedy, A. R., Cox, P. J., Evans, J. S. O., Steele, G., Cosgrove, S. D. & Frampton, C. S. (2005). *J. Appl. Cryst.* **38**, 249-259.

Preparing input file

Preparing input file in EXPO2014

To run EXPO2014 you need of an input file (*.exp). Two possibilities are available:
1) to create a new input file; 2) to use an already existing input file.
Possibility n.1): creation of an input file by selecting, *via* graphic interface, the creation of a new project by the menu "File">"New".
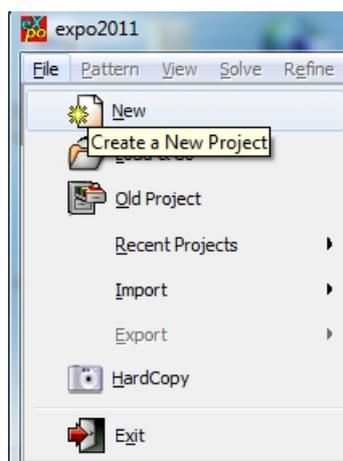
The required information are here listed.

1. You must specify the structure name. This name will be used to create the name of output file.
2. Load the profile counts file. The following powder diffraction data file can be imported: ASCII XY files (*.xy), CIF powder files (*.rtv, *.cif), GSAS files (*.qda), CCDC Mercury

xye (*.xye). Use the extension *.pow to indicate an ASCII xy file or an ASCII file with 2theta min, 2theta step, 2theta max in the first row and counts in free format in the subsequent lines.

3. Select the 'Range' button if you want to change the 2theta min and/or the 2theta max.
4. Input the wavelength and the type of radiation source (X rays source is the default choice).
5. To carry out the space group determination it is necessary to activate the check button 'Find Space Group'.
6. For indexing, it is necessary to click on button 'Cell Determination by N-TREOR'.
7. You must provide the content of unit cell. This information is important for the space group determination and for the structure solution by Direct Methods. A valid molecular formula for the cell content must meet the following criteria:
   o Elements are represented by valid atomic symbols (IUPAC-nomenclature) each optionally followed by a number.
   o It is possible to write the formula in lower-case or upper-case but please note that 'Sb' will be interpreted as one antimony but 'SB' will be one antimony and one boron. Spaces are also permitted (eg. 'S B', one antimony and one boron). We suggest to write the two-letter elements with the second letter in lower-case.
   o Valid operators are: **( ) [ ] { }**

Examples of valid formula: `[C3H5(OH)3]4, C 28 O 8 S 4 h 24`

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 EXPO input file (*.exp) will be created in the same folder of powder data file and automatically loaded by the program for the indexing. The button 'Open' (10) is used to open (and possibly modify with a text editor) an already existing input file.

An already existing *.exp input file can be loaded from menu 'File' > 'Load & Go'.
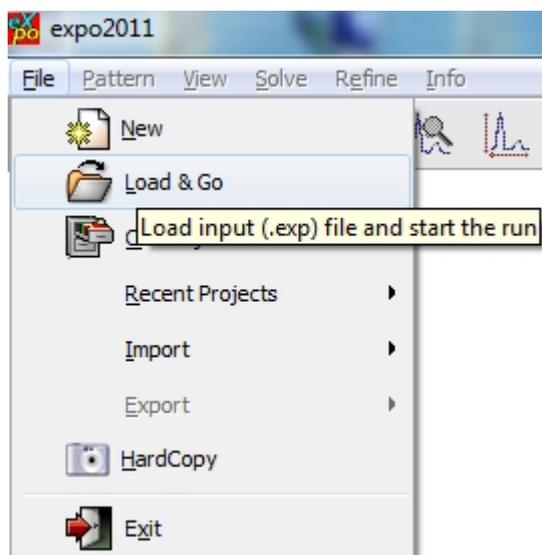
Format of diffraction data

EXPO2014 recognizes some formats of diffraction data that should be ASCII text files:
1. Files with extension `.xy` or `.xye` containing 2theta in the first column and the intensities in the second one. The format of counts can vary. # could be used for comment

line.

```
#This is a comment
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
```

Additional column after the second will be ignored. Optionally, the wavelength can be included in the file as the first line.

2. Files type with extension `.dat` containing 2theta minimum, 2theta step, 2theta maximum in the first line. Counts are in the successive lines separated by blanks. The number of counts for line and the format of counts can vary.

```
# example 1 for ascii file with start,step,end,intensity
6 0.020 90.0
149. 161. 151. 136. 144. 151. 150. 149. 117. 138.
157. 120. 143. 156. 110. 142. 147. 147. 124. 128.
130. 144. 134. 159. 145. 150. 138. 149. 134. 159.
142. 139. 135. 136. 147. 139. 155. 133. 140. 133.
134. 158. 165. 134. 156. 139. 122. 141. 158. 132.
150. 138. 134. 149. 147. 150. 154. 165. 145. 140.
..............................................
```

3. Additional readable ASCII file formats are: GSAS (`*.gda`), CIF (`*.rtv, *.cif`)

If the file has the extension `.pow` or the extension is unknown, the program tries to read according to the format of `.xy` or `.dat` file

---

Indexing by EXPO2014

Indexing by EXPO2014

In powder diffraction the first necessary step for solving crystal structure is the indexing, i.e. finding the unit cell parameters and reconstructing the three dimensional lattice from the one-dimensional information of the powder diffraction pattern. The indexing 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 provides a graphical interface for two popular and freely available indexing program: DICVOL (Boultif & Louër, 2004), McMaille (Le Bail, 2004).

Starting indexing by EXPO2014

The simplest method to perform indexing by EXPO2014 is input a powder diffraction data file by the menu **File > Import Diffraction Pattern.** Select the appropriate radiation type and the used radiation wavelength.
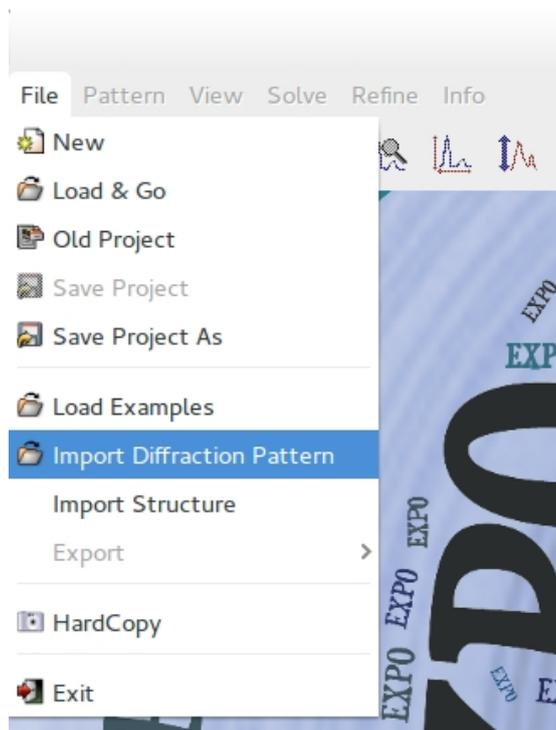
In section [format of diffraction data](#) the possible formats of the powder diffraction data file required by EXPO2014 are described. As example you can input the diffraction data `merca.xy` which is located in the directory examples and supply wavelength value 1.54056 Å.

Then select the indexing program from **Pattern > Indexing**. If you choose N-TREOR09 the dialogue box shown in the following figure opens.
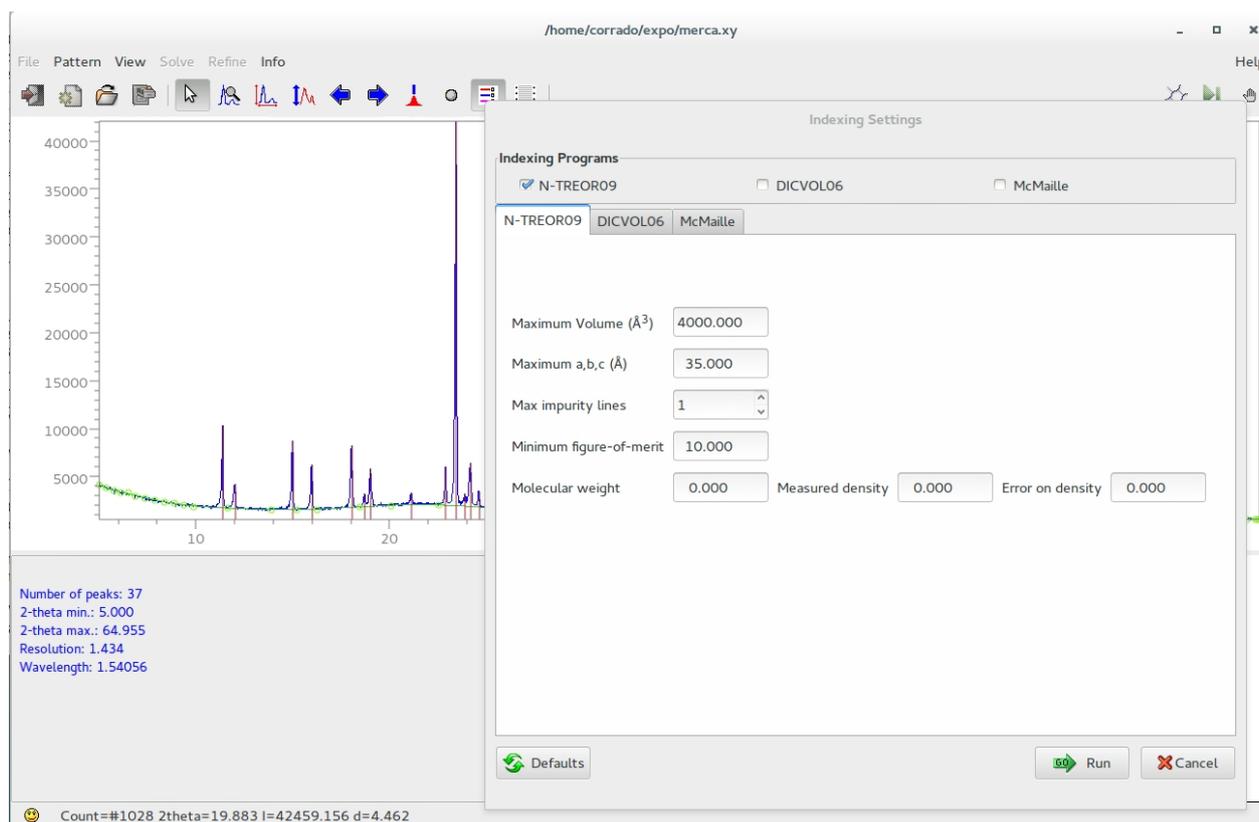
The user can modify the default values of the maximum cell volume, the maximum cell axial lengths, the maximum number of allowed impurity lines, the minimum de Wolff figure of merit (M20). If known you can enter the measured density with the corresponding error. Molecular weight is not used unless density is not zero.

If you tick DICVOL06 in the frame 'Indexing Programs' you can also activate the interface for indexing by DICVOL06. In this case EXPO2014 creates a DICVOL06 input file and automatically will run the program. 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` where `expo_install_dir` denotes the directory where you have installed the package. Under Linux OS, the Windows Emulator Wine is required to run DICVOL06. Wine is available in the software center of the most common linux distributions. You should read general explanation and suggestions about indexing in the DICVOL06 documentation available at http://cds.dl.ac.uk/cds/peakquest/dicvol06.html.

EXPO2014 also provides interface to McMaille version 4.00 (http://www.cristal.org/McMaille/McMaille-v4.html).

Clicking on **Pattern > Indexing >** and selecting the indexing program EXPO2014 automatically performs a peak search if peaks have not been already detected. If you think that the automatically located peaks are correct click on button **Run**. On the contrary, if you need to improve the selection of peaks press 'Cancel' and read carefully the next section.

Selecting Peaks for Indexing

Accurate and reliable peak search is a fundamental stage for a correct powder pattern indexing. EXPO2014 uses the second derivative method in peak recognition. This method is actually a combination of background subtraction and smoothing, followed by the calculation of derivatives. The automatic peak search is extremely sensitive to noise and may suffer from some defects due in particular to the misinterpretation of background that makes inefficient the algorithm. Moreover, the algorithm usually discards broad peaks at higher angle and in strong overlap.

In order to check the reliability of the automatic peak search results, use the buttons to explore the pattern. If the automatic peak search outcome is not reliable, you can modify the peak search conditions in different ways:

1. By changing the curve describing the background. This is possible adding or removing background points by clicking on the button and 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.
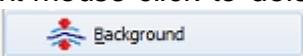
2. By choosing the tool bar button to add peaks by left click of mouse and remove

peaks by right click.

3.By the menu **Pattern > Modify 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. Threshlod' will be selected.
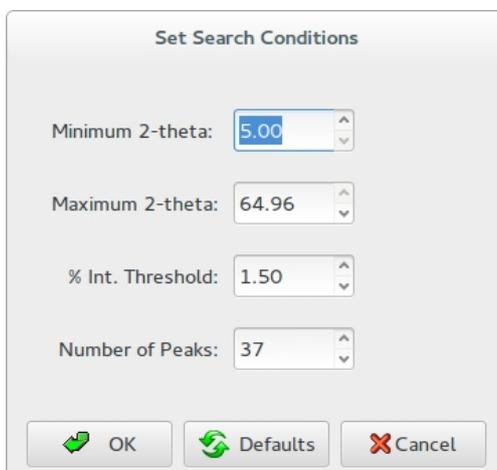
4.Use the button **Pattern > Peak List** to access to a list of values associated to each peak: 2q, *d*, peak intensity calculated removing the contribution of background, full width at half maximum (FWHM). This window can be used to edit the 2q or d values and to delete peaks. If you click on a row in the list, the corresponding peaks will be marked with tick on the pattern.

5.Use **Pattern > Modify Peaks > Load Peaks / Save peaks** to load or save peaks as 2theta or d values.
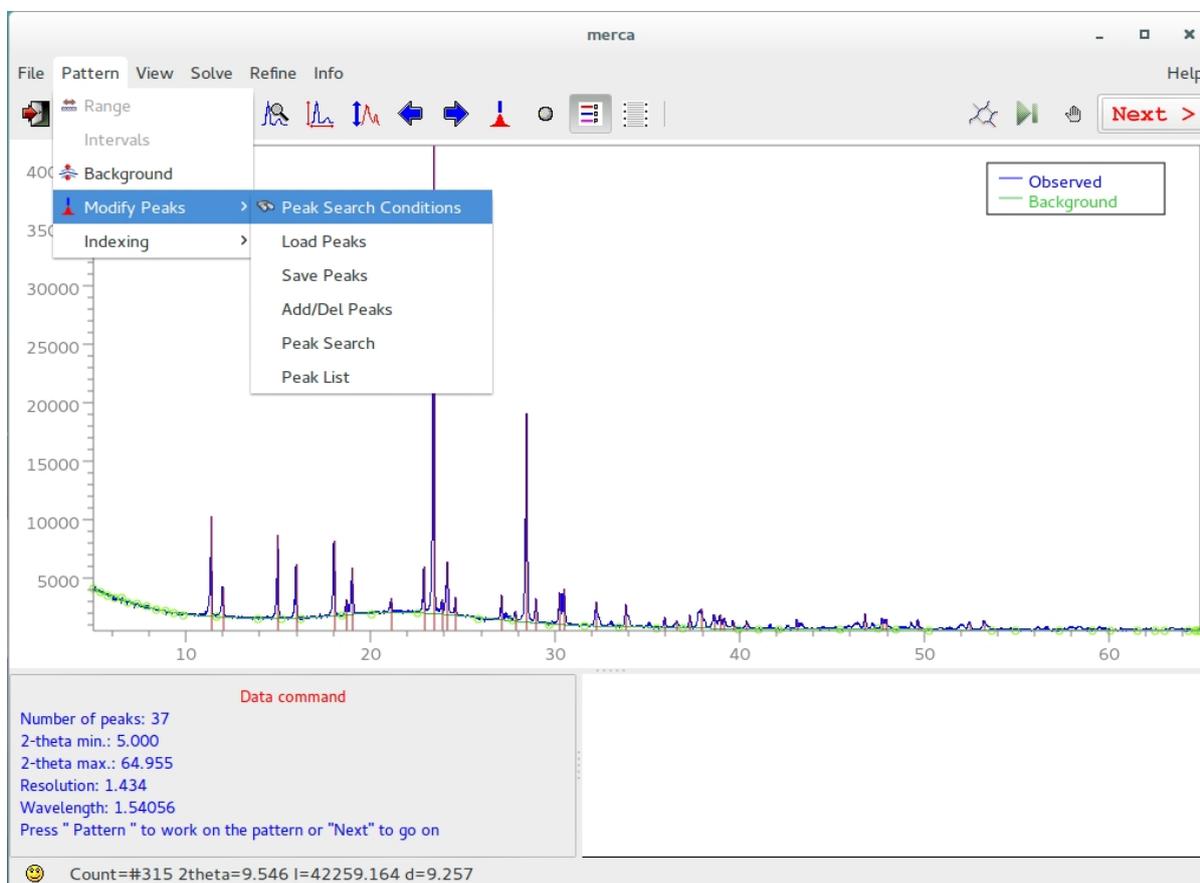
During the peak search remember that the number of peaks must be not too large; and the indexing problem can be generally solved with the 25 first peaks (even very weak peaks included). Peaks at low angle not clearly resolved and with an appreciable peak width should be not included to avoid large errors in peak position.

In case of failure try to improve the indexing 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 better indexing solution. The failure could be due to the presence of impurity peaks and/or to the absence of one or more useful peaks. Sometimes could be very useful to reduce the intensity threshold when weaker peaks are missed. On the other hand if you suspect the presence of an impurity and if the impurity phase is a minority phase in the specimen, it can be useful to reduce the selection to the most intense peaks, e.g. 16-20 peaks may be sufficient. Indexing program as N-TREOR09 and DICVOL06 are able to successfully index even if some peaks are omitted but they can fail if a lot of dubious peaks are included.

Submit the selected peaks to a indexing program by click on **Run** in the window 'Indexing Settings', accessible by **Pattern > Indexing**

Running Indexing in EXPO2014

The next step is the automatic refinement of peak positions and profile parameters (a Pearson VII function is used). At the end of the peak search a file named `structure_name.pea`, containing the refined d values associated to the selected peaks, is automatically created by the program. This file can be used in the future to skip the peak search procedure (see the directive `extefile`) if a new indexing process is performed. Additional input and output files can be found in the working directory: `structure_name.new`, `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.

At the end of the indexing procedure you can see a list containing, for each solution, the program used for indexing (**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, the crystal system. In addition, the figure of merit FOMnew (Altomare et al., 2009) and the most probable extinction symbol are available when N-TREOR09 is used. McM20 is displayed in case of indexing performed by McMaille.

Plausible cell parameters

Select cell

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

OK   Cancel   Export

Sometimes more plausible cells are suggested, they are ranked according to a figure of merit for each indexing program. When the quality of data is not good, uncertainty in the indexing procedure is usually manifested by the generation of more than one solution. If several unit cells are possible, the different solutions 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 determined cell might be wrong.

The quality and the reliability of indexing, therefore, is routinely characterized by means of various numerical figures of merit. The figures of merit reported by EXPO2014 are M20, FOMnew only for N-TREOR09 and McM20 only for McMaille. Consult the output file of the specific program (Pattern > Indexing > Output) to additional figures of merit (e.g. F20) for each solution. The more reliable indexing should yield the higher value of figure of merit. 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 correct solution is usually characterize by a high figure of merit and all diffraction peaks (or most of them) should be indexed.

If a program fails in indexing a powder pattern, it is possible that another program can succeed. Similar results obtained by different programs strengthen the confidence in the correctness of the cell. Because N-TREOR09 and DICVOL06 run usually much faster than McMaille, it is better to apply them first.

N-TREOR09 includes some features in the default strategy: an automatic correction of the 2theta zero-shift is made; it may successfully index powder diffraction patterns containing a few impurity peaks; 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 and therefore, indexing in the lower symmetry unit cells will not be performed.
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 long time depending on the indexing program.

Once the cell parameters calculation 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 `(C7H6O2S)4`. For cimetidine compound the cell content is `(C10H16N6S)4`

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

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.

2. Volume per Atom: ratio between the cell volume and number of non-hydrogen atoms. Its value is expected to be close to 18 $Å^3$ (18 $Å^3$ rule) but sensible deviation from this ideal value can be observed due to the presence of H-bonding or large void space (e.g. 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.

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

Indexing from Expo Input File (`.exp`)

An alternative way to perform indexing by EXPO2014 is to edit a new input file or modify an existing input file with a text editor and save it with the extension `.exp`.

An example of input file for indexing of powder pattern of cimetidine is available in the directory examples (file `cime.exp`). Its content is the following one:

```
%structure cime
%job cimetidine (C10H16N6S) -- Synchrotron data
%initialize
%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 command `%data`. Add `%continue` if you want to continue the crystal structure determination process. You can read the description of the rest of instructions of the input file by looking at the section commands of EXPO2014 and their use of this manual.

Add the command `%dicvol` or `%mcmaille` to activate the indexing by the program DICVOL06 and McMaille, respectively.

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

Once the input file has been created, load the .exp input file from the menu **File > Load & Go**.
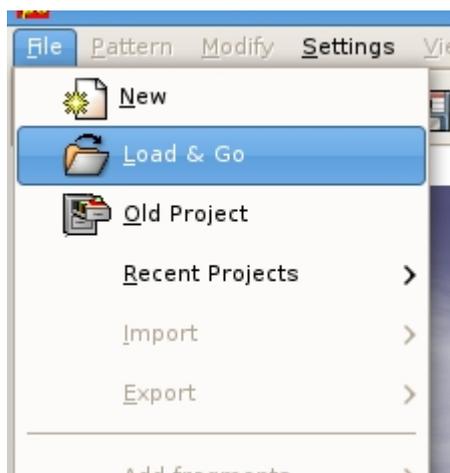
Use the button **Next >** in the toolbar to carry out the steps of the indexing procedure.

Indexing Using the Command Line

To execute the indexing procedure from command line, type the name of powder diffraction data file and add the option `-index=indexing program`. You can specify the program(s) to use for indexing with n for N-TREOR09, d for DICVOL06 and m for McMaille after the argument `-index=`.

```
expo merca.xy -index=ndm
```

the indexing of powder data `merca.xy` will be performed by using N-TREOR09 (`n`), DICVOL06 (`d`), McMaille (`m`).

Add `-wavel=wavelength` to define the wavelength of radiation source. If `-wavel` is omitted the default wavelength of Cu K-alpha 1 (1.54056 Å) is used.

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

You can also add the option `-auto` to suppress the interaction with graphical interface and run the program in automatic mode, `-nogui` to run the program N-TREOR09 without graphical interface.

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

Output of indexing

The output file contains details about indexing. Use **Info > View output file** to access to the output file or open the file `structure_name.out` with a text editor or click on the button ⠿ from the main menu.
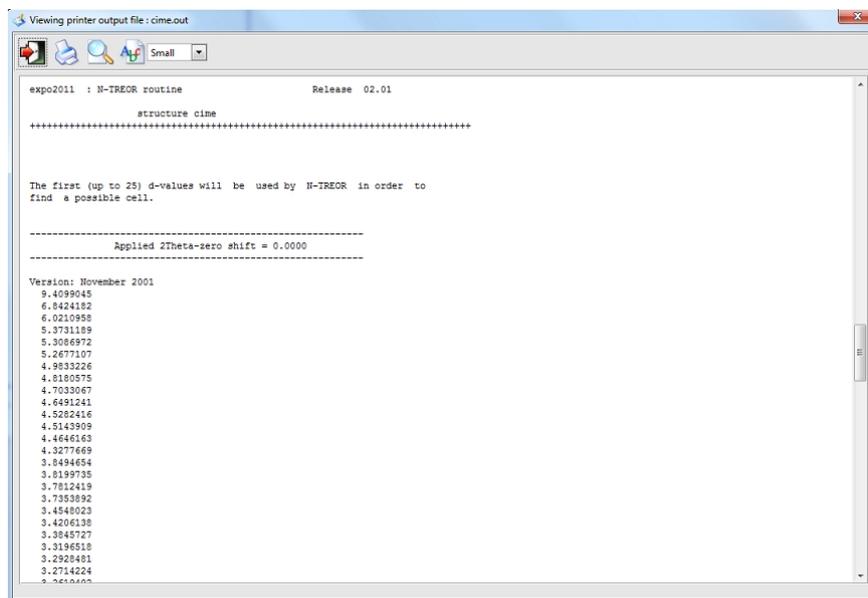
Additional output files are created by the specific indexing programs and are accessible through the menu **Pattern > Indexing > Output**

The indexing directives

Directives in data command

The following directives must be added after the command %data in the input file to activate some specific features useful for the indexing procedure

extefile file_name

This directive is used to supply an external file containing 'd' or '2theta' values of peak positions. In the latter case the additional `2-theta` directive has to be used. An example of the use of this directive for the indexing of the pattern of structure cime:

```
%structure cime
%job structure cime
%init
%data
```

```
     pattern cime.dat
     synchr
     wave 1.52904
     extefile cime.pea
%ntreor
%continue
```

`cime.pea` is an ASCII file containing the observed 'd' values associated to the selected peaks. In this case peak search and peak refinement will be skipped. When the directive `extefile` is omitted the program saves the *d* values associated to the peaks in a file with extension `.pea` before carrying out the indexing process.

2-theta

This directive is used combined with the `extefile` directive to supply to the indexing program N-TREOR09 an external file containing 2theta values of the peak positions.

search 2theta-min 2theta-max peak_number threshold

This directive is used to specify the conditions of peak search. The user has to indicate four parameters: the search range of the peaks in 2theta (`2theta-min` and `2theta-max`), the number of peaks (`peak_number`), the intensity threshold (`threshold`). Negative or zero values are ignored, it is not necessary to specify all values.

E.g.
To search the 25 most intense peaks in the range 5-80 add the following line after `%data` command:
```
search 5 80 25
```

To search the 20 most intense peaks:
```
search 0 0 20
```

To apply a threshold of 2.5
```
search 0 0 0 2.5
```

<p align="center">**Directives in ntreor command**</p>

The N-TREOR09 directives must be added after the command `%ntreor` in the input file to enable some specific features of the indexing procedure

All the directives of the indexing program TREOR (Werner, Eriksson & Westdahl, 1985) are still valid for the `%ntreor` command and have to be used with the same syntax (see the N-TREOR manual for the details).

The default choices can be changed *via* suitable directives.

An example of application of N-TREOR09 directives is the following one:

```
%structure diltia
%data
     pattern pd_0029.xye
     wavelength 1.54056
     extefile diltia.pea
%ntreor
     CEM = 50,
```

`CEM = n,` is a directive setting to n the maximum allowed cell axis

In a default indexing run, N-TREOR09 searches for cells with a maximum cell volume of 4000 $Å^3$ and a maximum cell axis of 35 Å. Due to this limits the correct cell is not found in a default indexing run for compounds having larger cell axis values. This is the case of Diltiazem hydrochloride (`diltia`) for which one of the maximum axis value is 42.190 Å. The directive 'CEM = 50' activates N-TREOR09 for a search for a cell having a maximum axis value of 50 Å.

Contact

For suggestions and bugs contact annagrazia.moliterni@ic.cnr.it.

References

Altomare, A., Campi, G., Cuocci, C., Eriksson, L., Giacovazzo, C., Moliterni, A., Rizzi, R. & Werner, P.-E. (2009). *Advances in powder diffraction pattern indexing: N-TREOR09. J. Appl. Cryst.* **42**, 768-775.

Altomare, A., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Rizzi, R. & Werner, P.-E. (2000). *New techniques for indexing: N-TREOR in EXPO. J. Appl. Cryst.* **33**, 1180-1186.

Boultif, A., Louër, D., (2004). *Powder pattern indexing with the dichotomy method. J. Appl. Cryst.* **37**, 724-731.

Le Bail, A., (2004). *Monte Carlo indexing with McMaille. Powder Diffraction* **19**, 249-254.

De Wolff, P. M.(1968). *A simplified criterion for the reliability of a powder pattern indexing. J. Appl. Cryst.* 1, 108-113

Werner, P.-E., Eriksson, L., & Westdahl, M. (1985). *TREOR, a semi-exhaustive trial-and-error powder indexing program for all symmetries. J. Appl. Cryst.* **18**, 367-370.

Space group determination

Space group determination by EXPO2014

In the crystal structure solution process the determination of the space group usually follows the indexing step. This task is carried out by a careful investigation of the experimental intensities, aimed at finding the correct 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 uncertain background definition. As a consequence, the judge about the 'presence' or 'absence' of the 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 z intensities (as extracted by the Le Bail method) to calculate the probability of each extinction symbol compatible with the crystal system suggested by N-TREOR09. A suitable weight w is associated to the intensity of each reflection, depending on the expected reliability of its estimate. The algorithm provides a probability for each extinction symbol of the considered crystal system. It 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).

Preparing input file for space group determination in EXPO2014

To run EXPO2014 for space group determination you need to create an input file (*.exp). You can create an input file for space group determination by graphic interface by the menu "File" > "New".
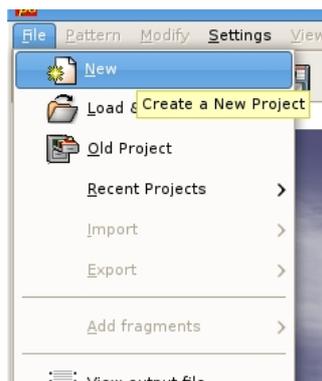
You must specify the structure name, the profile counts filename and its format, the type of radiation source. Click on button 'Find Space Group', fill the frame 'Cell Parameters' and 'Cell Content'. The following picture is an example of input file for the crystal structure determination in case of cimetidine structure from the indexed pattern.

Once the button 'Save' has been selected, an input file cime.exp will be created and automatically loaded by the program for the space group determination.

```
%structure cime
%job structure cime
```

```
%init
%data
    pattern cime.pow
    filetype counts
    wave 1.52904
    findspace
    cell 10.69808 18.81786 6.82267 90.00000 111.28911
90.00000
    cont  h 64 c 40 n 24 s 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 indexing of the experimental pattern. To know the meaning of the other lines of the input file, see the chapter "command and their use" of this manual.

Alternatively you can edit a new input file or modify an exiting input file by 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'.
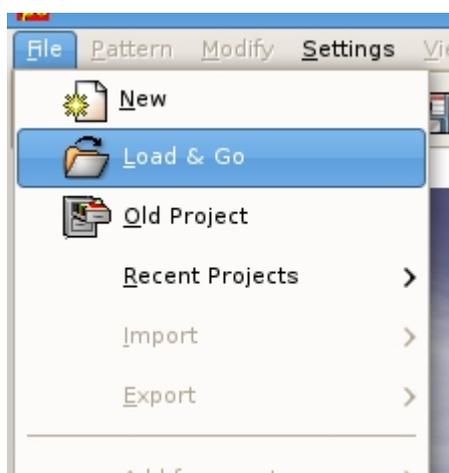
Space group determination by EXPO2013

The different steps of the algorithm are here schematized.

1) The full pattern decomposition is performed, 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 mm for the tetragonal, P 6/m mm for the trigonal-hexagonal, Pm-3m for the cubic systems).

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

3) The z statistic is used to define the probability of the extinction symbol.

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

Use the button **Next >** on the toolbar to carry out to next steps.

At the end of the procedure the following window will be displayed.
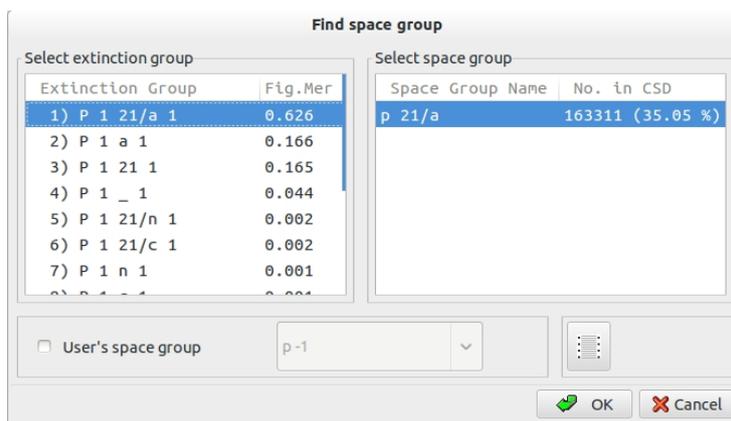
The window contains the extinction symbols ranked according to their calculated probability. When the user clicks on one of them, the corresponding list of compatible space groups is supplied. The frequency of the selected space group in the Cambridge Structural Database (CSD) is reported in terms of number of occurrences in the CSD and percentage. This information can be useful when different space groups are associated to an extinction group.

Press the button 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 reflection, the Miller index, the symmetry operator responsible for the extinction and the corresponding extinction condition, the probability for 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 a list will be marked on the pattern.

The reflections that are expected to be absent according to the selected extinction symbol are marked by a vertical bars situated below the observed pattern, just above the set of bars defining the reflection positions. 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.

Press the 'OK' button to accept the selected space group and continue with the structure solution process.
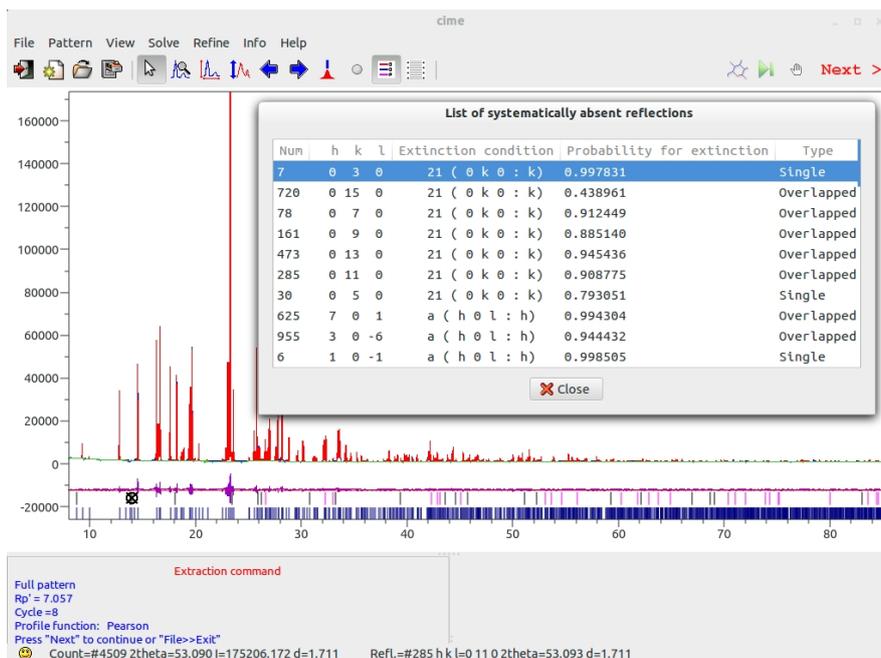
Output of space group determination

The output file contains details about the space group determination procedure. Use 'Info' > 'View output file' to access to the output file or open the file *structure_name.out* with a text editor. At the end of the procedure the 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 for skip the space group determination and to perform the structure solution with Direct Methods.

Contact

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., 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

Structure Solution in EXPO2014 program

Typical X-ray diffraction experiments provide structure factor moduli, while the relative phases are lost. Recovery of the phase information is crucial for crystal structure solution and is referred in crystallography as the *phase problem*. In EXPO2014 program, the default strategy used to solve this problem is based on the application of the Direct Methods (DM) theory (Giacovazzo, 1998). 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 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 will be 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](#)

[When default EXPO2014 fails: strategies](#)

[Suggestions](#)
[Contact and References](#)

Load input file for Structure Solution with 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".
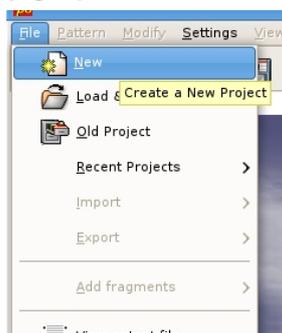
You must specify the structure name, the profile count filename and its format, the type of radiation source. It is supposed that the cell parameters and the space group have been determined before so fill the frame 'Cell Parameters' and 'Space Group'. Activate the check button 'Structure Solution' if you intend to perform the crystal structure solution process with Direct Methods. Otherwise use the button 'Open' to import an existing input 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 will be created and automatically loaded by the program for the structure solution process by DM.

```
%structure cime
%job structure cime
%init
%data
pattern cime.pow
content s 4 c 40 n 24 h 64
wave 1.52904
cell 10.6986  18.8181  6.8246  90.000  111.284  90.000
space p 21/n
sync
filetype counts
%continue
```

You can known the meaning of lines of the input file by looking at the chapter "command and their use" of this manual.

Alternatively you can edit a new input file or modify an exiting input file 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'.
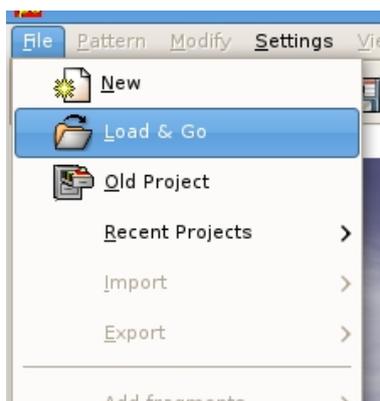
Press the button **Next >** in the toolbar and the following steps of the structure solution process will be automatically performed:

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

2) The integrated intensities are normalized *via* Wilson method (Wilson,1942);

3) Triplet invariants are calculated and their phases are estimated *via* P_{10} formula (Cascarano *et al.*, 1984);

4) Phasing process: more plausible sets of phases are generated whose reliability is assessed *via* the CFOM figure of merit (Cascarano *et al.*, 1992);

5) Twenty sets of phases with the largest CFOM values are stored and ranked in decreasing CFOM order. The first set of phases is automatically selected and used for computing an electron density map *via* an E-Fourier synthesis (E-map): the most intense peaks in the E-map are located and chemically interpreted. The obtained structure model is optimized and completed *via* the automatic structure model optimization.

More information about the structure solution process by Direct Methods are in the chapter 'The steps of structure solution procedure'.
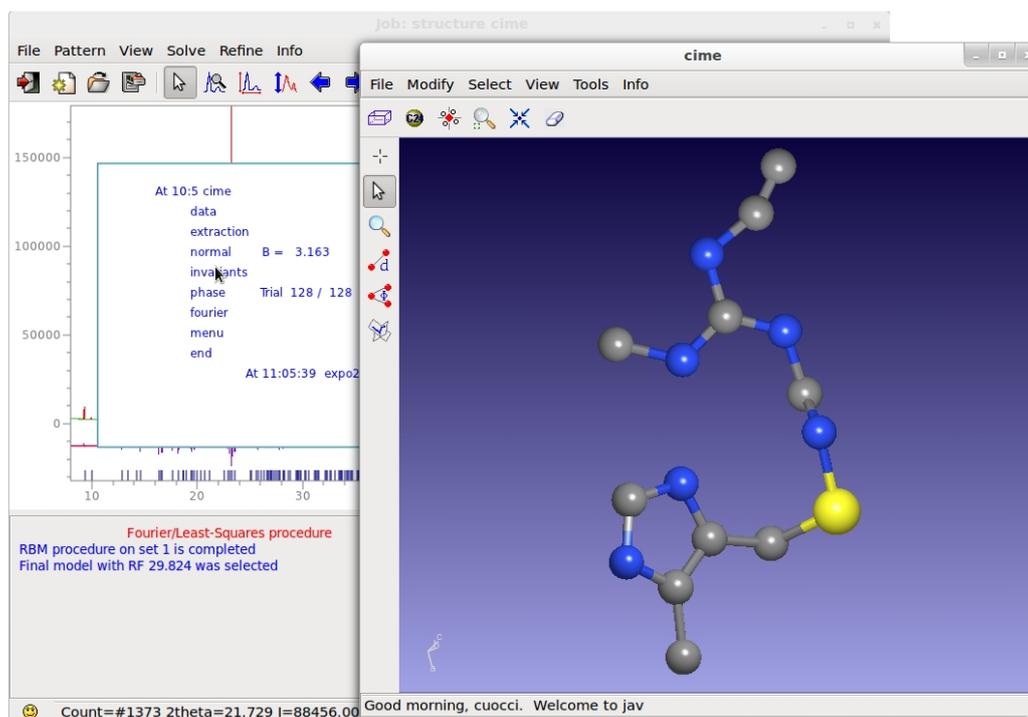
When default EXPO2014 fails: strategies

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

Unfortunately not always you can 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.

Explore trials

In a typical Direct Methods procedure more sets of phases (*trials*) are generated and ranked according to the combined CFOM figure of merit (estimating the quality of the trial) and only the best trial is used to calculate a Fourier map. Because of the unavoidable errors on the phases, the figure of merit can fail to evaluate the best trial, so it is strictly suggested to develop the other trials by menu 'Solve' > 'Explore trials'.
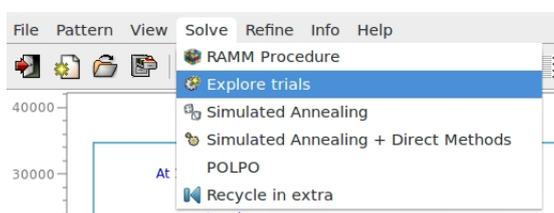
The following window will be opened (the procedure has been applied to 2-Mercaptobenzoic acid compound, whose input file is in the directory 'examples').
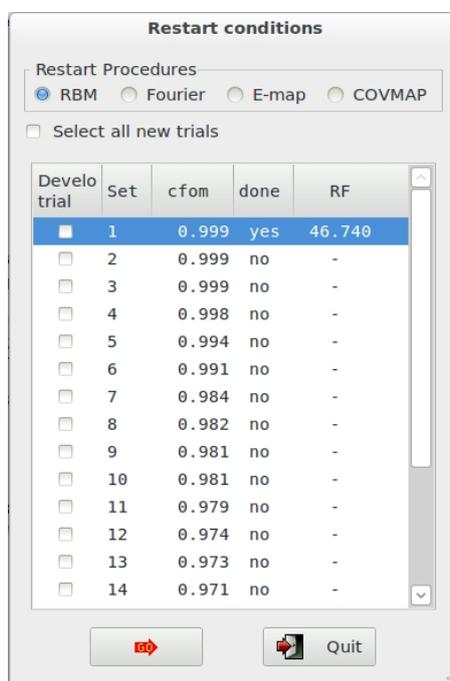
and the the CFOM value can be read. The highest CFOM value could not correspond to the correct solution while subsequent different trials may be successful. This is particularly true when several nearly equivalent CFOM's are available for different trials as in the picture. If the trial has been already developed (done='yes'), you can read the R_F figure of merit value that quantifies the agreement between the structure factors extracted from the experimental pattern and calculated by the model.

Check the button in the first column to decide which trial will be developed and press the

button [GO] If you check the button 'Select all new trials', all the trial 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.

2)Fourier: classical Fourier refinement, default choice for inorganic compounds.

3)E-map: the first electron density map calculated directly after the phasing process by Direct Methods.

4)COVMAP: the covariance principle based completion.

At the end of the procedure, all the trials will appear ordered according to the R_F 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
%Initialize
%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
%extra
%normal
%invar
%phase
%alltrials
```

Structure model optimization

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 are incorrectly positioned or completely false. In this situation the user can try to improve the quality of the Fourier map, as 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, 2010a, 2010b)for reducing, in the electron density map, the errors caused by the limited experimental resolution: peak broadening, peak shift, intensity distorsion. 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 complexityThe structure model obtained at the end of the default strategy can be partly correct with some atoms in correct position but other atoms are incorrectly positioned or completely false. In this situation the user can try to optimize del model by using different optimization strategies.

It is possible to graphically choice different RBM procedures for model optimization by the menu 'Resolution Bias Modification (RBM)', *via* 'Refine' button from the upper EXPO2014 menu.
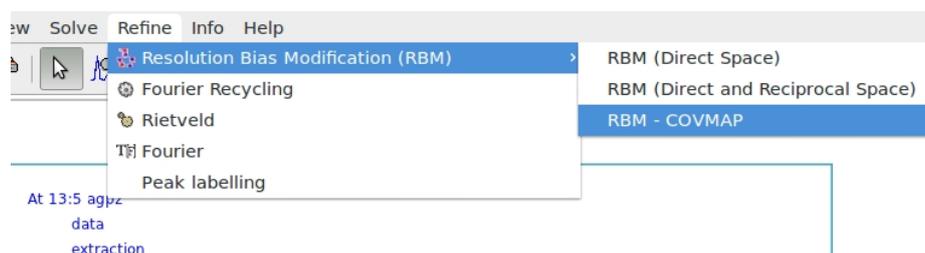
Apply the RAMM (RAndom Model based Method) procedure

The new method RAMM (Altomare *et al*., 2013) has been developed and implemented in the EXPO2014 computing 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 (or by Patterson methods) in order to obtain 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 EXPO: *wLSQ* , *RBM* and *COVMAP* which are able to lead to 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'.
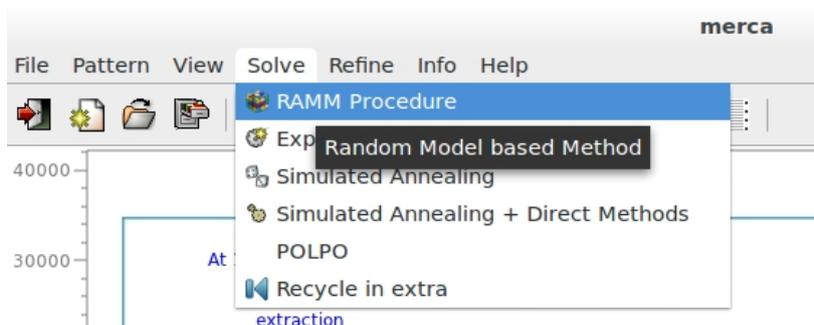
The procedure can take from some minutes to several hours depending on the complexity of problem (number of atoms, quality of data, data resolution). If more than one plausibile solution is found they will be ordered according to the R-structure factor in an interactive list.
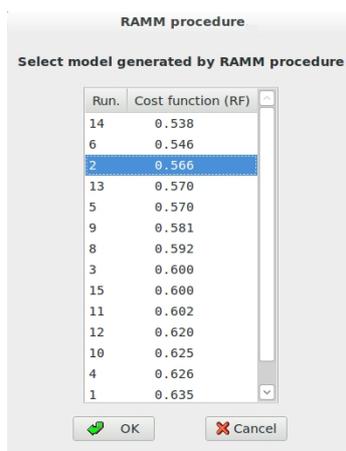
The selected model will be visualized on the molecular viewer. Press 'OK' to accept the selected model.
When Direct Methods procedure fails, the RAMM approach can be a very useful tool to find the correct solution.

The use of the command `%randomsolve` in the input file automatically activate the 'Explore trials' approach. An example of input file is here reported:

```
%Structure merca
%Initialize
%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
```

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 some other secondary strategies.

About 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. 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 (a graphic option must be selected). Click on 'Restart' > 'Recycle in extra'.

- The use of the 'RANDOM' directive in the EXTRACTION module 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 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 activates the procedure).

About the optimization of 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 (directive 'RANDOM' in the PHASE module).

- Second representation formula for triplet invariants proved much more powerful than first representation formula therefore it is used as a default for estimating triplets. In case of failure, the user can choose P3 formula (directive 'COCHRAN' in the INVARIANT module) 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 'PHASE' module).

- 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 (directive 'PSEUDO' in the NORMAL module).

- the value of NREF (number of reflections actively used in the phasing process) is fixed by EXPO2014. 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 NORMAL module).

- High (or low) resolution reflections may occasionally play a too important role in the first steps of convergence and divergence process. Fixing a thermal factor lower (or larger) than that automatically provided by the normalization routine may successfully change convergence and divergence process (directive 'BFACTOR' followed by the isotropic thermal factor in the NORMAL module).

- An alternative space group should be carefully considered (directive 'SPACEGROUP' in the DATA module).

- Sometimes the structure model is partial. Successive combined Fourier map calculations (F_o-F_c, $2F_o$-F_c or F_o) and Rietveld refinements may be conveniently selected by graphical options.

Contact

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., Giacovazzo, C., Moliterni, A.G.G., Rizzi, R. (2006.) *Powder diffraction: the new automatic least-squares Fourier recycling procedure in EXPO2005. J. Appl. Cryst.* **39** (2006) 558-562.

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

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A. & Rizzi, R. (2008*b*). *Correcting resolution bias in electron density maps of organic molecules derived by direct methods from powder data. J. Appl. Cryst.* **41**, 592-599.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2009) *Correcting electron-density resolution bias in reciprocal space. Acta Cryst.* A**65**, 183-189.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2010a) *The dual-space resolution bias correction algorithm: applications to powder data. J. Appl. Cryst.* **43,** 798-804.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. *The dual-space resolution bias correction in EXPO2010. Z. Kristallogr.* **225,** 548-551.

A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni and R. Rizzi (2012). *Covariance and*

correlation estimation in electron-density maps. Acta Cryst. A**68**, 244-255.

A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni and R. Rizzi (2013). *RAMM: a new random-model-based method for solving ab initio crystal structure using the EXPO package. J. Appl. Cryst.* **46,**

Cascarano, G., Giacovazzo, Camalli, M., Spagna, R., Burla, M. C, Nunzi, A. & Polidori, G. (1984). *The method of representations of structure seminvariants. The strengthening of triplet relationships. Acta Cryst.* A**40**, 278-283.

Cascarano G., Giacovazzo C. & Guagliardi A. (1992). *Improved figures of merit for direct methods. Acta Cryst.* A**48**, 859-865.

Giacovazzo, C. (1998). *Direct Phasing in Crystallography*. Oxford: IUCr/Oxford University Press.

Wilson, A. J. C. (1942). *Determination of Absolute from Relative X-Ray Intensity Data. Nature, Load.*

150, 152.

All steps of structure solution procedure

The steps of structure solution procedure

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)Estimate of structure invariants (`%invariants` command)
4)Application of the tangent formula (`%phase` command)
5)Crystal structure completion and refinement (`%fourier` command)

Extraction of the integrated intensities

Extraction of the integrated intensities ('extraction' command)

To calculate the integrated intensities, the program uses the Le Bail algorithm (Le Bail *et al.,* 1988) and combines it with a least squares procedure minimizing the quantity:

$$\Sigma_i\, w_i\, (y_{0i} - y_{ci})^2 \qquad\qquad\qquad\qquad\qquad (1)$$

The summation in (1) goes over all the data points,
$w_i = 1\,/y_i^2$
y_{oi} = observed intensity at the i-th step
y_{ci} = calculated intensity at the i-th step

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

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 is described by the polynomial function (Young, 1993)

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

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

$2\theta_M$ is the 2θ associated to the minimum intensity in the range in which the background is estimated.
From the pattern a standard 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:

$$\frac{\Gamma(\beta_h)}{\Gamma(\beta_h - 0.5)} \left(\frac{C_2}{M}\right)^{1/2} \frac{2}{H_h} \left(1 + 4C_2\, X_{i_h}^2\right)^{-\beta_h} \quad \textit{Pearson VII}$$

$$\frac{C_0^{1/2}}{\sqrt{\pi H_h}} \exp\left(-C_0 X_{i_h}^2\right)$$

Gaussian

$$\frac{C_1^{1/2}}{\pi H'_h} \left(1 + C_1 X_{i_h}^2\right)^{-1}$$

Lorentzian

$$\eta \frac{C_1^{1/2}}{\pi H_h} \left(1 + C_1 X_{i_h}^2\right)^{-1} + (1 - \eta_h) \frac{C_0^{1/2}}{\pi^{1/2} H_h} \exp\left(-C_0 X_{i_h}^2\right)$$

with $0 \le \eta \le 1$

Pseudo-Voigt

where

$$C_0 = 4 \ln 2 \quad C_1 = 4$$

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

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

$$\eta_h = \eta_0 + \eta_1 \cdot tg\,\theta_h + \eta_2 \cdot tg^2\,\theta_h$$

$$\beta_h = \beta_0 + \frac{\beta_1}{2\theta_h} + \frac{\beta_2}{(2\theta_h)^2}$$

The default is the Pearson VII function. In this case

$$A\left(2\theta_i - 2\theta_{\mathbf{h}}\right) = \frac{\alpha \cdot (2\theta_i - 2\theta_{\mathbf{h}})^3}{\left(\frac{H_{\mathbf{h}}^2}{C_{2_{\mathbf{h}}}} + (2\theta_i - 2\theta_{\mathbf{h}})^2\right)^{3/2}} + 1$$

otherwise

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

The peak shape, the asymmetry, the cell and the background parameters are all refined by the program. To overcome the tendency of the Le Bail algorithm to equipartition the intensities of strongly overlapping reflections, a new approach was introduced. It is based on the random variation of the reflection intensities belonging to a given cluster and on the subsequent selection of the good fit feasible trials.

After a first default run of EXTRA, it is possible to run it again starting from the refined variables in the previous run. In this case the preliminary standard peak procedure and the interval by interval extraction procedure are skipped. The related command is % RECYCLE.

<div align="center">

Directives in extraction command

</div>

The following directives must be added after the command '%extraction' in the input file to activate some specific features of the procedure

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.

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).

FAST

This directive is used if the graphic interface has been requested by the "window" command. In this case, the user graphic interaction in a default run is avoided.

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 .

FRECYCLE *filename.ext*

If the "filename" is available from a previous run of EXPO2014 (or EXTRA), or from any source containing hkl and the corresponding F-squared values, Le Bail algorithm is started again for obtaining new estimates.

GAUSSIAN

To use only if the gaussian function is preferred.

INTERVAL *n*

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

LORENTZIAN

To use only if the lorentzian function is preferred.

LOWB

To reduce the background by a constant factor equal to the minimum value observed count.

NCYCLE *n*

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

NOBA

To not refined the background

NOCELL

To skip the cell refinement.

NOSINGLE

This directive is used if the preliminary fitting of the standard peak (automatically selected by the program) is omitted. In this case a preliminary fitting of the first interval is carried out.

OPTI

To continue the refinement procedure for each interval.

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 selected by the program.

PVOIGT

To use only if the Pseudo-Voigt function is preferred.

RANDOM *p*

To activate the random procedure in the integrated intensity extraction step. *p* is the coefficient for the reflection overlapping definition. Default value = 0.5

SMOO

To activate the procedure for calculating the background by following (Brückner, 2000). In this case the background is not refined.

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.

SLOW

This directive can be used if the graphic interface has been requested by the "window" command. In this case the program stands by after some relevant calculations and waits for the user graphic interaction to go on.

VERBOSE

This directive is used 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.

WIDTH *x*

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

ZEROPROFILE *x*

x is the 2q shift value supplied by the user (to be substracted from the 2q observed positions). In this case the automatic calculation of the 2q shift using the preliminary fitting of the standard peak is skipped by the program).

EXAMPLES

Example 1

A Pseudo-Voigt function will be used and the background in each interval will be described by a constant. The counts file format is not the default one and the graphic window is suppressed.

```
%structure mes
%job MES - data from home diffractometer
%initialize
%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
  filetype double
%extraction
  pvoigt
  polynomial 0
%end
```

Example 2

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

```
%structure yono
%job YONO - Synchrotron data
%initialize
%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
    fstarting yono.ext
    ncycle 40
%end
```

Example 3

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

```
%window
%structure agpz
%job AGPZ - data from home diffractometer
%init
%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
```

Example 4

Neutron data and a deuterium atom in the cell content are used. Fast graphic interaction for the integrated intensities extraction process is activated.

```
%structure baco
%job BACO - Neutron data
%initialize
%data
    neutron
    isotope H 0.67
    pattern baco.pow
    cell 10.0632 7.9330 6.8487 90.0 122.3423 90.0
    spacegroup C 2/m
    content Ba 4 C 8 O 20 H 8
    wavelength 1.909
%extraction
    fast
%continue
```

References

Le Bail, A., Duroy, H., Fourquet, J.L. (1988_). Ab-initio structure determination of LiSbWO$_6$ by X-ray_
powder diffraction. Math. Res. Bull. **23**, 447-452.

Normalization of the integrated intensities

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

In this module diffraction intensities are normalized using the Wilson method. 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. 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.

Directives in normal command

The following directives must be added after the command '%normal' in the input file to activate some specific features of the procedure

BFACTOR *x*

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

NOPREFERRED

If this directive is given the program does not apply the correction for preferred orientation.

NOPSEUDO

If this directive is given the program does not recycle the PSEUDO information in a new extraction process.

NREFLECTIONS *n*

Number of active reflections with 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 smallest E-values (maximum E-value 0.3) to calculate psi-zero triplets. 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 not be 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 the n(1,1) can be set to the class number and the other values must be omitted.

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

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

where m is an integer number.

THRP *x*

Threshold value of the mean fractionary scattering power (M.F.S.P.) to recycle the pseudo-translational symmetry information in a new extraction process.

EXAMPLE

Example 1

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

```
%structure agpz
%job AGPZ - data from home diffractometer
%initialize
%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
   filetype xy
%extraction
%normal
   nopseudo
%continue
```

References

Cascarano G., Giacovazzo C. & Luic' M. (1988a). *Acta Cryst.* **A44**, 176-183.

Cascarano G., Giacovazzo C. & Luic' M. (1988b). *Acta Cryst.* **A44**, 183-188

Fan H., Yao J. & Qian J. (1988). *Acta Cryst.* **A44**, 688-691.

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.

Estimate of structure invariants

Estimate of structure invariants ('%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). 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.
Active triplets may be estimated according to Cochran's (1955) P3 distribution: the concentration parameter of the von Mises distribution is then:

$$C = \frac{2 \cdot |E_h \cdot E_k \cdot E_{h-k}|}{\sqrt{N}} \tag{3}$$

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

$$G = C\,(1 + q) \tag{4}$$

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 'NORMAL module' is available. Then triplet phases are no longer expected to be around zero (see quoted references) and may lie anywhere between 0 and 2p.
The method of the joint probability distribution functions (Cascarano *et al.,* 1991; Carrozzini *et al.,* 1997) can be applied to obtain an estimate of normalized structure factors, that will be used as a source of supplementary prior information in the extraction process (if directive ESTIMATION is used).

Directives in invariants command

The following directives must be added after the command '%invariants' in the input file to activate some specific features of the procedure

GMIN *x*
Positive triplets with G < x are not actively used.
Default value x=0.6 (in any case 0.2 < x < / = 1.0 is accepted range of value.

COCHRAN

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

CORRECTION *x*

If a PSEUDOTRANSLATION procedure is on, a correction factor depending on x is applied to the probabilistic argument of weak-weak-weak triplets when they are present. The default value is calculated by program.
x = -1 eliminates weak-weak-weak triplets,
x = 1 corresponds to the largest confidence in them.

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.
Default = minimum value of E for the strongest reflections selected by normalization routine.

ESTIMATION

To run the procedure which estimates the |E|'s for all the reflections and use them as a prior information in a new extraction process.
The procedure will not be applied if PSEUDOTRANSLATIONAL information has been recycled.

NPSI0

To not calculate PSI-0 triplets (Default uses them).

NRPSIZERO *n*

The number of weakest reflections for the psi-zero search.
Default = number computed by normalization routine.
The value will substitute the default one in the direct access file.

NRTRIPLETS *n*

The number of strongest reflections for the sigma2 search. (Up to Emin = 1.0)
Default = number computed by normalization routine (nstrong).
The value will substitute the default one in the direct access file.
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 EXPO 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.

YPSIE
To calculate PSI-E triplets (Default doesn't use them).

EXAMPLE

Example 1

Use of the triplets estimation as prior information for a new extraction process.

```
%structure mes
%job MES - data from home diffractometer
%initialize
%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
   filetype double
%extraction
%normal
   nopseudo
%invariants
   estimation
%continue
```

References

Altomare A., Cascarano G., Giacovazzo C. & Viterbo D. (1991). *Acta Cryst.* **A47**, 744-748.

Carrozzini B., Giacovazzo C., Guagliardi A., Rizzi R., Burla M.C. & Polidori G. (1997). *J. Appl. Cryst.* **30**, 92-97

Cascarano G., Giacovazzo C., Burla M.C., Nunzi A. & Polidori G. (1984). *Acta Cryst.* **A40**, 389-394.

Cascarano, G., Giacovazzo, C., Guagliardi, A. & Steadman, N. (1991). *Acta Cryst.* **A47**, 480-484.

Cochran W.(1955). *Acta Cryst.* **8**, 473-478.

Giacovazzo C. (1976). *Acta Cryst.* **A32**, 958-966.

Application of the tangent formula

Application of the tangent formula ('%phase' command)

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 according to

$$\langle \alpha_h \rangle = \sum_j \left[G_j \cdot D_1 (G_j) \cdot D_1 ((\alpha_k)) \right] \cdot D_1 ((\alpha_{h-k})) \quad (5)$$

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

$$D_1(G_j) = \frac{I_1(G_j)}{I_0(G_j)} \quad (6)$$

I_1 and I_0 represent modified Bessel functions of order one and zero respectively. The summation in (5) is over all relationships defining the reflection h. If P3 formula is used the default choice is

$$\langle \alpha_h \rangle = \sum_j \left[C_j \cdot D_1(C_j) \right] \quad (7)$$

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 < a >.

- <u>PHASE EXTENSION AND REFINEMENT</u> -

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, psi-zero triplets and the most reliable two-phase structure seminvariants. In the weighting scheme the experimental distributions of the alpha parameters are forced to match with the theoretical ones (Burla *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 then CFOM > 0.3 may characterize the correct solution.

Directives in phase command

The following directives must be added after the command '%phase' in the input file to activate some specific features of the procedure

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 100). Only if RANDOM directive is on.

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 = 10).

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.

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)

SEED *n*

Random generation seed. n must be an integer large value and must be odd.

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 = 5.

(Not consistent with RANDOM directive)

TABLE

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

TEST

To restart the random number generator using the default seed. Only if RANDOM directive is on.

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).

YNGQ

To actively use negative quartets in the phasing process. (Default does not use them).

EXAMPLE

Example 1

The program restarts from the PHASE procedure.

A random approach is used and the best 100 sets of phases, over 250 trials, are retained in the direct access file. No Fourier is required.

%structure nizr

%job nizr structure

%init

%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

filetype double

synch

%extraction

%normal

%invar
%phase
  random
  maxtrials 250
  minfom 1.0 100
%end

References

Baggio R., Woolfson M.M., Declerq J.P. & Germain G. (1978). *Acta Cryst.* **A34**, 883-892.

Burla M.C., Cascarano G., Giacovazzo C., Nunzi A. & Polidori G. (1987). *Acta Cryst.* **A43**, 370-374.

Burla M.C., Cascarano G. & Giacovazzo C. (1992). *Acta Cryst.* **A48**, 906-912.

Cascarano G., Giacovazzo C. & Viterbo D. (1987). *Acta Cryst.* **A43**, 22-29.

Cascarano G., Giacovazzo C. & Guagliardi A. (1992). *Acta Cryst.* **A48**, 859-865.Germain G., Main P. & Woolfson M.M. (1970). *Acta Cryst.* **B26**, 274-285.

Main P. (1978). *Acta Cryst.* **A34**, 31-38.

White P.S. & Woolfson M.M. (1975). *Acta Cryst.* **A31**, 53-56.

Crystal structure completion and refinement

Crystal structure completion and refinement ('%fourier' command)

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, 1978; Main *et al.* 1980).

Several additional features have been introduced in the present version:
a) From the solution with the highest CFOM value a Fourier map may be calculated by involving a supplemental number of reflections which are phased through the so-called psi-E triplets (Altomare *et al.* 1991)
b) Special positions are handled, peaks very close to symmetry elements are moved onto symmetry elements, the site symmetry is defined and the atomic occupancy factor is calculated; key numbers for designating free, coupled or fixed positional parameters for least-squares subroutines are also calculated together with symmetry conditions on the thermal ellipsoid.
c) Cycles of structure factor calculation -least squares- (2Fo -Fc) Fourier synthesis are automatically performed in order to provide a more complete and refined crystal structure. Due to reflection overlapping the results of the refinement must carefully be considered by the user who will take always into account the ratio 'number of parameters to refine/number of independent observations'.
The program provides the following R-factors which can be used as reliability parameters:

$$R(f) = \frac{\sum |F_0 - F_c|}{\sum |F_0|} \tag{8}$$

$$Rw(f) = \frac{\sum |F_0 - F_c| \cdot F_0^2}{\sum |F_0| \cdot F_0^2} \tag{9}$$

$$R(f^2) = \frac{\sum |F_0^2 - F_c^2|}{\sum |F_0^2|} \tag{10}$$

An alternative approach, 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), has been introduced in EXPO and modified to improve the phasing process from powder data. It combines an Electron Density Modification procedure with Fourier map calculations.

Directives in fourier command

The following directives must be added after the command '%fourier' in the input file to activate some specific features of the procedure

COORDINATION *El dmin dmax n*

The minimum and maximum value of the bond distance are assigned to element El. n is the coordination type.

Allowed values for bond angles are:

n = 1   55 - 180 Degrees
n = 2   85 - 145 Degrees

EDMO

This directive is used to enable a direct space procedure consisting of an Electron Density Modification (EDM) process followed by successive Fourier maps using (2Fobs -Fcalc) coefficients. Except for the Fourier directive 'set', all the rest of Fourier recycling directives are not valid for EDM procedure.

2EOEC

This directive enables to use (2Eobs-Ecalc) as Fourier coefficients in the Fourier step (if EDM procedure is applied).

EOBS

This directive enables to use the observed normalised structure factors (Eobs) as Fourier coefficient in the Fourier step (if EDM procedure is applied).

FRAGMENT *string*

Used to supply a known fragment to be completed. String is the name of the file in which, for each atom, are stored: Element X, Y, Z, isotropic thermal factor.

FWHM *D*

This directive is used in order to change the value of D that is the parameter for the condition $2q_1 - 2q_2 < D*FWHM$.
Default value is approximately 0.9 depending on the ratio between the number of observations and the number of parameters to refine.

GRID *x*

The grid on which the Fourier-map is evaluated will have a spacing of approximately x angstroms. The default spacing is computed by the program (1/3 Å).

INTEGRATION *x n*

x is the radius of integration in the peak search routine. (Default value 0.5). n may be 1 to use the gaussian fitting of the fourier peak in the map or 2 to use the peak average density. (Default value 2).

LAYX

This directive can be used only if MAP directive is given. The Fourier-map will be printed in section of constant x.

LAYY

This directive can be used only if MAP directive is given. The Fourier-map will be printed in section of constant y.

LAYZ

This directive can be used only if MAP directive is given. The Fourier-map will be printed in section of constant z.

LEVEL *n*

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

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

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

LIMITS *x1 x2 x3*

The E-map will be printed 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 MAP directive).

MAP

If MAP is specified, the E-map will be printed on the lineprinter. 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 map. The default is the number of peaks which

$$\text{å}\,(occ(i)\text{*}m) = 1.3 \text{ * } n$$

where occ is the crystallographic site occupancy factor, m is the number of symmetry operators, n is the number of non-hydrogen atoms in the unit cell.

PSE

To actively use PSI-E triplets in these procedures (Default doesn't use them). (This directive can be used only if associated to YPSIE directive of INVARIANTS routine).

RADIUS *El x*

The value x of the radius used to compute the connectivity is assigned to element El.

RECYCLE *n*

Used to stop the recycling procedure after cycle number n. If n is equal to zero only the E-map will be calculated.

SET *n*

The serial number of the set from which to calculate the E-map. If 'SET n' is not specified, the set with the highest combined figure of merit will be used, and then if FOURIER routine is re-run, the set with the next highest combined FOM will be used.

SOGF *x*

The percentage of the largest |F| value reflections actively used in the Fourier recycling routine.

The reflections with the smallest |F| values corresponding to the rejected percentage have a zero weight value in the refinement process. (Default value 100.0).

SOGI *x*

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

STOPF *x*

This directive enables to stop the Fourier step at the cycle number x (if EDM procedure is applied). The default value of x is 10 (the maximum allowed value).

THRE *x*

x is the threshold for the E value of the reflections actively used in the Electron Density Modification step. The default value is 0.8.

THRM *x*

x is the fraction of the Fourier map points actively used in the Electron Density Modification step. The default value is 0.025.

THRW *x*

x is the threshold of the Bessel weight of the reflections actively used in the Fourier step (if EDM procedure is applied). The default value is 0.5.

EXAMPLES

Example 1

Only the E-map corresponding to the set number 7 and the complete FOURIER/LEAST-SQUARES procedure for set number 10 are requested by the user. The binary file "vfi.bin" must exist.

```
%structure nbpo
%fourier
   set 7
   recyc 0
%fourier
   set 10
%continue
```

Directives in label command

Direct methods applied to powder diffraction data often provide electron density maps of which the quality is usually affected by systematic and/or random phase errors, by amplitude truncation effects in the series representation of the electron density, etc. The frequent incorrect labelling of the peaks can strongly affect the efficiency of the procedures used for crystal structure refinement. An algorithm has been introduced in EXPO aiming at using the crystal chemical information that is usually available *a priori* to improve the peak labelling (Altomare *et al.,* 2001)

The following directives must be added after the command '%label' in the input file to activate some specific features.

DLIGHT *d tol*

d is the expected heavy atom 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 peaks to relabel using El element;

El is the atomic species to assign to the peaks.

Only 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).

EXAMPLE

Example 1

Use of the label change procedure to relabel the peaks of an electron density map obtained by applying the FOURIER procedure. The binary file "vfi.bin" must exist.

% window

% structure vfi

% job VFI - Synchrotron data

% changelab

    label 3 al 3 p

    coord 4 4

    dwei 3.1 0.5

    dlig 1.9 0.5

% continue

Directives in polyhedra command

The following directives must be added after the command '%polyhedra' in the input file to activate some specific features of the polyhedra procedure

MISSING *n*1 *n2 El tetrahedron (octahedron) d tol1 tol2*

n1 and n2 are the two order numbers (taken from an external file or from a Fourier map) of

the cations between which the missing cation must be located;

El is the missing cation label tetrahedron (octahedron): 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 distance and the tetrahedron (octahedron) angle respectively. Default values are 0.3.

OCTAHEDRON *El d tol1 tol2*

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

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

EXAMPLES

Example 1

The POLPO1 procedure is applied to locate the light atoms around each specified heavy atoms. The crox.fra file contains the cations coordinates.
The binary file 'crox.bin' must exist.

```
% window
%structure crox
%job CROX - home diffractometer data
%polyhedra
    fragment crox.fra
    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
```

Example 2

The POLPO2 procedure is applied to locate one or more missing cations and surronding anions. The sapo.fra file contains the located cations coordinates.
The binary file 'sapo.bin' must exist.

```
% window
%structure sapo
%job SAPO - data from home diffractometer
%polyhedra
    fragment sapo.fra
    tetrahedron Si 1 1.6 0.2 0.2
    tetrahedron Si 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

Directives in Patterson command

The following directives must be added after the command '%patterson' in the input file to activate some specific features of the patterson procedure.

E**2 (or E*E)

Coefficients to be used in the Patterson synthesis. This is the default.

E*F (or F*E)

Coefficients to be used in the Patterson synthesis.

F**2 (or F*F)

Coefficients to be used in the Patterson synthesis.

INVERSE

To activate the procedure of calculating the |F| values from an inverted suitably modified Patterson map to use as starting point for the Le Bail algorithm in the extraction routine.

MAP, LAYX, LAYY, LAYZ, PEAKS, LIMIT, GRID: See FOURIER directives.

NPAT _n_

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

References

Altomare A., Cascarano G., Giacovazzo C. & Viterbo D. (1991). *Acta Cryst.* **A47**, 744-748.
Main P. (1978). *Acta Cryst.* **A34**, 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.* **A33**, 977-979.

Structure solution by Direct Space methods

Direct Space algorithms in EXPO2014

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. 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 and angles are usually known and kept fixed while only the torsion angles are varied during the procedure.

Two Direct Space algorithms are now 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.

The procedure is supported by the same graphic interface of Simulated Annealing and 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 other cases the classical SA approach is used.

[Preparing Input file](#)

[Graphic Interface of Simulated Annealing](#)

[Output file](#)

[When Direct Space approach fails](#)

[The directives in Simulated Annealing](#)

[Contact and References](#)

Preparing Input file

To run EXPO2014 for structure solution by Direct Space approaches, first you need to create the input file. 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 check button 'Structure Solution', select 'Simulated Annealing' and import the starting model from 'fragment filename'. The following picture is an example for the crystal structure determination of the paracetamol molecule. The profile counts file (paracetamol.xy) is in the directory 'examples'.

When you press the button 'Save' an input file paracetamol.exp (shown below) will be created and automatically loaded by the program for the structure solution process by Simulated Annealing.

```
%Structure paracetamol
%Initialize
%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
```

Otherwise you can import an existing input file by the button 'Open' on the 'New Project' dialog window or selecting 'File' > 'Load & GO' in the main menu. For example you could load the file `paracetamol.exp` already existing in the example folder.
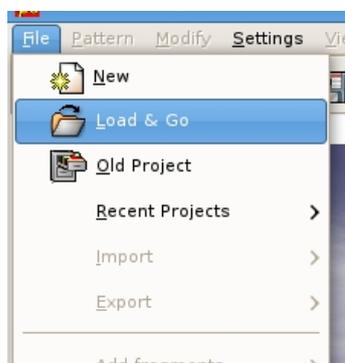
`%sannel` is a command to access to the Simulated Annealing graphic interface.

`%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 <u>Usage of command `%fragment`</u>). 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). A complete list of the supported input file can be find in section <u>Import</u> 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**.
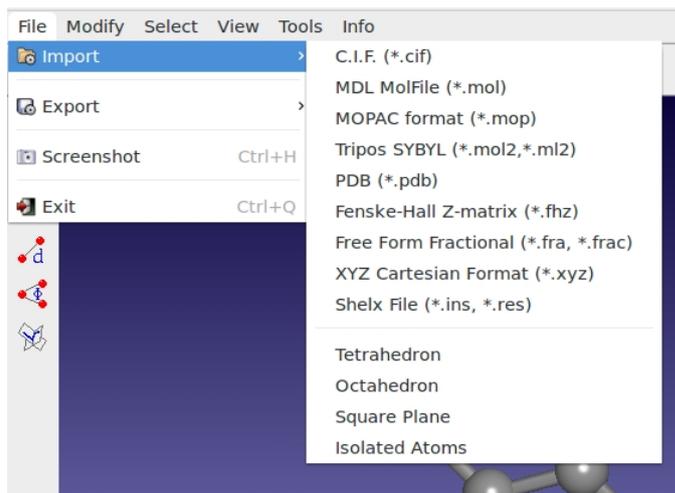
Modify > Add fragments to add fragment to an existing partial structural model.

Instead of command `%sannel`, you can access directly to the graphic interface from menu "Solve" > "Simulated Annealing".
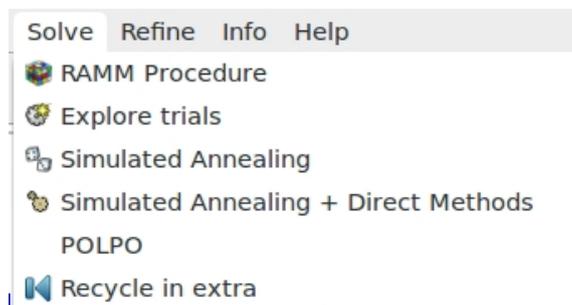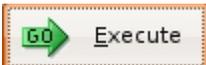
Press the button in the dialog window to run the Simulated Annealing procedure.

Consider the following two strategies on the preparing starting model for direct-space solution.

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 literature. If a new struc

Some free database where look up molecules are:

1. NIST Chemistry WebBook: http://webbook.nist.gov/chemistry/

2. Drugbank: http://www.drugbank.ca/

3. PubChem: https://pubchem.ncbi.nlm.nih.gov/

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 by **File > Import Fragment > From SMILES** or **Modify > Add Fragments > From SMILES**

2) Starting molecular model can be created by geometry optimization using quantum-chemistry package, e.g., MOPAC, Gaussian, Gamess, NWChem. Some free available software that can be used to sketch molecules, optimize the geometry by force field 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.

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:

or *H*BB-BC algorithm:

Cost function: 2 cost functions can be selected: R weighted profile (default cost function), R-Bragg intensities.

Resolution: defines the maximum resolution used by Simulated Annealing procedure.

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):
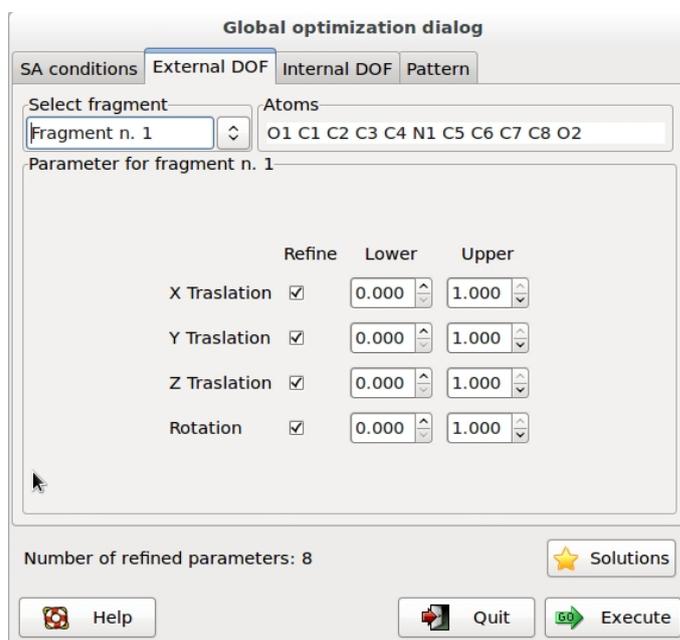
Select fragment: selects the fragment and visualizes the corresponding structure information.

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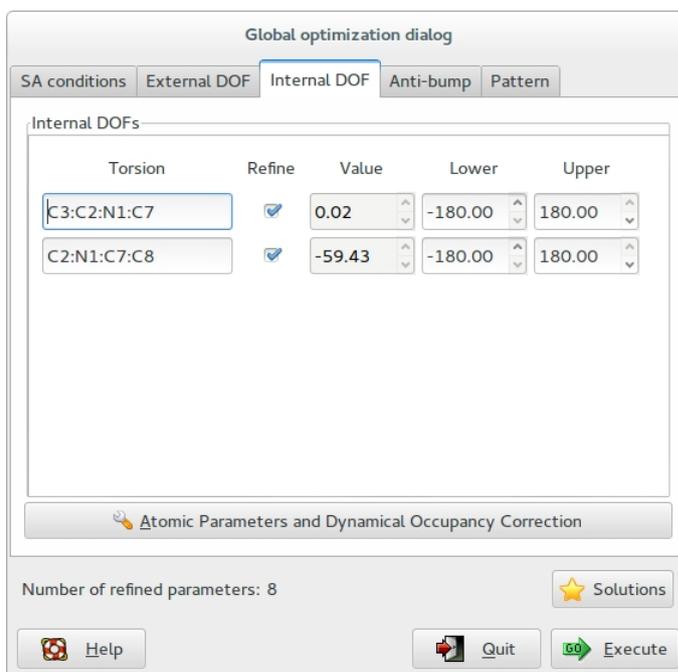
Internal DOFs: list of the torsions associated to each refinable internal DOF. Check the parameters to refine, enter the lower and upper bounds of parameters.

Press the button '**Atomic parameters**' to access to the following dialog window:
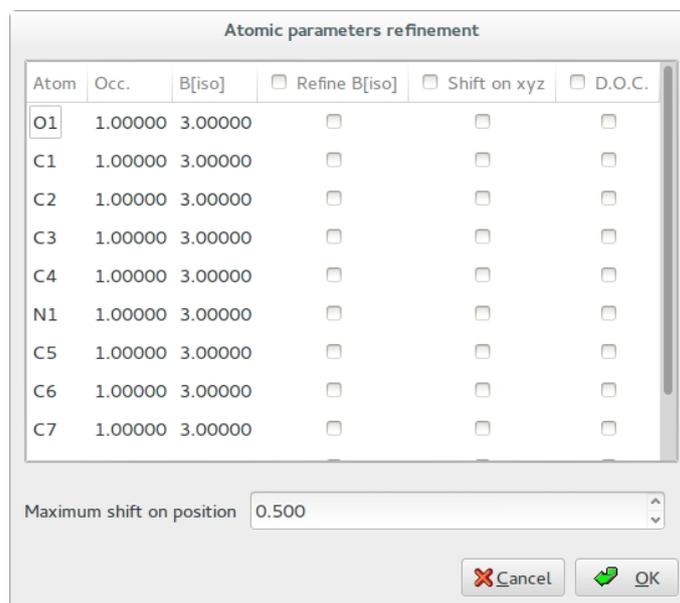
Thermal parameters: check the buttons on column '**B[iso]**' to refine the thermal parameters of atoms. Click on label '**B[iso]**' to select all checks in the column. If the thermal parameters are not indicated in the imported fragment file, the program assigns the default values of B=3.0 for non-hydrogen 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 on the JAV viewer.
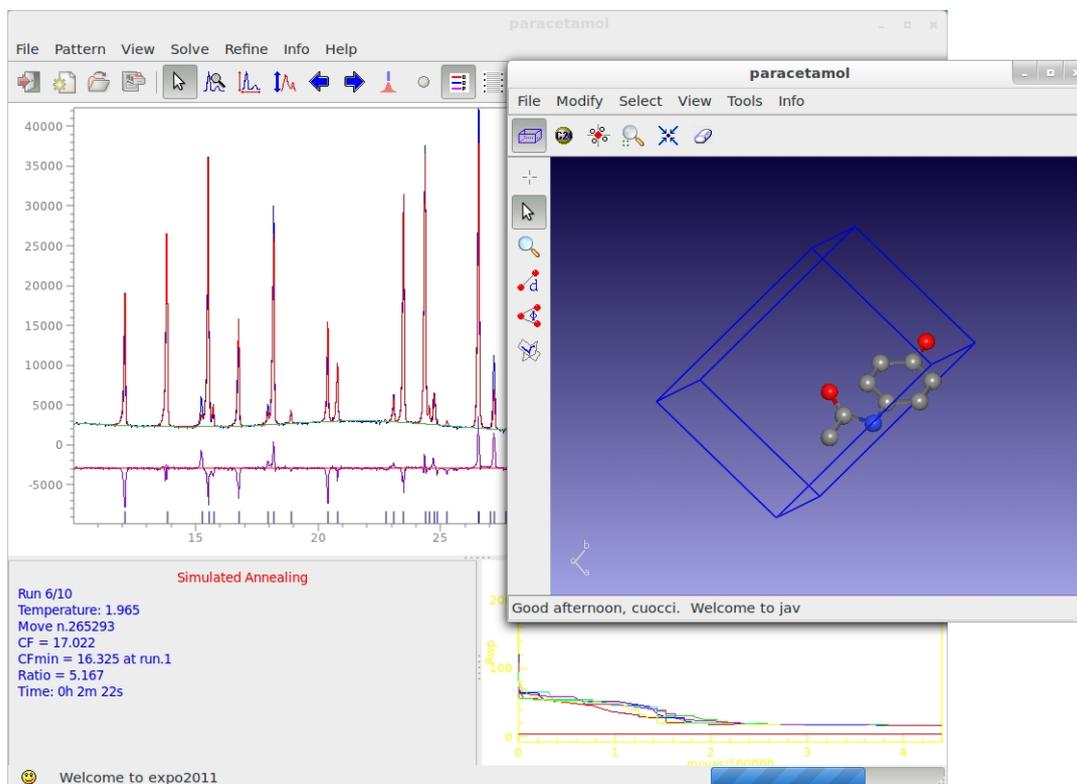
During the Direct Space runs, three buttons are active on the toolbar of the main window:
 opens the JAV viewer for crystal structure visualization.
 skips the current run.

🖐 stops the procedure.

Output file

The output file generated from the procedure, contains information on:
1) the starting structure
2) the volume per atom
2) the connectivity
3) the internal and external DOFs
4) the refined parameters by SA
5) each SA or HBB-BC run
6) the summary of all the SA or HBB-BC runs
7) the selected final structure

At the end of each run the structure coordinates are saved in CIF file with name created by project name with suffix _best1,_best2, ... (e.g. 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 Direct Space approach fails

When Direct Space approaches fail:

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

2) 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 for atom in the output file (about 15-20 Ang/atom).

3) The assumption about thermal factors is invalid. Check thermal factors from similar structures.

4) 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.

5) The default conditions, for complex structure, cannot be appropriate. Increase the number of moves and/or runs, try with slower temperature reduction.

The directives in Simulated Annealing

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`.

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".

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

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

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.

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` directive is used the specified rotation will be included in the panel 'Internal DOF' of the graphical interface.

E.g.
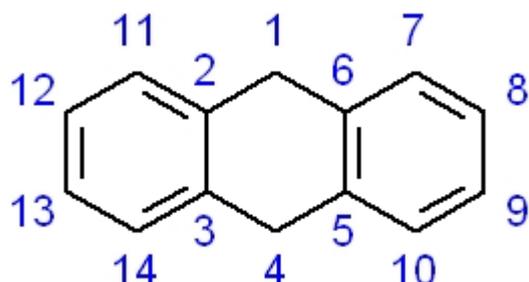

```
rotate C1 C4 C5 C6 C7 C8 C9 C10
```

the phenyl ring C5-C6-C7-C8-C9-C10 will be rigidly rotated around the axis C1-C4

bump
Automatic generation of anti-bumping restraints extended to all C, N, O and S atoms

res `atom1 atom2 target_dist  tol weight`
Apply restraints between 2 atoms. `atom1` and `atom2` are the labels of atoms. The other parameters are optional and can be omitted.
`target_dist` is the ideal distance between the pair of atoms, when omitted the distance is automatically deduced by using an internal table of distances.
`tol` is a permitted tolerance, when omitted the default value is 0.2
`weight` is a user supplied weight. A default weight is specified as 100.

fixrotation `atom1 atom2`
To fix the internal DOF. `atom1` and `atom2` are the 2 atoms defining 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 in conjugated acyclic systems, triple bonds, nitrilic triple bonds, etc.

fftranslation `atom1 atom2 ....`
To fix the translation of fragments containing `atom1, atom2,` ... It is enough to specify at least one of atom belonging to the molecular fragment.

E.g.

```
fftranslation Ag1 Ag2
```

If `Ag1` and `Ag2` are in 2 different fragment, the traslation of fragment containing `Ag1` will be fixed and the traslation of other one will be fixed.

ffrotation `atom1 atom2 ...`
To fix the rotation of fragments containing `atom1, atom2,` ... It is enough to specify at least one of atom belonging to the molecular fragment.

refinetf
To refine the thermal parameters.

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.

cost_function n

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

randomize n

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

rotat n1 n2

To specify the internal DOF using the input file.

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.

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

Usage of command `%fragment`

The command `%fragment` can be used to import molecular starting model. This command 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 use 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)
%Initialize
%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).

Contact

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

References

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

Direct Methods combined with Simulated Annealing

Direct Methods combined with Simulated Annealing in the EXPO2014 program

The information contained in the electron density maps provided by Direct Methods has been combined with the Simulated Annealing technique in order to solve organic crystal structures from powder diffraction data. In this procedure each peak of the Fourier map is in turn associated to all the atoms of the imported model (Altomare *et al.,* 2008). For each association (the atom of the structural model is moved on the peak) the position

of the model is fixed, while its orientation and internal conformation have to be found by SA runs.

Run of Direct Methods combined with Simulated Annealing

After a run of EXPO2014, the procedure can be graphically activated by 'Simulated Annealing+Direct Methods' option of 'Solve' button in the upper menu of the EXPO2014 window.

Graphic Interface of Direct Methods combined with Simulated Annealing

The dialog window of Direct Methods combined with Simulated Annealing procedure in EXPO2014 is composed by 3 pages containing the general settings of the procedure.

The first page '**SA conditions**' contains the general settings of the algorithm:
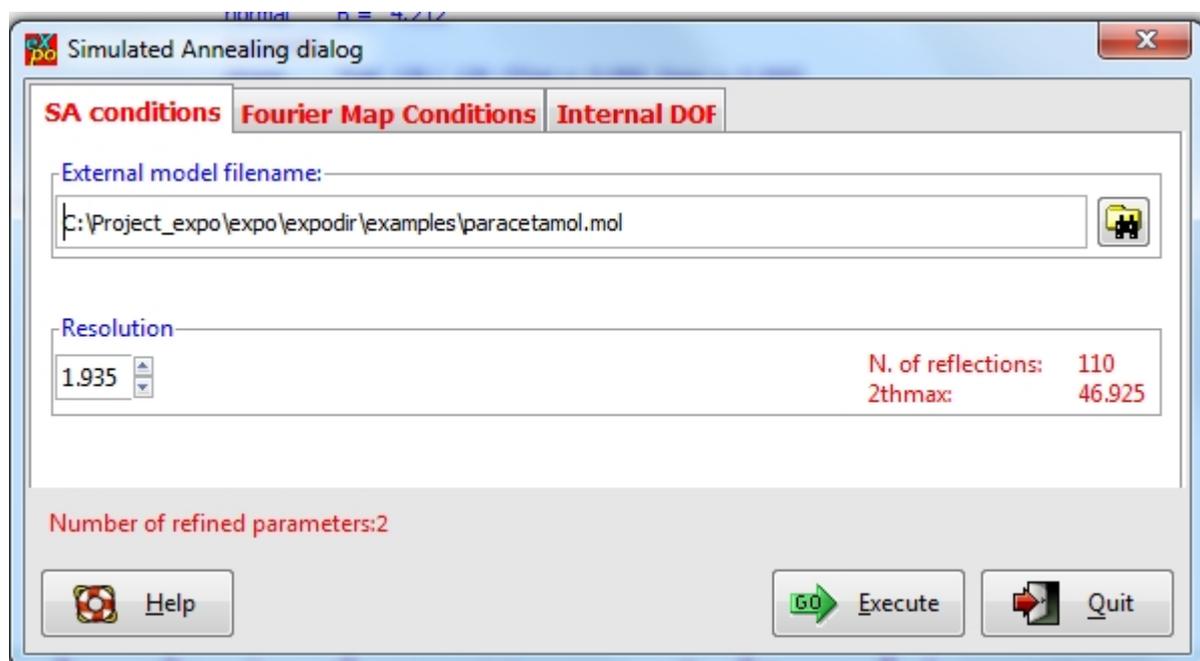
External model filename: the name of the imported starting structural model must be

specified. Some file types can be imported: C.I.F. file (*.cif), Expo fragment or Free Fractional Format (*.fra), XYZ format (*.xyz), MOPAC file (*.mop), MDL MolFile (*.mol), Protein Data Bank file (*.pdb), Tripos SYBYL (*.mol2, *.ml2), Fenske-Hall Z-Matrix (*.zmt) and Shelx File (*.ins).

Refer to openbabel wiki or http://www.ch.ic.ac.uk/chemime for informations about common molecular file formats.

Refer to http://en.wikipedia.org/wiki/Molecule_editor and http://www.ccp14.ac.uk/solution/2d_3d_model_builders/index.html for an exhaustive list of programs for building 3D starting models. Suggested free available programs for building 3D models are ACD/ChemSketch (Windows), Avogadro (all platforms), Marvin (all platforms), Ghemical (Linux). NIST Chemistry WebBook and Drugbank are useful links to look up fragments.

Resolution: defines the maximum resolution used by Simulated Annealing + Direct Methods procedure.

The second page '**Fourier map condition**' contains the information about the Fourier map at which the imported starting model will be associated:
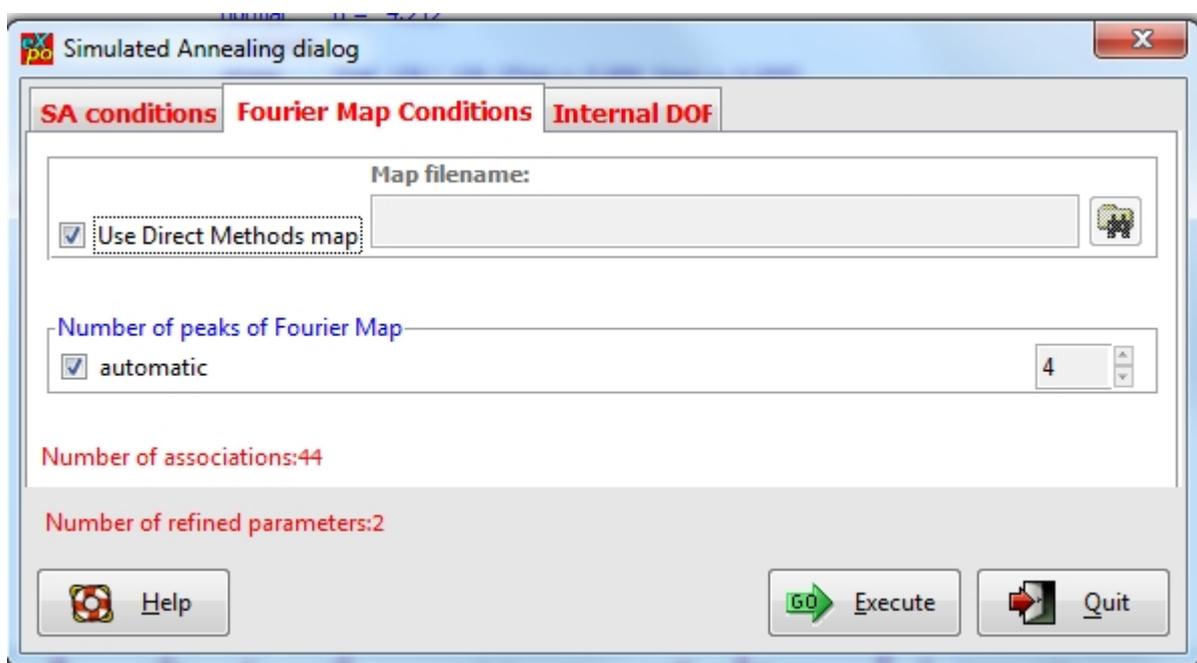
Use Direct Methods map: if selected, the program will use the atomic positions of the Fourier map obtained at the end of the current run of EXPO2014.

Map filename: it is activated only if 'Use Direct Methods map' is de-selected. The filename of the external Fourier map (in term of atomic positions) must be supply. Some file types can be imported: C.I.F. file (*.cif), Expo fragment or Free Fractional Format (*.fra), etc.

Number of peaks of Fourier Map: the number of atomic positions selected from the map. If 'automatic' is selected, the program will use the 40% of the electron density peaks (those with the largest intensity) to associate in turn to each model atom. This number can be reduced by the user if 'automatic' is de-selected.

The third page '**Internal DOF**' contains the information about the internal degrees of freedom (DOFs):

Torsion: List of torsions associated to each refinable internal DOF. The order number of each atom involved in the torsion angle to vary are also shown. The internal DOFs can be or not refined (check Refine).

Lower-Upper: range of variation of the torsion angle. 180 means that the allowed range is (x-180, x+180), where x is the value of the torsion angle in the initial model.

A visual match between the observed and calculated powder diffraction pattern is plotted when the procedure is running. When Direct Methods combined with Simulated Annealing procedure are running the progress of structure solution is monitored and the user can examine:
1) the graph of the minimum value of the cost function (CF) vs. the number of moves;
2) the match between the observed and calculated pattern.

Use the buttons ⌕ ⌕ ⌕ to zoom on graph.

During the Direct Methods combined with Simulated Annealing procedure runs two buttons on the toolbar are activated:

⬡ opens the JAV viewer for crystal structure visualization.

✋ stops the procedure.

Press the button ⟦GO⟧ Execute to run the Simulated Annealing+Direct Methods procedure.

Output of Direct Methods combined with Simulated Annealing

The output file contains information on:
1) the selected peaks of the Fourier Map imported or obtained at the end of a run of EXPO2014;
2) the imported starting structural model;
3) the table of connectivity;

4) the list of internal degree of freedoms (DOFs) to be varied;

6) the information about each association: peak of the map-atom of the model;

7) the summary of the ten best associations which are refined in sequence;

8) the selected best final structure.

About the construction of the models

Refer to openbabel wiki or http://www.ch.ic.ac.uk/chemime for information about common molecular file formats.

Refer to http://en.wikipedia.org/wiki/Molecule_editor and http://www.ccp14.ac.uk/solution/2d_3d_model_builders/index.html for an exhaustive list of programs for building 3d molecules. Suggested free available programs for building 3D molecules are ACD/ChemSketch (windows), Avogadro (linux), Marvin (all platforms), Ghemical (linux) NIST Chemisty webbook and Drugbank.

Contact

For suggestions and bugs please contact rosanna.rizzi@ic.cnr.it.

References

A. Altomare, R. Caliandro, C. Cuocci, C. Giacovazzo, A. G. G. Moliterni, R. Rizzi and C. Platteau. *'Direct methods and simulated annealing: a hybrid approach for powder diffraction data'. J. Appl. Cryst. (2008).* **41**, 56-61.

Rietveld Refinement

Rietveld Refinement in EXPO2014

Structure refinement by Rietveld method (Rietveld, 1969) is available in 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, 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 the 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, Thompson-Cox-Hastings 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 χ^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 (R_b) which use the structure factor moduli extracted from the experimental profile. In addition, graphical tools are available for visualizing: 1) the observed, calculated and difference pattern and the cumulative χ^2 value; 2) the evolution of the structure model during the refinement in real time.

$$R_p = \frac{\sum_i^N |y_i - y_{ci}|}{\sum_i y_i}$$

$$R_{pw} = \left[\frac{\sum_i^N w_i (y_i - y_{ci})^2}{\sum_i^N w_i y_i^2} \right]^{1/2}$$

$$R'_p = \frac{\sum_i^N |y_i - y_{ci}|}{\sum_i^N (y_i - y_{bi})} \qquad\qquad R'_{pw} = \left[\frac{\sum_i^N w_i\left(y_i - y_{ci}\right)^2}{\sum_i^N w_i\left(y_i - y_{bi}\right)^2}\right]^{1/2}$$

$$R'_{p\,p} = \frac{\sum_i^{N_\rho} |y_i - y_{ci}|}{\sum_i^{N_\rho} (y_i - y_{bi})} \qquad\qquad R'_{pw\,\rho} = \left[\frac{\sum_i^{N_\rho} w_i\left(y_i - y_{ci}\right)^2}{\sum_i^{N_\rho} w_i\left(y_i - y_{bi}\right)^2}\right]^{1/2}$$

$$R_{\text{exp}} = \left[\frac{(N-P)}{\sum_i^N w_i y_i^2}\right]^{1/2} \qquad\qquad R'_{\text{exp}} = \left[\frac{(N-P)}{\sum_i^N w_i\left(y_i - y_{bi}\right)^2}\right]^{1/2}$$

where:
N is the total number of counts in the pattern and P is the number of refined parameters.

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
%Initialize
%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 a command to access to the Rietveld Refinement graphical interface.
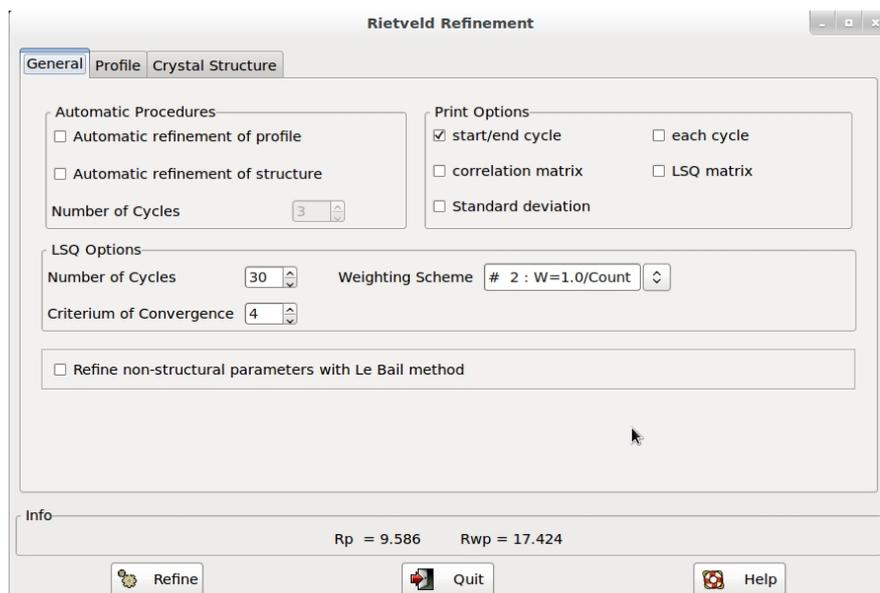
Alternatively, when a structure model is available, you can access directly to the graphic interface from menu "Refine" > "Rietveld"
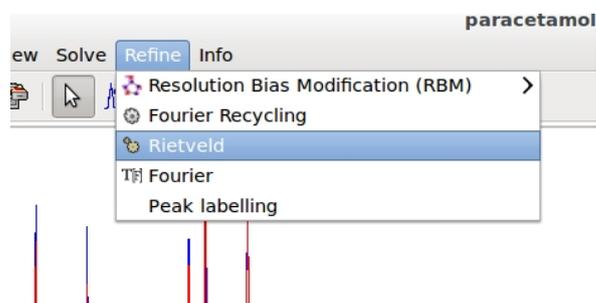
Constraints

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 deduce and impose 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, occA and occB 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 'Crystal Structure' section is possible to access to a dialogue window to define group of ADPs that can have the same shift.
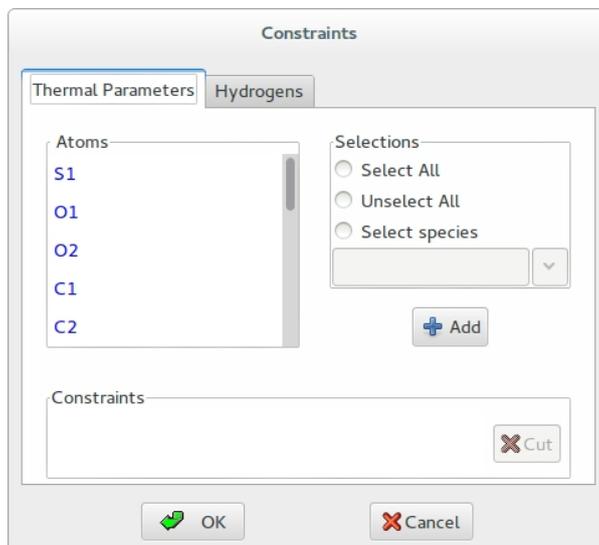
Contraints are applied to H atoms by using the **riding model**. This technique is used to move H atoms synchronously with the C atoms to which they are bonded, thereby preserving the bond length and direction. The isotropic displacement parameters Uiso of the hydrogens are constrained to be 1.2 times of that of the heavy atom to which they are attached. The factor 1.2 can be changed by graphic interface.
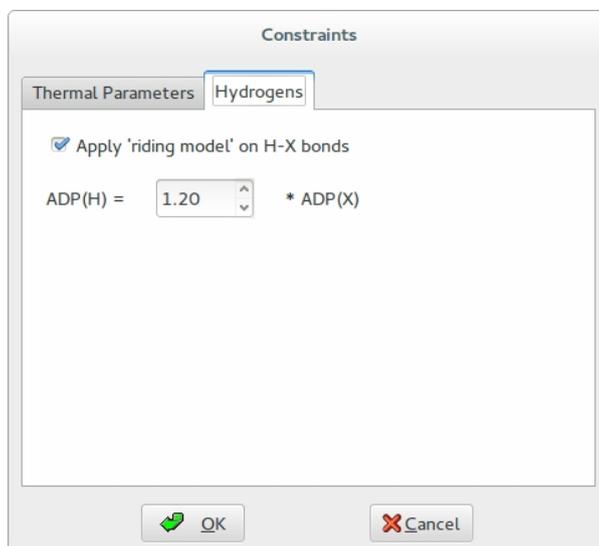
Contact

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

Rietveld, H.M.: A profile refinement method for nuclear and magnetic structures. *J. Appl. Cryst.* **2** (1969) 65–71.
McCusker, L.B.; Von Dreele, R.B.; Cox, D.E.; Louër, D.; Scardi, P.: Rietveld refinement guidelines. *J. Appl. Cryst.* (1999). **32**, 36-50
The Rietveld Method. Edited by R. A. Young, Oxford University Press, Oxford (1995)

Commands of EXPO2014 and their use

Commands of EXPO2014 and their use

The input file name for EXPO2014 is constituted by the name of the structure to analyze followed by the suffix *.exp*, i.e.

name.exp

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

For the default run see Example 1.

EXPO recognizes the following commands:

%WINDOW	Graphic window is required
%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 files adding the appropriate extension to the structure name. The file names are: string.bin -> binary file string.out -> output file If this command is not used the default string "STRUCT" (instead of the name of the structure) is used to create file names
%JOB caption	A caption is printed in the output
%INITIALIZE	Initialize the direct access file (to override previous results and data)
%DATA	Data input routine
%NTREOR	Indexing routine
%EXTRACTION	Integrated intensities extraction routine
%NORMAL	Normalization routine

%INVARIANTS	Invariants routine
%PHASE	Converge-tangent routine
%FOURIER	Fourier/Least-Squares routine
%MENU	
%END	End of the input file
%CONTINUE	The program runs in default conditions from the last given command up to the end
%PATTERSON	Patterson map calculation routine
%REFINE	Fourier/Least-Squares restart routine
%RECYCLE	Extraction recycle routine
%SANNEL	direct space routines
%SDIRECT	Direct Methods combined with direct space routine
%EXPORT filename.xxx	Export files (xxx=hkl,cif,xyz,frac)

Directives are described below, in the sections dedicated to the various routines.

All commands and directives are in free format (between columns 1-80) and are cases independent. Only the first four characters are significant. The keywords can start in any position. If the first non-blank character is ">", then the record is interpreted as a comment. EXPO2014 preserves intermediate results in the file named **name.bin** (*binary file or project file*). For example, if invariant estimates have been already obtained during a previous run of EXPO2014, in a new run the command %INVARIANTS can be omitted.

Commands can be given in any order, under the following conditions:

- first routine used must be DATA, if it has not been used in a previous run;

- EXTRACTION routine must be used to extract integrated intensities from the diffraction pattern, if a reflections file is not supplied by the user;

- NORMAL must be used if reflections file has been supplied or EXTRACTION step has been performed;

- INVARIANT routine has no meaning if diffraction moduli are not normalized;

- PHASE routine has no meaning if no triplets have been calculated;

- FOURIER routine cannot run before PHASE routine (unless a fragment information is supplied by the user; in this case only DATA, EXTRACTION and FOURIER commands are needed).

The minimal information needed by EXPO2014 for a full run (from indexing to structure solution) is constituted by (see Example 1):

- cell parameters (not necessary if N-TREOR program is used)

- space group symbol (not necessary if the Space Group Determination procedure is applied by using "findspace" directive)

- cell content

- diffraction pattern (or reflections if an external reflection file is provided by any profile fitting program)

- wavelength

- type of radiation (if neutron or synchrotron data have been used; home diffractometer data is the default).

EXPO2014 prints results in a file named **name**.out.

Add '%EXPORT filename.xxx' in the input file to export files. Filename is the name of file to export and xxx define the type of file. Use xxx=cif, xyz, frac to export structure model respectively in cif file, xyz cartesian coordinates, fractional coordinates; use xxx=hkl to export file with hkl, extracted structure factors, calculated structure factors, phases from the obtained model.

In the following example, the command export before '%data' is used to export hkl and cif file and it will be processed only at the end of structure solution process with Direct Methods.

```
%structure cime
%job structure cime
%init
%export cime.hkl
%export cime.cif
%data
pattern cime.pow
cont s 4 c 40 n 24 h 64
wave 1.52904
cell    10.6986  18.8181 6.8246      90.000  111.284   90.000
space p 21/n
sync
%continue
```

An other example in case of structure solution with Simulated Annealing

```
%Structure paracetamol
%Job paracetamol
%Initialize%Initialize
%export paracetamol.hkl
%export paracetamol.cif
%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
%continue

Description of EXPO2014 commands

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:

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

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 < a >.

- 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.* **A47**, 744-748.

Altomare, A., Campi, G., Cuocci, C., Eriksson, L., Giacovazzo, C., Moliterni, A., Rizzi, R. & Werner, P.-E. (2009). *J. Appl. Cryst.* **42**, 768-775.35,182-184.

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.* **A34**, 883-892.

Burla M.C., Cascarano G. & Giacovazzo C. (1992). *Acta Cryst.* **A48**, 906-912

Burzlaff H. & Hountas A. (1982). *J. Appl. Cryst.* **15**, 464-467.

Cascarano G., Giacovazzo C. & Luic' M. (1988a). *Acta Cryst.* **A44**, 176-183.

Cascarano G., Giacovazzo C. & Luic' M. (1988b). *Acta Cryst.* **A44**, 183-188

Fan H., Yao J. & Qian J. (1988). *Acta Cryst.* **A44**, 688-691.

Germain G., Main P. & Woolfson M.M. (1970). *Acta Cryst.* **B26**, 274-285.

Giacovazzo C. (1976). *Acta Cryst.* **A32**, 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.* **A34**, 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.* **A33**, 977-979.

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

White P.S. & Woolfson M.M. (1975). *Acta Cryst.* **A31**, 53-56.

Wilson, A. J. C. (1942). *Determination of Absolute from Relative X-Ray Intensity Data. Nature, Load.*

150, 152.

Preparation of data

Preparation of data for the DATA routine

This routine reads the basic crystallographic information 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).

ALPHA

It is used to start the alpha2 stripping.

CELL *a b c* a b g

cell dimensions a, b, and c are in angstroms, a , b and g in degrees.

CONTENTS *El1 n1 El2 n2 El3 n3 ....*

Unit cell contents. Eli is the chemical symbol of atomic species i, ni is the corresponding number of atoms in the unit cell (to a maximum of 8 atomic species). 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, etc. in a file (see notes on implementation). Neutron scattering factors are coded in the program for elements up to Pu (Z=94).

EQUIPORTION

If this directive is given the overall intensity of a group of overlapping reflections is equiportioned among the reflections in the group (equal E-values are obtained).

EXTEFILE *string*

This directive is used to supply an external file containing 'd' or '2q' values of peak positions. In the latter case the additional '2-THETA' directive has to be used.
If this directive is used the program will skip the automatic "*peaksearch* " procedure.

FILETYPE *string*

This directive is used to specify the format of input file containing the diffraction data. *string* defines the type of file and can be 'counts', 'xy', 'double', 'gsas'.

If the directive FILETYPE *string* is absent or *string* is 'counts' the first three number at the beginning of the counts file must be the 2theta minimum, 2theta step and 2theta maximum. The following lines must be counts. The number of counts for row or the format of counts can vary.

#Example 1: use, in the input file, the directive **'FILETYPE counts'** or nothing

```
 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   17902

....

#Example 2: use, in the input file, directive **'FILETYPE counts'** or nothing

6   0.01   55.98

412.98

404.51

412.40

405.57

403.30

405.94

404.89

389.39

396.55

408.00

403.24

385.44

396.46

393.20

390.52

401.58

....

If *string* is 'xy' the counts file contains 2theta in the first column and the intensities in the second one. The format of counts can vary.

#The use of directive **'FILETYPE xy'**

4.700	4769.000	69.05795
4.714	4860.000	69.71370
4.729	4794.000	69.23872
4.743	4857.000	69.69218
4.758	4837.000	69.54855
4.772	4891.000	69.93568
4.787	4788.000	69.19537
4.801	4886.000	69.89993
4.816	4720.000	68.70226
4.830	4688.000	68.46897
4.845	4690.000	68.48357
4.859	4867.000	69.76389
4.874	4800.000	69.28204

4.888   4863.000   69.73521

....

If string is 'double' the first line in the counts file contains 2theta minimum, 2theta step, 2theta maximum. The rest of file contains intensities and standard deviations on separated lines.

#The use of directive **'FILETYPE double'**

```
 3.000   0.005  41.600
3567.2 3405.3 3370.1 3432.4 3476.2 3526.9 3354.7 3596.6 3476.8 3676.7
 67.27  65.44  65.16  65.78  66.12  66.55  64.64  67.14  66.08  67.88
3623.0 3703.6 3555.1 3500.3 3558.6 3584.8 3606.2 3556.8 3710.3 3704.8
 67.46  68.33  66.89  66.12  66.77  67.00  67.75  66.77  68.73  68.03
3721.0 3670.7 3775.1 3796.4 3820.9 3880.8 3809.1 3922.7 3963.5 3919.6
 68.76  67.82  68.60  69.04  69.28  70.05  69.39  69.91  70.44  70.11
4062.6 4104.5 4148.8 4219.5 4221.6 4376.3 4294.3 4439.1 4349.1 4500.5
 71.69  71.85  71.87  72.61  72.61  73.94  73.66  74.44  74.18  75.39
4449.2 4704.1 4833.7 4851.3 4892.2 5105.9 5143.3 5161.8 5272.0 5418.5
 74.34  77.34  78.13  78.37  78.02  79.82  80.53  80.55  81.09  81.93
5543.3 5635.8 5918.0 6007.9 6406.6 6395.2 6515.0 6753.3 7255.9 7464.6
 83.36  84.18  86.37  86.95  89.58  89.58  90.60  91.99  95.62  96.94
7811.5 8415.6 8597.6 9251.3 9995.7 10714.5 11097.2 11067.4 10678.3 9864.0
 98.89 102.81 103.75 107.71 111.43 115.71 117.76 117.40 115.54 110.72
```

....

In the previous cases comment records can appear at the beginning of data file if the first character is '#' or '!' or '>'. Examples in directory examples.

'FILETYPE gsas' can be used to read counts files in GSAS standard format with keyword TYPE='STD' and TYPE='ESD'.

FINDSPACE

This directive is used in order to find a possible space group when the cell parameters are known.

FWHM *x*

This directive supplies the value of *x* for the overlapping calculation. Two reflections are assumed to be in overlapping if $2q_1 - 2q_2 < x * FWHM$. Default is $x = 0.1$.

HISTOGRAM *n*

n is the number of bars in the histogram for the alpha2 stripping procedure (only if ALPHA directive is used).

ISOTOPE *El sf*

To modify the value of the neutron scattering factor for element El using the value sf. (It must be used if NEUTRON directive is given).

MOREPEAKS

This directive is used together with the peaksearch directive in order to add more low intensity peaks to the set of peaks already detected via the peaksearch directive.

NEUTRON

This directive is used to specify that neutron radiation has been used.

NOREFINE

This directive is used to skip the cluster refinement step in the peak search procedure.

PATTERN *filename*

Filename is the name of the profile counts file (up to 40 characters). It is used to change the default. [Default is name.pow where name is defined using the command %STRUCTURE].

RANGE *thmin thmax*

Only intensities included in the range thmin-thmax will be considered; thmin and thmax are the minimum and maximum 2q value of the powder pattern respectively.

REF2 *string*

This directive is used to supply an external reflection file. EXPO reads directly the output provided by any profile fitting program. HKL, FWHM and F**2 must be supplied [default format is (3i4,2f15.4)]. If a comment line is at the beginning of the reflection file, the first character must be '>'. If this directive is used, the directives 'pattern' and 'range' must not be introduced.

SHIFT *sx sy sz*

Origin shift components if desired by the user. They must be integral multiples of 1/24.

SPACEGROUP *string*

String is the symbol of the space group, according to International Tables (1974). Blanks

are necessary among the terms constituting the space group symbol (see examples at the end of this manual).

SYNCHROTRON

This directive is used to specify that synchrotron radiation has been used.

STEP *step_value*

You can use this directive to specify that your data were collected at fixed step. This directive is activated when the data file contains two columns (2theta values and intensities) and you are using the directive 'FILETYPE xy'. If you use the directive STEP, the first column is ignored and the program automatically recreates the 2theta values by using *step_value* for the step and the first 2theta value in the data file.

THCORR *x*

x is the Dq_0 value that is used to correct by zeropoint shift the set of 2q peak positions. The shift value will be equal to $2(n-1)Dq_0$, where n is the sequential number of the N-TREOR trial.

If this directive is not introduced the Dq_0 default value is the experimental 2q step.

2-THETA

This directive is used combined with the EXTEFILE directive to supply to the indexing program N-TREOR09 an external file containing 2q values of peak positions.

WAVELENGTH *w*

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

References
Burzlaff H. & Hountas A. (1982). *J. Appl. Cryst.* **15**, 464-467.

Examples

Examples of input for EXPO2014

Example 1

The following example shows the default use of EXPO2014 in case of indexed pattern. Most of the structures can be solved in this way. Diffraction data are in the file crox.pow. The content of the file crox.exp is

```
%structure crox
%job CROX- data from home diffractometer
%initialize
%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
```

Default "profile counts file" name: crox.pow

crox.pow content:

```
 6.0000 0.02  100.0
   1166.    1179.    1224.    1172.    1221.    1159.    1204.    1347.
1295.    1146.
   1353.    1188.    1214.    1250.    1253.    1240.    1285.    1344.
1314.    1324.
   1438.    1444.    1463.    1522.    1664.    1548.    1836.    1865.
1913.    2160.
   2367.    2412.    2856.    3186.    3681.    4280.    4941.    5883.
7298.    8978.
  11248.   14375.   20187.   25377.   23771.   15690.    8496.    6385.
5553.    5022.
   4643.    4371.    4202.    3840.    3684.    3296.    3199.    3047.
2833.    2710.
   2567.    2334.    2412.    2156.    2163.    2134.    1981.    1981.
2033.    1901.
   1849.    1768.    1761.    1709.    1654.    1616.    1580.    1629.
1561.    1609.
   1658.    1609.    1493.    1334.    1298.    1331.    1305.    1289.
1292.    1214.
   1149.    1182.    1081.    1146.    1094.    1065.    1030.    1055.
981.    1020.
   1046.     981.     971.    1049.     975.     965.     900.     929.
861.     858.
```

```
    874.    929.    835.    793.    839.    787.    907.    858.
822.    826.
```

Example 2

The Ka 2 stripping with a 3 bars histogram is requested.

```
%structure dada
%job DADA- data from home diffractometer
%initialize
%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.pow
     alpha2
     histogram 3
%continue
```

dada.pow content is:

```
   10.00 0.02   95.00
        457.0
        437.0
        421.0
        475.0
        469.0
        425.0
        465.0
        405.0
        447.0
        461.0
        433.0
        433.0
        433.0
        415.0
        453.0
        451.0
        429.0
        453.0
        427.0
```

Example 3

Data are from neutron source.

```
%structure cfcl
%job CFCL - data from neutron source
%initialize
```

```
%data
    pattern cfcl.pow
    cell 10.168 14.964 5.100 90.000 90.000 90.000
    content C 8 F 16 Cl 16
    spacegroup f d d 2
    wavelength 1.595
    neutron
%continue
```

Example 4

A Pseudo-Voigt function will be used and the background in each interval will be described by a constant. The counts file format is not the default one and the graphic window is suppressed.

```
%structure mes
%job MES - data from home diffractometer
%initialize
%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
%end
```

Example 5

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

```
%structure yono
%job YONO - Synchrotron data
%initialize
%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.ext
      ncycle 40
%end
```

Example 6

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

```
%structure agpz
%job AGPZ - data from home diffractometer
%init
%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
```

Example 7

Neutron data and a deuterium atom in the cell content are used. Fast graphic interaction for the integrated intensities extraction process is activated.

```
%structure baco
%job BACO - Neutron data
%initialize
%data
     neutron
     isotope H 0.67
     pattern baco.pow
     cell 10.0632 7.9330 6.8487 90.0 122.3423 90.0
```

```
    spacegroup C 2/m
    content Ba 4 C 8 O 20 H 8
    wavelength 1.909
%extraction
    fast
%continue
```

Example 8

The program restarts from the PHASE procedure.

A random approach is used and the best 100 sets of phases, over 250 trials, are retained in the direct access file. No Fourier is required.

```
%structure nizr
%phase
    random
    maxtrials 250
    minfom 1.0 100
%end
```

Example 9

The program restarts from the FOURIER procedure.

Only the E-map corresponding to the set number 7 and the complete FOURIER/LEAST-SQUARES procedure for set number 10 are requested by the user.

```
%structure nbpo
%fourier
    set 7
    recyc 0
%fourier
    set 10
%continue
```

Example 10

To view only the structure previously produced by FOURIER routine.

```
%window
%structure crox
%menu
%end
```

Example 11

Use of the pseudotranslational symmetry (if it is present) as prior information for a new extraction process is inhibited. Use of nopse directive after command %normal.

```
%structure agpz

%job AGPZ - data from home diffractometer
%initialize
%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
     nopseudo
%continue
```

Example 12

Use of the triplets estimation as prior information for a new extraction process.

```
%structure mes
%job MES - data from home diffractometer
%initialize
%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
     nopseudo
%invariants
     estimation
%continue
```

Example 13

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

```
%structure lasi
%job lasi - Neutron data
%initialize
%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
     nopse
%patterson
     inverse
%continue
```

Example 14

A known fragment is used to complete the structure by applying the FOURIER/LEAST-SQUARES procedure. The binary file "sapo.bin" must exist.

```
%structure sapo
%fourier
     fragment SAPO.FRA
%continue
```

Coordinates are in the file "SAPO.FRA" which contains

Si .00000 .00000 .00000

Example 15

The indexing process is performed.

```
%structure gapo
%job GAPO - Synchrotron data
%initialize
%data
     wave 1.24012
     pattern gapo.pow
     synchrotron
%ntreor
%end
```

Example 16

If a set of d (or 2 q) values is already available, the automatic peak search procedure is skipped:

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

```
      VOL = 6000, CEM = 40,
%end
```

'mes.pea' is an ASCII file containing the 2q peak positions.

The N-TREOR09 directives 'VOL = 6000' and 'CEM = 40' activate a search for a cell having a maximum volume of 6000 Å^3 and maximum axis value of 40 Å.

Example 17

Use of the change label procedure to relabel the peaks of an electron density map obtained by applying the FOURIER procedure. The binary file "vfi.bin" must exist.

```
%structure vfi
%job VFI - Synchrotron data
%changelab
      label 3 al 3 p
      coord 4 4
      dwei 3.1 0.5
      dlig 1.9 0.5
%continue
```

Example 18

The POLPO1 procedure is applied to locate the light atoms around each specified heavy atoms. The crox.fra file contains the cations coordinates.
The binary file 'crox.bin' must exist.

```
%structure crox
%job CROX - home diffractometer data
%polyhedra
      fragment crox.fra
      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
```

Example 19

The POLPO2 procedure is applied to locate one or more missing cations and surronding anions. The sapo.fra file contains the located cations coordinates.
The binary file 'sapo.bin' must exist.

```
%structure sapo

%job SAPO - data from home diffractometer
%polyhedra
      fragment sapo.fra
      tetrahedron Si 1 1.6 0.2 0.2
      tetrahedron Si 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
```

Example 20

To find the space group.

```
%structure cf3br
%job CF3BR - Neutron data
%init
%data
    pattern cf3br.pow
    cell 8.146 5.85 7.962 90. 111.722 90.
    cont c 4 br 4 f 12
    wave 1.911
    neutro
    findspace
%continue
```

Table 1

Classes of reflections corresponding to the low index pseudo-translational 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

Crystal Structure Visualization

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).
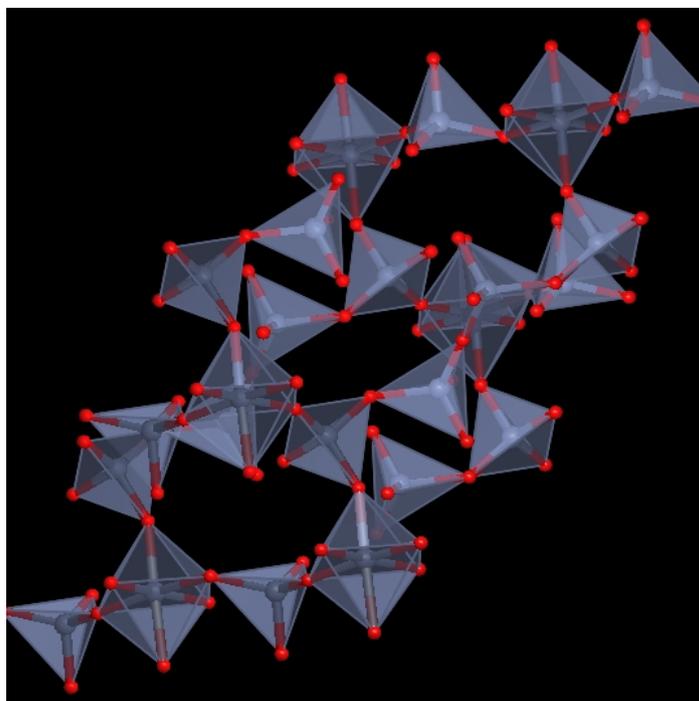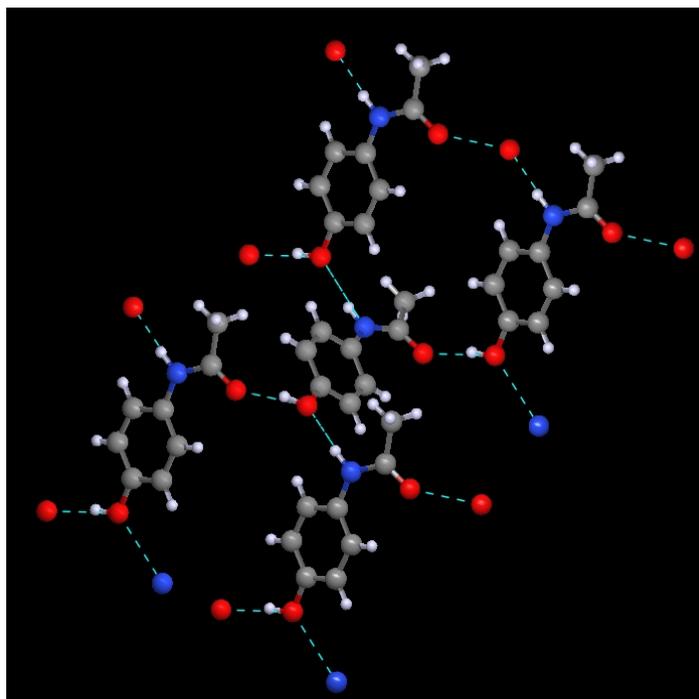
How to

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

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

Import	Opens a coordinate file
Export	Saves a coordinate file
Screens hot	Produces an image file of Jav screen
Exit	Closes the viewer

Import
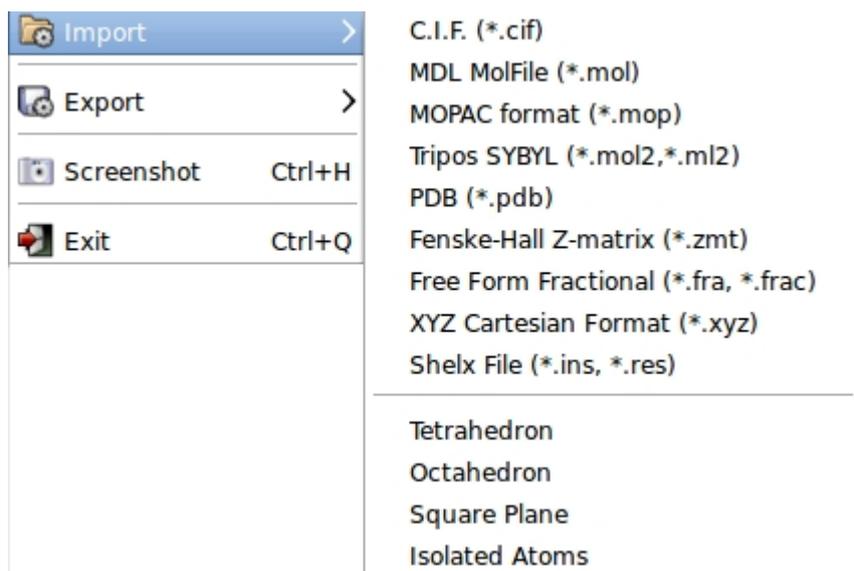
It is possible to read crystal structure from the following types of file:

- 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/

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.

Export
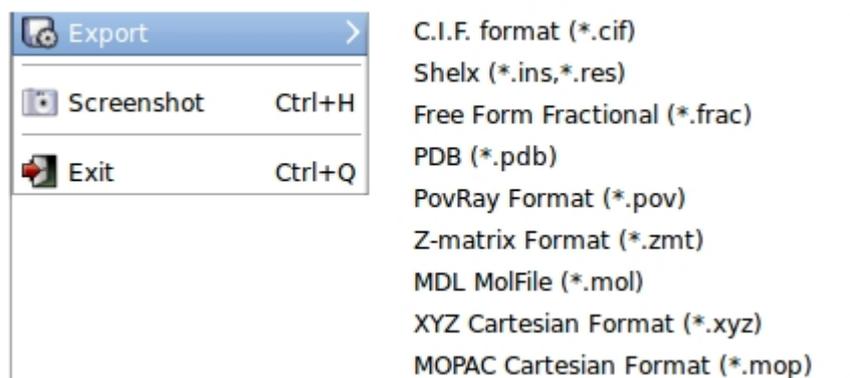
Information about the currently displayed crystal structure can be in following text files:

- 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/

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)

Exit

This is used to exit the program.

Modify

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

Modify
↩ Undo import	Ctrl+Z	**Undo/Redo**
↪ Redo	Ctrl+Y	
🗑 Delete Selection		**Delete Selection**
✳ Delete Bonds		**Delete Bonds**
⊗ Delete All Type ...	>	**Delete All Type**
Change Species	>	**Change Species**
Rename		**Rename**
Renumber		**Renumber**
Sort by Species		**Sort by Species**
Add Bonds		
📝 Edit structure		**Edit structure**
📁 Add fragments	>	**Add fragments**
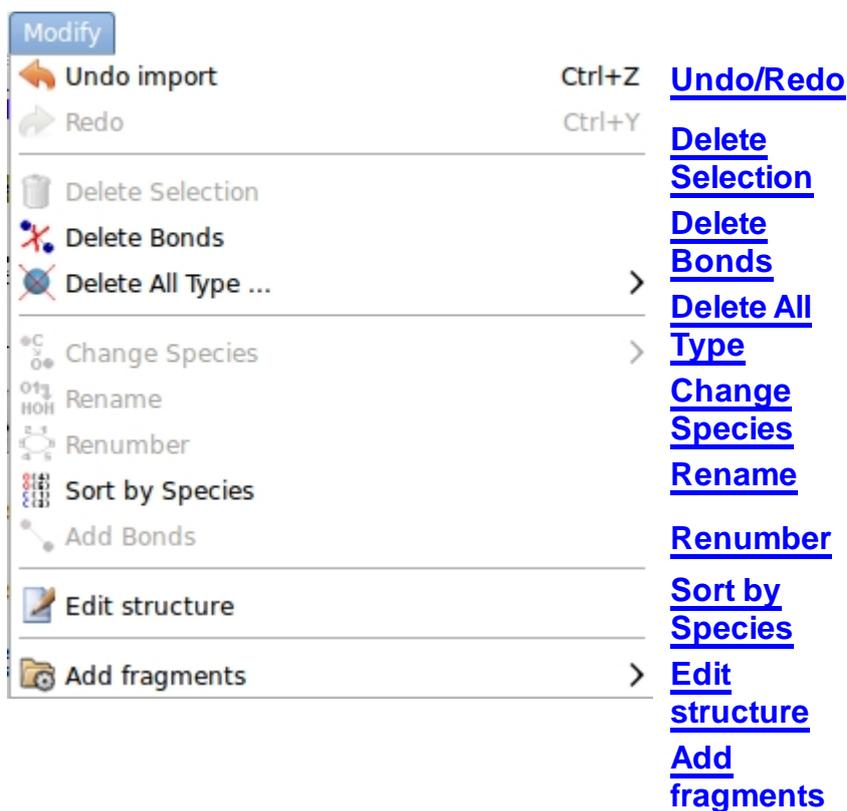
Created with the Personal Edition of HelpNDoc: Full-featured Kindle eBooks generator

Select

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

Select
-¦- Selection Mode	**Selection Mode**
🔍 Zoom Mode	**Zoom Mode**
↖ Manipulation Mode	**Manipulation Mode**
◪ Invert selection	**Invert selection**
✕ Select None	**Select none**
✔ Select Type ... >	**Select Type**
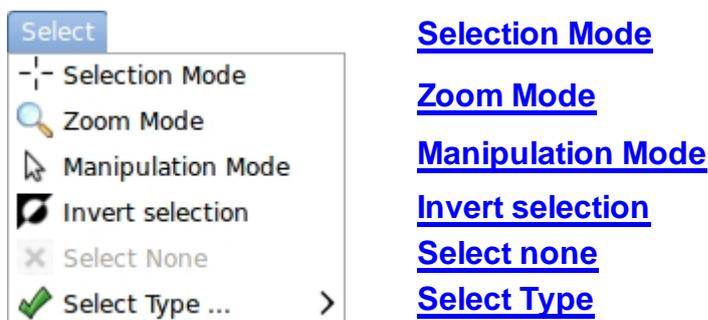
Created with the Personal Edition of HelpNDoc: Easily create EBooks

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.
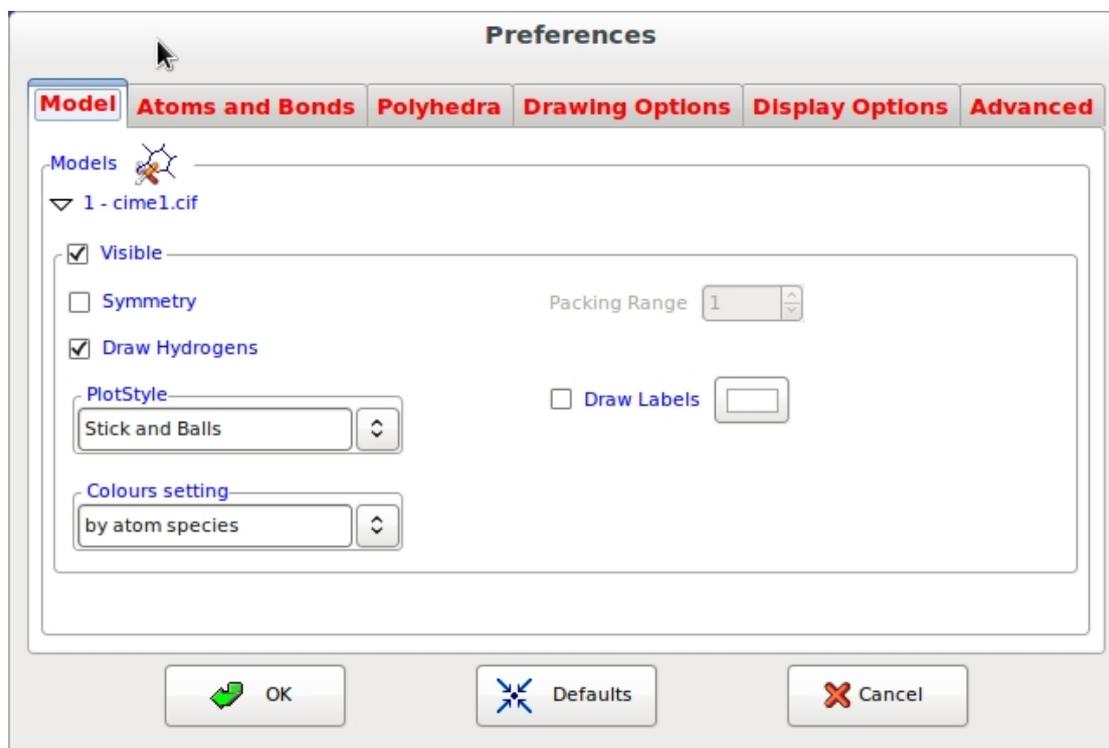
In order to get a better view of a model it is possible to tune the radius of the spheres

representing the atoms and the radius of the cylinders representing the bonds. The atom colors can be modified.
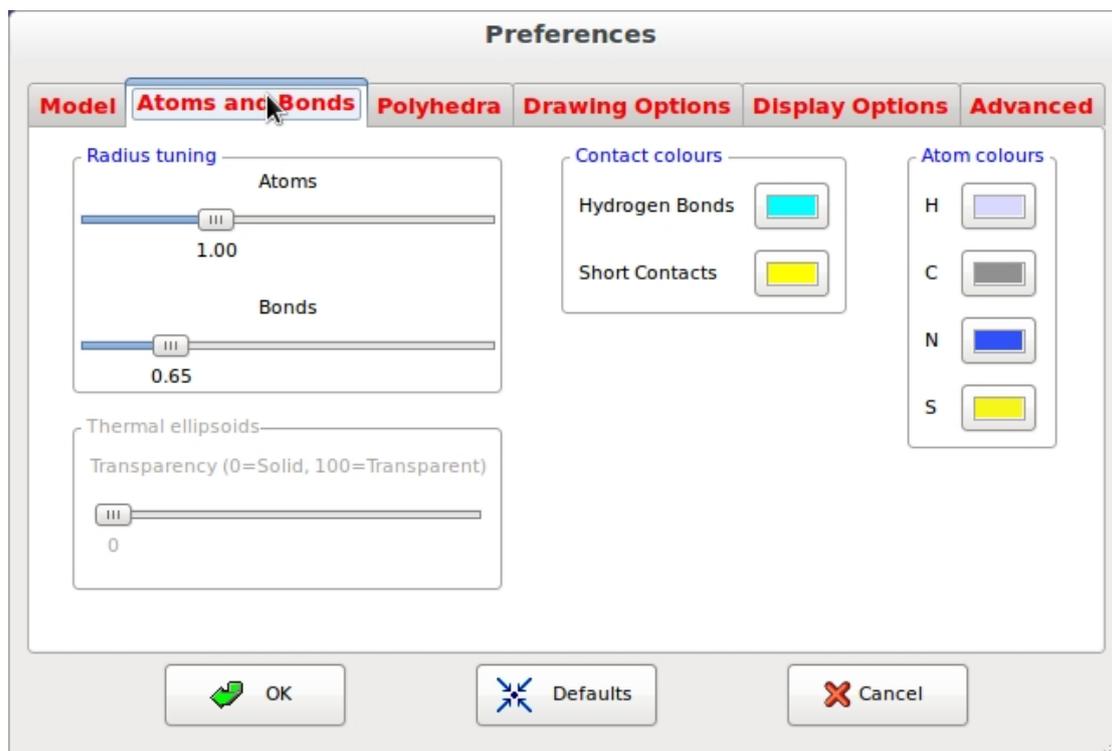
Polyhedra Properties
This window provides some tools to set the properties of the polyhedra; it is possible to visualize or to hide some of them, it is possible to change the colors and the transparency, to show or to hide the edges etc.
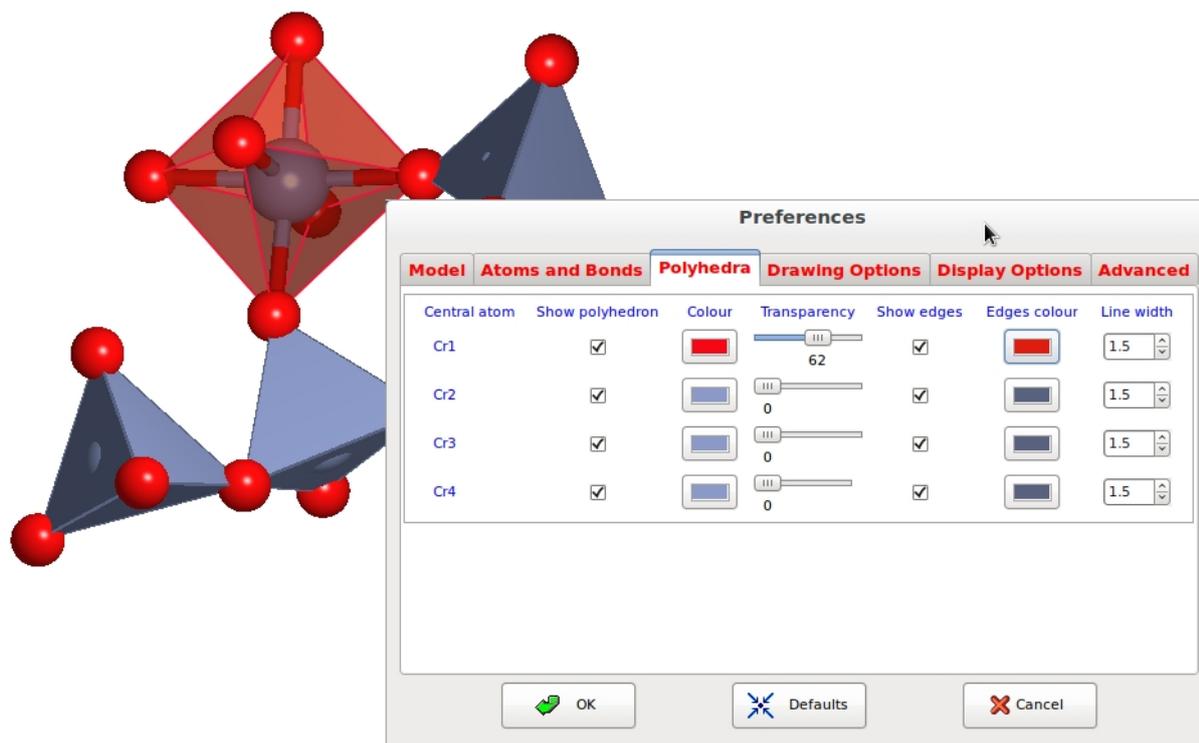
View plane

Set the preferred view plane.
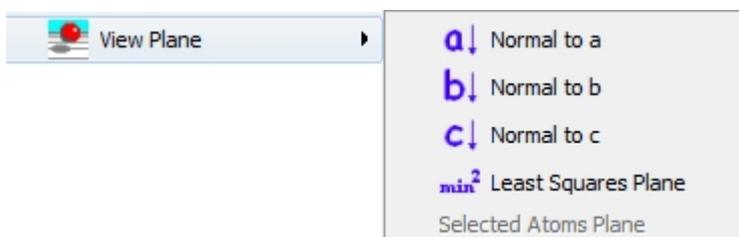
Tools

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

Tools

✏ Add Hydrogens   ›

🔧 Build Bonds

🔧 Bond Settings

↻ Complete

↻ Regularize

♟ Enantiomorph

🔧 Transform

    Optimize Geometry   ›

    Optimize Crystal Structure by DFT-D

⚹ Overlay Structures

ℹ Overlay Info

    Centroids

    Powder Diffraction Pattern
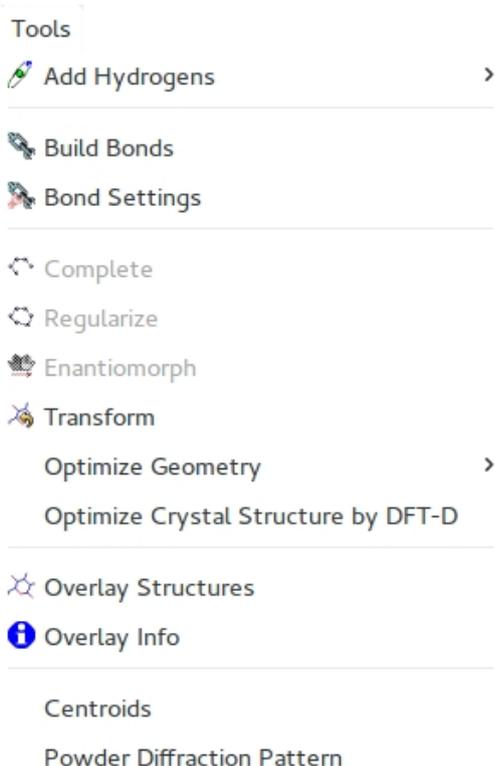
Geometry Optimization

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).
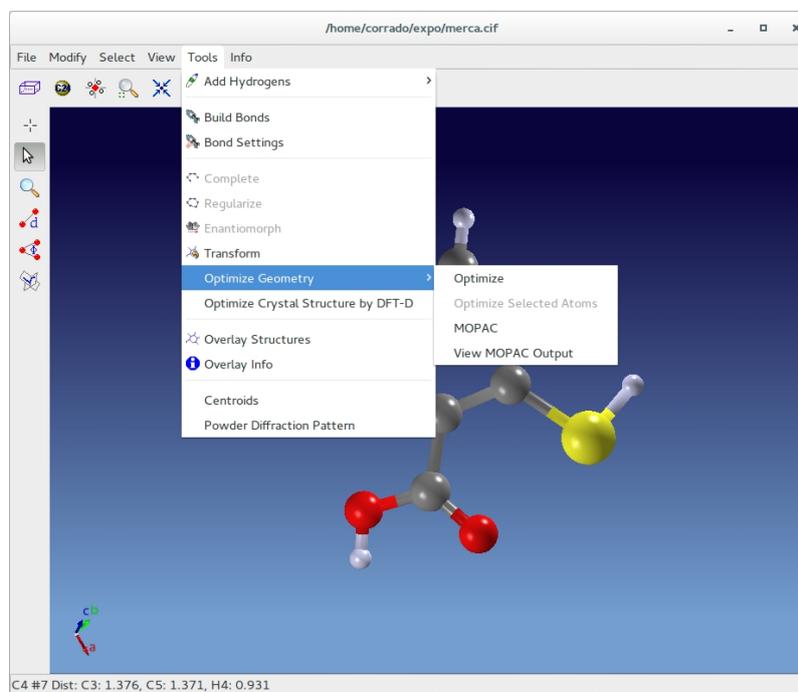
Geometry optimization by molecular-mechanics force fields

In molecular mechanics, 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 and the specified set of equations and parameters is called the force field. The constants in this equation are obtained either from spectroscopic data or *ab initio* calculations. A set of equations with their associated constants is called a force field. 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:

$$f_i = - \partial E_{tot}/\partial q_i$$

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**.

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:
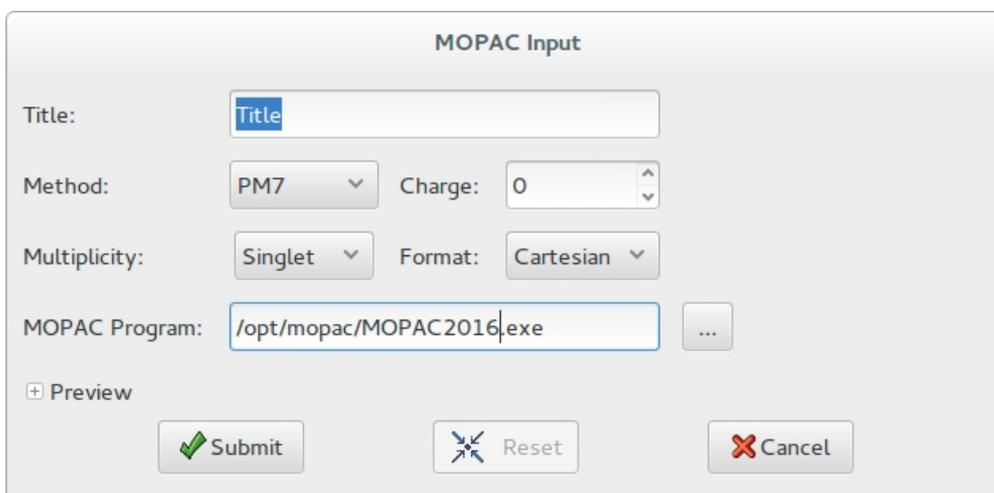
Title: define the title of the calculations, *e.g.*, the name of molecule. It is reported in the second line of the MOPAC input file

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 may 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.

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 > Import 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.

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

Info

The sub-menu Info is shown on the left. It allows:
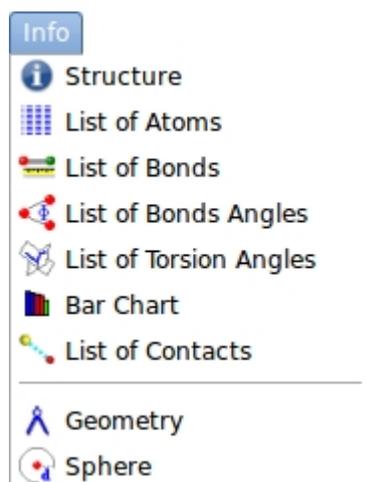
Structure	Info on the structure.	
List of Atoms	List of Atoms.	
List of Bonds	List of Bond Distances.	
List of Bond Angles	List of Bond Angles.	
List of Torsion Angles	List of Torsion Angles.	
Bar Chart	Histogram of the peak intensities.	
List of Contacts		
Geometry	Some geometric calculations can be 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.
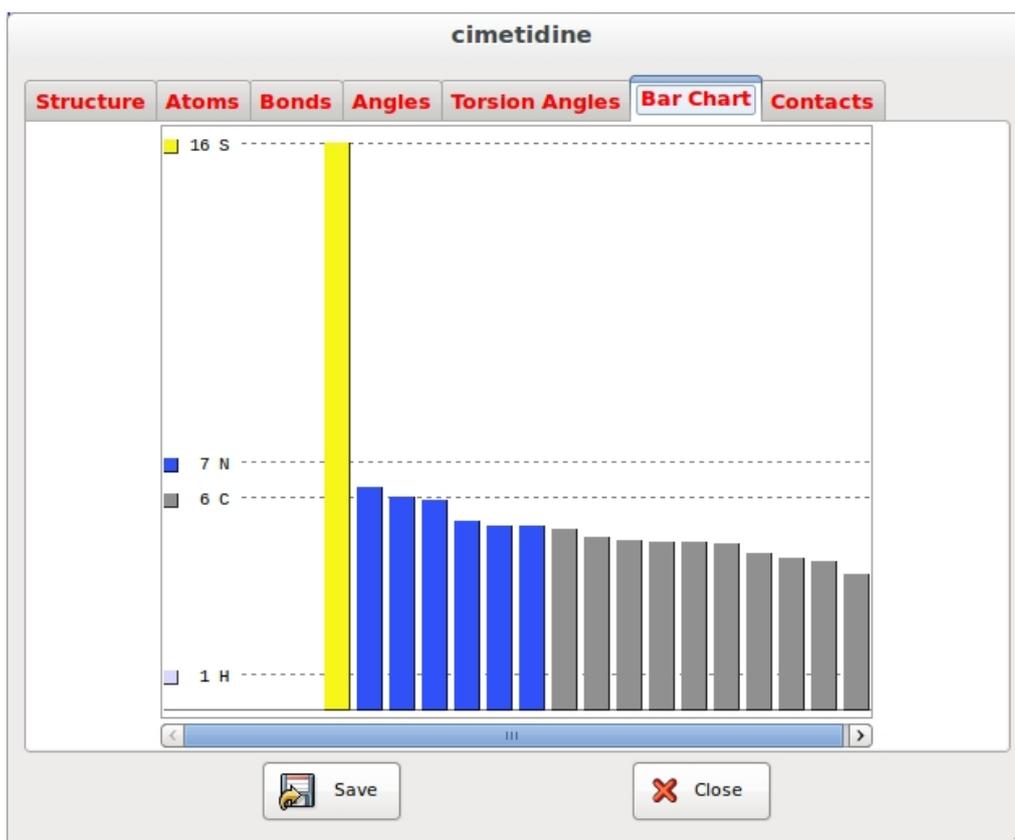
List of Contacts

The tab Contacts contains a list of non bonded contacts.

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
```