

APPENDIX 10. *Highest quality large custom databases from CIF files*

Recently, we have added to XPowderX (Ver 2019.06.10 and later, www.xpowder.com) a procedure that allows us to simultaneously convert collections of to approximately 20000 crystal structure files to an single database file containing the *ascii* diffractogram data of compounds selected according to your own research topics. The original files must be placed in a single folder and must include the description of the crystalline structures of the selected compounds (as examples, cif files of Crystallography Open Database -COD-, AMCSD or those generated by the programs that calculate such structures, such as the .RES of the XSHel program of Sheldric,1963, etc).

The number of records of the databases that can be created is unlimited. For example, if you include the CIF files between '9000000.cif' and '9014008.cif' of the folder number 9 of COD that contains the mineral structures, you can generate in a single session, a database of 16731 records containing exclusively mineral diffractograms (See figure 10.3A). It can also include mineral diffractograms that are scattered by other COD folders different from 9. In this case, the number of mineral diffractograms generated is close to 25,000.

The diffractograms thus generated may come from different sources (CIF, RES, etc.). You can also create one or unlimited databases where diffractograms of different substances (organic, inorganic, complex, organometallic, minerals, proteins, pigments, medicines, explosives, cements, etc.) can coexist. It is also possible to create custom databases. For example, a database can be created containing only Silicates or compounds containing Fe and Cu, large molecules, and, or, small molecules, etc.

Also, and as examples, you can obtain databases that only contain compounds where Fe and Mg appear simultaneously (Fe <AND> Mg) or optionally simultaneously (Fe <OR> Mg) by using appropriate Boolean operators.

The diffractograms are generated with the Lorentz-Polarization, multiplicity of hkl, effects of the monochromator crystal (optional) and automatic slit (optional) corrections. Also, corrections for 'Anomalouss scattering' and 'Absorption' can also be optionally made.

The program recognizes non-standard orientations of the space groups, as well as the different axial crossings of the trigonal and hexagonal lattices.

The time to generate a database depends on the size of the unit-cells and the number of calculated patterns. As an example, the mineral database shows in figure 10.1A takes about two hours to be generated.

In summary, each researcher can generate with a single click of the left mouse button, a personal database of diffractograms according to their research needs.

REMARKS

A) It is preferable not to use either the desktop or long names for the data entry and exit folders, as conflicts may arise when the input files of various operating systems are processed.

B) Remember to mention the authors of the COD files. The most frequent references include:

COD reference database 2012

Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P. & Le Bail, A.; Crystallography Open Database - an open-access collection of crystal structures, *J. Appl. Cryst.*, 2009, 42, 726-729.

When your research relies on the subset of COD containing mineral data from AMCSD (the range 9* datasets), then please mention this source database:

Downs, R. T. & Hall-Wallace, M. (2003). "The American Mineralogist Crystal Structure Database". *American Mineralogist* 88, 247-250."

INSTRUCTIONS.

1. Download the databases of crystalline structures in the usual directions (see below, COD) and unzip the files in any folder on your computer.
2. Click on 'menu/PXRD database CIF Gener' as figure 10.1A shows

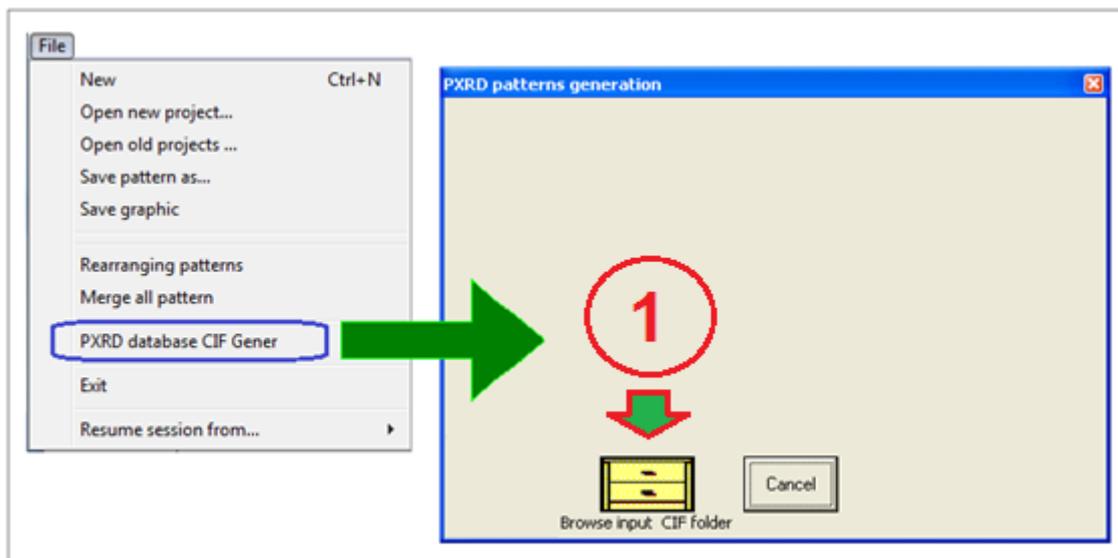


Figure 10.1A. Access tools for 'PXRD database patterns generation'. Files with the selected structures must be placed in a single folder, which must be located through the browser (1).

3. Browse the folder that contains the files with the crystalline structures  as figure 10.2A shows.

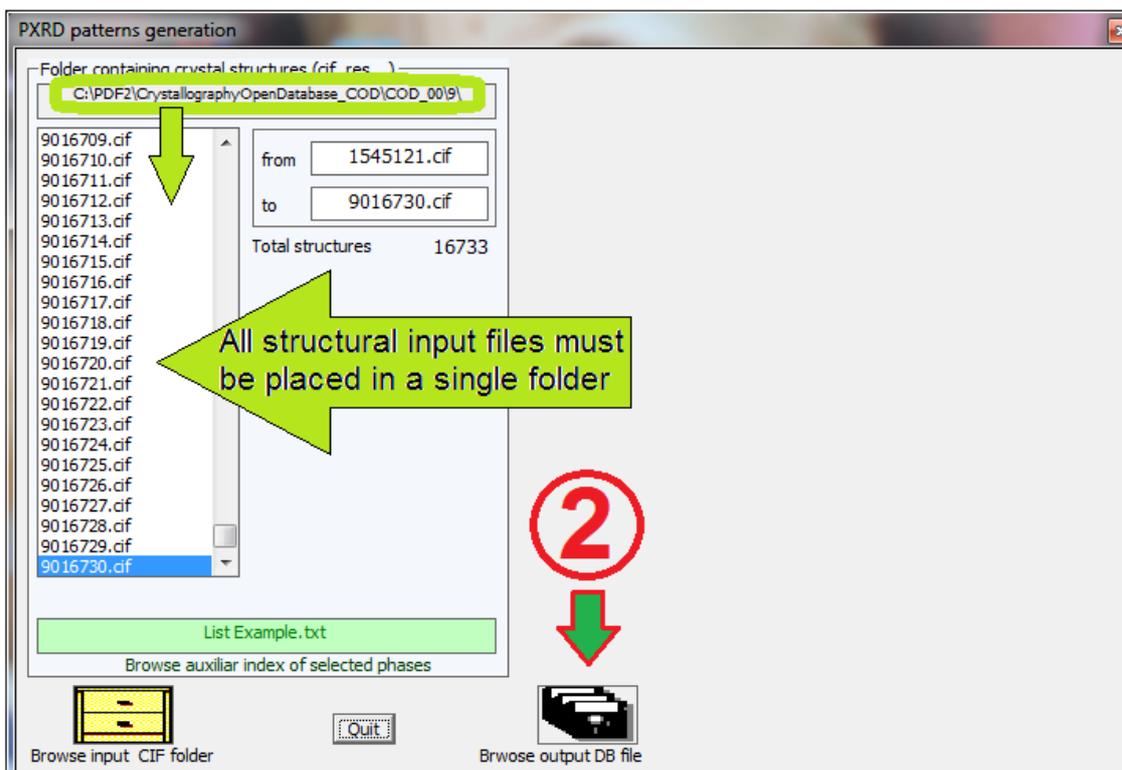


Figure 10.2A. The database that will be generated using the 'structure input files of the single input folder' must be selected using the browser (2).

4. Select the folder where the database will be generated.  as figure 10.3A shows

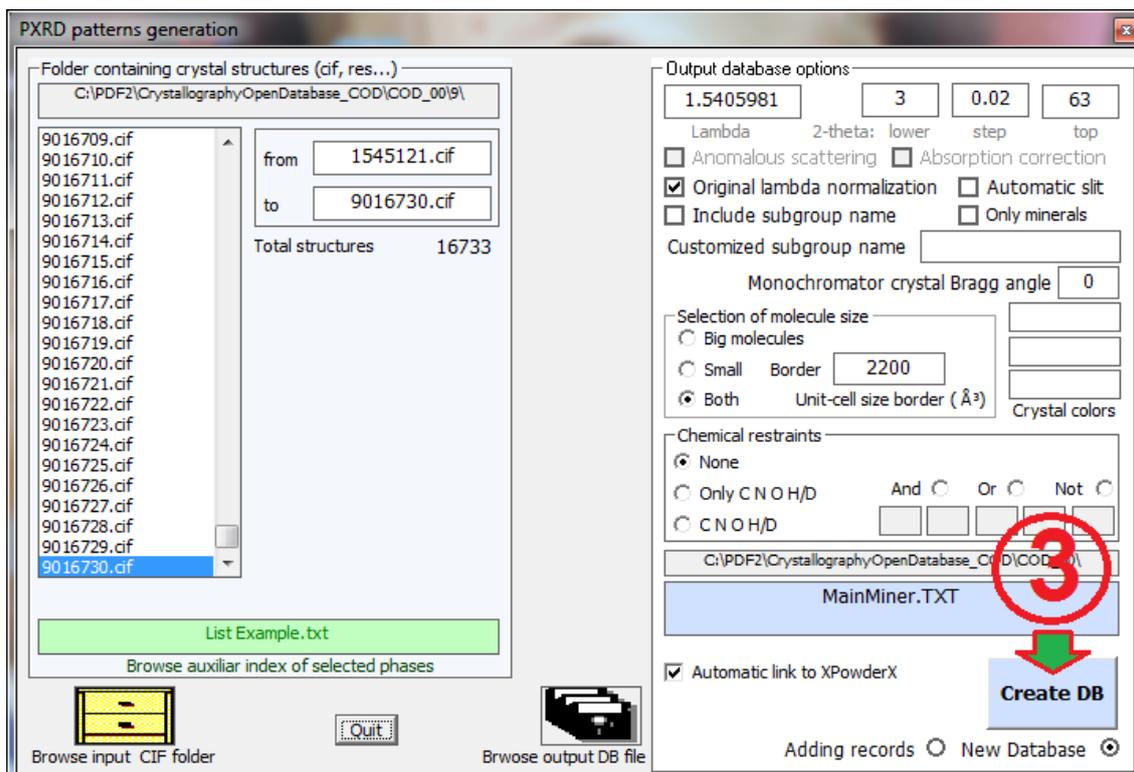
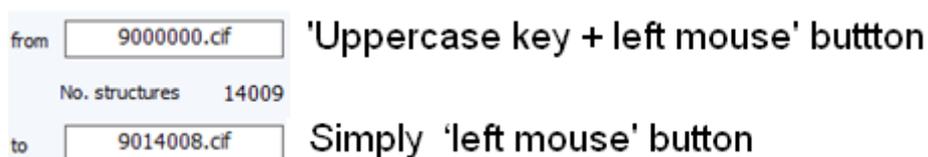


Figure 10.3A. Clicking (3), the diffractogram database of all the structures contained in the incoming folder will be generated. The entire process can take several hours.

The diffractograms will be generated sequentially between the limits indicated in the 'from' and 'to' boxes. **The entire process can take several hours although it can be done in several sessions, using properly the option (8) instead of (9) (figure 10.4A) in successive sessions.**

The number of input files can be reduced by modifying the upper and lower limits of the contents of the input folder. To do this, just click the item on the left list ('uppercase key +left mouse button' and simply 'left mouse button' to modify the lower and upper indexes respectively.



It is possible to use some options (Name of database file output, create or complete a previous one, correction for monochromator crystal or automatic slit, 2-theta limits and step, output lambda fixed value, normalization of input lambda values, etc), or filters to reduce the number of generated diffractograms (Unit-Cell volume, Chemical composition when available, etc).

5. Auxiliary lists containing structures.

5.1. Name phase list: A good option to create good quality databases is to do a list of structures with the names of the phases that should be included in the database, as shown in the example (4) in Figure 10.4A. These lists must be saved in a text file with any of the

extensions 'txt', 'tab' or 'cho'. The example (4) contains a list of names of minerals in alphabetical order, whose structures must exist in the input folder, as 'cif' files and it can be found in the subfolder 'Auxiliary' within the XPowderX folder (Mineral_List_Example.txt). When the 'Create DB' command is executed, the program will only select the structural files of the input folder, but will only create the diffractograms included in the list. The procedure is slow and may take several hours, but the results are excellent.

5.2. CIF file names list: Faster than the previous option is to create a list with the files containing the selected crystalline phases, as shown in (5) (Figure 10.4A). You can also include the name of the phase separated by a comma as in (6), but you cannot mix both formats in the same file. Both lists type must be saved in a text file with any of the extensions 'STC'.

Using both (5.2) options requires knowing the names of the CIF files that contain each crystalline phase, but this task is easy by creating a small auxiliary input / output program that provides this information. You can also use the tools that COD provides in this regard. As an example, in the subfolder 'Auxiliary' you can find a file called "MainMinerals.STC" that includes the names of 3818 CIF files of COF of the most frequent minerals

6. Clicking (3) **Create DB** in figure 10.3A, the diffractogram database of all the structures contained in the incoming folder will be generated. The entire process can take several hours.

7. If box (7) (figure 10.4A) was checked, after the creation of the new database, in addition to it being registered in a text file with the extension 'TXT', the program will automatically create the necessary index files to automatically link it to XPowder and add at the end of those that were previously installed on your computer. From this point, you can use this database without further delay.

Of course, the database file 'TXT' can be deleted at any time or, on the contrary, installed again with specific formats on it or other computers you own, without having to re-carry out the actions explained in this appendix 10 (See also Appendix 2, p 42).

Notice: It is preferable not to use either the desktop or long names for the data entry and exit folders, as conflicts may arise when the input files of various operating systems are processed.

Main links for the free database COD (Crystallography Open Database)

1 Home page

<http://www.crystallography.net/cod/>

2, Searching

<http://www.crystallography.net/cod/search.html>

(Selects 'restriction / Search/'. Then <http://www.crystallography.net/cod/result.php> 'list of COD numbers')

3. The most up-to-date version of PCOD can be found on

<http://cod.ensicaen.fr/pcod/> , but we also maintain a mirror.

4. The whole CIF tree is available at <http://www.crystallography.net/pcod/cif/>

<http://www.cryst.ehu.es/cgi-bin/cryst/programs/nph-wp-list>

<http://www.cryst.ehu.es/#retrievaltop>

5. Searching for selected phases <http://www.crystallography.net/cod/search.html>

6. All CIF COD files

<http://www.crystallography.net/archives/cod-cifs-mysql.zip>

<http://www.crystallography.net/archives/cod-cifs-mysql.tgz>

<http://www.crystallography.net/archives/cod-cifs-mysql.txz>

7. The current version and the older snapshots are available at

<http://www.crystallography.net/archives/>

8. Finally, using the **Web browser**, the whole CIF tree [1/ to 9/ CIF COD files (9= minerals)] can be fetched at:

<http://www.crystallography.net/cif/>

Partial database output example

```

=====
_END_
  Galena
'Pb S'
Ramsdell, L. S.. (1925) ' The crystal structures of some metallic sulfides Locality: Colorado, USA'.
American Mineralogist 281 - 304 Vol 10
# CIF file is available in the Crystallography Open Database (COD),
  _database_code_amcsd 0000002
  _database_code_COD 9000001
CELL PARAMETERS: 5.9300 5.9300 5.9300 90.000 90.000 90.000
SPACE GROUP: F m -3 m
# X-Ray wavelength in cif: 1.540598
# Anode in cif:
X-RAY WAVELENGTH: 1.540598
# Space Group Number = 225
# Standard Sp Group N = 225
# Laue Group Number = 11
# HM Laue Group = m3m
# Crystal system number = 7
# Crystal system = cubic
# Cell formula units = 4
# F000 = 392
_chemical_formula_sum 'Pb S'
_exptl_crystal_density_meas 7.621
_exptl_crystal_density_diffn 7.621
# Unit-Cell volume = 208.5280
# Top Intensity = 4327866
MAX. ABS. INTENSITY / VOLUME**2: 99.528
  2-THETA INTENSITY D-SPACING H K L
  26.00 591440 3.4243 1 1 1(8)
  30.12 4327866 2.9646 2 0 0(6)
  43.12 3252299 2.0962 2 2 0(12)
  51.04 433949 1.7880 3 1 1(24)
  53.48 1166073 1.7120 2 2 2(8)
  62.62 560991 1.4823 4 0 0(6)
=====
_END_
  Cinnabar
'Hg S'
Ramsdell, L. S.. (1925) ' The crystal structures of some metallic sulfides Locality: unknown'.
American Mineralogist 281 - 304 Vol 10
# CIF file is available in the Crystallography Open Database (COD),
  _database_code_amcsd 0000006
  _database_code_COD 9000004
CELL PARAMETERS: 4.1500 4.1500 9.5100 90.000 90.000 120.000
SPACE GROUP: P 32 2 1
# X-Ray wavelength in cif: 1.540598
# Anode in cif:
X-RAY WAVELENGTH: 1.540598
# Space Group Number = 154
# Standard Sp Group N = 154
# Laue Group Number = 7
# HM Laue Group = -32m
# Crystal system number = 5
# Crystal system = trigonal
# Crystal axis = Hexagonal Lattice
# Cell formula units = 3
# F000 = 288
_chemical_formula_sum 'Hg S'
_exptl_crystal_density_meas 8.171
_exptl_crystal_density_diffn 8.171
# Unit-Cell volume = 141.8430
# Top Intensity = 2680871
MAX. ABS. INTENSITY / VOLUME**2: 133.248
  2-THETA INTENSITY D-SPACING H K L
  24.76 5722 3.5929 1 0 0(6)
  26.50 1504649 3.3608 1 0 1(12) /0 1 1(12)
  31.16 2680871 2.8680 1 0 2(12) /0 1 2(12)
  37.82 6852 2.3769 1 0 3(12) /0 1 3(12)
  43.58 545091 2.0751 1 1 0(6)
  45.72 955495 1.9829 1 0 4(12) /0 1 4(12)
. . .
. . .
. . .

```