# Short guide for using the program MolPDF

**Juan Rodriguez-Carvajal and Aleksei Bytchkov**

**ILL and ESRF, November 2016**

This version of the program **MolPDF** (Version 0.7 October 2016) corresponds to a preliminary version before developing a new one including rigid body refinement.

At present the program can work with several patterns and several phases. The program is able to refine the atom positions of a crystal structure even in the case of nanoparticles with not well-resolved diffraction pattern, together with a set of distance and angles restraints. The program works with at least an input file with extension *.cfl (CFL file in the following). The optimization procedure used for optimization is the Levenberg-Marquardt (LM) algorithm that works between the gradient steeped descendent and the Gauss-Newton algorithms, allowing a robust refinement even in the case the initial parameter are relatively far from the optimum.

## Getting started

The program can be invoked from the command line using one of the following commands:

```
Prompt> MolPDF my_CFL_file.cfl
```

```
Prompt> MolPDF my_CFL_file.cfl calc
```

```
Prompt> MolPDF my_CFL_file.cfl calcp
```

In the first case a refinement or a calculation will be performed depending on the content of the CFL file. If the keyword `JUST_CALC` appears in the CFL file only a calculation is performed and the optimization procedure is not invoked. A series of files will be produced (`my_CFL_file.out, my_CFL_file.prf, my_CFL_file_new.cfl, my_CFL_file.cif, my_CFL_file.vesta, etc.`) depending of the content of the CFL file.

In the second case just a calculation is done producing a PRF file (`my_CFL_file.prf`), the normal output file (`my_CFL_file.out`) and a file containing all the distance pairs in the calculation (`my_CFL_file_pairs.lst`).

In the third case the same files than in the second case are produced and another one with the calculated PDF file (`my_CFL_file_cmp_gr.dat`) with a format similar to the XYDATA of input files for FullProf (`INSTR=10`).

The program can also be invoked from the **FullProf Suite** toolbar, provided that a CFL is loaded in the toolbar.

## Preparing a CFL file for MolPDF

A minimal CFL file (a single phase and a single pattern) can be created using the interface existing in **WinPLOTR-2006** in the menu `Ext.Applications>Prepare a CFL file for Calc_PDF`. Clicking this option opens a dialog in which one can import a PCR, a CIF, a SHELX (or even a pre-existing CFL file) that is loaded in the dialog. The user can modify the options of the dialog (in particular the name of the output CFL file) and run **Calc_PDF** (which is MolPDF invoked as in the third case) from the dialog. The user can edit the CFL file and adapt it to his (her) needs.

### Structure of the CFL file for working with MolPDF

The CFL file can be read by whatever program based in CrysFML (Crystallographic Fortran Modules Library) but for working specifically with **MolPDF** several prescriptions have to be respected. In general the file is free format and comments provided by the uses can be conserved when **MolPDF** runs and creates a new CFL file with the last values of the refined parameters. The CFL file contains generally keywords and numerical values that are passed to the program for doing different actions. Default values are used if some keywords are not given.

The first information needed concerns the characteristics of the pattern(s) and this is given in one of the PATTERNs blocks. The total number of patterns is the number of existing PATTERN-blocks in the CFL file. These blocks have to appear before any information about the crystal structure of the different phases appear in the CFL file. The structure of such blocks is as follows in this example of two patterns:

```
! Patterns and general info block
PATTERN_01    0.50    ! Starting block for pattern 1 and weight
TITLE Mixture rutile-anatase neutrons
GLOBAL_ITERATIONS 2   ! The call to LM-procedure is done two times
NO_WRITE_SIGMA   ! Standard deviations not written in the new CFL
#JUST_CALC   ! If this keyword appears, only a calculation is done
PDFGETFILE 2phases_N.xys     ! Name of the file containing G(r)
SCALE_FACTOR      1.00000
PDF_RANGE    0.50 0.01 29.99   !rmin, step, rmax
QMAX   30.000000
DAMP      0.01200   ! Value of Qdamp
BROAD     0.00200   ! Value of QBroad
JOBTYPE Neutrons    ! Radiation use to obtain the pattern
PDFFORMAT XYDATA    ! Format of the PDF file (XYDATA, PDFGUI, ASCII)
gvary scalef   ! The scale factor of the current pattern is refined
#gvary qdamp   ! qdamp is not refined. Preserved comment
gfix qbroad    ! qbroad is fixed


PATTERN_02    0.50
TITLE  mixture rutile-anatase x-rays
SCALE_FACTOR      1.00000
PDF_RANGE    0.50 0.01 29.99
QMAX   30.000000
DAMP      0.01200
BROAD      0.00020
JOBTYPE X-rays
PDFGETFILE 2phases_X.xys
PDFFORMAT XYDATA
gvary scalef
#gvary qdamp
#gvary qbroad
```

The words written in bold are keywords and they are followed by numerical or character values. The comments starting with "#" are preserved in the new CFL file generated by the program, those starting with "!" are generally not preserved. They are used above for explaining the meaning of the keywords. For instance the keyword **NO_WRITE_SIGMA** is used to avoid the writing of the standard deviations in the new created CFL file. This is important in cases where one wants to preserve the precision to re-start a refinement with starting Chi2 value similar to that of the previous

refinement. At the end of a refinement it is advisable to remove this keyword to get a short file with a summary of the results.

The second set of blocks is that describing the different crystallographic phases contributing to the patterns, these blocks are called PHASE-blocks. One of these blocks is described below.

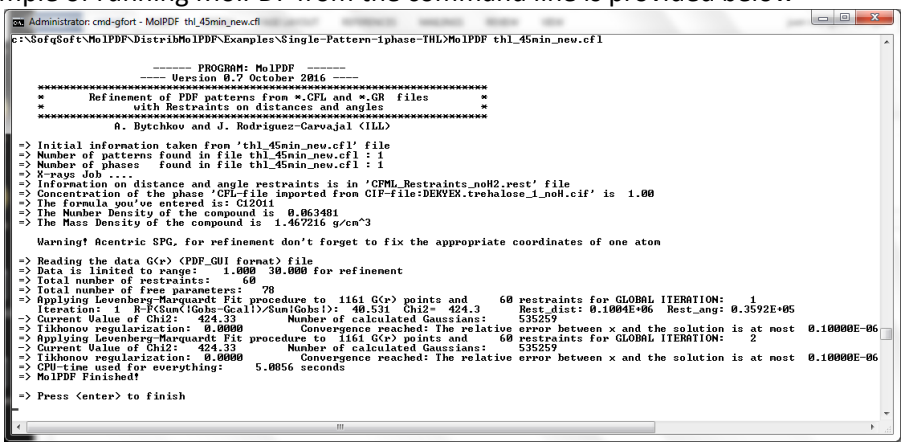
```
! Phases info block

PHASE_01       1.00000  ! Starting PHASE-block 1, a single phase with xfract=1.00
TITLE CFL-file imported from CIF-file:DEKYEX.trehalose_1_noH.cif
!  Automatically generated CFL file (Write_CFL)
!         a          b         c       alpha    beta      gamma
Cell  12.99258   8.26937  6.80399   90.0000  98.30946  90.0000
!  Space Group #   4
Spgr  P 21
!    Atom  Type   x/a      y/b      z/c     Biso    Occ  Spin  Charge   Info
Atom     C1  C   0.18603  0.00000  0.16143 2.31145 1.00000 0.00  0.00    #
Atom     C2  C   0.17997 -0.18142  0.21273 2.31145 1.00000 0.00  0.00    #
Atom     C3  C   0.20466 -0.28519  0.03736 2.31145 1.00000 0.00  0.00    #
Atom     C4  C   0.12594 -0.24031 -0.15089 2.31145 1.00000 0.00  0.00    #
Atom     C5  C   0.13156 -0.05400 -0.19307 2.31145 1.00000 0.00  0.00    #
Atom     C6  C   0.04399  0.00152 -0.36281 2.31145 1.00000 0.00  0.00    #
Atom     C7  C   0.30652  0.18891  0.08704 2.31145 1.00000 0.00  0.00    #
Atom     C8  C   0.42021  0.21044  0.05258 2.31145 1.00000 0.00  0.00    #
Atom     C9  C   0.50030  0.19545  0.24760 2.31145 1.00000 0.00  0.00    #
Atom    C10  C   0.46551  0.29971  0.41394 2.31145 1.00000 0.00  0.00    #
Atom    C11  C   0.35060  0.26109  0.42918 2.31145 1.00000 0.00  0.00    #
Atom    C12  C   0.30706  0.38388  0.57185 2.31145 1.00000 0.00  0.00    #
Atom     O1  O   0.28820  0.02650  0.13050 2.12377 1.00000 0.00  0.00    #
Atom     O2  O   0.25658 -0.21301  0.37598 2.12377 1.00000 0.00  0.00    #
Atom     O3  O   0.17658 -0.45240  0.07213 2.12377 1.00000 0.00  0.00    #
Atom     O4  O   0.15411 -0.32471 -0.33266 2.12377 1.00000 0.00  0.00    #
Atom     O5  O   0.11503  0.04226 -0.00628 2.12377 1.00000 0.00  0.00    #
Atom     O6  O  -0.05346 -0.05860 -0.33857 2.12377 1.00000 0.00  0.00    #
Atom     O7  O   0.44564  0.09441 -0.09441 2.12377 1.00000 0.00  0.00    #
Atom     O8  O   0.59920  0.24553  0.21166 2.12377 1.00000 0.00  0.00    #
Atom     O9  O   0.52363  0.25492  0.59430 2.12377 1.00000 0.00  0.00    #
Atom    O10  O   0.29007  0.29020  0.24291 2.12377 1.00000 0.00  0.00    #
Atom    O11  O   0.20759  0.34297  0.60945 2.12377 1.00000 0.00  0.00    #
SUPCEL    1 1 1 n   ! The unit cell is the crystallographic cell: 1x1x1, n: no write
CHEM_FORM  C12O11   ! Chemical formula
DELTAS      1.39180     0.00000  ! Standard sharpening factors delta1 and delta2
SPHSIZE     5  2.27288  ! Degree of the size-distribution and Size parameter
MODEL_SHARPF   Exponent   Intra_Dist   Out_Dist   ! Molecular sharpening model, mode & output
SHARPF      0.10034    ! Molecular sharpening coefficient
RESTR_FILE   CFML_Restraints_noH2.rest   ! Name of the restraints file
gvary delta1    ! delta1 is refined
gvary cell      ! cell parameters are refined
gvary sharpf    ! Molecular sharpening parameter is refined
gvary size      ! Size parameter is refined
vary xyz        ! Atom coordinates are refined
vary Biso       ! Isotropic temperature factors are refined
fix y_C1        ! The y-coordinate of the C1 atom is kept fixed (Acentric group)
equal B_C1 B_C2 B_C3 B_C4 B_C5 B_C6 B_C7 B_C8 B_C9 B_C10 B_C11 B_C12    !constraints in Biso
equal B_O1 B_O2 B_O3 B_O4 B_O5 B_O6 B_O7 B_O8 B_O9 B_O10 B_O11          !constraints in Biso
FST_CMD  conn C C 0 1.8   !Commands for VESTA program
FST_CMD  conn C O 0 1.8
```

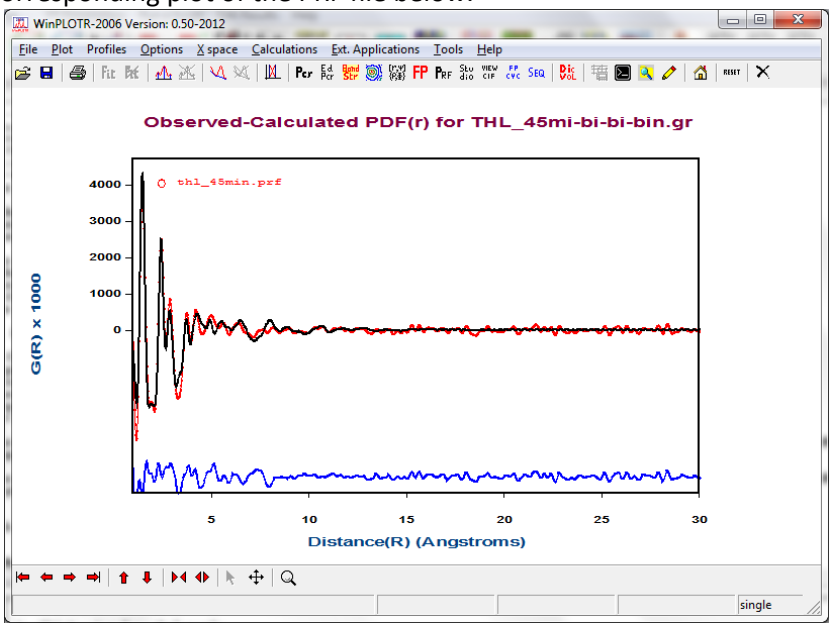
There are a number of blocks equal to the number of phases contributing to the patterns. The number following the keyword **PHASE_n**, is the weight fraction of the phase and it is a refinable parameter (only the first n-1 fractions are refinable, the last one is calculated so that the sum of all of them is one). These parameters are called **xfract** and if one needs to refine them it is necessary to use the directive **gvary xfract** in the respective PHASE-block. The explanation of the different keywords are provided in the comments starting with "!". The keyword **MODEL_SHARPF** admits three character values that are 1: **Exponent** or **Factor**, 2: **Restr_Dist**, **Intra_Dist** or **All_Dist**, 3: **Out_Dist** or nothing. The first value is the model used for the Molecular sharpening that may be an exponential model or a factor (see the document "Computational aspects of PDF refinement.pdf"), the second value tells to the program to which distances the molecular sharpening is applied (the distances written in the restraints file, all intra-molecular distances of all distances) and the third value (if present) makes the output of all distances to which the molecular

sharpening is applied in an output file called normally `my_CFL_file_dist_sharp.out` in which the applied model for sharpening is written as well as the initial value of the sharpening factor $s_{mol}$ applied to each distance.

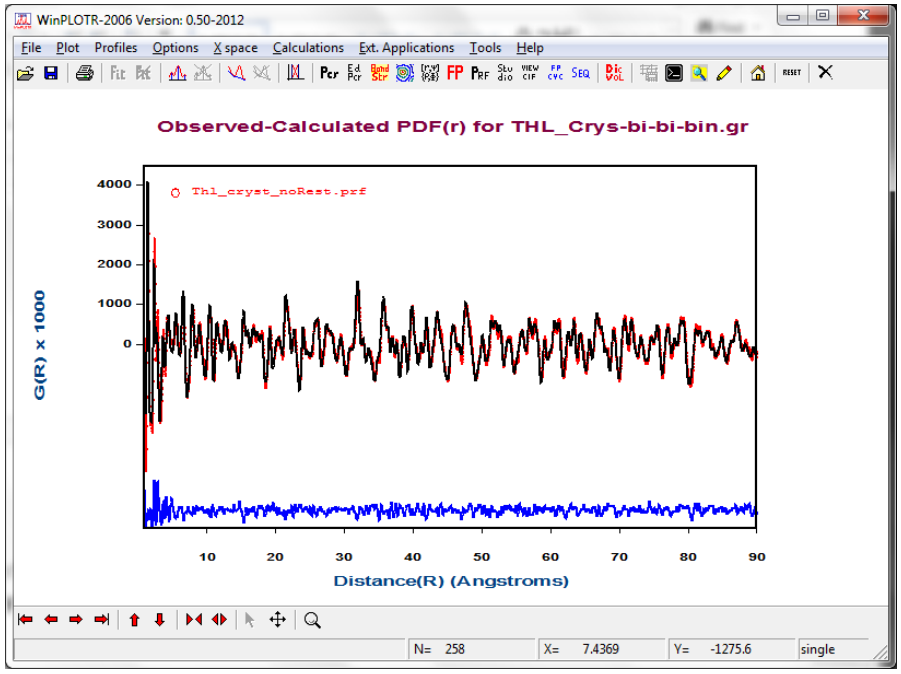
An example of running MolPDF from the command line is provided below



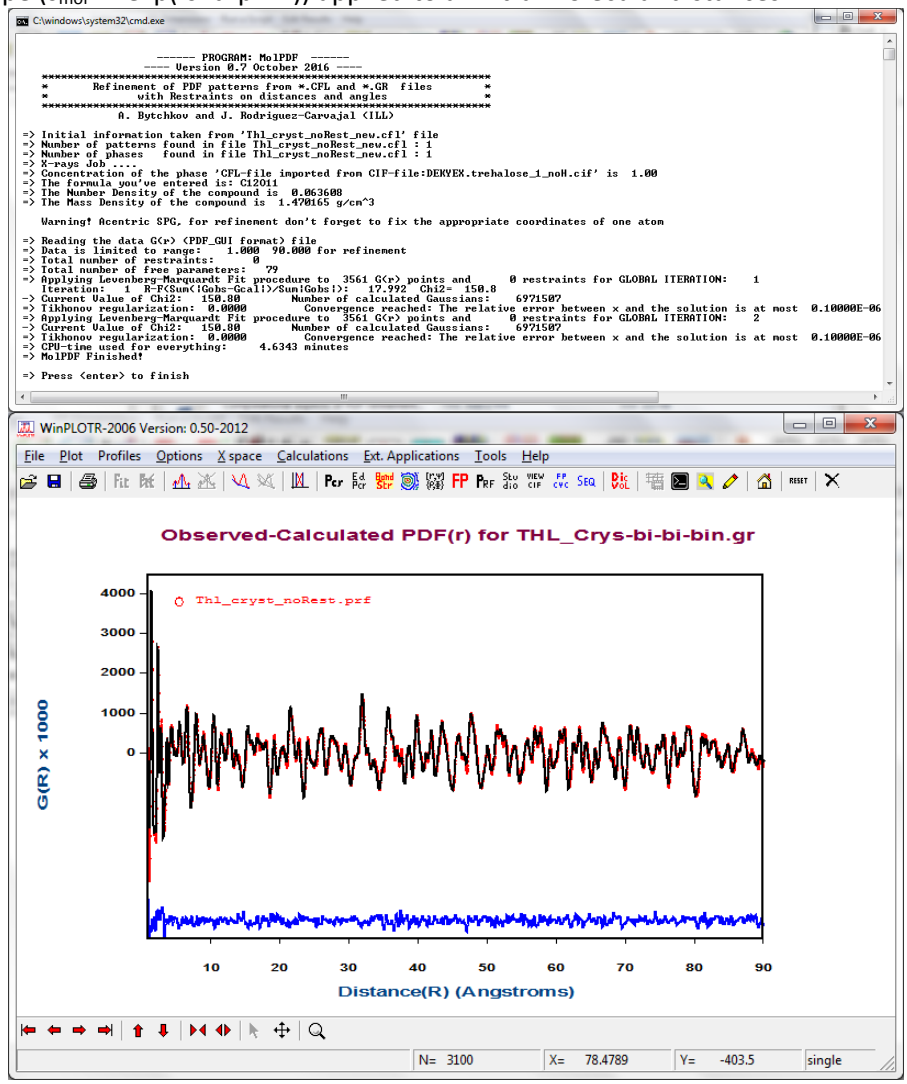

The above figure shows a refinement that is not good. The used model cannot be improved with the parameters used in the CFL file. In fact there is a mixture of phases that is not treated in the CFL file. See the corresponding plot of the PRF file below.




A much better refinement is shown in the following images corresponding to the refinement in a big range of the PDF (90 angstroms) of crystalline trehalose without restraints. The first image correspond to a refinement without the molecular sharpening factor.




The second image correspond to the same refinement but using the molecular sharpening factor of Exponent-type ($s_{mol}$=1-exp(-sharpf . r)) applied to all intra-molecular distances.



## Projected short-term improvements for MolPDF

Here we describe a series of improvement for working with MolPDF that will be introduced progressively.

**1**: Currently it is not possible to make constraints between different phases or different patterns. All constraints are intra-phase or intra-pattern. We envisage introducing general constraints between structural parameters of different phases using the same mechanism (through the directive "`equal`") as that for constraints within a phase.

**2:** Planar and torsion restraints have still to be implemented.

**3**: In some cases it may be useful to describe the damping of the signal as due to a bi-modal distribution of sizes. This allows the possibility to describe within a single phase cases that need normally two phases with different sizes.

**4**: Another improvement that may be needed is the use of general linear restraints between any kinds of free parameters.

**5**: Presently the selection of intra-molecular distances for applying a sharpening factor is done with the atoms within the asymmetric unit. In case there is more than a molecule the sharpening factors in also applied to inter-molecule distances. We envisage a mechanism for distinguishing the molecules by selecting set of atoms belonging to one or other molecule. For instance, atoms between C1 and C10, may be designed C1-C10 and an appropriate keyword like MOL-1, MOL-2, etc...

**6**: Create an option to use simulated annealing as method of optimization.