

# Manual to PROTPOW version 90\_01

## Contents

General description of the program.....	2
Initiation file (.ini-file) .....	3
Graphical interface and hot keys .....	3
File formats.....	4
File Menu .....	5
Settings Menu .....	10
Modify Menu.....	11
Multiple pattern Menu .....	12
Background Menu .....	13
Peak fit Menu .....	14
Plot Menu.....	16
Window menu .....	17

## General description of the program

PROTPOW is a program for visualisation and handling powder diffraction patterns and has been optimised for protein powder purposes. PROTPOW calculates powder patterns from coordinate files (PDB-format) including correction for bulk-solvent, or from structure factor files (CIF-format), and is thereby well-suited for identification of different crystalline protein polymorphs. Calculation of multiple patterns with different solvent correction parameters are also provided. Full pattern profile fits can be carried out for determination of unit cell and peak function parameters, and a powder pattern can be directly calculated using the found parameters.

PROTPOW can display up to five different powder patterns and perform different arithmetic operations or scale them relatively to each other for comparison. The program conducts a series of different straightforward corrections e.g. absorption corrections, corrections for Lorentz- and polarisation, conversion of wavelength and zero-shift calibration. For handling of background subtraction both manual and automated processes are available.

When using the program for publications, please cite one of the following articles:

Original publication on PROTPOW:

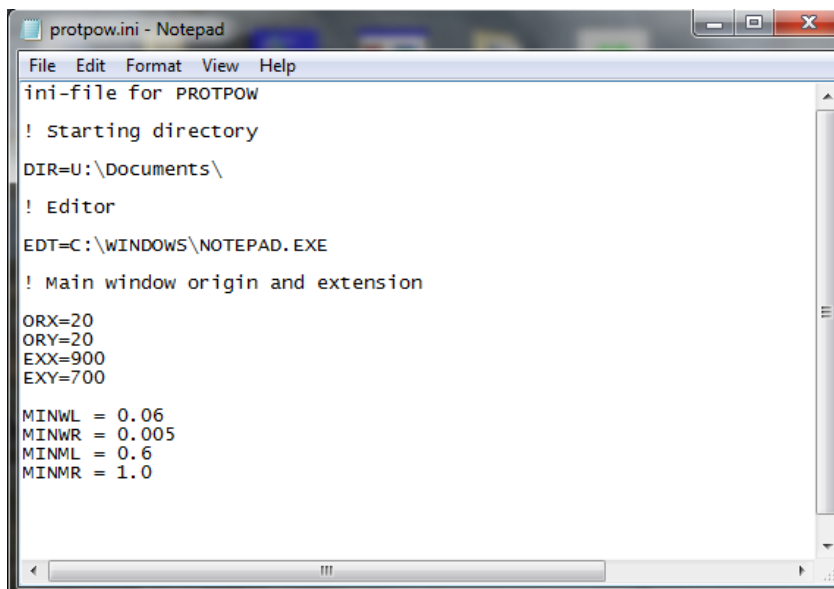
Hartmann, C. G., Nielsen, O. F., Ståhl, K. & Harris, P. (2010). *J. Appl. Cryst.* **43**, 876–882.

Recent publication on the method:

Ståhl, K. Frankær, C. G., Petersen, J. & Harris, P. (2013). *Powder Diffraction* **28**, S458–S469.

## Initiation file (.ini-file)

The program automatically generates an initiation file, which will be located in the same directory as PROTPOW. In the .ini-file you can instruct the program about editor, your preferred starting directory and window size (ORX, ORY, EXX, EXY), as well as default peak shape parameters (MINWL, MINWR, MINML, MINMR – see “Calculate pattern”). Locate the Notepad.exe or some other editor and specify it in the .ini-file. The paths to the four most recently opened files are also saved in the .ini-file.

A screenshot of a Notepad window titled 'protpow.ini - Notepad'. The window contains the following text:

```
File Edit Format View Help
ini-file for PROTPOW
! starting directory
DIR=U:\Documents\
! Editor
EDT=C:\WINDOWS\NOTEPAD.EXE
! Main window origin and extension
ORX=20
ORY=20
EXX=900
EXY=700
MINWL = 0.06
MINWR = 0.005
MINML = 0.6
MINMR = 1.0
```

Figure 1: Initiation file.

## Graphical interface and hot keys

The program runs with a graphical interface where the pattern is viewed and an editor interface where output comments are written. F3 and F4 show both interfaces at the same time: on top of each other (F3) or side-by-side (F4). Dragging with the mouse you can zoom on the pattern plot. Using the arrow keys you can move an expanded window left or right, zoom out (up) or zoom in (down). The full pattern is restored with F1. Hot key for undo is Ctrl+Z.

## File formats

PROTPOW is able to handle different types of pattern formats:

- .gdf (Huber): A standard Huber format.
- .txt (Bruker): A standard Bruker format.
- .uxd (Bruker): A standard Bruker format.
- .udf (Panalytical): A standard Panalytical format.
- .xrdml (Panalytical): A standard Panalytical format.
- .xy (General): A two column file listing the  $2\theta$ -values in the first column and the intensities in the second.
- .std (General): Gives the  $2\theta$  start,  $2\theta$  step size,  $2\theta$  end in the first line separated with comma. The intensities are listed from line two with 10 per line.
- .CIF (General): The file contains a header using the standard CIF code words, where the file name and directory, the  $2\theta$  start,  $2\theta$  end and the  $2\theta$  step size are specified, followed by the intensities.

## File Menu

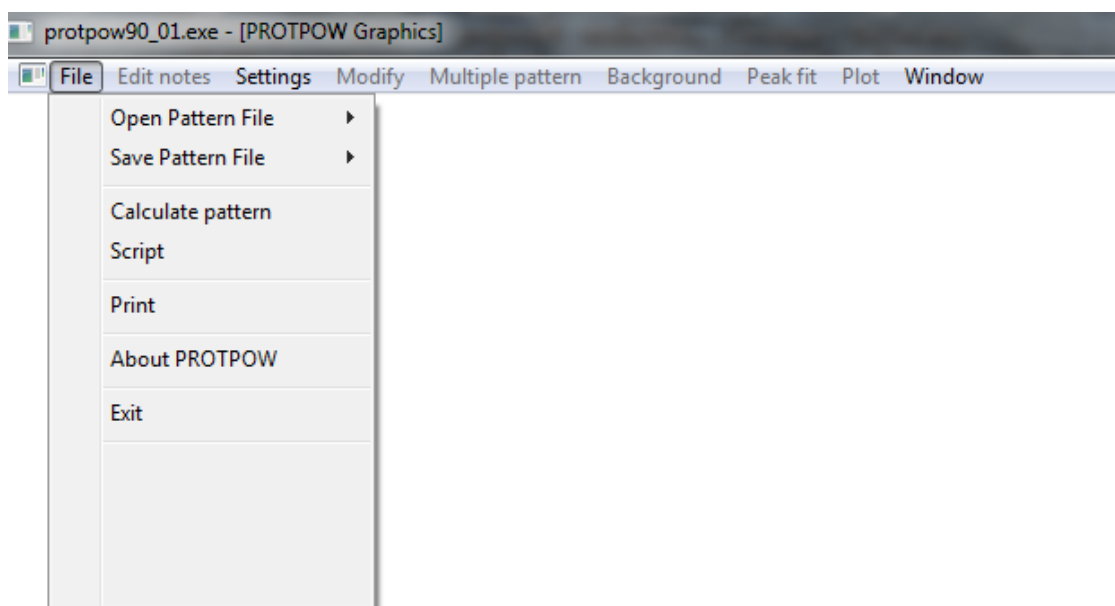


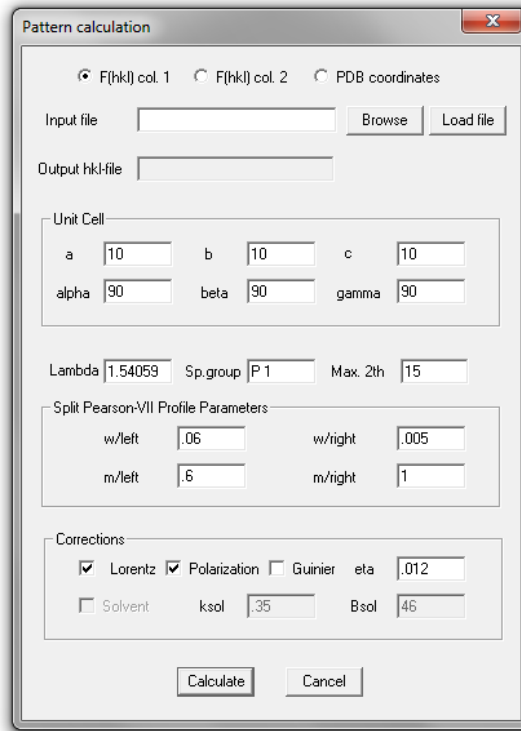
Figure: 2 File menu.

**Open Pattern File:** PROTPOW is able to open powder diffraction data in different kinds of formats: .gdf (Huber), .txt and .uxd (Bruker), .udf and .xrxml (Panalytical) and the general formats: .xy, .std and .CIF. (see "File formats")

**Save Pattern File:** PROTPOW can save powder diffraction data using the following formats: .gdf, .xy, .std and .CIF. The  $2\theta$ -range and the  $2\theta$ -step have to be specified along with the type of pattern: Input pattern (default), calculated pattern, background. When increasing the step size the data point in the saved pattern file will be extrapolated.

**Calculate pattern:** The calculation pattern menu is shown in Figure 3. The calculation of a powder patterns uses either structure factors or coordinates. The structure factor input file must be in cif-format (.hkl-file) listing the hkl-reflections either calculated or observed (col. 1 and col. 2). The coordinate input file format is pdb.

State the unit cell parameters and space group either manually or press "Load file" to insert those stated in the input file. For structure factor files (.hkl-files) the parameters are stated in the header. For pdb-files the parameters read are those stated after the code word CRYST1 following the general pdb-format. Hereafter, state the wavelength  $\lambda$  and maximum  $2\theta$ -value.



**Figure 3: Calculate pattern menu.**

Peak profile parameters controlling the peak width and shape of an asymmetric split Pearson VII function, can be stated if known (e.g. from a full-pattern profile fit). Otherwise a parameter set must be found and optimized for each experimental setup. The default values can be specified in the .ini-file. The split Pearson VII function is defined as:

$$P_l = \frac{C_l}{H_l} \cdot \left( 1 + 4 \left( 2^{\frac{1}{m_l}} - 1 \right) \frac{(2\theta - 2\theta_B)^2}{H_l^2} \right)^{-m_l}$$

Where  $m_l$  is the shape exponent for the left side contribution,  $2\theta_B$  is the scattering angle at the Bragg reflection maximum,  $C_l$  and  $H_l$  are given as:

$$C_l = \frac{2\sqrt{m_l} \cdot \sqrt{2^{\left(\frac{1}{m_l}\right)} - 1}}{\sqrt{m_l - \frac{1}{2}} \cdot \sqrt{\pi}}$$

$$H_l = \sqrt{w_l}$$

Where  $w_l$  is the FWHM similar to the  $w$  Cagliotti parameter. The right side contribution  $P_r$  is analogously given using  $m_r$  and  $w_r$ . Continuous peak profiles are obtained by normalization of the left and right side contributions.

Finally specify whether the following post-calculation corrections should be applied. Each multiplies the raw calculated pattern by the following factors:

Lorentz:	$\frac{1}{\sin(2\theta)} \frac{1}{\sin(\theta)}$
Polarization:	$1 + \cos^2(2\theta)$
Guinier ( $\theta_G = 45^\circ$ ):	$\frac{1}{\cos(\theta_G - 2\theta)}$
Additional Lorentz correction, $\eta$ :	$\frac{\sin^2(\theta)}{(\sin^2(\theta) + \frac{\lambda^2 \eta^2}{12})}$

For further details see Hartmann *et al.*, 2010.

When performing calculations from pdb-coordinate files it is necessary to apply solvent correction. The contribution  $\mathbf{F}_{\text{sol}}$  from disordered solvent in the channels surrounding the protein molecules are added to the structure factors calculated for the macromolecule  $\mathbf{F}_{\text{mol}}$ . Hence solvent correction is not applicable for calculation from .hkl-files.

In the correction the macromolecular and the solvent regions are respectively defined by a mask function.

- 1) *Defining a grid*: A grid is defined over the asymmetric unit. The maximum grid size is 1.0 Å, which is then reduced by division by the grid reduction factor (default is 1.33, resulting in grid size of 0.75 Å). Higher values of grid reduction factors will slow the calculations.
- 2) *Calculate Van der Waals radii*: The extension of all ordered atoms described in the pdb-file (i.e. those belonging to the protein, ligands and ordered solvent molecules) is defined by their Van der Waals radii.
- 3) *Initial definition of the regions*: A sphere with radius  $R_{\text{probe}}$  (default is 1.11 Å) is rolled over surface of the defined atoms. The volume (i.e. grid points) occupied by  $R_{\text{vdw}} + R_{\text{probe}}$  is defined as belonging to the macromolecule.
- 4) *Re-definition of the regions*: A sphere with radius  $R_{\text{shrink}}$   $R_{\text{probe}}$  (default is 0.9 Å) is rolled on the inside of the macromolecular region. The grid points touched by  $R_{\text{shrink}}$  is redefined to belong to the solvent region. This two-step procedure allows smoothing of smaller pockets and channels in the protein as defined by its atom's Van der Waals radii.
- 5) *Insertion of flat solvent density*: A flat solvent density level ( $k_{\text{sol}}$ ) is inserted in the solvent region and the steepness of the boundary between the two regions is given by  $B_{\text{sol}}$ . Default average values are 0.35 e/Å<sup>3</sup> and 46 Å<sup>2</sup> for  $k_{\text{sol}}$  and  $B_{\text{sol}}$ , respectively (Fokine & Urzhumtsev, 2002).

When calculating from coordinate files a hkl-file is first written containing structure factors in which the contribution for the bulk-solvent has been added. This file is afterwards used for calculation of patterns. This means that it is not necessary to calculate patterns from scratch if e.g. the profile parameters should be changed, as the solvent correction calculation might be time consuming in particular for large protein structures.

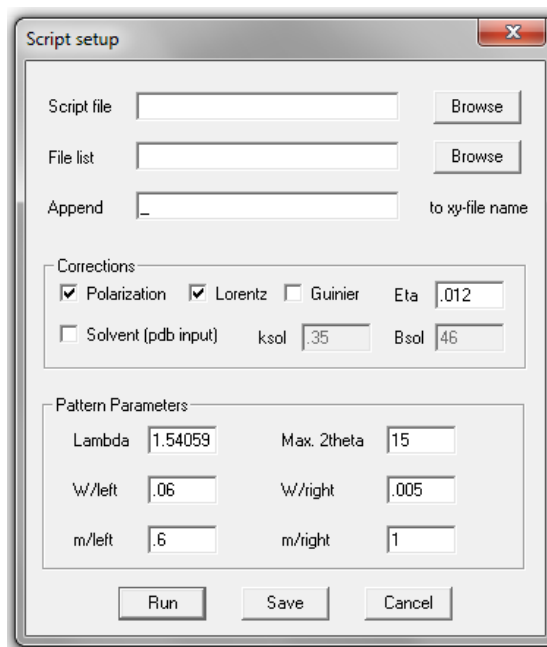


Figure 4: Script setup menu.

Script:

Feature for calculating multiple patterns. A procedure can be saved as a script file by clicking “Save”. Such procedures can be loaded into “Script file” – but are not mandatory. In “File list” a plain text file with the extension .mlt is uploaded. This file contains all relevant pdb-files or hkl-files (one file name on each row, see Figure 5), and is located in the same directory as the pdb-files. A text string can be appended to the file names in “Append”. Pattern parameters and correction are given. When calculating from hkl-files, solvent correction cannot be performed. Unit cell parameters and space groups are read from the pdb- and/or hkl-files.

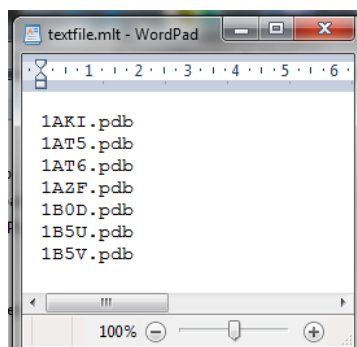


Figure 5: Format of “file list”-file (.mlt).



Print: Prints the displayed part of the powder pattern in portrait view with file details.

About PROTPOW: Details about the program.

Exit: Exits the program

The four most recently used files, which paths are saved in the .ini-file, can be accessed from the file menu.

## Settings Menu

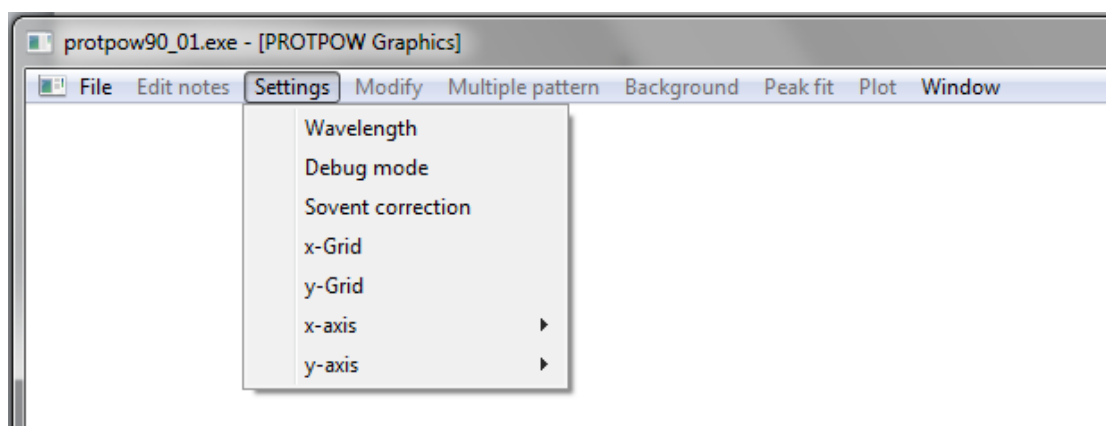


Figure 6: Settings menu.

Wavelength:	Set the operational/reference wavelength
Debug mode:	For internal use. May help finding where the program crashes.
Solvent correction:	Set the solvent correction parameters $R_{\text{probe}}$ , $R_{\text{shrink}}$ and Grid reduction factor, as defined under "Calculate pattern".
x-Grid:	Shows a grid perpendicular the x-axis
y-Grid:	Shows a grid perpendicular the y-axis
x-axis:	Choose between $2\theta$ , $d$ or $Q$ -values
y-axis:	Choose between linear or logarithmic scale

## Modify Menu

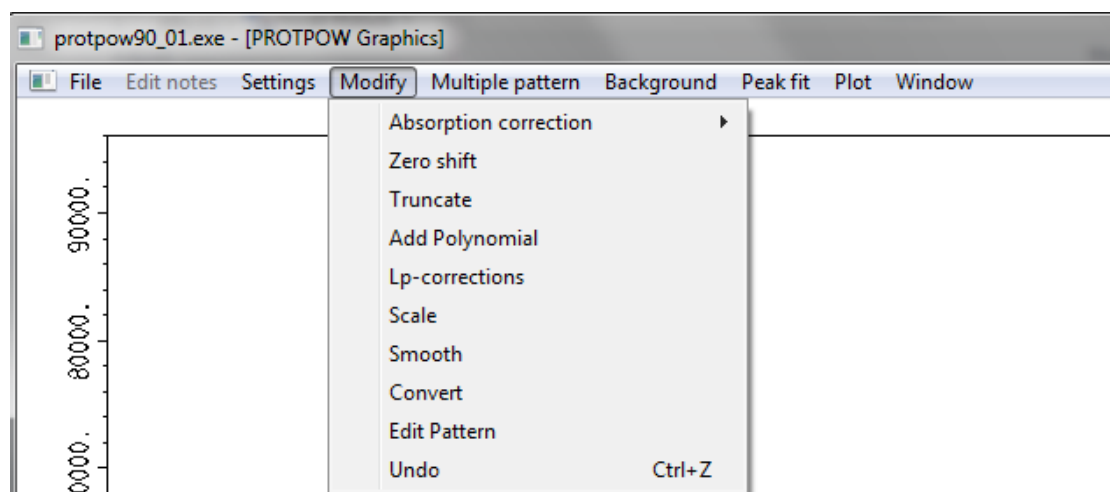


Figure 7: Modify menu.

Absorption correction:	For capillary state the absorption coefficient $\mu$ , and the radius $R$ . For flat plates state the absorption coefficient $\mu$ , the thickness $x$ , and the inclination angle.
Zero shift	Translation along the x-axis for calibration of the zero point.
Truncate	Truncates the pattern and sets the step size between the data points.
Add polynomial	Addition of a polynomial $A_0 + A_1 x + A_2 x^2 + A_3 x^3 + A_4 x^4 + A_5 x^5$ . Maximal order is fifth order.
Lp-corrections	Forwards and backwards corrections for Lorentz, polarisation and Guinier tangent angle are conducted by choosing different combinations of the given factors.
Scale	Manually scaling of a pattern.
Smooth	Smoothens the pattern. The higher the value the more smoothing effect is obtained
Convert	Conversion from one wavelength to another.
Edit Pattern	Manually editing the pattern file
Undo	Undo the previous action

## Multiple pattern Menu

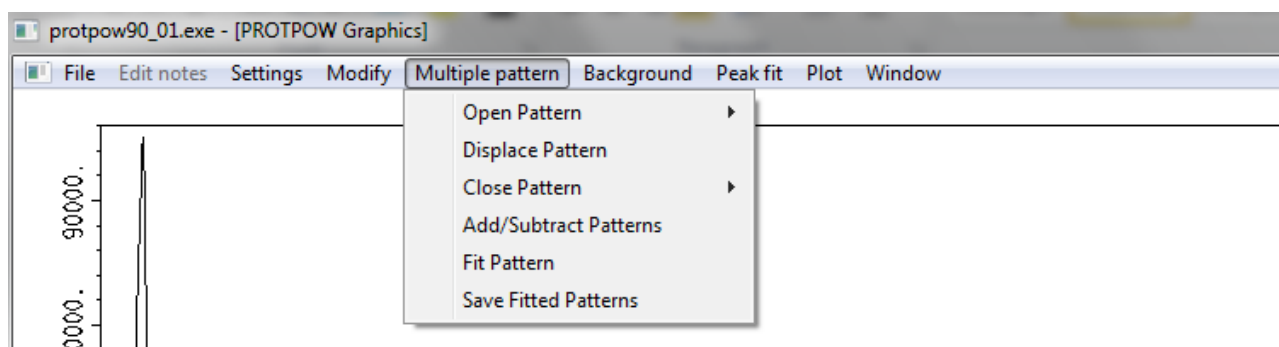
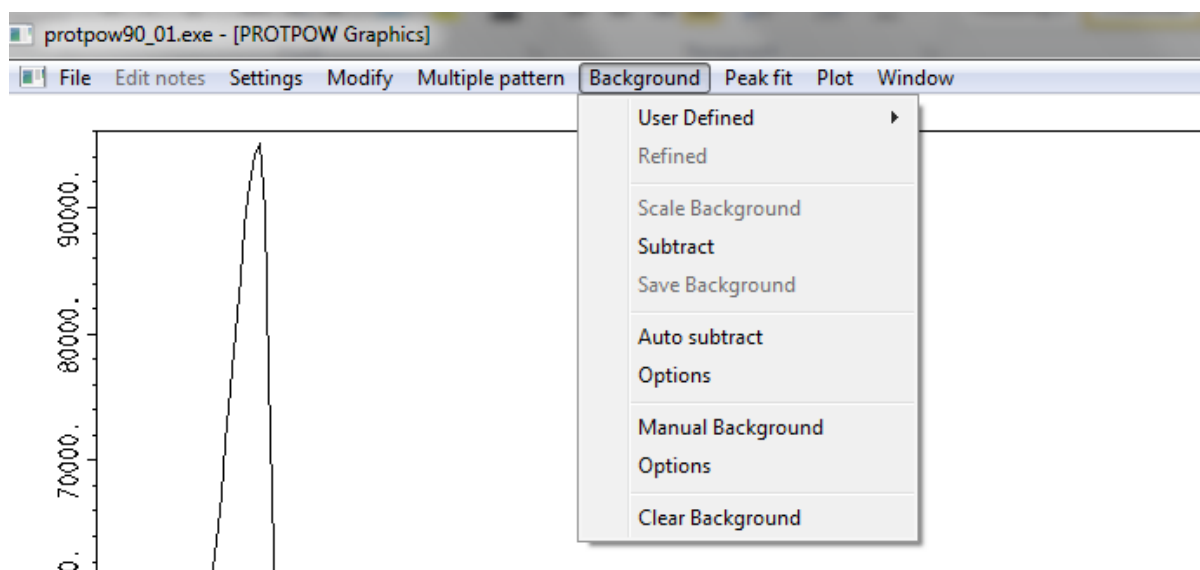


Figure 8: Multiple pattern menu.

- Open pattern:** Opens another powder pattern which is showed in the same window as the original pattern. PROTPOW can handle up to five open patterns simultaneously, i.e. four multiple patterns.
- Displace Pattern:** Translation of the opened patterns along the  $x$ -axis ( $2\theta$ ) or the  $y$ -axis (Intensity). The values given determine the spacing between each pattern.
- Close pattern:** Closes a powder pattern.
- Add/Subtract Patterns:** Adds/subtracts two patterns together/from each other. There is an option for scaling one of the patterns simultaneously. This feature works with two open patterns only.
- Fit Pattern:** Fits two patterns to each other. This feature only works with two open patterns. The scale factor can be refined, and an  $n^{\text{th}}$  order Chebyshev polynomial ( $n$  is an integer between 1 and 14) can be included in order to simulate a background. Furthermore zero point correction can be included along with a revised Lorentz correction which may influence a protein powder pattern (for details see Hartmann *et al.* (2010) *J. Appl. Cryst.* **43**, 876). It is not possible to refine the latter two parameters, but optimisation can be carried out iteratively by comparing the profile  $R$ -factor which is evaluated after each step. Profile  $R$ -factors are shown to the left in the lower status bar.
- Save Fitted Patterns:** Saves all the active patterns in one file. The file is in .xy-format with  $x$ -coordinates in the first column and  $y_1, y_2 \dots y_n$  in the following columns, and the background in the last column.

## Background Menu



**Figure 9: Background menu.**

- User Defined: Read an experimental background powder pattern
- Scale Background: Scales an automatic chosen background level
- Subtract: Execute the subtraction of a background.
- Save Background: Saves the background in a .bxy-file with  $2\theta$ -values in the first column, brutto intensity in the second and the background intensity in the third column.
- Auto subtract: Automatic generation of the background level using “Robust smoothing” (Brückner, J. Appl. Cryst. (2000), 33, 977-979). The signal below this line is interpreted as background and subtracted automatically.
- Options: Setting the smoothening window and number of smoothening cycles which are the two parameters defining the sensitivity of the automatic background correction
- Manual Background: Opens a tool where points can be picked manually. The background level is defined by connecting the points picked. Subtraction is executed by choosing “Subtract” in the Background menu.
- Options: Sets the way the manually picked points should be connected: Linear steps, fitting a spline or a Chebyshev polynomial.
- Clear Background: Clears the actual background.

## Peak fit Menu

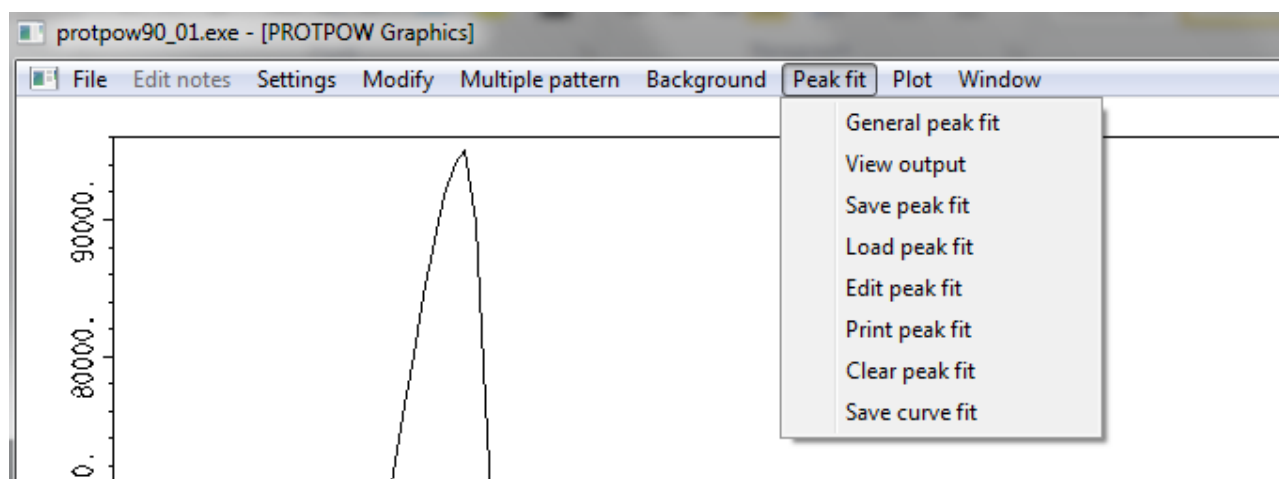


Figure 10: Peak fit menu.

General peak fit: Opens a GUI for full pattern profile fits, Figure 11. Firstly, the structure factor file (CIF-format) is loaded for cross-validation of the scaling of the intensities. Use the "Browse" button. Select the corrections that should be applied (see "Calculate pattern"). An estimate of the scale may be entered, and refined later. Click "Load" for copying the unit cell parameters and Laue class given in the .hkl-file loaded, or enter them manually.

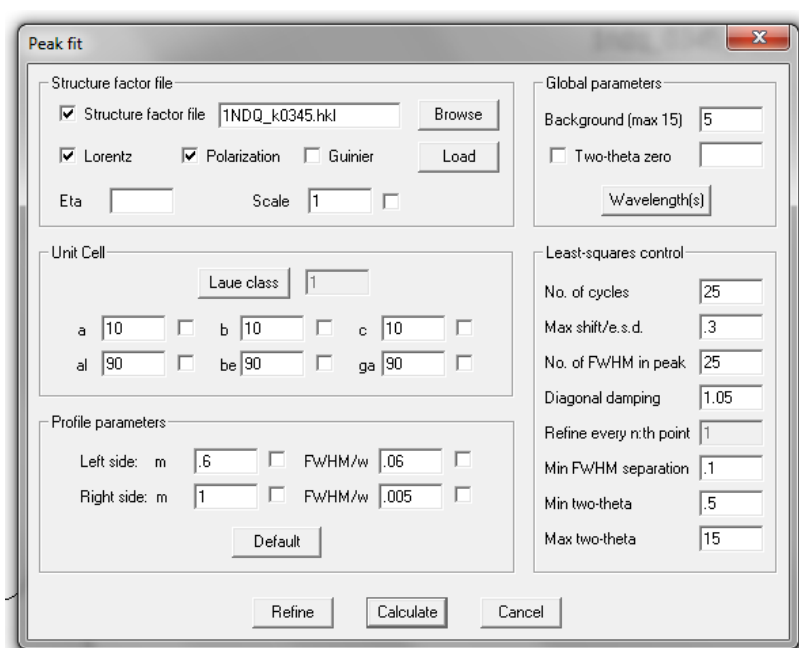


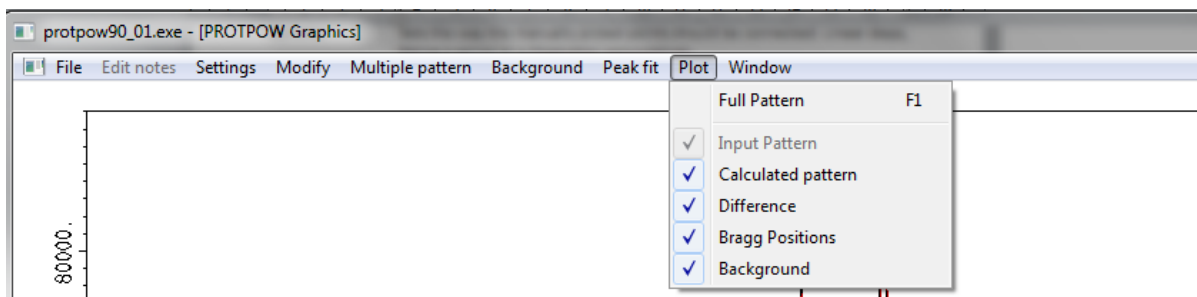
Figure 11: General peak fit menu.

Peak profile parameters ( $w_l$ ,  $w_r$ ,  $m_l$  and  $m_r$ ) that take into account asymmetric effects are given as a split Pearson VII function describing respectively the left and right side contribution are entered. Default values are given in the .ini-file. Global parameters are stated at the top right box: number of coefficients/polynomial order for simulation of the background with a Chebyshev polynomial and the zero point correction. Parameters concerning the fitting procedure (e.g. iteration cycles) are stated in the bottom right box along with the  $2\theta$ -range.

The full pattern profile fit is performed step by step: Start with realistic values of the unit cell parameters and perform a calculation. Start to slowly refine the parameters one by one by ticking off the small tick-boxes next to the values. Start with the scale factor, then cell parameters and zero-point correction, and finally background and profile parameters. When changing the  $2\theta$ -range it is necessary to do another calculation before refinement.

- View output: Shows the output-file "%refine.out" from the refinement, which is written to the working directory. This file is overwritten every time a refinement is performed.
- Save peak fit: Saves the parameters from the fit in a .fit-file.
- Load peak fit: Loads the parameters from a saved fit.
- Edit peak fit: Opens an editor for editing in the .fit-file.
- Print peak fit: Prints the .fit-file.
- Clear peak fit: Clears the parameters
- Save curve fit: When a good fit is obtained the curves can be saved as an .xy file. The first column is the  $2\theta$ -values, second column is the fitted pattern inclusive background, third column is the original pattern, fourth column is the separate background as modelled by the Chebyshev polynomial.

## Plot Menu



**Figure 12: Plot menu.**

Full Pattern: Plots the full pattern e.g. after a zoom.

The following functions are accessible to control the plot:

Input Pattern: Plots the full pattern.

Calculated pattern: Plots the calculated pattern.

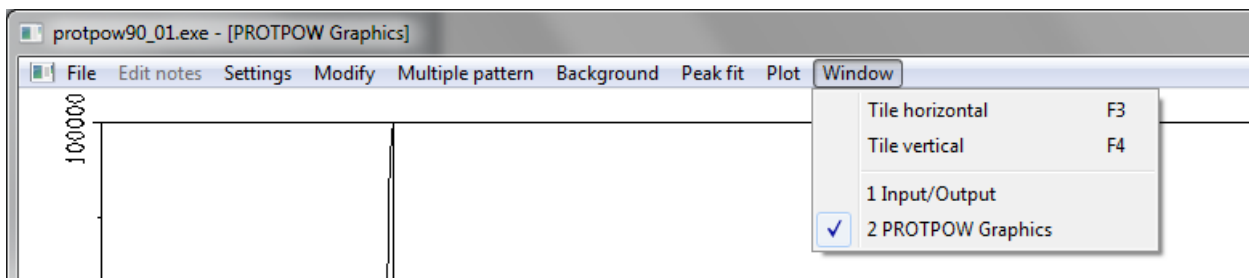
Difference: Plots the difference between the full pattern and the background.

Bragg positions: Plots the positions of possible Bragg peaks.

Background: Plots the background



## Window menu



**Figure 13: Window menu.**

- Tile horizontal: Show both the input/output and the graphics window on top of each other.
- Tile vertical: Show both the input/output and the graphics window side-by-side.
- 1 Input/Output: Input/output window on top.
- 2 PROTPOW Graphics: Graphics window on top.