

NEW PROGRAM IN WOLFRAM MATHEMATICA: CALCULATE PEAKS EASILY

A.S. Kazantsev¹, Yu.L. Ryzhykau¹, A.V. Vlasov¹, A.I. Kuklin^{1,2}

¹Moscow Institute of Physics and Technology, Dolgoprudny, Russian Federation

²Joint Institute of Nuclear Research, Dubna, Russian Federation

E-mail: kazantsev@phystech.edu

Experimental data obtained by small-angle scattering may contain diffraction peaks. Simple and fast calculation of its parameters is an important task for structure-obtaining workflow. We developed a program in Wolfram Mathematica. The program provides an interface for loading and processing sets of points with errors using predefined but mutable parameters (Figure 1).

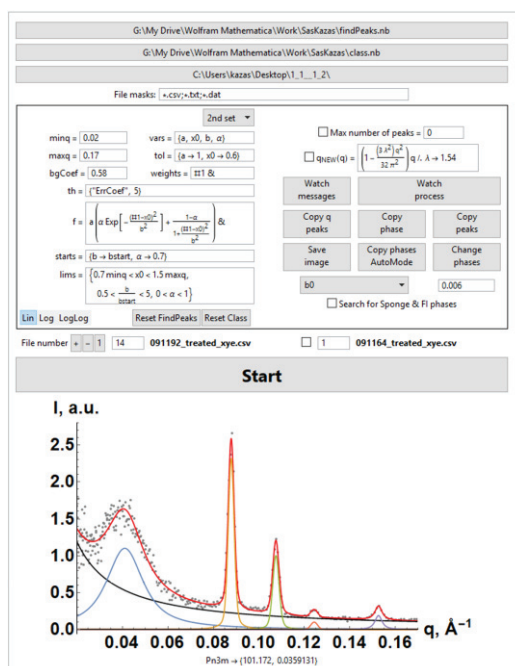


Figure 1. The interface of the program. The upper part contains parameters of loading and processing, the lower one – a processed picture. Legend of the picture: gray dots are experimental data, black line is a background signal, calculated by the program, multicolor curves are peaks, calculated by the program too, and the red line is the final approximation.

The program was initially developed for faster defining of lipidic cubic phase types. Therefore, it does two main things: calculation of peaks and definition of exact phase set by proportions of peaks' parameters.

Calculation of peaks is conducted by approximation of experimental data by Goal fit function:

$$(A \cdot x^b) + \sum_i PeakFunc(x)$$

“PeakFunc” is a peak function; its form may be defined by user. Number of peak functions in the sum equals to number of peaks, found by the program. $A \cdot x^b$ – background, which parameters are calculated by the program too, as described below.

The pipeline of approximation of experimental data consists of the following.

- Make approximation of experimental data by background.
- Subtract $k \cdot$ background ($k \leq 1$ and user-defined) from data.
- Search for groups of experimental points, which intensities are higher than user-defined threshold.
- Repeat until there are no such groups:

- For each group:
 - select a point of maximal intensity;
 - make parabolic approximation of nearest to it points;
 - approximate nearest points by user-defined “PeakFunc”, using parabolic approximation params as start values.
- Nonlinear Model Fit of the Goal fit function; start values are obtained from the last approximation for all peaks.

Approximation of experimental data has calculation parameters. They are presented in Wolfram Language style. The description of the most important of the parameters is presented below.

A peak function is customizable. A user can change the predefined function (gaussian) to his own one. A function should contain parameter “a” as height, “b” as width, and “x0” as a peak center. A function may contain any other parameters. All function parameters should be presented in “vars” parameter.

A user is able to define custom constraints for peak function variables in “lims” parameter.

The default value is $\{0.7\text{minq} < x0 < 1.5\text{maxq}, 0.5 < \frac{b}{\text{bstart}} < 5\}$. In this parameter several special symbols may be used: “minq” and “maxq” are the abscissa limits for loaded data, and “bstart”, which is a first (parabolic) approximation for peak width.

The “starts” parameter defines which values the main approximation starts from for all parameters except “a” and “x0”.

The second part of the program is definition of lipidic mesophase set. The main characteristic for phases’ definition is proportions between coordinates of peak centers (q-ratio) and between their heights [1-3]. These proportions reveal the crystallographic group of lipidic cubic phases (LCP) or type of non-crystal structure, and their lattice parameters. The program is able to define a set of different types of mesophases in a sample. In addition, search for Sponge and Fluid Isotropic phases (wide peaks) and for $L\alpha$ (laminar) phase with only 1 peak is also provided by the program.

To understand the significance of the program, it was compared to known analogs. In the part of peak centers calculation, OriginPro 2016 was chosen. Test on a standard experimental data, obtained from Pn3m LCP, failed: we were not able to tune approximation process correctly. In general, all attempts to approximate at least five-peaks-data with OriginPro were not successful. In the part of lipidic mesophase set definition, comparison was conducted with [3]. The crucial difference is that comparator cannot distinguish multiple mesophases.

The program is not limited by small-angle diffraction experimental data. Fill free to use it for any data containing peaks no more than 20.

Thus, our novel program fills empty space in data processing necessities.

[1] Hahn T. International tables for crystallography Volume A: Space Group Symmetry. 2002.911 p.

[2] Cherezov V. The Cherezov Lab - LCP Resources: Useful Equations [Electronic resource]. URL: http://cherezov.usc.edu/lcp_equations.htm (accessed: 26.04.2017).

[3] Joseph J.S. et al. Characterization of lipid matrices for membrane protein crystallization by high-throughput small angle X-ray scattering // Methods. Elsevier Inc., 2011. Vol. 55, № 4. P. 342–349.