Manual for AutoFIDEL v. 0.4

Jonas Nyman

March 11, 2021

AutoFIDEL is a Python 3 program for matching PXRD-patterns to crystal structures. It is intended to be used for solving the structure from a PXRD-pattern by comparisons to CSP-generated structure candidates, without the need for indexing.

The script requires the CCDC Python API and the NumPy and SciPy libraries.

The program calculates a similarity score between PXRD patterns by means of a cross-correlation function. The theory is described elsewhere [1, 2]. The similarity score takes values between zero and unity.

The lattice parameters of the crystal structures are optimized with respect to the powder pattern similarity. The lattice parameters are allowed to change by $\pm 5\%$, which approximately corresponds to the discrepancies that can be expected due to thermal expansion and errors in computational methods.

It is crucial to have good experimental PXRD-data. Data collected in transmission mode is strongly recommended, since preferred orientation cannot be accounted for. The experimental PXRD-patterns must be background-corrected and any spurious peaks at very small 2θ should be removed. The patterns may be truncated at $2\theta = 4$ or 5°, provided there are no peaks there.

The upper bound on the PXRD data is 35° by default, but can be set manually with the '-t' option. The upper bound should be set to include all the large peaks in the "fingerprint region" of the powder pattern, but not higher. The 2θ step size in the PXRD-pattern should preferably be smaller than 0.15° and equidistant.

The program takes the following arguments:

-p Input PXRD-pattern file, either an .xy or .xye file. The X-ray wavelength is read from this file.

-c CIF file containing one or several candidate crystal structures.

-s Optimize only those structures with at least this much initial similarity. The default value is 0.3. To make a single-point calculation without optimization, set this to 1.0.

-t Upper 2θ cutoff in degrees. PXRD-data above this value are not used. The default value is 35.

The optimization within AutoFIDEL only optimizes the cell parameters, so the optimized structures will have (very slightly) distorted molecular geometries. Distances between atoms typically change by on the order of 1/100 Å. The structures are good starting points for indexing, Rietveldt refinement or further geometry optimization with fixed-cell periodic DFT-D or CrystalOptimizer.

For versions up to 0.32, the similarity score could appear to decrease during optimization. The similarity score could also exceed 1.0 for purely numerical reasons. These artifacts have been corrected in this version.

The optimization is internally performed in two stages, first a "coarse" optimization is done with a wide window, meaning that the peak positions are allowed to be relatively far apart, yet still give a high score. This ensures that we find the matching structure even if the initial similarity is poor. After the first optimization, the peak positions for the correct structure should essentially overlap with the experimental pattern, so the window size is reduced, and the similarity scores are optimized again, with a very narrow window. This ensures that we get a high score only for those structures that really have a good pattern similarity. The L-BFGS-B method is used for the optimization, meaning that a good starting guess is needed and there is no guarantee that the global optimum is found.

The optimization algorithm works best for primitive ("most orthogonal") unit cells. It may also be helpful to try different settings such as $P2_1/a$ vs. $P2_1/c$. The setting can be changed in Mercury, and the most orthogonal unit cell can be obtained with PLATON's LEPAGE command.

The program writes a small log file (autofidel.log) with results showing the most plausible matches, and a CIF file (matches.cif) with the optimized matching structures.

A final similarity score below 0.75 is almost certainly not the correct structure, provided good quality, background-corrected PXRD-data is used. Preferred orientation or contamination can however cause even the correct structure to get a low score.

A final score above 0.8 is indicative of a possibly correct match. Structures with final scores above 0.8 should be further investigated, for instance by trying alternative settings of the cell, which may allow the optimization to reach the global optimum. Structures with scores above 0.8 should be used in attempts to index the pattern. A successful indexing and Pawley refinement is a definitive confirmation that an essentially correct structure has been found.

If no candidate is a definitive match, it can be useful to look at how much similarity the structures gained during the optimization. Incorrect structures will typically not increase much in similarity score, while the correct structure tends to have the highest gain. The structures with the largest gains are now reported in the log file. If the structure with the highest score is also the structure with the highest gain, that is strongly indicative of a possible match.

Two small utility scripts are included. The script equalize.py makes experimental powder patterns strictly equidistant in 2θ. It is not necessary to use it, but the cross correlation and similarity scores will not be exact if the pattern is not equidistant.

The script xy2xye.py converts .xy patterns to the .xye file format. Again, it is not necessary to use it, AutoFIDEL can read either file format. It is useful if only .xy data is available and the pattern needs to be background-corrected in DASH.

The 10 polymorphs of galunisertib and their experimental powder patterns are included in the Examples folder for testing purposes. AutoFIDEL correctly identifies all of them. Three polymorphs of ROY are also included. The ON and Y PXRD patterns are simulated from the experimental single-crystal structures, so the similarity score should reach unity. In all cases, the provided unit cells have been distorted so as to showcase the effect of the optimization.

References

[1] René de Gelder, Ron Wehrens, and Jos A Hageman. A generalized expression for the similarity of spectra: application to powder diffraction pattern classification. *J. Comput. Chem.*, 22(3):273–289, 2001.

[2] Stefan Habermehl, Philipp Moerschel, Pierre Eisenbrandt, Sonja M Hammer, and Martin U Schmidt. Structure determination from powder data without prior indexing, using a similarity measure based on cross-correlation functions. *Acta Cryst. B*, 70(2):347–359, 2014.