

## HIGH THROUGHPUT POWDER DIFFRACTION - THE SEARCH FOR POLYMORPHS

Chris Gilmore and Gordon Barr, Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland, UK.

High throughput powder X-ray diffraction is now an established technique for searching for polymorphs, screening salts and optimizing crystallization procedures for the development of new drugs. Such an experiment produces copious quantities of data which need to be processed in an automatic or semi-automatic way in order to fully benefit from the automation procedures, and to ensure that the data are fully investigated (there is a limit to how long any operator can look at powder patterns on a screen!). A suitable procedure will be demonstrated using the SNAP-1D program, but the principles can be applied anywhere.

SNAP utilizes the full diffraction profiles of patterns, unlike other programs which usually represent patterns as a list of d-spacings and corresponding peak intensities. This is important when you are searching for changes in diffraction pattern that are relatively small or the pattern is dominated by one or two low-angle invariant peaks. Powder patterns can be imported from a variety of file formats into a proprietary SNAP database. Patterns can be processed to remove background artifacts and to reduce noise using wavelets. An automatic peak-picking algorithm may also be employed. Pre-processing data in this way needs care if important detail is not to be lost. The program then performs two main functions:

1. The pattern-matching module answers the question: To what patterns in an existing database is a new unknown pattern most similar? User-controlled combinations of both non-parametric and parametric statistical tests are utilized to answer this. Matching results can be displayed in numerical or graphical form. This process can be used for comparison of patterns, or in sample quality control, for example.
2. In quantitative mode, given a mixture pattern known to consist of a combination of other patterns already in the program database, SNAP-1D can produce a list of suggested component phases and their relative weight fractions in the mixture. This is done using a variation of a generalized inverse linear least squares technique that is much more straightforward to use than Rietveld refinement-based analysis systems. Identified components can be examined in a variety of different ways.

What happens when there is no reference sample(s)? Other techniques of multivariate data analysis can then be applied including metric multidimensional scaling, minimum spanning trees, clustering, silhouettes, principal component analysis with their associated visualization tools. These techniques will be discussed. Up to 1000 patterns can be processed at one time in less than 15 minutes on a PC.