

## **APPLICATION OF HIGH-THROUGHPUT CLUSTER ANALYSIS TO MULTIPLE DATA TYPE – COMBINING DATASETS**

G.J. Cunningham, G. Barr, W. Dong, C.J. Gilmore, WestCHEM, Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland; C. Frampton, Pharmorphix Limited, 250 Cambridge Science Park, Milton Road, Cambridge, CB4 0WE

A method has been developed<sup>1</sup> to pattern match full profile high-throughput powder diffraction data. This uses Pearson and Spearman correlation coefficients to measure the similarity of the profiles of each pattern with every other pattern to create a pattern-pattern correlation matrix. This is then used to partition the patterns into similar groups using a variety of cluster analysis methods. These techniques have been integrated into the program PolySNAP 2<sup>2</sup>. The methodology was originally developed solely for powder X-ray Diffraction (PXRD) data, but has been extended to Raman, differential scanning calorimetry (DSC) and infrared (IR) data.

The correlation matrixes from two or more datasets can be combined to give a new correlation matrix. This new correlation matrix can then be used to partition the datasets into clusters. Getting the same results by combining datasets as is given by the individual datasets allows the user to have much greater confidence in their results. Combining a 'poorer' quality dataset and a 'good' quality dataset can also allow the user to gain a better understanding of what the individual datasets are showing.

The effect of combining PXRD, Raman, DSC and IR data will be studied using data from three polymorphs of sulfathiazole, two polymorphs of carbamazepine and a single polymorph of piroxicam.

1. C. J. Gilmore, G. Barr, J. Paisley (2004). "High-throughput powder diffraction. I. A new approach to qualitative and quantitative powder diffraction pattern analysis using full pattern profiles." Journal of Applied Crystallography **37**: 231-242
2. G. Barr, W. Dong, C. J. Gilmore (2004). "PolySNAP: a computer program for analysing high-throughput powder diffraction data." Journal of Applied Crystallography **37**: 658-664.