

W0295

dSNAP: Correlating Database Mining with Chemistry. C.J.Gilmore, G. Barr, W. Dong, A. Parkin, C.C. Wilson. Dept. of Chemistry, Univ. of Glasgow, University Ave., Glasgow G12 8QQ, UK.

With the explosion in high quality structural determinations in the area of small molecule crystallography, the problem of efficient and meaningful mining of the data in, for example, the Cambridge Structural Database (CSD) is very much a live one for structural chemists. The available databases represent an enormously powerful resource for this subject, but faced with a data set of more than 280,000 structures such as is found in the CSD, the attempt to extract meaningful chemical information can be daunting.

The poster describes a novel, efficient and highly visual method of identifying structural correlations in fragments extracted from mining the CSD. The method is based on cluster analysis and multivariate data analysis of the distance matrix that describes the geometry of the matched fragments from a database search. It is a simplified version of the procedure used for matching and classifying powder patterns developed by Gilmore *et al.* [1,2]. The procedure allows clustering of the extracted data, which is found to be sensitive to small but significant data variations. The method will be illustrated by several applications in which the method is found to be conspicuously successful. The associated software that performs these calculations and is called *dSNAP* and we will demonstrate this to any interested parties.

[1] Barr, G. Dong, W. & Gilmore, C.J., (2004). *J. Appl. Cryst.* 37, 243-252.

[2] Barr, G. Dong, W. & Gilmore, C.J., (2004). *J. Appl. Cryst.* 37, 658-664.