m34.p08

Open Source Software for small molecule Crystallography

Mustapha Sadki[(1)], David Watkin[(1)], Luc Bourhis[(2)], Horst. Puschmann[(2)], Judith Howard[(2)]

[(1)] Inorganic Chemistry Laboratory University Of Oxford South Parks Road OX1 3QR UK [(2)] Chemistry Department, Durham University Durham DH1 3LE. UK

Keywords: crystallography open source software, mathematical software systems, programming crystallographic software

The ''Next Generation of small molecule Crystallographic Software'' is a joint project between the University of Oxford and Durham University, with an external management advisory group and scientific advisory committee of experienced senior crystallographers from many Universities. The consortium was set up to react to the challenges facing current crystallographic software and the knowledge contained within them. To date, almost all widely used small molecule programs are written by single individuals or small groups to software standards dating from the 1970's. The algorithm details are not documented and this software is neither extendable nor supportable.
There is a risk of loss of technical knowledge as authors retire or leave the field, hence the pressing need to bridge the gap between previous and future generations of crystallographers. The aim of the project is to develop Open Source software solutions for Crystallography, which we consider to be a natural way for a project to evolve, since we adhere to software culture that encourages code-sharing. This will ensure the emergence of new science and will complement existing macromolecular crystallographic developments within the domain of small molecule crystallography. To this end, we will implement a pilot design and develop a new program to modern software standards, unhindered by legacy code, supported by detailed documentation, and which will guarantee maintenance and sustainability. It will provide a software kernel on which other researches can build new applications.
The initial phase of the project will include all the fields covered by mainstream crystallographic software, such as CRYSTALS, GSAS, JANA, PLATON, SHELX, TOPAS, XP, and include data quality analysis, model building, electron density maps and related analysis, refinement, structure analyses and validation. The presentation will explain why it is important to develop such a project for crystallography now, and what the key challenges will be in completing it.
The development team recognises that while current crystallographers are generally content with existing software, the next generation will requite something altogether more integrated and sophisticated, and their aim is to anticipate these needs. After introducing our conceptual view of the project, we will show how we intend to develop methods and algorithms within the new software architecture and at the same time benefit from existing programs.
Other parties will be strongly encouraged to contribute their knowledge and ideas to the main development code base. As an Open Source software project, the code will be openly available to all for re-use in order to develop their own extensions.

m34.p09

dSNAP: Working towards automatic interpretation of structural chemistry

D. Sneddon, G. Barr, A. Collins, W. Dong, C.J. Gilmore, A. Parkin, C.C. Wilson

WestCHEM, Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland.

Keywords: data representation, computational analysis of crystallographic data, clustering

Cluster analysis is a well-established tool in statistics, but one that is used surprisingly little in crystallography. We have established its use [1] in analysing the results of database searches using the Cambridge Structural Database (CSD).
CSD searches can produce thousands of 'hits' especially if a simple fragment is used, and as a result processing and interpreting the results becomes a considerable task, and one in which it is easy to make mistakes. Cluster analysis using dendrograms, metric multidimensional scaling and suitable visualization tools can reduce the workload to a few hours with minimal user intervention, and thus minimal user bias. The real beauty of the method is in its interactivity and scalability it allows you to go from an overview of the entire dataset and easily spot any outliers or errors, to the ability to 'drill down' within a cluster, looking at more and more detailed differences between different fragments. The methods are implemented in the computer program *d*SNAP [1]. It is not confined to the use of bonded atoms but works equally well with non-bonded interactions, or, indeed, a mixture of both. For example, we have used the method to study and classify interactions of carboxylic acids with primary amides and with other carboxylic acid groups [2].
The results from these analyses still require detailed manual interpretation by an experienced structural analyst to extract meaningful chemical information. To help remove this final manual step, we are investigating new methods of automating this interpretation. We will present some of the latest developments in the *d*SNAP methodology, with examples being used to illustrate these methods with real chemical examples.
The *d*SNAP software is available **free of charge** to all interested researchers from the Bruker-AXS website at http://www.bruker-axs.de

[1] G. Barr, W. Dong, C. J. Gilmore, A. Parkin and C. C. Wilson,. *J. Appl. Cryst*,. 2005, 38, 833-841.
[2] A. Parkin, G. Barr, W. Dong, C.J. Gilmore and C. C. Wilson,. *Cryst. Eng. Comm.*, 2006, 8, 257-264.