

Back to (data) basics – how crystallographic databases can help solve new structures

N. T. Johnson¹, S. B. Wiggin¹, C. A. Tovee¹, P. I. Andrews¹, M. P. Lightfoot¹

¹CCDC, 12 Union Road, Cambridge, CB2 1EZ, UK

njohnson@ccdc.cam.ac.uk

For over 60 years, crystallographic databases have existed to draw together knowledge from published crystal structures. The foresight of our academic forebearers and continued effort from the community has gifted us ever growing collections of structural data, holding a wealth of knowledge that researchers utilise every day. These curated resources are not only a record of what has been studied, they are also valuable resources when solving new structures. Focussing on the Cambridge Structural Database[1], a database of over 1.3 million small molecule organic and metal organic crystal structures, this talk will describe the many ways databases can be used throughout a crystallographic experiment, highlighting existing tools and new developments to aid structural scientists.

Before an experiment even begins, searching crystallographic databases, whether using web, desktop or programmatic services, allows the scientific landscape to be surveyed to gather information to help with planning a study. During data collection, an initial unit cell can be searched against databases to help identify if the desired compound has been crystallised, before a full data collection takes place. Comparing structure solutions to existing structures or investigating individual bond lengths, angles and torsions against similar structures may help to identify any potential structural errors. The structural information within databases can also be leveraged to create knowledge bases, that are then used to help interpret structures[2].

When looking towards the sharing of results, many databases have visualisation programs that enable scientists to make images ready for publication[3]. Depositing structures in databases prior to publication allows researchers to enrich entries with additional information (e.g. properties, recrystallisation solvent), as well as providing persistent identifiers for structures to enable readers to find data associated with a publication. The CCDC also provides an easy route to disseminate structures without needing an accompanying publication – *CSD Communications*. Furthermore, databases can be a useful tool in the education of future researchers. A world without crystallographic databases would be a very different place. This talk celebrates our communities' efforts to collect and share crystal structures for the common good.

[1] Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst.*, **B72**, 171–179

[2] Wood, P. A., Olsson, T. S. G., Cole, J. C., Cottrell, S. J., Feeder, N., Galek, P. T. A., Groom, C. R. & Pidcock, E. (2013). *CrystEngComm*, **15**, 65–72

[3] C. F. Macrae, I. Sovago, S. J. Cottrell, P. T. A. Galek, P. McCabe, E. Pidcock, M. Platings, G. P. Shields, J. S. Stevens, M. Towler and P. A. Wood, (2020). *J. Appl. Cryst.*, **53**, 226-235