MS46 Reproducibility in crystallography

MS46-02
Reliability and Reproducibility in Small-Molecule Crystallography
A. Linden [1]
[1] *University of Zurich, Department of Chemistry - Zurich (Switzerland)*

Abstract
Most reported small-molecule crystal structures have only been determined once from one crystal selected from a crystallization batch, often the first attempt. Multiple determinations of the structures of some compounds are found in the CSD, but these are relatively rare compared with the size of the database. How, then, do we assess the reliability or each of these "one-off" results and how reproducible is that result? Validation via *checkCIF* and by other means goes a long way towards verifying the veracity of the archived results. The inclusion of refinement instructions and reflection data or structure factors in the deposited record allows reviewers and readers to know exactly how the authors developed their model and, if interested, to investigate and tinker with the structure further from that point. However, even though CCD, image plate, pixel array and other area detectors have been commonplace since the early 1990s, up to now, in the single-crystal world, there is usually no public record of the original frame images from the diffractometer. Even if a laboratory is lovingly archiving their frames locally, do they still have a functioning version of the software or hardware to be able to reprocess them? One of the challenges with establishing repositories for frame data is to have enough metadata to know exactly how the original data reduction was performed. There are many options in modern data reduction software and it is hard to know how the original scientist tweaked the options to optimise the outcome. Very likely identical strategies are not used from one structure to the next. Without such knowledge, exact reproducibility might not be achievable. Nonetheless, the overall structure result may well be sufficiently reproducible and reliable for the purpose of the study, even if not quite identical. "Fit for purpose" is a key consideration and users of deposited data must remain cognisant of the intentions of the original investigation and not try to extract more significance out of a result than might be possible given the data collected at the time. Despite best efforts, some crystal structures can never entirely unambiguously answer even the original question behind the study. At the same time, it is important that the original report or deposited data is properly annotated with details of any special data handling or model refinement strategies that were used and any caveats about the sample, data and results that future users of the data should be aware of. With automated CIF generation software, attention to this last aspect is sometimes insufficient.