Harvesting should be a Complete and Repeatable Record of Experiment

Thomas Terwilliger, *Bioscience Division, Los Alamos National Laboratory, Mail Stop M888, Los Alamos, NM 87545 USA*. E-mail: terwilliger@lanl.gov

The procedural and crystallographic data that are harvested and stored as a record of a crystallographic experiment should be sufficient to (1) repeat the analysis and (2) to estimate uncertainties, preferably in a completely automated fashion. It is highly likely that automated methods will be constantly developed over the next years and that there will be repeatedly be utility to a re-analysis of all structures currently in the PDB. Such analyses will be possible ony if the primary data and any associated knowledge about the experiment necessary to fully and automatically repeat the structure determination and analysis process are available. Further, even in the absence of full redeterminations of structure, re-estimates of errors and uncertainties (as in structure validation) will be done repeatedly on the entire PDB. The final structure and structure factors are not sufficient to repeat either the structure determination or the the error analysis (for example a free-R analysis is impossible if a test set is not specified; model bias can never be fully removed if an atom is placed and later removed). The data harvested from an experiment should therefore be a complete record of the process used to determine the structure, in a form that can be automatically read, repeated, and modified.

Keywords: PDB, LIMS, harvesting