## A Dictionary Approach to Translate Memory Variables from Crystallography Software to mmCIF Items

Heping Zheng, Wladek Minor, Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908, USA. E-mail: dust@iwonka.med.virginia.edu

A major obstacle in building CIF output from crystallography software is to address the relation between the information software can supply and the information mmCIF required. Generally the building process is time-consuming and even more effort is necessary to maintain the source code due to constant changes of the software. We present here, a dictionary-based approach, and the tool used to build such a dictionary. In this approach, the memory-to-mmCIF relation is classified as equivalence, conversion, constant, source conversion, comment, pending or unknown. Each mmCIF item is subject to classification by the developer's examination with the assistance from a domain expert. The CIF Translator Dictionary (CTD) builder is utilizing a dump of all global variables with its value in memory as source of information. This memory dump is in STAR format and allow the CTD developer to do realtime tracking of related variables in memory. Generally it is possible to fetch related variable names in 2 to 5 memory scan by a domain expert. Andfter addressing the relationship between these variables with mmCIF item, a CTD entry will be generated automatically for simple relation, or more infomation will be acquired for complicate relation.

To test the effectiveness of this approach, HKL2000 CTD is built in its initial stage. Automatic completion from HKL2000 memory is performed without human intervention. For more specific tuning toward publication quality CIF after autofill, HKL2000-CIF is also designed as a CIF editor featuring entities editing and providing an evolving amount of wizard procedures that assist further manual examination, and validation before final submission.

Keywords: CIF, CTD, mmCIF