How the RCSB Validates PDB Structures

Helen M. Berman, *Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey*. E-mail: berman@rcsb.rutgers.edu

The RCSB validates PDB data while considering the evolving nature of data and nomenclature standards. In order to provide the community with high quality data, the RCSB Protein Data Bank (www.pdb.org) has developed a number of tools that support the deposition and processing of X-ray and NMR structures and that are based upon the mmCIF dictionary [1]. A key feature is the Validation Suite [2], which produces a validation report highlighting close contacts, bond and angle deviations, chirality problems, missing and extra atoms and residues, and distant waters. Authors are encouraged to validate their structures before data deposition.

As the number of structures being determined is constantly increasing, the automation of data validation is extremely important. To reach this end, there needs to be consistency in the syntax and representation of incoming structure data. Tools are being developed to aid this process.

The RCSB also collaborates with wwPDB members to validate the entire PDB archive, and to distribute these data in a way that is most useful to members of the community.

The RCSB PDB is funded by NSF, NIGMS, Office of Science DOE, NLM, NCI, NCRR, NIBIB, and NINDS.

[1] Bourne P.E., Berman H.M., Watenpaugh K., Westbrook J.D., Fitzgerald P.M.D., *Meth. Enz.*, **1997**, 277, 571. [2] Westbrook J., Feng Z., Burkhardt K., Berman H.M., *Meth. Enz.* **2003**, 374, 370.

Keywords: structure validation, structure analysis, mmCIF