Keynote Lectures

[KN5] The Brain as a Validation Tool

Zbigniew Dauter

National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA.
E-mail: dauter@anl.gov

The Protein Data Bank (PDB) serves as an archive of macromolecular structures acquired by various biophysical methods. Currently it contains about 90,000 atomic models, majority of which were obtained by X-ray crystallography, but ca. 10,000 by NMR and a few hundred by other methods. Before submission to the PDB each of these structures is validated by one of the dedicated programs, automatically checking for correctness of various statistical, geometrical, symmetry and other aspects of the investigated models. The programs used for validation are highly sophisticated and very useful, also as tools during the process of model building and refinement. However, macromolecular crystallography is a very interdisciplinary branch of science, integrating such fields as physics (diffraction experiments), chemistry (investigated compounds), biology (source of samples), and medicine (ultimate goal of the research), with advanced computing (need for fast computers and elaborate graphical tools) and mathematics (symmetry theory and other sophisticated algorithms). The task of validation programs is therefore enormous, if one takes into account how widely different all these structures are in terms of diffraction data resolution, crystal symmetry, chemical properties and biological function. It is not surprising that among many PDB structures there are some examples containing certain erroneously interpreted aspects in terms of physics, chemistry, biology, or symmetry. Such errors may result from the wrong interpretation of results provided by various computer programs or from the lack of experience in one of the mentioned above branches of science. Usually such misinterpretations do not invalidate the biological conclusions, but sometimes may have serious negative consequences. The presence of small amount of errors among so many crystal structures in the PDB is a "statistically inevitable" consequence of the enormous success of macromolecular crystallography, where the availability of very powerful and highly automatic programs makes it possible for investigators less experienced in the underlining methodology to relatively easily solve structures of the molecules of interest. Many examples of misinterpretations could be avoided by critical analysis of all available information by the knowledgeable specialist. In that respect a common sense provided by the human brain can be treated as a valuable validation tool.

Keywords: validation; computer programs, PDB